

# FROM THE FERMI SCALE TO COSMOLOGY

EDITED BY: Alberto Salvio and Francesco Sannino

PUBLISHED IN: Frontiers in Astronomy and Space Sciences and  
Frontiers in Physics



# frontiers

## Frontiers eBook Copyright Statement

The copyright in the text of individual articles in this eBook is the property of their respective authors or their respective institutions or funders. The copyright in graphics and images within each article may be subject to copyright of other parties. In both cases this is subject to a license granted to Frontiers.

The compilation of articles constituting this eBook is the property of Frontiers.

Each article within this eBook, and the eBook itself, are published under the most recent version of the Creative Commons CC-BY licence.

The version current at the date of publication of this eBook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

When exercising any right under the CC-BY licence, Frontiers must be attributed as the original publisher of the article or eBook, as applicable.

Authors have the responsibility of ensuring that any graphics or other materials which are the property of others may be included in the CC-BY licence, but this should be checked before relying on the CC-BY licence to reproduce those materials. Any copyright notices relating to those materials must be complied with.

Copyright and source acknowledgement notices may not be removed and must be displayed in any copy, derivative work or partial copy which includes the elements in question.

All copyright, and all rights therein, are protected by national and international copyright laws. The above represents a summary only. For further information please read Frontiers' Conditions for Website Use and Copyright Statement, and the applicable CC-BY licence.

ISSN 1664-8714

ISBN 978-2-88963-205-3

DOI 10.3389/978-2-88963-205-3

## About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

## Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

## Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

## What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: [researchtopics@frontiersin.org](mailto:researchtopics@frontiersin.org)

# FROM THE FERMI SCALE TO COSMOLOGY

Topic Editors:

**Alberto Salvio**, University of Rome and INFN Tor Vergata, Italy

**Francesco Sannino**, University of Southern Denmark, Denmark

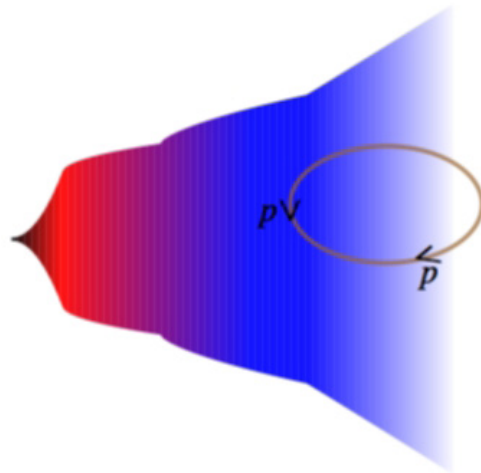


Image: Alberto Salvio

**Citation:** Salvio, A., Sannino, F., eds. (2019). From the Fermi Scale to Cosmology. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88963-205-3

# Table of Contents

<b>04</b>	<b><i>Editorial: From the Fermi Scale to Cosmology</i></b>
	Alberto Salvio and Francesco Sannino
<b>07</b>	<b><i>Higgs Inflation</i></b>
	Javier Rubio
<b>26</b>	<b><i>Cosmological Aspects of Higgs Vacuum Metastability</i></b>
	Tommi Markkanen, Arttu Rajantie and Stephen Stopyra
<b>60</b>	<b><i>The Top-Quark Mass: Challenges in Definition and Determination</i></b>
	Gennaro Corcella
<b>76</b>	<b><i>Several Problems in Particle Physics and Cosmology Solved in One SMASH</i></b>
	Guillermo Ballesteros, Javier Redondo, Andreas Ringwald and Carlos Tamarit
<b>95</b>	<b><i>Impact of Cosmological and Astrophysical Constraints on Dark Matter Simplified Models</i></b>
	Chiara Arina
<b>126</b>	<b><i>Neutrino Mass Ordering From Oscillations and Beyond: 2018 Status and Future Prospects</i></b>
	Pablo F. de Salas, Stefano Gariazzo, Olga Mena, Christoph A. Ternes and Mariam Tórtola
<b>176</b>	<b><i>An Asymptotically Safe Guide to Quantum Gravity and Matter</i></b>
	Astrid Eichhorn
<b>205</b>	<b><i>Quadratic Gravity</i></b>
	Alberto Salvio
<b>229</b>	<b><i>Euclidean Wormholes, Baby Universes, and Their Impact on Particle Physics and Cosmology</i></b>
	Arthur Hebecker, Thomas Mikhail and Pablo Soler



# Editorial: From the Fermi Scale to Cosmology

Alberto Salvio<sup>1\*</sup> and Francesco Sannino<sup>2</sup>

<sup>1</sup> Physics Department, University of Rome, INFN Tor Vergata, Rome, Italy, <sup>2</sup> CP3-Origins, University of Southern Denmark, Odense, Denmark

**Keywords:** standard model and beyond, cosmology-theory, gravitation, field theory, astrophysics

## Editorial on the Research Topic

### From the Fermi Scale to Cosmology

In July 2012 (Aad et al., 2012; Chatrchyan et al., 2012) the Large Hadron Collider (LHC) experimental collaborations ATLAS and CMS announced the discovery of the Higgs boson, the long sought elementary particle of the Standard Model (SM) of particle interactions. The discovery crowned the SM as the most successful description of natural phenomena at the shortest distance ever tested by humans.

The discovery left still a number of unanswered questions since the SM remains an incomplete description of nature. In fact, it fails to account for neutrino masses, the particle nature of dark matter (DM) and last but not the least the observed asymmetry between matter and antimatter in the universe (BAU).

Intriguingly, some of the issues above, specifically DM and BAU, lie at the interface between particle physics (encoded in the SM) and cosmology. This has naturally intensified the efforts to arrive at a deeper understanding of the connection between the physics of very short distances and the one dealing with very large ones. For example, the precise knowledge of the SM parameters allows us to make sensible predictions in cosmology.

Of course, in order for this to be feasible the SM should be embedded in a theory of gravity. Einstein's theory, a.k.a General Relativity (GR), is extremely successful and able to describe gravity and its interactions with matter from the largest scales almost down to the extremely small Planck length  $l_P \sim 10^{-33}$  cm. However, when approaching  $l_P$ , GR cannot be treated classically and needs to be quantized. If we were to naively quantize GR one would discover that it is not renormalizable, like the pion Lagrangian. This fact tells us that it is only a low energy description of physical phenomena and that below or at around the Planck length the true elementary nature will appear. Currently a satisfactory quantum theory of gravity is still missing and both theoretical and experimental inputs are needed to determine its nature.

*The purpose of the Research Topic "From the Fermi Scale to Cosmology" is to bring to light the latest exciting ideas and results that are meant to elucidate the relation between the LHC collider program and astrophysics, paying attention to their impact on cosmology.*

Perhaps a natural subject along these lines is the potential cosmological role of the Higgs boson. In Bezrukov and Shaposhnikov (2008, 2014), it was proposed that the Higgs could play the role of the inflaton. This remains an interesting possibility and therefore Rubio provided an in depth overview on this subject. Although this is an economical and predictive model of the early universe its viability is intimately related to the nature of the electroweak vacuum in the SM. Its fate is not yet sealed since it could still be absolutely stable, unstable, or metastable with a preference for the last case according to the latest determinations of the top mass  $m_t$ . It is for this reason that in recent years much work has been devoted to understand whether the metastable case can yield a consistent cosmology and, if so, what are the physical implications. A contribution

## OPEN ACCESS

### Edited and reviewed by:

Jan De Boer,  
University of Amsterdam, Netherlands

### \*Correspondence:

Alberto Salvio  
alberto.salvio@cern.ch

### Specialty section:

This article was submitted to  
High-Energy and Astroparticle  
Physics,  
a section of the journal  
Frontiers in Astronomy and Space  
Sciences

**Received:** 02 August 2019

**Accepted:** 03 September 2019

**Published:** 18 September 2019

### Citation:

Salvio A and Sannino F (2019)  
Editorial: From the Fermi Scale to  
Cosmology.  
Front. Astron. Space Sci. 6:61.  
doi: 10.3389/fspas.2019.00061

by Markkanen et al. presents a comprehensive review of the implications of the Higgs vacuum metastability for cosmology along with a pedagogical discussion of the related theoretical topics, including renormalization group improvement, quantum field theory in curved spacetime and vacuum decay in field theory.

However, it is still possible, although unlikely, that we do not leave in a metastable vacuum, but rather in an absolutely stable one because the uncertainty on  $m_t$  is still fairly large. Corcella, in his review on the definition and determination of  $m_t$ , discusses the main strategies to extract  $m_t$  at the LHC and the interpretation of the measurements in terms of well-posed top-mass definitions, taking particular care about renormalon ambiguities, progress in Monte Carlo event generators for top physics and theoretical uncertainties.

Of course, even if we were certain about the fate of the SM vacuum, it would still be possible (and actually likely) that new degrees of freedom (elementary particles) with masses below the Planck mass are present in nature. In this case, many particle physics models of the early universe can be investigated. An interesting possibility is that the inflaton is identified with the radial scalar of an axion model. The axion (Weinberg, 1978; Wilczek, 1978) is independently motivated by the strong CP problem (Peccei and Quinn, 1977a,b) and is a viable DM candidate. Ballesteros et al. contribute to this Research Topic with an overview on this possibility, which, along the lines of a previous work (Salvio, 2015), is combined with a model of neutrino masses that is capable of accounting for BAU.

About DM, Arina explains the impact of cosmological and astrophysical constraints on DM, focusing on simplified models. She classifies the models that have been analyzed so far and for each of them she reviews in detail the complementarity of relic density, direct, and indirect searches with respect to the LHC searches. She also discusses the capabilities of each type of search to identify regions where individual approaches to dark matter detection are the most relevant to constrain the model parameter space. Finally, a critical overview on the validity of the dark matter simplified models is provided and the caveats for the interpretation of the experimental results extracted for these models are discussed.

As mentioned above, besides DM, neutrino masses furnish further evidence for BSM. One of the currently most exciting research areas regarding neutrino masses is about their ordering. de Salas et al. provide us with a review on the status of our understanding of neutrino mass ordering and future prospects. Quite interestingly the Bayesian analysis to the 2018 publicly

available oscillation and cosmological data sets provides strong evidence for the normal neutrino mass ordering vs. the inverted scenario, with a significance of 3.5 standard deviations.

Beyond experimental facts theoretical considerations further motivate extending or modifying the SM and GR. For example the lack of predictivity of GR at length scales smaller than  $l_p$  calls for new ideas to address this issue. Focusing on scenarios that maintain the GR relativistic field theoretic structure while remaining in four dimensions, this editorial features two reviews [written by Eichhorn and Salvio (one of the editors of this Research Topic)] on the recent efforts to address the lack of GR predictability. Both approaches essentially require gravitational and non-gravitational interactions to flow to a UV fixed point. Eichhorn's review focuses on Weinberg's asymptotic safety (Weinberg, 1977, 1980) where an infinite number of terms are added to the Einstein-Hilbert action to ensure renormalizability. Only very recently asymptotic safety has been rigorously established in four dimensions for gauge-Yukawa theories (Litim and Sannino, 2014). A complementary point of view is described in Salvio's review of quadratic gravity (adding only the terms quadratic in the curvature to the Einstein-Hilbert action renders gravity renormalizable).

In all these scenarios, as well as in GR, non-perturbative gravitational effects, known as gravitational instantons, can have deep implications for both particle physics and cosmology. A review by Hebecker et al. revisits some of the more recent discussions of the phenomenological relevance of gravitational instantons. In particular, these instantons are expected to break the shift symmetries of axions or Goldstone bosons non-perturbatively.

Clearly the subjects of SM physics and beyond, quantum gravity and the interplay with cosmology are vast and this Research Topic is by no mean meant as an exhaustive review. The choice of the topics reflect partially our interests and partially the limited time we had at our disposal for a more in depth review. Additionally, we also wished to guide the reader through a less dispersive choice of arguments. We sincerely hope that this Research Topic will inspire the readers to tackle and solve many of the exciting unsolved issues discussed by our colleagues in their reviews. Finally, it is our pleasure to thank the colleagues that accepted being part of this effort for their excellent contributions.

## AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

## REFERENCES

- Aad, G., Abajyan, T., Abbott, B., Abdallah, J., Abdel Khalek, S., Abdelalim, A. A., et al. (2012). Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Phys. Lett. B* 716, 1–29. doi: 10.1016/j.physletb.2012.08.020
- Bezrukov F., and Shaposhnikov, M. (2014). Higgs inflation at the critical point. *Phys. Lett. B* 734, 249–254. doi: 10.1016/j.physletb.2014.05.074
- Bezrukov, F. L., and Shaposhnikov, M. (2008). The standard model higgs Boson as the inflaton. *Phys. Lett. B* 659, 703–706. doi: 10.1016/j.physletb.2007.11.072
- Chatrchyan, S., Khachatryan, V., Sirunyan, A. M., Tumasyan, A., Adam, W., Aguilo, E., et al. (2012). Observation of a new Boson at a mass of 125 GeV with the CMS experiment at the LHC. *Phys. Lett. B* 716, 30–61. doi: 10.1016/j.physletb.2012.08.021
- Litim, D. F., and Sannino, F. (2014). Asymptotic safety guaranteed. *J. High Energy Phys.* 1412:178 doi: 10.1007/JHEP12(2014)178

- Peccei R. D., and Quinn, H. R. (1977a). CP conservation in the presence of instantons. *Phys. Rev. Lett.* 38:1440. doi: 10.1103/PhysRevLett.38.1440
- Peccei R. D., and Quinn, H. R. (1977b). Constraints imposed by CP conservation in the presence of instantons. *Phys. Rev. D* 16, 1791–1797. doi: 10.1103/PhysRevD.16.1791
- Salvio, A. (2015). A simple motivated completion of the standard model below the planck scale: axions and right-handed neutrinos. *Phys. Lett. B* 743 428–434. doi: 10.1016/j.physletb.2015.03.015
- Weinberg, S. (1978). A new light Boson? *Phys. Rev. Lett.* 40:223. doi: 10.1103/PhysRevLett.40.223
- Weinberg, S. (1977). *Understanding the Fundamental Constituents of Matter*. ed A. Zichichi. New York, NY: Plenum Press.
- Weinberg, S. (1980). *General Relativity: An Einstein Centenary Survey*. eds S. W. Hawking and W. Israel (Cambridge: Cambridge University Press), 790–831.
- Wilczek, F. (1978). Problem of strong  $P$  and  $T$  invariance in the presence of instantons. *Phys. Rev. Lett.* 40:279. doi: 10.1103/PhysRevLett.40.279
- Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Salvio and Sannino. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Higgs Inflation

Javier Rubio\*

Institut für Theoretische Physik, Ruprecht-Karls-Universität Heidelberg, Heidelberg, Germany

The properties of the recently discovered Higgs boson together with the absence of new physics at collider experiments allows us to speculate about consistently extending the Standard Model of particle physics all the way up to the Planck scale. In this context, the Standard Model Higgs non-minimally coupled to gravity could be responsible for the symmetry properties of the Universe at large scales and for the generation of the primordial spectrum of curvature perturbations seeding structure formation. We overview the minimalistic Higgs inflation scenario, its predictions, open issues and extensions and discuss its interplay with the possible metastability of the Standard Model vacuum.

**Keywords:** higgs, inflation, higgs inflation, vacuum stability, scale-invariance

## 1. INTRODUCTION AND SUMMARY

Inflation is nowadays a well-established paradigm (Starobinsky, 1980; Guth, 1981; Mukhanov and Chibisov, 1981; Albrecht and Steinhardt, 1982; Linde, 1982, 1983) able to explain the flatness, homogeneity and isotropy of the Universe and the generation of the primordial density fluctuations seeding structure formation (Hawking, 1982; Starobinsky, 1982; Sasaki, 1986; Mukhanov, 1988). In spite of the phenomenological success, the inflaton's nature remains unknown and its role could be played by any particle physics candidate able to imitate a slowly-moving scalar field in the very early Universe.

In spite of dedicated searches, the only outcome of the Large Hadron Collider (LHC) experiments till date is a scalar particle with mass (Aad et al., 2012; Chatrchyan et al., 2012; Tanabashi et al., 2018)

$$m_H = 125.18 \pm 0.16 \text{ GeV} \quad (1.1)$$

and properties similar to those of the Standard Model (SM) Higgs. The mass value (1.1) is certainly particular since it allows to extend the SM up to the Planck scale without leaving the perturbative regime (Shaposhnikov and Wetterich, 2010). The main limitation to this appealing scenario is the potential instability of the electroweak vacuum at high energies. Roughly speaking, the value of the Higgs self-coupling following from the SM renormalization group equations decreases with energy till a given scale and starts increasing thereafter. Whether it stays positive all the way up to the Planck scale, or turns negative at some intermediate scale  $\mu_0$  depends, mainly, on the interplay between the Higgs mass  $m_H$  and the top quark Yukawa coupling  $y_t$  extracted from the reconstructed Monte-Carlo top mass in collider experiments (Butenschoen et al., 2016), cf. **Figure 1**. Neglecting the effect of gravitational corrections, the critical value  $y_t^{\text{crit}}$  separating the region of absolute stability from the metastability/instability<sup>1</sup> regions is given by (Bezrukov and Shaposhnikov, 2015b)

$$y_t^{\text{crit}} = 0.9244 \pm 0.0012 \frac{m_H/\text{GeV} - 125.7}{0.4} + 0.0012 \frac{\alpha_s(m_Z) - 0.01184}{0.0007}, \quad (1.2)$$

<sup>1</sup>The metastability region is defined as the parameter space leading to vacuum instability at energies below the Planck scale but with an electroweak vacuum lifetime longer than the age of the Universe.

## OPEN ACCESS

### Edited by:

Francesco Sannino,  
University of Southern Denmark,  
Denmark

### Reviewed by:

Sayantan Choudhury,  
Max-Planck-Institut für  
Gravitationsphysik, Germany  
Zhenbin Wu,  
University of Illinois at Chicago,  
United States

### \*Correspondence:

Javier Rubio  
j.rubio@thphys.uni-heidelberg.de

### Specialty section:

This article was submitted to  
High-Energy and Astroparticle  
Physics,  
a section of the journal  
Frontiers in Astronomy and Space  
Sciences

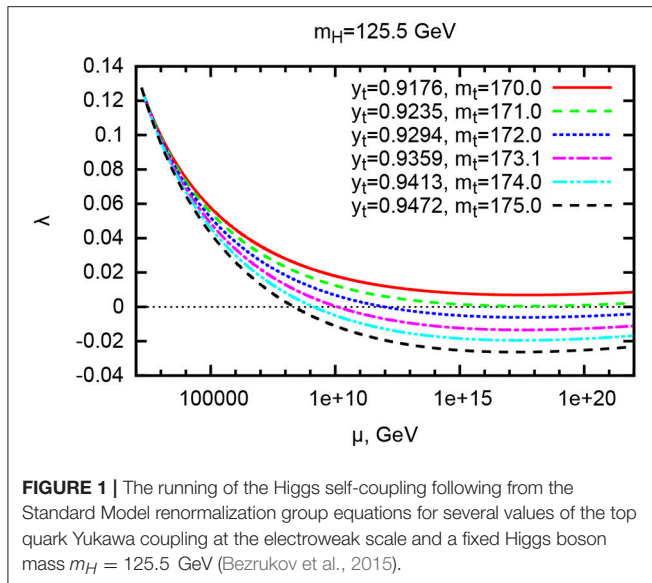
**Received:** 10 July 2018

**Accepted:** 20 December 2018

**Published:** 22 January 2019

### Citation:

Rubio J (2019) Higgs Inflation.  
Front. Astron. Space Sci. 5:50.  
doi: 10.3389/fspas.2018.00050



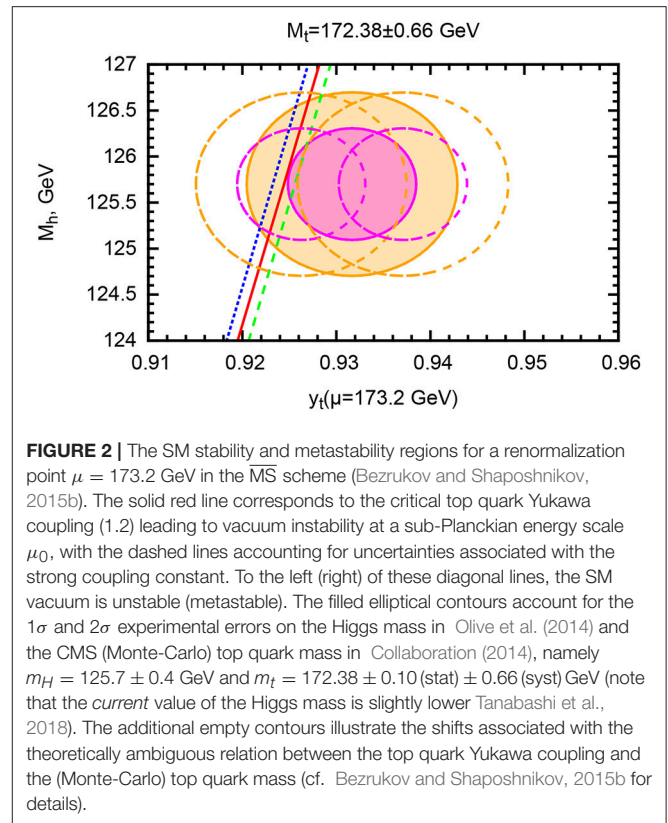
with  $\alpha_s(m_Z)$  the strong coupling constant at the Z boson mass. Within the present experimental and theoretical uncertainties (Bezrukov and Shaposhnikov, 2015b; Butenschoten et al., 2016; Espinosa et al., 2017), the SM is compatible with vacuum instability, metastability and absolute stability (Bezrukov et al., 2012; Degraasi et al., 2012; Buttazzo et al., 2013; Espinosa et al., 2015; Espinosa, 2016), with the separation among the different cases strongly depending on the ultraviolet completion of gravity (Branchina and Messina, 2013; Branchina et al., 2014, 2015), cf. Figure 2.

In the absence of physics beyond the SM, it is certainly tempting to identify the recently discovered Higgs boson with the inflaton condensate. Unfortunately, the Higgs self-interaction significantly exceeds the value  $\lambda \sim 10^{-13}$  leading to a sufficiently long inflationary period without generating an excessively large spectrum of primordial density perturbations (Linde, 1983). The situation is unchanged if one considers the renormalization group enhanced potential following from the extrapolation of the SM renormalization group equations to the inflationary scale (Isidori et al., 2008; Fairbairn et al., 2014; Hamada et al., 2014a). The simplest way out is to modify the Higgs field kinetic term in the large-field regime. In Higgs inflation<sup>2</sup> this is done by including a direct coupling to the Ricci scalar  $R$  (Bezrukov and Shaposhnikov, 2008), namely<sup>3</sup>

$$\delta S = \int d^4x \sqrt{-g} \left[ \xi H^\dagger H R \right], \quad (1.3)$$

<sup>2</sup>An alternative possibility involving a derivative coupling among the Einstein tensor and the Higgs kinetic term was considered in Germani and Kehagias (2010b), Germani and Kehagias (2010a), and Fumagalli et al. (2018).

<sup>3</sup>Prior to Bezrukov and Shaposhnikov (2008), the effect of non-minimal couplings had been extensively studied in the literature (see for instance Minkowski, 1977; Smolin, 1979; Zee, 1979; Spokoiny, 1984; Futamase and Maeda, 1989; Salopek et al., 1989; Fakir and Unruh, 1990a,b; Makino and Sasaki, 1991; Fakir et al., 1992; Cervantes-Cota and Dehnen, 1995a,b; Kaiser, 1995; Komatsu and Futamase, 1998), but never with the SM Higgs as the inflaton.



with  $H$  the Higgs doublet and  $\xi$  a dimensionless constant to be fixed by observations. The inclusion of the non-minimal coupling (1.3) can be understood as an inevitable consequence of the quantization of the SM in a gravitational background, where this term is indeed required for the renormalization of the energy-momentum tensor (Callan et al., 1970; Birrell and Davies, 1984).

When written in the Einstein frame, the Higgs inflation scenario displays two distinct regimes. At low energies, it approximately coincides with the SM minimally coupled to gravity. At high energies, it becomes a chiral SM with no radial Higgs component (Dutta et al., 2008; Bezrukov and Shaposhnikov, 2009). In this latter regime, the effective Higgs potential becomes exponentially flat, allowing for inflation with the usual chaotic initial conditions. The associated predictions depend only on the number of e-folds of inflation, which is itself related to the duration of the heating stage. As the type and strength of the interactions among the Higgs field and other SM particles is experimentally known, the duration of this entropy production era can be computed in detail (Bezrukov et al., 2009a, 2011a; Garcia-Bellido et al., 2009; Repond and Rubio, 2016; Ema et al., 2017a), leading to precise inflationary predictions in perfect agreement with observations (Akrami et al., 2018).

The situation becomes more complicated when quantum corrections are included. The shape of the inflationary potential depends then on the values of the Higgs mass and top Yukawa coupling *at the inflationary scale*. In addition to the plateau already existing at tree-level (Bezrukov and Shaposhnikov, 2014; Enckell et al., 2016; Fumagalli and Postma, 2016; Fumagalli,

2017), the renormalization group enhanced potential can develop a quasi-inflection point along the inflationary trajectory (Allison, 2014; Bezrukov and Shaposhnikov, 2014; Hamada et al., 2014b; Bezrukov et al., 2015, 2018; Rubio, 2015; Enckell et al., 2016; Fumagalli and Postma, 2016; Rasanen and Wahlman, 2017) or a hilltop (Fumagalli and Postma, 2016; Rasanen and Wahlman, 2017; Enckell et al., 2018), with different cases giving rise to different predictions.

Although a precise measurement of the inflationary observables could be understood as an interesting consistency check between cosmological observations and particle physics experiments (Barvinsky et al., 2008, 2012; Espinosa et al., 2008; Bezrukov and Shaposhnikov, 2009; Bezrukov et al., 2009b, 2012; De Simone et al., 2009; Popa and Caramete, 2010; Salvio, 2013), the low- to high-energy connection is subject to unavoidable ambiguities related to the non-renormalizability of the model (Barbon and Espinosa, 2009; Burgess et al., 2009, 2010; Bezrukov et al., 2011b, 2015, 2018; George et al., 2014, 2016; Rubio, 2015; Enckell et al., 2016; Fumagalli and Postma, 2016). In particular, the finite parts of the counterterms needed to renormalize the tree-level action lead to localized jumps in the SM renormalization group equations when connected to the chiral phase of Higgs inflation. The strength of these jumps encodes the remnants of the ultraviolet completion and cannot be determined within effective field theory approach (Bezrukov et al., 2011b, 2015; Burgess et al., 2014). If the finite parts are significantly smaller than the associated coupling constants, Higgs inflation leads to a direct connection among the SM parameters measured at collider experiments and the large scale properties of the Universe, provided that the former do not give rise to vacuum instability. On the contrary, if the jumps in the coupling constants are large, the relation between high- and low-energy physics is lost, but Higgs inflation can surprisingly occur even if the electroweak vacuum is not completely stable (Bezrukov et al., 2015).

In this article we review the minimalistic Higgs inflation scenario, its predictions, open issues and extensions, and discuss its interplay with the potential metastability of the SM vacuum. The paper is organized as follows:

- The general framework is introduced in section 2. To illustrate the effect of non-minimal couplings, we consider an induced gravity scenario in which the effective Newton constant is completely determined by the Higgs vacuum expectation value. Having this toy model in mind, we reformulate Higgs inflation in the so-called Einstein frame in which the coupling to gravity is minimal and all non-linearities appear in the scalar sector of the theory. After emphasizing the pole structure of the Einstein-frame kinetic term and its role in the asymptotic flatness of the Higgs inflation potential, we compute the tree-level inflationary observables and discuss the decoupling properties of the SM degrees of freedom.
- The limitations of Higgs inflation as a fundamental theory are reviewed in section 3. In particular, we present a detailed derivation of the cutoff scales signaling the violation of perturbative unitarity in different scattering processes and advocate the interpretation of Higgs inflation as an effective field theory to be supplemented by an infinite set

of higher dimensional operators. Afterwards, we adopt a self-consistent approach to Higgs inflation and formulate the set of assumptions leading to a controllable relation between low- and high-energy observables. Based on the resulting framework, we analyze the contribution of quantum corrections to the renormalization group enhanced potential and their impact on the inflationary observables. We finish this section by discussing the potential issues of Higgs inflation with the metastability of the SM vacuum.

- Some extensions and alternatives to the simplest Higgs inflation scenario are considered in section 4. In particular, we address the difference between the metric and Palatini formulations of the theory and its extension to a fully scale invariant framework (Shaposhnikov and Zenhausern, 2009; Blas et al., 2011; Garcia-Bellido et al., 2011, 2012; Bezrukov et al., 2013; Rubio and Shaposhnikov, 2014; Karananas and Rubio, 2016; Trashorras et al., 2016; Casas et al., 2017, 2018). The inflationary predictions in these models are put in one to one correspondence with the pole structure of the Einstein-frame kinetic term, allowing for an easy comparison with the results of the standard Higgs inflation scenario.

Overall, we intend to complement the existing monographs in the literature (Bezrukov, 2013; Bezrukov and Shaposhnikov, 2015a; Moss, 2015) by i) providing a further insight on the classical formulation of Higgs inflation and by ii) focusing on the uncertainties associated with the non-renormalizability of the theory and their impact on model building.

## 2. GENERAL FRAMEWORK

The inflationary paradigm is usually formulated in terms of conditions on the local flatness on an arbitrary potential, which can in principle contain a large number of extrema and slopes (Artymowski and Rubio, 2016). This flatness is usually related to the existence of some approximate shift-symmetry, which, for the purposes of Higgs inflation, is convenient to reformulate as a non-linear realization of approximate scale-invariance.

### 2.1. Induced Gravity

Let us start by considering an *induced gravity* scenario

$$S = \int d^4x \sqrt{-g} \left[ \frac{\xi h^2}{2} R - \frac{1}{2} (\partial h)^2 - \frac{\lambda}{4} h^4 - \frac{1}{4} \mathcal{F}_{\mu\nu} \mathcal{F}^{\mu\nu} - \frac{g^2}{4} h^2 B_\mu B^\mu - i \bar{\psi} \not{D} \psi - \frac{y}{\sqrt{2}} h \bar{\psi} \psi \right], \quad (2.1)$$

involving a scalar field  $h$ , a vector field  $B_\mu$  and a fermion field  $\psi$ , with interactions similar to those appearing in the SM of particle physics when written in the unitary gauge  $H = (0, h/\sqrt{2})^T$ . The quantity  $\mathcal{F}_{\mu\nu} \mathcal{F}^{\mu\nu}$  stands for the standard  $B_\mu$  kinetic term, which for simplicity we take to be Abelian. In this toy model, the effective Newton constant is induced by the scalar field expectation value,

$$G_{N,\text{eff}} \equiv \frac{1}{8\pi\xi h^2}. \quad (2.2)$$

In order for  $G_{N,\text{eff}}$  to be well-behaved, the non-minimal coupling  $\xi$  is restricted to take positive values. This condition is equivalent to require the semi-positive definiteness of the scalar field kinetic term, as can be easily seen by performing a field redefinition  $h^2 \rightarrow h^2/\xi$ .

An important property of the *induced gravity* action (2.1) is its invariance under scale transformations

$$x^\mu \rightarrow \tilde{x}^\mu = \alpha x^\mu, \quad \varphi(x) \rightarrow \tilde{\varphi}(\tilde{x}) = \alpha^{\Delta_\varphi} \varphi(\alpha x). \quad (2.3)$$

Here  $\alpha$  is an arbitrary constant,  $\varphi(x)$  compactly denotes the various fields in the model and  $\Delta_\varphi$ 's are their corresponding scaling dimensions. The consequences of this dilatation symmetry are more easily understood in a minimally-coupled frame displaying the standard Einstein-Hilbert term. This *Einstein frame* is achieved by a Weyl redefinition of the metric<sup>4</sup>

$$g_{\mu\nu} \rightarrow \Theta g_{\mu\nu}, \quad \Theta \equiv \frac{F_\infty^2}{h^2}, \quad F_\infty \equiv \frac{M_P}{\sqrt{\xi}}, \quad (2.4)$$

together with a Weyl rescaling of the vector and fermion fields,

$$A_\mu \rightarrow \Theta^{-1/2} A_\mu, \quad \psi \rightarrow \Theta^{-3/4} \psi. \quad (2.5)$$

After some trivial algebra we obtain an Einstein-frame action<sup>5</sup>

$$S = \int d^4x \sqrt{-g} \left[ \frac{M_P^2}{2} R - \frac{1}{2} M_P^2 K(\Theta) (\partial\Theta)^2 - \frac{\lambda}{4} F_\infty^4 - \frac{1}{4} \mathcal{F}_{\mu\nu} \mathcal{F}^{\mu\nu} - \frac{g^2}{4} F_\infty^2 B_\mu B^\mu - i\bar{\psi} \not{D} \psi - \frac{y}{\sqrt{2}} F_\infty \bar{\psi} \psi \right], \quad (2.6)$$

containing a non-canonical term for the  $\Theta$  field. The coefficient of this kinetic term,

$$K(\Theta) \equiv \frac{1}{4|a|\Theta^2}, \quad (2.7)$$

involves a quadratic pole at  $\Theta = 0$  and a constant

$$a \equiv -\frac{\xi}{1+6\xi} < 0, \quad (2.8)$$

varying between zero at  $\xi = 0$  and  $-1/6$  when  $\xi \rightarrow \infty$ . The  $\Theta$ -field kinetic term can be made canonical by performing an additional field redefinition,

$$\Theta^{-1} = \exp\left(\frac{2\sqrt{|a|}\phi}{M_P}\right), \quad (2.9)$$

<sup>4</sup>In spite of its extensive use in the literature, we avoid referring to point-wise rescalings of the metric as “conformal transformations.” For a comprehensive discussion on the differences between Weyl and conformal symmetries, see for instance Karananas and Monin (2016a,b).

<sup>5</sup>In order to keep the notation as simple as possible, we will not introduce different notations for the quantities defined in different Weyl-related frames. In particular, the implicit Lorentz contractions in this article should be understood to be performed with the metric of the frame under consideration.

mapping the vicinity of the pole at  $\Theta = 0$  to  $\phi \rightarrow \infty$ . The resulting action

$$S = \int d^4x \sqrt{-g} \left[ \frac{M_P^2}{2} R - \frac{1}{2} (\partial\phi)^2 - \frac{\lambda}{4} F_\infty^4 - \frac{1}{4} \mathcal{F}_{\mu\nu} \mathcal{F}^{\mu\nu} - \frac{g^2}{4} F_\infty^2 B_\mu B^\mu - i\bar{\psi} \not{D} \psi + \frac{y}{\sqrt{2}} F_\infty \bar{\psi} \psi \right], \quad (2.10)$$

is invariant under shift transformations  $\phi \rightarrow \phi + C$ , with  $C$  a constant. The exponential mapping in Equation (2.9) indicates that such translational symmetry is nothing else than the non-linear realization of the original scale invariance we started with in Equation (2.1) (Csaki et al., 2014). The Einstein-frame transition in Equation (2.4) is indeed equivalent to the spontaneous breaking of dilatations, since we implicitly required the field  $h$  to acquire a non-zero expectation value. The canonically normalized scalar field  $\phi$  is the associated Goldstone boson and as such it is completely decoupled from the matter fields  $B_\mu$  and  $\psi$ . The non-minimal coupling to gravity effectively replaces  $h$  by  $F_\infty$  in all dimension-4 interactions involving conformal degrees of freedom. Note, however, that this decoupling statement does not apply to scale-invariant extensions including additional scalar fields (Kaiser, 2010; Garcia-Bellido et al., 2011; Bezrukov et al., 2013; Kaiser et al., 2013; Kaiser and Sfakianakis, 2014; Karananas and Rubio, 2016) or other non-conformal interactions such as  $R^2$  terms (Starobinsky, 1980; Gorbunov and Panin, 2011, 2012; Gorbunov and Tokareva, 2013).

## 2.2. Higgs Inflation From Approximate Scale Invariance

Although the toy model presented above contains many of the key ingredients of Higgs inflation, it is not phenomenologically viable. In particular, the Einstein-frame potential is completely shift-symmetric and does not allow for inflation to end. On top of this limitation, the scalar field  $\phi$  is completely decoupled from all conformal fields, excluding the possibility of entropy production and the eventual onset of a radiation-dominated era. All these phenomenological limitations are intrinsically related to the *exact* realization of scale invariance and as such they should be expected to disappear once a (sizable) dimensionfull parameter is included into the action. This is precisely what happens in Higgs inflation. The total *Higgs inflation* action (Bezrukov and Shaposhnikov, 2008)

$$S = \int d^4x \sqrt{-g} \left[ \frac{M_P^2}{2} R + \xi H^\dagger H R + \mathcal{L}_{\text{SM}} \right], \quad (2.11)$$

contains two dimensionfull parameters: the reduced Planck  $M_P \equiv 1/\sqrt{8\pi G_N} = 2.435 \times 10^{18}$  GeV and the Higgs vacuum expectation value  $v_{\text{EW}} \simeq 250$  GeV responsible for the masses within the SM Lagrangian density  $\mathcal{L}_{\text{SM}}$ . Among these two scales, the Planck mass is the most important one at the large field values relevant for inflation. To illustrate how the inclusion of  $M_P$  modifies the results of the previous section, let us consider

the graviscalar part of Equation (2.11) in the unitary gauge  $H = (0, h/\sqrt{2})^T$ , namely

$$S = \int d^4x \sqrt{-g} \left[ \frac{M_P^2 + \xi h^2}{2} R - \frac{1}{2} (\partial h)^2 - U(h) \right], \quad (2.12)$$

with

$$U(h) = \frac{\lambda}{4} (h^2 - v_{EW}^2)^2, \quad (2.13)$$

the usual SM symmetry-breaking potential. As in the induced gravity scenario, the inclusion of the non-minimal coupling to gravity changes the strength of the gravitational interaction and makes it dependent on the Higgs field,

$$G_{N,\text{eff}} = \frac{G_N}{1 + 8\pi\xi G_N h^2}. \quad (2.14)$$

In order for the graviton propagator to be well-defined at all  $h$  values, the non-minimal coupling  $\xi$  must be positive<sup>6</sup>. If  $\xi \neq 0$ , this requirement translates into a weakening of the effective Newton constant at increasing Higgs values. For non-minimal couplings in the range  $1 \ll \xi \ll M_P^2/v_{EW}^2$ , this effect is important in the large-field regime  $h \gg M_P/\sqrt{\xi}$ , but completely negligible otherwise.

As we did in section 2.1, it is convenient to reformulate Equation (2.12) in the Einstein frame by performing a Weyl transformation  $g_{\mu\nu} \rightarrow \Theta g_{\mu\nu}$  with

$$\Theta^{-1} = 1 + \frac{h^2}{F_\infty^2}, \quad F_\infty \equiv \frac{M_P}{\sqrt{\xi}}. \quad (2.15)$$

In the new frame, all the non-linearities associated with the non-minimal Higgs-gravity interaction are moved to the scalar sector of the theory,

$$S = \int d^4x \sqrt{-g} \left[ \frac{M_P^2}{2} R - \frac{1}{2} M_P^2 K(\Theta) (\partial\Theta)^2 - V(\Theta) \right], \quad (2.16)$$

which contains now a *non-exactly flat* potential

$$V(\Theta) \equiv U(\Theta) \Theta^2 = \frac{\lambda F_\infty^4}{4} \left[ 1 - \left( 1 + \frac{v_{EW}^2}{F_\infty^2} \right) \Theta \right]^2, \quad (2.17)$$

and a non-canonical kinetic sector resulting from the rescaling of the metric determinant and the non-homogeneous part of the Ricci scalar transformation. The kinetic function

$$K(\Theta) \equiv \frac{1}{4|a|\Theta^2} \left( \frac{1 - 6|a|\Theta}{1 - \Theta} \right), \quad (2.18)$$

shares some similarities with that in Equation (2.7). In particular, it contains two poles located respectively at  $\Theta = 0$  and  $\Theta = 1$ .

<sup>6</sup>Models with negative  $\xi$  have been considered in the literature (Herranen et al., 2014; Kamada, 2015a; Figueroa et al., 2018). In this type of scenarios the gravitational instability at large field values can be avoided by replacing the quadratic coupling  $\xi h^2$  by a designed function  $\xi f(h)$  remaining smaller than  $M_P^2$  during the whole field regime.

The first pole is an *inflationary pole*, like the one appearing in the induced gravity scenario. This pole leads to an enhanced friction for the  $\Theta$  field around  $\Theta = 0$  and allows for inflation to happen even if the potential  $V(\Theta)$  is not sufficiently flat. The second pole is a *Minkowski pole* around which the Weyl transformation equals one and the usual SM action is approximately recovered. To see this explicitly, we carry out an additional field redefinition<sup>7</sup>,

$$\frac{1}{M_P^2} \left( \frac{d\phi}{d\Theta} \right)^2 = K(\Theta), \quad (2.19)$$

to recast Equation (2.16) in terms of a canonically normalized scalar field  $\phi$ . This differential equation admits an exact solution (Garcia-Bellido et al., 2009)

$$\frac{\sqrt{|a|}\phi}{M_P} = \text{arcsinh} \sqrt{\frac{1 - \Theta}{(1 - 6|a|\Theta)}} - \sqrt{6|a|} \text{arcsinh} \sqrt{\frac{6|a|(1 - \Theta)}{1 - 6|a|}}. \quad (2.20)$$

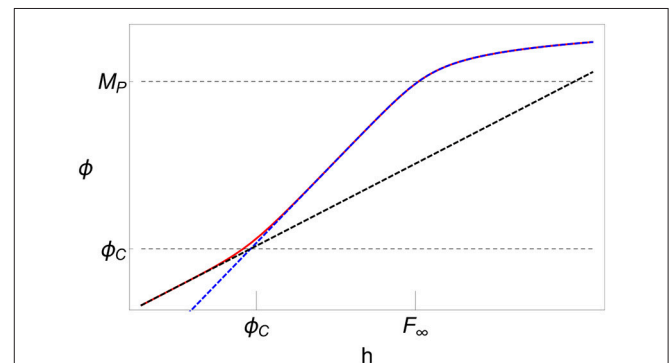
In terms of the original field  $h$ , we can distinguish two asymptotic regimes

$$\phi \simeq \begin{cases} h & \text{for } \phi < \phi_C, \\ \frac{M_P}{2\sqrt{|a|}} \log \left( 1 + \frac{h^2}{F_\infty^2} \right) & \text{for } \phi > \phi_C, \end{cases} \quad (2.21)$$

separated by a critical value

$$\phi_C \equiv \frac{2M_P(1 - 6|a|)}{\sqrt{|a|}}. \quad (2.22)$$

The comparison between these approximate expressions and the exact field redefinition in Equation (2.20) is shown in **Figure 3**.



**FIGURE 3** | Comparison between the approximate expressions in Equation (2.21) (dashed black and blue lines) and the exact solution (2.20) (solid red). Below the critical scale  $\phi_C$ , Higgs inflation coincides, up to highly suppressed corrections, with the SM minimally coupled to gravity. Above that scale, the Higgs field starts to decouple from the SM particles. The decoupling becomes efficient at a scale  $F_\infty$ , beyond which the model can be well approximated by a chiral SM with no radial Higgs component.

<sup>7</sup>Note that all equations till this point hold even if the non-minimal coupling  $\xi$  is field-dependent (Ezquiaga et al., 2018; Masina, 2018).

The large hierarchy between the transition scale  $\phi_C$  and the electroweak scale allows us to identify in practice the vacuum expectation value  $v_{EW}$  with  $\phi = 0$ . In this limit, the Einstein-frame potential (2.17) can be rewritten as

$$V(\phi) \simeq \frac{\lambda}{4} F^4(\phi), \quad (2.23)$$

with

$$F(\phi) \equiv \begin{cases} \phi & \text{for } \phi < \phi_C, \\ F_\infty \left(1 - e^{-\frac{2\sqrt{|a|}\phi}{M_P}}\right)^{\frac{1}{2}} & \text{for } \phi > \phi_C. \end{cases} \quad (2.24)$$

At  $\phi < \phi_C$  we recover the usual Higgs potential (up to highly suppressed corrections, cf. section 3.1). At  $\phi > \phi_C$  the Einstein-frame potential becomes exponentially stretched and approaches the asymptotic value  $F_\infty$  at  $\phi > M_P/(2\sqrt{|a|})$ . The presence of  $M_P$  in Equation (2.11) modifies also the decoupling properties of the Higgs field as compared to those in the induced gravity scenario. In particular, the masses of the intermediate gauge bosons and fermions in the Einstein-frame<sup>8</sup>,

$$m_B^2(\phi) \equiv \frac{g^2}{4} F^2(\phi), \quad m_F(\phi) \equiv \frac{y}{\sqrt{2}} F(\phi), \quad (2.25)$$

coincide with the SM masses in the small field regime ( $\phi < \phi_C$ ) and evolve toward constant values proportional to  $F_\infty$  in the large-field regime ( $\phi > M_P/(2\sqrt{|a|})$ ). The transition to the Einstein-frame effectively replaces  $h$  by  $F(\phi)$  in all (non-derivative) SM interactions. This behavior allows us to describe the Einstein-frame matter sector in terms of a chiral SM with vacuum expectation value  $F(\phi)$  (Dutta et al., 2008; Bezrukov and Shaposhnikov, 2009).

### 2.3. Tree-Level Inflationary Predictions

The flattening of the Einstein-frame potential (2.23) due to the  $\Theta = 0$  pole allows for inflation with the usual slow-roll conditions even if the potential  $V(\Theta)$  is not sufficiently flat. Let us compute the inflationary observables in the corresponding region  $\phi > \phi_C$ , where

$$V(\phi) \simeq \frac{\lambda F_\infty^4}{4} \left(1 - e^{-\frac{2\sqrt{|a|}\phi}{M_P}}\right)^2. \quad (2.26)$$

The statistical information of the primordial curvature fluctuations generated by a single-field model like the one under consideration is mainly encoded in the two-point correlation functions of scalar and tensor perturbations, or equivalently in their Fourier transform, the power spectra. Following the standard approach (Mukhanov et al., 1992), we parameterize these spectra in an almost scale-invariant form,

$$P_s = A_s \left(\frac{k}{k_*}\right)^{n_s-1}, \quad P_t = A_t \left(\frac{k}{k_*}\right)^{n_t}, \quad (2.27)$$

<sup>8</sup>Here we use a compact notation for the gauge boson couplings, namely  $g = g_2$  and  $g_2 \cos \theta_w$  for the  $B = W^\pm$  and  $Z$  bosons respectively, with  $g_1$  and  $g_2$  the gauge couplings of the  $U(1)_Y$  and  $SU(2)_L$  SM groups and  $\theta_w = \tan^{-1}(g_1/g_2)$  the weak mixing angle. The coupling  $y$  denotes a generic Yukawa coupling.

and compute the inflationary observables

$$A_s = \frac{1}{24\pi^2 M_P^4} \frac{V}{\epsilon}, \quad n_s = 1 + 2\eta - 6\epsilon, \quad r \equiv \frac{A_t}{A_s} = -8n_t = 16\epsilon, \quad (2.28)$$

with

$$\epsilon \equiv \frac{M_P^2}{2} \left(\frac{V'}{V}\right)^2, \quad \eta \equiv M_P^2 \frac{V''}{V}, \quad (2.29)$$

the first and second slow-roll parameters and the primes denoting derivatives with respect to  $\phi$ . The quantities in (2.28) should be understood as evaluated at a field value  $\phi_* \equiv \phi(N_*)$ , with

$$N_* = \frac{1}{M_P} \int_{\phi_E}^{\phi_*} \frac{d\phi}{\sqrt{2\epsilon}} = \frac{1}{8|a|} \left( e^{2\sqrt{|a|}\phi/M_P} - \frac{2\sqrt{|a|}\phi}{M_P} \right) \Big|_{\phi_E}^{\phi_*} \quad (2.30)$$

the e-fold number at which the reference scale  $k_*$  in Equation (2.27) exits the horizon, i.e.  $k_* = a_* H_*$ . Here,

$$\phi_E = \frac{M_P}{2\sqrt{|a|}} \ln \left(1 + 2\sqrt{2|a|}\right), \quad (2.31)$$

stands for the field value at the end of inflation, which is defined, as usual, by the condition  $\epsilon(\phi_E) \equiv 1$ . Equation (2.30) admits an exact inversion,

$$e^{2\sqrt{|a|}\phi_*/M_P} = -\mathcal{W}_{-1} \left[ -e^{-8|a|\bar{N}_*} \right], \quad (2.32)$$

with  $\mathcal{W}_{-1}$  the lower branch of the Lambert function and

$$\bar{N}_* \equiv N_* + \frac{1}{8|a|} \left( e^{2\sqrt{|a|}\phi_E/M_P} - \frac{2\sqrt{|a|}\phi_E}{M_P} \right), \quad (2.33)$$

a rescaled number of e-folds. Inserting Equation (2.32) into (2.28) we get the following analytical expressions for the primordial scalar amplitude,

$$A_s = \frac{\lambda(1 - 6|a|)^2}{12\pi^2 |a|} \frac{(1 + \mathcal{W}_{-1})^4}{(8|a|\mathcal{W}_{-1})^2}, \quad (2.34)$$

its spectral tilt,

$$n_s = 1 - 16|a| \frac{1 - \mathcal{W}_{-1}}{(1 + \mathcal{W}_{-1})^2}, \quad (2.35)$$

and the tensor-to-scalar ratio

$$r = \frac{128|a|}{(1 + \mathcal{W}_{-1})^2}. \quad (2.36)$$

At large  $|a|\bar{N}_*$ , these predictions display an interesting attractor behavior, very similar to that appearing in  $\alpha$ -attractor scenarios (Ferrara et al., 2013; Kallosh et al., 2013; Galante et al., 2015) (see also Artymowski and Rubio, 2016). Indeed, by taking into account the lower bound on the Lambert function (Chatzigeorgiou, 2016),

$$\mathcal{W}_{-1}[-e^{-8|a|\bar{N}_*}] > -8|a|\bar{N}_* - \sqrt{2(8|a|\bar{N}_* - 1)}, \quad (2.37)$$

we can obtain the approximate expressions<sup>9</sup>

$$n_s \simeq 1 - \frac{2}{\bar{N}_*}, \quad r \simeq \frac{2}{|a|\bar{N}_*^2}. \quad (2.38)$$

at  $8|a|\bar{N}_* \gg 1$ . The free parameter  $|a|$  (or equivalently the non-minimal coupling  $\xi$ ) can be fixed by combining Equation (2.34) with the normalization of the primordial spectrum at large scales (Akrami et al., 2018),

$$\log(10^{10} A_s) \simeq 3.094 \pm 0.034. \quad (2.39)$$

Doing this, we get a relation

$$\xi \simeq 800\bar{N}_*\sqrt{\lambda}, \quad (2.40)$$

among the non-minimal coupling  $\xi$ , the number of e-folds  $\bar{N}_*$  and the Higgs self-coupling  $\lambda$ .

The precise value of the number of e-folds in Equations (2.38), (2.40) depends on the whole post-inflationary expansion and, in particular, on the duration of the heating stage. As the strength of the interactions among the Higgs field and the SM particles is experimentally known, the entropy production following the end of inflation can be computed in detail (Bezrukov et al., 2009a; Garcia-Bellido et al., 2009; Repond and Rubio, 2016)<sup>10</sup>. The depletion of the Higgs-condensate is dominated by the non-perturbative production of massive intermediate gauge bosons, which, contrary to the SM fermions, can experience bosonic amplification. Once created, the  $W^\pm$  and  $Z$  bosons can decay into lighter SM fermions with a decay probability proportional to the instantaneous expectation value of the Higgs field  $\phi(t)$ . The onset of the radiation-domination era is determined either by i) the time at which the Higgs amplitude approaches the critical value  $\phi_C$  where the effective potential becomes quartic or by ii) the moment at which the energy density into relativistic fermions approaches that of the Higgs condensate; whatever happens first. The estimates in Garcia-Bellido et al. (2009), Bezrukov et al. (2009a), and Repond and Rubio (2016) provide a range

$$10^{13} \text{ GeV} \lesssim T_H \lesssim 2 \times 10^{14} \text{ GeV}, \quad (2.41)$$

with the lower and upper bounds associated respectively with the cases i) and ii) above. For the upper limit of this narrow window, we have  $\bar{N}_* \simeq N_* \simeq 59$  and we can rewrite Equation (2.40) as a relation between  $\xi$  and  $\lambda$ ,

$$\xi \simeq 47200\sqrt{\lambda}. \quad (2.42)$$

Note that a variation of the Higgs self-coupling in this equation can be compensated by a change of the *a priori* unknown non-minimal coupling to gravity. For the tree-level value  $\lambda \sim \mathcal{O}(1)$ , the non-minimal coupling must be significantly larger than one, but still much smaller than the value  $\xi \sim M_P^2/v_{EW}^2 \sim 10^{32}$  leading to sizable modifications of the effective Newton constant

at low energies. In this regime, the parameter  $|a|$  is very close to its maximum value  $1/6$ . This effective limit simplifies considerably the expression for the critical scale  $\phi_C$  separating the low- and high-energy regimes,

$$\phi_C \simeq \sqrt{\frac{2}{3}} \frac{M_P}{\xi}, \quad (2.43)$$

and collapses the inflationary predictions to the attractor values (Bezrukov and Shaposhnikov, 2008)

$$n_s \simeq 1 - \frac{2}{\bar{N}_*} \simeq 0.966, \quad r \simeq \frac{12}{\bar{N}_*^2} \simeq 0.0034, \quad (2.44)$$

in very good agreement with the latest results of the Planck collaboration (Akrami et al., 2018). Note that, although computed in the Einstein frame, these predictions could have been alternatively obtained in the non-minimally coupled frame (2.12), provided a suitable redefinition of the slow-roll parameters in order to account for the Weyl factor relating the two frames (Makino and Sasaki, 1991; Fakir et al., 1992; Komatsu and Futamase, 1999; Flanagan, 2004; Tsujikawa and Gumjudpai, 2004; Koh, 2006; Chiba and Yamaguchi, 2008, 2013; Weenink and Prokopec, 2010; Postma and Volponi, 2014; Ren et al., 2014; Jarv et al., 2015a,b, 2017; Burns et al., 2016; Kuusk et al., 2016; Karam et al., 2017; Karamitsos and Pilaftsis, 2018a,b).

### 3. EFFECTIVE FIELD THEORY INTERPRETATION

The presence of gravity makes Higgs inflation perturbatively non-renormalizable (Barbon and Espinosa, 2009; Burgess et al., 2009, 2010; Bezrukov et al., 2011b) and forbids its interpretation as an ultraviolet complete theory. The model should be therefore understood as an effective description valid up to a given cut-off scale  $\Lambda$  (Bezrukov et al., 2011b; George et al., 2016). This cutoff could either indicate the onset of a strongly coupled regime to be studied within the model by non-perturbative techniques (such as resummations, lattice simulations or functional renormalization studies) (Aydemir et al., 2012; Calmet and Casadio, 2014; Saltas, 2016; Escrivà and Germani, 2017) or the appearance of new degrees of freedom beyond the initially-assumed SM content (Giudice and Lee, 2011; Barbon et al., 2015).

#### 3.1. The Cutoff Scale

*A priori*, the cutoff scale of Higgs inflation could coincide with the Planck scale, where gravitational effects should definitely taken into account. Although quite natural, the identification of these two energy scales may not be theoretically consistent, since other interactions could lead to violations of tree-level unitarity at a lower energy scale. An estimate<sup>11</sup> of the cutoff scale can be obtained by expanding the fields around their background values,

<sup>9</sup>Note that the expressions contain  $\bar{N}_*$  rather than  $N_*$ .

<sup>10</sup>This allows, for instance, to distinguish Higgs inflation from  $R^2$  Starobinsky inflation (Starobinsky, 1980; Bezrukov and Gorbunov, 2012).

<sup>11</sup>This procedure does not take into account possible cancellations among scattering diagrams, as those taking place, for instance, in models involving a *singlet* scalar field not minimally coupled to gravity (Hertzberg, 2010).

such that all kind of higher dimensional operators appear in the resulting action (Bezrukov et al., 2011b; Ferrara et al., 2011). The computation is technically simpler in the original frame (2.11). In order to illustrate the procedure let us consider the graviscalar sector in Equation (2.12). Expanding the fields around their background values  $\bar{g}_{\mu\nu}$  and  $\bar{h}$ ,

$$g_{\mu\nu} = \bar{g}_{\mu\nu} + \gamma_{\mu\nu}, \quad h = \bar{h} + \delta h, \quad (3.1)$$

we obtain the following quadratic Lagrangian density for the perturbations  $\gamma_{\mu\nu}$  and  $\delta h$

$$\mathcal{L}^{(2)} = \frac{M_P^2 + \xi \bar{h}^2}{8} (\gamma^{\mu\nu} \square \gamma_{\mu\nu} + 2 \partial_\nu \gamma^{\mu\nu} \partial^\rho \gamma_{\mu\rho} - 2 \partial_\nu \gamma^{\mu\nu} \partial_\mu \gamma - \gamma \square \gamma) - \frac{1}{2} (\partial_\mu \delta h)^2 + \xi \bar{h} (\partial_\lambda \partial_\rho \gamma^{\lambda\rho} - \square \gamma) \delta h, \quad (3.2)$$

with  $\gamma = \bar{g}^{\mu\nu} \gamma_{\mu\nu}$  denoting the trace of the metric excitations. For non-vanishing  $\xi$ , the last term in this equation mixes the trace of the metric perturbation with the scalar perturbation  $\delta h$  (Barvinsky et al., 2008, 2009; De Simone et al., 2009). To identify the different cutoff scales one must first diagonalize the kinetic terms. This can be done by performing a redefinition of the perturbations  $(\gamma_{\mu\nu}, \delta h) \rightarrow (\hat{\gamma}_{\mu\nu}, \hat{\delta h})$  with

$$\gamma_{\mu\nu} = \frac{1}{\sqrt{M_P^2 + \xi \bar{h}^2}} \hat{\gamma}_{\mu\nu} - \frac{2\xi \bar{h} \bar{g}_{\mu\nu}}{\sqrt{(M_P^2 + \xi \bar{h}^2)(M_P^2 + (1 + 6\xi)\xi \bar{h}^2)}} \delta \hat{h}, \quad (3.3)$$

$$\delta h = \sqrt{\frac{M_P^2 + \xi \bar{h}^2}{M_P^2 + (1 + 6\xi)\xi \bar{h}^2}} \delta \hat{h}. \quad (3.4)$$

Once Equation (3.2) has been reduced to a diagonal form, we can proceed to read the cutoff scales. The easiest one to identify is that associated with purely gravitational interactions,

$$\Lambda_P(\bar{h}) \equiv \sqrt{M_P^2 + \xi \bar{h}^2}, \quad (3.5)$$

which coincides with the effective Planck scale in Equation (2.12). For scalar-graviton interactions, the leading-order higher-dimensional operator is  $(\delta \hat{h})^2 \square \hat{\gamma} / \Lambda_S(\bar{h})$ , where

$$\Lambda_S(\bar{h}) \equiv \frac{M_P^2 + (1 + 6\xi)\xi \bar{h}^2}{\xi \sqrt{M_P^2 + \xi \bar{h}^2}}. \quad (3.6)$$

Although we have focused on the graviscalar sector of the theory, the lack of renormalizability associated with the non-minimal coupling to gravity permeates all SM sectors involving the Higgs field. One could study, for instance, the scattering of intermediate  $W^\pm$  and  $Z$  bosons. Since we are working in the unitary gauge, it is sufficient to consider the longitudinal polarization. The modification of the Higgs kinetic term at large field values changes the delicate pattern of cancellations in the SM and leads to a tree-level unitarity violation at a scale

$$\Lambda_G(\bar{h}) \equiv \frac{\sqrt{M_P^2 + \xi(1 + 6\xi)\bar{h}^2}}{\sqrt{6\xi}}. \quad (3.7)$$

Note that the above scales depend on the background field  $\bar{h}$ . For small field values ( $\bar{h} \lesssim M_P/\xi$ ), the cutoffs (3.5), (3.6) and (3.7) coincide with those obtained by naively expanding the theory around the electroweak scale, namely  $\Lambda_P \simeq M_P$ ,  $\Lambda_S \simeq \sqrt{6}\Lambda_G \simeq M_P/\xi$  (Barbon and Espinosa, 2009; Burgess et al., 2009, 2010; Hertzberg, 2010; Atkins and Calmet, 2011). At large field values, ( $\bar{h} \gtrsim M_P/\xi$ ), the suppression scale depends on the particular process under consideration. For  $M_P/\xi \ll \bar{h} \ll M_P/\sqrt{\xi}$  the graviscalar cutoff  $\Lambda_S$  grows quadratically till  $\bar{h} \simeq M_P/\sqrt{\xi}$ , where it becomes linear in  $\bar{h}$  and traces the *dynamical* Planck mass in that regime,  $\Lambda_P \simeq \sqrt{\xi} \bar{h}$ . On the other hand, the gauge cutoff  $\Lambda_G$  smoothly interpolates between  $\Lambda_G \sim M_P/\xi$  at  $\bar{h} \lesssim M_P/\xi$  and  $\Lambda_G \sim g \bar{h}$  at  $\bar{h} \gtrsim M_P/\xi$ . Note that all cutoffs scales become linear in  $\bar{h}$  at  $\bar{h} \gtrsim M_P/\xi$ . This means that any operator  $\Delta \mathcal{L}$  constructed out of them, the Higgs field and some Wilson coefficients  $c_n$  approaches a scale-invariant form at large field values, namely

$$\Delta \mathcal{L} \equiv \sum_n \frac{c_n \mathcal{O}_n[\bar{h}]}{[\Lambda(\bar{h})]^{n-4}} \simeq \sum_n \frac{c_n \mathcal{O}_n[\bar{h}]}{(\sqrt{\xi} \bar{h})^{n-4}} \sim \sum_n \frac{c_n}{(\sqrt{\xi})^{n-4}} h^4, \quad n > 4. \quad (3.8)$$

## 3.2. Relation Between High- and Low-Energy Parameters

In what follows we will assume that the ultraviolet completion of the theory respects the original symmetries of the tree-level action, and in particular the approximate scale invariance of Equation (2.12) in the large-field regime and the associated shift-symmetry of its Einstein-frame formulation. This strong assumption forbids the generation of dangerous higher-dimensional operators that would completely spoil the predictivity of the model. In some sense, this requirement is not very different from the one implicitly assumed in other inflationary models involving trans-Planckian field displacements.

The minimal set of higher-dimensional operators to be included on top of the tree-level action is the one generated by the theory itself via radiative corrections (Bezrukov et al., 2011b, 2015). The cancellation of the loop divergences stemming from the original action requires the inclusion of an infinite set of counterterms with a very specific structure. As in any other *non-renormalizable* theory, the outcome of this subtraction procedure *depends* on the renormalization scheme, with different choices corresponding to different assumptions about the ultraviolet completion of the theory. Among the different subtractions setups, a dimensional regularization scheme involving a *field-dependent* subtraction point (Bezrukov and Shaposhnikov, 2009)

$$\mu^2 \propto M_P^2 + \xi h^2, \quad (3.9)$$

fits pretty well with the approximate scale-symmetry of Equation (2.11) at large-field values<sup>12</sup>. Given this frame and

<sup>12</sup>The use of other schemes such as Pauli-Villars regularization or standard dimensional regularization with *field-independent* subtraction point leads to dilatation-symmetry breaking and the consequent bending of the Higgs inflation plateau due to radiative corrections (see for instance Barvinsky et al., 2008, 2009, 2012; De Simone et al., 2009).

scheme, the minimal set of higher-dimensional operators generated by the theory can be computed in any Weyl-related frame provided that all fields and dimensionfull parameters are appropriately rescaled. The computation becomes particularly simple in the Einstein-frame, where the Weyl-rescaled renormalization point  $\mu^2\Theta$  coincides with the standard field-independent prescription of renormalizable field theories,  $\mu^2\Theta \propto M_p^2$ . A general counter-term in dimensional regularization contains a finite part  $\delta\mathcal{L}$  and a divergent part in the form of a pole in  $\epsilon = (4-d)/2$ , with  $d$  the dimension of spacetime. The coefficient of the pole is chosen to cancel the loop divergences stemming from the original action. Once this divergent part is removed, we are left with the finite contribution  $\delta\mathcal{L}$ . The strength of this term encodes the remnants of a particular ultraviolet completion and cannot be determined within the effective field theory approach (Bezrukov et al., 2011b, 2015; Burgess et al., 2014). From a quantitative point of view, the most relevant  $\delta\mathcal{L}$  terms are related to the Higgs and top-quark interactions. In the Einstein-frame at one loop, they take the form (Bezrukov et al., 2015)

$$\begin{aligned}\delta\mathcal{L}_1^F &= \left[ \delta\lambda_a \left( F'^2 + \frac{1}{3} F'' F \right)^2 - \delta\lambda_b \right] F^4, \\ \delta\mathcal{L}_1^\psi &= [\delta y_a F'^2 F + \delta y_b F'' (F^4)'] \bar{\psi} \psi,\end{aligned}\quad (3.10)$$

where the primes denote again derivatives with respect to  $\phi$ . Note that these operators differ, as expected, from those appearing in the tree-level action. This means that, while the contribution  $\delta\lambda_b$  can be removed by a self-coupling redefinition, the finite parts  $\delta\lambda_a, \delta y_a$  and  $\delta y_b$  should be promoted to new couplings constants. Once the associated operators are added to the tree-level action, the re-evaluation of radiative corrections will generate additional contributions beyond the original one-loop result. These contributions come together with new finite parts that must be again promoted to novel couplings with their own renormalization group equations. The iteration of this scheme leads to a renormalized action including an infinite set of higher-dimensional operators constructed out of the function  $F$  and its derivatives. For small field values, the function  $F$  becomes approximately linear ( $F \approx \phi, F' = 1$ ) and one recovers the SM non-minimally coupled to gravity up to highly suppressed interactions. In this limit, the coefficients of the infinite set of counterterms can be eliminated by a redefinition of the low energy couplings, as happens in a renormalizable theory. When evolving toward the inflationary region, the function  $F$  becomes approximately constant ( $F_\infty = F_\infty, F' = 0$ ) and some of the previously absorbed finite parts are dynamically subtracted. The unknown finite parts modify therefore the running of the SM couplings at the transition region  $\phi_C < \phi < \sqrt{3/2} M_p$ , such that the SM masses at the electroweak scale cannot be unambiguously related to their inflationary counterparts without a precise knowledge of the ultraviolet completion (Hertzberg, 2012; Burgess et al., 2014; Bezrukov et al., 2015).

If the finite contributions are of the same order as the loops generating them, the tower of higher dimensional operators generated by radiative corrections can be truncated (Bezrukov

et al., 2015). In this case, the effect of the 1-loop threshold corrections can be imitated by an effective change<sup>13</sup> (Bezrukov et al., 2015)

$$\begin{aligned}\lambda(\mu) &\rightarrow \lambda(\mu) + \delta\lambda_a \left[ \left( F'^2 + \frac{1}{3} F'' F \right)^2 - 1 \right], \\ y_t(\mu) &\rightarrow y_t(\mu) + \delta y_a [F'^2 - 1],\end{aligned}\quad (3.11)$$

with  $\lambda(\mu)$  and  $y_t(\mu)$  given by the SM renormalization group equations. We emphasize, however, that the truncation of the renormalization group equations is not essential for most of the results presented below, since, within the self-consistent approach to Higgs inflation, the functional form of the effective action is almost insensitive to it (Bezrukov et al., 2011b, 2015, 2018).

### 3.3. Potential Scenarios and Inflationary Predictions

To describe the impact of radiative corrections on the inflationary predictions, we will make use of the renormalization group enhanced potential. This is given by the one in Equation (2.26) but with the Higgs self-coupling  $\lambda$  replaced by its corresponding running value  $\lambda(\phi)$ ,

$$V(\phi) = \frac{\lambda(\phi)}{4} F^4(\phi). \quad (3.12)$$

Note that we are not promoting the non-minimal coupling  $\xi$  within  $F(\phi)$  to a running coupling  $\xi(\phi)$ —as done, for instance, in Ezquiaga et al. (2018)—but rather assuming it to be constant during inflation. This is indeed a reasonable approximation since the one-loop beta function determining the running of  $\xi$  (Yoon and Yoon, 1997; Bezrukov and Shaposhnikov, 2009),

$$\beta_\xi(\mu) = \mu \frac{\partial}{\partial \mu} \xi = -\frac{1}{16\pi^2} \xi \left( \frac{3}{2} g'^2 + 3g^2 - 6y_t^2 \right), \quad (3.13)$$

is rather small for realistic values of the couplings constant at the inflationary scale,  $\beta_\xi \propto \mathcal{O}(10^{-2})$  (Bezrukov et al., 2018; Masina, 2018) (see also Salvio, 2018).

Although, strictly speaking, the renormalization group enhanced potential is not gauge invariant, the gauge dependence is small during slow-roll inflation, especially in the presence of extrema (Cook et al., 2014; Espinosa et al., 2015, 2017). In the vicinity of the minimum of  $\lambda(\phi)$ , we can use the approximation (Bezrukov and Shaposhnikov, 2014)

$$\lambda(\phi) = \lambda_0 + b \log^2 \left( \frac{m_t(\phi)}{q} \right), \quad (3.14)$$

with the parameters  $\lambda_0, q$  and  $b$  depending on the inflationary values of the Einstein-frame Higgs and top quark masses, according to the fitting formulas (Bezrukov and Shaposhnikov, 2014)

$$\lambda_0 = 0.003297 [(m_H^* - 126.13) - 2(m_t^* - 171.5)],$$

<sup>13</sup>This replacement implicitly neglects the running of the finite parts  $\delta\lambda_a$  and  $\delta y_a$  in the transition region  $\phi_C < \phi < \sqrt{3/2} M_p$ .

$$q = 0.3M_P \exp \left[ 0.5(m_H^* - 126.13) - 0.03(m_t^* - 171.5) \right],$$

$$b = 0.00002292 - 1.12524 \times 10^{-6} [(m_H^* - 126.13) - 1.75912(m_t^* - 171.5)], \quad (3.15)$$

with  $m_H^*$  and  $m_t^*$  in GeV. As seen in the last expression, the parameter  $b$ , standing for the derivative of the beta function for  $\lambda$  at the scale of inflation, is rather insensitive to the Higgs and top quark mass values at that scale and can be well-approximated by  $b \simeq 2.3 \times 10^{-5}$ . The choice

$$\frac{m_t(\phi)}{q} = \alpha \cdot \frac{y_t}{\sqrt{2}} \frac{F(\phi)}{q} \equiv \frac{\sqrt{\xi} F(\phi)}{\kappa M_P}, \quad (3.16)$$

with  $\alpha = 0.6$  optimizes the convergence of perturbation theory (Bezrukov and Shaposhnikov, 2009; Bezrukov et al., 2009b), while respecting the asymptotic symmetry of the tree-level action (2.12) and its non-linear shift-symmetric Einstein-frame realization. In the second equality, we have introduced an effective parameter  $\kappa$  to facilitate the numerical computation of the inflationary observables.

A simple inspection of Equations (3.12) and (3.14) allows us to distinguish three regimes:

- i) *Non-critical regime/Universal*: If  $\lambda_0 \gg b/(16\kappa)$ , the effective potential (3.12) is almost independent of the radiative logarithmic correction and can be well approximated by its tree-level form (2.26). Consequently, the inflationary observables retain their tree-level values (Bezrukov and Shaposhnikov, 2014; Enckell et al., 2016; Fumagalli and Postma, 2016; Bezrukov et al., 2018), cf. **Figure 4**.
- ii) *Critical regime*: If  $\lambda_0 \gtrsim b/(16\kappa)$ , the first two derivatives of the potential are approximately zero,  $V' \simeq V'' \simeq 0$ , leading to the appearance of a *quasi*-inflection point at

$$\phi_I = \sqrt{\frac{3}{2}} \log \left( \frac{\sqrt{e}}{\sqrt{e}-1} \right) M_P. \quad (3.17)$$

Qualitatively, the vast majority of inflationary e-folds in this scenario takes place in the vicinity of the inflection point  $\phi_I$ , while the inflationary observables depend on the form of the potential as some value  $\phi_* > \phi_I$ .

Given the small value of the Higgs self-coupling in this scenario,  $\lambda_0 \sim \mathcal{O}(10^{-6})$ , the nonminimal coupling  $\xi$  can be significantly smaller than in the universal regime,  $\xi \sim \mathcal{O}(10)$ , while still satisfying the normalization condition (2.39) (Allison, 2014; Bezrukov and Shaposhnikov, 2014; Hamada et al., 2014b, 2015). This drastic decrease of the non-minimal coupling alleviates the tree-level unitary problems discussed in section 3.1 by raising the cutoff scale.

For small  $\xi$  values, the tensor-to-scalar ratio can be rather large,  $r \sim \mathcal{O}(10^{-1})$  (Allison, 2014; Bezrukov and Shaposhnikov, 2014; Hamada et al., 2014b, 2015) (see also Masina, 2018). Note, however, that although CMB data seems to be consistent with the primordial power spectrum at large scales, the simple expansion in (2.27) cannot accurately describe its global behavior since the running of the spectral tilt  $\alpha_s \equiv d \ln n_s / d \ln k$  and its scale dependence  $\beta_s \equiv d^2 \ln n_s / d \ln k^2$  also become considerably large, cf. **Figure 5**.

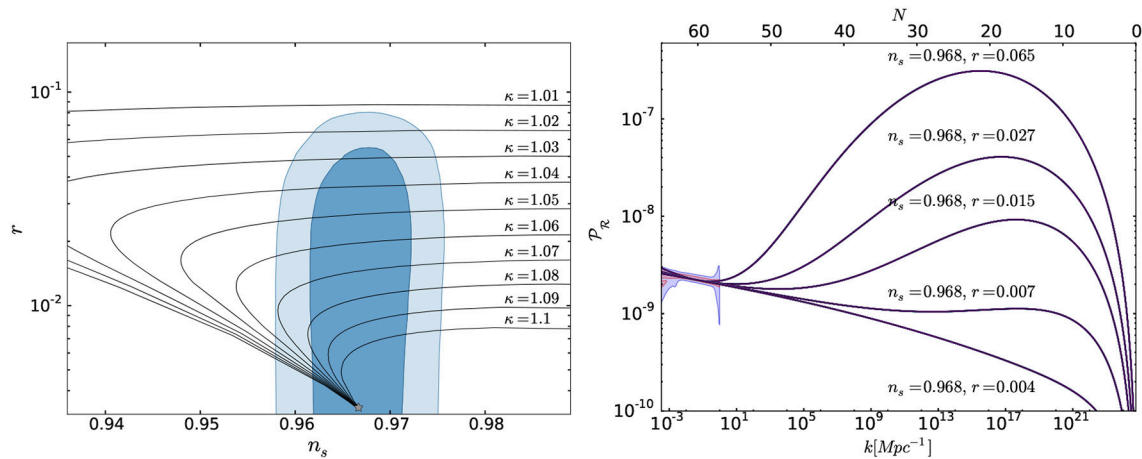
The non-monotonic evolution of the slow-roll parameter  $\epsilon$  in the vicinity of the inflection point leads to the enhancement of the spectrum of primordial density fluctuations at small and intermediate scales. It is important to notice at this point that the standard slow-roll condition may break down if the potential becomes extremely flat and the inertial contribution in the equation of motion for the inflation field is not negligible as compared with the Hubble friction (Garcia-Bellido and Ruiz Morales, 2017; Germani and Prokopec, 2017; Kannike et al., 2017). In this regime, even the classical treatment is compromised since stochastic effects can no longer be ignored (Starobinsky and Yokoyama, 1994; Vennin and Starobinsky, 2015; Pattison et al., 2017; Ezquiaga and Garcia-Bellido, 2018).

- If we restrict ourselves to situation in which the slow-roll approximation is satisfied during the whole inflationary trajectory (Bezrukov et al., 2018)<sup>14</sup>, the height and width of the generated bump at fixed spectral tilt are correlated with the tensor-to-scalar ratio  $r$ , cf. **Figure 4**. Contrary to some claims in the literature (Ezquiaga et al., 2018), the maximum amplitude of the power-spectrum compatible with the 95% C.L. Planck  $n_s - r$  contours (Bezrukov et al., 2018) is well below the critical threshold  $\mathcal{P}_{\mathcal{R}}^{\max} \simeq 10^{-2} - 10^{-3}$  needed for primordial black hole formation (Bird et al., 2016; Carr et al., 2016, 2017) (see however, Ezquiaga and Garcia-Bellido, 2018; Rasanen and Tomberg, 2018). This conclusion is unchanged if one considers the effect of non-instantaneous threshold corrections (Bezrukov et al., 2018), which could potentially affect the results given the numerical proximity of the inflection point (3.17) to the upper boundary of the transition region,  $\phi \simeq \sqrt{3/2} M_P$ .
- iii) *Hilltop regime*: If  $\lambda \lesssim b/(16\kappa)$  the potential develops a new minimum at large field values (Fumagalli, 2017; Rasanen and Wahlman, 2017). This minimum is separated from the electroweak minimum by a local maximum where hilltop inflation can take place (Boubekeur and Lyth, 2005; Barenboim et al., 2016). This scenario is highly sensitive to the initial conditions since the inflaton field must start on the electroweak vacuum side and close enough to the local maximum in order to support an extended inflationary epoch. On top of that, the fitting formulas in (3.15) may not be accurate enough for this case, since they are based on an optimization procedure around the  $\lambda(\phi)$  minimum. The tensor-to-scalar ratio in this scenario differs also from the universal/non-critical Higgs inflation regime, but contrary to the critical case, it is decreased to  $2 \times 10^{-5} < r < 1 \times 10^{-3}$ , rather than increased (Fumagalli and Postma, 2016; Rasanen and Wahlman, 2017).

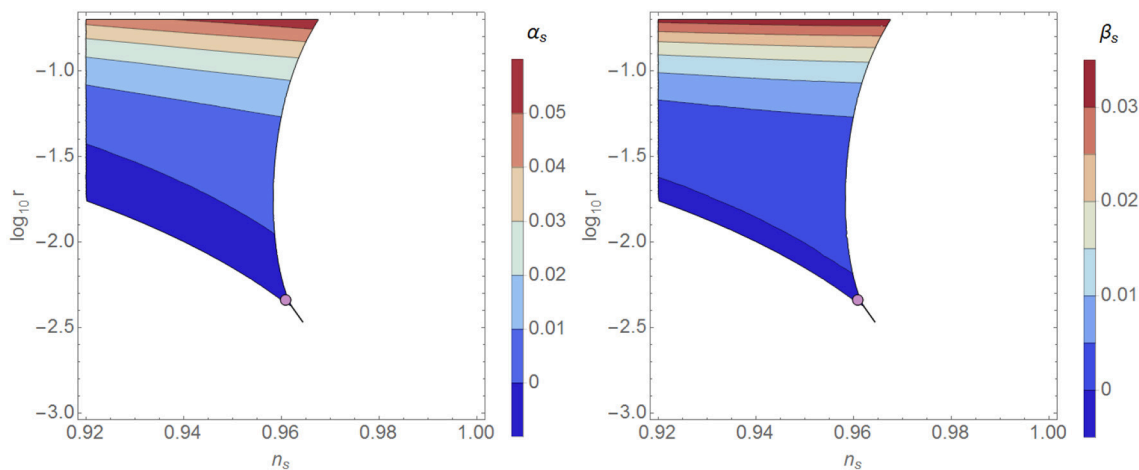
### 3.4. Vacuum Metastability and High-Temperature Effects

The qualitative classification of scenarios and predictions presented in the previous section depends on the *inflationary values* of the Higgs and top quark masses and holds

<sup>14</sup>The onset of the slow-roll regime prior to the arrival of the field to the inflection point and its dependence on pre-inflationary conditions was studied in Salvio (2018), where a robust inflationary attractor was shown to exist.



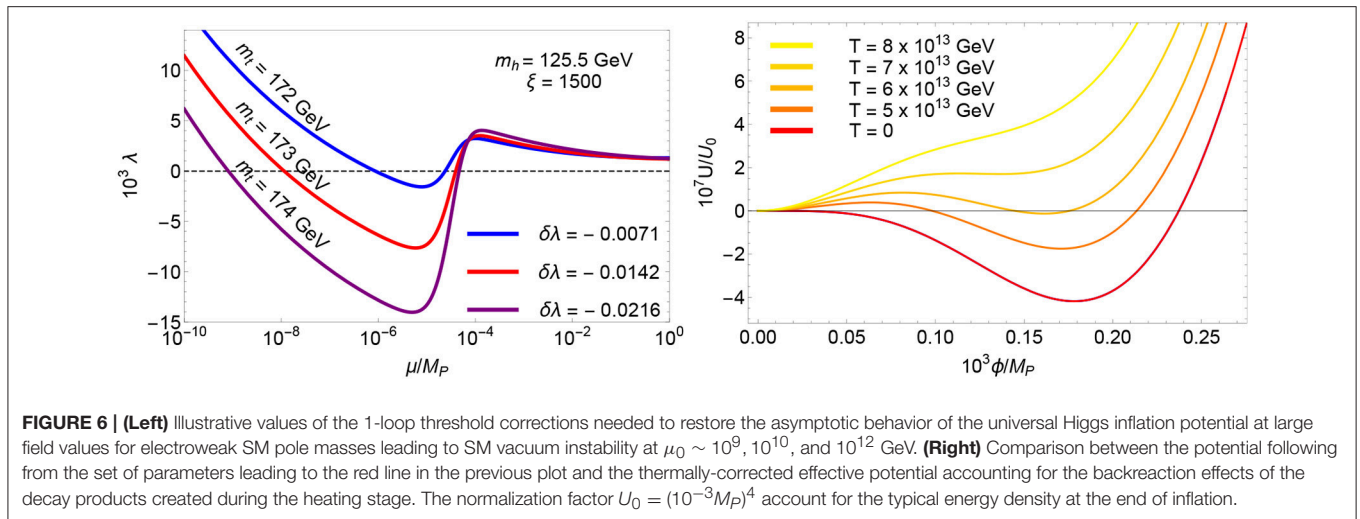
**FIGURE 4 | (Left)** The tensor-to-scalar ratio  $r$  and the spectral tilt  $n_s$  following from the effective potential (3.12) (Bezrukov et al., 2018). The non-minimal coupling  $\xi$  varies between 10 and 100 along the lines of constant  $\kappa$ , with larger values corresponding to smaller tensor-to-scalar ratios. The star in the lower part of the plot stands for the universal values in Equation (2.44). The blue contours indicate the latest 68 and 95% C.L. Planck constraints on the  $r$ - $n_s$  plane (Akrami et al., 2018). **(Right)** The power spectrum  $\mathcal{P}_R$  as a function of the number of e-folds before the end of inflation and the associated comoving scale  $k$  in inverse megaparsecs (Bezrukov et al., 2018). The monotonic curve at the bottom of the plot corresponds to the *universal/non-critical Higgs inflation* scenario. The upper non-monotonic curves are associated with different realizations of the *critical Higgs inflation* scenario. The shaded regions stand for the latest 68 and 95% C.L. constraints provided by the Planck collaboration (Akrami et al., 2018).



**FIGURE 5 | (Left)** Spectral tilt running  $\alpha_s \equiv d \ln n_s / d \ln k$  in critical Higgs inflation as a function of the tensor-to-scalar ratio  $r$  and the spectral-tilt  $n_s$  (Rasanen and Wahlman, 2017). **(Right)** Scale dependence of the spectral-tilt  $\beta_s \equiv d^2 \ln n_s / d \ln k^2$  in the same case (Rasanen and Wahlman, 2017). The purple dots indicate the universal/non-critical Higgs inflation regime. The boundaries on the right-hand side of the figures correspond to the constraint on the number of e-folds following the heating estimates in Garcia-Bellido et al. (2009), Bezrukov et al. (2009a), and Repond and Rubio (2016). For lower heating efficiencies, the boundaries move to the left, decreasing the spectral tilt but not significantly affecting the tensor-to-scalar ratio (Rasanen and Wahlman, 2017).

independently of the value of their electroweak counterparts. In particular, any pair of couplings following from the SM renormalization group equations can be connected to a well-behaved pair of couplings in the chiral phase by a proper choice of the unknown threshold corrections. This applies also if the SM vacuum is not completely stable. Some examples of the 1-loop threshold correction  $\delta\lambda_a$  needed to restore the universal/non-critical Higgs inflation scenario beyond  $\mu_0 \sim 10^9, 10^{10}$ , and  $10^{12}$  GeV are shown in **Figure 6**. For a detailed scan of the parameter space, see Enckell et al. (2016); Fumagalli and Postma (2016).

The non-trivial interplay between vacuum stability and threshold corrections generates an additional minimum at large field values. Provided the usual chaotic initial conditions, the Higgs field will evolve in the trans-Planckian field regime, inflating the Universe while moving toward smaller field values. Since the new minimum is significantly wider and deeper than the electroweak one, it seems likely that the Higgs field will finish its post-inflationary evolution there. Note, however, that this conclusion is strongly dependent on the ratio of the Higgs energy density to the second minimum depth. If this ratio is large,



the entropy production at the end of inflation may significantly modify the shape of the potential, triggering its stabilization and allowing the Higgs field to evolve toward the desired electroweak vacuum (Bezrukov et al., 2009a).

The one-loop finite temperature corrections to be added on top of the Einstein-frame renormalization group enhanced potential take the form (Linde, 1979)

$$\Delta V = T \sum_{i=B,F} \int \frac{d^3 k}{(2\pi)^3 a^3} \ln \left[ 1 \pm \exp \left( -\frac{k^2/a^2 + m_i^2}{T} \right) \right], \quad (3.18)$$

with the plus and minus signs corresponding respectively to fermions and bosons and  $m_{B,F}$  standing for the Einstein-frame masses in Equation (2.25). The most important contributions in Equation (3.18) are associated with the top quark and the electroweak bosons, with the corresponding coupling constants  $y_t$  and  $g$  evaluated at  $\mu_{y_t} = 1.8 T$  and  $\mu_g = 7 T$ , in order to minimize the radiative corrections (Kajantie et al., 1996).

A detailed analysis of the universal/non critical Higgs inflation scenario reveals that the temperature of the decay products generated during the heating stage exceeds generically the temperature at which the unwanted secondary vacuum at large field values disappears (Bezrukov et al., 2009a; Garcia-Bellido et al., 2009), see Figure 6<sup>15</sup>. The stabilization becomes favored for increasing  $\mu_0$  values<sup>16</sup>, and holds even if this scale is as low as  $10^{10}$  GeV (Bezrukov et al., 2009a). The thermally-corrected potential enables the Higgs field to relax to the SM vacuum. After the heating stage, the temperature decreases as the Universe expands and the secondary minimum reappears, first as a local minimum and eventually as the global one. When that happens, the Higgs field is already trapped in the electroweak vacuum. Although the barrier separating the two minima prevents a direct decay, the Higgs field could still tunnel to the global minimum.

<sup>15</sup>A detailed scan of the parameter space assuming instantaneous conversion of the inflaton energy density into a thermal bath was performed in Enckell et al. (2016).

<sup>16</sup>The larger  $\mu_0$  is, the shallower and narrower the “wrong” minimum becomes, cf. Figure 6

The probability for this to happen is, however, very small and the lifetime of SM vacuum significantly exceeds the life of the Universe (Anderson, 1990; Arnold and Vokos, 1991; Espinosa and Quiros, 1995; Espinosa et al., 2008). Universal/non-critical Higgs inflation with a graceful exit can therefore take place for electroweak SM pole masses leading to vacuum metastability at energies below the inflationary scale (Bezrukov et al., 2009a).

The situation changes completely if one considers the critical Higgs inflation scenario. In this case, the energy of the Higgs condensate is comparable to the depth of the secondary minimum and symmetry restoration does not take place. Unless the initial conditions are extremely fine-tuned, the Higgs field will relax to the minimum of the potential at Planckian values, leading with it to the inevitably collapse of the Universe (Felder et al., 2002). The success of critical Higgs inflation requires therefore the absolute stability of the electroweak vacuum (Bezrukov et al., 2009a).

## 4. VARIATIONS AND EXTENSIONS

Many variations and extensions of Higgs inflation have been considered in the literature (see for instance Ben-Dayana and Einhorn, 2010; Lerner and McDonald, 2010, 2011; Arai et al., 2011; Giudice and Lee, 2011; Kamada et al., 2011, 2012; Einhorn and Jones, 2012; Greenwood et al., 2013; Kanemura et al., 2013; Steinwachs, 2013; Choudhury et al., 2014; He and Xianyu, 2014; Oda and Tomoyose, 2014a,b; Xianyu and He, 2014; Cai et al., 2015; Ellis et al., 2015, 2016; Kamada, 2015b; Lazarides and Pallis, 2015; Okada and Shafi, 2015; Calmet and Kuntz, 2016; Ge et al., 2016; Takahashi and Takahashi, 2016; van de Bruck and Longden, 2016; Ema, 2017; Ema et al., 2017b; Marian et al., 2017; Okada and Raut, 2017; Chen et al., 2018; He et al., 2018). In what follows we will restrict ourselves to those proposals that are more closely related to the minimalistic spirit of the original scenario. In particular, we will address a Palatini formulation of Higgs inflation and the embedding of the model to a fully scale invariant framework.

## 4.1. Palatini Higgs Inflation

In the usual formulation of Higgs inflation presented in section 2.2, the action is minimized with respect to the metric. This procedure implicitly assumes the existence of a Levi-Civita connection depending on the metric tensor and the inclusion of a York-Hawking-Gibbons term ensuring the cancellation of a total derivative term with no-vanishing variation at the boundary (York, 1972; Gibbons and Hawking, 1977). One could alternatively consider a Palatini formulation of gravity in which the metric tensor and the connection are treated as independent variables and no additional boundary term is required to obtain the equations of motion (Ferraris et al., 1982). Roughly speaking, this formulation corresponds to assuming an ultraviolet completion involving different gravitational degrees of freedom.

Although the metric and Palatini formulations of General Relativity give rise to the same equations of motion (Ferraris et al., 1982), this is not true for scalar-tensor theories as Higgs inflation. To see this explicitly let us consider the Higgs inflation action in Equation (2.12) with  $R = g^{\mu\nu} R_{\mu\nu}(\Gamma, \partial\Gamma)$  and  $\Gamma$  a non-Levi-Civita connection<sup>17</sup>. Performing a Weyl rescaling of the metric  $g_{\mu\nu} \rightarrow \Theta g_{\mu\nu}$  with  $\Theta$  given by Equation (2.15) we obtain an Einstein-frame action

$$S = \int d^4x \sqrt{-g} \left[ \frac{M_P^2}{2} g^{\mu\nu} R_{\mu\nu}(\Gamma) - \frac{1}{2} M_P^2 K(\Theta) (\partial\Theta)^2 - V(\Theta) \right], \quad (4.1)$$

containing a potential (2.17) and a non-canonical kinetic term with

$$K(\Theta) \equiv \frac{1}{4|a|\Theta^2} \left( \frac{1}{1-\Theta} \right), \quad (4.2)$$

and

$$|a| \equiv \xi. \quad (4.3)$$

Note that the kinetic function (4.2) differs from that obtained in the metric formulation, see Equation (2.18). In particular, it does not contain the part associated with the non-homogeneous transformation of the Ricci scalar, since  $R = R(\Gamma)$  is now invariant under Weyl rescalings. For the purposes of inflation, this translates into a modification of the residue of the inflationary pole at  $\Theta = 0$  with respect to the metric case. While the metric value of  $|a|$  in Equation (2.18) is bounded from above [cf. Equation (2.8)], it can take positive arbitrary values in the Palatini formulation [cf. Equation (4.3)]. Performing a field redefinition

$$\frac{1}{M_P^2} \left( \frac{d\phi}{d\Theta} \right)^2 = K(\Theta) \quad \longrightarrow \quad h(\phi) = F_\infty \sinh \left( \frac{\sqrt{a}\phi}{M_P} \right), \quad (4.4)$$

<sup>17</sup>For a recent generalization built from the Higgs, the metric and the connection and involving only up to two derivatives (see Rasanen, 2018).

to canonically normalize the  $h$ -field kinetic term, we can rewrite the graviscalar action (4.1) at  $\phi \gg v_{EW}$  as

$$S = \int d^4x \sqrt{-g} \left[ \frac{M_P^2}{2} R - \frac{1}{2} (\partial\phi)^2 - V(\phi) \right], \quad (4.5)$$

with

$$V(\phi) = \frac{\lambda}{4} F^4(\phi), \quad F(\phi) \equiv F_\infty \tanh \left( \frac{\sqrt{a}\phi}{M_P} \right). \quad (4.6)$$

The comparison of the latest expression with Equation (2.24) reveals some important differences between the metric and Palatini formulations. In both cases, the effective Einstein-frame potential smoothly interpolates between a low-energy quartic potential and an asymptotically flat potential at large field values. Note, however, that the transition in the Palatini case is rather direct and does not involve the quadratic piece appearing in the metric formulation. On top of that, the flatness of the asymptotic plateau is different in the two cases, due to the effective change in  $|a|$ . The Palatini dependence  $|a| = \xi$  has a strong impact on the inflationary observables. In the large-field regime they read

$$n_s \simeq 1 - \frac{2}{\bar{N}_*}, \quad r \simeq \frac{2}{\xi \bar{N}_*^2}, \quad (4.7)$$

with

$$\bar{N}_* \equiv N_* + \frac{1}{16|a|} \cosh \left( \frac{2\sqrt{a}\phi_E}{M_P} \right), \quad (4.8)$$

a rescaled number of e-folds and

$$\phi_E = \frac{M_P}{2\sqrt{a}} \operatorname{arcsinh}(\sqrt{32a}), \quad (4.9)$$

the inflaton value at the end of inflation ( $\epsilon(\phi_{\text{end}}) \equiv 1$ ), with  $\phi_E = \sqrt{3/2} \operatorname{arcsinh}(4/\sqrt{3}) M_P$  corresponding to the  $\xi \rightarrow \infty$  limit and  $\phi_E = 2\sqrt{2} M_P$  to the end of inflation in a minimally coupled  $\lambda\phi^4$  theory. A relation between the non-minimal coupling  $\xi$ , the self-coupling  $\lambda$  and the number of e-folds  $\bar{N}_*$  can be obtained by taking into account the amplitude of the observed power spectrum in Equation (2.39),

$$\xi \simeq 3.8 \times 10^6 \bar{N}_*^2 \lambda. \quad (4.10)$$

A simple inspection of Equation (4.7) reveals that the predicted tensor-to-scalar ratio in Palatini Higgs inflation is within the reach of current or future experiments (Matsumura et al., 2016) only if  $\xi \lesssim 10$ , which, assuming  $\bar{N} \simeq 59$ , requires a very small coupling  $\lambda \lesssim 10^{-9}$ . For a discussion of unitarity violations in the Palatini formulation, see Bauer and Demir (2011).

## 4.2. Higgs-Dilaton Model

The existence of robust predictions in (non-critical) Higgs inflation is intimately related to the emerging dilatation symmetry of its tree-level action at large field values. The uplifting of Higgs inflation to a completely scale-invariant

setting was considered in Shaposhnikov and Zenhausern (2009), Garcia-Bellido et al. (2011), Blas et al. (2011), Bezrukov et al. (2013), Garcia-Bellido et al. (2012), Rubio and Shaposhnikov (2014), Trashorras et al. (2016), Karananas and Rubio (2016), Casas et al. (2017), and Casas et al. (2018). In the unitary gauge  $H = (0, h/\sqrt{2})^T$ , the graviscalar sector of the Higgs-Dilaton model considered in these papers takes the form

$$S = \int d^4x \sqrt{-g} \left[ \frac{\xi_h h^2 + \xi_\chi \chi^2}{2} R - \frac{1}{2} (\partial h)^2 - \frac{1}{2} (\partial \chi)^2 - V(h, \chi) \right], \quad (4.11)$$

with

$$U(h, \chi) = \frac{\lambda}{4} (h^2 - \alpha \chi^2)^2 + \beta \chi^4 \quad (4.12)$$

a scale-invariant version of the SM symmetry-breaking potential and  $\alpha, \beta$  positive dimensionless parameters. The existence of a well-defined gravitational interactions at all field values requires the non-minimal gravitational couplings to be positive-definite, i.e.  $\xi_h, \xi_\chi > 0$ . In the absence of gravity, the ground state of Equation (4.11) is determined by the scale-invariant potential (4.12). For  $\alpha \neq 0$  and  $\beta = 0$ , the vacuum manifold extends along the flat directions  $h_0 = \pm \alpha \chi_0$ . Any solution with  $\chi_0 \neq 0$  breaks scale symmetry spontaneously and induces non-zero values for the effective Planck mass and the electroweak scale<sup>18</sup>. The relation between these highly hierarchical scales is set by fine-tuning  $\alpha \sim v^2/M_P^2 \sim 10^{-32}$ . For this small value, the flat valleys in the potential  $U(h, \chi)$  are essentially aligned and we can safely approximate  $\alpha \simeq 0$  for all inflationary purposes.

To compare the inflationary predictions of this model with those of the standard Higgs-inflation scenario, let us perform a Weyl rescaling  $g_{\mu\nu} \rightarrow M_P^2/(\xi_h h^2 + \xi_\chi \chi^2) g_{\mu\nu}$  followed by a field redefinition (Casas et al., 2017)

$$\gamma^{-2} \Theta \equiv \frac{(1 + 6\xi_h)h^2 + (1 + 6\xi_\chi)\chi^2}{\xi_h h^2 + \xi_\chi \chi^2},$$

$$\exp \left[ \frac{2\gamma\Phi}{M_P} \right] \equiv \frac{a}{\bar{a}} \frac{(1 + 6\xi_h)h^2 + (1 + 6\xi_\chi)\chi^2}{M_P^2}, \quad (4.13)$$

with

$$\gamma \equiv \sqrt{\frac{\xi_\chi}{1 + 6\xi_\chi}}, \quad a \equiv -\frac{\xi_h}{1 + 6\xi_h}, \quad \bar{a} \equiv a \left( 1 - \frac{\xi_\chi}{\xi_h} \right). \quad (4.14)$$

After some algebra, we obtain a rather simple Einstein-frame action (Karananas and Rubio, 2016; Casas et al., 2017)

$$S = \int d^4x \sqrt{-g} \left[ \frac{M_P^2}{2} R - \frac{1}{2} M_P^2 K(\Theta) (\partial \Theta)^2 - \frac{1}{2} \Theta (\partial \Phi)^2 - U(\Theta) \right], \quad (4.15)$$

<sup>18</sup> Among the possible values of  $\beta$  in the presence of gravity, the case  $\beta = 0$  seems also preferred (Allen and Folacci, 1987; Garcia-Bellido et al., 2011; Jalmuzna et al., 2011, see also Antoniadis et al., 1986, 2007; Tsamis and Woodard, 1993, 1995; Polyakov, 2010; Wetterich, 2017).

containing a potential

$$U(\Theta) = U_0(1 - \Theta)^2, \quad U_0 \equiv \frac{\lambda M_P^4}{4} \left( \frac{1 + 6\bar{a}}{\bar{a}} \right)^2, \quad (4.16)$$

and a non-canonical, albeit diagonal, kinetic sector. The kinetic function for the  $\Theta$  field,

$$K(\Theta) = \frac{1}{4|\bar{a}|\Theta^2} \left( \frac{c}{|\bar{a}|\Theta - c} + \frac{1 - 6|\bar{a}|\Theta}{1 - \Theta} \right), \quad (4.17)$$

contains two “inflationary” poles at  $\Theta = 0$  and  $\Theta = c/|\bar{a}|$  and a “Minkowski” pole at  $\Theta = 1$ , where the usual SM action is approximately recovered. As in the single field case, the “Minkowski” pole does not play a significant role during inflation and can be neglected for all practical purposes. Interestingly, the field-derivative space becomes in this limit a maximally symmetric hyperbolic manifold with Gaussian curvature  $a < 0$  (Karananas and Rubio, 2016).

Inflation takes place in the vicinity of the inflationary poles. During this regime, the kinetic term of the  $\Phi$ -field is effectively suppressed and the dilaton rapidly approaches a constant value  $\Phi = \Phi_0$  (Garcia-Bellido et al., 2011). This effective freezing is an immediate consequence of scale invariance. As in the single field case, the shift symmetry  $\Phi \rightarrow \Phi + C$  in Equation (4.15) allows us to interpret  $\Phi$  as the dilaton or Goldstone of dilatations. As first shown in Garcia-Bellido et al. (2011), the equation of motion for this field coincides with the scale-current conservation equation, effectively restricting the evolution to constant  $\Phi$  ellipsoidal trajectories in the  $\{h, \chi\}$  plane. Given this emergent *single-field* dynamics, no non-gaussianities nor isocurvature perturbations are significantly generated during inflation (Garcia-Bellido et al., 2011). If the  $\Theta$  variable is dominated by the Higgs component ( $\xi_h \gg \xi_\chi$ ), the spectral tilt and the tensor-to-scalar ratio take the compact form

$$n_s \simeq 1 - 8 c \coth(4cN_*) , \quad r \simeq \frac{32 c^2}{|a|} \text{csch}^2(4cN_*) , \quad (4.18)$$

with  $|a| \simeq 1/6$  in order to satisfy the normalization condition (2.39). Note that these expressions rapidly converge to the Higgs inflation values (2.38) for  $4cN_* \ll 1$ . For increasing  $c$  and fixed  $N_*$ , the spectral tilt decreases linearly and the tensor-to-scalar ratio approaches zero.

## 5. CONCLUDING REMARKS

Before the start of the LHC, it was widely believed that we would find a plethora of new particles and interactions that would reduce the Standard Model to a mere description of Nature at energies below the TeV scale. From a bottom-up perspective, new physics was typically advocated to cure the divergences associated with the potential growth of the Higgs self-coupling at high energies. The finding of a relatively light Higgs boson in the Large Hadron Collider concluded the quest

of the Standard Model spectrum while demystifying the concept of naturalness and the role of fundamental scalar fields in particle physics and cosmology. The Standard Model is now a confirmed theory that could stay valid till the Planck scale and provide a solid theoretical basis for describing the early Universe.

The Higgs field itself could lead to inflation if a minimalistic coupling to the Ricci scalar is added to the Standard Model action. The value of this coupling can be fixed by the normalization of the spectrum of primordial density perturbations, leaving a theory with no free parameters at tree level. On top of that, the experimental knowledge of the Standard Model couplings reduces the usual uncertainties associated with the heating stage and allows us to obtain precise predictions in excellent agreement with observations. Note, however, the mere existence of gravity makes the theory non-renormalizable and forces its interpretation as an effective field theory. Even in a self-consistent approach to Higgs inflation, the finite parts of the counterterms needed to make the theory finite obscure the connection between low- and high-energy observables. If these unknown coefficients are small, Higgs inflation provides an

appealing relation between the Standard Model parameters and the properties of the Universe at large scales. If they are large, this connection is lost but Higgs inflation can surprisingly take place even when the Standard Model vacuum is not completely stable.

## AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

## ACKNOWLEDGMENTS

The author acknowledges support from the Deutsche Forschungsgemeinschaft through the Open Access Publishing funding programme of the Baden-Württemberg Ministry of Science, Research and Arts and the Ruprecht-Karls-Universität Heidelberg as well as through the project TRR33 The Dark Universe. He thanks Guillem Domenech, Georgios Karananas and Martin Pauly for useful comments and suggestions on the manuscript.

## REFERENCES

- Aad, G., Abajyan, T., Abbott, B., Abdallah, J., Abdel Khalek, S., Abdelalim, A. A., et al. (2012). Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Phys. Lett. B* 716, 1–29. doi: 10.1016/j.physletb.2012.08.020
- Akrami, Y., Arroja, F., Ashdown, M., Aumont, J., Baccigalupi, C., Ballardini, M., et al. (2018). Planck 2018 results. X. Constraints on inflation. arXiv:1807.06211.
- Albrecht, A., and Steinhardt, P. J. (1982). Cosmology for grand unified theories with radiatively induced symmetry breaking. *Phys. Rev. Lett.* 48, 1220–1223. doi: 10.1103/PhysRevLett.48.1220
- Allen, B., and Folacci, A. (1987). The massless minimally coupled scalar field in De Sitter space. *Phys. Rev. D* 35:3771. doi: 10.1103/PhysRevD.35.3771
- Allison, K. (2014). Higgs xi-inflation for the 125–126 GeV Higgs: a two-loop analysis. *J. High Ener. Phys.* 02:040. doi: 10.1007/JHEP02(2014)040
- Anderson, G. W. (1990). New cosmological constraints on the Higgs Boson and top quark Masses. *Phys. Lett. B* 243, 265–270. doi: 10.1016/0370-2693(90)90849-2
- Antoniadis, I., Iliopoulos, J., and Tomaras, T. N. (1986). Quantum instability of De Sitter space. *Phys. Rev. Lett.* 56:1319. doi: 10.1103/PhysRevLett.56.1319
- Antoniadis, I., Mazur, P. O., and Mottola, E. (2007). Cosmological dark energy: prospects for a dynamical theory. *New J. Phys.* 9:11. doi: 10.1088/1367-2630/9/1/011
- Arai, M., Kawai, S., and Okada, N. (2011). Higgs inflation in minimal supersymmetric SU(5) GUT. *Phys. Rev. D* 84:123515. doi: 10.1103/PhysRevD.84.123515
- Arnold, P. B., and Vokos, S. (1991). Instability of hot electroweak theory: bounds on  $m(H)$  and  $M(t)$ . *Phys. Rev. D* 44, 3620–3627. doi: 10.1103/PhysRevD.44.3620
- Artymowski, M., and Rubio, J. (2016). Endlessly flat scalar potentials and  $\alpha$ -attractors. *Phys. Lett. B* 761, 111–114. doi: 10.1016/j.physletb.2016.08.024
- Atkins, M., and Calmet, X. (2011). Remarks on higgs inflation. *Phys. Lett. B* 697, 37–40. doi: 10.1016/j.physletb.2011.01.028
- Aydemir, U., Anber, M. M., and Donoghue, J. F. (2012). Self-healing of unitarity in effective field theories and the onset of new physics. *Phys. Rev. D* 86:014025. doi: 10.1103/PhysRevD.86.014025
- Barbon, J. L. F., Casas, J. A., Elias-Miro, J., and Espinosa, J. R. (2015). Higgs inflation as a mirage. *J. High Ener. Phys.* 09:027. doi: 10.1007/JHEP09(2015)027
- Barbon, J. L. F., and Espinosa, J. R. (2009). On the naturalness of higgs inflation. *Phys. Rev. D* 79:081302. doi: 10.1103/PhysRevD.79.081302
- Barenboim, G., Park, W.-I., and Kinney, W. H. (2016). Eternal hilltop inflation. *J. Cosmol. Astropart. Phys.* 1605:030. doi: 10.1088/1475-7516/2016/05/030
- Barvinsky, A. O., Kamenshchik, A. Yu., Kiefer, C., Starobinsky, A. A., and Steinwachs, C. (2009). Asymptotic freedom in inflationary cosmology with a non-minimally coupled Higgs field. *J. Cosmol. Astropart. Phys.* 0912:003. doi: 10.1088/1475-7516/2009/12/003
- Barvinsky, A. O., Kamenshchik, A. Yu., Kiefer, C., Starobinsky, A. A., and Steinwachs, C. F. (2012). Higgs boson, renormalization group, and naturalness in cosmology. *Eur. Phys. J. C* 72:2219. doi: 10.1140/epjc/s10052-012-2219-3
- Barvinsky, A. O., Kamenshchik, A. Yu., and Starobinsky, A. A. (2008). Inflation scenario via the Standard Model Higgs boson and LHC. *J. Cosmol. Astropart. Phys.* 0811:021. doi: 10.1088/1475-7516/2008/11/021
- Bauer, F., and Demir, D. A. (2011). Higgs-palatini inflation and unitarity. *Phys. Lett. B* 698, 425–429. doi: 10.1016/j.physletb.2011.03.042
- Ben-Dayan, I., and Einhorn, M. B. (2010). Supergravity higgs inflation and shift symmetry in electroweak theory. *J. Cosmol. Astropart. Phys.* 1012:002. doi: 10.1088/1475-7516/2010/12/002
- Bezrukov, F. (2013). The Higgs field as an inflaton. *Class. Quant. Grav.* 30:214001. doi: 10.1088/0264-9381/30/21/214001
- Bezrukov, F., Gorbunov, D., and Shaposhnikov, M. (2009a). On initial conditions for the Hot Big Bang. *J. Cosmol. Astropart. Phys.* 0906:029. doi: 10.1088/1475-7516/2009/06/029
- Bezrukov, F., Gorbunov, D., and Shaposhnikov, M. (2011a). Late and early time phenomenology of Higgs-dependent cutoff. *J. Cosmol. Astropart. Phys.* 1110:001. doi: 10.1088/1475-7516/2011/10/001
- Bezrukov, F., Kalmykov, M. Yu., Kniehl, B. A., and Shaposhnikov, M. (2012). Higgs boson mass and new physics. *J. High Ener. Phys.* 10:140. doi: 10.1007/JHEP10(2012)140
- Bezrukov, F., Karananas, G. K., Rubio, J., and Shaposhnikov, M. (2013). Higgs-Dilaton Cosmology: an effective field theory approach. *Phys. Rev. D* 87:096001. doi: 10.1103/PhysRevD.87.096001
- Bezrukov, F., Magnin, A., Shaposhnikov, M., and Sibiryakov, S. (2011b). Higgs inflation: consistency and generalisations. *J. High Ener. Phys.* 01:016. doi: 10.1007/JHEP01(2011)016
- Bezrukov, F., Pauly, M., and Rubio, J. (2018). On the robustness of the primordial power spectrum in renormalized Higgs inflation. *J. Cosmol. Astropart. Phys.* 1802:40. doi: 10.1088/1475-7516/2018/02/040

- Bezrukov, F., Rubio, J., and Shaposhnikov, M. (2015). Living beyond the edge: higgs inflation and vacuum metastability. *Phys. Rev. D* 92:83512. doi: 10.1103/PhysRevD.92.083512
- Bezrukov, F., and Shaposhnikov, M. (2009). Standard Model Higgs boson mass from inflation: Two loop analysis. *J. High Ener. Phys.* 7:89. doi: 10.1088/1126-6708/2009/07/089
- Bezrukov, F. and Shaposhnikov, M. (2014). Higgs inflation at the critical point. *Phys. Lett. B* 734, 249–254. doi: 10.1016/j.physletb.2014.05.074
- Bezrukov, F., and Shaposhnikov, M. (2015a). Inflation, LHC and the Higgs boson. *Comptes Rendus Physique* 16, 994–1002. doi: 10.1016/j.crhy.2015.08.005
- Bezrukov, F., and Shaposhnikov, M. (2015b). Why should we care about the top quark Yukawa coupling? *J. Exp. Theor. Phys.* 120, 335–343. doi: 10.1134/S1063776115030152
- Bezrukov, F. L., and Gorbunov, D. S. (2012). Distinguishing between  $R^2$ -inflation and Higgs-inflation. *Phys. Lett. B* 713, 365–368. doi: 10.1016/j.physletb.2012.06.040
- Bezrukov, F. L., Magnin, A., and Shaposhnikov, M. (2009b). Standard model higgs boson mass from inflation. *Phys. Lett. B* 675, 88–92. doi: 10.1016/j.physletb.2009.03.035
- Bezrukov, F. L., and Shaposhnikov, M. (2008). The standard model higgs boson as the inflaton. *Phys. Lett. B* 659, 703–706. doi: 10.1016/j.physletb.2007.11.072
- Bird, S., Cholis, I., Munoz, J. B., Ali-Haïmoud, Y., Kamionkowski, M., Kovetz, E. D., et al. (2016). Did LIGO detect dark matter? *Phys. Rev. Lett.* 116:201301. doi: 10.1103/PhysRevLett.116.201301
- Birrell, N. D., and Davies, P. C. W. (1984). *Quantum Fields in Curved Space*. Cambridge Monographs on Mathematical Physics. Cambridge: Cambridge University Press.
- Blas, D., Shaposhnikov, M., and Zenhausern, D. (2011). Scale-invariant alternatives to general relativity. *Phys. Rev. D* 84:044001. doi: 10.1103/PhysRevD.84.044001
- Boubekeur, L. and Lyth, D. H. (2005). Hilltop inflation. *J. Cosmol. Astropart. Phys.* 0507:010. doi: 10.1088/1475-7516/2005/07/010
- Branchina, V., and Messina, E. (2013). Stability, higgs boson mass and new physics. *Phys. Rev. Lett.* 111:241801. doi: 10.1103/PhysRevLett.111.241801
- Branchina, V., Messina, E., and Platania, A. (2014). Top mass determination, Higgs inflation, and vacuum stability. *J. High Ener. Phys.* 09:182. doi: 10.1007/JHEP09(2014)182
- Branchina, V., Messina, E., and Sher, M. (2015). Lifetime of the electroweak vacuum and sensitivity to Planck scale physics. *Phys. Rev. D* 91:013003. doi: 10.1103/PhysRevD.91.013003
- Burgess, C. P., Lee, H. M., and Trott, M. (2009). Power-counting and the validity of the classical approximation during inflation. *J. High Ener. Phys.* 09:103. doi: 10.1088/1126-6708/2009/09/103
- Burgess, C. P., Lee, H. M., and Trott, M. (2010). Comment on higgs inflation and naturalness. *J. High Ener. Phys.* 07:007. doi: 10.1007/JHEP07(2010)007
- Burgess, C. P., Patil, S. P., and Trott, M. (2014). On the predictiveness of single-field inflationary models. *J. High Ener. Phys.* 06:010. doi: 10.1007/JHEP06(2014)010
- Burns, D., Karamitsos, S., and Pilaftsis, A. (2016). Frame-covariant formulation of inflation in scalar-curvature theories. *Nucl. Phys. B* 907, 785–819. doi: 10.1016/j.nuclphysb.2016.04.036
- Butenschoen, M., Dehnadi, B., Hoang, A. H., Mateu, V., Preisser, M., and Stewart, I. W. (2016). Top quark mass Calibration for Monte Carlo Event Generators. *Phys. Rev. Lett.* 117:232001. doi: 10.1103/PhysRevLett.117.232001
- Buttazzo, D., Degrandi, G., Giardino, P. P., Giudice, G. F., Sala, F., Salvio, A., et al. (2013). Investigating the near-criticality of the Higgs boson. *J. High Ener. Phys.* 12:089. doi: 10.1007/JHEP12(2013)089
- Cai, R.-G., Guo, Z.-K., and Wang, S.-J. (2015). Higgs inflation in Gauss-Bonnet braneworld. *Phys. Rev. D* 92:063514. doi: 10.1103/PhysRevD.92.063514
- Callan, C. G. Jr., Coleman, S. R., and Jackiw, R. (1970). A New improved energy - momentum tensor. *Ann. Phys.* 59, 42–73. doi: 10.1016/0003-4916(70)90394-5
- Calmet, X., and Casadio, R. (2014). Self-healing of unitarity in Higgs inflation. *Phys. Lett. B* 734, 17–20. doi: 10.1016/j.physletb.2014.05.008
- Calmet, X., and Kuntz, I. (2016). Higgs Starobinsky Inflation. *Eur. Phys. J. C* 76:289. doi: 10.1140/epjc/s10052-016-4136-3
- Carr, B., Kuhnel, F., and Sandstad, M. (2016). Primordial black holes as dark matter. *Phys. Rev. D* 94:083504. doi: 10.1103/PhysRevD.94.083504
- Carr, B., Raidal, M., Tenkanen, T., Vaskonen, V., and Veermae, H. (2017). Primordial black hole constraints for extended mass functions. *Phys. Rev. D* 96:023514. doi: 10.1103/PhysRevD.96.023514
- Casas, S., Karananas, G. K., Pauly, M., and Rubio, J. (2018). Scale-invariant alternatives to general relativity. III. The inflation–dark-energy connection, arXiv:1811.05984.
- Casas, S., Pauly, M., and Rubio, J. (2017). Higgs-dilaton cosmology: an inflation–dark-energy connection and forecasts for future galaxy surveys. *Phys. Rev. D* 97:043520. doi: 10.1103/PhysRevD.97.043520
- Cervantes-Cota, J. L., and Dehnen, H. (1995a). Induced gravity inflation in the standard model of particle physics. *Nucl. Phys. B* 442, 391–412. doi: 10.1016/0550-3213(95)00128-X
- Cervantes-Cota, J. L., and Dehnen, H. (1995b). Induced gravity inflation in the SU(5) GUT. *Phys. Rev. D* 51, 395–404. doi: 10.1103/PhysRevD.51.395
- Chatrchyan, S., Khachatryan, V., Sirunyan, A. M., Tumasyan, A., Adam, W., Aguilo, E., et al. (2012). Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Phys. Lett. B* 716, 30–61. doi: 10.1016/j.physletb.2012.08.021
- Chatzigeorgiou, I. (2016). Bounds on the lambert function and their application to the outage analysis of user cooperation. *CoRR*, abs/1601.04895.
- Chen, H.-Y., Gogoladze, I., Hu, S., Li, T., and Wu, L. (2018). Natural Higgs Inflation, Gauge Coupling Unification, and Neutrino Masses, arXiv:1805.00161.
- Chiba, T., and Yamaguchi, M. (2008). Extended slow-roll conditions and rapid-roll conditions. *J. Cosmol. Astropart. Phys.* 0810:021. doi: 10.1088/1475-7516/2008/10/021
- Chiba, T., and Yamaguchi, M. (2013). Conformal-frame (in)dependence of cosmological observations in scalar-tensor theory. *J. Cosmol. Astropart. Phys.* 1310:040. doi: 10.1088/1475-7516/2013/10/040
- Choudhury, S., Chakraborty, T., and Pal, S. (2014). Higgs inflation from new Kähler potential. *Nucl. Phys. B* 880, 155–174. doi: 10.1016/j.nuclphysb.2014.01.002
- Collaboration, C. (2014). *Combination of the CMS Top-quark Mass Measurements From Run 1 of the LHC, Report number CMS-PAS-TOP-14-015*.
- Cook, J. L., Krauss, L. M., Long, A. J., and Sabharwal, S. (2014). Is Higgs inflation ruled out? *Phys. Rev. D* 89:103525. doi: 10.1103/PhysRevD.89.103525
- Csaki, C., Kaloper, N., Serra, J., and Terning, J. (2014). Inflation from broken scale invariance. *Phys. Rev. Lett.* 113:161302. doi: 10.1103/PhysRevLett.113.161302
- De Simone, A., Hertzberg, M. P., and Wilczek, F. (2009). Running inflation in the standard model. *Phys. Lett. B* 678, 1–8. doi: 10.1016/j.physletb.2009.05.054
- Degrassi, G., Di Vita, S., Elias-Miro, J., Espinosa, J. R., Giudice, G. F., Isidori, G., et al. (2012). Higgs mass and vacuum stability in the Standard Model at NNLO. *J. High Ener. Phys.* 08:098. doi: 10.1007/JHEP08(2012)098
- Dutta, S., Hagiwara, K., Yan, Q.-S., and Yoshida, K. (2008). Constraints on the electroweak chiral Lagrangian from the precision data. *Nucl. Phys. B* 790, 111–137. doi: 10.1016/j.nuclphysb.2007.08.017
- Einhorn, M. B., and Jones, D. R. T. (2012). GUT scalar potentials for higgs inflation. *J. Cosmol. Astropart. Phys.* 1211:049. doi: 10.1088/1475-7516/2012/11/049
- Ellis, J., He, H.-J., and Xianyu, Z.-Z. (2015). New Higgs Inflation in a No-Scale Supersymmetric SU(5) GUT. *Phys. Rev. D* 91:021302. doi: 10.1103/PhysRevD.91.021302
- Ellis, J., He, H.-J., and Xianyu, Z.-Z. (2016). Higgs inflation, reheating and gravitino production in no-scale supersymmetric GUTs. *J. Cosmol. Astropart. Phys.* 1608:068. doi: 10.1088/1475-7516/2016/08/068
- Ema, Y. (2017). Higgs scalaron mixed inflation. *Phys. Lett. B* 770, 403–411. doi: 10.1016/j.physletb.2017.04.060
- Ema, Y., Jinno, R., Mukaida, K., and Nakayama, K. (2017a). Violent preheating in inflation with nonminimal coupling. *J. Cosmol. Astropart. Phys.* 1702:045. doi: 10.1088/1475-7516/2017/02/045
- Ema, Y., Karciauskas, M., Lebedev, O., Rusak, S., and Zatta, M. (2017b). Higgs-Inflaton Mixing and Vacuum Stability, arXiv:1711.10554.
- Enckell, V.-M., Enqvist, K., and Nurmi, S. (2016). Observational signatures of Higgs inflation. *J. Cosmol. Astropart. Phys.* 1607:047. doi: 10.1088/1475-7516/2016/07/047
- Enckell, V.-M., Enqvist, K., Rasanen, S., and Tomberg, E. (2018). Higgs inflation at the hilltop. *J. Cosmol. Astropart. Phys.* 1806:005. doi: 10.1088/1475-7516/2018/06/005

- Escrivà, A., and Germani, C. (2017). Beyond dimensional analysis: Higgs and new Higgs inflations do not violate unitarity. *Phys. Rev. D* 95:123526. doi: 10.1103/PhysRevD.95.123526
- Espinosa, J. R. (2016). Implications of the top (and Higgs) mass for vacuum stability. *PoS TOP2015*:043. doi: 10.22323/1.257.0043
- Espinosa, J. R., Garny, M., Konstandin, T., and Riotto, A. (2017). Gauge-independent scales related to the standard Model Vacuum Instability. *Phys. Rev. D* 95:056004. doi: 10.1103/PhysRevD.95.056004
- Espinosa, J. R., Giudice, G. F., Morgante, E., Riotto, A., Senatore, L., Strumia, A., et al. (2015). The cosmological Higgstory of the vacuum instability. *J. High Ener. Phys.* 09:174. doi: 10.1007/JHEP09(2015)174
- Espinosa, J. R., Giudice, G. F., and Riotto, A. (2008). Cosmological implications of the Higgs mass measurement. *J. Cosmol. Astropart. Phys.* 0805:2. doi: 10.1088/1475-7516/2008/05/002
- Espinosa, J. R., and Quiros, M. (1995). Improved metastability bounds on the standard model Higgs mass. *Phys. Lett. B* 353, 257–266. doi: 10.1016/0370-2693(95)00572-3
- Ezquiaga, J. M., and Garcia-Bellido, J. (2018). Quantum diffusion beyond slow-roll: implications for primordial black-hole production. *J. Cosmol. Astropart. Phys.* 1808:018. doi: 10.1088/1475-7516/2018/08/018
- Ezquiaga, J. M., Garcia-Bellido, J., and Ruiz Morales, E. (2018). Primordial black hole production in critical higgs inflation. *Phys. Lett. B* 776, 345–349. doi: 10.1016/j.physletb.2017.11.039
- Fairbairn, M., Grothaus, P., and Hogan, R. (2014). The problem with false vacuum higgs inflation. *J. Cosmol. Astropart. Phys.* 1406:039. doi: 10.1088/1475-7516/2014/06/039
- Fakir, R., Habib, S., and Unruh, W. (1992). Cosmological density perturbations with modified gravity. *Astrophys. J.* 394:396. doi: 10.1086/171591
- Fakir, R., and Unruh, W. G. (1990a). Improvement on cosmological chaotic inflation through nonminimal coupling. *Phys. Rev. D* 41, 1783–1791. doi: 10.1103/PhysRevD.41.1783
- Fakir, R., and Unruh, W. G. (1990b). Induced gravity inflation. *Phys. Rev. D* 41, 1792–1795. doi: 10.1103/PhysRevD.41.1792
- Felder, G. N., Frolov, A. V., Kofman, L., and Linde, A. D. (2002). Cosmology with negative potentials. *Phys. Rev. D* 66:023507. doi: 10.1103/PhysRevD.66.023507
- Ferrara, S., Kallosh, R., Linde, A., Marrani, A., and Van Proeyen, A. (2011). Superconformal symmetry, NMSSM, and inflation. *Phys. Rev. D* 83:025008. doi: 10.1103/PhysRevD.83.025008
- Ferrara, S., Kallosh, R., Linde, A., and Porrati, M. (2013). Minimal supergravity models of inflation. *Phys. Rev. D* 88:085038. doi: 10.1103/PhysRevD.88.085038
- Ferraris, M., Francaviglia, M., and Reina, C. (1982). Variational formulation of general relativity from 1915 to 1925 Palatini's method discovered by Einstein in 1925. *Gen. Relat. Gravit.* 14, 243–254. doi: 10.1007/BF00756060
- Figueroa, D. G., Rajantie, A., and Torrenti, F. (2018). Higgs field-curvature coupling and postinflationary vacuum instability. *Phys. Rev. D* 98:023532. doi: 10.1103/PhysRevD.98.023532
- Flanagan, E. E. (2004). The Conformal frame freedom in theories of gravitation. *Class. Quant. Grav.* 21:3817. doi: 10.1088/0264-9381/21/15/N02
- Fumagalli, J. (2017). Renormalization group independence of cosmological attractors. *Phys. Lett. B* 769, 451–459. doi: 10.1016/j.physletb.2017.04.017
- Fumagalli, J., Mooij, S., and Postma, M. (2018). Unitarity and predictiveness in new Higgs inflation. *J. High Ener. Phys.* 03:038. doi: 10.1007/JHEP03(2018)038
- Fumagalli, J., and Postma, M. (2016). UV (in)sensitivity of higgs inflation. *J. High Ener. Phys.* 05:049. doi: 10.1007/JHEP05(2016)049
- Futamase, T., and Maeda, K.-i. (1989). Chaotic inflationary scenario in models having nonminimal coupling with curvature. *Phys. Rev. D* 39, 399–404. doi: 10.1103/PhysRevD.39.399
- Galante, M., Kallosh, R., Linde, A., and Roest, D. (2015). Unity of cosmological inflation attractors. *Phys. Rev. Lett.* 114:141302. doi: 10.1103/PhysRevLett.114.141302
- Garcia-Bellido, J., Figueroa, D. G., and Rubio, J. (2009). Preheating in the standard model with the higgs-inflaton coupled to gravity. *Phys. Rev. D* 79:063531. doi: 10.1103/PhysRevD.79.063531
- Garcia-Bellido, J., Rubio, J., and Shaposhnikov, M. (2012). Higgs-dilaton cosmology: are there extra relativistic species? *Phys. Lett. B* 718, 507–511. doi: 10.1016/j.physletb.2012.10.075
- Garcia-Bellido, J., Rubio, J., Shaposhnikov, M., and Zenhausern, D. (2011). Higgs-dilaton cosmology: from the early to the late universe. *Phys. Rev. D* 84:123504. doi: 10.1103/PhysRevD.84.123504
- Garcia-Bellido, J., and Ruiz Morales, E. (2017). Primordial black holes from single field models of inflation. *Phys. Dark Univ.* 18, 47–54. doi: 10.1016/j.dark.2017.09.007
- Ge, S.-F., He, H.-J., Ren, J., and Xianyu, Z.-Z. (2016). Realizing dark matter and higgs inflation in light of LHC diphoton excess. *Phys. Lett. B* 757, 480–492. doi: 10.1016/j.physletb.2016.04.008
- George, D. P., Mooij, S., and Postma, M. (2014). Quantum corrections in Higgs inflation: the real scalar case. *J. Cosmol. Astropart. Phys.* 1402:024. doi: 10.1088/1475-7516/2014/02/024
- George, D. P., Mooij, S., and Postma, M. (2016). Quantum corrections in Higgs inflation: the standard model case. *J. Cosmol. Astropart. Phys.* 1604:006. doi: 10.1088/1475-7516/2016/04/006
- Germani, C., and Kehagias, A. (2010a). Cosmological perturbations in the new higgs inflation. *J. Cosmol. Astropart. Phys.* 1005:019. doi: 10.1088/1475-7516/2010/05/019
- Germani, C., and Kehagias, A. (2010b). New model of inflation with non-minimal derivative coupling of standard model higgs boson to gravity. *Phys. Rev. Lett.* 105:011302. doi: 10.1103/PhysRevLett.105.011302
- Germani, C., and Prokopec, T. (2017). On primordial black holes from an inflection point. *Phys. Dark Univ.* 18, 6–10. doi: 10.1016/j.dark.2017.09.001
- Gibbons, G. W., and Hawking, S. W. (1977). Action integrals and partition functions in quantum gravity. *Phys. Rev. D* 15, 2752–2756. doi: 10.1103/PhysRevD.15.2752
- Giudice, G. F., and Lee, H. M. (2011). Unitarizing higgs inflation. *Phys. Lett. B* 694, 294–300. doi: 10.1016/j.physletb.2010.10.035
- Gorunov, D., and Tokareva, A. (2013).  $R^2$ -inflation with conformal SM Higgs field. *J. Cosmol. Astropart. Phys.* 1312:021. doi: 10.1088/1475-7516/2013/12/021
- Gorunov, D. S., and Panin, A. G. (2011). Scalaron the mighty: producing dark matter and baryon asymmetry at reheating. *Phys. Lett. B* 700, 157–162. doi: 10.1016/j.physletb.2011.04.067
- Gorunov, D. S., and Panin, A. G. (2012). Free scalar dark matter candidates in  $R^2$ -inflation: the light, the heavy and the superheavy. *Phys. Lett. B* 718, 15–20. doi: 10.1016/j.physletb.2012.10.015
- Greenwood, R. N., Kaiser, D. I., and Sfakianakis, E. I. (2013). Multifield dynamics of higgs inflation. *Phys. Rev. D* 87:064021. doi: 10.1103/PhysRevD.87.064021
- Guth, A. H. (1981). The inflationary universe: a possible solution to the horizon and flatness problems. *Phys. Rev. D* 23, 347–356. doi: 10.1103/PhysRevD.23.347
- Hamada, Y., Kawai, H., and Oda, K.-Y. (2014a). Minimal Higgs inflation. *Prog. Theor. Exp. Phys.* 2014:023B02. doi: 10.1093/ptep/ptt116
- Hamada, Y., Kawai, H., Oda, K.-Y., and Park, S. C. (2014b). Higgs inflation is still alive after the results from BICEP2. *Phys. Rev. Lett.* 112:241301. doi: 10.1103/PhysRevLett.112.241301
- Hamada, Y., Kawai, H., Oda, K.-Y., and Park, S. C. (2015). Higgs inflation from Standard Model criticality. *Phys. Rev. D* 91:053008. doi: 10.1103/PhysRevD.91.053008
- Hawking, S. W. (1982). The development of irregularities in a single bubble inflationary universe. *Phys. Lett. B* 115B:295. doi: 10.1016/0370-2693(82)90373-2
- He, H.-J., and Xianyu, Z.-Z. (2014). Extending higgs inflation with TeV Scale new physics. *J. Cosmol. Astropart. Phys.* 1410:019. doi: 10.1088/1475-7516/2014/10/019
- He, M., Starobinsky, A. A., and Yokoyama, J. (2018). Inflation in the mixed Higgs- $R^2$  model. *J. Cosmol. Astropart. Phys.* 1805:064. doi: 10.1088/1475-7516/2018/05/064
- Herranen, M., Markkanen, T., Nurmi, S., and Rajantie, A. (2014). Spacetime curvature and the Higgs stability during inflation. *Phys. Rev. Lett.* 113:211102. doi: 10.1103/PhysRevLett.113.211102
- Hertzberg, M. P. (2010). On Inflation with Non-minimal Coupling. *J. High Ener. Phys.* 11:023. doi: 10.1007/JHEP11(2010)023
- Hertzberg, M. P. (2012). Can inflation be connected to low energy particle physics? *J. Cosmol. Astropart. Phys.* 1208:008. doi: 10.1088/1475-7516/2012/08/008
- Isidori, G., Rychkov, V. S., Strumia, A., and Tetradis, N. (2008). Gravitational corrections to standard model vacuum decay. *Phys. Rev. D* 77:025034. doi: 10.1103/PhysRevD.77.025034

- Jal muzna, J., Rostworowski, A., and Bizon, P. (2011). A comment on AdS collapse of a scalar field in higher dimensions. *Phys. Rev. D* 84:085021. doi: 10.1103/PhysRevD.84.085021
- Jarv, L., Kannike, K., Marzola, L., Racioppi, A., Raidal, M., Runkla, M., et al. (2017). Frame-independent classification of single-field inflationary models. *Phys. Rev. Lett.* 118:151302. doi: 10.1103/PhysRevLett.118.151302
- Jarv, L., Kuusk, P., Saal, M., and Vilson, O. (2015a). Invariant quantities in the scalar-tensor theories of gravitation. *Phys. Rev. D* 91:024041. doi: 10.1103/PhysRevD.91.024041
- Jarv, L., Kuusk, P., Saal, M., and Vilson, O. (2015b). Transformation properties and general relativity regime in scalar-tensor theories. *Class. Quant. Grav.* 32:235013. doi: 10.1088/0264-9381/32/23/235013
- Kaiser, D. I. (1995). Primordial spectral indices from generalized Einstein theories. *Phys. Rev. D* 52, 4295–4306. doi: 10.1103/PhysRevD.52.4295
- Kaiser, D. I. (2010). Conformal transformations with multiple scalar fields. *Phys. Rev. D* 81:084044. doi: 10.1103/PhysRevD.81.084044
- Kaiser, D. I., Mazenc, E. A., and Sfakianakis, E. I. (2013). Primordial bispectrum from multifield inflation with nonminimal couplings. *Phys. Rev. D* 87:064004. doi: 10.1103/PhysRevD.87.064004
- Kaiser, D. I., and Sfakianakis, E. I. (2014). Multifield inflation after Planck: the case for nonminimal couplings. *Phys. Rev. Lett.* 112:011302. doi: 10.1103/PhysRevLett.112.011302
- Kajantie, K., Laine, M., Rummukainen, K., and Shaposhnikov, M. E. (1996). Generic rules for high temperature dimensional reduction and their application to the standard model. *Nucl. Phys. B* 458, 90–136. doi: 10.1016/0550-3213(95)00549-8
- Kallosh, R., Linde, A., and Roest, D. (2013). Superconformal inflationary  $\alpha$ -Attractors. *J. High Ener. Phys.* 11:198. doi: 10.1007/JHEP11(2013)198
- Kamada, K. (2015a). Inflationary cosmology and the standard model Higgs with a small Hubble induced mass. *Phys. Lett. B* 742, 126–135. doi: 10.1016/j.physletb.2015.01.024
- Kamada, K. (2015b). On the strong coupling scale in Higgs G-inflation. *Phys. Lett. B* 744, 347–351. doi: 10.1016/j.physletb.2015.04.012
- Kamada, K., Kobayashi, T., Takahashi, T., Yamaguchi, M., and Yokoyama, J. (2012). Generalized higgs inflation. *Phys. Rev. D* 86:023504. doi: 10.1103/PhysRevD.86.023504
- Kamada, K., Kobayashi, T., Yamaguchi, M., and Yokoyama, J. (2011). Higgs G-inflation. *Phys. Rev. D* 83:083515. doi: 10.1103/PhysRevD.83.083515
- Kanemura, S., Matsui, T., and Nabeshima, T. (2013). Higgs inflation in a radiative seesaw model. *Phys. Lett. B* 723, 126–131. doi: 10.1016/j.physletb.2013.05.002
- Kannike, K., Marzola, L., Raidal, M., and Veermae, H. (2017). Single field double inflation and primordial black holes. *J. Cosmol. Astropart. Phys.* 1709:020. doi: 10.1088/1475-7516/2017/09/020
- Karam, A., Pappas, T., and Tamvakis, K. (2017). Frame-dependence of higher-order inflationary observables in scalar-tensor theories. *Phys. Rev. D* 96:064036. doi: 10.1103/PhysRevD.96.064036
- Karamitsos, S., and Pilaftsis, A. (2018a). Frame covariant nonminimal multifield inflation. *Nucl. Phys. B* 927, 219–254. doi: 10.1016/j.nuclphysb.2017.12.015
- Karamitsos, S., and Pilaftsis, A. (2018b). “On the cosmological Frame Problem,” in *17th Hellenic School and Workshops on Elementary Particle Physics and Gravity (CORFU2017) Corfu, Greece, September 2-28, 2017*. Available online at: <https://pos.sissa.it/318/036>
- Karananas, G. K., and Monin, A. (2016a). Weyl and Ricci gauging from the coset construction. *Phys. Rev. D* 93:064013. doi: 10.1103/PhysRevD.93.064013
- Karananas, G. K., and Monin, A. (2016b). Weyl vs. Conformal. *Phys. Lett. B* 757, 257–260. doi: 10.1016/j.physletb.2016.04.001
- Karananas, G. K., and Rubio, J. (2016). On the geometrical interpretation of scale-invariant models of inflation. *Phys. Lett. B* 761, 223–228. doi: 10.1016/j.physletb.2016.08.037
- Koh, S. (2006). Non-gaussianity in nonminimally coupled scalar field theory. *J. Korean Phys. Soc.* 49, S787–S790. Available online at: <http://inspirehep.net/record/693829>
- Komatsu, E., and Futamase, T. (1998). Constraints on the chaotic inflationary scenario with a nonminimally coupled ‘inflaton’ field from the cosmic microwave background radiation anisotropy. *Phys. Rev. D* 58:023004. doi: 10.1103/PhysRevD.58.023004
- Komatsu, E., and Futamase, T. (1999). Complete constraints on a nonminimally coupled chaotic inflationary scenario from the cosmic microwave background. *Phys. Rev. D* 59:064029. doi: 10.1103/PhysRevD.59.064029
- Kuusk, P., Jarv, L., and Vilson, O. (2016). Invariant quantities in the multiscalar-tensor theories of gravitation. *Int. J. Mod. Phys. A* 31:1641003. doi: 10.1142/S0217751X16410037
- Lazarides, G., and Pallis, C. (2015). Shift symmetry and higgs inflation in supergravity with observable gravitational waves. *J. High Ener. Phys.* 11:114. doi: 10.1007/JHEP11(2015)114
- Lerner, R. N., and McDonald, J. (2010). A unitarity-conserving higgs inflation model. *Phys. Rev. D* 82:103525. doi: 10.1103/PhysRevD.82.103525
- Lerner, R. N., and McDonald, J. (2011). Distinguishing higgs inflation and its variants. *Phys. Rev. D* 83:123522. doi: 10.1103/PhysRevD.83.123522
- Linde, A. D. (1979). Phase transitions in gauge theories and cosmology. *Rept. Prog. Phys.* 42:389. doi: 10.1088/0034-4885/42/3/001
- Linde, A. D. (1982). A new inflationary universe scenario: a possible solution of the horizon, flatness, homogeneity, isotropy and primordial monopole problems. *Phys. Lett. B* 108, 389–393. doi: 10.1016/0370-2693(82)91219-9
- Linde, A. D. (1983). Chaotic inflation. *Phys. Lett. B* 129, 177–181. doi: 10.1016/0370-2693(83)90837-7
- Makino, N., and Sasaki, M. (1991). The density perturbation in the chaotic inflation with nonminimal coupling. *Prog. Theor. Phys.* 86, 103–118. doi: 10.1143/ptp/86.1.103
- Marian, I. G., Defenu, N., Trombettoni, A., and Nandori, I. (2017). Pseudo Periodic Higgs Inflation, arXiv:1705.10276.
- Masina, I. (2018). Ruling out critical higgs inflation? *Phys. Rev. D* 98:043536. doi: 10.1103/PhysRevD.98.043536
- Matsumura, T., Akiba, Y., Arnold, K., Borrill, J., Chendra, R., Chinone, Y., et al. (2016). LiteBIRD: mission overview and focal plane layout. *J. Low. Temp. Phys.* 184, 824–831. doi: 10.1007/s10909-016-1542-8
- Minkowski, P. (1977). On the spontaneous origin of Newton’s constant. *Phys. Lett. B* 71, 419–421. doi: 10.1016/0370-2693(77)90256-8
- Moss, I. G. (2015). Higgs boson cosmology. *Contemp. Phys.* 56, 468–476. doi: 10.1080/00107514.2015.1058543
- Mukhanov, V. F. (1988). Quantum theory of gauge invariant cosmological perturbations. *Sov. Phys. JETP* 67, 1297–1302.
- Mukhanov, V. F., and Chibisov, G. V. (1981). Quantum fluctuations and a nonsingular universe. *JETP Lett.* 33, 532–535.
- Mukhanov, V. F., Feldman, H. A., and Brandenberger, R. H. (1992). Theory of cosmological perturbations. Part 1. Classical perturbations. Part 2. Quantum theory of perturbations. Part 3. Extensions. *Phys. Rept.* 215, 203–333. doi: 10.1016/0370-1573(92)90044-Z
- Oda, I., and Tomoyose, T. (2014a). Conformal Higgs inflation. *J. High Ener. Phys.* 09:165. doi: 10.1007/JHEP09(2014)165
- Oda, I., and Tomoyose, T. (2014b). Quadratic chaotic inflation from Higgs inflation. *Adv. Stud. Theor. Phys.* 8:551. doi: 10.12988/astp.2014.4572
- Okada, N., and Raut, D. (2017). Inflection-point Higgs inflation. *Phys. Rev. D* 95:035035. doi: 10.1103/PhysRevD.95.035035
- Okada, N., and Shafi, Q. (2015). Higgs inflation, seesaw physics and Fermion dark matter. *Phys. Lett. B* 747, 223–228. doi: 10.1016/j.physletb.2015.06.001
- Olive, K. A., Agashe, K., Amsler, C., Antonelli, M., Arguin, J.-F., Asner, D. M., et al. (2014). Review of particle physics. *Chin. Phys. C* 38:090001. doi: 10.1088/1674-1137/38/9/090001
- Pattison, C., Vennin, V., Assadullahi, H., and Wands, D. (2017). Quantum diffusion during inflation and primordial black holes. *J. Cosmol. Astropart. Phys.* 1710:046. doi: 10.1088/1475-7516/2017/10/046
- Polyakov, A. M. (2010). Decay of vacuum energy. *Nucl. Phys. B* 834, 316–329. doi: 10.1016/j.nuclphysb.2010.03.021
- Popa, L. A., and Caramete, A. (2010). Cosmological constraints on higgs boson mass. *Astrophys. J.* 723, 803–811. doi: 10.1088/0004-637X/723/1/803
- Postma, M., and Volponi, M. (2014). Equivalence of the Einstein and Jordan frames. *Phys. Rev. D* 90:103516. doi: 10.1103/PhysRevD.90.103516
- Rasanen, S. (2018). Higgs inflation in the Palatini formulation with kinetic terms for the metric, arXiv:1811.09514.
- Rasanen, S., and Tomberg, E. (2018). Planck scale black hole dark matter from Higgs inflation, arXiv:1810.12608.
- Rasanen, S., and Wahlman, P. (2017). Higgs inflation with loop corrections in the Palatini formulation. *J. Cosmol. Astropart. Phys.* 1711:047. doi: 10.1088/1475-7516/2017/11/047

- Ren, J., Xianyu, Z.-Z., and He, H.-J. (2014). Higgs gravitational interaction, weak Boson scattering, and Higgs inflation in Jordan and Einstein frames. *J. Cosmol. Astropart. Phys.* 1406:032. doi: 10.1088/1475-7516/2014/06/032
- Repond, J., and Rubio, J. (2016). Combined Preheating on the lattice with applications to Higgs inflation. *J. Cosmol. Astropart. Phys.* 1607:043. doi: 10.1088/1475-7516/2016/07/043
- Rubio, J. (2015). Higgs inflation and vacuum stability. *J. Phys. Conf. Ser.* 631:012032. doi: 10.1088/1742-6596/631/1/012032
- Rubio, J., and Shaposhnikov, M. (2014). Higgs-Dilaton cosmology: universality versus criticality. *Phys. Rev. D* 90:027307. doi: 10.1103/PhysRevD.90.027307
- Salopek, D. S., Bond, J. R., and Bardeen, J. M. (1989). Designing density fluctuation spectra in inflation. *Phys. Rev. D* 40:1753. doi: 10.1103/PhysRevD.40.1753
- Saltaş, I. D. (2016). Higgs inflation and quantum gravity: an exact renormalisation group approach. *J. Cosmol. Astropart. Phys.* 1602:048. doi: 10.1088/1475-7516/2016/02/048
- Salvio, A. (2013). Higgs inflation at NNLO after the Boson discovery. *Phys. Lett. B* 727, 234–239. doi: 10.1016/j.physletb.2013.10.042
- Salvio, A. (2018). Initial conditions for critical Higgs inflation. *Phys. Lett. B* 780, 111–117. doi: 10.1016/j.physletb.2018.03.009
- Sasaki, M. (1986). Large scale quantum fluctuations in the inflationary universe. *Prog. Theor. Phys.* 76:1036. doi: 10.1143/PTP.76.1036
- Shaposhnikov, M., and Wetterich, C. (2010). Asymptotic safety of gravity and the Higgs boson mass. *Phys. Lett. B* 683, 196–200. doi: 10.1016/j.physletb.2009.12.022
- Shaposhnikov, M., and Zenhausern, D. (2009). Scale invariance, unimodular gravity and dark energy. *Phys. Lett. B* 671, 187–192. doi: 10.1016/j.physletb.2008.11.054
- Smolin, L. (1979). Towards a theory of space-time structure at very short distances. *Nucl. Phys. B* 160, 253–268. doi: 10.1016/0550-3213(79)90059-2
- Spokoiny, B. L. (1984). Inflation and generation of perturbations in broken symmetric theory of gravity. *Phys. Lett. B* 147, 39–43. doi: 10.1016/0370-2693(84)90587-2
- Starobinsky, A. A. (1980). A new type of isotropic cosmological models without singularity. *Phys. Lett. B* 91, 99–102. doi: 10.1016/0370-2693(80)90670-X
- Starobinsky, A. A. (1982). Dynamics of phase transition in the new inflationary universe scenario and generation of perturbations. *Phys. Lett. B* 117, 175–178. doi: 10.1016/0370-2693(82)90541-X
- Starobinsky, A. A., and Yokoyama, J. (1994). Equilibrium state of a selfinteracting scalar field in the De Sitter background. *Phys. Rev. D* 50, 6357–6368. doi: 10.1103/PhysRevD.50.6357
- Steinwachs, C. F. (2013). *Non-minimal Higgs Inflation and Frame Dependence in Cosmology*. Cham: Springer. doi: 10.1007/978-3-319-01842-3
- Takahashi, F., and Takahashi, R. (2016). Renormalization group improved Higgs inflation with a running kinetic term. *Phys. Lett. B* 760, 329–334. doi: 10.1016/j.physletb.2016.07.009
- Tanabashi, M., Hagiwara, K., Hikasa, K., Nakamura, K., Sumino, Y., Takahashi, F., et al. (2018). Review of Particle Physics. *Phys. Rev. D* 98:030001. doi: 10.1103/PhysRevD.98.030001
- Trashorras, M., Nesseris, S., and Garcia-Bellido, J. (2016). Cosmological constraints on Higgs-dilaton inflation. *Phys. Rev. D* 94:063511. doi: 10.1103/PhysRevD.94.063511
- Tsamis, N. C., and Woodard, R. P. (1993). Relaxing the cosmological constant. *Phys. Lett. B* 301, 351–357. doi: 10.1016/0370-2693(93)91162-G
- Tsamis, N. C., and Woodard, R. P. (1995). Strong infrared effects in quantum gravity. *Annals Phys.* 238, 1–82. doi: 10.1006/aphy.1995.1015
- Tsujikawa, S., and Gumjudpai, B. (2004). Density perturbations in generalized Einstein scenarios and constraints on nonminimal couplings from the cosmic microwave background. *Phys. Rev. D* 69:123523. doi: 10.1103/PhysRevD.69.123523
- van de Bruck, C., and Longden, C. (2016). Higgs inflation with a Gauss-Bonnet term in the Jordan frame. *Phys. Rev. D* 93:063519. doi: 10.1103/PhysRevD.93.063519
- Vennin, V., and Starobinsky, A. A. (2015). Correlation functions in stochastic inflation. *Eur. Phys. J. C* 75:413. doi: 10.1140/epjc/s10052-015-3643-y
- Weenink, J., and Prokopec, T. (2010). Gauge invariant cosmological perturbations for the nonminimally coupled inflaton field. *Phys. Rev. D* 82:123510. doi: 10.1103/PhysRevD.82.123510
- Wetterich, C. (2017). Graviton fluctuations erase the cosmological constant. *Phys. Lett. B* 773, 6–19. doi: 10.1016/j.physletb.2017.08.002
- Xianyu, Z.-Z., and He, H.-J. (2014). Asymptotically safe Higgs inflation. *J. Cosmol. Astropart. Phys.* 1410:083. doi: 10.1088/1475-7516/2014/10/083
- Yoon, Y., and Yoon, Y. (1997). Asymptotic conformal invariance of SU(2) and standard models in curved space-time. *Int. J. Mod. Phys. A* 12, 2903–2914. doi: 10.1142/S0217751X97001602
- York Jr, J. W. (1972). Role of conformal three geometry in the dynamics of gravitation. *Phys. Rev. Lett.* 28, 1082–1085. doi: 10.1103/PhysRevLett.28.1082
- Zee, A. (1979). A broken symmetric theory of gravity. *Phys. Rev. Lett.* 42:417. doi: 10.1103/PhysRevLett.42.417

**Conflict of Interest Statement:** The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Rubio. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Cosmological Aspects of Higgs Vacuum Metastability

Tommi Markkanen<sup>1</sup>, Arttu Rajantie<sup>1\*</sup> and Stephen Stopyra<sup>1,2</sup>

<sup>1</sup> Department of Physics, Imperial College London, London, United Kingdom, <sup>2</sup> Department of Physics and Astronomy, University College London, London, United Kingdom

## OPEN ACCESS

### Edited by:

Alberto Salvio,  
European Organization for Nuclear  
Research (CERN), Switzerland

### Reviewed by:

Kazuharu Bamba,  
Fukushima University, Japan  
Vyacheslav Ivanovich Dokuchaev,  
Institute for Nuclear Research (RAS),  
Russia  
Yohei Ema,  
Deutsches Elektronen-Synchrotron,  
Helmholtz-Gemeinschaft Deutscher  
Forschungszentren (HZ), Germany

### \*Correspondence:

Arttu Rajantie  
a.rajantie@imperial.ac.uk

### Specialty section:

This article was submitted to  
Cosmology,  
a section of the journal  
Frontiers in Astronomy and Space  
Sciences

**Received:** 18 September 2018

**Accepted:** 21 November 2018

**Published:** 18 December 2018

### Citation:

Markkanen T, Rajantie A and  
Stopyra S (2018) Cosmological  
Aspects of Higgs Vacuum  
Metastability.  
Front. Astron. Space Sci. 5:40.  
doi: 10.3389/fspas.2018.00040

The current central experimental values of the parameters of the Standard Model give rise to a striking conclusion: metastability of the electroweak vacuum is favored over absolute stability. A metastable vacuum for the Higgs boson implies that it is possible, and in fact inevitable, that a vacuum decay takes place with catastrophic consequences for the Universe. The metastability of the Higgs vacuum is especially significant for cosmology, because there are many mechanisms that could have triggered the decay of the electroweak vacuum in the early Universe. We present a comprehensive review of the implications from Higgs vacuum metastability for cosmology along with a pedagogical discussion of the related theoretical topics, including renormalization group improvement, quantum field theory in curved spacetime and vacuum decay in field theory.

**Keywords:** Higgs boson, vacuum stability, quantum tunneling, quantum field theory, cosmological inflation

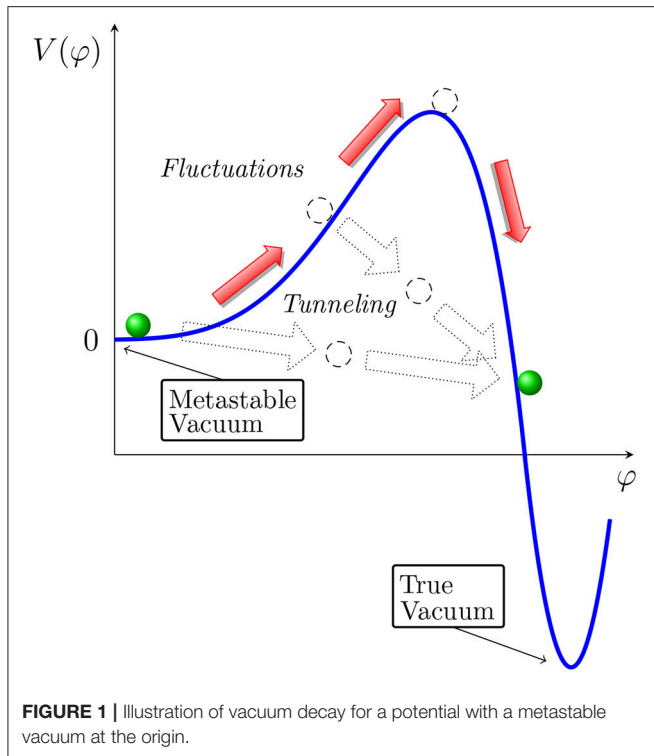
## 1. INTRODUCTION

One of the most striking results of the discovery of Higgs boson (Aad et al., 2012; Chatrchyan et al., 2012) has been that its mass lies in a regime that predicts the current vacuum state to be a false vacuum, that is, there is a lower energy vacuum state available to which the electroweak vacuum can decay into (Degrassi et al., 2012; Buttazzo et al., 2013). That this was a possibility in the Standard Model (SM) has been known for a long time (Hung, 1979; Sher, 1993; Casas et al., 1996; Isidori et al., 2001; Ellis et al., 2009; Elias-Miro et al., 2012). The precise behavior of the Higgs potential is sensitive to the experimental inputs, in particular the physical masses for the Higgs and the top quark and also physics beyond the SM. The current best estimates of the Higgs and top quark masses (Tanabashi et al., 2018),

$$M_h = 125.18 \pm 0.16 \text{ GeV}, \quad M_t = 173.1 \pm 0.9 \text{ GeV}, \quad (1.1)$$

place the Standard Model squarely in the metastable region.

As in any quantum system, there are three main ways in which the vacuum decay can happen. They are illustrated in **Figure 1**. If the system is initially in the false vacuum state, the transition would take place through quantum tunneling. On the other hand, if there is sufficient energy available, for example in a thermal equilibrium state, it may be possible for the system to move classically over the barrier. The third way consists of quantum tunneling from an excited initial state. This is often the dominant process if the temperature is too low for the fully classical process. All three mechanisms can be relevant for the decay of the electroweak vacuum state, and their rates depending on the conditions. In each of them, the transition happens initially locally in a small volume, nucleating a small bubble of the true vacuum. The bubble then starts to expand, reaching the speed of light very quickly, any destroying everything in its way.



If the Universe was infinitely old, even an arbitrarily low vacuum decay rate would be incompatible with our existence. The implications of vacuum metastability can therefore only be considered in the cosmological context, taking into account the finite age and the cosmological history of the Universe. Although the vacuum decay rate is extremely slow in the present day, that was not necessarily the case in the early Universe. High Hubble rates during inflation and high temperatures afterwards could have potentially increased the rate significantly. Therefore the fact that we still observe the Universe in its electroweak vacuum state allows us to place constraints on the cosmological history, for example the reheat temperature and the scale of inflation, and on Standard Model parameters, such as particle masses and the coupling between the Higgs field and spacetime curvature.

In this review we discuss the implications of Higgs vacuum metastability in early Universe cosmology and describe the current state of the literature. We also discuss all the theoretical frameworks, with detailed derivations, that are needed for the final results. This article complements earlier comprehensive reviews of electroweak vacuum metastability (Sher, 1989; Schrempf and Wimmer, 1996), which focus on the particle physics aspects rather than the cosmological context, and the recent introductory review (Moss, 2015) that explores the role of the Higgs field in cosmology more generally.

In section 2 we present renormalization group improvement in flat space by using the Yukawa theory as an example before discussing the full SM. Section 3 contains an overview of quantum field theory on curved backgrounds relevant for our purposes, including the modifications to the SM. In section 4 we go through the various ways vacuum decay can occur. In section

5 we discuss the connection to cosmology and in section 6 we present our concluding remarks.

Our sign conventions for the metric and curvature tensors are  $(-, -, -)$  in the classification of Misner et al. (1973) and throughout we will use units where the reduced Planck constant, the Boltzmann constant and the speed of light are set to unity,  $\hbar \equiv k_B \equiv c \equiv 1$ . The reduced Planck mass is given by Newton's constant as

$$M_P \equiv (8\pi G)^{-1/2} \approx 2.435 \times 10^{18} \text{ GeV}. \quad (1.2)$$

We will use  $\varphi$  for the vacuum expectation value (VEV) of a spectator field (usually the Higgs),  $\phi$  for the inflaton and  $\Phi$  for the SM Higgs doublet. The inflaton potential is  $U(\phi)$  and the Higgs potential  $V(\varphi)$ . The physical Higgs and top masses read  $M_h$  and  $M_t$ .

## 2. EFFECTIVE POTENTIAL IN FLAT SPACETIME

### 2.1. Example: Yukawa Theory

The possibility of quantum corrections destabilizing a classically stable vacuum has been known for quite some time (Krive and Linde, 1976; Krasnikov, 1978; Maiani et al., 1978; Politzer and Wolfram, 1979; Cabibbo et al., 1979; Hung, 1979). Although our focus will be strictly on the SM, one should keep in mind that the instability that potentially arises in the SM is only a specific example of a more general phenomenon that could manifest in a variety of other theories of elementary particles. For this reason all the essential features of the vacuum instability in the SM can be illustrated with the simple Yukawa theory, which we will now discuss before moving on to the full Standard Model in section 2.3.

The action containing a massless, quartically self-interacting scalar field  $\varphi$  Yukawa-coupled to a massless Dirac fermion  $\psi$  is

$$S = \int d^4x \left[ \frac{1}{2} \partial_\mu \varphi \partial^\mu \varphi - \frac{\lambda}{4} \varphi^4 + \bar{\psi} \not{\partial} \psi - g \varphi \bar{\psi} \psi \right]. \quad (2.1)$$

Classically, the potential for the scalar field is simply

$$V_{\text{cl}}(\varphi) = \frac{\lambda}{4} \varphi^4, \quad (2.2)$$

which quite trivially has a well-defined state of lowest energy at the origin.

When quantized the potential for the field  $\varphi$  becomes modified by quantum corrections

$$V(\varphi) = V_{\text{cl}}(\varphi) + \text{quantum corrections}, \quad (2.3)$$

which may be investigated within the usual framework of quantum field theory (Peskin and Schroeder, 1995). Importantly, it has been for a long time understood that in some instances predictions in a quantum theory can deviate significantly from those of the classical case. A prime example of such behavior

is radiatively induced symmetry breaking (Coleman Weinberg, 1973).

In the one-loop approximation the result for the quantum corrected or *effective* potential for the Yukawa model has the form (see for example, Markkanen et al., 2018)

$$V_{\text{eff}}(\varphi) = \frac{\lambda(\mu)}{4} \varphi^4(\mu) + \frac{1}{64\pi^2} \left[ M_\varphi^4(\mu) \left( \log \frac{M_\varphi^2(\mu)}{\mu^2} - \frac{3}{2} \right) - 4M_\psi^4(\mu) \left( \log \frac{M_\psi^2(\mu)}{\mu^2} - \frac{3}{2} \right) \right] + \dots, \quad (2.4)$$

with

$$M_\varphi^2(\mu) \equiv 3\lambda(\mu)\varphi^2(\mu); \quad M_\psi^2(\mu) \equiv g^2(\mu)\varphi^2(\mu). \quad (2.5)$$

In the above we have explicitly denoted the dependence on the *renormalization scale*  $\mu$ , which is an arbitrary energy scale, which one needs to choose in order to define the renormalized parameters of the theory. There is also a similar dependence in  $\varphi(\mu)$  which now refers to the renormalized one-point function of the quantized field, which is related to the bare field via the field renormalization constant (Peskin and Schroeder, 1995)

$$\varphi_{\text{bare}} = \sqrt{Z(\mu)}\varphi(\mu). \quad (2.6)$$

In the one-loop effective potential (2.4), the contribution from the fermion  $\psi$  comes with a minus sign. For sufficient high values of  $g$ , it can overtake the classical contribution and lead to a region with negative potential energy. In the limit of large field values  $\varphi \rightarrow \infty$ , one may write the potential as

$$V_{\text{eff}}(\varphi \rightarrow \infty) \rightarrow \varphi^4 \frac{9\lambda^2 - 4g^4}{32\pi^2} \log\left(\frac{\varphi}{\mu}\right) + \dots, \quad (2.7)$$

implying that if

$$\lambda < \lambda_{\text{cr}} \equiv \frac{2g^2}{3}, \quad (2.8)$$

the potential has a barrier and starts to decrease without bound at high field values (Krive and Linde, 1976). When  $\lambda$  is larger than the critical threshold  $\lambda_{\text{cr}}$  the quantum correction approaches  $+\infty$  indicating that an arbitrary small deviation from  $\lambda_{\text{cr}}$  leads either to  $+\infty$  or  $-\infty$  at large enough field values.

Hence we have seen that in the Yukawa theory the low-field vacuum will be separated by a barrier from an infinitely deep well on the other side. Even if the barrier is very robust, after a sufficiently long time the system initialized in the classical vacuum must eventually make a transition to the other side of the barrier and evolve toward the state of minimum energy.

A potential unbounded from below is a problematic concept and it is often assumed that, perhaps due to non-perturbative physics invisible to a loop expansion, some mechanism reverses the behavior of the potential at very high energies. This means that the minimum energy is in fact bounded from below, and

the effect of the quantum corrections is to generate second local minimum beyond the barrier as depicted in **Figure 1**. In theories containing  $U(1)$  gauge fields, such as the SM, the reversal of the potential can be shown to happen and the issue of an infinitely deep well does not arise. In the effective theory framework, which arguably is the correct way of viewing the SM, this issue is also not present as one will always encounter a finite scale beyond which the calculation becomes unreliable. Indeed, gravitational corrections are a prime example of a modification that is expected to become significant at large field values.

From a practical point of view, whether or not the potential is infinitely or deep of has a second or more accurately a *true* minimum beyond the barrier is not important for the generic prediction that the vacuum at the origin should eventually decay if the potential possesses regions with lower energy than at the origin.

However, conclusions based on the behavior of the perturbative one-loop result (2.4) may be premature. This is because for very large field values the logarithms become non-perturbatively large making the loop expansion invalid: generically one would expect higher powers of the logarithmic contributions in the square brackets of Equation (2.7) to be generated by higher orders in the expansion, as for example is evident in the results of Chung et al. (1999). Concretely, for our Yukawa theory (2.1) this requirement means that we can only draw conclusions in the region where

$$\frac{4g^4}{64\pi^2} \log\left(\frac{g^2\varphi^2}{\mu^2}\right) \lesssim 1 \quad \text{and} \quad \frac{9\lambda^2}{64\pi^2} \log\left(\frac{3\lambda\varphi^2}{\mu^2}\right) \lesssim 1. \quad (2.9)$$

In principle, the smaller the logarithms the more accurate the result.

## 2.2. Renormalization Group Improvement

By making use of renormalization group (RG) techniques it is possible to improve the accuracy of an existing perturbative expression such that the issue of large logarithms may be avoided (Kastening, 1992; Bando et al., 1993a,b; Ford et al., 1993).

Demanding that the effective potential (2.4) does not depend on the renormalization scale  $\mu$  gives rise to the Callan-Symanzik equation (Callan, 1970; Symanzik, 1970, 1971)

$$\frac{d}{d\mu} V_{\text{eff}}(\varphi) = 0 \Leftrightarrow \left\{ \mu \frac{\partial}{\partial \mu} + \beta_\lambda \frac{\partial}{\partial \lambda} + \beta_g \frac{\partial}{\partial g} - \gamma \varphi \frac{\partial}{\partial \varphi} \right\} V_{\text{eff}}(\varphi) = 0, \quad (2.10)$$

where we have defined the beta functions and the anomalous dimension in the usual manner

$$\beta_{c_i} \equiv \mu \frac{\partial c_i}{\partial \mu}, \quad \gamma \equiv \mu \frac{\partial \log \sqrt{Z}}{\partial \mu}, \quad (2.11)$$

with  $\gamma$  from the field renormalization constant in (2.6), which has a dependence on the renormalization scale  $Z \equiv Z(\mu)$ . Deriving the beta functions and the anomalous dimension for the Yukawa theory is a well-known calculation (see for example, Bando et al.,

1993a) and here we simply state the results

$$16\pi^2\beta_{m^2} = m^2(6\lambda + 4g^2), \quad (2.12)$$

$$16\pi^2\beta_\lambda = 18\lambda^2 + 8g^2\lambda - 8g^4, \quad (2.13)$$

$$16\pi^2\beta_g = 5g^3, \quad (2.14)$$

$$16\pi^2\gamma = 2g^2, \quad (2.15)$$

where for completeness we have included the beta function also for a mass parameter of the scalar field.

The beta functions tell us how the values of the renormalized parameters “run,” i.e., depend on the scale choice  $\mu$ . For example, assuming renormalized coupling value  $g(\mu_0)$  at some scale choice  $\mu_0$ , one may solve the running of the Yukawa coupling  $g(\mu)$  from Equation (2.14),

$$g^2(\mu) = \frac{g^2(\mu_0)}{1 - \frac{5g^2(\mu_0)}{8\pi^2} \log(\mu/\mu_0)}. \quad (2.16)$$

This shows that increasing  $\mu$  leads to a larger  $g(\mu)$ , and that the coupling  $g(\mu)$  appears to diverge at scale

$$\mu = \mu_0 \exp\left(\frac{8\pi^2}{5g^2(\mu_0)}\right), \quad (2.17)$$

which is known as the Landau pole (Landau, 1955). However, well before the Landau pole is reached, the loop expansion ceases to be valid. For more information on the effect of running couplings we refer the reader to more or less any textbook on quantum field theory (for example Cheng and Li, 1984; Peskin and Schroeder, 1995).

Even though the full effective potential  $V_{\text{eff}}(\varphi)$  has to be independent of the scale choice  $\mu$ , for any finite-order perturbative result that is only true up to neglected higher-order terms. This means that some scale choices will work better than the others, and by a judicious choice, one can improve the accuracy of the perturbative result. In general, one would choose the scale  $\mu$  to optimize the perturbative expansion in such a way that the loop corrections are small as indicated in Equation (2.9). However, for the effective potential (2.7), the loop corrections depend on the field value  $\varphi$ . Therefore each given choice of scale would only work well over a relatively narrow range of field values.

To ensure that Equation (2.9) remains satisfied at any field values, one can take this approach further and make the renormalization scale a function of the field  $\varphi$ ,

$$\mu = \mu_*(\varphi), \quad (2.18)$$

so that the expansion is optimized at all field values. This procedure is generically called *renormalization group improvement* (RGI)<sup>1</sup>. This way one can define the renormalization group improved (RGI) effective potential as

$$V_{\text{RGI}}(\varphi) \equiv V_{\text{eff,RG}}(\varphi, \mu_*(\varphi)). \quad (2.19)$$

<sup>1</sup>In our work the improvement is understood to come from the specific step of optimizing the expansion via a particular choice of  $\mu$ . In some works, it simply means making use of running couplings.

One should note that in this expression  $\varphi$  refers to the field defined at the field-dependent renormalization scale,  $\varphi = \varphi(\mu_*(\varphi))$  (for more discussion, see Markkanen et al., 2018), and that at any finite order in perturbation theory the resulting function  $V_{\text{RGI}}(\varphi)$  depends on the choice of the function  $\mu_*(\varphi)$ .

In principle, one could choose  $\mu_*$  in such a way that the loop correction vanishes exactly. For the one-loop potential (2.7) in the Yukawa theory, this would give

$$\mu_*^{\text{exact}}(\varphi) = e^{-3/4} \left( \frac{3\lambda}{g^2} \right)^{\frac{9\lambda^2}{18\lambda^2 - 8g^4}} g\varphi, \quad (2.20)$$

where both the couplings  $g$  and  $\lambda$  and the field  $\varphi$  are renormalized at scale  $\mu_*^{\text{exact}}(\varphi)$ , and therefore the equation defines the scale  $\mu_*(\varphi)$  implicitly. With this choice, the RGI effective potential  $V_{\text{RGI}}(\varphi)$  is given simply by the tree-level potential with  $\varphi$ -dependent couplings,

$$V_{\text{RGI}}(\varphi) = \frac{1}{4} \lambda(\mu_*^{\text{exact}}(\varphi)) \varphi^4. \quad (2.21)$$

In more realistic theories it is often impractical to choose  $\mu_*(\varphi)$  that cancels the loop correction exactly (Markkanen et al., 2018). Instead, one chooses some simpler function that keeps the loop correction sufficiently small. The most common choice in the literature is simply

$$\mu_*(\varphi) = \varphi. \quad (2.22)$$

Because the loop correction in Equation (2.7) does not vanish for this scale choice it should still be included in the effective potential. It is nevertheless, fairly common to make the further approximation of dropping it, and writing the tree-level RGI effective potential simply as

$$V_{\text{RGI}}^{\text{tree}}(\varphi) = \frac{\lambda(\varphi)}{4} \varphi^4. \quad (2.23)$$

For weak couplings this is not a good approximation, though. Equation (2.20) shows that the loop correction vanishes for  $\mu_* \approx g\varphi$ , and therefore a good approximation to RGI effective potential is

$$V_{\text{RGI}}(\varphi) \approx \frac{1}{4} \lambda(g\varphi) \varphi^4 = \frac{1}{4g^4} \lambda(g\varphi) (g\varphi)^4 = \frac{1}{g^4} V_{\text{RGI}}^{\text{tree}}(g\varphi). \quad (2.24)$$

From this we can see that the use of the tree-level RGI potential (2.23) with the scale choice (2.22) gets the barrier position wrong by a factor of  $g$  and the barrier height by a factor of  $g^4$ . Therefore one should either keep the one-loop correction, or use a more accurate scale choice.

From the beta function (2.13) for  $\lambda$ , we see that if  $g^2 \gg \lambda$ ,  $\lambda$  can become negative at high scales  $\mu$ . It is conventional to define the *instability scale*  $\mu_\Lambda$  as the scale where this happens,

$$\lambda(\mu_\Lambda) = 0. \quad (2.25)$$

If  $\mu_\Lambda < \infty$ , the effective potential (2.21) becomes negative at high field values, too, implying an instability. Again, the root

cause is a negative contribution from the fermions, this time in the beta functions.

The solution for the running  $\lambda(\mu)$  can be obtained analytically, but is unfortunately quite complicated (see e.g., Bando et al., 1993a). However, it is easy to see that the critical value of the coupling, below which the instability appears, is

$$\lambda_{\text{cr}} = \frac{1 + \sqrt{145}}{18} g^2. \quad (2.26)$$

Close to this critical value one may provide relatively simple analytical results. Suppose we have initial conditions given at some reference scale  $\mu_0$  for the running parameters  $g(\mu_0)$  and  $\lambda(\mu_0)$  the latter of which we parameterize as a fixed value  $\lambda_{\text{cr}}$  and a perturbation  $\delta\lambda$  as

$$g(\mu_0); \quad \lambda(\mu_0) \equiv \lambda_{\text{cr}} - \delta\lambda. \quad (2.27)$$

By solving Equations (2.13) and (2.14) explicitly one may show that  $\lambda(\mu)$  has the following expansion

$$\frac{\lambda(\mu)}{g^2(\mu)} = \frac{\lambda_{\text{cr}}}{g^2(\mu_0)} - \frac{\delta\lambda}{g^2(\mu_0)} \left( \frac{g^2(\mu)}{g^2(\mu_0)} \right)^{\sqrt{29/5}} + \mathcal{O}(\delta\lambda^2). \quad (2.28)$$

From Equation (2.16) it is apparent that, because  $g(\mu)$  is a monotonically increasing function of  $\mu$ , the RGI effective potential (2.21) is unbounded from below at large field values, for an arbitrarily small positive perturbation  $\delta\lambda > 0$ . For comparison, the threshold (2.8) in the unimproved case was  $\lambda_{\text{cr}}/g^2 = 2/3$ , somewhat lower than the RGI result (2.26).

The above makes apparent a very important generic feature: renormalization group improvement can lead to conclusions that are qualitatively different from the unimproved results. In particular, sizes of couplings deemed as well-behaved and hence giving rise to a stable potential may in fact reveal to result in an instability by the RG improved results. This also implies that close to the critical value the higher loop corrections become quite important as even a small change may tilt the conclusion from stable to unstable, or vice versa. This is also suggested by the fact that the couplings run very gradually and the precise value of the instability scale is very sensitive to small corrections: even a tiny change in the initial values or the running may change  $\mu_\Lambda$  by several orders of magnitude. These features are illustrated in the example below.

For concreteness, let us consider a numerical example that highlights the importance of renormalization group improvement. Specifically, we choose the Yukawa theory with a negligible mass parameter and with the initial conditions defined at the renormalization scale  $\mu_0$  as

$$g(\mu_0) = \frac{1}{\sqrt{2}}; \quad \lambda(\mu_0) = \frac{1 + \sqrt{145}}{36} - 10^{-2} \approx 0.352. \quad (2.29)$$

which from (2.27) can be seen to correspond to a choice that is below the critical value by

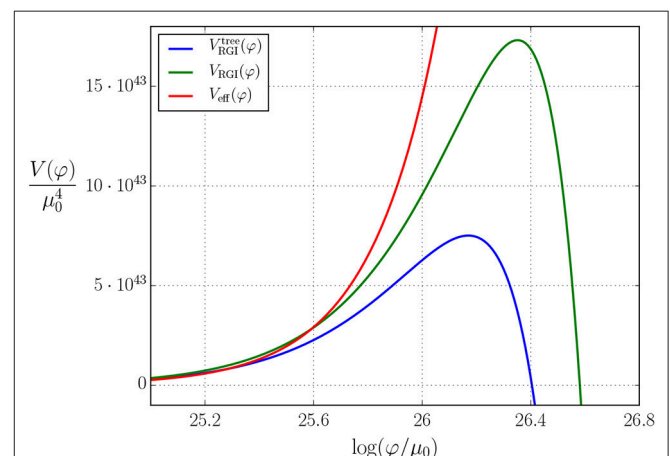
$$\delta\lambda = 10^{-2}. \quad (2.30)$$

Since Equation (2.29) satisfies  $\lambda(\mu_0) > \frac{2}{3}g^2(\mu_0)$  the unimproved effective potential (2.8) implies no instability. This is however not the case after renormalization group improvement as shown in **Figure 2**. We must however make sure that the above scale is such that all parameters remain perturbative, in particular for the Yukawa theory we need to check that the  $g$ -coupling is sufficiently small. For our parametrization this can be loosely expressed as  $2g^2(\mu_\Lambda) \lesssim 4\pi$  and perturbativity is easily demonstrated with the help of Equation (2.16). This check is quite important since if  $g(\mu)$  reaches a large value before  $\mu_\Lambda$ , it will render the entire derivation inconsistent.

What is also apparent from **Figure 2** that there is a clear difference between the tree-level RGI approximation (2.23) and the full RGI result (2.19), when using the simple non-exact scale choice (2.22). In many applications this would result in a non-negligible inaccuracy, but as shown in Equation (2.24), it changes the barrier position by a factor  $O(g)$  and height by  $O(g^4)$ , which can be important for vacuum stability. This sensitivity to quantum corrections and the choice of  $\mu_*$  comes from the fact that the instability occurs precisely at the point where the tree-level contribution vanishes.

### 2.3. Effective Potential in the Standard Model

The SM has a far richer particle content than the simple Yukawa theory of section 2.1, but the main reason for the possible vacuum instability remains the same: Quantum corrections from the fermions contribute with a minus sign and if significant enough can lead to the formation of regions with lower potential energy than the electroweak vacuum. In the SM the effect is mostly due to the top quark, because it is by far the heaviest and thus has the largest Yukawa coupling. As discussed in section 4, general field theory principles then dictate that after a sufficiently long time has passed the system should relax into the configuration



**FIGURE 2** | Behavior of the 1-loop RGI effective potential (2.19) (green), the tree-level RGI effective potential (2.23) (blue), and the non-improved result (2.4) (red) with the choices (2.29) at the reference scale  $\mu_0$ . The RGI scale choice was  $\mu_*(\varphi) = \varphi$ .

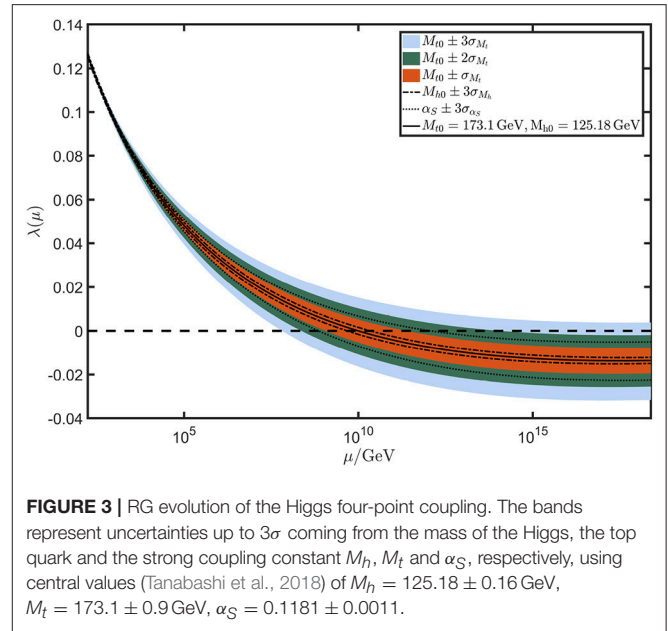
with the lowest energy resulting in the decay of the electroweak vacuum.

Through increasing experimental accuracy and improved analytic estimates in recent years it has become apparent that the central values for the couplings of the SM allow extrapolation to energy scales close to the Planck scale and that they are in fact incompatible with the situation where the electroweak vacuum would be the state of lowest energy. Some important early works addressing the question of vacuum instability are Krive and Linde (1976), Krasnikov (1978), Maiani et al. (1978), Politzer and Wolfram (1979), Hung (1979), and Cabibbo et al. (1979). The full body of work studying aspects of the vacuum instability is vast (to say the least) and includes Linde (1980), Lindner (1986), Lindner et al. (1989), Arnold (1989), Arnold and Vokos (1991), Ellwanger and Lindner (1993), Ford et al. (1993), Sher (1993), Altarelli and Isidori (1994), Casas et al. (1995), Espinosa and Quiros (1995), Casas et al. (1996), Hambye and Riesselmann (1997), Nie and Sher (1999), Frampton et al. (2000), Isidori et al. (2001), Gonderinger et al. (2010), Ellis et al. (2009), Holthausen et al. (2012), Elias-Miro et al. (2012), Chen and Tang (2012), Elias-Miro et al. (2012), Rodejohann and Zhang (2012), Bezrukov et al. (2012), Datta and Raychaudhuri (2013), Alekhin et al. (2012), Chakraborty et al. (2013), Anchordoqui et al. (2013), Masina (2013), Chun et al. (2012), Chung et al. (2013), Gonderinger et al. (2012), Degraasi et al. (2012), Buttazzo et al. (2013), Bhupal Dev et al. (2013), Nielsen (2012), Tang (2013), Klinkhamer (2013), He et al. (2013), Chun et al. (2013), Jegerlehner (2014), Branchina and Messina (2013), Di Luzio et al. (2016), Martin (2014), Gies et al. (2014), Branchina and Messina (2017), Eichhorn et al. (2015), Antipin et al. (2013), Chao et al. (2012), Spencer-Smith (2014), Chetyrkin and Zoller (2012), Chetyrkin and Zoller (2013), Gabrielli et al. (2014), Branchina et al. (2015), Bednyakov et al. (2015), Branchina et al. (2014), Bednyakov et al. (2013), Bednyakov et al. (2013), Bednyakov et al. (2014), Kobakhidze and Spencer-Smith (2013), Salvio et al. (2016), Chigusa et al. (2018), Chigusa et al. (2017), Garg et al. (2017), Khan and Rakshit (2015), Khan and Rakshit (2014), Liu and Zhao (2013), Bambhaniya et al. (2017), Schrempp and Wimmer (1996), Sher (1989), and Moss (2015).

The modern high precision era of vacuum instability investigations can be thought to have been initiated by the detailed analyses performed in Degraasi et al. (2012) and Buttazzo et al. (2013), which presented the first complete next-to-next-to-leading order analysis of the Standard Model Higgs potential and the running couplings.

The current state-of-the-art calculation for the running of Standard Model parameters uses two-loop matching conditions, three-loop RG evolution and pure QCD corrections to four-loop order (Bednyakov et al., 2015). The running of the Higgs self-coupling  $\lambda$  is shown in **Figure 3** for the central mass values (1.1), together with bands showing the effects of the estimated errors in the parameter values. For the central mass values (1.1), the instability scale (2.25), defined by  $\lambda(\mu_\Lambda) = 0$ , is

$$\mu_\Lambda = 9.92 \times 10^9 \text{ GeV}. \quad (2.31)$$



**FIGURE 3 |** RG evolution of the Higgs four-point coupling. The bands represent uncertainties up to  $3\sigma$  coming from the mass of the Higgs, the top quark and the strong coupling constant  $M_h$ ,  $M_t$  and  $\alpha_S$ , respectively, using central values (Tanabashi et al., 2018) of  $M_h = 125.18 \pm 0.16 \text{ GeV}$ ,  $M_t = 173.1 \pm 0.9 \text{ GeV}$ ,  $\alpha_S = 0.1181 \pm 0.0011$ .

This depends sensitively on the top and Higgs masses: At  $1\sigma$  the range is  $1.16 \times 10^9 \text{ GeV} < \mu_\Lambda < 2.37 \times 10^{11} \text{ GeV}$ , and the case in which  $\lambda(\mu)$  is never negative is still included within  $3\sigma$  uncertainty. Using the three-loop running, and including the one-loop correction in the RGI effective potential with the scale choice  $\mu_*(\varphi) = \varphi$ , the top of the potential barrier lies at

$$\varphi_{\text{bar}} = 4.64 \times 10^{10} \text{ GeV}, \quad (2.32)$$

and the barrier height is

$$\begin{aligned} \Delta V(\varphi_{\text{bar}}) &= V(\varphi_{\text{bar}}) - V(\varphi_F) = 3.46 \times 10^{38} \text{ GeV}^4 \\ &= (4.31 \times 10^9 \text{ GeV})^4. \end{aligned} \quad (2.33)$$

For comparison, the tree-level RGI form (2.23), which means dropping the one-loop correction and is common in the literature, would give a significantly lower position for the top of the potential barrier,  $\varphi_{\text{bar}} = 7.70 \times 10^9 \text{ GeV}$ . Using the unimproved one-loop effective potential with parameters renormalized at the electroweak scale gives as even lower value  $\varphi_{\text{bar}} = 5.78 \times 10^4 \text{ GeV}$ . This demonstrates that, as discussed in section 2.2, the use of renormalization group improvement and the inclusion of at least the one-loop correction in the RGI effective potential are both crucial for accurate results.

A slightly more formal issue that must also be kept in mind is that the barrier position  $\varphi_{\text{bar}}$  is in fact gauge dependent and strictly speaking has limited physical significance (Andreassen et al., 2014; Di Luzio and Mihaila, 2014; Espinosa et al., 2017, 2016). The value of the potential at its extrema are however gauge independent as demanded by the famous Nielsen identity (Nielsen, 1975). In the simplest approximation the probability of vacuum decay involves only the values of the potential at the extrema and subtleties involving gauge dependence are evaded. Furthermore, more precise calculations of the rate of vacuum

decay, since it is a physical process, can be expected to always be cast into a gauge-invariant form (Plascencia and Tamarit, 2016).

### 3. FIELD THEORY IN EXPANDING UNIVERSE

#### 3.1. Spectator Field on a Curved Background

In the extreme conditions of the early Universe, gravity plays a significant role. In order to investigate the consequences from Higgs metastability we must therefore make use of an approach that incorporates also gravitational effects. This can be achieved in the framework of quantum field theory in a curved spacetime. The study of quantum fields theory on curved backgrounds is hardly a recent endeavor. For a thorough discussion on the subject we refer the reader to the standard textbooks, such as Birrell and Davies (1984), Mukhanov and Winitzki (2007), and Parker and Toms (2009).

As a representative model we choose an action consisting only of a self-interacting scalar field

$$S = \int d^4x \sqrt{|g|} \left[ \frac{1}{2} \nabla_\mu \varphi \nabla^\mu \varphi - \frac{1}{2} m^2 \varphi^2 - \frac{\xi}{2} R \varphi^2 - \frac{\lambda}{4} \varphi^4 \right], \quad (3.1)$$

where the curved background is visible in the metric dependence of integration measure,  $\sqrt{|g|}$ , the covariant derivative  $\nabla_\mu$  and in the appearance of the non-minimal coupling  $\xi$  that connects the field to the scalar curvature of gravity  $R$ . The necessity of an operator  $\propto R\varphi^2$  was discovered already in Tagirov (1973), Callan et al. (1970), and Chernikov and Tagirov (1968), the reasons for which we will elaborate in section 3.5. It will turn out to be a key ingredient for the implications of the vacuum (in)stability in the early Universe.

Since our discussion assumes a classical curved background with no fluctuations of the metric  $g^{\mu\nu}$  some effects visible in a complete quantum gravity approach are possibly missed. For energy scales below the Planck threshold and for spectator fields with a negligible effect on the evolution of the background modifications from quantum gravity are expected to be suppressed. For the case of a quasi de Sitter background this was verified in detail in Markkanen et al. (2017) for the SM Higgs. The reason why quantum gravity is not relevant for a potential SM metastability can be understood from the simple fact that the instability scale (see section 2.3) is significantly lower than the Planck mass

$$\frac{\mu_\Lambda}{M_P} \approx 10^{-8}. \quad (3.2)$$

#### 3.2. Homogeneous and Isotropic Spacetime

From the cosmological point of view it is often sufficient to consider the special case of a homogeneous and isotropic spacetime with the Friedmann–Lemaître–Robertson–Walker (FLRW) line-element given in cosmic time as

$$ds^2 = dt^2 - a(t)^2 d\mathbf{x}^2, \quad (3.3)$$

where  $a(t) \equiv a$  is the scale factor describing cosmic acceleration. We will furthermore assume that the energy and pressure densities of the background,  $\rho$  and  $p$ , are connected via the constant equation of state parameter  $w$  as

$$p = w\rho; \quad \rho = T_{00}, \quad p = T_{ii}/a(t)^2, \quad (3.4)$$

where  $T_{\mu\nu}$  is the energy-momentum tensor of the background. With the line-element (3.3) the Einstein equation reduces to the Friedmann equations

$$\begin{cases} 3H^2 M_P^2 &= 3\left(\frac{\dot{a}}{a}\right)^2 M_P^2 &= \rho \\ -(3H^2 + 2\dot{H})M_P^2 &= -\left[\left(\frac{\dot{a}}{a}\right)^2 + 2\frac{\ddot{a}}{a}\right] M_P^2 &= p = w\rho \end{cases}, \quad (3.5)$$

which allow one to easily find expressions for the Hubble rate  $H \equiv \dot{a}/a$  and the scale factor as functions of  $w$

$$\begin{aligned} a &= \left(\frac{t}{t_0}\right)^{\frac{2}{3(w+1)}}, \quad H = \frac{2}{3(w+1)t}, \quad \text{for } w \neq -1 \\ a &= e^{Ht}, \quad H = H_0, \quad \text{for } w = -1. \end{aligned} \quad (3.6)$$

For the purposes of this discussion the most important quantity characterizing gravitational effects will be the scalar curvature of gravity  $R$ , which may be written as a function of the equation of state parameter and the Hubble rate

$$R = 6\left[\left(\frac{\dot{a}}{a}\right)^2 + \frac{\ddot{a}}{a}\right] = 3(1 - 3w)H^2. \quad (3.7)$$

#### 3.3. Amplified Fluctuations

Let us then concentrate on a free quantum theory by setting  $\lambda = 0$ . For this case the action (3.1) leads to the equation of motion

$$(\square + m^2 + \xi R)\hat{\varphi} = 0, \quad (3.8)$$

whose solutions, as usual, can be expressed as a mode expansion

$$\hat{\varphi} = \int \frac{d^3\mathbf{k} e^{i\mathbf{k}\cdot\mathbf{x}}}{\sqrt{(2\pi)^3 a^2}} \left[ \hat{a}_{\mathbf{k}} f_{\mathbf{k}}(\eta) + \hat{a}_{-\mathbf{k}}^\dagger f_{\mathbf{k}}^*(\eta) \right], \quad (3.9)$$

with  $[\hat{a}_{\mathbf{k}}, \hat{a}_{\mathbf{k}'}^\dagger] = \delta^{(3)}(\mathbf{k} - \mathbf{k}')$ ,  $[\hat{a}_{\mathbf{k}}, \hat{a}_{\mathbf{k}'}] = [\hat{a}_{\mathbf{k}}^\dagger, \hat{a}_{\mathbf{k}'}^\dagger] = 0$ , where  $\mathbf{k}$  is the co-moving momentum and  $k \equiv |\mathbf{k}|$ . In the above we have also made use of conformal time defined as

$$\eta = \int^t \frac{dt'}{a(t')} \quad \Rightarrow \quad ds^2 = a^2(d\eta^2 - d\mathbf{x}^2). \quad (3.10)$$

From Equations (3.8) and (3.9) we may write down the mode equation

$$f_{\mathbf{k}}''(\eta) + \left[ \mathbf{k}^2 + a^2 \mathcal{M}^2 \right] f_{\mathbf{k}}(\eta) = 0, \quad (3.11)$$

where the primes denote derivatives with respect to conformal time and we have defined the effective mass

$$\mathcal{M}^2 \equiv m^2 + \left( \xi - \frac{1}{6} \right) R. \quad (3.12)$$

Equation (3.11) may be interpreted as that of a harmonic oscillator with a time-dependent mass. The crucial point is that for many cosmologically relevant combinations of  $m$ ,  $\xi$  and  $w$  the  $\mathcal{M}^2$ -contribution is in fact negative. A prime example would be a massless minimally coupled scalar field during cosmological inflation for which  $m = 0$ ,  $\xi = 0$  and  $w = -1$  giving  $\mathcal{M}^2 = -2H^2$ . If  $\mathcal{M}^2 < 0$  it is a simple matter to show that the modes with  $(\mathbf{k}/a)^2 + \mathcal{M}^2 < 0$  contain an exponentially growing branch, which implies that a large field fluctuation can be generated. This effect coming from an imaginary mass-like contribution is sometimes called tachyonic or spinodal instability/amplification (Felder et al., 2001). We note that even if no tachyonicity occurs, a large fluctuation can nonetheless be generated if there is a rapid i.e., a non-adiabatic change in  $\mathcal{M}$ .

A more precise way of understanding the generation of a large fluctuation is by calculating the infrared (IR) portion of the variance i.e., a loop with a low-momentum cut-off  $\Lambda_{\text{IR}}$ . This shows that in many situations that can broadly be characterized as having  $\mathcal{M}^2 \lesssim 0$  the result diverges (Markkanen, 2018)<sup>2</sup>

$$\langle \hat{\phi}^2 \rangle_{\Lambda_{\text{IR}}} \propto \int_0^{\Lambda_{\text{IR}}} dk k^2 |f_k(\eta)|^2 \xrightarrow{t \rightarrow \infty} \infty; \quad \text{for } \lambda = 0. \quad (3.13)$$

When the theory is not free interactions will via backreaction prevent the generation of arbitrary large fluctuations. In practice one may understand this as the emergence of positive mass-like contributions from the interactions making the field heavy and thus preventing tachyonic or non-adiabatic amplification. The functioning of this mechanism usually allows a significant  $\langle \hat{\phi}^2 \rangle$  term indicating that quite generally an IR divergence in the free theory implies a large fluctuation when interactions are included.

This rather simple discussion leads to an important implication in regards the vacuum instability problem in the cosmological setting: even if in flat space the decay of a metastable vacuum is enormously unlikely, this may not have been the case during the earlier cosmological epochs when a transition over the barrier can be induced by a large fluctuation generated by the dynamics on a curved background.

### 3.4. Quantum Theory in de Sitter Space

Even in the simple special case of a de Sitter background it is difficult to perform analytic calculations for an interacting quantum theory. This is mostly due to the non-trivial infrared behavior of quantum fields in de Sitter space (Allen, 1985; Sasaki et al., 1993; Mukhanov et al., 1997; Abramo et al., 1997; Prokopec et al., 2003; Onemli and Woodard, 2004; Losic and Unruh, 2005; Enqvist et al., 2008). A manifestation of this is the lack of a perturbative expansion based on a non-interacting propagator due to the infrared divergence as described in Equation (3.13). The infrared properties of de Sitter space

have attracted significant attention over the years and we refer the interested reader to the review (Seery, 2010) for more information and references.

One popular way forward is to use techniques based on the so-called two-particle-irreducible (2PI) diagrams, which are essentially non-perturbative resummations of distinct classes of Feynman diagrams. The 2PI approach is attractive in that it is derivable via first principles from quantum field theory without any approximations. Hence, in principle it can be used up to arbitrary accuracy. Unfortunately, only the leading terms that come by the Hartree approximation are analytically tractable. Applications of 2PI techniques to de Sitter space include Riotto and Sloth (2008), Tranberg (2008), Arai (2012), Serreau (2011), Garbrecht et al. (2014), Herranen et al. (2014), and Gautier and Serreau (2015).

A non-perturbative framework for calculating quantum correlators in de Sitter space was laid out in Starobinsky (1986) and Starobinsky and Yokoyama (1994). This technique is generally known as the stochastic formalism and is surprisingly straightforward calculationally. It is based on the insight that to a good approximation in de Sitter space one may neglect the quantum nature of the problem and devise a set-up in which the correlators may be calculated from a classical probability distribution  $P(t, \varphi)$ . If the scalar field  $\hat{\phi}$  is light,  $m \ll H$ , coarse graining over horizon sized patches allows one to approximate its dynamics with a Langevin equation

$$\dot{\phi} = -\frac{V'(\varphi)}{3H} + f(t), \quad (3.14)$$

where  $V(\varphi)$  is the classical potential and  $f(t)$  is a white noise term satisfying

$$\langle f(t')f(t) \rangle = \frac{H^3}{4\pi^2} \delta(t' - t). \quad (3.15)$$

The reason why the “hat” notation has been dropped from  $\varphi$  is that Equation (3.14) contains only classical stochastic quantities i.e., the quantum features are no longer visible.

The Langevin Equation (3.14) can be cast in the form of a Fokker-Planck equation for the probability density  $P(t, \varphi)$  (Starobinsky and Yokoyama, 1994)

$$\dot{P}(t, \varphi) = \frac{1}{3H} \frac{\partial}{\partial \varphi} [P(t, \varphi) V'(\varphi)] + \frac{H^3}{8\pi^2} \frac{\partial^2}{\partial \varphi^2} P(t, \varphi). \quad (3.16)$$

After a sufficiently long time has passed one would expect that  $P(t, \varphi)$  reaches a constant equilibrium distribution. When  $\dot{P}(t, \varphi) = 0$ , Equation (3.16) has a simple analytic solution as

$$P_{\text{eq}}(\varphi) = N \exp \left\{ -\frac{8\pi^2 V(\varphi)}{3H^4} \right\}, \quad (3.17)$$

where  $N$  is a normalization factor.

As an example, for a theory with only a quartic term  $V(\varphi) = (\lambda/4)\varphi^4$ , which in many cases is the relevant approximation for

<sup>2</sup>For example in de Sitter space one has  $|f_{k \rightarrow 0}(\eta)|^2 \propto H^2 k^{-3}$  for  $m = \xi = 0$ .

the SM Higgs in the early Universe, this results in the equilibrium probability distribution

$$P_{\text{eq}}(\varphi) = \left( \frac{32\pi^2\lambda}{3H^4} \right)^{1/4} \frac{1}{\Gamma(1/4)} \exp \left\{ -\frac{2\pi^2\lambda\varphi^4}{3H^4} \right\}. \quad (3.18)$$

The corresponding field variance becomes

$$\langle \hat{\varphi}^2 \rangle = \sqrt{\frac{3}{2\pi^2}} \frac{\Gamma(\frac{3}{4})}{\Gamma(\frac{1}{4})} \frac{H^2}{\sqrt{\lambda}} \approx 0.132 \frac{H^2}{\sqrt{\lambda}}. \quad (3.19)$$

This means that the Higgs field develops a non-zero value  $\varphi \sim \lambda^{-1/4}H$ , which is sometimes called a condensate (Kunimitsu and Yokoyama, 2012; Enqvist et al., 2013, 2014; Kusenko et al., 2015; Enqvist et al., 2016; Pearce et al., 2015; Freese et al., 2018; Hardwick, 2018).

The central assumption that leads to the stochastic description is that the effect of the ultraviolet physics on the infrared behavior can be described as a white noise term in the Langevin Equation (3.14). The ultraviolet modes also contribute to the effective potential  $V(\varphi)$  in the Fokker-Planck Equation (3.16), as was discussed in section 2.2 in flat space. These are two separate effects, which both need to be included in the calculation (Markkanen et al., 2018). Especially when investigating the vacuum stability of the SM it is therefore imperative that the quantum corrections are incorporated in the stochastic approach, for example by making use of the RGI effective potential as the input in Equation (3.16).

### 3.5. Curvature Corrections to the Effective Potential

It is clearly evident from the derivations of section 3.3 that a scalar field in curved spacetime feels the curvature of the background. It then follows that also the effective potential must receive a contribution from curvature. In order to reliably investigate the implications from the SM metastability in the early Universe these contributions then must be included in a discussion of quantum corrections to the potential.

Investigations of the effective potential on a curved background have been performed by a number of authors in a variety of models (Ford and Toms, 1982; Toms, 1982, 1983; Hu and O'Connor, 1984; Buchbinder et al., 1985; Odintsov, 1991; Buchbinder et al., 1992; Elizalde and Odintsov, 1994a, 1993, 1994b; Kirsten et al., 1993; Odintsov, 1993; Elizalde and Odintsov, 1994c; George et al., 2012; Czerwińska et al., 2015; Bounakis and Moss, 2018). However, the derivation of the effective potential for the full SM in curved spacetime was only recently carried out in Markkanen et al. (2018).

Deriving the effective potential for a quantized scalar field on a curved background is naturally much more difficult than in flat space: for many backgrounds even the case of a free scalar field admits no closed form solutions for the mode equation (Birrell and Davies, 1984). Another complication that arises is that choosing the boundary condition i.e., the specific quantum state in which the effective potential is calculated is far from obvious. This is due to the fact that in curved space the concept of a particle and hence the vacuum state is no longer well-defined

globally, but depends on the specific dynamics and perceptions of a given particular observer (Gibbons and Hawking, 1977).

However, even on an arbitrary curved background some things remain universal: renormalizability of a quantum field theory imposes the requirement that all quantum states should have coinciding divergences. From this it follows that it is possible to derive an effective potential retaining terms only originating from the very high ultraviolet (UV), which is a contribution that is always present irrespective of the quantum state one is interested in. Such an expression would then allow one to determine all the generated operators and their respective runnings, as RG effects are ultimately the result of UV physics.

Let us once more study the Yukawa theory of section 2.1 only this time in curved spacetime and without neglecting the mass parameter for the scalar. In curved spacetime the action reads

$$S = \int d^4x \sqrt{|g|} \left[ \frac{1}{2} \nabla_\mu \varphi \nabla^\mu \varphi - \frac{1}{2} m^2 \varphi^2 - \frac{\xi}{2} R \varphi^2 - \frac{\lambda}{4} \varphi^4 + \bar{\psi} \not{\nabla} \psi - g \varphi \bar{\psi} \psi \right]. \quad (3.20)$$

The most convenient way of deriving the effective potential is the Heat Kernel method reviewed in Avramidi (2000), see also Buchbinder et al. (1992). This approach has been known for a long time, see Schwinger (1951), DeWitt (1964), Seeley (1967), Gilkey (1975), Minakshisundaram and Pleijel (1949), and Hadamard (1923) for early work. We will make use of the resummed form of the Heat Kernel expansion derived in Parker and Toms (1985) and Jack and Parker (1985), which for the action (3.20) gives (for details, see Markkanen et al., 2018)

$$V_{\text{eff}}(\varphi) = \frac{1}{2} m^2 \varphi^2 + \frac{\xi}{2} R \varphi^2 + \frac{\lambda}{4} \varphi^4 + V_\varphi^{(1)}(\varphi) + V_\psi^{(1)}(\varphi), \quad (3.21)$$

where the one-loop quantum corrections from the scalar and the fermion,  $V_\varphi^{(1)}(\varphi)$  and  $V_\psi^{(1)}(\varphi)$ , read

$$V_\varphi^{(1)}(\varphi) = \frac{\mathcal{M}_\varphi^4}{64\pi^2} \left[ \log \left( \frac{|\mathcal{M}_\varphi^2|}{\mu^2} \right) - \frac{3}{2} \right] + \frac{\frac{1}{90} (R_{\mu\nu\delta\eta} R^{\mu\nu\delta\eta} - R_{\mu\nu} R^{\mu\nu})}{64\pi^2} \log \left( \frac{|\mathcal{M}_\varphi^2|}{\mu^2} \right), \quad (3.22)$$

and

$$V_\psi^{(1)}(\varphi) = -\frac{4\mathcal{M}_\psi^4}{64\pi^2} \left[ \log \left( \frac{|\mathcal{M}_\psi^2|}{\mu^2} \right) - \frac{3}{2} \right] + \frac{\frac{1}{90} ((7/2) R_{\mu\nu\delta\eta} R^{\mu\nu\delta\eta} + 4R_{\mu\nu} R^{\mu\nu})}{64\pi^2} \log \left( \frac{|\mathcal{M}_\psi^2|}{\mu^2} \right), \quad (3.23)$$

respectively, and the curved space effective masses  $\mathcal{M}_\varphi^2$  and  $\mathcal{M}_\psi^2$  are now

$$\mathcal{M}_\varphi^2 \equiv m^2 + 3\lambda\varphi^2 + (\xi - 1/6) R; \quad \mathcal{M}_\psi^2 \equiv g^2\varphi^2 + R/12. \quad (3.24)$$

The  $R^{\mu\nu}$  and  $R^{\mu\nu\alpha\beta}$  are the Ricci and Riemann tensors, respectively. We have introduced the absolute values in the logarithms to ensure that the result is never complex. A complex effective potential in flat space can be interpreted as a finite lifetime of the quantum state (Weinberg and Wu, 1987), but this is ultimately an infrared effect and hence not correctly represented in an UV expansion. Therefore, the effective potential in curved space derived with the Heat Kernel expansion correctly represents the local physics and can for example be used to determine the running of parameters in curved space and the possible generation of new operators (see the next section), but in order to answer questions about vacuum decay one needs additional technology, which is discussed in section 4.

What the above clearly shows is that on a curved background a highly non-trivial dependence on the curvature emerges: A curved spacetime leads to the generation of additional operators that couple to the scalar field. Importantly, the non-minimal term  $\propto R\varphi^2$  directly coupling  $\varphi$  to  $R$  is not the only one, but terms  $\propto R^2$ ,  $R_{\mu\nu}R^{\mu\nu}$ , and  $R_{\mu\nu\delta\eta}R^{\mu\nu\delta\eta}$  are also unavoidable and they couple to the scalar field via the logarithmic loop contributions. These terms are not necessarily small, for example in de Sitter space with a constant Hubble rate  $H$  the various curvature contributions may be written as

$$R = 144H^4, \quad R_{\mu\nu}R^{\mu\nu} = 36H^4, \quad R_{\mu\nu\delta\eta}R^{\mu\nu\delta\eta} = 24H^4, \quad (3.25)$$

and in the early Universe the Hubble rate can be several orders of magnitude larger than any mass parameter of the SM. Simply put, since curvature is felt by the scalar field its inclusion in the calculation is vital for making robust predictions because the scale provided by  $H$  often is the largest scale of the problem.

### 3.6. Running Couplings in Curved Space

The basic principles laid out in the flat space analysis of section 2.2 remain unchanged when the background is no longer flat: Demanding a result independent of the renormalization scale  $\mu$  leads to the Callan-Symanzik equation from which the beta functions may be solved given the anomalous dimension  $\gamma$ . Since  $\gamma$  is a dimensionless number it will receive no contributions from constants associated with the curvature of space such as  $\xi$ . Otherwise parameters only visible in the action when the background is curved would nonetheless influence the RG running of, say,  $\lambda$ . Similar arguments imply that all beta functions present in flat space remain unchanged when the background is curved.

As one may see from Equations (3.21)–(3.23) operators that are not present in the tree-level action (3.20) are generated by the loop correction. This means that even if one renormalizes these terms to zero, they may resurface via RG running. Ultimately, this is the reason behind the non-minimal term  $\propto R\varphi^2$  already in (3.20). For the same reason in our theory we must include the following purely gravitational action

$$S_g = - \int d^4x \sqrt{|g|} \left[ V_\Lambda - \kappa R + \alpha_1 R^2 + \alpha_2 R_{\mu\nu}R^{\mu\nu} + \alpha_3 R_{\mu\nu\delta\eta}R^{\mu\nu\delta\eta} \right]. \quad (3.26)$$

A straightforward application of the Callan-Symanzik Equation (2.10) with Equation (2.15) for Equations (3.21)–(3.23) gives the beta functions for the Yukawa theory

$$\begin{aligned} \beta_\xi &= \frac{(\xi - 1/6)}{16\pi^2} (6\lambda + 4g^2); & \beta_{V_\Lambda} &= \frac{m^4/2}{16\pi^2}; & \beta_\kappa &= -\frac{m^2(\xi - 1/6)}{16\pi^2}; \\ \beta_{\alpha_1} &= \frac{(\xi - 1/6)^2/2 - 1/72}{16\pi^2}; & \beta_{\alpha_2} &= \frac{1/60}{16\pi^2}; & \beta_{\alpha_3} &= \frac{1/40}{16\pi^2}. \end{aligned} \quad (3.27)$$

which along with Equations (2.12)–(2.15) provide a complete set of RG equations for the Yukawa theory in curved spacetime.

A crucial difference to the flat space case arises when implementing renormalization group improvement. In section 2.2 we exploited the fact that the full quantum result must be independent of the renormalization scale  $\mu$  in order to optimize the perturbative expansion. Namely, we made the choice (2.18) in order to keep the logarithms small also at large scales. In curved space the logarithms in the loop corrections (3.22) and (3.23) have dependence on the scalar curvature  $R$ , and therefore it must be included in the optimization. The one-loop calculation shows that the exact scale choice that would fully cancel the loop correction is not possible across the whole range of field values (Markkanen et al., 2018). Instead, a sensible choice for the optimized scale  $\mu_*$  is a linear combination of  $\varphi^2$  and  $R$  i.e.,

$$\mu_*^2(\varphi, R) = a\varphi^2 + bR, \quad (3.28)$$

where the parameters  $a$  and  $b$  are chosen in such a way that the logarithms remain under control.

Equation (3.28) highlights an often neglected effect arising in curved spaces after renormalization group improvement: In a curved background the optimal scale choice depends significantly on curvature. This phenomenon may be characterized as *curvature induced running* and was recently studied in detail for the full SM in Markkanen et al. (2018). In situations where the curvature of the background is significant it can give the dominant contribution to the scale. Considering the metastability of the SM in the early Universe this in fact is often the case as during and after inflation one may have a Hubble rate much larger than the instability scale,  $H \gg \mu_\Lambda$ .

### 3.7. The Standard Model

The Standard Model particle content can be expressed with the Lagrangian

$$\mathcal{L}_{\text{SM}} = \mathcal{L}_{\text{YM}} + \mathcal{L}_{\text{F}} + \mathcal{L}_{\Phi} + \mathcal{L}_{\text{GF}} + \mathcal{L}_{\text{GH}}. \quad (3.29)$$

The first three terms in Equation (3.29) describe the contributions coming from the gauge fields, the fermions and the Higgs doublet  $\Phi$  whose one point function we write as  $\langle \hat{\Phi} \rangle \equiv \varphi$ , from now on dropping the hats. The “GF” and “GH” are the gauge fixing and ghost Lagrangians, respectively. Here we show explicitly only the Higgs contribution (for the full result see Markkanen et al., 2018)

$$\mathcal{L}_{\Phi} = (D_\mu \Phi)^\dagger (D^\mu \Phi) + m^2 \Phi^\dagger \Phi - \xi R \Phi^\dagger \Phi - \lambda (\Phi^\dagger \Phi)^2, \quad (3.30)$$

with the SM covariant derivative

$$D_\mu = \nabla_\mu - ig\tau^a A_\mu^a - ig'YA_\mu; \quad \tau^a = \sigma^a/2, \quad (3.31)$$

where  $\nabla_\mu$  contains the connection appropriate for Einsteinian gravity,  $g, g'$  ( $A_\mu^a$  and  $A_\mu$ ) are the  $SU(2)$  and  $U(1)$  gauge couplings (fields)  $\tau^a$  and  $Y$ , the corresponding generators, and  $\sigma^a$  the Pauli matrices.

As de Sitter space is the most important application of our results here we show the perturbative 1-loop correction for the SM in a spacetime with an equation of state  $w = -1$  i.e., a constant Hubble rate  $H$  (see section 3.2)

$$V_{\text{SM}}^{(1)}(\varphi, \mu) = \frac{1}{64\pi^2} \sum_{i=1}^{31} \left\{ n_i \mathcal{M}_i^4 \left[ \log \left( \frac{|\mathcal{M}_i^2|}{\mu^2} \right) - d_i \right] + n'_i H^4 \log \left( \frac{|\mathcal{M}_i^2|}{\mu^2} \right) \right\}, \quad (3.32)$$

where the sum is over all degrees of freedom of the SM, which may be found in **Tables 1, 2**. The masses are defined as

$$\begin{aligned} m_h^2 &= -m^2 + 3\lambda\varphi^2, & m_t^2 &= \frac{y_t^2}{2}\varphi^2, & m_W^2 &= \frac{g^2}{4}\varphi^2, \\ m_Z^2 &= \frac{g^2 + (g')^2}{4}\varphi^2, & m_\chi^2 &= -m^2 + \lambda\varphi^2. \end{aligned} \quad (3.33)$$

and the  $\zeta_i$  are the gauge fixing parameters.

The flat space beta functions have of course been known for some time, see for example Ford et al. (1993) and Buttazzo et al. (2013). The complete set of SM beta functions to 1-loop order was however first calculated only in Markkanen et al. (2018). The 1-loop SM beta functions for couplings associated with gravity

coming from the action (3.26),  $\xi, V_\Lambda, \kappa, \alpha_1, \alpha_2$  and  $\alpha_3$  can be solved from Equation (3.32) and read

$$16\pi^2\beta_\xi = \left( \xi - \frac{1}{6} \right) \left[ 12\lambda + 2Y_2 - \frac{3(g')^2}{2} - \frac{9g^2}{2} \right], \quad (3.34)$$

$$16\pi^2\beta_{V_\Lambda} = 2m^4, \quad (3.35)$$

$$16\pi^2\beta_\kappa = 4m^2 \left( \xi - \frac{1}{6} \right), \quad (3.36)$$

$$16\pi^2\beta_{\alpha_1} = 2\xi^2 - \frac{2\xi}{3} - \frac{277}{144}, \quad (3.37)$$

$$16\pi^2\beta_{\alpha_2} = \frac{571}{90}, \quad (3.38)$$

$$16\pi^2\beta_{\alpha_3} = -\frac{293}{720}, \quad (3.39)$$

where

$$Y_2 \equiv 3(y_u^2 + y_c^2 + y_t^2) + 3(y_d^2 + y_s^2 + y_b^2) + (y_e^2 + y_\mu^2 + y_\tau^2), \quad (3.40)$$

with  $y_i$  being a Yukawa coupling for a fermion type  $i$ .

Much like in the flat space case in Equation (2.19) we can write the RGI effective potential by choosing an optimized scale  $\mu_*(\varphi, R)$  in such a way that the loop correction is small (Markkanen et al., 2018). In curved space in addition to the Lagrangian from Equation (3.30) we must include the purely gravitational terms from Equation (3.26) in addition to the one loop contributions (3.32) giving rise to

$$\begin{aligned} V_{\text{RGI}}^{\text{SM}}(\varphi) &= \frac{\xi(\mu_*)}{2} R\varphi^2 + \frac{\lambda(\mu_*)}{4} \varphi^4 + \alpha_1(\mu_*) R^2 \\ &+ \alpha_2(\mu_*) R_{\mu\nu} R^{\mu\nu} + \alpha_3(\mu_*) R_{\mu\nu\delta\eta} R^{\mu\nu\delta\eta} \\ &+ V_{\text{RGI}}^{(1),\text{SM}}(\varphi, \mu_*), \end{aligned} \quad (3.41)$$

where  $\mu_*$  generally depends on both  $\varphi$  and  $R$ , and we have assumed  $|R| \gg |m_h^2|$ , which is usually true for the SM Higgs in the early Universe.

**TABLE 1 |** The 1-loop effective potential (3.32) contributions with tree-level couplings to the Higgs.

$\Psi$	$i$	$n_i$	$d_i$	$n'_i$	$\mathcal{M}_i^2$
$W^\pm$	1	2	3/2	-34/15	$m_W^2 + H^2$
	2	6	5/6	-34/5	$m_W^2 + H^2$
	3	-2	3/2	4/15	$m_W^2 - 2H^2$
$Z^0$	4	1	3/2	-17/15	$m_Z^2 + H^2$
	5	3	5/6	-17/5	$m_Z^2 + H^2$
	6	-1	3/2	2/15	$m_Z^2 - 2H^2$
$q$	7–12	-12	3/2	38/5	$m_q^2 + H^2$
$l$	13–15	-4	3/2	38/15	$m_l^2 + H^2$
$h$	16	1	3/2	-2/15	$m_h^2 + 12(\xi - 1/6)H^2$
$\chi_W$	17	2	3/2	-4/15	$m_\chi^2 + \zeta_W m_W^2 + 12(\xi - 1/6)H^2$
$\chi_Z$	18	1	3/2	-2/15	$m_\chi^2 + \zeta_Z m_Z^2 + 12(\xi - 1/6)H^2$
$c_W$	19	-2	3/2	4/15	$\zeta_W m_W^2 - 2H^2$
$c_Z$	20	-1	3/2	2/15	$\zeta_Z m_Z^2 - 2H^2$

$\Psi$  stands for  $W^\pm$  and  $Z^0$  bosons, the 6 quarks  $q$ , the 3 charged leptons  $l$ , the Higgs  $h$ . The Goldstone bosons are  $\chi_W$  and  $\chi_Z$  and ghosts  $c_W$  and  $c_Z$ . The masses may be found in Equation (3.33).

**TABLE 2 |** Contributions to the effective potential (3.32) with no coupling to the Higgs at tree-level.

$\Psi$	$i$	$n_i$	$d_i$	$n'_i$	$\mathcal{M}_i^2$
$\gamma$	21	1	3/2	-17/15	$H^2$
	22	3	5/6	-17/5	$H^2$
	23	-1	3/2	2/15	$-2H^2$
$g$	24	8	3/2	-136/15	$H^2$
	25	24	5/6	-136/5	$H^2$
	26	-8	3/2	16/15	$-2H^2$
$\nu$	27–29	-2	3/2	19/15	$H^2$
$c_\gamma$	30	-1	3/2	2/15	$-2H^2$
$c_g$	31	-8	3/2	16/15	$-2H^2$

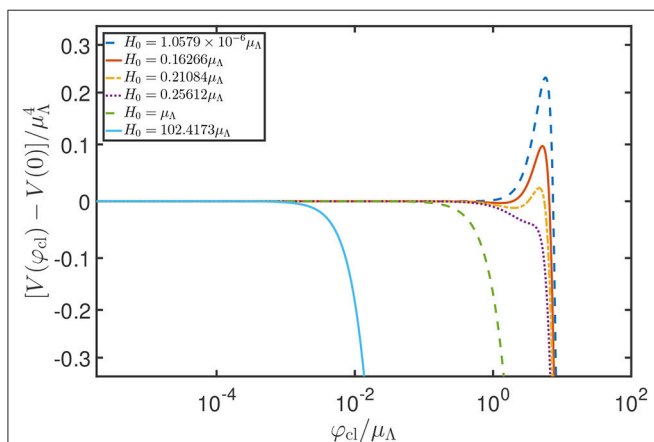
The  $\Psi$  include the photon  $\gamma$ , the 8 gluons  $g$ , the 3 neutrinos  $\nu$  and the respective ghosts  $c_\gamma$  and  $c_g$ .

When the Hubble rate is above electroweak scales it is quite obvious that the highly non-trivial curvature dependence apparent in Equation (3.32) and also in Equation (3.41) with the optimized scale (3.28) cannot be neglected: it is just as, if not more, important as what would have been obtained by using only a flat space derivation. The most obvious difference is the emergence of the direct non-minimal coupling between the Higgs and the scalar curvature  $R$ . Due to the curvature dependence of the optimized renormalization scale in curved space (3.28), which can be traced back to the curvature dependence of the one-loop correction (3.32), the generation of the non-minimal coupling in the current cosmological paradigm is unavoidable. It will be sourced by the changing Hubble rate  $H$ . Furthermore, as can be read from the beta function (3.34),  $\xi = 0$  is not a fixed point of the RG flow. Depending on the sign of  $\xi R$ , the non-minimal coupling can have a stabilizing or destabilizing effect, which can be very significant in the early Universe.

In **Figure 4** we illustrate the behavior of effective potential for the full SM in de Sitter space including the one loop quantum correction (3.32). We have chosen to set the renormalized non-minimal coupling  $\xi$  to zero at the electroweak scale. We use the field renormalized at the physical top mass

$$\varphi_{\text{cl}} \equiv \varphi(M_t) = \sqrt{\frac{Z(\mu_*)}{Z(M_t)}} \varphi(\mu_*), \quad (3.42)$$

as the x-axis. It is clearly evident that in curved space the potential may have drastically different predictions to flat space. As can be read off from the beta function for  $\xi$  (3.34), if  $\xi = 0$  at some low scale, it will run to negative values at high scales. Furthermore, since in de Sitter space  $R = 12H^2 > 0$ , a negative  $\xi$  can prevent the emergence of a potential barrier, even if robustly present on a flat background, as visible in **Figure 4**.



**FIGURE 4 |** The one-loop RGI effective potential for the full SM in de Sitter space with  $\xi = 0$  at the electroweak scale, in units of the instability scale (2.25), using the optimized scale choice (3.28). The x axis is given by the field renormalized at the physical top mass,  $\varphi_{\text{cl}} \equiv \varphi(M_t)$ . The disappearance of the potential barrier at large Hubble rates can be traced back to the RG running of the non-minimal coupling  $\xi$ . Figure taken from Markkanen et al. (2018).

## 4. VACUUM DECAY

### 4.1. Quantum Tunneling and Bubble Nucleation

The main mechanism behind vacuum decay in the Standard Model is essentially a direct extension of ordinary quantum tunneling to quantum field theories. In ordinary quantum mechanics, the wave-function for particles trapped by a potential barrier can penetrate the classically forbidden region of the barrier, leading to a non-zero probability to be found on the other side. The transition rate for particles of energy  $E$  incident on a barrier described by potential  $W(x)$  can be estimated using the WKB method (Coleman, 1985),

$$T = \exp \left( -2 \int_{x_1(E)}^{x_2(E)} dx \sqrt{2(W(x) - E)} \right), \quad (4.1)$$

where  $x_1, x_2$  are the turning points of the potential. As is clear from this expression, the tunneling rate is suppressed by wide and tall barriers.

Although Equation (4.1) can in principle be evaluated directly, we will follow a different approach that readily generalizes to quantum field theories (Coleman, 1977; Brown and Weinberg, 2007). The idea is to use the equation of motion,

$$\frac{d^2 x}{dt^2} = -W'(x) \rightarrow \frac{1}{2} \left( \frac{dx}{dt} \right)^2 + W(x) = E. \quad (4.2)$$

The region  $(x_1, x_2)$  is classically forbidden, since  $W(x) - E > 0$  there. We can apply a trick, however, by analytically continuing time to an imaginary value:  $\tau = it$ , which gives a *Euclidean* equation of motion,

$$\frac{d^2 x}{d\tau^2} = +W'(x) \Rightarrow \frac{1}{2} \left( \frac{dx}{d\tau} \right)^2 - W(x) = -E. \quad (4.3)$$

The most notable feature of these equations is that the potential has effectively been inverted. This means that we can find a classical solution that rolls through the barrier between the turning points  $x_1$  and  $x_2$ . If we can find this solution, it allows us to re-express the integral in Equation (4.1) as

$$\begin{aligned} 2 \int_{x_1}^{x_2} dx \frac{dx}{d\tau} &= 2 \int_{\tau_1}^{\tau_2} d\tau \left( \frac{dx}{d\tau} \right)^2 \\ &= 2 \int_{\tau_1}^{\tau_2} d\tau \left[ \frac{1}{2} \left( \frac{dx}{d\tau} \right)^2 + W(x) - E \right] \\ &= S_E[x_B(\tau)] - S_E[x_{\text{fv}}(\tau)], \end{aligned} \quad (4.4)$$

where  $S_E$  is the *Euclidean* action corresponding to Equation (4.3)

$$S_E[x(\tau)] = \int d\tau \left[ \frac{1}{2} \left( \frac{dx}{d\tau} \right)^2 + W(x) \right], \quad (4.5)$$

while  $x_B(\tau)$  is a bounce solution of the Euclidean equations of motion satisfying  $x'(\tau_1) = x'(\tau_2) = 0$ , and  $x_{\text{fv}}(\tau)$  is a constant

solution, sitting in the false vacuum with energy  $E$ . The “bounce” solution is so named because we see, by energy conservation, that it starts at  $x_1$ , rolls down the inverted potential before “bouncing” off  $x_2$  and rolling back. By finding this solution and evaluating its action, we can compute the rate for tunneling through a barrier.

This argument generalized straightforwardly to many-body quantum systems, where we use the action

$$S_E[q_i(\tau)] = \int d\tau \left[ \sum_i \frac{1}{2} \left( \frac{dq_i}{d\tau} \right)^2 + W(q_i) \right]. \quad (4.6)$$

With more than one degree of freedom, however, there are actually an infinite number of paths that  $q_i(\tau)$  could take when passing through the barrier, corresponding to an infinite number of solutions. However, since the decay rate is exponentially dependent on the action,  $\Gamma \propto e^{-S_E[q_i]}$ , it is clear that only the solution with smallest Euclidean action will contribute significantly, as this will dominate the decay rate (in other words, the tunneling takes the “path of least resistance”).

The generalization from a many body system,  $q_i$ , to a quantum field theory with scalar field  $\varphi(x)$  is then straightforward,

$$S_E[\varphi(x)] = \int d^4x \left[ \frac{1}{2} \partial_\mu \varphi \partial^\mu \varphi + V(\varphi) \right]. \quad (4.7)$$

The integral here is over flat four-dimensional Euclidean space, and note that the opposite sign of the potential leads to an opposing sign in the equations of motion,

$$-\nabla_\mu \nabla^\mu \varphi + V'(\varphi) = 0. \quad (4.8)$$

Although it is tempting to interpret  $V(\phi)$  as the potential to be tunneled through, this is only somewhat true. The analog of  $W(q_i)$  in Equation (4.6) is a functional of the field configuration  $\varphi(x)$ , given by an integral over three-dimensional space,

$$U[\varphi(x)] = \int d^3x \left[ \frac{1}{2} (\nabla \varphi)^2 + V(\varphi) \right], \quad (4.9)$$

where  $\nabla \varphi$  represents the *spatial* derivative of the field. In the analogy with quantum mechanics, this term should be considered part of the potential, as its many body equivalent is a nearest-neighbor interaction between adjacent degrees of freedom,  $q_i, q_{i\pm 1}$ . This means, in particular, while in quantum mechanics, the particle emerges after tunneling at a point  $x_2$  that has the same potential energy,  $W(x_1) = W(x_2)$ , in quantum field theory, the field emerges *lower* down the potential  $V$ .

In a field theory, the analog of  $x_2$  is a field configuration,  $\varphi(x)$ , given by slicing the bounce solution at its mid-way point. This is a nucleated “true-vacuum” bubble, whose decay rate is determined by the Euclidean action of the bounce solution,  $\varphi_B$ . As we will see in section 4.7, the dominant Euclidean solutions have  $O(4)$  symmetry, which means that the bubble nucleates with  $O(3,1)$  symmetry. This causes it to expand at near the speed of light, resulting in the space around a nucleation point being converted to the true vacuum, releasing energy into the bubble wall. Apart from the destruction that this would unleash, and the

different masses of fundamental particles in the bubble interior, the result is also gravitational collapse of the bubble (Coleman and De Luccia, 1980), making its nucleation in our past light-cone completely incompatible with the trivial observation that the vacuum has not decayed (yet).

In cosmological applications, but also other areas, it is also important to consider the effect of thermally induced fluctuations over the barrier. Brown and Weinberg (2007) describe how thermal effects can be included in the above argument. At non-zero temperature, we must integrate over the possible excited states, and the decay exponent which depends on energy,

$$T \propto \int dE e^{-\beta E} e^{-B(E)}, \quad (4.10)$$

where  $B(E)$  is the (energy dependent) difference in Euclidean action between the bounce solution and the excited state of energy  $E$ . This integral is dominated by the energy that minimizes the exponent  $\beta E + B(E)$ , which is easily shown to satisfy

$$\beta = 2(\tau_2(E) - \tau_1(E)), \quad (4.11)$$

where  $\tau_1, \tau_2$  are the initial and final values in imaginary time of the (energy dependent) bounce solution. In other words, the bounce solution is periodic in imaginary time, with period controlled by the temperature.

In quantum field theory, the decay rate per unit volume and time of a metastable vacuum decays was first discussed by Coleman (Coleman, 1977; Callan and Coleman, 1977), and is given by

$$\Gamma = A \exp(-B), \quad A = \left( \frac{B}{2\pi} \right)^2 \left| \frac{\det'(S''[\varphi_B])}{\det(S''[\varphi_{fv}])} \right|^{-\frac{1}{2}}, \quad (4.12)$$

where

$$B = S[\varphi_B] - S[\varphi_{fv}] \quad (4.13)$$

is the difference between the Euclidean action of a so called bounce solution  $\varphi_B$  of the Euclidean (Wick rotated) equations of motion, and the action of the constant solution  $\varphi_{fv}$  which sits in the false vacuum.  $S''$  denotes the second functional derivative of the Euclidean action of a given solution, and  $\det'$  denotes the functional determinant after extracting the four zero-mode fluctuations which correspond to translations of the bounce (these are responsible for the formula giving a decay rate per unit volume). Precise calculations of the pre-factor  $A$  in the Standard Model were performed in Isidori et al. (2001), and involve computing the fluctuations around the bounce solution of all fields that couple to the Higgs. This requires renormalizing the loop corrections, and also to avoid double-counting, expanding around the tree-level bounce, rather than the bounce in the loop corrected potential.

In the gravitational case, the prefactor  $A$  is harder to compute. The main issue is that it includes both Higgs and *gravitational* fluctuations, and without a way of renormalizing the resulting graviton loops, the calculation becomes much harder. Various attempts have been made to do this using the fluctuations

discussed in section 4.5 (see Dunne and Wang, 2006; Lee and Weinberg, 2014; Koehn et al., 2015 for example), but a full description, especially for the Standard Model case, is not yet available.

In most cases, it is reasonable to estimate the prefactor  $A$  using dimensional analysis. Because  $A$  has dimension four, one would expect

$$A \sim \mu^4, \quad (4.14)$$

where  $\mu$  the characteristic energy scale of the instanton solution. Due to the exponential dependence on the decay exponent,  $B$ , this will not lead to large errors, and therefore we will use this result in the absence of more accurate estimates.

## 4.2. Asymptotically Flat Spacetime at Zero Temperature

In flat Minkowski space, the bounce solution corresponds to a saddle point of the Euclidean action,

$$S[\varphi] = \int d^4x \left[ \frac{1}{2} \partial_\mu \varphi \partial^\mu \varphi + V(\varphi) \right], \quad (4.15)$$

with one negative eigenvalue (see section 4.5). Since Equation (4.12) depends exponentially on the bounce action, only the lowest action bounce solutions will contribute. In flat space, it is always the case that the lowest action solution has  $O(4)$  symmetry (Coleman et al., 1978). This means that the equations of motion for the bounce can be reduced to

$$\ddot{\varphi} + \frac{3}{r} \dot{\varphi} - V'(\varphi) = 0, \quad (4.16)$$

subject to the boundary conditions  $\dot{\varphi}(0) = 0$  and  $\varphi(r \rightarrow \infty) \rightarrow \varphi_{\text{fv}}$ . These ensure that the bounce action is finite and thus gives non-zero contribution to the decay rate. There are always trivial solutions corresponding to the minima of the potential  $V(\varphi)$ , but they do not contribute to vacuum decay because they have no negative eigenvalues.

For example, in a theory with a constant negative quartic coupling, that is,

$$V(\varphi) = -|\lambda| \frac{\varphi^4}{4}, \quad (4.17)$$

there exists the Lee-Weinberg or Fubini bounce (Fubini, 1976; Lee and Weinberg, 1986). This is a solution of the form:

$$\varphi_{\text{LW}}(r) = \sqrt{\frac{2}{|\lambda|}} \frac{2r_B}{r_B^2 + r^2}, \quad (4.18)$$

where the arbitrary parameter  $r_B$  characterizes the size of the bounce (and thus the nucleated bubble). This arbitrary parameter appears in the theory because the potential Equation (4.17) is conformally invariant, and thus bounces of all scales contribute equally with action

$$S[\varphi_{\text{LW}}] = \frac{8\pi^2}{3|\lambda|}. \quad (4.19)$$

In fact, similar bounces contribute approximately in the Standard Model, where the running of the couplings breaks this approximate conformal symmetry, so that bounces of order the scale at which  $\lambda$  is most negative (which is the minimum of the  $\lambda(\mu)$  running curve) dominate the decay rate (Isidori et al., 2001).

The complete calculation would also include gravity, and would therefore involve finding the corresponding saddle point of the action

$$S[\varphi, g_{\mu\nu}] = \int d^4x \left[ \frac{1}{2} \nabla_\mu \varphi \nabla^\mu \varphi + V(\varphi) - \frac{M_{\text{P}}^2}{2} R \right], \quad (4.20)$$

where  $R$  is the Ricci scalar. The leading gravitational correction to Equation (4.19) is Isidori et al. (2008)

$$\Delta S_{\text{gravity}} = \frac{256\pi^3}{45(r_B M_{\text{P}} \lambda)^2}. \quad (4.21)$$

Another approach is to solve the bounce equations numerically, which makes it possible to use the exact field and Einstein equations and the full effective potential. The difference is a second order correction (Isidori et al., 2008). Using the tree-level RGI effective potential (2.23), the full numerical result including gravitational effects for  $M_t = 173.34$  GeV,  $M_h = 125.15$  GeV,  $\alpha_s(M_z) = 0.1184$  and minimal coupling  $\xi = 0$  is Rajantie and Stopyra (2017)

$$B_{\text{grav}} = 1808.3. \quad (4.22)$$

A non-minimal value of the Higgs curvature coupling  $\xi$  changes the action and the shape of the bounce solution (and thus the scale that dominates tunneling) (Isidori et al., 2008; Czerwinska et al., 2016; Rajantie and Stopyra, 2017; Salvio et al., 2016; Czerwinska et al., 2017). **Figure 5** shows the bounce action  $B$  as a function of  $\xi$ , computed numerically in Rajantie and Stopyra (2017). As the plot shows, the action is smallest near the conformal value  $\xi = 1/6$ . For  $\xi \approx 1/6$ , the result agrees well with the perturbative calculation (Salvio et al., 2016),

$$\Delta S_{\text{gravity}} = \frac{32\pi^2(1 - 6\xi)^2}{45(r_B M_{\text{P}} \lambda)^2}. \quad (4.23)$$

For comparison, for the same parameters, the numerically computed decay exponent in flat space is (Rajantie and Stopyra, 2017)

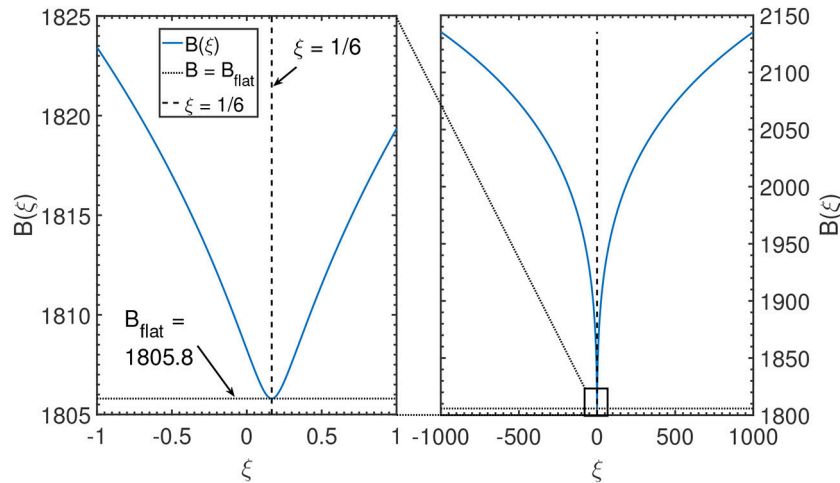
$$B_{\text{flat}} = 1805.8, \quad (4.24)$$

which is very close to the full gravitational result with the conformal coupling  $\xi = 1/6$ . The analytical approximation (4.19) using  $\mu_{\text{min}} = 2.79 \times 10^{17}$  GeV gives

$$S[\varphi_{\text{LW}}] = 1804.5. \quad (4.25)$$

Calculations of the prefactor  $A$  show that the decay rate (4.12) is well approximated by Isidori et al. (2001)

$$\Gamma \sim \mu_{\text{min}}^4 e^{-B} \sim 10^{-716} \text{ GeV}^4, \quad (4.26)$$



**FIGURE 5** | Plot of the decay rate for a flat false vacuum for different values of the non-minimal coupling,  $\xi$ . The minimal action is obtained close to the conformal value  $\xi = 1/6$ , and agrees well with the flat space result (4.24). Originally published in Rajantie and Stopyra (2017).

where the numerical value corresponds to the action (4.22). This agrees with the estimate from dimensional analysis (4.14). Note, however, that the rate is very sensitive to the top quark and Higgs boson masses, and also to higher-dimensional operators (Branchina and Messina, 2013; Branchina et al., 2015).

The presence of a small black hole can catalyze vacuum decay and make it significantly faster (Gregory et al., 2014; Burda et al., 2015a,b, 2016; Tetradis, 2016). The action of the vacuum decay instanton in the presence of a seed black hole is given by

$$B = \frac{M_{\text{seed}}^2 - M_{\text{remnant}}^2}{2M_{\text{p}}^2}, \quad (4.27)$$

where  $M_{\text{seed}}$  and  $M_{\text{remnant}}$  are the masses of the seed black hole and the left over remnant black hole. For black holes of mass  $M_{\text{seed}} \lesssim 10^5 M_{\text{p}} \approx 1 \text{ g}$  the vacuum decay rate becomes unsuppressed. This can be interpreted (Tetradis, 2016; Mukaida and Yamada, 2017) as a thermal effect due to the black hole temperature  $T_{\text{seed}} = M_{\text{p}}^2/M_{\text{seed}}$ . The catalysis of vacuum decay does not necessarily rule out cosmological scenarios with primordial black holes, because positive values of non-minimal coupling  $\xi$  would suppress the vacuum decay in the presence of a black hole (Canko et al., 2018).

### 4.3. Non-zero Temperature

The presence of a heat bath with non-zero temperature has a significant impact on the vacuum decay rate  $\Gamma$  (Anderson, 1990; Arnold and Vokos, 1991). On one hand, the thermal bath modifies the effective potential of the Higgs field, and on the other hand, as discussed in section 4.1, it modifies the process itself because it can start from an excited state rather than the vacuum state.

At one-loop level, the finite-temperature effective potential can be written as Arnold and Vokos (1991)

$$V_{\text{eff}}(T, \varphi) = V_{\text{eff}}(\varphi) + T \sum_i n_i \int \frac{d^3 k}{(2\pi)^3} \ln \left[ 1 \mp e^{-\sqrt{k^2 + \mathcal{M}_i^2}/T} \right], \quad (4.28)$$

where  $n_i$  and  $\mathcal{M}_i^2$  are given in Table 1 (taking  $H = 0$ ). In the high-temperature limit,  $T \gg M_h$ , this can be approximated by

$$V_{\text{eff}}(T, \varphi) \approx V_{\text{eff}}(0, \varphi) + \frac{1}{2} \gamma^2 T^2 \varphi^2, \quad (4.29)$$

where

$$\gamma^2 \approx \frac{1}{12} \left( \frac{3}{4} g^2 + \frac{9}{4} g'^2 + 3y_t^2 + 6\lambda \right). \quad (4.30)$$

Therefore the thermal fluctuations give rise to a positive contribution to the quadratic term. This raises the height of the potential barrier, and therefore would appear to suppress the decay rate.

At non-zero temperatures the decay process is described by a periodic instanton solution with period  $\beta$  in the Euclidean time direction. In the high-temperature limit, the solution becomes independent of the Euclidean time, and has the interpretation of a classical sphaleron configuration. The instanton action is therefore given by

$$B(T) = E_{\text{sph}}(T)/T, \quad (4.31)$$

where  $E_{\text{sph}}$  is the energy of the sphaleron, which is the three-dimensional saddle point configuration analogous to the Coleman bounce (4.16), and satisfies the equation

$$\ddot{\varphi} + \frac{2}{r} \dot{\varphi} - V'(\varphi) = 0. \quad (4.32)$$

Using the approximation of constant negative  $\lambda$ , the action is Arnold and Vokos (1991)

$$B(T) = \frac{E_{\text{sph}}(T)}{T} \approx 18.9 \frac{\gamma}{|\lambda|}. \quad (4.33)$$

Because  $\gamma \ll 1$ , this is smaller than the zero-temperature action (4.19). Therefore the net effect of the non-zero temperature is to increase the vacuum decay rate compared to the zero-temperature case.

More accurately, the sphaleron solutions have been calculated numerically in Delle Rose et al. (2016) and Salvio et al. (2016). At high temperatures  $T \gtrsim 10^{16}$  GeV, the action is roughly

$$B(T \gtrsim 10^{16} \text{ GeV}) \sim 300. \quad (4.34)$$

When the temperature decreases, the action increases, so that  $B(10^{14} \text{ GeV}) \sim 400$ .

Salvio et al. (2016) obtained fully four-dimensional instanton solutions numerically, without assuming independence on the Euclidean time, and found that the three-dimensional sphaleron solutions have always the lowest action and are therefore the dominant solutions. They also showed that including the two-loop corrections to the quadratic term (4.30) or the one-loop correction to the Higgs kinetic term gives only small correction to the action.

Taking also the prefactor into account, the vacuum decay rate at non-zero temperature is (Espinosa et al., 2008; Delle Rose et al., 2016)

$$\Gamma(T) \approx T^4 \left( \frac{B(T)}{2\pi} \right)^{3/2} e^{-B(T)}. \quad (4.35)$$

#### 4.4. Vacuum Decay in de Sitter Space

In extending from flat space to curved space, the theorem (Coleman et al., 1978) that guarantees  $O(4)$  symmetry of the bounce no longer applies. There is some evidence, however, that in background metrics that do respect this symmetry,  $O(4)$  symmetric solutions should still dominate (Masoumi and Weinberg, 2012). This would include the special case of particular interest in this review - an inflationary, or de Sitter background<sup>3</sup>. A Wick rotated metric can be placed in a co-ordinate system that makes the  $O(4)$  symmetry of the bounce immediately manifest,

$$ds^2 = d\chi^2 + a^2(\chi) d\Omega_3^2, \quad (4.36)$$

where  $\chi$  is a radial variable,  $d\Omega_3^2$  is the 3-sphere metric, and  $a^2(\chi)$  is a scale factor that physically describes the radius of curvature of a surface at constant  $\chi$ . The bounce equations of motion then take the form (Coleman and De Luccia, 1980)

$$\ddot{\varphi} + \frac{3\dot{a}}{a} \dot{\varphi} - V'(\varphi) = 0 \quad (4.37)$$

$$\dot{a}^2 = 1 - \frac{a^2}{3M_{\text{P}}^2} \left( -\frac{\dot{\varphi}^2}{2} + V(\varphi) \right). \quad (4.38)$$

<sup>3</sup>In principle, inflation is not exact de Sitter, and so the background does not respect exact  $O(4)$  symmetry if Euclideanised, but for slow roll inflation models, it is a reasonable approximation to make.

We will consider the case in which the false vacuum has a positive energy density,  $V(\varphi_{\text{fv}}) > 0$ , and therefore non-zero Hubble rate

$$H^2 = \frac{V(\varphi_{\text{fv}})}{3M_{\text{P}}^2}. \quad (4.39)$$

The boundary conditions the bounce solution must satisfy require special attention:  $a(0) = 0$  is required because of the definition of  $a(\chi)$  as a radius of curvature of a surface of constant  $\chi$ , while we require  $\dot{\varphi}(0) = \dot{\varphi}(\chi_{\text{max}}) = 0$ , where  $\chi_{\text{max}} > 0$  is defined by  $a(\chi_{\text{max}}) = 0$ . These boundary conditions avoid the co-ordinate singularities at  $\chi = 0, \chi_{\text{max}}$  giving infinite results, but allow for the peculiar property that the bounces are compact, and do not approach the false vacuum anywhere.

One way of understanding this peculiar feature was discussed by Brown and Weinberg (2007). They considered vacuum decay in de Sitter space, specifically the static patch co-ordinates where the metric takes the form

$$dS_n^2 = -(1 - H^2 r^2) dt^2 + (1 - H^2 r^2)^{-1} dr^2 + r^2 d\Omega_{n-2}^2, \quad (4.40)$$

where  $d\Omega_{n-2}^2$  is the  $n - 2$ -sphere metric (in this case,  $n = 4$ ). The important feature of these co-ordinates is that they are valid only up to the horizon at  $r = 1/H$ . The Euclidean action can then be re-written as

$$S_E = \int_{-\frac{\pi}{H}}^{\frac{\pi}{H}} d\tau \int d^3x \sqrt{\det h} \left[ \frac{1}{2} (1 - H^2 r^2)^{-\frac{1}{2}} \left( \frac{d\varphi}{d\tau} \right)^2 + \frac{1}{2} (1 - H^2 r^2)^{\frac{1}{2}} h^{ij} \partial_i \varphi \partial_j \varphi + (1 - H^2 r^2)^{\frac{1}{2}} V(\varphi) \right], \quad (4.41)$$

where  $h_{ij}$  is the remaining spatial metric. Brown and Weinberg interpreted this to mean that tunneling takes place on a *compact* Euclidean space, with a curved three-dimensional geometry. This compactness condition is reflected in the boundary conditions  $\dot{\varphi}(0) = \dot{\varphi}(\chi_{\text{max}})$ , which inevitably produce a compact bounce solution. They observed that the same effect would be seen in considering a spatially curved universe with this same spatial geometry, but with a non-zero temperature,

$$T_{\text{GH}} = \frac{H}{2\pi}. \quad (4.42)$$

This corresponds to the Gibbons-Hawking temperature of de Sitter space (Gibbons and Hawking, 1977), and implies that bounces in de Sitter space may have a thermal interpretation.

The simplest solution of Equations (4.37) and (4.38) is the Hawking-Moss solution (Hawking and Moss, 1982). This is a constant solution, for which  $\varphi = \varphi_{\text{bar}}$  sits at the top of the barrier for the entire Euclidean period, and the scale factor is given by

$$a(\chi) = \frac{1}{H_{\text{HM}}} \sin(H_{\text{HM}} \chi), \quad H_{\text{HM}}^2 = \frac{V(\varphi_{\text{bar}})}{3M_{\text{P}}^2}. \quad (4.43)$$

Hence  $\chi_{\text{max}} = \pi/H_{\text{HM}}$ . The action difference of Equation (4.13) is then easily computed analytically to be

$$B_{\text{HM}} = 24\pi^2 M_{\text{P}}^4 \left( \frac{1}{V(\varphi_{\text{fv}})} - \frac{1}{V(\varphi_{\text{bar}})} \right). \quad (4.44)$$

A particularly important limit is that in which  $\Delta V(\varphi_{\text{bar}}) = V(\varphi_{\text{bar}}) - V(\varphi_{\text{fv}}) \ll V(\varphi_{\text{fv}})$ . In that case, Equation (4.44) is approximately

$$B_{\text{HM}} = \frac{8\pi^2 \Delta V(\varphi_{\text{bar}})}{3H^4}, \quad (4.45)$$

where  $H^2 = V(\varphi_{\text{fv}})/3M_{\text{p}}^2$  is the background Hubble rate. The prefactor (4.14) in the decay rate can be expected to be at the scale of the Hubble, and therefore the vacuum decay rate due to the Hawking-Moss instanton can be approximated by

$$\Gamma(H) \sim H^4 e^{-B_{\text{HM}}(H)} \quad (4.46)$$

Equation (4.45) has a simple thermal interpretation: It is the ratio of the energy required to excite an entire Hubble volume,  $4\pi/3H^3$  from the false vacuum to the top of the barrier, divided by the background Gibbons-Hawking temperature (4.42). Therefore it can be understood as Boltzmann suppression in classical statistical physics.

The bounce equations (4.37) and (4.38) also often have Coleman-de Luccia (CdL) instantons, in which the field increases monotonically from  $\varphi(0) < \varphi_{\text{bar}}$  to  $\varphi(\chi_{\text{min}}) > \varphi_{\text{bar}}$ . For low false vacuum Hubble rates,  $H \ll \mu_{\text{min}}$ , a CdL solution can be found as a perturbative correction to Equation (4.18), with the action (Shkerin and Sibiryakov, 2015)

$$B_{\text{CdL}} \approx \frac{8\pi^2}{3|\lambda(\mu_{\text{min}})|} \left[ 1 + 36 \left( \xi - \frac{1}{6} \right) \frac{H^2}{\mu_{\text{min}}^2} \ln \frac{\mu_{\text{min}}}{H} \right], \quad (4.47)$$

Numerical HM and CdL bounce solutions in the Standard Model were found in Rajantie and Stoparra (2018) and the corresponding actions are shown in **Figure 6**, for the parameters  $M_h = 125.15 \text{ GeV}$ ,  $M_t = 173.34 \text{ GeV}$ ,  $\alpha_s = 0.1184$ . We can see that at low Hubble rates, the CdL solution has a lower action than the HM solution. For example, for the case of background Hubble rate  $H = 1.1937 \times 10^8 \text{ GeV}$ , the numerical result is  $B_{\text{CdL}} = 1805.8$  in a fixed de Sitter background metric, and  $B_{\text{CdL}} = 1808.26$  including gravitational back-reaction. The CdL action is also almost independent of the Hubble rate.

On the other hand, the Hawking-Moss action (4.44) decreases rapidly as the Hubble rate increases. It crosses below  $B_{\text{CdL}}$  at Hubble rate (Rajantie and Stoparra, 2018)

$$H_{\text{cross}} = 1.931 \times 10^8 \text{ GeV}. \quad (4.48)$$

At Hubble rates below this,  $H > H_{\text{cross}}$  vacuum decay is dominated by the Coleman-de Luccia instanton, which describes quantum tunneling through the potential barriers, whereas above this,  $H > H_{\text{cross}}$ , the dominant process is the Hawking-Moss instanton. This is discussed further in section 4.6.

In addition to the HM and CdL solutions, one may also find oscillating solutions (Hackworth and Weinberg, 2005; Weinberg, 2006; Lee et al., 2015, 2017), which cross the top of the barrier  $\varphi_{\text{bar}}$  multiple times between  $\chi = 0$  and  $\chi = \chi_{\text{max}}$ , and additional CdL-like solutions with higher action (Hackworth and Weinberg, 2005; Rajantie and Stoparra, 2018). The latter were

found numerically in the Standard Model in Rajantie and Stoparra (2018). Because these solutions have a higher action than the HM and CdL solutions, they are highly subdominant as vacuum decay channels. Oscillating solutions also have more than one negative eigenvalues (Dunne and Wang, 2006; Lavrelashvili, 2006).

## 4.5. Negative Eigenvalues

In order for a stationary point of the action to describe vacuum decay, it has to have precisely one negative eigenvalue. The reason is that the decay rate of a metastable vacuum is determined by the imaginary part of the energy as computed by the effective action (Callan and Coleman, 1977), and thus only solutions that contribute an imaginary part to the vacuum energy will contribute to metastability.

This requirement comes in via the functional determinant which encodes the quantum corrections to the bounce solution. This functional determinant is given by a product over the eigenvalues for fluctuations around the relevant bounce solution. In flat space, these all satisfy (Callan and Coleman, 1977)

$$-\nabla_\mu \nabla^\mu \delta\varphi + V''(\varphi_B) \delta\varphi = \lambda \delta\varphi, \quad (4.49)$$

where  $\varphi_B$  is the solution expanded around. The  $O(4)$  symmetric bounce solutions in flat space can be shown to have at least one negative eigenvalue, since they possess zero modes corresponding to translations of the bounce around the space-time. In fact, there must only be one such eigenvalue. Solutions with more negative eigenvalues do not contribute to tunneling rates, because while they are stationary points of the Euclidean action, they are not minima of the barrier penetration integral (4.1) obtained from the WKB approximation (Coleman, 1985).

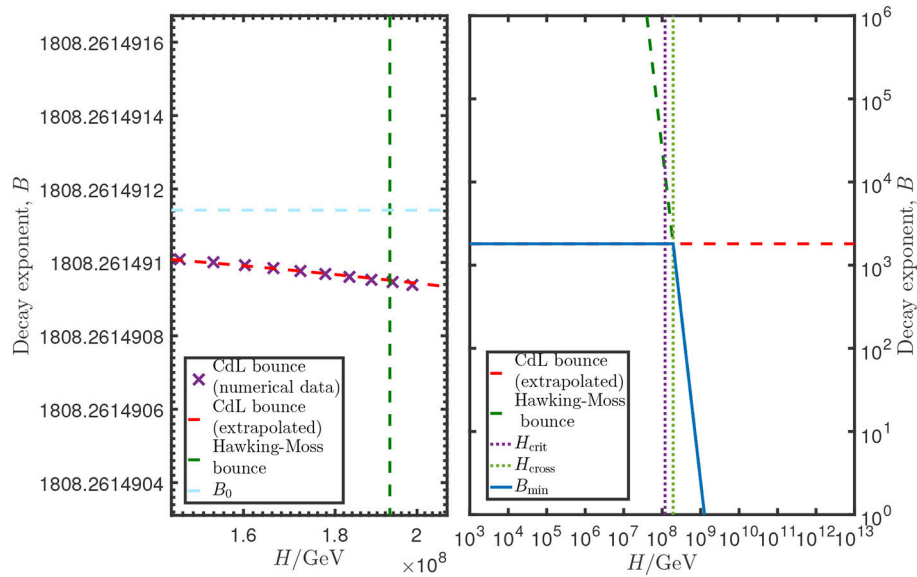
The situation is somewhat different in the gravitational case, however, due to the fact that in addition to the scalar field, we can also consider metric fluctuations about a bounce solution. A quadratic action for fluctuations about a bounce in curved space was first derived by Lavrelashvili et al. (1985) and has been considered by several authors (Lavrelashvili, 2006; Lee and Weinberg, 2014; Koehn et al., 2015). This takes the gauge invariant form

$$\mathcal{L}^{(2)}(\zeta_l, \dot{\zeta}_l) = \frac{a^3(\chi) (1 - \frac{1}{3}l(l+2))}{2(Q - \frac{1}{3}\dot{a}^2 l(l+2))} [\dot{\zeta}_l^2(\chi) + f(a, \phi) \zeta_l^2(\chi)], \quad (4.50)$$

where

$$Q = 1 - \frac{a^2(\chi) V(\varphi)}{3M_{\text{p}}^2}, \quad (4.51)$$

and  $f$  is a complicated function of  $a$  and  $\varphi$  which can be found in Lee and Weinberg (2014), Lavrelashvili (2006), and Koehn et al. (2015). The analysis of this Lagrangian is complicated, but some conclusions can be drawn. To begin with, it is possible to argue that expanded around a CdL bounce solution, this action *always* has an infinite number of negative eigenvalues. This is the so called “negative mode problem” (Lavrelashvili, 2006; Lee and Weinberg, 2014; Koehn et al., 2015). The argument, as expressed



**FIGURE 6 |** CdL bounce decay exponent plotted against the Hawking-Moss solution in the Standard Model with  $M_t = 173.34$  GeV,  $M_h = 125.15$  GeV,  $\alpha_S(M_Z) = 0.1184$ . The critical values  $H_{\text{crit}} = 1.193 \times 10^8$  GeV,  $H_{\text{cross}} = 1.931 \times 10^8$  GeV are also plotted, along with  $B_0$ , the bounce action obtained at  $H = 0$ .

in Lee and Weinberg (2014), is that we can re-write  $Q$  using Equation (4.38) as

$$Q = \dot{a}^2 - \frac{a^2 \dot{\varphi}^2}{3M_{\text{P}}^2}. \quad (4.52)$$

Note that the bounce always has a point satisfying  $\dot{a} = 0$ , which is the largest value obtained by  $a(\chi)$ . Consequently, there is always a region where  $Q$  is negative, so for the  $l = 0$  modes it is possible to construct a negative kinetic term in Equation (4.50). This means that sufficiently rapidly varying fluctuations will have their action unbounded below, so there is an infinite tower of high frequency, rapidly oscillating fluctuations that all have negative eigenvalues. Note that for  $l = 1$  the quadratic Lagrangian is zero (these are the zero-modes associated to translations of the bounce), and for  $l > 1$ , both numerator and denominator in Equation (4.50) are negative, thus the kinetic terms are always positive. Since  $Q = 1$  in flat space (obtained by taking the  $M_{\text{P}} \rightarrow \infty$  limit), it is clear that these “rapidly oscillating” modes are somehow associated to the gravitational sector.

At first, this seems concerning, however, it was pointed out in Lee and Weinberg (2014) that these high frequency oscillations are inherently associated with quantum gravity contributions, and thus may not affect tunneling. If we focus on the “slowly varying” modes, the structure of these is much more similar to the analogous flat space bounces. The conclusion we should draw then, is that a solution is relevant only if there is a single *slowly varying* negative eigenvalue.

## 4.6. Hawking-Moss/Coleman-de Luccia Transition

As discussed in section 4.4, there are two types of solutions that contribute to vacuum decay in de Sitter space. The first is the Hawking-Moss solution (4.43), and the second is the Coleman-de Luccia solution, which crosses the barrier once. By considering the negative eigenvalues of the HM solution, one gains insight into which solutions exist and contribute to vacuum decay at a given Hubble rate.

The eigenvalues of the Hawking-Moss solution are Lee and Weinberg (2014)

$$\lambda_N = \frac{V''(\varphi_{\text{bar}})}{H_{\text{HM}}^2} + N(N+3), \quad (4.53)$$

and their degeneracy is Rubin and Ordonez (1983)

$$D_N(4, 0) = \frac{(N+1)(N+2)(2N+3)}{6}. \quad (4.54)$$

Because  $V''(\varphi_{\text{bar}}) < 0$ , the  $N = 0$  mode is self evidently negative, and has degeneracy 1. Higher modes will all be positive if and only if

$$\lambda_1 = \frac{V''(\varphi_{\text{bar}})}{H_{\text{HM}}^2} + 4 > 0. \quad (4.55)$$

This imposes a lower bound on  $H_{\text{HM}}$ , below which the Hawking-Moss solution has multiple negative eigenvalues. Hence, it cannot contribute to vacuum decay for Hubble rates below the critical threshold (Coleman, 1985; Brown and Weinberg, 2007). An alternative way of expressing this is in terms of a critical Hubble rate. If we define  $H^2 = V(\varphi_{\text{FV}})/3M_{\text{P}}^2$  to be the background Hubble

rate in the false vacuum, then the condition for Hawking-Moss solutions to contribute to vacuum decay is  $H > H_{\text{crit}}$  where

$$H_{\text{crit}}^2 = -\frac{V''(\varphi_{\text{bar}})}{4} - \frac{\Delta V(\varphi_{\text{bar}})}{3M_{\text{P}}^2}. \quad (4.56)$$

Here,  $\Delta V(\varphi) \equiv V(\varphi) - V(\varphi_{\text{fv}})$ . However, the second term generally only contributes significantly if the difference in height between the top of the barrier and the false vacuum is comparable to the Planck Mass. For most potentials, only the second derivative at the top of the barrier matters.

At low Hubble rates,  $H < H_{\text{crit}}$ , the Hawking-Moss solution does not contribute to vacuum decay, but on the other hand, a CdL solution is *guaranteed* to exist (Balek and Demetrian, 2004). In most potentials, the CdL solution smoothly merges into the Hawking-Moss solution as the Hubble rate approached  $H_{\text{crit}}$  from below, and the Hawking-Moss solution becomes relevant (Balek and Demetrian, 2005; Hackworth and Weinberg, 2005). Close to the critical Hubble rate,  $H \sim H_{\text{crit}}$ , one can define the quantity (Tanaka and Sasaki, 1992; Balek and Demetrian, 2005; Joti et al., 2017)

$$\Delta \equiv -\frac{1}{14} \left[ V^{(4)}(\varphi_{\text{bar}}) - \frac{(V^{(3)}(\varphi_{\text{HM}}))^2}{3V^{(2)}(\varphi_{\text{bar}})} - \frac{8V^{(2)}(\varphi_{\text{HM}})}{3M_{\text{P}}^2} \right], \quad (4.57)$$

which divides potentials into two classes (Balek and Demetrian, 2005; Rajantie and Stopyra, 2018). Those with  $\Delta < 0$  are “typical” potentials, for which the perturbative solution only exists for  $H < H_{\text{crit}}$  (Balek and Demetrian, 2005), while those with  $\Delta > 0$  only have perturbative solutions for  $H > H_{\text{crit}}$ . When a perturbative solution exists, its action is given by Balek and Demetrian (2005)

$$B_{\text{CdL}} = B_{\text{HM}} + \frac{2\pi^2(\varphi_0 - \varphi_{\text{HM}})^4 \Delta}{15H_{\text{HM}}^4}, \quad (4.58)$$

where  $\varphi_0$  is the true vacuum side value of the bounce (which approaches  $\varphi_{\text{HM}}$  in the  $H \rightarrow H_{\text{crit}}$  limit) and  $\varphi_{\text{bar}}$  is the top of the barrier.

Hence one can see that if  $\Delta < 0$ , a CdL solution with lower action,  $B_{\text{CdL}} < B_{\text{HM}}$ , exists for  $H < H_{\text{crit}}$ , and approaches the Hawking-Moss solution smoothly as  $H \rightarrow H_{\text{crit}}$ , until it vanishes at  $H_{\text{crit}}$ . At the same point, the second eigenvalue of the HM solution turns positive, and therefore the HM solution starts to contribute to vacuum decay.

On the other hand, if  $\Delta > 0$ , which is the case for the Standard Model Higgs potential (Rajantie and Stopyra, 2018), the perturbative CdL solution exists only for  $H > H_{\text{crit}}$ . Below  $H_{\text{crit}}$ , the HM solution has two negative eigenvalues, which means that it does not contribute to vacuum decay. Instead, the relevant solution is the CdL solution, which also has a lower action (see **Figure 6**). When the Hubble rate is increased, a second, perturbative CdL solution appears smoothly at  $H = H_{\text{crit}}$ , at the same as the second eigenvalue of the HM solution becomes positive. At  $H > H_{\text{crit}}$  there are, therefore, at least two distinct CdL solutions, and in fact, numerical calculations indicate that there are at least four (Rajantie and Stopyra, 2018). For the parameters used in **Figure 6**, the critical Hubble rate is  $H_{\text{crit}} = 1.193 \times 10^8 \text{ GeV}$ .

## 4.7. Evolution of Bubbles After Nucleation

The bounce solution  $\varphi_B$  determines the field configuration to which the vacuum state tunnels (Callan and Coleman, 1977; Brown and Weinberg, 2007), and therefore sets the initial conditions for its later evolution. It is the equivalent of the second turning point on the true vacuum side,  $x_2$ , appearing in Equation (4.1). In ordinary quantum mechanics, a particle with energy  $E$  emerges on the true vacuum side of the barrier at  $x_2(E)$  after tunneling. This is related to the bounce solution, which starts at  $x_1$ , rolls until reaching  $x_2$ , and then bounces back to  $x_1$ , thus  $x_2$  represents a slice of the bounce solution half way through.

In complete analogy, the field emerges at a configuration corresponding to a slice half way through the bounce solution (in Euclidean time). In flat space tunneling, the bounce is  $\varphi_B(\chi)$  where  $\chi^2 = \tau^2 + r^2$ , and thus touches the false vacuum at  $\tau \rightarrow \pm\infty$ . Hence the mid-way points occurs at  $\tau = 0$  and the solution emerges with  $\phi(x, 0) = \varphi_B(r)$ . One can then use this as an initial condition at  $t = 0$  for the Lorentzian field equations,

$$\nabla_\mu \nabla^\mu \varphi + V'(\varphi) = 0. \quad (4.59)$$

However, this is not really necessary, as the  $O(4)$  symmetry of the bounce solution carries over into  $O(3, 1)$  solution (Callan and Coleman, 1977), and thus the solution can be read off as

$$\varphi(x, t) = \varphi_B(\sqrt{r^2 - t^2}) \quad \text{for } r > t. \quad (4.60)$$

From this one can see that the bubble wall is moving outwards asymptotically at the speed of light. The inside of the light cone corresponds to an anti-de Sitter spacetime collapsing into a singularity (Espinosa et al., 2008, 2015; Burda et al., 2016; East et al., 2017).

The situation in de Sitter space is considerably more complicated, but the conclusion is the same (Brown and Weinberg, 2007). First, de Sitter bounces can be thought of as bounces at finite temperature on a curved *spatial* background described by constant time slices of the static patch of de Sitter space,

$$ds^2 = -(1 - H^2 r^2) dt^2 + (1 - H^2 r^2)^{-1} dr^2 + r^2 d\Omega_2^2. \quad (4.61)$$

The temperature in this case is the Gibbons-Hawking temperature (4.42) of de Sitter space. Bounces at finite temperature  $\beta = 1/k_B T$  correspond to periodic bounces in Euclidean space (Brown and Weinberg, 2007), with period  $\tau_{\text{period}} = \beta$ . In this case, the bounce starts at the false vacuum at  $\tau = -\pi/H$ , hits its mid-point at  $\tau = 0$ , and returns to the false vacuum side at  $\tau = \pi/H$ . Thus, the  $\tau = 0$  hypersurface describes the final state of the field after tunneling.

Analytic continuation of the metric back to real space can be performed using the approach of Burda et al. (2016). The  $O(4)$  symmetric Euclidean metric is of the form

$$ds^2 = d\chi^2 + a^2(\chi)[d\psi^2 + \sin^2 \psi d\Omega_2^2], \quad (4.62)$$

where in the de Sitter case,

$$a(\chi) = \frac{1}{H} \sin(H\chi). \quad (4.63)$$

Since it is straightforward to analytically continue the flat space metric back to real space via the transformation  $\tau = it$ , then the same thing can be done with any conformally flat metric, by changing variables to  $\tilde{\tau}, \tilde{r}$  such that

$$ds^2 = \frac{a^2(\chi)}{f^2(\chi)} [d\tilde{\tau}^2 + d\tilde{r}^2 + \tilde{r}^2 d\Omega_2^2], \quad (4.64)$$

which is achieved by choosing  $f(\chi)$  such that  $f'(\chi) = f/a, f(0) = 0$ . In the de Sitter case, this means

$$f(\chi) = C \frac{\sin(H\chi)}{1 + \cos(H\chi)} = C \tan(H\chi/2), \quad (4.65)$$

where  $C$  is an arbitrary constant - we can choose it to be 1. This co-ordinate system is obtained from the  $O(4)$  symmetric co-ordinates via

$$\tilde{\tau} = f(\chi) \cos(\psi), \quad (4.66)$$

$$\tilde{r} = f(\chi) \sin(\psi). \quad (4.67)$$

One then transforms back to real space exactly as in flat space, via  $\tilde{\tau} = it$ . The co-ordinate  $\chi$  is then related to  $\tilde{t}$  and  $\tilde{r}$  via

$$\chi = f^{-1}(\sqrt{\tilde{r}^2 - \tilde{t}^2}). \quad (4.68)$$

It should be noted that  $\tilde{t}, \tilde{r}$  as defined only cover the  $\tilde{r} > \tilde{t}$  portion of de Sitter space. Because the metric is manifestly conformally flat in these co-ordinates, we can see that this corresponds to the portion of de Sitter space *outside* the light-cone, which lies at  $\tilde{r} = \pm \tilde{t}$ .

Doing this for de Sitter yields the real space metric

$$ds^2 = \frac{4}{H^2[1 + \tilde{r}^2 - \tilde{t}^2]^2} [-d\tilde{t}^2 + d\tilde{r}^2 + \tilde{r}^2 d\Omega_2^2], \quad (4.69)$$

which at first glance, is not obviously de Sitter space. However, the transformation

$$t = \frac{1}{2H} \log \left| \frac{1 - \tilde{r}^2 + 2\tilde{t} + \tilde{t}^2}{1 - \tilde{r}^2 - 2\tilde{t} + \tilde{t}^2} \right|, \quad (4.70)$$

$$r = \frac{2\tilde{r}}{H(1 + \tilde{r}^2 - \tilde{t}^2)}, \quad (4.71)$$

can be readily shown to yield Equation (4.61), thus this is indeed a valid analytic continuation of the Euclidean 4-sphere back to de Sitter space.

To describe the subsequent evolution of the bubble, it is argued in Burda et al. (2016) that  $\phi(r, t) = \phi_B(\chi(r, t))$  matches the symmetry of the  $O(4)$  symmetric bounce, just as in flat space, with  $\chi(r, t)$  defined by Equation (4.68). As mentioned before, this describes only the evolution of the scalar field outside the light-cone. For  $\tilde{r} < \tilde{t}$ , it is necessary to solve the Euclidean equations directly. That calculation demonstrates explicitly that the formation of a singularity in the negative-potential region is inevitable (Burda et al., 2016), confirming previous calculations using the thin wall approximation in Coleman and De Luccia (1980).

As for the evolution outside the light-cone, it can be seen that, much as in flat space, a point of constant field value  $\varphi_0$  corresponding to  $\chi_0$  where  $\varphi_0 = \varphi(\chi_0)$ , satisfies

$$\tilde{r}(\tilde{t}) = \sqrt{\tilde{t}^2 + f^2(\chi_0(\phi_0))}, \quad (4.72)$$

which means that it rapidly approaches the speed of light as  $\tilde{t} \rightarrow \infty$ . Thus, just as in flat space, the bubble expands outwards at the speed of light.

Even if the bubble wall moves outward at the speed of light, it does not necessarily grow to fill the whole Universe, if it is trapped behind an event horizon. Scenarios in which bubbles of true vacuum form primordial black holes have been discussed (Hook et al., 2015; Kearney et al., 2015; Espinosa et al., 2018a,b). This can happen if inflation ends before the space inside the bubble hits the singularity. When the Universe reheats, thermal corrections (4.28) stabilize the Higgs potential, preventing the collapse. The bubble then collapses into a black hole, and the primordial black holes produced in this way could potentially constitute part or all of the dark matter in the Universe (Espinosa et al., 2018a). This scenario requires fine tuning to avoid the singularity or new heavy degrees of freedom that modify the potential at high field values (Espinosa et al., 2018b). The same scenario can also produce potentially observable gravitational waves (Espinosa et al., 2018).

## 5. COSMOLOGICAL CONSTRAINTS

### 5.1. Cosmological History

For the Universe to be currently in a metastable state rather than in its true ground state, it is not enough that the decay rate is slow today. The Universe also had to somehow end up in the metastable electroweak-scale state, and the decay rate had to be sufficiently slow in the past for the Universe to stay there through the whole history of the Universe. The former requirement depends on the initial conditions of the Universe, which are often assumed to involve Planck-scale field values, and therefore one needs to explain how the Higgs could have relaxed into the electroweak-scale vacuum without getting trapped into the negative-energy true vacuum. The latter condition, the survival of the current metastable state through the history of the Universe, requires that no bubbles of true vacuum were nucleated in our past light cone (Espinosa et al., 2008). This is because, once nucleated, a bubble of true vacuum expands at the speed of light and destroys everything in its way. If even a single bubble had nucleated at any time, anywhere in our past lightcone, it would have already hit us.

To describe the history of the Universe, we approximate it with the FLRW metric (3.3). The scale factor  $a(t)$  satisfies the Friedmann Equation (3.5)

$$H^2 \equiv \frac{\dot{a}^2}{a^2} = \frac{\rho}{3M_{\text{p}}^2}, \quad (5.1)$$

where  $\rho$  is the energy density of the Universe. When the dominating energy forms can be described by ideal fluids, one can write an equation of state  $p = w\rho$ , which relates the pressure

$p$  to the energy density  $\rho$  through the equation of state parameter  $w$ . From the first law of thermodynamics it then follows that the energy density scales with the expansion of the Universe as

$$\rho \propto a^{-3(1+w)}. \quad (5.2)$$

Observations indicate the Universe currently contains three forms of energy: radiation ( $w = 1/3$ ), matter ( $w = 0$ ) and dark energy, which we assume to be a cosmological constant with  $w = -1$ . The total energy density can be therefore written as a function of the scale factor as

$$\rho(a) = \rho_{\text{tot}}^0 \left( \Omega_{\Lambda} + \Omega_{\text{mat}} \left( \frac{a_0}{a} \right)^3 + \Omega_{\text{rad}} \left( \frac{a_0}{a} \right)^4 \right), \quad (5.3)$$

where  $\Omega_{\Lambda} = 0.69$ ,  $\Omega_{\text{mat}} = 0.31$  and  $\Omega_{\text{rad}} = 5.4 \times 10^{-5}$  are the observed energy fractions of cosmological constant, matter and radiation, respectively (Tanabashi et al., 2018),  $\rho_{\text{tot}}^0$  is the current total energy density, and  $a_0$  is the current value of the scale factor. It is common to choose  $a_0 = 1$  but we include it explicitly for clarity. The Universe is therefore currently dominated by dark energy, but in the past it was dominated by matter and, at even earlier times, by radiation. Observations also show that in its very early stages, before radiation-dominated epoch, the Universe went through a period of accelerating expansion known as inflation, during which the equation of state was, again,  $w \approx -1$ .

To find the expected number of bubbles in the past lightcone, it is convenient to write the FLRW metric in terms of the conformal time  $\eta$  as in Equation (3.10). In these coordinates, light satisfies  $|d\vec{r}/d\eta| = 1$ , so if we denote the current conformal time by  $\eta_0$ , the comoving radius of our past light cone at conformal time  $\eta$  is  $r(\eta) = \eta_0 - \eta$ .

The dependence of the scale factor on the conformal time is determined by the Friedmann Equation (3.5), which in terms of the conformal time is

$$\left( \frac{da}{d\eta} \right)^2 = \frac{\rho a^4}{3M_{\text{P}}^2}. \quad (5.4)$$

Using Equation (5.3) one finds that the conformal time since the end of inflation is

$$\eta_0 - \eta_{\text{inf}} = \frac{1}{H_0} \int_0^{a_0} \frac{da}{\sqrt{\Omega_{\Lambda} a^4 + \Omega_{\text{mat}} a_0^3 a + \Omega_{\text{rad}} a_0^4}} \approx 3.21 (a_0 H_0)^{-1}. \quad (5.5)$$

The bubble nucleation rate  $\Gamma$  may have been very different in different stages of the early evolution of the Universe. It depends on the curvature of spacetime and temperature, and also potentially on any perturbations or non-equilibrium processes that could catalyze or trigger the decay process and therefore it is function of the scale factor,  $\Gamma = \Gamma(a)$ . This allows us to write an expression for the expected number of bubbles  $\langle \mathcal{N} \rangle$  in our past

lightcone (after some initial time  $\eta_{\text{ini}}$ ) as

$$\begin{aligned} \langle \mathcal{N} \rangle &= \int_{\text{past}} d^4x \sqrt{-g} \Gamma(x) = \int_{\eta_{\text{ini}}}^{\eta_0} d\eta a(\eta)^4 \frac{4\pi r(\eta)^3}{3} \Gamma(a(\eta)) \\ &= \frac{4\pi}{3} \int_{\eta_{\text{ini}}}^{\eta_0} d\eta a(\eta)^4 (\eta_0 - \eta)^3 \Gamma(a(\eta)) \\ &= \frac{4\pi}{3} \int_0^{a_0} da (\eta_0 - a)^3 \frac{a^2}{H(a)} \Gamma(a). \end{aligned} \quad (5.6)$$

If this number is much greater than one, it would be unlikely that our part of the Universe could have survived until today, and therefore our existence requires

$$\langle \mathcal{N} \rangle \lesssim 1. \quad (5.7)$$

## 5.2. Late Universe

Let us first consider the post-inflationary Universe described by the energy density (5.3) and assume that the bubble nucleation rate  $\Gamma(a)$  in the past was at least as high as its current Minkowski space value  $\Gamma_0$ , i.e.,  $\Gamma(a) \geq \Gamma_0$ . In this case the expected number of bubbles is

$$\langle \mathcal{N} \rangle_{\text{post}} \geq \Gamma_0 \mathcal{V}_{\text{post}} = \frac{4\pi}{3} \Gamma_0 \int_{\eta_{\text{inf}}}^{\eta_0} d\eta (\eta_0 - \eta)^3 a(\eta)^4 \approx 0.125 \Gamma_0 H_0^{-4}. \quad (5.8)$$

Hence, the constraint on the nucleation rate  $\Gamma_0$  from the post-inflationary era is

$$\Gamma_0 \lesssim 8.0 H_0^4. \quad (5.9)$$

Using Equation (4.26) and  $H_0 \approx 70 \text{ km/s/Mpc}$ , this translates to a bound

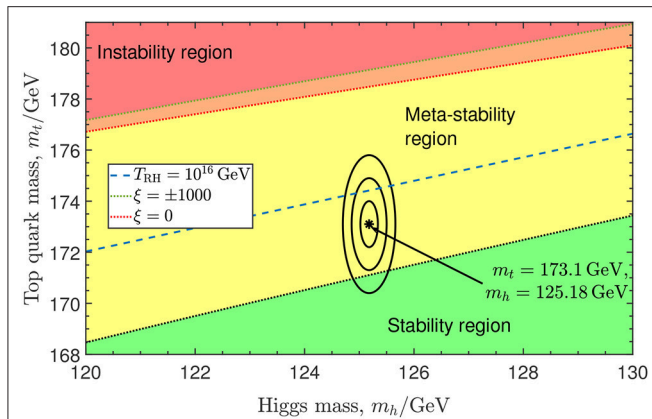
$$B \gtrsim 540 \quad (5.10)$$

on the bounce action.

By calculating the nucleation rate  $\Gamma_0$ , theories can be divided into categories: *stable*, *metastable* and *unstable*. If the rate exceeds the bound (5.9), the Universe would not have survived until the present day, and hence the vacuum is said to be unstable. If the rate is non-zero but satisfies Equation (5.9), the vacuum would not have decayed by the present time but would decay in the future, and hence it is said to be metastable. Finally, if the decay rate is strictly zero, which is the case when the current vacuum state is the global minimum of the potential, then the vacuum is said to be stable.

**Figure 7** shows the stability diagram of the Standard Model based on (Rajantie and Stopyra, 2017) (see section 4.2 for discussion), in terms of the Higgs mass  $M_h$ , top mass  $M_t$ , for three different values of the non-minimal coupling  $\xi$ . The ellipses show the 68%, 95%, and 99% contours based on the experimental and theoretical uncertainties in the masses.

It is worth mentioning that one could invoke the anthropic principle to evade the bound (5.9). Even if the expected number of bubbles  $\langle \mathcal{N} \rangle$  is large, there is always a non-zero probability that no bubbles were nucleated. Life can obviously only exist in those parts of the Universe that have no bubble nucleation event in their past light cone, and therefore that is necessarily



**FIGURE 7 |** Stability diagram of the Standard Model vacuum state in the pole masses  $M_t$ ,  $M_h$  of the top quark and Higgs boson, respectively. Ellipses show the  $1\sigma$ ,  $2\sigma$ ,  $3\sigma$  confidence intervals for  $M_t$  and  $M_h$  around their central values from Tanabashi et al. (2018). In the green region, the current vacuum is absolutely stable, in the yellow region it satisfies the bound (5.9), and in the red region it is so unstable that it would not have survived until the present day. The instability boundary includes gravitational backreaction (Rajantie and Stoprya, 2017) and is shown for  $\xi = 0$  and  $\xi = \pm 1000$  of the non-minimal curvature coupling. The blue dashed line shows the instability bound (5.62) obtained by taking the thermal history of the Universe into account (Delle Rose et al., 2016) and assuming a high reheating temperature  $T_{RH} = 10^{16}$  GeV. For lower reheating temperatures, the instability bound becomes weaker, and approaches the red dotted line as  $T_{RH} \rightarrow 0$ .

what we observe, no matter how low the probability is a priori. One can therefore argue that observations do not require  $\langle \mathcal{N} \rangle \lesssim 1$ . However, the anthropic argument does not rule out bubbles hitting us in the future, and therefore, if the Universe survives for a further period of time, that imposes a bound that is not subject to the anthropic principle. For this, the quantity that matters is the time derivative of the expected number of bubbles,

$$\frac{d\langle \mathcal{N} \rangle}{dt} = \frac{4\pi}{a_0} \Gamma_0 \int_{\eta_{ini}}^{\eta_0} d\eta a(\eta)^4 (\eta_0 - \eta)^2. \quad (5.11)$$

This imposes constraints that are numerically weaker but cannot be avoided by anthropic reasoning. To be concrete, one can carry out an experiment by waiting for a period of time  $t_{exp}$ , for example 1 year. If, at the end of the time period, the experimenter has not been hit by a bubble wall, this gives a constraint

$$t_{exp} \frac{d\langle \mathcal{N} \rangle}{dt} \lesssim 1. \quad (5.12)$$

For the post-inflationary Universe this is

$$t_{exp} \frac{d\langle \mathcal{N} \rangle}{dt} = (t_{exp} H_0) \times 4.91 \Gamma_0 H_0^{-4}, \quad (5.13)$$

and for  $t_{exp} = 1 \text{ yr}$ , one obtains the bound

$$\Gamma_0 \lesssim 2.9 \times 10^{10} H_0^4, \quad \text{or} \quad B \gtrsim 520. \quad (5.14)$$

This is weaker than Equation (5.9), but because of the very strong dependence of  $\Gamma_0$  on the top and Higgs masses, it does not change the stability constraints on them significantly.

### 5.3. Inflation

Although most of the spacetime volume of our past lightcone comes from the late times, the vacuum decay rate  $\Gamma(a)$  was much higher in the very early Universe. Depending on the cosmological scenario, it can be high enough to violate the bound (5.7), and this can be used to constrain theories.

The earliest stage in the evolution of the Universe that we have evidence for is inflation, a period of accelerating expansion, which made the Universe spatially flat, homogeneous and isotropic and also generated the initial seeds for structure formation. In simplest models of inflation, the energy density driving it is in the form of the potential energy  $V(\phi)$  of a scalar field  $\phi$  known as the inflaton. The inflaton field is nearly homogeneous, and satisfies the equation of motion

$$\ddot{\phi} + 3H\dot{\phi} + V'(\phi) = 0. \quad (5.15)$$

During inflation the potential satisfies the slow-roll conditions,

$$\epsilon \equiv \frac{M_{Pl}^2}{2} \left( \frac{V'}{V} \right)^2 \ll 1, \quad \text{and} \quad -1 \ll \eta \equiv M_{Pl}^2 \left( \frac{V''}{V} \right) \ll 1. \quad (5.16)$$

These conditions guarantee the existence of a solution in which the first term in Equation (5.15) is subdominant, and the inflaton field rolls slowly down the potential  $V(\phi)$ . As a consequence, the energy density  $\rho \approx V(\phi)$  and the Hubble rate are approximately constant.

The Hubble rate during inflation,  $H_{inf}$ , is largely unknown. Observationally it is constrained from above by the limits on primordial B-mode polarization in the cosmic microwave background radiation. This gives an upper bound  $r < 0.09$  on the tensor-to-scalar ratio (Ade et al., 2016), which implies  $H_{inf} \lesssim 3.3 \times 10^{-5} M_{Pl} \approx 8.0 \times 10^{13} \text{ GeV}$  at the time when the observable scales left the horizon. In a realistic inflationary model, the Hubble rate decreases with time, and would therefore be lower at the end of inflation. Although there are models in which the Hubble rate is well below the tensor bound, it is generally expected to be close to it, and in the simplest single-field inflation models it even exceeds it. It is therefore considered to be likely that the Hubble rate was significantly higher than the Higgs mass  $m_H \approx 125 \text{ GeV}$ .

The minimal inflationary model is Higgs inflation (Bezrukov et al., 2008), in which the non-minimal curvature coupling of the Higgs field is large,  $\xi \sim -49000\sqrt{\lambda}$ . This allows it to play the role of the inflaton, without the need for a separate inflaton field. During inflation, the Higgs field has a large value  $\phi \sim M_{Pl}/|\xi|$ , which means that the existence of a negative-energy minimum would appear to pose a problem for the scenario, because if the Higgs field gets trapped there, it would lead to a rapid collapse of the Universe instead of inflation. However, inclusion of higher-dimensional operators and finite temperature effects can avoid this problem (Bezrukov et al., 2015). Of course, if the actual top and Higgs masses lie in the stable region (see Figure 7), no problem arises. Furthermore, if they are just below the stability boundary, the effective Higgs potential would have an inflection point which would allow the scenario known as critical Higgs inflation (Bezrukov and Shaposhnikov, 2014; Hamada et al.,

2015, 2014), in which the Higgs field values are significantly lower than in conventional Higgs inflation. In the following our focus will be on the conventional scenario in which the inflaton is a separate field, and therefore we will not discuss Higgs inflation in detail. A thorough and up-to-date review of Higgs inflation, covering also the vacuum stability issues, is given in Rubio (2018).

Even in the scenario in which the inflaton is not the Standard Model Higgs field, one could expect on general grounds that the natural initial value for the Higgs field is at the Planck scale  $\varphi \sim M_{\text{P}}$  (Lebedev and Westphal, 2013). In that case the existence of a negative-energy true vacuum between the electroweak and Planck scales would appear to be a problem, just like in Higgs inflation. Therefore one either has to assume special initial conditions that guarantee  $\varphi \ll \varphi_{\text{bar}}$  everywhere, or find a mechanism that allows the Higgs field to roll to small values without getting trapped in the negative energy true vacuum.

In addition, even if that problem is solved, one still needs to avoid the nucleation bubbles of true vacuum, and hence satisfy the bound (5.7). Approximating inflation with a de Sitter space with constant Hubble rate  $H_{\text{inf}}$ , the expected number of bubbles (5.6) in our past lightcone originating from inflation is

$$\langle \mathcal{N} \rangle \approx \Gamma_{\text{inf}} \mathcal{V}_{\text{inf}}, \quad (5.17)$$

where  $\Gamma_{\text{inf}}$  is the vacuum decay rate, and  $\mathcal{V}_{\text{inf}}$  is the volume of the inflationary part of our past light cone. One can write this as

$$\begin{aligned} \mathcal{V}_{\text{inf}} &\approx \frac{4\pi}{9} [a_{\text{inf}}^3 H_{\text{inf}}^3 (\eta_0 - \eta_{\text{inf}})^3 + 3N_{\text{tot}}] H_{\text{inf}}^{-4} \\ &\approx \frac{4\pi}{9} \left[ 33.2 \times \left( \frac{a_{\text{inf}} H_{\text{inf}}}{a_0 H_0} \right)^3 + 3N_{\text{tot}} \right] H_{\text{inf}}^{-4}, \end{aligned} \quad (5.18)$$

where  $a_{\text{inf}}$  is the scale factor at the end of inflation,  $H_{\text{inf}}$  is the Hubble rate during inflation, and  $N_{\text{tot}}$  is the total number of e-foldings of inflation. In principle, if inflation lasted for an infinite amount of time, the volume of the inflationary past light cone would be infinite. In practice, inflation has a finite duration in most models, and the first term usually dominates in Equation (5.18).

The factor  $(a_{\text{inf}} H_{\text{inf}} / a_0 H_0)$  is the ratio of the comoving Hubble lengths today and at the end of inflation. It can be expressed as

$$\frac{a_{\text{inf}} H_{\text{inf}}}{a_0 H_0} = e^N, \quad (5.19)$$

where  $N$  is the number of e-foldings from the moment the largest observable scales left the horizon during inflation, to the end of inflation. It depends somewhat on the cosmological history, but is approximately (Liddle and Leach, 2003)

$$N \approx 60 + \ln \frac{V_{\text{inf}}^{1/4}}{10^{16} \text{ GeV}}. \quad (5.20)$$

This means that the spacetime volume of the inflationary past light cone is

$$\mathcal{V}_{\text{inf}} \approx 46 e^{3N} H_{\text{inf}}^{-4}. \quad (5.21)$$

From Equation (5.7), one then obtains a bound on the decay rate during inflation

$$\Gamma_{\text{inf}} \lesssim 0.02 e^{-3N} H_{\text{inf}}^4 \sim 10^{-80} \left( \frac{V_{\text{inf}}^{1/4}}{10^{16} \text{ GeV}} \right)^{-3} H_{\text{inf}}^4. \quad (5.22)$$

In the literature, the vacuum stability during inflation is often discussed in terms of the survival probability  $P_{\text{survival}}$ , which can be defined either as the fraction of volume that remains in the metastable vacuum at the end of inflation, or as the probability that a given Hubble volume remains in the metastable vacuum until the end of inflation. This is related to  $\langle \mathcal{N} \rangle$  by

$$\langle \mathcal{N} \rangle \approx e^N (1 - P_{\text{survival}}), \quad (5.23)$$

and therefore the bound (5.9) can be written as

$$1 - P_{\text{survival}} \lesssim e^{-3N}. \quad (5.24)$$

One can use the bounds (5.22) or (5.24) to constrain the Hubble rate during inflation  $H_{\text{inf}}$  and other parameters of the theory. This computation can be done in two ways, either using the instanton calculation of the tunneling rate discussed in section 4, or using the stochastic Starobinsky-Yokoyama approach discussed in section 3.4. The instanton calculation includes both quantum tunneling and classical excitation, and it can incorporate interactions and gravitational backreaction at short distances. Because it requires analytic continuation, it only works with constant Hubble rate  $H_{\text{inf}}$ , but it can still be expected to be a good approximation when the Hubble rate is slowly varying. In contrast, the stochastic approach can describe a time-dependent Hubble rate and gives a more detailed picture of the time evolution, but it includes only the classical excitation process and does not include interactions on sub-Hubble scales.

In the stochastic approach, the dynamics is described by either the Langevin Equation (3.14), or by the Fokker-Planck Equation (3.16), which gives the time evolution of the one-point probability distribution  $P(t, \varphi)$  of the Higgs field  $\varphi$ .

If the Higgs field is assumed to vanish initially,  $\varphi = 0$ , the probability distribution grows initially as

$$P(h, t) = \sqrt{\frac{2\pi}{H^3 t}} \exp\left(-\frac{2\pi^2 \varphi^2}{H^3 t}\right). \quad (5.25)$$

This is obtained by ignoring the Higgs potential  $V(\varphi)$ , which should be a good approximation at early times.

After some time the potential becomes important and starts to limit this growth. If the Hubble rate  $H$  is constant, the field approaches asymptotically the equilibrium distribution (3.17), and it is also a good approximation if the Hubble rate is varying sufficiently slowly. Considering the tree-level potential  $V(\varphi) = \lambda \varphi^4 / 4$  with constant  $\lambda > 0$ , the typical (rms) value of the field is given by Equation (3.19) as

$$\varphi_* \approx 0.363 \lambda^{-1/4} H \approx 0.605 H, \quad (5.26)$$

where the last expression is for the experimental value of the Higgs self coupling  $\lambda \approx 0.13$ . If  $H \gtrsim 10^{10} \text{ GeV}$ , these field

values are beyond the position (2.32) of the maximum of the potential. This means that for such values of the Hubble rate, inflationary fluctuations of the Higgs field would be able to throw the Higgs field over the potential barrier, triggering the vacuum instability (Espinosa et al., 2008; Lebedev and Westphal, 2013; Gabrielli et al., 2014; Kobakhidze and Spencer-Smith, 2013; Bhattacharya et al., 2014; Herranen et al., 2014; Fairbairn and Hogan, 2014; Hook et al., 2015; Kehagias and Riotto, 2014; Kobakhidze and Spencer-Smith, 2014; Enqvist et al., 2014; Kamada, 2015; Kearney et al., 2015; Shkerin and Sibiryakov, 2015). This would place a rough upper bound on the Hubble rate,

$$H \lesssim 10^{10} \text{ GeV}. \quad (5.27)$$

To make the bound more precise, Espinosa et al. (2008) solved the equation for the initial state  $P(0, \varphi) = \delta(\varphi)$ , with the boundary condition  $P(\varphi_{\text{bar}}, t) = 0$  to account for the destruction of any Hubble volume where  $\varphi > \varphi_{\text{bar}}$ . They then defined the survival probability of the vacuum as

$$P_{\text{survival}}(t) = \int_{-\varphi_{\text{bar}}}^{\varphi_{\text{bar}}} P(h, t). \quad (5.28)$$

Because of the boundary conditions, the survival probability is not conserved but decreases with time, and from the late-time asymptotic decay,

$$P_{\text{survival}} \sim e^{-\gamma t}, \quad (5.29)$$

one can determine the vacuum decay rate  $\Gamma \approx \gamma H^3$ . This way, they found the decay rate per unit time to be

$$\Gamma \sim \frac{H^6}{32\varphi_{\text{bar}}^2}, \quad \text{if } H \gtrsim \varphi_{\text{bar}}, \quad (5.30)$$

$$\Gamma \sim \lambda^{5/4} \varphi_{\text{bar}}^3 H \exp\left(-\frac{8\pi^2 V(\varphi_{\text{bar}})}{3H^4}\right), \quad \text{if } H \lesssim \varphi_{\text{bar}}. \quad (5.31)$$

One can see immediately that high Hubble rates,  $H \gtrsim \varphi_{\text{bar}}$ , are ruled out by the bound (5.22). The relevant result is therefore Equation (5.31). Comparing with Equation (5.22) one obtains the constraint

$$H \lesssim \left(\frac{8\pi^2}{9N} V(\varphi_{\text{bar}})\right)^{1/4}. \quad (5.32)$$

The numerical value of this constraint depends on the number of e-foldings  $N$  and, in particular, the height of the potential barrier, which is highly dependent on the precise Higgs and top masses. The bound on the ratio  $H/\varphi_{\text{bar}}$  is much less sensitive to the mass values, and therefore also quote the bounds in units of  $\varphi_{\text{bar}}$  rather than GeV. To obtain indicative bounds in physical units, one can use the central estimate for  $\varphi_{\text{bar}}$  in Equation (2.32). Using  $N = 60$ , the bound (5.32) becomes

$$H \lesssim 0.067 \varphi_{\text{bar}}. \quad (5.33)$$

The same result can also be obtained using the instanton approach (Kobakhidze and Spencer-Smith, 2013), which gives the decay rate (4.46),

$$\Gamma_{\text{inf}} \sim H_{\text{inf}}^4 e^{-B(H_{\text{inf}})}, \quad (5.34)$$

where  $B(H_{\text{inf}})$  is the relevant instanton action in de Sitter space with Hubble rate  $H_{\text{inf}}$ . The bound (5.22) can therefore be expressed as

$$B(H_{\text{inf}}) \gtrsim 3N + 4 \approx 180. \quad (5.35)$$

Figure 6 shows that for Hubble rates near  $\varphi_{\text{bar}}$ , the relevant instanton solution is the Hawking-Moss instanton, whose action (4.45) agrees with the exponent in Equation (5.31) in the limit where the barrier height is much less than the false vacuum energy. The instanton and Fokker-Planck calculations are therefore in good agreement in this case.

As discussed in section 4.6, the relevant instanton for lower Hubble rates,  $H < H_{\text{cross}}$ , is the Coleman-de Luccia solution (Rajantie and Stopyra, 2018). However, this is below the bound (5.32) and the Coleman-de Luccia action is very high,  $B \sim 1800$ , so that it gives a negligible decay rate, and therefore this does not change the bound (5.32).

There has been some debate about the correct field value used for the boundary condition (5.28) in the Fokker-Planck calculation. Hook et al. (2015) applied the boundary condition  $P(t, \varphi_{\text{cl}}) = 0$  at  $\varphi = \varphi_{\text{cl}}$ , determined from the condition

$$-V'(\varphi_{\text{cl}}) = \frac{3H^3}{2\pi}. \quad (5.36)$$

This condition means that at  $h > h_{\text{cl}}$  the classical motion of the field due to the potential gradient dominates over the quantum noise. Therefore it allows field trajectories that cross the top of the barrier but return to the metastable side because of the quantum fluctuations. This leads to a slower decay rate in the case of the high Hubble rate,

$$\Gamma \approx \frac{H^6}{32\varphi_{\text{cl}}^2}, \quad \text{for } H \gtrsim \varphi_{\text{bar}}. \quad (5.37)$$

East et al. (2017) considered the cutoff point the value  $\varphi_{\text{sr}}$ , where

$$\varphi_{\text{sr}} = -\frac{V'(\varphi_{\text{sr}})}{3H^2}. \quad (5.38)$$

This is the value above which the Higgs field no longer satisfies the slow roll condition and therefore the stochastic approach fails. The choice of the boundary condition becomes less important when  $H \ll \varphi_{\text{bar}}$ , and therefore it does not affect the bound Equation (5.32) very much. By solving the Fokker-Planck equation numerically, the authors obtained the bound

$$H \lesssim 0.067 \varphi_{\text{bar}}, \quad (5.39)$$

which coincides numerically with Equation (5.32).

There are aspects of physics that are not included in the approximations leading to the bound (5.32), and which can therefore provide a way to evade the bound. First, the high spacetime curvature  $R = 12H^2$  during inflation modifies the effective potential both at the tree level through the non-minimal coupling  $\xi$  and through the curvature-dependence of the loop

corrections. The non-minimal coupling gives rise to an effective curvature-dependent mass term (3.12),

$$\mathcal{M}^2 = m^2 + 12 \left( \xi - \frac{1}{6} \right) H^2. \quad (5.40)$$

If  $\xi$  is positive, it increases the potential height between the electroweak and true vacua and helps to stabilize the electroweak vacuum even if the Hubble rate is well above the bound (5.32) (Espinosa et al., 2008; Kehagias and Riotto, 2014; Herranen et al., 2014). On the other hand, negative values of  $\xi$  make the vacuum less stable. For  $\xi < 0$ , Joti et al. (2017) obtained the bound

$$H \lesssim \frac{0.005}{\sqrt{-\xi}} \varphi_{\text{bar}}. \quad (5.41)$$

The stabilizing effects of the non-minimal coupling have also been discussed in Kamada (2015), Espinosa et al. (2015), Shkerin and Sibiryakov (2015), Kohri and Matsui (2017), Kohri and Matsui (2016), Kawasaki et al. (2016), Calmet et al. (2018), and Markkanen et al. (2018)

The curvature dependent loop corrections mean that the non-minimal coupling  $\xi$  runs with the renormalization scale, and even it is zero at low energies, it runs to a negative value  $\xi \approx -0.03$  at the relevant scales for the instability  $\mu_\Lambda \sim 10^{10}$  GeV (Herranen et al., 2014). Curvature contributions to the loop corrections to the rest of the effective potential can be approximated using renormalization group improvement (Herranen et al., 2014), by choosing the renormalization scale as  $\mu_* \approx H$  when  $H \gtrsim \varphi$ , rather than  $\mu_* \approx \varphi$  which had been used previously. Using the curvature-dependent renormalization scale, such as Equation (3.28), has become the norm in the more recent literature (Kearney et al., 2015; East et al., 2017; Rodriguez-Roman and Fairbairn, 2018). Having  $\mu_* \sim H$  means that for sufficiently high Hubble rates the effective coupling becomes negative,  $\lambda(\mu_*) < 0$ , and the potential barrier disappears completely, unless  $\xi$  is sufficiently large. Both of these effects, running  $\xi$  and the curvature-dependent renormalization scale, tend to de-stabilize the vacuum. Taking them into account gives the bound (Herranen et al., 2014)

$$\xi \gtrsim 0.06 \quad \text{for} \quad H \gtrsim \varphi_{\text{bar}}. \quad (5.42)$$

The full curvature-dependent effective potential was computed at one-loop order in Markkanen et al. (2018), and confirms this expectation. The stability bounds as a function of the Hubble rate  $H$  and the non-minimal coupling  $\xi$  are shown in **Figure 8**. For comparison, the bound from particle collider experiments is  $|\xi| \lesssim 2.6 \times 10^{15}$  (Atkins and Calmet, 2013).

A sufficiently large positive non-minimal coupling  $\xi$  can also avoid the Higgs field initial condition problem. It was found in Calmet et al. (2018) that if

$$\xi \gtrsim H/10^{-4} M_{\text{P}}, \quad (5.43)$$

the positive curvature contribution to the effective potential allows the Higgs field to roll from Planck-scale values to its

electroweak minimum during inflation without getting trapped into the negative-energy true vacuum.

The bound (5.32) also does not take into account any direct coupling between the Higgs and the inflaton field  $\phi$ . Although a direct coupling is not radiatively generated, in general it is possible and the precise form it would have and its effects on vacuum stability depend on the details of the inflaton sector. The simplest example is a coupling of the form  $\lambda_{\phi h} \phi^2 h^2$  in chaotic inflation with a quadratic potential. During inflation, the inflaton field has a high value  $\phi \gtrsim M_{\text{P}}$ , and therefore the coupling produces an effective mass term for the Higgs field,

$$\mathcal{M}^2 = m^2 + \lambda_{\phi h} \phi^2. \quad (5.44)$$

Coupling values  $\lambda_{\phi h} \lesssim 10^{-6}$  would not spoil the flatness of the inflaton potential (Lebedev and Westphal, 2013; Gross et al., 2016), and if  $\lambda_{\phi h} \gtrsim 10^{-10}$ , it would stabilize the vacuum during inflation and allow the Higgs field to roll to its current small field values even if starts from a Planck-scale value at the beginning of inflation (Lebedev and Westphal, 2013; Fairbairn and Hogan, 2014; Gross et al., 2016). This coupling has also been discussed in Kamada (2015). Considering the non-minimal curvature coupling  $\xi$  and the direct Higgs-inflaton coupling  $\lambda_{\phi h}$  together, Ema et al. (2017) finds the constraint

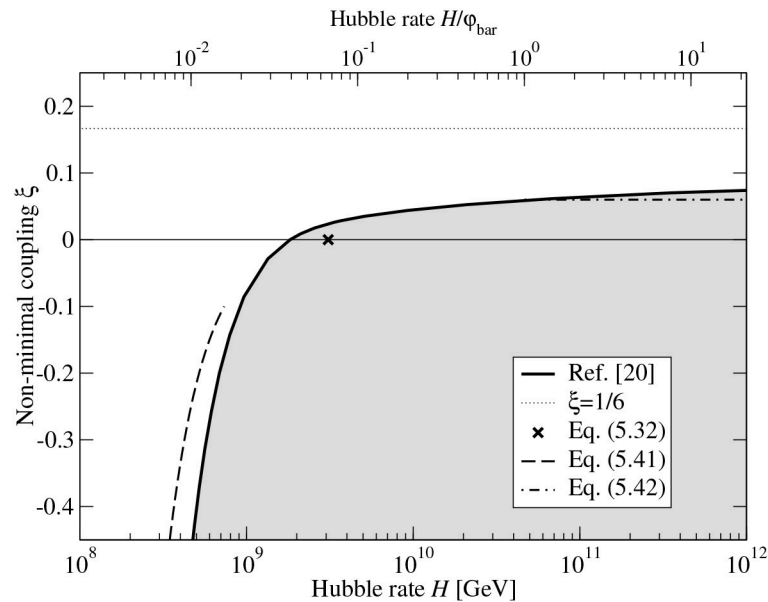
$$10^{-10} \lesssim \lambda_{\phi h} + 10^{-10} \xi \lesssim 10^{-6}, \quad (5.45)$$

in the quadratic chaotic inflation model.

Other forms of the Higgs-inflaton coupling have been considered in Bhattacharya et al. (2014), Hook et al. (2015), Ballesteros and Tamarit (2015), and Cline and Espinosa (2018). There are also other effects that could potentially stabilize the vacuum state during inflation. Non-zero temperature  $T \gtrsim 6 \times 10^{13}$  GeV during inflation (Fairbairn and Hogan, 2014), moduli fields (Ema et al., 2016), coupling to a spectator scalar field (Gong and Kitajima, 2017; Han et al., 2018), or top quark production (Rodriguez-Roman and Fairbairn, 2018) could all generate an effective stabilizing term in the effective potential.

## 5.4. Reheating

The end of inflation can be defined as the point at which the Universe no longer undergoes accelerated expansion, which occurs when  $w = -1/3$ . This marks the beginning of the so-called *reheating* phase during which the energy density stored as potential energy gets converted into the hot thermal plasma of the Big Bang. If the acceleration is sourced by a slowly rolling inflaton  $\phi$ , during reheating the slow-roll conditions cease to hold and the inflaton will begin a phase where its (average) kinetic energy is comparable to its potential energy. This usually manifests as coherent oscillations around the minimum of the potential. Reheating is said to be completed when the energy density of the hot Big Bang overtakes that of the inflaton sector, which often proceeds via direct couplings allowing the inflaton to decay into SM constituents. It is however worth pointing out that it is perfectly possible to have successful reheating without any couplings between the inflaton and the SM sector, for examples of such models see Figueroa and Byrnes (2017), Tenkanen and



**FIGURE 8 |** Stability bounds on the non-minimal coupling  $\xi$  (renormalized at the electroweak scale) and the Hubble rate during inflation  $H_{\text{inf}}$ . The colored area shows the unstable region based on the numerical results from Markkanen et al. (2018), the cross corresponds to Equation (5.32), the dashed line to Equation (5.41) and the dash-dotted line to Equation (5.42). The bottom axis refers to units calculated using the barrier position from Equation (2.32).

Vaskonen (2016), Dimopoulos and Markkanen (2018), and Haro (2018).

An inflaton field coherently oscillating around the minimum of its potential may source a very potent non-perturbative amplification of quantum modes, which takes place during the early stages of reheating and is hence often referred to as *preheating* (Kofman et al., 1994, 1997). If a phase of preheating occurs, it does not lead to the completion of reheating as the created particles tend to shut off any non-perturbative behavior through backreaction and a perturbative decay channel is often required to ensure the complete decay of the inflaton.

From the point of view of a possible vacuum destabilization, preheating is a crucial epoch because vacuum decay is potentially induced by a large amplification of the Higgs field (Herranen et al., 2015). It is important to note that at the time of preheating, the Universe has not yet reheated to a high temperature, and therefore the thermal effects discussed in section 4.3 cannot stabilize the vacuum state.

Let us proceed to consider the familiar Lagrangian appropriate for the Higgs doublet in curved space (3.30). We consider Hubble rates well above the electroweak scale,  $H \gg M_h$ , and therefore we can neglect the tree-level mass parameter, and use the action

$$S = \int d^4x \sqrt{|g|} \left[ \frac{1}{2} \nabla_\mu \varphi \nabla^\mu \varphi - \frac{\xi}{2} R \varphi^2 - \frac{\lambda}{4} \varphi^4 \right]. \quad (5.46)$$

We also assume a single-field model of inflation with a canonical kinetic term and the potential  $U(\phi)$ . The inflaton  $\phi$  is taken to dominate the energy density of the Universe completely and because of this the Higgs field may be considered as a

subdominant spectator that can be neglected in the Einstein equation. Using then

$$\rho = \frac{1}{2} \dot{\phi}^2 + U(\phi); \quad p = \frac{1}{2} \dot{\phi}^2 - U(\phi), \quad (5.47)$$

in the Friedmann equations (3.5), we can solve for the Ricci scalar  $R$

$$R = 6 \left[ \left( \frac{\dot{a}}{a} \right)^2 + \frac{\ddot{a}}{a} \right] = \frac{1}{M_{\text{Pl}}^2} [4U(\phi) - \dot{\phi}^2]. \quad (5.48)$$

After inflation ends, the inflaton field  $\phi$  rolls down its potential, and initially oscillates coherently about its minimum  $\phi_{\text{min}}$ , until it eventually decays. We assume that the inflaton potential vanishes at the minimum,  $U(\phi_{\text{min}})$ , as is usually the case. We can see from Equation (5.48) that during every oscillation, when  $\phi \approx \phi_{\text{min}}$ , the Ricci scalar becomes negative,  $R < 0$ . This, in turn, means that the non-minimal term  $\sim \xi R \varphi^2$  gives rise to a tachyonic mass term (3.12) for the Higgs field. As already discussed in section 3.3, this gives rise to significant excitation of the field. The fact that the non-minimal term can lead to extremely efficient particle creation during preheating was first discussed in Bassett and Liberati (1998) and Tsujikawa et al. (1999).

Particle creation from a periodically tachyonic effective mass was analyzed in detail in Dufaux et al. (2006) where it was named *tachyonic resonance*. It is much more extreme than the resonant effects usually taking place during preheating. Hence a dangerous fluctuation of the Higgs field can be generated during a single oscillation of the inflaton.

For concreteness, we now focus on the case of a quadratic inflaton potential

$$U(\phi) = \frac{1}{2}m^2\phi^2. \quad (5.49)$$

Although as a complete model of inflation, this is not compatible with observations (Akrami et al., 2018), it approximates the shape of the potential around the minimum in general single-field models. The behavior of the inflaton field during its coherent oscillations can be approximately written as  $\phi = \phi_0(t) \cos(mt)$  where  $\phi_0$  is a slowly changing amplitude  $\phi_0(t) = \sqrt{6H(t)M_P/m}$  (Kofman et al., 1994, 1997).

We will focus only on a very brief time period immediately after inflation, when no thermalization has yet taken place. In cosmic time the properly normalized mode is obtained from Equation (3.9) as  $f(\eta) \rightarrow f(t)/\sqrt{a}$  giving the mode equation

$$\ddot{f}(t) + 3H\dot{f}(t) + \left[ \frac{k^2}{a^2} - \frac{9}{4}H^2 - \frac{3}{2}\dot{H} + \xi R \right] f(t) = 0. \quad (5.50)$$

By using the Friedmann equations (3.5) in this approximation the mode equation can be cast in the Mathieu form (Bassett and Liberati, 1998; Herranen et al., 2015)

$$\frac{d^2 f(t)}{dz^2} + \left[ A_k - 2q \cos(2z) \right] f(t) = 0, \quad z = mt, \quad (5.51)$$

$$A_k = \frac{k^2}{a^2 m^2} + \xi \frac{\phi_0^2}{2M_P^2}, \quad q = \frac{3\phi_0^2}{4M_P^2} \left( \frac{1}{4} - \xi \right).$$

Making use of the analysis in Dufaux et al. (2006) we can derive an analytical result for the occupation number of the Higgs field  $n_k$  after the first oscillation

$$n_k = e^{2X_k}, \quad X_k = \int_{\Delta z} \Omega_k dz \approx \sqrt{\xi} \frac{\phi_0}{M_P} \approx \sqrt{\xi}, \quad (5.52)$$

where  $\Omega^2 \equiv -\omega^2$ . The  $\omega^2$  is the term in the square brackets in Equation (5.51) and  $\Delta z$  covers the time period when  $\omega^2 < 0$ . Including only the IR modes  $k < aH$ , neglecting the expansion of space, the self-interaction and furthermore assuming  $\xi \gtrsim 1$  we can estimate the generated Higgs fluctuations at horizon scale,  $\langle \hat{\phi}^2 \rangle_{aH}$ , after the first oscillation of the inflaton as Herranen et al. (2015)

$$\langle \hat{\phi}^2 \rangle_{aH} \approx \int_0^{aH} \frac{dk k^2}{2\pi^2 a^3} 2|f(t)|^2 n_k \approx \left( \frac{H}{2\pi} \right)^2 \frac{2 \exp \left\{ \sqrt{\xi} \frac{2\phi_0}{M_P} \right\}}{3\sqrt{3\xi}}. \quad (5.53)$$

If  $\phi_0 \sim M_P$ , as in chaotic inflation, one can see from Equation (5.53), that the Higgs fluctuations are exponentially amplified if  $\xi \gtrsim 1$ . The fluctuation  $\Delta\phi \sim$  can become larger than the position of the potential barrier in the SM

$$\Delta\phi \equiv \sqrt{\langle |\hat{\phi}|^2 \rangle_{aH}} = \sqrt{4\langle \hat{\phi}^2 \rangle_{aH}} \gtrsim \varphi_{\text{bar}}. \quad (5.54)$$

Note that a large and positive  $\xi$  gives rise to a destabilizing effect after inflation. This is opposite to what happens during inflation

when it suppresses fluctuations by effectively making the field heavy (see section 5.3).

In general, once a significant particle density is produced it tends to work against any further particle production (Kofman et al., 1997). For the Higgs the main backreaction comes from the self-interaction term, which contributes to the effective mass (3.12), along with the curvature terms visible in (5.50), as

$$\mathcal{M}^2 = -\frac{9}{4}H^2 - \frac{3}{2}\dot{H} + \xi R + 6\lambda\langle \hat{\phi}^2 \rangle, \quad (5.55)$$

very similarly as we derived in the 1-loop approximation for a scalar singlet in Equation (3.24) of section 3.5. In order for tachyonic particle creation to take place one must have  $\xi|R| \gtrsim 6\lambda\langle \hat{\phi}^2 \rangle$  for  $\xi \gtrsim 1$ . However, in section (3.6) it was shown that the Hubble rate contributes to the RG scale through curvature induced running (see Equation 3.28). If  $H \gtrsim \varphi_{\text{bar}}$ , the four-point coupling is negative, implying that the backreaction in fact enhances the instability, and will not suppress tachyonic particle creation even if a large variance is generated.

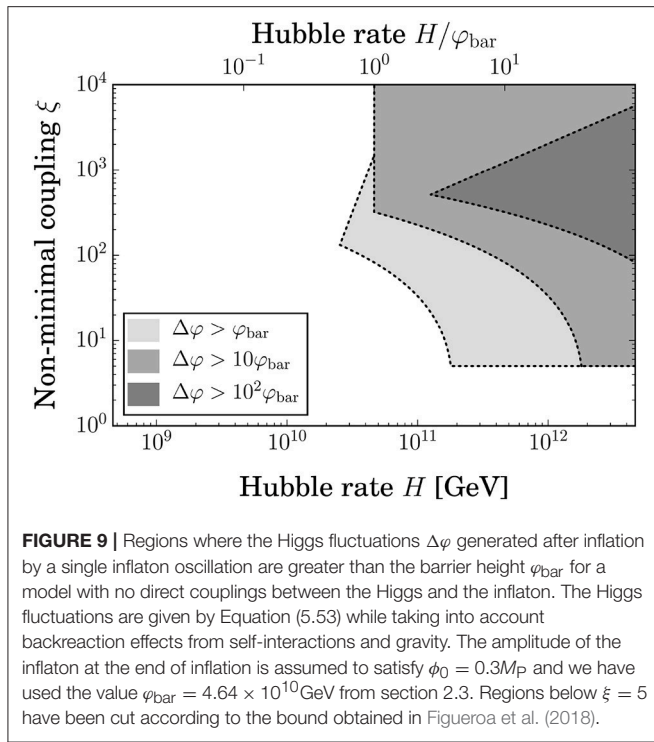
Backreaction also arises from the gravitational disturbance of the generated particle density. In order to reach this threshold one must create enough particles such that their energy density approaches  $3H^2 M_P^2$ . The relevance of gravitation backreaction we can estimate from the approximate energy density for the Higgs (Herranen et al., 2015)

$$\rho_{\text{Higgs}} \approx 24\xi H^2 \langle \hat{\phi}^2 \rangle + 6\lambda \langle \hat{\phi}^2 \rangle^2. \quad (5.56)$$

When  $\rho_{\text{Higgs}} \sim 3H^2 M_P^2$  the Higgs starts to influence the dynamics of spacetime requiring a non-linear analysis. Below we will assume that when the gravitational backreaction threshold is reached the particle production will seize.

More detailed calculations of the process have been carried out using linearized approximation (Kohri and Matsui, 2016) and lattice field theory simulations (Ema et al., 2016; Figueroa et al., 2018). The most detailed analysis, carried out in Figueroa et al. (2018), used the tree-level RGI effective potential with three-loop running, and considered different top quark masses. The main conclusion was that the instability is triggered with high probability for  $\xi \gtrsim 4-5$ , for a top quark mass  $M_t \approx 173.3\text{GeV}$ . This implies an upper bound on  $\xi$  in the context of quadratic chaotic inflation.

The regions where a dangerously large fluctuation of the Higgs is generated after a single oscillation of the inflaton are shown in **Figure 9**. The rather complicated shapes are the result of the interplay of the variance (5.53) and the constraints coming from self interactions and gravitational backreaction. In **Figure 9** we have assumed that the amplitude of the inflaton at the end of inflation satisfies  $\phi_0 = 0.3M_P$ . While this is true for the quadratic (chaotic) model of inflation, it is not true generically. Since Equation (5.53) is exponentially dependent on  $\phi_0$  the predictions are very sensitive to the specifics of inflation. Similarly, the duration of reheating plays a crucial role and for prolonged reheating a possible instability may be further enhanced. The derivation of Equation (5.53) is based on the adiabatic approximation (Dufaux et al., 2006), which can be shown to break down for small  $\xi$  (Postma and van de Vis, 2017).



Furthermore, very little particle creation is expected when close to the (approximately) conformally invariant point  $\xi = 1/6$ . For these reasons and the lattice results of Figueroa et al. (2018) we have conservatively cut out regions with  $\xi \leq 5$ .

As discussed in the previous section, the set-up in Equation (5.46) assuming that the inflaton is decoupled from the SM is in many ways the minimal one. Couplings between the inflaton and the SM sector may of course be introduced or even required by a specific reheating model. Vacuum stability during preheating in models with no non-minimal coupling but with direct couplings between the inflaton and the Higgs was investigated in Ema et al. (2017), Gross et al. (2016), and Enqvist et al. (2016). In particular, in Ema et al. (2017) it was shown that in some cases vacuum decay during preheating may take place also for low-scale inflation.

In Kohri and Matsui (2016), Ema et al. (2017), Ema et al. (2017), and Ema et al. (2016) both the non-minimal coupling and direct Higgs-inflaton couplings were considered. In a sense in this case the Higgs fluctuations are sourced in a complicated manner by the interplay of tachyonic resonance (Dufaux et al., 2006) and the (usual) parametric resonance (Kofman et al., 1997). For the precise coupling ranges where significant particle production takes place and possible implications for instability, see Ema et al. (2017). We also point out that particle creation resulting from the non-adiabatic change in the background curvature when inflation ends, already shown in Ford (1987), can be enough to probe the unstable region of the effective Higgs potential (Herranen et al., 2015).

## 5.5. Hot Big Bang

After reheating, the Universe entered a thermal radiation-dominated state, in which vacuum decay rate

can be approximated by the thermal rate (4.35) at the relevant temperature, and the Hubble rate was given by the equation

$$H(T)^2 = g_*(T) \frac{\pi^2}{90} \frac{T^4}{M_{\text{P}}^2}, \quad (5.57)$$

where  $g_*(T)$  is the effective number of degrees of freedom and has the value  $g_*(T) = 106.75$  in the Standard Model at high temperatures.

Using Equation (5.6) one can write the expected number of true vacuum bubbles in our past light cone from this era as Espinosa et al. (2008); Salvio et al. (2016)

$$\frac{d\langle \mathcal{N} \rangle}{d \ln T} = \frac{4\pi}{3} \left( \frac{g_{*S}^0}{g_*(T)} \right) \left( \frac{T_0}{T} \right)^3 \frac{(\eta_0 - \eta(T))^3}{H(T)} \Gamma(T). \quad (5.58)$$

where  $g_{*S}^0 = 3.94$  is the effective number of entropy degrees of freedom today. Using Equations (5.5) and (5.57) this becomes

$$\frac{d\langle \mathcal{N} \rangle}{d \ln T} \approx 1.49 \frac{M_{\text{P}}}{H_0^3} \left( \frac{T_0}{T} \right)^3 \frac{\Gamma(T)}{T^2}, \quad (5.59)$$

If the Universe reheated instantaneously after inflation, the reheat temperature  $T_{\text{RH}}$  to which the Universe equilibrate, is related to the Hubble rate at the end of inflation  $H_{\text{inf}}$  through where  $g_* \geq 106.75$  is the effective number of degrees of freedom. Because the rate decreases when the temperature decreases, Equation (5.59) is dominated by high temperatures  $T \sim T_{\text{RH}}$ . Therefore one can approximate

$$\langle \mathcal{N} \rangle \approx \frac{M_{\text{P}} T_0^3}{H_0^3} \frac{\Gamma(T_{\text{RH}})}{T_{\text{RH}}^5} \approx \frac{M_{\text{P}} T_0^3}{H_0^3 T_{\text{RH}}} e^{-B(T_{\text{RH}})}. \quad (5.60)$$

Requiring that  $\langle \mathcal{N} \rangle \ll 1$  leads to the bound

$$B(T_{\text{RH}}) \gtrsim 3 \ln \frac{T_0}{H_0} + \ln \frac{M_{\text{P}}}{T_{\text{RH}}} \approx 202 + \ln \frac{M_{\text{P}}}{T_{\text{RH}}}, \quad (5.61)$$

which is satisfied by the numerical result (4.34) for the current central Higgs and top mass values, and therefore it does not imply a bound on the reheat temperature.

As at zero temperature, the vacuum stability depends sensitively on the top and Higgs masses. A detailed analysis (Delle Rose et al., 2016) based on integrating Equation (5.59) gives an upper bound on the top quark mass,

$$\frac{M_t}{\text{GeV}} < 0.283 \left( \frac{\alpha_s - 0.1184}{0.0007} \right) + 0.4612 \frac{M_h}{\text{GeV}} + 1.907 \log_{10} \frac{T_{\text{RH}}}{\text{GeV}} + \frac{1.2 \times 10^3}{0.323 \log_{10} \frac{T_{\text{RH}}}{\text{GeV}} + 8.738}. \quad (5.62)$$

In practice, reheating is not instantaneous, and there may have been a period when the Standard Model degrees of freedom were in thermal equilibrium but were not the dominant energy component. In the scenario in which the inflaton

field decays slowly and dominates the energy density of the Universe for an extended period, the maximum temperature is (Espinosa et al., 2008; Elias-Miro et al., 2012; Delle Rose et al., 2016)

$$T_{\max} = \left(\frac{3}{8}\right)^{2/5} \left(\frac{40}{\pi^2}\right)^{1/8} \frac{g_*^{1/8}(T_{\text{RH}})}{g_*^{1/4}(T_{\max})} (M_{\text{P}} H_{\text{inf}} T_{\text{RH}}^2)^{1/4}. \quad (5.63)$$

Because the Universe was not radiation-dominated Equation (5.59) does not describe the period when  $T \gtrsim T_{\text{RH}}$ . Instead, one has

$$\frac{\langle \mathcal{N} \rangle}{d \ln T} \approx \frac{M_{\text{P}}}{T^2} H_0^3 \left(\frac{T_0}{T_{\text{RH}}}\right)^3 \left(\frac{T_{\text{RH}}}{T}\right)^{10} \Gamma(T). \quad (5.64)$$

However, this has only a small effect on the numerical bounds (Delle Rose et al., 2016).

In **Figure 7**, the blue dashed line shows the bound (5.62) calculated with a reheat temperature  $T_{\text{RH}} \sim 10^{16}$  GeV. As can be seen, the inclusion of the thermal history of the Universe reduces the allowed mass range compared with the zero-temperature bounds. The central experimental values are still allowed, but the instability boundary lies within two standard deviations from them.

## 6. CONCLUDING REMARKS

The current experimental data show that with a high likelihood, the electroweak vacuum state of the Standard Model is metastable. Even though the vacuum state could be stabilized by new physics beyond the Standard Model, and even in the Standard Model parameters corresponding to a stable vacuum are still allowed by experimental errors, it is important to study the implications of the possible metastability. That allows one to understand whether the metastability is compatible with observations, and if so, what constraints it places on the parameters of the theory.

If the electroweak vacuum really is metastable, then bubbles of the true, negative-energy vacuum can be nucleated by quantum tunneling or classical excitation, as discussed in section 4. Once a bubble has formed, it expands at the speed of light, destroying everything in its way. This clearly has not happened yet in our part of the Universe, which means there has not been a single bubble nucleation event in our whole past light cone. In section 5, we showed how the likelihood of this can be estimated by computing the nucleation rate and integrating it over the past light cone. Because the past light cone includes all of the different cosmological eras, and the nucleation rate and its dependence on theory parameters is different in each era, this provides a rich set of constraints on both the cosmological history and on the Standard Model parameters.

In this review, we have focussed on four different cosmological eras: inflation, preheating, hot radiation-dominated phase, and the late Universe. Vacuum stability in the late Universe one obtains constraints on the Higgs and top masses, and they are made tighter by considering the hot radiation-dominated phase, as summarized in **Figure 7**. Survival of the vacuum through inflation and the subsequent preheating phase constrains the Hubble rate during inflation and the Higgs-curvature coupling  $\xi$  (**Figures 8, 9**), as well as other aspects of inflationary models. A demonstration of the power of these considerations is that for quadratic chaotic inflation, the non-minimal coupling is constrained to be within the range  $0.06 \lesssim \xi \lesssim 5$ , which is 15 orders of magnitude stronger than the experimental bounds from the Large Hadron Collider (Atkins and Calmet, 2013). Cosmological vacuum decay has a unique connection to gravity via the early Universe, which opens up an observational window to particle physics well beyond what colliders can achieve.

In this work we have reviewed the, already rather significant, body of work investigating the cosmological consequences of the SM Higgs possessing a metastable potential. We have also discussed the relevant theoretical frameworks required for such studies. The multidisciplinary nature of the problem is perhaps one of the reasons behind the ongoing significant interest as particle physics, quantum field theory and gravity all play a prominent role. Although the specifics of the theory behind early Universe dynamics are not currently known what has become quite apparent is that a metastable Higgs potential generically leads to non-trivial constraints, which are completely invisible to colliders.

On the other hand, despite the large number of existing studies, much remains to be explored. For example, at the moment very few works exist that go beyond the simple quadratic model of inflation. This is equally true for the inflationary and reheating epochs. There is also a great deal of scope for improving calculation techniques in order to obtain more precise and robust constraints, for example by going beyond the semiclassical approximation or fully including gravitational effects. The work on the cosmological aspects of Higgs vacuum metastability is only starting.

## AUTHOR CONTRIBUTIONS

Initially, sections 2 and 3 were drafted by TM, section 4 by SS, and section 5 by AR, but subsequently all authors edited and contributed to the whole manuscript.

## ACKNOWLEDGMENTS

The authors are grateful to numerous individuals for discussion and collaboration on this topic. AR and TM were funded by the STFC grant ST/P000762/1, and SS was funded by the Imperial College President's PhD Scholarship and the UCL Cosmoparticle Initiative.

## REFERENCES

- Ade, P. A. R., Ahmed, Z., Aikin, R. W., Alexander, K. D., Barkats, D., Benton, S. J., et al. (2012). Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Phys. Lett. B* 716, 1–29. doi: 10.1016/j.physletb.2012.08.020
- Abramo, L. R. W., Brandenberger, R. H., and Mukhanov, V. F. (1997). The energy - momentum tensor for cosmological perturbations. *Phys. Rev. D* 56, 3248–3257. doi: 10.1103/PhysRevD.56.3248
- Aad, G., Abajyan, T., Abbott, B., Abdallah, J., Abdel Khalek, S., Ali Abdelalim, A., et al. (2016). Improved constraints on cosmology and foregrounds from BICEP2 and keck array cosmic microwave background data with inclusion of 95 GHz band. *Phys. Rev. Lett.* 116:031302. doi: 10.1103/PhysRevLett.116.031302
- Akrami, Y., Arroja, F., Ashdown, M., Aumont, J., Baccigalupi, C., Ballardini, M., et al. (2018). Planck 2018 results. X. Constraints on inflation. *arXiv [Preprint]:1807.06211*.
- Alekhin, S., Djouadi, A., and Moch, S. (2012). The top quark and Higgs boson masses and the stability of the electroweak vacuum. *Phys. Lett. B* 716, 214–219. doi: 10.1016/j.physletb.2012.08.024
- Allen, B. (1985). Vacuum states in de sitter space. *Phys. Rev. D* 32:3136. doi: 10.1103/PhysRevD.32.3136
- Altarelli, G., and Isidori, G. (1994). Lower limit on the Higgs mass in the standard model: an update. *Phys. Lett. B* 337, 141–144. doi: 10.1016/0370-2693(94)91458-3
- Anchordoqui, L. A., Antoniadis, I., Goldberg, H., Huang, X., Lust, D., Taylor, T. R., et al. (2013). Vacuum stability of standard model<sup>++</sup>. *J. High Energy Phys.* 2:74. doi: 10.1007/JHEP02(2013)074
- Anderson, G. W. (1990). New cosmological constraints on the higgs boson and top quark masses. *Phys. Lett. B* 243, 265–270. doi: 10.1016/0370-2693(90)90849-2
- Andreassen, A., Frost, W., and Schwartz, M. D. (2014). Consistent Use of the Standard Model Effective Potential. *Phys. Rev. Lett.* 113:241801. doi: 10.1103/PhysRevLett.113.241801
- Antipin, O., Gillioz, M., Krog, J., Mølgaard, E., and Sannino, F. (2013). Standard model vacuum stability and weyl consistency conditions. *J. High Energy Phys.* 8:34. doi: 10.1007/JHEP08(2013)034
- Arai, T. (2012). Nonperturbative infrared effects for light scalar fields in de sitter space. *Class. Quant. Grav.* 29:215014. doi: 10.1088/0264-9381/29/21/215014
- Arnold, P. B. (1989). Can the electroweak vacuum be unstable? *Phys. Rev. D* 40:613. doi: 10.1103/PhysRevD.40.613
- Arnold, P. B., and Vokos, S. (1991). Instability of hot electroweak theory: bounds on m(H) and M(t). *Phys. Rev. D* 44, 3620–3627. doi: 10.1103/PhysRevD.44.3620
- Atkins, M., and Calmet, X. (2013). Bounds on the nonminimal coupling of the higgs boson to gravity. *Phys. Rev. Lett.* 110:051301. doi: 10.1103/PhysRevLett.110.051301
- Avramidi, I. G. (2000). Heat kernel and quantum gravity. *Lect. Notes Phys. Monogr.* 64, 1–149. doi: 10.1007/3-540-46523-5\_1
- Balek, V., and Demetrian, M. (2004). A Criterion for bubble formation in de Sitter universe. *Phys. Rev. D* 69:063518. doi: 10.1103/PhysRevD.69.063518
- Balek, V., and Demetrian, M. (2005). Euclidean action for vacuum decay in a de Sitter universe. *Phys. Rev. D* 71:023512. doi: 10.1103/PhysRevD.71.023512
- Ballesteros, G., and Tamari, C. (2015). Higgs portal valleys, stability and inflation. *J. High Energy Phys.* 9:210. doi: 10.1007/JHEP09(2015)210
- Bambhaniya, G., Bhupal Dev, P., Goswami, S., Khan, S., and Rodejohann, W. (2017). Naturalness, vacuum stability and leptogenesis in the minimal seesaw model. *Phys. Rev. D* 95:095016. doi: 10.1103/PhysRevD.95.095016
- Bando, M., Kugo, T., Maekawa, N., and Nakano, H. (1993a). Improving the effective potential: multimass scale case. *Prog. Theor. Phys.* 90, 405–418. doi: 10.1143/ptp/90.2.405
- Bando, M., Kugo, T., Maekawa, N., Nakano, H. (1993b). Improving the effective potential. *Phys. Lett. B* 301, 83–89. doi: 10.1016/0370-2693(93)90725-W
- Bassett, B. A., and Liberati, S. (1998). Geometric reheating after inflation. *Phys. Rev. D* 58:021302. doi: 10.1103/PhysRevD.58.021302
- Bednyakov, A. V., Kniehl, B. A., Pikelner, A. F., and Veretin, O. L. (2015). Stability of the electroweak vacuum: gauge independence and advanced precision. *Phys. Rev. Lett.* 115:201802. doi: 10.1103/PhysRevLett.115.201802
- Bednyakov, A. V., Pikelner, A. F., and Velizhanin, V. N. (2013). Yukawa coupling beta-functions in the Standard Model at three loops. *Phys. Lett. B* 722, 336–340. doi: 10.1016/j.physletb.2013.04.038
- Bednyakov, A. V., Pikelner, A. F., and Velizhanin, V. N. (2013). Higgs self-coupling beta-function in the Standard Model at three loops. *Nucl. Phys. B* 875, 552–565. doi: 10.1016/j.nuclphysb.2013.07.015
- Bednyakov, A. V., Pikelner, A. F., and Velizhanin, V. N. (2014). Three-loop Higgs self-coupling beta-function in the Standard Model with complex Yukawa matrices. *Nucl. Phys. B* 879, 256–267. doi: 10.1016/j.nuclphysb.2013.12.012
- Bezrukov, F., Kalmykov, M. Yu., Kniehl, B. A., and Shaposhnikov, M. (2012). Higgs boson mass and new physics. *J. High Energy Phys.* 10:140. doi: 10.1007/JHEP10(2012)140
- Bezrukov, F., Rubio, J., and Shaposhnikov, M. (2015). Living beyond the edge: higgs inflation and vacuum metastability. *Phys. Rev. D* 92:083512. doi: 10.1103/PhysRevD.92.083512
- Bezrukov, F., and Shaposhnikov, M. (2014). Higgs inflation at the critical point. *Phys. Lett. B* 734, 249–254. doi: 10.1016/j.physletb.2014.05.074
- Bezrukov, F. L., and Shaposhnikov, M. (2008). The standard model higgs Boson as the inflaton. *Phys. Lett. B* 659, 703–706. doi: 10.1016/j.physletb.2007.11.072
- Bhattacharya, K., Chakraborty, J., Das, S., and Mondal, T. (2014). Higgs vacuum stability and inflationary dynamics after BICEP2 and PLANCK dust polarisation data. *J. Cosmol. Astropart. Phys.* 1412:1. doi: 10.1088/1475-7516/2014/12/001
- Bhupal Dev, P. S., Ghosh, D. K., Okada, N., and Saha, I. (2013). 125 GeV Higgs Boson and the type-II seesaw model. *J. High Energy Phys.* 3:150. doi: 10.1007/JHEP03(2013)150
- Birrell, N. D., and Davies, P. C. W. (1984). *Quantum Fields in Curved Space*. Cambridge Monographs on Mathematical Physics. Cambridge: Cambridge Univ. Press.
- Bounakis, M., and Moss, I. G. (2018). Gravitational corrections to Higgs potentials. *J. High Energy Phys.* 4:71. doi: 10.1007/JHEP04(2018)071
- Branchina, V., and Messina, E. (2013). Stability, higgs boson mass and new physics. *Phys. Rev. Lett.* 111:241801. doi: 10.1103/PhysRevLett.111.241801
- Branchina, V., and Messina, E. (2017). Stability and UV completion of the Standard Model. *EPL* 117:61002. doi: 10.1209/0295-5075/117/61002
- Branchina, V., Messina, E., and Platania, A. (2014). Top mass determination, Higgs inflation, and vacuum stability. *J. High Energy Phys.* 9:182. doi: 10.1007/JHEP09(2014)182
- Branchina, V., Messina, E., and Sher, M. (2015). Lifetime of the electroweak vacuum and sensitivity to Planck scale physics. *Phys. Rev. D* 91:013003. doi: 10.1103/PhysRevD.91.013003
- Brown, A. R., and Weinberg, E. J. (2007). Thermal derivation of the Coleman-De Luccia tunneling prescription. *Phys. Rev. D* 76:064003. doi: 10.1103/PhysRevD.76.064003
- Buchbinder, I. L., and Odintsov, S. D. (1985). Effective potential and phase transitions induced by curvature in gauge theories in curved space-time. *Yad. Fiz.* 42, 1268–1278.
- Buchbinder, I. L., Odintsov, S. D., and Shapiro, I. L. (1992). *Effective Action in Quantum Gravity*. Abingdon: Taylor & Francis Group, LLC.
- Burda, P., Gregory, R., and Moss, I. (2015a). Gravity and the stability of the Higgs vacuum. *Phys. Rev. Lett.* 115:071303. doi: 10.1103/PhysRevLett.115.071303
- Burda, P., Gregory, R., and Moss, I. (2015b). Vacuum metastability with black holes. *J. High Energy Phys.* 8:114. doi: 10.1007/JHEP08(2015)114
- Burda, P., Gregory, R., and Moss, I. (2016). The fate of the Higgs vacuum. *J. High Energy Phys.* 6:25. doi: 10.1007/JHEP06(2016)025
- Buttazzo, D., Degrandi, G., Giardino, P. P., Giudice, G. F., Sala, F., Salvio, A., et al. (2013). Investigating the near-criticality of the Higgs boson. *J. High Energy Phys.* 12:89. doi: 10.1007/JHEP12(2013)089
- Cabibbo, N., Maiani, L., Parisi, G., and Petronzio, R. (1979). Bounds on the fermions and Higgs boson masses in grand unified theories. *Nucl. Phys. B* 158, 295–305. doi: 10.1016/0550-3213(79)90167-6
- Callan, C. G. Jr. (1970). Broken scale invariance in scalar field theory. *Phys. Rev. D* 2, 1541–1547. doi: 10.1103/PhysRevD.2.1541
- Callan, C. G. Jr., and Coleman, S. R. (1977). The Fate of the False Vacuum. 2. First Quantum Corrections. *Phys. Rev. D* 16, 1762–1768. doi: 10.1103/PhysRevD.16.1762
- Callan, C. G. Jr., Coleman, S. R., and Jackiw, R. (1970). A new improved energy - momentum tensor. *Ann. Phys.* 59, 42–73. doi: 10.1016/0003-4916(70)90394-5
- Calmet, X., Kuntz, I., Moss, I. G. (2018). Non-Minimal Coupling of the Higgs Boson to Curvature in an Inflationary Universe. *Found. Phys.* 48, 110–120. doi: 10.1007/s10701-017-0131-2

- Canko, D., Gialamas, I., Jelic-Cizmek, G., Riotto, A., and Tetradis, N. (2018). On the catalysis of the electroweak vacuum decay by black holes at high temperature. *Eur Phys J. C* 78:328. doi: 10.1140/epjc/s10052-018-5808-y
- Casas, J. A., Espinosa, J. R., and Quiros, M. (1995). Improved Higgs mass stability bound in the standard model and implications for supersymmetry. *Phys. Lett. B* 342, 171–179. doi: 10.1016/0370-2693(94)01404-Z
- Casas, J. A., Espinosa, J. R., and Quiros, M. (1996). Standard model stability bounds for new physics within LHC reach. *Phys. Lett. B* 382, 374–382. doi: 10.1016/0370-2693(96)00682-X
- Chakraborty, J., Das, M., and Mohanty, S. (2013). Constraints on TeV scale Majorana neutrino phenomenology from the Vacuum Stability of the Higgs. *Mod. Phys. Lett. A* 28:1350032. doi: 10.1142/S0217732313500326
- Chao, W., Gonderinger, M., and Ramsey-Musolf, M. J. (2012). Higgs vacuum stability, neutrino mass, and dark matter. *Phys. Rev. D* 86:113017. doi: 10.1103/PhysRevD.86.113017
- Chatrchyan, S., et al. (2012). Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Phys. Lett. B* 716, 30–61. doi: 10.1016/j.physletb.2012.08.021
- Chen C.-S., and Tang, Y. (2012). Vacuum stability, neutrinos, and dark matter. *J. High Energy Phys.* 4:19. doi: 10.1007/JHEP04(2012)019
- Cheng T. P., and Li, L. F. (1984). *Gauge Theory of Elementary Particle Physics*. Oxford: Oxford University Press.
- Chernikov, N. A., and Tagirov, E. A. (1968). Quantum theory of scalar fields in de Sitter space-time. *Ann. Inst. H. Poincaré Phys. Theor.* A9:109.
- Chetyrkin, K. G., and Zoller, M. F. (2012). Three-loop  $\beta$ -functions for top-Yukawa and the Higgs self-interaction in the Standard Model. *J. High Energy Phys.* 6:33. doi: 10.1007/JHEP06(2012)033
- Chetyrkin, K. G., and Zoller, M. F. (2013).  $\beta$ -function for the Higgs self-interaction in the Standard Model at three-loop level. *J. High Energy Phys.* 4:91. doi: 10.1007/JHEP04(2013)091
- Chigusa, S., Moroi, T., and Shoji, Y. (2017). State-of-the-art calculation of the decay rate of electroweak vacuum in the standard model. *Phys. Rev. Lett.* 119:211801. doi: 10.1103/PhysRevLett.119.211801
- Chigusa, S., Moroi, T., and Shoji, Y. (2018). Decay Rate of Electroweak Vacuum in the Standard Model and Beyond. *Phys. Rev. D* 97:116012. doi: 10.1103/PhysRevD.97.116012
- Chun, E. J., Jung, S., and Lee, H. M. (2013). Radiative generation of the Higgs potential. *Phys. Lett. B* 725, 158–163. doi: 10.1016/j.physletb.2013.06.055
- Chun, E. J., Lee, H. M., and Sharma, P. (2012). Vacuum stability, perturbativity, EWPd and higgs-to-diphoton rate in Type II Seesaw Models. *J. High Energy Phys.* 11:106. doi: 10.1007/JHEP11(2012)106
- Chung, D. J. H., Long, A. J., and Wang, L.-T. (2013). 125 GeV Higgs boson and electroweak phase transition model classes. *Phys. Rev. D* 87:023509. doi: 10.1103/PhysRevD.87.023509
- Chung, J. M., and Chung, B. K. (1999). Renormalization group improvement of the effective potential in massive  $\phi^4$  theory. *Phys. Rev. D* 60:105001. doi: 10.1103/PhysRevD.60.105001
- Cline, J. M., and Espinosa, J. R. (2018). Axionic landscape for Higgs coupling near-criticality. *Phys. Rev. D* 97:035025. doi: 10.1103/PhysRevD.97.035025
- Coleman, S. (1985). *Aspects of Symmetry*. Cambridge: Cambridge University Press
- Coleman, S. R. (1977). The fate of the false vacuum. 1. Semiclassical theory. *Phys. Rev. D* 15, 2929–2936. doi: 10.1103/PhysRevD.15.2929
- Coleman, S. R., and De Luccia, F. (1980). Gravitational effects on and of vacuum decay. *Phys. Rev. D* 21:3305. doi: 10.1103/PhysRevD.21.3305
- Coleman, S. R., Glaser, V., and Martin, A. (1978). Action minima among solutions to a class of euclidean scalar field equations. *Commun. Math. Phys.* 58, 211–221. doi: 10.1007/BF01609421
- Coleman, S. R., and Weinberg, E. J. (1973). Radiative corrections as the origin of spontaneous symmetry breaking. *Phys. Rev. D* 7, 1888–1910. doi: 10.1103/PhysRevD.7.1888
- Czerwinska, O., Lalak, Z., Lewicki, M., and Olszewski, P. (2016). The impact of non-minimally coupled gravity on vacuum stability. *J. High Energy Phys.* 10:4. doi: 10.1007/JHEP10(2016)004
- Czerwinska, O., Lalak, Z., Lewicki, M., and Olszewski, P. (2017). Non-minimally coupled gravity and vacuum stability. *PoS CORFU2016:64*. doi: 10.22323/1.292.0064
- Czerwińska, O., Lalak, Z., and Nakonieczny, L. (2015). Stability of the effective potential of the gauge-less top-Higgs model in curved spacetime. *J. High Energy Phys.* 11:207. doi: 10.1007/JHEP11(2015)207
- Datta, A., and Raychaudhuri, S. (2013). Vacuum stability constraints and LHC searches for a model with a universal extra dimension. *Phys. Rev. D* 87:035018. doi: 10.1103/PhysRevD.87.035018
- Degrassi, G., Di Vita, S., Elias-Miro, J., Espinosa, J. R., Giudice, G. F., Isidori, G., et al. (2012). Higgs mass and vacuum stability in the Standard Model at NNLO. *J. High Energy Phys.* 8:98. doi: 10.1007/JHEP08(2012)098
- Delle Rose, L., Marzo, C., and Urbano, A. (2016). On the fate of the Standard Model at finite temperature. *J. High Energy Phys.* 5:50. doi: 10.1007/JHEP05(2016)050
- DeWitt, B. S. (1964). Dynamical theory of groups and fields. *Conf. Proc. C* 630701, 585–820.
- Di Luzio, L., Isidori, G., and Ridolfi, G. (2016). Stability of the electroweak ground state in the Standard Model and its extensions. *Phys. Lett. B* 753, 150–160. doi: 10.1016/j.physletb.2015.12.009
- Di Luzio, L., and Mihaila, L. (2014). On the gauge dependence of the Standard Model vacuum instability scale. *J. High Energy Phys.* 6:79. doi: 10.1007/JHEP06(2014)079
- Dimopoulos, K., and Markkanen, T. (2018). Non-minimal gravitational reheating during kination. *J. Cosmol. Astropart. Phys.* 1806:021.
- Dufaux, J. F., Felder, G. N., Kofman, L., Peloso, M., and Podolsky, D. (2006). Preheating with trilinear interactions: tachyonic resonance. *J. Cosmol. Astropart. Phys.* 607:6. doi: 10.1088/1475-7516/2006/07/006
- Dunne, G. V., and Wang, Q.-H. (2006). Fluctuations about cosmological instantons. *Phys. Rev. D* 74:024018. doi: 10.1103/PhysRevD.74.024018
- East, W. E., Kearney, J., Shakya, B., Yoo, H., and Zurek, K. M. (2017). Spacetime dynamics of a higgs vacuum instability during inflation. *Phys. Rev. D* 95:023526. doi: 10.1103/PhysRevD.95.023526
- Eichhorn, A., Gies, H., Jaeckel, J., Plehn, T., Scherer, M. M., and Sondenheimer, R. (2015). The higgs mass and the scale of new physics. *J. High Energy Phys.* 4:22. doi: 10.1007/JHEP04(2015)022
- Elias-Miro, J., Espinosa, J. R., Giudice, G. F., Isidori, G., Riotto, A., and Strumia, A. (2012). Higgs mass implications on the stability of the electroweak vacuum. *Phys. Lett. B* 709, 222–228. doi: 10.1016/j.physletb.2012.02.013
- Elias-Miro, J., Espinosa, J. R., Giudice, G. F., Lee, H. M., and Strumia, A. (2012). Stabilization of the electroweak vacuum by a scalar threshold effect. *J. High Energy Phys.* 6:31. doi: 10.1007/JHEP06(2012)031
- Elizalde, E., and Odintsov, S. D. (1993). Renormalization group improved effective potential for gauge theories in curved space-time. *Phys. Lett. B* 303, 240–248. doi: 10.1016/0370-2693(93)91427-O
- Elizalde, E., and Odintsov, S. D. (1994a). Renormalization group improved effective Lagrangian for interacting theories in curved space-time. *Phys. Lett. B* 321, 199–204. doi: 10.1016/0370-2693(94)90464-2
- Elizalde, E., and Odintsov, S. D. (1994b). Renormalization group improved effective potential for interacting theories with several mass scales in curved space-time. *Z. Phys. C* 64, 699–708. doi: 10.1007/BF01957780
- Elizalde, E., and Odintsov, S. D. (1994c). Renormalization group improved effective potential for finite grand unified theories in curved space-time. *Phys. Lett. B* 333, 331–336. doi: 10.1016/0370-2693(94)90151-1
- Ellis, J., Espinosa, J. R., Giudice, G. F., Hoecker, A., and Riotto, A. (2009). The probable fate of the standard model. *Phys. Lett. B* 679, 369–375. doi: 10.1016/j.physletb.2009.07.054
- Ellwanger, U., and Lindner, M. (1993). Constraints on new physics from the Higgs and top masses. *Phys. Lett. B* 301, 365–371. doi: 10.1016/0370-2693(93)91164-I
- Ema, Y., Karciauskas, M., Lebedev, O., Rusak, S., and Zatta, M. (2017). Higgs-inflaton mixing and vacuum stability. *arXiv [Preprint]:1711.10554*.
- Ema, Y., Karciauskas, M., Lebedev, O., and Zatta, M. (2017). Early universe higgs dynamics in the presence of the Higgs-inflaton and non-minimal Higgs-gravity couplings. *J. Cosmol. Astropart. Phys.* 1706:54. doi: 10.1088/1475-7516/2017/06/054
- Ema, Y., Mukaida, K., and Nakayama, K. (2016). Electroweak vacuum stabilized by moduli during/after inflation. *Phys. Lett. B* 761, 419–423. doi: 10.1016/j.physletb.2016.08.046
- Ema, Y., Mukaida, K., and Nakayama, K. (2016). Fate of electroweak vacuum during preheating. *J. Cosmol. Astropart. Phys.* 1610:43. doi: 10.1088/1475-7516/2016/10/043
- Ema, Y., Mukaida, K., and Nakayama, K. (2017). Electroweak vacuum metastability and low-scale inflation. *J. Cosmol. Astropart. Phys.* 1712:30. doi: 10.1088/1475-7516/2017/12/030

- Enqvist, K., Karčiauskas, M., Lebedev, O., Rusak, S., and Zatta, M. (2016). Postinflationary vacuum instability and Higgs-inflaton couplings. *J.Cosmol. Astropart. Phys.* 1611:25. doi: 10.1088/1475-7516/2016/11/025
- Enqvist, K., Meriniemi, T., and Nurmi, S. (2013). Generation of the Higgs Condensate and Its Decay after Inflation. *J.Cosmol. Astropart. Phys.* 1310:57. doi: 10.1088/1475-7516/2013/10/057
- Enqvist, K., Meriniemi, T., and Nurmi, S. (2014). Higgs dynamics during inflation. *J.Cosmol. Astropart. Phys.* 1407:25. doi: 10.1088/1475-7516/2014/07/025
- Enqvist, K., Nurmi, S., Podolsky, D., and Rigopoulos, G. I. (2008). On the divergences of inflationary superhorizon perturbations. *J.Cosmol. Astropart. Phys.* 804:25. doi: 10.1088/1475-7516/2008/04/025
- Enqvist, K., Nurmi, S., Rusak, S., and Weir, D. (2016). Lattice Calculation of the Decay of Primordial Higgs Condensate. *J.Cosmol. Astropart. Phys.* 1602:57. doi: 10.1088/1475-7516/2016/02/057
- Espinosa, J. R., Garny, M., and Konstandin, T. (2016). Interplay of infrared divergences and gauge-dependence of the effective potential. *Phys. Rev. D* 94:055026. doi: 10.1103/PhysRevD.94.055026
- Espinosa, J. R., Garny, M., Konstandin, T., and Riotto, A. (2017). Gauge-independent scales related to the standard model vacuum instability. *Phys. Rev. D* 95:056004. doi: 10.1103/PhysRevD.95.056004
- Espinosa, J. R., Giudice, G. F., Morgante, E., Riotto, A., Senatore, L., Strumia, A. et al. (2015). The cosmological Higgstory of the vacuum instability. *J. High Energy Phys.* 9:174. doi: 10.1007/JHEP09(2015)174
- Espinosa, J. R., Giudice, G. F., and Riotto, A. (2008). Cosmological implications of the Higgs mass measurement. *J.Cosmol. Astropart. Phys.* 805:2. doi: 10.1088/1475-7516/2008/05/002
- Espinosa, J. R., and Quiros, M. (1995). Improved metastability bounds on the standard model Higgs mass. *Phys. Lett. B* 353, 257–266. doi: 10.1016/0370-2693(95)00572-3
- Espinosa, J. R., Racco, D., and Riotto, A. (2018). A cosmological signature of the SM higgs instability: gravitational waves. *arXiv [Preprint]:1804.07732*.
- Espinosa, J. R., Racco, D., and Riotto, A. (2018a). Cosmological signature of the standard model higgs vacuum instability: primordial black holes as dark matter. *Phys. Rev. Lett.* 120:121301. doi: 10.1103/PhysRevLett.120.121301
- Espinosa, J. R., Racco, D., and Riotto, A. (2018b). Primordial black holes from higgs vacuum instability: avoiding fine-tuning through an ultraviolet safe mechanism. *arXiv [Preprint]:1804.07731*.
- Fairbairn, M., and Hogan, R. (2014). Electroweak vacuum stability in light of BICEP2. *Phys. Rev. Lett.* 112:201801. doi: 10.1103/PhysRevLett.112.201801
- Felder, G., Kofman, L., and Linde, A. D. (2001). Tachyonic instability and dynamics of spontaneous symmetry breaking. *Phys. Rev. D* 64:123517. doi: 10.1103/PhysRevD.64.123517
- Figueroa, D. G., and Byrnes, C. T. (2017). The standard model higgs as the origin of the hot big bang. *Phys. Lett. B* 767, 272–277. doi: 10.1016/j.physletb.2017.01.059
- Figueroa, D. G., Rajantie, A., and Torrenti, F. (2018). Higgs field-curvature coupling and postinflationary vacuum instability. *Phys. Rev. D* 98:023532. doi: 10.1103/PhysRevD.98.023532
- Ford, C., Jones, D. R. T., Stephenson, P. W., and Einhorn, M. B. (1993). The Effective potential and the renormalization group. *Nucl. Phys. B* 395, 17–34. doi: 10.1016/0550-3213(93)90206-5
- Ford, L. H. (1987). Gravitational particle creation and inflation. *Phys. Rev. D* 35:2955. doi: 10.1103/PhysRevD.35.2955
- Ford, L. H., and Toms, D. J. (1982). Dynamical symmetry breaking due to radiative corrections in cosmology. *Phys. Rev. D* 25:1510. doi: 10.1103/PhysRevD.25.1510
- Frampton, P. H., Hung, P. Q., and Sher, M. (2000). Quarks and leptons beyond the third generation. *Phys. Rept.* 330:263. doi: 10.1016/S0370-1573(99)00095-2
- Freese, K., Sfakianakis, E. I., Stengel, P., and Visinelli, L. (2018). The higgs boson can delay Reheating after Inflation. *J.Cosmol. Astropart. Phys.* 1805:67. doi: 10.1088/1475-7516/2018/05/067
- Fubini, S. (1976). A new approach to conformal invariant field theories. *Nuovo Cim. A* 34:521. doi: 10.1007/BF02785664
- Gabrielli, E., Heikinheimo, M., Kannike, K., Racioppi, A., Raidal, M., and Spethmann, C. (2014). Towards completing the standard model: vacuum stability, EWSB and dark matter. *Phys. Rev. D* 89:015017. doi: 10.1103/PhysRevD.89.015017
- Garbrecht, B., Rigopoulos, G., and Zhu, Y. (2014). Infrared correlations in de Sitter space: field theoretic versus stochastic approach. *Phys. Rev. D* 89:063506. doi: 10.1103/PhysRevD.89.063506
- Garg, I., Goswami, S., and Khan, N. (2017). Electroweak vacuum stability in presence of singlet scalar dark matter in TeV scale seesaw models. *Phys. Rev. D* 96:055020. doi: 10.1103/PhysRevD.96.055020
- Gautier, F., and Serreau, J. (2015). Scalar field correlator in de Sitter space at next-to-leading order in a  $1/N$  expansion. *Phys. Rev. D* 92:105035. doi: 10.1103/PhysRevD.92.105035
- George, D. P., Mooij, S., and Postma, M. (2012). Effective action for the Abelian Higgs model in FLRW. *J.Cosmol. Astropart. Phys.* 1211:43. doi: 10.1088/1475-7516/2012/11/043
- Gibbons, G. W., and Hawking, S. W. (1977). Cosmological Event Horizons, Thermodynamics, and Particle Creation. *Phys. Rev. D* 15, 2738–2751. doi: 10.1103/PhysRevD.15.2738
- Gies, H., Gneiting, C., and Sondenheimer, R. (2014). Higgs Mass Bounds from Renormalization Flow for a simple Yukawa model. *Phys. Rev. D* 89:045012. doi: 10.1103/PhysRevD.89.045012
- Gilkey, P. B. (1975). The spectral geometry of a Riemannian manifold. *J. Diff. Geom.* 10, 601–618. doi: 10.4310/jdg/1214433164
- Gonderinger, M., Li, Y., Patel, H., and Ramsey-Musolf, M. J. (2010). Vacuum stability, perturbativity, and scalar singlet dark matter. *JHEP* 1:053. doi: 10.1103/PhysRevD.86.043511
- Gonderinger, M., Lim, H., and Ramsey-Musolf, M. J. (2012). Complex scalar singlet dark matter: vacuum stability and phenomenology. *Phys. Rev. D* 86:043511. doi: 10.1103/PhysRevD.86.043511
- Gong, J.-O., and Kitajima, N. (2017). Cosmological stochastic Higgs field stabilization. *Phys. Rev. D* 96:063521. doi: 10.1103/PhysRevD.96.063521
- Gregory, R., Moss, I. G., and Withers, B. (2014). Black holes as bubble nucleation sites. *J. High Energy Phys.* 3:81. doi: 10.1007/JHEP03(2014)081
- Gross, C., Lebedev, O., and Zatta, M. Higgs-inflaton coupling from reheating and the metastable Universe. *Phys. Lett. B* (2016) 753, 178–181. doi: 10.1016/j.physletb.2015.12.014
- Hackworth, J. C., and Weinberg, E. J. (2005). Oscillating bounce solutions and vacuum tunneling in de Sitter spacetime. *Phys. Rev. D* 71:044014. doi: 10.1103/PhysRevD.71.044014
- Hadamard, J. (1923). *Lectures on Cauchy's Problem*. New Haven, CT: Yale Univ. Press.
- Hamada, Y., Kawai, H., Oda, K.-Y., and Park, S. C. (2014). Higgs inflation is still alive after the results from BICEP2. *Phys. Rev. Lett.* 112:241301. doi: 10.1103/PhysRevLett.112.241301
- Hamada, Y., Kawai, H., Oda, K.-Y., and Park, S. C. (2015). Higgs inflation from standard model criticality. *Phys. Rev. D* 91:053008. doi: 10.1103/PhysRevD.91.053008
- Hambye, T., and Riesselmann, K. (1997). Matching conditions and Higgs mass upper bounds revisited. *Phys. Rev. D* 55, 7255–7262. doi: 10.1103/PhysRevD.55.7255
- Han, C., Pi, S., and Sasaki, M. (2018). Quintessence saves higgs instability. *arXiv [Preprint]:1809.05507*.
- Hardwick, R. J. (2018). Multiple spectator condensates from inflation. *J.Cosmol. Astropart. Phys.* 1805:54. doi: 10.1088/1475-7516/2018/05/054
- Haro, J. (2018). Different reheating mechanisms in quintessence inflation. *arXiv [Preprint]:1807.07367*.
- Hawking, S. W., and Moss, I. G. (1982). Supercooled phase transitions in the very early universe. *Phys. Lett. B* 110, 35–38. doi: 10.1016/0370-2693(82)90946-7
- He, X.-G., Phoon, H., Tang, Y., and Valencia, G. (2013). Unitarity and vacuum stability constraints on the couplings of color octet scalars. *J. High Energy Phys.* 5:26. doi: 10.1007/JHEP05(2013)026
- Herranen, M., Markkanen, T., Nurmi, S., and Rajantie, A. (2014). Spacetime curvature and the Higgs stability during inflation. *Phys. Rev. Lett.* 113:211102. doi: 10.1103/PhysRevLett.113.211102
- Herranen, M., Markkanen, T., Nurmi, S., and Rajantie, A. (2015). Spacetime curvature and Higgs stability after inflation. *Phys. Rev. Lett.* 115:241301. doi: 10.1103/PhysRevLett.115.241301
- Herranen, M., Markkanen, T., and Tranberg, A. (2014). Quantum corrections to scalar field dynamics in a slow-roll space-time. *J. High Energy Phys.* 5:26. doi: 10.1007/JHEP05(2014)026
- Holthausen, M., Lim, K. S., and Lindner, M. (2012). Planck scale Boundary Conditions and the Higgs Mass. *J. High Energy Phys.* 2:37. doi: 10.1007/JHEP02(2012)037
- Hook, A., Kearney, J., Shakya, B., and Zurek, K. M. (2015). Probable or improbable universe? Correlating electroweak vacuum instability with the scale of inflation. *J. High Energy Phys.* 1:61. doi: 10.1007/JHEP01(2015)061

- Hu, B. L., and O'Connor, D. J. (1984). Effective Lagrangian for  $\lambda\phi^4$  Theory in Curved Space-time With Varying Background Fields: Quasilocal Approximation. *Phys. Rev. D* 30:743. doi: 10.1103/PhysRevD.30.743
- Hung, P. Q. (1979). Vacuum instability and new constraints on fermion masses. *Phys. Rev. Lett.* 42:873. doi: 10.1103/PhysRevLett.42.873
- Isidori, G., Ridolfi, G., and Strumia, A. (2001). On the metastability of the standard model vacuum. *Nucl. Phys. B* 609, 387–409. doi: 10.1016/S0550-3213(01)00302-9
- Isidori, G., Rychkov, V. S., Strumia, A., and Tetradis, N. (2008). Gravitational corrections to standard model vacuum decay. *Phys. Rev. D* 77:025034. doi: 10.1103/PhysRevD.77.025034
- Jack, I., and Parker, L. (1985). Proof of summed form of proper time expansion for propagator in curved space-time. *Phys. Rev. D* 31:2439. doi: 10.1103/PhysRevD.31.2439
- Jegerlehner, F. (2014). The standard model as a low-energy effective theory: what is triggering the Higgs mechanism? *Acta Phys. Polon. B* 45:1167. doi: 10.5506/APhysPolB.45.1167
- Joti, A., Katsis, A., Loupas, D., Salvio, A., Strumia, A., Tetradis, N., et al. (2017). (Higgs) vacuum decay during inflation. *J. High Energy Phys.* 7:58. doi: 10.1007/JHEP07(2017)058
- Kamada, K. (2015). Inflationary cosmology and the standard model Higgs with a small Hubble induced mass. *Phys. Lett. B* 742, 126–135. doi: 10.1016/j.physletb.2015.01.024
- Kastening, B. M. (1992). Renormalization group improvement of the effective potential in massive  $\phi^4$  theory. *Phys. Lett. B* 283, 287–292. doi: 10.1016/0370-2693(92)90021-U
- Kawasaki, M., Mukaida, K., and Yanagida, T. T. (2016). Simple cosmological solution to the Higgs field instability problem in chaotic inflation and the formation of primordial black holes. *Phys. Rev. D* 94:063509. doi: 10.1103/PhysRevD.94.063509
- Kearney, J., Yoo, H., and Zurek, K. M. (2015). Is a higgs vacuum instability fatal for high-scale inflation? *Phys. Rev. D* 91:123537. doi: 10.1103/PhysRevD.91.123537
- Kehagias, A., and Riotto, A. (2014). Remarks about the tensor mode detection by the BICEP2 collaboration and the super-planckian excursions of the inflaton field. *Phys. Rev. D* 89:101301. doi: 10.1103/PhysRevD.89.101301
- Khan, N., and Rakshit, S. (2014). Study of electroweak vacuum metastability with a singlet scalar dark matter. *Phys. Rev. D* 90:113008. doi: 10.1103/PhysRevD.90.113008
- Khan, N., and Rakshit, S. (2015). Constraints on inert dark matter from the metastability of the electroweak vacuum. *Phys. Rev. D* 92:055006. doi: 10.1103/PhysRevD.92.055006
- Kirsten, K., Cognola, G., and Vanzo, L. (1993). Effective lagrangian for selfinteracting scalar field theories in curved space-time. *Phys. Rev. D* 48, 2813–2822. doi: 10.1103/PhysRevD.48.2813
- Klinkhamer, F. R. (2013). Standard model higgs field and energy scale of gravity. *JETP Lett.* 97, 297–300. doi: 10.1134/S002136401306009X
- Kobakhidze, A., and Spencer-Smith, A. (2013). Neutrino masses and higgs vacuum stability. *J. High Energy Phys.* 8:36. doi: 10.1007/JHEP08(2013)036
- Kobakhidze, A., and Spencer-Smith, A. (2013). Electroweak vacuum (In)Stability in an inflationary universe. *Phys. Lett. B* 722, 130–134. doi: 10.1016/j.physletb.2013.04.013
- Kobakhidze, A., and Spencer-Smith, A. (2014). The Higgs vacuum is unstable. *arXiv [Preprint]:1404.4709*.
- Koehn, M., Lavrelashvili, G., Lehnert, J.-L. (2015). Towards a solution of the negative mode problem in quantum tunnelling with gravity. *Phys. Rev. D* 92:023506. doi: 10.1103/PhysRevD.92.023506
- Kofman, L., Linde, A. D., and Starobinsky, A. A. (1994). Reheating after inflation. *Phys. Rev. Lett.* 73:3195–3198. doi: 10.1103/PhysRevLett.73.3195
- Kofman, L., Linde, A. D., and Starobinsky, A. A. (1997). Towards the theory of reheating after inflation. *Phys. Rev. D* 56, 3258–3295. doi: 10.1103/PhysRevD.56.3258
- Kohri, K., and Matsui, H. (2016). Higgs vacuum metastability in primordial inflation, preheating, and reheating. *Phys. Rev. D* 94:103509. doi: 10.1103/PhysRevD.94.103509
- Kohri, K., and Matsui, H. (2017). Electroweak vacuum instability and renormalized higgs field vacuum fluctuations in the inflationary universe. *J. Cosmol. Astropart. Phys.* 1708:11. doi: 10.1088/1475-7516/2017/08/011
- Krasnikov, N. V. (1978). Restriction of the fermion mass in gauge theories of weak and electromagnetic interactions. *Yad Fiz.* 28, 549–551.
- Krive, I. V., and Linde, A. D. (1976). On the vacuum stability problem in gauge theories. *Nucl. Phys. B* 117, 265–268. doi: 10.1016/0550-3213(76)90573-3
- Kunimitsu, T., and Yokoyama, J. (2012). Higgs condensation as an unwanted curvaton. *Phys. Rev. D* 86:083541. doi: 10.1103/PhysRevD.86.083541
- Kusenko, A., Pearce, L., and Yang, L. (2015). Postinflationary Higgs relaxation and the origin of matter-antimatter asymmetry. *Phys. Rev. Lett.* 114:061302. doi: 10.1103/PhysRevLett.114.061302
- Landau, L. (1955). *Niels Bohr and the Development of Physics: Essays Dedicated to Niels Bohr on the Occasion of His Seventieth Birthday*. Reading, PA: McGraw-Hill.
- Lavrelashvili, G. (2006). The Number of negative modes of the oscillating bounces. *Phys. Rev. D* 73:083513. doi: 10.1103/PhysRevD.73.083513
- Lavrelashvili, G. V., Rubakov, V. A., and Tinyakov, P. G. (1985). Tunneling transitions with gravitation: breaking of the quasiclassical approximation. *Phys. Lett. B* 161, 280–284. doi: 10.1016/0370-2693(85)90761-0
- Lebedev, O., and Westphal, A. (2013). Metastable electroweak vacuum: implications for inflation. *Phys. Lett. B* 719, 415–418. doi: 10.1016/j.physletb.2012.12.069
- Lee, B.-H., Lee, W., Ro, D., and Yeom, D.-H. (2015). Oscillating Fubini instantons in curved space. *Phys. Rev. D* 91:124044. doi: 10.1103/PhysRevD.91.124044
- Lee, B.-H., Lee, W., Ro, D., and Yeom, D.-H. (2017). Tunneling without a barrier: fubini instantons in curved space. in *Proceedings, 14th Marcel Grossmann Meeting on Recent Developments in Theoretical and Experimental General Relativity, Astrophysics, and Relativistic Field Theories (MG14) (In 4 Volumes): July 12-18, 2015 (Rome)*, 2766–2768.
- Lee, H., and Weinberg, E. J. (2014). Negative modes of Coleman-De Luccia bounces. *Phys. Rev. D* 90:124002. doi: 10.1103/PhysRevD.90.124002
- Lee, K.-M., and Weinberg, E. J. (1986). Tunneling without barriers. *Nucl. Phys. B* 267, 181–202. doi: 10.1016/0550-3213(86)90150-1
- Liddle, A. R., and Leach, S. M. (2003). How long before the end of inflation were observable perturbations produced?. *Phys. Rev. D* 68:103503. doi: 10.1103/PhysRevD.68.103503
- Linde, A. D. (1980). Vacuum instability, cosmology and constraints on particle masses in the Weinberg-Salam model. *Phys. Lett. B* 92, 119–121. doi: 10.1016/0370-2693(80)90318-4
- Lindner, M. (1986). Implications of triviality for the standard model. *Z. Phys. C* 31:295. doi: 10.1007/BF01479540
- Lindner, M., Sher, M., and Zaglauer, H. W. (1989). Probing vacuum stability bounds at the fermilab collider. *Phys. Lett. B* 228, 139–143. doi: 10.1016/0370-2693(89)90540-6
- Liu, C., and Zhao, Z.-H. (2013).  $\theta_{13}$  and the Higgs mass from high scale supersymmetry. *Commun. Theor. Phys.* 59, 467–471. doi: 10.1088/0253-6102/59/4/14
- Losic, B., and Unruh, W. G. (2005). Long-wavelength metric backreactions in slow-roll inflation. *Phys. Rev. D* 72:123510. doi: 10.1103/PhysRevD.72.123510
- Maiani, L., Parisi, G., and Petronzio, R. (1978). Bounds on the number and masses of quarks and leptons. *Nucl. Phys. B* 136, 115–124. doi: 10.1016/0550-3213(78)90018-4
- Markkanen, T. (2018). Light scalars on cosmological backgrounds. *J. High Energy Phys.* 1:116. doi: 10.1007/JHEP01(2018)116
- Markkanen, T., Nurmi, S., and Rajantie, A. (2017). Do metric fluctuations affect the Higgs dynamics during inflation? *J. Cosmol. Astropart. Phys.* 1712:26. doi: 10.1088/1475-7516/2017/12/026
- Markkanen, T., Nurmi, S., Rajantie, A., and Stopyra, S. (2018). The 1-loop effective potential for the Standard Model in curved spacetime. *J. High Energy Phys.* 6:40. doi: 10.1007/JHEP06(2018)040
- Martin, S. P. (2014). Three-loop Standard Model effective potential at leading order in strong and top Yukawa couplings. *Phys. Rev. D* 89:013003. doi: 10.1103/PhysRevD.89.013003
- Masina, I. (2013). Higgs boson and top quark masses as tests of electroweak vacuum stability. *Phys. Rev. D* 87:053001. doi: 10.1103/PhysRevD.87.053001
- Masoumi, A., and Weinberg, E. J. (2012). Bounces with  $O(3) \times O(2)$  symmetry. *Phys. Rev. D* 86:104029. doi: 10.1103/PhysRevD.86.104029
- Minakshisundaram, S., and Pleijel, A. (1949). Some properties of the eigenfunctions of the laplace-operator on riemannian manifolds. *Can. J. Math.* 1, 242–256. doi: 10.4153/CJM-1949-021-5
- Misner, C. W., Thorne, K. S., and Wheeler, J. A. (1973). *Gravitation*. San Francisco, CA: W. H. Freeman.

- Moss, I. G. (2015). Higgs boson cosmology. *Contemp. Phys.* 56, 468–476. doi: 10.1080/00107514.2015.1058543
- Mukaida, K., and Yamada, M. (2017). False vacuum decay catalyzed by black holes. *Phys. Rev. D* 96:103514. doi: 10.1103/PhysRevD.96.103514
- Mukhanov, V., and Winitzki, S. (2007). *Introduction to Quantum Effects in Gravity*. Cambridge: Cambridge University Press.
- Mukhanov, V. F., Abramo, L. R. W., and Brandenberger, R. H. (1997). On the Back reaction problem for gravitational perturbations. *Phys. Rev. Lett.* 78:1624–1627. doi: 10.1103/PhysRevLett.78.1624
- Nie, S., and Sher, M. (1999). Vacuum stability bounds in the two Higgs doublet model. *Phys. Lett. B* 449, 89–92. doi: 10.1016/S0370-2693(99)00019-2
- Nielsen, H. B. (2012). PREDICTED the higgs mass. *Bled Workshops Phys.* 13, 94–126.
- Nielsen, N. K. (1975). On the Gauge Dependence of Spontaneous Symmetry Breaking in Gauge Theories. *Nucl. Phys. B* 101, 173–188. doi: 10.1016/0550-3213(75)90301-6
- Odintsov, S. D. (1991). Renormalization group, effective action and grand unification theories in curved space-time. *Fortsch. Phys.* 39, 621–641. doi: 10.1002/prop.2190390803
- Odintsov, S. D. (1993). Two loop effective potential in quantum field theory in curved space-time. *Phys. Lett. B* 306, 233–236. doi: 10.1016/0370-2693(93)90073-Q
- Onemli, V. K., and Woodard, R. P. (2004). Quantum effects can render  $w < -1$  on cosmological scales. *Phys. Rev. D* 70:107301. doi: 10.1103/PhysRevD.70.107301
- Parker, L., and Toms, D. J. (1985). New form for the coincidence limit of the Feynman propagator, or heat kernel, in curved space-time. *Phys. Rev. D* 31:953. doi: 10.1103/PhysRevD.31.953
- Parker, L. E., and Toms, D. (2009). *Quantum Field Theory in Curved Spacetime*. Cambridge Monographs on Mathematical Physics. Cambridge: Cambridge University Press.
- Pearce, L., Yang, L., Kusenkov, A., and Peloso, M. (2015). Leptogenesis via neutrino production during Higgs condensate relaxation. *Phys. Rev. D* 92:023509. doi: 10.1103/PhysRevD.92.023509
- Peskin M. E., and Schroeder, D. V. (1995). *An Introduction to Quantum Field Theory*. Reading, PA: Addison-Wesley.
- Plascencia, A. D., and Tamarit, C. (2016). Convexity, gauge-dependence and tunneling rates. *J. High Energy Phys.* 10:99. doi: 10.1007/JHEP10(2016)099
- Politzer, H. D., and Wolfram, S. (1979). Bounds on particle masses in the Weinberg-Salam Model. *Phys. Lett. B* 82, 242–246. doi: 10.1016/0370-2693(79)90746-9
- Postma, M., and van de Vis, J. (2017). Electroweak stability and non-minimal coupling. *J. Cosmol. Astropart. Phys.* 1705:4. doi: 10.1088/1475-7516/2017/05/004
- Prokopec, T., Tornkvist, O., and Woodard, R. P. (2003). One loop vacuum polarization in a locally de Sitter background. *Ann. Phys.* 303, 251–274. doi: 10.1016/S0003-4916(03)00004-6
- Rajantie, A., and Stoprya, S. (2017). Standard model vacuum decay with gravity. *Phys. Rev. D* 95:025008. doi: 10.1103/PhysRevD.95.025008
- Rajantie, A., and Stoprya, S. (2018). Standard model vacuum decay in a de Sitter Background. *Phys. Rev. D* 97:025012. doi: 10.1103/PhysRevD.97.025012
- Riotto, A., and Sloth, M. S. (2008). On Resumming Inflationary Perturbations beyond One-loop. *J. Cosmol. Astropart. Phys.* 804:30. doi: 10.1088/1475-7516/2008/04/030
- Rodejohann, W., and Zhang, H. (2012). Impact of massive neutrinos on the Higgs self-coupling and electroweak vacuum stability. *J. High Energy Phys.* 6:22. doi: 10.1007/JHEP06(2012)022
- Rodriguez-Roman, D., and Fairbairn, M. (2018). Gravitationally produced top quarks and the stability of the electroweak vacuum during inflation. *arXiv [Preprint]*:1807.02450.
- Rubin, M. A., and Ordonez, C. R. (1983). Eigenvalues and degeneracies for N-dimensional tensor spherical harmonics. *J. Math. Phys.* 25, 2888–2894.
- Rubio, J. (2018). Higgs inflation. *arXiv [Preprint]*:1807.02376.
- Salvio, A., Strumia, A., Tetradias, N., and Urbano, A. (2016). On gravitational and thermal corrections to vacuum decay. *J. High Energy Phys.* 9:54. doi: 10.1007/JHEP09(2016)054
- Sasaki, M., Suzuki, H., Yamamoto, K., and Yokoyama, J. (1993). Superexpansionary divergence: breakdown of perturbative quantum field theory in space-time with accelerated expansion. *Class. Quant. Grav.* 10:L55–L60. doi: 10.1088/0264-9381/10/5/003
- Schrempp, B., and Wimmer, M. (1996). Top quark and Higgs boson masses: interplay between infrared and ultraviolet physics. *Prog. Part. Nucl. Phys.* 37, 1–90. doi: 10.1016/0146-6410(96)00059-2
- Schwinger, J. (1951). On gauge invariance and vacuum polarization. *Phys. Rev.* 82:664. doi: 10.1103/PhysRev.82.664
- Seeley, R. T. (1967). Complex powers of an elliptic operator. *Proc. Symp. Pure Math.* 10, 288–307. doi: 10.1090/pspum/010/0237943
- Seery, D. (2010). Infrared effects in inflationary correlation functions. *Class. Quant. Grav.* 27:124005. doi: 10.1088/0264-9381/27/12/124005
- Serreau, J. (2011). Effective potential for quantum scalar fields on a de Sitter geometry. *Phys. Rev. Lett.* 107:191103. doi: 10.1103/PhysRevLett.107.191103
- Sher, M. (1989). Electroweak Higgs potentials and vacuum stability. *Phys. Rept.* 179, 273–418. doi: 10.1016/0370-1573(89)90061-6
- Sher, M. (1993). Precise vacuum stability bound in the standard model. *Phys. Lett. B* 317, 159–163. doi: 10.1016/0370-2693(93)91586-C
- Shkerin, A., and Sibiryakov, S. (2015). On stability of electroweak vacuum during inflation. *Phys. Lett. B* 746, 257–260. doi: 10.1016/j.physletb.2015.05.012
- Spencer-Smith, A. (2014). Higgs vacuum stability in a mass-dependent renormalisation scheme. *arXiv [Preprint]*:1405.1975.
- Starobinsky, A. A. (1986). Stochastic de sitter (inflationary) stage in the early universe. *Lect. Notes Phys.* 246, 107–126. doi: 10.1007/3-540-16452-9\_6
- Starobinsky, A. A., and Yokoyama, J. (1994). Equilibrium state of a selfinteracting scalar field in the De Sitter background. *Phys. Rev. D* 50, 6357–6368. doi: 10.1103/PhysRevD.50.6357
- Symanzik, K. (1970). Small distance behavior in field theory and power counting. *Commun. Math. Phys.* 18, 227–246. doi: 10.1007/BF01649434
- Symanzik, K. (1971). Small distance behavior analysis and Wilson expansion. *Commun. Math. Phys.* 23, 49–86. doi: 10.1007/BF01877596
- Tagirov, E. A. (1973). Consequences of field quantization in de Sitter type cosmological models. *Ann. Phys.* 76, 561–579. doi: 10.1016/0003-4916(73)90047-X
- Tanabashi, M., et al. (2018). Review of particle physics. *Phys. Rev. D* 98:030001. doi: 10.1103/PhysRevD.98.030001
- Tanaka, T., and Sasaki, M. (1992). False vacuum decay with gravity: negative mode problem. *Prog. Theor. Phys.* 88, 503–528. doi: 10.1143/ptp/88.3.503
- Tang, Y. (2013). Vacuum stability in the standard model. *Mod. Phys. Lett. A* 28:1330002. doi: 10.1142/S0217732313300024
- Tenkanen, T., and Vaskonen, V. (2016). Reheating the standard model from a hidden sector. *Phys. Rev. D* 94:083516. doi: 10.1103/PhysRevD.94.083516
- Tetradis, N. (2016). Black holes and Higgs stability. *J. Cosmol. Astropart. Phys.* 1609:36. doi: 10.1088/1475-7516/2016/09/036
- Toms, D. J. (1982). Renormalization of interacting scalar field theories in curved space-time. *Phys. Rev. D* 26:2713. doi: 10.1103/PhysRevD.26.2713
- Toms, D. J. (1983). The effective action and the renormalization group equation in curved space-time. *Phys. Lett. B* 126, 37–40. doi: 10.1016/0370-2693(83)90011-4
- Tranberg, A. (2008). Quantum field thermalization in expanding backgrounds. *J. High Energy Phys.* 11:37. doi: 10.1088/1126-6708/2008/11/037
- Tsujikawa, S., Maeda, K.-I., and Torii, T. (1999). Resonant particle production with nonminimally coupled scalar fields in preheating after inflation. *Phys. Rev. D* 60:063515. doi: 10.1103/PhysRevD.60.063515
- Weinberg, E. J. (2006). New bounce solutions and vacuum tunneling in de Sitter spacetime. *AIP Conf. Proc.* 805, 259–265. doi: 10.1063/1.2149708
- Weinberg, E. J., and Wu, A.-Q. (1987). Understanding complex perturbative effective potentials. *Phys. Rev. D* 36:2474. doi: 10.1103/PhysRevD.36.2474

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Markkanen, Rajantie and Stoprya. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# The Top-Quark Mass: Challenges in Definition and Determination

Gennaro Corcella\*

Laboratori Nazionali di Frascati, INFN, Frascati, Italy

The top-quark mass is a parameter of paramount importance in particle physics, playing a crucial role in the electroweak precision tests and in the stability of the Standard Model vacuum. I will discuss the main strategies to extract the top-quark mass at the LHC and the interpretation of the measurements in terms of well-posed top-mass definitions, taking particular care about renormalon ambiguities, progress in Monte Carlo event generators for top physics and theoretical uncertainties.

**Keywords:** colliders, heavy quarks, Monte Carlo generators, QCD calculations, standard model

## 1. INTRODUCTION

The mass of the top quark is a fundamental parameter of the Standard Model, since it enters in the electroweak precision tests [1] and constrained the mass of the Higgs boson even before its discovery at the LHC. It plays a role in Higgs inflation model (see [2, 3] for some recent work on the subject), while the property of the electroweak vacuum to lie on the boundary between stability and metastability regimes [4] does depend on the actual values and definitions of top and Higgs masses used in the computation<sup>1</sup>. Also, in the determination of the lifetime of the Universe, undertaken in Andreassen et al. [5], part of the uncertainty is related to the top-quark mass.

In such calculations, one typically assumes that the measured top-quark mass, whose current world average reads  $m_t = [173.34 \pm 0.27(\text{stat}) \pm 0.71(\text{syst})]$  GeV [6], corresponds to the pole mass and eventually adds errors of the order of few hundreds MeV to account for possible deviations from this identification. For instance, possible changes of the central value or of the uncertainty on  $m_t$  may affect the results in Degrand et al. [4], to the point of even moving the vacuum position inside the stability or instability regions. It is therefore of paramount importance determining  $m_t$  at the LHC with the highest possible precision, estimating reliably all sources of uncertainty and eventually interpreting the results in terms of field-theory mass definitions.

More generally, the top-quark mass is determined by comparing experimental data with theory predictions, so that the measured mass has to be identified with the parameter  $m_t$  employed in the calculations. From the viewpoint of the techniques used in the extraction, one usually labels as “standard measurements” those relying on the direct reconstruction of the top-decay products by means of the template, matrix-element or ideogram methods, and as “alternative measurements” the top-mass determinations which use suitably defined observables, such as total production cross section or peaks/endpoints of differential distributions. It is remarkable noticing that, up to now, such classes of mass determinations have never been combined.

From the theory side, as most top-mass extractions use Monte Carlo shower codes, one traditionally defines “Monte Carlo mass” the quantity which is determined. On the other hand,

## OPEN ACCESS

### Edited by:

Alberto Salvio,  
European Organization for Nuclear  
Research (CERN), Switzerland

### Reviewed by:

Frank Krauss,  
Durham University, United Kingdom  
Choong Sun (C. S.) Kim,  
Yonsei University, South Korea

### \*Correspondence:

Gennaro Corcella  
gennaro.corcella@lnf.infn.it

### Specialty section:

This article was submitted to  
High-Energy and Astroparticle  
Physics,  
a section of the journal  
Frontiers in Physics

**Received:** 28 January 2019

**Accepted:** 25 March 2019

**Published:** 24 April 2019

### Citation:

Corcella G (2019) The Top-Quark  
Mass: Challenges in Definition and  
Determination. *Front. Phys.* 7:54.  
doi: 10.3389/fphy.2019.00054

<sup>1</sup> Strictly speaking the stability of the electroweak vacuum also depends on whether there is New Physics up to the Planck scale or not. Degrand et al. [4] assumes that the Standard Model is valid up to the Planck scale; other alternatives are discussed in Branchina et al. [3].

one refers to pole- or  $\overline{\text{MS}}$ -mass extraction whenever a measurement is compared with a fixed-order, possibly resummed QCD calculation employing a given field-theory mass definition. The distinction between Monte Carlo and well-posed mass definitions like the pole mass has been the core of several discussions within the top-quark physics community, as we have authors trying to quantify the discrepancy between such masses, finding results of the order of a few hundreds of MeV [7–11] and others who instead present arguments against the classification of some measurements as Monte Carlo mass determinations [12, 13] and try to interpret them still as pole-mass extractions, with an uncertainty which depends on the specific measurement strategy and details of the event generation. Furthermore, as will be discussed later, even the so-called pole or  $\overline{\text{MS}}$  mass determinations are not completely Monte Carlo independent, since the evaluation of the experimental acceptance depends, though quite mildly, on the shower code which is employed and on the implemented mass parameter.

Another issue that was often used to argue against the employment of the pole mass has been the infrared renormalon ambiguity [14, 15], namely the factorial growth of the coefficients of the expansion in powers of the strong coupling of the heavy-quark self energy, whenever it is expressed in terms of the pole mass. However, recent work on this topic [16, 17] showed that, using the 4-loop relation between pole and (renormalon-free)  $\overline{\text{MS}}$  masses [18], the renormalon ambiguity is actually of the order at most of 250 MeV, hence smaller than the current error on the top mass. Furthermore, although the projections for the future high-energy and high-luminosity runs of the LHC aim at even lower uncertainties, it should always be reminded that the top quark is an unstable particle with a width of the order of 1 GeV which, as long as it is included in the computation, acts as a cutoff for radiation off top quarks<sup>2</sup>.

In the following, I shall give an overview of the up-to-date top-mass determinations and, above all, I will try to stress the main points of the existing controversies concerning mass definitions and interpretation of the LHC measurements, as well as the sources of theory uncertainty. In section 2 I shall review the heavy-quark mass definitions; in section 3 I will discuss the renormalon ambiguity; in section 4 the main strategies to measure the top mass will be presented. The interpretation of the measurements and the theoretical uncertainties will be investigated in section 5, while section 6 will contain some final remarks.

## 2. TOP-QUARK MASS DEFINITIONS

Heavy-quark mass definitions are related to how one subtracts the ultraviolet divergences in the renormalized heavy-quark self energy  $\Sigma^R$ . Higher-order corrections to the self energy are typically calculated in dimensional regularization, with  $d = 4 - 2\epsilon$  dimensions. At one loop in QCD, for a heavy quark with four-momentum  $p$  and bare mass  $m_0$ , the renormalized self energy

reads:

$$\Sigma^R(m_0, p, \mu) = \frac{i\alpha_S}{4\pi} \left\{ \left[ \frac{1}{\epsilon} - \gamma + \ln 4\pi + A(m_0, p, \mu) \right] \not{p} - \left[ 4 \left( \frac{1}{\epsilon} - \gamma + \ln 4\pi \right) + B(m_0, p, \mu) \right] m_0 \right\} + i[(Z_2 - 1)\not{p} - (Z_2 Z_m - 1)m_0] + \mathcal{O}(\alpha_S^2), \quad (1)$$

where  $Z_2$  and  $Z_m$  are the wave-function and mass renormalization constants, respectively,  $\gamma = 0.577216\dots$  the Euler–Mascheroni constant and  $\mu$  is the renormalization scale<sup>3</sup>. The functions  $A$  and  $B$  in Equation (1) depend on  $p$ ,  $m_0$  and  $\mu$  and are independent of  $\epsilon$ . The bare heavy-quark propagator is  $S^0(p) = i/(\not{p} - m_0)$ , while the renormalized  $S^R$  reads, in terms of the renormalized self energy:

$$S^R(p, \mu) = \frac{i}{\not{p} - m_0 - i\Sigma^R(m_0, p, \mu)}. \quad (2)$$

The on-shell renormalization scheme, leading to the pole mass, is defined so that the self energy and its partial derivative with respect to  $\not{p}$  vanish whenever  $\not{p} = 0$ :

$$\Sigma^R \Big|_{\not{p}=0} = 0; \quad \frac{\partial \Sigma^R}{\partial \not{p}} \Big|_{\not{p}=0} = 0. \quad (3)$$

The minimal-subtraction ( $\overline{\text{MS}}$ ) scheme is indeed typical of dimensional regularization and fixes  $Z_2$  and  $Z_m$  in order to subtract just the contributions  $\sim \frac{1}{\epsilon} - \gamma + \ln 4\pi$  in Equation (1)<sup>4</sup>.

Since pole and  $\overline{\text{MS}}$  masses are the most popular top-mass schemes, hereafter I will devote some discussion on such definitions. In the on-shell (o.s.) and  $\overline{\text{MS}}$  schemes  $S^R(p)$  can then be expressed in terms of pole and  $\overline{\text{MS}}$  masses, respectively, as follows:

$$S_{\text{o.s.}}^R(p) \simeq \frac{i}{\not{p} - m_{\text{pole}}}, \quad S_{\overline{\text{MS}}}^R(p, \mu) \simeq \frac{i}{\not{p} - m_{\overline{\text{MS}}}(\mu) - (A - B)m_{\overline{\text{MS}}}(\mu)}. \quad (4)$$

From Equation (4), one can learn that  $m_{\text{pole}}$  is still the pole of the propagator, even after the renormalization procedure, which is in agreement with the intuitive notion of the mass of a free particle, whereas  $m_{\overline{\text{MS}}}(\mu)$  may be quite far from the pole. Also, unlike the pole mass, the  $\overline{\text{MS}}$  mass depends on the renormalization scale  $\mu$ . The relation between top-quark pole ( $m_{t,\text{pole}}$ ) and  $\overline{\text{MS}}$  ( $\bar{m}_t(\bar{m}_t)$ )

<sup>2</sup>The latest Particle Data Group [19] quotes a top width  $\Gamma_t = (1.41^{+0.19}_{-0.15})$  GeV.

<sup>3</sup>In  $d$  dimensions, the coupling  $g_S$ , related to  $\alpha_S$  via  $\alpha_S = g_S^2/(4\pi)$ , gets mass dimension  $\epsilon$ , i.e.,  $g_S \rightarrow g_S \mu_\epsilon^\epsilon$ ,  $\mu_\epsilon$  being a regularization scale. After adding suitable counter-terms,  $\Sigma^R$  is eventually expressed in terms of the renormalization scale  $\mu$ .  
<sup>4</sup>Alternatively to working in  $d$  dimensions, one can use a mass regularization scheme, giving the gluon a fictitious mass  $\lambda$ . The renormalized self energy with a gluon mass  $\lambda$  can be obtained from Equation (1) by means of the replacement:  $1/\epsilon - \gamma + \ln[(4\pi\mu^2)/m_0^2] \rightarrow \ln(\lambda^2/m_0^2)$ .

masses was calculated up to four loops in Marquard et al. [18] and reads:

$$\begin{aligned} m_{t,\text{pole}} &= \bar{m}_t(\bar{m}_t) \left[ 1 + 0.4244 \alpha_S + 0.8345 \alpha_S^2 + 2.375 \alpha_S^3 \right. \\ &\quad \left. + (8.615 \pm 0.017) \alpha_S^4 + \mathcal{O}(\alpha_S^5) \right] \\ &= [163.508 + 7.529 + 1.606 + 0.496 + (0.195 \pm 0.0004)] \\ &\quad \text{GeV}. \end{aligned} \quad (5)$$

The last term in Equation (5) yields an uncertainty of about 200 MeV on the pole- $\overline{\text{MS}}$  conversion. Beyond four loops, one can find in Kataev and Molokoedov [20] the dependence of the five- and six-loop corrections to the pole- $\overline{\text{MS}}$  relation on the number of light flavors.

As discussed in the introduction, higher-order corrections to the self energy, when expressed in terms of the pole mass, lead to infrared renormalons [14], namely the factorial growth of the coefficients of  $\alpha_S^n$ : we shall discuss recent calculations on renormalons in the next section. For the time being, I just point out that the  $\overline{\text{MS}}$  mass is renormalon-free and it is therefore a so-called short-distance mass, well defined in the infrared regime. However, differently from the pole mass, it is not a suitable mass definition at threshold, as it exhibits corrections  $(\alpha_S/v)^k$ ,  $v$  being the top velocity, that are large in the threshold limit  $v \rightarrow 0$ . On the contrary, the  $\overline{\text{MS}}$  mass is appropriate to describe processes far from threshold, i.e., at scales  $Q \gg m_t$  for top quarks, since, by setting the renormalization scale  $\mu \simeq Q$ , one is capable of resumming large logarithms  $\ln(Q^2/m_t^2)$  in the mass definition itself. As will be highlighted in the next section, Equation (5), relating the pole mass to the renormalon-free  $\overline{\text{MS}}$  one, can be used as a starting point to evaluate the renormalon ambiguity in the top pole mass.

Another mass definition, which has been employed especially in the framework of Soft Collinear Effective Theory (SCET), is the so-called MSR mass, which was introduced to interpolate between pole and  $\overline{\text{MS}}$  masses [7]. Such a mass, labeled as  $m_t^{\text{MSR}}(R, \mu)$  for top quarks, besides the renormalization scale  $\mu$ , depends on an extra scale  $R$ , in such a way that:

$$\begin{aligned} m_t^{\text{MSR}}(R) &\rightarrow m_{t,\text{pole}} \text{ for } R \rightarrow 0 \text{ and } m_t^{\text{MSR}}(R) \rightarrow \bar{m}_t(\bar{m}_t) \text{ for} \\ &R \rightarrow \bar{m}_t(\bar{m}_t). \end{aligned} \quad (6)$$

The MSR mass can be related to any other mass definitions, such as the pole mass, by means of a counterterm like:

$$m_{t,\text{pole}} = m_t^{\text{MSR}}(R, \mu) + \delta m_t(R, \mu), \quad (7)$$

where the  $\mu$ -dependence of  $m_t^{\text{MSR}}(R, \mu)$  follows renormalization group equations. As will be argued in the following, the MSR mass has often been adopted in the literature to connect the top-mass measurements with well-defined top-mass definitions, with  $R \sim \mathcal{O}(1 \text{ GeV})$ .

For the sake of generality, although the present review will be mostly devoted to hadron-collider top-mass determinations, I wish to remind some other top mass definitions which are often employed in analyses on the  $m_t$  extraction at future lepton colliders. In fact, physical observables at threshold, such the  $t\bar{t}$

cross section in  $e^+e^-$  collisions at  $\sqrt{s} \simeq 2m_t$ , require suitable mass schemes. One of such definitions is the 1S mass, defined as half the mass of a fictitious  $\Upsilon(1S)$  resonance, made up of a bound  $t\bar{t}$  state [21]:

$$m_{t,1S} = \frac{1}{2} \{m[\Upsilon(1S)]\}. \quad (8)$$

The 1S mass reads, in terms of the pole mass (Hoang et al. [22]):

$$m_{t,1S} = m_{t,\text{pole}} (1 - \Delta^{\text{LL}} - \Delta^{\text{NLL}} - \Delta^{\text{NNLL}}). \quad (9)$$

The explicit expression of the  $\Delta$  terms can be found in Hoang et al. [22], where the threshold  $e^+e^- \rightarrow t\bar{t}$  cross section was computed in the next-to-next-to-leading logarithmic approximation, and the superscripts LL, NLL and NNLL refer to the resummation of large logarithms of the top velocity  $v$ , which are large in the regime  $v \sim \alpha_S \ll 1$  and  $\alpha_S \ln v \sim 1$ .

The potential-subtracted (PS) mass is instead constructed in terms of the  $t\bar{t}$  Coulomb potential, in such a way that contributions below a factorization scale  $\mu_F$  are subtracted off, as to suppress renormalons [23]:

$$m_{\text{PS}}(\mu_F) = m_{\text{pole}} - \frac{1}{2} \int_{|q| < \mu_F} \frac{d^3q}{(2\pi)^3} \tilde{V}(q). \quad (10)$$

In Equation (10)  $\tilde{V}(q)$  is the transform in momentum space of the  $t\bar{t}$  Coulomb potential. The PS mass is a threshold mass too, particularly suitable to deal with  $t\bar{t}$  production at energies slightly above  $2m_t$ . The relation between PS and pole top-quark masses is given by the following equation [24]:

$$m_{t,\text{PS}}(\mu_F) = m_{t,\text{pole}} - \frac{4}{3\pi} \alpha_S(\mu_F) \mu_F + \mathcal{O}(\alpha_S^2). \quad (11)$$

More recently, the theoretical error on the possible extraction of 1S and PS masses in  $e^+e^-$  collisions just above the  $t\bar{t}$  threshold was estimated. In detail, by using a NNLL threshold resummation of the ratio  $R = \sigma(e^+e^- \rightarrow t\bar{t})/\sigma(e^+e^- \rightarrow \mu^+\mu^-)$ , the 1S mass can be extracted with an uncertainty about 40 MeV [25], whereas, by employing a fixed-order NNNLO calculation, the PS mass can be determined with an error below 50 MeV [26]. It will be of course desirable to combine such fixed-order and resummed computations to possibly decrease further such an uncertainty.

Another threshold mass definition is the renormalon-subtracted (RS) mass, which removes from the pole mass the pure renormalon contribution [27]. The RS mass was determined in Pineda [27] after constructing its Borel transform and reads, in terms of the pole mass:

$$m_{t,\text{RS}} = m_{t,\text{pole}} - \sum_{n=0}^{\infty} N_n \mu_F \alpha_S^{n+1}(\mu_F) \sum_{k=0}^{\infty} c_k \frac{\Gamma(n+1+b-k)}{\Gamma(1+b-k)}, \quad (12)$$

where the expression for the coefficients  $N_n$  and  $c_k$  are given in Pineda [27] and  $b$  can be expressed in terms of the QCD  $\beta$ -function as  $b = \beta_1/(2\beta_0^2)$ . Potential-, renormalon-subtracted and 1S top-quark masses were related to the  $\overline{\text{MS}}$  mass in Marquard et al. [18] with four-loop accuracy in the conversion.

The uncertainty in the conversion was gauged about 7, 11 and 23 MeV for PS, RS and 1S masses, respectively.

Finally, the so-called kinetic mass was defined in Bigi et al. [28] for the purpose of improving the convergence of the perturbative expansion of the semileptonic  $B$ -meson decay width. It was constructed by subtracting from the pole mass the HQET (Heavy Quark Effective Theory) matrix elements, denoted by  $\bar{\Lambda}(\mu)$  in Bigi et al. [28], expressing the shift between pole and meson masses. The kinetic bottom-quark mass reads, up to terms suppressed as the inverse of the quark/meson mass:

$$m_{b,\text{kin}}(\mu_F) = m_B - \bar{\Lambda}(\mu_F) + \mathcal{O}\left(\frac{1}{m_B}\right). \quad (13)$$

In Hoang et al. [24], the kinetic mass was generalized to  $t\bar{t}$  bound states, obtaining the following expansion in terms of the pole mass:

$$m_{t,\text{kin}}(\mu_F) = m_{t,\text{pole}} - \frac{16}{9\pi} \alpha_S(\mu_F) \mu_F + \mathcal{O}(\alpha_S^2). \quad (14)$$

As underlined before, the 1S, PS, and RS masses are threshold masses which, unlike the pole mass, do not exhibit the renormalon ambiguity. Recent calculations aimed at estimating the renormalon uncertainty in the pole mass will be the topic of next section.

### 3. THE RENORMALON AMBIGUITY IN THE TOP MASS

Problems with the renormalized heavy-quark self energy, when expressed in terms of the pole mass, were first understood in Beneke and Braun [14] and Beneke [15]. In fact, after including higher-order contributions in the strong coupling constant, the renormalized heavy-quark self energy exhibits the following expansion in powers of  $\alpha_S$ :

$$\Sigma^R(m_{\text{pole}}, m_{\text{pole}}) \approx m_{\text{pole}} \sum_n \alpha_S^{n+1} (2b_0)^n n!, \quad (15)$$

where  $b_0$  is first  $\beta$ -function coefficient entering in the  $\overline{\text{MS}}$  strong coupling constant<sup>5</sup>. From Equation (15), one learns that the coefficients of the expansion grow like  $n!$  at order  $\alpha_S^{n+1}$ .

After re-expressing  $\alpha_S$  in terms of the  $\beta$  function and of the QCD scale  $\Lambda$ , and inserting  $\Sigma^R$  in the on-shell propagator as in Equation (4), one will get a correction to the pole mass:

$$\Delta m_{\text{pole}} \simeq \mathcal{O}(\Lambda), \quad (16)$$

which is the renowned renormalon ambiguity in  $m_{\text{pole}}$ , i.e., an uncertainty of the order of the QCD scale in the pole-mass definition. This result can be related to the fact that a quark is not a free parton, but has to be confined into a hadron: in fact, one can prove that the renormalon uncertainty is due to

<sup>5</sup>We recall that, e.g., at LO in the  $\overline{\text{MS}}$  scheme, it is  $\alpha_S(Q^2) = 1/[b_0 \ln(Q^2/\Lambda^2)]$ ,  $\Lambda$  being the QCD scale. For  $Q^2 \sim \Lambda^2$  one hits the well-known Landau pole and perturbative QCD can no longer be applied.

the gluon self coupling, while it is not present when dealing with leptons. Therefore, the pole mass behaves like a physical mass for electrons or muons, whereas for heavy quarks it is not a short-distance mass, because of infrared renormalon effects, and one should choose on a case-by-case basis whether the pole mass or other definitions are adequate to describe a given physical process.

In order to quantify the renormalon ambiguity in the pole mass, one can employ the relation between pole and  $\overline{\text{MS}}$  masses, relying on the fact that the  $\overline{\text{MS}}$  mass is unaffected by renormalons. Equation (5) can be parametrized to all orders as in Beneke et al. [16]:

$$m_{\text{pole}} = \bar{m}(\mu_m) \left[ 1 + \sum_{n=1}^{\infty} c_n(\mu, \mu_m, \bar{m}(\mu_m)) \alpha_S^n(\mu) \right], \quad (17)$$

with  $\bar{m}(\mu_m)$  being the  $\overline{\text{MS}}$  mass at some scale  $\mu_m$  and  $\mu$  the renormalization scale at which the strong coupling is evaluated. The dominant renormalon divergence implies that the coefficients  $c_n$  in the asymptotic expansion have to satisfy the following relation at large  $n$ :

$$c_n(\mu, \mu_m, m(\mu_m)) \rightarrow N \frac{\mu}{m(\mu_m)} c_n^{\text{as}} \text{ for } n \rightarrow \infty. \quad (18)$$

The expression for the asymptotic coefficients  $c_n^{\text{as}}$  can be found in Beneke et al. [16] and is consistent with the fact that the renormalon factorial growth is due to the low-momentum region in the higher-order loop corrections to the heavy-quark self energy. The calculation of the normalization coefficient  $N$  is non trivial: in Beneke et al. [16]  $N$  was extracted after fitting the third- and fourth-order coefficient in the exact four-loop  $\overline{\text{MS}}$ -pole mass conversion and amounts to  $N \simeq 0.976 \dots$  for  $N_C = 3$  number of colors.

Furthermore, an alternative and possibly better method to deal with factorially divergent series consists in using the Borel transform, which, for a function  $f(\alpha_S)$  reads:

$$f(\alpha_S) = \sum_{n=0}^{\infty} c_n \alpha_S^{n+1}; \quad B[f](t) = \sum_{n=0}^{\infty} c_n \frac{t^n}{n!}, \quad (19)$$

which implies

$$f(\alpha_S) = \int_0^{\infty} e^{-t/\alpha_S} B[f](t) dt. \quad (20)$$

The evaluation of the Borel integral in Equation (20) depends on a prescription: one typically takes its principal value and, following the so-called “Im/Pi” method, the uncertainty is estimated as the modulus of the imaginary part, arising from the integration above and below the singular cuts in the complex plane, divided by  $\pi$ . In fact, in Beneke et al. [16] the asymptotic expansion of the pole mass with respect to the  $\overline{\text{MS}}$  one was computed as an inverse Borel transform, by using the Im/Pi method for the error, considering only three light flavors and accounting for charm and bottom masses. The final result is that

the leading renormalon ambiguity amounts to about 110 MeV for top as well as bottom and charm pole masses.

A different strategy to gauge the renormalon ambiguity was instead tackled in Hoang et al. [17], where the MSR mass  $m_{\text{MSR}}(R)$  was used. In the relation between  $m_{\text{pole}}$  and  $m_{\text{MSR}}(R)$ ,

$$m_{\text{pole}} = m_{\text{MSR}}(R) + R \sum_{n=1}^{\infty} a_n \left[ \frac{\alpha_S(R)}{4\pi} \right]^n, \quad (21)$$

the scale  $R$  is set to the  $\overline{\text{MS}}$  top mass  $m_t(\bar{m}_t)$  and the series in Equation (21) is truncated at some fixed order  $n$ . A value  $n_{\text{min}}$  is determined in such a way to minimize the difference  $\Delta(n) = m_{\text{pole}}(n) - m_{\text{pole}}(n-1)$  and a number  $f$  slightly above unity is defined. The set  $\{n\}_f$  is thus constructed in such a way that  $\Delta(n) \leq f\Delta(n_{\text{min}})$ : the midpoint of  $m_{\text{pole}}(n)$  within  $\{n\}_f$  is then chosen as the central value and half of the variation range of  $m_{\text{pole}}(n)$  as an estimate of the ambiguity, accounting for the running of the renormalization scale as well. After observing that the results depend on  $f$  rather mildly, in Hoang et al. [17]  $f = 5/4$  was chosen, yielding an ambiguity about 253 MeV in the pole mass. Both in Beneke et al. [16] and Hoang et al. [17], some thorough discussion is devoted to the inclusion of charm and bottom masses. The results of 110 and 253 MeV would go down to 70 [16] and 180 [17] MeV if one treated charm and bottom quarks as massless. Some attempts to relate the different methods adopted in Beneke et al. [16] and Hoang et al. [17] were made in Nason [13]. In fact, the result in Beneke et al. [16] can be obtained even following the method in Hoang et al. [17], but taking as central value half the sum of all  $\Delta(n)$  and setting  $f = 1 + 1/(4\pi)$  in the uncertainty evaluation.

In the following, no strong statement supporting the calculation in Beneke et al. [16] or Hoang et al. [17] will be made. I just wish to point out that, on the one hand, as long as the uncertainties in the top-mass measurement stay around 500 GeV, both renormalon determinations are smaller and should not play any role in supporting the use of a given mass definition. This may not be the case if, in future perspective, one ideally aims at precisions about 200–300 MeV. However, as will be underlined when dealing with Monte Carlo modeling and theoretical errors, recent implementations of top production and decay in shower codes include width effects [29], in such a way that the top width, about 1.4 GeV and well above the energy range of both renormalon estimates, acts as a cutoff for the radiation off top quarks<sup>6</sup>. Of course, if one considers observables relying on top decays ( $t \rightarrow bW$ ), the  $b$ -quark is allowed to emit soft radiation down to the shower cutoff and, in principle, in quantities depending on  $b$ -jets one may have to deal with renormalons.

A careful exploration of renormalon effects in observables depending on the top mass was carried out in Ferrario Ravasio et al. [30]. The authors found that the  $\overline{\text{MS}}$  mass is a better definition for quantities like the total  $t\bar{t}$  cross section, while using the pole mass would lead to a linear renormalon and an ambiguity of  $\mathcal{O}(100 \text{ MeV})$  on the  $m_t$  extraction.

<sup>6</sup>This would not be the case in codes or calculations which instead neglect width effects and interference between top-production and decay phases. In this case, even top quarks are capable of radiating down to the infrared cutoff.

Indications in favor of such a short-distance mass were also given whenever final-state jets are reconstructed using algorithms with a large jet radius  $R$ . As for the reconstructed top mass from, e.g., the  $b$ -jet+ $W$  invariant mass, in the pole-mass scheme a linear renormalon correction is present, whose coefficient is nevertheless pretty small if one employs a large  $R$  in the  $b$ -jet definition. Finally, leptonic observables exhibit a linear renormalon with both mass definitions, as long as one works in the narrow-width approximation. On the contrary, there are no linear renormalons if one adopts a short-distance mass and includes the finite top width.

## 4. TOP-QUARK MASS EXTRACTION AT LHC

Top-quark mass determinations at hadron colliders are classified as standard or alternative measurements and, according to the decay modes of the two  $W$ 's in top decays, as measurements in the dilepton, lepton+jet or all-hadronic channels. Standard top-mass analyses are based on the direct reconstruction of top-decay final states and compare observables, such as the  $b$ -jet+lepton invariant-mass distribution, with the predictions yielded by the Monte Carlo codes. So-called alternative measurements use instead other observables, such as total/differential cross sections or distribution peaks/endpoints. Since, as will be detailed in the following, Monte Carlo codes are of paramount importance for most top-mass analyses, I shall first sketch their main features, and then review the experimental methods to extract  $m_t$ .

### 4.1. Monte Carlo Generators for Top Physics

The last couple of decades has seen a tremendous progress in the implementation of Monte Carlo event generators, besides the reknowned general-purpose HERWIG [31, 32] and PYTHIA [33, 34], in such a way that several reliable programs are currently available for the top-mass analyses. On the one hand, strategies to match NLO calculations with parton showers were developed, on the other one a number of so-called matrix-element generators were released. In fact, matrix-element generators simulate multi-leg amplitudes and are interfaced to HERWIG or PYTHIA for shower and hadronization: besides top-quark signals, they are very useful to simulate backgrounds with high jet multiplicities, such as  $W/Z + n$  jets, which would be poorly described by HERWIG or PYTHIA for  $n > 1$ .

Regarding top phenomenology, standard Monte Carlo programs simulate both top production and decays using leading order (LO) matrix elements, multi-parton emission in the soft or collinear limit and the interference between top-production and decay stages is neglected (narrow-width approximation). HERWIG parton showers satisfy angular ordering [35, 36], with the latest version even allowing the option of dipole-like evolution [37]; PYTHIA cascades are instead ordered in transverse momentum<sup>7</sup>. Matrix-element corrections to parton showers are implemented for top decays [38, 39], but not

<sup>7</sup>The old PYTHIA 6 code also implements virtuality ordering, with the option to veto non-angular-ordered emissions.

for production, and the total production cross section and top-decay width are still calculated at LO. Hadronization is included by adopting the cluster model [40], based on color pre-confinement, in HERWIG and the string model [41] in PYTHIA. The underlying event used to be described assuming soft collisions between the proton spectators and tuning the model parameters to minimum-bias events at small transverse momentum. Nevertheless, all modern codes implement it through multiple scatterings strongly ordered in transverse momentum: the underlying event is thus a secondary collision, whose transverse momentum is much lower than the primary hard scattering [42, 43].

Among the new generation of Monte Carlo programs, SHERPA [44] can also be considered a multi-purpose code, in the light of the wide spectrum of processes which it is capable of simulating. In detail, matrix elements are computed by means of the AMEGIC++ [45] and COMIX [46] codes, while the interface to one-loop generators, implemented along the lines of [47], allows one to include NLO QCD and possibly electroweak corrections. Parton showers are then accounted for according to the dipole formalism developed in Catani and Seymour [48], underlying event and hadronization follow the multiple-scattering and cluster models in PYTHIA and HERWIG, respectively.

For the purpose of the matching of NLO matrix elements and multi-parton cascades, NLO+shower programs, such as MadGraph5\_aMC@NLO [49, 50] and POWHEG [51], implement NLO hard-scattering amplitudes, but still depend on HERWIG and PYTHIA for parton cascades and non-perturbative phenomena. The earlier versions of such NLO+shower algorithms only included NLO corrections to  $t\bar{t}$  production, while (LO) top decays and hadronization were still handled in the parton shower approximation. The later implementation of POWHEG [29] includes in the  $b\bar{b}4\ell$  code both top production and decay at NLO, accounting for the interference between top production and decay stages, as well as non-resonant contributions leading to  $(W^+b)(W^-\bar{b})$  final states<sup>8</sup>. As for MadGraph5\_aMC@NLO, strictly speaking, top decays are still at LO, however spin correlations are included through the MadSpin package [53] and, as discussed in Frederix et al. [54], they account for a significant part of the NLO corrections. For the purpose of HERWIG, it has its own implementation of NLO+shower merging/matching [55, 56], working for top-quark production and decay in the narrow-width approximation [57].

Regarding matrix-element generators, suitable codes to describe top-quark signals and backgrounds are, among others, ALPGEN [58], MCFM [59], CalcHEP [60], HELAC [61], and WHIZARD [62]. In particular, ALPGEN and CalcHEP simulate multi-parton final states at LO and can be interfaced to HERWIG or PYTHIA for shower and hadronization. HELAC and WHIZARD have been lately provided with NLO corrections [63, 64] and matching to shower and hadronization codes as well. MCFM is a NLO parton-level Monte Carlo code: top

production and decay are handled at NLO, in the narrow-width approximation.

Before concluding this subsection, it is worthwhile saying a few words on the precision of the predictions yielded by Monte Carlo codes. As observed before, parton showers simulate multiple radiation in the soft or collinear approximation and, in general, the accuracy of a prediction depends on the specific observable under investigation. Although total cross sections and widths are (N)LO, for most quantities Monte Carlo predictions are equivalent to leading-logarithmic resummations, i.e., they resum double soft and collinear logarithms, and include some classes of subleading logarithms, i.e., only soft- or collinear-enhanced<sup>9</sup>. Catani et al. [66] even proved that, in Deep Inelastic Scattering and Drell–Yan processes at large values of the Bjorken  $x$ , the HERWIG algorithm is capable of capturing all next-to-leading logarithms, i.e., all single logarithms, enhanced for soft or collinear emission, as long as one rescales the QCD scale  $\Lambda$  to a Monte Carlo value, labeled  $\Lambda_{MC}$ .<sup>10</sup>

## 4.2. Standard and Alternative Top-Mass Measurements

In this subsection I shall briefly present the main strategies to measure the top mass at hadron colliders in  $t\bar{t}$  events, taking particular care about the analyses carried out at the LHC.

### 4.2.1. Direct Reconstruction Methods

Strategies based on the direct reconstruction of the top-decay products, namely the template, matrix-element and ideogram methods, have been traditionally classified as standard top-mass determinations. As for ATLAS, the most up-to-date measurements are given at 8 TeV and  $19.7 \text{ fb}^{-1}$  in Aaboud et al. [67–69] for dilepton, lepton+jets and all-hadronic modes, respectively. Regarding CMS, at the moment even results at  $\sqrt{s} = 13 \text{ TeV}$  and  $\mathcal{L} = 35.9 \text{ fb}^{-1}$  are available and are reported in Sirunyan et al. [70] (dileptons), [71] (lepton+jets) and [72] (all hadrons). Regarding these analyses and summing in quadrature systematic and statistical errors, CMS quotes uncertainties about 0.73 GeV for dileptons, 0.62 GeV for leptons+jets and 0.61 GeV for the all-hadron channel. As for ATLAS, the uncertainties are 0.84 GeV (dileptons), 0.91 GeV (lepton+jets) and 0.73 GeV (all jets). The standard top-mass measurements have been the basis to determine the world average [6], already presented in the introduction, which, after summing statistical and systematic errors in quadrature, yields an overall uncertainty about 800 MeV. Work toward an updated world average is currently under way. The LHC collaborations have nevertheless released their own combined measurements using 7 and 8 TeV data together: details on such studies can be found in Aaboud et al. [69] and Khachatryan et al. [73] for ATLAS and CMS, respectively. Both analyses yield a total error about 0.5 GeV, hence an overall

<sup>8</sup>See also Heinrich et al. [52] for an independent investigation of NLO and top-width effects on the top-mass determination.

<sup>9</sup>A notable exception is given by leading non-global logarithms, sensitive to a limited portion of the phase space, which, as discussed in Banfi et al. [65], are partially accounted for by the angular-ordered showers of HERWIG, while they are mostly absent in virtuality- or transverse-momentum-ordered PYTHIA.

<sup>10</sup>With respect to  $\Lambda$  in the  $\overline{\text{MS}}$  scheme it is  $\Lambda_{MC} = \Lambda \exp[K/(4\pi b_0)]$ , with  $K = N_C(67/18 - \pi^2/6) - 5N_f/9$ , with  $N_C$  and  $N_f$  being the number of colors and active flavors, respectively.

precision on the top mass around 0.3%. **Figure 1** summarizes the state of the art on top-mass measurements carried out at the LHC, including the world average, as well as ATLAS and CMS combinations.

As discussed in the introduction, since the standard  $m_t$ -reconstruction methods rely on the use of Monte Carlo generators, such measurements are usually quoted as “Monte Carlo mass” and much debate has been taking place on whether the extracted mass can be related to any well-posed definition, with some calculable uncertainty, such as the pole mass. The ongoing discussion on the theoretical interpretation of the measured top mass will be the main topic of next session. Before moving to this issue, it is worthwhile reviewing the so-called “alternative” strategies, making use of total/differential cross sections, endpoints or other kinematic properties of  $t\bar{t}$  final states.

#### 4.2.2. Total and Differential $t\bar{t}$ Cross Section

The total  $t\bar{t}$  cross section was calculated in QCD in the NNLO+NNLL approximation in Czakon et al. [74]<sup>11</sup> and was used to determine  $m_t$  by ATLAS in Aad et al. [75] (7 and 8 TeV data) and by CMS in Khachatryan et al. [76] (7 and 8 TeV) and Sirunyan et al. [77] (13 TeV). Since the calculation in Czakon et al. [74] employed the pole mass definition, the results in Khachatryan et al. [73], Aad et al. [75], and Sirunyan et al. [77] are quoted as pole mass measurements. Although to some extent this is mostly correct, it should always be reminded that even those analyses are not completely independent of the shower generator, and therefore of its mass parameter, which is still used to evaluate the acceptance. Nevertheless, it was proved that such a sensitivity is rather mild. Overall, the errors in Aad et al. [75], Khachatryan et al. et al. [76], and Sirunyan et al. [77] are larger than those in the standard methods, as they are about 2.5 GeV; however, they are expected to decrease thanks to the higher statistics foreseen in the LHC future runs. After the computation of the total cross section, even differential distributions were calculated at NNLO in Czakon et al. [78], still using the top pole mass: this computation was used by the D0 Collaboration [79] to extract the top mass at the Tevatron accelerator, namely  $\sqrt{s} = 1.96$  TeV and  $\mathcal{L} = 9.7$  fb<sup>-1</sup>. The error on this measurement is about 2.5 GeV, hence competitive with those obtained at the LHC from the total production cross section.

Dowling and Moch [80] explored the extraction of the top mass by using the NNLO total  $t\bar{t}$  cross section and NLO differential distributions, such as transverse momentum, rapidity and  $t\bar{t}$  invariant mass, expressed in terms of pole and  $\overline{\text{MS}}$  masses. Overall, Dowling and Moch [80] found that using the running mass yields a milder scale dependence of such observables; nevertheless, implementing the full NNLO differential cross section or the four-loop pole- $\overline{\text{MS}}$  mass conversion, along the lines of Czakon et al. [78] and Marquard et al. [18], respectively, will be obviously very useful to shed light on the scale dependence.

Still on the  $t\bar{t}$  total cross section, it is worthwhile pointing out the recent work carried out to merge NNLO QCD and NLO

electroweak corrections in Czakon et al. [81]. Such a computation was then used to predict the top-quark charge asymmetry at Tevatron and LHC and the electroweak corrections exhibited a remarkable impact, say about 20%, on the forward-backward asymmetry. It will be clearly very interesting determining the top pole mass from differential distributions, along the lines of [79], including electroweak contributions as well.

#### 4.2.3. $t\bar{t}j$ Cross Section

The top mass was also extracted from the measurement of the  $t\bar{t} + 1$  jet cross section, which has a stronger sensitive to  $m_t$  than the inclusive  $t\bar{t}$  rate. In Alioli et al. [82], the NLO  $t\bar{t}j$  cross section was calculated using POWHEG and its pole mass implementation, matched to PYTHIA. Detector and shower/hadronization effects were unfolded in order to recover the pure NLO  $t\bar{t}j$  cross section. From the experimental viewpoint, the approach proposed in Alioli et al. [82] was followed in Aad et al. [83] by ATLAS (7 TeV and 5 fb<sup>-1</sup>) and by CMS in Collaboration [84] (8 TeV and 19.7 fb<sup>-1</sup>). The error on  $m_t$  extracted from the  $t\bar{t}j$  cross section is slightly smaller than from the inclusive  $t\bar{t}$  one, but still much above the direct-reconstruction measurements. Such mass determinations are referred to as pole mass measurements, since this is the mass definition employed by POWHEG, while the PYTHIA mass parameter used in the parton shower has a mild effect in the determination of the acceptance. Fuster et al. [85] used the running  $\overline{\text{MS}}$  top mass in the calculation of the NLO  $t\bar{t}j$  rate and, after comparing with the cross section measurements, obtained results which are, within the errors, in agreement with the pole mass yielded by the approach in Alioli et al. [82].

Other so-called alternative methods to reconstruct  $m_t$  rely on the kinematic properties of top-decay final states: since they are based on the comparison with Monte Carlo predictions, the measured  $m_t$  has to be identified with the mass parameter in the shower code. Overall, such techniques yield uncertainties in the mass about the order of magnitude of those relying on the total cross section, say about 1 GeV or above.

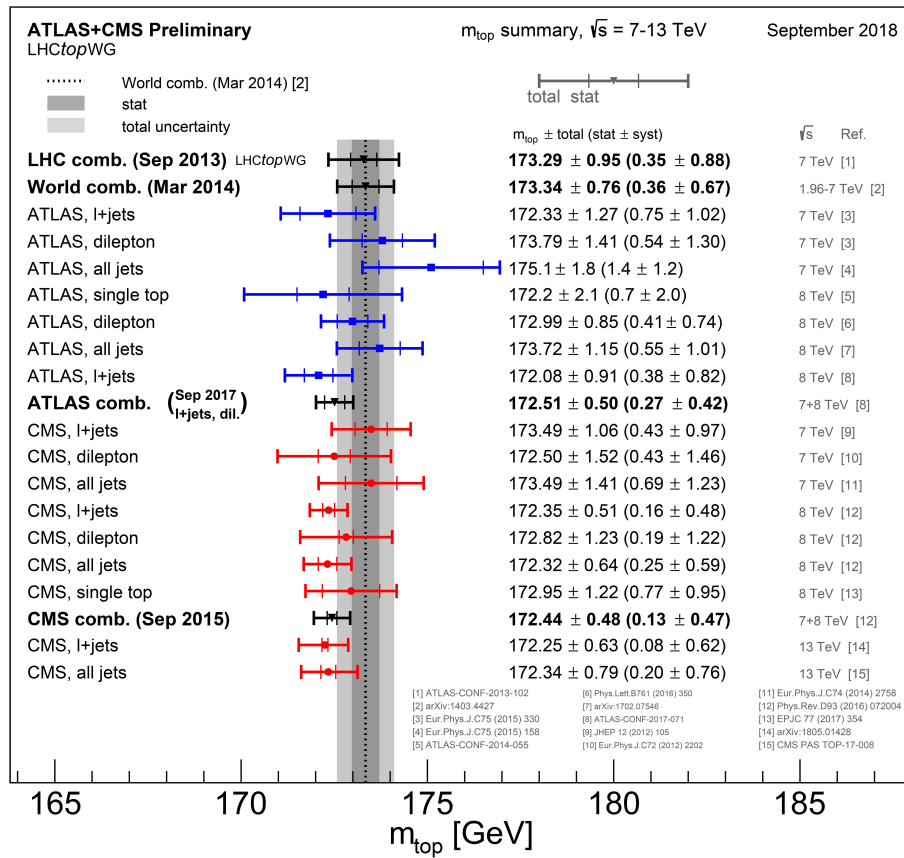
#### 4.2.4. Peak of the $b$ -Jet Energy Spectrum

It was observed that the peak of the  $b$ -jet energy in top decay at LO is independent of the boost from the top to the laboratory frame, as well as of the production mechanism [86]. The CMS Collaboration did measure the top mass from the  $b$ -jet energy peak data at 8 TeV and 19.7 fb<sup>-1</sup> [87], by using POWHEG and MadGraph to simulate top production and decay, and PYTHIA for parton shower, hadronization and underlying event. The resulting uncertainties are 1.17 GeV (statistics) and 2.66 GeV (systematics).

#### 4.2.5. $m_{b\ell}$ , $m_{b\ell\nu}$ and Stranverse Mass $m_{T2}$

The  $b$ -jet+lepton invariant-mass ( $m_{b\ell}$ ) spectrum was used by CMS to reconstruct  $m_t$  in the dilepton channel in CMS Collaboration [88], at 8 TeV and 19.7 fb<sup>-1</sup>. The data were compared with the MadGraph+PYTHIA simulation, yielding a measurement consistent with the world average and an uncertainty about 1.3 GeV. In CMS Collaboration [88], for the sake of comparison, even the NLO code MCFM was used to predict the  $m_{b\ell}$  distribution. More recently, in Sirunyan et al. [89]

<sup>11</sup>At NNLO the  $t\bar{t}$  cross section is  $\mathcal{O}(\alpha_s^4)$ , whereas the threshold logarithms which are resummed in Czakon et al. [74] are  $\sim \alpha_s^n [\ln^m(1-z)/(1-z)]_+$ , with  $z = m_t^2/\hat{s}$ ,  $\hat{s}$  being the partonic center-of-mass energy and  $m \leq 2n - 1$ .



**FIGURE 1 |** Summary of the top-mass analyses at the LHC, accounting for the world average and the ATLAS and CMS combinations as well.

CMS extracted  $m_t$  even from the so-called transverse mass  $m_{T2}$  [90] and from  $m_{b\ell\nu}$ , which accounts for the neutrino missing transverse momentum as well. The sensitivity of these observables to  $m_t$  yields an uncertainty about 180 MeV (statistics) and 900 MeV (systematics).

#### 4.2.6. Endpoint Method

Another method to measure  $m_t$  consists of using the endpoints of distributions sensitive to  $m_t$ , namely the endpoints of  $m_{b\ell}$ ,  $\mu_{bb}$  and  $\mu_{\ell\ell}$ , where  $b$  is a  $b$ -flavored jet, and  $\mu_{bb}$  and  $\mu_{\ell\ell}$  generalizations of the  $b\bar{b}$  and  $\ell^+\ell^-$  invariant masses in the dilepton channel, as described in Chatrchyan et al. [91] (CMS, 7 TeV and 5 fb $^{-1}$ ). Since  $b$ -flavored jets can be calibrated directly from data, the endpoint strategy is claimed to minimize the Monte Carlo error on  $m_t$ , which is mostly due to color reconnection, namely the formation of a  $B$  hadron by combining a  $b$  quark in  $t$  decay with an antiquark from  $\bar{t}$  decay or initial-state radiation. Constraining the neutrino and  $W$  masses to their world-average values, this method leads to uncertainties about 900 MeV (statistics) and 2 GeV (systematics).

#### 4.2.7. Leptonic Observables

Purely leptonic observables in the dilepton channel, such as the Mellin moments of lepton energies or transverse momenta, were proposed to measure  $m_t$ , since in this way

one can escape the actual reconstruction of the top quarks [92]. However, this method still yields uncertainties due to hadronization, production mechanism, Lorentz boost from the top to the laboratory frame, as well as missing higher-order corrections. Preliminary analyses have been carried out in CMS Collaboration [93] (CMS, based on LO MadGraph) and Aaboud et al. [68] (ATLAS, based on the MCFM NLO parton-level code [94]) using data at 8 TeV and 19.7 fb $^{-1}$  and are expected to be improved by matching NLO amplitudes with shower/hadronization generators. For the time being, the uncertainties quoted in Collaboration [93] are 1.1 GeV (statistics), 0.5 GeV (experimental systematics) and 2.5–3.1 GeV (theoretical systematics), whereas in Aaboud et al. [68] they read 0.9, 0.8, and 1.2 GeV, respectively. CMS Collaboration [93] also quotes an uncertainty  $^{+0.8}_{-0.0}$  GeV due to the description of the top-quark transverse momentum. In fact, previous CMS analyses had displayed a mismodeling of the top  $p_T$  simulated by MadGraph+PYTHIA, and therefore CMS Collaboration [93] reweighted the transverse momentum to match the measured one.

#### 4.2.8. $J/\psi$ Method

Final states with  $J/\psi$  mesons were exploited by the CMS Collaboration in Khachatryan et al. [95] to measure  $m_t$ , using data collected at 8 TeV and a luminosity about 19.7 fb $^{-1}$ . In

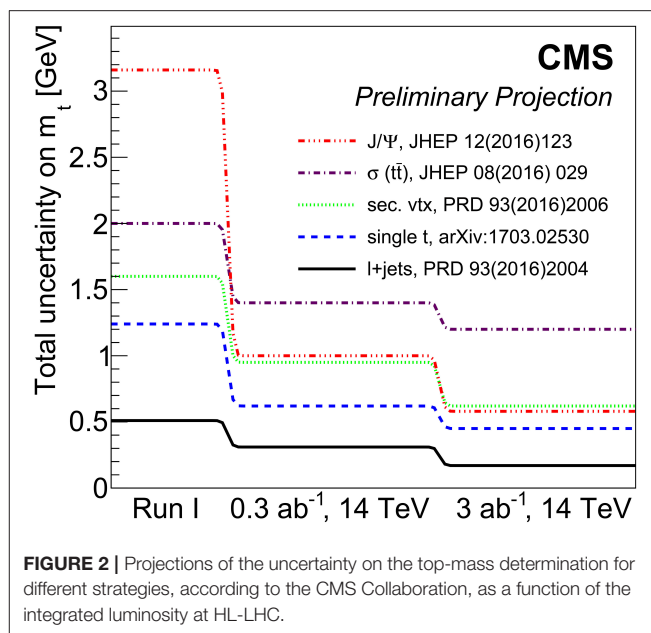
this work, one explores  $t \rightarrow bW$  processes where  $b$ -flavored hadrons decay into states containing a  $J/\psi$ , the  $J/\psi$  decays according to  $J/\psi \rightarrow \mu^+\mu^-$  pair and the  $W$  bosons undergo the leptonic transition  $W \rightarrow \ell\nu$ . The top mass is then extracted by fitting the invariant mass distributions  $m_{\mu\mu}$  or  $m_{J/\psi\ell}$ , as well as the transverse momentum of the  $J/\psi$ . The analysis was carried out by using the MadGraph code, interfaced with PYTHIA, while, for the sake of estimating the theoretical error, POWHEG and SHERPA were employed as well. Overall, the statistical uncertainty in the investigation [95] amounts to 3 GeV, while the systematic error to 0.9 GeV. The conclusion of Khachatryan et al. [95] is that, since the systematic uncertainties are of different origin from those entering in the measurements based on direct reconstruction and given the higher statistics which are foreseen, the  $J/\psi$  method should ultimately be worth to be included in the combination with the extractions from matrix-element or template strategies.

#### 4.2.9. Final-State Charged Particles

A novel technique was presented by the CMS Collaboration in Khachatryan et al. [96], where  $m_t$  is measured by exploiting the kinematic properties of final-state charged particles. The observable used in this analysis is the mass  $m_{s\ell}$  of the secondary vertex-lepton system, namely the invariant mass of a system made of the charged lepton in  $W$  decays and charged hadrons in a jet originating from a common secondary vertex. Using only charged particles, in fact, reduces the overall acceptance uncertainty, whereas this method is obviously dependent on the modeling of top decays and bottom hadronization. The investigation was undertaken using MadGraph+PYTHIA to simulate the signal, POWHEG and SHERPA to estimate the uncertainty due to the matrix-element generation and hadronization, respectively. The final error on the measurement of  $m_t$  from charged particles is then 200 MeV (statistics) and  $^{+1.58}_{-0.97}$  GeV (systematics), by using data sets of 8 TeV collisions and a luminosity of  $19.7 \text{ fb}^{-1}$ .

#### 4.2.10. Perspectives at High Luminosity

The perspectives for the top-mass determination at High Luminosity (HL) LHC were debated in Azzi et al. [97], where the HL-LHC will collide protons at 14 TeV and accumulate an integrated luminosity of  $3,000 \text{ fb}^{-1}$ . In the report [97] the ATLAS Collaboration presented a projection for the accuracy on  $m_t$  using samples of events in the lepton+jets mode and  $J/\psi \rightarrow \mu^+\mu^-$  decays in the final state, along the lines of CERN [98]. The expected statistical and systematic uncertainties amount to 0.14 and 0.48, GeV respectively. As for CMS, the potentials for the top-mass extraction at HL-LHC are detailed in CERN [99] and summarized in **Figure 2**: one can learn that all uncertainties will tremendously decrease at HL-LHC. In particular, one expects an error which ranges from about 0.2 GeV (0.1%) for direct reconstruction in the lepton+jets channel to 1.2 GeV (0.7%) from the total  $t\bar{t}$  NNLO cross section. It is remarkable that the uncertainty from  $J/\psi$  final states will go down to about 0.6 GeV (0.35%).



## 5. INTERPRETATION OF THE TOP-MASS MEASUREMENTS AND THEORETICAL UNCERTAINTIES

The nature of the reconstructed top-quark mass and its possible relations with field-theory mass definitions has lately become the topic of a very lively debate (see, e.g., the reviews in Nason [13], Hoang [100], and Corcella [101]). I shall first overview the main issues concerning the  $m_t$  interpretation and then discuss the dominant sources of theoretical uncertainty.

### 5.1. Measured Mass Ad Theoretical Definitions

The discussion on the identification of the measured quantity is mostly based on the claim that Monte Carlo codes are LO, while well-posed field-theory mass definitions need at least a NLO computation. Although it is certainly true that, referring to standard codes, total cross sections are LO, event shapes and differential distributions go well beyond LO and account for a resummation of enhanced logarithms. NLO+shower codes like POWHEG and MC@NLO yield NLO total cross sections, adopting the top pole mass in the computation, while the differential spectra rely on the shower approximation and on the modeling of hadronization and underlying event. Nevertheless, it is indeed cumbersome interpreting the reconstructed top mass in terms of theoretical definitions or, in other words, scrutinizing all possible sources of uncertainties which may prevent such an identification. As far as this controversy is concerned, one can basically follow two mainstream viewpoints.

On the one hand, there are authors [7–11] who claim that the measured quantity cannot be directly associated with any field-theory mass definition and therefore one must stick to the notion of Monte Carlo mass. Along this point of view, much

work has been undertaken in order to relate the Monte Carlo mass to definitions like the pole mass: the quoted discrepancies between Monte Carlo and pole masses have through the years ranged from few hundreds MeV to, in the most extreme case, almost 1 GeV. If this were indeed the case, it would be an uncertainty comparable or even larger than the current errors on the directly reconstructed top mass. On the other hand, we have authors [12, 13] who instead argue against the use of the Monte Carlo mass and claim that, under given circumstances, the reconstructed mass should actually mimic the pole mass. According to this viewpoint, instead of constructing other mass definitions to properly interpret the measurements, the effort should rather be devoted to carefully estimate the theoretical uncertainties, of both perturbative and non-perturbative nature, in the identification of the measured quantity with the pole mass. In the following, I will briefly review the work carried out in this respect.

As far as I know, the pioneering work on relating the measured mass to the pole mass was carried out in Fleming et al. [7] and Hoang and Stewart [8]. First, Fleming et al. [7] defined, for the case study of  $e^+e^- \rightarrow t\bar{t}$  collisions, the SCET (MSR-like) short-distance jet mass  $m_J(\mu)$ , associated with the collinear jet function and corresponding to the MSR mass at a scale about the top width, i.e.,  $R = \Gamma_t$ . Then,  $m_J(\mu)$  was related to the pole mass by means of the following equation:

$$m_J(\mu) = m_{\text{pole}} - \frac{\alpha_S(\mu)C_F\Gamma_t}{\pi} \left( \ln \frac{\mu}{\Gamma_t} + \frac{3}{2} \right) + \mathcal{O}(\alpha_S^2). \quad (22)$$

Setting, e.g.,  $\mu \simeq 1$  GeV, then the jet mass differs from the pole mass by about 200 MeV at  $\mathcal{O}(\alpha_S)$ . It is also remarkable that the correction is of order  $\mathcal{O}(\alpha_S\Gamma_t)$ , which confirms the intuition that the top width has to play a role in the uncertainty in the measured mass. Later on, Hoang and Stewart [8] did define a Monte Carlo mass and, relying on the standard shower implementations, stated that the extracted top mass could be interpreted as the jet mass evaluated at a scale of the order of the shower cutoff  $Q_0$ , i.e.,  $m_t^{\text{MC}} \simeq m_t^{\text{MSR}}(Q_0)$ . Hoang and Stewart [8] set  $Q_0 = (3_{-2}^{+6})$  GeV and  $m_t^{\text{MSR}}(Q_0)$  to the value of the (Tevatron-based) top-mass world average at that time, and got a consistent value of the  $\overline{\text{MS}}$  mass  $\bar{m}_t(\bar{m}_t)$ , by using renormalization group evolution equations.

More recently, Butenschoen et al. [9] compared PYTHIA with a SCET computation in the NLO approximation, resumming soft- and collinear-enhanced contributions to NLL or even NNLL accuracy. As in Hoang and Stewart [8], the SCET resummed calculation employed the MSR mass  $m_t^{\text{MSR}}(R)$ , with  $R \sim \Gamma_t$  and  $m_t^{\text{MSR}}(R) \rightarrow m_{t,\text{pole}}$  for  $R \rightarrow 0$ . The PYTHIA mass parameter was then calibrated to reproduce the SCET prediction for the 2-jettiness  $\tau_2$ , after running the code for several center-of-mass energies and a few values of the top mass. The result of Butenschoen et al. [9] is that the PYTHIA mass is consistent, within the errors, with the MSR mass evaluated at a scale of 1 GeV. Using instead the pole mass in the computation yields a shift with respect to the PYTHIA  $m_t$  about 600–900 MeV, according to whether the Monte Carlo results are compared with a NLL or NNLL resummation. The work in Butenschoen

et al. [9] was extended to  $pp$  collisions in Hoang et al. [10], where the extraction of  $m_t$  from boosted top jets with light soft-drop grooming was proposed<sup>12</sup>. By comparing the NLL resummation for the groomed top-jet mass with PYTHIA, the pole mass was found about 400–700 MeV below the calibrated Monte Carlo mass, depending on the energy of the  $pp$  collision and non-perturbative parameters contained in the resummation. Still on this subject, Hoang et al. [11] explores the dependence of  $m_t$  on the parton shower cutoff, referring to the HERWIG 7 angular-ordered cascade. By working in the quasi-collinear limit, with boosted massive quarks in the NLL approximation, the authors of Hoang et al. [11] stated that the mass parameter in a Monte Carlo code should be identified with a cutoff-dependent, coherent-branching (CB) mass, labeled as  $m_t^{\text{CB}}(Q_0)$ . Such a coherent-branching mass is a low-scale short-distance mass, free from renormalon corrections, related to the pole mass by a relation like:

$$m_{t,\text{CB}}(Q_0) = m_{t,\text{pole}} - \frac{2}{3}\alpha_S(Q_0)Q_0 + \mathcal{O}(\alpha_S^2Q_0). \quad (23)$$

Expressing in Equation (23)  $\alpha_S$  in terms of the Monte Carlo QCD scale  $\Lambda_{\text{MC}}$  defined in Catani et al. [66] and setting  $Q_0 = 1.25$  GeV, like the shower cutoff of HERWIG 7, the shift between pole and CB masses amounts to about 500 MeV. Using instead the standard  $\overline{\text{MS}}$  scheme for  $\alpha_S$  yields a discrepancy of the order of 300 MeV. Concerning the calibration of the Monte Carlo mass parameter, another approach was suggested in Kieseler et al. [103]: one measures an observable, e.g., a total or differential cross section, ignoring anything on the event generation, and, by comparing the data with the simulation, calibrates both observable and  $m_t$ . The finding of Kieseler et al. [103] is that, given the current precision on the inclusive  $t\bar{t}$  rate, the uncertainty on this calibration is roughly 2 GeV.

As anticipated above, other authors, such as Nason [12, 13], claim that it is not really necessary to introduce the Monte Carlo mass concept to interpret measurements relying on final-state direct reconstruction. The starting point is the observation that, in the narrow-width approximation and assuming that one is able to catch all final-state radiation, the invariant mass of top-decay products in  $t \rightarrow bWX$ ,  $X$  being some extra radiation off top and bottom quarks, should mimic the on-shell top mass, i.e., the pole mass. Effects due to the top final width, parton emission which is not included in the reconstruction, contamination from initial-state radiation and non-perturbative phenomena, such as color reconnection or underlying event, clearly spoil the direct identification of the invariant mass of top-decay final states with the pole mass. However, in the perspective of Nason [12, 13] rather than a genuine shift of the measured mass with respect to the pole mass, such effects are seen as uncertainties, of either perturbative or non-perturbative nature, in the identification of the extracted mass as pole mass.

Although such approaches may sound pretty different, work toward a possible compromise was carried out in Azzi et al. [97], in such a way to guide the top-quark community and avoid

<sup>12</sup>Soft-drop grooming is a jet-substructure technique, which recursively removes soft wide-angle radiation from a jet. See [102] for details.

confusion or statements claiming a sort of ignorance on the nature of the measured top-quark mass. Though starting from different perspectives, all those papers agree that the measured  $m_t$  can be connected to the pole mass by means of a relation like:

$$m_t = m_{t,\text{pole}} + \delta m_t \pm \Delta m_t, \quad (24)$$

where  $\delta m_t$  is a possible shift between measured and pole masses and  $\Delta m_t$  is an uncertainty. According to Nason [12, 13], which basically discourage the use of the concept of Monte Carlo mass, the extracted mass through top-decay final-state reconstruction mimics the pole mass, up to some computable uncertainty. In this approach  $\delta m_t \simeq 0$ , while  $\Delta m_t$  is a theoretical (Monte Carlo based) error that, in measurements employing event generators, should be estimated, e.g., varying shower/hadronization parameters, confronting different models (cluster and string models for hadronization are a typical example) or changing the analysis details (for final-state jets, increasing/decreasing the jet radius leads to accounting for more or less gluon radiation). In the view of Nason [12, 13], the uncertainty  $\Delta m_t$  in the identification of the measurements with the pole mass should be of the order of the hadronization scale, i.e.,  $\mathcal{O}(\Lambda)$ . On the contrary, in the work carried out in Hoang and Stewart [8], Butenschoen et al. [9], Hoang et al. [10], and Hoang et al. [11]  $m_t$  is labeled as Monte Carlo mass and  $\delta m_t$  is an actual discrepancy with respect to the pole mass, typically about  $\mathcal{O}[Q_0\alpha_S(Q_0)]$  as in Equation (23), while  $\Delta m_t$  is still an uncertainty, which one can estimate by varying the parameters or options in the codes and computations employed in the comparison.

Therefore, the disagreement among most authors of the relevant literature on the interpretation of the top-mass measurement is conceptually relevant, but in practice concerns whether one should calculate an actual discrepancy  $\delta m_t$  or not, as well as the meaning of  $\Delta m_t$  and its numerical magnitude. In Hoang and Stewart [8], Butenschoen et al. [9], Hoang et al. [10], and Hoang et al. [11] different values for  $\delta m_t$  and  $\Delta m_t$  have been quoted, which is reasonable, since, as also advocated in Nason [12] for the purpose of the uncertainty, any possible relation between the pole mass and the measured quantity has to depend on the observable which is used to extract  $m_t$ , on the details of the analysis, such as the imposed cuts, the energy of the collider and whether it runs, e.g.,  $e^+e^-$  or  $pp$  modes. Moreover, since such determinations are based on a comparison between Monte Carlo results with resummed calculations, with  $m_t$  being a tunable parameter,  $\delta m_t$  and  $\Delta m_t$  also depend on the accuracy of the resummations, e.g., NLL or NNLL. As discussed above,  $\delta m_t$  is about 200 MeV in Hoang and Stewart [8], in the range 600–900 MeV in Butenschoen et al. [9], 400–700 MeV in Hoang et al. [10] and 300–500 MeV in Hoang et al. [11]. The uncertainty  $\Delta m_t$  in the relation of Equation (24) was estimated to be roughly 250 MeV in Hoang et al. [11] and 280–380 MeV in Butenschoen et al. [9]. Nason [12, 13] do not contain an explicit calculation of  $\Delta m_t$ , but rather propose a method to compute it, e.g., by varying Monte Carlo perturbative and non-perturbative parameters or, in a POWHEG-like implementation, switching NLO and width effects on or off. Of course, it will be very interesting to follow

such an approach and compare the results with the numbers obtained in Hoang and Stewart [8], Butenschoen et al. [9], Hoang et al. [10], and Hoang et al. [11]. One may already guess that, since Nason [12, 13] do not account for any explicit discrepancy  $\delta m_t$ , one may likely get a larger uncertainty  $\Delta m_t$  when following this approach. Furthermore, it will be crucial understanding how much, for a given observable, any shift/uncertainty of the measured mass with respect to the pole mass depends on the specific shower code and, e.g., one finds an impact of the late implementation of NLO corrections and width effects along the lines of [29].

## 5.2. Theoretical Uncertainties in the Top Mass Determination

For the sake of a precise determination of the top-quark mass, a reliable estimate of the theoretical error is of paramount importance. In the top-mass world-average extraction, i.e., The ATLAS et al. [6], based on the so-called standard measurements, the overall theory uncertainty accounts for about 540 MeV of the total 710 MeV systematics. In particular, The ATLAS et al. [6] distinguishes the contributions due to Monte Carlo generators, radiation effects, color reconnection and parton distribution functions (PDFs).

The Monte Carlo systematics is due to the differences in the implementation of parton showers, matrix-element matching, width effects, hadronization and underlying event in the various programs available to describe top-quark production and decay. There is no unique way to estimate this uncertainty, though, and each collaboration even follows different prescription according to the analysis. One can either compare two different generators, which are considered appropriate for a given analysis and have been properly tuned to some data sets, or choose one single code and explore how its predictions fare with respect to variations of its parameters. For example, in The ATLAS et al. [6] CDF compares HERWIG and PYTHIA, while D0 uses ALPGEN+PYTHIA and ALPGEN+HERWIG; both Tevatron experiments use MC@NLO to gauge the overall impact of NLO corrections. At the LHC, ATLAS compares MC@NLO with POWHEG for the NLO contributions and PYTHIA with HERWIG for shower and hadronization; CMS instead confronts LO MadGraph with NLO POWHEG.

The radiation uncertainty gauges the effect of initial- and final-state radiation on the top mass and is typically obtained by varying in suitable ranges the relevant parameters in the parton-shower generators. Concerning PDFs, there are distinct strategies to evaluate the induced error on  $m_t$  in the different experiments, although using two different sets or a given set but with different parametrizations are common trends. More generally, the choice of the PDF set in analyses based on event generators has also been the topic of several discussions: as pointed out before, although Monte Carlo codes yield LO or NLO total cross sections, differential spectra go beyond such approximations and include the resummation of classes of enhanced logarithmic terms. An attempt to propose some improved sets of parton distribution functions for standard parton shower generators was presented in Sherstnev and Thorne [104].

Among the sources of theoretical uncertainty and possible shifts between measured and pole masses, color reconnection should deserve some special attention. In fact, it accounts for about 310 MeV in the world average presented in The ATLAS et al. [6]. Also, the very fact that, for example, a bottom quark in top decay ( $t \rightarrow bW$ ) can be color-connected to an initial-state antiquark does not have its counterpart in  $e^+e^-$  annihilation and therefore its modeling in Monte Carlo event generators may need retuning at hadron colliders. Investigations on the impact of color reconnection on  $m_t$  were undertaken in Argyropoulos and Sjöstrand [105] and Corcella [106], in the frameworks of PYTHIA and HERWIG, respectively. In particular, Corcella [106] addresses this issue by simulating fictitious top-flavored hadrons  $T$  in HERWIG and comparing final-state distributions, such as the  $BW$  invariant mass, with standard  $t\bar{t}$  events. In fact, in the top-hadron case, assuming  $T$  decays according to the spectator model, the  $b$  quark is forced to connect its color with the spectator or with antiquarks in its own shower, namely  $b \rightarrow bg$ , followed by  $g \rightarrow q\bar{q}$ , and color reconnection is suppressed. The analysis in Corcella [106] is still ongoing and, in future perspectives, it may also serve to address the error on the identification of the measured mass with the pole mass. In fact, in the event samples simulated in Corcella [106] the Monte Carlo (HERWIG) mass is the mass of a heavy hadron, which can be related to any definition of the heavy-quark (top for  $T$  mesons) mass definition by means of lattice, potential models or Non Relativistic QCD. In Argyropoulos and Sjöstrand [105], color reconnection is instead investigated within the Lund string model, tuned to charged-particle multiplicity or transverse-momentum data. Several possible models for color reconnection were investigated and the yielded uncertainty on the top mass varied between 200 and 500 MeV, depending on the chosen framework.

Another non-perturbative phenomenon which plays a role in the theoretical error is bottom-quark fragmentation, i.e., the hadronization of bottom quarks in top decays into  $b$ -flavored mesons or baryons. The usual way to deal with it consists in tuning the Monte Carlo fragmentation parameters to precise  $e^+e^- \rightarrow b\bar{b}$  data and then using the best parametrizations to describe bottom-quark hadronization in top decays. This approach was followed, e.g., in Corcella and Drollinger [107] and Corcella and Mescia [108], where data from DELPHI [109] SLD [110], OPAL [111] and ALEPH [112] were employed to tune the parameters of HERWIG [31] and PYTHIA [33]. In particular, Corcella and Mescia [108] used such a tuning to predict the  $B$ -hadron+lepton invariant mass  $m_{B\ell}$  in  $t\bar{t}$  events at LHC. A possible extraction of  $m_t$  using this observable exhibited a large discrepancy between the two event generators, which was explained as due to the different quality of the  $e^+e^-$  fits, with HERWIG being only marginally consistent with the data. More recent modeling and fits, such as the so-called Monash [113] or A14 [114], or using the dipole-like shower implementation in Cormier et al. [57] are expected to give a better description of bottom fragmentation in top decays. Investigations on the uncertainties using these implementations are currently in progress; it will be very interesting, in particular, exploring bottom-quark fragmentation by using NLO+shower codes, such

as POWHEG and aMC@NLO, interfaced to HERWIG or PYTHIA. In fact, it is mandatory to understand whether the Monte Carlo default parameterizations or tunings like those in Skands et al. [113] and Collaboration [114] work well at the LHC even when the hard scattering is at NLO, or one would rather need to refit the Monte Carlo parameters. In general, although the approach followed in Corcella and Mescia [108] relies on the universality of the hadronization transition, it is not absolutely guaranteed that models which reproduce  $e^+e^-$  data work equally well in a colored environment like  $t\bar{t}$  events at the LHC, where initial-state radiation, color reconnection and underlying event play a role. Therefore, tuning shower and hadronization parameters to LHC data should become a ultimate goal.

From this viewpoint, more recently, Corcella et al. [115] reconsidered the issue of the dependence of  $m_t$  on Monte Carlo parameters, suggesting a possible *in-situ* calibration of the shower codes using top events in the dilepton channel, and taking particular care about observables sensitive to  $b$ -fragmentation in top decays. In particular, Corcella et al. [115] extended the work in Corcella and Mescia [108] exploring top-decay observables in terms of  $B$ -hadrons, instead of  $b$ -jets, so that one should deal with fragmentation uncertainties, rather than with the jet-energy scale. For instance, if  $\langle O \rangle$  is the average value of a given observable  $O$  and  $\theta$  a generic generator parameter, then one can write the following relations:

$$\frac{dm_t}{m_t} = \Delta_O^m \frac{d\langle O \rangle}{\langle O \rangle} ; \quad \frac{d\langle O \rangle}{\langle O \rangle} = \Delta_\theta^O \frac{d\theta}{\theta} \Rightarrow \frac{dm_t}{m_t} = \Delta_\theta^m \frac{d\theta}{\theta}, \quad (25)$$

where we defined  $\Delta_\theta^m = \Delta_O^m \Delta_\theta^O$ . Therefore, if one aims at, e.g., an error of 500 MeV on  $m_t$ , namely  $dm_t/m_t < 0.003$ , one should also have  $\Delta_\theta^m(d\theta/\theta) < 0.003$ . Corcella et al. [115] then identifies some so-called calibration observables, which depend on the shower/hadronization parameters but are rather insensitive to the top mass. Examples of such quantities are, e.g., the ratios of  $B$ -hadron to  $b$ -jet ( $b$ ) transverse momenta  $p_{T,B}/p_{T,b}$ , of invariant masses  $m_{B\bar{b}}/m_{b\bar{b}}$  ( $\bar{b}$  being a jet containing a  $\bar{B}$  hadron), the azimuthal separations and invariant opening angles  $\Delta\phi(b\bar{b})$ ,  $\Delta\phi(B\bar{B})$ ,  $\Delta R(b\bar{b})$ ,  $\Delta R(B\bar{B})$ <sup>13</sup>. Then, imagining that one could ideally tune the parameters to measurements of the calibration observables, other quantities can be explored to extract  $m_t$ , such as the  $B$ -hadron energy and transverse momentum  $E_B$  and  $p_{T,B}$ , or the invariant masses  $m_{B\ell}$ ,  $m_{\ell\bar{\ell}}$  and  $m_{B\bar{B}\ell\bar{\ell}}$ . The conclusion of this exploration is that, in order to achieve a 0.3% precision on the top mass, one needs to determine the strong coupling constant at 1% accuracy and other parameters, such as the shower cutoff, the gluon and quark effective masses or the hadronization parameters at 10%. Overall, Corcella et al. [115] proposes a method to tune directly Monte Carlo generators to data from top events at the LHC, which, whenever top-production data were to become precise enough, should be preferable to the use of fits to  $e^+e^-$  data, in such a way to avoid all uncertainties and ambiguities in the application of  $e^+e^-$ -based fits to hadron collisions.

<sup>13</sup>For two particles at  $\Delta\phi$  and  $\Delta\eta$  distances in azimuth and rapidity, the invariant opening angle is defined as  $\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}$ .

## 6. CONCLUSIONS

I discussed some challenging issues regarding the determination and interpretation of the top quark mass at hadron colliders. I reviewed the main top mass definitions, pointing out their most notable features and taking particular care about the pole and  $\overline{MS}$  masses. I described recent calculations for the purpose of the renormalon ambiguity in the pole mass in the infrared regime, yielding uncertainties about 100–250 MeV, which, for the time being, are below the current error on the top mass. Also, such estimates are well below the top-width energy scale, about 1.4 GeV.

The most relevant features of Monte Carlo codes for top-quark phenomenology were then presented, stressing the late implementation of NLO corrections and interference effects between top-production and decay phases. Even the standard shower codes are nevertheless beyond LO in the differential distributions which account for classes of enhanced soft/collinear logarithms to all orders.

The main experimental methods to measure the top mass were discussed, pointing out the differences among the so-called standard and alternative measurements and the magnitude of the quoted uncertainties. For the time being, although the alternative measurements provide an excellent ground to reconstruct  $m_t$  using the kinematic properties of the final states and, in some cases, they are even capable of minimizing the impact of the chosen Monte Carlo generator, the standard methods are still those which yield the lowest uncertainty. This will also be the case in the future LHC runs, albeit the higher statistics are expected to decrease the errors in the alternative strategies too.

Much space was then devoted to the present debate on the interpretation of the measurements and whether one should relate the extracted  $m_t$  to some alternative mass definition or rather express it in terms of the pole mass, up to some uncertainty. A common feature of both attitudes is nonetheless that there is no universal relation between the measured mass and any field-theory definition, but it depends on the considered observable and on the type of Monte Carlo shower code or QCD calculation which is employed in the comparison. There have been many investigations to relate the measured  $m_t$  to short-distance masses by comparing Monte Carlo predictions with SCET resummed computations: the obtained shift with respect to the pole mass was eventually derived and is of the order of a few hundreds of MeV, depending on the specific analysis and accuracy of the calculation. On the other hand, work is in

progress to explore the sources of errors which, on the top of the theoretical systematics, affect the straightforward identification of the top mass in direct-reconstruction analyses as a pole mass, such as color reconnection. Although the starting point of such approaches are conceptually different, a compromise can be reached and it will be very appealing applying the ongoing work on color-reconnection and bottom-fragmentation uncertainties to the interpretation of the top-mass measurements in terms of well-defined field-theory quantities.

Finally, referring in particular to the world-average analysis, the contributions to the quoted theoretical error were debated, along with the current work aimed at obtaining even more reliable estimates of such uncertainties. Furthermore, it was discussed the possibility to use top-quark events and suitable calibration observables to fit Monte Carlo parameters, which will probably be the way to follow in future perspectives, once the data become precise enough to compete with  $e^+e^-$  experiments for the purpose of the tuning of event generators.

In summary, top-quark phenomenology at the LHC, especially in the high-luminosity perspective, has become precision physics and the smallness of the current and foreseen uncertainties in the top-mass measurement are a clear example of such a level of accuracy. However, for the sake of a robust and reliable top-mass determination, much work is still necessary, in order to understand better and possibly reduce the sources of uncertainties. In particular, progress in Monte Carlo studies and QCD calculations for top production and decay, as well as in theoretical work concerning top-mass definitions, should definitely be encouraged. As pointed out many times in this review, investigations along these lines are already in progress, in such a way that one can feel confident that the theoretical and experimental efforts will eventually converge to match the precisions which are expected in the future LHC runs and ultimately at HL-LHC.

## AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

## ACKNOWLEDGMENTS

I acknowledge discussions with André Hoang and Paolo Nason, co-authors of the top-mass section in Azzi et al. [97], on the interpretation of the top-mass measurements.

## REFERENCES

1. Baak M, Cúth J, Haller J, Hoecker A, Kogler R, Mönig K, et al. The global electroweak fit at NNLO and prospects for the LHC and ILC. *Eur Phys J.* (2014) **C74**:3046. doi: 10.1140/epjc/s10052-014-3046-5
2. Masina I, Notari A. Standard model false vacuum inflation: correlating the tensor-to-scalar ratio to the top quark and Higgs boson masses. *Phys Rev Lett.* (2012) **108**:191302. doi: 10.1103/PhysRevLett.108.191302
3. Branchina V, Messina E, Platania A. Top mass determination, Higgs inflation, and vacuum stability. *J High Energy Phys.* (2014) **9**:182. doi: 10.1007/JHEP09(2014)182
4. Degrandi G, Di Vita S, Elias-Miro J, Espinosa JR, Giudice GF, Isidori G, et al. Higgs mass and vacuum stability in the Standard Model at NNLO. *J High Energy Phys.* (2012) **8**:098. doi: 10.1007/JHEP08(2012)098
5. Andreassen A, Frost W, Schwartz MD. Scale invariant instantons and the complete lifetime of the standard model. *Phys Rev.* (2018) **D97**:056006. doi: 10.1103/PhysRevD.97.056006
6. The ATLAS, CDF, CMS, D0 Collaborations. First combination of Tevatron and LHC measurements of the top-quark mass. *arXiv:1403.4427*.
7. Fleming S, Hoang AH, Mantry S, Stewart IW. Jets from massive unstable particles: top-mass determination. *Phys Rev.* (2008) **D77**:074010. doi: 10.1103/PhysRevD.77.074010

8. Hoang AH, Stewart IW. Top mass measurements from jets and the tevatron top-quark mass. *Nucl Phys Proc Suppl.* (2008) **185**:220–6. doi: 10.1016/j.nuclphysbps.2008.10.028
9. Butenschoen M, Dehnadi B, Hoang AH, Mateu V, Preisser M, Stewart IW. Top quark mass calibration for Monte Carlo event generators. *Phys Rev Lett.* (2016) **117**:232001. doi: 10.1103/PhysRevLett.117.232001
10. Hoang AH, Mantry S, Pathak A, Stewart IW. Extracting a Short Distance Top Mass with Light Grooming. *arXiv:1708.02586* (2017).
11. Hoang AH, Plätzer S, Samitz D. On the cutoff dependence of the quark mass parameter in angular ordered parton showers. *J High Energy Phys.* (2018) **10**:200. doi: 10.1007/JHEP10(2018)200
12. Nason P. Theory summary. *PoS.* (2016) arXiv:1602.00443. doi: 10.22323/1.257.0056
13. Nason P. The top mass in hadronic collisions. In: Levy A, Forte S, Ridolfi G, editors. *From my Vast Repertoire: Guido Altarelli's Legacy*. Singapore: World Scientific (2019). p. 123–51.
14. Beneke M, Braun VM. Heavy quark effective theory beyond perturbation theory: Renormalons, the pole mass and the residual mass term. *Nucl Phys.* (1994) **B426**:301–43. doi: 10.1016/0550-3213(94)90314-X
15. Beneke M. More on ambiguities in the pole mass. *Phys Lett.* (1995) **B344**:341–7. doi: 10.1016/0370-2693(94)01505-7
16. Beneke M, Marquard P, Nason P, Steinhauser M. On the ultimate uncertainty of the top quark pole mass. *Phys Lett.* (2017) **B775**:63–70. doi: 10.1016/j.physletb.2017.10.054
17. Hoang AH, Lepenik C, Preisser M. On the light massive flavor dependence of the large order asymptotic behavior and the ambiguity of the pole mass. *J High Energy Phys.* (2017) **9**:099. doi: 10.1007/JHEP09(2017)099
18. Marquard P, Smirnov AV, Smirnov VA, Steinhauser M. Quark mass relations to four-loop order in perturbative QCD. *Phys Rev Lett.* (2015) **114**:142002. doi: 10.1103/PhysRevLett.114.142002
19. Tanabashi M, Nagoya U, Nagoya KMI. Review of particle physics. *Phys Rev.* (2018) **D98**:030001. doi: 10.1103/PhysRevD.98.030001
20. Kataev AL, Molokodov VS. Dependence of five- and six-loop estimated QCD corrections to the relation between pole and running masses of heavy quarks on the number of light flavors. *JETP Lett.* (2018) **108**:777–82. doi: 10.1134/S0021364018240050
21. Hoang AH, Teubner T. Top quark pair production close to threshold: top mass, width and momentum distribution. *Phys Rev.* (1999) **D60**:114027.
22. Hoang AH, Manohar AV, Stewart IW, Teubner T. The threshold  $t\bar{t}$  cross-section at NNLL order. *Phys Rev.* (2002) **D65**:014014. doi: 10.1103/PhysRevD.65.014014
23. Beneke M. A Quark mass definition adequate for threshold problems. *Phys Lett.* (1998) **B434**:115–25. doi: 10.1016/S0370-2693(98)00741-2
24. Hoang AH, Melnikov K, Teubner T, Yelkhovsky A, Yakovlev O, Beneke M, et al. Top - anti-top pair production close to threshold: synopsis of recent NNLO results. *Eur Phys Jdirect.* (2000) **2**:3. doi: 10.1007/s1010500c0003
25. Hoang AH, Stahlhofen M. The top-antitop threshold at the ILC: NNLL QCD uncertainties. *J High Energy Phys.* (2014) **5**:121. doi: 10.1007/JHEP05(2014)121
26. Beneke M, Kiyo Y, Marquard P, Penin A, Piclum J, Steinhauser M. Next-to-next-to-next-to-leading order QCD prediction for the top antitop S-wave pair production cross section near threshold in  $e^+e^-$  annihilation. *Phys Rev Lett.* (2015) **115**:192001. doi: 10.1103/PhysRevLett.115.192001
27. Pineda A. Determination of the bottom quark mass from the  $\Upsilon(1S)$  system. *J High Energy Phys.* (2001) **6**:022. doi: 10.1088/1126-6708/2001/06/022
28. Bigi Y II, Shifman MA, Uraltsev N, Vainshtein AI. High power  $n$  of  $m_b$  in beauty widths and  $n = 5 \rightarrow \infty$  limit. *Phys Rev.* (1997) **D56**:4017–30.
29. Jezo T, Lindert JM, Nason P, Oleari C, Pozzorini S. An NLO+PS generator for  $t\bar{t}$  and  $Wt$  production and decay including non-resonant and interference effects. *Eur Phys J.* (2016) **C76**:691. doi: 10.1140/epjc/s10052-016-4538-2
30. Ferrario Ravasio S, Nason P, Oleari C. All-orders behaviour and renormalons in top-mass observables. *J High Energy Phys.* (2018). **2018**:203. doi: 10.1007/JHEP01(2019)203
31. Corcella G, Knowles IG, Marchesini G, Moretti S, Odagiri K, Richardson P, et al. HERWIG 6: an event generator for hadron emission reactions with interfering gluons (including supersymmetric processes). *J High Energy Phys.* (2001) **1**:010. doi: 10.1088/1126-6708/2001/01/010
32. Bellm J, Gieseke S, Grellscheid D, Plätzer S, Rauch M, Reuschle C, et al. Herwig 7.0/Herwig++ 3.0 release note. *Eur Phys J.* (2016) **C76**:196. doi: 10.1140/epjc/s10052-016-4018-8
33. Sjöstrand T, Mrenna S, Skands PZ. PYTHIA 6.4 physics and manual. *J High Energy Phys.* (2006) **5**:026. doi: 10.1088/1126-6708/2006/05/026
34. Sjöstrand T, Ask S, Christiansen JR, Corke R, Desai N, Ilten P, et al. An introduction to PYTHIA 8.2. *Comput Phys Commun.* (2015) **191**:159–77. doi: 10.1016/j.cpc.2015.01.024
35. Marchesini G, Webber BR. Simulation of QCD jets including soft gluon interference. *Nucl Phys.* (1984) **B238**:1–29. doi: 10.1016/0550-3213(84)90463-2
36. Gieseke S, Stephens P, Webber B. New formalism for QCD parton showers. *J High Energy Phys.* (2003) **12**:045. doi: 10.1088/1126-6708/2003/12/045
37. Plätzer S, Gieseke S. Dipole showers and automated NLO matching in Herwig++. *Eur Phys J.* (2012) **C72**:2187. doi: 10.1140/epjc/s10052-012-2187-7
38. Corcella G, Seymour MH. Matrix element corrections to parton shower simulations of heavy quark decay. *Phys Lett.* (1998) **B442**:417–26.
39. Norrbin E, Sjöstrand T. QCD radiation off heavy particles. *Nucl Phys.* (2001) **B603**:297–342. doi: 10.1016/S0550-3213(01)00099-2
40. Webber BR. A QCD model for jet fragmentation including soft gluon interference. *Nucl Phys.* (1984) **B238**:492–528.
41. Andersson B, Gustafson G, Ingelman G, Sjöstrand T. Parton fragmentation and string dynamics. *Phys Rept.* (1983) **97**:31–145.
42. Butterworth JM, Forshaw JR, Seymour MH. Multiparton interactions in photoproduction at HERA. *Z Phys.* (1996) **C72**:637–46.
43. Sjöstrand T, Skands PZ. Multiple interactions and the structure of beam remnants. *J High Energy Phys.* (2004) **3**:053. doi: 10.1088/1126-6708/2004/03/053
44. Gleisberg T, Hoeche S, Krauss F, Schonherr M, Schumann S, Siegert F, et al. Event generation with SHERPA 1.1. *J High Energy Phys.* (2009) **2**:007. doi: 10.1088/1126-6708/2009/02/007
45. Krauss F, Kuhn R, Soff G. AMEGIC++ 1.0: a Matrix element generator in C++. *J High Energy Phys.* (2000) **2**:044. doi: 10.1088/1126-6708/2002/02/044
46. Gleisberg T, Hoeche S. Comix, a new matrix element generator. *J High Energy Phys.* (2008) **12**:039. doi: 10.1088/1126-6708/2008/12/039
47. Binoth T, Boudjema F, Dissertori G, Lazopoulos A, Denner A, Dittmaier S, et al. A Proposal for a standard interface between Monte Carlo tools and one-loop programs. *Comput Phys Commun.* (2010) **181**:1612–22. doi: 10.1016/j.cpc.2010.05.016
48. Catani S, Seymour MH. The Dipole formalism for the calculation of QCD jet cross-sections at next-to-leading order. *Phys Lett.* (1996) **B378**:287–301. doi: 10.1016/0370-2693(96)00425-X
49. Frixione S, Webber BR. Matching NLO QCD computations and parton shower simulations. *J High Energy Phys.* (2002) **6**:029. doi: 10.1088/1126-6708/2002/06/029
50. Alwall J, Frederix R, Frixione S, Hirschi V, Maltoni F, Mattelaer O, et al. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *J High Energy Phys.* (2014) **7**:079. doi: 10.1007/JHEP07(2014)079
51. Alioli S, Nason P, Oleari C, Re E. A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX. *J High Energy Phys.* (2010) **6**:043. doi: 10.1007/JHEP06(2010)043
52. Heinrich G, Maier A, Nisius R, Schlenk J, Schulze M, Scyboz L, et al. NLO and off-shell effects in top quark mass determinations. *J High Energy Phys.* (2017). **2017**:129. doi: 10.1007/JHEP07(2018)129
53. Artoisenet P, Frederix R, Mattelaer O, Rietkerk R. Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations. *J High Energy Phys.* (2013) **3**:015. doi: 10.1007/JHEP03(2013)015
54. Frederix R, Frixione S, Papanastasiou AS, Prestel S, Torrielli P. Off-shell single-top production at NLO matched to parton showers. *J High Energy Phys.* (2016) **6**:027. doi: 10.1007/JHEP06(2016)027
55. Plätzer S. Controlling inclusive cross sections in parton shower + matrix element merging. *J High Energy Phys.* (2013) **08**:114. doi: 10.1007/JHEP08(2013)114
56. Bellm J, Gieseke S, Plätzer S. Merging NLO multi-jet calculations with improved unitarization. *Eur Phys J.* (2018) **C78**:244. doi: 10.1140/epjc/s10052-018-5723-2

57. Cormier K, Plätzer S, Reuschle C, Richardson P, Webster S. Parton shower and matching uncertainties in top quark pair production with Herwig 7. (2018).
58. Mangano ML, Moretti M, Piccinini F, Pittau R, Polosa AD. ALPGEN, a generator for hard multiparton processes in hadronic collisions. *J High Energy Phys.* (2003) 7:001. doi: 10.1088/1126-6708/2003/07/001
59. Campbell JM, Ellis RK. MCFM for the Tevatron and the LHC. *Nucl Phys Proc Suppl.* (2010) 205-6:10–5. doi: 10.1016/j.nuclphysbps.2010.08.011
60. Belyaev A, Christensen ND, Pukhov A. CalcHEP 3.4 for collider physics within and beyond the Standard Model. *Comput Phys Commun.* (2013) 184:1729–69. doi: 10.1016/j.cpc.2013.01.014
61. Kanaki A, Papadopoulos CG. HELAC: a Package to compute electroweak helicity amplitudes. *Comput Phys Commun.* (2000) 132:306–15. doi: 10.1016/S0010-4655(00)00151-X
62. Kilian W, Ohl T, Reuter J. WHIZARD: simulating multi-particle processes at LHC and ILC. *Eur Phys J.* (2011) C71:1742. doi: 10.1140/epjc/s10052-011-1742-y
63. Bevilacqua G, Czakon M, Garzelli MV, van Hameren A, Kardos A, Papadopoulos CG, et al. HELAC-NLO. *Comput. Phys. Commun.* (2013) 184: 986–97. doi: 10.1016/j.cpc.2012.10.033
64. Weiss C, Chokoufe Nejad B, Kilian W, Reuter J. Automated NLO QCD corrections with WHIZARD. *PoS.* (2015) EPS-HEP2015:466. doi: 10.22323/1.234.0466
65. Banfi A, Corcella G, Dasgupta M. Angular ordering and parton showers for non-global QCD observables. *J High Energy Phys.* (2007) 3:050. doi: 10.1088/1126-6708/2007/03/050
66. Catani S, Webber BR, Marchesini G. QCD coherent branching and semiinclusive processes at large  $x$ . *Nucl Phys.* (1991) B349:635–54. doi: 10.1016/0550-3213(91)90390-J
67. Aaboud M, Aad G, Abbott B, Abdallah J, Abidinov O, Abeloos B, et al. Measurement of the top quark mass in the  $t\bar{t} \rightarrow$  dilepton channel from  $\sqrt{s} = 8$  TeV ATLAS data. *Phys Lett.* (2016) B761:350–71. doi: 10.1016/j.physletb.2016.08.042
68. Aaboud M, Aad G, Abbott B, Abdallah J, Abidinov O, Abeloos B, et al. Top-quark mass measurement in the all-hadronic  $t\bar{t}$  decay channel at  $\sqrt{s} = 8$  TeV with the ATLAS detector. *J High Energy Phys.* (2017) 9:118. doi: 10.1007/JHEP09(2017)118
69. Aaboud M, Aad G, Abbott B, Abdallah J, Abidinov O, Abeloos B, et al. Measurement of the top quark mass in the  $t\bar{t} \rightarrow$  lepton+jets channel from  $\sqrt{s} = 8$  TeV ATLAS data and combination with previous results. *Eur Phys J.* (2018) C79:290. doi: 10.1140/epjc/s10052-019-6757-9
70. Sirunyan AM, Tumasyan A, Adam W, Ambrogio F, Asilar E, Bergauer T, et al. Measurement of the  $t\bar{t}$  production cross section, the top quark mass, and the strong coupling constant using dilepton events in  $pp$  collisions at  $\sqrt{s} = 13$  TeV. (2018).
71. Sirunyan AM, Tumasyan A, Adam W, Ambrogio F, Asilar E, Bergauer T, et al. Measurement of the top quark mass with lepton+jets final states using  $pp$  collisions at  $\sqrt{s} = 13$  TeV. *Eur Phys J.* (2018) C78:891. doi: 10.1140/epjc/s10052-018-6332-9
72. Sirunyan AM, Tumasyan A, Adam W, Ambrogio F, Asilar E, Bergauer T, et al. Measurement of the top quark mass in the all-jets final state at  $\sqrt{s} = 13$  TeV and combination with the lepton+jets channel. *Eur Phys J.* (2018). doi: 10.1140/epjc/s10052-019-6788-2
73. Khachatryan V, Sirunyan AM, Tumasyan A, Adam W, Asilar E, Bergauer T. Measurement of the top quark mass using proton-proton data at  $\sqrt{s} = 7$  and 8 TeV. *Phys Rev D.* (2016) D93:072004. doi: 10.1103/PhysRevD.93.072004
74. Czakon M, Fiedler P, Mitov A. Total Top-Quark pair-production cross section at hadron colliders through  $O(\alpha_s^4)$ . *Phys Rev Lett.* (2013) 110:252004. doi: 10.1103/PhysRevLett.110.252004
75. Aad G, Abbott B, Abdallah J, Abdel Khalek S, Abidinov O, Aben R. Measurement of the  $t\bar{t}$  production cross-section using  $e\mu$  events with b-tagged jets in  $pp$  collisions at  $\sqrt{s} = 7$  and 8 TeV with the ATLAS detector. *Eur Phys J.* (2014) C74:3109. doi: 10.1140/epjc/s10052-014-3109-7
76. Khachatryan V, Sirunyan AM, Tumasyan A, Adam W, Asilar E, Bergauer T. Measurement of the  $t\bar{t}$  production cross section in the  $e\mu$  channel in proton-proton collisions at  $\sqrt{s} = 7$  and 8 TeV. *J High Energy Phys.* (2016) 1608:029. doi: 10.1007/JHEP08(2016)029
77. Sirunyan AM, et al. Measurement of the  $t\bar{t}$  production cross section using events with one lepton and at least one jet in  $pp$  collisions at  $\sqrt{s} = 13$  TeV. *J High Energy Phys.* (2017) 2017:51. doi: 10.1007/JHEP09(2017)051
78. Czakon M, Heymes D, Mitov A. High-precision differential predictions for top-quark pairs at the LHC. *Phys Rev Lett.* (2016) 116:082003. doi: 10.1103/PhysRevLett.116.082003
79. Czakon M, Fiedler P, Heymes D, Mitov A. Measurement of the Pole Mass of the Top Quark using Differential  $t\bar{t}$  Cross Sections in  $p\bar{p}$  Collisions at  $\sqrt{s} = 1.96$  TeV (2016). Available online at: <http://lss.fnal.gov/archive/2016/conf/fermilab-conf-16-383-ppd.pdf>.
80. Dowling M, Moch SO. Differential distributions for top-quark hadro-production with a running mass. *Eur Phys J.* (2014) C74:3167. doi: 10.1140/epjc/s10052-014-3167-x
81. Czakon M, Heymes D, Mitov A, Pagani D, Tsirikos I, Zaro M. Top-pair production at the LHC through NNLO QCD and NLO EW. *J High Energy Phys.* (2017) 10:186. doi: 10.1007/JHEP10(2017)186
82. Alioli S, Fernandez P, Fuster J, Irlas A, Moch SO, Uwer P, et al. A new observable to measure the top-quark mass at hadron colliders. *Eur Phys J.* (2013) C73:2438. doi: 10.1140/epjc/s10052-013-2438-2
83. Aad G, et al. Determination of the top-quark pole mass using  $t\bar{t} + 1$ -jet events collected with the ATLAS experiment in 7 TeV  $pp$  collisions. *JHEP.* (2015) 10:121. doi: 10.1007/JHEP10(2015)121
84. Collaboration C. Determination of the normalised invariant mass distribution of  $t\bar{t}$ -jet and extraction of the top quark mass (2016).
85. Fuster J, Irlas A, Melini D, Uwer P, Vos M. Extracting the top-quark running mass using  $t\bar{t} + 1$ -jet events produced at the Large Hadron Collider. *Eur Phys J.* (2017) C77:794. doi: 10.1140/epjc/s10052-017-5354-z
86. Agashe K, Franceschini R, Kim D, Schulze M. Top quark mass determination from the energy peaks of  $b$ -jets and  $B$ -hadrons at NLO QCD. *Eur Phys J.* (2016) C76:636. doi: 10.1140/epjc/s10052-016-4494-x
87. CMS Collaboration. Measurement of the Top-Quark Mass from Thebjet Energy Spectrum. (2015). CMS-PAS-TOP-15-002.
88. CMS Collaboration. Determination of the top-Quark Mass from the  $m_{lb}$  Distribution in Dileptonic  $t\bar{t}$  Events at  $\sqrt{s} = 8$  TeV. (2014). CMS-PAS-TOP-14-014.
89. Sirunyan AM, Tumasyan A, Adam W, Asilar E, Bergauer T, Brandstetter J. Measurement of the top quark mass in the dileptonic  $t\bar{t}$  decay channel using the mass observables  $M_{b\bar{b}}$ ,  $M_{T2}$ , and  $M_{b\bar{b}v}$  in  $pp$  collisions at  $\sqrt{s} = 8$  TeV. *Phys Rev.* (2017) D96:032002. doi: 10.1103/PhysRevD.96.032002
90. Lester CG, Summers DJ. Measuring masses of semi-invisibly decaying particles pair produced at hadron colliders. *Phys Lett.* (1999) B463:99–103. doi: 10.1016/S0370-2693(99)00945-4
91. Chatrchyan S, et al. Measurement of masses in the  $t\bar{t}$  system by kinematic endpoints in  $pp$  collisions at  $\sqrt{s} = 7$  TeV. *Eur Phys J.* (2013) C73:2494. doi: 10.1140/epjc/s10052-013-2494-7
92. Frixione S, Mitov A. Determination of the top quark mass from leptonic observables. *J High Energy Phys.* (2014) 2014:012. doi: 10.1007/JHEP09(2014)012
93. CMS Collaboration. Determination of the top Quark Mass from Leptonic Observables Using  $e\mu$ +jets Final States Selected in Proton-Proton Collisions at  $\sqrt{s} = 8$  TeV. (2016). CMS-PAS-TOP-16-002.
94. Campbell JM, Ellis RK, Giele WT. A multi-threaded version of MCFM. *Eur Phys J.* (2015) C75:246. doi: 10.1140/epjc/s10052-015-3461-2
95. Khachatryan V, Sirunyan AM, Tumasyan A, Adam W, Asilar E, Bergauer T. Measurement of the mass of the top quark in decays with a  $J/\psi$  meson in  $pp$  collisions at 8 TeV. *J High Energy Phys.* (2016) 12:123. doi: 10.1007/JHEP12(2016)123
96. Khachatryan V, et al. Measurement of the top quark mass using charged particles in  $pp$  collisions at  $\sqrt{s} = 8$  TeV. *Phys Rev.* (2016) D93:092006. doi: 10.1103/PhysRevD.93.092006
97. Azzi P, et al. Standard model physics at the HL-LHC and HE-LHC. (2019).
98. The ATLAS collaboration. Prospects for measurement of the top quark mass using  $t\bar{t}$  events with  $J/\psi \rightarrow \mu^+\mu^-$  decays with the upgraded ATLAS detector at the High Luminosity LHC.
99. CERN. ECFA 2016: Prospects for Selected Standard Model Measurements With the CMS Experiment at the High-Luminosity LHC. Geneva: CERN. Available online at: <https://cds.cern.ch/record/2262606>.

100. Hoang AH. The top mass: interpretation and theoretical uncertainties. In: *Proceedings, 7th International Workshop on Top Quark Physics (TOP2014)*. (2014) (Cannes).
101. Corcella G. Interpretation of top-quark mass measurements: a theory overview. In: *8th International Workshop on Top Quark Physics (TOP2015) - Session 10: Top Mass and Combinations*. (2015). doi: 10.22323/1.257.0037
102. Larkoski AJ, Marzani S, Soyez G, Thaler J. Soft drop. *J High Energy Phys.* (2014) **05**:146. doi: 10.1007/JHEP05(2014)146
103. Kieseler J, Lipka K, Moch SO. Calibration of the top-quark monte carlo mass. *Phys Rev Lett.* (2016) **116**:162001. doi: 10.1103/PhysRevLett.116.162001
104. Sherstnev A, Thorne RS. Parton distributions for LO generators. *Eur Phys J.* (2008) **C55**:553–75. doi: 10.1140/epjc/s10052-008-0610-x
105. Argyropoulos S, Sjöstrand T. Effects of color reconnection on  $t\bar{t}$  final states at the LHC. *J High Energy Phys.* (2014) **11**:043. doi: 10.1007/JHEP11(2014)043
106. Corcella G. Hadronization systematics and top mass reconstruction. *EPJ Web Conf.* (2014) **80**:00019. doi: 10.1051/epjconf/20148000019
107. Corcella G, Drollinger V. Bottom-quark fragmentation: Comparing results from tuned event generators and resummed calculations. *Nucl Phys.* (2005) **B730**:82–102. doi: 10.1016/j.nuclphysb.2005.09.030
108. Corcella G, Mescia F. A phenomenological study of bottom quark fragmentation in top quark decay. *Eur Phys J.* (2010) **C65**:171–80.
109. Abreu P, et al. Tuning and test of fragmentation models based on identified particles and precision event shape data. *Z Phys.* (1996) **C73**:11–60.
110. Abe K, Abe K, Abe T, Adam I, Akimoto H, Aston D. Precise measurement of the  $b$  quark fragmentation function in  $Z^0$  boson decays. *Phys Rev Lett.* (2000) **84**:4300–4. doi: 10.1103/PhysRevLett.84.4300
111. Abbiendi G, Ainsley C, Akesson PF, Alexander G, Allison J, Amaral P. Inclusive analysis of the  $b$  quark fragmentation function in  $Z$  decays at LEP. *Eur Phys J.* (2003) **C29**:463–78. doi: 10.1140/epjc/s2003-01229-x
112. Heister A, Schael S, Barate R, De Bonis I, Decamp D, Goy C. Study of the fragmentation of  $b$  quarks into  $B$  mesons at the  $Z$  peak. *Phys Lett.* (2001) **B512**:30–48. doi: 10.1016/S0370-2693(01)00690-6
113. Skands P, Carrazza S, Rojo J. Tuning PYTHIA 8.1: the Monash 2013 Tune. *Eur Phys J.* (2014) **C74**:3024. doi: 10.1140/epjc/s10052-014-3024-y
114. Collaboration A. ATLAS pythia 8 Tunes to 7 TeV data. (2014).
115. Corcella G, Franceschini R, Kim D. Fragmentation Uncertainties in Hadronic Observables for Top-quark Mass Measurements. *Nucl Phys.* (2018) **B929**:485–526. doi: 10.1016/j.nuclphysb.2018.02.012

**Conflict of Interest Statement:** The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Corcella. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Several Problems in Particle Physics and Cosmology Solved in One SMASH

Guillermo Ballesteros<sup>1,2</sup>, Javier Redondo<sup>3,4</sup>, Andreas Ringwald<sup>5\*</sup> and Carlos Tamarit<sup>6</sup>

<sup>1</sup> Instituto de Física Teórica, UAM-CSIC, Madrid, Spain, <sup>2</sup> Departamento de Física Teórica, Universidad Autónoma de Madrid, Madrid, Spain, <sup>3</sup> Departamento de Física Teórica, Universidad de Zaragoza, Zaragoza, Spain, <sup>4</sup> Max-Planck-Institut für Physik, München, Germany, <sup>5</sup> Deutsches Elektronen-Synchrotron DESY, Hamburg, Germany, <sup>6</sup> Physik Department T70, Technische Universität München, Garching, Germany

## OPEN ACCESS

### Edited by:

Francesco Sannino,  
University of Southern Denmark,  
Denmark

### Reviewed by:

Sayantan Choudhury,  
Max-Planck-Institut für  
Gravitationsphysik, Germany  
Michael Andreas Schmidt,  
University of New South Wales,  
Australia

### \*Correspondence:

Andreas Ringwald  
andreas.ringwald@desy.de

### Specialty section:

This article was submitted to  
High-Energy and Astroparticle  
Physics,  
a section of the journal  
Frontiers in Astronomy and Space  
Sciences

**Received:** 11 April 2019

**Accepted:** 11 July 2019

**Published:** 30 July 2019

### Citation:

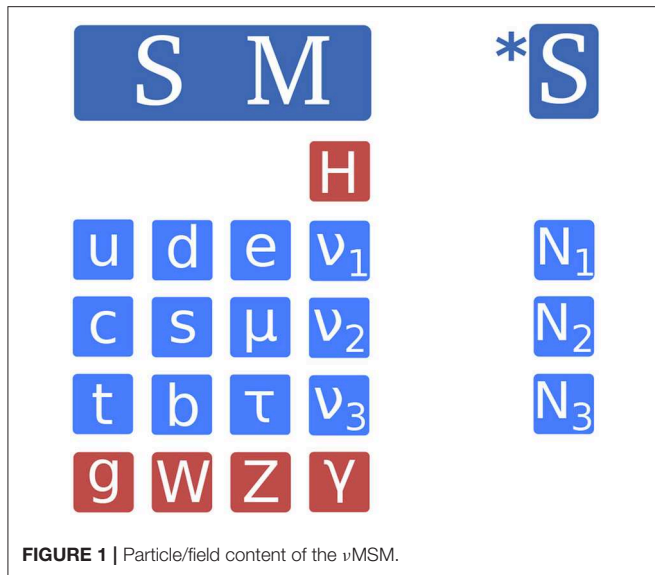
Ballesteros G, Redondo J, Ringwald A  
and Tamarit C (2019) Several  
Problems in Particle Physics and  
Cosmology Solved in One SMASH.  
Front. Astron. Space Sci. 6:55.  
doi: 10.3389/fspas.2019.00055

The Standard Model (SM) of particle physics is a big success. However, it lacks explanations for cosmic inflation, the matter-anti-matter asymmetry of the Universe, dark matter, neutrino oscillations, and the feebleness of CP violation in the strong interactions. The latter may be explained by introducing an exotic vector-like quark which is charged under a chiral global  $U(1)$  Peccei-Quinn (PQ) symmetry which is spontaneously broken by the vacuum expectation value of a complex SM singlet scalar field—the PQ field. Moreover, the pseudo Nambu-Goldstone boson of this breaking—the axion—may play the role of the dark matter. Furthermore, the modulus of the PQ field is a candidate for driving inflation. Furthermore, three extra SM singlet neutrinos are added who acquire their Majorana mass from the breaking of the PQ symmetry and which explain the small masses of the active neutrinos and their oscillations by the seesaw mechanism. The resulting extension of the SM which has been dubbed SMASH—for SM-Axion-Seesaw-Higgs portal inflation—solves the five aforementioned problems in one stroke. We review how this works in SMASH and discuss its further predictions and tests in astrophysics, cosmology, and laboratory experiments. Furthermore, we consider and comment on variants of SMASH.

**Keywords:** inflation, matter anti-matter asymmetry, dark matter, neutrino masses and mixing, strong CP problem

## 1. INTRODUCTION

The SM is arguably the most successful theory in physics. It describes the known particles and their interactions remarkably well. No significant deviations from the theoretical predictions of the SM have been found so far at precision collider experiments and the like. On the other hand, it is generally agreed that there are a number of fundamental problems in particle physics and cosmology which require new physics beyond the SM. In fact, there is compelling evidence that nearly 85% of the matter in the Universe is non-baryonic. This evidence is supported by observations on many scales, ranging from the shapes of the rotation curves of spiral galaxies to the temperature fluctuations of the cosmic microwave background (CMB). Furthermore, the SM cannot explain the exponential expansion of the very early Universe called inflation which is required to explain the isotropic, Gaussian and nearly scale-invariant temperature fluctuations of the CMB. Moreover, the CP violation within the SM is too feeble to explain the asymmetry between the fraction of the baryonic matter and anti-matter in the Universe. Furthermore, the SM does not feature masses for the active



neutrinos, while the observed flavor oscillations of the active neutrinos require tiny neutrino masses. Last, but not least, there is the strong CP problem: the SM has no explanation for the smallness of the  $\bar{\theta}$ -angle of quantum chromodynamics (QCD) which induces CP-violation in flavor-diagonal interactions. In fact, the non-observation of an electric dipole moment of the neutron places a very strong upper limit on the angle,  $|\bar{\theta}| < 10^{-10}$ , requiring an extreme fine-tuning which cannot even be justified on the basis of anthropic arguments.

Three of these problems can be tackled simultaneously in the Neutrino Minimal SM ( $\nu$ MSM) (Asaka, 2005; Asaka et al., 2005): a remarkably simple extension of the SM by three right-handed singlet neutrinos  $N_i$  (cf. **Figure 1**), having Dirac masses  $m_D = Fv/\sqrt{2}$  arising from Yukawa couplings  $F$  with the Higgs ( $H$ ) and lepton ( $L_i$ ) doublets, as well as explicit Majorana masses  $M$ ,

$$\mathcal{L} \supset -[F_{ij}L_i\epsilon HN_j + \frac{1}{2}M_{ij}N_iN_j], \quad (1)$$

(in Weyl spinor notation). In the seesaw limit,  $M \gg m_D$ , the neutrino mass spectrum splits into a light set given by the eigenvalues  $m_1 < m_2 < m_3$  of the matrix

$$m_\nu = -m_D M^{-1} m_D^T, \quad (2)$$

with the eigenstates corresponding mainly to mixings of the active left-handed neutrinos  $\nu_\alpha$ , and a heavy set given by the eigenvalues  $M_1 < M_2 < M_3$  of the matrix  $M$ , with the eigenstates corresponding to mixings of the sterile right-handed neutrinos  $N_i$ . The neutrino mass and mixing problem is thus solved by the usual type-I seesaw mechanism (Minkowski, 1977; Gell-Mann et al., 1979; Yanagida, 1979; Mohapatra and Senjanovic, 1980). Intriguingly, the baryogenesis and dark matter problems can be solved simultaneously if  $M_1 \sim \text{keV}$  and  $M_2 \sim M_3 \sim \text{GeV}$ . In fact, in this case  $N_{2,3}$  create flavored lepton asymmetries from CP-violating oscillations in the early Universe, which generate

the baryon asymmetry of the Universe via ARS leptogenesis (Akhmedov et al., 1998). The lightest sterile neutrino  $N_1$  can act as dark matter, with the correct relic abundance achieved through freeze-in production, resonantly enhanced with the MSW effect (Wolfenstein, 1978, 1979; Mikheyev and Smirnov, 1985). Moreover, it was argued in Bezrukov and Shaposhnikov (2008) that the puzzle of inflation can be solved even in the SM by allowing a non-minimal coupling of the Higgs field to the Ricci scalar,

$$S \supset - \int d^4x \sqrt{-g} \xi_H H^\dagger H R, \quad (3)$$

which promotes the Higgs field to an inflaton candidate.

However, the viability of the  $\nu$ MSM as a minimal model of particle cosmology is threatened by several facts. First of all, recent findings in astrophysics have seriously constrained the parameter space for  $N_1$  as a dark matter candidate (Schneider, 2016; Perez et al., 2017). Secondly, the generically large value of the non-minimal coupling  $\xi_H \sim 10^5 \sqrt{\lambda_H}$ , where  $\lambda_H$  is the Higgs self-coupling, required to fit the amplitude of the scalar perturbations inferred from the cosmic microwave background (CMB) temperature fluctuations, imply that perturbative unitarity breaks down at the scale  $M_P/\xi_H \sim 10^{14}$  GeV, where  $M_P = 1/\sqrt{8\pi G}$  is the reduced Planck mass, making the inflationary predictions unreliable (Barbon and Espinosa, 2009; Burgess et al., 2009). Even more, successful inflation cannot happen in this context if the quartic coupling  $\lambda_H$  in the Higgs potential:

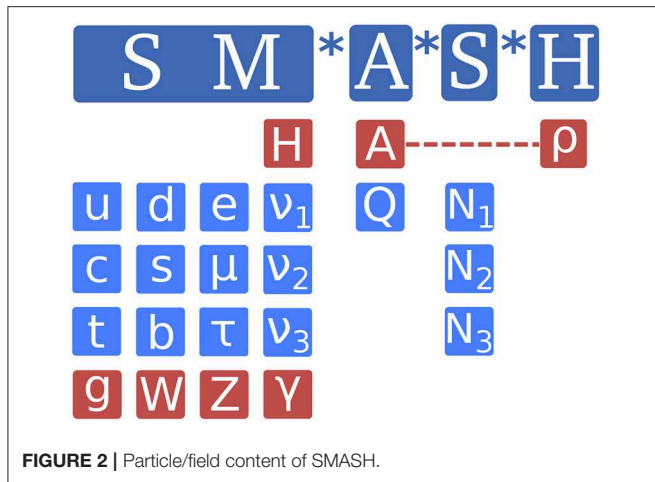
$$V(H) = \lambda_H \left( H^\dagger H - \frac{v^2}{2} \right)^2,$$

runs negative at large (Planckian) field values due to loop corrections involving the top quark. In fact, the central values of the strong gauge coupling and the Higgs and top quark masses imply that  $\lambda_H$  becomes negative at a field value corresponding to an energy scale  $\Lambda_I \sim 10^{11}$  GeV. This is much lower than what is required for Higgs inflation and thus inconsistent with it. However, given the current experimental uncertainties, a definite conclusion cannot yet be drawn (see e.g., Buttazzo et al., 2013; Bednyakov et al., 2015).

These obstacles of the  $\nu$ MSM can be neatly circumvented<sup>1</sup> in SMASH-type (Ballesteros et al., 2017a,b; Ernst et al., 2018) extensions of the SM which are built around the axion for the solution of the strong CP problem (Peccei and Quinn, 1977; Weinberg, 1978; Wilczek, 1978), as well as for dark matter, and allow inflation to be driven by (a mixture of the modulus of the Higgs field with) the modulus of the Peccei-Quinn field – sometimes called saxion field (Pi, 1984; Fairbairn et al., 2015).

This review is organized as follows. In section 2 we describe a number of Peccei-Quinn-type extensions of the  $\nu$ MSM: bottom-up constructions featuring KSVZ- and DFSZ-type axions (cf. sections 2.1, 2.2, respectively) and top-down constructions based

<sup>1</sup>Higgs inflation can also be realized in supergravity extensions of the SM (see for example Ben-Dayana and Einhorn, 2010; Choudhury et al., 2014; Pallis, 2017, 2018a,b; Pallis and Shafi, 2018).



on non-supersymmetric grand unification (cf. section 2.3). Section 3 is devoted to inflation, while stability is analyzed in section 4. Reheating is reviewed in section 5, dark matter in section 6, and baryogenesis in section 7. Conclusions are drawn in section 8.

## 2. SMASH AND ITS VARIANTS

In this section we will describe a number of extensions of the SM which exploit the Peccei-Quinn (PQ) mechanism (Peccei and Quinn, 1977) to solve the strong CP problem and thus have the potential to solve the big five problems of particle physics and cosmology in one smash.

### 2.1. SMASH

The model with smallest field content—dubbed here and in the following SMASH—is based on a KSVZ-type axion model (Kim, 1979; Shifman et al., 1980): a SM-singlet complex scalar field  $\sigma$ , which features a (spontaneously broken) global  $U(1)_{\text{PQ}}$  symmetry, and a vector-like colored Dirac fermion  $Q$ , which transforms as  $(3, 1, -1/3)$  or, alternatively, as  $(3, 1, 2/3)$  under the SM gauge group  $SU(3)_C \times SU(2)_L \times U(1)_Y$  and which transforms chirally under  $U(1)_{\text{PQ}}$ , are added to the field content of the  $\nu\text{MSM}$  (cf. Figure 2). The scalar potential, which relates the Higgs field  $H$  to  $\sigma$ , is assumed to have the general form

$$V(H, \sigma) = \lambda_H \left( H^\dagger H - \frac{v^2}{2} \right)^2 + \lambda_\sigma \left( |\sigma|^2 - \frac{v_\sigma^2}{2} \right)^2 + 2\lambda_{H\sigma} \left( H^\dagger H - \frac{v^2}{2} \right) \left( |\sigma|^2 - \frac{v_\sigma^2}{2} \right), \quad (4)$$

with  $\lambda_H, \lambda_\sigma > 0$  and  $\lambda_{H\sigma}^2 < \lambda_H \lambda_\sigma$ , in order to ensure that both the electroweak symmetry and the PQ symmetry are broken in

<sup>2</sup>These hypercharge assignments ensure that  $Q$  can mix with the right-handed SM down-type quarks or up-quarks, respectively, allowing its decay to the latter, thereby evading overabundance problems (Nardi and Roulet, 1990; Berezhiani et al., 1992).

the vacuum; i.e., the minimum of the scalar potential is attained at the vacuum expectation values (VEVs)

$$\langle H^\dagger H \rangle = v^2/2, \quad \langle |\sigma|^2 \rangle = v_\sigma^2/2, \quad (5)$$

where  $v = 246 \text{ GeV}$ . The PQ symmetry breaking scale  $v_\sigma$  is assumed to be much larger than the Higgs VEV  $v$ . Correspondingly, the particle excitation of the modulus  $\rho$  of  $\sigma$ , cf.

$$\sigma(x) = \frac{1}{\sqrt{2}} [v_\sigma + \rho(x)] e^{iA(x)/v_\sigma}, \quad (6)$$

gets a large mass

$$m_\rho = \sqrt{2\lambda_\sigma} v_\sigma + \mathcal{O}\left(\frac{v}{v_\sigma}\right), \quad (7)$$

while the particle excitation  $A$  of the angular degree of freedom of  $\sigma$ —which is dubbed “axion” in the context of the PQ solution of the strong CP problem (Weinberg, 1978; Wilczek, 1978)—is a massless Nambu-Goldstone (NG) boson,  $m_A = 0$ .

However, due to the assumed chiral transformation of the new vector-like fermion  $Q$ , the  $U(1)_{\text{PQ}}$  symmetry is broken due to the gluonic triangle anomaly,

$$\partial_\mu J_{U(1)_{\text{PQ}}}^\mu \supset -\frac{\alpha_s}{8\pi} G_{\mu\nu}^c \tilde{G}^{c,\mu\nu}. \quad (8)$$

Under these circumstances, the NG field

$$\theta(x) \equiv \frac{A(x)}{f_A}, \quad \text{with } f_A \equiv v_\sigma, \quad (9)$$

acts as a space-time dependent  $\theta$ -angle in QCD. In fact, the anomaly ensures that, at energies above the scale of QCD,  $\Lambda_{\text{QCD}}$ , but far below the scale of PQ symmetry breaking,  $v_\sigma$ , that is after integrating out the saxion  $\rho$  and the vector-like quark  $Q$ , which also gets a large mass from its Yukawa coupling with the PQ scalar,

$$m_Q = \frac{y}{\sqrt{2}} v_\sigma + \mathcal{O}\left(\frac{v}{v_\sigma}\right), \quad (10)$$

the effective Lagrangian of the axion has the form

$$\mathcal{L}_\theta = \frac{f_A^2}{2} \partial_\mu \theta \partial^\mu \theta - \frac{\alpha_s}{8\pi} \theta(x) G_{\mu\nu}^c \tilde{G}^{c,\mu\nu}. \quad (11)$$

Correspondingly, the  $\bar{\theta}$ -angle in QCD can be eliminated by a shift  $\theta(x) \rightarrow \theta(x) - \bar{\theta}$ . At energies below  $\Lambda_{\text{QCD}}$ , the effective potential of the shifted field, which for convenience we again denote by  $\theta(x)$ , will then coincide with the vacuum energy of QCD as a function of  $\bar{\theta}$

$$V(\theta) \equiv -\frac{1}{\mathcal{V}} \ln \frac{Z(\bar{\theta})}{Z(0)} \Big|_{\bar{\theta}=\theta} \simeq \Sigma_0 (m_u + m_d) \left( 1 - \frac{\sqrt{m_u^2 + m_d^2 + 2m_u m_d \cos \theta}}{m_u + m_d} \right), \quad (12)$$

where  $\mathcal{V}$  is the Euclidean space-time volume,  $Z(\bar{\theta})$  is the partition function of QCD, and  $\Sigma_0 = -\langle \bar{u}u \rangle = -\langle \bar{d}d \rangle$  is the chiral condensate (Vecchia and Veneziano, 1980; Leutwyler and Smilga, 1992). Notably, CP is conserved in the vacuum, since  $V(\theta)$  has an absolute minimum at  $\theta = 0$  and thus the vacuum expectation value of  $\theta$  vanishes,  $\langle \theta \rangle = 0$  (Vafa and Witten, 1984). Expanding the potential around zero and using

$$m_\pi^2 = \frac{\Sigma_0}{f_\pi^2} (m_u + m_d) + \mathcal{O}(m^2), \quad (13)$$

one finds the mass of the axion as the coefficient of the quadratic term,

$$m_A \equiv \frac{\sqrt{\chi_0}}{f_A} \simeq \frac{m_\pi f_\pi}{f_A} \frac{\sqrt{m_u m_d}}{m_u + m_d}, \quad (14)$$

where  $\chi_0$  is the topological susceptibility in QCD,  $m_\pi = 135$  MeV the neutral pion mass,  $f_\pi \approx 92$  MeV its decay constant, and  $m_u, m_d$  are the masses of the lightest quarks, with ratio  $z = m_u/m_d \approx 0.56$ . A recent determination in next-to-leading order (NLO) chiral perturbation theory (Grilli di Cortona et al., 2016) yielded  $\chi_0 = [75.5(5)\text{MeV}]^4$ , which agrees beautifully with the result from lattice QCD,  $\chi_0 = [75.6(1.8)(0.9)\text{MeV}]^4$  (Borsanyi et al., 2016), resulting in<sup>3</sup>

$$m_A = 57.0(7) \left( \frac{10^{11} \text{GeV}}{f_A} \right) \mu\text{eV}. \quad (15)$$

Moreover, also couplings to the photon and the nuclei are inherited from the axion's mixing with the pion. The full low energy Lagrangian of the axion with photons ( $F_{\mu\nu}$ ), nucleons,  $\psi_N = p, n$ , electrons ( $e$ ) and active neutrinos ( $\nu_i$ ) has the generic form

$$\begin{aligned} \mathcal{L}_A = & \frac{1}{2} \partial_\mu A \partial^\mu A - V(A) - \frac{\alpha}{8\pi} C_{A\gamma} \frac{A}{f_A} F_{\mu\nu} \tilde{F}^{\mu\nu} \\ & + \frac{1}{2} C_{AN} \frac{\partial_\mu A}{f_A} \bar{\psi}_N \gamma^\mu \gamma_5 \psi_N \\ & + \frac{1}{2} C_{Ae} \frac{\partial_\mu A}{f_A} \bar{\psi}_e \gamma^\mu \gamma_5 \psi_e + \frac{1}{2} C_{Av} \frac{\partial_\mu A}{f_A} \bar{\nu}_i \gamma^\mu \gamma_5 \nu_i, \end{aligned} \quad (16)$$

where  $V(A) = V(\theta = A/f_A)$ . The dimensionless coupling to photons,  $C_{A\gamma}$ , involves a model-independent part from the mixing with the pion and a model-dependent part depending of the electric charge of  $Q$ . It is given in **Table 1** for the two variants of SMASH. Similarly, the proton and neutron have a model-independent part and a model dependent contribution

**TABLE 1** | Axion predictions for two SMASH variants exploiting distinct vector-like quarks transforming as  $R_Q$  under the SM gauge group factors  $SU(3)_C \times SU(2)_L \times U(1)_Y$ : Axion decay constant  $f_A$ , coupling to the photon  $C_{A\gamma}$ , and tree-level couplings to quarks and charged leptons  $C_{Ai}$ ,  $i = u, \dots, t, e, \dots, \tau$ .

Model	$R_Q$	$f_A$	$C_{A\gamma}$	$C_{Ai}$
SMASH(d)	$(3, 1, -\frac{1}{3})$	$v_\sigma$	$\frac{2}{3} - 1.92(4)$	0
SMASH(u)	$(3, 1, +\frac{2}{3})$	$v_\sigma$	$\frac{8}{3} - 1.92(4)$	0

that arises from possible axion-quark couplings of the form  $(C_{Aq}/2)(\partial_\mu A/f_A) \bar{\psi}_q \gamma^\mu \gamma_5 \psi_q$  in the high-energy theory

$$\begin{aligned} C_{Ap} &= -0.47(3) + 0.88(3)C_{Au} - 0.39(2)C_{Ad} - 0.038(5)C_{As} \\ &\quad - 0.012(5)C_{Ac} - 0.009(2)C_{Ab} - 0.0035(4)C_{At}, \\ C_{An} &= -0.02(3) + 0.88(3)C_{Ad} - 0.39(2)C_{Au} - 0.038(5)C_{As} \\ &\quad - 0.012(5)C_{Ac} - 0.009(2)C_{Ab} - 0.0035(4)C_{At}, \end{aligned} \quad (17)$$

as found in the state-of-the-art calculation (Grilli di Cortona et al., 2016). In SMASH, all the axion-quark and axion-charged-lepton couplings vanish at tree level (cf. **Table 1**).

To avoid strong bounds from laboratory experiments and stellar astrophysics, the axion decay constant  $f_A$  has to be much larger than the electroweak scale (Tanabashi et al., 2018), notably  $f_A \gtrsim 10^8$  GeV from the measured duration of the neutrino signal of supernova 1987A (Raffelt, 2008; Fischer et al., 2016; Chang et al., 2018).

Optionally, one may unify the PQ symmetry with a lepton number symmetry by assigning PQ charges also to the leptons and sterile neutrinos (Shin, 1987; Dias et al., 2014). In this case, the latter get their Majorana masses also from PQ symmetry breaking,

$$M_{ij} = \frac{Y_{ij}}{\sqrt{2}} v_\sigma + \mathcal{O}\left(\frac{v}{v_\sigma}\right), \quad (18)$$

where  $Y_{ij}$  are Yukawa couplings, and the mass scale of the active neutrinos is determined by the PQ scale,

$$m_\nu = 0.04 \text{ eV} \left( \frac{10^{11} \text{GeV}}{v_\sigma} \right) \left( \frac{-F Y^{-1} F^T}{10^{-4}} \right). \quad (19)$$

Moreover, the axion  $A$  is in this case at the same time the majoron  $J$ : the NG boson arising from the breaking of the global lepton number symmetry (Chikashige et al., 1981; Gelmini and Roncadelli, 1981; Schechter and Valle, 1982). This leads to a non-zero tree-level coupling of the  $A/J$  to the active neutrinos,  $(-1/4)(\partial_\mu A/f_A) \bar{\nu}_i \gamma^\mu \gamma_5 \nu_i$  and to possibly sizeable loop-induced couplings to SM quarks and charged leptons from the loop involving the sterile neutrinos  $N_i$  (Shin, 1987; Pilaftsis, 1994). To lowest order in the seesaw limit,  $m_D/M_M \ll 1$ , they are given by Garcia-Cely and Heeck (2017)

$$C_{aq} \simeq \frac{1}{8\pi^2} T_3^q \text{tr} \kappa, \quad C_{a\ell} \simeq -\frac{1}{16\pi^2} (\text{tr} \kappa - 2\kappa_{\ell\ell}), \quad (20)$$

<sup>3</sup>Very recently, Gorghetto and Villadoro (2019) improved the theoretical prediction of  $\chi_0$  by including  $\mathcal{O}(\alpha)$  and NNLO corrections in the chiral expansion, resulting in  $\chi_0^{1/4} = 75.44(34)$  MeV, corresponding to  $m_A = 56.91(51) \mu\text{eV}$  ( $10^{11} \text{GeV}/f_A$ ), almost coinciding with the previous NLO result.

where  $T_3^d = -\frac{1}{2} = -T_3^u$  and the dimensionless hermitian  $3 \times 3$  matrix  $\kappa$  is defined as

$$\kappa \equiv \frac{m_D m_D^\dagger}{v^2} = \frac{FF^\dagger}{2}. \quad (21)$$

Intriguingly, a KSVZ-type axion/majoron with  $f_A \sim 10^8$  GeV may explain the  $\sim 3\sigma$  hint of an anomalously large energy loss of helium burning stars, red giants and white dwarfs, if  $|\kappa - 2\kappa_{ee}|$  is of order unity (Giannotti et al., 2017).

## 2.2. 2hdSMASH

A less minimal variant of SMASH—dubbed 2hdSMASH—exploits DFSZ-type axion models (Zhitnitsky, 1980; Dine et al., 1981): in those the SM Higgs sector is extended by two Higgs doublets,  $H_u$  and  $H_d$ , whose vacuum expectation values  $v_u$  and  $v_d$  give masses to up-type and down-type quarks, respectively. There are two possibilities, named 2hdSMASH(d) or 2hdSMASH(u), according to whether leptons couple to  $H_d$ , which occurs in familiar Grand Unified Theories (GUTs), or to  $H_u$ . The  $n_f = 6$  SM model quarks are assumed to carry PQ charges such that the gluonic triangle anomaly arises from them alone,

$$\partial_\mu J_{U(1)_{PQ}}^\mu \supset -n_f \frac{\alpha_s}{8\pi} G_{\mu\nu}^c \tilde{G}^{c,\mu\nu}. \quad (22)$$

The low-energy Lagrangian of a DFSZ-type PQ extension of the SM is identical to that of a 2 Higgs Doublet Model (2HDM), augmented by seesaw-generated neutrino masses (Equation 2), and the one of a DFSZ-type axion. The DFSZ axion properties are given in **Table 2**. In this case, there are tree-level couplings to quarks and leptons. In fact, the anomalous stellar energy losses mentioned above can be alternatively explained by a DFSZ-type axion with  $f_A \gtrsim 10^8$  GeV and  $\tan\beta \equiv v_u/v_d \sim 1$  (Giannotti et al., 2017).

Again, optionally the PQ symmetry may be unified with a lepton number symmetry (Langacker et al., 1986; Volkas et al., 1988; Clarke and Volkas, 2016), in which case the active neutrino mass scale is determined by the PQ scale and the DFSZ axion is at the same time a Majoron.

## 2.3. gutSMASH

As commented in the previous section, the model 2hdSMASH(d) can be embedded into a GUT. The simplest unified group is  $SU(5)$  (Georgi and Glashow, 1974; Georgi, 1975), with each generation of fermions (not including right-handed neutrinos) fitting into the representations  $10_F$  and  $\bar{5}_F$ , with  $SU(5)$  broken into the SM group by the VEV of a scalar in the  $24_H$ , and with the electroweak breaking carried out by two scalars in the  $5_H$ . It was realized early on that  $SU(5)$  GUTs can accommodate an axion with a decay constant  $f_A$  tied to the unification scale (Wise et al., 1981). However, minimal non-supersymmetric  $SU(5)$  GUTs are incompatible with proton decay limits, because the  $SU(2)$  and  $U(1)$  gauge couplings meet at too low a scale. However, there are viable extensions in which particles in additional  $SU(5)$  multiplets appropriately modify the running of the gauge couplings so as to yield successful unification compatible with proton decay limits. The extension proposed

**TABLE 2 |** DFSZ-type axion predictions: axion decay constant  $f_A$ , coupling to the photon  $C_{A\gamma}$ , and tree-level couplings to quarks and charged leptons  $C_{Ai}$ ,  $i = u, \dots, t, e, \dots, \tau$ , with  $\tan\beta \equiv v_u/v_d$ .

Model	$f_A$	$C_{A\gamma}$	$C_{Au}$	$C_{Ad}$	$C_{Ae}$
2hdSMASH(d)	$v_\sigma/6$	$\frac{8}{3} - 1.92(4)$	$\frac{1}{3} \cos^2 \beta$	$\frac{1}{3} \sin^2 \beta$	$\frac{1}{3} \sin^2 \beta$
2hdSMASH(u)	$v_\sigma/6$	$\frac{8}{3} - 1.92(4)$	$\frac{1}{3} \cos^2 \beta$	$\frac{1}{3} \sin^2 \beta$	$\frac{1}{3} \cos^2 \beta$

in Bajc and Senjanovic (2007) and further studied in Bajc et al. (2007), Luzio and Mihaila (2013) makes use of a fermionic multiplet in the  $24_F$ , which contains right-handed neutrinos getting a mass from the VEV of the  $24_H$ , which breaks  $SU(5)$  into the SM. This generates masses for the light neutrinos through a combination of the type I and III seesaw mechanisms, and also allows for baryogenesis from leptogenesis. When extending this viable  $SU(5)$  model to accommodate a global PQ symmetry with its corresponding axion (Di Luzio et al., 2018), one has a SMASH-type construction with the complex scalar in the  $24_H$  containing the axion and acting as a Majoron. The Lagrangian of this model, which we will refer to as miniSU(5)PQ, contains the following interactions (written only schematically),

$$\mathcal{L} \supset \bar{5}_F 10_F 5_H'^* + 10_F 10_F 5_H + \bar{5}_F 24_F 5_H + \text{Tr} 24_F^2 24_H'^* + 5_H'^* 24_H^2 5_H + 5_H'^* 5_H \text{Tr}(24_H^2) + \text{h.c.}, \quad (23)$$

which enforce the PQ charge assignments in **Table 3**.

The axion decay constant is related to the unification scale  $v_U$  as  $f_A = v_U/11$ , while the axion couplings to nucleons and leptons are given in **Table 4**.

The unification scale turns out to be highly constrained and grows with decreasing mass of the light fermion triplet contained in  $24_F$ . This is due to the fact that increasing the unification scale requires a larger deviation in the running of the  $SU(2)$  and  $U(1)$  gauge couplings with respect to the SM case, which can only be achieved if the extra particles with electroweak charges in the  $24_F$  multiplet become lighter. The light electroweak triplets can be probed by LHC searches (Arhrib et al., 2010; Sirunyan et al., 2017), which then give upper bounds for  $v_U \propto f_A$ . On the other hand, proton decay experiments, such as Super-Kamiokande (Abe et al., 2017) constrain the unification scale from below. Given the relation (15) between  $f_A$  and the axion mass, this results in a remarkably constrained window of allowed values of  $m_A$ :

$$m_A \in [4.8, 6.6] \text{ neV}. \quad (24)$$

The upper limit can be relaxed to  $m_A < 330$  neV when allowing for fine-tuning in the flavor structure of the model so as to close as many decay channels for the proton as possible (Dorsner and Fileviez Perez, 2005). The above axion mass window can be targeted in a complementary manner by future high-energy colliders (Ruiz, 2015; Cai et al., 2018), proton decay experiments, such as Hyper-Kamiokande (Abe et al., 2011), as well as direct axion dark matter searches with CASPER-Electric (Budker et al., 2014; Jackson Kimball et al., 2017) and ABRACADABRA (Kahn et al., 2016).

**TABLE 3 |** Field content and PQ charge assignments in the PQ-extended  $SU(5)$  model of Di Luzio et al. (2018).

Model	$\bar{5}_F$	$10_F$	$24_F$	$5_H$	$5'_H$	$24_H$
miniSU(5)PQ	1	1	1	−2	2	2

**TABLE 4 |** Axion predictions in  $SU(5) \times U(1)_{PQ}$  (Di Luzio et al., 2018) and  $SO(10) \times U(1)_{PQ}$  models (Ernst et al., 2018): axion decay constant  $f_A$ , coupling to the photon  $C_{A\gamma}$ , and tree-level couplings to quarks and charged leptons  $C_{Ai}$ ,  $i = u, \dots, t, e, \dots, \tau$ .

Model	$f_A$	$C_{A\gamma}$	$C_{Au}$	$C_{Ad}$	$C_{At}$
miniSU(5)PQ	$v_U/11$	$\frac{8}{3} - 1.92(4)$	$\frac{2}{11} \cos^2 \beta$	$\frac{2}{11} \sin^2 \beta$	$\frac{2}{11} \sin^2 \beta$
miniSO(10)PQ	$v_U/3$	$\frac{8}{3} - 1.92(4)$	$\frac{1}{3} \cos^2 \beta$	$\frac{1}{3} \sin^2 \beta$	$\frac{1}{3} \sin^2 \beta$
gutSMASH	$v_\sigma/3$	$\frac{8}{3} - 1.92(4)$	$\frac{1}{3} \cos^2 \beta$	$\frac{1}{3} \sin^2 \beta$	$\frac{1}{3} \sin^2 \beta$

In the  $SU(5)$  theory,  $\tan \beta = v_H/v'_H$ , while for the  $SO(10)$  models  $\tan^2 \beta = ((v_U^{10})^2 + (v_U^{126})^2)/((v_U^{10})^2 + (v_U^{126})^2)$ .

The smallness of the axion mass in this model implies that the axion can be identified with dark matter only if the Peccei-Quinn symmetry is broken before or during inflation and not restored afterwards, as reviewed in section 6. On the other hand, the large value of  $f_A$  implies that inflation can source large axionic isocurvature fluctuations which may be in conflict with observations (cf. section 6).

Compared to  $SU(5)$  GUTs, theories based on the  $SO(10)$  group (Fritzsch and Minkowski, 1975; Georgi, 1975) can yield viable unification patterns without the need to either consider supersymmetric extensions or to add additional fermion multiplets beyond those containing the SM fermions. Moreover, right-handed neutrinos are automatically incorporated, as these occur automatically with the rest of the SM quarks and leptons if one considers three spinorial representations  $16_F$  of  $SO(10)$ . The latter can have the following Yukawa couplings with scalar Higgses in the  $10_H$  and  $\overline{126}_H$  representations,

$$\mathcal{L}_Y = 16_F (Y_{10} 10_H + \tilde{Y}_{10} 10_H^* + Y_{126} \overline{126}_H) 16_F + \text{h.c.}, \quad (25)$$

which can give rise to the seesaw mechanism (Gell-Mann et al., 1979). Moreover, a PQ symmetry, under which the fields transform as

$$16_F \rightarrow 16_F e^{i\alpha}; \quad 10_H \rightarrow 10_H e^{-2i\alpha}; \quad \overline{126}_H \rightarrow \overline{126}_H e^{-2i\alpha}, \quad (26)$$

can be motivated independently from the strong CP problem: it forbids the second term in the Yukawa interactions (25), thereby crucially improving the economy and predictivity of the models (Babu and Mohapatra, 1993; Bajc et al., 2006).

Adding a further Higgs representation, say  $210_H$ , the  $SO(10)$  symmetry can be broken at the unification scale  $M_U$  by the VEV of the  $210_H$  to the Pati-Salam gauge group  $SU(4)_C \times SU(2)_L \times SU(2)_R$ , which is broken to the SM gauge group  $SU(3)_C \times SU(2)_L \times U(1)_Y$  at the scale of  $B-L$  breaking  $M_{BL}$  (which is thus

**TABLE 5 |** Field content and PQ charge assignments in two distinct  $SO(10) \times U(1)_{PQ}$  models (Ernst et al., 2018).

Model	$16_F$	$\overline{126}_H$	$10_H$	$210_H$	$\sigma$
miniSO(10)PQ	1	−2	−2	4	−
gutSMASH	1	−2	−2	0	4

the seesaw scale) by the VEV of the  $\overline{126}_H$ , which itself is broken at the weak scale  $M_Z$  by the VEV of the  $10_H$ ,

$$SO(10) \xrightarrow{M_U - 210_H} 4_C 2_L 2_R \xrightarrow{M_{BL} - \overline{126}_H} 3_C 2_L 1_Y \xrightarrow{M_Z - 10_H} 3_C 1_{em}.$$

Unfortunately, the minimal PQ symmetry (26) leads to a decay constant  $f_A = v/3$  (Holman et al., 1983; Mohapatra and Senjanovic, 1983; Altarelli and Meloni, 2013; Ernst et al., 2018), which is clearly experimentally excluded. The simplest way to remedy this problem is to associate a PQ charge also to the  $210_H$ ,

$$210_H \rightarrow 210_H e^{4i\alpha}. \quad (27)$$

We dub this model miniSO(10)PQ—for Minimal  $SO(10) \times U(1)_{PQ}$  model—and summarize the field content and PQ charge assignments in the first row of Table 5. Its axion properties are given in Table 4.

The photon and fermion couplings are the same as for 2hdSMASH(d), although the microscopic origin of the parameter  $\beta$  differs, as it is determined by the VEVs of four Higgses, as opposed to two in DFSZ models. Moreover, as in miniSU(5)PQ, the decay constant in miniSO(10)PQ is proportional to the scale of grand unification,  $f_A = v_U/3$ , which is determined by the requirement of gauge coupling unification. Therefore, this model is more predictive in the axion sector than SMASH or 2hdSMASH, yet less predictive than miniSU(5)PQ due to the additional freedom inherent in having a multi-step breaking of the grand unified group—as opposed to the single-step breaking in the  $SU(5)$  case—as well as due to the additional threshold corrections that can arise from the greater number of particles included in the  $SO(10)$  multiplets. Allowing for a reasonable range of scalar threshold corrections and taking into account constraints from black hole superradiance (Arvanitaki et al., 2015) and proton decay, the axion decay constant and mass is predicted to lie in the range (Ernst et al., 2018).

$$2.6 \times 10^{15} \text{ GeV} < f_A < 3.0 \times 10^{17} \text{ GeV}, \\ 1.9 \times 10^{-11} \text{ eV} < m_A < 2.2 \times 10^{-9} \text{ eV}. \quad (28)$$

As in the miniSU(5)PQ model, such light axion can only be compatible with dark matter with a pre-inflationary breaking of the PQ symmetry, and isocurvature constraints can be important. In fact, a one-step breaking model analogous to miniSU(5)PQ can also be realized in  $SO(10)$  by breaking the group at a high scale not just with the  $210_H$ , but with the added effect of a non-zero VEV in a  $45_H$  scalar multiplet (Boucenna et al., 2019). In this model, successful unification with a proton lifetime in reach of Hyper-Kamiokande is achieved by ensuring that the octets and triplets inside the  $210_H$  remain light, in analogy with

the light triplets in miniSU(5)PQ. The PQ charge of the  $210_H$  is now zero, while the  $45_H$  is assigned charge 4, which still gives a GUT-scale axion with a low mass and thus affected by isocurvature constraints.

Such constraints can be definitely evaded in the  $SO(10) \times U(1)_{PQ}$  variant dubbed gutSMASH whose field content and PQ charge assignments are specified in the second row of **Table 5**. In this model the  $210_H$  has no PQ charge. Instead, it features a further complex singlet scalar  $\sigma$  which is charged under the PQ symmetry. Its VEV determines the PQ symmetry breaking scale (see also Babu and Khan, 2015; Boucenna and Shafi, 2018) and the axion decay constant turns out to be  $f_A = v_\sigma/3$  (Ernst et al., 2018) (cf. second row of **Table 4**), which is a free parameter of the model.

### 3. INFLATION

In SMASH and its variants, introduced in the last section, there are two or more scalar fields that in principle could have driven primordial inflation. Let us look into this issue in some detail.

In SMASH, the modulus of the complex PQ field,  $\rho^2 = 2|\sigma|^2$ , or a mixture of it with  $h$ , the neutral component of the Higgs doublet in the unitary gauge,  $H^t = (0, h)/\sqrt{2}$ , is a viable inflaton candidate. It was pointed out in Bezrukov and Shaposhnikov (2008) that a non-minimal coupling of the Higgs,  $H$ , to the Ricci scalar  $R$  [cf. Equation (3)], would allow  $h$  to play that role, in a model that is since dubbed Higgs inflation. Indeed, after scalar and metric field redefinitions into the so-called Einstein frame, this kind of coupling flattens any quartic potential, making it convex and asymptotically flat at large field values (Salopek et al., 1989), approaching a plateau-like form which is preferred by CMB measurements (Akrami et al., 2018). However, as mentioned in the Introduction, a large value of the non-minimal coupling  $\xi_H$ —as required to fit the amplitude of the primordial scalar fluctuations ( $\xi_H \sim 5 \times 10^4 \sqrt{\lambda_H}$ ) for the central value of the top quark mass (Tanabashi et al., 2018) (see also Figure 14 of Aaboud et al., 2018)—implies that perturbative unitarity breaks down at a scale  $M_P/\xi_H$ , well below the Higgs field values during inflation  $h \sim M_P/\sqrt{\xi_H}$  and comparable to the scale given by the fourth square root of the potential (Barbon and Espinosa, 2009; Burgess et al., 2009). See Bezrukov and Shaposhnikov (2014) and Hamada et al. (2014) for the statistically disfavored possibility of reducing  $\xi_H$  by considering significantly smaller top masses.

This problem can be eliminated in Hidden Scalar Inflation (HSI) (Pi, 1984; Fairbairn et al., 2015; Boucenna and Shafi, 2018) or Higgs-Hidden Scalar inflation (HHSI) (Ballesteros et al., 2017a,b), which exploit a non-minimal coupling analogous to the previous one:

$$S \supset - \int d^4x \sqrt{-g} \xi_\sigma \sigma^* \sigma R. \quad (29)$$

Such couplings are not *ad-hoc*, since they are generated radiatively in a Friedman-Robertson-Walker background. For negligible  $\xi_H$ , slow-roll inflation with a tree-level asymptotically flat potential can thus happen along two different directions in field space: the  $\rho$ -direction for  $\lambda_{H\sigma} > 0$  (HSI) and the line

$h/\rho = \sqrt{-\lambda_{H\sigma}/\lambda_H}$  for  $\lambda_{H\sigma} < 0$  (HHSI) (cf. **Figure 3**). In both cases, inflation can be described in the Einstein frame by a single canonically normalized field  $\chi$  with potential

$$\tilde{V}(\chi) = \frac{\lambda}{4} \rho(\chi)^4 \left( 1 + \xi_\sigma \frac{\rho(\chi)^2}{M_P^2} \right)^{-2}, \quad (30)$$

where

$$\lambda \equiv \begin{cases} \lambda_\sigma, & \text{for HSI,} \\ \lambda_\sigma \left( 1 - \frac{\lambda_{H\sigma}^2}{\lambda_\sigma \lambda_H} \right), & \text{for HHSI.} \end{cases} \quad (31)$$

The field  $\chi$  is the solution of  $\Omega^2 d\chi/d\rho \simeq (b\Omega^2 + 6\xi_\sigma^2 \rho^2/M_P^2)^{1/2}$ , with  $\Omega \simeq 1 + \xi_\sigma \rho^2/M_P^2$  being the Weyl transformation into the Einstein frame and  $b = 1$  (for HSI) or  $b = 1 + |\lambda_{H\sigma}/\lambda_H|$  (for HHSI). We will see in the next section that vacuum stability requires a small value of  $|\lambda_{H\sigma}| \lesssim 10^{-6}$  and consequently  $b \sim 1$  in HHSI, which makes practically impossible distinguishing between HSI and HHSI from the measurements of the CMB power spectra. However, even a small Higgs component in the inflaton is a key aspect for reheating, which sets apart both possibilities, as we will discuss later.

**Figure 4** from Ballesteros et al. (2017b) shows the agreement of the non-minimally coupled potential (30) with the CMB at the pivot scale  $0.05 \text{ Mpc}^{-1}$  (Ade et al., 2016a,b), summarized in the amplitude of scalar perturbations  $A_s$ , the spectral index  $n_s$ , and the tensor-to-scalar ratio  $r$ ,

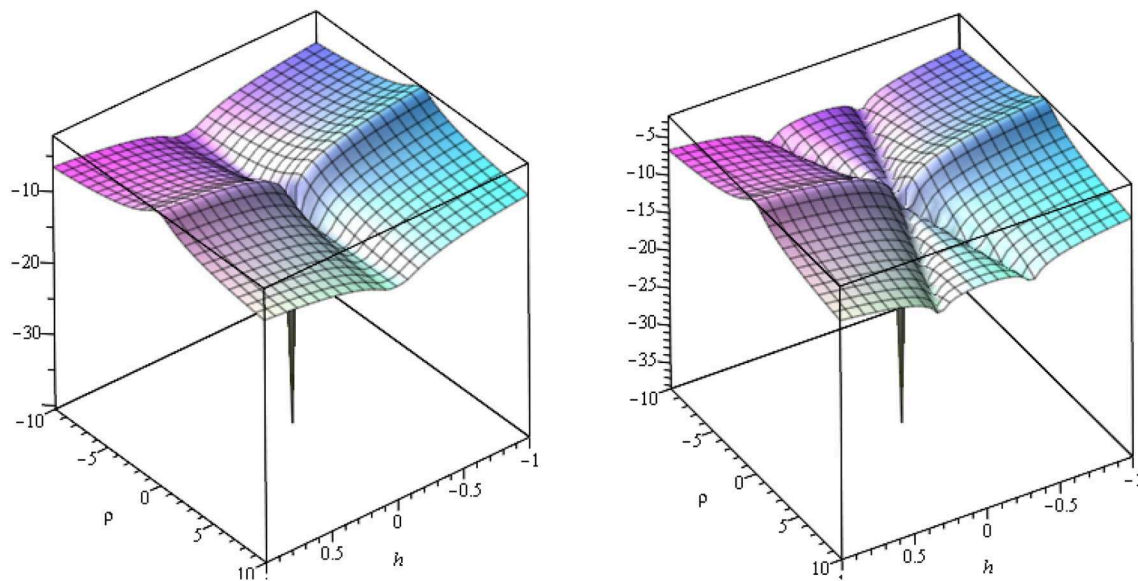
$$A_s = (2.207 \pm 0.103) \times 10^{-9}, \quad (32)$$

$$n_s = 0.969 \pm 0.004, \quad (33)$$

$$r < 0.07. \quad (34)$$

Current constraints from the latest Planck analysis (July 2018) are very similar to the ones quoted above (Akrami et al., 2018). Importantly, the effective quartic coupling  $\lambda$  has to be small enough,  $\lambda \lesssim 10^{-10}$ , so that the required non-minimal coupling to fit the amplitude of primordial scalar perturbations is at most  $\xi_\sigma \lesssim 1$  [cf. **Figure 4** (up right)]. In this region of parameter space, the perturbative consistency of HSI and HHSI is guaranteed and superior to Higgs Inflation, which necessarily operates at large  $\xi_H$  for the measured value of the top mass, since in this latter case the value of  $\lambda_H$  as determined from the measured Higgs mass is sizable (Ballesteros et al., 2017b). The predictions of the potential (30) in the case  $\lambda = \lambda_\sigma$  (or  $b \rightarrow 1$  in HHSI) for the tensor-to-scalar ratio  $r$  vs. the scalar spectral index  $n_s$  are shown in **Figure 5** for various values of  $\xi_\sigma$ . The requirement of predictive inflation, free of unitarity problems, demands  $r \gtrsim 0.01$ , which will be probed by the next generation of CMB experiments, such as CMB-S4 (Abazajian, 2016), LiteBird (Matsumura et al., 2014), and the Simons Observatory (Aguirre et al., 2019). Since in SMASH and its extensions the particle content is known, the reheating process can be computed in detail. This allows to constrain  $n_s$  and  $r$  to a narrow band, unlike for generic inflationary potentials devoid of a connection to the SM.

The generalization of Equation (31) to the case of a 2HDM—as relevant for the 2hdSMASH model—or to even



**FIGURE 3 |** Decadic log of the scalar potential (4) in the Einstein frame ( $\xi_H \ll \xi_\sigma$ ), as a function of  $h$  and  $\rho$ , all in units of  $M_P$ , supporting, for  $\lambda_{H\sigma} > 0$ , pure Hidden Scalar Inflation (HSI) (left), and, for  $\lambda_{H\sigma} < 0$ , Higgs-Hidden Scalar Inflation (HHSI) (right) [taken and used with permission from Ballesteros et al. (2017b)]. Inflation proceeds along one of the valleys. The couplings have been chosen such that the amplitude of primordial scalar perturbation is properly normalized.

more scalars—as relevant for gutSMASH—has not been worked out yet in full generality. For the related non-minimal Higgs Inflation in the 2HDM (see Gong et al., 2012). However, as far as HSI inflation is concerned, i.e., as long as the non-minimal couplings of all scalars apart from the saxion can be neglected, it is clear that the relevant potential for inflation is—in the Einstein frame—identical in SMASH HSI. Correspondingly, in this case, the same inflationary predictions as in SMASH HSI apply also for 2hdSMASH and gutSMASH HSI.

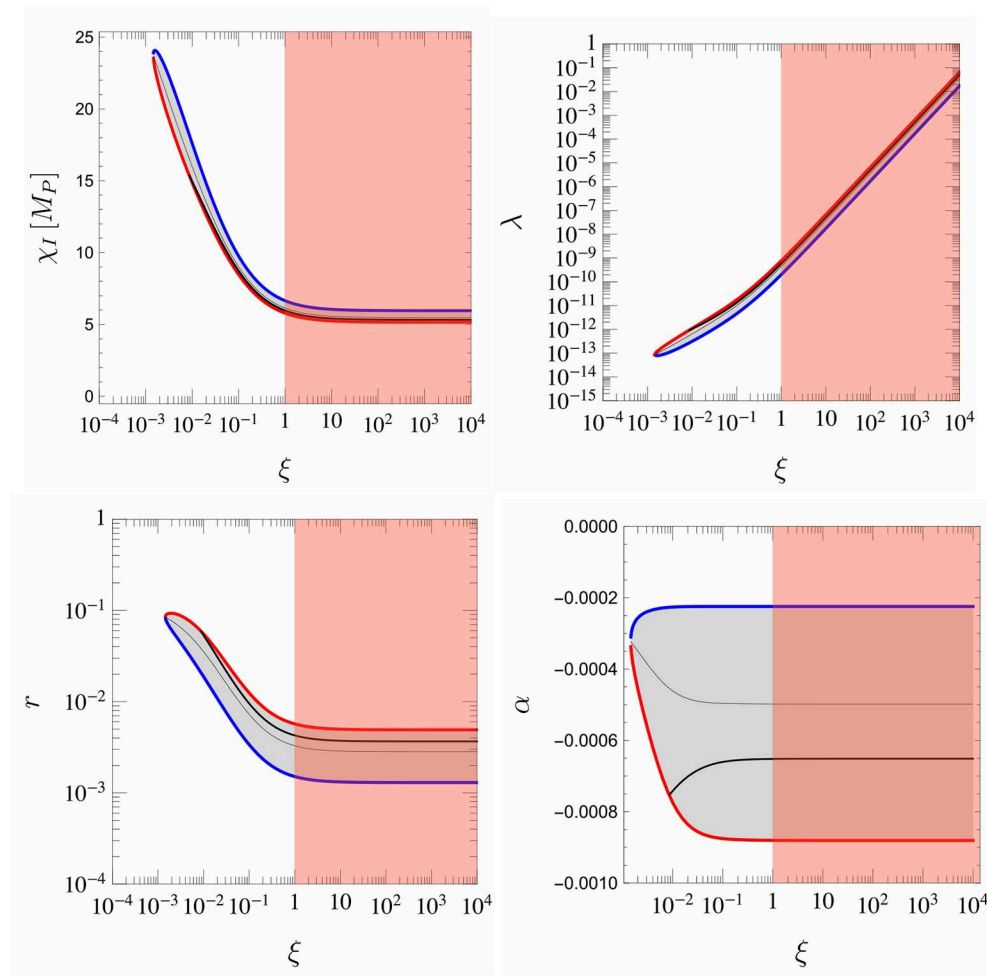
## 4. STABILITY

Primordial inflation of the kind described in the previous section is driven by a positive potential energy and Planckian field excursions. Therefore, a consistent realization within SMASH-type models requires a positive effective potential all the way up to the Planck scale. Although classical dynamics during inflation only requires a positive effective potential along the inflationary trajectory, instabilities in other regions of field space are dangerous because the fields can end up trapped in them as a result of the quantum fluctuations generated during inflation. To avoid this issue altogether we can demand a strictly positive potential in all field directions. Such requirement of (absolute) stability is threatened in the SM by loop corrections to the Higgs potential due to the top quark. When capturing virtual corrections by means of an RG-improved effective potential with parameters that run with the field scale ( $\mu \propto h$ ), an instability arises for the preferred values of the Higgs and top masses as a result of negative contributions to the beta function of the Higgs quartic coupling. In SMASH(d/u) (cf. Table 1)—with a

portal interaction between the Higgs and the complex scalar  $\sigma$  containing the axion—one can circumvent this problem thanks to the threshold stabilization mechanism pointed out in Elias-Miro et al. (2012) and Lebedev (2012). In the presence of the Higgs portal coupling, with the  $\sigma$  field acquiring a large VEV, the relation between the Higgs mass and the Higgs quartic coupling is altered with respect to that in the SM, such that the quartic can be larger in SMASH than in the SM. At an appropriate matching scale  $\mu_0$ , the couplings in SMASH and the SM are related as

$$\lambda_H(\mu_0) = \lambda_H^{\text{SM}}(\mu_0) + \delta(\mu_0), \quad \delta \equiv \frac{\lambda_{H\sigma}^2(\mu_0)}{\lambda_\sigma(\mu_0)}. \quad (35)$$

Despite the larger value of  $\lambda_H$ , stabilization is a bit subtle because, as expected from the decoupling of the massive  $\sigma$  field at low scales, the SM potential with its corresponding quartic can always be recovered in an appropriate region of field space. For  $\lambda_{H\sigma} > 0$  this region is of limited extent and can be made not to reach the SM instability scale. Then the potential in the SM-like region can stay positive, while outside of it the larger value of  $\lambda_H$  can ensure stability up to Planckian scales. Stabilization is then a tree-level effect and requires a small enough  $v_\sigma$  (which is harder to realize in GUT models), in order to ensure that the SM-like region does not go beyond the scale of the SM instability. For  $\lambda_{H\sigma} < 0$  on the other hand the SM-like region of the potential extends to arbitrary scales, and stabilization must crucially rely on loop effects that correct the running of the effective quartic coupling in the SM-like region. Stability can be achieved thanks to the positive contributions to the beta function of  $\lambda_H$  proportional to  $\lambda_H$  itself, which can counter-balance the negative corrections from



**FIGURE 4 |** Confidence level (C.L.) contours (95% C. L.) of the parameters in the scalar potential (30), as a function of the non-minimal coupling parameter  $\xi = \xi_\sigma$ , yielding inflation constrained by Planck 2015 observations at the pivot scale  $0.05 \text{ Mpc}^{-1}$  [taken and used with permission from Ballesteros et al. (2017b)]. **(Upper left)** The canonical inflaton value  $\chi_I$ . **(Upper right)** The value of the quartic coupling. **(Lower left)** The predicted tensor-to-scalar ratio. **(Lower right)** The running of the spectral index. The best fit for a given  $\xi_\sigma$  is drawn as a thin black line, while the minimum and maximum values of  $n_s$  are drawn as red and blue curves, respectively, corresponding to a redder or bluer primordial spectrum of curvature perturbations. The thicker black line displays the predictions when accounting for the HHSI prediction of a universe expanding, immediately after inflation, according to the equation of state of radiation domination. The region  $\xi_\sigma > 1$ , where perturbative unitarity fails in SMASH, is shaded.

the top quark: while in the SM the effect of the  $\lambda_H$ -dependent corrections is sub-dominant, this changes in SMASH due to the larger values of  $\lambda_H$  ensured by the modified matching in Equation (35).

Of course, one also needs to guarantee stability in the  $\sigma$  direction, which can again be endangered by fermion loops, this time coming from the RH neutrinos and the exotic quark  $Q$ . In this case stability can be achieved by demanding sufficiently small Yukawas.

After accounting for the previous effects, we have found that for the SMASH model stability in the Higgs direction can be achieved if the threshold parameter  $\delta$  in Equation (35) is roughly between  $10^{-3}$  and  $10^{-2}$  (for  $\lambda_{H\sigma} > 0$ ) or  $10^{-3}$  and  $10^{-1}$  (for  $\lambda_{H\sigma} < 0$ ), depending on the top mass. On the other hand, stability in the  $\sigma$  direction restricts the Yukawa couplings of the RH neutrinos and  $Q$  to

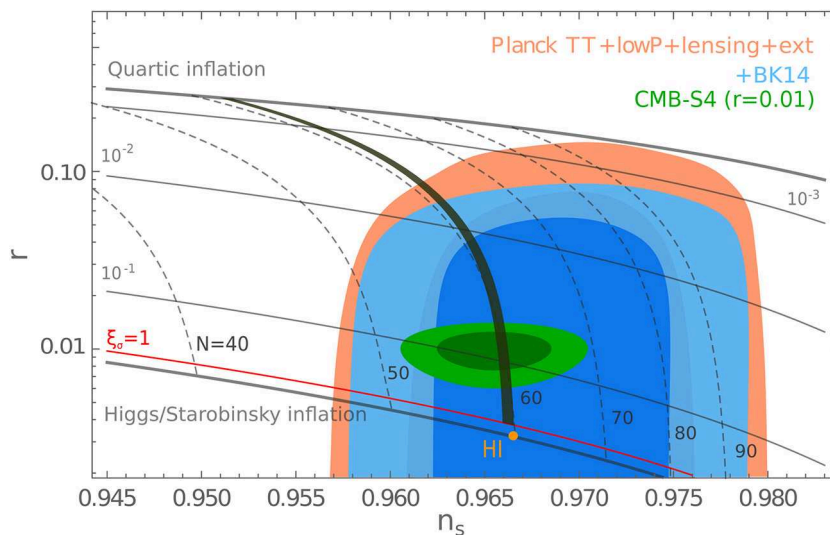
$$\sum_i Y_{ii}^4 + 6y^4 \lesssim \frac{16\pi^2 \lambda_\sigma}{\log\left(\frac{30M_P}{\sqrt{2}\lambda_\sigma v_\sigma}\right)}, \quad (36)$$

in the case that the Peccei-Quinn symmetry is extended to a lepton symmetry. Otherwise, the contribution of the Yukawas  $Y_{ii}$  on the left-hand side of Equation (36) is absent.

A stability analysis for 2hdSMASH and gutSMASH models is of course more involved due to the extra scalars and has not been done in full generality yet.

## 5. REHEATING

After inflation, the background scalar fields that drove the accelerated expansion will typically oscillate around a minimum of the potential, and throughout these oscillations they will lose energy by producing SM particles that reheat into a plasma which



**FIGURE 5 |** Predictions for the potential of Equation (30) in the  $r$  vs.  $n_s$  plane with a pivot scale of  $0.002 \text{ Mpc}^{-1}$  [adapted and used with permission from Ballesteros et al. (2017b)]. Contours of constant  $\xi_\sigma$  are shown as black solid lines. The SMASH prediction accounting for a consistent reheating history is given by the thick black line, while the thin dotted lines give isocontours of the number of e-folds that ignore reheating constraints. Also shown are the 68 and 95% C.L. regions at  $0.002 \text{ Mpc}^{-1}$  of Ade et al. (2016b) and the projected sensitivity of CMB-S4 (Abazajian, 2016) (in green). The line labeled as “Quartic inflation” shows the prediction for a quartic potential (corresponding to the limit  $\xi_\sigma \rightarrow 0$ ), while we also show a black solid line corresponding to the limit  $\xi_\sigma \rightarrow \infty$ , in which the dynamics is analogous to that in the Starobinsky (Starobinsky, 1980) and Higgs inflation (HI) (Bezrukov and Shaposhnikov, 2008) models. The HI result of Gorbunov and Tokareva (2013) is indicated as a point on this line.

ends up dominating the energy density of the universe. This reheating process was studied in detail in SMASH (Ballesteros et al., 2017b), and arises from the coupled dynamics of the field  $\sigma$  containing the axion, the Higgs and the weak gauge bosons. As long as the relevant dynamics only involves Higgses and a complex singlet, and all the other scalar fields remain heavy and decoupled, we expect that some of the features of reheating in SMASH may apply for other variants as well. Differences may arise due to choosing different parameters or from the presence of additional fields with non-trivial dynamics. For example, stability requirements end up enforcing some kinematic blockings in SMASH which could be lifted in other scenarios. And within GUT models, the presence of multiple components within the GUT multiplets containing the axion or Higgses could have non-trivial consequences.

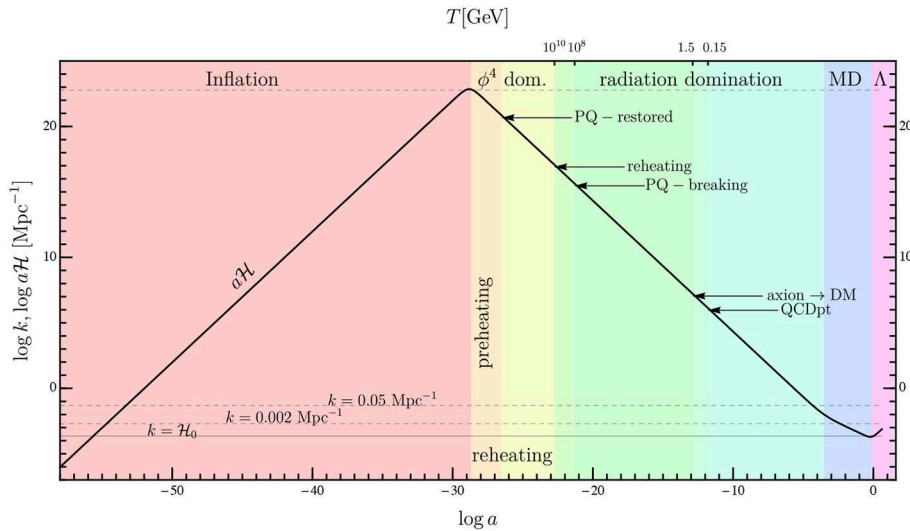
Within the SMASH model, slow-roll inflation ends for  $\rho \sim \mathcal{O}(M_P)$ , when the inflaton field starts undergoing Hubble-damped oscillations in a quartic potential (for such field values and for  $\xi_\sigma \lesssim 1$ , as required for predictive inflation, the effect of the non-minimal gravitational coupling can be ignored). These oscillations source a stress-energy tensor whose time-average mimics a radiation fluid. Hence, radiation domination starts right after inflation, and lasts through the phase of reheating in which the oscillating fields trigger the production of SM particles and the energy of the inflaton is transferred into the SM plasma. This post-inflationary history in a radiation-domination era (see **Figure 6** for a summary of the cosmological history of SMASH) fixes the relation between the scales of the matter perturbations we observe in the Universe today and the size of the primordial fluctuations which gave rise to them, when they outgrew the

Hubble horizon and became frozen until their later horizon re-entry. This relation between scales determines the number of e-folds between a perturbation’s horizon crossing and the end of inflation, which fixes the thick black lines in **Figures 4, 5** as the prediction for the parameters in SMASH.

In order to understand the process of particle production from the oscillating background field, one has to account for non-perturbative parametric resonance effects (Kofman et al., 1997; Tkachev et al., 1998). When the background field changes slowly in time—away from successive crossings of the origin—one can describe the fields through an adiabatic approximation in which particle number is well-defined, and conserved. However, during the crossings the adiabatic approximation breaks down and the appropriately matched adiabatic solutions separated by a crossing have different particle numbers. This particle production is dominated by bosonic fields, and can be understood as a resonance effect accounting for many-body bosonic interactions. The oscillating field may be thought of as a condensate of scalar particles with energy equal to the oscillating frequency, which for a quartic potential goes as

$$\omega = \sqrt{\lambda} \phi_0, \quad (37)$$

with  $\phi_0$  the oscillating amplitude. In SMASH, the relevant effective quartic for the inflationary background is determined by  $\lambda_\sigma$ —see Equation (31)—which is fixed to  $\lambda_\sigma \lesssim 10^{-10}$  by inflationary constraints. In turn, the inflaton condensate couples to Higgs particles with an effective mass dominated by background-dependent contributions, going as  $\sqrt{\lambda_{H\sigma}}|\phi|$ .



**FIGURE 6** | Expansion and thermal history of the Universe in SMASH HHSI [taken and used with permission from Ballesteros et al. (2017b)].

Stability constraints on the  $\delta$  parameter of Equation (35) typically imply  $\lambda_{H\sigma} \gg \lambda_\sigma$ , so that the background Higgs mass is on average much larger than the energy of the particles in the condensate, and Higgs production is blocked except during crossings ( $\phi = 0$ ). Due to this, non-perturbative particle production is dominated by the growth of perturbations of the field  $\sigma$  itself, for both the real and imaginary part. This effect, confirmed by lattice simulations (Ballesteros et al., 2017b), breaks the coherence of the oscillating background and leads to a non-perturbative restoration of the PQ symmetry, as the phase of  $\sigma$  ends up taking random values across the Universe. The loss of coherence of  $\phi$  ends up further blocking the production of Higgs particles, as  $|\phi|$  stops having an oscillatory behavior and the Higgs mass always remains above the frequency of the condensate.

In HSI, the Higgs is the only field that couples directly to the inflaton and the production of SM particles is quenched by this effect. The energy of the inflaton gets evenly distributed between the modulus and the phase of  $\sigma$ , and lattice simulations show that the axion excitations generated in this preheating phase are highly relativistic (Ballesteros et al., 2017b). Reheating into SM particles only becomes possible when the  $\sigma$  fluctuations redshift below the scale  $f_A$ , the PQ symmetry becomes broken and the  $\rho$  field acquires a mass that finally allows the decay into Higgses. This late decay results in a low reheating temperature of around  $T \sim 10^7$  GeV, while the initial production of relativistic axions results in an unacceptable amount of dark radiation at late times, predicting an increase in the effective number of relativistic degrees of freedom of  $\Delta N_{\nu}^{\text{eff}} = \mathcal{O}(1)$ , which is ruled out by the Planck constraint  $N_{\nu}^{\text{eff}} = 3.04 \pm 0.18$  at 68% CL (Ade et al., 2015).

In HHSI on the other hand the inflaton is an admixture of  $H$  and  $\sigma$ . This mixing endows the inflaton with a tree-level coupling to gauge bosons. Again, the gauge bosons in the Higgs background acquire oscillating masses  $m_W \sim gH \sim$

$g\sqrt{|\lambda_{H\sigma}|/(2\lambda_H)}\phi$  whose average is typically above the frequency of the condensate, but which become zero at the inflaton's crossings of the origin. Crucially, since as argued before the growth of Higgs perturbations is thwarted by the fast production of  $\sigma$  excitations, the Higgs component of the background does not lose coherence and continues to oscillate, which keeps the production of electroweak gauge bosons open during crossings. The decay rate of the gauge bosons is fast enough to essentially deplete their population between crossings, so that the boson production is never resonantly enhanced. Nevertheless, a thermal feedback mechanism takes place which enhances the rate of extraction of energy from the inflaton into the SM plasma. The decay products of the gauge bosons quickly reach a thermal bath, which may in turn produce gauge bosons by inverse decays near the crossings. Away from them, the extra bosons gain energy from the condensate as their mass grows with increasing  $|\phi|$ , and this energy is transferred into the SM plasma when the massive gauge bosons decay. Modeling this dynamics using Boltzmann equations and energy conservation constraints, one can predict a reheating temperature in HHSI near  $10^{10}$  GeV. This implies a thermal restoration of the PQ symmetry, as the critical temperature  $T_c$  for the PQ transition goes as

$$\frac{T_c}{v_\sigma} \simeq \frac{2\sqrt{6\lambda_\sigma}}{\sqrt{8(\lambda_\sigma + \lambda_{H\sigma}) + \sum_i Y_{ii}^2 + 6y^2}}, \quad (38)$$

and  $T_c$  is below  $10^{10}$  GeV for the preferred SMASH parameters. Moreover, the reheating temperature is also enough to guarantee that the axion population reaches thermal equilibrium, so that its abundance is no longer fixed by the earlier non-perturbative production. In this way the HSI problem with  $\Delta N_{\text{eff}}$  is avoided, and one predicts a modest amount of cosmic axion background radiation (CAB) corresponding to  $\Delta N_{\nu}^{\text{eff}} \simeq 0.03$ , which may be

probed with future CMB and large scale structure observations (Baumann et al., 2018).

Within GUT variants, the gutSMASH model with  $f_A$  independent of the unification scale could feature similar dynamics as SMASH in appropriate regions of parameter space. On the other hand, for the miniSO(10)PQ model the large  $f_A \gtrsim 2.6 \times 10^{15}$  GeV can give rise to important differences.<sup>4</sup> For example, if the reheating temperature is comparable to that in SMASH, the large value of  $f_A$  might mean that a thermal restoration of the PQ symmetry can be avoided, since the critical temperature is proportional to the VEV of the PQ field [see Equation (38)]. This can be a nice feature of the model, as for large  $f_A$  one should avoid a post-inflationary restoration of the PQ symmetry in order to avoid overclosure of the Universe by axion dark matter, as reviewed in the next section. However, this still leaves open the possibility of a non-thermal restoration of the PQ symmetry due to the preheating dynamics. Luckily, the large value of  $f_A$  can again come to the rescue. The large growth of perturbations in the inflaton field can be hampered for large  $f_A$  because the modulus of the field can become quickly trapped around the minimum before the fluctuations in the angular component grow large enough so as to restore the PQ symmetry. Once trapped in the minimum, the  $\rho$  fluctuations become massive and can decay quickly into SM particles, so that the growth of angular perturbations is expected to stop. With the results of the lattice simulations in SMASH (Ballesteros et al., 2017b), one can do a simple extrapolation to estimate the time at which the redshifting oscillations of the field reach a maximum of the order of a given value of  $f_A$ . If the time is below the onset of the parametric growth of the angular perturbations, one then expects that PQ restoration will be avoided. Such estimate gives that the PQ restoration might be avoided for  $f_A \gtrsim 4 \times 10^{16}$  GeV, which is in the allowed window of Equation (28) and raises the hope that the miniSO(10)PQ model could have a viable parameter space with a consistent cosmological history compatible with pre-inflationary axion dark matter.

## 6. DARK MATTER

The most important prediction of SMASH is that the PQ symmetry is broken after inflation. In the post-inflationary scenario, dark matter is produced by the re-alignment mechanism (Abbott and Sikivie, 1983; Dine and Fischler, 1983; Preskill et al., 1983) and the decay of topological defects (axion strings and domain walls) (Kawasaki et al., 2015). In models, in which the axion decay constant is an integer fraction of the PQ symmetry breaking scale,  $f_A = v_{PQ}/N_{DW}$ , with  $N_{DW} > 1$ , and in which the PQ symmetry is exact, there are  $N_{DW}$  degenerate CP-preserving vacua and domain walls develop between them when the axion field becomes non-relativistic; i.e., when at some temperature  $T_1$  the Hubble scale becomes of the order of the axion mass:  $\mathcal{H}(T_1) \sim m_A(T_1)$ . Since there is no preferred vacuum, the system of strings and walls is predicted to continue a scaling regime where the energy in domain-walls soon exceeds the observations. Therefore those models have to be

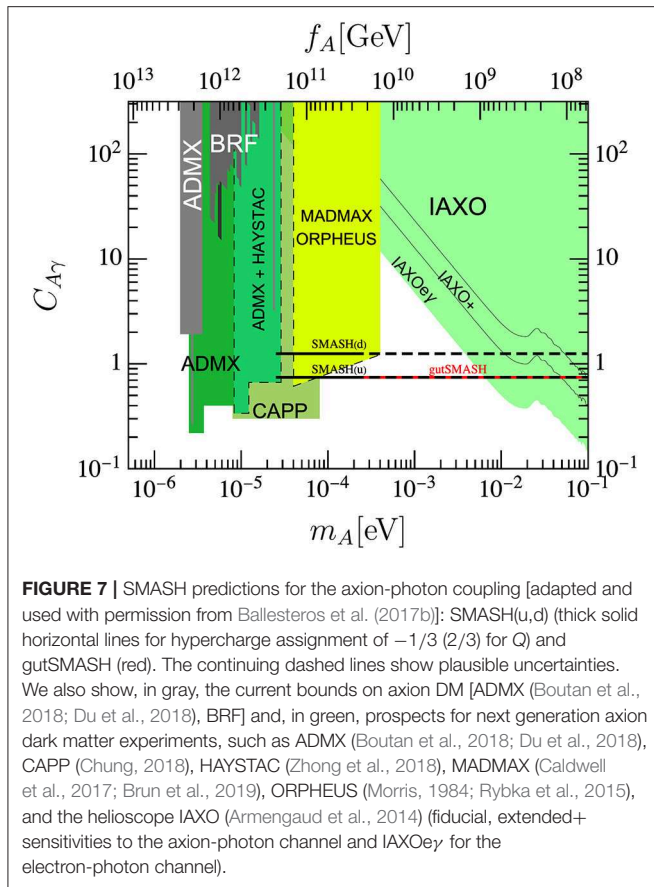
discarded (Sikivie, 1982) and  $N_{DW}$  can only be 1 in SMASH. This is the main motivation for introducing just one extra heavy quark in SMASH. The alternative models with larger values of  $N_{DW}$  [e.g., 2hdSMASH, with  $N_{DW} = 6$  (cf. **Table 2**), miniSU(5)PQ, with  $N_{DW} = 11$ , and miniSO(10)PQ and gutSMASH, with  $N_{DW} = 3$  (cf. **Table 4**)] can only become viable in scenarios in which the PQ symmetry is not exact—so that the degeneracy of the CP-preserving vacua can be lifted, and the domain-walls become unstable—or when the PQ symmetry is broken before or during inflation, never to be restored afterwards. In such a situation the energy density stored in the domain walls is simply diluted away by the exponential expansion of the universe during inflation.

Owing to the post-inflationary scenario, the original SMASH model becomes extremely predictive, at least in theory. In principle the axion DM abundance in this scenario is calculable by performing numerical simulations of the axion-string-wall network. The physics determining axion DM depends crucially on  $m_A$ . Uncertainties from the unknown initial conditions of the axion field are averaged away over many causal domains. Since there is no other cold DM candidate in the model, axions should provide all the observed CDM abundance and the theoretical relation  $\Omega_A h^2(m_A) = 0.12$  allows to obtain the required value of  $m_A$  (and thus  $f_A$ ). Unfortunately, there is a long-standing controversy regarding the calculation of  $\Omega_A h^2 = \Omega_A h^2(m_A)$ . Because of the large dynamical range required ( $f_A/\mathcal{H}(T_1) \sim 10^{30}$  from string cores to the horizon size) an extrapolation is mandatory and different authors have argued differently on how to perform it. Recently, a new method has been developed to endow the strings with the physically motivated effective tension,  $\propto \log f_A/\mathcal{H}$ , (if not the energy distribution around the string) and has lead to a very precise prediction,  $m_A \simeq (26.2 \pm 3.4) \mu\text{eV}$  (Klaer and Moore, 2017). The axion DM mass results so small because much of the network energy is radiated in hard axions (which count less for DM) and other hard quanta of the several extra fields that need being introduced. A recent detailed study of the string-network evolution (Gorghetto et al., 2018) has clarified substantially the results from standard numerical simulations and challenged the results of Kawasaki et al. (2015). The authors disregard the effective model of Klaer and Moore (2017) and highlight the huge uncertainty in the extrapolation to physical string-tensions.

When SMASH was proposed, the most detailed numerical simulations (Kawasaki et al., 2015) were pointing to  $m_A \sim 100 \mu\text{eV}$  and the uncertainties were revised to  $50 \mu\text{eV} \lesssim m_A \lesssim 200 \mu\text{eV}$  (Borsanyi et al., 2016; Ballesteros et al., 2017b). This corresponded to the range  $3 \times 10^{10} \text{ GeV} \lesssim f_A \lesssim 1.2 \times 10^{11} \text{ GeV}$ . According to the latest results, the lower limit on  $m_A$  could be a factor 2 smaller but the upper limit could be much greater. The next years might be decisive in resolving this controversy as new simulation techniques develop.

Most importantly, this axion dark matter mass window will be probed in the upcoming decade by axion dark matter direct detection experiments, such as ADMX (Boutan et al., 2018; Du et al., 2018), CAPP (Chung, 2018), HAYSTAC (Zhong et al., 2018), RADES (Melcón et al., 2018), MADMAX (Caldwell et al., 2017; Brun et al., 2019), ORPHEUS (Morris, 1984; Rybka

<sup>4</sup>Similar considerations apply for the miniSU(5)PQ model.



et al., 2015), and others (cf. **Figure 7**). A review on axion DM experiments can be found in Irastorza and Redondo (2018).

As anticipated earlier, non-minimal versions of SMASH where the degeneracy between  $N_{\text{DW}}$  vacua is broken are in principle possible and can be viable. Indeed, the degeneracy breaking generates a pressure between vacua that leads to the early collapse of the wall network (Sikivie, 1982). Ringwald and Saikawa (2016) studies how fundamental discrete symmetries can be invoked to protect the PQ symmetry from too large a breaking and estimates reasonable phenomenological parameters. This mechanism allows to avoid the domain wall problem for models like an extension of SMASH by further heavy quarks, 2hdSMASH and gutSMASH within a well-motivated framework. The price is however the non-minimality of the extra fields and the discrete symmetry. The best candidates tend to be  $Z_N$  symmetries with large  $N \sim 9, 10$  and point to axion masses in the meV mass ballpark. These predictions do not include the latest results about the string-network evolution that we mentioned above.

If the axion mass is around the meV ballpark, IAXO (Armengaud et al., 2014) could find the concomitant flux of solar axions but direct DM detection will be very difficult. The solar signal can be however used to pinpoint the axion mass and couplings (Dafni et al., 2019; Jaeckel and Thormaehlen, 2019), thus constraining the SMASH scenario and ease the search for DM.

The post-inflationary scenario typically favored in SMASH has many interesting phenomenological consequences. A large part of the DM is thought to be in the form of axion miniclusters (Kolb and Tkachev, 1993, 1994), small DM halos of typical radius  $\sim 10^{12}$  cm and mass  $\sim 10^{-12} M_\odot$  that form around matter-radiation equality with large densities  $\sim 10^7$  GeV/cm<sup>3</sup>. A recent study shows that smaller and denser objects are also unavoidable and more numerous (Vaquero et al., 2018). Axion miniclusters could be identified with pico-, femto- (Kolb and Tkachev, 1996; Zurek et al., 2007), and micro-lensing (Fairbairn et al., 2017, 2018, see also Katz et al., 2018). In many cases they will develop solitonic cores, sometimes called dilute axion stars (Visinelli et al., 2018) when considered in isolation. Most axion miniclusters survive until today and are so small that a direct encounter with the Earth is very rare. However, some others are tidally disrupted into streams whose encounters with the Earth can be more frequent and profitable for direct detection (Dokuchaev et al., 2017). The encounters of axion miniclusters/axion stars with the magnetic fields of compact objects has been speculated to be the origin of some fast-radio-bursts (Iwazaki, 2015; Tkachev, 2015).

In general, it is unfortunately impossible to predict whether SMASH variants will always realize the post-inflationary scenario. There is a strong tendency for this to be the case also in 2hdSMASH and gutSMASH if all the couplings are small and the inflaton is related to the PQ field. The addition of extra fields or non-minimal couplings could affect the isocurvature constraints from Planck and the reheating temperature. For the miniSU(5)PQ and miniSO(10)PQ models, as commented in section 5, the large values of  $f_A$  could in principle prevent the restoration of the PQ symmetry—as needed for the extremely light axion to remain compatible with dark matter—but dedicated studies are needed.

In the pre-inflationary scenario, the PQ symmetry would not be restored and the initial condition of the axion field would be an homogeneous local-Universe-wide value that could be anthropically selected for a very broad range of decay constants (Tegmark et al., 2006). For the axion to furnish all dark matter and  $f_A \lesssim 3 \times 10^{17}$  GeV, the initial mis-alignment angle  $\theta_I$  has to satisfy (Ballesteros et al., 2017b)

$$\theta_{I,c} \approx 0.0006 \times \left( \frac{f_A}{3 \times 10^{17} \text{ GeV}} \right)^{-0.504}. \quad (39)$$

We conclude this section discussing DM isocurvature bounds. If the PQ scalar is responsible for inflation one expects that the axion, its angular degree of freedom, gets its quantum fluctuations stretched to superhorizon length scales. Since axions constitute the DM, these fluctuations would get imprinted in the temperature anisotropies of the CMB as an isocurvature component, which is severely constrained by the data (Ade et al., 2015). The isocurvature bound gets translated into an upper bound on the Hubble expansion rate  $\mathcal{H}_I$  during inflation (and in turn on the tensor-to-scalar ratio,  $r$ ) as a function of  $f_A$ . Since there is an upper limit on  $r$  from the CMB (see section

3), this means a maximum possible value of  $f_A$ . Notice that this bound also depends on the initial axion mis-alignment angle, which together with  $f_A$  is the relevant parameter that determines the DM abundance in this scenario of PQ breaking during inflation. In scenarios in which the reheating temperature is such that the PQ symmetry becomes restored, the field values of the axion become processed by the thermal (or non-thermal) sub-horizon dynamics and all field perturbations end up being determined by a unique effective temperature scale and of the curvature type; thus, no isocurvature perturbations are generated.

In SMASH and its variants, the energy scale of inflation is mostly determined by the non-minimal coupling  $\xi$ , which imposes a lower bound on  $r$  (see **Figure 4**). The PQ symmetry is broken during inflation due to the time-dependent value of  $\rho$ —the modulus of the PQ scalar—which is not at the minimum of its potential, and thus the usual isocurvature bounds do not apply directly (see also Fairbairn et al., 2015). The reason can be understood by noticing that during inflation the effective  $f_A$  “seen” by the fluctuations in the direction orthogonal to the inflationary trajectory is actually the instantaneous value of the inflation field. Indeed, the “effective” value of  $f_A$  relevant to the isocurvature bounds is larger than the low-energy value of  $f_A$  (the one determined by the minimum of the PQ potential, entering into the axion mass equation) thanks to the non-minimal coupling and thus the ensuing constraints get weaker. A detailed calculation shows that the maximum allowed value of  $f_A$  is  $\sim 10^{14}$  GeV (Ballesteros et al., 2017b). This constraint, together with the fact that the PQ symmetry is always restored for  $f_A \lesssim 4 \times 10^{16}$  GeV, implies that the only viable SMASH realizations are those with PQ restoration after inflation, so that the DM abundance comes not only from oscillations of the axion field but also from the decay of topological defects, as discussed above.

The previous isocurvature bound in principle rules out the viability of miniSU(5)PQ or miniSO(10)PQ, with  $f_A$  tied to the unification scale. However, there is a possibility that the bound may be circumvented if one accounts for the fact that the axion field is not really massless during inflation, in contrast to what was assumed when deriving the isocurvature bound described above. During inflation the scalar fields do not sit at their minimum and Goldstone’s theorem does not apply; a detailed study of the evolution of the axion mass during and after inflation is needed. Moreover, in these models additional fields exist, which opens the possibility for additional paths in field space and further suppression of the bounds.

## 7. BARYOGENESIS

In SMASH models, the presence of right-handed neutrinos with masses proportional to the axion decay constant allows to explain the baryon asymmetry of the Universe through the mechanism of thermal leptogenesis (Fukugita and Yanagida, 1986). This relies on out-of-equilibrium, CP-violating decays of heavy RH neutrinos, which generate a net lepton asymmetry

which is partly converted into a baryon asymmetry by non-perturbative sphaleron processes that violate baryon plus lepton number. In SMASH-type models in which the PQ symmetry is restored thermally, such as the HHSI variant of SMASH, the RH neutrinos are massless after reheating, and are expected to acquire thermal equilibrium abundances. After the PQ phase transition they gain a mass, and as long the latter is smaller than the critical temperature of the transition, the massive RH neutrinos will typically re-enter equilibrium (Shuve and Tamarit, 2017) and decay at later times, generating the asymmetry after inverse decays become Boltzmann suppressed. This scenario is realized in SMASH, where demanding a stabilized potential in the  $\sigma$  direction, and assuming a hierarchy of Yukawas  $Y_{22} = Y_{33} = \kappa Y_{11}$  and  $y = Y_{11}$ , one has

$$\frac{T_c}{M_1} \gtrsim \frac{1}{\pi} \sqrt{\left(\frac{2+6\kappa^4}{7+2\kappa^2}\right) \log\left(\frac{30M_P}{\sqrt{2\lambda_\sigma} f_A}\right)}, \quad (40)$$

which follows from Equations (36) and (38) and is above 1 for typical SMASH parameters, including the case of near degenerate RH neutrinos with  $\kappa \approx 1$ .

In SMASH realizations in which the PQ symmetry is not restored thermally, as could be the case in models with very large  $f_A$ , such as GUT variants with  $f_A$  correlated with the unification scale, notably miniSU(5)PQ and miniSO(10)PQ<sup>5</sup>, the RH neutrinos are massive after reheating, but a thermal initial abundance can still be achieved for a reheating temperature above the RH masses. In this case the asymmetry will again be generated during late-time decays. A thermal initial abundance might not be achieved if the Yukawas of the RH neutrinos are very small, but in these so-called “weak washout” scenarios one can still produce an asymmetry from the out-of-equilibrium production and decays of RH neutrinos.

In the vanilla realizations of thermal leptogenesis with hierarchical RH neutrinos, the requirement of a large enough source of CP-violation in RH neutrino decays gives a lower bound  $M_1 \gtrsim 5 \times 10^8$  GeV (Casas and Ibarra, 2001; Giudice et al., 2004; Buchmuller et al., 2005). However, since the RH neutrino masses are proportional to their Yukawas with the field  $\sigma$ , and since these couplings tend to generate destabilizing corrections for the potential in the  $\sigma$  direction, having such heavy RH neutrinos can be in conflict with the requirement of stability. For example, in SMASH the stability bound in Equation (36) for a hierarchical  $N_i$  spectrum ( $M_3 = M_2 = 3M_1$ ) requires  $M_1 \lesssim 10^8 (\lambda/10^{-10})^{1/4} (v_\sigma/10^{11} \text{ GeV})$  GeV, which is just borderline compatible with the leptogenesis bound. Nevertheless, leptogenesis can occur for smaller masses with a mild resonant enhancement (Pilaftsis and Underwood, 2004) for a less hierarchical RH neutrino spectrum, which relaxes the stability bound and ensures that all the RH neutrinos remain in equilibrium after the phase transition. The estimated level of degeneracy needed in order to reconcile leptogenesis with the stability bound is of the order of 4%.

<sup>5</sup>Note that in order to avoid problems like monopole production, the reheating temperature in GUTs should be below the unification scale.

Finally, even though the RH neutrino masses are typically expected to be heavy, as they are proportional to the axion decay constant, fine-tuned values of the Yukawa couplings still allow for  $O(\text{GeV})$  masses. In such cases one recovers the  $\nu\text{MSM}$  at low energies, and even though lepton number violation is suppressed due to the small masses of the RH neutrinos, the baryon asymmetry can arise as a result of out-of-equilibrium oscillations of the right-handed neutrinos (Akhmedov et al., 1998). These give rise to flavored lepton asymmetries, which may even add up to zero initially, but as long as one flavor is out-of-equilibrium the washout will be incomplete and a net asymmetry will survive.

## 8. CONCLUSIONS

We have provided an overview of SMASHy extensions of the Standard Model which feature a new mass scale  $\nu_\sigma$ —of the order of  $10^{11}$  GeV in the simplest models, but which could also be tied to a Grand Unification scale around  $10^{16}$  GeV—and provide a falsifiable framework that addresses the following problems in particle physics and cosmology: inflation, baryogenesis, neutrino masses, dark matter and the strong CP problem. In addition, these models stabilize the electroweak vacuum. Whenever the dynamics of the most economical model (Ballesteros et al., 2017a,b), called SMASH in this review, is also realized in other extensions (as may happen if the additional fields remain decoupled during inflation and reheating), the models reviewed here predict a tensor-to-scalar-ratio  $r \gtrsim 0.004$ , a running of the spectral index  $\alpha \gtrsim -8 \times 10^{-4}$  (see **Figures 4, 5**), and a deviation in the effective number of relativistic neutrino species  $\Delta N_{\text{eff}} \sim 0.03$ , values which can be probed in future CMB experiments, such as CMB-S4, LiteBIRD, and the Simons Observatory. The SMASH model predicts a lower bound on the axion mass  $m_A \gtrsim 25 \mu\text{eV}$ , in the reach of future axion experiments, such as CAPP, MADMAX, ORPHEUS, and IAXO (see **Figure 7**). Given that the axion population in the model, constituting the totality of the DM, arises from the re-alignment mechanism and from the decay of topological defects (due to the post-inflationary breaking of the PQ symmetry), a large fraction of it may be in axion miniclusters, whose abundance may be tested via lensing studies of different astrophysical sources.

The models surveyed here revolve around the idea of exploiting the complex scalar field that implements the PQ symmetry and solves the strong CP problem. The axion—the angular part of this field—dynamically relaxes the theta parameter of QCD to a small maximum value, compatible with the upper bounds on the neutron electric dipole moment. On the other hand, the oscillations of the axion around the minimum of its potential constitute a condensate that can explain the nature of DM.

The modulus of the PQ scalar is instead the key ingredient for successful inflation. The inflationary sector of SMASH (which also contains a small Higgs component) predicts a primordial spectrum in agreement with the CMB, reheats the Universe efficiently and leads to a small relic abundance of thermal axions which may be identified through a determination of the effective number of relativistic species at early times. The coupling

between the Higgs doublet and the PQ scalar is instrumental for the stabilization of the effective potential at large field values, which in the SM is threatened by the large effect on the running of the Higgs quartic coupling coming from the top Yukawa. This interplay between inflation and stability set apart SMASHy extensions of the SM from models which utilize the Higgs alone to drive inflation (an idea that has more severe consistency issues related to the breakdown of perturbative unitarity).

The small masses of the light neutrinos are explained via the see-saw mechanism, adding three extra right-handed neutrinos whose heavy masses are induced by the VEV,  $\nu_\sigma$ , of the PQ scalar, which is proportional to the axion decay constant  $f_A$ . These heavy neutrinos can also explain the matter/anti-matter asymmetry of the Universe via thermal leptogenesis. The particle content of SMASH is illustrated in **Figure 2**. In addition to the PQ scalar and the three right handed neutrinos, the model features a heavy vector-like quark  $Q$  which is required for the KSVZ-like implementation of the PQ symmetry. At sufficiently low energy the model reduces to the SM augmented by small neutrino masses and the axion,  $A$ .

Possible extensions of the minimal SMASH model include implementations in Two-Higgs-Doublet models featuring a DFSZ axion, as well as embeddings of the latter into  $SU(5)$  and  $SO(10)$  GUTs. As long as one of the Higgses and the extra particles in the GUT multiplets are decoupled during inflation, one can expect to recover the inflationary predictions in SMASH. A similar post-inflationary history may be also recovered for an axion decay scale as in SMASH, i.e., near  $10^{11}$  GeV. However, for GUTs with the axion scale  $f_A$  tied to the unification scale, as in the miniSU(5)PQ and miniSO(10)PQ models, there can be important differences. First, isocurvature axion perturbations generated during inflation might be incompatible with Planck limits; although Ballesteros et al. (2017b) discarded  $f_A > 1.4 \times 10^{14}$  GeV on this account, the bound neglected the non-zero mass of the axion during inflation (arising from the fact that the scalar field is not at its minimum), and this needs to be accounted for. On the other hand, a large  $f_A$  is only compatible with axion dark matter in a scenario in which the PQ symmetry is not restored after inflation. Although dedicated lattice simulations are still lacking, there are indications that such behavior is possible, as very large values of  $f_A$  change the reheating dynamics and quench the generation of axion perturbations.

Given the lack of compelling new physics signals at the LHC, the idea of attempting to tackle several fundamental physics problems together in a simple (but coordinated) manner is appealing. Perhaps, one of the main take home messages from the SMASHy extensions of the SM that we have reviewed here is that the QCD axion might be a hint not only to dark matter, but also to inflation. In our opinion, it is interesting to continue exploring in the future possible connections between seemingly unrelated problems in particle physics and cosmology.

There exist other recent proposals which are also inspired by minimality and try to address simultaneously several of the SM standing issues. We will mention some of them briefly in the following. The model of Salvio (2015) has the same particle content as the one proposed in Dias et al. (2014) (and the same as in SMASH). It also attempts to address the same five problems

of the SM as SMASH, but it differs from it mainly regarding the heavy neutrino masses (which are not sourced by the VEV of the PQ scalar) and, also inflation, which in this case is driven by the Higgs and thus generically suffers from the unitarity issue. It has been recently argued in Salvio (2019) that the model can be safe from this problem if the top and Higgs masses are tuned in such a way that the quartic Higgs coupling relevant at the energies of inflation is very small. The proposal of Ema et al. (2017) aims to explain—in addition to the issues that SMASHy extensions of the SM deal with—the flavor structure of masses and mixings in the SM. The model differs from SMASH at several points. For example, the origin of the  $SU(3)$  anomaly of the PQ symmetry is unspecified. A key assumption in the model is a pole in the kinetic term of the new scalar field, which leads to an asymptotically flat potential after canonical normalization (see e.g., Galante et al., 2015). It has been argued that this kind of Lagrangian also suffers from an early breakdown of perturbative unitarity, and thus of consistency (Kehagias et al., 2014). The same idea of using a single  $U(1)$  symmetry for the flavor and the strong CP problems was independently proposed in Calibbi et al. (2017), although this paper does not deal with inflation nor with the matter/anti-matter asymmetry. A very different kind of proposal has been recently put forward in Gupta et al. (2019). This model aims to solve the same problems as SMASH, except inflation, and in addition it tackles the hierarchy problem. It does so by means of the relaxion mechanism (Graham et al., 2015) (for the hierarchy problem) and the Barr-Nelson mechanism (Barr, 1984; Nelson, 1984) (for the strong CP problem). Baryogenesis is triggered in this case by oscillations of the relaxion field around its final minimum.

## REFERENCES

- Aaboud, M., Aad, G., Abbott, B., Abidinov, O., Abeloos, B., Abhayasinghe, D. K., et al. (2018). Measurement of the top quark mass in the  $t\bar{t} \rightarrow \text{lepton} + \text{jets}$  channel from  $\sqrt{s} = 8$  TeV ATLAS data and combination with previous results. *Eur. Phys. J.* 79:290. doi: 10.1140/epjc/s10052-019-6757-9
- Abazajian, K. N. (2016). *CMB-S4 Science Book, 1st Edn*. doi: 10.2172/1352047
- Abbott, L. F., and Sikivie, P. (1983). A cosmological bound on the invisible axion. *Phys. Lett. B* 120, 133–136.
- Abe, K., Abe, T., Aihara, H., Fukuda, Y., Hayato, Y., Huang, K., et al. (2011). Letter of intent: the hyper-kamiokande experiment—detector design and physics potential—*arXiv:1109.3262*.
- Abe, K., Haga, Y., Hayato, Y., Ikeda, M., Iyogi, K., Kameda, J., et al. (2017). Search for proton decay via  $p \rightarrow e^+ \pi^0$  and  $p \rightarrow \mu^+ \pi^0$  in 0.31 megaton·years exposure of the Super-Kamiokande water Cherenkov detector. *Phys. Rev. D* 95:012004. doi: 10.1103/PhysRevD.95.012004
- Ade, P. A. R., Aghanim, N., Arnaud, M., Arroja, F., Ashdown, M., Aumont, J., et al. (2016a). Planck 2015 results. XX. Constraints on inflation. *Astron. Astrophys.* 594:A20. doi: 10.1051/0004-6361/201525898
- Ade, P. A. R., Aghanim, N., Arnaud, M., Ashdown, M., Aumont, J., Baccigalupi, C., et al. (2015). Planck 2015 results. XIII. Cosmological parameters. *Astron. Astrophys.* 594:A13. doi: 10.1051/0004-6361/201525830
- Ade, P. A. R., Ahmed, Z., Aikin, R. W., Alexander, K. D., Barkats, D., Benton, S. J., et al. (2016b). Improved constraints on cosmology and foregrounds from BICEP2 and keck array cosmic microwave background data with inclusion of 95 GHz band. *Phys. Rev. Lett.* 116:031302. doi: 10.1103/PhysRevLett.116.031302
- Aguirre, J., Aguirre, J., Ahmed, Z., Aiola, S., Ali, A., Alonso, D., et al. (2019). The Simons Observatory: science goals and forecasts. *JCAP* 1902:056. doi: 10.1088/1475-7516/2019/02/056
- In summary, we are living in interesting times for particle physics and cosmology, in which simple ideas blended together are providing new theoretical insights and unveiling possible connections between different problems.
- ## AUTHOR CONTRIBUTIONS
- AR determined the table of content and wrote mainly sections 1, 2. GB wrote section 3. JR concentrated on section 6. CT on sections 4, 5, 7. All authors contributed equally on section 8.
- ## ACKNOWLEDGMENTS
- Many thanks to Anne Ernst and Luca Di Luzio for the great collaboration on SMASHy extensions of the SM. The work of GB is funded by a Contrato de Atracción de Talento (Modalidad 1) de la Comunidad de Madrid (Spain), with number 2017-T1/TIC-5520. It has also been supported by MINECO (Spain) under contract FPA2016-78022-P and the Spanish MINECO's Centro de Excelencia Severo Ochoa Program under the grants SEV-2012-0249 and SEV-2016-0597. GB and CT acknowledge support from the Collaborative Research Centres SFB676 and SFB1258 of the Deutsche Forschungsgemeinschaft (DFG), respectively. The work of AR is partly supported by the DFG under Germany's Excellence Strategy—EXC 2121 Quantum Universe—390833306. JR is supported by the Ramon y Cajal Fellowship 2012-10597, the grant FPA2015-65745-P (MINECO/FEDER), the EU through the ITN Elusives H2020-MSCA-ITN-2015/674896 and the Deutsche Forschungsgemeinschaft under grant SFB1258 as a Mercator Fellow.
- Akhmedov, E. K., Rubakov, V. A., and Smirnov, A. Y. (1998). Baryogenesis via neutrino oscillations. *Phys. Rev. Lett.* 81, 1359–1362. doi: 10.1103/PhysRevLett.81.1359
- Akrami, Y., Arroja, F., Ashdown, M., Aumont, J., Baccigalupi, C., Ballardini, M., et al. (2018). Planck 2018 results. X. Constraints on inflation. *arXiv:1807.06211*.
- Altarelli, G., and Meloni, D. (2013). A non supersymmetric  $SO(10)$  grand unified model for all the physics below  $M_{GUT}$ . *JHEP* 1308:021. doi: 10.1007/JHEP08(2013)021
- Arhrib, A., Bajc, B., Ghosh, D. K., Han, T., Huang, G. Y., Puljak, I., et al. (2010). Collider signatures for heavy lepton triplet in Type I + III seesaw. *Phys. Rev. D* 82:053004. doi: 10.1103/PhysRevD.82.053004
- Armengaud, E., Avignone, F. T., Betz, M., Brax, P., Brun, P., Cantatore, G., et al. (2014). Conceptual design of the International Axion Observatory (IAXO). *JINST* 9:T05002. doi: 10.1088/1748-0221/9/05/T05002
- Arvanitaki, A., Baryakhtar, M., and Huang, X. (2015). Discovering the QCD axion with black holes and gravitational waves. *Phys. Rev. D* 91:084011. doi: 10.1103/PhysRevD.91.084011
- Asaka, T., Blanchet, S., and Shaposhnikov, M. (2005). The nuMSM, dark matter and neutrino masses. *Phys. Lett. B* 631, 151–156. doi: 10.1016/j.physletb.2005.09.070
- Asaka, T., and Shaposhnikov, M. (2005). The nuMSM, dark matter and baryon asymmetry of the universe. *Phys. Lett. B* 620, 17–26. doi: 10.1016/j.physletb.2005.06.020
- Babu, K. S., and Khan, S. (2015). Minimal nonsupersymmetric  $SO(10)$  model: gauge coupling unification, proton decay, and fermion masses. *Phys. Rev. D* 92:075018. doi: 10.1103/PhysRevD.92.075018
- Babu, K. S., and Mohapatra, R. N. (1993). Predictive neutrino spectrum in minimal  $SO(10)$  grand unification. *Phys. Rev. Lett.* 70, 2845–2848. doi: 10.1103/PhysRevLett.70.2845

- Bajc, B., Melfo, A., Senjanovic, G., and Vissani, F. (2006). Yukawa sector in non-supersymmetric renormalizable  $SO(10)$ . *Phys. Rev. D* 73:055001. doi: 10.1103/PhysRevD.73.055001
- Bajc, B., Nemevsek, M., and Senjanovic, G. (2007). Probing seesaw at LHC. *Phys. Rev. D* 76:055011. doi: 10.1103/PhysRevD.76.055011
- Bajc, B., and Senjanovic, G. (2007). Seesaw at LHC. *JHEP* 0708:014. doi: 10.1088/1126-6708/2007/08/014
- Ballesteros, G., Redondo, J., Ringwald, A., and Tamarit, C. (2017a). Unifying inflation with the axion, dark matter, baryogenesis and the seesaw mechanism. *Phys. Rev. Lett.* 118:071802. doi: 10.1103/PhysRevLett.118.071802
- Ballesteros, G., Redondo, J., Ringwald, A., and Tamarit, C. (2017b). Standard model–axion–seesaw–Higgs portal inflation. Five problems of particle physics and cosmology solved in one stroke. *JCAP* 1708:001. doi: 10.1088/1475-7516/2017/08/001
- Barbon, J. L. F., and Espinosa, J. R. (2009). On the naturalness of Higgs inflation. *Phys. Rev. D* 79:081302. doi: 10.1103/PhysRevD.79.081302
- Barr, S. M. (1984). Solving the strong CP problem without the Peccei–Quinn symmetry. *Phys. Rev. Lett.* 53:329. doi: 10.1103/PhysRevLett.53.329
- Baumann, D., Green, D., and Wallisch, B. (2018). Searching for light relics with large-scale structure. *JCAP* 1808:029. doi: 10.1088/1475-7516/2018/08/029
- Bednyakov, A. V., Kniehl, B. A., Pikelner, A. F., and Veretin, O. L. (2015). Stability of the electroweak vacuum: gauge independence and advanced precision. *Phys. Rev. Lett.* 115:201802. doi: 10.1103/PhysRevLett.115.201802
- Ben-Dayan, I., and Einhorn, M. B. (2010). Supergravity Higgs inflation and shift symmetry in electroweak theory. *JCAP* 1012:002. doi: 10.1088/1475-7516/2010/12/002
- Bereziani, Z. G., Sakharov, A. S., and Khlopov, M. Y. (1992). Primordial background of cosmological axions. *Sov. J. Nucl. Phys.* 55, 1063–1071.
- Bezrukov, F., and Shaposhnikov, M. (2014). Higgs inflation at the critical point. *Phys. Lett. B* 734, 249–254. doi: 10.1016/j.physletb.2014.05.074
- Bezrukov, F. L., and Shaposhnikov, M. (2008). The standard model Higgs boson as the inflaton. *Phys. Lett. B* 659, 703–706. doi: 10.1016/j.physletb.2007.11.072
- Borsanyi, S., Fodor, Z., Guenther, J., Kampert, K.-H., Katz, S. D., Kawanai, T., et al. (2016). Calculation of the axion mass based on high-temperature lattice quantum chromodynamics. *Nature* 539, 69–71. doi: 10.1038/nature20115
- Boucenna, S. M., Ohlsson, T., and Pernow, M. (2019). A minimal non-supersymmetric  $SO(10)$  model with Peccei–Quinn symmetry. *Phys. Lett. B* 792, 251–257. doi: 10.1016/j.physletb.2019.03.045
- Boucenna, S. M., and Shafi, Q. (2018). Axion inflation, proton decay and leptogenesis in  $SU(5) \times U(1)_{PQ}$ . *Phys. Rev. D* 97:075012. doi: 10.1103/PhysRevD.97.075012
- Boutan, C., Jones, M., LaRoque, B. H., Oblath, N. S., Cervantes, R., Du, N., et al. (2018). Piezoelectrically tuned multimode cavity search for axion dark matter. *Phys. Rev. Lett.* 121:261302. doi: 10.1103/PhysRevLett.121.261302
- Brun, P., Caldwell, A., Chevalier, L., Dvali, G., Freire, P., Garutti, E., et al. (2019). A new experimental approach to probe QCD axion dark matter in the mass range above  $40 \mu\text{eV}$ . *Eur. Phys. J. C* 79:186. doi: 10.1140/epjc/s10052-019-6683-x
- Buchmüller, W., Di Bari, P., and Plumacher, M. (2005). Leptogenesis for pedestrians. *Ann. Phys.* 315, 305–351. doi: 10.1016/j.aop.2004.02.003
- Budker, D., Graham, P. W., Ledbetter, M., Rajendran, S., and Sushkov, A. (2014). Proposal for a cosmic axion spin precession experiment (CASPER). *Phys. Rev. X* 4:021030. doi: 10.1103/PhysRevX.4.021030
- Burgess, C. P., Lee, H. M., and Trost, M. (2009). Power-counting and the validity of the classical approximation during inflation. *JHEP* 0909:103. doi: 10.1088/1126-6708/2009/09/103
- Buttazzo, D., Degrandi, G., Giardino, P. P., Giudice, G. F., Sala, F., Salvio, A., et al. (2013). Investigating the near-criticality of the Higgs boson. *JHEP* 1312:089. doi: 10.1007/JHEP12(2013)089
- Cai, Y., Han, T., Li, T., and Ruiz, R. (2018). Lepton number violation: seesaw models and their collider tests. *Front. Phys.* 6:40. doi: 10.3389/fphys.2018.00040
- Caldwell, A., Dvali, G., Majorovits, B., Millar, A., Raffelt, G., Redondo, J., et al. (2017). Dielectric haloscopes: a new way to detect axion dark matter. *Phys. Rev. Lett.* 118:091801. doi: 10.1103/PhysRevLett.118.091801
- Calibbi, L., Goertz, F., Redigolo, D., Ziegler, R., and Zupan, J. (2017). Minimal axion model from flavor. *Phys. Rev. D* 95:095009. doi: 10.1103/PhysRevD.95.095009
- Casas, J. A., and Ibarra, A. (2001). Oscillating neutrinos and  $\mu \rightarrow e, \gamma$ . *Nucl. Phys. B* 618, 171–204. doi: 10.1016/S0550-3213(01)00475-8
- Chang, J. H., Essig, R., and McDermott, S. D. (2018). Supernova 1987A constraints on sub-GeV dark sectors, millicharged particles, the QCD axion, and an axion-like particle. *J. High Energy Phys.* 2018:51. doi: 10.1007/JHEP09(2018)051
- Chikashige, Y., Mohapatra, R. N., and Peccei, R. D. (1981). Are there real goldstone bosons associated with broken Lepton number? *Phys. Lett. B* 98, 265–268.
- Choudhury, S., Chakraborty, T., and Pal, S. (2014). Higgs inflation from new Kähler potential. *Nucl. Phys. B* 880, 155–174. doi: 10.1016/j.nuclphysb.2014.01.002
- Chung, W. (2018). CULTASK, axion experiment at CAPP in Korea. doi: 10.3204/DESY-PROC-2017-02/woohyun\_chung
- Clarke, J. D., and Volkas, R. R. (2016). Technically natural nonsupersymmetric model of neutrino masses, baryogenesis, the strong CP problem, and dark matter. *Phys. Rev. D* 93:035001. doi: 10.1103/PhysRevD.93.035001
- Dafni, T., O'Hare, C. A. J., Lalic, B., Galán, J., Iguaz, F. J., Irastorza, I. G., et al. (2019). Weighing the solar axion. *Phys. Rev. D* 99:035037. doi: 10.1103/PhysRevD.99.035037
- Di Luzio, L., and Mihaila, L. (2013). Unification scale vs. electroweak-triplet mass in the  $SU(5) + 24_F$  model at three loops. *Phys. Rev. D* 87:115025. doi: 10.1103/PhysRevD.87.115025
- Di Luzio, L., Ringwald, A., and Tamarit, C. (2018). Axion mass prediction from minimal grand unification. *Phys. Rev. D* 98:095011. doi: 10.1103/PhysRevD.98.095011
- Di Vecchia, P., and Veneziano, G. (1980). Chiral dynamics in the large  $n$  limit. *Nucl. Phys. B* 171, 253–272. doi: 10.1016/0550-3213(80)90370-3
- Dias, A. G., Machado, A. C. B., Nishi, C. C., Ringwald, A., and Vaudrevange, P. (2014). The quest for an intermediate-scale accidental axion and further ALPs. *JHEP* 1406:037. doi: 10.1007/JHEP06(2014)037
- Dine, M., and Fischler, W. (1983). The not so harmless axion. *Phys. Lett. B* 120, 137–141.
- Dine, M., Fischler, W., and Srednicki, M. (1981). A simple solution to the strong CP problem with a harmless axion. *Phys. Lett. B* 104, 199–202. doi: 10.1016/0370-2693(81)90590-6
- Dokuchaev, V. I., Eroshenko, Y. N., and Tkachev, I. I. (2017). Destruction of axion miniclusters in the galaxy. *J. Exp. Theor. Phys.* 125:434. doi: 10.1134/S1063776117080039
- Dorsner, I., and Fileviez Perez, P. (2005). How long could we live? *Phys. Lett. B* 625:88. doi: 10.1016/j.physletb.2005.08.039
- Du, N., Force, N., Khatawada, R., Lentz, E., Ottens, R., Rosenberg, L. J., et al. (2018). A search for invisible axion dark matter with the axion dark matter experiment. *Phys. Rev. Lett.* 120:151301. doi: 10.1103/PhysRevLett.120.151301
- Elias-Miro, J., Espinosa, J. R., Giudice, G. F., Lee, H. M., and Strumia, A. (2012). Stabilization of the electroweak vacuum by a scalar threshold effect. *JHEP* 1206:031. doi: 10.1007/JHEP06(2012)031
- Ema, Y., Hamaguchi, K., Moroi, T., and Nakayama, K. (2017). Flaxion: a minimal extension to solve puzzles in the standard model. *JHEP* 1701:096. doi: 10.1007/JHEP01(2017)096
- Ernst, A., Ringwald, A., and Tamarit, C. (2018). Axion predictions in  $SO(10) \times U(1)_{PQ}$  models. *JHEP* 1802:103. doi: 10.1007/JHEP02(2018)103
- Fairbairn, M., Hogan, R., and Marsh, D. J. E. (2015). Unifying inflation and dark matter with the Peccei–Quinn field: observable axions and observable tensors. *Phys. Rev. D* 91:023509. doi: 10.1103/PhysRevD.91.023509
- Fairbairn, M., Marsh, D. J. E., and Quevillon, J. (2017). Searching for the QCD axion with gravitational microlensing. *Phys. Rev. Lett.* 119:021101. doi: 10.1103/PhysRevLett.119.021101
- Fairbairn, M., Marsh, D. J. E., Quevillon, J., and Rozier, S. (2018). Structure formation and microlensing with axion miniclusters. *Phys. Rev. D* 97:083502. doi: 10.1103/PhysRevD.97.083502
- Fischer, T., Chakraborty, S., Giannotti, M., Mirizzi, A., Payez, A., and Ringwald, A. (2016). Probing axions with the neutrino signal from the next galactic supernova. *Phys. Rev. D* 94:085012. doi: 10.1103/PhysRevD.94.085012
- Fritzsch, H., and Minkowski, P. (1975). Unified interactions of leptons and hadrons. *Ann. Phys.* 93, 193–266. doi: 10.1016/0003-4916(75)90211-0
- Fukugita, M., and Yanagida, T. (1986). Baryogenesis without grand unification. *Phys. Lett. B* 174, 45–47.
- Galante, M., Kallosh, R., Linde, A., and Roest, D. (2015). Unity of cosmological inflation attractors. *Phys. Rev. Lett.* 114:141302. doi: 10.1103/PhysRevLett.114.141302

- Garcia-Cely, C., and Heeck, J. (2017). Neutrino lines from Majoron dark matter. *JHEP* 1705:102. doi: 10.1007/JHEP05(2017)102
- Gell-Mann, M., Ramond, P., and Slansky, R. (1979). Complex spinors and unified theories. *Conf. Proc. C 790927*, 315–321.
- Gelmini, G. B., and Roncadelli, M. (1981). Left-handed neutrino mass scale and spontaneously broken Lepton number. *Phys. Lett. B* 99, 411–415.
- Georgi, H. (1975). The state of the art—gauge theories. *AIP Conf. Proc.* 23, 575–582. doi: 10.1063/1.2947450
- Georgi, H., and Glashow, S. L. (1974). Unity of all elementary particle forces. *Phys. Rev. Lett.* 32, 438–441. doi: 10.1103/PhysRevLett.32.438
- Giannotti, M., Irastorza, I. G., Redondo, J., Ringwald, A., and Saikawa, K. (2017). Stellar recipes for axion hunters. *JCAP* 1710:010. doi: 10.1088/1475-7516/2017/10/010
- Giudice, G. F., Notari, A., Raidal, M., Riotto, A., and Strumia, A. (2004). Towards a complete theory of thermal Leptogenesis in the SM and MSSM. *Nucl. Phys. B* 685, 89–149. doi: 10.1016/j.nuclphysb.2004.02.019
- Gong, J. O., Lee, H. M., and Kang, S. K. (2012). Inflation and dark matter in two Higgs doublet models. *JHEP* 1204:128. doi: 10.1007/JHEP04(2012)128
- Gorbunov, D., and Tokareva, A. (2013).  $R^2$ -inflation with conformal SM Higgs field. *JCAP* 1312:021. doi: 10.1088/1475-7516/2013/12/021
- Gorghetto, M., Hardy, E., and Villadoro, G. (2018). Axions from strings: the attractive solution. *JHEP* 1807:151. doi: 10.1007/JHEP07(2018)151
- Gorghetto, M., and Villadoro, G. (2019). Topological susceptibility and QCD axion mass: QED and NNLO corrections. *JHEP* 1903:033. doi: 10.1007/JHEP03(2019)033
- Graham, P. W., Kaplan, D. E., and Rajendran, S. (2015). Cosmological relaxation of the electroweak scale. *Phys. Rev. Lett.* 115:221801. doi: 10.1103/PhysRevLett.115.221801
- Grilli di Cortona, G., Hardy, E., Pardo Vega, J., and Villadoro, G. (2016). The QCD axion, precisely. *JHEP* 1601:034. doi: 10.1007/JHEP01(2016)034
- Gupta, R. S., Reiness, J. Y., and Spannowsky, M. (2019). All-in-one Relaxion, a unified solution to five BSM puzzles. *arXiv:1902.08633*.
- Hamada, Y., Kawai, H., Oda, K. Y., and Park, S. C. (2014). Higgs Inflation is Still Alive after the Results from BICEP2. *Phys. Rev. Lett.* 112:241301. doi: 10.1103/PhysRevLett.112.241301
- Holman, R., Lazarides, G., and Shafi, Q. (1983). Axions and the dark matter of the universe. *Phys. Rev. D* 27:995. doi: 10.1103/PhysRevD.27.995
- Irastorza, I. G., and Redondo, J. (2018). New experimental approaches in the search for axion-like particles. *Prog. Part. Nucl. Phys.* 102, 89–159. doi: 10.1016/j.pnpnp.2018.05.003
- Iwazaki, A. (2015). Axion stars and fast radio bursts. *Phys. Rev. D* 91:023008. doi: 10.1103/PhysRevD.91.023008
- Jackson Kimball, D. F., Afach, S., Aybas, D., Blanchard, J. W., Budker, D., Centers, G., et al. (2017). Overview of the cosmic axion spin precession experiment (CASPER). *arXiv:1711.08999*.
- Jaeckel, J., and Thormaehlen, L. J. (2019). Distinguishing axion models with IAXO. *J. Cosmol. Astropart. Phys.* 2019:39. doi: 10.1088/1475-7516/2019/03/039
- Kahn, Y., Safdi, B. R., and Thaler, J. (2016). Broadband and resonant approaches to axion dark matter detection. *Phys. Rev. Lett.* 117:141801. doi: 10.1103/PhysRevLett.117.141801
- Katz, A., Kopp, J., Sibiryakov, S., and Xue, W. (2018). Femtolensing by dark matter revisited. *JCAP* 1812:005. doi: 10.1088/1475-7516/2018/12/005
- Kawasaki, M., Saikawa, K., and Sekiguchi, T. (2015). Axion dark matter from topological defects. *Phys. Rev. D* 91:065014. doi: 10.1103/PhysRevD.91.065014
- Kehagias, A., Moradinezhad Dizgah, A., and Riotto, A. (2014). Remarks on the Starobinsky model of inflation and its descendants. *Phys. Rev. D* 89:043527. doi: 10.1103/PhysRevD.89.043527
- Kim, J. E. (1979). Weak interaction singlet and strong CP invariance. *Phys. Rev. Lett.* 43:103.
- Klaer, V. B., and Moore, G. D. (2017). The dark-matter axion mass. *JCAP* 1711:049. doi: 10.1088/1475-7516/2017/11/049
- Klaer, V. B., and Moore, G. D. (2017). How to simulate global cosmic strings with large string tension. *JCAP* 1710:043. doi: 10.1088/1475-7516/2017/10/043
- Kofman, L., Linde, A. D., and Starobinsky, A. A. (1997). Towards the theory of reheating after inflation. *Phys. Rev. D* 56, 3258–3295. doi: 10.1103/PhysRevD.56.3258
- Kolb, E. W., and Tkachev, I. I. (1993). Axion miniclusters and Bose stars. *Phys. Rev. Lett.* 71, 3051–3054. doi: 10.1103/PhysRevLett.71.3051
- Kolb, E. W., and Tkachev, I. I. (1994). Large amplitude isothermal fluctuations and high density dark matter clumps. *Phys. Rev. D* 50, 769–773. doi: 10.1103/PhysRevD.50.769
- Kolb, E. W., and Tkachev, I. I. (1996). Femtolensing and picolensing by axion miniclusters. *Astrophys. J.* 460:L25. doi: 10.1086/309962
- Langacker, P., Peccei, R. D., and Yanagida, T. (1986). Invisible axions and light neutrinos: are they connected? *Mod. Phys. Lett. A* 1:541. doi: 10.1142/S0217732386000683
- Lebedev, O. (2012). On stability of the electroweak vacuum and the Higgs portal. *Eur. Phys. J. C* 72:2058. doi: 10.1140/epjc/s10052-012-2058-2
- Leutwyler, H., and Smilga, A. V. (1992). Spectrum of Dirac operator and role of winding number in QCD. *Phys. Rev. D* 46, 5607–5632. doi: 10.1103/PhysRevD.46.5607
- Matsumura, T., Akiba, Y., Borrill, J., Chinone, Y., Dobbs, M., Fuke, H., et al. (2014). Mission design of LiteBIRD. *J. Low. Temp. Phys.* 176, 733–740. doi: 10.1007/s10909-013-0996-1
- Melcón, A. Á., Cuendisb, S. A., Cogollo, C., Díaz-Morcillo, A., Döbrich, B., Gallego, J. D., et al. (2018). Axion searches with microwave filters: the RADES project. *JCAP* 1805:040. doi: 10.1088/1475-7516/2018/05/040
- Mikheyev, S. P., and Smirnov, A. Y. (1985). Resonance amplification of oscillations in matter and spectroscopy of solar neutrinos. *Sov. J. Nucl. Phys.* 42, 913–917.
- Minkowski, P. (1977).  $\mu \rightarrow e\gamma$  at a rate of one out of  $10^9$  muon decays? *Phys. Lett. B* 67, 421–428.
- Mohapatra, R. N., and Senjanovic, G. (1980). Neutrino mass and spontaneous parity violation. *Phys. Rev. Lett.* 44:912.
- Mohapatra, R. N., and Senjanovic, G. (1983). The superlight axion and neutrino masses. *Z. Phys. C* 17, 53–56. doi: 10.1007/BF01577819
- Morris, D. E. (1984). *An Electromagnetic Detector For Relic Axions*. doi: 10.2172/6446664
- Nardi, E., and Roulet, E. (1990). Are exotic stable quarks cosmologically allowed? *Phys. Lett. B* 245, 105–110. doi: 10.1016/0370-2693(90)90172-3
- Nelson, A. E. (1984). Naturally weak CP violation. *Phys. Lett. B* 136, 387–391. doi: 10.1016/0370-2693(84)92025-2
- Pallis, C. (2017). Linking Starobinsky-type inflation in no-scale supergravity to MSSM. *JCAP* 1404:024. doi: 10.1088/1475-7516/2014/04/024
- Pallis, C. (2018a). Gravitational waves,  $\mu$  term and leptogenesis from  $B - L$  Higgs inflation in supergravity. *Universe* 4:13. doi: 10.3390/universe4010013
- Pallis, C. (2018b).  $B - L$  Higgs inflation in supergravity with several consequences. *PoS CORFU 2017:086*. doi: 10.22323/1.318.0086
- Pallis, C., and Shafi, Q. (2018). Induced-gravity GUT-scale Higgs inflation in supergravity. *Eur. Phys. J. C* 78:523. doi: 10.1140/epjc/s10052-018-5980-0
- Peccei, R. D., and Quinn, H. R. (1977). CP conservation in the presence of instantons. *Phys. Rev. Lett.* 38, 1440–1443.
- Perez, K., Ng, K. C. Y., Beacom, J. F., Hersch, C., Horiuchi, S., and Krivonos, R. (2017). Almost closing the  $\nu$ MSM sterile neutrino dark matter window with NuSTAR. *Phys. Rev. D* 95:123002. doi: 10.1103/PhysRevD.95.123002
- Pi, S. Y. (1984). Inflation without tears. *Phys. Rev. Lett.* 52, 1725–1728. doi: 10.1103/PhysRevLett.52.1725
- Pilaftsis, A. (1994). Astrophysical and terrestrial constraints on singlet Majoron models. *Phys. Rev. D* 49, 2398–2404. doi: 10.1103/PhysRevD.49.2398
- Pilaftsis, A., and Underwood, T. E. J. (2004). Resonant leptogenesis. *Nucl. Phys. B* 692, 303–345. doi: 10.1016/j.nuclphysb.2004.05.029
- Preskill, J., Wise, M. B., and Wilczek, F. (1983). Cosmology of the invisible axion. *Phys. Lett. B* 120, 127–132.
- Raffelt, G. G. (2008). Astrophysical axion bounds. *Lect. Notes Phys.* 741, 51–71. doi: 10.1007/978-3-540-73518-2\_3
- Ringwald, A., and Saikawa, K. (2016). Axion dark matter in the post-inflationary Peccei-Quinn symmetry breaking scenario. *Phys. Rev. D* 93:085031. doi: 10.1103/PhysRevD.93.085031
- Ruiz, R. (2015). QCD corrections to pair production of type III seesaw leptons at Hadron colliders. *JHEP* 1512:165. doi: 10.1007/JHEP12(2015)165
- Rybka, G., Wagner, A., Brill, A., Ramos, K., Percival, R., and Patel, K. (2015). Search for dark matter axions with the Orpheus experiment. *Phys. Rev. D* 91:011701. doi: 10.1103/PhysRevD.91.011701

- Salopek, D. S., Bond, J. R., and Bardeen, J. M. (1989). Designing density fluctuation spectra in inflation. *Phys. Rev. D* 40, 1753–1788. doi: 10.1103/PhysRevD.40.1753
- Salvio, A. (2015). A simple motivated completion of the standard model below the Planck scale: axions and right-handed neutrinos. *Phys. Lett. B* 743, 428–434. doi: 10.1016/j.physletb.2015.03.015
- Salvio, A. (2019). Critical Higgs inflation in a viable motivated model. *Phys. Rev. D* 99:015037. doi: 10.1103/PhysRevD.99.015037
- Schechter, J., and Valle, J. W. F. (1982). Neutrino decay and spontaneous violation of Lepton number. *Phys. Rev. D* 25:774.
- Schneider, A. (2016). Astrophysical constraints on resonantly produced sterile neutrino dark matter. *JCAP* 1604:059. doi: 10.1088/1475-7516/2016/04/059
- Shifman, M. A., Vainshtein, A. I., and Zakharov, V. I. (1980). Can confinement ensure natural CP invariance of strong interactions? *Nucl. Phys. B* 166, 493–506.
- Shin, M. (1987). Light neutrino masses and strong CP problem. *Phys. Rev. Lett.* 59, 2515–2518. doi: 10.1103/PhysRevLett.59.2515
- Shuve, B., and Tamarit, C. (2017). Phase transitions and Baryogenesis from decays. *JHEP* 1710:122. doi: 10.1007/JHEP10(2017)122
- Sikivie, P. (1982). Of axions, domain walls and the early universe. *Phys. Rev. Lett.* 48, 1156–1159. doi: 10.1103/PhysRevLett.48.1156
- Sirunyan, A. M., Tumasyan, A., Adam, W., Ambrogio, F., Asilar, E., Bergauer, T., et al. (2017). Search for evidence of the type-III seesaw mechanism in multilepton final states in proton-proton collisions at  $\sqrt{s} = 13$  TeV. *Phys. Rev. Lett.* 119:221802. doi: 10.1103/PhysRevLett.119.221802
- Starobinsky, A. A. (1980). A new type of isotropic cosmological models without singularity. *Phys. Lett. B* 91, 99–102. doi: 10.1016/0370-2693(80)90670-X
- Tanabashi, M., Hagiwara, K., Hikasa, K., Nakamura, K., Sumino, Y., Takahashi, F., et al. (2018). Review of particle physics. *Phys. Rev. D* 98:030001. doi: 10.1103/PhysRevD.98.030001
- Tegmark, M., Aguirre, A., Rees, M., and Wilczek, F. (2006). Dimensionless constants, cosmology and other dark matters. *Phys. Rev. D* 73:023505. doi: 10.1103/PhysRevD.73.023505
- Tkachev, I., Khlebnikov, S., Kofman, L., and Linde, A. D. (1998). Cosmic strings from preheating. *Phys. Lett. B* 440, 262–268. doi: 10.1016/S0370-2693(98)01094-6
- Tkachev, I. I. (2015). Fast radio bursts and axion miniclusters. *JETP Lett.* 101:1. doi: 10.1134/S0021364015010154
- Vafa, C., and Witten, E. (1984). Parity conservation in QCD. *Phys. Rev. Lett.* 53:535. doi: 10.1103/PhysRevLett.53.535
- Vaquero, A., Redondo, J., and Stadler, J. (2018). Early seeds of axion miniclusters. *JCAP* 4:012. doi: 10.1088/1475-7516/2019/04/012
- Visinelli, L., Baum, S., Redondo, J., Freese, K., and Wilczek, F. (2018). Dilute and dense axion stars. *Phys. Lett. B* 777, 64–72. doi: 10.1016/j.physletb.2017.12.010
- Volkas, R. R., Davies, A. J., and Joshi, G. C. (1988). Naturalness of the invisible axion model. *Phys. Lett. B* 215, 133–138. doi: 10.1016/0370-2693(88)91084-2
- Weinberg, S. (1978). A new light boson? *Phys. Rev. Lett.* 40:223.
- Wilczek, F. (1978). Problem of strong p and t invariance in the presence of instantons. *Phys. Rev. Lett.* 40, 279–282.
- Wise, M. B., Georgi, H., and Glashow, S. L. (1981). SU(5) and the invisible axion. *Phys. Rev. Lett.* 47, 402–404. doi: 10.1103/PhysRevLett.47.402
- Wolfenstein, L. (1978). Neutrino oscillations in matter. *Phys. Rev. D* 17:2369. doi: 10.1103/PhysRevD.17.2369
- Wolfenstein, L. (1979). Neutrino oscillations and stellar collapse. *Phys. Rev. D* 20, 2634–2635. doi: 10.1103/PhysRevD.20.2634
- Yanagida, T. (1979). Horizontal symmetry and masses of neutrinos. *Conf. Proc. C* 7902131:95.
- Zhitnitsky, A. R. (1980). On possible suppression of the axion hadron interactions. *Sov. J. Nucl. Phys.* 31:260.
- Zhong, L., Al Kenany, S., Backes, K. M., Brubaker, B. M., Cahn, S. B., Carosi, G., et al. (2018). Results from phase 1 of the HAYSTAC microwave cavity axion experiment. *Phys. Rev. D* 97:092001. doi: 10.1103/PhysRevD.97.092001
- Zurek, K. M., Hogan, C. J., and Quinn, T. R. (2007). Astrophysical effects of scalar dark matter miniclusters. *Phys. Rev. D* 75:043511. doi: 10.1103/PhysRevD.75.043511

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Ballesteros, Redondo, Ringwald and Tamarit. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Impact of Cosmological and Astrophysical Constraints on Dark Matter Simplified Models

Chiara Arina\*

*Centre for Cosmology, Particle Physics and Phenomenology, Université catholique de Louvain, Louvain-la-Neuve, Belgium*

## OPEN ACCESS

### Edited by:

Alberto Salvio,  
European Organization for Nuclear  
Research (CERN), Switzerland

### Reviewed by:

Stefano Profumo,  
University of California, Santa Cruz,  
United States  
Marco Regis,  
Università degli Studi di Torino, Italy

### \*Correspondence:

Chiara Arina  
chiara.arina@uclouvain.be

### Specialty section:

This article was submitted to  
High-Energy and Astroparticle  
Physics,  
a section of the journal  
Frontiers in Astronomy and Space  
Sciences

**Received:** 11 May 2018

**Accepted:** 10 August 2018

**Published:** 11 September 2018

### Citation:

Arina C (2018) Impact of Cosmological  
and Astrophysical Constraints on Dark  
Matter Simplified Models.  
Front. Astron. Space Sci. 5:30.  
doi: 10.3389/fspas.2018.00030

Studies of dark matter models lie at the interface of astrophysics, cosmology, nuclear physics, and collider physics. Constraining such models entails the capability to compare their predictions to a wide range of observations. In this review, we present the impact of global constraints to a specific class of models, called dark matter simplified models. These models have been adopted in the context of collider studies to classify the possible signatures due to dark matter production, with a reduced number of free parameters. We classify the models that have been analyzed so far and for each of them we review in detail the complementarity of relic density, direct, and indirect searches with respect to the LHC searches. We also discuss the capabilities of each type of search to identify regions where individual approaches to dark matter detection are the most relevant to constrain the model parameter space. Finally we provide a critical overview on the validity of the dark matter simplified models and discuss the caveats for the interpretation of the experimental results extracted for these models.

**Keywords:** dark matter theory, particle dark matter, direct searches of dark matter, indirect searches of dark matter, beyond standard model physics, LHC phenomenology

## 1. INTRODUCTION

The presence of dark matter, postulated at the beginning of last century (Jeans, 1922; Kapteyn, 1922; Oort, 1932; Zwicky, 1933; see Bertone and Hooper, 2016; de Swart et al., 2017 for a review), has been nowadays confirmed by several observations in cosmology and astrophysics. Besides precision measurements on its abundance from the cosmic microwave background and large scale structures, which state  $\Omega_{\text{DM}} h^2 = 0.1198 \pm 0.00015$  (Planck Collaboration et al., 2016), there is only gravitational evidence for this dark component while its nature and properties are completely unknown. Baryons can constitute only the 4% of the total energy content of the universe, not enough to explain the entire matter content of the universe ( $\sim 30\%$ ). This fact supports a non-baryonic origin for the dark matter particles, most likely arising in models beyond the Standard Model (SM) of particle physics, as SM neutrinos were relativistic in the early universe. Several theoretically motivated extensions of the SM, such as supersymmetry or universal extra-dimensions, provide dark matter candidates which fall into the category of WIMPs (Weakly Interacting Massive Particles). These particles are usually neutral, stable at least on cosmological scale, and with a mass in the GeV-TeV energy range. In this review we will comply with the WIMP paradigm and use WIMPs and dark matter interchangeably, even though other possibilities exist (see e.g., Marsh, 2016; Drewes et al., 2017) and the references therein.

With the generic hypothesis that WIMPs interact with the SM particles, a multitude of experimental approaches have been undertaken to detect it. These methods range from dark matter searches in underground detectors (Akerib et al., 2017; Amole et al., 2017; Aprile et al., 2017a) via the scattering of WIMPs off nuclei (direct detection), to observations of gamma rays, cosmic rays, and neutrinos, produced by dark matter annihilation in astrophysical environments (indirect detection) (see e.g., Aartsen et al., 2016; Aguilar et al., 2016; Albert et al., 2017; Ambrosi et al., 2017), and dedicated searches for missing energy signals at colliders (see e.g., Abercrombie et al., 2015; Boveia et al., 2016) (production). Yet, despite the enormous experimental effort, the dark matter detection remains a challenge and our understanding of dark matter properties limited, hence WIMP models can span many orders of magnitude in dark matter masses and interaction strengths. This makes it difficult to efficiently study all possible scenarios and models. It is necessary to find a strategy to combine the maximum amount of available experimental information in the most efficient way to: (i) carve out the dark matter models which are inconsistent with experimental observations; (ii) to highlight the most promising regions for discovery in the model parameter space, in the light of the near future dark matter search program; (iii) to highlight the complementarity among the diverse dark matter search methods. Dark matter simplified models (DMsims from hereafter) represent a convenient framework where to achieve these objectives, and will be the main focus of the review.

In these past few years, the dark matter program at the LHC has set the trend to follow the avenue of DMsims (Abdallah et al., 2015; Abercrombie et al., 2015; Boveia et al., 2016; Albert et al., 2017a), as compared to the Effective Field Theory (EFT) approach or as compared to the study of complete dark matter models. EFT states that the dark matter is the only accessible particle at our experiment, while all the other states that might characterize the dark sector are kinematically unaccessible. This is a valid framework when the masses of all particles mediating the interaction between the dark matter and the SM particles are assumed to be larger than the energy scale of the process. The limitations of this approach, at least as far as the LHC searches are concerned (Goodman and Shepherd, 2011; March-Russell et al., 2012; Shoemaker and Vecchi, 2012; Buchmueller et al., 2014; Busoni et al., 2014a,b,c; Bell et al., 2015; De Simone and Jacques, 2016), have now been recognized by the theoretical and experimental communities. Basically as soon as the momentum transfer of the process is near or larger than the mass of the mediator, EFT breaks down and the micro-physics describing the process needs to be taken into account. As far as it concerns dark matter direct detection, the momentum transfer is about a few MeV, hence EFT is a well-defined framework that can be used unless the mediator mass is of the order the MeV. Dark matter indirect detection lies in between the two cases described above and will be discussed in details in the paper. Notice that nowadays EFT at the LHC is a useful tool to grasp complementary information for instance for high scale (Belyaev et al., 2017) or for strongly interacting (Bruggisser et al., 2016) dark matter models.

The opposite approach with respect to EFT stands in considering UV (ultraviolet) complete theoretical models, motivated for instance by solving the hierarchy or the little hierarchy problems, such as supersymmetric models. These models have been and still are being extensively investigated in dedicated study programs, by both the theoretical and experimental communities. The complication arising from such models is the large number of free parameters: at present the dark matter data have not enough constraining power (the only measurement so far being the dark matter relic density) to select specific values of these free parameters of the theory space, hence it is common to end up with degeneracies among the parameters. Conversely, complete models usually feature complex dark sectors with interesting correlations among observables that cannot be reproduced by the EFT or simple models.

These simple models, called DMsims, are constituted by the addition to the SM particle content of a dark matter candidate which communicates with at least the SM quarks via one mediator. This minimalistic construction consists in expanding the EFT interaction by introducing a new state that mediates the interactions of the dark matter (and of the dark sector<sup>1</sup>) with the SM. Simplified models are typically characterized by three or four free parameters: the dark matter mass  $m_{\text{DM}}$ , the WIMP-SM  $g_{\text{DM}}$ , and mediator-SM  $g_{\text{SM}}$  couplings (or equivalently the coupling WIMP-SM-mediator  $y$ ) and the mediator mass  $m_{\text{med}}$ . So far, they have proven useful to categorize the dark matter searches at the LHC and to set up an easy framework for comparison with direct and indirect searches of dark matter. There are however several caveats emerging from the use of DMsims in relation with LHC searches and direct/indirect dark matter searches, which are currently driving these models, which might seem purely phenomenological constructions, into more natural bottom-up theoretical models (Albert et al., 2017b).

The rest of this review is organized as follows. Section 2 provides a general overview on the dark matter searches, ranging from cosmology to collider. Section 3 presents the state of art of current DMsims, with respect to all the dark matter searches presented in the previous section. A special focus is given to the cosmological and astrophysical constraints, as collider constraints are described in depth in many reviews and recommendation papers (see e.g., Abdallah et al., 2015; Abercrombie et al., 2015; Boveia et al., 2016; De Simone and Jacques, 2016; Arcadi et al., 2017; Kahlhoefer, 2017; Morgante, 2018) and the references therein. In particular section 3.1 considers  $s$ -channel mediator models and distinguishes the case of spin-0, spin-1, and spin-2 bosons, whereas section 3.2 reviews the status of  $t$ -channel models. Section 4 discusses the theoretical caveats of DMsims, while section 5 presents potential avenues for the future. We have tried to present the material in a self-contained form as much as possible, so that the review might serve as an introduction for the beginner and as a reference guide for the practitioner.

<sup>1</sup>If the dark sector consists of more than one new mediating particle, DMsims take into account the effect only of the lightest state that can be produced at the LHC.

## 2. OVERVIEW ON DARK MATTER SEARCHES

### 2.1. Cosmological Constraints, Astrophysical, and Direct Searches

In this section we provide the theoretical basic ingredients to compute cosmological, astrophysical, or scattering signals from the DMsims. For each type of search we discuss whether it is pertinent to use the EFT approximation or if the micro-physics processes should be fully taken into account. A summary plot is provided in **Figure 1**. We also discuss the theoretical assumptions and uncertainties related to each type of search, and how these are interpreted in terms of DMsims. Finally we briefly review the cosmological constraints on dark matter as well as the several astroparticle searches that are currently running. These constraints will be used to assess the complementarity of searches on the DMsimp parameter space in section 3.

#### 2.1.1. Dark Matter Relic Density

In the standard scenario, dark matter is considered a non-relativistic thermal relic in the early universe, which freezes out at  $x_f = T/m_{\text{DM}}$  (with  $T$  being the temperature of the thermal bath). Its relic abundance is given roughly by the following approximate solution of the evolution equation (see e.g., Srednicki et al., 1988; Kolb and Turner, 1990):

$$\Omega_{\text{DM}} h^2 \propto \frac{0.2 \times 10^{-9} \text{GeV}}{\langle \sigma v \rangle}, \quad (1)$$

where  $\Omega_{\text{DM}}$  is as usual the ratio between the dark matter energy density and the critical density of the universe,  $h$  is the reduced Hubble parameter ( $h = H_0/100 \text{ km s}^{-1} \text{ Mpc}^{-1}$ , with  $H_0$  being the Hubble constant today), and  $\sigma v$  is the thermally averaged annihilation cross section. The interaction of the particles needs to be extremely weak in order to achieve  $\Omega_{\text{DM}} h^2 \sim 0.1$ . To have an idea of the size of the coupling leading to the correct relic density, it is instructive to see what happens by considering a coupling with the strength of the weak interaction,  $\sigma v \sim G_F^2 m_{\text{DM}}^2 \sim 10^{-9} \text{GeV}^{-2} \sim 3 \times 10^{-26} \text{cm}^3/\text{s}$  ( $G_F$  is the Fermi constant): this is just right in the ballpark to achieve the measured  $\Omega_{\text{DM}} h^2$  for particles with masses around 100 GeV. Hence WIMPs have extremely small but non-vanishing interaction couplings with the SM; the size of the couplings depend on the mass of the WIMP and is typically of the order or smaller than the weak interaction for dark matter particles in the GeV-TeV range<sup>2</sup> to account for all the relic density. This is what is intended with the WIMP paradigm.

The inverse proportionality between  $\Omega_{\text{DM}}$  and the thermal averaged cross section dictates that: (i) the stronger the interaction rate is, the more depleted is the dark matter number density and as a consequence its relic abundance is too low (“under-abundant” dark matter, namely it contributes to  $\Omega_{\text{DM}} h^2$

by some %); (ii) annihilation processes for WIMPs are less efficient, the dark matter particles freeze out at early time and at present time have a significant abundance that matches  $\Omega_{\text{DM}} h^2$ ; (iii) the dark matter particles are too feebly interacting, hence they decouple too early and over-close the universe (“over-abundant” dark matter).

If  $\langle \sigma v \rangle$  varies slowly with energy, it can be expanded in plane waves Srednicki et al. (1988):

$$\langle \sigma v \rangle = \langle a + bv^2 + cv^4 + \dots \rangle = a + \frac{3}{2} \frac{b'}{x} + \frac{15}{8} \frac{c}{x^2} + \dots, \quad (2)$$

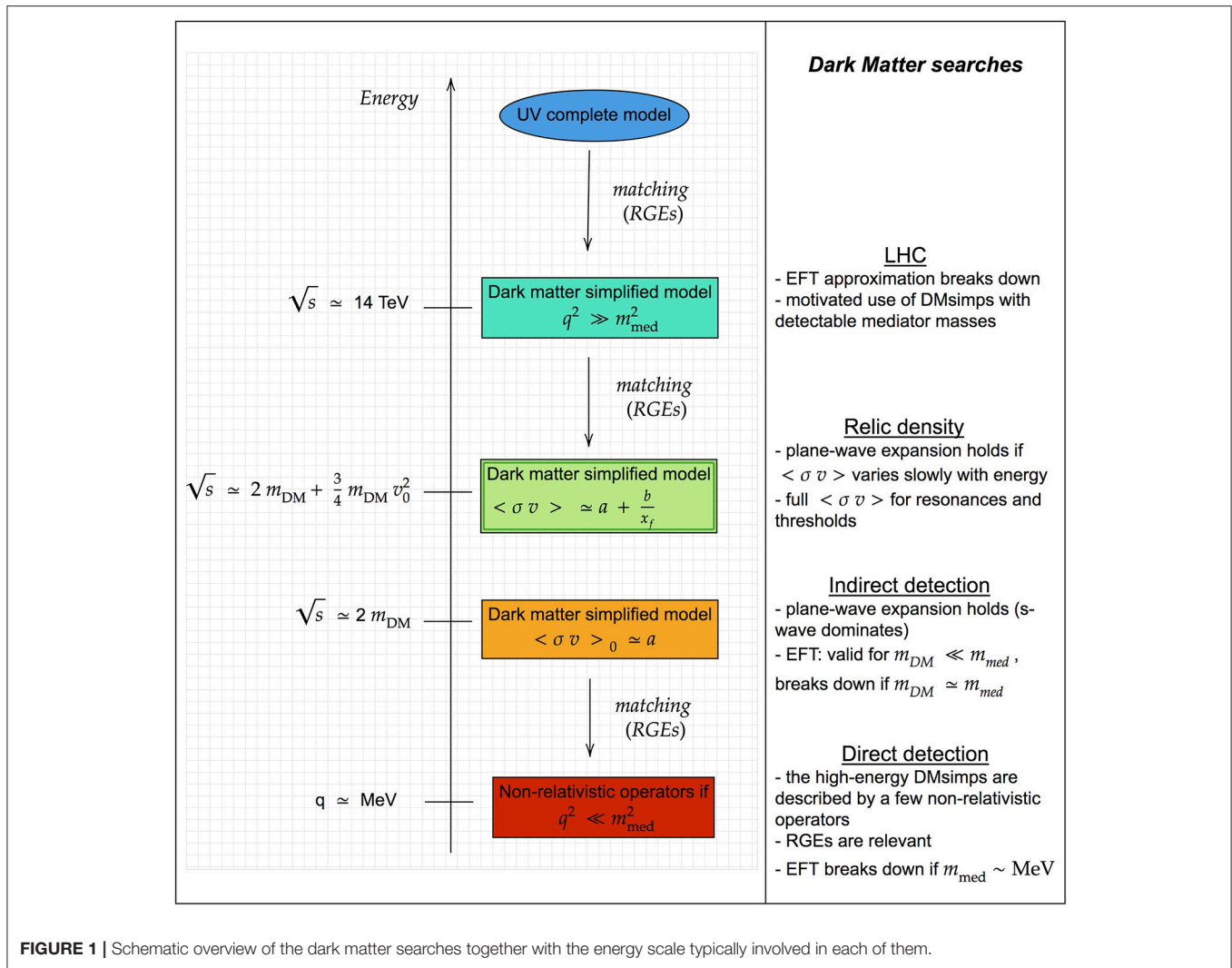
where  $b = 3/2b'$ . Typically freeze out occurs at  $x_f \sim 20 - 30$  leading to a most probable velocity  $v_0$  of the order of  $0.25c$ : corrections proportional to  $x^{-1}$  are indeed relevant with respect to the  $a$  term and need to be taken into account (notice that the back of the envelop estimate in Equation (1) is valid only for a pure  $s$ -wave  $\langle \sigma v \rangle$ ). Additionally, there might be selection rules at play that make the  $s$ -wave term vanishing. This occurs for several DMsims, as we will discuss in the next section, which end up having  $p$ -wave dominated annihilation cross sections.

There are circumstances in which the non-relativistic expansion of  $\langle \sigma v \rangle$  in Equation (2) breaks down (Gondolo and Gelmini, 1991; Griest and Seckel, 1991; Edsjo and Gondolo, 1997):

- **Resonant annihilation:** The annihilation cross section is not a smooth function of the centre-of-mass-energy  $s$  in the vicinity of an  $s$ -channel resonance. For  $m_{\text{DM}} \leq 2m_{\text{med}}$  the additional kinetic energy provided by the thermal bath brings  $s$  on top of the resonance and the annihilation cross section increases drastically. Conversely, for  $m_{\text{DM}} > 2m_{\text{med}}$  the additional kinetic energy brings  $s$  even further away from the resonance, hence the annihilation cross section decreases quickly.
- **Opening of new annihilation channels:** a fraction of dark matter particles might have a kinetic energy, given by their thermal distribution in the early universe, sufficient to annihilate into heavier particles, which are above the threshold. This again leads to a rapid enhancement of  $\langle \sigma v \rangle$ .
- **Co-annihilation:** if there are one or more particles heavier than the dark matter but close in mass (roughly speaking the difference in mass should be at most 10% of the dark matter mass), these are present as well in the thermal bath in the early universe and contribute to the relic abundance of the dark matter with annihilation processes among themselves and with the dark matter. These processes should be taken into account in the Boltzmann equation that leads to  $\Omega_{\text{DM}} h^2$  with an effective  $\langle \sigma v \rangle$ , the weighted sum over all annihilation processes (see for details Edsjo and Gondolo, 1997).

In these cases it is necessary to rely on the full computation of the thermally averaged cross section without approximations as well as to solve the complete Boltzmann equation to evaluate precisely  $\Omega_{\text{DM}} h^2$ . This is the standard procedure encoded in the public tools for dark matter [see e.g., micrOMEGAs (Belanger et al., 2018), DarkSUSY (Bringmann et al., 2018), and MadDM (Ambrogio et al., 2018), etc.]. As we will see in section 3, the model parameter space of DMsims often features the

<sup>2</sup>An upper bound on the WIMP mass of  $\mathcal{O}(100)\text{TeV}$  stems from the requirement of perturbative unitarity (Griest and Kamionkowski, 1990), while a lower bound is much more debated and model dependent. In certain models  $m_{\text{DM}} > 10 \text{ GeV}$  not to spoil recombination (Planck Collaboration et al., 2016), for other models  $m_{\text{DM}} > 4 \text{ GeV}$  (Lee and Weinberg, 1977), etc.



correct relic density in a tuned-region  $m_{DM} \sim 2m_{med}$ , relying on resonant annihilation, and several threshold openings are at play in setting  $\Omega_{DM}h^2$ . This is schematically resumed in **Figure 1**.

The dark matter relic density is the only precision measurement we have so far. As already anticipated in the introduction, it has been measured with great accuracy by the Planck satellite (Planck Collaboration et al., 2016). This measurement, combined with large scale structure data, gives:  $\Omega_{DM}h^2 = 0.1198 \pm 0.00015$ . The experimental error is at the level of ‰, two orders of magnitude smaller than the associated theoretical error, typically quoted around  $\mathcal{O}(10\%)$ .

There are a certain number of caveats when considering the relic density as a constraint for DMsims, which have to be taken into account in the interpretation of the complementarity of searches:

- DMsims provide an extension of the SM particle content into the most minimalistic dark sector possible, constituted solely by the dark matter and an extra mediator. If the dark/new

physics sector contains more particles, two types of processes can alter the relic density value: (i) there are additional mediators, opening up new annihilation channels including resonance effects; (ii) co-annihilation (Edsjo and Gondolo, 1997), if there are particles heavier but close in mass with the dark matter mass ( $\Delta m \lesssim \mathcal{O}(10\%)$ ). The region allowed by relic density in the DMsimp set up should be considered then as a subset of the whole allowed model parameter space.

- DMsims focus particularly on studying and constraining the dark matter-quark couplings, which are accessible at the LHC. However if the dark matter couples to other SM species, additional annihilation diagrams can have a significant impact on the model parameter space allowed by the relic density constraint by opening up new annihilation thresholds. Other couplings, such as dark matter—lepton couplings, start to be considered as well (Albert et al., 2017a) in the context of di-lepton searches. In these cases the interpretation of the allowed relic density regions becomes more robust (cfr. the other caveats).

- The constraint on  $\Omega_{\text{DM}} h^2$  relies on the assumption that the dark matter is a thermal relic. Other viable assumptions to bring  $\Omega_{\text{DM}} h^2$  to the observed value, are for instance: (i) the dark matter is non-thermally produced; (ii) the cosmological evolution of our universe is rather different than the one described by the Standard Cosmological model. For example, late-time entropy injection (Bramante and Unwin, 2017) can decrease the dark matter relic density, while late gravitino decays in supersymmetric theories can increase the neutralino relic abundance (Allahverdi et al., 2012).

In section 3, we will discuss the combination of dark matter searches in full generality, with and without considering the relic density as relevant constraint. Notice that all caveats described above spoil the model-independent approach of DMsims, as they rely on the specificity of the dark matter model.

### 2.1.2. Dark Matter Direct Detection

As the dark matter particles move in the Milky Way halo, it is worthwhile to explore the possibility to detect them. This can be done directly in underground terrestrial detectors, sensitive to the nuclear recoil caused by the passing wind of dark matter particles. From a theoretical point of view, in direct detection, the crucial quantity is the scattering cross section of dark matter particles off a nucleon, in a deeply non-relativistic regime. Indeed the momentum transfer in the collision is of the order of a few to tens of MeV, as the speed of the incoming WIMP is of the order of  $v \sim 10^{-3}c$ . As a consequence, direct detection can be safely treated in term of EFT<sup>3</sup>, except when the mediator mass connecting the dark matter and the SM quarks becomes of the order of the momentum transfer ( $m_{\text{med}}^2 \sim q^2 \sim (O)(10\text{MeV})$ ), as resumed in **Figure 1**.

It has been shown that the scattering process between the dark and ordinary matter can be expressed in terms of a limited number of relativistic degrees of freedom, which give rise to a basis of non-relativistic operators. As a matter of fact, any process of elastic scattering between the dark matter and the nucleon can be expressed as a combination of this basis in a unique way, irrespective of the details of the high-energy dark matter model. This basis is constituted by 12 operators, here we report the most relevant for the discussion of section 3 using the notation of (Cirelli et al., 2013):

$$\begin{aligned} \mathcal{O}_1^{\text{NR}} &= 1, & \mathcal{O}_4^{\text{NR}} &= \mathbf{s}_{\text{DM}} \cdot \mathbf{s}_N, \\ \mathcal{O}_6^{\text{NR}} &= (\mathbf{s}_{\text{DM}} \cdot \mathbf{q})(\mathbf{s}_N \cdot \mathbf{q}), & \mathcal{O}_8^{\text{NR}} &= \mathbf{s}_{\text{DM}} \cdot \mathbf{v}^\perp, \\ \mathcal{O}_9^{\text{NR}} &= i \mathbf{s}_{\text{DM}} \cdot (\mathbf{s}_N \times \mathbf{q}), \end{aligned} \quad (3)$$

Starting from the DMsimp Lagrangian, which describes the interaction of the dark matter with the quarks, it is necessary first to determine the dark matter-nucleon effective Lagrangian. Secondly, the elastic scattering occurs with the whole nucleus, due to the small WIMP speed in the galactic halo. Therefore, one needs to properly take into account the composite structure of the nucleus which results in the appearance of nuclear form factors in the cross section. Nuclear form factors parametrize

<sup>3</sup>This approximation is satisfied by the DMsimp framework, which typically features mediators heavier than GeV.

**TABLE 1** | List of direct detection EFT operators WIMP-nuclei for fermionic and scalar dark matter arising from the DMsimp high-energy interaction Lagrangians discussed in the paper.

Dark matter candidate	EFT operator	Matching
Fermionic	$\bar{X}X\bar{N}N$	$4m_{\text{DM}}m_N\mathcal{O}_1^{\text{NR}}$
	$i\bar{X}\gamma_5 X\bar{N}N$	$-4m_N\mathcal{O}_{11}^{\text{NR}}$
	$i\bar{X}X\bar{N}\gamma_5 N$	$4m_{\text{DM}}\mathcal{O}_{10}^{\text{NR}}$
	$i\bar{X}\gamma_5 X i\bar{N}\gamma_5 N$	$4\mathcal{O}_6^{\text{NR}}$
	$\bar{X}\gamma^\mu X\bar{N}\gamma_\mu N$	$8m_{\text{DM}}(m_N\mathcal{O}_8^{\text{NR}} + \mathcal{O}_9^{\text{NR}})$
	$i\bar{X}\gamma^\mu \gamma_5 X\bar{N}\gamma_\mu N$	$8m_N(-m_{\text{DM}}\mathcal{O}_8^{\text{NR}} + \mathcal{O}_9^{\text{NR}})$
	$i\bar{X}\gamma^\mu X\bar{N}\gamma_\mu \gamma_5 N$	$-16m_{\text{DM}}m_N\mathcal{O}_4^{\text{NR}}$
	$i\bar{X}\gamma^\mu \gamma_5 X i\bar{N}\gamma_\mu \gamma_5 N$	$32m_{\text{DM}}m_N\mathcal{O}_4^{\text{NR}}$
Scalar	$\Phi^*\Phi\bar{N}N$	$2m_{\text{DM}}\mathcal{O}_1^{\text{NR}}$
	$i\Phi^*\Phi\bar{N}\gamma_5 N$	$2\mathcal{O}_{10}^{\text{NR}}$

We provide the matching between these EFT operators and the non-relativistic (NR) operators in the third column. The WIMP-parton coefficients and the transformations from parton level to nuclear EFT operators can be found in e.g., Del Nobile et al. (2013). The dark matter particle is denoted by  $X$  if fermionic and by  $\Phi$  if scalar, while the nucleus is denoted by  $N$  and has a mass  $m_N$ . For both Majorana fermions and real scalars the vector operators vanish, reducing the list of relevant relativistic operators.

the loss of coherence in the scattering with increasing exchanged momentum. In **Table 1**, we provide the list of non-relativistic operators relevant for the DMsims presented in section 3 and their matching with the matrix element involving the whole nucleus. We refer to Cirigliano et al. (2012), Cirelli et al. (2013), Fitzpatrick et al. (2013), and De Simone and Jacques (2016) for the rigorous definition of the non-relativistic operator basis and for the detailed direct detection analyses<sup>4</sup>.

Concerning the experimental state of art for direct detection, a huge experimental effort has been deployed in the past years, that features nowadays more than 10 different experiments currently running toward unprecedented sensitivities. Several orders of magnitude in the WIMP-nucleus elastic interaction have been constrained by past and current experiments. As far as it concerns spin-independent elastic scattering, which occurs when the dark matter interacts with all the nucleons (it is proportional to the atomic number of the nucleus,  $A^2$ ), the most notable experiments are XENON1T (Aprile et al., 2017a), LUX (Akerib et al., 2017), and PANDAX-II (Fu et al., 2017) for intermediate WIMP masses, CDMSlite (SuperCDMS Collaboration et al., 2018) and CRESST-II (Angloher et al., 2016) at low WIMP masses. XENON1T excludes at 90% confidence level (CL) WIMP-nucleon cross sections of about  $8 \times 10^{-47} \text{cm}^2$  for dark matter masses of 30 GeV. The usual spin-independent scattering cross section corresponds to the operator  $\mathcal{O}_1^{\text{NR}}$  of **Table 1**. If present in the underlying particle physics model, this operator dominates over all other non-relativistic operators. Spin-dependent scattering occurs when the dark matter interacts with the spin of the unpaired proton or neutron of the nucleus. PICO 60 Amole et al. (2017) detains the most constraining bound for spin-dependent

<sup>4</sup>On a side note, except for (Cirelli et al., 2013), the publicly available dark matter numerical tools do not use the general description of direct detection in terms of non-relativistic operators, at the best of our knowledge at the time of writing.

scattering on proton so far. Only a few experiments are sensitive to the spin-dependent interaction on neutron (mostly dual phase xenon or noble liquid/gas detectors) and the strongest exclusion bound is held by the LUX Collaboration et al. (2017) experiment. The spin-dependent operator currently considered by the experimental collaborations is  $\mathcal{O}_4^{\text{NR}}$ . Exclusion limits for the other operators are provided in Cirelli et al. (2013), even though at present these exclusion bounds are a bit outdated. On the experimental side, the XENON collaboration has started to use the non-relativistic operator description and has released exclusion limits based on the XENON100 data (Aprile et al., 2017b).

Direct detection is affected by several astrophysical uncertainties related for instance to the description of the dark matter velocity distribution at the Sun position and to the local dark matter density. There are two different approaches to deal with these uncertainties: either perform a likelihood analysis and marginalize or profile over them (see e.g., Strigari and Trotta, 2009; Arina et al., 2011; Bertone et al., 2012; Arina, 2014), either use the so-called halo-independent method (see e.g., Fox et al., 2011; Gondolo and Gelmini, 2012; Del Nobile et al., 2013). In most of the analyses described in section 3, astrophysical uncertainties are not taken into account, hence we will not consider this matter any further.

### 2.1.3. Dark Matter Indirect Detection

Dark matter indirect detection relies on the principle that dark matter particles in galactic halos annihilate into SM particles. These SM particles subsequently undergo decays, showering and hadronization and lead to a continuum flux of cosmic rays, gamma rays, and neutrinos. In the case where the dark matter annihilates via loop-induced processes into a pair of photons or a photon and a boson, the signal is characterized by a sharp spectral feature such as a gamma-ray line. Dark matter annihilation takes place in several astrophysical environments and at different epochs in the evolution of the universe, from cosmological down to solar system scales. As dark matter indirect detection encompasses a large variety of searches, in this review we describe only the searches that have been directly used as complementary probes together with LHC dark matter searches and/or direct detection to constrain DMsimps. Those involve mainly gamma rays, neutrinos and anti-protons at galactic scales. For a detailed review on dark matter indirect detection we refer the reader to (e.g., Cirelli, 2015; Gaskins, 2016; Slatyer, 2017).

Before going into the details of the specific searches and theoretical predictions, let us mention two generic features concerning the flux of particles produced by dark matter annihilation. This quantity is proportional to

1.  $\langle\sigma v\rangle_0$ . This is defined as the velocity averaged annihilation cross section computed at present time. Annihilation in galactic halos occurs in a highly non-relativistic regime with an centre-of-mass-energy provided by  $\sqrt{s} = 2m_{\text{DM}}$  as the typical mean velocities characterizing the dark matter halo are negligible. For instance in the Milky Way the most probable velocity of dark matter particles is  $v_0 \sim 10^{-3}c \sim 230 \text{ km/s}$  (Schoenrich et al., 2010), while it is even lower

in dwarf Spheroidal galaxies (dSphs),  $v_0 \sim 10^{-5}c \sim 8 \text{ km/s}$  (Bonnivard et al., 2015), hence in indirect searches the non-relativistic expansion of  $\langle\sigma v\rangle_0$  in plane waves is a fairly good approximation. The dominant term that is in the reach of current astrophysical probe is the  $s$ -wave:  $\langle\sigma v\rangle_0 \simeq a$ . If this term is absent due to some selection rule, the model is most likely unconstrained from indirect detection. Notice that the EFT approach remains valid and can be used for  $m_{\text{DM}} \ll m_{\text{med}}$ . This is summarized in **Figure 1**.

2.  $dN_f/dE_f \equiv \sum_i B_i dN_f^i/dE_f$ . This is defined as the energy spectrum of the particle species  $f$  (with  $f = \gamma, \nu_l, e^+, \bar{p}$ , and  $l$  is the neutrino flavor,  $l = e, \mu, \tau$ ) at production where annihilation occurred. The index  $i$  runs over all possible annihilation final states of the dark matter model, each of them with a branching ratio  $B_i$ . The final states are typically SM pairs of particles, however new particles beyond the SM can appear as well, which will subsequently decay into SM particles. We will see in section 3 that this option is realized in several DMsimps.

Typically the experimental searches present the limits in a model-independent way, supposing a branching ratio of 100% into one species of SM particles and assuming that  $\Omega_{\text{DM}} h^2$  matches the observed value. To compare a specific dark matter model with the experimental exclusion limits, the most rigorous procedure is to recompute the upper bound for that particular model by means of the experimental likelihoods. If this is not possible, one can combine the experimental exclusion bounds after having rescaled them by the appropriate branching ratio. This procedure should be a good approximation provided the energy spectrum of the specific model does not differ too much from the energy spectrum for which each respective upper bound has been computed. The micrOMEGAs and DarkSUSY numerical tools rely on tabulated energy spectra for all possible SM final states and for dark matter masses ranging from 5 to 100 TeV. The MadDM tool (Ambrogio et al., 2018) allows to generate the energy spectrum in both model-independent and model-dependent ways for any possible dark matter mass.

Similarly to direct detection, indirect detection is affected by astrophysical uncertainties related to the dark matter density distribution in galactic halos, by the propagation parameters for cosmic rays, etc. Whenever relevant, we will discuss the comparison between different dark matter searches and the indirect detection limits based on different assumptions on the astrophysics.

### Gamma-ray searches

The gamma-ray flux from dark matter annihilation from a direction  $\psi$  in the sky, averaged over an opening angle  $\Delta\psi$ , is given by:

$$\frac{d\Phi}{dE_\gamma}(E_\gamma, \psi) = \frac{\langle\sigma v\rangle_0}{2m_\chi^2} \sum_i B_i \frac{dN_\gamma^i}{dE_\gamma} \frac{1}{4\pi} \int_{\psi} \frac{d\Omega}{\Delta\psi} \int_{\text{los}} \rho^2(\psi, l) dl. \quad (4)$$

For dark matter particles with distinct particle and antiparticle Equation (4) is multiplied by an additional factor of 1/2. The two integrals, over the angle and the line of sight (los), define the astrophysical  $J$  factor ( $J \equiv \int_{\psi} d\Omega/\Delta\psi \int_{\text{los}} \rho^2(\psi, l) dl$ ). The  $J$  factor encodes the information about the astrophysical environment (experimental window) where annihilation occurs (is sought) and the dark matter density profile.

We start by illustrating the gamma-ray constraints from dSphs, which are dark matter dominated objects (Mateo, 1998; Weisz et al., 2011; Brown et al., 2012; Coureau et al., 2014). The Fermi-LAT satellite looks for a gamma-ray emission from these Milky Way satellite galaxies, and so far, no excess in gamma rays has been observed.<sup>5</sup> Hence the Fermi-LAT collaboration has set upper bounds at 95% CL on the continuum prompt photon flux produced by dark matter annihilation (Ackermann et al., 2015a; Albert et al., 2017). From these bounds, it has publicly released upper limits for the annihilation rate into  $b\bar{b}$  and  $\tau^+\tau^-$  final states as a function of the dark matter mass. The  $b\bar{b}$  channel is an example of “soft” channel that produces photons mostly from the decay of neutral pions produced in hadronisation, while the  $\tau^+\tau^-$  is a “hard” channel that generates photons from final state radiation, scaling as  $E^{-1}$ , on top of the photons coming from  $\pi^0$  decays. The Fermi-LAT team has performed a stacked likelihood analysis for 45 dSphs. The resulting profile function for each dSph has been released publicly and can be used to compare for instance DMsims with dSphs data from the 6 years Fermi-LAT data (Pass 8 event reconstruction algorithm)<sup>6</sup>. These likelihood functions have been implemented in the last MacDM version (see Ambrogi et al., 2018) for details, and can be used for any generic dark matter model. Gamma-ray constraints from dSphs are subject to astrophysical uncertainties mostly related to the determination of the  $J$  factor. These uncertainties are particularly large in the case of the latest dSphs discovered, which are ultra-faint dwarf galaxies, for details (see Bonnavard et al., 2015). In addition to the prompt photon flux, there are also contribution from inverse Compton scattering generated by charged propagating particles. These are often neglected while computing the exclusion limits on the dark matter annihilation rate, however could have an impact for  $m_{\text{DM}} \geq 100$  GeV. Hence the exclusion bounds for large dark matter masses should be regarded as conservative.

Another search, used in the complementarity framework of DMsims, looks for gamma-ray spectral features toward the Galactic Centre. These spectral features encompass gamma-ray lines, narrow boxes (see e.g., Ibarra et al., 2015a) and sharp edges in the prompt photon energy spectrum coming for instance from internal bremsstrahlung processes (see e.g., Giacchino et al., 2013; Toma, 2013). The most constraining exclusion limits on the dark matter annihilation rate into gamma-ray

lines are provided by the Fermi-LAT satellite (Ackermann et al., 2015b) for  $m_{\text{DM}} < 500$  GeV and the HESS telescope for dark matter masses up to 25 TeV (Abramowski et al., 2013; Abdalla et al., 2016). These searches suffer of large astrophysical uncertainties related to the dark matter density profile, included in the  $J$  factor, and to the background modeling of the Galactic Centre<sup>7</sup>.

### Neutrino searches

If dark matter particles scatter in heavy astrophysical bodies such as the Sun, they can lose enough energy to become gravitationally trapped inside it. Dark matter particles start to accumulate in the center of these celestial bodies, where subsequently dark matter annihilation sets in (see e.g., Steigman et al., 1978; Press and Spergel, 1985; Silk et al., 1985; Gould, 1987; Ritz and Seckel, 1988; Kamionkowski, 1991; Jungman et al., 1996; Bergstrom et al., 1997; Gondolo et al., 2004; Blennow et al., 2008; Peter, 2009; Sivertsson and Edsjo, 2012). In the Sun, constituted primarily by hydrogen, the dark matter capture occurs mainly by spin-dependent elastic scattering [even though the spin-independent scattering on nucleons,  $\sigma_n^{\text{SI}}$ , can also play a role, as it is enhanced by the  $A^2$  term for heavy nuclei (Gondolo et al., 2004)]. The Sun is opaque to all dark matter annihilation products but neutrinos, which can escape the Sun surface and be detected by Earth based telescopes such as IceCube and Super-Kamiokande (Choi et al., 2015). The annihilation rate can become large enough to lead to an equilibrium between dark matter capture and annihilation. In this case  $\langle\sigma v\rangle_0$  and the elastic cross section on proton,  $\sigma_p^{\text{SD}}$ , become two related quantities that can be trade one for the other. This assumption is used for computing experimentally the exclusion bounds on the WIMP-nucleon elastic cross section. The IceCube collaboration has set stringent upper limits, competitive with those of direct detection searches for spin-dependent scattering (Aartsen et al., 2013, 2016), by the non observation of GeV-TeV scale neutrinos coming from the Sun direction. The exclusion bounds publicly released, at 90% CL, are based on IceCube data with 79 strings including DeepCore and are given for the following final states, “hard” channels ( $W^+W^-$ ,  $\tau^+\tau^-$ ,  $ZZ$ ,  $\nu\bar{\nu}$ ) and “soft” channels ( $b\bar{b}$ ,  $t\bar{t}$ ,  $gg$  and  $hh$ ).

The equilibrium assumption helps in the interpretation and comparison of dark matter exclusion limits coming from direct and indirect detection in terms of WIMP-quark coupling; this is particularly appreciable for DMsimp models, which often do include only these couplings. There is however an emergent caveat: direct detection experiments have pushed the upper bound on the spin-independent and spin-dependent cross-section to lower and lower values for which the equilibrium assumption starts to break down (Arina et al., 2017). Depending then on the size of  $\langle\sigma v\rangle_0$  and  $\sigma_n^{\text{SI,SD}}$ , the useful representation of exclusion bounds in terms of elastic scattering might not provide anymore a correct physical interpretation.

<sup>5</sup>There are four dSphs recently discovered by DES (Abbott et al., 2005), which, taken individually, show a slight excess over the background, of the order of  $2\sigma$ . Other analyses (see e.g., Geringer-Sameth et al., 2015; Hooper and Linden, 2015) have pointed out similarly a possible excess over the background. The excess disappears once the data are stacked with the other dSph data.

<sup>6</sup>Dataset (2015). Available online at: [https://www-glast.stanford.edu/pub\\_data/1203/](https://www-glast.stanford.edu/pub_data/1203/).

<sup>7</sup>In this review we do not consider the Galactic Center excess at low dark matter masses. For details, we refer the reader to (e.g., Gaskins, 2016) and the references therein.

### Anti-proton searches

Searches for dark matter annihilation products in local charged cosmic-ray fluxes can be highly sensitive, especially due to the low backgrounds for antimatter produced by astrophysical processes. A major challenge for these searches is the identification of the locations of the sources of cosmic rays due to their propagation throughout the Milky Way, conversely to the case of gamma rays and neutrinos, which do not diffuse and trace their source. Anti-protons have been recognized as important messengers not only for cosmic ray physics but constitute one of the primary channels in the dark matter searches (Silk and Srednicki, 1984; Silk et al., 2010). This idea has been further reinforced by the data released recently by the AMS 02 satellite (Aguilar et al., 2016), which have an amazing statistical precision and extend up to 450 GeV. The authors of (Giesen et al., 2015) and (Cuoco et al., 2018) have provided an analysis of these data in terms of exclusion limits for the dark matter velocity averaged annihilation cross section as a function of  $m_{\text{DM}}$  at 95% CL for the  $b\bar{b}$ ,  $g\bar{g}$ ,  $q\bar{q}$ ,  $t\bar{t}$ ,  $\mu^+\mu^-$ ,  $W^+W^-$ ,  $h\bar{h}$ , and  $\gamma\gamma$  final states. These bounds (especially  $b\bar{b}$ ) are used to assess the constraining power of anti-proton searches for DMsims in some of the analyses presented in section 3.

The exclusion limits on the dark matter annihilation rate from anti-protons suffer of very large astrophysical uncertainties. The exclusion limits can fluctuate upwards or downwards by one order of magnitude at low dark matter masses, mainly because of uncertainties in the propagation parameters in our galaxy and of solar modulation. The choice of the dark matter density profile is not the main cause of the lack of precision. For details we refer to Cirelli (2015), Giesen et al. (2015), Cuoco et al. (2018) and the references therein.

## 2.2. LHC Dark Matter Searches

In this section we summarize very briefly the main dark matter searches pursued by the LHC experimental collaborations. For a detailed information, we refer the reader to (e.g., Abercrombie et al., 2015; De Simone and Jacques, 2016; Albert et al., 2017a; Kahlhoefer, 2017; Morgante, 2018) and the references therein.

During the LHC Run 2, ATLAS and CMS have gone the avenue of dark matter simplified models to classify and categorize all possible final states arising in the dark matter search program. This method has been validated by the Dark Matter forum (Abercrombie et al., 2015) and further supported by the LHC Dark Matter Working Group, established as the successor of the Dark Matter Forum<sup>8</sup>.

The main bulk of dark matter searches at colliders is constituted by signatures with missing transverse energy ( $\cancel{E}_T$ ) in the final state, due to the pair-produced dark matter particles which elusively leave the detector with no trace. The mediator, produced by Drell-Yan or gluon fusion and decaying invisibly into a pair of dark matter particles, can be looked for by tagging an energetic jet, coming typically from initial state radiation, which balances the missing energy from the final state. These are the most relevant searches for DMsims undertook so far

by the ATLAS and CMS collaborations and are called mono- $X + \cancel{E}_T$  searches, where  $X$  stems for a jet, a photon, a vector boson, a Higgs, and multi-jets (from 2 to 6 jets) +  $\cancel{E}_T$ . All these searches require  $2m_{\text{DM}} < m_{\text{med}}$  and possible that the mediator has a large branching ratio into dark matter and SM particles (large  $g_{\text{DM}}$  and especially large  $g_{\text{SM}}$ ). Once these conditions are met, the searches are not very sensitive to the actual mass of the dark matter particle. This is the reason why LHC searches are more sensitive to very light dark matter masses, close the  $\mathcal{O}(1)$  GeV with respect to direct detection searches (Boveia et al., 2016). Additionally to mono- $X + \cancel{E}_T$  searches, a certain number of DMsims can be constrained by recasting searches in supersymmetric simplified models or by  $t\bar{t} + \cancel{E}_T$  searches.

Both the experimental and theoretical communities have recognized that resonance searches for the mediator can be as powerful as the  $\cancel{E}_T$  signals in DMsims, or in some case be even more constraining (see e.g., Arina et al., 2016; Albert et al., 2017a). These searches are based on the principle that, after its production by proton collisions, the mediator does not necessarily decay into dark matter particles but can decay back into SM final states. This is always the case for  $m_{\text{med}} < 2m_{\text{DM}}$ , as the invisible decay channel is closed; it is also satisfied for  $g_{\text{SM}} > g_{\text{DM}}$ , condition that leads to a small branching ratio into dark matter particles and a large branching ratio into visible SM species. Besides the two requirements above these searches as well are not very sensitive to the dark matter exact mass value. In general the most relevant resonance searches, depending on the specific of the DMsimp, are  $t\bar{t}$ , 4 tops, di-photons, dileptons, and di-jets. The sensitivity of each search depends on the specificity of the DMsimp under investigation. For instance, di-jet signals are irrelevant for scalar mediators, while  $t\bar{t}$  pair production and di-photons reveal very useful (Arina et al., 2016). Conversely spin-1 mediators are easily probed via di-jets and mono- $X$  signatures (Chala et al., 2015; du Pree et al., 2016).

Notice that the discovery of an anomalous signals in a mono- $X + \cancel{E}_T$  search at the LHC would not imply the discovery of dark matter, contrary to the case of direct and indirect detection searches. Hence a potential discovery at colliders needs to be supported by further evidence in direct or indirect searches, in order to fully identify the dark matter candidate. On the other hand, in case of new findings, LHC is able to provide an accurate characterization of the new mediator particle, while direct and indirect detection are more loosely sensitive to it.

## 3. CURRENT STATUS OF DARK MATTER SIMPLIFIED MODELS

Since the start of the LHC Run 2 and the publication of the DM forum recommendations (Abercrombie et al., 2015), the number of works studying DMsims has increased exponentially. DMsims have been adopted for their minimalistic structure to provide the SM with a dark matter particle, in the sense that they represent the minimal extension of the EFT approach used in the LHC Run 1 dark matter searches. The EFT operators are opened up by introducing a particle mediating the interaction between the dark matter and the SM particles (the so-called mediator).

<sup>8</sup>We chose not to provide any reference here for the specific searches conducted by ATLAS and CMS, and to provide the references in the next section, referring to the data sets actually used in the analyses discussed in this review.

They are simple enough to allow the LHC experimental collaborations to categorize all possible dark matter signals they can give rise to. A general classification stems from the class of vertices that characterize the model: Lagrangians featuring WIMP-WIMP-mediator and SM-SM-mediator type interactions identify models with an  $s$ -channel mediator, while Lagrangians characterized by WIMP-SM-mediator interactions define a  $t$ -channel mediator. In  $s$ -channel models, the mediator is always a color singlet, while in  $t$ -channel models it can be either a colored particle or a color singlet (even though this second possibility, is less appealing for the collider phenomenology). Nonetheless, the definition of DMsimp is not unique, especially as far as it concerns the mediator nature. Some works consider Higgs portal models as part of the DMsimp category (see e.g., Abdallah et al., 2015; De Simone and Jacques, 2016), while others do not include the SM Higgs boson in this context (Abercrombie et al., 2015; Boveia et al., 2016). For the rest of the section we will use the definition of DMsimp as provided in Abercrombie et al. (2015) and Boveia et al. (2016):

- There can be only one new mediator at a time that defines the interaction between the dark matter and the SM quarks. Namely the dark matter and the mediator are the only particle accessible by current experiments. The presence of additional new particles in the dark sector is assumed not to modify sensibly the physics described by DMsims. This assumption allows to introduce a very limited set of new free parameters (typically four). The mediator can have spin-0, spin-1/2, spin-1, and spin-2. The category of scalar mediators, however, does not include the Higgs boson (and no mixing with it is considered). We will comment on Higgs portal models in section 4.
- The new interaction should not violate the exact and approximate accidental global symmetries of the SM. For instance this means that baryon and lepton number conservation of the SM should be preserved by this interaction. Additionally, the new mediating particle can produce large flavor violating effects. By enforcing that the flavor structure of the couplings between the dark matter and the ordinary particles follows the same structure as in the SM, it is ensured that DMsims do not violate flavor constraints. This assumption is called Minimal Flavor Violation (MFV) (D'Ambrosio et al., 2002), for a detailed discussion (see e.g., Abdallah et al., 2015).
- Another recommendation concerns the nature of the dark matter particle. It is suggested to consider Dirac fermionic candidates only, because LHC searches are rather insensitive to the spin of the dark matter particles. As the  $E_T$  searches are based on cut-and-count analyses, minor changes in the kinematic distributions of the visible particle are expected to have little effect on these analyses, besides the fact the Majorana particles forbid some processes allowed for Dirac particles. However, whenever possible, we will review cases that go beyond the Dirac fermionic dark matter assumption, as the dark matter annihilation and elastic scattering cross sections do depend on the dark matter spin. Different selection rules are at play depending whether the dark matter is a

real scalar, a complex scalar, a Dirac or Majorana fermion, leading to suppressions or enhancements of direct or indirect detection signals. These selection rules change drastically the complementarity picture of dark matter searches and need to be considered and investigated further. **Table 2** provides a summary of the sensitivity of each dark matter search as a function of the DMsimp and of the spin of the dark matter particle, considered in this review.

Most of DMsims considered in this review have been implemented in FeynRules (Alloul et al., 2014) and are publicly available for download in the repository of the DMsimp framework<sup>9</sup>. DMsims for  $s$ -channel mediators include three different choices for the spin of the WIMP (Dirac fermion, real scalar and complex scalar for spin-0 and spin-1 mediators, and real scalar, Dirac fermion and vector dark matter for spin-2 mediators). Typically, the numerical tools used to compute the dark matter relic density and astrophysical constraints are micrOMEGAs (Belanger et al., 2018) and MadDM (Ambroggi et al., 2018). In the MadGraph\_aMC@NLO platform (Alwall et al., 2011, 2014), one-loop and NLO (next-to-leading order) computations in QCD and EW interactions can be automatically performed in models beyond the SM. This framework allows to compute accurate and precise predictions for production cross sections and distributions of dark matter particles produced at the LHC in association for instance with a mono-jet, mono-photon, mono-Z or mono-Higgs (see e.g., Backović et al., 2015; Mattelaer and Vryonidou, 2015; Arina et al., 2016; Das et al., 2017). It is known that higher order QCD corrections impact not only the production rate but also the shape of the distributions. Most of  $s$ -channel DMsims do include NLO corrections to the matrix elements and parton shower matching and merging. Indeed these higher order terms pertain only to the initial state and originate only from SM processes, hence they can be factorized with respect to the leading order (LO) process accounting for the production of the uncolored mediator and dark matter particles. Conversely the implementation of NLO corrections into  $t$ -channel DMsimp is much more involved, due to the colored nature of the mediator, which do not allow anymore to factorize initial and final state corrections. Typically  $t$ -channel DMsims are LO models, unless stated otherwise. The NLO DMsims [implemented with NLOCT (Degrande, 2015)] are also available at the DMsimp framework webpage<sup>9</sup>. As far as it concerns the DMsimp predictions for relic density, direct and indirect detection, NLO corrections are typically not considered. The automatization of loop-induced, one-loop and NLO processes is currently under development in a future release of MadDM, which is now a MadGraph\_aMC@NLO plugin and hence inherits all its features, including the capabilities of automatically generate the above-mentioned processes for dark matter observables.

As the literature about DMsims is vast, we consider and discuss only a few selected representative papers, whereas we try to be as exhaustive as possible with the references. In the

<sup>9</sup>Dataset (2015). Available online at: <http://feynrules.irmp.ucl.ac.be/wiki/DMsimp>.

**TABLE 2** | Schematic summary of the complementarity of dark matter searches for the DMSimps, taking into account the spins and nature of both mediator and dark matter particles.

Y Spin	DM spin	DD		ID ( $\sigma v$ ) <sub>0</sub>		LHC searches	
		SI	SD	s-channel	t-channel	$\mathcal{E}_T$	Resonance
S	S	OK	NO	helicity suppressed $\propto m_f^2$	s-wave	large $g_{DM} \cdot g_{SM}$	OK
	D	OK	NO	p-wave	p-wave	large $g_{DM} \cdot g_{SM}$	OK
P	D	NO	NO	s-wave	p-wave	large $g_{DM} \cdot g_{SM}$	OK
S-P	D	OK	NO	p-wave	s-wave	large $g_{DM} \cdot g_{SM}$	OK
		(if $g_{SM}$ is large)					
V	S	OK	NO	p-wave	/	OK	OK
	D	OK	NO	s-wave	p-wave	OK	OK
A	D	OK	OK	helicity suppressed	s-wave	OK	OK
	M	NO	OK	helicity suppressed	s-wave	OK	OK
2	S	NA	NO	s-wave	s-wave	large $g_{DM} \cdot g_{SM}$	OK
	F	NA	NA	p-wave	s-wave	large $g_{DM} \cdot g_{SM}$	OK
	V	NA	NA	s-wave	s-wave	large $g_{DM} \cdot g_{SM}$	OK

In the table, S, scalar; P, pseudo-scalar; V, vector; A, axial-vector; F, fermion; D, Dirac; M, Majorana; DM, dark matter; Y, mediator; DD, direct detection; SI, spin-independent; SD, spin-dependent; ID, indirect detection. OK means that the corresponding signal is in the reach of current and near future experiments, while NO means that the predictions are far below the experimental sensitivities, and NA means that there are no actual studies to assess the experimental reach, to the best of our knowledge. The analytic expressions for the annihilation and scattering cross sections can be found (e.g., Lee et al., 2014a; De Simone and Jacques, 2016; Albert et al., 2017a). For each DMSimp, the minimal model is considered, which entitles only couplings between the mediator and the SM quarks, as described in section 3 of this review. The only exception is the spin-2 model, where the mediator communicates with all SM fields.

following sections we provide the interaction Lagrangian for DMSimps we consider and the relevant details for the analyses we review. We take into consideration in general only mediator-quark couplings; couplings to leptons or other SM particles are switched on whenever relevant.

### 3.1. s-Channel Mediator Models

#### 3.1.1. Spin-0 Mediator

The material presented in this section is based on these selected reference papers (Haisch and Re, 2015; Arina et al., 2016; Banerjee et al., 2017), as they nicely exemplify the main features of scalar and pseudo-scalar mediators in the s-channel by performing comprehensive studies of the model, including astrophysical and cosmological dark matter searches.

We focus on the case of Dirac dark matter ( $X$ ), with spin-0 mediator ( $Y_0$ ) coupling to the matter fields of the SM (the dependence on the dark matter spin is briefly summarized in Table 2). The interaction Lagrangians is defined as:

$$\mathcal{L}_X^{Y_0} = \bar{X}(g_{DM}^S + ig_{DM}^P \gamma_5) X Y_0, \quad (5)$$

and

$$\mathcal{L}_{SM}^{Y_0} = \sum_{ij} \left[ \bar{d}_i \frac{y_{ij}^d}{\sqrt{2}} (g_{dij}^S + ig_{dij}^P \gamma_5) d_j + \bar{u}_i \frac{y_{ij}^u}{\sqrt{2}} (g_{uij}^S + ig_{uij}^P \gamma_5) u_j \right] Y_0, \quad (6)$$

where  $d$  and  $u$  denote down- and up-type quarks, respectively, ( $i, j=1,2,3$ ) are flavor indices,  $g_{DM}^{S/P}$  are the scalar/pseudo-scalar WIMP- $Y_0$  couplings. Following the prescriptions of MFV, the couplings of the mediator to the SM particles are proportional to the particle masses and normalized to the SM Yukawa couplings,

$y_{ii}^f = \sqrt{2} m_f / v$  and  $v$  being the Higgs vacuum expectation value, and all flavor off-diagonal couplings are set to zero. This choice of couplings ensures that: (i) the structure of flavor changing neutral current processes of the SM is preserved by the introduction of new physics; (ii) that all flavor violating transitions are governed by the Cabibbo-Kobayashi-Maskawa matrix.

The pure scalar and pure pseudo-scalar mediator scenarios, which we will review in the rest of the section, are given by setting the parameters in the Lagrangians (5) and (6) to:

$$g_{DM}^S \equiv g_{DM} \quad \text{and} \quad g_{DM}^P = 0, \quad (7)$$

$$g_{u_{ii}}^S = g_{d_{ii}}^S \equiv g_{SM} \quad \text{and} \quad g_{u_{ii}}^P = g_{d_{ii}}^P = 0, \quad (8)$$

and

$$g_{DM}^S = 0 \quad \text{and} \quad g_{DM}^P \equiv g_{DM}, \quad (9)$$

$$g_{u_{ii}}^S = g_{d_{ii}}^S = 0 \quad \text{and} \quad g_{u_{ii}}^P = g_{d_{ii}}^P \equiv g_{SM}, \quad (10)$$

respectively. With the simplification of a single universal coupling for the SM- $Y_0$  interactions, the model has only four independent parameters, i.e., two couplings and two masses:

$$\{g_{SM}, g_{DM}, m_{DM}, m_{med}\}. \quad (11)$$

The MFV assumption implies that we can even further neglect the contributions of all quarks but the top-quark in the model, as it has the largest Yukawa coupling. This is certainly an optimal approximation for LHC studies, while dark matter searches are sensitive to all quark flavors. The assumption however that  $g_{SM} \equiv g_{u33}^{S/A}$  provides a good description of the phenomenology of the model, as the inclusion of all other quark flavors has the effect of

globally decreasing the value of  $g_{\text{SM}}$  needed to achieve the same cross section.

The Lagrangians of Equations (5) and (6) induce dimension-five couplings of the mediator to gluons and photons via top-quark loop diagrams. These loop-induced operators are relevant for both astrophysical and collider searches for dark matter. For a scalar  $Y_0$ , the couplings of the mediator to gluons and photons are given, at the leading order, by the effective operators:

$$\mathcal{L}_g^{Y_0} = -\frac{1}{4} \frac{g_g(Q^2)}{v} G_{\mu\nu}^a G^{a,\mu\nu} Y_0 \quad \text{and} \quad \mathcal{L}_\gamma^{Y_0} = -\frac{1}{4} \frac{g_\gamma(Q^2)}{v} F_{\mu\nu} F^{\mu\nu} Y_0, \quad (12)$$

with the effective couplings being

$$g_g(Q^2) = g_{\text{SM}} \frac{\alpha_s}{3\pi} \frac{3}{2} F_S\left(\frac{4m_t^2}{Q^2}\right) \quad \text{and} \quad g_\gamma(Q^2) = g_{\text{SM}} \frac{8\alpha_e}{9\pi} \frac{3}{2} F_S\left(\frac{4m_t^2}{Q^2}\right), \quad (13)$$

where  $Q^2$  denotes the virtuality of the  $s$ -channel resonance, while  $F_S$  is the one-loop form factor

$$F_S(x) = x \left[ 1 + (1-x) \arctan^2\left(\frac{1}{\sqrt{x-1}}\right) \right]. \quad (14)$$

Similar expressions can be retrieved for the pseudo-scalar case (see e.g., Haisch and Re, 2015; Arina et al., 2017). Because of the hierarchy between the strong and the electromagnetic couplings ( $\alpha_s^2/\alpha_e^2 \sim 100$ ), the  $Y_0$  partial width into a pair of gluons is always larger than the one into a pair of photons. The expressions for tree level and loop-induced partial widths are provided in Arina et al. (2016).

Let us first discuss the case of pure scalar  $Y_0$  and summarize briefly all the relevant LHC and dark matter searches to constrain its parameter space:

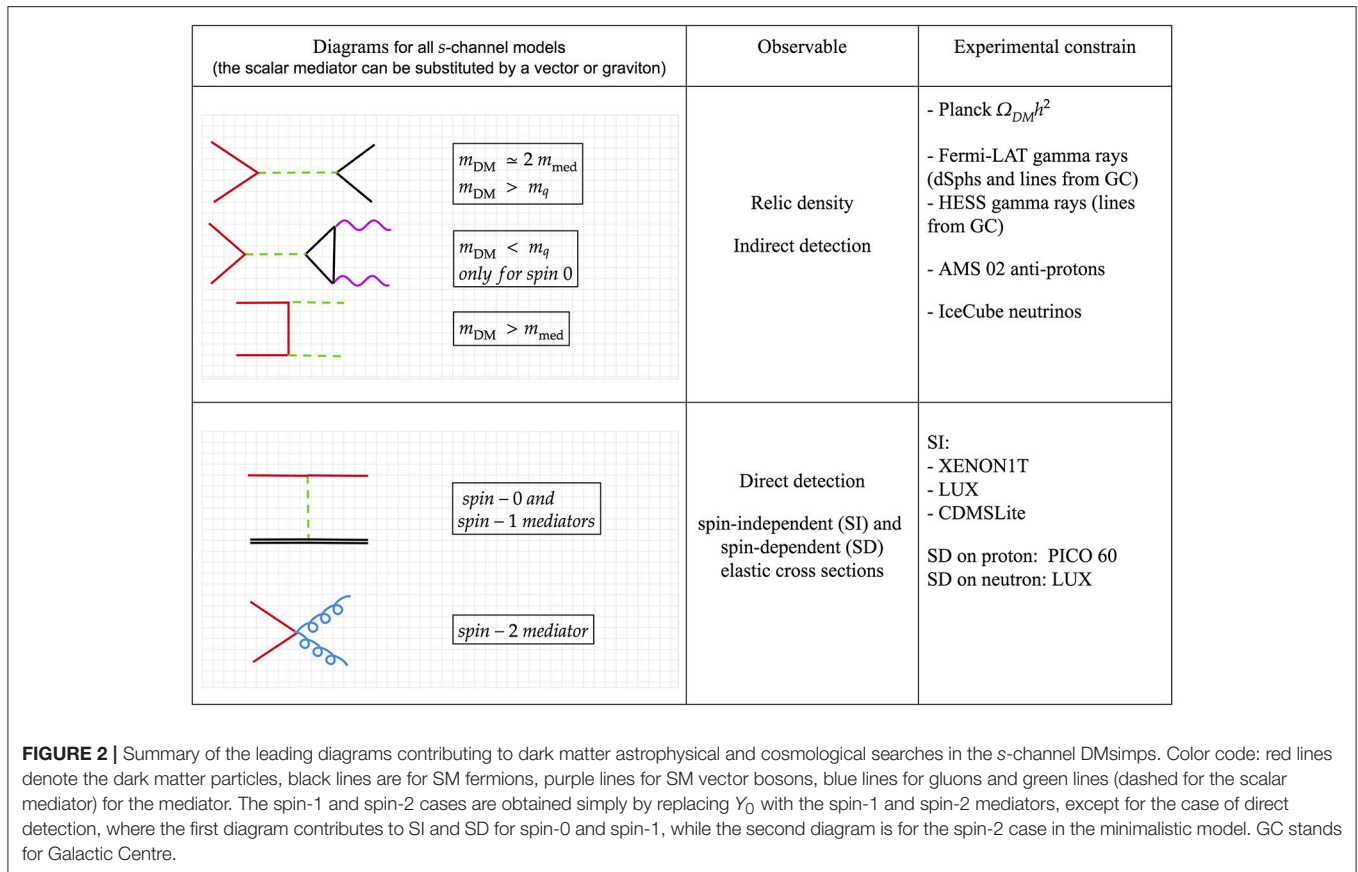
- **LHC  $\cancel{E}_T$  searches.** As this DMsimp features Yukawa-type couplings, the most relevant tree-level process at the LHC is dark matter pair production associated with a top-quark pair (CMS Collaboration, 2014). Similarly to Higgs production, at one loop, gluon fusion gives rise to  $\cancel{E}_T$  + jet signatures (Khachatryan et al., 2015a), mono- $Z$  (Khachatryan et al., 2016) and mono- $h$  (Aad et al., 2016), which are phenomenologically relevant.
- **LHC mediator searches.** The mediator is produced in association with top-quark pairs (Aad et al., 2015a), or via the loop-induced gluon fusion process. These searches are relevant for mediators produced on-shell, or close to on-shell, which decay back into top pairs if kinematically allowed, or a pair of jets (CMS, 2015) or photons (Khachatryan et al., 2015b). For the heavy mediator case, the four-top final state (Khachatryan et al., 2014) can be also relevant.
- **Relic density.** The dark matter achieves the correct relic density in three separated regions. If  $m_{\text{DM}} > m_{\text{med}}$  the relic density is set by the  $t$ -channel annihilation into a pair of mediators. Above the top threshold, resonant annihilation into top-quark pairs is efficient enough to lead to the correct value for  $\Omega_{\text{DM}} h^2$ . For  $m_{\text{DM}} < m_t$  the resonant annihilation into a pairs of gluon leads to the correct relic density for a very fine tuned part of the parameter space. This is due to the very small decay width into gluons.

- **Indirect detection.** All annihilation processes are  $p$ -wave suppressed, hence all fluxes of gamma rays, cosmic rays and neutrinos produced by this model are well below the present and future reach of indirect detection probes.

- **Direct detection.** The interaction Lagrangians in Equations (5) and (6), after some manipulations to express it in terms of nucleus instead of nucleons, reduces to the operator  $\bar{X}X\bar{N}N$ . This is equivalent to the  $\mathcal{O}_1^{\text{NR}}$  operator (see Table 1), which corresponds to the usual spin-independent interaction. The scalar DMsimp is hence highly constrained by the XENON1T and LUX experimental upper bounds.

All the leading order relevant diagrams for  $Y_0$  and dark matter production at the LHC and dark matter annihilation/scattering in astroparticle experiments are summarized in Figures 2, 3.

The result of the comprehensive studies are presented in Figures 4–6, from Haisch and Re (2015), Arina et al. (2016), and Ambroggi et al. (2018), assuming a narrow width approximation. Figure 4 illustrates the mono-jet +  $\cancel{E}_T$  constraints on fixed slices of the model parameter space (red regions). It is clear that mono-jets +  $\cancel{E}_T$  searches constrain the model parameter space for large values of the  $Y_0$ -SM coupling,  $g_{\text{SM}} \geq 3.5$ . The same couplings contribute to the direct detection signal,  $\sigma_n^{\text{SI}} \propto g_{\text{SM}}^2 g_{\text{DM}}^2 / m_{\text{med}}^4$ , and lead to large elastic scattering cross sections, already excluded by LUX (blue solid line). Also shown is the EFT limit, which sets in for heavy mediators. Notice that mono-jets (and mono- $X$ ) +  $\cancel{E}_T$  searches are sensitive to the region  $m_{\text{med}} > 2m_{\text{DM}}$ , where typically the dark matter over-closes the universe, if considered as a pure thermal relic. Figure 5A, illustrates the reach of the  $t\bar{t}$  +  $\cancel{E}_T$  search at 8 TeV, where NLO simulations, that reduce the theoretical errors, are used. Similarly to the case of mono-jets +  $\cancel{E}_T$ , the mediator should be heavier than twice the dark matter mass, in order to be able to decay into invisible states; and the constraints are sensitive to large  $g_{\text{SM}}$  couplings. In Figure 5B we show the behavior of the relic density calculation for a 2D scan over the mass parameters and couplings fixed at 1 [this is one of the benchmark point recommended by the LHC DM working group (Boveia et al., 2016)]. The black line represents the values of masses that achieve the correct  $\Omega_{\text{DM}} h^2$ , the blue region denotes under-abundant dark matter (mostly leaving in the region  $m_{\text{DM}} > m_{\text{med}}$  and dominated by the  $t$ -channel annihilation into mediator pairs), while the gray region stands for over-abundant dark matter (mostly covering the region  $m_{\text{med}} > 2m_{\text{DM}}$ , where  $\cancel{E}_T$  searches are relevant). Figure 6 illustrates a comprehensive parameter space sampling of the model, with the assumption that the dark matter is a thermal relic and constitutes 100% of the matter content of the universe. Couplings are freely varied in between  $10^{-4}$  and  $\pi$ . The relic density measurement rules out completely the region sensitive to  $\cancel{E}_T$  searches, while direct detection disfavors at 90% CL regions with a light mediator for a wide range of  $m_{\text{DM}}$ . Resonance searches are relevant and constrain the region  $m_{\text{DM}} > m_t$ . Di-photons are sensitive to the parameter space  $m_{\text{med}} < 2m_t$ , while the  $t\bar{t}$  and 4 top searches are sensitive to  $m_{\text{med}} > 2m_t$ . A summary of the search sensitivities is provided in Table 2.

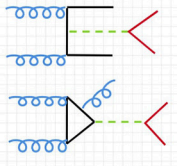
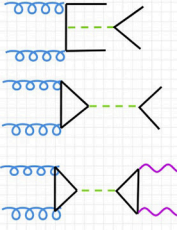
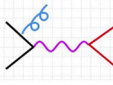
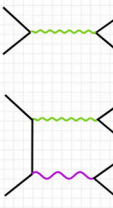
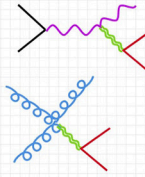
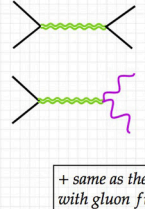


Moving on to the pure pseudo-scalar case (Banerjee et al., 2017), the relevant LHC and dark matter searches are:

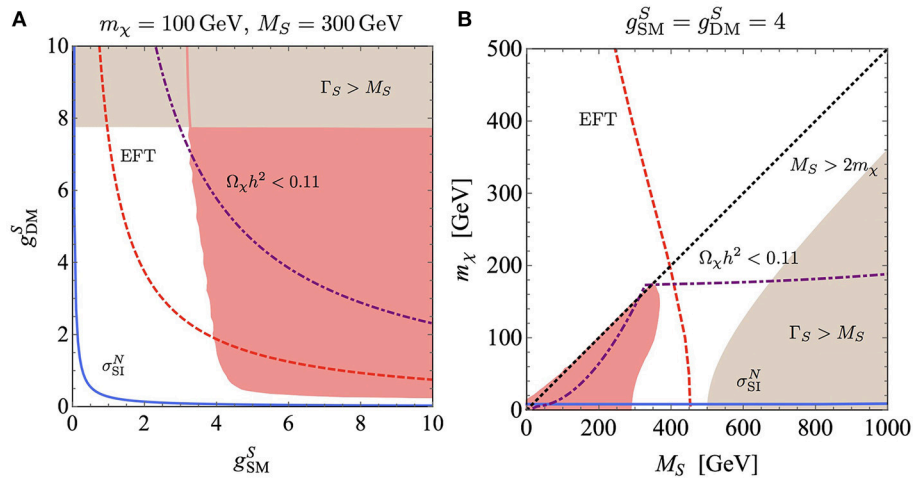
- **LHC  $\cancel{E}_T$  searches.** These are the same as for the scalar case.
- **LHC mediator searches.** These are the same as for the scalar case. By switching on the couplings to leptons, an additional relevant search is the production via gluon fusion or in association with a pair of bottom-quarks, of the mediator decaying into a pair of  $\tau$  leptons ( $A \rightarrow \tau^+ \tau^-$ ) (CMS Collaboration, 2016d). This holds for a scalar  $Y_0$  as well.
- **Indirect detection.** The annihilation channels with  $Y_0$  exchanged in the s-channel are s-wave dominated (i.e.,  $X\bar{X} \rightarrow gg, t\bar{t}$ ), hence the pseudo-scalar mediator model can be constrained by gamma-ray and cosmic-ray searches.
- **Direct detection.** Direct detection is not sensitive to pseudo-scalar mediators. This can be understood by looking at Table 1: the high-energy Lagrangians of the pure pseudo-scalar case are mapped into  $\mathcal{O}_6^{NR}$ . This non-relativistic operator is suppressed by the momentum transfer to the fourth power, hence the current direct detection experiments are insensitive to it, unless the mediator is of the order of the MeV Arina et al. (2015).

The result of the analysis are illustrated in Figure 7 from Banerjee et al. (2017). Figure 7A shows all astrophysical and cosmological

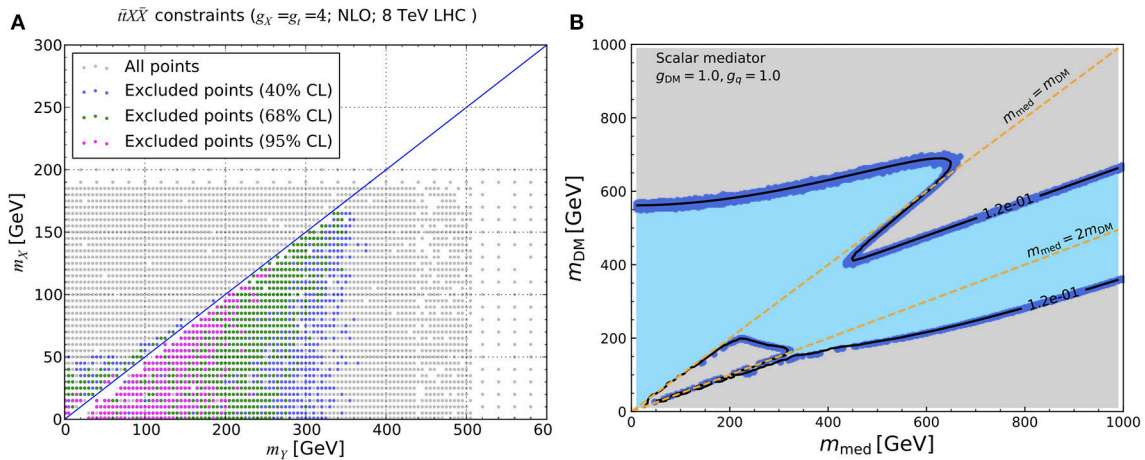
constraints for the dark matter: Fermi-LAT exclusion limits from dSphs are more stringent than both anti-proton bounds (as well as more robust in terms of astrophysical uncertainties) and gamma-ray line searches. Figure 7B shows the most stringent dark matter constraints combined with the LHC searches. A thermal relic scenario lives in the narrow band in between the black and the red solid lines. It is a narrow region because it is dominated by resonant s-channel annihilation, which is fine tuned however occurs in all dark matter models featuring an s-channel mediator.  $\cancel{E}_T$  searches probe a region which is already challenged by the Fermi-LAT dSph constraints. On the other hand, di-photons,  $t\bar{t}$  and  $\tau$  leptons can probe the mediator mass as low as 100 GeV and challenge the left-hand side region where dark matter is a viable thermal relic. The projection for the exclusion bounds coming from the Fermi-LAT satellite after 15 years of operation (red dashed line) shows that these data can basically probe the whole parameter space of the model (everything on the left hand side of the curve is excluded). Notice that additional dark states and mediators can affect the relic density and indirect detection regions. However the changes are supposed to go both in the same directions, hence the region allowed by Planck and Fermi-LAT will remain narrow. LHC bounds for  $m_{\text{med}} < 2m_t$  can change sensibly if additional scalars are introduced, as new decay channels will become available; conversely the constraints for  $m_{\text{med}} > 2m_t$  are robust and will be qualitatively unaltered.

Relevant diagrams for mediator and/or dark matter production at the LHC for all $s$ -channel models	Observable	Experimental search
SPIN-0 MEDIATOR		
 $2 m_{\text{DM}} < m_{\text{med}}$ $2 m_{\text{DM}} < m_{\text{med}}$	Missing energy searches (dark matter production)	<ul style="list-style-type: none"> <li>- <math>t\bar{t}</math> + MET</li> <li>- mono-<math>j</math> + MET</li> <li>- mono-<math>Z</math> + MET</li> <li>- mono-<math>h</math> + MET</li> </ul>
 $m_{\text{med}} > 2 m_t$ $m_{\text{med}} > 2 m_{g,l}$ $m_{\text{med}} < 2 m_{g,l}$ $m_{\text{med}} < 2 m_{\text{DM}}$	Mediator into visible SM particles	Resonance searches: <ul style="list-style-type: none"> <li>- <math>t\bar{t}</math></li> <li>- 4 tops</li> <li>- di-jets</li> <li>- di-leptons</li> <li>- di-photons</li> </ul>
SPIN-1 MEDIATOR		
 $2 m_{\text{DM}} < m_{\text{med}}$	Missing energy searches (dark matter production)	- mono- $j$ + MET
 $m_{\text{med}} > 2 m_{qq}$ $m_{\text{med}} > 2 m_{qq}$ (the associated boson vector decays leptonically)	Mediator into visible SM particles	- di-jets
SPIN-2 MEDIATOR		
 $2 m_{\text{DM}} < m_{\text{med}}$ $2 m_{\text{DM}} < m_{\text{med}}$	Missing energy searches (dark matter production)	<ul style="list-style-type: none"> <li>- mono-<math>j</math> + MET</li> <li>- jets + MET</li> </ul>
 $m_{\text{med}} > 2 m_{ll,qq}$ $m_{\text{med}} > 2 m_{WW,ZZ,hh}$ and $m_{\text{med}} < 2 m_{gg,ll}$ + same as the two diagrams above with gluon fusion instead of Drell – Yan	Mediator into visible SM particles	Resonance searches: <ul style="list-style-type: none"> <li>- di-jets</li> <li>- di-leptons</li> <li>- di-photons</li> <li>- WW</li> <li>- ZZ</li> <li>- hh</li> </ul>

**FIGURE 3 |** Schematic of leading order diagrams contributing to dark matter and mediator searches at the LHC in the  $s$ -channel DMSimps. MET stands for missing transverse energy. The color code is as in **Figure 2**.



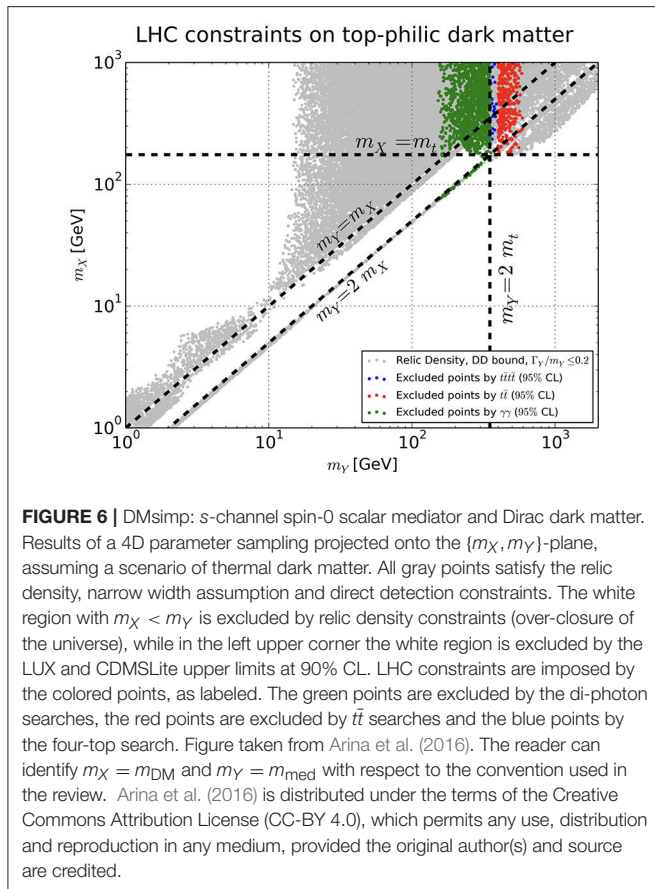
**FIGURE 4 |** DMsimp: s-channel spin-0 scalar mediator and Dirac dark matter. **(A)** Present mono-jet exclusion region at 95% CL (red contour and region within) for scalar mediators in a 2D scan of the parameter space in the  $(g_{\text{DM}}^S, g_{\text{SM}}^S)$ -plane. The fixed values of the two parameters over which the scan is not performed are indicated in each panel. For comparison, we show the region  $\Gamma_S > M_S$  (brown, with  $\Gamma_S$  being the mediator width), the LUX 90% CL exclusion limits on  $\sigma_{\text{SI}}^{\text{SI}}$  (solid blue curve, excludes above and on the right of the curve), the parameter space for under-abundant dark matter ( $\Omega_{\text{DM}} h^2 < 0.11$ , dot-dashed purple line), the EFT limit (red dashed line) and the region for which  $M_S > 2m_\chi$  (black dotted line). **(B)** Same as **(A)** in the  $(m_\chi, M_S)$ -plane. Figures taken from Haisch and Re (2015). The reader can identify  $g_{\text{DM}}^S = g_{\text{DM}}$ ,  $g_{\text{SM}}^S = g_{\text{SM}}$ ,  $m_\chi = m_{\text{DM}}$  and  $M_S = m_{\text{med}}$  with respect to the convention used in the review. Haisch and Re (2015) is distributed under the terms of the Creative Commons Attribution License (CC-BY 4.0), which permits any use, distribution and reproduction in any medium, provided the original author(s) and source are credited.



**FIGURE 5 |** DMsimp: s-channel spin-0 scalar mediator and Dirac dark matter. **(A)** Constraints from the CMS 8TeV  $t\bar{t} + \cancel{E}_T$  analysis in the  $(m_\chi, m_\gamma)$ -plane. The top and dark matter couplings to the mediator are set to 4, as labeled. The next to leading order (NLO) exclusions are shown. Figure taken from Arina et al. (2016). The reader can identify  $m_\chi = m_{\text{DM}}$  and  $m_\gamma = m_{\text{med}}$  with respect to the convention used in the review. **(B)** Dark matter relic density in the  $(m_{\text{DM}}, m_{\text{med}})$ -plane. The gray region denotes over-abundant dark matter, while the light blue region is for under-abundant dark matter. The black solid line/dark blue points denote the parameter space for which the dark matter has the correct relic density. The orange dashed lines stand for  $m_{\text{med}} = m_{\text{DM}}$  and  $m_{\text{med}} = 2m_{\text{DM}}$ , as labeled. The couplings are fixed at the values labeled in the plot. Figure taken from Ambrogio et al. (2018). The reader can identify  $g_q = g_{\text{SM}}$  with respect to the convention used in the review. Arina et al. (2016) is distributed under the terms of the Creative Commons Attribution License (CC-BY 4.0), which permits any use, distribution and reproduction in any medium, provided the original author(s) and source are credited, while (Ambrogio et al., 2018) is distributed with a non-exclusive arXiv license.

Other studies of the spin-0 case are for instance (Buckley et al., 2015; Harris et al., 2015; Dolan et al., 2016; du Pree et al., 2016), while details on loop-induced process for mono-jet + MET can be found in Haisch et al. (2013), Buckley et al. (2015), Harris

et al. (2015), Haisch and Re (2015), and Backović et al. (2015). Leptonic couplings have been introduced in e.g., Albert et al. (2017a). Similarly,  $Y_0$  couplings to the SM gauge bosons are discussed in Neubert et al. (2015).



### 3.1.2. Spin-1 Mediator

The material discussed in this section is based on these selected (Chala et al., 2015; Carpenter et al., 2016; Heisig et al., 2016; Albert et al., 2017a), that exhaustively exemplify the main features of vector and axial-vector mediators in the  $s$ -channel and perform comprehensive studies of the model, including astrophysical and cosmological dark matter searches.

The interaction Lagrangian of a spin-1 mediator ( $Y_1$ ) with a Dirac fermion dark matter particle ( $X$ ) is given by:

$$\mathcal{L}_X^{Y_1} = \bar{X} \gamma_\mu (g_{DM}^V + g_{DM}^A \gamma_5) X Y_1^\mu, \quad (15)$$

and with quarks by:

$$\mathcal{L}_{SM}^{Y_1} = \sum_{ij} \left[ \bar{d}_i \gamma_\mu (g_{dij}^V + g_{dij}^A \gamma_5) d_j + \bar{u}_i \gamma_\mu (g_{uij}^V + g_{uij}^A \gamma_5) u_j \right] Y_1^\mu, \quad (16)$$

where  $g_{DM}^{V/A}$  and  $g_{u/dij}^{V/A}$  are the vector/axial-vector couplings of the dark matter and quarks with  $Y_1$ . For a Majorana dark matter candidate the vector coupling is not allowed.

The pure vector and pure axial-vector mediator scenarios are obtained by setting the parameters in the Lagrangians (15) and (16) to

$$g_X^V \equiv g_{DM} \quad \text{and} \quad g_X^A = 0, \quad (17)$$

$$g_{uii}^V = g_{dii}^V \equiv g_{SM} \quad \text{and} \quad g_{uii}^A = g_{dii}^A = 0 \quad (18)$$

and

$$g_X^V = 0 \quad \text{and} \quad g_{X_D}^A \equiv g_{DM}, \quad (19)$$

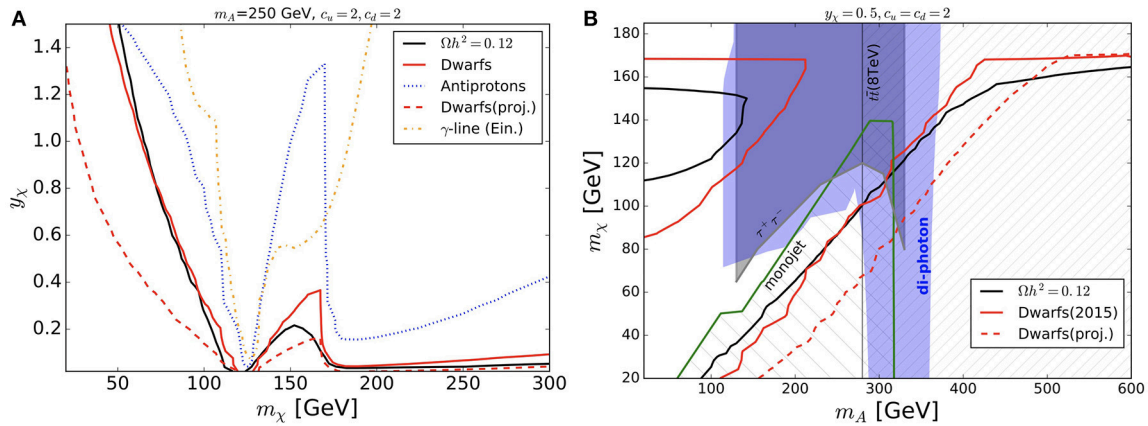
$$g_{uii}^V = g_{dii}^V = 0 \quad \text{and} \quad g_{uii}^A = g_{dii}^A \equiv g_{SM}, \quad (20)$$

respectively, where we assume quark couplings to the mediator to be flavor universal and set all flavor off-diagonal couplings to zero. Similarly to the case of spin-0 mediator, this model has only four free parameters, defined as in Equation (28). The universality assumption of the couplings is also justified by gauge invariance, which sets very tight constraints on the relation among couplings (see e.g., Bell et al., 2015). Even though the Lagrangians presented above do not preserve gauge invariance, the assumption of having different couplings to up- and down-type quarks, as e.g., in Chala et al. (2015), can lead to artificial enhanced cross sections which are not representative of gauge invariant theories.

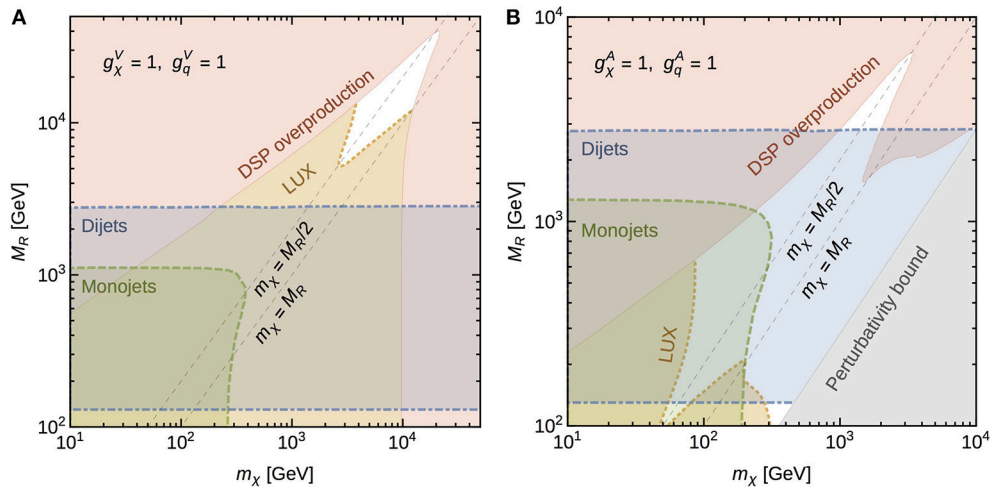
In this model the couplings to leptons are not considered, hence it can be seen as a lepto-phobic  $Z'$  model (see e.g., Duerr and Fileviez Perez, 2015). Leptonic couplings are indeed very tightly constrained by di-lepton resonant searches (Dudas et al., 2009; Arcadi et al., 2014; Lebedev and Mambrini, 2014) and can be switched off to allow to have large quark couplings.

Let us first discuss the complementarity of searches for the case of a pure vectorial  $Z'$  model, hence the dark matter candidate can only be a Dirac fermion (Chala et al., 2015; Carpenter et al., 2016; du Pree et al., 2016).

- **LHC  $\cancel{E}_T$  searches.** ATLAS and CMS searches for jets in association with  $\cancel{E}_T$  (due to initial state radiation of a gluon) place strong constraints on this model (Aad et al., 2015b; Khachatryan et al., 2015a).
- **LHC mediator searches.** The di-jet final state is a very important complementary channel, as it has been pointed out in Chala et al. (2015). Di-jets can be produced via  $Y_1$  Drell-Yan process or via associated production. Stringent bounds for di-jet invariant mass above 1 TeV are provided by ATLAS (ATLAS Collaboration, 2013; Aad et al., 2015c) and CMS (Khachatryan et al., 2015c), while complementary and equally tight bounds for smaller masses are provided by the UA2 (Alitti et al., 1993) experiment and the Tevatron CDF experiment (Aaltonen et al., 2009).
- **Relic density.** The dark matter achieves the correct relic density in a small narrow band for fixed couplings. If  $m_{DM} > m_{med}$  the relic density is set by the  $t$ -channel annihilation into a pair of mediators, which is an  $s$ -wave process proportional to  $g_{DM}^4$ . For  $g_{DM} \sim 1$  this cross section is small and the dark matter is under-abundant. For the benchmark points chosen by the LHC Dark matter working group (Albert et al., 2017a), the correct relic density is achieved by the exchange in the  $s$ -channel of a  $Y_1$ , leading to resonant annihilation into quark pairs, which is also  $s$ -wave. Of course, the introduction of leptonic couplings can change this classification.



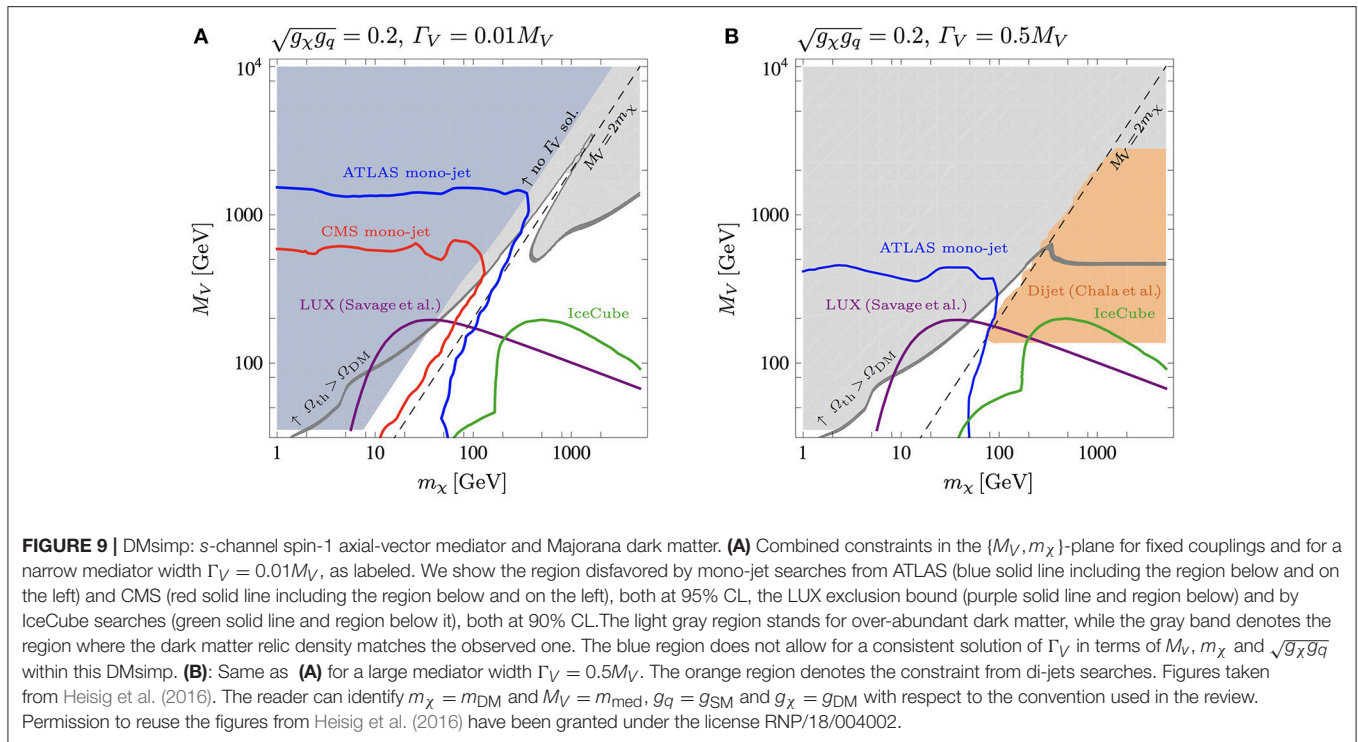
**FIGURE 7 |** DMsimp: s-channel spin-0 pseudo-scalar mediator and Dirac dark matter. **(A)** Dark matter constraints on the model parameter space in the  $\{y_\chi, m_\chi\}$ -plane. The other parameters are fixed as labeled. Below the black line the universe is over-closed, while the region above the red solid line is excluded by the Fermi-LAT dSph gamma-ray searches. The region above the dotted blue line is disfavored by AMS 02 anti-proton measurements, whereas the region above the yellow dot-dashed line is excluded at 95% CL by gamma-ray line searches from the Galactic Center. The red dashed curve denotes the expected sensitivity of the Fermi-LAT searches in dSphs after 15 years of data. **(B)** Dark matter and collider searches presented in the  $\{m_\chi, m_A\}$ -plane. The other parameters are fixed as labeled. If considered as thermal relic the dark matter allowed region is contained in between the red and black solid lines. The shaded regions are excluded by LHC searches at 95% CL: mono-jet (hatched green),  $A \rightarrow \tau^+\tau^-$  (gray), di-photons (blue), and  $t\bar{t}$  (hatched gray). Figures taken from Banerjee et al. (2017). The reader can identify  $m_\chi = m_{\text{DM}}$  and  $m_A = m_{\text{med}}$ ,  $c_u = c_d = g_{\text{SM}}$  and  $y_\chi = g_{\text{DM}}$  with respect to the convention used in the review. Banerjee et al. (2017) is distributed under the terms of the Creative Commons Attribution License (CC-BY 4.0), which permits any use, distribution and reproduction in any medium, provided the original author(s) and source are credited.



**FIGURE 8 |** DMsimp: s-channel spin-1 vector **(A)** and axial-vector **(B)** mediator and Dirac dark matter. **(A)** Combined constraints at 95% CL from the LUX experiment (orange dotted line and orange shaded region), from mono-jet searches (green dashed line and green shaded region) and di-jets (blue dot-dashed line and region in between) in the  $\{M_R, m_\chi\}$ -plane for fixed couplings, as labeled. We also show the region that over-closes the universe (red) and the region excluded by perturbativity (gray). **(B)** Same as **(A)**. The reader can identify  $m_\chi = m_{\text{DM}}$  and  $M_R = m_{\text{med}}$ ,  $g_\chi^{V/A} = g_{\text{DM}}$ , and  $g_q^{V/A} = g_{\text{SM}}$  with respect to the convention used in the review. Figures taken from Chala et al. (2015), distributed under the terms of the Creative Commons Attribution License (CC-BY 4.0), which permits any use, distribution and reproduction in any medium, provided the original author(s) and source are credited.

- **Indirect detection.**  $\langle\sigma v\rangle_0$  receives contributions from the same channels that fix the relic density. For the details on the annihilation cross section we refer to Albert et al. (2017a). However in the literature, at the best of our knowledge, there are no results on constraints from Fermi-LAT dSph gamma-ray searches that include the  $t$ -channel term.

- **Direct detection.** The interaction Lagrangians in Equations (15) and (16) are equivalent to  $\mathcal{O}_1^{\text{NR}}$ , see Table 1. This non-relativistic operator describes the usual spin-independent elastic scattering off nuclei. The vector model is hence highly constrained by the XENON1T and LUX experimental upper bounds.



The leading order relevant diagrams for  $Y_1$  and dark matter production at the LHC and dark matter annihilation/scattering in astroparticle experiments are summarized in **Figures 2, 3** (the same holds for the pure axial-vector mediator).

**Figure 8A**, from Chala et al. (2015), shows the complementarity of collider, cosmological and direct detection searches, with fixed couplings  $g_{SM} = g_{DM} = 1$ . Basically the whole parameter space of the model is strongly disfavored by the current limits of direct detection experiments. Di-jets and mono-jets have a rather smaller impact on the model parameter space. Notice however that, contrary to the case of spin-0 mediator, collider searches are sensitive to smaller values of  $g_{SM}$ , even of the order of  $\mathcal{O}(0.1)$ . Mono-X searches are more sensitive to the region for which  $m_{med} > 2m_{DM}$ , in which the DMsimp features over-abundant dark matter. This assumption can be circumvented by invoking for instance dark matter non thermal production or entropy injection. Conversely, di-jet constraints are rather independent of the dark matter mass and cover all dark matter regions. Constraints from Fermi-LAT dSphs have been discussed in Carpenter et al. (2016): the parameter space of the model is most restricted for  $m_{med} \sim 2m_{DM}$ , because of the enhancement in  $\langle \sigma v \rangle_0$  due to the resonance. If the vector mediator is much heavier than the dark matter, the total annihilation cross section drops and the parameter space becomes suddenly less constrained. This can be understood by the fact that annihilation occurs far away from the resonance, hence  $\langle \sigma v \rangle_0$  decreases quickly.

Moving to the axial-vector case, the dark matter can be either Dirac or Majorana. The most relevant dark matter searches are (Chala et al., 2015; du Pree et al., 2016; Heisig et al., 2016):

- **LHC  $E_T$  searches and mediator searches.** These are exactly the same as in the pure vector case described above.
- **Relic density.** The  $s$ -channel process is helicity suppressed if  $g_{DM}^V = 0$ , namely it is proportional to  $m_q^2$ , while the  $t$ -channel is  $s$ -wave, taken properly into account in the analysis in Albert et al. (2017a).
- **Indirect detection.** In the analyses performed so far there are no bounds from gamma-ray or cosmic-ray searches because the  $t$ -channel process has not been properly taken into account. However, relevant constraints for the model parameter space arise from the IceCube upper limits on  $\sigma_p^{SD}$ , where all annihilation processes contributing to  $\langle \sigma v \rangle_0$  have been properly taken into account.
- **Direct detection.** Spin-independent elastic scattering is superseded by the ordinary spin-dependent elastic scattering (corresponding to  $\mathcal{O}_4^{NR}$  in **Table 1**). This operator is less constrained by direct detection experiments with respect to  $\mathcal{O}_1^{NR}$ . The most constraining experiment is LUX for  $\sigma_n^{SD}$ .

The right panel of **Figure 8** from Chala et al. (2015), describes the complementarity of collider, cosmological and direct detection searches, with fixed couplings  $g_{SM} = g_{DM} = 1$ , for the axial-vector model. The impact of the LUX exclusion limit is rather reduced with respect to the pure vector case. Hence collider bounds have a nice degree of complementarity for this model, disfavoring the majority of the viable parameter space. The gray region is excluded by the perturbativity bound, obtained by imposing  $m_{med} > g_{DM}^4 m_{DM} \sqrt{4\pi}$ , which comes from the requirement that the annihilation cross section remains well-behaved at large dark matter masses. **Figure 9**, from Heisig et al.

(2016), shows the impact of the IceCube bounds on the model parameter space for fixed product of the couplings and for a narrow  $Y_1$  width (**Figure 9A**) and for a large mediator width (**Figure 9B**), as  $\Gamma_{Y_1} \equiv \Gamma_V$  is taken as a free parameter. In the very narrow width approximation, di-jet constraints are irrelevant, while mono-jet +  $\cancel{E}_T$  searches are much less affected by changes in the mediator width. The exclusion bound stemming from LUX does not depend on the mediator width, and remains unchanged in the two cases and constrain the DMSimp parameter space where dark matter is either a thermal relic or under-abundant. The IceCube exclusion limit has a subtle dependence on the mediator width, as the annihilation rate is sensitive to both the  $s$ -channel process, which depends on  $g_{SM} \times g_{DM}$ , and on the  $t$ -channel process, which depends only on  $g_{DM}$ , for  $m_{DM} \geq m_{med}^{10}$ . In **Figure 9A**, IceCube and LUX probe a complementary region, in which  $m_{DM} > m_{med}$ , with respect to LHC searches. LUX constraints are relevant at intermediate dark matter masses, while IceCube lower limits overtake all other constraints at large dark matter masses. In case of a large mediator width, the IceCube bound overlaps with the di-jet constraints. From a refined analysis on di-jets in Fairbairn et al. (2016), it has been shown that for  $m_{med} < 3$  TeV and  $\Gamma_{Y_1} > 0.25 m_{med}$ , the collider constraints disfavor the possibility that the WIMP-quark interactions are responsible for setting the dark matter relic density. A summary of the search sensitivities and their dependency on the dark matter nature is provided in **Table 2**.

The LHC Dark Matter working group has suggested to consider leptonic couplings as well (Albert et al., 2017a). These should be however at least one order of magnitude smaller than the mediator-quark couplings, to not completely exclude the model. Interestingly couplings to neutrinos would also be present because of gauge invariance requirements; these couplings will supply an additional  $\cancel{E}_T$  channel with the consequences of enhancing certain mono- $X$  +  $\cancel{E}_T$  signals.

Other studies of the spin-1 DMSimps are for instance (Buchmueller et al., 2015; Harris et al., 2015; Jacques and Nordstrom, 2015; Bell et al., 2016a; Brennan et al., 2016; du Pree et al., 2016; Fairbairn et al., 2016; Jacques et al., 2016). The latter papers in the list already consider a gauge invariant completion of the  $Z'$  model, instead of the DMSimp Lagrangians in Equations (15) and (16). This issue will be discussed in section 4.

### 3.1.3. Spin-2 Mediator

The material presented in this section is based on these selected (Lee et al., 2014a; Kraml et al., 2017; Zhang et al., 2017), as they exemplify the main features of a spin-2 mediator in the  $s$ -channel as compared with LHC searches and indirect detection searches. The literature on spin-2 mediator is rather reduced with respect to the spin-0 and spin-1 cases. Relevant works are provided by these (Garcia-Cely and Heeck, 2016; Dillon and Sanz, 2017;

Dillon et al., 2017; Rueter et al., 2017; Zhu and Zhang, 2017; Yang and Li, 2018).

Even though the exchange of a graviton in the  $s$ -channel is not considered in the recommendations of the LHC Dark Matter working group (Boveia et al., 2016), it entails several features in common with the DMSimp philosophy. It is possible to build a dark matter simplified model out of a gravity-mediated dark matter model proposed in Lee et al. (2014b), even though it requires a dedicated validation work, as such model is, in general, not renormalizable. This type of models have as well driven a lot of attention at the time of the 750 GeV excess in the di-photon channel (see e.g., Han et al., 2016; Martini et al., 2016; Arun and Saha, 2017) and the references therein.

The definition of the model follows the approach of DMSimps. We consider dark matter particles which interact with the SM particles via an  $s$ -channel spin-2 mediator. The interaction Lagrangian of a spin-2 mediator ( $Y_2$ ) with the dark matter ( $X$ ) is given by Lee et al. (2014b):

$$\mathcal{L}_X^{Y_2} = -\frac{1}{\Lambda} g_X^T T_{\mu\nu}^X Y_2^{\mu\nu}, \quad (21)$$

where  $\Lambda$  is the scale parameter of the theory,  $g_X^T$  is the coupling between  $Y_2$  and the dark matter, and  $T_{\mu\nu}^X$  is the energy-momentum tensor of the dark matter field. The energy-momentum tensors of the dark matter are:

$$T_{\mu\nu}^{X_R} = -\frac{1}{2} g_{\mu\nu} (\partial_\rho X_R \partial^\rho X_R - m_X^2 X_R^2) + \partial_\mu X_R \partial_\nu X_R, \quad (22)$$

$$T_{\mu\nu}^{X_D} = -g_{\mu\nu} (\bar{X}_D i \gamma_\rho \partial^\rho X_D - m_X \bar{X}_D X_D) + \frac{1}{2} g_{\mu\nu} \partial_\rho (\bar{X}_D i \gamma^\rho X_D) + \frac{1}{2} \bar{X}_D i (\gamma_\mu \partial_\nu + \gamma_\nu \partial_\mu) X_D - \frac{1}{4} \partial_\mu (\bar{X}_D i \gamma_\nu X_D) - \frac{1}{4} \partial_\nu (\bar{X}_D i \gamma_\mu X_D), \quad (23)$$

$$T_{\mu\nu}^{X_V} = -g_{\mu\nu} (-\frac{1}{4} F_{\rho\sigma} F^{\rho\sigma} + \frac{m_X^2}{2} X_V \rho X_V^\rho) + F_{\mu\rho} F_\nu^\rho + m_X^2 X_V \mu X_V^\nu \quad (24)$$

where  $F_{\mu\nu}$  is the field strength tensor. We consider three dark matter spins: a real scalar ( $X_R$ ), a Dirac fermion ( $X_D$ ), and a vector ( $X_V$ ). The interaction Lagrangian with the SM particles is:

$$\mathcal{L}_{SM}^{Y_2} = -\frac{1}{\Lambda} \sum_i g_i^T T_{\mu\nu}^i Y_2^{\mu\nu}, \quad (25)$$

where  $i$  denotes the SM fields: the Higgs doublet ( $H$ ), quarks ( $q$ ), leptons ( $\ell$ ), and  $SU(3)_C$ ,  $SU(2)_L$  and  $U(1)_Y$  gauge bosons ( $g, W, B$ ). Following (Ellis et al., 2013; Englert et al., 2013), the phenomenological coupling parameters are defined as:

$$g_i^T = \{g_H^T, g_q^T, g_\ell^T, g_g^T, g_W^T, g_B^T\} \quad (26)$$

without assuming any UV complete model. Notice that the interaction Lagrangian in Equation (25) defines couplings of the graviton with all SM fields. This hypothesis is more generic with respect to the standard assumptions of the DMSimps, where the mediator interacts only with the quark sector. The energy-momentum tensors of the SM fields are similar to Equation (22) and their explicit expression is provided in e.g., Das et al. (2017).

Complying with the DMSimp idea, it is instructive to consider universal couplings between the spin-2 mediator and the SM particles:

$$g_H^T = g_q^T = g_\ell^T = g_g^T = g_W^T = g_B^T \equiv g_{SM}. \quad (27)$$

<sup>10</sup>The exclusion bounds are not rescaled, as the authors assume that the dark matter makes up 100% of the matter content of the universe in the white region. Thermal production is then supplemented by some other mechanism to achieve the observed value of  $\Omega_{DM} h^2$ .

With this simplification, the model has only four independent parameters<sup>11</sup>, two masses and two couplings, as for the other DMsims considered so far:

$$\{m_X, m_Y, g_{DM}/\Lambda, g_{SM}/\Lambda\}. \quad (28)$$

This scenario with a universal coupling to SM particles is realized, e.g., in the original Randall–Sundrum model of localized gravity (Randall and Sundrum, 1999). With this choice of couplings the mediator decays mainly into gluons and light quarks, while the di-photon branching ratio is only  $\sim 5\%$ . The decay into top-quarks or vector bosons is relevant when kinematically allowed. As already discussed in the case of spin-1 mediator, the  $Y_2$ -neutrino coupling leads to  $\cancel{E}_T$  signals that are independent of the decays into dark matter particles and provide additional  $\cancel{E}_T$  channels for the mono-X signals.

In the following, to exemplify the complementarity of dark matter searches, we will focus on vectorial dark matter.

- **LHC  $\cancel{E}_T$  searches.** The  $Y_2$  production is mostly initiated by gluon fusion at low masses, which suppresses mono-photon, mono-Z and mono-W signals, as they can occur only in quark initiated processes. Hence the most constraining missing energy searches for the spin-2 model are a single mono-jet +  $\cancel{E}_T$  (ATLAS Aaboud et al., 2016a) and 2–6 jets +  $\cancel{E}_T$  (ATLAS Aaboud et al., 2016b).
- **LHC mediator searches.** Resonance searches from LHC Run 2 data [ATLAS (ATLAS Collaboration, 2016a,b,c,d,e,f; CMS Collaboration, 2016a) and CMS (CMS Collaboration, 2015, 2016b,c; Khachatryan et al., 2017a,b; Sirunyan et al., 2017a)] give strong constraints on the graviton mass in between few hundreds of GeV and several TeV. The considered final states are  $jj, ll, \gamma\gamma, W^+W^-, ZZ, hh, b\bar{b}, t\bar{t}$ .
- **Relic density.** The dark matter can achieve the correct relic density via the  $s$ -channel exchange of a graviton, especially in the region  $m_{\text{med}} \sim 2m_{\text{DM}}$ , and via  $t$ -channel annihilation into a pairs of  $Y_2$ , which subsequently decay into SM particles, in the region  $m_{\text{DM}} < m_{\text{med}}$ . Both annihilation channels are  $s$ -wave in the case of vectorial dark matter. The analytic expression for these channels are provided in Lee et al. (2014a).
- **Indirect detection.** Annihilation via  $s$ -channel into SM particles with  $Y_2$  exchange can produce both a continuum photon spectrum and gamma-ray lines. Both signals can be constrained by Fermi-LAT and HESS spectral feature searches at the Galactic Centre and by Fermi-LAT dSph exclusion limits. Additionally the  $t$ -channel annihilation process can give rise to box-shaped gamma-ray signatures (see e.g., Ibarra et al., 2015a), which are however only poorly constrained by Fermi-LAT searches for spectral features toward the Galactic Centre (Lee et al., 2014a).
- **Direct detection.** The WIMP-gluon interaction is relevant for direct searches: this coupling generates a twist-2 operator which induces a spin-independent cross-section dark matter-nucleon. This cross section can be in tension with the

XENON1T for dark matter masses below roughly 400 GeV (see Chu et al., 2012; Lee et al., 2014b) for the case of scalar dark matter. However, we couldn't find a dedicated analysis illustrating how direct detection impacts the whole DMsimp spin-2 parameter space. The elastic cross section WIMP-nuclei can receive additional contributions in non minimalistic models (Lee et al., 2014a).

The diagrams for dark matter annihilation are illustrated in **Figure 1**, while the mediator production at the LHC is shown in **Figure 3**.

At present, to the best of our knowledge, there are actually no comprehensive studies of spin-2 models, which encompass both LHC and dark matter searches, except for Zhang et al. (2017). We however refrain from using their results to illustrate the main features of this model, as they perform a random scan of the full parameter space. While this is certainly instructive, it is not necessarily the most optimal pedagogical approach to begin with. For the sake of the discussion, we choose to show 2D parameter scans, even though they do not show the complementarity of searches.

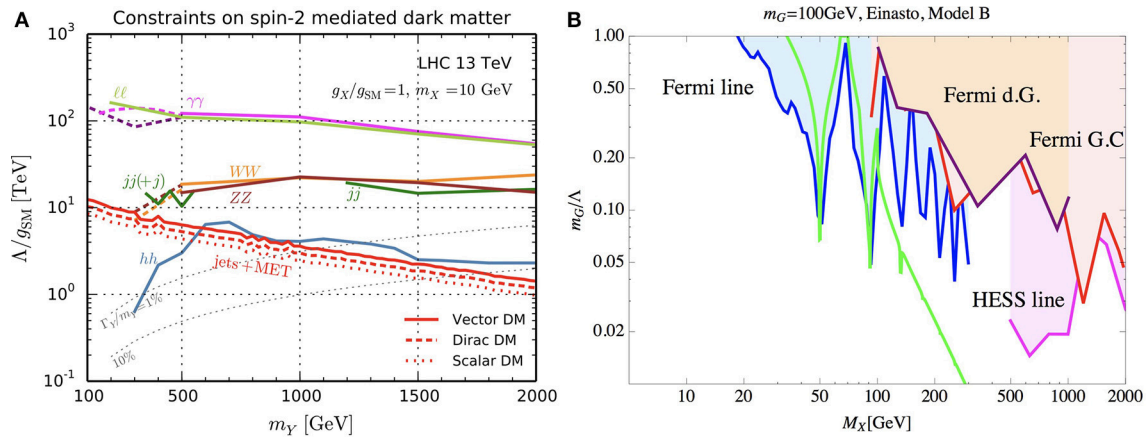
**Figure 10**, from Lee et al. (2014a); Kraml et al. (2017), resumes the constraints on slices of the DMsimp parameter space stemming from LHC searches for a massive graviton, **Figure 10A**, and the dark matter gamma-ray searches, **Figure 10B**. From **Figure 10A**, we clearly see that the di-photon and the di-lepton resonance searches provide the most stringent limit in the whole mediator mass range, constraining  $\Lambda/g_{\text{SM}} > 100$  TeV for graviton masses below 1 TeV. These searches are rather independent on the exact dark matter mass value. Mono-jets +  $\cancel{E}_T$  searches become competitive for large values of  $g_{\text{SM}}$  and, if the  $Y_2$  decays into  $\gamma\gamma$  and  $ll$ , are heavily suppressed. In **Figure 10B**, we show the impact of gamma-ray searches. For  $m_{\text{DM}} < m_{\text{med}}$ , the exclusion limits from gamma-ray lines provided by Fermi-LAT disfavor at 95% CL the model parameter space compatible with the thermal relic assumptions, as the dark matter annihilates mainly into  $gg$  and  $\gamma\gamma$ . For  $m_{\text{DM}} > m_{\text{med}}$  the thermal relic scenario is compatible with Fermi-LAT dSph upper limits and with the HESS gamma-ray line searches, which are the most sensitive constraints for large dark matter masses.

Fermionic and scalar dark matter particles are more loosely constrained by current gamma-ray searches with respect to vectorial dark matter particles, as  $\langle\sigma v\rangle_0$  is suppressed by  $p$ -wave or  $d$ -wave respectively. LHC constraints are less sensitive to the dark matter spin. The sensitivity to the dark matter particle nature depends on the hierarchy between  $g_{\text{DM}}$  and  $g_{\text{SM}}$ : for  $g_{\text{DM}} \sim g_{\text{SM}}$  only jets +  $\cancel{E}_T$  searches can differentiate among the spin of the dark matter candidate; for  $g_{\text{SM}} \gg g_{\text{DM}}$  all searches become sensitive to the dark matter nature. It turns out that the vectorial case is the most constrained model, while the scalar DMsimp is the less constrained and the fermionic case lies in between.

### 3.2. $t$ -Channel Mediator Models

In this section we discuss the phenomenology of  $t$ -channel DMsims and their current state of art with respect to the experimental situation.  $t$ -channel models couple directly the dark

<sup>11</sup>We have dropped the superscript  $T$  for simplicity.



**FIGURE 10 |** DMSimp:  $s$ -channel spin-2 mediator. **(A):** Summary of the 13 TeV LHC constraints in the  $(\Lambda/g_{\text{SM}}, m_Y)$ -plane. The other parameters are fixed as labeled. The differences among the different dark matter spins is not visible in the limits from the resonance searches (as labeled in the plots), conversely to the case of jets +  $\cancel{E}_T$  searches (red lines as labeled). Regions below each line are disfavored at 95% CL. Information on the mediator width-to-mass ratio is provided by the gray dotted lines. Figure taken from Kraml et al. (2017). The reader can identify  $m_Y = m_{\text{med}}$ ,  $g_X = g_{\text{DM}}$  and  $m_X = m_{\text{DM}}$  with respect to the convention used in the review. **(B):** Gamma-ray bounds from Fermi-LAT (d.G., line, G.C.) and HESS telescope (lines) are shown in case of vector dark matter in the  $(m_G/\Lambda, M_X)$ -plane, for a fixed graviton mass as labeled. Couplings are not universal, but fixed at  $g_X = 1$ ,  $g_Y = g_g = g_\gamma = 0.3$  and  $g_h = 0$ , and  $m_G/\Lambda$  corresponds to the  $Y_2$ -WIMP coupling  $g_{\text{DM}}$ . The green line corresponds to the region of parameter space achieving the correct  $\Omega_{\text{DM}} h^2$ . Figure taken from Lee et al. (2014a). The reader can identify  $M_X = m_{\text{DM}}$  and  $m_G = m_{\text{med}}$  with respect to the convention used in the review, while d.G. and G.C. stand for dSphs and Galactic Center respectively. Both Kraml et al. (2017) and Lee et al. (2014a) are distributed under the terms of the Creative Commons Attribution License (CC-BY 4.0), which permits any use, distribution and reproduction in any medium, provided the original author(s) and source are credited.

matter sector with the SM fermions (primarily quarks), leading to a different phenomenology with respect to  $s$ -channel models. The fields in the dark sector are both odd under a  $Z_2$  symmetry to ensure the stability of the dark matter candidate, while in  $s$ -channel models the mediator is usually assumed to be even under the  $Z_2$  symmetry<sup>12</sup>. As a consequence, LHC searches are always characterized by  $\cancel{E}_T$  signals, as the mediator is produced each time in combination with a dark matter particle. In order to connect the dark matter via  $t$ -channel with SM quarks there are two main possibilities: scalar dark matter and fermionic mediator, or fermionic dark matter and scalar mediator. The dark matter cannot have color charge, hence the mediator has to be colored. Additionally, to comply to MFV, either the mediator or the dark matter should have a flavor index. Here we assume to be the former case. For uncolored mediator models see (Garny et al., 2015), while for flavored dark matter we refer to Agrawal et al. (2012, 2014); Kile (2013). From the point of view of QCD corrections, the  $t$ -channel and  $s$ -channel models are very different, as in the former the mediator can be either neutral or colored, rendering more involved the treatment of NLO corrections. This has not been yet fully investigated in the literature, due to its complexity.

Among the vast literature on  $t$ -channel models (see e.g., Blumlein et al., 1997; Cao et al., 2009; Bell et al., 2011, 2012, 2016b; Barger et al., 2012; Bai and Berger, 2013, 2014; Garny et al., 2013, 2014, 2018; Giacchino et al., 2013, 2014, 2016; Toma, 2013; An et al., 2014; Chang et al., 2014a,b; Ibarra

et al., 2014, 2015b; Papucci et al., 2014; Yu, 2014; Abdallah et al., 2015; Abercrombie et al., 2015; Brennan et al., 2016; Bringmann et al., 2016; Carpenter et al., 2016; De Simone and Jacques, 2016; Goyal and Kumar, 2016; El Hedri et al., 2017), we choose to present the results obtained in Colucci et al. (2018) for the case of scalar dark matter and fermionic mediator, which is the most updated analysis at the time of writing. For the case of fermionic dark matter and scalar mediator we discuss the results presented in Garny et al. (2015), which is a comprehensive review paper focusing on  $t$ -channel simplified models alike the supersymmetric one.

Let us first discuss the case of scalar dark matter candidate  $S$  and a vector-like fermionic mediator  $T$ . We assume that the dark matter is a  $SU(2)_L$  singlet, hence it cannot couple at tree level with the weak gauge bosons. Consequently the dark matter hyper-charge is zero, in order to obtain an electrically neutral particle. The dark matter can couple to either right-handed or left-handed SM fermions. Here we assume a couplings with right-handed quarks, in particular only with the third generation. The main reason is dictated by the fact that right-handed couplings to quarks play a major role for the LHC and direct detection phenomenology and the Yukawa of the top is the largest coupling. The mediator  $T$  should be a color triplet, have opposite hyper-charge with respect to the right-handed quarks and be a singlet under  $SU(2)_L$ . The interaction Lagrangian between WIMPs and the SM quarks is then given by Colucci et al. (2018):

$$\mathcal{L}_S^T = y \bar{S} T P_R t + \text{h.c.}, \quad (29)$$

<sup>12</sup>In DMSimp  $s$ -channel models, the mediator cannot be odd under the  $Z_2$  otherwise for  $m_{\text{med}} < m_{\text{DM}}$  it would be playing the role of dark matter candidate.

where  $P_R$  is the right-handed chirality projector, and we have neglected the quartic term connecting the dark matter particle with the SM Higgs doublet, in the spirit of DMsimp construction. With these assumptions the model has only three free parameters:

$$\{m_{\text{DM}}, m_{\text{med}}, \gamma\}. \quad (30)$$

This model considers top-philic dark matter, which might seem more *ad-hoc* than a generic framework where the dark matter couples to all generations. However, this is enough to comprehend all the relevant phenomenology, as in the limit  $m_{\text{DM}} > m_t$ , the results are strictly equivalent as for the case in which the dark matter couples to the light quark (or lepton) generations only. Moreover, at energies comparable with the top mass, the computation of QCD and bremsstrahlung corrections are much more involved than in the chiral limit, hence it is relevant to have the most general framework where to treat them. We are not providing any detail on this part and refer to (Bringmann et al., 2017; Colucci et al., 2018) the interested reader. Notice that if the dark matter was coupled to all three quark generations with three different vector-like fermionic mediators, MFV requirements would enforce the three mediator masses to be equal, as well as their couplings with WIMPs and quarks.

The dark matter constraints for this model are:

- **LHC searches with  $\cancel{E}_T$ .** There are two types of searches particularly relevant for this model: (i) supersymmetric searches of scalar top partners [LEP (Abbiendi et al., 2002) and LHC (CMS Collaboration, 2017; Sirunyan et al., 2018)], recasted to constrain the vector-like fermionic mediator of the model, which is strongly interacting and leads to mainly  $t\bar{t} + \cancel{E}_T$  signals; (ii) the usual dark matter searches characterized by a mono-jet +  $\cancel{E}_T$  (Aaboud et al., 2016a, 2018; Sirunyan et al., 2017b,c) (actually the most updated mono-j searches do include more than one hard jet). NLO QCD corrections and matching with the parton showers have been taken into account, in order to comply with the state-of-art modeling of the LHC signals for the  $s$ -channel case.
- **Relic density.** There are several annihilation processes contributing to  $\Omega_{\text{DM}} h^2$ , depending on the model parameter space region. For  $m_{\text{DM}} \gg m_t$  the chiral limit is valid and virtual internal bremsstrahlung (VIB) adds a significative contribution to the tree level leading order  $t$ -channel diagram, which is helicity suppressed and the first non-zero term depends on  $v^4$  ( $d$ -wave). Decreasing  $m_{\text{DM}}$  just above the top threshold, the tree level  $t$ -channel diagram, which is  $s$ -wave, is the leading contribution to  $\langle \sigma v \rangle_0$ . Below the top mass, loop-induced processes into  $\gamma\gamma$  and  $gg$  can play a role (similarly to the spin-0 top-philic dark matter presented in section 3.1.1), while, for  $m_{\text{DM}} \lesssim m_t$  the off-shell decay  $t^* \rightarrow Wb$  is relevant. Additionally, if the dark matter and the mediator masses are close in mass (within 10%) co-annihilation between  $S$  and  $T$  is also relevant, as well as  $T$  annihilations.
- **Indirect detection.** For  $m_{\text{DM}} < m_t$ , annihilation via the loop induced process into pairs of gluons dominates. This leads to a prompt photon spectrum. The  $\gamma\gamma$  final state is

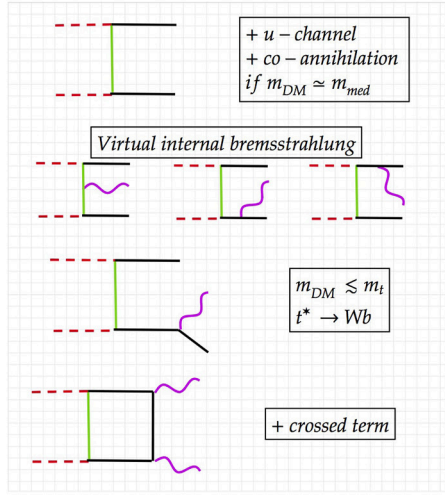
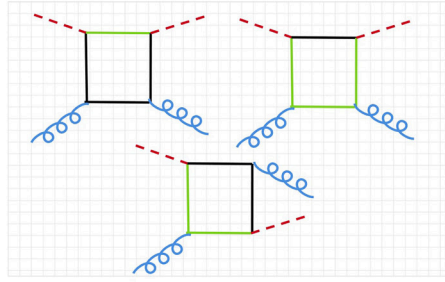
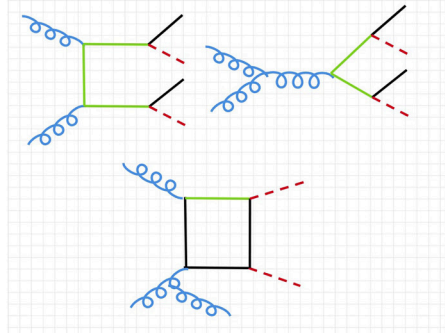
subdominant with respect to the  $gg$  final state as already discussed in section 3.1.1, however it gives rise to box-shaped gamma-ray signals (the width of the box depends on the mass hierarchy between  $S$  and  $T$ : if they are quasi degenerate the box is very narrow, otherwise it is a wide box). For  $m_{\text{DM}} \sim m_t$ , the dominant annihilation channel is the tree level  $t$ -channel,  $SS \rightarrow t\bar{t}$ , which leads to a continuum spectrum of prompt photons, detectable by the Fermi-LAT dSph searches. The same process can be constrained with the anti-proton data released by AMS 02. VIB with the emission of a photon, a gluon or a weak boson, has been demonstrated to be the dominant contribution in the chiral limit ( $m_{\text{DM}} \gg m_t$ ) (see e.g., Bell et al., 2011; Giacchino et al., 2013, 2014, 2016; Bringmann et al., 2016, 2017). The emission of an additional vector boson lifts the helicity suppression and gives rise to sizeable  $\langle \sigma v \rangle_0$ . If  $S$  and  $T$  are nearly degenerate in mass, the  $SS \rightarrow t\bar{t}\gamma$  process dominates the VIB contribution. This photon emission gives rise to a sharp spectral feature, that can be constrained with current gamma-ray line searches. Indeed, the present telescope resolution does not allow to discriminate among the sharp edge due to VIB or a true gamma-ray line (Garny et al., 2015). Direct annihilation of the dark matter into photon pairs via box diagram is on the same foot as VIB. On the other hand the annihilation process  $SS \rightarrow t\bar{t}g$  contributes to the continuum photon spectrum.

- **Direct detection.** An effective coupling WIMP-gluons generates a spin-independent contribution to the elastic scattering cross section, which is, for  $m_{\text{DM}} < m_t$ , in tension with the XENON1T bound. Conversely for  $m_{\text{DM}} > m_t$ ,  $\sigma_n^{\text{SI}}$  is negligible and below the neutrino background (Billard et al., 2014).

The relevant diagrams contributing to all dark matter searches in this DMsimp are shown in **Figure 11**, while the dependency on the dark matter spin is summarized in **Table 2**.

The results of the comprehensive dark matter study are illustrated in **Figure 12A**. Under the assumption that the dark matter is a thermal relic, the complementarity of dark matter searches is clearly shown in the plot. Direct detection experiments probe the region for  $m_{\text{DM}} < m_t$ , while Fermi-LAT, HESS and AMS 02 are sensitive to a mass range from roughly  $m_t$  up to 500 GeV. This shows that anti-matter constraints can be competitive with gamma-ray searches, modulo the larger astrophysical uncertainties. LEP searches constrain the most lightest values of  $m_{\text{DM}}$ , while CMS searches cover a parameter space orthogonal to indirect detection. In particular multi-jets +  $\cancel{E}_T$  searches loose quickly sensitivity with the increase of the dark matter mass, however the  $t\bar{t} + \cancel{E}_T$  searches are effective in the regime where the decay  $T \rightarrow St$  happens far from threshold. Notice that if the decay channel  $T \rightarrow St$  is closed, the mediator becomes long-lived. This case requires further dedicated studies.

The Majorana dark matter DMsimp exhibits only few differences with respect to the scalar dark matter model presented above. We summarize here the most important. Under the same assumptions made for the fermionic mediator, the interaction Lagrangian with only a single generation of light

Diagrams for $t$ -channel model (case of scalar dark matter and fermionic mediator)	Observable	Experimental constrain
 <p>+ <math>u</math>-channel + <math>co</math>-annihilation if <math>m_{DM} \approx m_{med}</math></p> <p>Virtual internal bremsstrahlung</p> <p><math>m_{DM} \lesssim m_t</math> <math>t^* \rightarrow Wb</math></p> <p>+ crossed term</p>	<p>Relic density (first three rows)</p> <p>Indirect detection (first, second rows and box diagrams)</p>	<p>- Planck <math>\Omega_{DM} h^2</math></p> <p>- Fermi-LAT gamma rays (dSphs and lines from GC)</p> <p>- HESS gamma rays (lines from GC)</p> <p>- AMS 02 anti-protons</p>
	<p>Direct detection spin-independent (SI) elastic cross section</p>	<p>SI: - XENON1T - LUX</p>
	<p>Missing energy searches</p>	<p>- <math>t\bar{t} + MET</math> - <math>mono-j + MET</math></p>

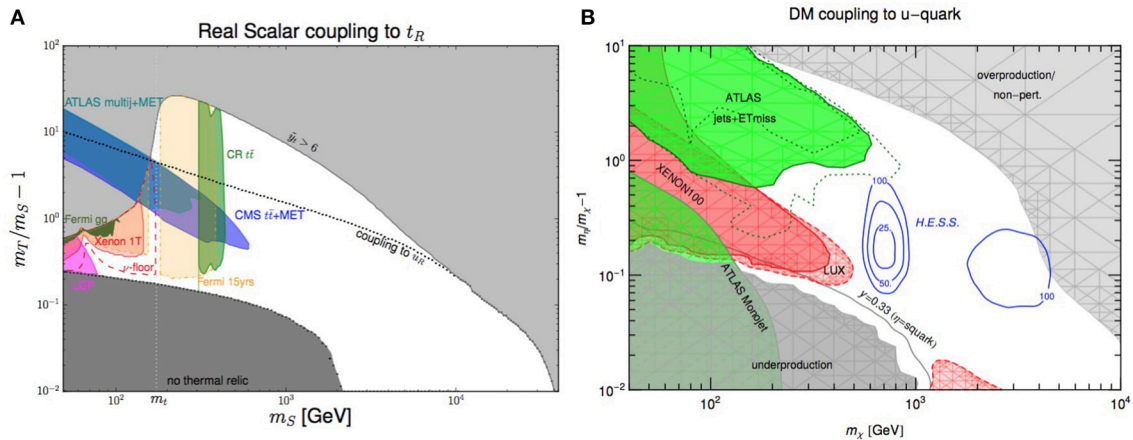
**FIGURE 11 |** Schematic of leading order diagrams contributing to all dark matter searches in the  $t$ -channel DMsims with scalar dark matter and fermionic mediator. The diagrams contributing to LHC searches are specifically drawn for the case of the top-philic model discussed in section 3.2 [for generic fermionic mediator the reader is referred to De Simone and Jacques (2016) and the references therein]. The case of Majorana dark matter and scalar mediator is easily obtained from the above diagrams. For fermionic dark matter there is an additional spin-dependent contribution to the direct detection elastic scattering cross section. MET stands for missing transverse energy. The color code is as in Figure 2.

quark (considering the model in Garny et al., 2015) is given by:

$$\mathcal{L}_S^T = y \tilde{T}^* \bar{X} P_R q + \text{h.c.}, \quad (31)$$

where now the dark matter field is denoted by  $X$  and the mediator by  $\tilde{T}$  and  $q$  is the light quark, which we assume to

be the  $u$  flavor for concreteness for the rest of the section. This Majorana model is very close to the simplified model considered in supersymmetric searches at the LHC, as it is implemented in the Minimal Supersymmetric Model with only light quarks and the neutralino, except that the coupling  $y$  is not fixed at the weak scale but can be varied freely.



**FIGURE 12 |** DMsimp:  $t$ -channel. **(A)** Combined constraints from direct and indirect detection and collider searches in the  $(\{m_T/m_S - 1\}, m_S)$ -plane. The gray regions stand for either under-abundant or for over-abundant dark matter for fermionic mediator and scalar WIMPs. The red region is excluded at 90% CL by XENON1T, while the red dashed line indicates the region of parameter above the neutrino floor hence detectable by direct detection. The green regions are excluded by Fermi-LAT dSph constraints on gamma rays and by anti-protons (Cuoco et al., 2018). The orange region denotes the expected sensitivity of 15 years of data taking by Fermi-LAT. The magenta and blue regions show constraints on scalar top production at LEP and at the LHC, while mono- $X + \cancel{E}_T$  searches disfavor the dark blue region at 95% CL. Figure taken from Colucci et al. (2018). The reader can identify  $m_T = m_{\text{med}}$  and  $m_S = m_{\text{DM}}$  with respect to the convention used in the review. **(B)** Same as **(A)** for Majorana dark matter and scalar mediator. The color code is: gray regions denote under- and over-abundant dark matter, while the green regions are disfavored at 95% CL by jet(s) searches +  $\cancel{E}_T$ . The red regions are excluded by XENON100 and LUX at 90% CL. The blue contour levels show the ratio between the excluded annihilation cross section and the thermal cross section. The regions inside these contour lines are disfavored if the gamma-ray flux from dark matter annihilation is enhanced relative to the Einasto profile (Einasto and Haud, 1989) by the corresponding factor (more cuspy profiles or presence of substructures). Figure taken from Garny et al. (2015). The reader can identify  $m_X = m_{\text{DM}}$  and  $m_\eta = m_{\text{med}}$  with respect to the convention used in the review. Both Colucci et al. (2018) and Garny et al. (2015) are distributed under the non-exclusive arXiv license.

- **LHC searches with  $\cancel{E}_T$ .** As for the fermionic dark matter case, the most stringent searches are given by jets +  $\cancel{E}_T$  and mono-jet +  $\cancel{E}_T$ . The first arises from the direct production of the colored mediator that further decays into the dark matter and light quarks. The latter stems from the loop-induced production of a dark matter pair that recoils against a jet.
- **Relic density.** The annihilation processes contributing to the dark matter relic density are analogous to the case of scalar dark matter in the chiral limit. In this limit, the  $XX \rightarrow u_R u_R$  process is  $p$ -wave suppressed.
- **Indirect detection.** This is completely analogous to the scalar dark matter case in the chiral limit, except that the  $t$ -channel tree level annihilation diagram is still helicity suppressed and the first non zero term in the chiral limit is  $p$ -wave. The authors in Garny et al. (2015) consider as well the exclusion limits on  $\sigma_p^{\text{SD}}$  stemming from IceCube. The vector boson generating the neutrino flux from the Sun arises from VIB:  $XX \rightarrow u_R u_R V$  and subsequently  $V = W, Z, h$  shower and hadronise and produce a continuum spectrum for the neutrinos. The IceCube bounds are less performant than direct detection searches to constrain the model parameter space, hence are not shown in the following.
- **Direct detection.** With respect to the diagrams shown in Figure 11, the elastic cross section off nucleus receives an additional contribution from the  $s$ -channel exchange of  $\tilde{T}$ , which is not present in the scalar dark matter case. There is a small contribution from the spin-independent operator, while the leading contributions to the elastic cross section are proportional to a combination of  $\mathcal{O}_4^{\text{NR}}$ ,  $\mathcal{O}_8^{\text{NR}}$ , and  $\mathcal{O}_9^{\text{NR}}$ , which

are spin-dependent operators. Still the most constraining bounds on the model come from spin-independent limits from the LUX experiment, as these are orders of magnitude more sensitive than the spin-dependent upper bounds.

The results of the comprehensive dark matter study are illustrated in Figure 12B. The picture is rather similar to the case of scalar dark matter, assuming a thermal dark matter scenario. Constraints from jets +  $\cancel{E}_T$  are relevant for large mass splitting between  $X$  and  $\tilde{T}$ , because after its production the mediator has a larger phase space for its decay into the dark matter and the light quark, leading to harder jets. Direct detection is sensitive to smaller mass splitting, while mono-jet +  $\cancel{E}_T$  searches are sensitive to the quasi degenerate region. Gamma rays probe the model parameter space in the intermediate  $m_{\text{DM}}$  mass range.

The Dirac dark matter DMsimp is different from the Majorana case reported above, as far as it concerns the dark matter studies. The elastic scattering cross section is dominated by spin-independent because of the contribution from vectorial currents, which are null in case of Majorana fermions. Hence, thermal Dirac dark matter models get strongly constrained by current direct detection experiments, which combined with LHC searches, completely disfavor the thermal hypothesis for such model (see De Simone and Jacques, 2016) for details.

As a concluding remark, in general a colored  $t$ -channel mediator scenario will be probed to a large extent by next generation experiments, assuming thermal dark matter production and perturbativity of the coupling.

## 4. CAVEATS OF DARK MATTER SIMPLIFIED MODELS

DMsims represent an improvement with respect to the use of EFT for collider dark matter searches during LHC Run 1. However, most of them are still considered not the ideal benchmark models over which categorize the dark matter searches and their complementarity. The main reason of concern is related to the fact that most of the DMsims are not gauge invariant, thus not renormalisable (see e.g., Bell et al., 2015, 2016a,b, 2017a; Englert et al., 2016; Haisch et al., 2016; Kahlhoefer et al., 2016). The most striking example is provided by the spin-1 mediator with axial-vector couplings to fermions. The interaction Lagrangians provided in section 3.1.2 are not gauge invariant unless a dark Higgs is introduced to give mass to the  $Z'$  mediator, and hence to unitarise its longitudinal component. As a consequence, the most minimalistic self consistent model would feature two mediators, an additional scalar along with the  $Z'$ . The presence of a second mediator would change the phenomenology of the model, which is not anymore well described by the single mediator assumption.

The  $s$ -channel scalar mediator case is not gauge invariant unless  $Y_0$  mixes with the Higgs boson, because the dark matter is a singlet under the SM gauge symmetries (see e.g., Lopez-Val and Robens, 2014; Khoze et al., 2015; Baek et al., 2016; Costa et al., 2016; Dupuis, 2016; Robens and Stefaniak, 2016; Wang et al., 2016; Balazs et al., 2017). The mixing with the Higgs boson introduces a major modification in the building of the next generation of DMsims, as the model parameter space then becomes constrained by measurements of the Higgs properties. This has motivated two types of scenarios: (i) models that communicate with the SM via the Higgs portal through the mixing parameters, or even models for which the scalar mediator is the Higgs itself; (ii) to avoid the tight constraints stemming from Higgs physics,  $Y_0$  mixes with an additional doublet similarly to a two Higgs doublet model. Likewise, pseudo-scalar DMsims (Goncalves et al., 2017) can be made theoretically consistent by promoting them to double mediator models. Two Higgs doublet models are well motivated theoretically, arising in several UV complete models such as supersymmetry, or other extensions of the SM (see e.g., Fayet, 1976; Gunion and Haber, 1986; Amaldi et al., 1991; Carena et al., 1996; Branco et al., 2012; Bhattacharyya and Das, 2016) and the references therein.

We will not discuss more in details here these issues and the proposed solutions. There is already a quite vast literature along the lines of the two Higgs doublet models and Higgs portals. The interested reader is referred to e.g., (Boveia et al., 2016; De Simone and Jacques, 2016; Duerr et al., 2016; Albert et al., 2017b; Baek et al., 2017; Bauer et al., 2017; Bell et al., 2017b; Ellis et al., 2017; Ko et al., 2017).

## 5. FUTURE PROSPECTS

The focus of this review has been to describe the state-of-art of dark matter simplified models, as defined by the LHC

Dark Matter Working group, with respect to the current dark matter searches. In particular we have discussed the degree of complementarity of LHC searches (mono-jet +  $\cancel{E}_T$ , jets +  $\cancel{E}_T$ , resonance searches), dark matter direct and indirect detection searches (gamma-ray, anti-matter, and neutrino searches) in several scenarios:  $s$ -channel mediator with spin-0 and spin-1 and Dirac dark matter,  $s$ -channel mediator with spin-2 and vectorial dark matter,  $t$ -channel mediator with either scalar dark matter and fermionic mediator or vice-versa. DMsims provide a simple framework where to define, categorize, and compare the current reach of dark matter searches, as well as the expected sensitivity of forthcoming experiments. These comprehensive analyses are a powerful tool to understand the dynamics underlying the various dark matter searches, modulo their interpretation being subject to the caveats described in the previous section. Keeping in mind the main assumption that the dark matter and the mediator are the only particles of the dark sector accessible at current and future experiments, we can formulate few general statements from the global analyses presented in this review, after having presented the forthcoming particles and astroparticle probes.

Experimentally, the close future is quite promising as there is a rich program expected to start soon and produce results in the next decade or so. Concerning the future of direct detection, starting from 2019, there are several experiments planned able to probe WIMP-nucleon cross section of the order of the neutrino floor ( $\sigma_n^{\text{SI}} \sim 10^{-48} \text{cm}^2$  for  $m_{\text{DM}} \sim 30 \text{GeV}$ ), see XENONnT (Aprile et al., 2016), LZ (Mount et al., 2017), and DARWIN (Aalbers et al., 2016). At low WIMP mass, around 3-4 GeV, exciting progresses are expected by SuperCDMS SNOLAB (Agnese et al., 2017), by CRESST III (Strauss et al., 2016), and by EDELWEISS-III (Arnaud et al., 2018), which can probe elastic spin-independent cross sections as low as  $10^{-44} \text{cm}^2$ . Concerning indirect detection, the Cherenkov Telescope Array (Acharya et al., 2013) (CTA) is one of the major advancements in the gamma-ray searches as it will be sensitive to the energy range in between 20 and 300 TeV. Starting from 2022, while operating, it will provide unprecedented complementary results to direct detection and LHC searches, as it will be sensitive to dark matter masses up to 100 TeV. These future probes, together with the LHC Run 3 foreseen for 2021, can vastly extend the coverage of the dark matter parameter space of simplified models (see e.g., Balazs et al., 2017; Baum et al., 2017; Bertone et al., 2017). More precisely:

- *$s$ -channel scalar mediator:* The spin-0 odd mediator can be probed either by direct detection, which is actually the most promising dark matter search for this model, either by LHC. Indirect detection searches are insensitive to this model, as  $\langle\sigma v\rangle$  is  $p$ -wave suppressed. XENONnT, with one year of exposure, can probe most of the dark matter region from 10 GeV up to roughly 200 GeV, with couplings  $g_{\text{DM}} \times g_{\text{SM}} > 10^{-3}$  (see Arina et al., 2016). Mono-X + missing energy searches can probe the region where the dark matter achieves the relic density by resonant annihilation up to mediator masses of 1 TeV and dark matter masses up to 500 GeV, for couplings of

order 1. More promising are LHC resonant searches, that can probe much higher mediator masses (see e.g., Balazs et al., 2017; Banerjee et al., 2017; Baum et al., 2017; Bertone et al., 2017).

This case is an example of how LHC searches for dark matter with mono-X and missing energy, direct dark matter searches and indirect dark matter searches in general probe regions of the model parameter space which are complementary to each other. Direct detection has better sensitivity than LHC searches in the intermediate dark matter mass range, while LHC performs better in the small dark matter range, where however the dark matter is often not viable as thermal relic (it is either over-abundant or under-abundant, depending on the model). The relic density constraint can be avoided by assuming for instance a non thermal dark matter scenario. Because of the complementarity of searches, for instance a non-detection in direct detection does not preclude a positive detection at the LHC or at gamma-ray telescopes.

- *s-channel pseudoscalar mediator*: This case has the same phenomenology as the scalar mediator, as far as it concerns the LHC searches. On the contrary of the scalar mediator case, it is completely elusive for direct detection searches but can be probed by indirect detection (Banerjee et al., 2017). Indirect dark matter searches extend to heavier dark matter masses with respect to LHC and direct detection exploring TeV candidates. The searches for gamma-rays from dSph galaxies after 15 years will cover basically all the parameter space where the dark matter has the correct relic density, except for the region where  $\Omega_{\text{DM}} h^2$  is obtained via resonant annihilation.
- *s-channel vector mediator*: This model, with Majorana dark matter, is an example of models that are already disfavoured as thermal relic by the combination of LHC and direct detection searches (Fairbairn et al., 2016). By reversing the argument, we can assert that if a signal is seen in the mono-X +  $\cancel{E}_T$  searches, the thermal dark matter hypothesis is under test. This can be solved: (i) by invoking a more complex dark sector, where co-annihilation and new annihilation channels can open up the thermal relic parameter space; (ii) dark matter is produced via additional non-thermal mechanisms to dilute/increase its relic abundance down/up to the observed value. If we do not consider the thermal relic hypothesis, the most LHC promising searches are dijets, that can probe vector mediator masses up to few TeV (Fairbairn et al., 2016) and dark matter direct detection. Conversely gamma-ray searches are poorly sensitive to the mode, the CTA telescope can probe regions of the model that are already excluded by the combination of di-jets searches and XENON1T (Balázs et al., 2017).
- *s-channel spin-2 mediator*: The spin-2 mediator model has been poorly investigated so far and deserves future careful comprehensive analyses. The literature available so far suggests that vectorial dark matter candidates in the TeV range can give rise to gamma-ray line signals partly in the reach of CTA (Zhang et al., 2017), if  $\langle \sigma v \rangle_{\gamma\gamma} > 5 \times 10^{-28} \text{ cm}^3/\text{s}$ . Current constraints seem to suggest that diphoton and dilepton

searches are potentially the best way to hunt for gravitons in DMsims at the LHC (Kraml et al., 2017).

- *t-channel mediator*: The *t*-channel model parameter space, under the hypothesis of thermal dark matter, will be almost entirely probed by future generation experiments. XENONnT can almost entirely probe the dark matter region for mediator masses below the top-quark mass, while the gamma-ray searches in dSphs performed by Fermi-LAT in 15 years can probe the model for mediator masses in between the top-quark mass and roughly 500 GeV. Larger mediator masses can be probed by mono-X searches of LHC Run 3 (Colucci et al., 2018).

It is crucial to keep looking for dark matter with a comprehensive approach relying in simplified bottom-up scenarios. Some theoretical shrewdnesses are in place. The use of gauge invariant models certainly constitutes a must, however theoretical predictions can be improved along other directions, which are often neglected. For instance, the wide separation of scales involved in constraining WIMP models, from the LHC to indirect detection and to direct detection, is often neglected. The authors in (D'Eramo and Procura, 2015) have shown that the running of EFT operators from the mediator mass scale to the nuclear scales probed by direct searches via one-loop Renormalisation Group Equations (RGEs) has an impact for models that would in generally not be constrained by direct detection searches because suppressed by the momentum transfer or by the WIMP velocity. These models can be excluded as a consequence of spin-independent couplings induced by SM loops.

In case of a positive signal at the LHC in a SM +  $\cancel{E}_T$  channel, the identification of the dark matter is non-trivial, as opposed to the characterization of the mediator. Luckily all dark matter searches, even though they feature a certain degree of complementarity, also probe common regions of the parameter space. In an optimistic scenario, a signal can be detected in multiple experiments allowing to pin point both the nature of the dark matter and the characteristics of the model.

Lastly, the Sun has been recently proposed as target to constrain a specific class of DMsims, in which the mediator is light (MeV range) and long-lived (Arina et al., 2017; Leane et al., 2017). LHC searches are insensitive to this type of mediators, which can however be observed in gamma rays. The Sun is opaque to all dark matter annihilation products but neutrinos and the neutral and weakly interacting mediators (the mechanism that produces the mediators inside the Sun is the same as for the neutrino signal). If these mediators are long-lived enough to decay outside the Sun, they could lead to characteristic gamma-ray signatures detectable within 10 years of Fermi-LAT mission, and in 1 year of full exposure of ground water Cherenkov telescopes (HAWC Abeyssekara et al., 2013 and LHAASO Zhen, 2014; He, 2016). Models with long-lived MeV mediators are actually very constrained by beam dump experiments and

cosmology (Arina et al., 2017). Their entire parameter space can be probed by next generation of intensity experiments, such as NA62 (Doebrich, 2017) and SHiP (Alekhnin et al., 2016). Hence dark matter simplified models not only serve as benchmark for high-energy studies but they can be exploited as a bridge relying the high-energy frontiers with the intensity frontiers.

## AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and approved it for publication.

## REFERENCES

- Aaboud, M., Aad, G., Abbott, B., Abdallah, J., Abidinov, O., Abeloos, B., et al. (2016a). Search for new phenomena in final states with an energetic jet and large missing transverse momentum in  $pp$  collisions at  $\sqrt{s} = 13$  TeV using the ATLAS detector. *Phys. Rev. D* 94:032005. doi: 10.1103/PhysRevD.94.032005
- Aaboud, M., Aad, G., Abbott, B., Abdallah, J., Abidinov, O., Abeloos, B., et al. (2016b). Search for squarks and gluinos in final states with jets and missing transverse momentum at  $\sqrt{s} = 13$  TeV with the ATLAS detector. *Eur. Phys. J. C* 76:392. doi: 10.1140/epjc/s10052-016-4184-8
- Aaboud, M., Aad, G., Abbott, B., Abdallah, J., Abidinov, O., Abeloos, B., et al. (2018). Search for dark matter and other new phenomena in events with an energetic jet and large missing transverse momentum using the ATLAS detector. *J. High Energy Phys.* 1:126. doi: 10.1007/JHEP01(2018)126
- Aad, G., Abbott, B., Abdallah, J., Abdel Khalek, S., Abidinov, O., Aben, R., et al. (2015b). Search for new phenomena in final states with an energetic jet and large missing transverse momentum in  $pp$  collisions at  $\sqrt{s} = 8$  TeV with the ATLAS detector. *Eur. Phys. J. C* 75:299. doi: 10.1140/epjc/s10052-015-3639-7
- Aad, G., Abbott, B., Abdallah, J., Abdel Khalek, S., Abidinov, O., Aben, R., et al. (2015c). Search for new phenomena in the dijet mass distribution using  $p - p$  collision data at  $\sqrt{s} = 8$  TeV with the ATLAS detector. *Phys. Rev. D* 91:052007. doi: 10.1103/PhysRevD.91.052007
- Aad, G., Abbott, B., Abdallah, J., Abidinov, O., Aben, R., Abolins, M., et al. (2015a). A search for  $t\bar{t}$  resonances using lepton-plus-jets events in proton-proton collisions at  $\sqrt{s} = 8$  TeV with the ATLAS detector. *J. High Energy Phys.* 8:148. doi: 10.1007/JHEP08(2015)148
- Aad, G., Abbott, B., Abdallah, J., Abidinov, O., Aben, R., Abolins, M., et al. (2016). Search for dark matter produced in association with a Higgs boson decaying to two bottom quarks in  $pp$  collisions at  $\sqrt{s} = 8$  TeV with the ATLAS detector. *Phys. Rev. D* 93:072007. doi: 10.1103/PhysRevD.93.072007
- Aalbers, J., Agostini, F., Alfonsi, M., Amaro, F. D., Amsler, C., Aprile, E., et al. (2016). DARWIN: towards the ultimate dark matter detector. *J. Cosmol. Astropart. Phys.* 1611:017. doi: 10.1088/1475-7516/2016/11/017
- Aaltonen, T., Adelman, J., Akimoto, T., Álvarez González, B., Amerio, S., Amidei, D., et al. (2009). Search for new particles decaying into dijets in proton-antiproton collisions at  $s^{1/2} = 1.96$ -TeV. *Phys. Rev. D* 79:112002. doi: 10.1103/PhysRevD.79.112002
- Aartsen, M. G., Abbasi, R., Abdou, Y., Ackermann, M., Adams, J., Aguilar, J. A., et al. (2013). Search for dark matter annihilations in the Sun with the 79-string IceCube detector. *Phys. Rev. Lett.* 110:131302. doi: 10.1103/PhysRevLett.110.131302
- Aartsen, M. G., Abraham, K., Ackermann, M., Adams, J., Aguilar, J. A., Ahlers, M., et al. (2016). Improved limits on dark matter annihilation in the Sun with the 79-string IceCube detector and implications for supersymmetry. *JCAP* 1604:022. doi: 10.1088/1475-7516/2016/04/022
- Abbiendi, G., Ainsley, C., Akesson, P., Alexander, G., Allison, J., Amaral, P., et al. (2002). Search for scalar top and scalar bottom quarks at LEP. *Phys. Lett. B* 545, 272–284. doi: 10.1016/S0370-2693(02)02808-3
- Abbott T, Abdalla, F. B., Aleksic, J., Allam, S., Amara, A., Bacon, D., et al. (2005). The dark energy survey. *arXiv: astro-ph/0510346*.

## FUNDING

This review article has been supported by the Innoviris grant ATTRACT Brains for Brussels 2015 (BECAP 2015-BB2B-4).

## ACKNOWLEDGMENTS

The author would like to acknowledge Fabio Maltoni, Luca Mantani, and Kentarou Mawatari for useful discussions about various aspects of the dark matter simplified models. She is also grateful to Jan Heisig for a careful reading of the manuscript and for providing useful comments.

- Abdalla, H., Abramowski, A., Aharonian, F., Ait Benkhali, F., Akhperjanian, A. G., Andersson, T., et al. (2016). H.E.S.S. limits on linelike dark matter signatures in the 100 GeV to 2 TeV energy range close to the Galactic Center. *Phys. Rev. Lett.* 117:151302. doi: 10.1103/PhysRevLett.117.151302
- Abdallah, J., Araujo, H., Arbey, A., Ashkenazi, A., Belyaev, A., Berger, J., et al. (2015). Simplified models for dark matter searches at the LHC. *Phys. Dark Univ.* 9–10, 8–23. doi: 10.1016/j.dark.2015.08.001
- Abercrombie, D., Akchurin, N., Akilli, E., Alcaraz Maestre, J., Allen, B., Alvarez Gonzalez, B., et al. (2015). *Dark Matter Benchmark Models for Early LHC Run-2 Searches: Report of the ATLAS/CMS Dark Matter Forum*.
- Abeysekara, A. U., Alfaro, R., Alvarez, C., Álvarez, J. D., Arceo, R., Arteaga-Velázquez, J. C., et al. (2013). Sensitivity of the high altitude water cherenkov detector to sources of multi-TeV gamma rays. *Astropart. Phys.* 50–52, 26–32. doi: 10.1016/j.astropartphys.2013.08.002
- Abramowski, A., Acero, F., Aharonian, F., Akhperjanian, A. G., Anton, G., Balenderan, S., et al. (2013). Search for photon-lineline signatures from dark matter annihilations with H.E.S.S. *Phys. Rev. Lett.* 110:041301. doi: 10.1103/PhysRevLett.110.041301
- Acharya, B. S. A. M., Croston, J. H., Granot, J., and Hardcastle, M. J. (2013). Introducing the CTA concept. *Astropart. Phys.* 43, 3–18. doi: 10.1016/j.astropartphys.2013.01.007
- Ackermann, M., Ajello, M., Albert, A., Anderson, B., Atwood, W. B., Baldini, L., et al. (2015b). Updated search for spectral lines from Galactic dark matter interactions with pass 8 data from the Fermi Large Area Telescope. *Phys. Rev. D* 91:122002. doi: 10.1103/PhysRevD.91.122002
- Ackermann, M., Albert, A., Anderson, B., Atwood, W. B., Baldini, L., Barbiellini, G., et al. (2015a). Searching for dark matter annihilation from milky way dwarf spheroidal galaxies with six years of Fermi Large Area Telescope Data. *Phys. Rev. Lett.* 115:231301. doi: 10.1103/PhysRevLett.115.231301
- Agnese, R., Anderson, A. J., Aramaki, T., Arnqvist, I., Baker, W., Barker, D., et al. (2017). Projected Sensitivity of the SuperCDMS SNOLAB experiment. *Phys. Rev. D* 95:082002. doi: 10.1103/PhysRevD.95.082002
- Agrawal, P., Batell, B., Hooper, D., and Lin, T. (2014). Flavored dark matter and the galactic center gamma-ray excess. *Phys. Rev. D* 90:063512. doi: 10.1103/PhysRevD.90.063512
- Agrawal, P., Blanchet, S., Chacko, Z., and Kilic, C. (2012). Flavored Dark Matter, and its implications for direct detection and colliders. *Phys. Rev. D* 86:055002. doi: 10.1103/PhysRevD.86.055002
- Aguilar, R., Ali Cavasonza, L., Alpat, B., Ambrosi, G., Arruda, L., Attig, N., et al. (2016). Antiproton flux, antiproton-to-proton flux ratio, and properties of elementary particle fluxes in primary cosmic rays measured with the alpha magnetic spectrometer on the international space station. *Phys. Rev. Lett.* 117:091103. doi: 10.1103/PhysRevLett.117.091103
- Akerib, D. S., Alsum, S., Araújo, H. M., Bai, X., Bailey, A. J., Balajthy, J., et al. (2017). Results from a search for dark matter in the complete LUX exposure. *Phys. Rev. Lett.* 118:021303. doi: 10.1103/PhysRevLett.118.021303
- Albert, A., Anderson, B., Bechtol, K., Drlica-Wagner, A., Meyer, M., Sánchez-Conde, M., et al. (2017). Searching for dark matter annihilation in recently discovered milky way satellites with Fermi-LAT. *Astrophys. J.* 834:110. doi: 10.3847/1538-4357/834/2/110

- Albert, A., Backović, M., Boveia, A., Buchmueller, O., Busoni, G., De Roeck, A., et al. (2017a). Recommendations of the LHC Dark Matter Working Group: comparing LHC searches for heavy mediators of dark matter production in visible and invisible decay channels. *arXiv:1703.05703*.
- Albert, A., Bauer, M., Brooke, J., Buchmueller, O., Cerdeno, D. G., Citron, M., et al. (2017b). Towards the next generation of simplified Dark Matter models. *Phys. Dark Univ.* 16, 49–70. doi: 10.1016/j.dark.2017.02.002
- Alekhin, S., Altmannshofer, W., Asaka, T., Batell, B., Bezrukov, F., Bondarenko, K., et al. (2016). A facility to Search for Hidden Particles at the CERN SPS: the SHiP physics case. *Rept. Prog. Phys.* 79:124201. doi: 10.1088/0034-4885/79/12/124201
- Alitti, J., Ambrosini, G., Ansari, R., Autiero, D., Bareyre, P., Bertram, I. A., et al. (1993). A Search for new intermediate vector mesons and excited quarks decaying to two jets at the CERN  $\bar{p}p$  collider. *Nucl. Phys.* B400, 3–24. doi: 10.1016/0550-3213(93)90395-6
- Allahverdi, R., Dutta, B., and Sinha, K. (2012). Non-thermal higgsino dark matter: cosmological motivations and implications for a 125 GeV Higgs. *Phys. Rev. D* 86:095016. doi: 10.1103/PhysRevD.86.095016
- Alloul, A., Christensen, N. D., Degrande, C., Duhr, C., and Fuks, B. (2014). FeynRules 2.0 - A complete toolbox for tree-level phenomenology. *Comput. Phys. Commun.* 185, 2250–2300. doi: 10.1016/j.cpc.2014.04.012
- Alwall, J., Frederix, R., Frixione, S., Hirschi, V., Maltoni, F., Mattelaer, O., et al. (2014). The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *J. High Energy Phys.* 7:079. doi: 10.1007/JHEP07(2014)079
- Alwall, J., Herquet, M., Maltoni, F., Mattelaer, O., and Stelzer, T. (2011). MadGraph 5: going beyond. *J. High Energy Phys.* 6:128. doi: 10.1007/JHEP06(2011)128
- Amaldi, U., de Boer, W., and Furstenau, H. (1991). Comparison of grand unified theories with electroweak and strong coupling constants measured at LEP. *Phys. Lett.* B260, 447–455. doi: 10.1016/0370-2693(91)91641-8
- Ambrogio, F., Arina, C., Backovic, M., Heisig, J., Maltoni, F., Mantani, L., et al. (2018). *MadDM v.3.0: A Comprehensive Tool for Dark Matter Studies*.
- Ambrosi, G., An, Q., Asfandiyarov, R., Azzarello, P., Bernardini, P., Bertucci, B., et al. (2017). Direct detection of a break in the teraelectronvolt cosmic-ray spectrum of electrons and positrons. *Nature* 552, 63–66. doi: 10.1038/nature24475
- Amole, C., Ardid, M., Arnquist, I. J., Asner, D. M., Baxter, D., Behnke, E., et al. (2017). Dark Matter search results from the PICO-60  $C_3F_8$  bubble chamber. *Phys. Rev. Lett.* 118:251301. doi: 10.1103/PhysRevLett.118.251301
- An, H., Wang, L. T., and Zhang, H. (2014). Dark matter with  $t$ -channel mediator: a simple step beyond contact interaction. *Phys. Rev. D* 89:115014. doi: 10.1103/PhysRevD.89.115014
- Angloher, G., Bento, A., Bucci, C., Canonica, L., Defay, X., Erb, A., et al. (2016). Results on light dark matter particles with a low-threshold CRESST-II detector. *Eur. Phys. J. C* 76:25. doi: 10.1140/epjc/s10052-016-3877-3
- Aprile, E., Aalbers, J., Agostini, F., Alfonsi, M., Amaro, F. D., Anthony, M., et al. (2016). Physics reach of the XENON1T dark matter experiment. *J. Cosmol. Astropart. Phys.* 1604:027. doi: 10.1088/1475-7516/2016/04/027
- Aprile, E., Aalbers, J., Agostini, F., Alfonsi, M., Amaro, F. D., Anthony, M., et al. (2017a). First Dark Matter Search Results from the XENON1T Experiment. *Phys. Rev. Lett.* 119:181301. doi: 10.1103/PhysRevLett.119.181301
- Aprile, E., Aalbers, J., Agostini, F., Alfonsi, M., Amaro, F. D., Anthony, M., et al. (2017b). Effective field theory search for high-energy nuclear recoils using the XENON100 dark matter detector. *Phys. Rev. D* 96:042004. doi: 10.1103/PhysRevD.96.042004
- Arcadi, G., Dutra, M., Ghosh, P., Lindner, M., Mambrini, Y., Pierre, M., et al. (2017). The waning of the WIMP? A review of models, searches, and constraints. *arXiv:1703.07364*.
- Arcadi, G., Mambrini, Y., Tytgat, M. H. G., and Zaldivar, B. (2014). Invisible  $Z'$  and dark matter: LHC vs LUX constraints. *J. High Energy Phys.* 3:134. doi: 10.1007/JHEP03(2014)134
- Arina, C. (2014). Bayesian analysis of multiple direct detection experiments. *Phys. Dark Univ.* 5–6, 1–17. doi: 10.1016/j.dark.2014.03.003
- Arina, C., Backović, M., Conte, E., Fuks, B., Guo, J., Heisig, J., et al. (2016). A comprehensive approach to dark matter studies: exploration of simplified top-philic models. *J. High Energy Phys.* 11:111. doi: 10.1007/JHEP11(2016)111
- Arina, C., Backović, M., Heisig, J., and Lucente, M. (2017). Solar  $\gamma$  rays as a complementary probe of dark matter. *Phys. Rev. D* 96:063010. doi: 10.1103/PhysRevD.96.063010
- Arina, C., Del Nobile, E., and Panci, P. (2015). Dark matter with pseudoscalar-mediated interactions explains the DAMA signal and the Galactic Center Excess. *Phys. Rev. Lett.* 114:011301. doi: 10.1103/PhysRevLett.114.011301
- Arina, C., Hamann, J., and Wong, Y. Y. Y. (2011). A Bayesian view of the current status of dark matter direct searches. *J. Cosmol. Astropart. Phys.* 1109:022. doi: 10.1088/1475-7516/2011/09/022
- Arnaud, Q., Armengaud, E., Augier, C., Benoît, A., Bergé, L., Billard, J., et al. (2018). Optimizing EDELWEISS detectors for low-mass WIMP searches. *Phys. Rev. D* 97:022003. doi: 10.1103/PhysRevD.97.022003
- Arun, M. T., and Saha, P. (2017). Gravitons in multiply warped scenarios: At 750 GeV and beyond. *Pramana* 88:93. doi: 10.1007/s12043-017-1387-y
- ATLAS Collaboration (2013). *Search for a Dijet Resonance Produced in Association with a Leptonically decaying  $W$  or  $Z$  Boson with the ATLAS Detector at  $\sqrt{s} = 8$  TeV*. ATLAS-CONF-2013-074. Geneva: CERN.
- ATLAS Collaboration (2016a). *Search for Pair Production of Higgs Bosons in the  $4b$ s with bar symbol Final State Using Proton-Proton Collisions at  $\sqrt{s} = 13$  TeV with the ATLAS Detector*. ATLAS-CONF-2016-049. Geneva: CERN.
- ATLAS Collaboration (2016b). *Search for Diboson Resonance Production in the  $lvqq$  Final State Using pp Collisions at  $\sqrt{s} = 13$  TeV With the ATLAS Detector at the LHC*. ATLAS-CONF-2016-062. Geneva: CERN.
- ATLAS Collaboration (2016c). *Search for New Phenomena in Dijet Events with the ATLAS Detector at  $\sqrt{s} = 13$  TeV with 2015 and 2016 Data*. ATLAS-CONF-2016-069. Geneva: CERN.
- ATLAS Collaboration (2016d). *Search for New Light Resonances Decaying to Jet Pairs and Produced in Association with a Photon or a Jet in Proton-Proton Collisions at  $\sqrt{s} = 13$  TeV with the ATLAS Detector*. ATLAS-CONF-2016-070. Geneva: CERN.
- ATLAS Collaboration (2016e). *Searches for Heavy  $ZZ$  and  $ZW$  Resonances in the  $llqq$  and  $vvqq$  Final States in pp Collisions at  $\sqrt{s} = 13$  TeV with the ATLAS Detector*. ATLAS-CONF-2016-082. Geneva: CERN.
- ATLAS Collaboration (2016f). *Search for Resonances in the Mass Distribution of Jet Pairs with One or Two Jets Identified as  $b$ -jets with the ATLAS Detector with 2015 and 2016 Data*. ATLAS-CONF-2016-060.
- Backović, M., Kraemer, M., Maltoni, F., Martini, A., Mawatari, K., and Pellen, M. (2015). Higher-order QCD predictions for dark matter production at the LHC in simplified models with  $s$ -channel mediators. *Eur. Phys. J. C* 75:482. doi: 10.1140/epjc/s10052-015-3700-6
- Baek, S., Ko, P., and Li, J. (2017). Minimal renormalizable simplified dark matter model with a pseudoscalar mediator. *Phys. Rev. D* 95:075011. doi: 10.1103/PhysRevD.95.075011
- Baek, S., Ko, P., Park, M., Park, W. I., and Yu, C. (2016). Beyond the Dark matter effective field theory and a simplified model approach at colliders. *Phys. Lett.* B756, 289–294. doi: 10.1016/j.physletb.2016.03.026
- Bai, Y., and Berger, J. (2013). Fermion portal Dark Matter. *J. High Energy Phys.* 11:171. doi: 10.1007/JHEP11(2013)171
- Bai, Y., and Berger, J. (2014). Lepton portal Dark Matter. *J. High Energy Phys.* 8:153. doi: 10.1007/JHEP08(2014)153
- Balázs, C., Conrad, J., Farmer, B., Jacques, T., Li, T., Meyer, M., et al. (2017). Sensitivity of the Cherenkov Telescope Array to the detection of a dark matter signal in comparison to direct detection and collider experiments. *Phys. Rev. D* 96:083002. doi: 10.1103/PhysRevD.96.083002
- Balazs, C., Fowlie, A., Mazumdar, A., and White, G. (2017). Gravitational waves at aLIGO and vacuum stability with a scalar singlet extension of the Standard Model. *Phys. Rev. D* 95:043505. doi: 10.1103/PhysRevD.95.043505
- Banerjee, S., Barducci, D., Bélanger, G., Fuks, B., Goudelis, A., and Zaldivar, B. (2017). Cornering pseudoscalar-mediated dark matter with the LHC and cosmology. *J. High Energy Phys.* 7:080. doi: 10.1007/JHEP07(2017)080
- Barger, V., Keung, W. Y., and Marfatia, D. (2012). Bremsstrahlung in dark matter annihilation. *Phys. Lett.* B707, 385–388. doi: 10.1016/j.physletb.2012.01.001
- Bauer, M., Haisch, U., and Kahlhoefer, F. (2017). Simplified dark matter models with two Higgs doublets: I. Pseudoscalar mediators. *J. High Energy Phys.* 5:138. doi: 10.1007/JHEP05(2017)138
- Baum, S., Catena, R., Conrad, J., Freese, K., and Krauss, M. B. (2017). Determining Dark Matter properties with a XENONnT/LZ signal and LHC-Run3 mono-jet searches. *arXiv:1709.06051*.
- Belanger, G., Boudjema, F., Goudelis, A., Pukhov, A., and Zaldivar, B. (2018). *micrOMEGAs5.0: freeze-in*.

- Bell, N. F., Busoni, G., and Sanderson, I. W. (2017b). Two higgs doublet dark matter portal. *arXiv: 1710.10764*.
- Bell, N. F., Cai, Y., Dent, J. B., Leane, R. K., and Weiler, T. J. (2015). Dark matter at the LHC: effective field theories and gauge invariance. *Phys. Rev. D* 92:053008. doi: 10.1103/PhysRevD.92.053008
- Bell, N. F., Cai, Y., and Leane, R. K. (2016a). Dark Forces in the Sky: Signals from  $Z'$  and the Dark Higgs. *J. Cosmol. Astropart. Phys.* 1608:001. doi: 10.1088/1475-7516/2016/08/001
- Bell, N. F., Cai, Y., and Leane, R. K. (2016b). Mono-W dark matter signals at the LHC: simplified model analysis. *J. Cosmol. Astropart. Phys.* 1601:051. doi: 10.1088/1475-7516/2016/01/051
- Bell, N. F., Cai, Y., and Leane, R. K. (2017a). Impact of mass generation for spin-1 mediator simplified models. *J. Cosmol. Astropart. Phys.* 1701:039. doi: 10.1088/1475-7516/2017/01/039
- Bell, N. F., Dent, J. B., Galea, A. J., Jacques, T. D., Krauss, L. M., and Weiler, T. J. (2011). W/Z Bremsstrahlung as the dominant annihilation channel for dark Matter, revisited. *Phys. Lett. B* 706, 6–12. doi: 10.1016/j.physletb.2011.10.057
- Bell, N. F., Dent, J. B., Galea, A. J., Jacques, T. D., Krauss, L. M., and Weiler, T. J. (2012). Searching for Dark Matter at the LHC with a Mono-Z. *Phys. Rev. D* 86:096011. doi: 10.1103/PhysRevD.86.096011
- Belyaev, A., Panizzi, L., Pukhov, A., and Thomas, M. (2017). Dark Matter characterization at the LHC in the Effective Field Theory approach. *J. High Energy Phys.* 4:110. doi: 10.1007/JHEP04(2017)110
- Bergstrom, L., Edsjo, J., and Gondolo, P. (1997). Indirect neutralino detection rates in neutrino telescopes. *Phys. Rev. D* 55, 1765–1770. doi: 10.1103/PhysRevD.55.1765
- Bertone, G., Bozorgnia, N., Kim, J. S., Liem, S., McCabe, C., Otten, S., et al. (2017). Identifying WIMP dark matter from particle and astroparticle data. *arXiv: 1712.04793*.
- Bertone, G., Cerdeno, D. G., Fornasa, M., Ruiz de Austri, R., Strege, C., and Trotta, R. (2012). Global fits of the cMSSM including the first LHC and XENON100 data. *J. Cosmol. Astropart. Phys.* 1201:015. doi: 10.1088/1475-7516/2012/01/015
- Bertone, G., and Hooper, D. (2016). A history of dark matter. *Rev. Mod. Phys.* *arXiv: 1605.04909*.
- Bhattacharyya, G., and Das, D. (2016). Scalar sector of two-Higgs-doublet models: a minireview. *Pramana* 87:40. doi: 10.1007/s12043-016-1252-4
- Billard, J., Strigari, L., and Figueroa-Feliciano, E. (2014). Implication of neutrino backgrounds on the reach of next generation dark matter direct detection experiments. *Phys. Rev. D* 89:023524. doi: 10.1103/PhysRevD.89.023524
- Blennow, M., Edsjo, J., and Ohlsson, T. (2008). Neutrinos from WIMP annihilations using a full three-flavor Monte Carlo. *J. Cosmol. Astropart. Phys.* 0801:021. doi: 10.1088/1475-7516/2008/01/021
- Blumlein, J., Boos, E., and Kryukov, A. (1997). Leptoquark pair production in hadronic interactions. *Z. Phys. C* 76, 137–153. doi: 10.1007/s002880050538
- Bonnivard, V., Combet, C., Daniel, M., Funk, S., Geringer-Sameth, A., Hinton, J. A., et al. (2015). Dark matter annihilation and decay in dwarf spheroidal galaxies: the classical and ultrafaint dSphs. *Mon. Not. Roy. Astron. Soc.* 453, 849–867. doi: 10.1093/mnras/stv1601
- Boveia, A., Buchmueller, O., Busoni, G., D'Eramo, F., De Roeck, A., De Simone, A., et al. (2016). Recommendations on presenting LHC searches for missing transverse energy signals using simplified s-channel models of dark matter. *arXiv: 1603.04156*.
- Bramante, J., and Unwin, J. (2017). Superheavy Thermal Dark Matter and Primordial Asymmetries. *J. High Energy Phys.* 2:119. doi: 10.1007/JHEP02(2017)119
- Branco, G., Ferreira, P. M., Lavoura, L., Rebelo, M. N., Sher, M., and Silva, J. P. (2012). Theory and phenomenology of two-Higgs-doublet models. *Phys. Rept.* 516, 1–102. doi: 10.1016/j.physrep.2012.02.002
- Brennan, A. J., McDonald, M. F., Gramling, J., and Jacques, T. D. (2016). Collide and conquer: constraints on simplified Dark Matter models using Mono-X Collider Searches. *J. High Energy Phys.* 5:112. doi: 10.1007/JHEP05(2016)112
- Bringmann, T., Calore, F., Galea, A., and Garmy, M. (2017). Electroweak and Higgs Boson Internal Bremsstrahlung: General considerations for Majorana dark matter annihilation and application to MSSM neutralinos. *J. High Energy Phys.* 9:041. doi: 10.1007/JHEP09(2017)041
- Bringmann, T., Edsjo, J., Gondolo, P., Ullio, P., and Bergstrom, L. (2018). DarkSUSY 6: an advanced tool to compute dark matter properties numerically. *arXiv: 1802.03399*.
- Bringmann, T., Galea, A. J., and Walia, P. (2016). Leading QCD corrections for indirect dark matter searches: a fresh look. *Phys. Rev. D* 93:043529. doi: 10.1103/PhysRevD.93.043529
- Brown, T. M., Tumlinson, J., Geha, M., Kirby, E. N., VandenBerg, D. A., Muñoz, R. R., et al. (2012). The primeval populations of the Ultra-Faint dwarf galaxies. *Astrophys. J.* 753:L21. doi: 10.1088/2041-8205/753/1/L21
- Bruggisser, S., Riva, F., and Urbano, A. (2016). The last gasp of dark matter effective theory. *J. High Energy Phys.* 11:069. doi: 10.1007/JHEP11(2016)069
- Buchmueller, O., Dolan, M. J., Malik, S. A., and McCabe, C. (2015). Characterising dark matter searches at colliders and direct detection experiments: Vector mediators. *J. High Energy Phys.* 1:037. doi: 10.1007/JHEP01(2015)037
- Buchmueller, O., Dolan, M. J., and McCabe, C. (2014). Beyond effective field theory for dark matter searches at the LHC. *J. High Energy Phys.* 1:025. doi: 10.1007/JHEP01(2014)025
- Buckley, M. R., Feld, D., and Goncalves, D. (2015). Scalar simplified models for Dark Matter. *Phys. Rev. D* 91:015017. doi: 10.1103/PhysRevD.91.015017
- Busoni, G., De Simone, A., Gramling, J., Morgante, E., and Riotto, A. (2014b). On the validity of the effective field theory for dark matter searches at the LHC, part II: complete analysis for the s-channel. *J. Cosmol. Astropart. Phys.* 1406:060. doi: 10.1088/1475-7516/2014/06/060
- Busoni, G., De Simone, A., Jacques, T., Morgante, E., and Riotto, A. (2014c). On the validity of the effective field theory for dark matter searches at the LHC part III: analysis for the t-channel. *J. Cosmol. Astropart. Phys.* 1409:022. doi: 10.1088/1475-7516/2014/09/022
- Busoni, G., De Simone, A., Morgante, E., and Riotto, A. (2014a). On the validity of the effective field theory for dark matter searches at the LHC. *Phys. Lett. B* 728, 412–421. doi: 10.1016/j.physletb.2013.11.069
- Cao, Q. H., Ma, E., and Shaughnessy, G. (2009). Dark Matter: the leptonic connection. *Phys. Lett. B* 673, 152–155. doi: 10.1016/j.physletb.2009.02.015
- Carena, M., Quiros, M., and Wagner, C. E. M. (1996). Effective potential methods and the Higgs mass spectrum in the MSSM. *Nucl. Phys. B* 461, 407–436. doi: 10.1016/0550-3213(95)00665-6
- Carpenter, L. M., Colburn, R., Goodman, J., and Linden, T. (2016). Indirect detection constraints on s and t channel simplified models of Dark Matter. *Phys. Rev. D* 94:055027. doi: 10.1103/PhysRevD.94.055027
- Chala, M., Kahlhoefer, F., McCullough, M., Nardini, G., and Schmidt-Hoberg, K. (2015). Constraining dark sectors with Monojets and Dijets. *J. High Energy Phys.* 7:089. doi: 10.1007/JHEP07(2015)089
- Chang, S., Edezhath, R., Hutchinson, J., and Luty, M. (2014a). Effective WIMPs. *Phys. Rev. D* 89:015011. doi: 10.1103/PhysRevD.89.015011
- Chang, S., Edezhath, R., Hutchinson, J., and Luty, M. (2014b). Leptophilic Effective WIMPs. *Phys. Rev. D* 90:015011. doi: 10.1103/PhysRevD.90.015011
- Choi, K., Abe, K., Haga, Y., Hayato, Y., Iyogi, K., Kameda, J., et al. (2015). Search for neutrinos from annihilation of captured low-mass dark matter particles in the Sun by Super-Kamiokande. *Phys. Rev. Lett.* 114:141301. doi: 10.1103/PhysRevLett.114.141301
- Chu, X., Hambye, T., Scarna, T., and Tytgat, M. H. G. (2012). What if Dark Matter gamma-ray lines come with Gluon Lines? *Phys. Rev. D* 86:083521. doi: 10.1103/PhysRevD.86.083521
- Cirelli, M. (2015). Status of indirect (and direct) dark matter searches. *arXiv: 1511.02031*.
- Cirelli, M., Del Nobile, E., and Panci, P. (2013). Tools for model-independent bounds in direct dark matter searches. *J. Cosmol. Astropart. Phys.* 1310:019. doi: 10.1088/1475-7516/2013/10/019
- Cirigliano, V., Graesser, M. L., and Ovanessian, G. (2012). WIMP-nucleus scattering in chiral effective theory. *J. High Energy Phys.* 10:025. doi: 10.1007/JHEP10(2012)025
- CMS (2015). *Search for Resonances Decaying to Dijet Final States at  $\sqrt{s} = 8$  TeV with Scouting Data*. CMS-PAS-EXO-14-005. Geneva: CERN.
- CMS Collaboration (2014). *Search for the Production of Dark Matter in Association with Top Quark Pairs in the Single-lepton Final State in pp Collisions at  $\sqrt{s} = 8$  TeV*. CMS-PAS-B2G-14-004.
- CMS Collaboration (2015). *Search for Massive Resonances Decaying into Pairs of boosted W and Z Bosons at  $\sqrt{s} = 13$  TeV*. CMS-PAS-EXO-15-002.
- CMS Collaboration (2016a). Search for new high-mass resonances in the dilepton final state using proton-proton collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector.

- CMS Collaboration (2016b). *Search for a High-Mass Resonance Decaying into a Dilepton Final State in  $13\text{ fb}^{-1}$  of  $pp$  Collisions at  $\sqrt{s} = 13\text{ TeV}$* . CMS-PAS-EXO-16-031.
- CMS Collaboration (2016c). *Search for  $t\bar{t}$  Resonances in Boosted Semileptonic Final States in  $pp$  Collisions at  $\sqrt{s} = 13\text{ TeV}$* . CMS-PAS-B2G-15-002.
- CMS Collaboration (2016d). *Search for a Neutral MSSMHiggs Boson Decaying into  $\tau\tau$  with  $12.9\text{ fb}^{-1}$  of Data at  $\sqrt{s} = 13\text{ TeV}$* . CMS-PAS-HIG-16-037.
- CMS Collaboration (2017). Search for supersymmetry in events with at least one soft lepton, low jet multiplicity, and missing transverse momentum in proton-proton collisions at  $\sqrt{s} = 13\text{ TeV}$ .
- Colucci, S., Fuks, B., Giacchino, F., Lopez Honorez, L., Tytgat, M. H. G., and VandeCasteele, J. (2018). Top-philic vector-like portal to scalar Dark Matter. *Phys. Rev. D* 98:035002. doi: 10.1103/PhysRevD.98.035002
- Costa, R., Muehleitner, M., Sampaio, M. O. P., and Santos, R. (2016). Singlet extensions of the standard model at LHC Run 2: benchmarks and comparison with the NMSSM. *J. High Energy Phys.* 6:034. doi: 10.1007/JHEP06(2016)034
- Courteau, S., Cappellari, M., de Jong, R. S., Dutton, A. A., Emsellem, E., Hoekstra, H., et al. (2014). Galaxy masses. *Rev. Mod. Phys.* 86, 47–119. doi: 10.1103/RevModPhys.86.47
- Cuomo, A., Heisig, J., Korsmeier, M., and Kraemer, M. (2018). Constraining heavy dark matter with cosmic-ray antiprotons. *J. Cosmol. Astropart. Phys.* 1804:004. doi: 10.1088/1475-7516/2018/04/004
- D'Ambrosio, G., Giudice, G. F., Isidori, G., and Strumia, A. (2002). Minimal flavor violation: an Effective field theory approach. *Nucl. Phys. B* 645, 155–187. doi: 10.1016/S0550-3213(02)00836-2
- Das, G., Degrande, C., Hirschi, V., Maltoni, F., and Shao, H. S. (2017). NLO predictions for the production of a spin-two particle at the LHC. *Phys. Lett. B* 770, 507–513. doi: 10.1016/j.physletb.2017.05.007
- De Simone, A., and Jacques, T. (2016). Simplified models vs. effective field theory approaches in dark matter searches. *Eur. Phys. J. C* 76, 367. doi: 10.1140/epjc/s10052-016-4208-4
- de Swart, J., Bertone, G., and van Dongen, J. (2017). How dark matter came to matter *Nat. Astron.* 1:0059. doi: 10.1038/s41550017-0059
- Degrande, C. (2015). Automatic evaluation of UV and R2 terms for beyond the Standard Model Lagrangians: a proof-of-principle. *Comput. Phys. Commun.* 197, 239–262. doi: 10.1016/j.cpc.2015.08.015
- Del Nobile, E., Gelmini, G., Gondolo, P., and Huh, J. H. (2013). Generalized halo independent comparison of direct dark matter detection data. *J. Cosmol. Astropart. Phys.* 1310:048. doi: 10.1088/1475-7516/2013/10/048
- D'Eramo, F., and Procura, M. (2015). Connecting Dark Matter UV complete models to direct detection rates via effective field theory. *J. High Energy Phys.* 4:054. doi: 10.1007/JHEP04(2015)054
- Dillon, B. M., Han, C., Lee, H. M., and Park, M. (2017). KK graviton resonance and cascade decays in warped gravity. *Int. J. Mod. Phys. A* 32:1745006. doi: 10.1142/S0217751X17450063
- Dillon, B. M., and Sanz, V. (2017). Kaluza-Klein gravitons at LHC2. *Phys. Rev. D* 96:035008. doi: 10.1103/PhysRevD.96.035008
- Doeblich, B. (2017). "Searches for very weakly-coupled particles beyond the Standard Model with NA62," in *Proceedings of the 13th "Patras" Workshop on Axions, WIMPs and WISPs, PATRAS 2017*.
- Dolan, M. J., Spannowsky, M., Wang, Q., and Yu, Z. H. (2016). Determining the quantum numbers of simplified models in  $t\bar{t}X$  production at the LHC. *Phys. Rev. D* 94:015025. doi: 10.1103/PhysRevD.94.015025
- Drewes, M., Lasserre, T., Merle, A., Mertens, S., Adhikari, R., Agostini, M., et al. (2017). A white paper on keV sterile neutrino dark matter. *JCAP* 1701: 025. doi: 10.1088/1475-7516/2017/01/025
- du Pree, T., Hahn, K., Harris, P., and Roskas, C. (2016). Cosmological constraints on dark matter models for collider searches. *arXiv: 1603.08525*.
- Dudas, E., Mambrini, Y., Pokorski, S., and Romagnoni, A. (2009). (In)visible Z-prime and dark matter. *J. High Energy Phys.* 8:014. doi: 10.1088/1126-6708/2009/08/014
- Duerr, M., and Fileviez Perez, P. (2015). Theory for baryon number and Dark Matter at the LHC. *Phys. Rev. D* 91:095001. doi: 10.1103/PhysRevD.91.095001
- Duerr, M., Kahlhoefer, F., Schmidt-Hoberg, K., Schwetz, T., and Vogl, S. (2016). How to save the WIMP: global analysis of a dark matter model with two s-channel mediators. *J. High Energy Phys.* 9:042. doi: 10.1007/JHEP09(2016)042
- Dupuis, G. (2016). Collider constraints and prospects of a scalar singlet extension to Higgs portal Dark Matter. *J. High Energy Phys.* 7:008. doi: 10.1007/JHEP07(2016)008
- Edsjo, J., and Gondolo, P. (1997). Neutralino relic density including coannihilations. *Phys. Rev. D* 56, 1879–1894.
- Einasto, J. and Haud, U. (1989). Galactic models with massive corona. I - Method. II - Galaxy. *Astron. Astrophys.* 223, 89–106.
- El Hedri, S., Kaminska, A., de Vries, M., and Zurita, J. (2017). Simplified phenomenology for colored dark sectors. *J. High Energy Phys.* 4:118. doi: 10.1007/JHEP04(2017)118
- Ellis, J., Fairbairn, M., and Tunney, P. (2017). Anomaly-Free Dark Matter Models are not so Simple. *J. High Energy Phys.* 8:053. doi: 10.1007/JHEP08(2017)053
- Ellis, J., Fok, R., Hwang, D. S., Sanz, V., and You, T. (2013). Distinguishing 'Higgs' spin hypotheses using  $\gamma\gamma$  and  $WW^*$  decays. *Eur. Phys. J. C* 73:2488. doi: 10.1140/epjc/s10052-013-2488-5
- Englert, C., Goncalves-Netto, D., Mawatari, K., and Plehn, T. (2013). Higgs quantum numbers in Weak Boson Fusion. *J. High Energy Phys.* 1:148. doi: 10.1007/JHEP01(2013)148
- Englert, C., McCullough, M., and Spannowsky, M. (2016). S-Channel Dark Matter simplified models and unitarity. *Phys. Dark Univ.* 14, 48–56. doi: 10.1016/j.dark.2016.09.002
- Fairbairn, M., Heal, J., Kahlhoefer, F., and Tunney, P. (2016). Constraints on Z? models from LHC dijet searches and implications for dark matter. *J. High Energy Phys.* 9:018. doi: 10.1007/JHEP09(2016)018
- Fayet, P. (1976). Supersymmetry and weak, electromagnetic and strong interactions. *Phys. Lett.* 64B:159. doi: 10.1016/0370-2693(76)90319-1
- Fitzpatrick, A. L., Haxton, W., Katz, E., Lubbers, N., and Xu, Y. (2013). The effective field theory of dark matter direct detection. *J. Cosmol. Astropart. Phys.* 1302:004. doi: 10.1088/1475-7516/2013/02/004
- Fox, P. J., Kribs, G. D., and Tait, T. M. P. (2011). Interpreting dark matter direct detection independently of the local velocity and density distribution. *Phys. Rev. D* 83:034007. doi: 10.1103/PhysRevD.83.034007
- Fu, C., Cui, X., Zhou, X., Chen, X., Chen, Y., Fang, D., et al. (2017). Spin-dependent weakly-interacting-massive-particle? Nucleon cross section limits from first data of PandaX-II experiment. *Phys. Rev. Lett.* 118:071301. doi: 10.1103/PhysRevLett.118.071301
- Garcia-Cely, C., and Heeck, J. (2016). Indirect searches of dark matter via polynomial spectral features. *J. Cosmol. Astropart. Phys.* 1608:023. doi: 10.1088/1475-7516/2016/08/023
- Garny, M., Heisig, J., Hufnagel, M., and Luel, B. (2018). Top-philic dark matter within and beyond the WIMP paradigm. *Phys. Rev. D* 97:075002. doi: 10.1103/PhysRevD.97.075002
- Garny, M., Ibarra, A., Pato, M., and Vogl, S. (2013). Internal bremsstrahlung signatures in light of direct dark matter searches. *J. Cosmol. Astropart. Phys.* 1312:046. doi: 10.1088/1475-7516/2013/12/046
- Garny, M., Ibarra, A., Rydbeck, S., and Vogl, S. (2014). Majorana Dark Matter with a coloured mediator: collider vs direct and indirect searches. *J. High Energy Phys.* 6:169. doi: 10.1007/JHEP06(2014)169
- Garny, M., Ibarra, A., and Vogl, S. (2015). Signatures of Majorana dark matter with t-channel mediators. *Int. J. Mod. Phys. D* 24:1530019. doi: 10.1142/S0218271815300190
- Gaskins, J. M. (2016). A review of indirect searches for particle dark matter. *Contemp. Phys.* 57, 496–525. doi: 10.1080/00107514.2016.1175160
- Geringer-Sameth, A., Walker, M. G., Koushiappas, S. M., Koposov, S. E., Belokurov, V., Torrealba, G., et al. (2015). Indication of gamma-ray emission from the newly discovered dwarf galaxy reticulum II. *Phys. Rev. Lett.* 115:081101. doi: 10.1103/PhysRevLett.115.081101
- Giacchino, F., Ibarra, A., Lopez Honorez, L., Tytgat, M. H. G., and Wild, S. (2016). Signatures from scalar Dark Matter with a vector-like quark mediator. *J. Cosmol. Astropart. Phys.* 1602:002. doi: 10.1088/1475-7516/2016/02/002
- Giacchino, F., Lopez-Honorez, L., and Tytgat, M. H. G. (2013). Scalar dark matter models with significant internal bremsstrahlung. *J. Cosmol. Astropart. Phys.* 1310:025. doi: 10.1088/1475-7516/2013/10/025
- Giacchino, F., Lopez-Honorez, L., and Tytgat, M. H. G. (2014). Bremsstrahlung and gamma ray lines in 3 scenarios of Dark Matter annihilation. *J. Cosmol. Astropart. Phys.* 1408:046. doi: 10.1088/1475-7516/2014/08/046
- Giesen, G., Boudaud, M., Genolini, Y., Poulin, V., Cirelli, M., Salati, P., et al. (2015). AMS-02 antiprotons, at last! Secondary astrophysical component and

- immediate implications for Dark Matter. *J. Cosmol. Astropart. Phys.* 1509:023. doi: 10.1088/1475-7516/2015/09/023
- Goncalves, D., Machado, P. A. N., and No, J. M. (2017). Simplified models for Dark Matter face their consistent completions. *Phys. Rev. D* 95:055027. doi: 10.1103/PhysRevD.95.055027
- Gondolo, P., Edsjo, J., Ullio, P., Bergstrom, L., Schelke, M., and Baltz, E. A. (2004). DarkSUSY: Computing supersymmetric dark matter properties numerically. *J. Cosmol. Astropart. Phys.* 0407:008. doi: 10.1088/1475-7516/2004/07/008
- Gondolo, P., and Gelmini, G. (1991). Cosmic abundances of stable particles: Improved analysis. *Nucl. Phys. B* 360, 145–179. doi: 10.1016/0550-3213(91)90438-4
- Gondolo, P., and Gelmini, G. B. (2012). Halo independent comparison of direct dark matter detection data. *J. Cosmol. Astropart. Phys.* 1212:015. doi: 10.1088/1475-7516/2012/12/015
- Goodman, J., and Shepherd, W. (2011). LHC bounds on UV-complete models of dark matter. *arXiv:1111.2359*.
- Gould, A. (1987). Resonant enhancements in WIMP capture by the Earth. *Astrophys. J.* 321: 571. doi: 10.1086/165653
- Goyal, A., and Kumar, M. (2016). Fermionic Dark Matter in a simple  $t$ -channel model. *J. Cosmol. Astropart. Phys.* 1611:001. doi: 10.1088/1475-7516/2016/11/001
- Griest, K., and Kamionkowski, M. (1990). Unitarity limits on the mass and radius of dark matter particles. *Phys. Rev. Lett.* 64:615.
- Griest, K., and Seckel, D. (1991). Three exceptions in the calculation of relic abundances. *Phys. Rev. D* 43, 3191–3203. doi: 10.1103/PhysRevD.43.3191
- Gunion, J. F., and Haber, H. E. (1986). Higgs Bosons in supersymmetric models. 1. *Nucl. Phys. B* 272:1. doi: 10.1016/0550-3213(93)90653-7
- Haisch, U., Kahlhoefer, F., and Tait, T. M. P. (2016). On Mono-W signatures in Spin-1 simplified models. *Phys. Lett. B* 760, 207–213. doi: 10.1016/j.physletb.2016.06.063
- Haisch, U., Kahlhoefer, F., and Unwin, J. (2013). The impact of heavy-quark loops on LHC dark matter searches. *J. High Energy Phys.* 7:125. doi: 10.1007/JHEP07(2013)125
- Haisch, U., and Re, E. (2015). Simplified dark matter top-quark interactions at the LHC. *J. High Energy Phys.* 6:078. doi: 10.1007/JHEP06(2015)078
- Han, C., Lee, H. M., Park, M., and Sanz, V. (2016). The diphoton resonance as a gravity mediator of dark matter. *Phys. Lett. B* 755, 371–379. doi: 10.1016/j.physletb.2016.02.040
- Harris, P., Khoze, V. V., Spannowsky, M., and Williams, C. (2015). Constraining dark sectors at colliders: beyond the effective theory approach. *Phys. Rev. D* 91:055009. doi: 10.1103/PhysRevD.91.055009
- He, H. (2016). *Design Highlights and Status of the LHAASO Project*. PoS ICRC2015.
- Heisig, J., Kraemer, M., Pellen, M., and Wiebusch, C. (2016). Constraints on Majorana Dark Matter from the LHC and IceCube. *Phys. Rev. D* 93:055029. doi: 10.1103/PhysRevD.93.055029
- Hooper, D., and Linden, T. (2015). On the gamma-ray emission from reticulum II and other dwarf galaxies. *J. Cosmol. Astropart. Phys.* 1509:016. doi: 10.1088/1475-7516/2015/09/016
- Ibarra, A., Lamperstorfer, A. S., López-Gehler, S., Pato, M., and Bertone, G. (2015a). On the sensitivity of CTA to gamma-ray boxes from multi-TeV dark matter. *J. Cosmol. Astropart. Phys.* 1509:048. doi: 10.1088/1475-7516/2015/09/048
- Ibarra, A., Pierce, A., Shah, N. R., and Vogl, S. (2015b). Anatomy of coannihilation with a scalar top partner. *Phys. Rev. D* 91:095018. doi: 10.1103/PhysRevD.91.095018
- Ibarra, A., Toma, T., Totzauer, M., and Wild, S. (2014). Sharp gamma-ray spectral features from scalar dark matter annihilations. *Phys. Rev. D* 90:043526. doi: 10.1103/PhysRevD.90.043526
- Jacques, T., Katz, A., Morgante, E., Racco, D., Rameez, M., and Riotto, A. (2016). Complementarity of DM searches in a consistent simplified model: the case of  $Z$ ? *J. High Energy Phys.* 10:071. doi: 10.1007/JHEP10(2016)071
- Jacques, T., and Nordstrom, K. (2015). Mapping monojet constraints onto Simplified Dark Matter Models. *J. High Energy Phys.* 6:142. doi: 10.1007/JHEP06(2015)142
- Jeans, J. H. (1922). The motions of stars in a Kapteyn Universe. *MNRAS* 82, 122–132. doi: 10.1093/mnras/82.3.122
- Jungman, G., Kamionkowski, M., and Griest, K. (1996). Supersymmetric dark matter. *Phys. Rept.* 267, 195–373. doi: 10.1016/0370-1573(95)00058-5
- Kahlhoefer, F. (2017). Review of LHC dark matter searches. *Int. J. Mod. Phys. A* 32:1730006. doi: 10.1142/S0217751X1730006X
- Kahlhoefer, F., Schmidt-Hoberg, K., Schwetz, T., and Vogl, S. (2016). Implications of unitarity and gauge invariance for simplified dark matter models. *J. High Energy Phys.* 2:016. doi: 10.1007/JHEP02(2016)016
- Kamionkowski, M. (1991). Energetic neutrinos from heavy neutralino annihilation in the sun. *Phys. Rev. D* 44, 3021–3042. doi: 10.1103/PhysRevD.44.3021
- Kapteyn, J. C. (1922). First attempt at a theory of the arrangement and motion of the sidereal system. *ApJS* 55:302. doi: 10.1086/142670
- Khachatryan, V., Sirunyan, A. M., Tumasyan, A., Adam, W., Ačılar, E., Bergauer, T., et al. (2015b). Search for diphoton resonances in the mass range from 150 to 850 GeV in pp collisions at  $\sqrt{s} = 8$  TeV. *Phys. Lett. B* 750, 494–519. doi: 10.1016/j.physletb.2015.09.062
- Khachatryan, V., Sirunyan, A. M., Tumasyan, A., Adam, W., Ačılar, E., Bergauer, T., et al. (2016). Search for dark matter and unparticles produced in association with a  $Z$  boson in proton-proton collisions at  $\sqrt{s} = 8$  TeV. *Phys. Rev. D* 93:052011. doi: 10.1103/PhysRevD.93.052011
- Khachatryan, V., Sirunyan, A. M., Tumasyan, A., Adam, W., Ačılar, E., Bergauer, T., et al. (2017a). Search for high-mass diphoton resonances in proton-proton collisions at 13 TeV and combination with 8 TeV search. *Phys. Lett. B* 767, 147–170. doi: 10.1016/j.physletb.2017.01.027
- Khachatryan, V., Sirunyan, A. M., Tumasyan, A., Adam, W., Ačılar, E., Bergauer, T., et al. (2017b). Search for heavy resonances decaying to tau lepton pairs in proton-proton collisions at  $\sqrt{s} = 13$  TeV. *J. High Energy Phys.* 2:048. doi: 10.1007/JHEP02(2017)048
- Khachatryan, V., Sirunyan, A. M., Tumasyan, A., Adam, W., Bergauer, T., Dragicevic, M., et al. (2014). Search for Standard Model Production of Four Top Quarks in the Lepton + Jets Channel in pp Collisions at  $\sqrt{s} = 8$  TeV. *J. High Energy Phys.* 11:154. doi: 10.1007/JHEP11(2014)154
- Khachatryan, V., Sirunyan, A. M., Tumasyan, A., Adam, W., Bergauer, T., Dragicevic, M., et al. (2015a). Search for dark matter, extra dimensions, and unparticles in monojet events in proton-proton collisions at  $\sqrt{s} = 8$  TeV. *Eur. Phys. J. C* 75:235. doi: 10.1140/epjc/s10052-015-3451-4
- Khachatryan, V., Sirunyan, A. M., Tumasyan, A., Adam, W., Bergauer, T., Dragicevic, M., et al. (2015c). Search for resonances and quantum black holes using dijet mass spectra in proton-proton collisions at  $\sqrt{s} = 8$  TeV. *Phys. Rev. D* 91:052009. doi: 10.1103/PhysRevD.91.052009
- Khoze, V. V., Ro, G., and Spannowsky, M. (2015). Spectroscopy of scalar mediators to dark matter at the LHC and at 100 TeV. *Phys. Rev. D* 92:075006. doi: 10.1103/PhysRevD.92.075006
- Kile, J. (2013). Flavored Dark Matter: a review. *Mod. Phys. Lett. A* 28:1330031. doi: 10.1142/S0217732313300310
- Ko, P., Natale, A., Park, M., and Yokoya, H. (2017). Simplified DM models with the full SM gauge symmetry: the case of  $t$ -channel colored scalar mediators. *J. High Energy Phys.* 1:086. doi: 10.1007/JHEP01(2017)086
- Kolb, E. W., and Turner, M. S. (1990). The Early Universe. *Front. Phys.* 69, 1–547.
- Kraml, S., Laa, U., Mawatari, K., and Yamashita, K. (2017). Simplified dark matter models with a spin-2 mediator at the LHC. *Eur. Phys. J. C* 77:326. doi: 10.1140/epjc/s10052-017-4871-0
- Leane, R. K., Ng, K. C. Y., and Beacom, J. F. (2017). Powerful solar signatures of long-lived dark mediators. *Phys. Rev. D* 95: 123016. doi: 10.1103/PhysRevD.95.123016
- Lebedev, O., and Mambrini, Y. (2014). Axial dark matter: the case for an invisible  $Z$ ? *Phys. Lett. B* 734, 350–353. doi: 10.1016/j.physletb.2014.05.025
- Lee, B. W., and Weinberg, S. (1977). Cosmological lower bound on heavy-neutrino masses. *Phys. Rev. Lett.* 39, 165–168.
- Lee, H. M., Park, M., and Sanz, V. (2014a). Gravity-mediated (or Composite) Dark Matter confronts astrophysical data. *J. High Energy Phys.* 5:063. doi: 10.1007/JHEP05(2014)063
- Lee, H. M., Park, M., and Sanz, V. (2014b). Gravity-mediated (or Composite) Dark Matter. *Eur. Phys. J. C* 74:2715. doi: 10.1140/epjc/s10052-014-2715-8
- Lopez-Val, D., and Robens, T. (2014).  $\Delta$  and the W-boson mass in the singlet extension of the standard model. *Phys. Rev. D* 90:114018. doi: 10.1103/PhysRevD.90.114018
- LUX Collaboration, Akerib, D. S., Alsum, S., Araújo, H. M., Bai, X., Bailey, A. J., et al. (2017). Limits on spin-dependent WIMP-nucleon cross section obtained from the complete LUX exposure. *Phys. Rev. Lett.* 118:251302. doi: 10.1103/PhysRevLett.118.251302

- March-Russell, J., Unwin, J., and West, S. M. (2012). Closing in on asymmetric dark matter I: model independent limits for interactions with quarks. *J. High Energy Phys.* 8:029. doi: 10.1007/JHEP08(2012)029
- Marsh, D. J. E. (2016). Axion cosmology. *Phys. Rept.* 643, 1–79. doi: 10.1016/j.physrep.2016.06.005
- Martini, A., Mawatari, K., and Sengupta, D. (2016). Diphoton excess in phenomenological spin-2 resonance scenarios. *Phys. Rev. D* 93:075011. doi: 10.1103/PhysRevD.93.075011
- Mateo, M. (1998). Dwarf galaxies of the Local Group. *Ann. Rev. Astron. Astrophys.* 36, 435–506. doi: 10.1146/annurev.astro.36.1.435
- Mattelaer, O., and Vryonidou, E. (2015). Dark matter production through loop-induced processes at the LHC: the s-channel mediator case. *Eur. Phys. J. C* 75:436. doi: 10.1140/epjc/s10052-015-3665-5
- Morgante, E. (2018). Simplified dark matter models. *arXiv: 1804.01245*.
- Mount, B. J., Hans, S., Rosero, R., Yeh, M., Chan, C., Gaitskill, R. J., et al. (2017). *LUX-ZEPLIN (LZ) Technical Design Report*.
- Neubert, M., Wang, J., and Zhang, C. (2015). Higher-order QCD predictions for dark matter production in mono-Z searches at the LHC. *J. High Energy Phys.* 2:082. doi: 10.1007/JHEP02(2016)082
- Oort, J. H. (1932). The force exerted by the stellar system in the direction perpendicular to the galactic plane and some related problems. *Bull. Astron. Inst. Netherlands* 6, 249.
- Papucci, M., Vichi, A., and Zurek, K. M. (2014). Monojet versus the rest of the world I: t-channel models. *J. High Energy Phys.* 11:024. doi: 10.1007/JHEP11(2014)024
- Peter, A. H. G. (2009). Dark matter in the solar system II: WIMP annihilation rates in the Sun. *Phys. Rev. D* 79:103532. doi: 10.1103/PhysRevD.79.103532
- Planck Collaboration, Ade, P. A. R., Aghanim, N., Arnaud, M., Ashdown, M., Aumont, J., et al. (2016). Planck 2015 results. XIII. Cosmological parameters. *Astron. Astrophys.* 594:A13. doi: 10.1051/0004-6361/201525830
- Press, W. H., and Spergel, D. N. (1985). Capture by the sun of a galactic population of weakly interacting massive particles. *Astrophys. J.* 296, 679–684. doi: 10.1086/163485
- Randall, L., and Sundrum, R. (1999). A Large mass hierarchy from a small extra dimension. *Phys. Rev. Lett.* 83, 3370–3373. doi: 10.1103/PhysRevLett.83.3370
- Ritz, S., and Seckel, D. (1988). Detailed neutrino spectra from cold dark matter annihilations in the Sun. *Nucl. Phys. B* 304, 877–908. doi: 10.1016/0550-3213(88)90660-8
- Robens, T., and Stefaniak, T. (2016). LHC benchmark scenarios for the real higgs singlet extension of the standard model. *Eur. Phys. J. C* 76:268. doi: 10.1140/epjc/s10052-016-4115-8
- Rueter, T. D., Rizzo, T. G., and Hewett, J. L. (2017). Gravity-mediated Dark Matter annihilation in the Randall-Sundrum Model. *J. High Energy Phys.* 10:094. doi: 10.1007/JHEP10(2017)094
- Schoenrich, R., Binney, J., and Dehnen, W. (2010). Local kinematics and the local standard of rest. *Mon. Not. Roy. Astron. Soc.* 403:1829. doi: 10.1111/j.1365-2966.2010.16253.x
- Shoemaker, I. M., and Vecchi, L. Unitarity and Monojet bounds on models for DAMA, CoGeNT, and CRESST-II. (2012). *Phys. Rev. D* 86:015023. doi: 10.1103/PhysRevD.86.015023
- Silk, J., Moore, B., Diemand, J., Bullock, J., Kaplinghat, M., Strigari, L., et al. (2010). *Particle Dark Matter: Observations, Models and Searches* Cambridge: Cambridge Univ. Press.
- Silk, J., Olive, K. A., and Srednicki, M. (1985). The Photino, the Sun and high-energy neutrinos. *Phys. Rev. Lett.* 55, 257–259. doi: 10.1103/PhysRevLett.55.257
- Silk, J., and Srednicki, M. (1984). Cosmic ray anti-protons as a probe of a photino dominated Universe. *Phys. Rev. Lett.* 53:624. doi: 10.1103/PhysRevLett.53.624
- Sirunyan, A.M., Tumasyan, A., Adam, W., Asilar, E., Bergauer, T., Brandstetter, J., et al. (2017c). Search for new physics in final states with an energetic jet or a hadronically decaying W or Z boson and transverse momentum imbalance at  $\sqrt{s} = 13$  TeV. *arXiv: 1712.02345*.
- Sirunyan, A. M., Tumasyan, A., Adam, W., AÇilar E, Bergauer T, Brandstetter J, et al. (2017a). Search for dijet resonances in proton-proton collisions at  $\sqrt{s} = 13$  TeV and constraints on dark matter and other models. *Phys. Lett. B* 769, 520–542. doi: 10.1016/j.physletb.2017.02.012
- Sirunyan, A. M., Tumasyan, A., Adam, W., Ambrogio, F., Asilar, E., Bergauer, T., et al. (2018). Search for top squarks and dark matter particles in opposite-charge dilepton final states at  $\sqrt{s} = 13$  TeV. *Phys. Rev. D* 97:032009. doi: 10.1103/PhysRevD.97.032009
- Sirunyan, A. M., Tumasyan, A., Adam, W., Asilar, E., Bergauer, T., Brandstetter, J., et al. (2017b). Search for dark matter produced with an energetic jet or a hadronically decaying W or Z boson at  $\sqrt{s} = 13$  TeV. *J. High Energy Phys.* 7:014. doi: 10.1007/JHEP07(2017)014
- Sivertsson, S., and Edsjo, J. (2012). WIMP diffusion in the solar system including solar WIMP-nucleon scattering. *Phys. Rev. D* 85:123514. doi: 10.1103/PhysRevD.85.123514
- Slatyer, T. R. (2017). “TASI lectures on indirect detection of dark matter,” in *Theoretical Advanced Study Institute in Elementary Particle Physics: Anticipating the Next Discoveries in Particle Physics (TASI 2016)* (Boulder, CO).
- Srednicki, M., Watkins, R., and Olive, K. A. (1988). Calculations of relic densities in the early universe. *Nucl. Phys. B* 310:693.
- Steigman, G., Sarazin, C. L., Quintana, H., and Faulkner, J. (1978). Dynamical interactions and astrophysical effects of stable heavy neutrinos. *Astron. J.* 83, 1050–1061. doi: 10.1086/112290
- Strauss, R., Angloher, G., Bento, A., Bucci, C., Canonica, L., Defay, X., et al. (2016). The cressT-III low-mass wimp detector. *J. Phys.* 718:042048. doi: 10.1088/1742-6596/718/4/042048
- Strigari, L. E., and Trotta, R. (2009). Reconstructing WIMP properties in direct detection experiments including galactic dark matter distribution uncertainties. *J. Cosmol. Astropart. Phys.* 0911:019. doi: 10.1088/1475-7516/2009/11/019
- SuperCDMS Collaboration, Agnese, R., Anderson, A. J., Aralis T., Aramaki, T., Arnquist, I. J., et al. (2018). Low-mass dark matter search with CDMSlite. *Phys. Rev. D* 97:022002. doi: 10.1103/PhysRevD.97.022002
- Toma, T. (2013). Internal bremsstrahlung signature of real scalar dark matter and consistency with thermal relic density. *Phys. Rev. Lett.* 111:091301. doi: 10.1103/PhysRevLett.111.091301
- Wang, Z. W., Steele, T. G., Hanif, T., and Mann, R. B. (2016). Conformal complex singlet extension of the standard model: scenario for Dark Matter and a Second Higgs Boson. *J. High Energy Phys.* 8:065. doi: 10.1007/JHEP08(2016)065
- Weisz, D. R., Dalcanton, J. J., Williams, B. F., Gilbert, K. M., Skillman, E. D., Seth, A. C., et al. (2011). The ACS Nearby Galaxy Survey Treasury VIII. The Global Star Formation Histories of 60 Dwarf Galaxies in the Local Volume. *Astrophys. J.* 739:5. doi: 10.1088/0004-637X/739/1/5
- Yang, D., and Li, Q. (2018). Probing the Dark Sector through Mono-Z Boson Leptonic Decays. *J. High Energy Phys.* 2:090. doi: 10.1007/JHEP02(2018)090
- Yu, J. H. (2014). Vector fermion-portal dark Matter: direct detection and galactic center gamma-ray excess. *Phys. Rev. D* 90:095010. doi: 10.1103/PhysRevD.90.095010
- Zhang, C., Cui, M. Y., Feng, L., Fan, Y. Z., and Ren, Z. Z. (2017).  $\gamma$ -ray emission signals in the massive graviton mediated dark matter model. *Nucl. Phys. B* 916, 208–218. doi: 10.1016/j.nuclphysb.2017.01.003
- Zhen, C. (2014). LHAASO: Science and Status. *Frascati Phys. Ser.* 58:331.
- Zhu, R., and Zhang, Y. (2017). Graviton-mediated dark matter model explanation the DAMPE electron excess and search at  $e^+e^-$  colliders. *arXiv: 1712.01143*.
- Zwicky, F. (1933). Spectral displacement of extra galactic nebulae. *Helv. Phys. Acta* 6, 110–127.

**Conflict of Interest Statement:** The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Arina. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Neutrino Mass Ordering From Oscillations and Beyond: 2018 Status and Future Prospects

Pablo F. de Salas, Stefano Gariazzo, Olga Mena\*, Christoph A. Ternes and Mariam Tórtola

*Instituto de Física Corpuscular, CSIC-Universitat de València, Valencia, Spain*

## OPEN ACCESS

### Edited by:

Alberto Salvio,  
European Organization for Nuclear  
Research (CERN), Switzerland

### Reviewed by:

Thomas Schwetz,  
Karlsruher Institut für Technologie  
(KIT), Germany  
Orlando Luis Goulart Peres,  
Universidade Estadual de Campinas,  
Brazil

### \*Correspondence:

Olga Mena  
omena@ific.uv.es

### Specialty section:

This article was submitted to  
High-Energy and Astroparticle  
Physics,  
a section of the journal  
Frontiers in Astronomy and Space  
Sciences

**Received:** 28 July 2018

**Accepted:** 12 September 2018

**Published:** 09 October 2018

### Citation:

de Salas PF, Gariazzo S, Mena O,  
Ternes CA and Tórtola M (2018)  
Neutrino Mass Ordering From  
Oscillations and Beyond: 2018 Status  
and Future Prospects.  
Front. Astron. Space Sci. 5:36.  
doi: 10.3389/fspas.2018.00036

The ordering of the neutrino masses is a crucial input for a deep understanding of flavor physics, and its determination may provide the key to establish the relationship among the lepton masses and mixings and their analogous properties in the quark sector. The extraction of the neutrino mass ordering is a data-driven field expected to evolve very rapidly in the next decade. In this review, we both analyse the present status and describe the physics of subsequent prospects. Firstly, the different current available tools to measure the neutrino mass ordering are described. Namely, reactor, long-baseline (accelerator and atmospheric) neutrino beams, laboratory searches for beta and neutrinoless double beta decays and observations of the cosmic background radiation and the large scale structure of the universe are carefully reviewed. Secondly, the results from an up-to-date comprehensive global fit are reported: the Bayesian analysis to the 2018 publicly available oscillation and cosmological data sets provides *strong* evidence for the normal neutrino mass ordering vs. the inverted scenario, with a significance of 3.5 standard deviations. This preference for the normal neutrino mass ordering is mostly due to neutrino oscillation measurements. Finally, we shall also emphasize the future perspectives for unveiling the neutrino mass ordering. In this regard, apart from describing the expectations from the aforementioned probes, we also focus on those arising from alternative and novel methods, as 21 cm cosmology, core-collapse supernova neutrinos and the direct detection of relic neutrinos.

**Keywords:** neutrino mass ordering, neutrino oscillations, neutrinoless double beta ( $0\nu\beta\beta$ ) decay, large scale structure formation, cosmic microwave Background (CMB), neutrino masses and flavor mixing

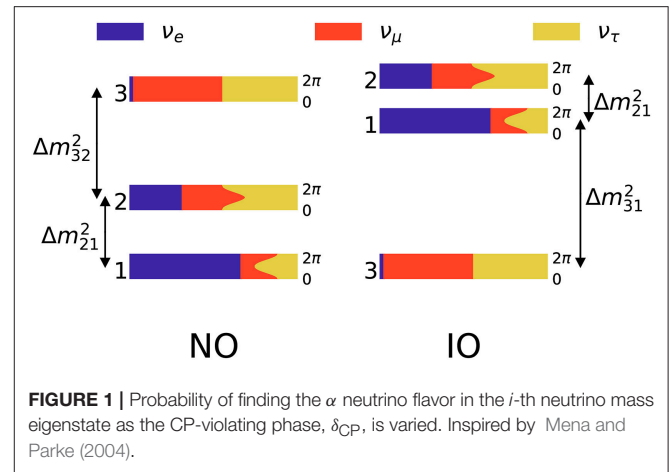
## 1. INTRODUCTION

The Royal Swedish Academy of Sciences decided to award the 2015 Nobel Prize in Physics to Takaaki Kajita and Arthur B. McDonald “for the discovery of neutrino oscillations, which shows that neutrinos have mass. [...] New discoveries about the deepest neutrino secrets are expected to change our current understanding of the history, structure and future fate of the Universe” (see Fukuda et al., 1998; Ahmad et al., 2001, 2002; Eguchi et al., 2003; Abe et al., 2011a; An et al., 2012) for essential publications. These discoveries robustly established that neutrinos are massive particles. However, neutrinos are massless particles in the Standard Model (SM) of particle physics: in the absence of any direct indication for their mass available at the time, they were introduced as fermions for which no gauge invariant renormalizable mass term can be constructed. As a consequence, in the SM there is neither mixing nor CP violation in the lepton sector. Therefore, neutrino oscillations and masses *imply the first known departure from the SM of particle physics*.

Despite the good precision that neutrino experiments have reached in the recent years, still many neutrino properties remain unknown. Among them, the neutrino character, Dirac vs. Majorana, the existence of CP violation in the leptonic sector, the absolute scale of neutrino masses, and the type of the neutrino mass spectrum. Future laboratory, accelerator and reactor, astrophysical and cosmological probes will address all these open questions, that may further reinforce the evidence for physics beyond the SM. The main focus of this review is, however, the last of the aforementioned unknowns. We will discuss what we know and how we could improve our current knowledge of the neutrino mass ordering.

Neutrino oscillation physics is only sensitive to the squared mass differences ( $\Delta m_{ij}^2 = m_i^2 - m_j^2$ ). Current oscillation data can be remarkably well-fitted in terms of two squared mass differences, dubbed as the solar mass splitting ( $\Delta m_{21}^2 \simeq 7.6 \times 10^{-5} \text{ eV}^2$ ) and the atmospheric mass splitting ( $|\Delta m_{31}^2| \simeq 2.5 \times 10^{-3} \text{ eV}^2$ ) (de Salas et al., 2018)<sup>1</sup>. Thanks to matter effects in the Sun, we know that  $\Delta m_{21}^2 > 0^2$ . Since the atmospheric mass splitting  $\Delta m_{31}^2$  is essentially measured only via neutrino oscillations in vacuum, which exclusively depend on its absolute value, its sign is unknown at the moment. As a consequence, we have two possibilities for the ordering of neutrino masses: *normal ordering* (NO,  $\Delta m_{31}^2 > 0$ ) or *inverted ordering* (IO,  $\Delta m_{31}^2 < 0$ ).

The situation for the mass ordering has changed a lot in the last few months. The 2017 analyses dealing with global oscillation neutrino data have only shown a mild preference for the normal ordering. Namely, the authors of Capozzi et al. (2017), by means of a frequentist analysis, found  $\chi_{\text{IO}}^2 - \chi_{\text{NO}}^2 = 3.6$  from all the oscillation data considered in their analyses. Very similar results were reported in the first version of de Salas et al. (2018)<sup>3</sup>, where a value of  $\chi_{\text{IO}}^2 - \chi_{\text{NO}}^2 = 4.3$  was quoted<sup>4</sup> (nuFIT)<sup>5</sup>. Furthermore, in Gariazzo et al. (2018a), the authors verified that the use of a Bayesian approach and the introduction of cosmological or neutrinoless double beta decay data did not alter the main result, which was a weak-to-moderate evidence for the normal neutrino mass ordering according to the Jeffreys' scale (see Table 2). The most recent global fit to neutrino oscillation data, however, reported a strengthened preference for normal ordering that is mainly due to the new data from the Super-Kamiokande Abe et al. (2018a), T2K Hartz (2017), and NOvA Radovic (2018) experiments. The inclusion of these new data in both the analyses of Capozzi et al. (2018a) and the 2018 update of de Salas et al. (2018)<sup>1</sup> increases the preference for normal ordering, which now lies mildly above the  $3\sigma$  level. In this review we will comment these new results (see section 2) and use them to perform an updated global



analysis, following the method of Gariazzo et al. (2018a) (see section 5).

The two possible hierarchical<sup>6</sup> neutrino mass scenarios are shown in Figure 1, inspired by Mena and Parke (2004), which provides a graphical representation of the neutrino flavor content of each of the neutrino mass eigenstates given the current preferred values of the oscillation parameters de Salas et al. (2018), see section 2. At present, even if the current preferred value of  $\delta_{\text{CP}}$  for both normal and inverted mass orderings lies close to  $3\pi/2$  de Salas et al. (2018), the precise value of the CP violating phase in the leptonic sector remains unknown. Consequently, in Figure 1, we have varied  $\delta_{\text{CP}}$  within its entire range, ranging from 0 to  $2\pi$ .

Given the two known mass splittings that oscillation experiments provide us, we are sure that at least two neutrinos have a mass above  $\sqrt{\Delta m_{21}^2} \simeq 8 \text{ meV}$  and that at least one of these two neutrinos has a mass larger than  $\sqrt{|\Delta m_{31}^2|} \simeq 50 \text{ meV}$ . For the same reason, we also know that there exists a lower bound on the sum of the three active neutrino masses ( $\sum m_\nu = m_1 + m_2 + m_3$ ):

$$\begin{aligned} \sum m_\nu^{\text{NO}} &= m_1 + \sqrt{m_1^2 + \Delta m_{21}^2} + \sqrt{m_1^2 + \Delta m_{31}^2}, \\ \sum m_\nu^{\text{IO}} &= m_3 + \sqrt{m_3^2 + |\Delta m_{31}^2|} + \sqrt{m_3^2 + |\Delta m_{31}^2| + \Delta m_{21}^2}, \end{aligned} \quad (1)$$

where the lightest neutrino mass eigenstate corresponds to  $m_1$  ( $m_3$ ) in the normal (inverted) ordering. Using the best-fit values for the neutrino mass splittings in Table 1 one finds that  $\sum m_\nu \gtrsim 0.06 \text{ eV}$  in normal ordering, while  $\sum m_\nu \gtrsim 0.10 \text{ eV}$  in inverted

<sup>1</sup> Valencia-Globalfit, 2018; Available online at: <http://globalfit.astroparticles.es/>.

<sup>2</sup> Note that the observation of matter effects in the Sun constrains the product  $\Delta m_{21}^2 \cos 2\theta_{12}$  to be positive. Therefore, depending on the convention chosen to describe solar neutrino oscillations, matter effects either fix the sign of the solar mass splitting  $\Delta m_{21}^2$  or the octant of the solar angle  $\theta_{12}$ , with  $\Delta m_{21}^2$  positive by definition.

<sup>3</sup> See the "July 2017" version in<sup>1</sup>.

<sup>4</sup> A somewhat milder preference in favor of normal mass ordering was obtained in the corresponding version of the analysis in Refs. Esteban et al. (2017)

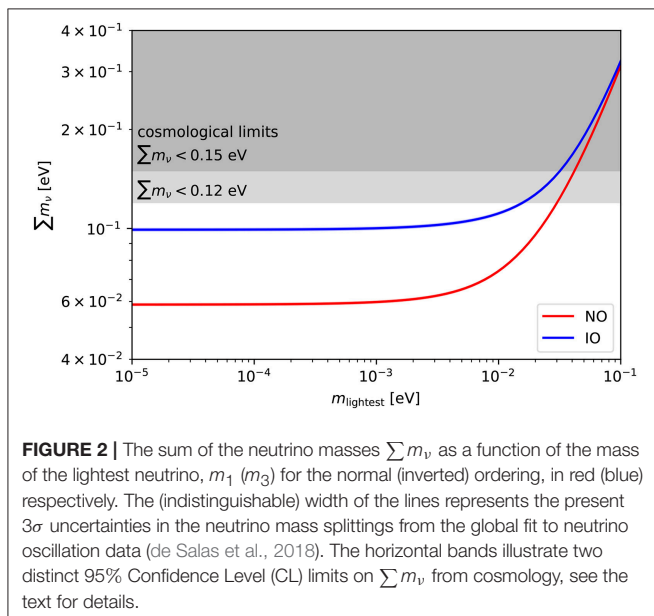
<sup>5</sup> NuFIT v3.2, <http://www.nu-fit.org/>.

<sup>6</sup> A clarification about the use of "hierarchy" and "ordering" is mandatory. One talks about "hierarchy" when referring to the absolute scales of neutrino masses, in the sense that neutrino masses can be distinguished and ranked from lower to higher. This does not include the possibility that the lightest neutrino mass is much larger than the mass splittings obtained by neutrino oscillation measurements, since in this case the neutrino masses are degenerate. On the other hand, the mass "ordering" is basically defined by the sign of  $\Delta m_{31}^2$ , or by the fact that the lightest neutrino is the most (least) coupled to the electron neutrino flavor in the normal (inverted) case.

**TABLE 1** | Neutrino oscillation parameters summary determined from the global analysis.

parameter	Best-fit $\pm 1\sigma$	$2\sigma$ range	$3\sigma$ range
$\Delta m_{21}^2$ [ $10^{-5}\text{eV}^2$ ]	$7.55^{+0.20}_{-0.16}$	7.20–7.94	7.05–8.14
$ \Delta m_{31}^2 $ [ $10^{-3}\text{eV}^2$ ] (NO)	$2.50 \pm 0.03$	2.44–2.57	2.41–2.60
$ \Delta m_{31}^2 $ [ $10^{-3}\text{eV}^2$ ] (IO)	$2.42^{+0.03}_{-0.04}$	2.34–2.47	2.31–2.51
$\sin^2 \theta_{12}/10^{-1}$	$3.20^{+0.20}_{-0.16}$	2.89–3.59	2.73–3.79
$\sin^2 \theta_{23}/10^{-1}$ (NO)	$5.47^{+0.20}_{-0.30}$	4.67–5.83	4.45–5.99
$\sin^2 \theta_{23}/10^{-1}$ (IO)	$5.51^{+0.18}_{-0.30}$	4.91–5.84	4.53–5.98
$\sin^2 \theta_{13}/10^{-2}$ (NO)	$2.160^{+0.083}_{-0.069}$	2.03–2.34	1.96–2.41
$\sin^2 \theta_{13}/10^{-2}$ (IO)	$2.220^{+0.074}_{-0.076}$	2.07–2.36	1.99–2.44
$\delta_{\text{CP}}/\pi$ (NO)	$1.32^{+0.21}_{-0.15}$	1.01–1.75	0.87–1.94
$\delta_{\text{CP}}/\pi$ (IO)	$1.56^{+0.13}_{-0.15}$	1.27–1.82	1.12–1.94

The results for inverted mass ordering were calculated with respect to this mass ordering.



**FIGURE 2** | The sum of the neutrino masses  $\sum m_\nu$ , as a function of the mass of the lightest neutrino,  $m_1$  ( $m_3$ ) for the normal (inverted) ordering, in red (blue) respectively. The (indistinguishable) width of the lines represents the present  $3\sigma$  uncertainties in the neutrino mass splittings from the global fit to neutrino oscillation data (de Salas et al., 2018). The horizontal bands illustrate two distinct 95% Confidence Level (CL) limits on  $\sum m_\nu$ , from cosmology, see the text for details.

ordering. **Figure 2** illustrates the values of  $\sum m_\nu$  as a function of the lightest neutrino mass for the two possible ordering schemes. We also show the two representative bounds on the sum of the neutrino masses from cosmology (discussed later in section 4) which is currently providing the strongest limits on  $\sum m_\nu$  thanks to the fact that neutrinos affect both the evolution of the cosmological background and perturbation quantities (see e.g., the excellent detailed reviews of Lesgourgues and Pastor, 2006, 2012, 2014; Lesgourgues et al., 2013; Lattanzi and Gerbino, 2018).

The state-of-knowledge of cosmological observations Ade et al. (2016b) points to a flat Universe whose mass-energy density includes 5% of ordinary matter (baryons), 22% non-baryonic *dark matter*, and that is dominated by the *dark energy*, identified as the motor for the accelerated expansion. This is the so-called  $\Lambda$ CDM Universe, which fits extremely well the Cosmic

Microwave Background (CMB) fluctuations, distant Supernovae Ia and galaxy clustering data.

Using the known neutrino oscillation parameters and the standard cosmological evolution, it is possible to compute the thermalization and the decoupling of neutrinos in the early universe (see e.g., Mangano et al., 2005; de Salas and Pastor, 2016). While neutrinos decoupled as ultra-relativistic particles, currently at least two out of the three neutrino mass eigenstates are non-relativistic. Neutrinos constitute the first and only known form of dark matter so far. Indeed, neutrinos behave as *hot* dark matter particles, possessing large thermal velocities, clustering only at scales below their free streaming scale, modifying the evolution of matter overdensities and suppressing structure formation at small scales. The CMB is also affected by the presence of massive neutrinos, as these particles may turn non-relativistic around the decoupling period. However, the strong degeneracy between the Hubble constant and the total neutrino mass requires additional constraints (from Baryon Acoustic Oscillations, Supernovae Ia luminosity distance data and/or direct measurements of the Hubble constant) to be added in the global analyses. In this regard, CMB lensing is also helpful and improves the CMB temperature and polarization constraints, as the presence of massive neutrinos modify the matter distribution along the line of sight through their free streaming nature, reducing clustering and, consequently, CMB lensing. The most constraining cosmological upper bounds to date on  $\sum m_\nu$  can be obtained combining CMB with different large scale structure observations and range from  $\sum m_\nu < 0.12$  eV to  $\sum m_\nu < 0.15$  eV at 95% CL (Palanque-Desabrouille et al., 2015; Cuesta et al., 2016; Di Valentino et al., 2016c; Giusarma et al., 2016; Vagnozzi et al., 2017, 2018; Lattanzi and Gerbino, 2018), as illustrated in **Figure 2**.

If the massive neutrino spectrum does not lie in the degenerate region, the three distinct neutrino masses affect the cosmological observables in a different way. For instance, the transition to the non-relativistic period takes place at different cosmic times, and the associated free-streaming scale is different for each of the neutrino mass eigenstates. However, the effect on the power spectrum is very small (permille level) and therefore an extraction of the neutrino mass hierarchy via singling out each of the massive neutrino states seems a very futuristic challenge. This will be possibly attainable only via huge effective volume surveys, as those tracing the 21 cm spin-flip transition in neutral hydrogen, see sections 6.4 and 6.5. On the other hand, should the cosmological measurements of  $\sum m_\nu$  be strong enough to rule out the  $\sum m_\nu$  parameter space corresponding to the inverted ordering (i.e., strong enough to establish in a very significant way that  $\sum m_\nu < 0.1$  eV), we would know that the neutrino mass ordering must be normal. A word of caution is needed here when dealing with Bayesian analyses, usually performed when dealing with cosmological data: a detection of the neutrino mass ordering could be driven by volume effects in the marginalization, and therefore the prior choice can make a huge difference, if data are not powerful enough (Schwetz et al., 2017).

Another way to probe the neutrino mass ordering, apart from direct determinations of the sign of the atmospheric mass splitting  $\Delta m_{31}^2$  in neutrino oscillation experiments and,

indirectly, from cosmological bounds on the sum of the neutrino masses, is *neutrinoless double  $\beta$  decay* (Rodejohann, 2011; Gomez-Cadenas et al., 2012; Vergados et al., 2012; Dell’Oro et al., 2016). This process is a spontaneous nuclear transition in which the charge of two isobaric nuclei would change by two units with the simultaneous emission of two electrons and without the emission of neutrinos. This process is only possible if the neutrino is a Majorana particle and an experimental signal of the existence of this process would constitute evidence of the putative Majorana neutrino character. The non-observation of the process provides bounds on the so-called *effective Majorana mass*  $m_{\beta\beta}$ , which is a combination of the (Majorana) neutrino masses weighted by the leptonic flavor mixing effects (see section 3). **Figure 3** illustrates the (Bayesian) 95.5% and 99.7% credible intervals for  $m_{\beta\beta}$  as a function of the lightest neutrino mass in the case of three neutrino mixing, considering a logarithmic prior on the lightest neutrino mass. The picture differs from the plot that is usually shown, which features an open band toward increasingly smaller values of  $m_{\beta\beta}$  for  $m_{\text{lightest}} \simeq 5$  meV, due to cancellations which depend on the values of the Majorana phases  $\alpha_i$  (see section 3). In the Bayesian sense of credible intervals, the values of  $\alpha_i$  which produce such a suppression of  $m_{\beta\beta}$  represent an extremely small fraction of the parameter space, which is therefore not relevant when computing the 95.5% and 99.7% credible intervals. In other words, given our knowledge of the neutrino mixing parameters, having  $m_{\beta\beta} \lesssim 2 \times 10^{-4}$  eV would require some amount of fine tuning in the Majorana phases. This figure is in perfect agreement with the results shown in Figure 1 of Agostini et al. (2017a), which shows that most of the allowed parameter space is not concentrated at small  $m_{\beta\beta}$  if one considers a linear prior on the lightest neutrino mass. We also show the most conservative version of some of the most competitive current limits, as those from KamLAND-Zen ( $m_{\beta\beta} < 61 - 165$  meV at 90% CL) Gando et al. (2016), GERDA Phase II ( $m_{\beta\beta} < 120 - 260$  meV at 90% CL) Agostini et al. (2018) and CUORE ( $m_{\beta\beta} < 110 - 520$  meV at 90% CL) Alduino et al. (2018a). Please note that a detection of the effective Majorana mass will not be sufficient to determine the mass ordering if the lightest neutrino mass is above  $\sim 40$  meV: in this case, indeed, the normal and the inverted ordering become indistinguishable from the point of view of neutrinoless double beta decay. Similarly to the case of the cosmological bounds on the neutrino mass  $\sum m_\nu$ , in which only constraining  $\sum m_\nu$  to be below 0.1 eV could be used to disfavor the inverted mass ordering, only a limit on  $m_{\beta\beta}$  below  $\sim 10$  meV could be used to rule out the inverted ordering scheme, and only assuming that neutrinos are Majorana particles.

Since neutrino oscillation measurements, cosmological observations and neutrinoless double beta decay experiments are cornering the inverted mass ordering region, it makes sense to combine their present results. Indeed, plenty of works have been recently devoted to test whether a preference for one mass ordering over the other exists, given current oscillation, neutrinoless double beta decay and cosmological data. A number of studies on the subject (Hannestad and Schwetz, 2016; Caldwell et al., 2017; Capozzi et al., 2017; Gerbino et al., 2017b; Wang and Xia, 2018) found that the preference for the normal vs. the inverted mass scenario is rather mild with

current data, regardless the frequentist vs. Bayesian approach. In the latter case, however, the results may be subject-dependent, as a consequence of different possible choices of priors and parameterizations when describing the theoretical model, for example in the case of sampling over the three individual neutrino mass states. Therefore, one must be careful when playing with different priors, as recently shown in Gariazzo et al. (2018a). The current status of the preference of normal vs. inverted ordering will be further investigated carefully throughout this review. Furthermore, as it will be carefully detailed in section 5, the Bayesian global fit to the 2018 publicly available oscillation and cosmological data points to a strong preference (3.5 standard deviations) for the normal neutrino mass ordering vs. the inverted one.

To summarize and conclude this introductory part, we resume that the current available methods to determine the neutrino mass ordering can be grouped as:

- a) neutrino oscillation facilities;
- b) neutrinoless double beta decay experiments, with the caveat that the results will only apply in case neutrinos are Majorana fermions;
- c) CMB and large scale structure surveys.

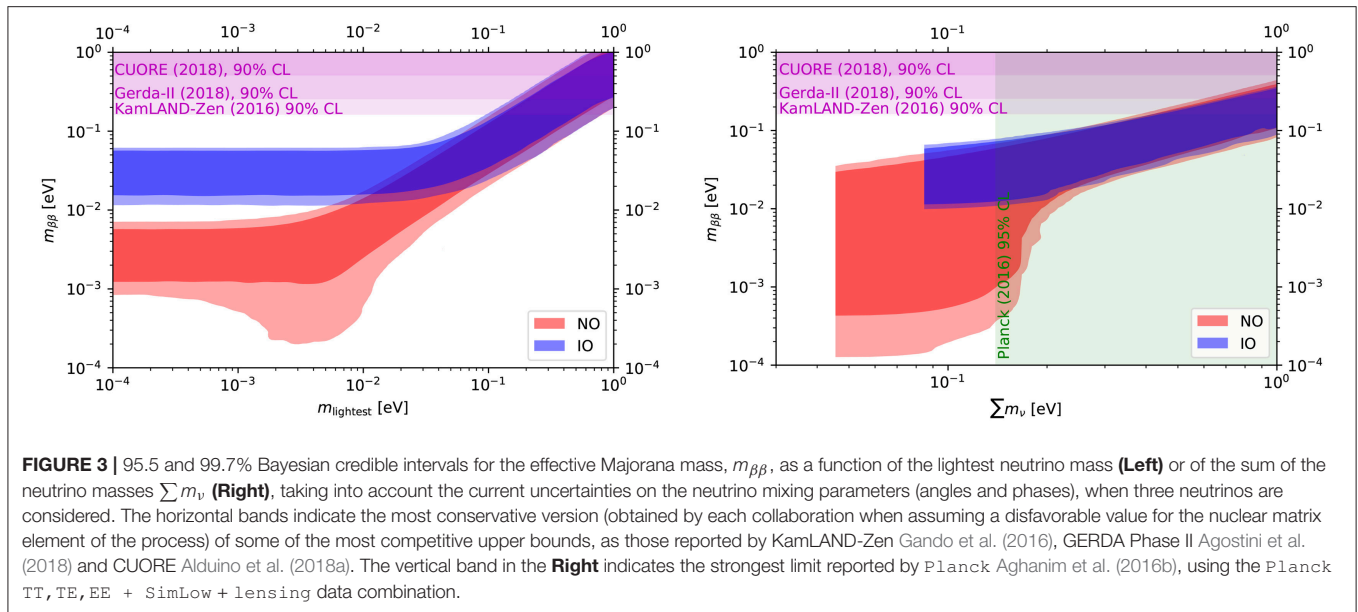
For each of these three categories we will review the current status and also analyse the future prospects, with a particular focus on the existing experiments which will be improved in the future and on new facilities which aim at determining the neutrino mass ordering in the next 22 years<sup>7</sup> In the second part of this review we will also focus on possible novel methods that in the future will enable us to determine the neutrino mass ordering, as for example future cosmological observations of the 21 cm line, the detection of neutrinos emitted by core-collapse supernovae, measurements of the electron spectrum of  $\beta$ -decaying nuclei and the direct detection of relic neutrinos.

We shall exploit the complementarity of both cosmology and particle physics approaches, profiting from the highly multidisciplinary character of the topic. We dedicate sections 2, 3, and 4 to explain the extraction of the neutrino mass ordering via neutrino oscillations,  $\beta$  and neutrinoless double  $\beta$  decays and cosmological observations, which will be combined in section 5 where we present the analysis of current data related to these three data sets. Future perspectives are described throughout section 6 and its subsections, while the final remarks will be outlined in section 7.

## 2. NEUTRINO OSCILLATIONS

Our current knowledge on the neutrino mass ordering comes mainly from the analysis of the available neutrino oscillation data. The sensitivity to the neutrino mass spectrum at oscillation experiments is mostly due to the presence of matter effects in the neutrino propagation. Therefore, one can expect that this sensitivity will increase with the size of matter effects, being larger for atmospheric neutrino experiments, where a fraction of

<sup>7</sup>See also the review Qian and Vogel (2015), focused mostly on neutrino oscillation perspectives.



neutrinos travel through the Earth. For long-baseline accelerator experiments, matter effects will increase with the baseline, while these effects will be negligible at short-baseline and medium-baseline experiments.

When neutrinos travel through the Earth, the effective matter potential due to the electron (anti)neutrino charged-current elastic scatterings with the electrons in the medium will modify the three-flavor mixing processes. The effect will strongly depend on the neutrino mass ordering: in the normal (inverted) mass ordering scenario, the neutrino flavor transition probabilities will get enhanced (suppressed). In the case of antineutrino propagation, instead, the flavor transition probabilities will get suppressed (enhanced) in the normal (inverted) mass ordering scenario. This is the Wolfenstein effect (Wolfenstein, 1978), later expanded by Mikheev and Smirnov (Mikheev and Smirnov, 1985, 1986), and commonly named as the Mikheev-Smirnov-Wolfenstein (MSW) effect (see e.g., Blennow and Smirnov, 2013 for a detailed description of neutrino oscillations in matter).

Matter effects in long-baseline accelerator or atmospheric neutrino oscillation experiments depend on the size of the effective mixing angle  $\theta_{13}$  in matter, which leads the transitions  $\nu_e \leftrightarrow \nu_{\mu,\tau}$  governed by the atmospheric mass-squared difference  $\Delta_{31} = \Delta m_{31}^2/2E$ . Within the simple two-flavor mixing framework, the effective  $\theta_{13}$  angle in matter reads as

$$\sin^2 2\theta_{13}^m = \frac{\sin^2 2\theta_{13}}{\sin^2 2\theta_{13} + \left(\cos 2\theta_{13} \mp \frac{\sqrt{2}G_F N_e}{\Delta_{31}}\right)^2}, \quad (2)$$

where the minus (plus) sign refers to neutrinos (antineutrinos) and  $N_e$  is the electron number density in the Earth interior. The neutrino mass ordering fixes the sign of  $\Delta_{31}$ , that is positive (negative) for normal (inverted) ordering: notice that, in the presence of matter effects, the neutrino (antineutrino) oscillation probability  $P(\nu_\mu \rightarrow \nu_e)$  [ $P(\bar{\nu}_\mu \rightarrow \bar{\nu}_e)$ ] gets enhanced if the

ordering is normal (inverted). Exploiting the different matter effects for neutrinos and antineutrinos provides therefore the ideal tool to unravel the mass ordering.

Matter effects are expected to be particularly relevant when the following resonance condition is satisfied:

$$\Delta m_{31}^2 \cos 2\theta_{13} = 2\sqrt{2}G_F N_e E. \quad (3)$$

The precise location of the resonance will depend on both the neutrino path and the neutrino energy. For instance, for  $\Delta m_{31}^2 \sim 2.5 \times 10^{-3} \text{ eV}^2$  and distances of several thousand kilometers, as it is the case of atmospheric neutrinos, the resonance effect is expected to happen for neutrino energies  $\sim 3 - 8 \text{ GeV}$ .

In the case of muon disappearance experiments, in the  $\sim \text{GeV}$  energy range relevant for long-baseline and atmospheric neutrino beams, the  $P_{\mu\mu}$  survival probabilities are suppressed (enhanced) due to matter effects if the ordering is normal (inverted). If the matter density is constant, the  $P_{\mu\mu}$  survival probability at terrestrial baselines<sup>8</sup> is given by

$$P_{\mu\mu} = 1 - \cos^2 \theta_{13}^m \sin^2 2\theta_{23} \times \sin^2 \left[ 1.27 \left( \frac{\Delta m_{31}^2 + A + (\Delta m_{31}^2)^m}{2} \right) \frac{L}{E} \right] - \sin^2 \theta_{13}^m \sin^2 2\theta_{23} \times \sin^2 \left[ 1.27 \left( \frac{\Delta m_{31}^2 + A - (\Delta m_{31}^2)^m}{2} \right) \frac{L}{E} \right] - \sin^4 \theta_{23} \sin^2 2\theta_{13}^m \sin^2 \left[ 1.27 (\Delta m_{31}^2)^m \frac{L}{E} \right], \quad (4)$$

<sup>8</sup>For an expansion including also the solar mixing parameters, see Ref. Akhmedov et al. (2004).

where  $A = 2\sqrt{2}G_F N_e E$ ,  $\theta_{13}^m$  is that of Equation (2) and

$$(\Delta m_{31}^2)^m = \Delta m_{31}^2 \sqrt{\sin^2 2\theta_{13} + \left( \cos 2\theta_{13} \mp \frac{2\sqrt{2}G_F N_e E}{\Delta m_{31}^2} \right)^2}. \quad (5)$$

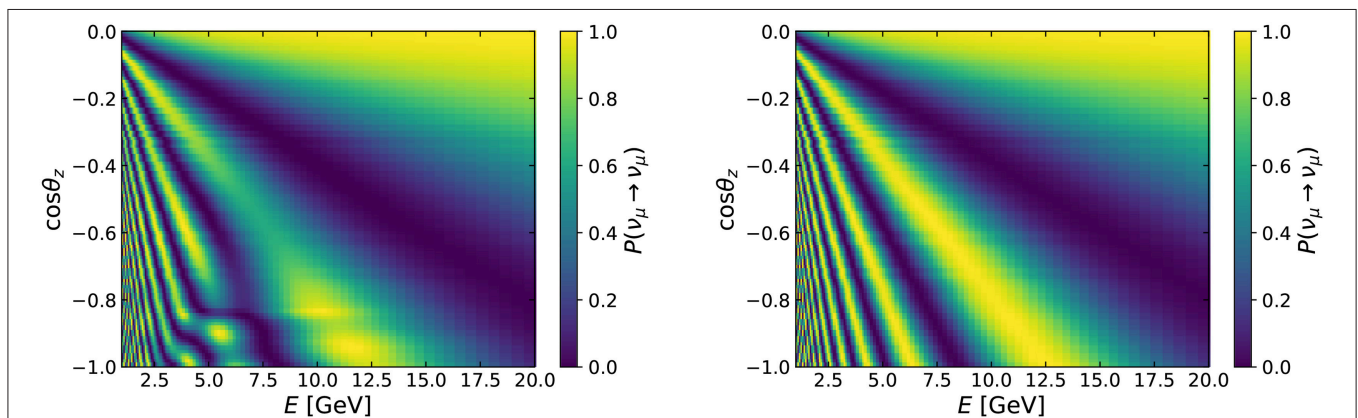
The dependence of the survival probability  $P_{\mu\mu}$  on the neutrino energy  $E$  and the cosine of the zenith angle  $\cos\theta_z$ , related to the distance the atmospheric neutrinos travel inside the Earth before being detected at the experiments, is shown in **Figure 4** for normal (**Left**) and inverted (**Right**) ordering. There, we can see that reconstructing the oscillation pattern at different distances and energies allows to determine the neutrino mass ordering (see also section 6.1).

Until very recently, oscillation experiments were not showing a particular preference for any of the mass orderings, not even when combined in a global analysis (see for instance Forero et al., 2014). Lately, however, the most recent data releases from some of the experiments have become more sensitive to the ordering of the neutrino mass spectrum. In particular, the long-baseline experiments T2K and NOvA on their own obtain a slight preference in favor of normal mass ordering, with  $\Delta\chi^2 \approx 4$  each (Hartz, 2017; Radovic, 2018). Note that these results have been obtained imposing a prior on the mixing angle  $\theta_{13}$ , according to its most recent determination at reactor experiments. Relaxing the prior on the reactor angle results in a milder preference for normal over inverted mass ordering. The latest atmospheric neutrino results from Super-Kamiokande also show some sensitivity to the neutrino mass ordering. In this case, the collaboration obtains a preference for normal ordering with  $\Delta\chi^2 \approx 3.5$ , without any prior on the reactor angle. Constraining the value of  $\theta_{13}$ , the preference for normal mass ordering increases up to  $\Delta\chi^2 \approx 4.5$  Abe et al. (2018a).

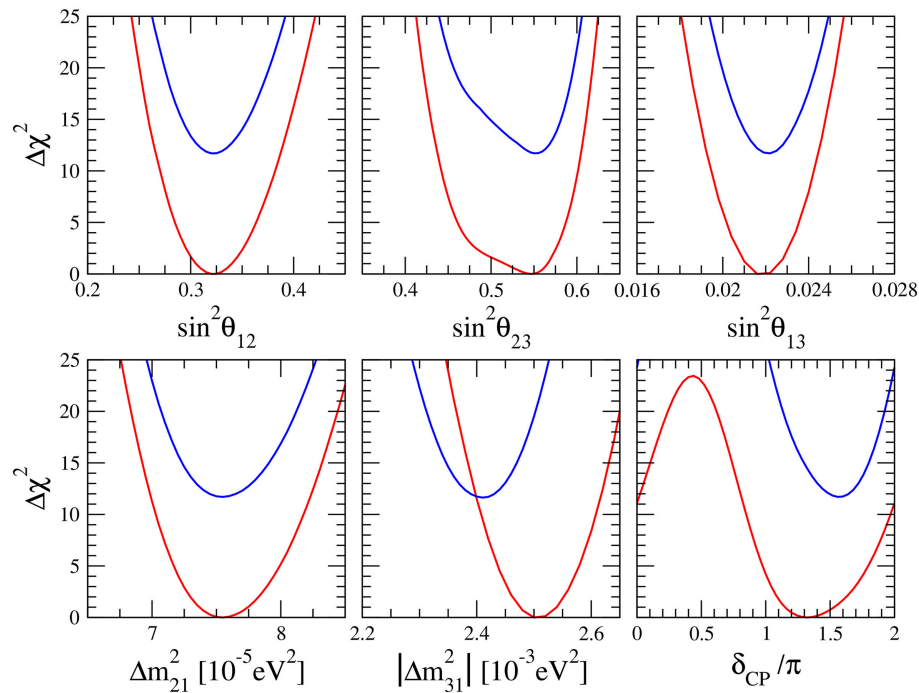
The full sensitivity to the ordering of the neutrino mass spectrum from oscillations is obtained after combining the data samples described above with all the available experimental results in a global fit (de Salas et al., 2018). This type of analysis exploits the complementarity among the different results as well as the correlations among the oscillation parameters to

obtain improved sensitivities on them. In the global analysis to neutrino oscillations, the parameters  $\sin^2 \theta_{12}$  and  $\Delta m_{21}^2$  are rather well measured by the solar experiments (Cleveland et al., 1998; Hosaka et al., 2006; Aharmim et al., 2008, 2010; Cravens et al., 2008; Abdurashitov et al., 2009; Kaether et al., 2010; Abe et al., 2011b; Bellini et al., 2014; Nakano, 2016) and the long-baseline reactor experiment KamLAND Gando et al. (2011). The short-baseline reactor neutrino experiments Daya Bay An et al. (2017), RENO Pac (2018) and Double Chooz Abe et al. (2014) are the most efficient ones in measuring the reactor angle  $\theta_{13}$  and also measure very well the atmospheric mass splitting,  $\Delta m_{31}^2$ . Notice however that the atmospheric mass splitting is best measured by the combined data from MINOS (beam and atmospheric) and MINOS+, as shown in Adam (2018). This mass splitting is also measured, together with the atmospheric angle  $\theta_{23}$ , by the atmospheric experiments IceCube-DeepCore Aartsen et al. (2015), ANTARES Adrian-Martinez et al. (2012) and Super-Kamiokande Abe et al. (2018a), where the latter shows some sensitivity to  $\theta_{13}$  and  $\delta_{CP}$ , too. The long-baseline accelerator experiments are also sensitive to these four parameters through their appearance and disappearance neutrino channels. Apart from the already mentioned T2K Hartz (2017) and NOvA Radovic (2018), the global fit also includes the previous experiments K2K Ahn et al. (2006) and MINOS Adamson et al. (2014).

The result of the global analysis is summarized in **Table 1** and **Figure 5**. Before discussing the sensitivity to the neutrino mass ordering, we shall briefly discuss some other features of this global fit. Notice first that now the best-fit value for the atmospheric mixing angle  $\theta_{23}$  lies in the second octant, although values in the first octant are still allowed with  $\Delta\chi^2 = 1.6$  (3.2) for normal (inverted) ordering. Therefore, the octant problem remains unsolved so far. Note also that, for the first time, the CP violating phase  $\delta_{CP}$  is determined with rather good accuracy. The best-fit values for this parameter lie close to maximal CP violation, being  $\delta_{CP} = 1.32\pi$  for normal ordering and  $\delta_{CP} = 1.56\pi$  for inverted ordering. As can be seen from the  $\Delta\chi^2$  profile in **Figure 5**, values around  $\delta_{CP} \approx 0.5\pi$  are now highly disfavored by data. Indeed, only around 50% of the



**FIGURE 4 |** Survival probability  $P_{\mu\mu}$ , as a function of the neutrino energy  $E$  and the cosine of the zenith angle  $\cos\theta_z$ , for normal (inverted) ordering in the (**Left**, **Right**).



**FIGURE 5 |** Summary of neutrino oscillation parameters, 2018. Red (blue) lines correspond to normal ordering (inverted ordering). Notice that the  $\Delta\chi^2$  profiles for inverted ordering are plotted with respect to the minimum for normal neutrino mass ordering.

parameter space remains allowed at the  $3\sigma$  level, roughly the interval  $[0.9\pi, 1.9\pi]$  for normal and  $[1.1\pi, 1.9\pi]$  for inverted ordering. In the case of normal ordering, CP conservation remains allowed at  $2\sigma$ , while it is slightly more disfavored for inverted ordering. For the remaining oscillation parameters, one clearly sees that neutrino oscillations are entering the precision era, with relative uncertainties on their determination of 5% or below. For a more detailed discussion about these parameters we refer the reader to de Salas et al. (2018) and<sup>1</sup>.

Concerning the neutrino mass ordering, we obtain a global preference of  $3.4\sigma$  ( $\Delta\chi^2 = 11.7$ ) in favor of normal ordering. This result emerges from the combination of all the neutrino oscillation experiments, as we explain in the following. Starting with long-baseline data alone, the inverted mass ordering is disfavored with  $\Delta\chi^2 = 2.0$ , when no prior is considered on the value of  $\theta_{13}$ . However, as explained above, the separate analysis of the latest T2K and NOvA data independently report a  $\Delta\chi^2 \approx 4$  among the two possible mass orderings when a prior on the reactor angle is imposed. This comes from the mismatch between the value of  $\theta_{13}$  preferred by short-baseline reactor and long-baseline accelerator experiments, which is more important for inverted ordering. Besides that, the combination of T2K and reactor data results in an additional tension relative to the preferred value of the atmospheric mass splitting  $\Delta m_{31}^2$ , which is again larger for the inverted mass ordering. This further discrepancy results in a preference for normal ordering with  $\Delta\chi^2 = 5.3$  for the combination of “T2K plus reactors” and  $\Delta\chi^2 = 3.7$  for the combination of “NOvA plus reactors”. From

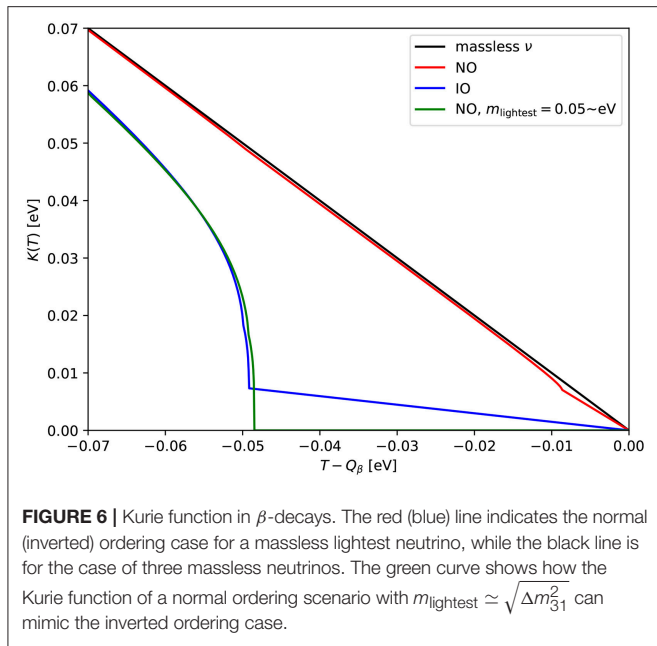
the combined analysis of all long-baseline accelerator and short-baseline reactor data one obtains a  $\Delta\chi^2 = 7.5$  between normal and inverted ordering, which corresponds to a preference of  $2.7\sigma$  in favor of normal mass ordering. By adding the atmospheric data to the neutrino oscillations fit, we finally obtain  $\Delta\chi^2 = 11.7^9$ , indicating a global preference for normal ordering at the level of  $3.4\sigma$ .

### 3. MASS ORDERING AND DECAY EXPERIMENTS

#### 3.1. Mass Ordering Through $\beta$ -Decay Experiments

The most reliable method to determine the absolute neutrino masses in a completely model independent way is to measure the spectrum of  $\beta$ -decay near the endpoint of the electron spectrum. The reason for this is related to the fact that, if neutrinos are massive, part of the released energy must go into the neutrino mass and the electron spectrum endpoint shifts to lower energies. When there are more than one massive neutrino, each of the separate mass eigenstates contributes to the suppression of the electron energy spectrum and it becomes possible to study the pattern of the neutrino masses. Nowadays none of the  $\beta$ -decay experiments can reach the energy resolution required to be able

<sup>9</sup>Note that this extra sensitivity comes essentially from Super-Kamiokande, since the effect of IceCube DeepCore and ANTARES is negligible in comparison.



to determine the mass hierarchy<sup>10</sup>, but we will explain in the following how, in principle, future experiments may aim at such result.

The best way to depict the effects of the separate mass eigenstates is to compute the Kurie function for  $\beta$ -decay. The complete expression can be written as (see e.g., Giunti and Kim, 2007):

$$K(T) = \left[ (Q_\beta - T) \sum_{i=1}^N |U_{ei}|^2 \sqrt{(Q_\beta - T)^2 - m_i^2} \Theta(Q_\beta - T - m_i) \right]^{1/2}, \quad (6)$$

where  $Q_\beta$  is the  $Q$ -value of the considered  $\beta$ -decay,  $T$  is the electron kinetic energy,  $\Theta$  is the Heaviside step function and  $|U_{ei}|^2$  is the mixing matrix element that defines the mixing between the electron neutrino flavor and the  $i$ -th mass eigenstate with a mass  $m_i$ . The standard scenario features  $N = 3$ , but the formula is valid also if a larger number of neutrinos exists (i.e., if there are sterile neutrinos, as explained for example in Gariazzo et al., 2016).

The Kurie function of Equation (6) is depicted in **Figure 6**, where we show in red (blue) the result obtained using a massless lightest neutrino and the current best-fit mixing angles and mass splittings for normal (inverted) ordering, as described in the previous section. As a reference, we also plot  $K(T)$  for a case with massless neutrinos only (in black). Should we consider higher values for the lightest neutrino mass, the detection of the mass ordering would be increasingly more difficult, since

<sup>10</sup>In the case of quasi-degenerate spectrum, the distortion of the spectrum will consist of just a bending and a shift of the end point, similar to the case of an electron neutrino with a given mass without mixing Farzan et al. (2001), and the ordering cannot be measured. Therefore, for future  $\beta$ -decay searches, measuring the neutrino mass ordering will be practically the same as measuring the neutrino mass hierarchy.

the separation of the mass eigenstates would decrease, eventually becoming negligible in the degenerate case. For this reason we will only discuss the case of a massless lightest neutrino from the perspective of the  $\beta$ -decay experiments.

Given the unitarity of the mixing matrix ( $\sum_{i=1}^N |U_{ei}|^2 = 1$ ), the normalization of the Kurie function is the same at  $Q_\beta - T \gg m_i$ . Since we are interested in the small differences which appear near the endpoint, the plot only focuses on the very end of the electron spectrum and the common normalization is not visible for the inverted ordering case. In the considered range, however, the effect of the different correspondence between the mass eigenstates and the mixing matrix elements introduces a difference which in principle would allow to determine the mass ordering through the observation of the  $\beta$ -decay spectrum. The observation of the kinks in the electron spectrum is very challenging, especially in the case of normal ordering, for which even the more pronounced kink (at  $Q_\beta - T \simeq \sqrt{\Delta m_{21}^2} \simeq 8$  meV) is barely visible in **Figure 6**. In the case of inverted ordering, since the mass difference between the two lightest mass eigenstates is the largest possible one ( $\sqrt{\Delta m_{31}^2} \simeq 50$  meV), and the lightest neutrino is the one with the smallest mixing with the electron neutrino, the amplitude of the kink is much larger. As a consequence, an experiment with enough energy resolution to measure the spectrum in the relevant energy range can directly probe the mass ordering observing the presence of a kink. Note that this measurement can be obtained even without a detection of the lightest neutrino mass. As we show in **Figure 6**, however, it is crucial to have a non-zero observation of the electron spectrum between  $Q_\beta$  and  $Q_\beta - \sqrt{\Delta m_{31}^2}$ , otherwise one could confuse the inverted ordering spectrum with a normal ordering spectrum obtained with a larger lightest neutrino mass  $m_{\text{lightest}} \simeq \sqrt{\Delta m_{31}^2}$  (green curve).

Another consideration is due. One could think to probe the neutrino mass ordering just using the fact that the expected number of events is smaller in the inverted ordering than in the normal ordering case. As we discussed above, this could be possible only if some independent experiment would be able to determine the mass of the lightest neutrino, in order to break the possible degeneracy between  $m_{\text{lightest}}$  and the mass ordering depicted by the blue and green curves in **Figure 6**, otherwise the conditions required to observe the electron spectrum between  $Q_\beta$  and  $Q_\beta - \sqrt{\Delta m_{31}^2}$  would be probably sufficient to guarantee a direct observation or exclusion of the kink. The best way to determine the neutrino mass ordering, however, may be to use an estimator which compares the binned spectra in the normal and inverted ordering cases, as proposed for example in Stanco et al. (2017) in the context of reactor neutrino experiments. The authors of the study, indeed, find that a dedicated estimator can enhance the detection significance with respect to a standard  $\chi^2$  comparison.

To conclude, today the status of  $\beta$ -decay experiments is far from the level of determining the mass ordering, since the energy resolution achieved in past and current measurements is not sufficient to guarantee a precise probe of the interesting part

of the spectrum. KATRIN, for example, aims at a sensitivity of 0.2 eV on the effective electron neutrino mass (Angrik et al., 2004; Sejersen Riis et al., 2011), only sufficient to probe the fully degenerate region of the neutrino mass spectrum.

### 3.2. Mass Ordering From Neutrinoless Double Beta Decay

In the second part of this section we shall discuss instead the perspectives from the neutrinoless double beta decay (see e.g., the reviews Gomez-Cadenas et al., 2012; Dell’Oro et al., 2016 and also Pascoli and Petcov (2002)), a process allowed only if neutrinos are Majorana particles (Schechter and Valle, 1982), since it requires the lepton number to be violated by two units. Neutrinoless double beta decay experiments therefore aim at measuring the life time  $T_{1/2}^{0\nu}$  of the decay, which can be written as:

$$\frac{1}{T_{1/2}^{0\nu}(\mathcal{N})} = G_{0\nu}^{\mathcal{N}} |\mathcal{M}_{\mathcal{N}}^{0\nu}|^2 \left( \frac{m_{\beta\beta}}{m_e} \right)^2, \quad (7)$$

where  $m_e$  is the electron mass,  $G_{0\nu}^{\mathcal{N}}$  is the phase space factor,  $\mathcal{M}_{\mathcal{N}}^{0\nu}$  is the nuclear matrix element (NME) of the neutrinoless double beta decay process,  $\mathcal{N}$  indicates the chemical element which is adopted to build the experiment and  $m_{\beta\beta}$  is the effective Majorana mass, see below. In case no events are observed, a lower bound on  $T_{1/2}^{0\nu}$  can be derived. Recent constraints on the neutrinoless double beta decay half-life come from the EXO-200 Albert et al. (2014), KamLAND-Zen Gando et al. (2016), CUORE Alduino et al. (2018a), Majorana Aalseth et al. (2018), CUPID-0 Azzolini et al. (2018), Gerda Agostini et al. (2018), and NEMO-3 Arnold et al. (2018) experiments. The strongest bounds to date on the half-life of the different isotopes are:  $T_{1/2}^{0\nu}(^{76}\text{Ge}) > 8.0 \times 10^{25}$  year from Gerda Agostini et al. (2018),  $T_{1/2}^{0\nu}(^{82}\text{Se}) > 2.4 \times 10^{24}$  year from CUPID-0 Azzolini et al. (2018),  $T_{1/2}^{0\nu}(^{130}\text{Te}) > 1.5 \times 10^{25}$  year from CUORE Alduino et al. (2018a) and  $T_{1/2}^{0\nu}(^{136}\text{Xe}) > 1.07 \times 10^{26}$  year from KamLAND-Zen, Gando et al. (2016).

The effective Majorana mass reads as:

$$m_{\beta\beta} = \sum_{k=1}^N e^{i\alpha_k} |U_{ek}|^2 m_k, \quad (8)$$

where  $N$  is the number of neutrino mass eigenstates, each with its mass  $m_k$ ,  $\alpha_k$  are the Majorana phases (one of which can be rotated away, so that there are  $N - 1$  independent phases), and  $U_{ek}$  represents the mixing between the electron flavor neutrino and the  $k$ -th mass eigenstate. Notice that the conversion between the half-life of the process and the effective Majorana mass depends on the NMEs (see e.g., Vergados et al., 2016; Engel and Menéndez, 2017), which are typically difficult to compute. Several methods can be employed and there is no full agreement between the results obtained with the different methods. As a consequence, the quoted limits on  $T_{1/2}^{0\nu}$  can be translated into limits on  $m_{\beta\beta}$  which depend on the NMEs. If the most conservative values for the NMEs are considered, none of the current constraints reaches the level required to start

constraining the inverted ordering in the framework of three neutrinos, see **Figure 3**.

If we compute  $m_{\beta\beta}$  as a function of the lightest neutrino mass with the current preferred values of the mixing parameters and in the context of three neutrinos, we discover that the value of  $m_{\beta\beta}$  depends on the mass ordering only for  $m_{\text{lightest}} \lesssim 40$  meV, see **Figure 3**. For this reason, neutrinoless double beta experiments can aim to distinguish the mass ordering only for the smallest values of the lightest neutrino mass. Please note that this also means that if the lightest neutrino has a mass above  $\sim 40$  meV, perfectly allowed by all the present constraints on the neutrino mass scale, the two mass orderings will never be distinguished in the context of neutrinoless double beta decay experiments.

When going to smaller  $m_{\text{lightest}}$ , the situation changes, as  $m_{\beta\beta}$  becomes independent of  $m_{\text{lightest}}$ . In the region  $m_{\text{lightest}} \lesssim 10$  meV, a difference between the two mass orderings appears, since the effective Majorana mass is constrained by the mass splittings to be larger than  $\sim 10$  meV for inverted ordering, while it must be below  $\sim 7$  meV for normal ordering. This means that experiments which can test the region  $m_{\beta\beta} < 10$  meV can rule out the inverted scenario. Note that a positive detection of  $T_{1/2}^{0\nu}$  in the range that corresponds to  $m_{\beta\beta} \gtrsim 10$  meV, on the other hand, would not give sufficient information to determine the mass ordering without an independent determination of  $m_{\text{lightest}}$ . To resume, in the context of three neutrino mixing, neutrinoless double beta decay experiments alone will be able to determine the neutrino mass ordering only ruling out the inverted scheme, that is to say if the ordering is normal and  $m_{\text{lightest}} \lesssim 10$  meV.

In any case, we should remember that if no neutrinoless double beta decay candidate event will ever be observed we will not have determined the mass ordering univocally: Dirac neutrinos escape the constraints from this kind of process, so that it would be still perfectly allowed to have an inverted ordering scheme and no Majorana fermions in the neutrino sector. Due to the presence of the Majorana phases in Equation (8), unfortunately, there is a small window for  $m_{\text{lightest}}$  in normal ordering that can correspond to almost vanishing values of  $m_{\beta\beta}$ , which will possibly never be observable. As we show in **Figure 3**, however, the region of parameter space where this happens has a very small volume if one considers the phases to vary between 0 and  $2\pi$ , so that the credible region for  $m_{\beta\beta}$  in a Bayesian context shows that it is rather unlikely to have  $m_{\beta\beta} \lesssim 2 \times 10^{-4}$  eV, as a significant amount of fine tuning would be needed in the (completely unknown) Majorana phases. Our statement, which arises from assuming a logarithmic prior on  $m_{\text{lightest}}$ , is in perfect agreement with the results of Agostini et al. (2017a), where a linear prior on  $m_{\text{lightest}}$  is assumed.

Please note that the situation depicted in **Figure 3** is only valid if there are only three neutrinos. If, as the current DANSS Alekseev et al. (2018) and NEOS Ko et al. (2017) experiments may suggest, a sterile neutrino with a mass around 1 eV exists (see e.g., Dentler et al., 2017, 2018; Gariazzo et al., 2017, 2018b), the situation would be significantly different. The allowed bands for  $m_{\beta\beta}$  as a function of the lightest neutrino mass when a light sterile neutrino is introduced are reported for example in Giunti and Zavanin (2015) (see also Gariazzo et al., 2016). In this three active plus one sterile neutrino case (3+1), the contribution of the

fourth neutrino mass eigenstate (mainly mixed with the sterile flavor) must be added in Equation (8), with the consequence that the allowed bands are located at higher  $m_{\beta\beta}$ . In **Figure 7**, adapted from Giunti (2017), we reproduce the dependence of the effective Majorana mass on the lightest neutrino mass when one assumes the 3+1 neutrino scenario, compared with the standard three neutrino case. As we can see, with the introduction of an extra sterile neutrino state,  $m_{\beta\beta}$  is significantly increased for the normal ordering case, reaching the level of the inverted ordering bands, which are less shifted toward higher values of  $m_{\beta\beta}$ . Furthermore, in the 3+1 scenario, also in the inverted ordering case it is possible to have accidental cancellations due to the three independent Majorana phases in Equation (8) (see the detailed discussion of Giunti and Zavanin, 2015), so that a non-detection of the neutrinoless double beta decay process would never be sufficient to rule out the inverted ordering. The opposite situation may occur in case the lightest neutrino mass will be independently constrained to be below  $\sim 10$  meV while  $m_{\beta\beta} \lesssim 10$  meV: in this case, however, we would rule out normal ordering. Consequently, if a light sterile neutrino exists, neutrinoless double beta experiments will never be able to determine the mass ordering if the mass ordering is normal, while some possibility remains if the ordering of the three active neutrino masses is inverted, provided that the lightest neutrino is very light and the Majorana phases are tuned enough. The KamLAND-Zen, Gerda, and CUORE experiments, using three different materials, may very soon start probing the inverted ordering region in the case of 3+1 neutrino mixing for all the possible values of the NMEs, see **Figure 7**, where the current KamLAND-Zen Gando et al. (2016) constraints are reported.

To conclude and summarize the current status: neutrinoless double beta decay cannot yet provide constraints on the neutrino mass ordering. Depending on the lightest neutrino mass and on the existence of a fourth (sterile) neutrino, it would be possible that not even far-future experiments could be able to reach this goal.

## 4. RESULTS FROM COSMOLOGY

Massive neutrinos affect the cosmological observables in different ways, that we shall summarize in what follows. For a comprehensive review of the effects of neutrino masses in cosmology, we refer the reader to the recent work presented in Lattanzi and Gerbino (2018).

A very important epoch when discussing the impact of massive neutrinos in the cosmological expansion history and in the perturbation evolution is the redshift at which neutrinos become non-relativistic. This redshift is given by

$$1 + z_{\text{nr},i} \simeq 1890 \left( \frac{m_i}{1 \text{ eV}} \right), \quad (9)$$

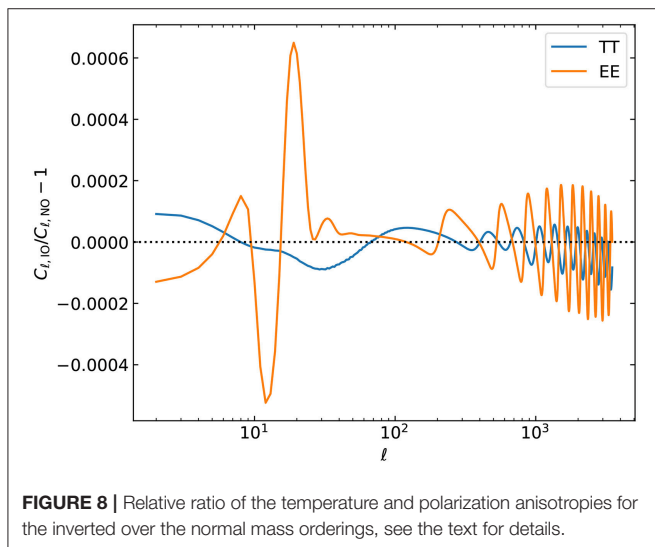
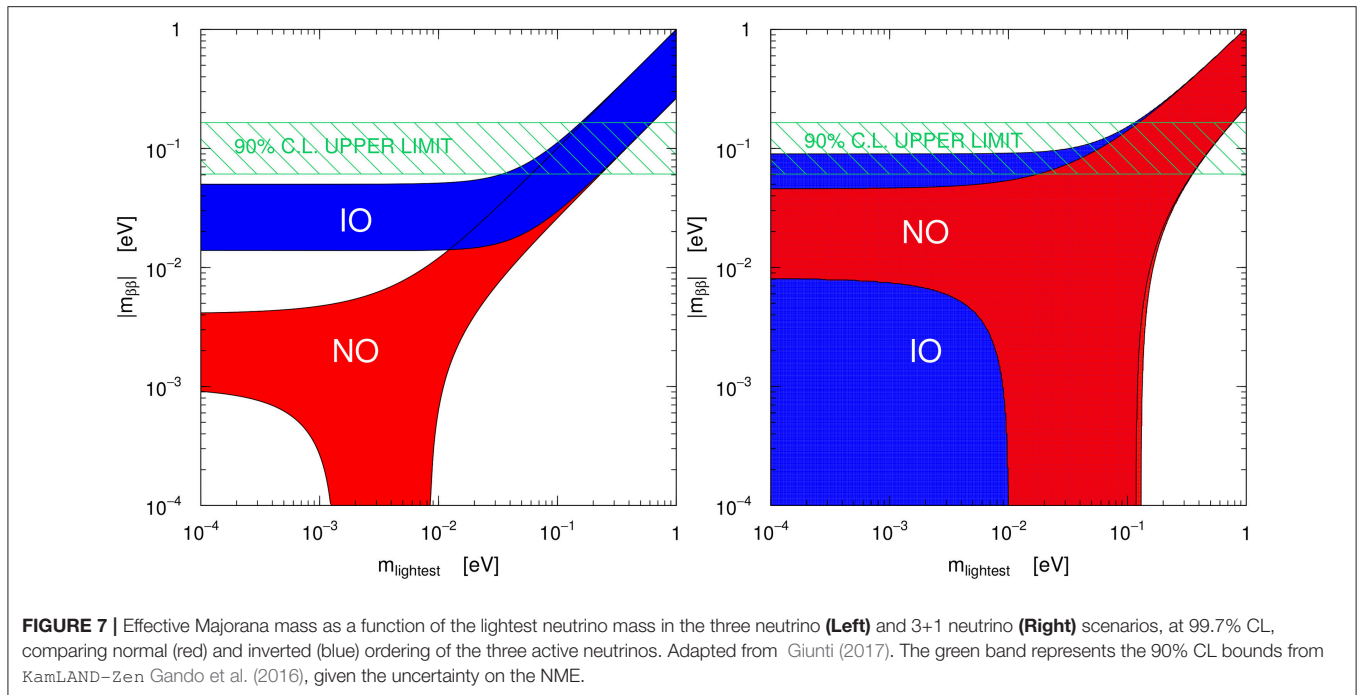
with  $m_i$  referring to the mass of each massive neutrino eigenstate. Current bounds on neutrino masses imply that at least two out of the three massive eigenstates became non-relativistic in the matter dominated period of the universe. As stated in the introductory section, and as we shall further illustrate along

this section, cosmological measurements are currently unable to extract individually the masses of the neutrino eigenstates and the ordering of their mass spectrum and, therefore, concerning current cosmological data, all the limits on the neutrino mass ordering will come from the sensitivity to the total neutrino mass  $\sum m_\nu$ . Consequently, in what follows, we shall mainly concentrate on the effects on the cosmological observables of  $\sum m_\nu$ , providing additional insights on the sensitivity to the ordering of the individual mass eigenstates whenever relevant.

### 4.1. CMB

There are several imprints of neutrino masses on the CMB temperature fluctuations pattern once neutrinos become non-relativistic: a shift in the matter-radiation equality redshift or a change in the amount of non-relativistic energy density at late times, both induced by the evolution of the neutrino background, that will, respectively, affect the angular location of the acoustic peaks and the slope of the CMB tail, through the *Late Integrated Sachs Wolfe (ISW) effect*. The former will mostly modify  $\Theta_s$ , i.e., the angular position of the CMB peaks, which is given by the ratio of the sound horizon and the angular diameter distance, both evaluated at the recombination epoch. Massive neutrinos enhance the Hubble expansion rate, with a consequent reduction of the angular diameter distance and an increase of  $\Theta_s$ , which would correspond to a shift of the peaks toward larger (smaller) angular scales (multipoles). The latter, the Late ISW effect, is related to the fact that the gravitational potentials are constant in a matter-dominated universe. The inclusion of massive neutrinos will delay the dark energy dominated period and consequently reduce the time variation of the gravitational potential at late times, suppressing the photon temperature anisotropies in the multipole region  $2 < \ell < 50$ . A very similar effect occurs at early times through the so-called Early ISW effect, which governs the height of the first CMB peak. Light active neutrino species, indeed, reduce the time variations of the gravitational potential also around the recombination period, due to the different evolution of these potentials in radiation/matter dominated epochs, leaving a signature on the CMB photon fluctuations when they become non-relativistic. Massive neutrinos will therefore decrease the temperature anisotropies by  $\Delta C_\ell / C_\ell \sim (m_{\nu,i} / 0.1 \text{ eV}) \%$  in the multipole range  $20 < \ell < 500$  (Lesgourgues and Pastor, 2012).

Unfortunately, the Late ISW effect affects the CMB spectrum in a region where cosmic variance does not allow for very accurate measurements. From what regards the other two effects, i.e., the shift in the location of the acoustic peaks and the Early ISW effect, they can both easily be compensated varying other parameters which govern the expansion of the universe. For example, within the minimal  $\Lambda$ CDM framework, the total amount of matter in the universe and the Hubble constant  $H_0$  can be tuned in order to compensate the effects of massive neutrinos. Therefore, CMB primary anisotropies alone can not provide very tight bounds on the neutrino masses, due to the strong parameter degeneracies. This automatically implies that CMB measurements alone are unable to extract any information concerning the neutrino mass ordering, as shown in **Figure 8**, obtained by means of the publicly available Boltzmann solver



Cosmic Linear Anisotropy Solving System (CLASS) (Blas et al., 2011; Lesgourgues, 2011a,b; Lesgourgues and Tram, 2014). In the figure we can notice that the difference between normal and inverted neutrino mass orderings, for  $\sum m_\nu = 0.12 \text{ eV}^{11}$  is almost negligible. Moreover, the largest differences appear in the multipole range where cosmic variance dominates.

<sup>11</sup>This is the most constraining 95% CL limit Palanque-Delabrouille et al. (2015) at present, excluding combinations of data sets that are in tension, and we have chosen it as the benchmark value in the following discussions throughout this review.

Among the secondary CMB anisotropies, i.e., those generated along the photon line of sight and not produced at recombination, there are two effects that can notably improve the sensitivity to the total neutrino mass  $\sum m_\nu$  from CMB observations. One of them is CMB lensing, that is, a distortion of the photon paths because of the presence of matter inhomogeneities along the line of sight. Due to such distortion, the CMB acoustic oscillation features will be smeared out in a scale-dependent way, mostly due to matter overdensities at  $z \lesssim 5$ . By measuring the non-gaussianities of CMB polarization and temperature maps it is possible to extract the power spectrum of the lensing potential. This, in turn, contains very useful information on the integrated matter distribution along the line of sight. Since massive neutrinos behave differently from a pure cold dark matter component, characterized by zero velocities, the small-scale structure suppression induced by the non-negligible neutrino dispersion velocities will decrease the CMB lensing signal expected in the absence of neutrino masses (Kaplighat et al., 2003; Song and Knox, 2004; Lesgourgues et al., 2006; Smith et al., 2006; de Putter et al., 2009; Allison et al., 2015), leaving unchanged the power spectrum of the lensing potential at large scales, and suppressing it at small scales. Furthermore, since CMB lensing involves high redshifts, non-linearities do not enter in the calculation of the matter density field. Therefore, CMB lensing enhances the capabilities to bound the neutrino masses using CMB data. In the future, this technique may even surpass weak lensing capabilities, based on statistical analyses of the ellipticity of remote galaxies, see below and section 6.4. Indeed, nowadays, measurements from the *Planck* satellite constrain the neutrino masses dominantly through CMB gravitational lensing. As stated in Ade et al. (2014), increasing the neutrino mass implies an increase on the expansion rate at redshifts  $z \geq 1$ ,

corresponding to a suppression of clustering at scales below the size of the horizon at the non-relativistic transition. This effect leads to a decrease in CMB lensing that, at multipoles  $\ell = 1,000$ , is  $\sim 10\%$  for  $\sum m_\nu = 0.66$  eV.

On the other hand, we have the reionization process in the late universe, when the first generation of galaxies emitted ultraviolet (UV) photons that ionized the neutral hydrogen, leading to the end of the so-called *dark ages*. Reionization increases the number density of free electrons  $n_e$  which can scatter the CMB with a probability given by a quantity named *reionization optical depth*,  $\tau$ , which can be computed as an integral over the line of sight of  $n_e$ . The consequence of an increase of  $\tau$  on the CMB temperature fluctuations is the suppression of the acoustic peaks by a factor  $\exp(-2\tau)$  at scales smaller than the Hubble horizon at the reionization epoch. Even if from the point of view of CMB temperature anisotropies this effect is highly degenerate with a change in the amplitude of the primordial power spectrum  $A_s$ , which governs the overall amplitude of the CMB spectra, reionization induces linear polarization on the CMB spectrum, leading to a “reionization bump” in the polarization spectra at large scales, which otherwise would vanish. Even if the reionization signal is rather weak, as it amounts to no more than  $\sim 10\%$  of the primary polarization signal (Aghanim et al., 2008), very accurate measurements of the reionization optical depth  $\tau$  sharpen considerably the CMB neutrino mass bounds (Vagnozzi et al., 2017), as they alleviate the degeneracy between  $A_s$  and  $\tau$  and consequently the existing one between  $\sum m_\nu$  and  $A_s$ .

## 4.2. Large Scale Structure of the Universe

The largest effect of neutrino masses on the cosmological observables is imprinted in the matter power spectrum (Bond et al., 1980; Hu et al., 1998). Neutrinos are *hot* dark matter particles and, therefore, due to the pressure gradient, at a given redshift  $z$ , the non-relativistic neutrino overdensities can only cluster at scales for which the wavenumber of perturbations is below the neutrino free streaming scale  $k_{fs}$  (i.e., at scales  $k < k_{fs}$ ), with

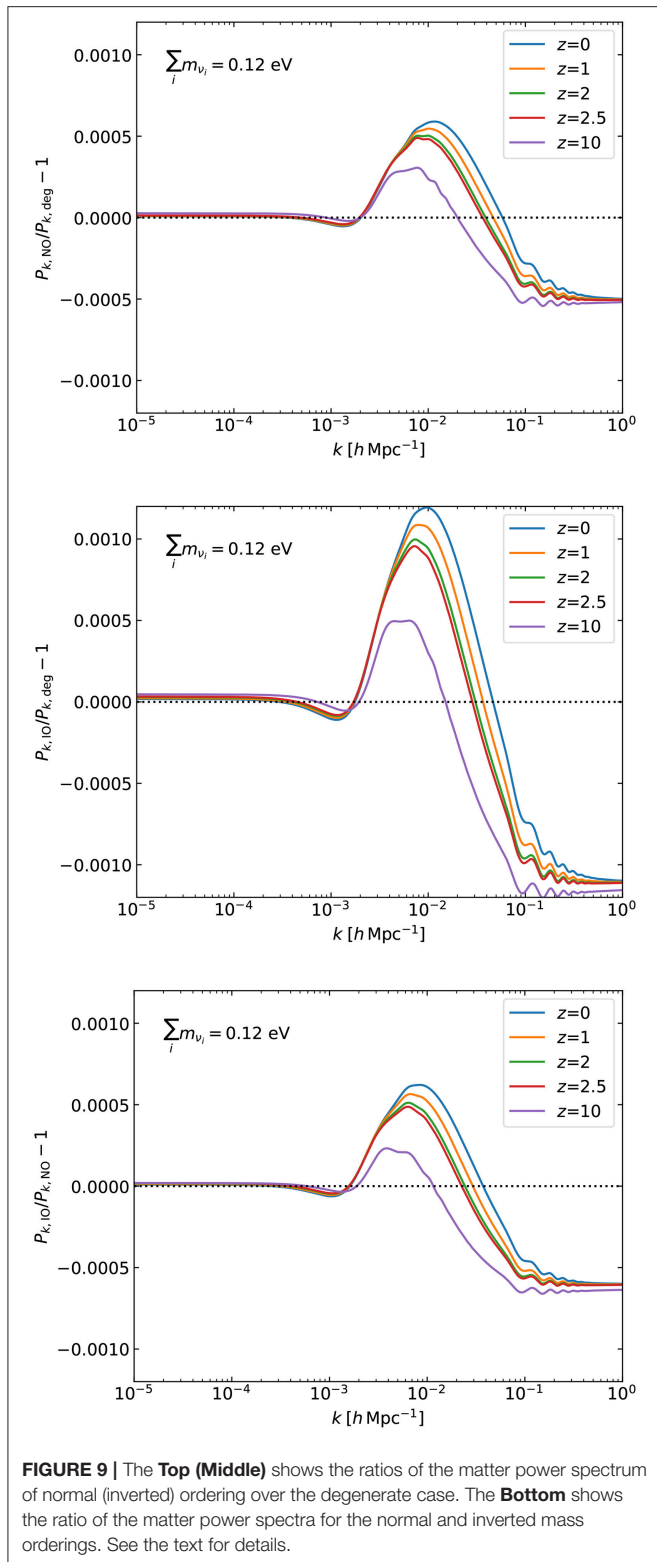
$$k_{fs}(z) = \frac{0.677}{(1+z)^{1/2}} \left( \frac{m_\nu}{1 \text{ eV}} \right) \sqrt{\Omega_m} h \text{ Mpc}^{-1}, \quad (10)$$

being  $\Omega_m$  the ratio of the total matter energy density over the critical density at redshift zero. The free-streaming nature of the neutrino will be directly translated into a suppression of the growth of matter fluctuations at small scales. One could then conclude that extracting the neutrino relic masses and their ordering is a straightforward task, once that measurements of the matter power spectrum at the relevant scales are available at a different number of redshifts. The former statement is incorrect, not only because it does not consider the existence of degeneracies with the remaining cosmological parameters, but also because a number of subtleties must be taken into account, as we shall explain in what follows. The decrease of the matter power spectrum due to the total neutrino mass  $\sum m_\nu$  is, in principle, currently measurable. Nonetheless, when fixing  $\sum m_\nu$ , the total mass could be splitted differently among the three neutrino mass eigenstates (i.e.,  $m_1$ ,  $m_2$  and  $m_3$ ), modifying slightly the relativistic to non-relativistic transition. This will

affect both the background evolution and the perturbation observables (Lesgourgues et al., 2004): the different free-streaming scales associated to each of the three neutrino mass eigenstates will be imprinted in the matter power spectrum. **Figure 9** shows the ratios of the matter power spectrum for normal over degenerate, inverted over degenerate, and inverted over normal neutrino mass spectra for a total neutrino mass of 0.12 eV. We illustrate such ratios at different redshifts. Notice that the differences among the possible neutrino mass schemes are tiny, saturating at the 0.06% level at  $k > 0.2h \text{ Mpc}^{-1}$ . Therefore, only very futuristic means of measuring the matter power spectrum could be directly sensitive to the neutrino mass ordering, and, eventually, be able to isolate each of the free-streaming scales associated to each individual neutrino mass eigenstate. We shall comment on these future probes in section 6.5.

Since, at present, matter power spectrum data constrain exclusively  $\sum m_\nu$ , it is only via these bounds, combined with CMB or other external data sets, that nowadays a limit on the neutrino mass ordering can be obtained, see section 4.3. Nevertheless, and as aforementioned, there are a number of problems which may interfere with a proper understanding of the scale-dependent neutrino mass suppression of clustering. The first of them is due to the fact that observations measure the spatial distributions (or their Fourier transforms, the power spectra) of galaxies, clusters, or quasars, e.g., of given tracers, mapping the large scale structure of the Universe at a number of redshifts, by measuring the growth of fluctuations at different scales. However, the matter distribution is not directly measured, i.e., it needs to be inferred from the tracers observed. A simple model of structure formation suggests that at large scales and, therefore, when the perturbation evolution is still in the linear regime, the galaxy power spectrum is related to the matter one by a constant named  $b$ , the light-to-mass bias (Desjacques et al., 2018). The galaxy bias can be determined either separately by independent methods or to be considered as an additional free parameter to be measured together with the neutrino mass  $\sum m_\nu$ . This approach has been followed in many studies in the past (Cuesta et al., 2016; Giusarma et al., 2016; Vagnozzi et al., 2017). However, when dealing with neutrino masses, the relationship between the tracers and the underlying matter field may be more complicated, as neutrinos themselves may induce scale-dependent features in the bias (Castorina et al., 2014; LoVerde and Zaldarriaga, 2014; Muñoz and Dvorkin, 2018) due to their free-streaming nature (see also the recent work of Giusarma et al., 2018 for a new method to extract a scale-dependent bias, based on the cross-correlation of CMB lensing maps and galaxy samples).

Another additional complication when extracting the neutrino mass from clustering observations arises from the presence of non-linearities at scales  $k \gtrsim k_{NL}^0 \equiv 0.1 - 0.2 h \text{ Mpc}^{-1}$  at  $z = 0$ . The effect of neutrino masses is very-well understood on linear scales, i.e., scales below  $k_{NL}$  at  $z = 0$  (or located at slightly larger values of  $k$  but at higher redshifts). Massive neutrinos induce a suppression in the linear matter power spectrum below their free streaming scale  $\Delta P/P \propto -8f_\nu$ , with  $f_\nu$  the fraction of matter in the form of massive neutrinos (Hu et al., 1998).



Accurate descriptions of the matter power spectrum in the non-linear regime are therefore mandatory in order to be sensitive to the full neutrino mass signature. This is particularly

important in the case of galaxy surveys, in which the information depends on the number of independent modes available, and where going to smaller scales (i.e., larger values of  $k$ ) has a profound impact on the sensitivity to neutrino masses. Several approaches have been followed in the literature to account for the effect of massive neutrinos in the non-linear regime, most of them relying on N-body cosmological simulations, which have been upgraded to include the effects of neutrino clustering in the evolution of the cosmological structures. Methods range from perturbative attempts (Saito et al., 2008; Brandbyge and Hannestad, 2009; Shoji and Komatsu, 2010; Ali-Haïmoud and Bird, 2012; Archidiacono and Hannestad, 2016; Upadhye et al., 2016; Dakin et al., 2017; Senatore and Zaldarriaga, 2017) to the fully non-linear inclusion (Villaescusa-Navarro et al., 2014a; Inman et al., 2015; Banerjee and Dalal, 2016; Dakin et al., 2017; Banerjee et al., 2018; Liu et al., 2018) of neutrinos as an extra set of particles. A conservative alternative consists on using exclusively power spectrum measurements within the linear regime (i.e.,  $k < 0.1 \text{ h Mpc}^{-1}$ ). Some of the cosmological constraints have also been obtained using the mildly non-linear regime ( $k < 0.2 \text{ h Mpc}^{-1}$ ) by means of the so-called Halofit formalism (Smith et al., 2003; Takahashi et al., 2012). The Halofit prescription models the non-linear matter power spectrum, and it has been calibrated against a wide range of CDM simulations. It has also been extended for massive neutrino cosmologies (Bird et al., 2012). Other predictions for the non-linear matter power spectrum include the Coyote emulator (Heitmann et al., 2014), which is based on a set of high-accuracy N-body simulations.

However, there is also another avenue to use large scale structure information, the *geometrical* approach, which exploits the so-called Baryon Acoustic Oscillations (BAO) rather than the measurements of the broad-band *shape* of the galaxy power spectrum. The BAO signal appears as a peak in the two-point mass correlation function corresponding to the distance a sound wave can travel in the photon-baryon fluid from very early in the universe until the drag epoch, when the baryon optical depth equals one. The BAO signature provides a standard ruler to measure the distance to various redshifts, and it can be measured either *along the line of sight*, in which the radial distance is inversely proportional to the Hubble expansion rate  $H(z)$ , or *across the line of sight*, in which case the angular distance is proportional to an integral of  $H(z)$ , the angular diameter distance  $d_A(z)$ . To use the BAO method, one must, therefore, extract the acoustic scale from the clustering of some tracer of the baryon distribution (galaxies, quasars). This is typically done statistically using the two-point correlation function of the spatial distribution of tracers, or from its Fourier transform, the power spectrum. From these functions, it is possible to measure two different quantities corresponding to the oscillations parallel and perpendicular to the line of sight, that is  $r_s H(z)$  and  $d_A(z)/r_s$ , with  $r_s$  the sound horizon at the drag epoch. Many of the BAO analyses to date have used spherically averaged clustering statistics, measuring an effective distance  $D_V \equiv (z d_A(z)^2 / H(z))^{1/3}$ , which is the volume-averaged distance. Some of the most recent BAO extractions by the Sloan Digital Sky Survey III (SDSS-III) (Eisenstein et al., 2011) Baryon Oscillation Spectroscopic Survey (BOSS) (Dawson et al., 2013)

have achieved, by measuring the clustering of 1.2 million galaxies with redshifts  $0.2 < z < 0.75$ , 1.8% precision on the radial BAO distance and 1.1% precision on the transverse distance in the  $z < 0.75$  redshift region (Beutler et al., 2017b; Ross et al., 2017; Vargas-Magaña et al., 2018). These results improve former determinations from previous data releases of BOSS and SDSS (Eisenstein et al., 2005; Anderson et al., 2013, 2014; Tojeiro et al., 2014; Ross et al., 2015) or other galaxy surveys (Percival et al., 2001; Cole et al., 2005; Beutler et al., 2011; Blake et al., 2011; Kazin et al., 2014), see also the recent works of Bautista et al. (2017b) for a 2.6% measurement of  $D_V$  at  $2.8\sigma$  significance with the extended Baryon Oscillation Spectroscopic Survey (eBOSS) from SDSS-IV Dawson et al. (2016). The Dark Energy Survey (DES) has also achieved a 4.4% accuracy on the measurement of  $d_A(z)/r_s$  at  $z = 0.81$  Abbott et al. (2017).

Galaxy clustering measurements can also be exploited to constrain, at a number of redshifts, the product of the linear growth rate  $f \times \sigma_8^{12}$ , by means of the so-called redshift space distortions, caused by galaxy peculiar velocities, see the recent analyses of Beutler et al. (2017a), Satpathy et al. (2016), and Sanchez et al. (2017).

Apart from the spatial distribution of galaxies, there are also other ways of mapping the large scale structure of the universe at different cosmic times. The Lyman- $\alpha$  forest power spectrum from distant quasars plays a major role for constraining the neutrino masses, as it is sensitive to smaller scales, where the effect of neutrino masses is more pronounced. We refer the reader to the seminal works of Croft et al. (1999), Seljak et al. (2005), Goobar et al. (2006), Seljak et al. (2006), Gratton et al. (2008), Fogli et al. (2008), and Viel et al. (2010). In addition, since the redshifts at which Lyman- $\alpha$  forest probes are sensitive to are higher than those corresponding to galaxy surveys, a fixed scale  $k$  will be closer to the linear regime in the Lyman- $\alpha$  case. An additional benefit of going to higher redshifts is that uncertainties related to the evolution of the dark energy fluid will be sub-dominant, as dark energy effects are expected to be more prominent at very low redshifts. However, modeling the neutrino mass effect in the Lyman- $\alpha$  forest power spectrum is highly non-trivial as it may strongly rely on hydrodynamical simulations (Viel et al., 2010). These numerical calculations try to properly account for the late time non-linear evolution of the intergalactic medium (IGM), including reionization processes (Viel et al., 2010; Villaescusa-Navarro et al., 2013). The BAO signature can also be measured in the flux correlation function of the Lyman- $\alpha$  forest of quasars, first detected at a mean redshift  $z = 2.3$  in Busca et al. (2013, see also Slosar et al., 2013; Font-Ribera et al., 2014a; Delubac et al., 2015; Bautista et al., 2017a; Bourboux et al., 2017 and Aubourg et al., 2015), in which joint constraints from the BAO signature from galaxies and quasars have been presented.

Galaxy clusters provide yet another test which allows us to trace the clustering of matter perturbations and, therefore, to test the suppression due to the presence of a non-zero  $\sum m_\nu$ . Galaxy clusters are, by far, the largest virialised objects in the universe,

providing a measurement of the so-called cluster number count function  $dN/dz$ . This function gives the number of clusters of a certain mass  $M$  within a redshift interval (bin)  $z + \delta z$  and, for a given survey:

$$\left. \frac{dN}{dz} \right|_{M > M_{\min}} = f_{\text{sky}} \frac{dV(z)}{dz} \int_{M_{\min}}^{\infty} dM \frac{dn}{dM}(M, z). \quad (11)$$

The quantity  $f_{\text{sky}} = \Delta\Omega/4\pi$  refers to the fraction of sky covered by the survey and the unit volume is given by

$$\frac{dV(z)}{dz} = \frac{4\pi}{H(z)} \int_0^z dz' \frac{1}{H(z')^2}. \quad (12)$$

While the redshift is relatively easy to measure, the main uncertainty of this method comes from the cluster mass estimates, determined through four main available methods: X-rays, velocity dispersion, Sunyaev-Zeldovich (SZ) effect<sup>13</sup> and weak lensing. The overall error in the cluster mass determination is usually around  $\Delta M/M \sim 10\%$ . Furthermore, in order to relate the cluster number count function to the underlying cosmological parameters, one needs as an input a mass function  $dn(z, M)/dM$  describing the abundance of virialised objects at a given redshift, usually obtained by means of  $N$ -body simulations (Tinker et al., 2008; Costanzi et al., 2013). This mass function depends on both the matter mass-energy density and on the standard deviation (computed in linear perturbation theory) of the density perturbations:

$$\sigma^2 = \frac{1}{2\pi^2} \int_0^{\infty} dk k^2 P(k) W^2(kR), \quad (13)$$

where  $P(k)$  is the matter power spectrum,  $W(kR)$  is the top-hat window function,  $R$  is the comoving fluctuation size, related to the cluster mass  $M$  as  $R = (3M/4\pi\rho_m)^{1/3}$ , and

$$W(kR) = \frac{3(\sin(kR) - (kR)\cos(kR))}{(kR)^3}. \quad (14)$$

There are still some degeneracies in the cosmological parameters probed by cluster surveys, whose results are reported by means of a relationship between the matter clustering amplitude  $\sigma_8$  (obtained from Equation 13), and the matter mass-energy density  $\Omega_m$  parameters. More concretely, cluster catalogs measure the so-called cluster normalization condition,  $\sigma_8\Omega_m^\gamma$ , where  $\gamma \sim 0.4$  (Allen et al., 2011; Weinberg et al., 2013). Current cluster catalogs include X ray clusters (see e.g., Hilton et al., 2018; Sohn et al., 2018 and references therein), the optically detected SDSS photometric redMaPPer cluster catalog (Roza and Rykoff, 2014; Rykoff et al., 2014; Roza et al., 2015) and the Planck SZ galaxy clusters (PSZ2) Ade et al. (2016a), which contains more than a thousand confirmed clusters. Other SZ cluster catalogs are those detected from the Atacama Cosmology Telescope (ACT) (Hilton et al., 2018) and from the South Pole Telescope (SPT) (de Haan et al., 2016).

<sup>12</sup>Here,  $\sigma_8$  corresponds to the normalization of the matter power spectrum on scales of  $8h^{-1}$  Mpc, see Equation (13).

<sup>13</sup>The thermal SZ thermal effect consists on a spectral distortion on CMB photons which arrive along the line of sight of a cluster.

Last but not least, weak lensing surveys are also an additional probe of the large scale structure effects of massive neutrinos (Cooray, 1999; Abazajian and Dodelson, 2003; Hannestad et al., 2006b; Kitching et al., 2008; De Bernardis et al., 2009; Ichiki et al., 2009; Tereno et al., 2009; Debono et al., 2010; Jimenez et al., 2010). Light rays from distant galaxies are bent by the matter density perturbations between the source galaxies and the observer, thereby inducing distortions in the observed images of the source galaxies (see the reviews Munshi et al., 2008; Kilbinger, 2015). Commonly, the deformations in the source galaxies are rather weak and to extract the lensing signature one needs a correlation among different galaxy images, the so-called shear-correlation functions. By measuring the angular correlation of these distortions, one can probe the clustering statistics of the intervening matter density field along the line of sight, without relying strongly on bias assumptions, setting therefore independent constraints on the neutrino masses. Weak lensing surveys usually report their cosmological constraints in terms of the clustering amplitude  $\sigma_8$  and the current matter energy density  $\Omega_m$ . More specifically, they make use of the combination  $S \equiv \sigma_8 \sqrt{\Omega_m}$  as an accurate description of the amplitude of structure growth in the universe. The most recent weak lensing cosmological analyses profiting of weak lensing data from DES and from the Kilo Degree Survey (KiDS), consisting of  $\sim 450 \text{ deg}^2$  of imaging data, are presented in Abbott et al. (2018), Hildebrandt et al. (2017), Joudaki et al. (2017), and Köhlinger et al. (2017), respectively.

### 4.3. Cosmological Bounds on Neutrino Masses and Their Ordering

In the following, we shall review the current cosmological bounds on neutrino masses and on their ordering, firstly in the standard  $\Lambda$ CDM scenario and then when considering extended cosmological models.

#### 4.3.1. Constraints Within the $\Lambda$ CDM Universe

Focusing on bounds exclusively from the Planck collaboration, making use of their CMB temperature anisotropies measurements in the multipole range  $\ell \lesssim 2500$  (Planck TT) and of their low-multipole (up to  $\ell = 29$ ) polarization data, lowP, a bound of  $\sum m_\nu < 0.72 \text{ eV}$  at 95% CL Ade et al. (2016b) is reported. When high-multipole (i.e., small scale,  $\ell > 30$ ) polarization measurements are included in the analyses (Planck TT, TE, EE + lowP), the quoted constraint is  $\sum m_\nu < 0.49 \text{ eV}$  at 95% CL. As the Planck TT, TE, EE data combination may still have some systematics due to temperature-to-polarization leakage (Ade et al., 2016b), the bounds including these measurements provide the less conservative approach when exploiting CMB data. In 2016, the Planck collaboration presented a series of new results based on a new analysis, in which the modeling and removal of unexplained systematics in the large angular polarization data were accounted for Aghanim et al. (2016b). The value of the optical depth  $\tau$  found in these refined analyses (using the SimLow likelihood) was smaller than that quoted in previous analyses (Ade et al., 2016b): while the lowP data was providing  $\tau = 0.067 \pm 0.022$ , the SimLow likelihood results in  $\tau = 0.055 \pm 0.009$ . The most important

**TABLE 2 |** Jeffreys' scale Jeffreys (1961) for estimating the strength of the preference for one model over the other (adapted from Trotta (2008)) when performing Bayesian model comparison analysis.

$ \ln B_{\text{NO},\text{IO}} $	Odds	Strength of evidence	$N\sigma$ for the mass ordering
$< 1.0$	$\lesssim 3:1$	inconclusive	$< 1.1\sigma$
$\in [1.0, 2.5]$	$(3 - 12):1$	weak	$1.1 - 1.7\sigma$
$\in [2.5, 5.0]$	$(12 - 150):1$	moderate	$1.7 - 2.7\sigma$
$\in [5.0, 10]$	$(150 - 2.2 \times 10^4):1$	strong	$2.7 - 4.1\sigma$
$\in [10, 15]$	$(2.2 \times 10^4 - 3.3 \times 10^6):1$	very strong	$4.1 - 5.1\sigma$
$> 15$	$> 3.3 \times 10^6:1$	decisive	$> 5.1\sigma$

The fourth column indicates the approximate correspondence between the quoted Bayes factor levels and the  $N\sigma$  probabilities computed for a Gaussian variable.

consequence of this lower value of  $\tau$  on the CMB bounds on  $\sum m_\nu$  is related to the degeneracy between the amplitude of the primordial power spectrum,  $A_s$ , and  $\tau$ , as already introduced in section 4.1: a lower value of  $\tau$  will imply a lower value of  $A_s$ , thus implying a lower overall normalization of the spectrum, leading therefore to tighter constraints on neutrino masses. The 95% CL limits of  $\sum m_\nu < 0.72 \text{ eV}$  and  $\sum m_\nu < 0.49 \text{ eV}$ , respectively from the Planck TT + lowP and Planck TT, TE, EE + lowP analyses, are updated to  $\sum m_\nu < 0.59 \text{ eV}$  and  $\sum m_\nu < 0.34 \text{ eV}$  when using Planck TT + SimLow and Planck TT, TE, EE + SimLow, respectively. These constraints are clearly located away from the region in which a preference for a given mass ordering (normal vs. inverted) may show up. Indeed, the CMB data alone were used by the authors of Gerbino et al. (2017b) which, by means of a novel approach to quantify the neutrino mass ordering, have shown that the odds favoring normal ordering vs. inverted ordering are 1:1 and 9:8 in the case of the Planck TT + lowP and Planck TT, TE, EE + lowP data combinations, respectively. These results point to an inconclusive strength of evidence, see Table 2. Based on a full Bayesian comparison analysis, Gariazzo et al. (2018a) has shown, using Planck TT, TE, EE + lowP measurements together with global neutrino oscillation data, that the Bayes factor for such a combination is  $\log(B_{\text{NO},\text{IO}}) \simeq 2.5$  for almost all the possible parameterizations and prior choices considered. This value of the Bayes factor, which only points to weak preference for normal ordering, is entirely due to neutrino oscillation data, in agreement with the results of Caldwell et al. (2017). Therefore, Planck temperature and polarization measurements alone can not further improve our current knowledge of the neutrino mass ordering from global oscillation data.

The CMB limits on neutrino masses can also include the lensing likelihood, which leads to  $\sum m_\nu < 0.59 \text{ eV}$  at 95% CL for the case of Planck TT, TE, EE + lowP + lensing measurements (Ade et al., 2016b). Notice that the bound with the lensing likelihood is less tight than that obtained without the lensing potential extraction ( $\sum m_\nu < 0.49 \text{ eV}$  at 95% CL from Planck TT, TE, EE + lowP). The reason is related to the fact that, while the Planck CMB power spectra favor a larger lensing amplitude, the lensing potential reconstructions prefer

a lower one. Since increasing the neutrino masses reduces the lensing amplitude, the one dimensional posterior distribution of  $\sum m_\nu$  arising from the combination of CMB temperature, polarization and lensing data sets shifts the neutrino mass constraints away from zero, so that less posterior volume is found near zero than when constraining  $\sum m_\nu$  only with CMB temperature and polarization data.

A significant strengthening on the aforementioned limits can be obtained by means of additional data sets, which help enormously in breaking the degeneracies which are allowed when only CMB data are considered. Among them, the one existing between  $\sum m_\nu$  and the Hubble constant  $H_0$  (see e.g. Giusarma et al. (2013b)). Large scale structure data from galaxy clustering are of great help in breaking degeneracies. When exploited in the geometrical (BAO) form, the Planck collaboration quotes 95% CL limits of  $\sum m_\nu < 0.17$  eV from the combination Planck TT,TE,EE + SimLow + lensing + BAO data Aghanim et al. (2016b)<sup>14</sup>. Concerning the neutrino mass ordering, the addition of BAO measurements to CMB Planck measurements leads to odds for the normal vs. the inverted ordering of 4:3 and of 3:2, in the case of the Planck TT + lowP + BAO and Planck TT,TE,EE + lowP + BAO respectively, suggesting only very mild evidence for the normal ordering case Gerbino et al. (2017b). These results confirmed the previous findings obtained in Hannestad and Schwetz (2016). The authors of Vagnozzi et al. (2017) reported odds for the normal vs. the inverted ordering of 2.4:1 from the combination of Planck TT,TE,EE + BAO plus the SimLow prior on the reionization optical depth, i.e.,  $\tau = 0.05 \pm 0.009$ . Notice that if data are not informative enough, the choice of prior on  $m_{\text{lightest}}$  will make a difference in the odds ratio<sup>15</sup>.

Another possible avenue to exploit galaxy clustering data is to use the information contained in the full-shape of the galaxy power spectrum (see e.g., Allen et al., 2003; Elgaroy and Lahav, 2003; Hannestad, 2003; Spergel et al., 2003; Barger et al., 2004; Crotty et al., 2004; Hannestad and Raffelt, 2004; Tegmark et al., 2004, 2006; Fogli et al., 2007, 2008; Hamann et al., 2007b, 2008, 2010a; de Putter et al., 2012, 2014; Riemer-Sørensen et al., 2012, 2014; Giusarma et al., 2013a, 2016; Zhao et al., 2013; Cuesta et al., 2016; Vagnozzi et al., 2017). Notice however that using BAO is currently a more robust method, as the effects of the galaxy bias and non-linearities are not as severe as in the *shape* approach. In the minimal  $\Lambda$ CDM scenario, the BAO *geometrical* approach can supersede the neutrino mass constraints obtained from the *shape* one (see e.g., Hamann et al., 2010a; Giusarma et al., 2013b). Indeed, a dedicated analysis has been devoted in Vagnozzi et al. (2017) to compare the constraining power of these two different approaches to large scale structure data with the SDSS-III BOSS measurements. The conclusions are that, even if the latest

measurements of the galaxy power spectrum map a large volume of our universe, the geometric approach is still more powerful, at least within the minimal  $\Lambda$ CDM +  $\sum m_\nu$  cosmology. The better performance of BAO measurements is partly due to the upper cutoff applied in the scale  $k$  of the power spectrum when dealing with shape analyses (mandatory to avoid non-linearities), and partly due to the fact that two additional nuisance parameters are further required to relate the galaxy power spectrum to the matter power one<sup>16</sup>. As an example, the 95% CL bound of  $\sum m_\nu < 0.118$  eV obtained with Planck TT,TE,EE + BAO plus SimLow is degraded to  $\sum m_\nu < 0.177$  eV when replacing part of the BAO data [more concretely, the high redshift BOSS CMASS Data Release 11 (DR11) sample by the full-shape power spectrum measurements from the BOSS CMASS Data Release 12 (DR12)].

An alternative tracer to map out the large scale structure in our universe and improve the CMB-only bounds on the sum of the three active neutrinos is the Lyman- $\alpha$  forest, leading to neutrino mass bounds which turn out to be among the most constraining ones. By means of the one-dimensional Lyman- $\alpha$  forest power spectrum extracted by Palanque-Delabrouille et al. (2013) and combining these measurements with Planck TT,TE,EE + lowP + BAO, the authors of Palanque-Delabrouille et al. (2015) find a 95% CL upper limit of  $\sum m_\nu < 0.12$  eV. It is also remarkable the fact that, even without the addition of CMB data, the combination of the Lyman- $\alpha$  forest power spectrum of Palanque-Delabrouille et al. (2013), together with those from the XQ-100 quasars at  $z \simeq 3.5 - 4.5$  and the high-resolution HIRES/MIKE spectrographs at  $z = 4.2$  and  $z = 4.6$  (Viel et al., 2013), is already able to provide a limit of  $\sum m_\nu < 0.8$  eV (Yeche et al., 2017), showing clearly the enormous potential of small-scale probes to extract the neutrino masses.

The degeneracies among  $\sum m_\nu$  and the other cosmological parameters that appear when considering CMB data only can also be strongly alleviated by the addition of Supernova Ia luminosity distance data and/or local measurements of the Hubble parameter<sup>17</sup>. Concerning Supernovae Ia data, the most complete photometric redshift calibrated sample joins the SuperNova Legacy Survey (SNLS) and SDSS supernova catalogs. This Joint Light-Curve Analysis (JLA) catalog (Betoule et al., 2013, 2014; Mosher et al., 2014) has been used by the Planck collaboration and by other analyses to improve the constraints on  $\sum m_\nu$ , being its impact particularly crucial in non-minimal cosmologies (Vagnozzi et al., 2018), as we shall explain toward the end of this section. Concerning the value of  $H_0$ , as there exists a strong anti-correlation between the Hubble constant and  $\sum m_\nu$  when considering CMB measurements, larger mean values of  $H_0$  will lead to tighter constraints on the neutrino mass and consequently on the inverted mass ordering. When performing combined analyses of CMB and  $H_0$  data, the

<sup>14</sup>The BAO measurements exploited by the Planck collaboration include the 6dF Galaxy Survey (6dFGS) Beutler et al. (2011), the BOSS LOWZ BAO extraction of the spherical averaged  $D_V/r_s$  (Anderson et al., 2014; Ross et al., 2015) and the BOSS CMASS-DR11 data of Anderson et al. (2014).

<sup>15</sup>See e.g., the work of the authors of Simpson et al. (2017) and the explanation of their results in Schwetz et al. (2017) and Gariazzo et al. (2018a). See also Long et al. (2018), Heavens and Sellentin (2018), Handley and Millea (2018) for useful discussions concerning the prior choice on the neutrino mass ordering extraction.

<sup>16</sup>As stated in Vagnozzi et al. (2017), in the future, a deeper understanding of the non-linear regime of the galaxy power spectrum with massive neutrinos included, plus a better understanding of the galaxy bias could change the constraining power of full-shape analyses vs. BAO ones.

<sup>17</sup>See Jackson (2007) and Freedman and Madore (2010) for dedicated reviews concerning the different possible local measurements of  $H_0$ . Among them, the one based on Cepheid variables.

2015 Planck release relies on the reanalysis Efstathiou (2014) of a former  $H_0$  measurement based on the Hubble Space Telescope (HST) ( $H_0 = (73.8 \pm 2.4) \text{ km s}^{-1} \text{ Mpc}^{-1}$ ; Riess et al., 2011), which was in mild ( $2.5\sigma$ ) tension with the value of the Hubble parameter derived from 2013 Planck CMB data Ade et al. (2014). This reanalysis (Efstathiou, 2014) considers the original Cepheid data of Riess et al. (2011) and uses a new geometric maser distance estimate to the active galaxy NGC 4,258 (Humphreys et al., 2013), which is used as a distance anchor to find a value of the Hubble constant  $H_0 = (70.6 \pm 3.3) \text{ km s}^{-1} \text{ Mpc}^{-1}$ <sup>18</sup>. The limit on the sum of the three active neutrino masses reported by the Planck collaboration using this value of  $H_0$  is  $\sum m_\nu < 0.23 \text{ eV}$  at 95% CL, when combined with Planck TT + lowP + lensing + BAO + SNIa data. Other estimates of the Hubble constant, however, exist. The 2.4% determination of Riess et al. (2016) profits from new, near-infrared observations of Cepheid variables, and it provides the value  $H_0 = (73.02 \pm 1.79) \text{ km s}^{-1} \text{ Mpc}^{-1}$  Riess et al. (2016). As the former mean  $H_0$  value is higher than the one considered by the Planck collaboration, it will lead to tighter limits on  $\sum m_\nu$ . Indeed, the work of Vagnozzi et al. (2017) quotes the 95% CL bounds of  $\sum m_\nu < 0.196 \text{ eV}$  and  $\sum m_\nu < 0.132 \text{ eV}$  when combining with external data sets using the priors  $H_0 = (70.6 \pm 3.3) \text{ km s}^{-1} \text{ Mpc}^{-1}$  and  $H_0 = (73.02 \pm 1.79) \text{ km s}^{-1} \text{ Mpc}^{-1}$ , respectively. Focusing on the less conservative choice  $H_0 = (73.02 \pm 1.79) \text{ km s}^{-1} \text{ Mpc}^{-1}$ , odds for the normal vs. the inverted neutrino mass ordering of 3.3:1 were found for both the Planck TT, TE, EE + BAO + SimLow +  $H_0$  and the Planck TT, TE, EE + BAO + SimLow +  $H_0$  + Planck SZ data sets (Vagnozzi et al., 2017). The 95% CL upper bounds on the neutrino mass for these two combinations are  $\sum m_\nu < 0.094 \text{ eV}$  and  $\sum m_\nu < 0.093 \text{ eV}$ , respectively. These results indicate, once again, very mild evidence for the normal mass ordering, even within these more aggressive and less conservative scenarios, in which the very tight limit on  $\sum m_\nu$  is mostly due to the tension between CMB and direct measurements of the Hubble constant  $H_0$ , together with the strong degeneracy between  $\sum m_\nu$  and  $H_0$ . Using these results, we stress that having an upper bound  $\sum m_\nu \lesssim 0.1 \text{ eV}$  at 95% CL is not equivalent to having a 95% CL preference for normal ordering: the probabilities for normal ordering and inverted ordering, as computed from the odds 3.3:1, are approximately 77 and 23% (see also section 5.1).

In general, the combination of data sets that are inconsistent is potentially dangerous. Apart from the constraining effect on the neutrino mass limits when considering a particular prior on the Hubble constant  $H_0$ , there have been also other related cases in which the neutrino masses were a tool to accommodate tensions among different data sets. For instance, in the case of galaxy cluster counts, a larger neutrino mass could in principle fit both CMB and low-redshift universe constraints on the power spectrum normalization  $\sigma_8$  Allen et al. (2003). The effect of combining CMB and BAO observations with clusters and/or

shear data is presented in Costanzi et al. (2014), where it is shown that the inclusion of either cluster or shear measurements in the Planck + BAO joint analysis indicates a preference for  $\sum m_\nu > 0$  at more than  $2\sigma$ . However, the authors clearly state that these results can not be interpreted as a claim for a cosmological detection of the neutrino mass, but rather as a remedy to palliate the existing tension between clusters/shear data and Planck/BAO observations.

Finally, weak lensing constraints from the Dark Energy Survey Year 1 results Abbott et al. (2018) (DES Y1), have also recently provided bounds on the sum of the total neutrino mass. Based on 1321 deg<sup>2</sup> imaging data, DES Y1 analyses exploit the galaxy correlation function (from 650,000 luminous red galaxies divided into five redshift bins) and the shear correlation function (from twenty-six million source galaxies from four different redshift bins) as well as the galaxy-shear cross-correlation. The 95% CL upper bound reported on the neutrino mass after combining their measurements with Planck TT, TE, EE + lowP + BAO + JLA is  $\sum m_\nu < 0.29 \text{ eV}$ ,  $\sim 20\%$  higher than without DES measurements. The reason for this higher value of  $\sum m_\nu$  is that the clustering amplitude in the case of DES Y1 is mildly below the one preferred by Planck measurements. Since larger values of the neutrino mass will decrease the value of the clustering amplitude, the upper limit on the total neutrino mass is loosened by  $\sim 20\%$  after the DES results are also considered.

#### 4.3.2. Extensions to the Minimal $\Lambda$ CDM Universe

So far we have discussed the neutrino mass and neutrino mass ordering sensitivities within the minimal  $\Lambda$ CDM universe. However, these limits will change when additional parameters are introduced in the analyses.

The first and most obvious scenario one can consider is to test the stability of the neutrino mass limits when new physics is added in the neutrino sector. As already mentioned in section 3.2, short baseline neutrino experiments indicate that a light sterile neutrino at the eV scale may exist. These extra sterile species will contribute to the effective number of relativistic degrees of freedom,  $N_{\text{eff}}$ , defined by

$$\rho_{\text{rad}} = \left( 1 + \frac{7}{8} \left( \frac{4}{11} \right)^{4/3} N_{\text{eff}} \right) \rho_\gamma, \quad (15)$$

where  $\rho_{\text{rad}}$  ( $\rho_\gamma$ ) is the total radiation (CMB photons) energy density. In the standard picture  $N_{\text{eff}} = 3.046$  (Mangano et al., 2005; de Salas and Pastor, 2016). This number accounts for the three active neutrino contribution and considers effects related to non-instantaneous neutrino decoupling and QED finite temperature corrections to the plasma evolution<sup>19</sup>. Variations in  $N_{\text{eff}}$ , apart from the light sterile neutrino, may be related to the existence of additional relativistic particles, as thermally-produced axions (see below). Analyses in which both the active neutrino masses and the number of additional massless or

<sup>18</sup>The final result of Efstathiou (2014) is however  $H_0 = (72.5 \pm 2.5) \text{ km s}^{-1} \text{ Mpc}^{-1}$ , when the combination of the  $H_0$  results obtained with three different distance estimators is performed. The value  $H_0 = (70.6 \pm 3.3) \text{ km s}^{-1} \text{ Mpc}^{-1}$  is the only one of the three which shows a milder tension with the  $H_0$  estimate from Planck.

<sup>19</sup>The work of de Salas and Pastor (2016), including three-flavor neutrino oscillations, has revisited previous calculations including all the proper collision integrals for both diagonal and off-diagonal elements in the neutrino density matrix and quotes the value of  $N_{\text{eff}} = 3.045$ .

massive species are varied simultaneously have been extensively carried out in the literature (Hamann et al., 2007a, 2010b, 2011; Giusarma et al., 2011, 2013b; Archidiacono et al., 2013a,b; Di Valentino et al., 2013; Riemer-Sorensen et al., 2013), showing that the bounds on the active neutrino mass are relaxed when additional sterile species are added to the fermion content of the SM of particle physics. The constraints on the total neutrino mass  $\sum m_\nu$  are less stringent than in the standard three neutrino case due to the large degeneracy between  $\sum m_\nu$  and  $N_{\text{eff}}$ , which arises from the fact that a number of massless or sub-eV sterile neutrino species contributing to the radiation content of the universe will shift both the matter-radiation equality era and the location of the CMB acoustic peaks. This effect could be compensated by enlarging the matter content of the universe, implying therefore that larger values for the neutrino masses could be allowed. Consequently, a priori, the constraints on  $\sum m_\nu$  when  $N_{\text{eff}}$  is also a free parameter in the analyses are not very competitive. Fortunately, CMB measurements from the Planck collaboration help enormously in sharpening the measurement of  $N_{\text{eff}}$ , especially when considering polarization measurements at small scales: including data at high multipoles, one obtains  $\Delta N_{\text{eff}} < 1$  at more than  $4\sigma$  significance from Planck CMB observations alone. Indeed, the limit on the sum of the three active neutrino species considering also additional radiation neutrino species (i.e., massless sterile neutrino species) is  $\sum m_\nu < 0.178$  eV at 95% CL from Planck TT, TE, EE + lowP + BAO data, very similar to the bound  $\sum m_\nu < 0.168$  eV at 95% CL arising from the very same dataset within the minimal  $\Lambda$ CDM scenario with three active massive neutrinos. Another possible way of relaxing (or even avoiding) the cosmological neutrino mass limits is via the addition of non-standard interactions in the active neutrino sector (Beacom et al., 2004; Fardon et al., 2004; Afshordi et al., 2005; Hannestad, 2005b; Bell et al., 2006; Brookfield et al., 2006a,b; Bjælde et al., 2008; Ichiki and Keum, 2008; Mota et al., 2008; Boehm et al., 2012; Archidiacono and Hannestad, 2014; Dvali and Funcke, 2016; Di Valentino et al., 2018b).

Furthermore, additional relics different from sterile neutrinos, as thermal axions (Peccei and Quinn, 1977a,b; Weinberg, 1978; Wilczek, 1978), contributing to both  $N_{\text{eff}}$  at early times and to the hot dark matter component in the late-time universe, suppress small-scale structure formation and show effects very similar to those induced by the (active) three massive neutrino species. Therefore, the cosmological bounds on the three active neutrino masses are modified in scenarios with thermal axions (see Hannestad et al., 2007, 2008, 2010; Melchiorri et al., 2007; Archidiacono et al., 2013c; Giusarma et al., 2014; Di Valentino et al., 2015, 2016a,b), as these two species have to share the allowed amount of dark matter. Nonetheless, there are non-negligible differences among neutrinos and thermal axions: (a) axions are colder than neutrinos, as they decouple earlier; (b) since the axion is a scalar particle, an axion mass larger than the neutrino one is required in order to make identical contributions to the current mass-energy density of the universe; (c) in the case of axions, the contribution to  $N_{\text{eff}}$  is related to their mass, while for neutrinos this is usually not true. Consequently, the bounds on the axion mass are always less constraining

than for the neutrino, and  $\sum m_\nu$  is slightly more constrained in scenarios in which thermal axions are also present. For instance, Di Valentino et al. (2016c) quotes  $\sum m_\nu < 0.175$  eV at 95% CL from the Planck TT, TE, EE + lowP + BAO data combinations when considering only neutrinos, while the analyses in Di Valentino et al. (2016b), including massive axions, find  $\sum m_\nu < 0.159$  eV and  $m_a < 0.763$  eV, both at 95% CL, for the very same data combination.

There are also other ways of relaxing the cosmological neutrino mass bounds, related either to the early or the late-time accelerating periods in the universe. In the former case one can play with inflationary processes. There have been a number of studies devoted to explore their degeneracies with the neutrino sector (see the recent works of Hamann et al., 2007b; Archidiacono et al., 2013b; Joudaki, 2013; de Putter et al., 2014; Canac et al., 2016; Di Valentino et al., 2016a; Gerbino et al., 2017a). The authors of Di Valentino et al. (2016a) have considered a non-standard and parametric form for the primordial power spectrum, parameterized with the PCHIP (piecewise cubic Hermite interpolating polynomial) formalism with twelve nodes between  $k_1 = 5 \times 10^{-6} \text{ Mpc}^{-1}$  and  $k_2 = 10 \text{ Mpc}^{-1}$  and derived the neutrino mass constraints within this more general scenario. When only Planck TT + lowP measurements were considered, the 95% CL mass bound of  $\sum m_\nu < 0.75$  eV obtained with the usual power-law description of the primordial power spectrum was relaxed to  $\sum m_\nu < 2.16$  eV. This large value is explained in terms of the strong degeneracy between  $\sum m_\nu$  and the PCHIP nodes corresponding to the wave-numbers where the contribution of the Early ISW effect is located, in such a way that the effect induced by a non-zero neutrino mass is easily compensated by an increase of the primordial power spectrum at these scales only. BAO information improves considerably the limits in the PCHIP prescription, but it is the addition of high- $\ell$  polarization data what further constrains the effect. The 95% CL upper limit in the PCHIP scenario from the Planck TT, TE, EE + lowP + BAO data combination is  $\sum m_\nu < 0.218$  eV, quite close to the bound found when the usual power-law description is applied ( $\sum m_\nu < 0.175$  eV). Reference Gerbino et al. (2017a) deals instead with the robustness of the constraints on the scalar spectral index  $n_s$  under several neutrino physics scenarios. The authors have explored the shifts induced in the inflationary parameters for different choices of the neutrino mass ordering, comparing the approximate massive neutrino case (one massive eigenstate plus two massless species when the total mass is close to the minimum allowed value by oscillation data, and three degenerate massive neutrinos otherwise) vs. the exact case (normal or inverted mass orderings). While the mass-ordering assumptions are not very significant when  $\sum m_\nu$  is fixed to its minimum value, there is a shift in  $n_s$  when  $\sum m_\nu$  is a free parameter, inherited from the strong degeneracies in the  $\sum m_\nu$ ,  $H_0$  and  $\Omega_m h^2$  parameter space. Fortunately, BAO measurements revert the  $\sum m_\nu$ - $n_s$  anti-correlation present with CMB data only, and the shift in the spectral index turns out to be negligible.

The other possibility is to play with the late-time acceleration period and study how the neutrino mass bounds change. The current accelerated expansion of the universe, explained in terms

of a cosmological constant in the minimal  $\Lambda$ CDM scenario, may be due to a dynamical dark energy fluid with a constant equation of state  $w \neq -1$  or a time-dependent  $w(z)$  (Chevallier and Polarski, 2001; Linder, 2003), or to quintessence models, based on the existence of a cosmic scalar field (Peebles and Ratra, 1988; Ratra and Peebles, 1988; Wetterich, 1995; Caldwell et al., 1998; Zlatev et al., 1999; Wang et al., 2000), which provide a dynamical alternative to the cosmological constant scenario with  $w = -1$ . It is naturally expected that the neutrino mass bounds will increase when enlarging the parameter space. Indeed, when the dark energy equation of state is allowed to vary within the phantom region  $w < -1$ , there is a very well-known degeneracy between the dark energy equation of state  $w$  and the sum of the three active neutrino masses, as first noticed in Hannestad (2005a) (see also La Vacca and Kristiansen, 2009; Archidiacono et al., 2013b; Joudaki, 2013; Lorenz et al., 2017; Sutherland, 2018; Vagnozzi et al., 2018)<sup>20</sup>. It has been pointed out that for very high neutrino masses only dark energy models lying within the phantom region will be allowed. The reason for that is the following: a larger  $\sum m_\nu$  can be compensated by a larger  $\Omega_m$ , which in turn can be compensated by a smaller equation of state of the dark energy component, i.e.,  $w < -1$ . Interestingly, the recent work of Vagnozzi et al. (2018) shows that the cosmological bounds on  $\sum m_\nu$  become more restrictive in the case of a dynamical dark energy component with  $w(z) \geq -1$ . Following the usual dynamical dark energy description, whose redshift dependence is described by the standard Chevallier-Polarski-Linder (CPL) parametrization (Chevallier and Polarski, 2001; Linder, 2003), the authors of Vagnozzi et al. (2018) have shown that the combination of Planck TT, TE, EE + BAO + JLA plus the SimLow prior on the reionization optical depth provides, at 95% CL,  $\sum m_\nu < 0.11$  eV in the CPL case when restricting  $w(z) \geq -1$  (within the physical, non-phantom region), while  $\sum m_\nu < 0.13$  eV in the  $\Lambda$ CDM case. When  $w(z)$  is also allowed to be in the phantom region ( $w(z) < -1$ ) within the CPL parameterization, the resulting 95% CL constraint on the three active neutrino masses is  $\sum m_\nu < 0.37$  eV. These results have a direct impact on the cosmological preference for a given neutrino mass ordering. Following Hannestad and Schwetz (2016) and Vagnozzi et al. (2017), it is found that the normal ordering is mildly preferred over the inverted one, with posterior odds 3 : 1 for the data combination quoted above when  $w(z) \geq -1$ . On the contrary, if there is no such a restriction and  $w(z)$  can also take values in the phantom region, the odds are 1 : 1. The odds in the non-phantom dynamical dark energy case show a mild preference for normal ordering. Therefore, if neutrino oscillation experiments or neutrinoless double beta decay searches find that the neutrino mass ordering is the inverted one, if the current cosmic acceleration is due to a dynamical dark energy component, one would require this component to be phantom.

As a final point in this section, we would like to note that also in scenarios in which the current accelerated expansion is explained by means of modifications of gravity at ultra-large

length scales, the cosmological limits on neutrino masses will differ from those in the standard  $\Lambda$ CDM model (see e.g., Huterer and Linder (2007), Baldi et al. (2014), Hu et al. (2015), Shim et al. (2014), Barreira et al. (2014), Bellomo et al. (2017), Peirone et al. (2018), Renk et al. (2017), and Dirian (2017)).

## 5. GLOBAL 2018 DATA ANALYSIS

In this section we shall combine the available measurements that allow us to constrain the neutrino mass ordering, updating the results presented in Gariazzo et al. (2018a).

### 5.1. Bayesian Model Comparison

Before performing the analysis, we will briefly summarize the method we will adopt to compare the two possible orderings.

We will follow a Bayesian approach to model comparison (see previous work suggesting the Bayesian method as the most suited one for the mass ordering extraction in Qian et al., 2012 and Blennow, 2014)<sup>21</sup>, which makes use of the Bayesian evidence  $Z$ . This quantity, which is also known as the marginal likelihood, is defined as the integral over the entire parameter space  $\Omega_{\mathcal{M}}$  of the prior  $\pi(\theta) \equiv p(\theta|\mathcal{M})$  times the likelihood  $\mathcal{L}(\theta) \equiv p(d|\theta, \mathcal{M})$ , where  $\theta$  is the set of parameters that describe the model  $\mathcal{M}$  and  $d$  represents the available data:

$$Z_{\mathcal{M}} = \int_{\Omega_{\mathcal{M}}} \mathcal{L}(\theta) \pi(\theta) d\theta. \quad (16)$$

The posterior probability of the model  $\mathcal{M}$  can be written in terms of its prior probability  $\pi(\mathcal{M})$  times the Bayesian evidence  $Z_{\mathcal{M}}$ :

$$p(\mathcal{M}|d) \propto Z_{\mathcal{M}} \pi(\mathcal{M}), \quad (17)$$

where the proportionality constant depends only on the data. In our case we will be interested in comparing normal ordering (NO) and inverted ordering (IO), which can be considered as two different competing models  $\mathcal{M}_1 \equiv \text{NO}$  and  $\mathcal{M}_2 \equiv \text{IO}$ . The ratio of the posterior probabilities of the two models can be written as

$$\frac{p(\text{NO}|d)}{p(\text{IO}|d)} = B_{\text{NO,IO}} \frac{\pi(\text{NO})}{\pi(\text{IO})}, \quad (18)$$

having defined the Bayes factor as

$$B_{\text{NO,IO}} = Z_{\text{NO}}/Z_{\text{IO}}. \quad (19)$$

Assuming the same prior probabilities for normal and inverted ordering, the Bayes factor is what determines the odds in favor of one of the competing models. In particular we will indicate the results in terms of its natural logarithm  $\ln B_{\text{NO,IO}}$ , which will be positive when data will prefer normal ordering and negative otherwise. Quantitatively, the preference is given in terms of posterior odds, which are always  $|B_{\text{NO,IO}}| : 1$  in favor of the preferred model. The strength of the preference can be

<sup>20</sup>Interacting dark energy models can also change the neutrino mass constraints (see e.g., Gavela et al., 2009; La Vacca et al., 2009; Lopez Honorez and Mena, 2010; Reid et al., 2010; Guo et al., 2018).

<sup>21</sup>We also refer the reader to Blennow et al. (2014), which provides a comprehensive study of the sensitivity reach to the mass ordering in the context of the frequentist approach.

also translated into an empirical scale, which in our case is summarized in the third column of **Table 2**.

Let us briefly discuss the correspondence of the quoted levels that classify the strength of the preference in favor of one of the competing models. In the case of the neutrino mass ordering, we have only two possibilities (normal or inverted), so that  $p(\text{NO}|d) + p(\text{IO}|d) = \pi(\text{NO}) + \pi(\text{IO}) = 1$ . If we assign the same prior probability to the two cases,  $\pi(\text{NO}) = \pi(\text{IO}) = 1/2$ , it is easy to compute the posterior probability for each of the two cases, which will be

$$p(\text{NO}|d) = B_{\text{NO,IO}} / (B_{\text{NO,IO}} + 1), \quad (20)$$

$$p(\text{IO}|d) = 1 / (B_{\text{NO,IO}} + 1), \quad (21)$$

having used Equations (18, 19). The confidence levels for the rejection of the disfavored (e.g., inverted) mass ordering will then be  $x = 100 \times (1 - |B_{\text{NO,IO}}|^{-1})\%$ . For example, a Bayes factor  $B_{\text{NO,IO}} = 10$  corresponds to a rejection of the inverted ordering at 90% CL. If, instead, we want to reproduce the probability levels  $P = \text{erf}(N/\sqrt{2})$  that are usually associated to the classical  $N\sigma$  levels for a Gaussian measurement, being erf the error function and considering, for example,  $N \in (1, 2, 3, 4, 5)$ , the corresponding Bayes factors  $B$  can be computed to be  $B = P/(1 - P)$ , which gives us  $\ln B_{N\sigma} \simeq 0.77, 3, 5.9, 9.7, 14.37$ . Therefore, our “strong”, “very strong” and “decisive” levels roughly correspond to the  $> 3\sigma$ ,  $> 4\sigma$  and  $> 5\sigma$  probabilities, as indicated in the fourth column of **Table 2**.

## 5.2. Parameterization and Data

Our two competing models are described by the same number of parameters, listed with their priors in **Table 3**: the three neutrino mixing angles ( $\sin^2 \theta_{12}$ ,  $\sin^2 \theta_{13}$ ,  $\sin^2 \theta_{23}$ ), the CP violating phase  $\delta_{\text{CP}}$  and the parameters associated with neutrino masses, neutrinoless double beta decay ( $0\nu\beta\beta$ ) and cosmology, as we shall describe now.

We consider in our analysis the parameterization that uses the two mass splittings ( $\Delta m_{21}^2$  and  $\Delta m_{31}^2$ ) and the lightest neutrino mass  $m_{\text{lightest}}$  with logarithmic priors. This parameterization, strongly motivated by the physical observables, was shown to provide the optimal strategy to successfully explore the neutrino parameter space (see Gariazzo et al., 2018a)<sup>22</sup>. Within the other possible choice, that is, within the parametrization that uses the three neutrino masses as free parameters, most of the parameter space at high neutrino masses is useless for the data fit. Therefore, this second possibility is penalized by the Occam’s razor and we shall not explore it here.

The neutrino mixing parameters are constrained using the same data we described in section 2. The complete oscillation data set is indicated with the label “OSC” in the following.

For the cosmological part, we will describe the universe using the  $\Lambda$ CDM model and its six parameters: the baryon and cold

dark matter densities,  $\Omega_b h^2$  and  $\Omega_c h^2$ ; the optical depth to reionization,  $\tau$ ; the angular scale of the acoustic peaks through  $\Theta_s$  and the amplitude  $\log(10^{10} A_s)$  and tilt  $n_s$  of the power spectrum of initial curvature perturbations. In addition, we add the effect of the three massive neutrinos computing the evolution of the cosmological observables assuming three independent mass eigenstates, which, in terms of the parameters involved in our analyses, read as  $m_1 = m_{\text{lightest}} \left( m_1 = \sqrt{m_{\text{lightest}}^2 + |\Delta m_{31}^2|} \right)$ ,  $m_2 = \sqrt{m_{\text{lightest}}^2 + \Delta m_{21}^2} \left( m_2 = \sqrt{m_{\text{lightest}}^2 + |\Delta m_{31}^2| + \Delta m_{21}^2} \right)$  and  $m_3 = \sqrt{m_{\text{lightest}}^2 + \Delta m_{31}^2} \left( m_3 = m_{\text{lightest}} \right)$  for normal (inverted) neutrino mass orderings.

When considering cosmological data, we will focus on the Planck measurements of the CMB spectrum and on the most recent results from BAO observations. For the former we consider the 2015 Planck release (Adam et al., 2016; Ade et al., 2016b) of the high- $\ell$  likelihood (Aghanim et al., 2016a), together with a prior on  $\tau$  as obtained in the 2016 intermediate results (Aghanim et al., 2016b). For the purposes of our analyses, this will be sufficient to mimic the final Planck release which is expected within the next few months. Complementary to the CMB, we include in our calculations the final constraints from the SDSS BOSS experiment, the DR12 release, in the form denoted as “final consensus” Alam et al. (2017), which provides constraints from observing 1.2 million massive galaxies in three separate bands at effective redshifts 0.38, 0.51, and 0.61, plus results from the 6DF survey at  $z = 0.106$  (Beutler et al., 2011) and from the SDSS DR7 MGS survey at  $z = 0.15$  (Ross et al., 2015). The combined dataset including the mentioned CMB and BAO data will be denoted as “Cosmo.”

In addition, we shall impose a prior on the Hubble parameter as obtained in the recent results from Riess et al. (2016):  $H_0 = (73.24 \pm 1.74) \text{ km s}^{-1} \text{ Mpc}^{-1}$ . We will denote the data combinations including this prior with the label “ $H_0$ .”

Finally, concerning neutrinoless double beta decay, we vary the two Majorana phases in the entire available range ( $0-2\pi$ ) and the NMEs according to the range allowed by recent theoretical calculations. We revised the NME ranges adopted in Gariazzo et al. (2018a), which were the ones suggested in Giuliani and Poves (2012). Here we use these new ranges:  $[3.3 - 5.7]$  for  $^{76}\text{Ge}$  and  $[1.5 - 3.7]$  for  $^{136}\text{Xe}$ , following the  $1\sigma$  range proposed in Vergados et al. (2016).

We use  $0\nu\beta\beta$  data from the  $^{136}\text{Xe}$  experiments KamLAND-Zen Gando et al. (2016) and EXO-200 Albert et al. (2014) and from the  $^{76}\text{Ge}$  experiment Gerda, for which we use the results in Agostini et al. (2017b), since the latest publication Agostini et al. (2018) does not contain enough information that allows us to parameterize a likelihood function. The most stringent bounds, anyways, still come from KamLAND-Zen, so that not including the new Gerda results does not affect significantly our results. For the very same reason we do not include the results of CUORE Alduino et al. (2018a), for which the uncertainty on the NME of  $^{130}\text{Te}$  is very large and the constraints corresponding to most of the values of  $\mathcal{M}_{130\text{Te}}^{0\nu}$  are much looser than the ones from KamLAND-Zen, and of CUPID-0 Azzolini et al. (2018), which establishes a much less

<sup>22</sup>As we are making use of logarithmic priors here, we shall not report the upper limits we obtain on the sum of the neutrino masses, as they will be much smaller than the usually quoted results due to the volume effects associated with the use of the logarithmic prior, that naturally leads to a preference for small neutrino masses.

**TABLE 3** | Neutrino, cosmological and  $0\nu\beta\beta$  parameters used in the analysis, with the adopted priors.

Neutrino mixing and masses		Cosmological		$0\nu\beta\beta$	
Parameter	Prior	Parameter	Prior	Parameter	Prior
$\sin^2 \theta_{12}$	0.1 – 0.6	$\Omega_b h^2$	0.019 – 0.025	$\alpha_2$	0 – $2\pi$
$\sin^2 \theta_{13}$	0.00 – 0.06	$\Omega_c h^2$	0.095 – 0.145	$\alpha_3$	0 – $2\pi$
$\sin^2 \theta_{23}$	0.25 – 0.75	$\Theta_s$	1.03 – 1.05	$\mathcal{M}_{76\text{Ge}}^{0\nu}$	3.3 – 5.7
$\delta_{\text{CP}}/\pi$	0 – 2	$\tau$	0.01 – 0.4	$\mathcal{M}_{136\text{Xe}}^{0\nu}$	1.5 – 3.7
$\Delta m_{21}^2/\text{eV}^2$	$5 \times 10^{-5} - 10^{-4}$	$n_s$	0.885 – 1.04		
$\Delta m_{31}^2/\text{eV}^2$	$1.5 \times 10^{-3} - 3.5 \times 10^{-3}$	$\log(10^{10} A_s)$	2.5 – 3.7		
$\log_{10}(m_{\text{lightest}}/\text{eV})$	–5 – 0				

All the priors are linear in the corresponding quantity.

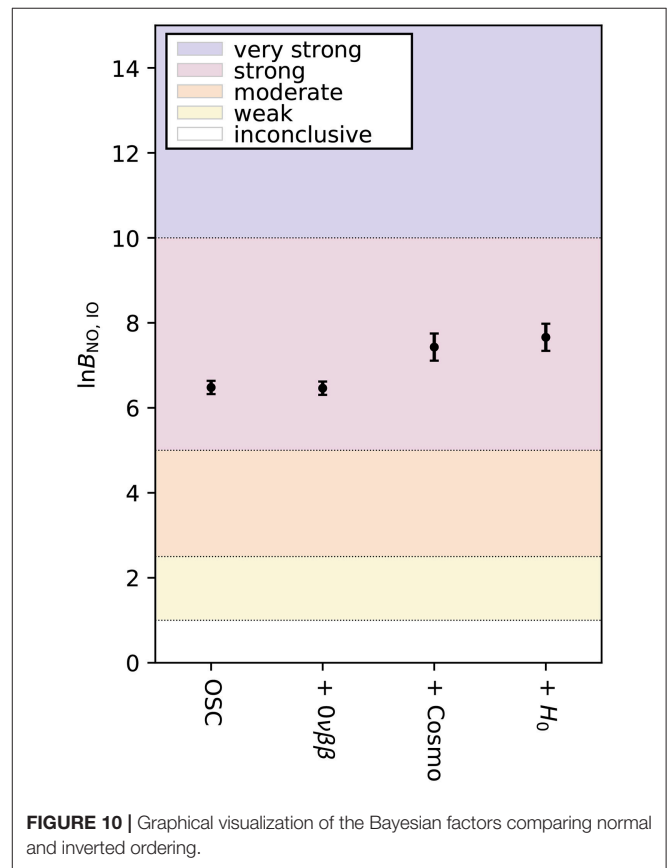
stringent limit on the  $^{82}\text{Se}$  half-life. The complete neutrinoless double beta set of data will be denoted as “ $0\nu\beta\beta$ .”

All the previously listed data are coded as likelihood quantities in a full Bayesian analysis. We compute the cosmological quantities using the Boltzmann solver CAMB Lewis et al. (2000), the likelihoods using the interface provided by CosmoMC Lewis and Bridle (2002), modified in order to take into account the oscillation and neutrinoless double beta decay data, while the calculation of the Bayesian evidence is committed to PolyChord Handley et al. (2015a,b).

### 5.3. Constraints on the Mass Orderings

The main results are depicted in **Figure 10**. The first data point corresponds to the Bayesian evidence from oscillation data only. Notice that the Bayes factor [ $\ln(B_{\text{NO,IO}}) = 6.5 \pm 0.2$  for concreteness] indicates *strong* evidence for the normal mass ordering *from oscillation data only*. This Bayes factor is translated into a  $\sim 3.2\sigma$  evidence favoring normal mass ordering. This result was expected in light of the results presented in section 2, arising from the frequentist joint analysis. There it was reported a  $\Delta\chi^2 = 11.7$  in favor of the normal mass ordering from the combination of all long baseline, reactor and atmospheric data, which corresponds, roughly, to  $\sim 3.4\sigma$ . Adding information from neutrinoless double beta decay searches does not affect the Bayesian analysis, as shown by the second data point in **Figure 10**, and as expected from previous work Gariazzo et al. (2018a).

Once CMB and BAO measurements are also added in the Bayesian analysis,  $\ln(B_{\text{NO,IO}}) = 7.4 \pm 0.3$  is obtained (see the third point in **Figure 10**), improving the significance of the preference for normal ordering from  $\sim 3.2\sigma$  to  $\sim 3.4\sigma$ . Notice that, even if the preference for the normal neutrino mass ordering is mostly driven by oscillation data, the information provided by cosmological observations is more powerful than that in the analysis carried out in Gariazzo et al. (2018a), as the Bayesian analyses here also include BAO measurements, together with CMB data. Indeed, from the two Bayes factors obtained considering oscillation data only [ $\ln(B_{\text{NO,IO}}) = 6.5 \pm 0.2$ ] and oscillation plus cosmological measurements [ $\ln(B_{\text{NO,IO}}) = 7.4 \pm 0.3$ ], it is straightforward to infer the probability odds for normal ordering arising exclusively from cosmology. By doing so, one obtains odds of 2.7 : 1 for the normal ordering against the

**FIGURE 10** | Graphical visualization of the Bayesian factors comparing normal and inverted ordering.

inverted one, in perfect agreement with the analyses of Vagnozzi et al. (2017), where odds of 2.4 : 1 with cosmological data only were reported when considering the very same data sets adopted here (albeit the odds were derived with an alternative method).

Finally, the addition of the prior on the Hubble constant raises the evidence for the normal ordering to  $\ln(B_{\text{NO,IO}}) = 7.7 \pm 0.3$  (i.e.,  $\sim 3.5\sigma$ ). This improvement is expected, as previously explained in section 4, since a prior on the Hubble constant breaks the degeneracy between  $\sum m_\nu$  and  $H_0$  and, therefore, sharpens the neutrino mass bounds from cosmology. By performing a similar exercise to the one previously quoted,

one finds that the odds for normal vs. inverted ordering from cosmology data only are 3.3 : 1 for the combination of CMB, BAO plus the  $H_0$  prior, again in excellent agreement with the results obtained in Vagnozzi et al. (2017).

## 6. FUTURE PROSPECTS

In this last section, we will explore the future prospects for the detection of the neutrino mass ordering. Let us clarify that many of the proposed methods are much less robust than the ones involving neutrino oscillations through matter (see section 6.1), and will likely give their first results much after the first experimental  $5\sigma$  determinations which are likely to be reached in the next 5 – 10 years. Many of the discussed methods, indeed, will give constraints on the neutrino mass ordering only as a secondary product of their operation and not as a main result, hence they are not optimized nor mainly focused on the mass ordering determination. Nevertheless, it is interesting to discuss these additional methods for different reasons. First of all, independent tests of the neutrino mass ordering from different methods are surely welcome to have more robust results. Secondly, the different methods can provide complementary information: if some inconsistencies or anomalies will appear, we will have new hints for our quest toward new physics beyond our current knowledge. In conclusion, even if the question regarding the neutrino mass ordering will be solved within the next few years by the currently running experiments or their immediate extensions, its study through the other methods we discuss here will be useful to shed more light on the topic and provide more interesting information on neutrino physics and beyond. This is why we do not focus only on neutrino oscillation experiments (section 6.1), which will probably provide the first and strongest results, but also on more exotic cases as determinations from decay experiments (sections 6.2 and 6.3) cosmological constraints (section 6.4), measurements from the 21 cm surveys (section 6.5), and probes which involve neutrinos emitted by core-collapse supernova explosions (section 6.6) or relic neutrinos from the early Universe (section 6.7).

### 6.1. Prospects From Oscillations

As we have seen in section 2, the combination of all current neutrino experiments leads to a preference for normal ordering of  $3.4\sigma$ , within the context of the latest frequentists global data analyses. The Bayesian analysis described in the previous section confirms these results, as we have reported a  $3.2\sigma$  evidence for normal mass ordering. In principle, one may expect to achieve further sensitivity on the neutrino mass ordering from more precise data by the current long-baseline and atmospheric neutrino experiments, since these experiments will still run for some time before the new experiments will take over. However, it is not easy to predict the final results of current experiments, since the sensitivity to the mass ordering is highly correlated to the true value of the CP phase  $\delta_{CP}$ . The NOvA experiment alone expects a  $3\sigma$  sensitivity for 30–50% of the values of  $\delta_{CP}$  by 2,024 Himmel (2018). If  $\delta_{CP} = 3\pi/2$ , the expected sensitivity would be higher than that and, then, a very strong result could be obtained by 2,024 Himmel (2018). Note,

however, that the NOvA sensitivity analysis considers a fixed value of  $\theta_{13}$  and does not marginalize over  $\Delta m_{31}^2$ . Upgrading T2K to T2K-II will improve the sensitivity substantially, since the experiment should gather around  $20 \times 10^{21}$  POT by 2026, which would be roughly 6 times the current amount of data<sup>23</sup>. Combining beam data from T2K with atmospheric data from SK can improve the sensitivity even further, as shown in Abe et al. (2018a). Performing a combined fit of T2K, NOvA and eventually SK could bring the sensitivity to the  $5\sigma$  level within a few years. In any case, apart from the combinations of different experiments, a very robust determination of the neutrino mass ordering from a single current experiment seems rather unlikely. Indeed, one of the main goals of the next-generation neutrino oscillation experiments, including new long-baseline, reactor, and atmospheric neutrino detectors, will be to perform the determination of the mass ordering by a single experiment. The upcoming facilities will be able to measure the neutrino mass ordering with astonishing precision. In this section we briefly discuss some of the proposed projects and their physics potential.

### Long-Baseline Experiments

The Deep Underground Neutrino Experiment (DUNE) (Acciarri et al., 2015, 2016a,b; Strait et al., 2016) will be a new long-baseline accelerator experiment, with a small near detector and a huge far detector with a fiducial mass of 40 kton located 1,300 km away from the neutrino source at Fermilab. With its powerful 1.1 MW beam, it will be exposed to around  $15 \times 10^{20}$  POTs (protons on target) per year, which will lead to a huge number of events and therefore to high precision measurements of the neutrino oscillation parameters. As explained in section 2, the presence of matter affects differently the neutrino appearance probabilities for normal and inverted mass orderings. DUNE, with the longest baseline ever for an accelerator neutrino experiment, will be able to measure the neutrino mass ordering with a significance above  $5\sigma$  for any set of the oscillation parameters ( $\theta_{23}, \delta_{CP}$ ) after 7 years of data taking. Note that this sensitivity could be further increased by using an improved energy reconstruction method, as shown in De Romeri et al. (2016). On the other hand, the sensitivities could also be biased by the potential presence of new physics beyond the SM, such as non-standard neutrino interactions (Miranda et al., 2006; Coloma, 2016; Coloma and Schwetz, 2016; de Gouvêa and Kelly, 2016; Forero and Huber, 2016; Masud and Mehta, 2016; Bakhti and Khan, 2017; Coloma et al., 2017; Deepthi et al., 2017; Forero and Huang, 2017; Farzan and Tortola, 2018), deviations from unitarity Blennow et al. (2017); Dutta et al. (2017); Escribuela et al. (2017) or the presence of light-sterile neutrinos (Berryman et al., 2015, 2016; Agarwalla et al., 2016; Coloma et al., 2018). Indeed, besides providing very precise information about the neutrino oscillation mechanism, the DUNE experiment will also be very useful to test different models for neutrino masses and mixings (Chatterjee et al., 2017a,b; Pasquini et al., 2017; Agarwalla et al., 2018b; Chakraborty et al., 2018; Srivastava et al., 2018a,b) as well as to check for various effects of new physics such as the

<sup>23</sup>We are not aware of any study showing the T2K or SK expectations to the mass ordering in the next few years.

ones mentioned above, neutrino decay scenarios (Coloma and Peres, 2017; Ascencio-Sosa et al., 2018; Choubey et al., 2018c), quantum decoherence Balieiro Gomes et al. (2018) or even CPT invariance (de Gouvêa and Kelly, 2017; Barenboim et al., 2018b,c) and Lorentz invariance (Barenboim et al., 2018a; Jurkovich et al., 2018).

There are also plans to build a larger version of the Super-Kamiokande detector, Hyper-Kamiokande Abe et al. (2018b), that will be very similar to its predecessor but with a fiducial mass of 560 kton, 25 times larger than Super-Kamiokande. The Hyper-Kamiokande detector will be a requirement for the upgrade of T2K, the T2HK (Tokai-to-Hyper-Kamiokande) experiment (Abe et al., 2015). The very massive detector together with the upgraded neutrino beam from J-PARC will guarantee a huge number of neutrino events and therefore larger statistics. As a consequence, T2HK will be able to determine the neutrino mass ordering after few years of running time with very high significance, as well as to explore new physics scenarios (see for instance Abe et al., 2017, 2018b; Agarwalla et al., 2018a). In combination with atmospheric data from Hyper-Kamiokande, a  $3\sigma$  rejection of the wrong mass ordering would be expected after 5 years of data taking. A third project has been proposed as an extension of T2HK to Korea, the T2HKK (Tokai-to-Hyper-Kamiokande-and-Korea) experiment (Abe et al., 2018c). This proposal includes a second far detector facility for the J-PARC neutrino beam, located at 1,000–1,300 km from the source. The longer path traveled within the Earth by the neutrinos detected in T2HKK will result in an enhanced sensitivity to the neutrino mass ordering if compared to T2HK alone.

The synergies and complementarities among the three long-baseline proposals above, DUNE, T2HK and T2HKK, have been discussed in Ballett et al. (2017). It is found that the combination of their experimental results may significantly mitigate the limitations of a given experiment, improving the precision in both the determination of the mass ordering and the measurement of CP violation.

Note that, although here we have focused on the long-baseline side of DUNE and Hyper-Kamiokande, they are actually designed as multi-purpose experiments, with a rich physics program aiming to study the neutrino oscillations with accelerator, atmospheric and solar neutrinos as well as to detect neutrinos from astrophysical sources and proton decay.

## Atmospheric Experiments

In atmospheric neutrino experiments, the sensitivity to the mass ordering comes from the matter effects that distort the pattern of neutrino oscillations inside Earth, see Equation (4). Based on the oscillatory pattern that depends on the reconstructed neutrino energy and zenith angle, an ideal experiment would observe a given number of events in each energy and zenith angle bin as shown in **Figure 11**. Comparing the observed two-dimensional histograms with the theoretical ones for normal (**Left**) or inverted ordering (**Right**) allows to determine the true mass ordering that

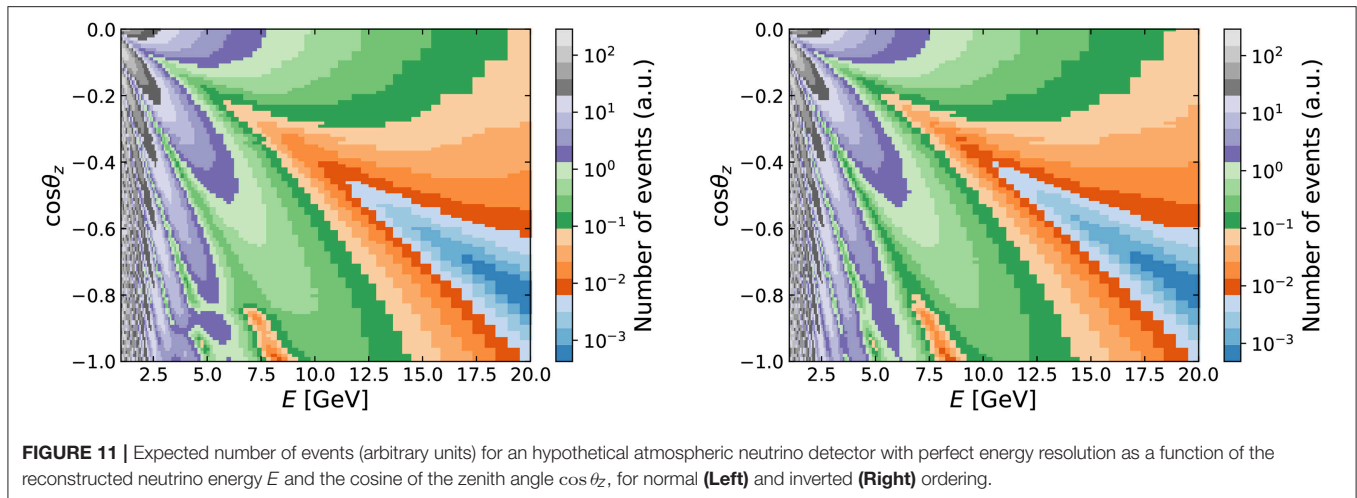
is realized in nature. In the following we list some of the future projects with this aim.

The Oscillation Research with Cosmics in the Abyss (ORCA) experiment Adrian-Martinez et al. (2016) will be a large neutrino telescope placed deep inside the Mediterranean sea. It will detect the Cherenkov light emitted by the muons and electrons created by the interactions of atmospheric neutrinos in the sea and that propagate into water. Unlike its precursor, ANTARES, with 12 lines and a separation of 70 meters between neighboring optical modules, ORCA will have 60 lines with modules separated by 9 m. Due to the matter effects on the propagation of atmospheric neutrinos, the ORCA experiment will be able to measure the neutrino mass ordering with very good precision. In particular, a  $3\sigma$  determination of the mass ordering can be expected after only 3 years of data taking, with even higher significance for the case in which nature has chosen normal ordering and the upper octant for the atmospheric mixing angle. Several studies have been performed in order to analyze the sensitivity of ORCA to the standard oscillation parameters (Ribordy and Smirnov, 2013; Yañez and Kouchner, 2015). Its potential to determine the Earth matter density through neutrino oscillation tomography Winter (2016) or to test new physics scenarios (Ge et al., 2017; Capozzi et al., 2018b) have also been extensively discussed.

PINGU (Precision IceCube Next Generation Upgrade) (Aartsen and Adelaide, 2014) is a planned upgrade of the IceCube DeepCore detector, an ice-Cherenkov neutrino telescope which uses the antarctic ice as a detection medium. The IceCube design aims at the detection of very high energy neutrinos, with an energy threshold above the relevant energy range for neutrino oscillations. However, the denser instrumented region DeepCore allows IceCube to decrease its energy threshold down to  $E_{\text{th}} = 6.3$  GeV. A further improvement with an even denser zone, PINGU, could lower  $E_{\text{th}}$  to only a few GeV. With this very low-energy threshold, one of the main purposes of PINGU is the determination of the neutrino mass ordering, with expected sensitivities similar to the ORCA experiment<sup>24</sup>. Besides that, PINGU is expected to have the best sensitivity to  $\nu_\tau$  appearance and to determine accurately the octant of the atmospheric mixing angle. The PINGU capabilities to detect high-energy supernova neutrinos (Murase, 2018), and to investigate scenarios beyond the Standard Model, such as non-standard interactions (Choubey and Ohlsson, 2014) or dark matter self-interactions (Chen et al., 2014; Robertson and Albuquerque, 2018) have been also analyzed in the literature.

The India-based Neutrino Observatory (INO) is a very ambitious project, aiming to detect atmospheric neutrinos with a 50 kton magnetized iron calorimeter (ICAL) (Ahmed et al., 2017). The most outstanding feature of the INO experiment will be its capability to distinguish neutrinos from antineutrinos in an event by event basis. As a result, the identification of the matter effects in the neutrino and antineutrino propagation will be much cleaner in comparison with the sea water/ice Cherenkov detectors. Indeed, one of the main scientific goals of INO will be the determination of the neutrino mass ordering (Ghosh

<sup>24</sup>The effect of statistic and systematic uncertainties on the PINGU sensitivity to the mass ordering has been presented in Capozzi et al. (2015).



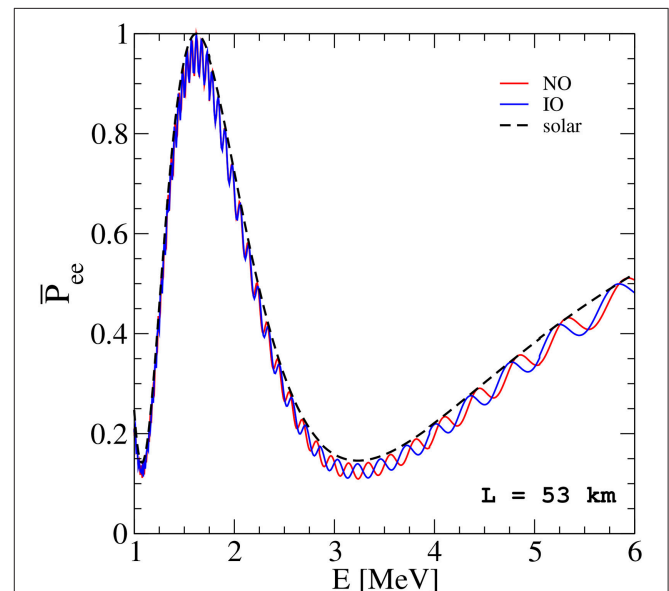
et al., 2013). According to the Physics White Paper of the ICAL (INO) Collaboration Ahmed et al. (2017), after 10 years run, INO will be able to identify the correct neutrino mass ordering with a significance larger than  $3\sigma$ . As the experiments discussed above, the atmospheric neutrino results from INO can also be used to test the presence of new physics beyond the SM, such as CPT- or Lorentz violation (Chatterjee et al., 2014), sterile neutrinos (Behera et al., 2017; Thakore et al., 2018), dark matter related studies Dash et al. (2016); Choubey et al. (2018a), non-standard neutrino interactions Choubey et al. (2015) or decaying neutrinos (Choubey et al., 2018b).

### Medium-Baseline Reactor Experiments

We have focused so far on extracting the neutrino mass ordering from matter effects in the neutrino propagation through the Earth interior. An alternative technique is that provided by medium-baseline reactor neutrino experiments Petcov and Piai (2002). For baselines of the order of 50 km, the survival probability for reactor antineutrinos exhibits a pattern that may allow the discrimination between normal and inverted mass orderings. Indeed, for such distances, the electron antineutrino survival probability is given by the following expression:

$$P_{\bar{\nu}_e \rightarrow \bar{\nu}_e} = 1 - \cos^4 \theta_{13} \sin^2 2\theta_{12} \sin^2 \Delta_{21} - \sin^2 2\theta_{13} [\sin^2 \Delta_{31} + \sin^2 \theta_{12} \sin^2 \Delta_{21} \cos 2\Delta_{31} \mp \frac{\sin^2 \theta_{12}}{2} \sin 2\Delta_{21} \sin 2|\Delta_{31}|], \quad (22)$$

where  $\Delta_{ij} = \frac{\Delta m_{ij}^2 L}{4E}$  and the minus (plus) sign in the last term corresponds to normal (inverted) mass ordering. This probability contains a main oscillatory term with a frequency given by the solar neutrino mass splitting  $\Delta m_{21}^2$ , plus an additional term whose frequency depends on the sign of the atmospheric splitting  $\Delta m_{31}^2$ , i.e., on the neutrino mass ordering. The effect of the ordering over the neutrino survival probability in a medium-baseline reactor experiment is illustrated in Figure 12. There, we depict in black the oscillatory term corresponding to the solar splitting frequency. The red (blue) line corresponds to the



**FIGURE 12 |** Electron antineutrino survival probabilities in a medium-baseline reactor experiment with  $L = 53$  km. The red (blue) line corresponds to normal (inverted) mass ordering using the best-fit values from Table 2, while the black line contains the main term in the survival probability, given by the solar mass splitting frequency by setting  $\Delta m_{31}^2 = 0$ .

full neutrino survival probability for normal (inverted) mass ordering. Note that this plot was obtained using the best-fit values from Table 1 for each ordering.

The Jiangmen Underground Neutrino Observatory (JUNO) An et al. (2016) is a 20 kton multi-purpose underground liquid scintillator detector. The site of the experiment, located 53 km away from the Yangjiang and Taishan nuclear power plants in China, was chosen to optimize its sensitivity to the neutrino mass ordering, one of its main physics goals. Like any other reactor neutrino experiment, JUNO will be sensitive to the disappearance of electron antineutrinos, with about  $10^5$  events

expected after 6 years of run time. From this high-statistics data sample, JUNO will try to reconstruct with extremely good precision the neutrino oscillation spectrum and to discriminate the different high-frequency behavior for normal and inverted mass ordering, as illustrated in Equation (22) and **Figure 12**. For a projected energy resolution of 3% at 1 MeV, JUNO will be able to establish the neutrino mass ordering at the level of  $3\text{--}4\sigma$  in 6 years. Its combination with the PINGU facility could lead to a high significance improvement of the individual capabilities of these two experiments (see Blennow and Schwetz, 2013).

Apart from the mass ordering, JUNO will also provide precision measurements of the solar oscillation parameters,  $\theta_{12}$  and  $\Delta m_{21}^2$ , with an accuracy of around 1%. In this sense, JUNO might help to solve the observed disagreement between the mass splitting measured at solar experiments and at the reactor experiment KamLAND. If the discrepancy persists after new measurements by JUNO and future solar results by Super-Kamiokande, it could be considered as an indication of new physics Farzan and Tortola (2018). Moreover, JUNO will be sensitive to different types of new physics scenarios beyond the SM, as studied in Khan et al. (2013), Bakhti and Farzan (2014), Girardi et al. (2014), Li and Zhao (2014), Ohlsson et al. (2014), Abrahão et al. (2015), Liao et al. (2017), Zhao et al. (2017), and Krnjaic et al. (2018).

In parallel to JUNO, there is a proposal to extend the already existing experiment RENO with a third medium-baseline detector located at a distance of 47 km. This new project is known as RENO-50 Kim (2015), given its location, at approximately 50 km from the Hanbit power plant, in South Korea. The detector would consist of a 18 kton ultra-low-radioactive liquid scintillator instrumented with 15,000 high quantum efficiency photomultiplier tubes. Using the same technique described above, RENO-50 will be able to determine the neutrino mass ordering as well as the solar oscillation parameters with extremely good precision. Conceived as multi-purpose detectors, JUNO and RENO-50 will have a wide physics program, including not only the observation of reactor and solar neutrinos, but also neutrinos from supernova bursts, the diffuse supernova neutrino background, atmospheric neutrinos and geoneutrinos.

## 6.2. Prospects From Beta-Decay Experiments

As already mentioned in section 3, the determination of the mass ordering through the observation of the energy spectrum near the endpoint of  $\beta$ -decay or similar will be extremely challenging, because an impressive energy resolution is required to distinguish the kink due to the second and third mass eigenstates in the spectrum. We list here the main projects that aim at detecting the neutrino mass in the future and comment their perspectives for the mass ordering determination.

The first experiment we will comment on is KATRIN, which has recently started operations and aims at a detection of the effective electron antineutrino mass with a sensitivity of 0.2 eV (Angrik et al., 2004; Sejersen Riis et al., 2011). The first results from KATRIN are expected in early 2019, but the final target statistics will be reached after 3 year of data taking. Thanks to the

detailed study of the detector systematics which has been carried out, it is possible that the final mass determination will reach a better sensitivity than the nominal one of 0.2 eV, eventually reaching something closer to 0.1 eV (Parno, 2018). Even with the more optimistic sensitivity, however, it will be impossible for KATRIN to determine the mass ordering.

Other tritium experiments exploiting different technologies include the Project-8 (Doe et al., 2013; Asner et al., 2015; Esfahani et al., 2017) experiment, which will use the Cyclotron Radiation Emission Spectroscopy (CRES) in order to determine the mass of the electron antineutrino. The technique consists in measuring the frequency of cyclotron radiation emitted by the electrons released during tritium decay and spiraling into a magnetic field. The frequency can then be related with the electron energy and consequently the energy spectrum can be determined. At the moment, Project-8 is in the calibration phase (phase-II) (Rybka, 2018) for a small prototype which will not have enough sensitivity to be competitive in the determination of the neutrino mass. Next phases include a large volume system using molecular tritium (phase-III), starting in 2020, which will be competitive in determining the neutrino mass and will serve as an intermediate step before moving to phase-IV, which will use atomic tritium, required in order to avoid uncertainties related to the existence of excited molecular tritium states. Project-8 in its atomic tritium phase is expected to reach the sensitivity  $m_{\bar{\nu}_e} \lesssim 40$  meV with an exposure of  $10 - 100 \text{ m}^3 \text{ year}$ , sufficient to probe the values of  $m_{\bar{\nu}_e}$  allowed in the context of inverted ordering (Esfahani et al., 2017), so that in case of no observation we will know that the ordering of neutrino masses must be normal.

Another interesting class of the experiments includes the HOLMES (Alpert et al., 2015; Giachero et al., 2017) and ECHO (Eliseev et al., 2015) experiments, which both aim at the determination of the electron neutrino mass through observations of the endpoint of the electron capture decay of  $^{163}\text{Ho}$ , which practically proceeds through the measurement of de-excitation transitions of the Dy atoms, which are produced in the process  $^{163}\text{Ho} + e^- \rightarrow ^{163}\text{Dy}^* + \nu_e$  (De Rujula and Lusignoli, 1982). As for the tritium  $\beta$ -decay, also the endpoint of the  $^{163}\text{Ho}$  electron capture spectrum depends on the value of the neutrino masses and, in principle, it would be possible to determine the mass ordering in this way. Besides the experimental and theoretical problems that the HOLMES and ECHO collaborations must face, however, it seems that the current technology is not yet at the level of precision required for the mass ordering determination. The HOLMES demonstrator, currently running, should reach a sensitivity of  $m_{\nu_e} \lesssim 10$  eV by the end of 2018, while the full-scale experiment, possibly starting in 2019, has a target sensitivity of  $m_{\nu_e} \lesssim 1$  eV (Gastaldo, 2018). ECHO, on the other hand, is running a first phase (ECHO-1k) which has also a target of  $m_{\nu_e} \lesssim 10$  eV in 1 year, while the full scale ECHO-100k will reach  $m_{\nu_e} \lesssim 1.5$  eV in 3 year of data taking, expected to start in 2019 (Gastaldo, 2018). Both results are impressive when compared with the current upper limit on the electron neutrino mass using the same isotope, which is 225 eV (Springer et al., 1987), more than two orders of magnitude larger.

Finally, to conclude this subsection we want to mention that the PTOLEMY proposal (Betts et al., 2013; Baracchini et al., 2018), aiming at the detection of the relic neutrino background and recently approved by the Scientific Committee of the Laboratori Nazionali del Gran Sasso (LNGS), will be able to study and possibly determine the mass ordering through the observation of the  $\beta$ -spectrum of tritium decay. PTOLEMY will be discussed later in section 6.7.

### 6.3. Prospects From Neutrinoless Double Beta Decay

We list here the future perspectives for neutrinoless double beta decay experiments in terms of sensitivity to the half-life for the processes of interest (where possible). As we already commented in section 3.2, the conversion between the half-life  $T_{1/2}^{0\nu}$  and the effective Majorana mass  $m_{\beta\beta}$  depends on the NME and the phase space factor of the process of interest, see Equation (7). In order to exclude the inverted ordering allowed range for  $m_{\beta\beta}$  (in case there is no sterile neutrino), one would need to constrain  $m_{\beta\beta} \lesssim 10$  meV, which corresponds to  $T_{1/2}^{0\nu} \simeq 1 \times 10^{28}$  year, with some dependence on the material (phase space and NME). This means that none of the current generation experiments will be able to reach the required sensitivity, and we will have to wait for next-generation upgrades and new projects. Many of the information listed in the following has been taken from Agostini et al. (2017a) and Giuliani (2018).

#### Current Generation Experiments

Let us firstly address the current generation of experiments, which at most will be able to start exploring the three-neutrino inverted mass ordering regime or to probe the upper range for  $m_{\beta\beta}$  allowed within the 3+1 neutrino scenario. The experiments will be listed in alphabetical order.

AMORE Alenkov et al. (2015) is an experiment devoted to determine the life-time of  $^{100}\text{Mo}$ . After a first pilot run, the current status (AMORE-I) is to test the technology with a  $^{100}\text{Mo}$  mass of 5–6 kg, in order to demonstrate the scalability before moving to the full scale (AMORE-II) detector, which will use 200 kg of material and is expected to start around 2020, with a final target sensitivity of  $T_{1/2}^{0\nu} \simeq 5 \times 10^{26}$  year.

CUORE Artusa et al. (2015) and Alduino et al. (2016, 2018b), already mentioned in section 3.2, works with  $^{130}\text{Te}$  and is already taking data with the full scale detector, which will have as ultimate sensitivity  $T_{1/2}^{0\nu} \simeq 9 \times 10^{25}$  year after 5 year of data taking (Adams et al., 2018; Ouellet, 2018).

The KamLAND-Zen experiment (Gando et al., 2016; Gando, 2018), after the previous successful data taking period, is now upgrading the detector for a new observation run with approximately 750 kg of  $^{136}\text{Xe}$  and a new balloon inside the KamLAND detector. The target sensitivity for the upcoming phase is around  $T_{1/2}^{0\nu} \simeq 5 \times 10^{26}$  year, a factor of five larger than the current limit (Gando et al., 2016).

A smaller experiment is NEXT (Martín-Albo et al., 2016), which is running background studies in the Canfranc laboratories in Spain. NEXT will use high pressure  $^{136}\text{Xe}$  TPCs, which will allow an impressive tracking of the emitted particles through scintillation and electroluminescence. A

prototype with 10 kg of natural Xenon will start data taking this year to demonstrate that the expected background control and particle tracking have been achieved. In 2019 NEXT is expected to start a new phase with 100 kg of  $^{136}\text{Xe}$ , which will reach  $T_{1/2}^{0\nu} \simeq 1 \times 10^{26}$  year with 5 year of data.

A similar project is called Panda-X-III (Chen et al., 2017), which is based in the Jinping underground laboratories in China. Panda-X-III will run the first phase using 200 kg of  $^{136}\text{Xe}$  to reach  $T_{1/2}^{0\nu} \simeq 1 \times 10^{26}$  year in 3 year.

Going to a different concept, SNO+ (Andringa et al., 2016) will feature a detector of 760 ton of ultra-pure liquid scintillator. SNO+ will be a multipurpose detector, as it will be capable of studying reactor, solar, supernova and geoneutrinos, and also to probe proton decay (Orebi Gann, 2018). After the background studies will be completed, a 0.5% loading will be performed, inserting  $^{130}\text{Te}$  in the detector to measure double beta decay processes. The target sensitivity after 5 year is  $T_{1/2}^{0\nu} \simeq 2 \times 10^{26}$  year. Future plans for the SNO+ experiment include the further  $^{130}\text{Te}$  loading to 1%, or even more, of the detector mass, with the advantage that increasing the  $^{130}\text{Te}$  amount will not influence the backgrounds but only the signal. The final target for this second phase is to reach  $T_{1/2}^{0\nu} \simeq 1 \times 10^{27}$  year, thus starting to cover the inverted ordering allowed range.

Let us finally comment the SuperNEMO experiment (Arnold et al., 2010; Patrick and Xie, 2017), which uses  $^{82}\text{Se}$  for its study. SuperNEMO is particularly interesting because it will be able to perform a full topological reconstruction of the events, which is extremely important in case of detection because it opens the possibility to directly test the mechanism that underlies neutrinoless double beta decay and, in principle, to determine the lepton-number violating process. A first demonstrator of about 7 kg is expected to start in 2018 and to reach  $T_{1/2}^{0\nu} \simeq 6 \times 10^{24}$  year with 2.5 year of data. The subsequent plans include an extension with a  $\sim 100$  kg scale detector with 20 modules, which will be able to probe  $T_{1/2}^{0\nu}$  up to  $1 \times 10^{26}$  year, and the possibility to use the  $^{150}\text{Nd}$  isotope, for two reasons: to have a more favorable phase space when converting  $T_{1/2}^{0\nu}$  to  $m_{\beta\beta}$  and to get rid of the Rn background which affects the  $^{82}\text{Se}$  measurements (Giuliani, 2018).

As a summary, some of the current generation experiments will be able to probe the inverted ordering range of  $m_{\beta\beta}$  within the standard three neutrino framework and assuming an exchange of light Majorana neutrinos. However, none of them will be able to rule out completely the inverted mass ordering, because of the uncertainty related to the NMEs.

#### Next Generation Experiments

The situation will be different for the following generation of experiments, which are mostly the natural evolution of current experiments to the ton-scale of decaying material. With the increased amount of material, a larger statistics will be achieved and stronger bounds, of the order of  $T_{1/2}^{0\nu} \simeq 1 \times 10^{28}$  year, will be feasible. We briefly discuss here the main current proposals for the next 10–20 years. The time schedules for these projects will be necessarily vague, as they will depend on the results of the present ones.

Let us start with CUPID (CUORE Upgrade with Particle ID) (Wang et al., 2015; Azzolini et al., 2018), which will be the evolution of the previously discussed CUORE experiment. The goal of CUPID is to use particle tracking in order to have a better discrimination of background and ultimately allow a background-free experiment: the target is  $< 0.1$  counts/(ton year) (Ouellet, 2018). A first demonstrator, named CUPID-0 (Azzolini et al., 2018), is already running with about 5 kg of  $^{82}\text{Se}$ , and already obtained the strongest-to-date constraint on the life-time on this isotope. In order to reach the target sensitivity  $T_{1/2}^{0\nu} \gtrsim 1 \times 10^{27}$  year, however, further improvement in the crystals quality and radio-purity is required. A full development plan for CUPID is currently under discussion.

Although not specifically designed for neutrinoless double beta decay searches, the DARWIN (DARK matter WImp search with liquid xenoN) experiment (Aalbers et al., 2016) will have sensitivity to a number of rare decay phenomena. The primary target of DARWIN is to perform direct detection of dark matter in a wide mass-range of the experimentally accessible parameter space for Weakly Interacting Massive Particles (WIMPs), to the level at which neutrino interactions with the target become an irreducible background (the so-called neutrino floor). The core of the detector will be a multi-ton liquid xenon time projection chamber. Having a large mass, low-energy threshold and ultra-low background level, DARWIN will also search for solar axions or galactic axion-like particles, measure the low-energy solar neutrino flux with  $< 1\%$  precision, observe coherent neutrino-nucleus interactions, detect galactic supernovae neutrinos and study the double beta decay of  $^{136}\text{Xe}$  (Aalbers et al., 2016). Even if it will be build using natural Xenon without isotope enrichment, DARWIN will contain 3.5 t of  $^{136}\text{Xe}$ . If the target energy resolution of  $1 - 2\%$  at 2.5 MeV will be achieved, the sensitivity of DARWIN will be  $T_{1/2}^{0\nu} \simeq 5.6 \times 10^{26}$  year with an exposure of 30 t yr (Aalbers et al., 2016). The estimated ultimate sensitivity, which will be achieved only with a complete mitigation of the material background and 140 t year of exposure, is claimed to be  $T_{1/2}^{0\nu} \simeq 8.5 \times 10^{27}$  year (Aalbers et al., 2016).

The successor of KamLAND-Zen, KamLAND2-Zen (Shirai, 2017; Gando, 2018; Giuliani, 2018) will benefit the upgrades of KamLAND into KamLAND2, including the improved light collection and better energy resolution guaranteed by the new photomultipliers, together with an increased amount of  $^{136}\text{Xe}$ , to reach at least 1 ton of material. These upgrades will be performed after the completion of KamLAND-Zen 800, expected to start this year. The target sensitivity after 5 year will be  $m_{\beta\beta} \lesssim 20 \text{ meV}^{25}$ , sufficient for “fully covering the inverted ordering region” (Shirai, 2017). Future studies will also test the possibility to accommodate scintillating crystals inside the detector and run a multi-isotope experiment.

Back to  $^{76}\text{Ge}$ -based experiments, the efforts of the Gerda and Majorana collaborations will join to work on the LEGEND (Large Enriched Germanium Experiment for Neutrinoless Double beta decay) experiment. Learning from both its precursors, LEGEND will need further background rejection and

will be built in different phases. The first module, LEGEND-200, made of 200 kg of Germanium and expected to start in 2021, will be built on top of the existing Gerda infrastructures and will have a target sensitivity  $T_{1/2}^{0\nu} \simeq 1 \times 10^{27}$  year in 5 year. The full scale detector, LEGEND-1000, consisting in several modules summing up to a total of 1 ton of material, will have as an ultimate goal  $T_{1/2}^{0\nu} \simeq 1 \times 10^{28}$  year in 10 year (Abgrall et al., 2017), giving a full coverage of the inverted mass ordering region.

Even larger in size, nEXO (Albert et al., 2018; Kharusi et al., 2018) will replace the EXO-200 experiment after its completion, expected this year. The new detector will use 5 ton of Xenon in order to reach  $T_{1/2}^{0\nu} \simeq 1 \times 10^{27}$  year with just 1 year of data and  $T_{1/2}^{0\nu} \simeq 1 \times 10^{28}$  year with the full statistics, after 10 year.

After the completion of the upcoming phase, NEXT-100 will be possibly upgraded into NEXT 2.0, which will need a 1.5 ton of Xenon to obtain the statistics for achieving  $T_{1/2}^{0\nu} \simeq 1 \times 10^{27}$  year after 5 year of running (Agostini et al., 2017a; Giuliani, 2018).

In the same way, the Panda-X-III collaboration is also planning a 1 ton scale phase II with a target of  $T_{1/2}^{0\nu} \simeq 1 \times 10^{27}$  year (Chen et al., 2017).

The last comment regards another interesting possibility related to the SNO+ experiment. The THEIA proposal (Orebi Gann, 2015) is a concept study for a gigantic detector of something around 30–100 kton of target material which will use water-based liquid scintillator. Such target allows to track both Cherenkov and delayed scintillation light, thus enabling high light yield and low-threshold detection with attenuation close to that of pure water. The result is that such a detector would be able to achieve excellent background rejection thanks to directionality, event topology, and particle ID, with very large statistics. Loading of metallic ions which can undergo neutrinoless double beta decay would enable to use the THEIA detector for studying the Dirac/Majorana nature of neutrinos. Given the size of the detector, a 0.5% loading will allow to store several tons of decaying material, which naturally result in huge statistics when compared with current experiments. A 3% loading with natural (not enriched) Tellurium will be sufficient to reach, assuming  $m_{\beta\beta} \simeq 15 \text{ meV}$ , a  $3\sigma$  discovery in 10 year (Alonso et al., 2014; Giuliani, 2018).

## 6.4. Prospects From Cosmology

There are a number of studies in the literature focused on forecasting the expected sensitivity from both future CMB and large scale structure surveys to the total neutrino mass  $\sum m_\nu$  (de Putter et al., 2009; Abazajian et al., 2011; Carbone et al., 2011, 2012; Hamann et al., 2012; Basse et al., 2014; Font-Ribera et al., 2014b; Allison et al., 2015; Archidiacono et al., 2017; Amendola et al., 2018; Di Valentino et al., 2018a; Sprenger et al., 2018).

Awaiting for very futuristic measurements which may allow for the extraction of each of the individual masses associated to the neutrino mass eigenstates (see section 4), the extraction of the neutrino mass ordering strongly relies on the error achieved on  $\sum m_\nu$  for a chosen *fiducial* value of the neutrino mass.

<sup>25</sup>The collaboration does not report the sensitivity in terms of the half-life of the decay.

A complete, updated and useful summary is provided in Table II of (Lattanzi and Gerbino, 2018), which shows the expected sensitivity  $[\sigma(\sum m_\nu)]$  from different future cosmological probes, assuming the fiducial value  $\sum m_\nu = 0.06$  eV. Nevertheless, the authors of Gerbino et al. (2017b) considered different fiducial values for the total neutrino mass and computed the odds for the normal vs. the inverted ordering for possible combinations of future cosmological probes including the current information from oscillation experiments. We shall comment on these results toward the end of this section.

#### 6.4.1. CMB Prospects

Two main missions are expected to lead the next decade generation of CMB experiments, albeit a number of other experiments are in progress between now and then. The latter list includes ground-based observatories as the ACT (Atacama Cosmology Telescope) (De Bernardis et al., 2016), the SPT-3G (South Pole Telescope-3G) (Benson et al., 2014), the Simons Array (Suzuki et al., 2016), CLASS (Essinger-Hileman et al., 2014), BICEP3 (Ahmed et al., 2014), and the Simons Observatory<sup>26</sup>. The two main missions are expected to be the CMB-Stage IV project (Abazajian et al., 2016) and CORE (Cosmic Origin Explorer) (Delabrouille et al., 2018). The former, the CMB-Stage IV project (Abazajian et al., 2016), expected to be *the definitive ground-based CMB experiment*, aims at 250000 detectors operating for 4 years, covering a 40% fraction of the sky. Depending on the beam size and on the effective noise temperature, CMB Stage IV could reach sensitivities of  $\sigma(\sum m_\nu) = 0.073 - 0.11$  eV, assuming  $\sum m_\nu = 0.058$  eV as the fiducial model and an external prior on the reionization optical depth of  $\tau = 0.06 \pm 0.01$ , see (Abazajian et al., 2016) for the precise configuration details. The latter, CORE, a medium-size space mission proposed to the European Space Agency (ESA) (Delabrouille et al., 2018), is expected to have an one order of magnitude larger number of frequency channels and a twice better angular resolution than Planck. With these improved capabilities, CORE could achieve a sensitivity of  $\sigma(\sum m_\nu) = 0.044$  eV (Di Valentino et al., 2018a; Lattanzi and Gerbino, 2018), for a fiducial total neutrino mass of 0.06 eV. As it is evident from these estimates, future CMB experiments alone will not be able to determine the neutrino mass ordering.

#### 6.4.2. Large Scale Structure Prospects

From the large scale structure perspective, in analogy to the future CMB probes, there are also two main surveys, DESI (Dark Energy Spectroscopic Instrument) (Levi et al., 2013; Aghamousa et al., 2016), a ground-based telescope which will improve the SDSS-III and IV legacies (BOSS Dawson et al., 2013 and eBOSS galaxy surveys Dawson et al., 2016), and the Euclid space mission (Amendola et al., 2018). The baseline design of DESI assumes that it will run over 5 years, covering 14000 deg<sup>2</sup> of the sky targeting four different tracers: Bright,

Luminous Red and Emission Line Galaxies plus quasars in the redshift interval ( $0.05 < z < 1.85$ ), and a Lyman- $\alpha$  survey in the  $1.9 < z < 4$  redshift interval. The expected error in  $\sum m_\nu$  from DESI and Planck data is 0.02 eV. This number corresponds, approximately, to a  $2\sigma$  determination of the neutrino mass ordering in case neutrinos have the minimal mass within the normal ordering scenario (Aghamousa et al., 2016). The authors of (Font-Ribera et al., 2014b) have also explored a number of possible combinations of DESI with other surveys. Namely, combining DESI measurements with the final results from DES, an error of 0.017 eV in  $\sum m_\nu$  could be achieved. Their most constraining result,  $\sigma(\sum m_\nu) = 0.011$  eV, however, arises from an extension of the DESI survey, together with data from Euclid and LSST (Large Synoptic Survey Telescope) (Ivezic et al., 2008; Abell et al., 2009) (see below). In case this small error is achieved, the neutrino mass ordering can be determined with a high accuracy, again assuming a massless lightest neutrino and normal ordering. Other analyses have also reduced the nominal  $\sigma(\sum m_\nu) = 0.02$  eV expected from the DESI survey replacing the Planck CMB information with that expected from the future CMB Stage IV (Abazajian et al., 2016) or CORE (Di Valentino et al., 2018a) probes.

Euclid, an ESA mission expected to be launched early in the upcoming decade, mapping  $\sim 15,000$  deg<sup>2</sup> of the sky, has also been shown to provide excellent capabilities to test the neutrino properties (Amendola et al., 2018). Euclid will focus on both galaxy clustering and weak lensing measurements, which, combined with Planck CMB data, will provide errors on the sum of the neutrino masses of  $\sigma(\sum m_\nu) = 0.04$  eV (Carbone et al., 2011) and  $\sigma(\sum m_\nu) = 0.05$  eV (Kitching et al., 2008), respectively, albeit exploiting the mildly non-linear regime could highly reduce these errors (Audren et al., 2013). While these errors are large to extract useful information concerning the neutrino mass ordering, the weak gravitational lensing abilities from Euclid have also been considered to extract the neutrino mass ordering when it lies far enough from the degenerate region (see e.g., Amendola et al., 2018). The addition of future CMB measurements, as those from CORE, could notably improve the expected Euclid sensitivity. The authors of Archidiacono et al. (2017) have shown that CMB measurements from CORE, combined with full shape measurements of the galaxy power spectrum and weak lensing data from Euclid, could reach  $\sigma(\sum m_\nu) = 0.014$  eV. This result clearly states the complementarity of cosmic shear and galaxy clustering probes, crucial to test the neutrino mass ordering. Further improved measurements of the reionization optical depth  $\tau$  could strengthen this bound and consequently the sensitivity to the ordering of the neutrino masses (Liu et al., 2016; Archidiacono et al., 2017; Sprenger et al., 2018), see the following section. Other future large scale structure surveys are the aforementioned LSST and WFIRST (Spergel et al., 2013, 2015), that will lead as well to accurate measurements of the total neutrino mass. Their combination with e.g., Euclid could provide an error of a few meV on the total neutrino mass,  $\sigma(\sum m_\nu) \lesssim 0.008$  eV (Jain et al., 2015).

<sup>26</sup>For a detailed study on the prospects from pre- and post-2020 CMB experiments on the extraction of cosmological parameters, including the total neutrino mass  $\sum m_\nu$ , see also (Errard et al., 2016).

The above neutrino mass (neutrino mass ordering) projected errors (sensitivities), even if strongly constraining, are highly dependent on the fiducial value of  $\sum m_\nu$ , in the sense that the majority of the forecasts (a) are usually carried out assuming the minimal neutrino mass allowed within the normal ordering scheme, i.e.,  $\sum m_\nu \simeq 0.06 \text{ eV}^{27}$ ; (b) the quoted sensitivities in the neutrino mass ordering are computed via an extrapolation of the error on the sum of neutrino masses rather than from proper Bayesian comparison tools. The authors of Gerbino et al. (2017b) found that a future CMB CORE-like satellite mission, even combined with a 1% measurement of the Hubble constant  $H_0$  and with the future DESI survey (Font-Ribera et al., 2014b; Aghamousa et al., 2016) can not extract the ordering if nature has chosen a value for the neutrino masses of  $\sum m_\nu = 0.1 \text{ eV}$ . Odds for the normal vs. the inverted ordering of 1:1 were reported (Gerbino et al., 2017b). When considering the minimum allowed value for the total neutrino mass set by neutrino oscillation experiments, i.e.,  $\sum m_\nu = 0.06 \text{ eV}$ , they quote odds of 3:2 (9:1) for the case in which CORE and the prior on  $H_0$  without (with) DESI measurements are considered<sup>28</sup>. Therefore, the next generation of CMB and large scale structure surveys will be sensitive to the mass ordering only if it is normal and the lightest neutrino mass is close to zero. The significance of such a measurement will crucially depend on how far  $\sum m_\nu$  lies from its minimum allowed value from oscillation probes.

## 6.5. Prospects From 21 cm Surveys

Cosmological measurements of the redshifted 21 cm hydrogen line provide a unique test of the Epoch of Reionization (EoR) and the “dark ages,” the period before the first stars formed. The 21 cm line is due to spin-flip transitions in neutral hydrogen between the more energetic triplet state and the ground singlet state, and its intensity depends on the ratio of the populations of these two neutral hydrogen hyperfine levels. At a given observed frequency  $\nu$ , the 21 cm signal can be measured in emission or in absorption against the CMB. The so-called differential brightness temperature  $\delta T_b$  therefore refers to the contrast between the temperature of the hydrogen clouds and that of the CMB, which, for small frequencies and up to first order in perturbation theory, reads as Madau et al. (1997); Furlanetto et al. (2006); Pritchard and Loeb (2012); Furlanetto (2015)

$$\delta T_b(\nu) \simeq 27 x_{\text{HI}} (1 + \delta_b) \left( 1 - \frac{T_{\text{CMB}}}{T_S} \right) \left( \frac{1}{1 + H^{-1} \partial \nu_r / \partial r} \right) \left( \frac{1+z}{10} \right)^{1/2} \left( \frac{0.15}{\Omega_m h^2} \right)^{1/2} \left( \frac{\Omega_b h^2}{0.023} \right) \text{ mK}, \quad (23)$$

where  $x_{\text{HI}}$  is the fraction of neutral hydrogen,  $\delta_b$  is the baryon overdensity,  $\Omega_b h^2$  and  $\Omega_m h^2$  the present baryon and matter contributions to the mass-energy budget of the Universe,  $H(z)$  the Hubble parameter and  $\partial \nu_r / \partial r$  the comoving peculiar velocity

gradient along the line of sight. Therefore, 21 cm cosmology aims to trace the baryon overdensities via transitions in neutral hydrogen.

There are a number of current and future experimental setups devoted to detect the 21 cm global signal averaged over all directions in the sky, as EDGES (Experiment to Detect the Reionization Step) (Bowman and Rogers, 2010), the future LEDA (Large Aperture Experiment to Detect the Dark Ages) (Greenhill and Bernardi, 2012) or DARE (Moon space observatory Dark Ages Radio Experiment) (Burns et al., 2012). The EDGES experiment has quoted the observation of an absorption profile located at a frequency of  $78 \pm 1 \text{ MHz}$ , corresponding to a redshift of  $z \sim 17$ , with an amplitude of about a factor of two larger than the maximum expected in the canonical  $\Lambda$ CDM framework (Bowman et al., 2018). This recent claim has led to a number of studies aiming either to explain the effect or to constrain some non-standard scenarios (Barkana, 2018; Barkana et al., 2018; Berlin et al., 2018; Cheung et al., 2018; Clark et al., 2018; Costa et al., 2018; D’Amico et al., 2018; Ewall-Wice et al., 2018; Falkowski and Petraki, 2018; Feng and Holder, 2018; Fialkov et al., 2018; Fraser et al., 2018; Hektor et al., 2018; Hill and Baxter, 2018; Hirano and Bromm, 2018; Kang, 2018; Liu and Slatyer, 2018; Mahdawi and Farrar, 2018; McGaugh, 2018; Mitridate and Podo, 2018; Muñoz and Loeb, 2018; Muñoz et al., 2018; Pospelov et al., 2018; Safarzadeh et al., 2018; Slatyer and Wu, 2018; Witte et al., 2018; Yang, 2018).

Fluctuations in the redshifted 21 cm signal can be used to compute the power spectrum of the differential brightness temperature. This is the major goal of experiments as GMRT (Giant Metrewave Radio Telescope) (Ananthakrishnan, 1995; Paciga et al., 2011), LOFAR (LOw Frequency ARray) (van Haarlem et al., 2013), MWA (Murchison Widefield Array) (Tingay et al., 2013) and PAPER (Precision Array for Probing the Epoch of Reionization) (Parsons et al., 2010; Ali et al., 2015; Pober et al., 2015), targeting statistical power-spectrum measurements of the 21 cm signal employing large radio interferometers. Even if current experiments have not yet detected the 21 cm cosmological signature, the PAPER collaboration has recently improved the previous upper limits at  $z = 8.4$  (Ali et al., 2015). Next decade, high-redshift 21 cm experiments include the SKA (Square Kilometre Array) (Mellema et al., 2013) and HERA (Hydrogen Epoch of Reionization Array) (Beardsley et al., 2015). A three-dimensional map of the 21 cm signal could also be obtained by means of the so-called intensity mapping technique, which measures the collective emission from neutral hydrogen in dense clumps, targeting large regions without resolving individual galaxies in the post-reionization era ( $z \lesssim 3$ ) (Chang et al., 2008; Loeb and Wyithe, 2008; Wyithe et al., 2008; Villaescusa-Navarro et al., 2014b). The experimental efforts for this technique include the GBT-HIM project, with the GBT (Green Bank Telescope) (Chang et al., 2016), CHIME (Canadian Hydrogen Intensity Mapping Experiment) (Newburgh et al., 2014), the Tianlai project (Chen and Xu, 2016) and SKA-mid frequency (Dewdney et al., 2015) (see e.g., Bull et al., 2015).

Despite the fact that the primary task of future 21 cm experiments is to improve our current knowledge of the reionization history, they provide as well an additional tool for

<sup>27</sup>The authors of Amendola et al. (2018) have nonetheless presented constraints for different fiducial models.

<sup>28</sup>For the CORE CMB mission, data were generated following Refs (Bond et al., 1997, 2000). For DESI, mock  $r_s H(z)$  and  $d_A(z)/r_s$  data were generated for the three DESI tracers in the  $0.15 < z < 1.85$  redshift range, accordingly to Font-Ribera et al. (2014b).

fundamental cosmology (Scott and Rees, 1990; Tozzi et al., 2000; Iliev et al., 2002; Barkana and Loeb, 2005a,b; McQuinn et al., 2006; Santos and Cooray, 2006; Bowman et al., 2007; Mao et al., 2008; Visbal et al., 2009; Clesse et al., 2012; Liu and Parsons, 2016; Liu et al., 2016), complementary to CMB missions and galaxy surveys. Indeed, 21 cm cosmological observations will play a very important role concerning neutrino physics. As previously stated, there are two types of experiments. First of all, we will have observations focused on the pre-reionization and EoR periods, that can probe very large volumes (where the non-linear scale is small). Remember that the largest signal from relic neutrino masses and their ordering appears at scales which, at the redshifts attainable at galaxy clustering surveys, lie within the mildly non-linear regime. Therefore, one needs to rely on either N-body simulations or on analytical approximations for the matter power spectrum to simulate the massive neutrino signature. EoR 21 cm experiments will achieve the scales required to observe the neutrino signature within the linear regime, avoiding the simulation problems described in section 4.2. In this regard, these probes may widely surpass the constraints on neutrino masses expected from even very large galaxy surveys (McQuinn et al., 2006; Mao et al., 2008; Pritchard and Pierpaoli, 2008; Tegmark and Zaldarriaga, 2009; Abazajian et al., 2011; Oyama et al., 2013, 2016; Shimabukuro et al., 2014). Furthermore, the neutrino constraints will be largely independent of the uncertainties in the dark energy fluid, which, as we have seen in section 4.3.2, have instead a non-negligible impact in lower redshift, galaxy survey measurements. This is a byproduct of using the 21 cm line to trace the matter overdensities: at redshifts  $z \lesssim 2$ , the universe starts to be dominated by the dark energy fluid and the growth of matter perturbations is modified depending on the dark energy equation of state  $w(z)$ , whose precise time-evolution remains unknown. Consequently, for a given perturbation in the matter fluid, a suppression in its structure growth could be either due to the presence of massive neutrinos or to an evolving dark energy fluid. Focusing at higher redshifts, the neutrino mass constraints from 21 cm probes will be largely independent of the uncertainties in the dark energy fluid properties.

Expectations from MWA, SKA and FFTT (Fast Fourier Transform Telescope) (Tegmark and Zaldarriaga, 2009) were considered in Mao et al. (2008). Focusing on 4000 h of observations of two areas in the sky in a range of  $z = 6.8 - 8.2$  (divided into three redshift bins) and a value of  $k_{\text{max}} = 2 \text{ Mpc}^{-1}$ , the reported errors on  $\sum m_\nu$  are 0.19 (0.027), 0.056 (0.017), 0.007 (0.003) for MWA, SKA and FFTT, respectively, in their middle (optimistic) scenarios<sup>29</sup>, when combined with Planck measurements. These forecasts were performed for a fiducial  $\Omega_b h^2 = 0.0875$ , which corresponds to a quite high value for the neutrino mass, lying in the fully degenerate neutrino mass spectrum.

The authors of Oyama et al. (2013) devoted a dedicated analysis to establish the potential for extracting the neutrino mass ordering combining the FFTT capabilities with future CMB

polarization measurements. Based exclusively on the induced effect of the neutrino mass ordering on the cosmic expansion rate, a robust 90% CL neutrino mass ordering extraction was reported if  $\sum m_\nu < 0.1 \text{ eV}$ , regardless the underlying true ordering (i.e., normal or inverted). In Oyama et al. (2016), the authors propose to combine ground-based CMB polarization observations, SKA Phase 2 and BAO measurements from DESI. With these data sets, a  $2\sigma$  extraction of the neutrino mass ordering seems feasible, unless the neutrino spectrum is degenerate. Notice that these results arise from the signature induced by the neutrino mass ordering in the cosmic expansion rate, as the minimum cutoff of the wavenumber in the 21 cm observations is  $k_{\text{min}} = 0.06h \text{ Mpc}^{-1}$ , while the wavenumber corresponding to the neutrino free-streaming scale is  $k_{\text{min}} \simeq 0.02h \text{ Mpc}^{-1}$  for a 0.05 eV massive neutrino.

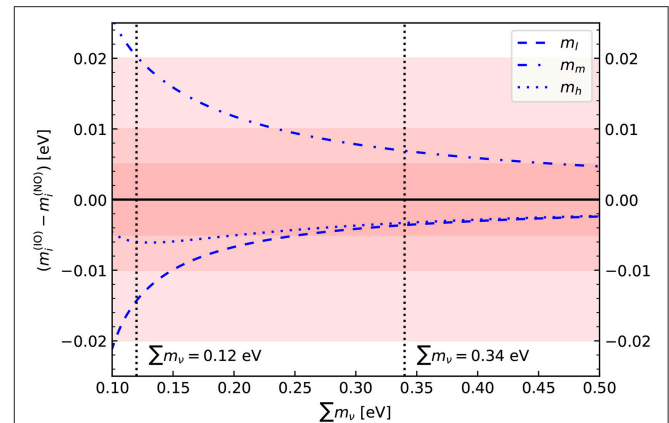
More futuristic 21 cm experiments, as FFTT, may open the possibility of going beyond measurements of the total neutrino mass  $\sum m_\nu$  and measure the individual neutrino masses, revealing the uniqueness of such experiments for constraining the neutrino properties. As shown in Figure 9 in section 4, the differences in the power spectra for the two possible mass orderings are tiny. Therefore, exquisite precision measurements are required to identify such signatures. Galaxy surveys, already discussed in the previous section, are limited by two facts. The first one is related to non-linearities, which will not allow for a measurement of the power spectrum at scales  $k > 0.2h \text{ Mpc}^{-1}$  at small redshifts, see section 4.2. Since the non-linear scale at  $z = 8$  is  $k \simeq 3h \text{ Mpc}^{-1}$ , both SKA and FFTT can measure the entire linear region and be more sensitive to the scale-dependent suppression, which is different in the two neutrino mass orderings. The second one is related to the fact that a galaxy survey requires a large number density of tracers to ensure a good sensitivity at small scales, while for 21 cm surveys, tracing the ubiquitous permeating hydrogen, a high-density antennae distribution will already warrant excellent small-scale sensitivities. One drawback of 21 cm probes are foregrounds, which should be kept under control.

The authors of Pritchard and Pierpaoli (2008) have studied the perspectives for extracting the individual neutrino masses with SKA and FFTT, finding that FFTT could be able to distinguish all the three neutrino masses from zero at the  $3\sigma$  level, due to its enormous effective volume (see Figure 3 of Pritchard and Pierpaoli, 2008). Extracting the neutrino mass ordering directly from the individual masses, however, was shown to be a very difficult achievement. Our calculations show that, for the total neutrino mass we use here as a reference,  $\sum m_\nu = 0.12 \text{ eV}$ , the differences among the lightest ( $l$ ), medium ( $m$ ) and heaviest ( $h$ ) neutrino mass eigenstates between the normal and inverted orderings are  $(|\Delta m_l|, |\Delta m_m|, |\Delta m_h|) = (0.015, 0.0209, 0.0059) \text{ eV}$ , which, especially for the case of  $|\Delta m_h| = 0.0059 \text{ eV}$ , are tiny and very difficult to resolve, even with very futuristic 21 cm measurements. While increasing the exposure of FFTT may improve its capabilities for this purpose (the error in the most optimistic FFTT scenario of Mao et al. (2008) is 0.003 eV), it seems an extremely challenging task. Figure 13 depicts the differences in the values of the three neutrino masses as a function of the total neutrino mass between

<sup>29</sup>These scenarios differ in the assumptions concerning the power modeling, the prior on the reionization history and the residual foregrounds cutoff scale, among other factors (see Tegmark and Zaldarriaga, 2009).

inverted and normal orderings. We show with a dashed vertical line our representative case  $\sum m_\nu = 0.12$  eV (the present most constraining 95% CL upper limit) and another one for  $\sum m_\nu = 0.34$  eV (the most recent 95% CL bound from the Planck collaboration after the removal of systematics in their polarization data at high angular scales Aghanim et al., 2016b). Notice that, as expected, the differences between the values of the three neutrino masses decrease with the total neutrino mass. In this regard, the lower the neutrino mass, the easier it could be to single out the three neutrino mass eigenstates, because they are more separated. However, an extraction of the mass ordering in the non-degenerate region via the values of the individual neutrino masses seems very difficult. Indeed, **Figure 13** illustrates the values of the individual neutrino masses for the heaviest, medium and lightest states for the normal and inverted orderings as a function of the total neutrino mass. The bands, from **Top to Bottom**, depict the errors  $\sigma(m_i) = 0.02$  eV and  $\sigma(m_i) = 0.01$  eV, together with the very futuristic FFTT one,  $\sigma(m_i) = 0.005$  eV. For an error of  $\sigma(m_i) = 0.02$  eV, there is no hope to disentangle the individual neutrino masses, as the error bands overlap for the heaviest, medium and lightest masses in all the parameter space. If instead one could achieve  $\sigma(m_i) = 0.01$  eV, a measurement of the individual neutrino masses in the non-degenerate region could be possible at the  $1 - 2\sigma$  level, but in order to unravel the ordering one would need very extreme conditions as, for instance, a value of  $\sum m_\nu$  very close to 0.1 eV independently determined with very small errors. The bottom plot in **Figure 14** shows the results if we assume the futuristic value of  $\sigma(m_i) = 0.005$ , expected to be achieved by FFTT. In this case, a measurement of the three neutrino masses will be achieved. Furthermore, in this (very optimistic) situation, the error bars will be, in principle, sufficiently small to detect the presence (or the lack) of two massive neutrino states with masses in the 0.02–0.03 eV range, required if the ordering is normal to explain  $\sum m_\nu \simeq 0.1$  eV, which would strongly confirm the normal (or inverted) neutrino mass ordering. If  $\sigma(m_i) = 0.005$ , the detection of the mass ordering will still be possible even if  $\sum m_\nu \lesssim 0.1$  eV, since the error on  $\sum m_\nu$  will allow to exclude the inverted ordering with great accuracy.

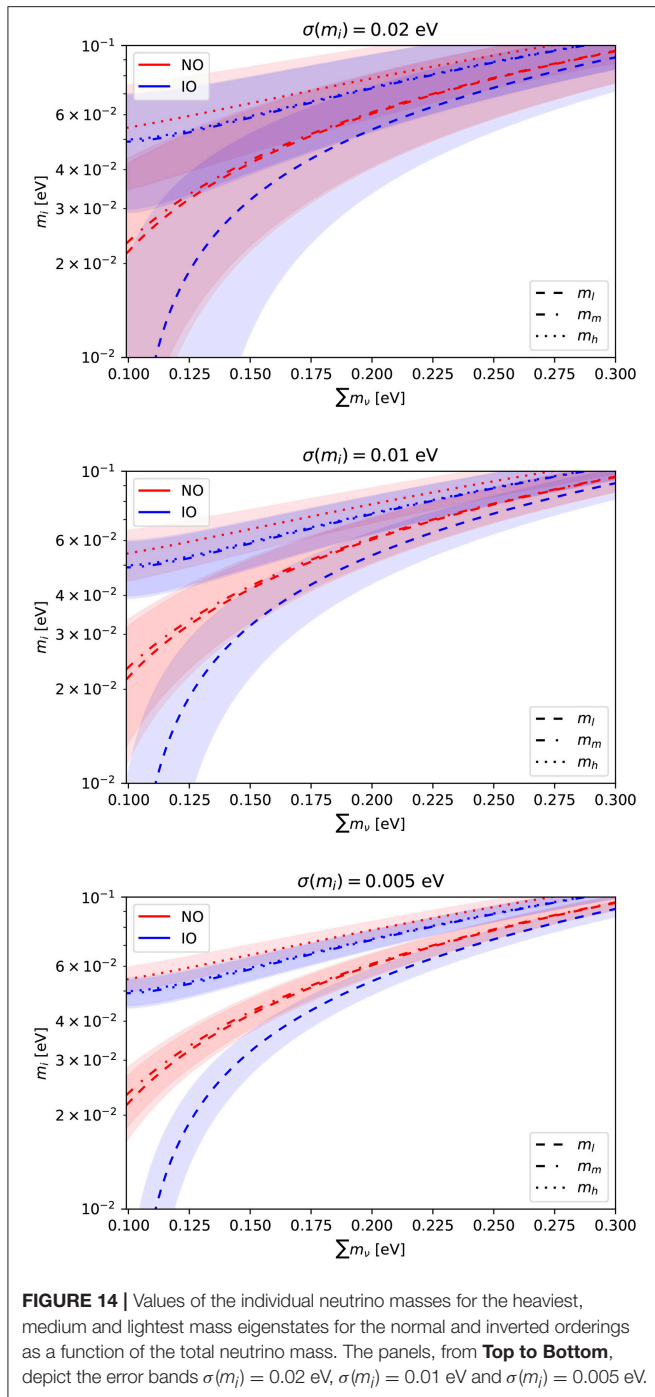
As already mentioned, another possibility is the so-called 21 cm intensity mapping, which will focus on low redshifts  $z \lesssim 3$  and will measure, with low angular resolution, the integrated 21 cm flux emitted from unresolved sources observing large patches of the sky. The lack of high angular resolution will result in a less precise measurement of non-linear scales. On the other hand, low angular resolution will imply a much faster survey. Future planned intensity mapping surveys are developed within the Phase 1 of the SKA experiment, which will include a wide and deep survey at low redshifts ( $z \lesssim 3$ , the SKA1-MID array) and a narrow and deep survey at higher redshift ( $3 \lesssim z \lesssim 6$ , the SKA1-LOW array), and within the Phase 2 of SKA (SKA2). Since, in some sense, these intensity mapping probes will be complementary to future planned optical surveys, as DESI or Euclid, it makes sense to combine their expected results. The intensity mapping technique, as galaxy clustering, is also affected by bias uncertainties and non-linearities at small scales.



**FIGURE 13 |** Differences in the masses for three neutrino mass eigenstates as a function of the total neutrino mass between inverted and normal orderings. The vertical dashed lines depict the value  $\sum m_\nu = 0.12$  eV and  $\sum m_\nu = 0.34$  eV, which are the present most constraining 95% CL limit on  $\sum m_\nu$  Palanque-Desabrouille et al. (2015) and the latest 95% CL bound quoted by the Planck collaboration (Aghanim et al., 2016b), respectively. Different shades of colored bands indicate the possible errors which could be achieved by future cosmological experiments on the determination of single neutrino masses: 0.02 eV, 0.01 eV or 0.005 eV.

Several studies have been carried out in the literature to unravel the perspectives of the intensity mapping technique in unveiling the neutrino properties. Some of them include the combination of the expectations from future large scale structure and intensity mapping surveys (Loeb and Wyithe, 2008; Visbal et al., 2009; Abazajian et al., 2011; Villaescusa-Navarro et al., 2015; Archidiacono et al., 2017; Sprenger et al., 2018). Notice that all these studies rely on different assumptions on the cosmological parameters, on the foregrounds and on the systematic uncertainties, therefore we can not do comparisons among them. Instead, we quote the most recent findings and the impact for an eventual future detection of the neutrino mass ordering.

The authors of Villaescusa-Navarro et al. (2015) found that, by combining SKA1-LOW with Planck measurements, the 95% CL error on  $\sum m_\nu$  could be  $\sim 0.089$  eV. It is remarkable that such a combination could potentially rule out the inverted ordering scenario, assuming that normal ordering is realized in nature. These authors also find that, under identical assumptions in the forecasted analyses, their combination of intensity mapping surveys (SKA1-LOW and MID) should be regarded as competitive with future spectroscopic surveys concerning neutrino mass properties. The authors of Archidiacono et al. (2017) showed that constraints of the future CORE CMB mission and galaxy redshift/weak lensing large scale structure surveys (as Euclid) on the neutrino mass can be improved if a prior on the reionization optical depth from 21 cm probes as HERA or SKA is also included. A prior of  $\sigma(\tau) = 0.001$  will reduce the freedom in the amplitude of the primordial power spectrum  $A_s$ , as CMB measurements mostly constrain the combination  $A_s \exp(-2\tau)$ , see section 4.1. Therefore, the direct correlation between  $\sum m_\nu$  and  $A_s$ , both modifying the amplitude of the



matter power spectrum (although the change induced by  $\sum m_\nu$  is, obviously, scale dependent), is largely affected by the presence of a precise determination of  $\tau$ . The  $1\sigma$  sensitivity they find for the combination of CORE, Euclid plus the prior on the optical depth from future 21 cm observations is  $\sigma(\sum m_\nu) = 0.012$  eV<sup>30</sup>.

<sup>30</sup>More recently, this very important synergy between Euclid and future 21 cm surveys, concretely with the intensity mapping survey SKA1, has been further assessed in Sprenger et al. (2018).

Nevertheless, as carefully detailed above, even if these tiny errors on  $\sum m_\nu$  will be reached and extrapolated to an error on the individual neutrino mass eigenstates, the possibility of extracting the neutrino mass ordering via singling out the neutrino mass eigenstates with cosmological observables remains unfeasible, unless very visionary scenarios, as FFTT under the most optimistic assumptions, are envisaged.

## 6.6. Prospects From Core-Collapse Supernova

Neutrinos from core-collapse supernovae offer an independent and complementary way to test neutrino physics. The existence of these neutrinos was robustly confirmed by the detection of 25 events from Supernova 1987A in the Large Magellanic Cloud (Alekseev et al., 1987; Bionta et al., 1987; Hirata et al., 1987), located at  $\sim 50$  kpc from our Milky Way galaxy. Such a detection allowed to set very compelling bounds on a number of neutrino properties (Schramm and Truran, 1990; Raffelt, 1999). Even if laboratory experiments have surpassed some of these limits, the eventual detection of supernovae neutrinos will still provide precious information about the details of the explosion process (see e.g., Janka, 2012; Mirizzi et al., 2016; Scholberg, 2018 and references therein), and also of neutrino mixing effects in dense media, see also Horiuchi and Kneller (2018).

Neutrino production in core-collapse supernovae occurs in a number of different stages. The first one is the *infall*, in which electron neutrinos are produced, confined, as a result of the process  $e^- + p \rightarrow n + \nu_e$ . When electrons are converted, the outwards pressure they generate disappears and the gravity forces are no more balanced: the core will start to collapse until its density reaches that of matter inside atomic nuclei, i.e., nuclear densities. Once these densities are reached, matter becomes incompressible, and a hydrodynamic shock is formed. As this shock wave propagates outwards, it heats up the nuclei and disintegrates them, releasing neutrinos. This initial neutrino release is commonly known as *neutronization burst*, and it is mainly composed of  $\nu_e$  and may last for a few tens of milliseconds. After the neutronization burst, the remnant proto-neutron star may evolve into a neutron star or collapse to a black hole, depending on the mass of the progenitor star. During this phase of *explosion and accretion*, which lasts for 1–2 s, the  $\nu_e$  contribution is still the dominant one, albeit there is also a contribution from other (anti)neutrino flavors, in particular  $\bar{\nu}_e$ . The neutrinos produced in the *cooling* stage give the main contribution to the total flux, as it is in this phase when the supernova releases its energy via all-flavor neutrino-antineutrino pair production, reaching its final cold state. This process lasts for about tens of seconds. The differences in the mean temperature of the neutrino fluxes of  $\nu_e$ ,  $\bar{\nu}_e$  and  $\nu_x$  ( $\bar{\nu}_x$ ) are due to the different medium opacity of each species. The larger the opacity, the lower the temperature that the (anti)neutrino will have at decoupling. The neutrino fluxes read as Scholberg (2018)

$$\phi(E_\nu) = N_0 \frac{(\alpha + 1)^{(\alpha+1)}}{\langle E_\nu \rangle \Gamma(\alpha + 1)} \left( \frac{E_\nu}{\langle E_\nu \rangle} \right)^\alpha \exp \left( -(\alpha + 1) \frac{E_\nu}{\langle E_\nu \rangle} \right), \quad (24)$$

where  $N_0$  is the total number of emitted neutrinos, and both  $\alpha$  and the mean energy  $\langle E_\nu \rangle$  are flavor dependent. The supernova neutrino energy spectra peaks around the 10 – 20 MeV region.

The most popular process for supernova neutrino detection is inverse beta decay on protons ( $\bar{\nu}_e + p \rightarrow n + e^+$ ). Other possibilities include elastic scattering on electrons ( $\nu + e^- \rightarrow \nu + e^-$ ), whose kinematics may provide information on the supernova location. Supernova neutrinos can also interact with nuclei via charged current or neutral current interactions, giving rise to charged leptons and/or excited nuclei which may provide flavor tagging. A very important process on argon nuclei is  $\nu_e + {}^{40}\text{Ar} \rightarrow e^- + {}^{40}\text{K}^*$ , which allows for electron neutrino tagging. In practice, water Cherenkov and scintillator detectors are mostly sensitive to electron antineutrinos via inverse beta decay, while the liquid argon technique mainly detects electron neutrinos. While other flavors may also be detected, the two processes above are the dominant ones. Large detector volumes (dozens of kilotons) are required to detect neutrinos from core-collapse supernovae located at  $\sim \mathcal{O}(10)$  kpc. A convenient way to scale the total number of supernova neutrino events in a detector of given effective mass is Beacom and Vogel (1998); Mena et al. (2007)

$$N = N_0 \left( \frac{E_B}{3 \times 10^{53} \text{ erg}} \right) \left( \frac{10 \text{ kpc}}{D_{\text{OS}}} \right)^2. \quad (25)$$

In the expression above,  $E_B$  is the gravitational binding energy of the collapsing star and  $D_{\text{OS}}$  the distance between the observer and the supernova. Assuming sensitivity to all reactions, the reference rate is  $N_0 = \mathcal{O}(10^4)$  for the Super-Kamiokande water Cherenkov detector with 32 kton and 5 MeV energy detection threshold. (Scholberg, 2012, 2018) give an estimate of the number of neutrino events for a number of ongoing and future facilities, based on different detection techniques: water Cherenkov (including also those with long string photosensors in ice, as Icecube and PINGU), liquid argon time projection chambers, and liquid scintillators. Upcoming neutrino detectors, already described in section 6.1 and crucial for oscillation physics measurements, such as the JUNO liquid scintillator (An et al., 2016), the liquid argon DUNE (Acciarri et al., 2015, 2016a,b; Strait et al., 2016) and the water Cherenkov Hyper-Kamiokande (Abe et al., 2015) can lead to a number of 6000, 3000 and 75000 supernova neutrino events respectively, assuming that the explosion occurs at 10 kpc from our position.

Flavor transitions inside a supernova have been carefully reviewed in Refs. Mirizzi et al. (2016); Scholberg (2018) (see also Lunardini and Smirnov, 2001a,b, 2003, 2004; Akhmedov et al., 2002). Here we summarize the most relevant results. As we have seen in section 2, when neutrinos propagate through matter their mixing effects undergo the so-called MSW mechanism, feeling a matter potential which is proportional to the electron number density  $N_e$ . If the supernova matter density has a profile which varies slowly, the neutrino matter eigenstates will propagate adiabatically and their final flavor composition will depend on the neutrino mass ordering, which will establish whether or not resonant transitions associated to each neutrino mass

squared difference (solar and atmospheric) take place<sup>31</sup>. In the normal ordering case, the neutrino fluxes will have a significantly transformed spectrum, while the electron antineutrino one will only be partially transformed ( $F_{\nu_e}^{\text{final}} = F_{\nu_x}^{\text{initial}}$  and  $F_{\bar{\nu}_e}^{\text{final}} = \cos^2 \theta_{12} F_{\bar{\nu}_e}^{\text{initial}} + \sin^2 \theta_{12} F_{\bar{\nu}_x}^{\text{initial}}$ ). In the inverted ordering case, the effects on the electron neutrino and antineutrino fluxes will be approximately the opposite ones ( $F_{\nu_e}^{\text{final}} = \sin^2 \theta_{12} F_{\nu_e}^{\text{initial}} + \cos^2 \theta_{12} F_{\nu_x}^{\text{initial}}$  and  $F_{\bar{\nu}_e}^{\text{final}} = F_{\bar{\nu}_x}^{\text{initial}}$ ). Once neutrinos exit from supernovae, they can still undergo flavor transitions if they traverse the Earth. Their final flavor composition at the detector location will again depend on the neutrino mass ordering, as matter effects in Earth depend on it (see e.g., Scholberg, 2018) and references therein.

Furthermore, *collective effects* from neutrino self-interactions, due to  $\nu_e + \bar{\nu}_e \rightarrow \nu_x + \bar{\nu}_x$  flavor processes, can lead to departures from the above summarized three-flavor oscillation picture (Hannestad et al., 2006a; Duan et al., 2007, 2010; Esteban-Pretel et al., 2007, 2008; Raffelt and Sigl, 2007; Mirizzi et al., 2016). The effective potential, proportional to the difference between the electron antineutrino and the muon/tau antineutrino fluxes, and inversely proportional to the supernova radius, should dominate over the standard matter one, leading to *spectral swaps or splits* (Raffelt and Smirnov, 2007a,b; Dasgupta et al., 2009, 2010). In the early stages, these self-interacting effects are sub-leading for mass ordering signatures, albeit we shall comment on possible non-thermal features in the neutrino or antineutrino spectra which depend on the mass ordering (Choubey et al., 2010).

In the following, we shall summarize the most relevant available methods to extract the neutrino mass ordering using the mentioned fluxes. For a recent and thorough review of the mass ordering signatures from supernovae neutrinos, we refer the reader to Scholberg (2018). The electron neutrinos produced in the neutronization burst undergo the MSW effect being fully (only partially) transformed, i.e.,  $F_{\nu_e}^{\text{final}} = F_{\nu_x}^{\text{initial}}$  ( $F_{\nu_e}^{\text{final}} = \sin^2 \theta_{12} F_{\nu_e}^{\text{initial}} + \cos^2 \theta_{12} F_{\nu_x}^{\text{initial}}$ ) if the mass ordering is normal (inverted), respectively. Therefore, detectors with good  $\nu_e$  tagging, such as liquid argon or water Cherenkov ones, will detect a neutronization burst only in the inverted neutrino mass ordering case. Concerning the accretion phase, and once electron antineutrinos are also produced, as they are almost unchanged in the MSW resonance, the largest signature is expected to occur for the normal ordering case for the three type of aforementioned detector types (liquid argon, water Cherenkov and scintillator), although the Icecube detector, with its excellent capabilities to reconstruct the time dependence of the signal, could also distinguish between the normal and inverted mass orderings (Ott et al., 2013). While a devoted study with precise and accurate mass ordering sensitivities attainable at these three detector types via supernova neutrinos is, to our knowledge, missing in the literature, we exploit the event rates during the accretion phase quoted for normal and inverted orderings in Scholberg (2018) for a supernova located at 10 kpc. For a 40 kton liquid argon detector,

<sup>31</sup>In case the matter potential inside the supernova suffers from discontinuities, the neutrino transitions will be non-adiabatic and the final flavor composition will depend on the precise matter profile.

374 kton water Cherenkov and 20 kton scintillator, the normal mass ordering could be extracted with  $\sim 2, 6$  and  $2\sigma$  significance, respectively, based on a pure statistical-error analysis.

On the other hand, collective effects, which lead to spectral swaps in the electron (anti)neutrino spectra, show very sharp features at fixed energy values which depend, among other factors, on the neutrino mass ordering. However, these signatures are not as robust as the ones existing in the neutronization and accretion phases. Finally, a very significant imprint of the neutrino mass ordering on the supernovae neutrino fluxes is that due to their propagation through the Earth interior, where the standard MSW effect will induce a few percent-level oscillatory pattern in the 10 – 60 MeV energy range, in the electron (anti)neutrino spectra in case of (normal) inverted mass ordering. The detection of these wiggles requires however excellent energy resolution.

## 6.7. Prospects From Relic Neutrino Direct Detection

In the early Universe, neutrinos decoupled from the cosmic plasma during the cool down, in a process similar to the one leading to the formation of the CMB but at an earlier time, when the universe was seconds to minutes old. These neutrinos have been free-streaming for such a long time that they have decohered and are currently propagating as mass eigenstates. The decoupling of neutrinos occurred just before  $e^\pm$  annihilated and reheated photons, leading to the following ratio between the photon ( $T_\gamma$ ) and neutrino ( $T_\nu$ ) temperatures, see Equation (15):

$$\frac{T_\nu}{T_\gamma} = \left(\frac{4}{11}\right)^{1/3}. \quad (26)$$

Today, the temperature of the neutrino background is  $T_\nu^0 \simeq 1.6 \times 10^{-4}$  eV. Their mean energy is  $\langle E_\nu \rangle \simeq 3 T_\nu \simeq 5 \times 10^{-4}$  eV, much smaller than the minimal mass of the second-to-lightest neutrino as required by flavor oscillations, so that at least two out of three neutrinos are non-relativistic today. The cosmic neutrino background (CνB) is the only known source of non-relativistic neutrinos and it has never been detected directly.

Apart from the imprints that relic neutrinos leave in the CMB (see section 4.1), which allow to have an indirect probe of their existence through the determination of  $N_{\text{eff}}$ , the direct detection of the CνB would offer a good opportunity to test neutrino masses and their ordering. Capturing relic neutrinos is not only rewarding from the point of view of what we can learn about neutrino properties, but also because it would be a further confirmation of the standard Big Bang cosmological model. Different ideas on how to achieve such a detection have been proposed (Weinberg, 1962; Weiler, 1982, 1984; Duda et al., 2001; Eberle et al., 2004; Barenboim et al., 2005; Gelmini, 2005; Ringwald, 2005; Cocco et al., 2007; Li, 2015; Vogel, 2015; Domcke and Spinrath, 2017), ranging from absorption dips in the ultra-high-energy (UHE) neutrino fluxes due to their annihilation with relic neutrinos at the  $Z$  boson resonance, to forces generated by coherent scattering of the relic bath on a pendulum and measured by laser interferometers. Most of these proposed methods are impractical from the experimental point

of view. The one exploiting UHE neutrinos Weiler (1982, 1984); Eberle et al. (2004); Barenboim et al. (2005) has two problems, one related with the fact that it is difficult to think about a source that produces such UHE neutrinos, of energies

$$E_\nu^{\text{res}} = \frac{m_Z^2}{2m_i} \simeq 4 \cdot 10^{22} \left(\frac{0.1 \text{ eV}}{m_i}\right) \text{ eV}, \quad (27)$$

and another one regarding the difficulties of detecting a large enough sample of UHE neutrinos in order to resolve the dips. The method based on interferometers (Domcke and Spinrath, 2017) is even more complicated to address. At interferometers, current sensitivities to accelerations are of the order of  $a \simeq 10^{-16} \text{ cm/s}^2$ , with an optimistic estimation of  $a \simeq 3 \cdot 10^{-18} \text{ cm/s}^2$  (Domcke and Spinrath, 2017) for the incoming generation. However, expected accelerations due to relic neutrino interactions are of the order of  $(10^{-27} - 10^{-33}) \text{ cm/s}^2$  (Duda et al., 2001; Domcke and Spinrath, 2017), many orders of magnitude below the sensitivity of the next-generation interferometers.

The most promising approach to detect relic neutrinos is to use neutrino capture in a  $\beta$ -decaying nucleus  $A$

$$\bar{\nu}_e^{(-)} + A \rightarrow e^\pm + A', \quad (28)$$

where the signal for a positive detection is a peak located about  $2m_\nu$  above the true  $\beta$ -decay endpoint (see below). In particular, tritium is considered as the best candidate since it has a high neutrino capture cross section, low  $Q$ -value and it is long-lived (Cocco et al., 2007; Blennow, 2008; Lazauskas et al., 2008; Faessler et al., 2011; Long et al., 2014). The proposal for an experiment chasing this purpose was made in Cocco et al. (2007). Currently, efforts are put for such experiment, the PonTecorvo Observatory for Light Early-Universe Massive-Neutrino Yield (PTOLEMY) (Betts et al., 2013; Baracchini et al., 2018), to be built. The experiment has recently been approved by the Scientific Committee of the Italian National Laboratories of Gran Sasso and, in the following months, the existing prototypes for various components are expected to be moved from Princeton, where the R&D has been performed up to now, to Gran Sasso. The idea is to implant the tritium source on graphene layers, to avoid the problems related to a gaseous source, then collect and measure the energy of the emitted electrons using a combination of MAC-E filter, radio-frequency tracking and micro-calorimetry to obtain a determination of the  $\beta$ -decay and neutrino capture spectrum of tritium with an energy resolution of the order  $\Delta \simeq 0.05 - 0.1$  eV.

The total expected event rate from relic neutrino capture for a PTOLEMY-like experiment, assuming the estimated tritium mass of 100 g, is

$$\Gamma_{\text{CνB}} = [n_0(v_{h_R}) + n_0(v_{h_L})] N_T \bar{\sigma} \sum_{i=1}^3 |U_{ei}|^2 f_c(m_i), \quad (29)$$

where  $n_0(v_{h_{R,L}})$  is the averaged number density of relic neutrinos with right (R) or left (L) helicity,  $N_T = M_T/m(^3\text{H})$  is the approximated number of tritium atoms in the source,  $\bar{\sigma} \simeq 3.834 \times 10^{-45} \text{ cm}^2$  (Long et al., 2014), and  $f_c(m_i)$  is a mass-dependent overdensity factor that accounts for the clustering of

relic neutrinos under the gravitational attraction of the matter potential (mostly from the dark matter halo) of our galaxy. This last factor was originally computed in Singh and Ma (2003) and Ringwald and Wong (2004) and later updated in de Salas et al. (2017) (see also Zhang and Zhang, 2018), where smaller masses were considered for the neutrinos, and the treatment of the matter potential of the Milky Way was improved. The values of  $f_c(m_\nu)$  range from 1.1 – 1.2 for a neutrino with  $m_\nu = 60$  meV to 1.7 – 2.9 for  $m_\nu = 150$  meV (de Salas et al., 2017).

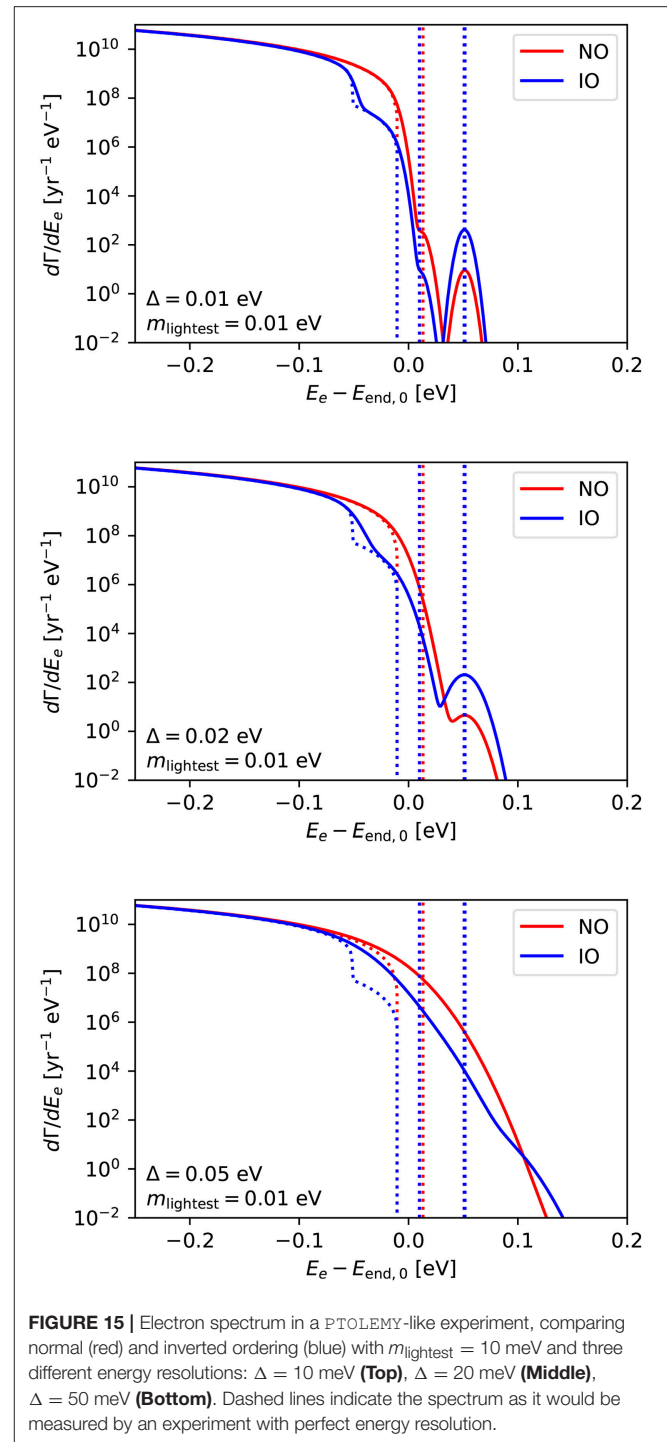
For unclustered neutrinos (i.e.,  $f_c = 1$ ) and 100 g of tritium, the expected number of events per year is Long et al. (2014)

$$\Gamma_{\text{CvB}}^{\text{D}} \simeq 4 \text{ year}^{-1}, \quad \Gamma_{\text{CvB}}^{\text{M}} = 2\Gamma_{\text{CvB}}^{\text{D}} \simeq 8 \text{ year}^{-1}, \quad (30)$$

where the upperscripts D and M stand for the possible Dirac and Majorana neutrino character. If neutrinos are Majorana particles, the expected number of events is doubled with respect to the Dirac case. The reason is related to the fact that, during the transition from ultra-relativistic to non-relativistic particles, helicity is conserved, but not chirality. The population of relic neutrinos is then composed by left- and right-helical neutrinos in the Majorana case, and only left-helical neutrinos in the Dirac case. Since the neutrino capture can only occur for left-chiral electron neutrinos, the fact that in the Majorana case the right-handed neutrinos can have a left-chiral component leads to a doubled number of possible interactions. While this means that in principle it is possible to distinguish the Dirac or Majorana neutrino nature with a precise determination of the event rate, there are two problems. First of all, even without assuming new physics, the factor of two coming from the neutrino nature is degenerate with the clustering factor, see Equation (29), so that a precise calculation of  $f_c$  is required to determine if neutrinos are Dirac or Majorana particles through the direct detection of relic neutrinos (de Salas et al., 2017). Moreover, non-standard interactions can increase the event rate in the Dirac case by a factor larger than two, canceling the difference with Majorana neutrinos in some scenarios (Arteaga et al., 2017).

Let us come back to the PTOLEMY proposal. Instead of considering the total event rate, for this kind of experiment it is much better to study the energy spectrum, as the direct detection of relic neutrinos can only be possible if one can distinguish the signal events due to neutrino capture from the background events due to the  $\beta$ -decay of tritium. A crucial issue for such an experiment, actually more important than the event rate, is therefore the energy resolution. In order to distinguish the peak due to the captured relic neutrinos from the  $\beta$ -decay background, a full-width half maximum (FWHM) energy resolution  $\Delta \lesssim 0.7 m_\nu$  is needed (Long et al., 2014). If neutrinos are non-degenerate in mass, the neutrino capture signal has a peak for each of the separate neutrino mass eigenstates. The full expression of the energy spectrum of neutrino capture, given an energy resolution  $\sigma = \Delta/\sqrt{8 \ln 2}$ , can be written as:

$$\frac{d\tilde{\Gamma}_{\text{CvB}}}{dE_e}(E_e) = \frac{1}{\sqrt{2\pi}\sigma} n_0 N_T \bar{\sigma} \sum_{i=1}^{N_\nu} |U_{ei}|^2 f_{c,i} \times \exp \left\{ -\frac{[E_e - (E_{\text{end}} + m_i + m_{\text{lightest}})]^2}{2\sigma^2} \right\}, \quad (31)$$



**FIGURE 15 |** Electron spectrum in a PTOLEMY-like experiment, comparing normal (red) and inverted ordering (blue) with  $m_{\text{lightest}} = 10$  meV and three different energy resolutions:  $\Delta = 10$  meV (**Top**),  $\Delta = 20$  meV (**Middle**),  $\Delta = 50$  meV (**Bottom**). Dashed lines indicate the spectrum as it would be measured by an experiment with perfect energy resolution.

where  $E_{\text{end}}$  is the energy of the  $\beta$ -decay endpoint,  $E_{\text{end}} = E_{\text{end},0} - m_{\text{lightest}}$ , being  $E_{\text{end},0}$  the endpoint energy when  $m_{\text{lightest}} = 0$ . If the energy resolution is good enough, the three peaks coming from the three neutrino mass eigenstates could be resolved, each of them with an expected number of events modulated by  $|U_{ei}|^2$ . This might lead to a positive detection of the neutrino mass ordering, since the electron-flavor component of  $\nu_1$  is

larger than the one of  $\nu_2$  and  $\nu_3$ , and therefore the furthest peak from the  $\beta$ -decay endpoint (again if neutrinos are non-degenerate) is enhanced if the ordering of neutrino masses is inverted. This can be seen in the three panels of **Figure 15**, which also show the effect of changing the mass ordering on the  $\beta$ -decay spectrum. Dashed lines represent the spectrum which would be determined by an experiment capable of measuring the  $\beta$  spectrum with zero energy uncertainty, while solid lines represent the shape of the spectrum that one would observe in a real experiment. We plot in red (blue) the spectrum obtained using normal (inverted) ordering, a FWHM resolution  $\Delta = 10$  meV (top),  $\Delta = 20$  meV (middle),  $\Delta = 50$  meV (bottom) and a lightest neutrino mass  $m_{\text{lightest}} = 10$  meV. As we can see from the figure, the kink commented in section 3 is clearly visible when one observes the huge number of events that come from the 100 g of decaying tritium with a sufficient energy resolution. While for distinguishing the relic neutrino events from the  $\beta$ -decay background and for having a direct detection of the C $\nu$ B the energy resolution is a crucial requirement, in principle even a worse energy resolution may allow to determine the neutrino mass scale and the mass ordering, thanks to the fact that we expect less events near the endpoint when the ordering is inverted. A direct observation of the amplitude of all the C $\nu$ B peaks, however, would give a much cleaner signal, because the peak corresponding to the heaviest neutrino would be always higher in the inverted ordering case, independently of any other factor.

In summary, the C $\nu$ B capture event rate in a PTOLEMY-like experiment (Equation (31)), even within SM physics and without considering non-standard interactions, depends on several main unknowns: *i*) the absolute neutrino mass, *ii*) the matter distribution (especially that of dark matter) in our galaxy, *iii*) the nature of neutrino masses (whether neutrinos are Dirac or Majorana particles), and *iv*) the true mass ordering. This last dependence is encoded in the  $|U_{ei}|^2$  factor in Equation (31) and it is only accessible if neutrinos are non-degenerate. A quantitative study on the PTOLEMY capabilities in determining the mass ordering has not been published yet, but a new Letter of Intent is in preparation (Baracchini et al., 2018)<sup>32</sup>.

## 7. SUMMARY

Identifying the neutrino mass ordering is one of the major pending issues to complete our knowledge of masses and mixings in the lepton sector. The two possibilities, normal vs. inverted, may result from very different underlying symmetries and therefore to single out the one realized in nature is a mandatory step to solve the flavor puzzle, i.e., to ensure a full theoretical understanding of the origin of particle masses and mixings. We have presented a comprehensive review on the current status and on future prospects of extracting the neutrino mass ordering via a number of different ongoing and upcoming observations. Furthermore, the most updated and complete result on the

preference for a given neutrino mass ordering from a Bayesian global fit to all 2018 publicly available neutrino data has also been presented.

Currently, among the three available methods to extract the neutrino mass ordering (oscillations, neutrinoless double beta decay searches and cosmological observations), the leading probe comes from oscillations in matter, measured at long-baseline accelerator or atmospheric neutrino beams in combination with reactor experiments. The latest frequentists global data analysis results in a preference for normal mass ordering with  $\Delta\chi^2 = 11.7$  ( $\sim 3.4\sigma$ ), mostly arising from the combination of the long-baseline T2K and NOvA data with reactor experiments (Daya Bay, RENO and Double Chooz), plus the latest atmospheric neutrino results from Super-Kamiokande. Similar results for the preference in favor of the normal mass ordering arise from other global fit analyses (Capozzi et al., 2018a).

Cosmological measurements are able to set indirect, albeit independent bounds on the neutrino mass ordering. Neutrinos affect Cosmic Microwave Background (CMB) primary anisotropies by changing the gravitational potential at the recombination period when they become non-relativistic. However, for sub-eV neutrino masses this effect is tiny and the most prominent effect on the CMB is via lensing, as neutrinos, having non-zero velocities, will reduce the lensing effect at small scales. Nevertheless, the largest impact of neutrinos in cosmology gets imprinted in the matter power spectrum. Once neutrinos become non-relativistic, their large velocity dispersions will prevent the clustering of matter inhomogeneities at all scales smaller than their free streaming length. At present, the cosmological constraints on the neutrino mass ordering come from the sensitivity to the total neutrino mass  $\sum m_\nu$  and not via the effects induced in the CMB and matter power spectrum by each of the individual neutrino masses  $m_i$ . Within the context of the minimal  $\Lambda$ CDM model with massive neutrinos, current cosmological probes cannot provide odds stronger than  $\sim 3:1$  in favor of normal ordering.

Neutrinoless double beta decay searches can also test the neutrino mass ordering if neutrinos are Majorana particles. However, present constraints on the so-called effective Majorana mass do not affect the overall Bayesian analyses.

All in all, the 2018 Bayesian global analysis, including all the neutrino oscillation data available before the Neutrino 2018 conference, results in a  $3.2\sigma$  preference for the normal neutrino mass ordering which, in Bayesian words, implies a *strong* preference for such a scenario. One can then combine the oscillation data with  $0\nu\beta\beta$  data from KamLAND-Zen, EXO-200 and Gerda and cosmological observations from Planck, SDSS BOSS, 6DF and SDSS DR7 MGS. Using this conservative cosmological data combination, the aforementioned preference becomes  $3.4\sigma$ , which raises to  $3.5\sigma$  if a prior on the Hubble parameter  $H_0$  from local measurements is considered in addition. This clearly states the current power of oscillation results when dealing with neutrino mass ordering extractions.

While in the very near perspective an improved sensitivity (i.e., above the  $3.5\sigma$  level) is expected mostly from more precise measurements of current long-baseline and atmospheric experiments, and, to a minor extent, from cosmological surveys

<sup>32</sup>We suggest the interested readers to look forward to the publication of this document, which will describe in more detail the physics reach and the technical characteristics of PTOLEMY.

(Planck, DES and eBOSS among others), there will be a number of planned experiments which will be crucial for extracting the neutrino mass ordering in the non-immediate future.

Of particular relevance are the upcoming neutrino oscillation facilities, as they will be able to measure the neutrino mass ordering with astonishing precision without relying on combinations of different data sets. Such is the case of the Deep Underground Neutrino Experiment (DUNE), that will be able to measure the neutrino mass ordering with a significance of  $5\sigma$  with 7 years of data. Atmospheric neutrino observatories as PINGU or ORCA will also mainly focus on the mass ordering measurement. Some of these future devices could also identify the neutrino mass ordering via the detection of matter effects in the neutrino fluxes emitted at the eventual explosion of a supernova in our galaxy or in its neighborhood. On the other hand, medium baseline reactor neutrino detectors such as JUNO or RENO will also be able to extract the neutrino mass ordering despite matter effects are negligible for these two experiments. They will focus instead on an extremely accurate measurement of the survival electron antineutrino probability.

Improved masses and detection techniques in neutrinoless double beta decay future searches could go down the 10 meV region in the effective Majorana mass  $m_{\beta\beta}$ , and they could be able to discard at some significance level the inverted mass ordering scenario, in the absence of a positive signal. These limits, however, will apply only in case neutrinos have a Majorana nature. Moreover, the determination of the neutrino mass ordering may be complicated by the presence of a light sterile neutrino at the eV scale, as currently suggested by the NEOS and DANSS results.

Concerning future cosmological projects, the combination of different probes will still be required. Near-future CMB and large scale structure surveys will only be sensitive to the neutrino mass ordering via their achieved error on  $\sum m_\nu$ . Furthermore, the accuracy in the extraction of the neutrino mass ordering will strongly depend on how far  $\sum m_\nu$  lies from the minimum allowed value from oscillation probes. The future CMB mission CORE plus the DESI galaxy survey could provide odds of 9:1 for normal neutrino mass ordering assuming  $\sum m_\nu = 0.056$  eV. Even if very futuristic surveys, based on the observation of the 21 cm redshifted line in neutral hydrogen, may be able to extract the individual values of the neutrino masses, their precision on the  $m_i$  values may not be enough to guarantee a direct determination of the neutrino mass ordering by these means, albeit they can achieve an accurate measurement of the ordering thanks to their unprecedented precision on  $\sum m_\nu$ .

Last, but not least, relic neutrino capture in tritium in a PTOLEMY-like experiment could also establish the neutrino mass ordering via an almost perfect energy reconstruction of the

$\beta$ -decay spectrum, ensured by the extremely large amount of tritium adopted. The detection is possible both from a kink in the  $\beta$ -decay spectrum which only appears if the ordering is inverted and from the peaks due to neutrino capture just above the endpoint.

All these future probes may either confirm or reject the current *strong* preference ( $3.5\sigma$ ) in favor of the normal neutrino mass ordering. Such a preference has kept gaining significance in the recent years, thanks to the fact that current neutrino oscillation experiments have enormously improved our knowledge of neutrino flavor physics.

## AUTHOR CONTRIBUTIONS

PFdS, CT, and MT have contributed mainly to review the current status of neutrino oscillations as well as to the summary of the prospects on the neutrino mass ordering from future neutrino oscillation experiments. SG has contributed to review beta and neutrinoless double beta decay current status and future prospects. SG has also led the 2018 Global Bayesian analysis results. PFdS and SG have also contributed to review the present and future cosmological bounds and have also reviewed the prospects from relic neutrino detection. OM has mainly contributed to review the cosmological bounds on the mass ordering and the perspectives from future 21 cm surveys and core-collapse supernovae.

## ACKNOWLEDGMENTS

We thank C. Giunti for providing us the two panels constituting **Figure 7** and M. Hirsch for his suggestions on how to improve the first version of the manuscript. Work supported by the Spanish grants FPA2015-68783-REDT, FPA2017-90566-REDC (Red Consolider MultiDark), FPA2017-85216-P, FPA2017-85985-P and SEV-2014-0398 (AEI/FEDER, UE, MINECO), and PROMETEOII/2014/050, PROMETEO/2018/165 and GV2016-142 (Generalitat Valenciana). SG receives support from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie individual grant agreement No. 796941. OM is also supported by the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreements No. 690575 and 674896 and acknowledges the hospitality of the Fermilab Theoretical Physics Department. PFdS and CAT are also supported by the MINECO fellowships FPU13/03729 and BES-2015-073593, respectively. MT acknowledges financial support from MINECO through the Ramón y Cajal contract RYC-2013-12438 as well as from the L'Oréal-UNESCO *For Women in Science* initiative.

## REFERENCES

- Aalbers, J., Agostinib, F., Alfonsic, M., Amaror, F. D., Amslerd, C., Aprilee, E., et al. (2016). DARWIN: towards the ultimate dark matter detector. *JCAP* 11:017. doi: 10.1088/1475-7516/2016/11/017
- Aalseth, C., Berkeley, U. C., and Richland, P. N. L. (2018). Search for Neutrinoless Double- $\beta$  Decay in  $^{76}\text{Ge}$  with the Majorana Demonstrator. *Phys. Rev. Lett.* 120:132502. doi: 10.1103/PhysRevLett.120.132502
- Aartsen, M., and Adelaide U. (2014). Letter of intent: the precision iceCube next generation upgrade (PINGU). *arXiv:1401.2046*. 136.

- Aartsen, M. G., Ackermann, M., Adams, J., Aguilar, J. A., Ahlers, M., Ahrens, M., et al. (2015). Determining neutrino oscillation parameters from atmospheric muon neutrino disappearance with three years of IceCube DeepCore data. *Phys. Rev. D* 91:072004. doi: 10.1103/PhysRevD.91.072004
- Abazajian, K., Calabrese, E., Cooray, A., De Bernardis, F., Dodelson, S., Friedland, A., et al. (2011). Cosmological and astrophysical neutrino mass measurements. *Astropart. Phys.* 35, 177–184. doi: 10.1016/j.astropartphys.2011.07.002
- Abazajian, K. N., Adshead, P., Ahmed, Z., Allen, S. W., Alonso, D., Arnold, K. S., et al. (2016). *CMB-S4 Science Book*, 1st Edn.
- Abazajian, K. N., and Dodelson, S. (2003). Neutrino mass and dark energy from weak lensing. *Phys. Rev. Lett.* 91:041301. doi: 10.1103/PhysRevLett.91.041301
- Abbott, T., Abdalla, F. B., Alarcon, A., Aleksic, J., Allam, S., Allen, S., et al. (2018). Dark energy survey year 1 results: cosmological constraints from galaxy clustering and weak lensing. *Phys. Rev. D* 98:043526. doi: 10.1103/PhysRevD.98.043526
- Abbott, T., Abdalla, F. B., Alarcon, A., Allam, S., Andrade-Oliveira, F., Annis, J., et al. (2017). Dark energy survey year 1 results: measurement of the baryon acoustic oscillation Scale in the distribution of galaxies to redshift 1. *Mon. Not. R. Astron. Soc.* D98:043526.
- Abdurashitov, J. N., Gavrin, V. N., Gorbachev, V. V., Gurkina, P. P., Ibragimova, T. V., Kalikhov, A. V. et al. (2009). Measurement of the solar neutrino capture rate with gallium metal. III: Results for the 2002–2007 data-taking period. *Phys. Rev. C* 80:015807. doi: 10.1103/PhysRevC.80.015807
- Abe, K., Abe, K., Ahn, S. H., Aihara, H., Aimi, A., Akutsu, R., et al. (2018c). Physics potentials with the second hyper-kamiokande detector in Korea. *PTEP* 2018:063C01. doi: 10.1093/ptep/pty044
- Abe, K., Abe, K., Aihara, H., Aimi, A., Akutsu, R., Andreopoulos, C., et al. (2018b). Hyper-kamiokande design report. *arXiv:1805.04163v1*.
- Abe, K., Abgrall, N., Ajima, Y., Aihara, H., Albert, J. B., Andreopoulos, C., et al. (2011a). Indication of electron neutrino appearance from an accelerator-produced Off-axis muon neutrino beam. *Phys. Rev. Lett.* 107:041801. doi: 10.1103/PhysRevLett.107.041801
- Abe, K., Aihara, H., Andreopoulos, C., Anghel, I., Ariga, A., Ariga, T., et al. (2015). Physics potential of a long-baseline neutrino oscillation experiment using a J-PARC neutrino beam and Hyper-Kamiokande. *Prog. Theor. Exp. Phys.* 2015:053C02. doi: 10.1093/ptep/ptv061
- Abe, K., Bronner, C., Haga, Y., Hayato, Y., Ikeda, M., Iyogi, K., et al. (2018a). Atmospheric neutrino oscillation analysis with external constraints in Super-Kamiokande I-IV. *Phys. Rev. D* 97:072001. doi: 10.1103/PhysRevD.97.072001
- Abe, K., Hayato, Y., Iida, T., Ikeda, M., Ishihara, C., Iyogi, K., et al. (2011b). Solar neutrino results in Super-Kamiokande-III. *Phys. Rev. D* 83:052010. doi: 10.1103/PhysRevD.83.052010
- Abe, Y., Asano, Y., Haba, N., and Yamada, T. (2017). Heavy neutrino mixing in the T2HK, the T2HKK and an extension of the T2HK with a detector at Oki Islands. *Eur. Phys. J. C* 77:851. doi: 10.1140/epjc/s10052-017-5294-7
- Abe, Y., dos Anjos, J. C., Barriere, J. C., Baussan, E., Bekman, I., Bergevin, M., et al. (2014). Improved measurements of the neutrino mixing angle  $\theta_{13}$  with the Double Chooz detector. *JHEP* 10:086. doi: 10.1007/JHEP10(2014)086
- Abell, P. A., Allison, J., Anderson, S. F., Andrew, J. R., Roger, J., Angel, P., et al. (2009). *LSST Science Book, Version 2.0*.
- Abgrall, N., Abramov, A., Abrosimov, N., Abt, I., Agostini, M., Agartioglu, M., et al. (2017). The large enriched germanium experiment for neutrinoless double beta decay (LEGEND). *AIP Conf. Proc.* 1894:020027. doi: 10.1063/1.5007652
- Abrahão, T., Minakata, H., Nunokawa, H., and Quiroga, A. A. (2015). Constraint on Neutrino Decay with Medium-Baseline Reactor Neutrino Oscillation Experiments. *J. High Energy Phys.* 11:001. doi: 10.1007/JHEP11(2015)001
- Acciarri, R., Acero, M. A., Adamowski, M., Adams, C., Adamson, P., Adhikari, S., et al. (2016a). Long-baseline neutrino facility (LBNF) and deep underground neutrino experiment (DUNE) Volume I. *arXiv:1601.05471v1*.
- Acciarri, R., Acero, M. A., Adamowski, M., Adams, C., Adamson, P., Adhikari, S., et al. (2015). Long-Baseline Neutrino Facility (LBNF) and Deep Underground Neutrino Experiment (DUNE): Conceptual Design Report, Volume 2: The Physics Program for DUNE at LBNF. *arXiv:1512.06148*.
- Acciarri, R., Acero, M. A., Adamowski, M., Adams, C., Adamson, P., Adhikari, S., et al. (2016b). Long-Baseline Neutrino Facility (LBNF) and Deep Underground Neutrino Experiment (DUNE) : Conceptual Design Report, Volume 4 The DUNE Detectors at LBNF. *arXiv:1601.02984*.
- Adam, A. (2018). “Recent results from MINOS and MINOS+,” in *XXVIII International Conference on Neutrino Physics and Astrophysics* (Heidelberg), Available online at: doi: 10.5281/zenodo.1286760
- Adam, R., Ade, P. A. R., Aghanim, N., Akrami, Y., Alves, M. I. R., Arnaud, M., et al. (2016). Planck 2015 results. I. Overview of products and scientific results. *Astron. Astrophys.* 594:A1. doi: 10.1051/0004-6361/201527101
- Adams, D., Adams, Q., Alduino, C., Alfonso, K., Avignone, F. T. III., Azzolini, O., et al. (2018). Update on the recent progress of the CUORE experiment. *arXiv:1808.10342v1*.
- Adamson, P., Anghel, I., Aurisano, A., Barr, G., Bishai, M., Blake, A., et al. (2014). Combined analysis of  $\nu_\mu$  disappearance and  $\nu_\mu \rightarrow \nu_e$  appearance in MINOS using accelerator and atmospheric neutrinos. *Phys. Rev. Lett.* 112:191801. doi: 10.1103/PhysRevLett.112.191801
- Ade, P., Aghanim, N., Arnaud, M., Ashdown, M., Aumont, J., Baccigalupi, C., et al. (2016a). Planck 2015 results. XXVII. The Second Planck Catalogue of Sunyaev-Zeldovich Sources. *Astron. Astrophys.* 594:A27. doi: 10.1051/0004-6361/201525823
- Ade, P. A. R., Aghanim, N., Armitage-Caplan, C., Arnaud, M., Ashdown, M., Atrio-Barandela, F., et al. (2014). Planck 2013 results. XVI. Cosmological parameters. *Astron. Astrophys.* 571:A16. doi: 10.1051/0004-6361/201321591
- Ade, P. A. R., Aghanim, N., Arnaud, M., Ashdown, M., Aumont, J., Baccigalupi, C., et al. (2016b). Planck 2015 results. XIII. Cosmological parameters. *Astron. Astrophys.* 594:A13. doi: 10.1051/0004-6361/201525830
- Adrian-Martinez, S., Ageron, M., Aharonian, F., Aiello, S., Albert, A., Ameli, F., et al. (2016). Letter of intent for KM3NeT 2.0. *J. Phys. G* 43, 084001. doi: 10.1088/0954-3889/43/8/084001
- Adrian-Martinez, S., Al Samarai, I., Albert, A., Andre, M., Anghinolfi, M., Anton, G., et al. (2012). Measurement of atmospheric neutrino oscillations with the ANTARES neutrino telescope. *Phys. Lett. B* 714, 224–230. doi: 10.1016/j.physletb.2012.07.002
- Afshordi, N., Zaldarriaga, M., and Kohri, K. (2005). On the stability of dark energy with mass-varying neutrinos. *Phys. Rev. D* 72:065024. doi: 10.1103/PhysRevD.72.065024
- Agarwalla, S. K., Chatterjee, S. S., and Palazzo, A. (2016). Physics Reach of DUNE with a Light Sterile Neutrino. *J. High Energy Phys.* 09:016. doi: 10.1007/JHEP09(2016)016
- Agarwalla, S. K., Chatterjee, S. S., and Palazzo, A. (2018a). Signatures of a Light Sterile Neutrino in T2HK. *J. High Energy Phys.* 04:091. doi: 10.1007/JHEP04(2018)091
- Agarwalla, S. K., Chatterjee, S. S., Petcov, S., and Titov, A. (2018b). Addressing Neutrino Mixing Models with DUNE and T2HK. *Eur. Phys. J. C* 78:286. doi: 10.1140/epjc/s10052-018-5772-6
- Aghamousa, A., Aguilar, J., Ahlen, S., Alam, S., Allen, L. E., Prieto, C. A., et al. (2016). The DESI experiment part I: Science, targeting, and survey design. *arXiv:1611.00036v2*.
- Aghanim, N., Arnaud, M., Ashdown, M., Aumont, J., Baccigalupi, C., Banday, A. J., et al. (2016a). Planck 2015 results. XI. CMB power spectra, likelihoods, and robustness of parameters. *Astron. Astrophys.* 594:A11. doi: 10.1051/0004-6361/201526926
- Aghanim, N., Ashdown, M., Aumont, J., Baccigalupi, C., Ballardini, M., Banday, A. J., et al. (2016b). Planck intermediate results. XLVI. Reduction of large-scale systematic effects in HFI polarization maps and estimation of the reionization optical depth. *Astron. Astrophys.* 596:A107. doi: 10.1051/0004-6361/201628890
- Aghanim, N., Majumdar, S., and Silk, J. (2008). Secondary anisotropies of the CMB. *Rept. Prog. Phys.* 71:066902. doi: 10.1088/0034-4885/71/6/066902
- Agostini, M., Allardt, M., Bakalyarov, A. M., Balata, M., Barabanov, I., Baudis, L., et al. (2017b). Background-free search for neutrinoless double- $\beta$  decay of  $^{76}\text{Ge}$  with GERDA. *Nature* 544, 47–52. doi: 10.1038/nature21717
- Agostini, M., Bakalyarov, A. M., Balata, M., Barabanov, I., Baudis, L., Bauer, C., et al. (2018). Improved Limit on Neutrinoless Double- $\beta$  Decay of  $^{76}\text{Ge}$  from GERDA Phase II. *Phys. Rev. Lett.* 120:132503. doi: 10.1103/PhysRevLett.120.132503

- Agostini, M., Benato, G., and Detwiler, J. (2017a). Discovery probability of next-generation neutrinoless double- $\beta$  decay experiments. *Phys. Rev. D* 96:053001. doi: 10.1103/PhysRevD.96.053001
- Aharmim, B., Ahmed, S. N., Amsbaugh, J. F., Anthony, A. E., Banar, J., Barros, N., et al. (2008). An independent measurement of the total active B-8 solar neutrino flux using an array of He-3 proportional counters at the sudbury neutrino observatory. *Phys. Rev. Lett.* 101:111301. doi: 10.1103/PhysRevLett.101.111301
- Aharmim, B., Ahmed, S. N., Anthony, A. E., Barros, N., Beier, E. W., Bellerive, A., et al. (2010). Low energy threshold analysis of the Phase I and Phase II data sets of the sudbury neutrino observatory. *Phys. Rev. C* 81:055504. doi: 10.1103/PhysRevC.81.055504
- Ahmad, Q. R., Allen, R. C., Andersen, T. C., Anglin, J. D., Bühler, G., Barton, J. C., et al. (2001). Measurement of the rate of  $\nu_e + d \rightarrow p + p + e^-$  interactions produced by  $^8\text{B}$  solar neutrinos at the Sudbury Neutrino Observatory. *Phys. Rev. Lett.* 87:071301. doi: 10.1103/PhysRevLett.87.071301
- Ahmad, Q. R., Allen, R. C., Andersen, T. C., Anglin, J. D., Bühler, G., Barton, J. C., et al. (2002). Direct evidence for neutrino flavor transformation from neutral current interactions in the Sudbury Neutrino Observatory. *Phys. Rev. Lett.* 89:011301. doi: 10.1103/PhysRevLett.89.011301
- Ahmed, S., Sajjad Athar, M., Hasan, R., Salim, M., Singh, S. K., Inbanathan, S. S. R., et al. (2017). Physics Potential of the ICAL detector at the India-based Neutrino Observatory (INO). *Pramana* 88:79. doi: 10.1007/s12043-017-1373-4
- Ahmed, Z., Amiri, M., Benton, S. J., Bock, J. J., Bowens-Rubin, R., Buder, I., et al. (2014). BICEP3: a 95GHz refracting telescope for degree-scale CMB polarization. 9153, 91531N. doi: 10.1117/12.2057224
- Ahn, M. H., Aliu, E., Andringa, S., Aoki, S., Aoyama, Y., Argyriades, J., et al. (2006). Measurement of neutrino oscillation by the K2K experiment. *Phys. Rev. D* 74:072003. doi: 10.1103/PhysRevD.74.072003
- Akhmedov, E. K., Johansson, R., Lindner, M., Ohlsson, T., and Schwetz, T. (2004). Series expansions for three flavor neutrino oscillation probabilities in matter. *J. High Energy Phys.* 04:078. doi: 10.1088/1126-6708/2004/04/078
- Akhmedov, E. K., Lunardini, C., and Smirnov, A. Y. (2002). Supernova neutrinos: difference of muon-neutrino - tau-neutrino fluxes and conversion effects. *Nucl. Phys.* B643, 339–366. doi: 10.1016/S0550-3213(02)00692-2
- Alam, S., Ata, M., Bailey, S., Beutler, F., Bizyaev, D., Blazek, J. A., et al. (2017). The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: cosmological analysis of the DR12 galaxy sample. *Mon. Not. Roy. Astron. Soc.* 470, 2617–2652. doi: 10.1093/mnras/stx721
- Albert, J. B., Anton, G., Arnquist, I. J., Badhrees, I., Barbeau, P., Beck, D., et al. (2018). Sensitivity and discovery potential of nEXO to neutrinoless double beta decay. *Phys. Rev. C* 97:065503. doi: 10.1103/PhysRevC.97.065503
- Albert, J. B., Auty, D. J., Barbeau, P. S., Beauchamp, E., Beck, D., Belov, V., Benitez-Medina, C., et al. (2014). Search for Majorana neutrinos with the first two years of EXO-200 data. *Nature* 510, 229–234. doi: 10.1038/nature13432
- Alduino, C., Alfonso, K., Andreotti, E., Arnaboldi, C., Avignone, F. T. III., Azzolini, O., et al. (2018a). First results from CUORE: a search for lepton number violation via  $0\nu\beta\beta$  Decay of  $^{130}\text{Te}$ . *Phys. Rev. Lett.* 120:132501. doi: 10.1103/PhysRevLett.120.132501
- Alduino, C., Alfonso, K., Artusa, D. R., Avignone, F. T. III., Azzolini, O., Balata, M., et al. (2016). CUORE-0 detector: design, construction and operation. *JINST* 11:P07009. doi: 10.1088/1748-0221/11/07/P07009
- Alduino, C., Alfonso, K., Avignone, F. T. III., Azzolini, O., Bari, G., Bellini, F., et al. (2018b). Study of rare nuclear processes with CUORE. *Int. J. Mod. Phys. A* 33:1843002. doi: 10.1142/S0217751X18430029
- Alekseev, E., Alekseeva, L., Volchenko, V., and Krivosheina, I. (1987). Possible detection of a neutrino signal on 23 february 1987 at the baksan undergroundscintillation telescope of the institute of nuclear research. *JETP Lett.* 45, 589–592.
- Alekseev, I., Belov, V., Brudanin, V., Danilov, M., Egorov, V., Filosofov, D., et al. (2018). Search for sterile neutrinos at the DANSS experiment. *High Energy Phys. Exp. arXiv:1804.04046v2*.
- Alenkov, V., Aryal, P., Beyer, J., Boiko, R. S., Boonin, K., Buzanov, O., et al. (2015). Technical design report for the AMoRE  $0\nu\beta\beta$  decay search experiment. *arXiv:1512.05957*.
- Ali, Z. S., Parsons, A. R., Zheng, H., Pober, J. C., Liu, A., Aguirre, J. E., et al. (2015). PAPER-64 constraints on reionization: the 21cm power spectrum at  $z = 8.4$ . *Astrophys. J.* 809:61. doi: 10.1088/0004-637X/809/1/61
- Ali-Haïmoud, Y., and Bird, S. (2012). An efficient implementation of massive neutrinos in non-linear structure formation simulations. *Mon. Not. R. Astron. Soc.* 428, 3375–3389. doi: 10.1093/mnras/sts286
- Allen, S., Schmidt, R., and Bridle, S. (2003). A Preference for a non-zero neutrino mass from cosmological data. *Mon. Not. Roy. Astron. Soc.*, 346:593.
- Allen, S. W., Evrard, A. E., and Mantz, A. B. (2011). Cosmological parameters from observations of galaxy clusters. *Ann. Rev. Astron. Astrophys.* 49, 409–470. doi: 10.1146/annurev-astro-081710-102514
- Allison, R., Caucal, P., Calabrese, E., Dunkley, J., and Louis, T. (2015). Towards a cosmological neutrino mass detection. *Phys. Rev. D*, 92(12):123535.
- Alonso, J., Barros, N., Bergevin, M., Bernstein, A., Bignell, L., Blucher, E., et al. (2014). Advanced scintillator detector concept (ASDC): a Concept Paper on the physics potential of water-based liquid scintillator. *e-Print: arXiv:1409.5864*.
- Alpert, B., Balata, M., Bennett, D., Biasotti, M., Boragno, C., Brofferio, C., et al. (2015). HOLMES-The electron capture decay of  $^{163}\text{Ho}$  to measure the electron neutrino mass with Sub-eV sensitivity. *Eur. Phys. J. C* 75:112. doi: 10.22323/1.244.0066
- Amendola, L., Appleby, S., Bacon, D., Baker, T., Baldi, M., Bartolo, N., et al. (2018). Cosmology and fundamental physics with the euclid satellite. *Living Rev. Rel.* 21:2. doi: 10.1007/s41114-017-0010-3
- An, F., An, G., An, Q., Antonelli, V., Baussan, E., Beacom, J., et al. (2016). Neutrino physics with JUNO. *J. Phys. G* 43:030401.
- An, F. P., Bai, J. Z., Balantekin, A. B., Band, H. R., Beavis, D., Beriguete, W., et al. (2012). Observation of electron-antineutrino disappearance at Daya Bay. *Phys. Rev. Lett.* 108:171803. doi: 10.1103/PhysRevLett.108.171803
- An, F. P., Balantekin, A. B., Band, H. R., Bishai, M., Blyth, S., Cao, D., et al. (2017). Measurement of electron antineutrino oscillation based on 1230 days of operation of the Daya Bay experiment. *Phys. Rev. D* 95:072006. doi: 10.1103/PhysRevD.95.072006
- Ananthakrishnan, S. (1995). The giant meterwave radio telescope/GMRT. *J. Astrophys. Astr. Suppl.* 16:427.
- Anderson, L., Aubourg, E., Bailey, S., Beutler, F., Bhardwaj, V., Blanton, M., et al. (2014). The clustering of galaxies in the SDSS-III baryon oscillation spectroscopic survey: baryon acoustic oscillations in the Data Releases 10 and 11 Galaxy samples. *Mon. Not. R. Astron. Soc.*, 441, 24–62. doi: 10.1093/mnras/stu523
- Anderson, L., Aubourg, E., Bailey, S., Bizyaev, D., Blanton, M., Bolton, A. S., et al. (2013). The clustering of galaxies in the SDSS-III baryon oscillation spectroscopic survey: baryon acoustic oscillations in the data release 9 spectroscopic galaxy sample. *Mon. Not. R. Astron. Soc.* 427, 3435–3467. doi: 10.1111/j.1365-2966.2012.22066.x
- Andringa, S., Arushanova, E., Asahi, S., Askins, M., Auty, D. J., Back, A. R., et al. (2016). Current status and future prospects of the SNO+ experiment. *Adv. High Energy Phys.* 2016:6194250. doi: 10.1155/2016/6194250
- Angrik, J., Armbrust, T., Beglarian, A., Besserer, U., Blumer, J., Bonn, J., et al. *KATRIN Design Report* (2004). Report NPI ASCR Řež, EXP-01/2005.
- Archidiacono, M., Brinckmann, T., Lesgourgues, J., and Poulin, V. (2017). Physical effects involved in the measurements of neutrino masses with future cosmological data. *JCAP* 02:052. doi: 10.1088/1475-7516/2017/02/052
- Archidiacono, M., Giusarma, E., Hannestad, S., and Mena, O. (2013a). Cosmic dark radiation and neutrinos. *Adv. High Energy Phys.*, 2013:191047.
- Archidiacono, M., Giusarma, E., Melchiorri, A., and Mena, O. (2013b). Neutrino and dark radiation properties in light of recent CMB observations. *Phys. Rev. D*, 87(10):103519.
- Archidiacono, M., and Hannestad, S. (2014). Updated constraints on non-standard neutrino interactions from Planck. *JCAP* 07:046. doi: 10.1088/1475-7516/2014/07/046
- Archidiacono, M., and Hannestad, S. (2016). Efficient calculation of cosmological neutrino clustering in the non-linear regime. *JCAP* 06:018. doi: 10.1088/1475-7516/2016/06/018
- Archidiacono, M., Hannestad, S., Mirizzi, A., Raffelt, G., and Wong, Y. Y. (2013c). Axion hot dark matter bounds after Planck. *JCAP* 10:020. doi: 10.1088/1475-7516/2013/10/020
- Arnold, R., Augier, C., Baker, J., Barabash, A. S., Basharina-Freshville, A., Bongrand, M., et al. (2010). Probing new physics models of neutrinoless double beta decay with superNEMO. *Eur. Phys. J. C* 70, 927–943. doi: 10.1140/epjc/s10052-010-1481-5

- Arnold, R., Augier, C., Barabash, A. S., Basharina-Freshville, A., Blondel, S., Blot, S., et al. (2018). Final results on  $^{82}\text{Se}$  double beta decay to the ground state of  $^{82}\text{Kr}$  from the NEMO-3 experiment. *e-Print: arXiv:1806.05553*.
- Arteaga, M., Bertuzzo, E., Perez-Gonzalez, Y. F., and Zukanovich Funchal, R. (2017). Impact of beyond the standard model physics in the detection of the cosmic neutrino background. *J. High Energy Phys.* 09:124. doi: 10.1007/JHEP09(2017)124
- Artusa, D., Artusa, D. R., Avignone, F. T. III., Azzolini, O., Balata, M., Banks, T. I., et al. (2015). Searching for neutrinoless double-beta decay of  $^{130}\text{Te}$  with CUORE. *Adv. High Energy Phys.* 2015:879871. doi: 10.1103/PhysRevLett.115.102502
- Ascencio-Sosa, M., Calatayud-Cadenillas, A., Gago, A., and Jones-Pérez, J. (2018). Matter effects in neutrino visible decay at future long-baseline experiments. *J. High Energy Phys. arXiv:1805.03279*.
- Asner, D. M., Bradley, S., Bautista, J. E., Beutler, F., Bhardwaj, V., Bizyaev, D., et al. (2015). Single electron detection and spectroscopy via relativistic cyclotron radiation. *Phys. Rev. Lett.* 114:162501. doi: 10.1103/PhysRevLett.114.162501
- Aubourg, E., Bailey, S., Bautista, J. E., Beutler, F., Bhardwaj, V., Bizyaev, D., et al. (2015). Cosmological implications of baryon acoustic oscillation measurements. *Phys. Rev. D* 92:123516. doi: 10.1103/PhysRevD.92.123516
- Audren, B., Lesgourgues, J., Bird, S., Haehnelt, M. G., and Viel, M. (2013). Neutrino masses and cosmological parameters from a Euclid-like survey: markov chain monte carlo forecasts including theoretical errors. *JCAP* 01:026. doi: 10.1088/1475-7516/2013/01/026
- Azzolini, O., Barrera, M. T., Beeman, J. W., Bellini, F., Beretta, M., Biassoni, M., et al. (2018). First result on the neutrinoless double beta decay of  $^{82}\text{Se}$  with CUPID-0. *Phys. Rev. Lett.* 120:232502. doi: 10.1103/PhysRevLett.120.232502
- Bakhti, P., and Farzan, Y. (2014). Shedding light on LMA-Dark solar neutrino solution by medium baseline reactor experiments: JUNO and RENO-50. *J. High Energy Phys.* 07:064. doi: 10.1007/JHEP07(2014)064
- Bakhti, P., and Khan, A. N. (2017). Sensitivities to charged-current nonstandard neutrino interactions at DUNE. *J. Phys. G* 44:125001. doi: 10.1088/1361-6471/aa9098
- Baldi, M., Villaescusa-Navarro, F., Viel, M., Puchwein, E., Springel, V., and Moscardini, L. (2014). Cosmic degeneracies – I. Joint N-body simulations of modified gravity and massive neutrinos. *Mon. Not. R. Astron. Soc.* 440, 75–88. doi: 10.1093/mnras/stu259
- Balheiro Gomes, G., Forero, D., Guzzo, M., De Holanda, P., and Oliveira, R. N. (2018). Quantum decoherence effects in neutrino oscillations at DUNE. *High Energy Phys. arXiv:1805.09818*.
- Ballett, P., King, S. F., Pascoli, S., Prouse, N. W., and Wang, T. (2017). Sensitivities and synergies of DUNE and T2HK. *Phys. Rev. D* 96:033003. doi: 10.1103/PhysRevD.96.033003
- Banerjee, A., and Dalal, N. (2016). Simulating nonlinear cosmological structure formation with massive neutrinos. *JCAP* 11:015. doi: 10.1088/1475-7516/2016/11/015
- Banerjee, A., Powell, D., Abel, T., and Villaescusa-Navarro, F. (2018). Reducing noise in cosmological N-body simulations with neutrinos. *JCAP arXiv:1801.03906*. doi: 10.1088/1475-7516/2018/09/028
- Baracchini, E. et al. (2018). PTOLEMY: a proposal for thermal relic detection of massive neutrinos and directional detection of MeV dark matter. *Cosmol. Nongalact. Astrophys. arXiv:1808.01892*.
- Barenboim, G., Masud, M., Ternes, C. A., and Tórtola, M. (2018a). Exploring the intrinsic Lorentz-violating parameters at DUNE. *High Energy Phys. arXiv:1805.11094*.
- Barenboim, G., Mena Quejedo, O., and Quigg, C. (2005). Diagnostic potential of cosmic-neutrino absorption spectroscopy. *Phys. Rev. D* 71:083002. doi: 10.1103/PhysRevD.71.083002
- Barenboim, G., Ternes, C. A., and Tórtola, M. (2018b). Neutrinos, DUNE and the world best bound on CPT invariance. *Phys. Lett. B* 780, 631–637. doi: 10.1016/j.physletb.2018.03.060
- Barenboim, G., Ternes, C. A., and Tórtola, M. (2018c). New physics vs new paradigms: distinguishing CPT violation from NSI. *High Energy Phys. arXiv:1804.05842*.
- Barger, V., Marfatia, D., and Tregre, A. (2004). Neutrino mass limits from SDSS, 2dFGRS and WMAP. *Phys. Lett. B* 595, 55–59. doi: 10.1016/j.physletb.2004.06.049
- Barkana, R. (2018). Possible interaction between baryons and dark-matter particles revealed by the first stars. *Nature* 555, 71–74. doi: 10.1038/nature25791
- Barkana, R., and Loeb, A. (2005a). A Method for separating the physics from the astrophysics of high-redshift 21 cm fluctuations. *Astrophys. J.* 624, L65–L68. doi: 10.1086/430599
- Barkana, R., and Loeb, A. (2005b). Probing the epoch of early baryonic infall through 21cm fluctuations. *Mon. Not. R. Astron. Soc.* 363, L36–L40. doi: 10.1111/j.1745-3933.2005.00079.x
- Barkana, R., Outmezguine, N. J., Redigolo, D., and Volansky, T. (2018). Signs of Dark Matter at 21-cm? *Cosmol. Nongalact. Astrophys. arXiv:1803.03091*.
- Barreira, A., Li, B., Baugh, C., and Pascoli, S. (2014). Modified gravity with massive neutrinos as a testable alternative cosmological model. *Phys. Rev. D* 90:023528. doi: 10.1103/PhysRevD.90.023528
- Basse, T., Bjælde, O. E., Hamann, J., Hannestad, S., and Wong, Y. Y. (2014). Dark energy properties from large future galaxy surveys. *JCAP* 05:021. doi: 10.1088/1475-7516/2014/05/021
- Bautista, J. E., Busca, N. G., Guy, J., Rich, J., Blomqvist, M., Bourboux, H. d. M. d., et al. (2017a). Measurement of BAO correlations at  $z = 2.3$  with SDSS DR12 Ly $\alpha$ -Forests. *Astron. Astrophys.* 603:A12. doi: 10.1051/0004-6361/201730533
- Bautista, J. E., Vargas-Magaña, M., Dawson, K. S., Percival, W. J., Brinkmann, J., Brownstein, J., et al. (2017b). The SDSS-IV extended baryon oscillation spectroscopic survey: baryon acoustic oscillations at redshift of 0.72 with the DR14 luminous red galaxy sample. *Astrophys. J.* 863:110. doi: 10.3847/1538-4357/aace5
- Beacom, J. F., Bell, N. F., and Dodelson, S. (2004). Neutrinoless universe. *Phys. Rev. Lett.* 93:121302. doi: 10.1103/PhysRevLett.93.121302
- Beacom, J. F., and Vogel, P. (1998). Mass signature of supernova muon-neutrino and tau-neutrino neutrinos in Super-Kamiokande. *Phys. Rev. D* 58:053010. doi: 10.1103/PhysRevD.58.053010
- Beardsley, A. P., Morales, M. F., Lidz, A., Malloy, M., and Sutter, P. M. (2015). Adding context to james webb space telescope surveys with current and future 21cm radio observations. *Astrophys. J.* 800:128. doi: 10.1088/0004-637X/800/2/128
- Behara, S. P., Ghosh, A., Choubey, S., Datar, V. M., Mishra, D. K., and Mohanty, A. K. (2017). Search for the sterile neutrino mixing with the ICAL detector at INO. *Eur. Phys. J. C* 77:307. doi: 10.1140/epjc/s10052-017-4851-4
- Bell, N. F., Pierpaoli, E., and Sigurdson, K. (2006). Cosmological signatures of interacting neutrinos. *Phys. Rev. D* 73:063523. doi: 10.1103/PhysRevD.73.063523
- Bellini, G., Benziger, J., Bick, D., Bonfini, G., Bravo, D., Avanzini, M. B., et al. (2014). Final results of Borexino Phase-I on low energy solar neutrino spectroscopy. *Phys. Rev. D* 89:112007. doi: 10.1103/PhysRevD.89.112007
- Bellomo, N., Bellini, E., Hu, B., Jimenez, R., Peña-Garay, C., and Verde, L. (2017). Hiding neutrino mass in modified gravity cosmologies. *JCAP* 02:043. doi: 10.1088/1475-7516/2017/02/043
- Benson, B., Ade, P. A. R., Ahmed, Z., Allen, S. W., Arnold, K., Austermann, J. E., et al. (2014). “SPT-3G: a next-generation cosmic microwave background polarization experiment on the south pole telescope,” *Proceedings Volume 9153, Millimeter, Submillimeter, and Far-Infrared Detectors and Instrumentation for Astronomy VII* (Montréal, QC).
- Berlin, A., Hooper, D., Krnjaic, G., and McDermott, S. D. (2018). Severely constraining dark matter interpretations of the 21-cm anomaly. *Phys. Rev. Lett.* 121:011102. doi: 10.1103/PhysRevLett.121.011102
- Berryman, J. M., de Gouvea, A., Kelly, K. J., and Kobach, A. (2015). Sterile neutrino at the deep underground neutrino experiment. *Phys. Rev. D* 92:073012. doi: 10.1103/PhysRevD.92.073012
- Berryman, J. M., de Gouvea, A., Kelly, K. J., Peres, O. L. G., and Tabrizi, Z. (2016). Large, extra dimensions at the deep underground neutrino experiment. *Phys. Rev. D* 94:033006. doi: 10.1103/PhysRevD.94.033006
- Betoule, M., Kessler, R., Guy, J., Mosher, J., Hardin, D., Biswas, R., et al. (2014). Improved cosmological constraints from a joint analysis of the SDSS-II and SNLS supernova samples. *Astron. Astrophys.* 568:A22. doi: 10.1051/0004-6361/201423413

- Betoule, M., Marriner, J., Regnault, N., Cuillandre, J.-C., Astier, P., Guy, J., et al. (2013). Improved photometric calibration of the SNLS and the SDSS supernova surveys. *Astron. Astrophys.* 552:A124. doi: 10.1051/0004-6361/201220610
- Betts, S., Blanchard, W. R., Carnevale, R. H., Chang, C., Chen, C., Chidzik, S., et al. (2013). "Development of a Relic Neutrino Detection Experiment at PTOLEMY: Princeton Tritium Observatory for Light, Early-Universe, Massive-Neutrino Yield." in *Proceedings, Community Summer Study 2013: Snowmass on the Mississippi (CSS2013)* (Minneapolis, MN).
- Beutler, F., Blake, C., Colless, M., Jones, D. H., Staveley-Smith, L., Campbell, L., Parker, Q., Saunders, W., and Watson, F. (2011). The 6dF galaxy survey: baryon acoustic oscillations and the local hubble constant. *Mon. Not. Roy. Astron. Soc.* 416, 3017–3032. doi: 10.1111/j.1365-2966.2011.19250.x
- Beutler, F., Seo, H.-J., Ross, A. J., McDonald, P., Saito, S., Bolton, A. S., et al. (2017b). The clustering of galaxies in the completed SDSS-III Baryon Oscillation spectroscopic survey: baryon acoustic oscillations in Fourier-space. *Mon. Not. R. Astron. Soc.* 464, 3409–3430. doi: 10.1093/mnras/stw2373
- Beutler, F., Seo, H.-J., Saito, S., Chuang, C.-H., Cuesta, A. J., Eisenstein, D. J., et al. (2017a). The clustering of galaxies in the completed SDSS-III baryon oscillation spectroscopic survey: anisotropic galaxy clustering in Fourier-space. *Mon. Not. R. Astron. Soc.* 466, 2242–2260. doi: 10.1093/mnras/stw3298
- Bionta, R., Blewitt, G., Bratton, C. B., Casper, D., Ciocio, A., Claus, R., Cortez, B., et al. (1987). Observation of a neutrino burst in coincidence with supernova SN 1987a in the large magellanic cloud. *Phys. Rev. Lett.* 58:1494. doi: 10.1103/PhysRevLett.58.1494
- Bird, S., Viel, M., and Haehnelt, M. G. (2012). Massive Neutrinos and the Non-linear Matter Power Spectrum. *Mon. Not. R. Astron. Soc.* 420, 2551–2561. doi: 10.1111/j.1365-2966.2011.20222.x
- Bjaelde, O. E. et al. (2008). Neutrino dark energy—revisiting the stability issue. *JCAP* 01:026. doi: 10.1088/1475-7516/2008/01/026
- Blake, C., Kazin, E., Beutler, F., Davis, T., Parkinson, D., Brough, S., et al. (2011). The WiggleZ Dark Energy Survey: mapping the distance-redshift relation with baryon acoustic oscillations. *Mon. Not. R. Astron. Soc.* 418, 1707–1724. doi: 10.1111/j.1365-2966.2011.19592.x
- Blas, D., Lesgourgues, J., and Tram, T. (2011). The cosmic linear anisotropy solving system (CLASS) II: approximation schemes. *JCAP* 07:034. doi: 10.1088/1475-7516/2011/07/034
- Blennow, M. (2008). Prospects for cosmic neutrino detection in tritium experiments in the case of hierarchical neutrino masses. *Phys. Rev. D* 77:113014. doi: 10.1103/PhysRevD.77.113014
- Blennow, M. (2014). On the Bayesian approach to neutrino mass ordering. *J. High Energy Phys.* 01:139. doi: 10.1007/JHEP01(2014)139
- Blennow, M., Coloma, P., Fernandez-Martinez, E., Hernandez-Garcia, J., and Lopez-Pavon, J. (2017). Non-Unitarity, sterile neutrinos, and Non-Standard neutrino Interactions. *J. High Energy Phys.* 04:153. doi: 10.1007/JHEP04(2017)153
- Blennow, M., Coloma, P., Huber, P., and Schwetz, T. (2014). Quantifying the sensitivity of oscillation experiments to the neutrino mass ordering. *J. High Energy Phys.* 03:028. doi: 10.1007/JHEP03(2014)028
- Blennow, M., and Schwetz, T. (2013). Determination of the neutrino mass ordering by combining PINGU and Daya Bay II. *J. High Energy Phys.* 09:089. doi: 10.1007/JHEP09(2013)089
- Blennow, M., and Smirnov, A. Y. (2013). Neutrino propagation in matter. *Adv. High Energy Phys.* 2013:972485. doi: 10.1155/2013/972485
- Boehm, C., Dolan, M. J., and McCabe, C. (2012). Increasing  $N_{\text{eff}}$  with particles in thermal equilibrium with neutrinos. *JCAP* 12:027. doi: 10.1088/1475-7516/2012/12/027
- Bond, J., Efstathiou, G., and Tegmark, M. (1997). Forecasting cosmic parameter errors from microwave background anisotropy experiments. *Mon. Not. R. Astron. Soc.* 291, L33–L41. doi: 10.1093/mnras/291.1.L33
- Bond, J., Jaffe, A. H., and Knox, L. (2000). Radical compression of cosmic microwave background data. *Astrophys. J.* 533:19. doi: 10.1086/308625
- Bond, J. R., Efstathiou, G., and Silk, J. (1980). Massive neutrinos and the large scale structure of the universe. *Phys. Rev. Lett.* 45, 1980–1984. doi: 10.1103/PhysRevLett.45.1980
- Bourboux, H. d. M. d., Le Goff, J.-M., Blomqvist, M., Busca, N. G., Guy, J., Rich, J., et al. (2017). Baryon acoustic oscillations from the complete SDSS-III Ly $\alpha$  quasar cross-correlation function at  $z = 2.4$ . *Astron. Astrophys.* 608:A130. doi: 10.1051/0004-6361/201731731
- Bowman, J. D., Morales, M. F., and Hewitt, J. N. (2007). Constraints on fundamental cosmological parameters with upcoming epoch of reionization observations. *Astrophys. J.* 661, 1–9. doi: 10.1086/516560
- Bowman, J. D., and Rogers, A. E. (2010). A lower limit of  $dz > 0.06$  for the duration of the reionization epoch. *Nature* 468, 796–798. doi: 10.1038/nature09601
- Bowman, J. D., Rogers, A. E. E., Monsalve, R. A., Mozdzen, T. J., and Mahesh, N. (2018). An absorption profile centred at 78 megahertz in the sky-averaged spectrum. *Nature* 555, 67–70. doi: 10.1038/nature25792
- Brandbyge, J., and Hannestad, S. (2009). Grid based linear neutrino perturbations in cosmological N-body simulations. *JCAP* 05:002. doi: 10.1088/1475-7516/2009/05/002
- Brookfield, A., van de Bruck, C., Mota, D., and Tocchini-Valentini, D. (2006a). Cosmology with massive neutrinos coupled to dark energy. *Phys. Rev. Lett.* 96:061301. doi: 10.1103/PhysRevLett.96.061301
- Brookfield, A. W., van de Bruck, C., Mota, D., and Tocchini-Valentini, D. (2006b). Cosmology of mass-varying neutrinos driven by quintessence: theory and observations. *Phys. Rev. D* 73:083515. doi: 10.1103/PhysRevD.73.083515
- Bull, P., Ferreira, P. G., Patel, P., and Santos, M. G. (2015). Late-time cosmology with 21cm intensity mapping experiments. *Astrophys. J.* 803:21. doi: 10.1088/0004-637X/803/1/21
- Burns, J. O., Lazio, T. J. W., Bale, S. D., Bowman, J. D., Bradley, R. F., Carilli, C. L., et al. (2012). Probing the first stars and black holes in the early universe with the dark ages radio explorer (DARE). *Adv. Space Res.* 49, 433–450. doi: 10.1016/j.asr.2011.10.014
- Busca, N. G., Delubac, T., Rich, J., Bailey, S., Font-Ribera, A., Kirkby, D., et al. (2013). Baryon acoustic oscillations in the Ly- $\alpha$  forest of BOSS quasars. *Astron. Astrophys.* 552:A96. doi: 10.1051/0004-6361/201220724
- Caldwell, A., Merle, A., Schulz, O., and Totzauer, M. (2017). A global bayesian analysis of neutrino mass data. *Phys. Rev. D* 96:073001. doi: 10.1103/PhysRevD.96.073001
- Caldwell, R., Dave, R., and Steinhardt, P. J. (1998). Cosmological imprint of an energy component with general equation of state. *Phys. Rev. Lett.* 80, 1582–1585. doi: 10.1103/PhysRevLett.80.1582
- Canac, N., Aslanyan, G., Abazajian, K. N., Easther, R., and Price, L. C. (2016). Testing for new physics: neutrinos and the primordial power spectrum. *JCAP* 09:022. doi: 10.1088/1475-7516/2016/09/022
- Capozzi, F., Di Valentino, E., Lisi, E., Marrone, A., Melchiorri, A., and Palazzo, A. (2017). Global constraints on absolute neutrino masses and their ordering. *Phys. Rev. D* 95:096014. doi: 10.1103/PhysRevD.95.096014
- Capozzi, F., Lisi, E., and Marrone, A. (2015). PINGU and the neutrino mass hierarchy: Statistical and systematic aspects. *Phys. Rev. D* 91:073011. doi: 10.1103/PhysRevD.91.073011
- Capozzi, F., Lisi, E., Marrone, A., and Palazzo, A. (2018a). Current unknowns in the three neutrino framework. *Prog. Part. Nucl. Phys.* 102, 48–72. doi: 10.1016/j.pnpnp.2018.05.005
- Capozzi, F., Shoemaker, I. M., and Vecchi, L. (2018b). Neutrino oscillations in dark backgrounds. *JCAP* 07:004. doi: 10.1088/1475-7516/2018/07/004
- Carbone, C., Fedeli, C., Moscardini, L., and Cimatti, A. (2012). Measuring the neutrino mass from future wide galaxy cluster catalogues. *JCAP* 03:023. doi: 10.1088/1475-7516/2012/03/023
- Carbone, C., Verde, L., Wang, Y., and Cimatti, A. (2011). Neutrino constraints from future nearly all-sky spectroscopic galaxy surveys. *JCAP* 03:030. doi: 10.1088/1475-7516/2011/03/030
- Castorina, E., Sefusatti, E., Sheth, R. K., Villaescusa-Navarro, F., and Viel, M. (2014). Cosmology with massive neutrinos II: on the universality of the halo mass function and bias. *JCAP* 02:049. doi: 10.1088/1475-7516/2014/02/049
- Chakraborty, K., Deepthi, K., Goswami, S., Joshupura, A. S., and Nath, N. (2018). Partial  $\mu$ - $\tau$  reflection symmetry and its verification at DUNE and hyperkamiokande. *High Energy Phys. arXiv:1804.02022*.
- Chang, C., Pujol, A., Gaztanaga, E., Amara, A., Refregier, A., Bacon, D., et al. (2016). Galaxy bias from the dark energy survey science verification data: combining galaxy density maps and weak lensing maps. *Mon. Not. R. Astron. Soc.* 459, 3203–3216. doi: 10.1093/mnras/stw861
- Chang, T.-C., Pen, U.-L., Peterson, J. B., and McDonald, P. (2008). Baryon acoustic oscillation intensity mapping as a test of dark energy. *Phys. Rev. Lett.* 100:091303. doi: 10.1103/PhysRevLett.100.091303

- Chatterjee, A., Gandhi, R., and Singh, J. (2014). Probing lorentz and CPT violation in a magnetized iron detector using atmospheric neutrinos. *J. High Energy Phys.* 06:045. doi: 10.1007/JHEP06(2014)045
- Chatterjee, S. S., Masud, M., Pasquini, P., and Valle, J. (2017a). Cornering the revamped BMV model with neutrino oscillation data. *Phys. Lett. B* 774, 179–182. doi: 10.1016/j.physletb.2017.09.052
- Chatterjee, S. S., Pasquini, P., and Valle, J. W. F. (2017b). Probing atmospheric mixing and leptonic CP violation in current and future long baseline oscillation experiments. *Phys. Lett. B* 771, 524–531. doi: 10.1016/j.physletb.2017.05.080
- Chen, C.-S., Lee, F.-F., Lin, G.-L., and Lin, Y.-H. (2014). Probing dark matter self-interaction in the sun with iceCube-PINGU. *JCAP* 10:049. doi: 10.1088/1475-7516/2014/10/049
- Chen, X. et al. (2017). PandaX-III: Searching for neutrinoless double beta decay with high pressure  $^{136}\text{Xe}$  gas time projection chambers. *Sci. China Phys. Mech. Astron.* 60:061011. doi: 10.1007/s11433-017-9028-0
- Chen, Y., and Xu, L. (2016). Galaxy clustering, CMB and supernova data constraints on  $\phi$  CDM model with massive neutrinos. *Phys. Lett. B* 752, 66–75. doi: 10.1016/j.physletb.2015.11.022
- Cheung, K., Kuo, J.-L., Ng, K.-W., and Tsai, Y.-L. S. (2018). The impact of EDGES 21-cm data on dark matter interactions. *arXiv:1803.09398*.
- Chevallier, M., and Polarski, D. (2001). Accelerating universes with scaling dark matter. *Int. J. Mod. Phys. D* 10, 213–224. doi: 10.1142/S0218271801000822
- Choubey, S., Dasgupta, B., Dighe, A., and Mirizzi, A. (2010). Signatures of collective and matter effects on supernova neutrinos at large detectors. *arXiv:1008.0308*.
- Choubey, S., Ghosh, A., Ohlsson, T., and Tiwari, D. (2015). Neutrino physics with non-standard interactions at INO. *J. High Energy Phys.* 12:126. doi: 10.1007/JHEP12(2015)126
- Choubey, S., Ghosh, A., and Tiwari, D. (2018a). Prospects of Indirect Searches for Dark Matter at INO. *JCAP* 05:006. doi: 10.1088/1475-7516/2018/05/006
- Choubey, S., Goswami, S., Gupta, C., Lakshmi, S., and Thakore, T. (2018b). Sensitivity to neutrino decay with atmospheric neutrinos at the INO-ICAL detector. *Phys. Rev. D* 97:033005. doi: 10.1103/PhysRevD.97.033005
- Choubey, S., Goswami, S., and Pramanik, D. (2018c). A Study of Invisible Neutrino Decay at DUNE and its Effects on  $\theta_{23}$  Measurement. *J. High Energy Phys.* 02:055. doi: 10.1007/JHEP02(2018)055
- Choubey, S., and Ohlsson, T. (2014). Bounds on non-standard neutrino interactions using PINGU. *Phys. Lett. B* 739, 357–364. doi: 10.1016/j.physletb.2014.11.010
- Clark, S., Dutta, B., Gao, Y., Ma, Y.-Z., and Strigari, L. E. (2018). 21cm limits on decaying dark matter and primordial black holes. *Phys. Rev. D* 98:043006. doi: 10.1103/PhysRevD.98.043006
- Clesse, S., Lopez-Honorez, L., Ringeval, C., Tashiro, H., and Tytgat, M. H. (2012). Background reionization history from omniscopes. *Phys. Rev. D* 86:123506. doi: 10.1103/PhysRevD.86.123506
- Cleveland, B., Daily, T., Davis, R. Jr., Distell, J. R., Lande, K., Lee, C. K., et al. (1998). Measurement of the solar electron neutrino flux with the Homestake chlorine detector. *Astrophys. J.* 496, 505–526. doi: 10.1086/305343
- Cocco, A. G., Mangano, G., and Messina, M. (2007). Probing low energy neutrino backgrounds with neutrino capture on beta decaying nuclei. *JCAP* 06:015. doi: 10.1088/1475-7516/2007/06/015
- Cole, S., Percival, W. J., Peacock, J. A., Norberg, P., Baugh, C. M., Frenk, C. S., et al. (2005). The 2dF galaxy redshift survey: power-spectrum analysis of the final dataset and Cosmological Implications. *Mon. Not. R. Astron. Soc.* 362, 505–534. doi: 10.1111/j.1365-2966.2005.09318.x
- Coloma, P. (2016). Non-standard interactions in propagation at the deep underground neutrino experiment. *J. High Energy Phys.* 03:016. doi: 10.1007/JHEP03(2016)016
- Coloma, P., Denton, P. B., Gonzalez-Garcia, M. C., Maltoni, M., and Schwetz, T. (2017). Curtailing the dark side in non-standard neutrino interactions. *J. High Energy Phys.* 04:116. doi: 10.1007/JHEP04(2017)116
- Coloma, P., Forero, D. V., and Parke, S. J. (2018). DUNE sensitivities to the mixing between sterile and tau neutrinos. *J. High Energy Phys.* 07:079. doi: 10.1007/JHEP07(2018)079
- Coloma, P., and Peres, O. L. G. (2017). Visible neutrino decay at DUNE. *J. High Energy Phys.* *arXiv:1705.03599*.
- Coloma, P., and Schwetz, T. (2016). Generalized mass ordering degeneracy in neutrino oscillation experiments. *Phys. Rev. D* 94:055005. doi: 10.1103/PhysRevD.95.079903
- Cooray, A. R. (1999). Weighing neutrinos: weak lensing approach. *Astron. Astrophys.* 348:31.
- Costa, A. A., Landim, R. C., Wang, B., and Abdalla, E. (2018). Interacting dark energy: possible explanation for 21-cm absorption at cosmic dawn. *Eur. Phys. J. C* 78:746. doi: 10.1140/epjc/s10052-018-6237-7
- Costanzi, M., Sartoris, B., Viel, M., and Borgani, S. (2014). Neutrino constraints: what large-scale structure and CMB data are telling us? *JCAP* 10:081. doi: 10.1088/1475-7516/2014/10/081
- Costanzi, M., Villaescusa-Navarro, F., Viel, M., Xia, J.-Q., Borgani, S., Castorina, E., et al. (2013). Cosmology with massive neutrinos III: the halo mass function and application to galaxy clusters. *JCAP* 12:012. doi: 10.1088/1475-7516/2013/12/012
- Cravens, J. et al. (2008). Solar neutrino measurements in Super-Kamiokande-II. *Phys. Rev. D* 78:032002.
- Croft, R. A., Hu, W., and Dave, R. (1999). Cosmological Limits on the Neutrino Mass from the Lya Forest. *Phys. Rev. Lett.* 83, 1092–1095. doi: 10.1103/PhysRevLett.83.1092
- Crotty, P., Lesgourgues, J., and Pastor, S. (2004). Current cosmological bounds on neutrino masses and relativistic relics. *Phys. Rev. D* 69:123007. doi: 10.1103/PhysRevD.69.123007
- Cuesta, A. J., Niro, V., and Verde, L. (2016). Neutrino mass limits: robust information from the power spectrum of galaxy surveys. *Phys. Dark Univ.* 13, 77–86. doi: 10.1016/j.dark.2016.04.005
- Dakin, J., Brandbyge, J., Hannestad, S., Haugbølle, T., and Tram, T. (2017).  $\nu$ CONCEPT: Cosmological neutrino simulations from the non-linear Boltzmann hierarchy. *arXiv:1712.03944*.
- D'Amico, G., Panci, P., and Strumia, A. (2018). Bounds on dark matter annihilations from 21 cm data. *Phys. Rev. Lett.* 121:011103. doi: 10.1103/PhysRevLett.121.011103
- Dasgupta, B., Dighe, A., Raffelt, G. G., and Smirnov, A. Y. (2009). Multiple spectral splits of supernova neutrinos. *Phys. Rev. Lett.* 103:051105. doi: 10.1103/PhysRevLett.103.051105
- Dasgupta, B., Mirizzi, A., Tamborra, I., and Tomas, R. (2010). Neutrino mass hierarchy and three-flavor spectral splits of supernova neutrinos. *Phys. Rev. D* 81:093008. doi: 10.1103/PhysRevD.81.093008
- Dash, N., Datar, V., and Majumder, G. (2016). Sensitivity for detection of decay of dark matter particle using ICAL at INO. *Pramana* 86, 927–937. doi: 10.1007/s12043-015-1094-5
- Dawson, K. S., Kneib, J.-P., Percival, W. J., Alam, S., Albareti, F. D., Anderson, S. F., et al. (2016). The SDSS-IV extended baryon oscillation spectroscopic survey: overview and early data. *Astron. J.* 151:44. doi: 10.3847/0004-6256/151/2/44
- Dawson, K. S., Schlegel, D. J., Ahn, C. P., Anderson, S. F., Aubourg, É., Bailey, S., et al. (2013). The baryon oscillation spectroscopic survey of SDSS-III. *Astron. J.* 145:10. doi: 10.1088/0004-6256/145/1/10
- De Bernardis, F., Kitching, T. D., Heavens, A., and Melchiorri, A. (2009). Determining the neutrino mass hierarchy with cosmology. *Phys. Rev. D* 80:123509. doi: 10.1103/PhysRevD.80.123509
- De Bernardis, F., Stevens, J. R., Hasselfield, M., Alonso, D., Bond, J. R., Calabrese, E., et al. (2016). “Survey strategy optimization for the atacama Cosmology telescope,” in *Proceedings Volume 9910, Observatory Operations: Strategies, Processes, and Systems VI; 991014* (Edinburgh).
- de Gouvêa, A., and Kelly, K. J. (2016). Non-standard neutrino interactions at DUNE. *Nucl. Phys. B* 908, 318–335. doi: 10.1016/j.nuclphysb.2016.03.013
- de Gouvêa, A., and Kelly, K. J. (2017). Neutrino vs. Antineutrino oscillation parameters at DUNE and hyper-kamiokande. *Phys. Rev. D* 96:095018. doi: 10.1103/PhysRevD.96.095018
- de Haan, T., Benson, B. A., Bleem, L. E., Allen, S. W., Applegate, D. E., Ashby, M. L. N., et al. (2016). Cosmological constraints from galaxy clusters in the 2500 square-degree SPT-SZ survey. *Astrophys. J.* 832:95. doi: 10.3847/0004-637X/832/1/95
- de Putter, R., Linder, E. V., and Mishra, A. (2014). Inflationary freedom and cosmological neutrino constraints. *Phys. Rev. D* 89:103502. doi: 10.1103/PhysRevD.89.103502
- de Putter, R., Mena, O., Giusarma, E., Ho, S., Cuesta, A., Seo, H.-J., et al. (2012). New neutrino mass bounds from sloan digital sky survey

- III data release 8 photometric luminous galaxies. *Astrophys. J.* 761:12. doi: 10.1088/0004-637X/761/1/12
- de Putter, R., Zahn, O., and Linder, E. V. (2009). CMB lensing constraints on neutrinos and dark energy. *Phys. Rev. D* 79:065033. doi: 10.1103/PhysRevD.79.065033
- De Romeri, V., Fernandez-Martinez, E., and Sorel, M. (2016). Neutrino oscillations at DUNE with improved energy reconstruction. *J. High Energy Phys.* 09:030. doi: 10.1007/JHEP09(2016)030
- De Rujula, A., and Lusignoli, M. (1982). Calorimetric measurements of  $^{163}\text{Ho}$  decay as tools to determine the electron neutrino mass. *Phys. Lett. B* 118:429. doi: 10.1016/0370-2693(82)90218-0
- de Salas, P., Forero, D., Ternes, C., Tórtola, M., and Valle, J. (2018). Status of neutrino oscillations 2018:  $3\sigma$  hint for normal mass ordering and improved CP sensitivity. *Phys. Lett. B* 782, 633–640. doi: 10.1016/j.physletb.2018.06.019
- de Salas, P. F., Gariazzo, S., Lesgourgues, J., and Pastor, S. (2017). Calculation of the local density of relic neutrinos. *JCAP* 09:034. doi: 10.1088/1475-7516/2017/09/034
- de Salas, P. F., and Pastor, S. (2016). Relic neutrino decoupling with flavour oscillations revisited. *JCAP* 07:051. doi: 10.1088/1475-7516/2016/07/051
- Debono, I., Rassat, A., Refregier, A., Amara, A., and Kitching, T. (2010). Weak lensing forecasts for dark energy, neutrinos and initial conditions. *Mon. Not. R. Astron. Soc.* 404, 110–119. doi: 10.1111/j.1365-2966.2010.16284.x
- Deepthi, K. N., Goswami, S., and Nath, N. (2017). Nonstandard interactions jeopardizing the hierarchy sensitivity of DUNE. *Phys. Rev. D* 96:075023. doi: 10.1103/PhysRevD.96.075023
- Delabrouille, J., de Bernardis, P., Bouchet, F. R., Achúcarro, A., Ade, P. A. R., Allison, R., et al. (2018). Exploring cosmic origins with CORE: survey requirements and mission design. *JCAP* 04:014. doi: 10.1088/1475-7516/2018/04/014
- Dell’Oro, S., Marcocci, S., Viel, M., and Vissani, F. (2016). Neutrinoless double beta decay: 2015 review. *Adv. High Energy Phys.* 2016:2162659. doi: 10.1155/2016/2162659
- Delubac, T., Bautista, J. E., Busca, N. G., Rich, J., Kirkby, D., Bailey, S., et al. (2015). Baryon acoustic oscillations in the Ly $\alpha$  forest of BOSS DR11 quasars. *Astron. Astrophys.* 574:A59. doi: 10.1051/0004-6361/201423969
- Dentler, M., Hernández-Cabezudo, A., Kopp, J., Machado, P. A. N., Maltoni, M., Martínez-Soler, I., et al. (2018). Updated global analysis of neutrino oscillations in the presence of eV-scale sterile neutrinos. *J. High Energy Phys.* 08:010. doi: 10.1007/JHEP08(2018)010
- Dentler, M., Hernández-Cabezudo, A., Kopp, J., Maltoni, M., and Schwetz, T. (2017). Sterile neutrinos or flux uncertainties?—status of the reactor anti-neutrino anomaly. *J. High Energy Phys.* 11:099. doi: 10.1007/JHEP11(2017)099
- Desjacques, V., Jeong, D., and Schmidt, F. (2018). Large-scale galaxy bias. *Phys. Rept.* 733:1–193. doi: 10.1016/j.physrep.2017.12.002
- Dewdney, P., Turner, W., Millenaar, R., McCool, R., Lazio, J., Cornwell, T. J., et al. (2015). *SKA Baseline Description*. Available online at: [https://www.skatelescope.org/wp-content/uploads/2014/03/SKA-TEL-SKO-0000308\\_SKA1\\_System\\_Baseline\\_v2\\_DescriptionRev01-part-1-signed.pdf](https://www.skatelescope.org/wp-content/uploads/2014/03/SKA-TEL-SKO-0000308_SKA1_System_Baseline_v2_DescriptionRev01-part-1-signed.pdf).
- Di Valentino, E., Brinckmann, T., Gerbino, M., Poulin, V., Bouchet, F. R., Lesgourgues, J., et al. (2018a). Exploring cosmic origins with cORE: cosmological parameters. *JCAP* 04:017. doi: 10.1088/1475-7516/2018/04/017
- Di Valentino, E., Gariazzo, S., Gerbino, M., Giusarma, E., and Mena, O. (2016a). Dark radiation and inflationary freedom after planck 2015. *Phys. Rev. D* 93:083523. doi: 10.1103/PhysRevD.93.083523
- Di Valentino, E., Gariazzo, S., Giusarma, E., and Mena, O. (2015). Robustness of cosmological axion mass limits. *Phys. Rev. D* 91:123505. doi: 10.1103/PhysRevD.91.123505
- Di Valentino, E., Giusarma, E., Lattanzi, M., Mena, O., Melchiorri, A., and Silk, J. (2016b). Cosmological Axion and neutrino mass constraints from Planck 2015 temperature and polarization data. *Phys. Lett. B* 752, 182–185. doi: 10.1016/j.physletb.2015.11.025
- Di Valentino, E., Giusarma, E., Mena, O., Melchiorri, A., and Silk, J. (2016c). Cosmological limits on neutrino unknowns versus low redshift priors. *Phys. Rev. D* 93:083527. doi: 10.1103/PhysRevD.93.083527
- Di Valentino, E., Hm, C. B. e., Hivon, E., and Bouchet, F. R. (2018b). Reducing the  $H_0$  and  $\sigma_8$  tensions with Dark Matter-neutrino interactions. *Phys. Rev. D* 97:043513. doi: 10.1103/PhysRevD.97.043513
- Di Valentino, E., Melchiorri, A., and Mena, O. (2013). Dark radiation sterile neutrino candidates after Planck data. *JCAP* 11:018. doi: 10.1088/1475-7516/2013/11/018
- Dirian, Y. (2017). Changing the prior: absolute neutrino mass constraints in nonlocal gravity. *Phys. Rev. D* 96:083513. doi: 10.1103/PhysRevD.96.083513
- Doe, P., Kofron, J., McBride, E. L., Robertson, R. G. H., Rosenberg, L. J., Rybka, G., S. et al. (2013). Project 8: determining neutrino mass from tritium beta decay using a frequency-based method. *Nuclear Exp.* arXiv:1309.7093.
- Domcke, V., and Spinrath, M. (2017). Detection prospects for the cosmic neutrino background using laser interferometers. *JCAP* 06:055. doi: 10.1088/1475-7516/2017/06/055
- Duan, H., Fuller, G. M., Carlson, J., and Qian, Y.-Z. (2007). Neutrino mass hierarchy and stepwise spectral swapping of supernova neutrino flavors. *Phys. Rev. Lett.* 99:241802. doi: 10.1103/PhysRevLett.99.241802
- Duan, H., Fuller, G. M., and Qian, Y.-Z. (2010). Collective neutrino oscillations. *Ann. Rev. Nucl. Part. Sci.* 60, 569–594. doi: 10.1146/annurev.nucl.012809.104524
- Duda, G., Gelmini, G., and Nussinov, S. (2001). Expected signals in relic neutrino detectors. *Phys. Rev. D* 64:122001. doi: 10.1103/PhysRevD.64.122001
- Dutta, D., Ghoshal, P., and Roy, S. (2017). Effect of non unitarity on neutrino mass hierarchy determination at DUNE, NOvA and T2K. *Nucl. Phys. B* 920, 385–401. doi: 10.1016/j.nuclphysb.2017.04.018
- Dvali, G., and Funcke, L. (2016). Small neutrino masses from gravitational  $\theta$ -term. *Phys. Rev. D* 93:113002. doi: 10.1103/PhysRevD.93.113002
- Eberle, B., Ringwald, A., Song, L., and Weiler, T. J. (2004). Relic neutrino absorption spectroscopy. *Phys. Rev. D* 70:023007. doi: 10.1103/PhysRevD.70.023007
- Efstathiou, G. (2014). H0 revisited. *Mon. Not. R. Astron. Soc.* 440, 1138–1152. doi: 10.1093/mnras/stu278
- Eguchi, K., Enomoto, S., Furuno, K., Goldman, J., Hanada, H., Ikeda, H. et al. (2003). First results from KamLAND: evidence for reactor anti-neutrino disappearance. *Phys. Rev. Lett.* 90:021802. doi: 10.1103/PhysRevLett.90.021802
- Eisenstein, D. J., Weinberg, D. H., Agol, E., Aihara, H., Prieto, C. A., Anderson, S. F., et al. (2011). SDSS-III: massive spectroscopic surveys of the distant universe, the milky way galaxy, and extra-solar planetary systems. *Astron. J.* 142:72. doi: 10.1088/0004-6256/142/3/72
- Eisenstein, D. J., Zehavi, I., Hogg, D. W., Scoccimarro, R., Blanton, M. R., Nichol, R. C., et al. (2005). Detection of the baryon acoustic peak in the large-scale correlation function of SDSS luminous red galaxies. *Astrophys. J.* 633, 560–574. doi: 10.1086/466512
- Elgaroy, O., and Lahav, O. (2003). Upper limits on neutrino masses from the 2dFGRS and WMAP: the role of priors. *JCAP* 04:004. doi: 10.1088/1475-7516/2003/04/004
- Eliseev, S., Blaum, K., Block, M., Chenmarev, S., Dorrer, H., Düllmann, C. E. et al. (2015). Direct measurement of the mass difference of  $^{163}\text{Ho}$  and  $^{163}\text{Dy}$  solves the Q-value puzzle for the neutrino mass determination. *Phys. Rev. Lett.* 115:062501. doi: 10.1103/PhysRevLett.115.062501
- Engel, J., and Menéndez, J. (2017). Status and future of nuclear matrix elements for neutrinoless double-beta decay: a review. *Rept. Prog. Phys.* 80:046301.
- Errard, J., Feeney, S. M., Peiris, H. V., and Jaffe, A. H. (2016). Robust forecasts on fundamental physics from the foreground-obscured, gravitationally-lensed CMB polarization. *JCAP* 03:052. doi: 10.1088/1475-7516/2016/03/052
- Escribuela, F. J., Forero, D. V., Miranda, O. G., Tortola, M., and Valle, J. W. F. (2017). Probing CP violation with non-unitary mixing in long-baseline neutrino oscillation experiments: DUNE as a case study. *New J. Phys.* 19:093005. doi: 10.1088/1367-2630/aa79ec
- Esfahani, A. A., Asner, D. M., Böser, S., Cervantes, R., Claessens, C., de Viveiros, L., et al. (2017). Determining the neutrino mass with cyclotron radiation emission spectroscopy-Project 8. *J. Phys. G* 44:054004. doi: 10.1088/1361-6471/aa5b4f
- Essinger-Hileman, T., Ali, A., Amiri, M., Appel, J. W., Araujo, D., Bennett, C. L., et al. (2014). “CLASS: the cosmology large angular scale surveyor,” in *Proceedings Millimeter, Submillimeter, and Far-Infrared Detectors and Instrumentation for Astronomy VII* (Montréal, QC). doi: 10.1117/12.2056701
- Esteban, I., Gonzalez-Garcia, M. C., Maltoni, M., Martínez-Soler, I., and Schwetz, T. (2017). Updated fit to three neutrino mixing: exploring the accelerator-reactor complementarity. *J. High Energy Phys.* 01:087. doi: 10.1007/JHEP01(2017)087

- Esteban-Pretel, A., Mirizzi, A., Pastor, S., Tomas, R., Raffelt, G. G., Serpico, P. D., G., et al. (2008). Role of dense matter in collective supernova neutrino transformations. *Phys. Rev. D* 78:085012. doi: 10.1103/PhysRevD.78.085012
- Esteban-Pretel, A., Pastor, S., Tomas, R., Raffelt, G. G., and Sigl, G. (2007). Decoherence in supernova neutrino transformations suppressed by deleptonization. *Phys. Rev. D* 76:125018. doi: 10.1103/PhysRevD.76.125018
- Ewall-Wice, A., Chang, T., Lazio, J., Dore, O., Seiffert, M., and Monsalve, R. (2018). Modeling the radio background from the first black holes at cosmic Dawn: implications for the 21 cm absorption amplitude. *arXiv:1803.01815*.
- Faessler, A., Hodak, R., Kovalenko, S., and Simkovic, F. (2011). Beta decaying nuclei as a probe of cosmic neutrino background. *J. High Energy Phys. arXiv:1102.1799*.
- Falkowski, A., and Petraki, K. (2018). 21cm absorption signal from charge sequestration. *J. High Energy Phys. arXiv:1803.10096*.
- Fardon, R., Nelson, A. E., and Weiner, N. (2004). Dark energy from mass varying neutrinos. *JCAP* 10:005. doi: 10.1088/1475-7516/2004/10/005
- Farzan, Y., Peres, O., and Smirnov, A. (2001). Neutrino mass spectrum and future beta decay experiments. *Nucl. Phys. B* 612, 59–97. doi: 10.1016/S0550-3213(01)00361-3
- Farzan, Y., and Tortola, M. (2018). Neutrino oscillations and Non-Standard Interactions. *Front. Phys.* 6:10. doi: 10.3389/fphy.2018.00010
- Feng, C., and Holder, G. (2018). Enhanced global signal of neutral hydrogen due to excess radiation at cosmic dawn. *Astrophys. J.* 858:L17. doi: 10.3847/2041-8213/aac0fe
- Fialkov, A., Barkana, R., and Cohen, A. (2018). Constraining baryon–dark matter scattering with the cosmic dawn 21-cm signal. *Phys. Rev. Lett.* 121:011101. doi: 10.1103/PhysRevLett.121.011101
- Fogli, G., Lisi, E., Marrone, A., Melchiorri, A., Palazzo, A., Serra, P., Silk, J., et al. (2007). Observables sensitive to absolute neutrino masses: a Reappraisal after WMAP-3y and first MINOS results. *Phys. Rev. D* 75:053001. doi: 10.1103/PhysRevD.75.053001
- Fogli, G., Lisi, E., Marrone, A., Melchiorri, A., Palazzo, A., Serra, P., Silk, J., et al. (2008). Observables sensitive to absolute neutrino masses. 2. *Phys. Rev. D* 78:033010. doi: 10.1103/PhysRevD.78.033010
- Font-Ribera, A., Kirkby, D., Busca, N., Miralda-Escudé, J., Ross, N. P., Slosar, A., et al. (2014a). Quasar-lyman  $\alpha$  forest cross-correlation from BOSS DR11 : baryon acoustic oscillations. *JCAP* 05:027. doi: 10.1088/1475-7516/2014/05/027
- Font-Ribera, A., McDonald, P., Mostek, N., Reid, B. A., Seo, H.-J., and Slosar, A. (2014b). DESI and other dark energy experiments in the era of neutrino mass measurements. *JCAP* 05:023. doi: 10.1088/1475-7516/2014/05/023
- Forero, D. V. and Huang, W.-C. (2017). Sizable NSI from the  $SU(2)_L$  scalar doublet-singlet mixing and the implications in DUNE. *J. High Energy Phys.* 03:018. doi: 10.1007/JHEP03(2017)018
- Forero, D. V., and Huber, P. (2016). Hints for leptonic CP violation or New Physics? *Phys. Rev. Lett.* 117:031801. doi: 10.1103/PhysRevLett.117.031801
- Forero, D. V., Tortola, M., and Valle, J. W. F. (2014). Neutrino oscillations refitted. *Phys. Rev. D* 90:093006. doi: 10.1103/PhysRevD.90.093006
- Fraser, S., Hektor, A., Hütsi, G., Kannike, K., Marzo, C., Marzola, L., et al. (2018). The EDGES 21 cm anomaly and properties of dark matter. *Phys. Lett. B* 785, 159–164. doi: 10.1016/j.physletb.2018.08.035
- Freedman, W. L., and Madore, B. F. (2010). The hubble constant. *Ann. Rev. Astron. Astrophys.* 48, 673–710. doi: 10.1146/annurev-astro-082708-101829
- Fukuda, Y., Hayakawa, T., Ichihara, E., Inoue, K., Ishihara, K., Ishino, H., et al. (1998). Evidence for oscillation of atmospheric neutrinos. *Phys. Rev. Lett.* 81, 1562–1567. doi: 10.1103/PhysRevLett.81.1562
- Furlanetto, S., Oh, S., and Briggs, F. (2006). Cosmology at low frequencies: the 21 cm transition and the high-redshift universe. *Phys. Rept.* 433, 181–301. doi: 10.1016/j.physrep.2006.08.002
- Furlanetto, S. R. (2015). “The 21-cm line as a probe of reionization,” in *Understanding the Epoch of Cosmic Reionization. Astrophysics and Space Science Library*, Vol 423, ed A. Mesinger (Cham: Springer), 247–280.
- Gando, A., Gando, Y., Ichimura, K., Ikeda, H., Kibe, Y., Kishimoto, Y., et al. (2011). Constraints on  $\theta_{13}$  from a three-flavor oscillation analysis of reactor antineutrinos at KamLAND. *Phys. Rev. D* 83:052002. doi: 10.1103/PhysRevD.83.052002
- Gando, A., Gando, Y., Ichimura, K., Ikeda, H., Kibe, Y., Kishimoto, Y., et al. (2016). Search for majorana neutrinos near the inverted mass hierarchy region with KamLAND-Zen. *Phys. Rev. Lett.* 117:082503. doi: 10.1103/PhysRevLett.117.082503
- Gando, A. “KamLAND-Zen,” in *XXVIII International Conference on Neutrino Physics and Astrophysics*, 4–9 June (2018), (Heidelberg). Available online at: URL: <http://doi.org/10.5281/zenodo.1286895>
- Gariazzo, S., Archidiacono, M., de Salas, P. F., Mena, O., Ternes, C. A., and Tórtola, M. (2018a). Neutrino masses and their ordering: global data, priors and models. *JCAP* 03:011. doi: 10.1088/1475-7516/2018/03/011
- Gariazzo, S., Giunti, C., Laveder, M., and Li, Y. F. (2017). Updated global 3+1 analysis of short-baseline neutrino oscillations. *J. High Energy Phys.* 06:135. doi: 10.1007/JHEP06(2017)135
- Gariazzo, S., Giunti, C., Laveder, M., and Li, Y. F. (2018b). Model-independent  $\bar{\nu}_e$  short-baseline oscillations from reactor spectral ratios. *Phys. Lett. B* 782, 13–21. doi: 10.1016/j.physletb.2018.04.057
- Gariazzo, S., Giunti, C., Laveder, M., Li, Y. F., and Zavanin, E. M. (2016). Light sterile neutrinos. *J. Phys. G*, 43:033001.
- Gastaldo, L. (2018). “Determining the electron neutrino mass with Ho-163,” in *XXVIII International Conference on Neutrino Physics and Astrophysics* (Heidelberg). Available online at: <https://doi.org/10.5281/zenodo.1286950>
- Gavela, M. B., Hernandez, D., Lopez Honorez, L., Mena, O., and Rigolin, S. (2009). Dark coupling. *JCAP* 07:034. doi: 10.1088/1475-7516/2009/07/034
- Ge, S.-F., Lindner, M., and Rodejohann, W. (2017). New physics and atmospheric neutrino trident production with PINGU and ORCA. *Phys. Lett. B* 772, 164–168. doi: 10.1016/j.physletb.2017.06.020
- Gelmini, G. B. (2005). Prospect for relic neutrino searches. *Phys. Scripta* T121, 131–136. doi: 10.1088/0031-8949/2005/T121/019
- Gerbino, M., Freese, K., Vagnozzi, S., Lattanzi, M., Mena, O., Giusarma, E., and Ho, S. (2017a). Impact of neutrino properties on the estimation of inflationary parameters from current and future observations. *Phys. Rev. D* 95:043512. doi: 10.1103/PhysRevD.95.043512
- Gerbino, M., Lattanzi, M., Mena, O., and Freese, K. (2017b). A novel approach to quantifying the sensitivity of current and future cosmological datasets to the neutrino mass ordering through Bayesian hierarchical modeling. *Phys. Lett. B* 775, 239–250. doi: 10.1016/j.physletb.2017.10.052
- Ghosh, A., Thakore, T., and Choubey, S. (2013). Determining the neutrino mass hierarchy with INO, T2K, NOvA and reactor Experiments. *J. High Energy Phys.* 04:009. doi: 10.1007/JHEP04(2013)009
- Giachero, A., Alpert, B., Becker, D., Bennett, D., Biasotti, M., Brofferio, C., et al. (2017). Measuring the electron neutrino mass with improved sensitivity: the HOLMES experiment. *JINST* 12:C02046. doi: 10.1088/1748-0221/12/02/C02046
- Girardi, I., Meloni, D., Ohlsson, T., Zhang, H., and Zhou, S. (2014). Constraining sterile neutrinos using reactor neutrino experiments. *J. High Energy Phys.* 08:057. doi: 10.1007/JHEP08(2014)057
- Giuliani, A. (2018). “The mid and long term future of neutrinoless double beta decay,” in *XXVIII International Conference on Neutrino Physics and Astrophysics* (Heidelberg). Available online at : <http://doi.org/10.5281/zenodo.1286915>
- Giuliani, A., and Poves, A. (2012). Neutrinoless double-beta decay. *Adv. High Energy Phys.* 2012:857016.
- Giunti, C. (2017). “Light sterile neutrinos and neutrinoless double-beta decay,” in *Proceedings, Matrix Elements for the Double beta decay Experiments (MEDEX'17)* (Prague) Vol. 1894, 020009.
- Giunti, C., and Kim, C. W. (2007). *Fundamentals of Neutrino Physics and Astrophysics*. Oxford: Oxford University Press.
- Giunti, C., and Zavanin, E. M. (2015). Predictions for neutrinoless double-beta decay in the 3+1 sterile neutrino scenario. *J. High Energy Phys.* 07:171. doi: 10.1007/JHEP07(2015)171
- Giusarma, E., Corsi, M., Archidiacono, M., de Putter, R., Melchiorri, A., Mena, O., et al. (2011). Constraints on massive sterile neutrino species from current and future cosmological data. *Phys. Rev. D* 83:115023. doi: 10.1103/PhysRevD.83.115023
- Giusarma, E., de Putter, R., Ho, S., and Mena, O. (2013a). Constraints on neutrino masses from Planck and Galaxy Clustering data. *Phys. Rev. D* 88:063515. doi: 10.1103/PhysRevD.88.063515

- Giusarma, E., De Putter, R., and Mena, O. (2013b). Testing standard and nonstandard neutrino physics with cosmological data. *Phys. Rev. D* 87:043515. doi: 10.1103/PhysRevD.87.043515
- Giusarma, E., Di Valentino, E., Lattanzi, M., Melchiorri, A., and Mena, O. (2014). Relic neutrinos, thermal axions and cosmology in early 2014. *Phys. Rev. D* 90:043507. doi: 10.1103/PhysRevD.90.043507
- Giusarma, E., Gerbino, M., Mena, O., Vagnozzi, S., Ho, S., and Freese, K. (2016). Improvement of cosmological neutrino mass bounds. *Phys. Rev. D* 94:083522. doi: 10.1103/PhysRevD.94.083522
- Giusarma, E., Vagnozzi, S., Ho, S., Ferraro, S., Freese, K., Kamen-Rubio, R., et al. (2018). Scale-dependent galaxy bias, CMB lensing-galaxy cross-correlation, and neutrino masses. *arXiv:1802.08694*.
- Gomez-Cadenas, J. J., Martin-Albo, J., Mezzetto, M., Monrabal, F., and Sorel, M. (2012). The Search for neutrinoless double beta decay. *Riv. Nuovo Cim.* 35, 29–98. doi: 10.1393/ncr/i2012-10074-9
- Goobar, A., Hannestad, S., Mortsell, E., and Tu, H. (2006). A new bound on the neutrino mass from the sdss baryon acoustic peak. *JCAP* 06:019. doi: 10.1088/1475-7516/2006/06/019
- Gratton, S., Lewis, A., and Efstathiou, G. (2008). Prospects for constraining neutrino mass using planck and lyman-alpha forest data. *Phys. Rev. D* 77:083507. doi: 10.1103/PhysRevD.77.083507
- Greenhill, L., and Bernardi, G. (2012). HI epoch of reionization arrays. *arXiv:1201.1700*.
- Guo, R.-Y., Zhang, J.-F., and Zhang, X. (2018). Exploring neutrino mass and mass hierarchy in the scenario of vacuum energy interacting with cold dark matter. *Chin. Phys. C* 42:095103. doi: 10.1088/1674-1137/42/9/095103
- Hamann, J., Hannestad, S., Lesgourgues, J., Rampf, C., and Wong, Y. Y. Y. (2010a). Cosmological parameters from large scale structure - geometric versus shape information. *JCAP* 07:022. doi: 10.1088/1475-7516/2010/07/022
- Hamann, J., Hannestad, S., Melchiorri, A., and Wong, Y. Y. Y. (2008). Nonlinear corrections to the cosmological matter power spectrum and scale-dependent galaxy bias: implications for parameter estimation. *JCAP* 07:017. doi: 10.1088/1475-7516/2008/07/017
- Hamann, J., Hannestad, S., Raffelt, G., and Wong, Y. Y. Y. (2007a). Observational bounds on the cosmic radiation density. *JCAP* 08:021. doi: 10.1088/1475-7516/2007/08/021
- Hamann, J., Hannestad, S., Raffelt, G. G., Tamborra, I., and Wong, Y. Y. Y. (2010b). Cosmology seeking friendship with sterile neutrinos. *Phys. Rev. Lett.* 105:181301. doi: 10.1103/PhysRevLett.105.181301
- Hamann, J., Hannestad, S., Raffelt, G. G., and Wong, Y. Y. Y. (2011). Sterile neutrinos with eV masses in cosmology: how disfavoured exactly? *JCAP* 09:034. doi: 10.1088/1475-7516/2011/09/034
- Hamann, J., Hannestad, S., Sloth, M. S., and Wong, Y. Y. Y. (2007b). How robust are inflation model and dark matter constraints from cosmological data? *Phys. Rev. D* 75:023522. doi: 10.1103/PhysRevD.75.023522
- Hamann, J., Hannestad, S., and Wong, Y. Y. Y. (2012). Measuring neutrino masses with a future galaxy survey. *JCAP* 11:052. doi: 10.1088/1475-7516/2012/11/052
- Handley, W., and Millea, M. (2018). Maximum entropy priors with derived parameters in a specified distribution. *e-Print: arXiv:1804.08143*.
- Handley, W. J., Hobson, M. P., and Lasenby, A. N. (2015a). PolyChord: nested sampling for cosmology. *Mon. Not. R. Astron. Soc.* 450, L61–L65. doi: 10.1093/mnras/slv047
- Handley, W. J., Hobson, M. P., and Lasenby, A. N. (2015b). PolyChord: next-generation nested sampling. *Mon. Not. R. Astron. Soc.* 453:4384. doi: 10.1093/mnras/stv1911
- Hannestad, S. (2003). Neutrino masses and the number of neutrino species from WMAP and 2dFGRS. *JCAP* 05:004. doi: 10.1088/1475-7516/2003/05/004
- Hannestad, S. (2005a). Neutrino masses and the dark energy equation of state - Relaxing the cosmological neutrino mass bound. *Phys. Rev. Lett.* 95:221301. doi: 10.1103/PhysRevLett.95.221301
- Hannestad, S. (2005b). Structure formation with strongly interacting neutrinos - Implications for the cosmological neutrino mass bound. *JCAP* 02:011.
- Hannestad, S., Mirizzi, A., Raffelt, G. G., and Wong, Y. Y. Y. (2008). Cosmological constraints on neutrino plus axion hot dark matter: Update after WMAP-5. *JCAP* 04:019. doi: 10.1088/1475-7516/2008/04/019
- Hannestad, S., Mirizzi, A., Raffelt, G. G., and Wong, Y. Y. Y. (2007). Cosmological constraints on neutrino plus axion hot dark matter. *JCAP* 08:015. doi: 10.1088/1475-7516/2007/08/015
- Hannestad, S., Mirizzi, A., Raffelt, G. G., and Wong, Y. Y. Y. (2010). Neutrino and axion hot dark matter bounds after WMAP-7. *JCAP* 08:001. doi: 10.1088/1475-7516/2010/08/001
- Hannestad, S., and Raffelt, G. (2004). Cosmological mass limits on neutrinos, axions, and other light particles. *JCAP* 04:008. doi: 10.1088/1475-7516/2004/04/008
- Hannestad, S., Raffelt, G. G., Sigl, G., and Wong, Y. Y. Y. (2006a). Self-induced conversion in dense neutrino gases: pendulum in flavour space. *Phys. Rev. D* 74:105010. doi: 10.1103/PhysRevD.76.029901
- Hannestad, S., and Schwetz, T. (2016). Cosmology and the neutrino mass ordering. *JCAP* 11:035. doi: 10.1088/1475-7516/2016/11/035
- Hannestad, S., Tu, H., and Wong, Y. Y. Y. (2006b). Measuring neutrino masses and dark energy with weak lensing tomography. *JCAP* 06:025. doi: 10.1088/1475-7516/2006/06/025
- Hartz, M. (2017). *T2K NEUTRINO OSCILLATION RESULTS WITH DATA UP TO 2017 SUMMER*. Available online at: <http://www.t2k.org/docs/talk/282>
- Heavens, A. F., and Sellentin, E. (2018). Objective Bayesian analysis of neutrino masses and hierarchy. *JCAP* 04:047. doi: 10.1088/1475-7516/2018/04/047
- Heitmann, K., Lawrence, E., Kwan, J., Habib, S., and Higdon, D. (2014). The coyote universe extended: precision emulation of the matter power spectrum. *Astrophys. J.* 780:111. doi: 10.1088/0004-637X/780/1/111
- Hektor, A., Hütsi, G., Marzola, L., Raidal, M., Vaskonen, V., and Veermäe, H. (2018). Constraining primordial black holes with the EDGES 21-cm absorption signal. *Phys. Rev. D* 98:023503. doi: 10.1103/PhysRevD.98.023503
- Hildebrandt, H., Viola, M., Heymans, C., Joudaki, S., Kuijken, K., Blake, C., et al. (2017). KiDS-450: cosmological parameter constraints from tomographic weak gravitational lensing. *Mon. Not. R. Astron. Soc.* 465:1454. doi: 10.1093/mnras/stw2805
- Hill, J. C., and Baxter, E. J. (2018). Can early dark energy explain EDGES? *JCAP* 08:037. doi: 10.1088/1475-7516/2018/08/037
- Hilton, M., Hasselfield, M., Sifón, C., Battaglia, N., Aiola, S., Bharadwaj, V., et al. (2018). The atacama cosmology telescope: the two-season ACTPol Sunyaev-Zel'dovich effect selected cluster catalog. *Astrophys. J. Suppl.* 235:20. doi: 10.3847/1538-4365/aaa6cb
- Himmel, A. (2018). *First Oscillation Results With Neutrino and Antineutrino Beams in NOvA*. Available online at: <http://nova-docdb.fnal.gov/cgi-bin/ShowDocument?docid=30273>
- Hirano, S., and Bromm, V. (2018). Baryon-dark matter scattering and first star formation. *Mon. Not. R. Astron. Soc.* 480:L85. doi: 10.1093/mnras/sly132
- Hirata, K., Kajita, T., Koshihara, M., Nakahata, M., Oyama, Y., Sato, N., et al. (1987). Observation of a Neutrino Burst from the Supernova SN 1987a. *Phys. Rev. Lett.* 58, 1490–1493. doi: 10.1103/PhysRevLett.58.1490
- Horiuchi, S., and Kneller, J. P. (2018). What can be learned from a future supernova neutrino detection? *J. Phys. G* 45:043002. doi: 10.1088/1361-6471/aaa90a
- Hosaka, J. et al. (2006). Solar neutrino measurements in super-Kamiokande-I. *Phys. Rev. D* 73:112001. doi: 10.1103/PhysRevD.73.112001
- Hu, B., Raveri, M., Silvestri, A., and Frusciante, N. (2015). Exploring massive neutrinos in dark cosmologies with EFTCAMB/ EFTCosmoMC. *Phys. Rev. D* 91:063524. doi: 10.1103/PhysRevD.91.063524
- Hu, W., Eisenstein, D. J., and Tegmark, M. (1998). Weighing neutrinos with galaxy surveys. *Phys. Rev. Lett.* 80, 5255–5258. doi: 10.1103/PhysRevLett.80.5255
- Humphreys, E. M. L., Reid, M. J., Moran, J. M., Greenhill, L. J., and Argon, A. L. (2013). Toward a new geometric distance to the active galaxy NGC 4258. III. Final results and the hubble constant. *Astrophys. J.* 775:13. doi: 10.1088/0004-637X/775/1/13
- Huterer, D., and Linder, E. V. (2007). Separating dark physics from physical darkness: minimalist modified gravity vs. dark energy. *Phys. Rev. D* 75:023519. doi: 10.1103/PhysRevD.75.023519
- Ichiki, K., and Keum, Y.-Y. (2008). Neutrino masses from cosmological probes in interacting neutrino dark-energy models. *J. High Energy Phys.* 06:058. doi: 10.1088/1126-6708/2008/06/058

- Ichiki, K., Takada, M., and Takahashi, T. (2009). Constraints on Neutrino Masses from Weak Lensing. *Phys. Rev. D* 79:023520. doi: 10.1103/PhysRevD.79.023520
- Iliev, I. T., Shapiro, P. R., Ferrara, A., and Martel, H. (2002). On the direct detectability of the cosmic dark ages: 21-cm emission from minihalos. *Astrophys. J.* 572:123. doi: 10.1086/341869
- Inman, D., Emberson, J. D., Pen, U.-L., Farchi, A., Yu, H.-R., and Harnois-Déraps, J. (2015). Precision reconstruction of the cold dark matter-neutrino relative velocity from  $N$ -body simulations. *Phys. Rev. D* 92:023502. doi: 10.1103/PhysRevD.92.023502
- Ivezic, Z., Tyson, J., Allsman, R., Andrew, J., and Angel, R. (2008). LSST: from science drivers to reference design and anticipated data products. *Astrophysics arXiv:0805.2366*.
- Jackson, N. (2007). The hubble constant. *Living Rev. Rel.* 10:4. doi: 10.12942/lrr-2007-4
- Jain, B., Spergel, D., Bean, R., Connolly, A., Dell'antonio, I., Frieman, J., et al. (2015). The whole is greater than the sum of the parts: optimizing the joint science return from LSST, Euclid and WFIRST. *arXiv:1501.07897*.
- Janka, H.-T. (2012). Explosion mechanisms of core-collapse supernovae. *Ann. Rev. Nucl. Part. Sci.* 62, 407–451. doi: 10.1146/annurev-nucl-102711-094901
- Jeffreys, H. (1961). *Theory of Probability*. The International series of monographs on physics. Clarendon Press.
- Jimenez, R., Kitching, T., Peña-Garay, C., and Verde, L. (2010). Can we measure the neutrino mass hierarchy in the sky? *JCAP* 05:035. doi: 10.1088/1475-7516/2010/05/035
- Joudaki, S. (2013). Constraints on neutrino mass and light degrees of freedom in extended cosmological parameter spaces. *Phys. Rev. D* 87:083523. doi: 10.1103/PhysRevD.87.083523
- Joudaki, S., Mead, A., Blake, C., Choi, A., de Jong, J., Erben, T., et al. (2017). KiDS-450: testing extensions to the standard cosmological model. *Mon. Not. R. Astron. Soc.* 471, 1259–1279. doi: 10.1093/mnras/stx998
- Jurkovich, H., Ferreira, C., and Pasquini, P. (2018). Shadowing neutrino mass hierarchy with lorentz invariance violation *High Energy Phys. arXiv:1806.08752*.
- Kaether, F., Hampel, W., Heusser, G., Kiko, J., and Kirsten, T. (2010). Reanalysis of the GALLEX solar neutrino flux and source experiments. *Phys. Lett. B* 685, 47–54. doi: 10.1016/j.physletb.2010.01.030
- Kang, Z. (2018). Post-recombination dark matter for the 21-cm signal. *arXiv:1803.04928*.
- Kaplinghat, M., Knox, L., and Song, Y.-S. (2003). Determining neutrino mass from the CMB alone. *Phys. Rev. Lett.* 91:241301. doi: 10.1103/PhysRevLett.91.241301
- Kazin, E. A., Koda, J., Blake, C., Padmanabhan, N., Brough, S., Colless, M., et al. (2014). The wiggleZ dark energy survey: improved distance measurements to  $z = 1$  with reconstruction of the baryonic acoustic feature. *Mon. Not. R. Astron. Soc.* 441, 3524–3542. doi: 10.1093/mnras/stu778
- Khan, A. N., McKay, D. W., and Tahir, F. (2013). Sensitivity of medium-baseline reactor neutrino mass-hierarchy experiments to nonstandard interactions. *Phys. Rev. D* 88:113006. doi: 10.1103/PhysRevD.88.113006
- Kharusi, S. A., Alamre, A., Albert, J. B., Alfari, M., Anton, G., Arnquist, I. J., et al. (2018). nEXO Pre-Conceptual Design Report. *arXiv:1805.11142*.
- Kilbinger, M. (2015). Cosmology with cosmic shear observations: a review. *Rept. Prog. Phys.* 78:086901. doi: 10.1088/0034-4885/78/8/086901
- Kim, S.-B. (2015). New results from RENO and prospects with RENO-50. *Nucl. Part. Phys. Proc.* 265–266, 93–98. doi: 10.1016/j.nuclphysbps.2015.06.024
- Kitching, T., Heavens, A., Verde, L., Serra, P., and Melchiorri, A. (2008). Finding evidence for massive neutrinos using 3D weak lensing. *Phys. Rev. D* 77:103008. doi: 10.1103/PhysRevD.77.103008
- Ko, Y. J., Kim, B. R., Kim, J. Y., Han, B. Y., Jang, C. H., Jeon, E. J., et al. (2017). Sterile neutrino search at the NEOS experiment. *Phys. Rev. Lett.* 118:121802. doi: 10.1103/PhysRevLett.118.121802
- Köhlinger, F., Viola, M., Joachimi, B., Hoekstra, H., van Uitert, E., Hildebrandt, H., et al. (2017). KiDS-450: the tomographic weak lensing power spectrum and constraints on cosmological parameters. *Mon. Not. R. Astron. Soc.* 471:4412. doi: 10.1093/mnras/stx1820
- Krnjaic, G., Machado, P. A. N., and Necib, L. (2018). Distorted neutrino oscillations from ultralight scalar dark matter. *Phys. Rev. D* 97:075017. doi: 10.1103/PhysRevD.97.075017
- La Vacca, G., Bonometto, S., and Colombo, L. (2009). Higher neutrino mass allowed if DM and DE are coupled. *New Astron.* 14, 435–442. doi: 10.1016/j.newast.2008.12.004
- La Vacca, G., and Kristiansen, J. (2009). Dynamical dark energy model parameters with or without massive neutrinos. *JCAP* 07:036. doi: 10.1088/1475-7516/2009/07/036
- Lattanzi, M., and Gerbino, M. (2018). Status of neutrino properties and future prospects - Cosmological and astrophysical constraints. *Front. Phys.* 5:70. doi: 10.3389/fphy.2017.00070
- Lazauskas, R., Vogel, P., and Volpe, C. (2008). Charged current cross section for massive cosmological neutrinos impinging on radioactive nuclei. *J. Phys. G* 35:025001. doi: 10.1088/0954-3899/35/2/025001
- Lesgourgues, J. (2011a). The cosmic linear anisotropy solving system (CLASS) I: overview. *arXiv:1104.2932*.
- Lesgourgues, J. (2011b). The cosmic linear anisotropy solving system (CLASS) III: comparison with CAMB for  $\Lambda$ CDM. *arXiv:1104.2934*.
- Lesgourgues, J., Mangano, G., Miele, G., and Pastor, S. (2013). *Neutrino Cosmology*. Cambridge, UK: Cambridge University Press. doi: 10.1017/CBO9781139012874
- Lesgourgues, J., and Pastor, S. (2006). Massive neutrinos and cosmology. *Phys. Rept.* 429, 307–379. doi: 10.1016/j.physrep.2006.04.001
- Lesgourgues, J., and Pastor, S. (2012). Neutrino mass from Cosmology. *Adv. High Energy Phys.* 2012:608515. doi: 10.1155/2012/608515
- Lesgourgues, J., and Pastor, S. (2014). Neutrino cosmology and Planck. *New J. Phys.* 16:065002. doi: 10.1088/1367-2630/16/6/065002
- Lesgourgues, J., Pastor, S., and Perotto, L. (2004). Probing neutrino masses with future galaxy redshift surveys. *Phys. Rev. D* 70:045016. doi: 10.1103/PhysRevD.70.045016
- Lesgourgues, J., Perotto, L., Pastor, S., and Piat, M. (2006). Probing neutrino masses with cmb lensing extraction. *Phys. Rev. D*, 73:045021. doi: 10.1103/PhysRevD.73.045021
- Lesgourgues, J., and Tram, T. (2014). Fast and accurate CMB computations in non-flat FLRW universes. *JCAP* 09, 032. doi: 10.1088/1475-7516/2014/09/032
- Levi, M., Bebek, C., Beers, T., Blum, R., Cahn, R., Eisenstein, D., et al. (2013). The DESI experiment, a Whitepaper for snowmass 2013. *arXiv:1308.0847*.
- Lewis, A., and Bridle, S. (2002). Cosmological parameters from CMB and other data: a monte carlo approach. *Phys. Rev. D* 66:103511. doi: 10.1103/PhysRevD.66.103511
- Lewis, A., Challinor, A., and Lasenby, A. (2000). Efficient computation of CMB anisotropies in closed FRW models. *Astrophys. J.* 538, 473–476. doi: 10.1086/309179
- Li, Y.-F. (2015). Detection prospects of the cosmic neutrino background. *Int. J. Mod. Phys. A* 30:1530031. doi: 10.1142/9789814704779\_0025
- Li, Y.-F. and Zhao, Z.-h. (2014). Tests of lorentz and CPT violation in the medium baseline reactor antineutrino experiment. *Phys. Rev. D* 90:113014. doi: 10.1103/PhysRevD.90.113014
- Liao, J., Marfatia, D., and Whisnant, K. (2017). Nonstandard interactions in solar neutrino oscillations with Hyper-Kamiokande and JUNO. *Phys. Lett. B* 771, 247–253. doi: 10.1016/j.physletb.2017.05.054
- Linder, E. V. (2003). Exploring the expansion history of the universe. *Phys. Rev. Lett.* 90:091301. doi: 10.1103/PhysRevLett.90.091301
- Liu, A., and Parsons, A. R. (2016). Constraining cosmology and ionization history with combined 21 cm power spectrum and global signal measurements. *Mon. Not. R. Astron. Soc.* 457, 1864–1877. doi: 10.1093/mnras/stw071
- Liu, A., Pritchard, J. R., Allison, R., Parsons, A. R., Seljak, U., and Sherwin, B. D. (2016). Eliminating the optical depth nuisance from the CMB with 21 cm cosmology. *Phys. Rev. D* 93:043013. doi: 10.1103/PhysRevD.93.043013
- Liu, H., and Slatyer, T. R. (2018). Too hot, too cold or just right? Implications of a 21-cm signal for dark matter annihilation and decay. *Phys. Rev. D* 98:023501. doi: 10.1103/PhysRevD.98.023501
- Liu, J., Bird, S., Matilla, J. M. Z., Hill, J. C., Haiman, Z., Madhavacheril, M. S., et al. (2018). MassiveNuS: cosmological massive neutrino simulations. *JCAP* 03:049. doi: 10.1088/1475-7516/2018/03/049

- Loeb, A., and Wyithe, S. (2008). Precise measurement of the cosmological power spectrum with a dedicated 21cm survey after reionization. *Phys. Rev. Lett.* 100:161301. doi: 10.1103/PhysRevLett.100.161301
- Long, A. J., Lunardini, C., and Sabancilar, E. (2014). Detecting non-relativistic cosmic neutrinos by capture on tritium: phenomenology and physics potential. *JCAP* 08:038. doi: 10.1088/1475-7516/2014/08/038
- Long, A. J., Raveri, M., Hu, W., and Dodelson, S. (2018). Neutrino mass priors for cosmology from random matrices. *Phys. Rev. D* 97:043510. doi: 10.1103/PhysRevD.97.043510
- Lopez Honorez, L., and Mena, O. (2010). Instabilities in dark coupled models and constraints from cosmological data. *AIP Conf. Proc.* 1241, 1016–1024. doi: 10.1063/1.3462595
- Lorenz, C. S., Calabrese, E., and Alonso, D. (2017). Distinguishing between Neutrinos and time-varying Dark Energy through Cosmic Time. *Phys. Rev. D* 96:043510. doi: 10.1103/PhysRevD.96.043510
- LoVerde, M., and Zaldarriaga, M. (2014). Neutrino clustering around spherical dark matter halos. *Phys. Rev. D* 89:063502. doi: 10.1103/PhysRevD.89.063502
- Lunardini, C., and Smirnov, A. (2001a). Neutrinos from SN1987A, earth matter effects and the LMA solution of the solar neutrino problem. *Phys. Rev. D* 63:073009. doi: 10.1103/PhysRevD.63.073009
- Lunardini, C., and Smirnov, A. (2001b). Supernova neutrinos: earth matter effects and neutrino mass spectrum. *Nucl. Phys.* B616:307–348.
- Lunardini, C., and Smirnov, A. Y. (2003). Probing the neutrino mass hierarchy and the 13 mixing with supernovae. *JCAP* 06:009. doi: 10.1088/1475-7516/2003/06/009
- Lunardini, C. and Smirnov, A. Y. (2004). Neutrinos from SN1987A: flavor conversion and interpretation of results. *Astropart. Phys.* 21, 703–720. doi: 10.1016/j.astropartphys.2004.05.005
- Madau, P., Meiksin, A., and Rees, M. J. (1997). 21-CM tomography of the intergalactic medium at high redshift. *Astrophys. J.* 475:429. doi: 10.1086/303549
- Mahdawi, M. S., and Farrar, G. R. (2018). Constraints on dark matter with a moderately large and velocity-dependent DM-Nucleon cross-section. *arXiv:1804.03073*.
- Mangano, G., Miele, G., Pastor, S., Pinto, T., Pisanti, O., and Serpico, P. D. (2005). Relic neutrino decoupling including flavor oscillations. *Nucl. Phys.* B729, 221–234. doi: 10.1016/j.nuclphysb.2005.09.041
- Mao, Y., Tegmark, M., McQuinn, M., Zaldarriaga, M., and Zahn, O. (2008). How accurately can 21 cm tomography constrain cosmology? *Phys. Rev. D* 78:023529. doi: 10.1103/PhysRevD.78.023529
- Martin-Albo, J., Muñoz Vidal, J., Ferrario, P., Nebot-Guino, M., Gómez-Cadenas, J. J., Álvarez, V., et al. (2016). Sensitivity of NEXT-100 to neutrinoless double beta decay. *J. High Energy Phys.* 05:159. doi: 10.1007/JHEP05(2016)159
- Masud, M., and Mehta, P. (2016). Nonstandard interactions and resolving the ordering of neutrino masses at DUNE and other long baseline experiments. *Phys. Rev. D* 94:053007. doi: 10.1103/PhysRevD.94.053007
- McGaugh, S. (2018). Strong hydrogen absorption at cosmic Dawn: the signature of a baryonic universe. *Res. Notes AAS* 2:37. doi: 10.3847/2515-5172/aab497
- McQuinn, M., Zahn, O., Zaldarriaga, M., Hernquist, L., and Furlanetto, S. R. (2006). Cosmological parameter estimation using 21 cm radiation from the epoch of reionization. *Astrophys. J.* 653, 815–830. doi: 10.1086/505167
- Melchiorri, A., Mena, O., and Slosar, A. (2007). An improved cosmological bound on the thermal axion mass. *Phys. Rev. D* 76:041303. doi: 10.1103/PhysRevD.76.041303
- Mellema, G., Koopmans, L., Abdalla, F., Bernardi, G., Ciardi, B., Daiboo, S., et al. (2013). Reionization and the cosmic dawn with the square kilometre array. *Exper. Astron.* 36, 235–318. doi: 10.1007/s10686-013-9334-5
- Mena, O., Mocioiu, I., and Quigg, C. (2007). Gravitational lensing of supernova neutrinos. *Astropart. Phys.* 28, 348–356. doi: 10.1016/j.astropartphys.2007.07.002
- Mena, O. and Parke, S. J. (2004). Unified graphical summary of neutrino mixing parameters. *Phys. Rev. D* 69:117301.
- Mikheev, S. P., and Smirnov, A. Yu. (1985). Resonance amplification of oscillations in matter and spectroscopy of solar neutrinos. *Sov. J. Nucl. Phys.* 42:913–917. doi: 10.1103/PhysRevD.69.117301
- Mikheev, S. P., and Smirnov, A. Yu. (1986). Resonant amplification of neutrino oscillations in matter and solar neutrino spectroscopy. *Nuovo Cim.* C9, 17–26. doi: 10.1007/BF02508049
- Miranda, O., Tortola, M., and Valle, J. (2006). Are solar neutrino oscillations robust? *J. Health Energy Phys.* 10:008. doi: 10.1088/1126-6708/2006/10/008
- Mirizzi, A., Tamborra, I., Janka, H.-T., Saviano, N., Scholberg, K., Bollig, R., et al. (2016). Supernova neutrinos: production, oscillations and detection. *Riv. Nuovo Cim.* 39, 1–112. doi: 10.1393/ncr/i2016-10120-8
- Mitridate, A., and Podo, A. (2018). Bounds on dark matter decay from 21 cm line. *JCAP* 05:069. doi: 10.1088/1475-7516/2018/05/069
- Mosher, J. et al. (2014). Cosmological parameter uncertainties from SALT-II type Ia supernova light curve models. *Astrophys. J.* 793:16. doi: 10.1088/0004-637X/793/1/16
- Mota, D., Pettorino, V., Robbers, G., and Wetterich, C. (2008). Neutrino clustering in growing neutrino quintessence. *Phys. Lett.* B663:160–164. doi: 10.1016/j.physletb.2008.03.060
- Muñoz, J. B., and Dvorkin, C. (2018). Efficient computation of galaxy bias with neutrinos and other relics. *Phys. Rev. D* 98:043503. doi: 10.1103/PhysRevD.98.043503
- Muñoz, J. B., Dvorkin, C., and Loeb, A. (2018). 21-cm fluctuations from charged dark matter. *arXiv:1804.01092*.
- Muñoz, J. B., and Loeb, A. (2018). A small amount of mini-charged dark matter could cool the baryons in the early Universe. *Nature* 557:684. doi: 10.1038/s41586-018-0151-x
- Munshi, D., Valageas, P., Van Waerbeke, L., and Heavens, A. (2008). Cosmology with weak lensing surveys. *Phys. Rept.* 462, 67–121. doi: 10.1016/j.physrep.2008.02.003
- Murase, K. (2018). New prospects for detecting high-energy neutrinos from nearby Supernovae. *Phys. Rev. D* 97:081301. doi: 10.1103/PhysRevD.97.081301
- Nakano, Y. (2016). Ph.D. Thesis, University of Tokyo. Available online at: [http://www-sk.icrr.u-tokyo.ac.jp/sk/\\_pdf/articles/2016/doc\\_thesis\\_nakano.pdf](http://www-sk.icrr.u-tokyo.ac.jp/sk/_pdf/articles/2016/doc_thesis_nakano.pdf)
- Newburgh, L. B., Addison, G. E., Amiri, M., Bandura, K., Bond, J. R., Connor, L., et al. (2014). Calibrating CHIME, a new radio interferometer to probe dark energy. *Proc. SPIE Int. Soc. Opt. Eng.* 9145: 4V. doi: 10.1117/12.2056962
- Ohlsson, T., Zhang, H., and Zhou, S. (2014). Nonstandard interaction effects on neutrino parameters at medium-baseline reactor antineutrino experiments. *Phys. Lett.* B728, 148–155. doi: 10.1016/j.physletb.2013.11.052
- Orebi Gann, G. (2018) “SNO+,” in *XXVIII International Conference on Neutrino Physics and Astrophysics* (Heidelberg) Available online at: <http://doi.org/10.5281/zenodo.1286908>
- Orebi Gann, G. D. (2015). Physics potential of an advanced scintillation detector: introducing THEIA *arXiv:1504.08284*.
- Ott, C., O'Connor, E., Gossan, S., Abdikamalov, E., Gamma, U., and Drasco, S. (2013). Core-collapse supernovae, neutrinos, and gravitational waves. *Nuclear Phys. B* 235–236, 381–387. doi: 10.1016/j.nuclphysb.2013.04.036
- Ouellet, J. (2018) “Latest results from the CUORE experiment,” in *XXVIII International Conference on Neutrino Physics and Astrophysics, 4–9 June 2018*, (Heidelberg) Available online at: <http://doi.org/10.5281/zenodo.1286904>
- Oyama, Y., Kohri, K., and Hazumi, M. (2016). Constraints on the neutrino parameters by future cosmological 21 cm line and precise CMB polarization observations. *JCAP* 02:008. doi: 10.1088/1475-7516/2016/02/008
- Oyama, Y., Shimizu, A., and Kohri, K. (2013). Determination of neutrino mass hierarchy by 21 cm line and CMB B-mode polarization observations. *Phys. Lett.* B718, 1186–1193. doi: 10.1016/j.physletb.2012.12.053
- Pac, M. Y. (2018). Recent results from RENO. *PoS NuFact2017*:038. doi: 10.22323/1.295.003
- Paciga, G., Chang, T.-C., Gupta, Y., Nityanada, R., Odegova, J., Pen, U.-L., et al. (2011). The GMRT epoch of reionization experiment: a new upper limit on the neutral hydrogen power spectrum at  $z \approx 8.6$ . *Mon. Not. R. Astron. Soc.* 413, 1174–1183. doi: 10.1111/j.1365-2966.2011.18208.x
- Palanque-Delabrouille, N., Yèche, C., Baur, J., Magneville, C., Rossi, G., Lesgourgues, J., Arnaud Borde, et al. (2015). Neutrino masses and cosmology with Lyman-alpha forest power spectrum. *JCAP* 11:011. doi: 10.1088/1475-7516/2015/11/011
- Palanque-Delabrouille, N., Yèche, C., Borde, A., Le Goff, J.-M., Rossi, G., Viel, M., et al. (2013). The one-dimensional Ly-alpha forest power spectrum from BOSS. *Astron. Astrophys.* 559:A85. doi: 10.1051/0004-6361/201322130
- Parno, D. (2018) “KATRIN: toward a high-precision neutrino-mass determination with tritium,” in *XXVIII International Conference on Neutrino Physics and*

- Astrophysics*, 4–9 June 2018, (Heidelberg). Available online at: <http://doi.org/10.5281/zenodo.1287933>
- Parsons, A. R., Backer, D. C., Bradley, R. F., Aguirre, J. E., Benoit, E. E., Carilli, C. L., et al. (2010). The precision array for probing the epoch of reionization: 8 station results. *Astron. J.* 139:1468. doi: 10.1088/0004-6256/139/4/1468
- Pascoli, S., and Petcov, S. (2002). The SNO solar neutrino data, neutrinoless double beta decay and neutrino mass spectrum. *Phys. Lett. B* 544, 239–250. doi: 10.1016/S0370-2693(02)02510-8
- Pasquini, P., Chulia, S. C., and Valle, J. W. F. (2017). Neutrino oscillations from warped flavor symmetry: predictions for long baseline experiments T2K, NOvA and DUNE. *Phys. Rev. D* 95:095030. doi: 10.1103/PhysRevD.95.095030
- Patrick, C., and Xie, F. (2017). Status of the superNEMO  $0\nu\beta\beta$  Experiment. *arXiv:1704.06670*.
- Peccei, R. D., and Quinn, H. R. (1977a). Constraints imposed by CP conservation in the presence of instantons. *Phys. Rev. D* 16, 1791–1797.
- Peccei, R. D., and Quinn, H. R. (1977b). CP Conservation in the Presence of Instantons. *Phys. Rev. Lett.* 38, 1440–1443.
- Peebles, P., and Ratra, B. (1988). Cosmology with a Time Variable Cosmological Constant. *Astrophys. J.* 325:L17. doi: 10.1086/185100
- Peirone, S., Frusciante, N., Hu, B., Raveri, M., and Silvestri, A. (2018). Do current cosmological observations rule out all Covariant Galileons? *Phys. Rev. D* 97:063518. doi: 10.1103/PhysRevD.97.063518
- Percival, W. J., Baugh, C. M., Bland-Hawthorn, J., Bridges, T., Cannon, R., Cole, S., et al. (2001). The 2dF galaxy redshift survey: the power spectrum and the matter content of the Universe. *Mon. Not. R. Astron. Soc.* 327:1297. doi: 10.1046/j.1365-8711.2001.04827.x
- Petcov, S., and Piai, M. (2002). The LMA MSW solution of the solar neutrino problem, inverted neutrino mass hierarchy and reactor neutrino experiments. *Phys. Lett. B* 533, 94–106. doi: 10.1016/S0370-2693(02)01591-5
- Pober, J. C., Ali, Z. S., Parsons, A. R., McQuinn, M., Aguirre, J. E., Bernardi, G., et al. (2015). PAPER-64 constraints on reionization II: the temperature of the  $z = 8.4$  intergalactic medium. *Astrophys. J.* 809:62. doi: 10.1088/0004-637X/809/1/62
- Pospelov, M., Pradler, J., Ruderman, J. T., and Urbano, A. (2018). Room for new physics in the rayleigh-jeans tail of the cosmic microwave background. *Phys. Rev. Lett.* 121:031103. doi: 10.1103/PhysRevLett.121.031103
- Pritchard, J. R., and Loeb, A. (2012). 21-cm cosmology. *Rept. Prog. Phys.* 75:086901. doi: 10.1088/0034-4885/75/8/086901
- Pritchard, J. R., and Pierpaoli, E. (2008). Constraining massive neutrinos using cosmological 21 cm observations. *Phys. Rev. D* 78:065009. doi: 10.1103/PhysRevD.78.065009
- Qian, X., Tan, A., Wang, W., Ling, J., McKeown, R., and Zhang, C. (2012). Statistical evaluation of experimental determinations of neutrino mass hierarchy. *Phys. Rev. D* 86:113011. doi: 10.1103/PhysRevD.86.113011
- Qian, X., and Vogel, P. (2015). Neutrino mass hierarchy. *Prog. Part. Nucl. Phys.* 83, 1–30. doi: 10.1016/j.pnpnp.2015.05.002
- Radovic, A. (2018). *JETP January 2018, NOvA Oscillation Results*. Available online at: <http://nova-docdb.fnal.gov/cgi-bin/ShowDocument?docid=25938>
- Raffelt, G. (1999). Stars as particle-physics laboratories. *AIP Conf. Proc.* 490, 125–162.
- Raffelt, G., and Sigl, G. (2007). Self-induced decoherence in dense neutrino gases. *Phys. Rev. D* 75:083002. doi: 10.1103/PhysRevD.75.083002
- Raffelt, G. G., and Smirnov, A. Y. (2007a). Adiabaticity and spectral splits in collective neutrino transformations. *Phys. Rev. D* 76:125008. doi: 10.1103/PhysRevD.76.125008
- Raffelt, G. G., and Smirnov, A. Y. (2007b). Self-induced spectral splits in supernova neutrino fluxes. *Phys. Rev. D* 76:081301. doi: 10.1103/PhysRevD.77.029903
- Ratra, B., and Peebles, P. (1988). Cosmological consequences of a rolling homogeneous scalar field. *Phys. Rev. D* 37:3406.
- Reid, B. A., Verde, L., Jimenez, R., and Mena, O. (2010). Robust neutrino constraints by combining low redshift observations with the CMB. *JCAP* 01:003. doi: 10.1088/1475-7516/2010/01/003
- Renk, J., Zumalacárregui, M., Montanari, F., and Barreira, A. (2017). Galileon gravity in light of ISW, CMB, BAO and  $H_0$  data. *JCAP* 10:020. doi: 10.1088/1475-7516/2017/10/020
- Ribordy, M., and Smirnov, A. Y. (2013). Improving the neutrino mass hierarchy identification with inelasticity measurement in PINGU and ORCA. *Phys. Rev. D* 87:113007. doi: 10.1103/PhysRevD.87.113007
- Riemer-Sørensen, S. et al. (2012). The wiggleZ dark energy survey: cosmological neutrino mass constraint from blue high-redshift galaxies. *Phys. Rev. D* 85:081101. doi: 10.1103/PhysRevD.85.081101
- Riemer-Sørensen, S., Parkinson, D., and Davis, T. M. (2014). Combining Planck data with large scale structure information gives a strong neutrino mass constraint. *Phys. Rev. D* 89:103505. doi: 10.1103/PhysRevD.89.103505
- Riemer-Sørensen, S., Parkinson, D., Davis, T. M., and Blake, C. (2013). Simultaneous constraints on the number and mass of relativistic species. *Astrophys. J.* 763:89. doi: 10.1088/0004-637X/763/2/89
- Riess, A. G., Macri, L., Casertano, S., Lampeitl, H., Ferguson, H. C., Filippenko, A. V., et al. (2011). A 3% solution: determination of the hubble constant with the hubble space telescope and wide field camera 3. *Astrophys. J.* 730:119. doi: 10.1088/0004-637X/730/2/119
- Riess, A. G., Macri, L. M., Hoffmann, S. L., Scolnic, D., Casertano, S., Filippenko, A. V., et al. (2016). A 2.4% determination of the local value of the hubble constant. *Astrophys. J.* 826:56. doi: 10.3847/0004-637X/826/1/56
- Ringwald, A. (2005). How to detect big bang relic neutrinos? *High Energy Phys.* 473–491.
- Ringwald, A. and Wong, Y. Y. Y. (2004). Gravitational clustering of relic neutrinos and implications for their detection. *JCAP* 12:005. doi: 10.1088/1475-7516/2004/12/005
- Robertson, D. S., and Albuquerque, I. F. M. (2018). Probing velocity dependent self-interacting dark matter with neutrino telescopes. *JCAP* 02:056. doi: 10.1088/1475-7516/2018/02/056
- Rodejohann, W. (2011). Neutrino-less double beta decay and particle physics. *Int. J. Mod. Phys. E* 20, 1833–1930. doi: 10.1142/S0218301311020186
- Ross, A. J., Beutler, F., Chuang, C.-H., Pellejero-Ibanez, M., Seo, H.-J., Vargas-Magana, M., et al. (2017). The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: observational systematics and baryon acoustic oscillations in the correlation function. *Mon. Not. R. Astron. Soc.* 464, 1168–1191. doi: 10.1093/mnras/stw2372
- Ross, A. J., Samushia, L., Howlett, C., Percival, W. J., Burden, A., and Manera, M. (2015). The clustering of the SDSS DR7 main Galaxy sample - I. A 4 per cent distance measure at  $z = 0.15$ . *Mon. Not. R. Astron. Soc.* 449, 835–847. doi: 10.1093/mnras/stv154
- Rozo, E., and Rykoff, E. S. (2014). redMaPPer II: X-ray and SZ performance benchmarks for the SDSS catalog. *Astrophys. J.* 783:80. doi: 10.1088/0004-637X/783/2/80
- Rozo, E., Rykoff, E. S., Bartlett, J. G., and Melin, J. B. (2015). redMaPPer – III. A detailed comparison of the Planck 2013 and SDSS DR8 redMaPPer cluster catalogues. *Mon. Not. R. Astron. Soc.* 450, 592–605. doi: 10.1093/mnras/stv605
- Rybka, G. (2018). “Project 8: progress towards using cyclotron radiation emission spectroscopy on atomic tritium for a neutrino mass measurement,” in *XXVIII International Conference on Neutrino Physics and Astrophysics*, 4–9 June 2018 (Heidelberg). Available online at: <http://doi.org/10.5281/zenodo.1286954>
- Rykoff, E., Rozo, E., Busha, M. T., Cunha, C. E., Finoguenov, A., Evrard, A., et al. (2014). redMaPPer I: algorithm and SDSS DR8 catalog. *Astrophys. J.* 785, 104. doi: 10.1088/0004-637X/785/2/104
- Safarzadeh, M., Scannapieco, E., and Babul, A. (2018). A limit on the warm dark matter particle mass from the redshifted 21 cm absorption line. *Astrophys. J.* 859:L18. doi: 10.3847/2041-8213/aac5e0
- Saito, S., Takada, M., and Taruya, A. (2008). Impact of massive neutrinos on nonlinear matter power spectrum. *Phys. Rev. Lett.* 100:191301. doi: 10.1103/PhysRevLett.100.191301
- Sanchez, A. G. et al. (2017). The clustering of galaxies in the completed SDSS-III Baryon oscillation spectroscopic survey: cosmological implications of the configuration-space clustering wedges. *Mon. Not. R. Astron. Soc.* 464, 1640–1658. doi: 10.1093/mnras/stw2443
- Santos, M. G., and Cooray, A. (2006). Cosmological and astrophysical parameter measurements with 21-cm anisotropies during the era of reionization. *Phys. Rev. D* 74:083517. doi: 10.1103/PhysRevD.74.083517
- Satpathy, S., Alam, S., Ho, S., White, M., Bahcall, N. A., Beutler, F., et al. (2016). BOSS DR12 combined galaxy sample: The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: On the

- measurement of growth rate using galaxy correlation functions. *Mon. Not. R. Astron. Soc.* doi: 10.1093/mnras/stx883
- Schechter, J., and Valle, J. W. F. (1982). Neutrinoless Double beta Decay in SU(2) x U(1) Theories. *Phys. Rev. D* 25:2951. doi: 10.1103/PhysRevD.25.2951
- Scholberg, K. (2012). Supernova neutrino detection. *Ann. Rev. Nucl. Part. Sci.* 62, 81–103. doi: 10.1146/annurev-nucl-102711-095006
- Scholberg, K. (2018). Supernova signatures of neutrino mass ordering. *J. Phys. G* 45:014002. doi: 10.1088/1361-6471/aa97be
- Schramm, D., and Truran, J. (1990). New physics from supernova SN1987A. *Phys. Rept.* 189, 89–126. doi: 10.1016/0370-1573(90)90020-3
- Schwetz, T., Freese, K., Gerbino, M., Giusarma, E., Hannestad, S., Lattanzi, M., et al. (2017). Comment on “strong evidence for the normal neutrino hierarchy”. *arXiv:1703.04585*.
- Scott, D., and Rees, M. J. (1990). The 21-cm line at high redshift: a diagnostic for the origin of large scale structure. *Mon. Not. R. Astron. Soc.* 247:510.
- Sejersens Riis, A., Hannestad, S., and Weinheimer, C. (2011). Analysis of simulated data for the KArlsruhe TRItium Neutrino experiment using Bayesian inference. *Phys. Rev. C* 84:045503. doi: 10.1103/PhysRevC.84.045503
- Seljak, U., Makarov, A., McDonald, P., Anderson, S., Bahcall, N., Brinkmann, J., et al. (2005). Cosmological parameter analysis including SDSS Ly-alpha forest and galaxy bias: Constraints on the primordial spectrum of fluctuations, neutrino mass, and dark energy. *Phys. Rev. D* 71:103515. doi: 10.1103/PhysRevD.71.103515
- Seljak, U., Slosar, A., and McDonald, P. (2006). Cosmological parameters from combining the Lyman-alpha forest with CMB, galaxy clustering and SN constraints. *JCAP* 10:014. doi: 10.1088/1475-7516/2006/10/014
- Senatore, L., and Zaldarriaga, M. (2017). The effective field theory of large-Scale structure in the presence of massive neutrinos. *arXiv:1707.04698*.
- Shim, J., Lee, J., and Baldi, M. (2014). Breaking the cosmic degeneracy between modified gravity and massive neutrinos with the cosmic web. *arXiv:1404.3639*.
- Shimabukuro, H., Ichiki, K., Inoue, S., and Yokoyama, S. (2014). Probing small-scale cosmological fluctuations with the 21 cm forest: Effects of neutrino mass, running spectral index, and warm dark matter. *Phys. Rev. D* 90:083003. doi: 10.1103/PhysRevD.90.083003
- Shirai, J. (2017). Results and future plans for the KamLAND-Zen experiment. *J. Phys.* 888, 012031. doi: 10.1088/1742-6596/888/1/012031
- Shoji, M., and Komatsu, E. (2010). Massive neutrinos in cosmology: analytic solutions and fluid approximation. *Phys. Rev. D* 81:123516. doi: 10.1103/PhysRevD.81.123516
- Simpson, F., Jimenez, R., Peña-Garay, C., and Verde, L. (2017). Strong evidence for the normal neutrino hierarchy. *JCAP* 06:029. doi: 10.1088/1475-7516/2017/06/029
- Singh, S., and Ma, C.-P. (2003). Neutrino clustering in cold dark matter halos : implications for ultrahigh-energy cosmic rays. *Phys. Rev. D* 67:023506. doi: 10.1103/PhysRevD.67.023506
- Slatyer, T. R., and Wu, C.-L. (2018). Early-universe constraints on dark matter-baryon scattering and their implications for a global 21cm signal. *Phys. Rev. D* 98:023013. doi: 10.1103/PhysRevD.98.023013
- Slosar, A., Iršič, V., Kirkby, D., Bailey, S., Busca, N. G., Delubac, T., et al. (2013). Measurement of baryon acoustic oscillations in the lyman-alpha forest fluctuations in boss data release 9. *JCAP* 04:026. doi: 10.1088/1475-7516/2013/04/026
- Smith, K. M., Hu, W., and Kaplinghat, M. (2006). Cosmological Information from Lensed CMB Power Spectra. *Phys. Rev. D* 74:123002. doi: 10.1103/PhysRevD.74.123002
- Smith, R. E., Peacock, J. A., Jenkins, A., White, S. D. M., Frenk, C. S., Pearce, F. R., et al. (2003). Stable clustering, the halo model and nonlinear cosmological power spectra. *Mon. Not. R. Astron. Soc.* 341:1311. doi: 10.1046/j.1365-8711.2003.06503.x
- Sohn, J., Chon, G., Böhringer, H., Geller, M. J., Diaferio, A., Hwang, H. S., et al. (2018). The hectoMAP cluster survey - II. X-ray clusters. *Astrophys. J.* 855:100. doi: 10.3847/1538-4357/aac7a
- Song, Y.-S., and Knox, L. (2004). Determination of cosmological parameters from cosmic shear data. *Phys. Rev. D* 70:063510. doi: 10.1103/PhysRevD.70.063510
- Spergel, D., Gehrels, N., Baltay, C., Bennett, D., Breckinridge, J., Donahue, M., et al. (2015). Wide-Field Infrared Survey telescope-astrophysics focused telescope assets WFIRST-AFTA 2015 Report. *arXiv:1503.03757*.
- Spergel, D., Gehrels, N., Breckinridge, J., Donahue, M., Dressler, A., Gaudi, B. S., et al. (2013). Wide-field Infrared survey telescope-astrophysics focused telescope assets WFIRST-AFTA Final Report. *arXiv:1305.5422*.
- Spergel, D., Verde, L., Peiris, H. V., Komatsu, E., Nolte, M. R., Bennett, C. L., et al. (2003). First year wilkinson microwave anisotropy probe (WMAP) observations: determination of cosmological parameters. *Astrophys. J. Suppl.* 148, 175–194. doi: 10.1086/377226
- Sprenger, T., Archidiacono, M., Brinckmann, T., Clesse, S., and Lesgourgues, J. (2018). Cosmology in the era of Euclid and the Square Kilometre Array. *arXiv:1801.08331*.
- Springer, P., Bennett, C., and Baisden, P. (1987). Measurement of the neutrino mass using the Inner Bremsstrahlung Emitted in the Electron-Capture Decay of <sup>163</sup>Ho. *Phys. Rev. A* 35, 679–689. doi: 10.1103/PhysRevA.35.679
- Srivastava, R., Ternes, C. A., Tórtola, M., and Valle, J. W. F. (2018a). Testing a lepton quarticity flavor theory of neutrino oscillations with the DUNE experiment. *Phys. Lett. B* 778, 459–463. doi: 10.1016/j.physletb.2018.01.014
- Srivastava, R., Ternes, C. A., Tórtola, M., and Valle, J. W. F. (2018b). Zooming in on neutrino oscillations with DUNE. *Phys. Rev. D* 97:095025. doi: 10.1103/PhysRevD.97.095025
- Stanco, L., Salamanna, G., Sawy, F., and Sirignano, C. (2017). A new way to determine the neutrino mass hierarchy at reactors. *arXiv:1707.07651*.
- Strait, J. et al. (2016). Long-baseline neutrino facility (LBNF) and deep underground neutrino experiment (DUNE). *arXiv:1601.05823*.
- Sutherland, W. (2018). The CMB neutrino mass/vacuum energy degeneracy: a simple derivation of the degeneracy slopes. *Mon. Not. R. Astron. Soc.* 477, 1913–1920. doi: 10.1093/mnras/sty687
- Suzuki, A. et al. (2016). The POLARBEAR-2 and the Simons Array Experiment. *J. Low. Temp. Phys.* 184, 805–810. doi: 10.1007/s10909-015-1425-4
- Takahashi, R., Sato, M., Nishimichi, T., Taruya, A., and Oguri, M. (2012). Revising the halo fit model for the nonlinear matter power spectrum. *Astrophys. J.* 761:152. doi: 10.1088/0004-637X/761/2/152
- Tegmark, M., Eisenstein, D., Strauss, M., Weinberg, D., Blanton, M., Frieman, J., et al. (2006). Cosmological constraints from the SDSS luminous red galaxies. *Phys. Rev. D* 74:123507. doi: 10.1103/PhysRevD.74.123507
- Tegmark, M., Strauss, M., Blanton, M., Abazajian, K., Dodelson, S., Sandvik, H., et al. (2004). Cosmological parameters from SDSS and WMAP. *Phys. Rev. D* 69:103501. doi: 10.1103/PhysRevD.69.103501
- Tegmark, M., and Zaldarriaga, M. (2009). The fast fourier transform telescope. *Phys. Rev. D* 79:083530. doi: 10.1103/PhysRevD.79.083530
- Tereno, I., Schimd, C., Uzan, J.-P., Kilbinger, M., Vincent, F. H., and Fu, L. (2009). CFHTLS weak-lensing constraints on the neutrino masses. *Astron. Astrophys.* 500, 657–665. doi: 10.1051/0004-6361/200811077
- Thakore, T., Devi, M. M., Kumar Agarwalla, S., and Dighe, A. (2018). Active-sterile neutrino oscillations at INO-ICAL over a wide mass-squared range. *J. High Energy Phys.* 08:022. doi: 10.1007/JHEP08(2018)022
- Tingay, S., Goeke, R., Bowman, J. D., Emrich, D., Ord, S. M., Mitchell, D. A., et al. (2013). The Murchison widefield array: the square kilometre array precursor at low radio frequencies. *Publ. Astron. Soc. Austral.* 30:7. doi: 10.1017/pasa.2012.007
- Tinker, J. L., Kravtsov, A. V., Klypin, A., Abazajian, K., Warren, M. S., Yepes, G., et al. (2008). Toward a halo mass function for precision cosmology: the Limits of universality. *Astrophys. J.* 688, 709–728. doi: 10.1086/591439
- Tojeiro, R., Ross, A. J., Burden, A., Samushia, L., Manera, M., Percival, W. J., et al. (2014). The clustering of galaxies in the SDSS-III baryon oscillation spectroscopic survey: galaxy clustering measurements in the low redshift sample of data release 11. *Mon. Not. R. Astron. Soc.* 440, 2222–2237. doi: 10.1093/mnras/stu371
- Tozzi, P., Madau, P., Meiksin, A., and Rees, M. J. (2000). Radio signatures of hi at high redshift: mapping the end of the “dark ages”. *Astrophys. J.* 528:597. doi: 10.1086/308196
- Trotta, R. (2008). Bayes in the sky: bayesian inference and model selection in cosmology. *Contemp. Phys.*, 49, 71–104. doi: 10.1080/00107510802066753
- Upadhye, A., Kwan, J., Pope, A., Heitmann, K., Habib, S., Finkel, H., and Frontiere, N. (2016). Redshift-space distortions in massive neutrino and evolving dark energy cosmologies. *Phys. Rev. D* 93:063515. doi: 10.1103/PhysRevD.93.063515

- Vagnozzi, S., Dhawan, S., Gerbino, M., Freese, K., Goobar, A., and Mena, O. (2018). Constraints on the sum of the neutrino masses in dynamical dark energy models with  $w(z) \geq -1$  are tighter than those obtained in  $\Lambda$ CDM. *arXiv:1801.08553*.
- Vagnozzi, S., Giusarma, E., Mena, O., Freese, K., Gerbino, M., Ho, S., and Lattanzi, M. (2017). Unveiling  $\nu$  secrets with cosmological data: neutrino masses and mass hierarchy. *Phys. Rev. D* 96:123503. doi: 10.1103/PhysRevD.96.123503
- van Haarlem, M., Wise, M. W., Gunst, A. W., Heald, G., McKean, J. P., Hessels, J. W. T., et al. (2013). LOFAR: the low-frequency ARray. *Astron. Astrophys.* 556:A2. doi: 10.1051/0004-6361/201220873
- Vargas-Magaña, M. et al. (2018). The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: theoretical systematics and Baryon Acoustic Oscillations in the galaxy correlation function. *Mon. Not. R. Astron. Soc.* 477, 1153–1188. doi: 10.1093/mnras/sty571
- Vergados, J., Ejiri, H., and Simkovic, F. (2012). Theory of neutrinoless double beta decay. *Rept. Prog. Phys.* 75:106301. doi: 10.1088/0034-4885/75/10/106301
- Vergados, J., Ejiri, H., and Šimkovic, F. (2016). Neutrinoless double beta decay and neutrino mass. *Int. J. Mod. Phys. E* 25:1630007. doi: 10.1142/S0218301316300071
- Viel, M., Becker, G. D., Bolton, J. S., and Haehnelt, M. G. (2013). Warm dark matter as a solution to the small scale crisis: new constraints from high redshift Lyman- $\alpha$  forest data. *Phys. Rev. D* 88:043502. doi: 10.1103/PhysRevD.88.043502
- Viel, M., Haehnelt, M. G., and Springel, V. (2010). The effect of neutrinos on the matter distribution as probed by the intergalactic medium. *JCAP* 06:015. doi: 10.1088/1475-7516/2010/06/015
- Villaescusa-Navarro, F. et al. (2014a). Cosmology with massive neutrinos I: towards a realistic modeling of the relation between matter, haloes and galaxies. *JCAP* 03:011. doi: 10.1088/1475-7516/2014/03/011
- Villaescusa-Navarro, F., Bull, P., and Viel, M. (2015). Weighing neutrinos with cosmic neutral hydrogen. *Astrophys. J.* 814:146. doi: 10.1088/0004-637X/814/2/146
- Villaescusa-Navarro, F., Viel, M., Datta, K. K., and Choudhury, T. R. (2014b). Modeling the neutral hydrogen distribution in the post-reionization Universe: intensity mapping. *JCAP* 09:050. doi: 10.1088/1475-7516/2014/09/050
- Villaescusa-Navarro, F., Vogelsberger, M., Viel, M., and Loeb, A. (2013). Neutrino signatures on the high transmission regions of the Lyman-alpha forest. *Mon. Not. R. Astron. Soc.* 431:3670. doi: 10.1093/mnras/stt452
- Visbal, E., Loeb, A., and Wyithe, J. B. (2009). Cosmological constraints from 21cm surveys after reionization. *JCAP* 10:030. doi: 10.1088/1475-7516/2009/10/030
- Vogel, P. (2015). How difficult it would be to detect cosmic neutrino background? *AIP Conf. Proc.* 1666, 140003. doi: 10.1063/1.4915587
- Wang, G., Chang, C. L., Yefremenko, V., Ding, J., Novosad, V., Bucci, C., et al. (2015). CUPID: CUORE (Cryogenic underground observatory for rare events) upgrade with particle IDentification. *arXiv:1504.03599*.
- Wang, L.-M., Caldwell, R., Ostriker, J., and Steinhardt, P. J. (2000). Cosmic concordance and quintessence. *Astrophys. J.* 530, 17–35. doi: 10.1086/308331
- Wang, S., and Xia, D.-M. (2018). Constraints on the sum of neutrino masses from Bayesian analysis of the latest cosmological data. *Chin. Phys. C* 42:065103.
- Weiler, T. J. (1982). Resonant absorption of cosmic ray neutrinos by the relic neutrino background. *Phys. Rev. Lett.* 49:234. doi: 10.1103/PhysRevLett.49.234
- Weiler, T. J. (1984). Big bang cosmology, relic neutrinos, and absorption of neutrino cosmic rays. *Astrophys. J.* 285:495. doi: 10.1086/162524
- Weinberg, D. H., Mortonson, M. J., Eisenstein, D. J., Hirata, C., Riess, A. G., and Rozo, E. (2013). Observational probes of cosmic acceleration. *Phys. Rept.* 530, 87–255. doi: 10.1016/j.physrep.2013.05.001
- Weinberg, S. (1962). Universal neutrino degeneracy. *Phys. Rev.* 128, 1457–1473. doi: 10.1103/PhysRev.128.1457
- Weinberg, S. (1978). A new light boson? *Phys. Rev. Lett.* 40, 223–226. doi: 10.1103/PhysRevLett.40.223
- Wetterich, C. (1995). The Cosmon model for an asymptotically vanishing time dependent cosmological 'constant'. *Astron. Astrophys.*, 301:321–328.
- Wilczek, F. (1978). Problem of strong p and t invariance in the presence of instantons. *Phys. Rev. Lett.* 40, 279–282. doi: 10.1103/PhysRevLett.40.279
- Winter, W. (2016). Atmospheric neutrino oscillations for earth tomography. *Nucl. Phys. B* 908, 250–267. doi: 10.1016/j.nuclphysb.2016.03.033
- Witte, S., Villanueva-Domingo, P., Gariazzo, S., Mena, O., and Palomares-Ruiz, S. (2018). EDGES result versus CMB and low-redshift constraints on ionization histories. *Phys. Rev. D*, 97:103533. doi: 10.1103/PhysRevD.97.103533
- Wolfenstein, L. (1978). Neutrino oscillations in matter. *Phys. Rev. D* 17, 2369–2374. doi: 10.1103/PhysRevD.17.2369
- Wyithe, S., Loeb, A., and Geil, P. (2008). Baryonic acoustic oscillations in 21cm emission: a probe of dark energy out to high redshifts. *Mon. Not. R. Astron. Soc.*, 383:1195. doi: 10.1111/j.1365-2966.2007.12631.x
- Yañez, J., and Kouchner, A. (2015). Measurement of atmospheric neutrino oscillations with very large volume neutrino telescopes. *Adv. High Energy Phys.*, 2015:271968. doi: 10.1155/2015/271968
- Yang, Y. (2018). The contributions of dark matter annihilation to the global 21cm spectrum observed by the EDGES experiment. *arXiv:1803.05803*.
- Yeche, C., Palanque-Delabrouille, N., Baur, J., and BourBoux, H. d. M. d. (2017). Constraints on neutrino masses from Lyman-alpha forest power spectrum with bOSS and XQ-100. *JCAP* 06:047. doi: 10.1088/1475-7516/2017/06/047
- Zhang, J., and Zhang, X. (2018). Gravitational clustering of cosmic relic neutrinos in the milky way. *Nat. Commun.* 9:1833. doi: 10.1038/s41467-018-04264-y
- Zhao, G.-B., Saito, S., Percival, W. J., Ross, A. J., Montesano, F., Viel, M., et al. (2013). The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: weighing the neutrino mass using the galaxy power spectrum of the CMASS sample. *Mon. Not. R. Astron. Soc.* 436, 2038–2053. doi: 10.1093/mnras/stt1710
- Zhao, J., Wen, L.-J., Wang, Y.-F., and Cao, J. (2017). Physics potential of searching for  $0\nu\beta\beta$  decays in JUNO. *Chin. Phys. C* 41:053001. doi: 10.1088/1674-1137/41/5/053001
- Zlatev, I., Wang, L.-M., and Steinhardt, P. J. (1999). Quintessence, cosmic coincidence, and the cosmological constant. *Phys. Rev. Lett.* 82, 896–899. doi: 10.1103/PhysRevLett.82.896

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 de Salas, Gariazzo, Mena, Ternes and Tórtola. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# An Asymptotically Safe Guide to Quantum Gravity and Matter

Astrid Eichhorn\*

*Institut für Theoretische Physik, Universität Heidelberg, Heidelberg, Germany*

Asymptotic safety generalizes asymptotic freedom and could contribute to understanding physics beyond the Standard Model. It is a candidate scenario to provide an ultraviolet extension for the effective quantum field theory of gravity through an interacting fixed point of the Renormalization Group. Recently, asymptotic safety has been established in specific gauge-Yukawa models in four dimensions in perturbation theory, providing a starting point for asymptotically safe model building. Moreover, an asymptotically safe fixed point might even be induced in the Standard Model under the impact of quantum fluctuations of gravity in the vicinity of the Planck scale. This review contains an overview of the key concepts of asymptotic safety, its application to matter and gravity models, exploring potential phenomenological implications and highlighting open questions.

## OPEN ACCESS

### Edited by:

Francesco Sannino,  
University of Southern Denmark,  
Denmark

### Reviewed by:

Borut Bajc,  
Jožef Stefan Institute (IJS), Slovenia  
Roberto Percacci,  
Scuola Internazionale Superiore di  
Studi Avanzati (SISSA), Italy

### \*Correspondence:

Astrid Eichhorn  
eichhorn@thphys.uni-heidelberg.de

### Specialty section:

This article was submitted to  
High-Energy and Astroparticle  
Physics,  
a section of the journal  
Frontiers in Astronomy and Space  
Sciences

**Received:** 17 October 2018

**Accepted:** 06 December 2018

**Published:** 22 January 2019

### Citation:

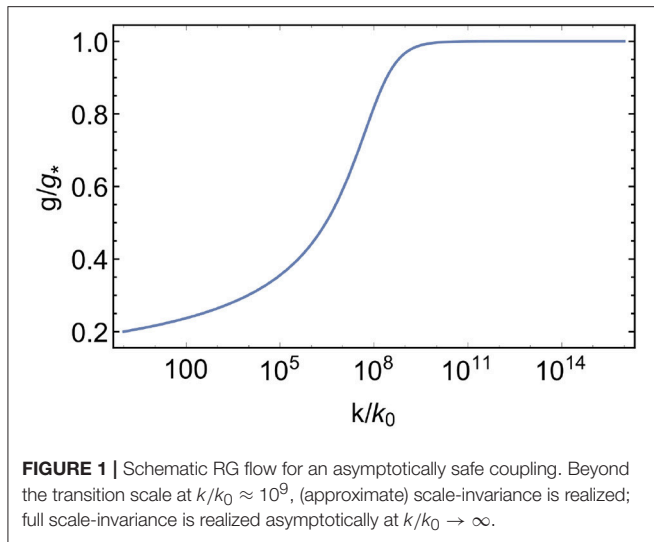
Eichhorn A (2019) An Asymptotically  
Safe Guide to Quantum Gravity and  
Matter.  
Front. Astron. Space Sci. 5:47.  
doi: 10.3389/fspas.2018.00047

**Keywords:** quantum gravity, beyond standard model, renormalization group, asymptotic safety, gauge theories

## 1. INVITATION TO ASYMPTOTIC SAFETY

Asymptotic safety (Weinberg, 1980) is a quantum-field theoretic paradigm providing an ultraviolet (UV) extension or completion for effective field theories. The high-momentum regime of an asymptotically safe theory is scale invariant, cf. **Figure 1**. It is governed by a fixed point of the Renormalization Group (RG) flow of couplings. As such, asymptotic safety is an example of a fruitful transfer of ideas from statistical physics to high-energy physics: In the former, interacting RG fixed points provide universality classes for continuous phase transitions (Wilson and Fisher, 1972; Zinn-Justin, 2002), in the latter these generalize asymptotic freedom to a scale-invariant UV completion with residual interactions. This paradigm is being explored for physics beyond the Standard Model in several promising ways. Following the discovery of perturbative asymptotic safety in weakly-coupled gauge-Yukawa models in four dimensions (Litim and Sannino, 2014), the search for asymptotically safe extensions of the Standard Model with new degrees of freedom close to the electroweak scale is ongoing. Mechanisms for asymptotic safety also exist in nonrenormalizable settings, making it a candidate paradigm for quantum gravity (Weinberg, 1980; Reuter, 1998). After the discovery of the Higgs boson (Aad, 2012; Chatrchyan, 2012), we know that the Standard Model can consistently be extended up to the Planck scale (Bezrukov et al., 2012; Buttazzo et al., 2013; Bezrukov and Shaposhnikov, 2015). Hence, the interplay of the Standard Model with quantum fluctuations of gravity within a quantum field theoretic setting is under active exploration.

This review aims at providing an introduction to asymptotic safety for non-experts, highlighting mechanisms that generate asymptotically safe physics, explaining how these could play a role in settings relevant for high-energy physics and discussing open questions of (potentially) asymptotically safe models. An extensive bibliography is intended to serve as a guide to further reading, providing more comprehensive and in-depth answers to many points only touched upon briefly here.



## 2. ASYMPTOTIC SAFETY - THE KEY IDEA

Quantum fluctuations induce a momentum-scale dependence in the couplings of a model, breaking scale invariance even in classically scale-invariant models. Scale invariance is restored at RG fixed points. These can be non-interacting, in which case the theory is asymptotically free, or interacting in at least one of the couplings, in which case the theory is asymptotically safe. Both fixed points underlie theories that are fundamental in a Wilsonian sense: For a theory that is discretized, e.g., on a lattice, an RG fixed point guarantees that a continuum limit exists. Scale-invariance protects the running couplings in a model from Landau poles which can signal a breakdown of a description of an interacting system by this model because of triviality<sup>1</sup>. Hence, the introduction of new physics is one viable theoretical option instead of a necessity.

Scale-invariance requires a fixed point in the dimensionless couplings  $g_i$ , obtained from their dimensionful counterparts  $\bar{g}_i$  with canonical dimension  $d_{\bar{g}_i}$  by multiplication with an appropriate power of the RG scale  $k$

$$g_i(k) = \bar{g}_i(k) k^{-d_{\bar{g}_i}}. \quad (1)$$

A scale-invariant point is a zero of all beta functions,

$$\beta_{g_i} = k \partial_k g_i(k) = 0 \quad \text{at} \quad g_i = g_{i*}. \quad (2)$$

<sup>1</sup>Models affected by the triviality problem can only hold up to arbitrarily high momentum scales if the coupling vanishes at all scales, rendering the models noninteracting, or trivial. Establishing triviality requires going beyond perturbation theory. Nonetheless, an intuitive understanding of the problem can be gained from perturbation theory, e.g., in scalar  $\lambda \phi^4$  theory in four dimensions. The one-loop beta function for the quartic self-interaction  $\lambda$  is  $\beta_\lambda = \# \lambda^2$ , where  $\# > 0$  holds. Integrating the beta function leads to a logarithmic divergence. Pushing the scale  $\Lambda$  of the divergence (the Landau pole) to infinity requires  $\lambda_0 = \lambda(k_0) = 0$ , since  $\Lambda/k_0 = e^{\frac{1}{\lambda_0}}$ .

Then, dimensionful couplings<sup>2</sup> scale with their canonical dimensionality, i.e.,  $\bar{g}_i(k) \sim k^{d_{\bar{g}_i}}$ , since  $g_i(k) = g_{i*} = \text{const}$  in a scale-invariant regime. This must hold for all couplings in the infinite-dimensional theory space, spanned by all interactions allowed by symmetries, including higher-order, i.e., canonically irrelevant interactions. Quantum fluctuations generically generate all interactions, as familiar from effective field theories. Moreover, there is no a priori physical argument to exclude higher-order terms from the dynamics. The restriction to perturbatively renormalizable terms that is commonly assumed is actually an automatic consequence of the universality class of the Gaussian, i.e., free fixed point which renders higher-order terms irrelevant for perturbative low-energy physics.

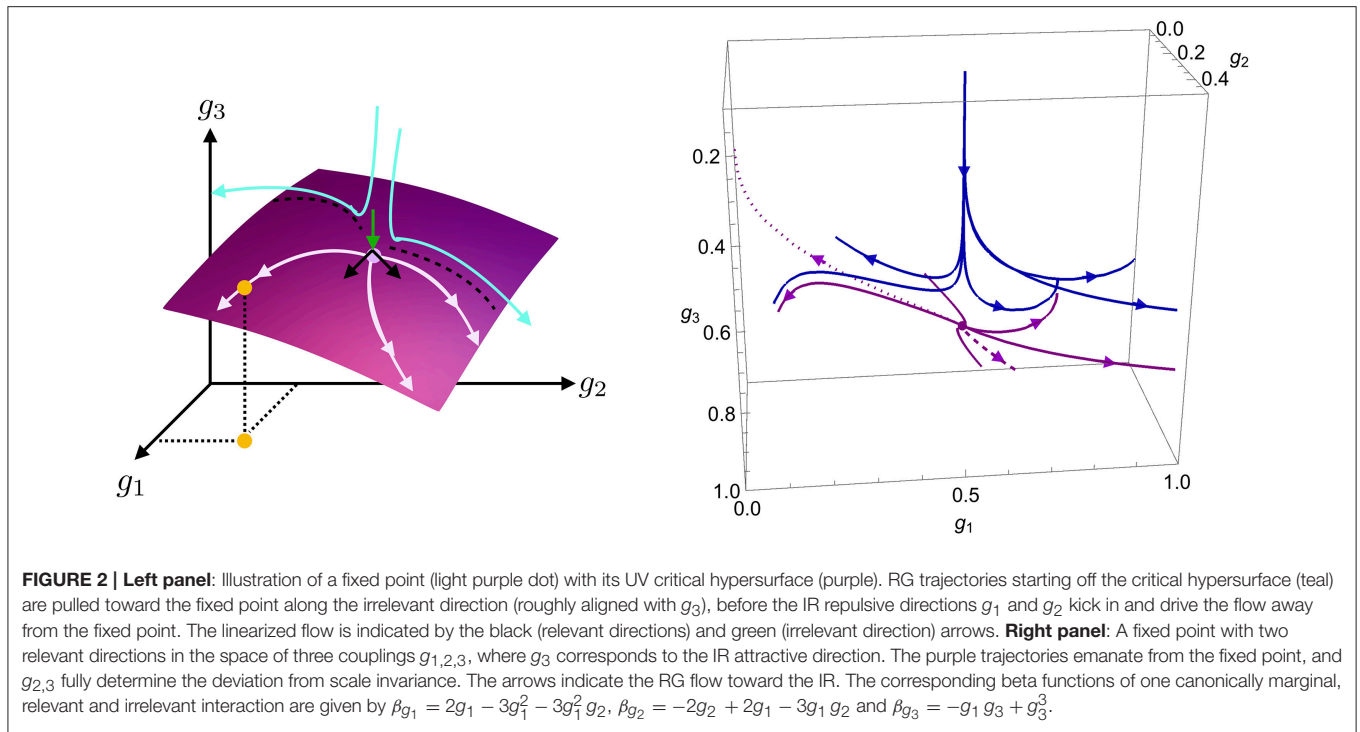
### 2.1. Predictivity in the Infinite-Dimensional Space of Couplings

The main consequence of an RG fixed point is not that it provides a fundamental theory—after all, experiments are limited to finite scales—but that it generates universal predictions for low-energy physics. It imposes relations between the couplings encoding the location of the UV-critical surface of the fixed point. This hypersurface is spanned by all couplings along which RG trajectories emanate from the fixed point as one lowers  $k$  toward the infrared (IR). The corresponding *relevant* directions parameterize the deviation from scale invariance, i.e., the flow toward the IR can only deviate from the fixed point along the relevant directions. They constitute free parameters, as a range of values of relevant couplings can be reached along different trajectories emanating from the fixed point, cf. **Figure 2**. It can be more intuitive to understand that a free parameter is associated to such a direction, as IR-repulsiveness equals UV-attractivity. Irrespective of its IR value, a UV-attractive coupling reaches the fixed point at high scales as one reverses the flow toward the UV. (Nevertheless, recall that although we measure physics at low energies and try to extrapolate toward viable UV physics, nature works the other way: IR physics emerges as a consequence of UV physics).

Toward the IR, the irrelevant, i.e., IR-attractive directions are automatically pulled toward the fixed point, cf. **Figure 2**. Accordingly, no free parameter is associated to them – this is the universality-generating property of an RG fixed point: Initializing the RG flow at some scale  $k_0$ , the flow maps a UV interval of values for an irrelevant coupling at  $k_0$  to a much smaller IR interval. The latter shrinks to zero as one takes  $k_0 \rightarrow \infty$ . As a one-coupling example with an IR attractive fixed point, consider

$$\beta_g = g(g - g_*), \quad (3)$$

<sup>2</sup>Weinberg (1980) motivates the focus on dimensionless couplings  $g_i$  instead of their dimensionful counterparts  $\bar{g}_i$  by requiring finiteness of observables. Measurable quantities at some energy scale  $E$ , e.g., a scattering cross-section  $\sigma$ , can be written as  $\sigma = E^\# f(g_i)$ , where  $\#$  is the canonical dimension of  $\sigma$ , multiplied by a function of the *dimensionless* couplings  $g_i$  that enter. Herein the RG scale is equated to a physical energy scale. If the dimensionless couplings diverge at a finite energy scale, this typically entails divergences in physical quantities.



with the solution

$$g(k) = \frac{g_*}{1 + \left(\frac{k}{k_0}\right)^{g_*} \left(\frac{g_*}{g_0} - 1\right)}, \quad (4)$$

where  $g(k_0) = g_0$ . As the initial scale  $k_0 \rightarrow \infty$ ,  $g(k) \rightarrow g_*$ . For a finite  $k_0$ , the difference  $g(k) - g_*$  goes to zero as  $k/k_0 \rightarrow 0$ . For a trajectory that emanates from the fixed point, there is no freedom of choice for the value for an irrelevant direction: the fixed-point requirement restricts the flow to lie within the critical hypersurface, resulting in completely determined values for the irrelevant directions. For instance, at the free fixed point, higher-order couplings do not play a role in the IR: the RG flow drives them toward zero for all perturbative initial conditions in the UV. This generates universality and independence of the IR physics from the UV physics in all but the (marginally) relevant couplings.

To determine the set of IR-repulsive (= UV attractive) directions, it suffices to examine the linearized flow about the fixed point<sup>3</sup> at  $\vec{g} = \vec{g}_*$ ,

$$\beta_{g_i} = \sum_j \frac{\partial \beta_{g_i}}{\partial g_j} \Big|_{\vec{g}=\vec{g}_*} (g_j - g_{j*}) + \mathcal{O}(g_j - g_{j*})^2. \quad (5)$$

In terms of the critical exponents<sup>4</sup>

$$\theta_I = -\text{eig} \mathcal{M}_{ij} = -\text{eig} \frac{\partial \beta_{g_i}}{\partial g_j} \Big|_{\vec{g}=\vec{g}_*}, \quad (6)$$

<sup>3</sup>To determine the basin of attraction of the fixed point, one numerically integrates the RG flow to generate full trajectories.

<sup>4</sup>The opposite sign convention is sometimes used in the literature.

and corresponding (right) eigenvectors  $V^I$ , the solution to Equation (5) is

$$g_i(k) = g_{i*} + \sum_I c_I V_i^I \left(\frac{k}{k_0}\right)^{-\theta_I}. \quad (7)$$

$k_0$  is an arbitrary reference scale and  $c_I$  are constants of integration. Typically, the set of couplings  $\vec{g}$  does not diagonalize the stability matrix  $\mathcal{M}_{ij}$  at  $\vec{g} = \vec{g}_*$  and the eigenvectors  $V^I$  are superpositions. As the stability matrix need not be symmetric, the eigenvalues need not be real. Their imaginary part results in a spiraling behavior of the flow in the vicinity of the fixed point, where the real part determines whether the spiraling is inwards or outwards. To determine the set of free parameters, it therefore suffices to consider the real parts. For the following discussion we will thus assume that the eigenvalues are real. For  $\theta_I > 0$ , the corresponding eigenvector  $V^I$  constitutes an IR repulsive direction: Toward the IR, the distance to the fixed-point regime grows, and the IR values of couplings appearing in  $V^I$  depend on  $c_I$ . Fixing this free parameter requires experimental input. Accordingly, predictivity requires a finite number of directions with  $\theta_I > 0$ .

In contrast, for  $\theta_I < 0$ , the IR values of couplings are independent of the corresponding  $c_I$ , cf. Equation (7): For  $\theta_I < 0$ , any deviation from the fixed-point value in  $V^I$  is washed out by the RG flow to the IR. Once a choice of coordinates in theory space is made, the critical hypersurface typically exhibits curvature. If the critical hypersurface had no curvature, the values of irrelevant couplings would be constant. Curvature of the critical hypersurface generates a scale dependence which is completely fixed once the values of all relevant couplings are

specified, cf. **Figure 2**. The values of the corresponding irrelevant couplings depend on the scale, but not independently of the relevant couplings, cf. green trajectory in **Figure 3**.

For asymptotic safety, the finite contribution to the  $\theta_I$  due to residual interactions at  $\bar{g}_*$  shifts the critical exponents away from the canonical dimensions of couplings, e.g.,

$$\theta_i = -\frac{\partial \beta_i}{\partial g_i} \Big|_{g_i=\bar{g}_*} = -\frac{\partial}{\partial g_i} (\partial_t \bar{g}_i k^{-d_{g_i}}) \Big|_{g_i=\bar{g}_*} = d_{\bar{g}_i} g_{i*} + \eta(g_{i*}^2), \quad (8)$$

for a coupling  $g_i$  that is an eigendirection of  $\mathcal{M}_{ij}$ . This can enhance the predictive power of interacting over free fixed points.

The interpretation of asymptotic safety as a way of imposing predictivity on a model specified by its field content and symmetries is crucial in the context of quantum gravity. The simplest interpretation of the Planck scale suggests that it acts as a minimal length, inducing discreteness for quantum-gravity models at the kinematical level. This might suggest that one need not search for a continuum limit in quantum gravity. Yet, by requiring a continuum limit one restricts the dynamics to a trajectory within the critical surface, leaving just a finite number of free parameters to determine the dynamics *at all scales*. In the presence of an explicit cutoff scale, the microscopic dynamics might be defined anywhere in the theory space, requiring specification of an infinite number of couplings for the UV dynamics (see also Eichhorn, 2018b). Similarly, predictivity *at high scales* breaks down in effective field theories. Moreover, *physical* discreteness can arise in quantum gravity even in a continuum theory, through the *dynamical* emergence of a scale (see e.g., Reuter and Schwindt, 2006; Percacci and Vacca, 2010), or through discreteness in the spectra of operators (Rovelli and Smolin, 1995; Ashtekar and Lewandowski, 1997).

## 2.2. Asymptotic Safety in a Nutshell

The development of the Standard Model was based on the principle of renormalizability. This is one way of implementing predictivity, i.e., constructing a low-energy theory with a finite number of free parameters. Yet, as, e.g.,  $\phi^4$  theory in 4 dimensions highlights, a perturbatively renormalizable theory is not guaranteed to exist as a fundamental theory in the Wilsonian sense, due to the triviality problem. Analogously, the Standard Model is actually expected to be an effective low-energy theory. Asymptotic safety is a paradigm that combines the requirement of predictivity with the possibility of obtaining a fundamental theory through an RG fixed point at high momenta with a finite number of relevant directions. The fixed point ensures nonperturbative renormalizability, while the finite dimensionality of the critical hypersurface guarantees predictivity of the model.

## 2.3. Non-fundamental Asymptotic Safety

Instead of providing a “fundamental” UV completion, asymptotic safety might serve as one step forward in our understanding of microscopic physics, with more fundamental physics to be discovered beyond. While providing a UV completion for some RG trajectories, a fixed point can simultaneously act as an IR attractor for a more fundamental

description. This follows, as a fixed point’s UV repulsive directions correspond to its IR attractive directions. Hence, it is a misconception that a fixed point is either UV or IR - whether trajectories emanate from it in the UV, or approach it in the IR depends on the *initial conditions* for the RG flow. Given *two* fixed points connected by an RG trajectory, the distinction into a UV and an IR fixed point (which is also expected to satisfy the *a-theorem* Cardy, 1988) follows from the trajectory.

For specificity, assume that a cutoff scale  $k_{UV}$  exists, such that for  $k > k_{UV}$  a (quasilocal) quantum field theoretic description is impossible or requires additional fields and/or symmetries. At  $k \leq k_{UV}$ , the dynamics can be described in the asymptotically safe theory space. Initial conditions for the RG flow are determined by the underlying fundamental model at  $k = k_{UV}$ . If they lie close to or on the *IR-critical surface* of a fixed point, the flow is attracted toward the fixed point along its IR-attractive directions. The flow is actually driven toward the *UV-critical surface*, cf. purple trajectories in **Figure 3**. Trajectories can even spend a large amount of RG “time” close to the fixed point. At  $k_{trans} < k_{UV}$  the effect of the IR-repulsive directions kicks in and the flow is driven away from the fixed point along its IR-repulsive directions. This trajectory will result in IR-values of couplings close to those of a “true” fixed-point trajectory (cf. **Figure 3** see Percacci and Vacca, 2010). The above is nothing but a detailed account of how a fixed point generates IR universality. Thus, asymptotically safe fixed points could generate universal IR predictions, even in the presence of  $k_{UV}$ .

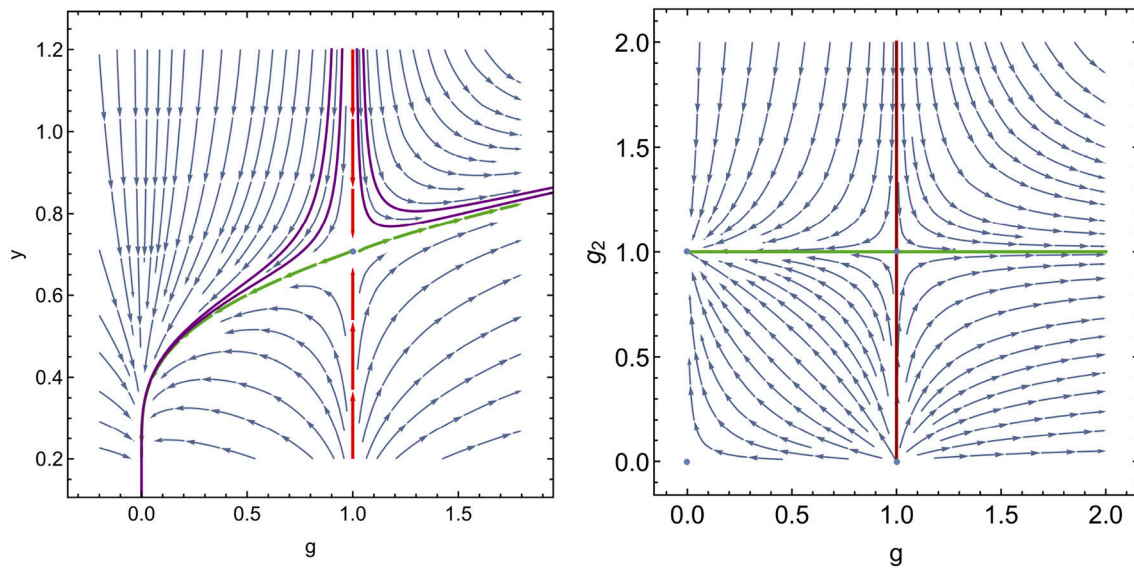
## 2.4. Mechanisms for and Selected Examples of Asymptotic Safety

A special case of an RG fixed point is that of an asymptotically free one. To generate it, antiscreening contributions have to dominate in the beta function of the respective coupling. In contrast, asymptotic safety is generated by several different mechanisms and can be realized both in the perturbative and the nonperturbative regime, i.e., with near-Gaussian or far-from-Gaussian critical exponents. As a second key difference, an interacting fixed point allows to combine finite, predictable IR values of couplings with UV completeness. For the free fixed point, finite IR values typically require the corresponding coupling to be an IR repulsive direction, i.e., relevant. This negates the possibility to predict the value of the coupling which remains a free parameter based on the free fixed point alone. (Of course, an interacting fixed point can dominate the flow in the IR, at which the coupling in question could be IR attractive. In this case it is again the universality class of the interacting fixed point which provides a prediction for a finite value of a coupling).

### 2.4.1. Canonical Scaling vs. Quantum Effects

This mechanism is available for couplings which are asymptotically free in their critical dimension  $d_{crit}$ , where they are dimensionless, i.e., their one-loop beta function is given by

$$\beta_{g_i} \Big|_{d=d_{crit}} = \beta_1 g_i^\#, \quad (9)$$



**FIGURE 3 | Left panel:** The beta functions  $\beta_g = 2g - 2g^2$  and  $\beta_y = -gy + 2y^3$  feature a fixed point at  $g = 1$ ,  $y = 1/\sqrt{2}$  that has one UV attractive and one IR attractive direction. The UV critical surface is indicated in green, the IR critical surface in red. The RG flow toward the IR is attracted toward the UV critical surface, such that the relation between  $g$  and  $y$  that parameterizes the UV critical surface is approximately realized also for trajectories (in purple) that start off the UV critical surface. **Right panel:** The flow described by  $\beta_g = 2g - 2g^2$  and  $\beta_{g_2} = -2g_2 + 2g_2^2$  features a fixed point at  $g = g_2 = 1$ , which is IR attractive in  $g_2$  and where the UV critical surface has no curvature. Therefore,  $g_2(k) = 1$  for the trajectories emanating from this fixed point.

with  $\beta_1 < 0$  and  $\# = 2, 3$ . In  $d = d_{\text{crit}} + \epsilon$ , the coupling is dimensionful,  $g_i = \bar{g}_i k^{c\epsilon}$ , where  $c > 0$  depends on the coupling under consideration. For  $\epsilon \ll 1$ , the one-loop beta function reads

$$\beta_{g_i} \Big|_{d=d_{\text{crit}}+\epsilon} = c\epsilon g_i + \beta_1 g_i^\# . \quad (10)$$

An interacting fixed point lies at

$$g_i^* = \left( -\frac{c\epsilon}{\beta_1} \right)^{1/(\#-1)} . \quad (11)$$

This mechanism is realized in Yang-Mills theory in  $d = 4 + \epsilon$  (Peskin, 1980), nonlinear sigma models in  $d = 2 + \epsilon$  (Polyakov, 1975; Bardeen et al., 1976; Friedan, 1980; Higashijima and Itou, 2003; Codello and Percacci, 2009; Fabbrichesi et al., 2011) and the Gross-Neveu model in  $d = 2 + \epsilon$  (Gawedzki and Kupiainen, 1985; Kikukawa and Yamawaki, 1990; de Calan et al., 1991; He et al., 1992; Hands et al., 1993; Braun et al., 2011).

For Yang-Mills theory, the  $\epsilon$ -expansion has been extended up to fourth order, indicating a fixed point in  $d = 5$  (Morris, 2005), cf. **Figure 4**, corroborating functional RG results (Gies, 2003), in contrast to lattice results (Knechtli and Rinaldi, 2016). For instance, consider SU(3) Yang-Mills, cf. **Figure 4**. The  $\epsilon$  expansion in Morris (2005) yields for  $\tilde{\alpha} = \frac{6}{(4\pi)^2} g^2$

$$\beta_{\tilde{\alpha}} = \epsilon \tilde{\alpha} - b_1 \tilde{\alpha}^2 - b_2 \tilde{\alpha}^3 - b_3 \tilde{\alpha}^4 - b_4 \tilde{\alpha}^5, \quad (12)$$

$$b_1 = 3.67, \quad b_2 = 5.67, \quad b_3 = 13.23, \quad b_4 = 39.43 + \frac{51.22}{9}, \quad (13)$$

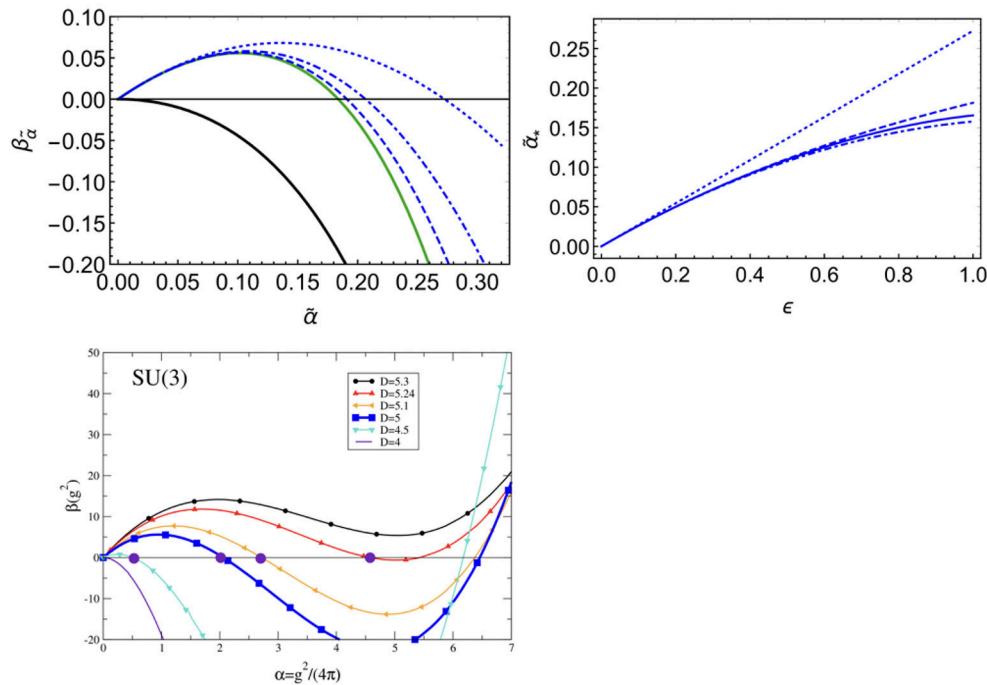
resulting in

$$\tilde{\alpha}_* = 0.272\epsilon - 0.115\epsilon^2 + 0.024\epsilon^3 - 0.016\epsilon^4. \quad (14)$$

Couplings which are marginally irrelevant in their critical dimension  $d_c$  can achieve interacting fixed points for  $d < d_c$ , where they correspond to irrelevant directions. In contrast to the case in  $d_c$ , where the free fixed point results in a vanishing coupling at all scales in order to be a UV fixed point (triviality problem), in  $d = d_c - \epsilon$ , the interacting fixed point requires a unique finite value of the coupling in the IR, corresponding to the fixed-point value, unless the UV critical hypersurface is curved. Thus, the interacting theory is UV complete for one unique value of the coupling. Conversely, asymptotically free trajectories reach the interacting fixed point in the IR.

For instance, for scalar theories, the marginally irrelevant nature of the quartic coupling in  $d = 4$  implies the existence of a fixed point in  $d = 4 - \epsilon$ . The well-known Wilson-Fisher fixed point is IR attractive in the quartic coupling (Wilson and Fisher, 1972) and serves as the IR endpoint of an asymptotically free trajectory. It has been characterized with various methods (Guida and Zinn-Justin, 1998; Campostrini et al., 1999; Pelissetto and Vicari, 2002; Canet et al., 2003; Litim and Zappala, 2011; El-Showk et al., 2012, 2014; Gliozzi and Rago, 2014) and serves as a benchmark example for many techniques. For  $d > 4$ , a possible fixed point (Fei et al., 2014) lies at negative quartic coupling, appearing to be at odds with a stable microscopic potential (Percacci and Vacca, 2014; Eichhorn et al., 2016).

Fixed points generated by such a mechanism are weakly coupled at small  $\epsilon$ , where the critical exponent is equal to minus the canonical dimension.



**FIGURE 4 | Left upper panel:** Based on results in Morris (2005), the beta function for SU(3) Yang-Mills theory in the epsilon expansion for  $\tilde{\alpha} = \frac{6}{(4\pi)^2} g^2$  for  $\epsilon = 0$  (black line),  $\epsilon = 1$  at one loop (dotted), two loops (dot-dashed), three loops (dashed) and four loops (green). **Right upper panel:** Fixed-point value for  $\alpha$  as a function of  $\epsilon$  up to  $\epsilon$  (dotted),  $\epsilon^2$  (dot-dashed),  $\epsilon^3$  (dashed), and  $\epsilon^4$  (continuous) emerge from the free fixed point at  $d \rightarrow d_{\text{crit}}$ , i.e.,  $\epsilon \rightarrow 0$ . **Lower panel:** Results from the FRG calculation taken from Gies (2003) for  $\alpha = \frac{g^2}{4\pi}$ .

A key example is gravity: Slightly above its critical dimension  $d = 2$ , where the Einstein action is purely topological, the beta function for the dimensionless Newton coupling  $G = G_N k^{d-2}$  at one loop reads (Weinberg, 1980; Gastmans et al., 1978; Christensen and Duff, 1978),

$$\beta_G = \epsilon G - \beta_1 G^2, \text{ such that } G_* = \frac{\epsilon}{\beta_1}, \quad \theta = -\epsilon, \quad (15)$$

where  $\beta_1$  depends on the parameterization of metric fluctuations  $h_{\mu\nu}$  around a background  $\bar{g}_{\mu\nu}$ . Note that in these calculations the Jacobian that arises in the path-integral measure from relating the different parameterizations is not taken into account. Specifically the functional RG in the limit  $d \rightarrow 2$  yields (Percacci and Vacca, 2015)

$$\beta_1 = -\frac{2(19 - 38\beta + 13\beta^2)}{3(1 - \beta^2)}, \quad (16)$$

for the linear parameterization  $g_{\mu\nu} = \bar{g}_{\mu\nu} + h_{\mu\nu}$ ,

$$\beta_1 = -\frac{2(25 - 38\beta + 19\beta^2)}{3(1 - \beta)^2}, \quad (17)$$

for the exponential parameterization  $g_{\mu\nu} = \bar{g}_{\mu\nu} \exp[h_{\mu\nu}]^K$ , where  $\beta$  is a gauge parameter, such that for  $\beta \rightarrow 0$  the result  $\beta_1 = -38/3$  is found (Tsao, 1977; Brown, 1977; Kawai and Ninomiya,

1990; Jack and Jones, 1991) for the linear parameterization and  $\beta_1 = -50/3$  for the exponential parameterization (David, 1988; Distler and Kawai, 1989; Kawai et al., 1993a,b, 1996; Aida et al., 1994; Nishimura et al., 1994; Aida and Kitazawa, 1997; Codello and D'Odorico, 2015). A continuous extension to  $d = 4$  might be possible (Falls, 2015, 2017).

#### 2.4.2. One-Loop vs. Higher-Loop

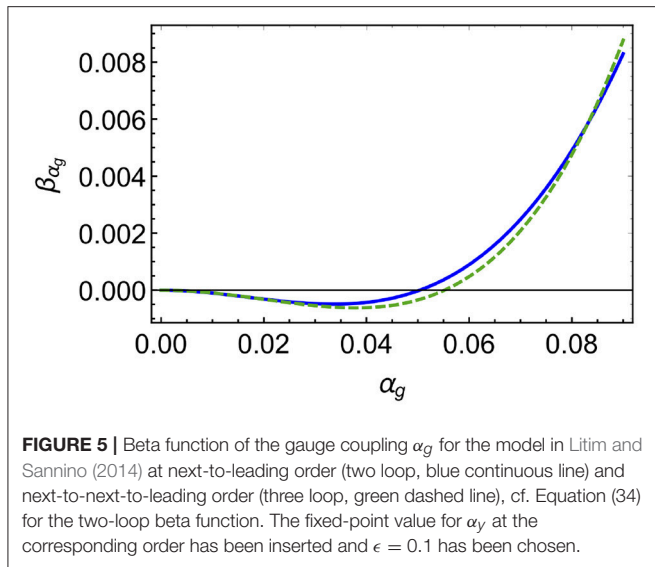
In perturbation theory, the signs of the one-loop and two-loop coefficients can differ, leading to a cancellation at a finite fixed-point value, schematically

$$\beta_{g_i} = \beta_1 g_i^{\#1} + \beta_2 g_i^{\#2} + \dots \quad (18)$$

with

$$\beta_{g_i} \Big|_{g_i=g_{i*}} = 0, \quad g_{i*} = \left( \frac{-\beta_1}{\beta_2} \right)^{\frac{1}{\#2 - \#1}}. \quad (19)$$

The fixed point is real for  $\text{sign}(\beta_2) = -\text{sign}(\beta_1)$ . For it to lie at small values, where higher-loop terms are small, we require  $|\beta_1| < |\beta_2|$ . Actually, the two-loop coefficient is a proxy for the higher-loop terms: the fixed point is generated by the competing signs of the one-loop vs. the “effective” two-loop term. As one extends the asymptotic perturbative series to higher loops, the fixed-point value shifts to compensate for the change, but as long as the sign of the “effective” two-loop term is unchanged, a fixed point exists, cf. **Figure 5**.



**FIGURE 5 |** Beta function of the gauge coupling  $\alpha_g$  for the model in Litim and Sannino (2014) at next-to-leading order (two loop, blue continuous line) and next-to-next-to-leading order (three loop, green dashed line), cf. Equation (34) for the two-loop beta function. The fixed-point value for  $\alpha_g$  at the corresponding order has been inserted and  $\epsilon = 0.1$  has been chosen.

The interacting fixed point is IR attractive (repulsive) for  $\beta_1 < 0 (> 0)$ . Additionally, a UV (IR) attractive fixed point lies at  $g_{i*} = 0$ . Therefore, a complete trajectory exists between the free (interacting) fixed point in the UV and the interacting (free) fixed point in the IR for  $\beta_1 < 0 (> 0)$ . The former case is known as the Banks-Zaks fixed point in the case of non-Abelian gauge theories (Caswell, 1974; Banks and Zaks, 1982). The latter underlies new developments in gauge-Yukawa models (Litim and Sannino, 2014), see section 3.

### 2.4.3. Competing Degrees of Freedom

In models with several degrees of freedom, a scale-invariant fixed-point regime can be achieved if the effect of different degrees of freedom can balance out - either within a perturbative expansion or at the nonperturbative level and for dimensionless as well as dimensionful couplings. Schematically,

$$\beta_g = \beta_g^{(\text{d.o.f.1})} - \beta_g^{(\text{d.o.f.2})}, \quad (20)$$

where, e.g., d.o.f.1 might be a bosonic and d.o.f.2 a fermionic contribution. ( $N = 4$ ) super Yang Mills could be seen as a special example (Sohnius, 1985).

A competition of fermionic and bosonic degrees of freedom also occurs in the beta function of a quartic scalar coupling which couples to fermions through a Yukawa coupling. This competition actually underlies Higgs mass bounds in the SM (Altarelli and Isidori, 1994; Hambye and Riessmann, 1997). Perturbatively, the Yukawa coupling in simple Yukawa models is UV unsafe. Hints for a nonperturbative fixed point have been found (Gies and Scherer, 2010; Gies et al., 2010), but been called into question in Vacca and Zambelli (2015) in extended truncations of the RG flow.

## 3. GAUGE-YUKAWA MODELS: ASYMPTOTIC SAFETY AT WEAK COUPLING

In  $d = 4$  dimensions, gauge-Yukawa models can exhibit perturbative asymptotic safety, discovered in Litim and Sannino (2014), achieved through a balance of one- vs. two-loop effects. We follow Litim and Sannino (2014) and consider a simple gauge theory with gauge coupling  $g$  with

$$\alpha_g = \frac{g^2}{(4\pi)^2}, \quad (21)$$

with 2-loop beta function

$$\beta_{\alpha_g} = (-B + C\alpha_g)\alpha_g^2. \quad (22)$$

An interacting fixed point lies at

$$\alpha_{g*} = \frac{B}{C}. \quad (23)$$

For the case  $B > 0$ ,  $C > 0$ , this is the Banks-Zaks fixed point (Banks and Zaks, 1982), which is IR attractive in the gauge coupling. Accordingly, a complete RG trajectory can be constructed, emanating from the free fixed point in the UV and ending in a conformal regime in the IR. This can be achieved within the conformal window, e.g.,  $11/2N_c < N_f < 34N_c^3/(13N_c^2 - 3)$  for  $N_f$  fermions in the fundamental representation of  $SU(N_c)$ , Rytlov and Shrock (2011); Pica and Sannino (2011), and Rytlov and Shrock (2016).

Asymptotic safety in the form of an IR-repulsive interacting fixed point occurs where asymptotic freedom is lost, i.e., the antiscreening effect of non-Abelian gauge bosons is overcompensated by the screening effect of charged matter. This requires  $B < 0$  (see Caswell, 1974; Tarasov and Vladimirov, 1977; Jones, 1982; Machacek and Vaughn, 1983), and accordingly  $C < 0$  for the coupling  $g$  to be real. As shown in Caswell (1974), see also Bond and Litim (2017), this is not possible to achieve with fermions only. Adding scalars to the model provides a Yukawa coupling

$$\alpha_y = \frac{y^2}{(4\pi)^2}, \quad (24)$$

that results in

$$C \rightarrow C - D\alpha_y. \quad (25)$$

This facilitates asymptotic safety. The one-loop Yukawa beta function reads

$$\beta_{\alpha_y} = \partial_t \alpha_y = (E\alpha_y - F\alpha_g)\alpha_y, \quad (26)$$

see Fischler and Oliensis (1982) and Machacek and Vaughn (1984) for two-loop results. The above system of beta functions

admits three solutions

$$\alpha_{g*} = 0, \alpha_{y*} = 0, \quad (27)$$

$$\alpha_{g*} = \frac{B}{C}, \alpha_{y*} = 0, \quad (28)$$

$$\alpha_{g*} = \frac{B}{C - D\frac{F}{E}}, \alpha_{y*} = \frac{B}{C - D\frac{F}{E}} \frac{F}{E}, \quad (29)$$

where appropriate conditions on the coefficients of the beta function ensure that fixed-point values are real. The second fixed point is a generalization of the Banks-Zaks fixed point. The fully interacting fixed point has one IR attractive and one IR-repulsive direction. The corresponding critical exponents are

$$\theta_{1,2} = \frac{-BE}{2(CE - DF)^2} \left( -BCE - CEF + DF^2 \pm \sqrt{B^2C^2E^2 - 2BF(CE - 2DF)(CE - DF) + F^2(CE - DF)^2} \right). \quad (30)$$

Perturbative asymptotic safety in four-dimensional gauge theories requires the presence of fermions *and* scalars (Litim and Sannino, 2014), providing a possible justification for the existence of fundamental scalars in nature. Moreover, (gravity-free) theories in four dimensions cannot exhibit weakly-coupled fixed points, i.e., arising from a balance of one-loop vs. two-loop effects, unless gauge interactions are present (Bond and Litim, 2017, 2018a). This explains why tentative proposals for interacting fixed points in four-dimensional fermion-scalar theories lie in a nonperturbative regime (Gies and Scherer, 2010; Eichhorn et al., 2018a).

As couplings can be rescaled arbitrarily (without an impact on the critical exponents), the fixed-point values of couplings do not automatically convey information on whether the fixed point is perturbative. To achieve strict perturbative control over the fixed point, the critical exponents should be arbitrarily close to the canonical ones. This can be achieved in the Veneziano limit which allows to continuously emerge the fixed point from the free one. Hence we now focus on an  $SU(N_c)$  gauge theory with  $N_f$  flavors of Dirac fermions in the fundamental representation to take the Veneziano-limit, Veneziano (1979),

$$N_f \rightarrow \infty, \quad N_c \rightarrow \infty, \quad \text{with } \epsilon = \frac{N_f}{N_c} - \frac{11}{2} \text{ finite.} \quad (31)$$

$C = 25$  (Caswell, 1974) holds in this limit without Yukawa interactions. The simplest way to add a  $N_f \times N_f$  matrix  $H$  of complex scalars is to have them uncharged under the gauge group,

$$\mathcal{L}_{H-\text{pot}} = -u \text{Tr} \left( H^\dagger H \right)^2 - v \left( \text{Tr} H^\dagger H \right)^2. \quad (32)$$

Then the two quartic couplings decouple from the beta functions for the gauge and Yukawa coupling at the above order in perturbation theory and in the Veneziano limit (Litim and Sannino, 2014, see Jack and Osborn, 1984; Machacek and Vaughn, 1985; Ford et al., 1992) for the two-loop beta functions.

In the limit (31), fixed-point values are controlled by  $\epsilon$  and remain perturbative for  $\epsilon \ll 1$  (Litim and Sannino, 2014). For a

study of gauge groups and representations for which such a fixed point exists (see Bond et al., 2018). Asymptotic safety is achieved in appropriately rescaled couplings that guarantee well-behaved large- $N$ -beta functions,

$$\tilde{\alpha}_y = \frac{y^2 N_c}{16\pi^2}, \quad \tilde{\alpha}_g = \frac{g^2 N_c}{16\pi^2}. \quad (33)$$

The beta functions read

$$\beta_{\tilde{\alpha}_g} = \tilde{\alpha}_g^2 \left( \frac{4}{3}\epsilon + \left( 25 + \frac{26}{3}\epsilon \right) \tilde{\alpha}_g - 2 \left( \frac{11}{2} + \epsilon \right)^2 \tilde{\alpha}_y \right), \quad (34)$$

$$\beta_{\tilde{\alpha}_y} = \tilde{\alpha}_y \left( (13 + 2\epsilon) \tilde{\alpha}_y - 6\tilde{\alpha}_g \right). \quad (35)$$

While the one- and two-loop contribution of the gauge coupling to  $\beta_{\tilde{\alpha}_g}$  are positive, the contribution of the Yukawa coupling is negative. Accordingly a finite fixed-point value of the Yukawa coupling induces a physically acceptable fixed point, i.e.,  $\tilde{\alpha}_{g*} > 0$ . In turn, the positive contribution  $\sim \tilde{\alpha}_y$  in  $\beta_{\tilde{\alpha}_y}$  can balance against the negative one  $\sim \tilde{\alpha}_g$  to generate a physically acceptable fixed point at  $\tilde{\alpha}_y > 0$ . This results in an interacting fixed point emerging from the Gaussian one, since  $\epsilon$  can become arbitrarily small for large enough numbers of fields,

$$\begin{aligned} \tilde{\alpha}_{g*} &= \frac{26\epsilon + 4\epsilon^2}{57 - 46\epsilon - 8\epsilon^2}, \\ \tilde{\alpha}_{y*} &= \frac{12\epsilon}{57 - 46\epsilon - 8\epsilon^2}. \end{aligned} \quad (36)$$

To leading order in  $\epsilon$ , the critical exponents are given by

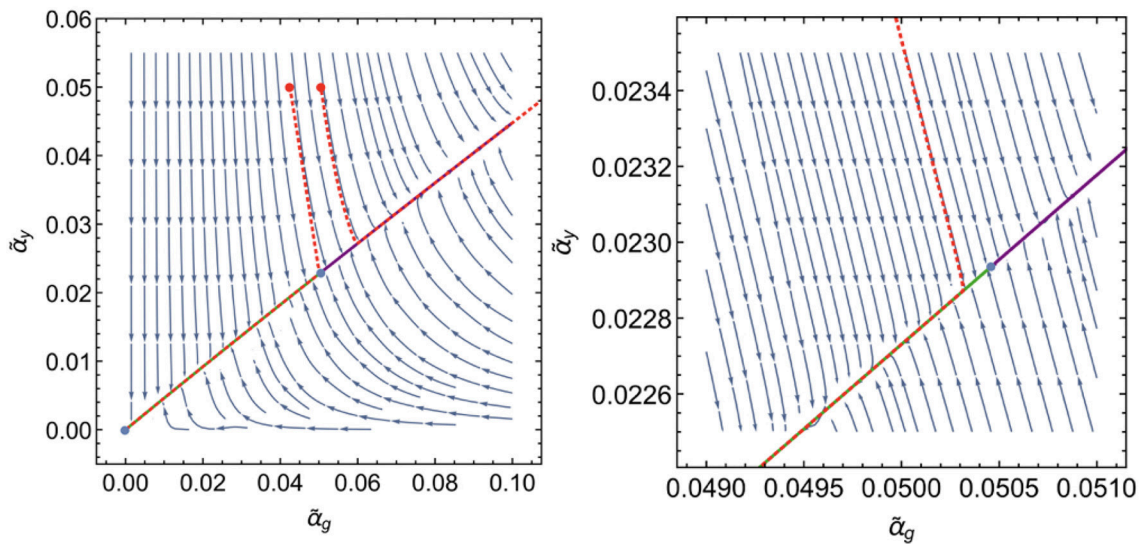
$$\theta_1 = \frac{104}{171}\epsilon^2, \quad \theta_2 = -\frac{52}{19}\epsilon, \quad (37)$$

which go back to the canonical, vanishing values for  $\epsilon \rightarrow 0$ . There is one IR repulsive and one IR attractive direction, fixing the Yukawa coupling at all scales in terms of the gauge coupling (or vice-versa). In other words, the value of one of the couplings in terms of the other is a prediction of the setting.

In a setting with “non-fundamental” asymptotic safety (with new physics kicking in at some  $k_{UV}$ ), it is important that the velocity of the flow in the IR-attractive direction is of order  $\epsilon$ , whereas it is of order  $\epsilon^2$  in the IR-repulsive direction. At the transition scale to the more fundamental description, initial conditions for the values of couplings are typically not the fixed-point values. Toward the IR, the flow is pulled toward the fixed point along the IR-attractive direction with a velocity  $\mathcal{O}(\epsilon)$  and repelled from the fixed point along the IR-repulsive direction with a velocity  $\mathcal{O}(\epsilon^2)$ . Accordingly, near-fixed-point scaling could determine the behavior of a larger class of trajectories, cf. **Figure 6**.

At the next order in the approximation, quartic scalar self-interactions have to be included. For the corresponding large  $N$  couplings

$$\alpha_h = \frac{u N_f}{16\pi^2}, \quad \alpha_v = \frac{v N_f}{16\pi^2}, \quad (38)$$



**FIGURE 6 |** The flow toward the IR from the fixed point in Equation (36) for  $\epsilon = 1/10$  features one strongly IR attractive and one weakly IR repulsive direction. The green and purple, thick, continuous lines are the two only “true” fixed-point trajectories. Initial conditions in the UV away from the fixed point (red dots) result in trajectories that are indistinguishably close to the fixed-point trajectories in the IR. The right panel shows a zoom into the vicinity of the fixed point, where the “non-fundamental” trajectory narrowly misses the fixed point, but approaches the critical hypersurface arbitrarily closely, resulting in universal predictions in the IR.

the one-loop beta functions are given by Jack and Osborn (1984); Machacek and Vaughn (1985), and Ford et al. (1992)

$$\beta_{\alpha_h} = -(11 + 2\epsilon)\tilde{\alpha}_y^2 + 4\alpha_h(\tilde{\alpha}_y + 2\alpha_h), \quad (39)$$

$$\beta_{\alpha_v} = 12\alpha_h^2 + 4\alpha_v(\alpha_v + 4\alpha_h + \tilde{\alpha}_y). \quad (40)$$

Due to the Yukawa coupling, fermionic fluctuations generate a scalar potential (cf. first term in Equation 39) and cannot be set to zero consistently if  $\tilde{\alpha}_y \neq 0$ ; therefore a nontrivial fixed point of the system  $\tilde{\alpha}_y, \tilde{\alpha}_g, \alpha_v, \alpha_h$  has to be found. To satisfy Weyl consistency conditions (see below), the beta function of the gauge coupling is extended to three-loop order and that of the Yukawa coupling to two-loop order, where there is also a contribution  $\sim \alpha_h$ . The double-trace coupling  $\alpha_v$  decouples from the remainder of the system. The system admits a joint, asymptotically safe fixed point at  $\alpha_{h*} > 0$ , and  $\alpha_{v*} < 0$  with  $\alpha_{h*} + \alpha_{v*} > 0$ , indicating a fixed-point potential that is bounded from below Litim and Sannino (2014). A study of the effective potential that includes quantum fluctuations at all scales on a trajectory emanating from the fixed point also indicates its stability (Litim et al., 2016). At the fixed point, the scalar couplings are irrelevant, therefore the full model only features one free parameter.

The inclusion of two-loop effects in the gauge coupling and one-loop effects in the Yukawa coupling (or three-loop in the gauge, two-loop in the Yukawa, and one loop in the scalar couplings) is suggested by Weyl consistency conditions (Jack and Osborn, 1990, 2014), which relate derivatives of beta functions. They arise by considering the model on a curved (but fixed) background and performing Weyl rescalings of the metric. As two subsequent Weyl rescalings commute, it follows that  $\frac{\partial \beta^i}{\partial g_j} =$

$\frac{\partial \beta^j}{\partial g_i}$ . Herein  $\beta^i = \chi^{ij}\beta_j$ , where  $\chi^{ij}$  is a metric in the space of couplings which depends on the couplings. An expression for  $\chi^{ij}$  for gauge-Yukawa models has been derived in Antipin et al. (2013).

These conditions should hold for the full RG flow and can be imposed on the perturbative expansion. For a discussion of the corresponding ordering scheme for beta functions as well as other systematic choices of perturbative orders in the context of gauge-Yukawa theories (see also Bond et al., 2018).

Residual interactions in canonically marginal couplings at an interacting fixed point provide finite contributions to beta functions of higher-order, canonically irrelevant couplings. Higher-order couplings in the scalar potential develop near-Gaussian fixed-point values of their own (Buyukbese and Litim, 2017). Accordingly, their scaling exponents follow the expectation that these couplings should remain irrelevant at a perturbative asymptotically safe fixed point.

The interacting fixed point in gauge-Yukawa systems constitutes a four-dimensional example of asymptotic safety, established within perturbation theory. It provides a new universality class which calls for an in-depth study of its possible extensions and generalizations.

The extension to a supersymmetric setting has been discussed in Intriligator and Sannino (2015); Bajc and Sannino (2016); Bond and Litim (2017); and Bajc and Dondi (2018). While perturbative asymptotic safety cannot be realized in the supersymmetric setting with a simple gauge-group (Intriligator and Sannino, 2015), it can exist in settings with semi-simple gauge groups (Bond and Litim, 2017; Bajc and Dondi, 2018). This highlights how an added symmetry can allow to derive strong no-go-theorems for asymptotic safety.

The fixed-point structure in gauge-Yukawa models is more intricate in a setting away from four dimensions (Codello et al., 2016) (or under the inclusion of potential quantum-gravity effects Christiansen et al., 2017), where the degeneracy of the free fixed point is lifted, and fixed-point collisions can occur.

Given that asymptotic safety appears in a range of gauge theories where asymptotic freedom is lost, the phase diagram of gauge theories could be richer than previously thought. In fact, indications for an interacting fixed point at leading order in  $1/N_f$  go back to Palanques-Mestre and Pascual (1984) and Gracey (1996), see Holdom (2011) for a recap and a discussion of higher orders in  $1/N_f$ . With a view toward the potential phenomenological importance of such fixed points (Antipin and Sannino, 2017; Antipin et al., 2018a), employ a resummation of the fermionic bubble diagrams that contribute at leading order in  $1/N_f$  to all orders in perturbation theory to the beta function for the non-Abelian gauge coupling. This provides indications for an interacting fixed point: In a  $1/N_f$  expansion, the first nontrivial order vanishes in the large  $N_f$  limit, unless there is a value of the coupling where it features a pole. In that case, depending on the sign of that contribution, a zero of the beta function can be generated. Indeed a corresponding pole can be found, providing an indication for a fixed point at a non-perturbatively large value of the gauge coupling. A similar resummation for the gauge contribution to the leading nontrivial order of the beta function of the Yukawa coupling has been performed in Kowalska and Sessolo (2018), see also Alanne and Blasi (2018) and Alanne and Blasi (2018).

The  $a$ -theorem (Cardy, 1988) has been explored in this setting (Antipin et al., 2018b; Dondi et al., 2018), showing that, as expected, the Jack and Osborn  $a$  function (Jack and Osborn, 2014) takes a larger value at the UV fixed point than at the IR fixed point.

These developments pave the way for asymptotically safe model building beyond the Standard Model (e.g., Abel and Sannino, 2017a,b; Bond et al., 2017; Mann et al., 2017; Molinaro et al., 2018).

### 3.1. Asymptotically Safe Phenomenology

The idea that scale-invariance is realized in physics beyond the Standard Model has received a lot of attention (see e.g., Meissner and Nicolai, 2007; Shaposhnikov and Zenhausern, 2009a,b; Holthausen et al., 2013; Khoze, 2013; Lindner et al., 2014; Gies and Zambelli, 2017; Lewandowski et al., 2018), mostly focusing on settings with classical scale invariance. It is therefore highly intriguing to explore whether extensions of the Standard Model are asymptotically safe along the lines in Litim and Sannino (2014), realizing quantum scale invariance. Measurements showing a decreasing SU(3) coupling as a function of energy only cover a finite energy range and hence do not exclude asymptotic safety.

Steps toward an asymptotically safe Standard Model include the observation that asymptotic safety can be achieved in semi-simple gauge groups (Esbensen et al., 2016; Bond and Litim, 2018b) and with chiral fermions (Mølgaard and Sannino, 2017). To render the non-Abelian Standard Model gauge couplings asymptotically safe, new fermionic states transforming in nontrivial representation of SU(2) and/or SU(3), have to

be added. Asymptotic safety might be achieved for one of the non-Abelian gauge couplings, with the others becoming asymptotically free, depending on the representation the new (vectorlike) fermions transform in Kowalska et al. (2017). The matching scale, essentially corresponding to the mass scale of the new fermions, which separates the regime of power-law running below the fixed point from the regime of logarithmic running in the Standard Model, adds new free parameters to these models.

Yet, the non-Abelian gauge groups of the Standard Model are SU(2) and SU(3), not SU( $N_c$ ) with  $N_c \rightarrow \infty$ , as required for the Veneziano limit. Accordingly, the addition of fermions to the Standard Model such that the non-Abelian gauge couplings, together with the BSM Yukawa coupling become asymptotically safe (Bond et al., 2017; Mann et al., 2017), is difficult to reconcile with a perturbative nature of the extension (Barducci et al., 2018), at least if one also insists on solving the U(1) triviality problem. This is evident, e.g., in the large values of the critical exponents that lead to a fast flow away from the fixed point toward the IR (see e.g., Bond et al., 2017; Mann et al., 2017). Hence, large  $N_f$  fixed points (Palanques-Mestre and Pascual, 1984; Gracey, 1996; Holdom, 2011; Antipin and Sannino, 2017; Antipin et al., 2018a; Kowalska and Sessolo, 2018) play a key role in these developments. Accommodating the Higgs at the correct mass is a challenge (Pelaggi et al., 2018). This could change in a grand unified setting (Molinaro et al., 2018), which could also become asymptotically safe (Abel and Sannino, 2017b).

Asymptotic safety beyond the Standard Model could have intriguing phenomenological consequences in astrophysics and cosmology (Nielsen et al., 2015). For instance, asymptotically safe dark matter could accommodate a significant running of the portal coupling to visible matter between the dark-matter-mass scale—relevant for thermal production of dark matter in the early universe—and the scale of direct detection experiments (Sannino and Shoemaker, 2015). In the WIMP-paradigm, the dark matter relic density is linked to the probability of direct detection, since the cross-section for dark-matter-annihilation into Standard Model particles is related to the cross-section for dark matter scattering off Standard Model particles. Hence, the lack of direct detection has put severe constraints on the paradigm (Tan et al., 2016; Akerib et al., 2017; Aprile et al., 2018). These might be circumvented by introducing additional fields, providing a new parameter that decreases the tension between direct experimental bounds on the cross-section and the relic-density constraints. Asymptotic safety could provide an alternative explanation (Sannino and Shoemaker, 2015): the value of the coupling at the higher scale is larger as it approaches an interacting fixed point. This might accommodate thermal production of the full dark matter relic density while being consistent with bounds from direct searches.

## 4. ASYMPTOTICALLY SAFE QUANTUM GRAVITY

### 4.1. Status of Asymptotic Safety in Gravity

Einstein gravity, quantized perturbatively, loses predictivity at (trans)planckian scales due to its perturbative

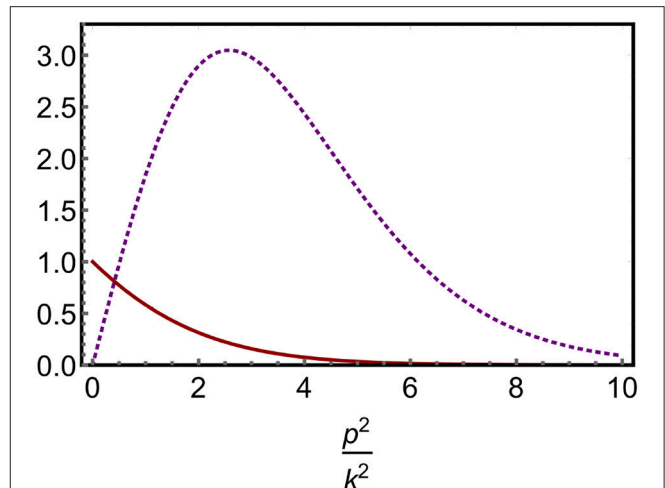
nonrenormalizability. Infinitely many free parameters are associated with counterterms required to absorb new divergences appearing at every loop order<sup>5</sup> (Deser and Nieuwenhuizen, 1974a,b; 't Hooft and Veltman, 1974; Goroff and Sagnotti, 1986; van de Ven, 1992). At momenta  $p$  below the Planck scale, only a finite number of the counterterms contribute (Donoghue, 1994a,b) if one assumes that the corresponding dimensionless couplings are all of order one. Then, higher-order terms are suppressed by  $(p/M_{\text{Planck}})^{\#}$ ,  $\# > 2$ . Thus, gravity and quantum physics are actually compatible, but a perturbative quantization only holds up to the Planck scale (Donoghue, 2012). The key challenge is to find an ultraviolet completion. The minimalistic and conservative nature of asymptotic safety as compared to many other approaches to quantum gravity make it a useful starting point for this endeavor: If this ansatz for quantum gravity fails, more radical notions on the quantum nature of spacetime are required.

As the free fixed point is IR attractive in the Newton coupling,<sup>6</sup> the first mechanism for asymptotic safety cf. Sec. 2.4.1, which is realized in  $d = 2 + \epsilon$  dimensions (Brown, 1977; Tsao, 1977; Christensen and Duff, 1978; Gastmans et al., 1978; Weinberg, 1980; David, 1988; Distler and Kawai, 1989; Kawai and Ninomiya, 1990; Jack and Jones, 1991; Kawai et al., 1993a,b, 1996; Aida et al., 1994; Nishimura et al., 1994; Aida and Kitazawa, 1997; Codello and D'Odorico, 2015), might also determine the fate of gravity in  $d = 4$  dimensions. The physical mechanism behind asymptotic safety in gravity (Nink and Reuter, 2013) is the antiscreening nature of metric fluctuations that shield the Newton coupling, similar to the effect of self-interacting gluons in the Yang-Mills vacuum.

An extension of the  $\epsilon$  expansion to higher order, combined with an appropriate resummation, could provide indications for or against a fixed point in four dimensions. This is also a goal of discrete approaches to the gravitational path-integral where spacetime configurations are constructed from scratch from microscopic building blocks: Causal (Ambjorn et al., 2000, 2001) (and possibly also Euclidean Laiho and Coumbe, 2011; Laiho et al., 2017) Dynamical Triangulations) (CDT) feature a higher-order phase transition (Ambjorn et al., 2011, 2012, 2017) facilitating a continuum limit. This could provide a universality class for quantum gravity. Complementary to Monte Carlo simulations of dynamical triangulations, an analytical approach

<sup>5</sup>The enhanced symmetry in supergravity rules out many of these counterterms, shifting the expected order of divergence in the maximally supersymmetric theory to higher orders (Bern et al., 2017, 2018).

<sup>6</sup>In the higher-derivative theory with the additional invariants  $R^2$  and  $R_{\mu\nu}R^{\mu\nu}$  the marginal couplings are asymptotically free (Stelle, 1977; Fradkin and Tseytlin, 1982; Avramidi and Barvinsky, 1985). Around flat space, this theory features a kinetic instability (see Salvio, 2018) for a review. Breaking Lorentz symmetry allows to use higher-order spatial derivatives while keeping the action at second order in time derivatives (Horava, 2009), resulting in perturbatively renormalizability. Yet, the projectable version propagates an additional scalar that becomes nonperturbative in the IR. As the non-projectable version features a larger number of couplings, asymptotic freedom has only been established in  $2 + 1$  (Barvinsky et al., 2017) dimensions and not  $3 + 1$ , as well as in the large  $N$  limit for  $N$  scalars coupled to Horava gravity (D'Odorico et al., 2014). Constraints from pulsars (Yagi et al., 2014) and gravitational waves from a neutron-star merger (Emir Gümürkçüoğlu et al., 2018) constrain these models.



**FIGURE 7 |** The regulator  $R_k(p^2)$  (continuous red line) acts as a suppression term for IR modes. In the flow equation Equation (43) its derivative with respect to  $k$  (dotted purple line) acts as a suppression for UV modes, as well, such that the main contribution to the scale dependence of the dynamics at  $k$  comes from modes at that momentum scale.

to search for a suitable continuum limit is based on tensor models (Ambjorn et al., 1991; Godfrey and Gross, 1991; Gross, 1992; Benedetti and Gurau, 2012; Gurau, 2016), see Sec. 4.2.4. Lattice studies based on Euclidean Regge calculus have also been put forward as indications for asymptotic safety (Hamber, 2009, 2015).

Intriguingly, perturbative techniques in  $d = 4$  yield indications for an asymptotically safe fixed point (Codello and Percacci, 2006; Niedermaier, 2009, 2010), providing a hint at a near-perturbative nature of asymptotically safe gravity.

Most of the compelling evidence for asymptotic safety in gravity comes from Euclidean functional RG (FRG) studies based on the Wetterich equation<sup>7</sup>. This framework provides beta functions for the dependence of couplings on the momentum scale  $k$ . The scale is introduced into the generating functional through an infrared cutoff function  $R_k(p^2)$ , called the regulator,

$$Z_k[J] = \int \mathcal{D}\varphi e^{-S[\varphi] - \frac{1}{2} \text{Tr} \varphi R_k(p^2) \varphi + \text{Tr} J \varphi}, \quad (41)$$

$$\Gamma_k[\phi] = \sup_J (\text{Tr} J \phi - \ln Z_k[J]) - \frac{1}{2} \text{Tr} \phi R_k(p^2) \phi, \quad \langle \varphi \rangle_k = \phi, \quad (42)$$

reducing to the standard definitions at  $k = 0$ .  $R_k(p^2)$  and its derivative  $k \partial_k R_k(p^2)$  are sketched in **Figure 7**. This setup provides a flow equation, the Wetterich equation (Wetterich, 1993), also Ellwanger (1994) and Morris (1994), pioneered for gauge theories in Reuter and Wetterich (1994) and gravity in Reuter (1998). The regulator acts as a simultaneous IR- and UV

<sup>7</sup>A variant of the Polchinski equation also provides support for the asymptotic-safety conjecture (de Alwis, 2018).

cutoff, such that the change of a coupling at scale  $k$  is mainly driven by quantum fluctuations at that scale:

$$\partial_t \Gamma_k = k \partial_k \Gamma_k = \frac{1}{2} \text{Tr} \left[ \left( \frac{\delta^2 \Gamma_k[\phi]}{\delta \phi^2} + R_k \right)^{-1} \partial_t R_k \right]. \quad (43)$$

For gravity, the covariant Laplacian  $\bar{\Delta}$  with respect to an auxiliary background metric  $\bar{g}_{\mu\nu}$  takes the role of the momentum  $p^2$ , Reuter (1998); Dou and Percacci (1998); Souma (1999), and Lauscher and Reuter (2002). For general introductions and reviews (see Berges et al., 2002; Polonyi, 2003; Pawłowski, 2007; Rosten, 2010; Braun, 2012; Delamotte, 2012; Gies, 2012, for gravity, see Reuter and Saueressig, 2012; Nink et al., 2013; Ashtekar et al., 2014; Percacci, 2017; Eichhorn, 2018a). The method is well-suited to models with dimensionful couplings, and therefore widely-used in condensed-matter physics and statistical physics (Kopietz et al., 2010; Metzner et al., 2012; Platt et al., 2013). Its relation to perturbation theory, which is straightforward at one loop, has been explored at higher loops in Papenbrock and Wetterich (1995) and Codello et al. (2014).

The FRG tracks the scale dependence of all couplings that are compatible with the symmetries, not just the perturbatively renormalizable interactions. For practical calculations, theory space is truncated to a (typically finite-dimensional) subspace, introducing a systematic error. To highlight that quantitative results can already be achieved in relatively small truncations, we provide the leading scaling exponents for the Ising model in **Table 1**.

For fixed points that arise via the mechanism in section 2.4.1, the scaling is near-canonical near the critical dimension, providing a systematic way to devise truncations that include all relevant couplings. There are indications that in quantum gravity four dimensions is close to two in the sense that the canonical dimension is a good predictor of relevance at the fixed point (Falls et al., 2013, 2016), enabling the setup of robust truncations by canonical power-counting. These indications require further confirmation, e.g., by including operators of the form  $R \square^n R$  (De Alwis, 2018).

Considerable evidence for the existence of the interacting Reuter fixed point has accumulated, starting from the seminal work (Dou and Percacci, 1998; Reuter, 1998; Souma, 1999; Lauscher and Reuter, 2002; Reuter and Saueressig, 2002), employing truncations of the form

$$\begin{aligned} \Gamma_k = & -\frac{1}{16\pi G_N} \int d^4x \sqrt{g} (R - 2\Lambda) + \Gamma_{k \text{ higher-order}} + \frac{1}{32\pi G_N \alpha} \\ & \int d^4x \sqrt{g} \bar{g}^{\mu\nu} \left( \bar{D}^\kappa h_{\mu\kappa} - \frac{1+\beta}{4} \bar{D}_\mu h \right) \left( \bar{D}^\lambda h_{\nu\lambda} - \frac{1+\beta}{4} \bar{D}_\nu h \right) \\ & - \sqrt{2} \int d^4x \sqrt{g} \bar{c}_\mu \left( \bar{g}^{\mu\rho} \bar{D}^\kappa g_{\rho\nu} D_\kappa + \bar{D}^\kappa g_{\kappa\nu} D_\rho \right) - \frac{1+\beta}{2} \bar{D}^\mu D_\nu c^\nu. \end{aligned} \quad (44)$$

The third term is a gauge fixing term with two parameters  $\alpha, \beta$  (see, e.g., Falkenberg and Odintsov, 1998; Gies et al., 2015; de Brito et al., 2018 for studies of the off-shell gauge dependence and Benedetti, 2012 for gauge-independent on-shell results) and the

**TABLE 1 |** Relevant and leading irrelevant critical exponent as well as the anomalous dimension for the Ising model obtained with the FRG in a derivative expansion to leading order (local potential approximation, LPA, to order  $2n$  in the field) and next-to-order (LPA') with field-independent anomalous dimension.

Truncation	$\nu = 1/\theta_1$	$\omega = -\theta_2$	$\eta$
LPA 2	1/2	1/3	0
LPA 3	0.729	1.07	0
LPA 4	0.651	0.599	0
LPA 5	0.645	0.644	0
LPA 6	0.65	0.661	0
LPA 7	0.65	0.656	0
LPA 8	0.65	0.654	0
LPA' 2	0.526	0.505	0.0546
LPA' 3	0.684	1.33	0.0387
LPA' 4	0.64	0.703	0.0433
LPA' 5	0.634	0.719	0.0445
LPA' 6	0.637	0.728	0.0443
LPA' 7	0.637	0.727	0.0443
LPA' 8	0.637	0.726	0.0443.

For the dimensionless potential  $u[\rho] = \sum_{i=2}^{\infty} \frac{\lambda_i}{i!} (\rho - \lambda_1)^i$  with  $\rho = \varphi^2/2$ , the flow equation from which the beta functions for the couplings  $\lambda_i$  are derived, reads  $\partial_t u[\rho] = -4u + (d-2+\eta)\rho u'[\rho] + \frac{1}{2(4\pi)^2} (1-\frac{\eta}{6}) \frac{1}{1+u'[\rho]+2\rho u''[\rho]}$ . The underlying derivation of the flow equation can be found (e.g., Berges et al., 2002; Delamotte, 2012) and the numerical evaluation of fixed-point values and critical exponents requires a basic numerical solver, such as Mathematica's FindRoot routine. At fourth order in the derivative expansion (Canet et al., 2003), one obtains  $\nu = 0.632$  and  $\eta = 0.033$  (see also Litim and Zappala, 2011) (compared to, e.g.,  $\nu = 0.6304$  and  $\eta = 0.0335$  from 7-loop studies Guida and Zinn-Justin, 1998).

third line is the corresponding Faddeev-Popov operator<sup>8</sup>. Barred quantities refer to a background metric  $\bar{g}_{\mu\nu}$  with respect to which the metric  $g_{\mu\nu}$  can be gauge fixed, and a local coarse-graining scheme is set up. The fluctuation field is

$$h_{\mu\nu} = g_{\mu\nu} - \bar{g}_{\mu\nu}. \quad (45)$$

A discussion of background-independence is given in section 4.2.3. All results below are in the background-approximation, where  $g_{\mu\nu} = \bar{g}_{\mu\nu}$  is used in the RG flow. Results from selected key truncations are summarized in **Table 2**.

For purposes of illustration, we also quote beta functions in the Einstein-Hilbert truncation with  $G = G_N k^2$  and  $\lambda = \Lambda/k^2$  from Codello et al. (2009), with anomalous dimensions  $\eta_{h(c)}$  for the metric (ghost) (e.g., in Donà et al., 2014), and for the functional  $f(\bar{R})$ , from Benedetti and Caravelli (2012) as found in Dietz and Morris (2013a).

<sup>8</sup>A nontrivial wave-function renormalization (Eichhorn and Gies, 2010; Groh and Saueressig, 2010) and ghost terms beyond the Faddeev-Popov are generated by the flow and have nonvanishing fixed-point values (Eichhorn, 2013).

**TABLE 2** | The operators beyond Einstein-Hilbert, the number of relevant/irrelevant directions, and the values of the positive critical exponents are indicated.

References	Gauge	Cutoff	Operators included beyond Einstein-Hilbert	# rel. dir.	# irrel. dir.	Re $\theta_1$	Re $\theta_2$	Re $\theta_3$
Reuter and Saueressig, 2002	$\alpha = 1, \beta = 0$	exp.	-	2	-	1.94	1.94	-
Litim, 2004	$\alpha = 0$	Litim (Litim, 2000, 2001)	-	2	-	1.67	1.67	-
Lauscher and Reuter, 2002	$\alpha = 0, \beta = 0$	exp.	$\sqrt{g}R^2$	3	0	28.8	2.15	2.15
Machado and Saueressig, 2008	$\beta = 1, \alpha = 0$	Litim	$\sqrt{g}R^2, \sqrt{g}R^3$	3	1	2.67	2.67	2.07
Codello et al., 2009	$\alpha = 1, \beta = 1$	Litim	$\sqrt{g}R^2, \sqrt{g}R^3$	3	1	2.71	2.71	2.07
Machado and Saueressig, 2008	$\beta = 1, \alpha = 0$	Litim	$\sqrt{g}R^2, \sqrt{g}R^6$	3	1	2.39	2.39	1.51
Codello et al., 2009	$\alpha = 1, \beta = 1$	Litim	$\sqrt{g}R^2, \dots, \sqrt{g}R^8$	3	6	2.41	2.41	1.40
Falls et al., 2013, 2016	$\alpha = 0, \beta = 0$	Litim	$\sqrt{g}R^2, \dots, \sqrt{g}R^{34}$	3	32	2.50	2.50	1.59
Benedetti et al., 2009	$\alpha = 0$ , h/o	Litim	$\sqrt{g}R^2, \sqrt{g}R_{\mu\nu}R^{\mu\nu}$	3	1	8.40	2.51	1.69
Gies et al., 2016	$\beta = \alpha = 1$	Litim	$\sqrt{g}C^{\mu\nu\kappa\lambda}C_{\kappa\lambda\rho\sigma}C^{\rho\sigma}_{\mu\nu}$	2	1	1.48	1.48	-

All truncations listed above, employing the linear parameterization and single-metric approximation (cf. section 4.2.3) feature an asymptotically safe fixed point with no more than three relevant directions. (All results in the literature for finite-dimensional truncations feature an asymptotically safe form in qualitative agreement with these results.)

$$\beta_G = 2G - \frac{G^2}{12 \cdot 4\pi} \left( \frac{52(4 - \eta_h)}{1 - 2\lambda} + 40(4 - \eta_c) \right),$$

$$\beta_\lambda = -2\lambda + \frac{G}{12 \cdot 4\pi} \left( \frac{20(6 - \eta_h)}{1 - 2\lambda} - 16(6 - \eta_c) \right) - \frac{G\lambda}{12 \cdot 4\pi} \left( \frac{52(4 - \eta_h)}{1 - 2\lambda} + 40(4 - \eta_c) \right), \quad (46)$$

$$\partial_t f = 4f - 2\tilde{R}f' + \frac{1}{384\pi^2} \left[ -20 \frac{\partial_t f' - 2\tilde{R}f'' + 8f'}{(\tilde{R} - 2)f' - 2f} - 36 - 12 - 5\tilde{R}^2 \right. \\ \left. + \frac{(\tilde{R}^4 - 54\tilde{R}^2 - 54)(\partial_t f'' - 2\tilde{R}f''') - (\tilde{R}^3 + 18\tilde{R}^2 + 12)(\partial_t f' - 2\tilde{R}f'' + 2f') - 36(\tilde{R}^2 + 2)(f' + 6f'')}{2(-9f'' + (\tilde{R} - 3)f' - 2f)} \right]. \quad (47)$$

Here,  $\tilde{R} = R/k^2$  is the dimensionless curvature and primes denote derivatives with respect to  $\tilde{R}$ . Equation (47) provides the beta functions for couplings of  $R^n$  upon a Taylor expansion in the curvature.

The Newton coupling and cosmological constant are relevant, cf. **Figure 8**. Accordingly, the IR value of the cosmological constant is unrestricted. The choice of different fixed-point trajectories in **Figure 8** results in different values of the dimensionful cosmological constant in the IR  $\Lambda_{\text{IR}}$ . To realize  $\Lambda_{\text{IR}}/M_{\text{Pl}}^2 \ll 1$ , a specific trajectory has to be chosen. The question, why this particular trajectory is realized, is the finetuning “problem.” Yet, any relevant coupling is actually linked to a similar question. For instance, the value of the QCD coupling at the electroweak scale would be different on other, also asymptotically free trajectories. The question how a more fundamental principle selects one out of many viable trajectories for the relevant couplings exists irrespective of whether the coupling is logarithmically or power-law sensitive to the momentum scale. (The need for successive tuning at each order in perturbation theory for the power-law case is a consequence of that particular approximation scheme, not a signature of a consistency problem of the theory).

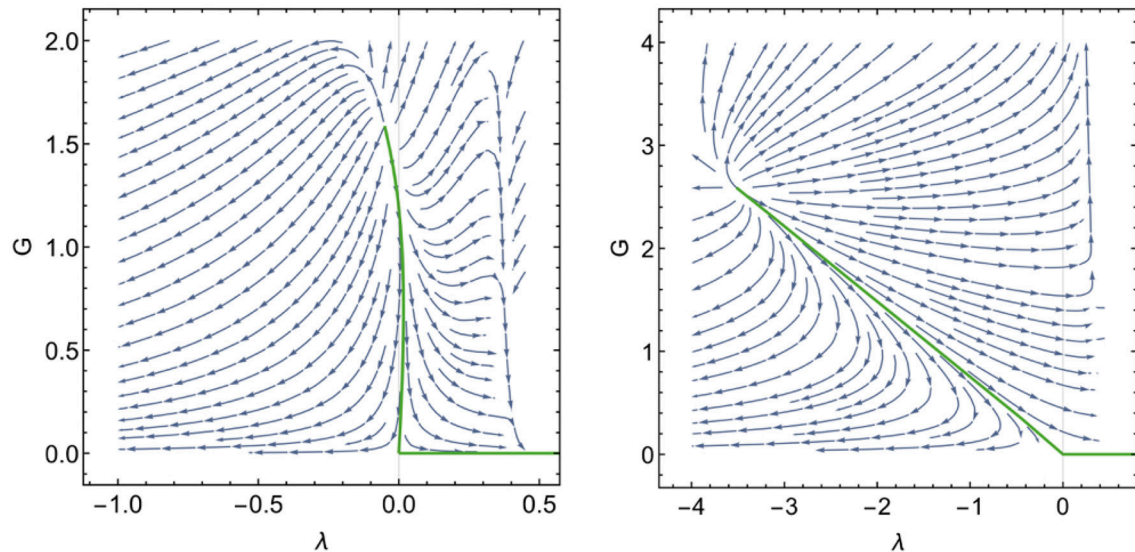
At the interacting Reuter fixed point, canonical ordering appears to hold<sup>9</sup>, cf. **Table 2**. This provides a scheme to set up consistent truncations: Assume for simplicity that the operators in **Table 2** diagonalize the stability matrix, resulting in critical exponents

$$\theta_i = d_{\tilde{g}_i} + \eta_i. \quad (48)$$

Unless the anomalous scaling contribution  $\eta_i$  would grow with the canonical dimension, the canonical dimension dominates for canonically highly irrelevant couplings, rendering them irrelevant at an interacting fixed point. In fact, already at the level of canonically marginal couplings of  $R^2$  and  $R_{\mu\nu}R^{\mu\nu}$ , there appears to be only one relevant direction<sup>10</sup>. All canonically irrelevant operators that have been examined are irrelevant at the fixed point. In Falls et al. (2013) and Falls et al. (2016), the normalized difference of canonical and quantum scaling dimension decreases with decreasing canonical dimension for  $R^n$ , cf. **Figure 9**. The near-Gaussian scaling spectrum (at higher orders in the curvature expansion) is also in line with the possibility to find indications of asymptotic safety from

<sup>9</sup>A combined truncation of Gies et al. (2016) with the operators in Benedetti et al. (2009) remains to be explored.

<sup>10</sup>Results in the exponential parameterization even yield one relevant direction less (Ohta et al., 2016; de Brito et al., 2018).



**FIGURE 8 |** The RG flow to the IR in the Einstein-Hilbert truncation in the setup discussed in Donà et al. (2014) for a type Ia cutoff features a trajectory—passing very close to the free fixed point—on which the dimensionful Newton coupling and cosmological constant reach constant values in the IR in agreement with measurements. Left panel: pure gravity case; right panel: including minimally coupled matter as in the Standard Model (4 scalars, 12 vectors, 45 Weyl fermions). The two eigendirections of the fixed point are superposition of  $G$  and  $\lambda$ .

perturbation theory (Codello and Percacci, 2006; Niedermaier, 2009, 2010).

Systematic truncation errors can be estimated given that in approximations schemes for QFTS, dependencies on unphysical parameters arise even at the level of observables. The better the approximation, the weaker such a dependence. Tests include gauge-parameter dependence (Gies et al., 2015), regulator dependence (Reuter and Saueressig, 2002) and dependence on the parameterization for metric fluctuations (Gies et al., 2015; Ohta et al., 2016; de Brito et al., 2018).

Going beyond finite-dimensional truncations, the closed fixed-point equation for  $f(\tilde{R})$ , e.g., Equation (47) has been investigated. Depending on the choice of regulator, it contains a varying number of fixed singularities, as the regulator introduces additional field-dependence in the background approximation. Thus, specific choices of the regulator allow for global solutions (Benedetti and Caravelli, 2012; Demmel et al., 2012; Dietz and Morris, 2013a,b; Demmel et al., 2015; Ohta et al., 2015) while others do not (Codello et al., 2009). One might conclude that extensions of the truncation are required, going beyond the background-approximation for  $f(\tilde{R})$  (see Christiansen et al., 2018, see section 4.2).

Many gravitational theories are classically dynamically equivalent to GR. Thus different theory spaces could allow for asymptotic safety (Krasnov and Percacci, 2018). For instance, the vielbein and the connection can be treated as independent variables (Daum and Reuter, 2012, 2013; Harst and Reuter, 2015, 2016), or torsion can be included (Pagani and Percacci, 2015; Reuter and Schollmeyer, 2016). The dimension of theory space and the number of relevant couplings decrease by one (Eichhorn, 2015) in unimodular gravity (Unruh, 1989; Finkelstein et al.,

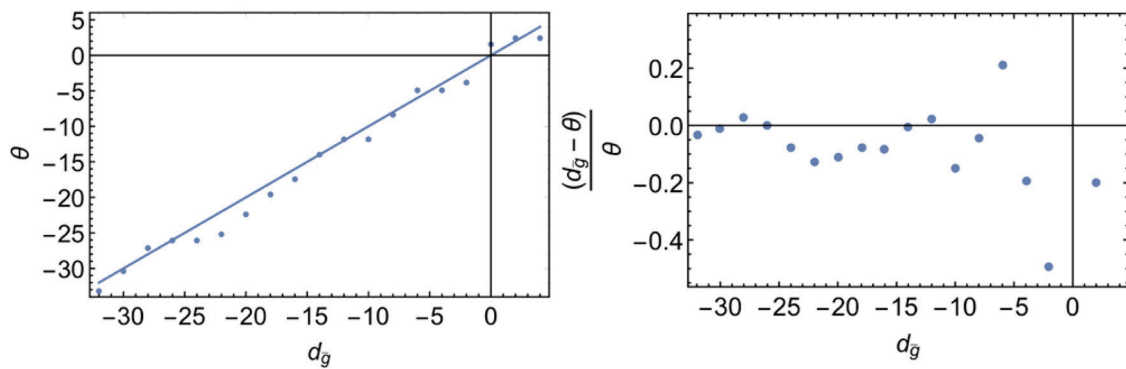
2001; Ellis et al., 2011), where the determinant of the metric is a fixed density, removing the cosmological constant from the action. Further, fluctuations in topology, dimensionality, signature etc. might be included in the gravitational path integral. The corresponding additional configurations either prevent the existence of a continuum limit/ RG fixed point, lead to an asymptotically safe fixed point in the same universality class as **Table 2**, or provide another gravitational universality class which differs in its physical implications and can therefore (in principle) be probed experimentally.

In two dimensions, the conformal field theory underlying asymptotic safety has been studied (Nink and Reuter, 2016). In  $d = 4$ , scale-invariance need not imply conformal invariance (in fact, sufficient conditions for this are not known). If it were possible to extend the conformal bootstrap program (Simmons-Duffin, 2016) to a gravitational setting, a search for the corresponding universality class with relevant directions according to **Table 2** might answer whether there is a conformal theory behind asymptotically safe gravity.

## 4.2. Open Questions & Future Perspectives

### 4.2.1. Lorentzian Signature

There is no simple Wick-rotation in quantum gravity, thus the above results do not directly imply Lorentzian asymptotic safety. In an ADM decomposition of the metric, the change of signature can be implemented by changing one parameter. This has been used in Manrique et al. (2011a) to find hints for asymptotic safety in a Lorentzian setting for the Einstein-Hilbert truncation. Further, RG flows in the ADM decomposition have been explored in Rechenberger and Saueressig (2013), Biemans et al. (2017), and Biemans et al. (2017). The FRG can be formulated in a Lorentzian



**FIGURE 9** | Data from Falls et al. (2016) on the critical exponents in a truncation  $\sum_n \sqrt{g} R^n$ , compared to the canonical dimension.

setting (Floerchinger, 2012), underlying the study of real-time correlators, e.g., in QCD (Pawlowski and Strodthoff, 2015).

Alternatively, a proposal (Eichhorn, 2018b) to search for Lorentzian asymptotic safety employs causal set quantum gravity. This is an intrinsically Lorentzian, discrete approach to quantum gravity, based on the path integral over all causal sets (Bombelli et al., 1987). Under the restriction to manifoldlike causal sets (implemented as a path integral over sprinklings Henson, 2006; Dowker, 2013) the space of couplings might feature a second-order phase transition (Surya, 2012; Glaser, 2018).

#### 4.2.2. Propagating Degrees of Freedom

Higher-order derivatives in QFTs on a flat background generically imply an instability in the kinetic term (Ostrogradsky, 1985; Woodard, 2015), translating into a violation of reflection positivity for the Euclidean propagator (Arici et al., 2017). In a quantum setting, the unboundedness of the Hamiltonian can be traded for unitarity violation through negative-norm states in the Hilbert space (Woodard, 2015).

In quantum gravity, an analysis of unitarity is presumably rather more subtle for several reasons.

Firstly, positivity violation in gauge-variant propagators occurs in unitary theories such as QCD (Cucchieri et al., 2005; Bowman et al., 2007). [A direct analogy with QCD has been proposed for (asymptotically free higher-derivative) gravity in Holdom and Ren, 2016a,b.] The physical “graviton” as the transverse traceless part of the metric propagator is defined perturbatively; but non-perturbatively no local separation of gauge and physical degrees of freedom is possible.

Secondly, an instability in the flat-space propagator is not in conflict with observations, given that the cosmological background appears to be FRW-like.

Thirdly, Ostrogradski instabilities occur under a crucial assumption, namely that of finitely many higher-order terms. Yet the case with infinitely higher order terms *can* feature a well-defined propagator, translating into a well-posed initial value problem at the level of the equations of motion (Barnaby and Kamran, 2008). Examples include string-field theory, see Barnaby and Kamran (2008) and references therein. Accordingly, truncated dynamics in asymptotically safe gravity might contain

spurious instabilities (just as an analysis of a truncated effective action for string theory would).

Fourth, even at the level of curvature-squared actions, the mass of the “ghost” (analyzed around *flat* space) runs as a function of momentum. Hence, such ghosts might not appear as physical states (see Floreanini and Percacci, 1995; Benedetti et al., 2009; Becker et al., 2017).

Finally, if asymptotic safety is “non-fundamental” (cf. section 2.3), the mass-scale of the ghosts (if these exist on physically relevant backgrounds) sets an upper bound on  $k_{UV}$ .

CDT satisfies reflection positivity (Ambjorn et al., 2000, 2001). Thus, its continuum limit, which might correspond to asymptotically safe gravity, inherits this property. As many other examples, this reinforces that the quest to understand quantum spacetime can be accelerated by searching for links between quantum-gravity approaches.

In addition to ghost-like states, higher-order gravity can (but again, need not) contain additional propagating degrees of freedom. These might be of phenomenological interest, e.g., driving inflation or leading to modifications of GR detectable in black holes and/or gravitational waves.

Determining the spectrum of propagating gravitational degrees of freedom in asymptotically safe gravity is an important outstanding question. A comprehensive answer in the FRG approach requires studying the full propagator (at  $k = 0$ , where all quantum fluctuations contribute) around a solution to the quantum equations of motion.

#### 4.2.3. Background Independence

Background independence is a key property of quantum gravity, meaning that all configurations in the path integral should be treated on an equal footing. This appears to be at odds with the introduction of a local coarse graining scheme, as this relies on a metric. Specifically, the regulator in the flow equation depends on a background metric  $\bar{g}_{\mu\nu}$ . Additionally, a local formulation of gauge theories requires gauge fixing to derive the propagator. The flow equation is based on a background gauge-fixing. Nevertheless, background independence can be achieved, if all backgrounds are treated on the same footing (Becker and Reuter,

**TABLE 3** | Fixed-point results from Becker and Reuter (2014) for the “dynamical” couplings in the Einstein-Hilbert truncation and their background counterparts.

$G^*$	$\lambda^*$	$G_B^*$	$\lambda_B^*$	$\theta_1$	$\theta_2$	$\theta_{B1}$	$\theta_{B2}$
0.70	0.21	8.2	-0.01	3.6+4.3i	3.6-4.3i	4	2

Critical exponents can be split into the two sectors, as the background couplings do not couple into the flow of the dynamical couplings and accordingly the stability matrix is upper/lower triangular in the background sector yielding canonical exponents.

2014), i.e., if  $g_{\mu\nu}$  and  $\bar{g}_{\mu\nu}$  are both kept as distinct arguments of the flowing action. In the limit  $k \rightarrow 0$ , where the regulator vanishes, setting  $g_{\mu\nu} = \bar{g}_{\mu\nu}$  yields an effective action that inherits diffeomorphism invariance and therefore background independence from the auxiliary background-diffeomorphism invariance that is kept intact for an appropriate choice of gauge fixing and regulator function. Therefore, ultimately we are interested in  $\Gamma_{k \rightarrow 0}[\bar{g}_{\mu\nu}, g_{\mu\nu} = \bar{g}_{\mu\nu}]$ , or  $\Gamma_{k \rightarrow 0}[\bar{g}_{\mu\nu}, h_{\mu\nu} = 0]$ , respectively. Crucially, the flow is driven by the fluctuation propagator,  $\left(\Gamma_k^{(0,2)}[\bar{g}_{\mu\nu}, g_{\mu\nu}]\right)^{-1} = \left(\frac{\delta^2}{\delta g_{\mu\lambda} \delta g_{\nu\sigma}} \Gamma_k[\bar{g}_{\mu\nu}, g_{\mu\nu}]\right)^{-1}$ , or, equivalently,  $\left(\Gamma_k^{(0,2)}[\bar{g}_{\mu\nu}, h_{\mu\nu}]\right)^{-1}$ . As the regulator and gauge fixing break the symmetry between  $g_{\mu\nu}$  and  $\bar{g}_{\mu\nu}$ , this is not the same as  $\left(\Gamma_k^{(2,0)}[\bar{g}_{\mu\nu}, g_{\mu\nu}]\right)^{-1}$ . Schematically,

$$\partial_t \Gamma_k[\Phi_{\text{bck}}, \Phi_{\text{phys}}] = \frac{1}{2} \text{Tr} \left[ \left( \frac{\delta^2 \Gamma_k[\Phi_{\text{bck}}, \Phi_{\text{phys}}]}{\delta \Phi_{\text{phys}}^2} + R_k[\Phi_{\text{bck}}] \right)^{-1} \partial_t R_k[\Phi_{\text{bck}}] \right]. \quad (49)$$

In the background approximation, one equates  $\Phi_{\text{phys}} = \Phi_{\text{bck}}$  after the derivation of  $\frac{\delta^2 \Gamma_k[\Phi_{\text{bck}}, \Phi_{\text{phys}}]}{\delta \Phi_{\text{phys}}^2}$ . Accordingly, projections on field monomials pick up the auxiliary background-field dependence of the regulator in this approximation.

As an intermediate step to obtaining an effective action that respects background independence, one has to derive the flow of the fluctuation field propagator (Christiansen et al., 2017, 2016, 2015; Denz et al., 2018) in a setting that makes explicit use of a background. Alternatively, one can map this to a “bimetric” truncation, in which the propagator of the full metric is distinguished from the background metric, and drives the RG flow (Manrique and Reuter, 2010; Manrique et al., 2011b,c; Becker and Reuter, 2014), see **Table 3**.

The fluctuation-field dynamics are not protected by an auxiliary diffeomorphism invariance (as the background dynamics is). Accordingly, the theory space is that of a spin-2-field, with (modified) Slavnov-Taylor identities relating different couplings as a consequence of the symmetry. In a vertex expansion, this results in distinct “avatars” of couplings. For instance, expanding the Einstein-Hilbert action to  $n$ th order in the fluctuation field results in  $n$  “avatars” of the Newton coupling and cosmological constant,  $\lambda_n$  and  $G_n$ . **Table 4** lists fixed-point results for these “avatars”. We use the notation  $\mu = -2\lambda_2$  and also provide the fluctuation field anomalous

dimension  $\eta_h$  and ghost anomalous dimension  $\eta_c$ . Where their full momentum dependence has been evaluated, as in Christiansen et al. (2017); Christiansen et al. (2015); and Denz et al. (2018), the numbers provided refer to anomalous dimensions at vanishing momentum. “Hybrid” calculations, which evaluate the anomalous dimensions of the fluctuation fields, but equate the background and fluctuation Newton couplings,  $G_B$  and cosmological constants,  $\Lambda_B$  are included.

The example of a background-deformed regularization for scalar field theory shows how the background dependence of the regulator can spoil the study of fixed-point results for the Wilson-Fisher fixed point (Bridle et al., 2014). A symmetry identity, namely the shift Ward-identity, follows from background independence. It is structurally similar to the flow equation and relates the background-field-dependence on  $\bar{\phi}$  and the fluctuation-field-dependence on  $\phi$  of the flowing action (Reuter and Wetterich, 1997; Litim and Pawłowski, 2002; Bridle et al., 2014; Safari, 2016),

$$\frac{\delta \Gamma_k}{\delta \bar{\phi}} - \frac{\delta \Gamma_k}{\delta \phi} = \frac{1}{2} \text{Tr} \left[ \left( \frac{\delta^2 \Gamma_k}{\delta \phi^2} + R_k[\bar{\phi}] \right)^{-1} \frac{\delta R_k[\bar{\phi}]}{\delta \bar{\phi}} \right]. \quad (50)$$

Imposing the shift Ward-identity allows to recover background-independent results (Bridle et al., 2014). In a similar spirit, studies imposing the shift Ward identity in background- approximations for gravity (where the analog of Equation (50) is supplemented by contributions from the gauge fixing sector) have been performed in Morris (2016); Percacci and Vacca (2017); Labus et al. (2016); Ohta (2017); and Nieto et al. (2017).

Dynamical triangulations are background-independent as there is no preferred configuration and even the foliation structure in CDTs appears to be dispensable (Jordan and Loll, 2013). Therefore, establishing whether a universal continuum limit exists in the same universality class (i.e., with matching physical critical exponents) as FRG studies indicate, tests background independence of asymptotically safe gravity. One can either approach this by the well-tested method of computer simulations, based on a Monte-Carlo approach, or explore tensor models (see Sec. 4.2.4).

#### 4.2.4. The RG Perspective on (Discrete) Quantum Gravity

The use of RG ideas in quantum gravity has been gaining traction in various forms over the last few years. Interacting fixed points play a role in several different approaches (see, e.g., Eichhorn and Kosłowski, 2013, 2018; Ambjorn et al., 2014; Benedetti et al., 2015; Bahr and Steinhaus, 2016, 2017; Dittrich et al., 2016a,b; Dittrich, 2017; Ben Geloun et al., 2018) and references therein. In particular, in models that introduce a discretization, RG tools enable searches for a universal continuum limit encoded in RG fixed points. As one example, consider tensor models. These are spacetime-free models which encode the gluing of fundamental building blocks of a triangulation in their combinatorics. They generate the sum over all simplicial pseudomanifolds (triangulations) through their Feynman-diagram expansion, thereby generalizing the success-story of matrix models (Di Francesco et al., 1995)

**TABLE 4 |** Fixed-point results for fluctuation couplings.

References	Gauge	Regulator	Bckr.	$\mu^*$	$\lambda_3$	$G_{3^*}$	$G_{4^*}$	$\eta_h$	$\eta_c$	$\text{Re } \theta_1$	$\text{Re } \theta_2$	$\text{Re } \theta_3$
Groh and Saueressig, 2010	$\beta = \alpha = 1$	Litim	sphere	$\Lambda_B = 0.14$	-	$G_B = 0.86$	-	-	-1.77	1.94	1.94	-
Eichhorn and Gies, 2010	$\beta = \alpha = 0$	exp.	flat/sphere	$\Lambda_B = 0.32$	-	$G_B = 0.29$	-	-	-0.78	2.03	2.03	-
Eichhorn and Gies, 2010	$\beta = \alpha = 1$	exp.	flat/sphere	$\Lambda_B = 0.48$	-	$G_B = 0.18$	-	-	-1.31	1.39	1.39	-
Christiansen et al., 2017	$\beta = 1, \alpha = 0$	Litim	flat	-0.49	-	0.83	-	0.5	-1.37	1.87	1.87	1.37
Codello et al., 2014	$\alpha = \beta = 1$	Litim	flat	$\Lambda_B = -0.06$	-	$G_B = 1.62$	-	0.69	-1.36	4.12	4.12	-
Christiansen et al., 2015	$\beta = 1, \alpha = 0$	Litim	flat	-0.59	0.11	0.66	-	$\eta_h(\rho^2)$	$\eta_c(\rho^2)$	1.4	1.4	-14
Denz et al., 2018	$\beta = 1, \alpha = 0$	Litim	flat	-0.45	0.12	0.83	0.57	$\eta_h(\rho^2)$	$\eta_c(\rho^2)$	4.7	2.0	2.0
Knorr and Lippoldt, 2017	$\beta = 1, \alpha = 0$	Litim	curved	0.20	-0.008	0.20	-	-	-	1.65	1.65	-5.43
Christiansen et al., 2018	$\beta = 1, \alpha = 0$	Litim	curved	-0.38	-0.12	0.60	-	-	-	2.1	2.1	-3.5

We caution that where several “avatars” of a coupling are present these are related by STIs. Accordingly not all critical exponents are physical.

to higher dimensions (Ambjorn et al., 1991; Godfrey and Gross, 1991; Gross, 1992). A universal continuum limit might exist if the couplings are tuned to critical values while the tensor size  $N$  is taken to infinity (Gurau, 2011). This limit corresponds to a fixed point of an abstract, non-local RG flow set up in the tensor size  $N$ , Brezin and Zinn-Justin (1992) and Eichhorn and Koslowski (2013). This coarse-graining flow goes from many degrees of freedom (large  $N$ ), to fewer degrees of freedom (small  $N$ ). It is background independent by making no reference to locality or spacetime. Therefore, if a viable fixed point, leading to a physically acceptable phase of spacetime (where the “emergent” spacetime is four dimensional at large scales) can be identified, this provides an indication for a universal continuum limit - i.e., asymptotic safety - in a background independent setting. In Eichhorn and Koslowski (2013) an FRG approach was proposed for matrix models and generalized for tensor models in Eichhorn and Koslowski (2018), also triggering activity in related group field theories (e.g., Benedetti et al., 2015; Ben Geloun et al., 2018).

#### 4.2.5. Toward Asymptotically Safe Phenomenology in Astrophysics and Cosmology

As a candidate for a model of quantum spacetime, asymptotic safety should explain the structure of spacetime in the very early universe (see Bonanno and Saueressig, 2017) for a review and in those regions of black-hole spacetimes that contain classical curvature singularities. Within the FRG language, the UV physics is encoded in the limit of the full effective action  $\Gamma_{k \rightarrow 0}$  in which *physical* scales, e.g., curvature scales, are taken to (trans)planckian values. External physical scales can act as an IR cutoff for quantum fluctuations, as is most easily seen for the external momenta in scattering processes. This motivates the use of “RG improvement” techniques that provide quantum-gravity “inspired” models. The RG-improvement is performed by upgrading all couplings to running couplings and subsequently identifying  $k$  with a physical scale of the system in question, either at the level of the action, the equations of motion or the classical solutions. In settings with a high degree of symmetry and correspondingly a single physical scale, the identification is unique and dictated by dimensional arguments (e.g.,  $k^2 \sim R$  is the unique choice for a deSitter-type setting). “RG improved” results indicate dimensional reduction of the spectral

dimension (Lauscher and Reuter, 2005; Reuter and Saueressig, 2011; Calcagni et al., 2013), singularity resolution in black holes (Bonanno and Reuter, 1999, 2000, 2006; Falls et al., 2012; Becker and Reuter, 2012; Falls and Litim, 2014; Koch and Saueressig, 2014a,b; Kofinas and Zarikas, 2015; Pawłowski and Stock, 2018; Adeifeoba et al., 2018), finite entanglement entropy (Pagani and Reuter, 2018) as well as an inflationary regime generated through quantum gravity effects (Bonanno and Reuter, 2008; Bonanno et al., 2011a,b; Reuter and Saueressig, 2013; Kofinas and Zarikas, 2016).

## 5. ASYMPTOTICALLY SAFE QUANTUM GRAVITY AND MATTER

Our universe contains gravitational and matter degrees of freedom which are coupled to each other. Thus, to understand the quantum structure of spacetime in our universe it is neither necessary nor sufficient to show consistency of quantum-gravity models disregarding matter. This does not imply that quantum gravity must be a unified theory of all interactions, or that it needs to contain matter as fundamental degrees of freedom. It simply means that at observationally accessible scales accessible, all degrees of freedom, gravitational and matter, must be accounted for and their predicted dynamics compatible with observations. As the Standard Model contains  $N_S = 4$  scalars,  $N_V = 12$  and  $N_D = 24$  fermions (including right-handed neutrinos) but there is only one metric, the microscopic gravitational dynamics might even be well-approximated by the dynamics obtained from an appropriate large  $N_i$  approximation.

The measured Higgs mass of  $M_h \approx 125$  GeV (Aad, 2012; Chatrchyan, 2012) lies within a narrow band where no new physics is required for the consistency of the Standard Model below the Planck scale. A Higgs mass higher than about 180 GeV (Hambye and Riessellmann, 1997) leads to Landau-pole type behavior in the quartic coupling below the Planck scale. In the absence of higher-order terms in the Higgs potential (see e.g., Branchina and Messina, 2013; Gies et al., 2014; Eichhorn et al., 2015) the lower bound on the Higgs mass from absolute vacuum stability lies at  $M_h = 129$  GeV in a three-loop study for  $\Delta_{NP} = M_{Pl}$ , (Bezrukov et al., 2012), rendering the electroweak

vacuum metastable. As its lifetime exceeds the age of the universe (Elias-Miro et al., 2012), see Markkanen et al. (2018) for a review, the next scale of new physics for the Standard Model could be the Planck scale. Such a “desert” provides an exciting opportunity for quantum gravity: The initial conditions for the RG flow of matter interactions are set by quantum gravity at the Planck scale. In the absence of the “desert,” new physics at intermediate scales could shield the quantum gravity scale from view. Conversely, in a “desert”-like setting, there is a direct link between Planck-scale physics and electroweak-scale physics.

## 5.1. Impact of Quantum Gravity on Matter

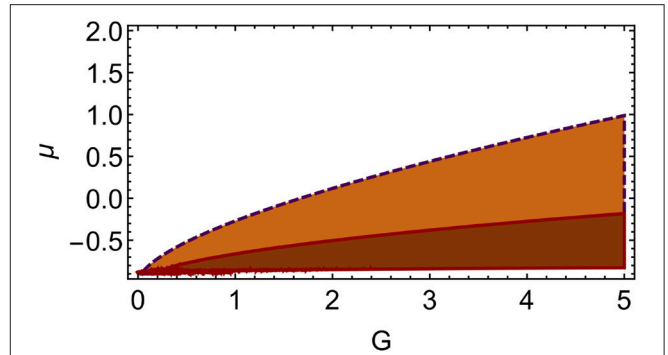
There are two effects of asymptotically safe gravity on matter in truncated FRG studies. Firstly, it generates nonzero fixed-point values for particular higher-order matter couplings, see section 5.1.1. Secondly, it impacts the scale dependence of the canonically marginal Standard Model couplings, see section 5.1.2. Both effects result in observational consistency constraints on the microscopic gravitational parameter space.

### 5.1.1. Matter Interacts in the Presence of Asymptotically Safe Quantum Gravity

The interacting nature of the asymptotically safe gravitational dynamics percolates into the matter sector. There cannot be UV fixed point with *all* matter interactions set to zero, Eichhorn and Gies (2011)<sup>11</sup> <sup>12</sup>. Interactions respecting the global symmetries of the kinetic terms for matter fields cannot be set to zero consistently (Eichhorn and Held, 2017). Finite contributions to their beta functions are generated by gravitational fluctuations. These prevent a free fixed point, as they are independent of the matter coupling and instead scale with the Newton coupling  $G$  (Eichhorn and Gies, 2011; Eichhorn, 2012; Eichhorn et al., 2016; Christiansen and Eichhorn, 2017; Eichhorn and Held, 2017; Eichhorn et al., 2018e). Thus, the free fixed point that exists in the limit of vanishing Newton coupling,  $G \rightarrow 0$ , is shifted to a finite value, the shifted Gaussian fixed point (sGFP). Matter couplings  $\bar{\chi}$  invariant under the global symmetries of the kinetic terms<sup>13</sup> feature canonical dimensions  $d_{\bar{\chi}} < 0$  in  $d = 4$ . Schematically, the FRG beta function reads

$$\beta_{\chi} = -d_{\bar{\chi}}\chi + \#_1 G_{\text{eff}}\chi + \#_2 G_{\text{eff}}^2 + \#_3 \chi^{\#}. \quad (51)$$

We focus on  $\# = 2$  (see Eichhorn and Gies, 2011; Eichhorn, 2012; Eichhorn et al., 2016; Christiansen and Eichhorn, 2017; Eichhorn



**FIGURE 10 |** The weak gravity bound in the  $(G, \mu = -2\Lambda)$  plane for the Yang-Mills system (from Christiansen and Eichhorn, 2017) in orange, bounded by the dark dashed line, and the weak gravity bound in scalar-fermion systems (from Eichhorn et al., 2016; Eichhorn and Held, 2017) in dark red, bounded by the red continuous line, lie close to each other.

and Held, 2017).  $G_{\text{eff}} = \frac{G}{1+\mu} - \frac{G}{(1+\mu^2)}$  parameterizes the effective strength of gravity fluctuations in the Einstein-Hilbert truncation (see Eichhorn and Held, 2017) for higher-order terms. The fixed points are

$$\chi_{1/2*} = \frac{d - \#_1 G_{\text{eff}} \pm \sqrt{-4\#_2 \#_3 G_{\text{eff}}^2 + (\#_1 G_{\text{eff}} - d_{\bar{\chi}})^2}}{2\#_3}, \quad (52)$$

such that  $\chi_{1*}$  is the sGFP. For  $\text{sign}\#_3 = \text{sign}\#_2$ , these two fixed points collide at

$$G_{\text{eff,crit}} = \frac{d_{\bar{\chi}}}{\#_1 - 2\sqrt{\#_2 \#_3}}. \quad (53)$$

Beyond, the sGFP is complex, thus  $G > G_{\text{eff,crit}}$  is inconsistent. As  $G_{\text{eff}}$  measures the effective strength of gravity fluctuations,  $G_{\text{eff,crit}}$  marks the (truncation dependent) weak-gravity bound. Once gravitational fluctuations exceed this bound, cf. **Figure 10**, they trigger novel divergences in matter couplings, restricting the viable microscopic parameter space to the remaining region. As the induced matter couplings are canonically irrelevant, they are power-law suppressed below the Planck scale and presumably irrelevant for particle physics at lower scales.

### 5.1.2. A Link That Could Matter: From the Planck Scale to the Electroweak Scale

Asymptotically safe gravity could uniquely fix the values of marginally irrelevant Standard Model (SM)-couplings (Abelian gauge couplings, Yukawas, Higgs quartic) at the Planck scale. This might allow to confront asymptotic safety with observations, since those couplings run logarithmically below the Planck scale, retaining a “memory” of their Planck-scale values.

For the marginal SM couplings  $g_{\text{SM}}$ , the quantum-gravity contribution to  $\beta_{g_{\text{SM}}}$  is linear in  $g_{\text{SM}}$ , as the gravitational RG flow cannot generate the SM interactions once they are set to zero due to their distinct symmetry structure (Eichhorn and Held, 2017). Technically, this is encoded in the diagrams underlying the FRG

<sup>11</sup>That four-fermion interactions are generated by quantum gravity fluctuations but remain finite implies that chiral symmetry, protecting the light fermions of the Standard Model, remains intact (see also Meibohm and Pawłowski, 2016; Eichhorn and Held, 2017). The effective background curvature in the UV can nevertheless break chiral symmetry (Gies and Martini, 2018).

<sup>12</sup>In  $d \neq 4$ , where specific matter models feature interacting fixed points, it is an intriguing question whether a new, combined universality class for matter and asymptotically safe gravity exists (see e.g., Elizalde et al., 1996; Percacci and Vacca, 2015; Labus et al., 2016).

<sup>13</sup>Notwithstanding arguments that suggest that quantum gravity should break global symmetries (Kallosh et al., 1995), studies of the FRG flow in truncations indicate the opposite result. This might be tied to the potential existence of black-hole remnants in asymptotic safety (Bonanno and Reuter, 2000; Falls et al., 2012). This implies that, e.g., Standard-Model couplings do not feature a contribution  $\sim \#_2$ , only a term  $\sim \#_1$ .

flow (see, e.g., Eichhorn et al., 2016; Christiansen and Eichhorn, 2017; Eichhorn and Held, 2017; Eichhorn, 2018; Eichhorn and Versteegen, 2018). Hence, the quantum-gravity contribution is

$$\beta_{g_{\text{SM}}}\bigg|_{\text{grav}} = -f_{g_{\text{SM}}} g_{\text{SM}}, \quad (54)$$

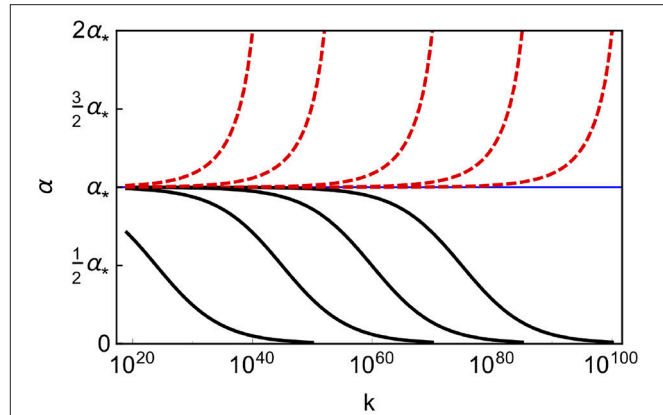
where  $f_{g_{\text{SM}}} \sim G$  is the contribution of metric fluctuations to the corresponding interaction vertex and additionally contains the gravity contribution to the anomalous dimensions of the matter fields. This contribution acts like a scaling dimension, i.e., like an effective change in spacetime dimensionality. For canonically irrelevant couplings, a UV completion requires  $f_{g_{\text{SM}}} > 0$ , resembling an effective *dimensional reduction*. It is unclear whether and how this fits with other indications for dimensional reduction in quantum gravity (Carlip, 2017).

Asymptotic freedom in non-Abelian gauge theories is a key cornerstone in the construction of the SM. This property could persist, as

$$\beta_g = -f_g g - \#_g g^3, \dots \quad (55)$$

for gauge couplings  $g$ , where  $f_g \geq 0$  holds in all FRG studies to date (Daum et al., 2010; Folkerts et al., 2012; Harst and Reuter, 2011; Christiansen and Eichhorn, 2017; Eichhorn and Versteegen, 2018; Christiansen et al., 2017).  $\#_g$  depends on the gauge group and matter content while the gravity contribution is blind to the internal index structure and accordingly gauge-group independent. Additional gravity contributions are indirect ones, arising through quantum-gravity-induced higher-order interactions which couple into the flow of the gauge coupling (Christiansen and Eichhorn, 2017) (note that the sign of the  $w$ -term in  $\eta$  is incorrect; accordingly this indirect contribution strengthens asymptotic freedom.)

The non-universality of beta functions, setting in at three loops for dimensionless couplings, starts at leading order for dimensionful couplings. Hence, the gravity contributions to beta functions in approximations differ in different schemes (see Robinson and Wilczek, 2006; Pietrykowski, 2007, 2013; Toms, 2007, 2008, 2009, 2010, 2011; Ebert et al., 2008, 2009; Mackay and Toms, 2010; Rodigast and Schuster, 2010; Anber et al., 2011; Felipe et al., 2011; Anber and Donoghue, 2012; Ellis et al., 2012; Narain and anishetty, 2013; Gonzalez-Martin and Martin, 2017) for perturbative studies. At the level of observables, such dependences must cancel. The same physics is encoded in different ways in distinct schemes. As the FRG is applicable to settings with dimensionful couplings (including a multitude of extensively probed universality classes in statistical physics), one could argue that it is well-suited to explore quantum gravity in simpler approximations. The non-universality of the gravity-contribution is reflected in the regulator-dependence of  $f_g$  in truncations: Within a background-field study,  $f_g = \frac{6}{\pi} G \Phi_1^1(0)$  (Daum et al., 2010), where  $\Phi_1^1(0) > 0$  always holds, but the value depends on the choice of regulator, e.g.,  $\Phi_1^1(0) = 1$  for the Litim-cutoff and  $\Phi_1^1(0) = \pi^2/6$  for the exponential cutoff. This dependence is expected to cancel against regulator-dependence of gravitational fixed-point values (at least at the level of physical observables).



**FIGURE 11 |** The transplanckian RG flow for  $\alpha = g_Y^2/(4\pi)$  described by Equation (56) features trajectories emanating from the free fixed point (black, continuous line), which approach the interacting fixed point at  $\alpha_*$ . One unique trajectory (blue, thin line) is the fixed-point trajectory for the interacting fixed point. UV unsafe trajectories are pulled toward the IR fixed point as well (red, dashed lines). Adapted from Eichhorn et al. (2018d).

For the Abelian gauge coupling the free fixed point is IR attractive in the absence of gravity, such that the observation of a nonvanishing Abelian gauge coupling in the IR presumably prevents an asymptotically free UV completion of the SM (Gell-Mann and Low, 1954). The quantum gravity contribution is the same as in the non-Abelian case (cf. Equation 55), thus

$$\beta_{g_Y} = -f_g g_Y + \frac{41}{6} \frac{g_Y^3}{16\pi^2} + \dots \quad (56)$$

Fixed points of Equation (56) lie at

$$g_{Y,*1} = 0, \quad g_{Y,*2} = \sqrt{\frac{f_g 6 \cdot 16\pi^2}{41}}. \quad (57)$$

The first is IR repulsive, the second IR attractive. If it lies at small enough values, then higher-order terms remain negligible and Equation (56) suffices to analyze the consequences. According to Equation (56), the IR repulsive fixed point at  $g_{Y,*1} = 0$  can be connected to a range of values for  $g_Y$  at the Planck scale<sup>14</sup>. However, no values above an upper bound,  $g_Y = g_{Y,*2}$ , can be reached, as  $g_{Y,*2}$  is IR attractive, cf. **Figure 11**. Only one unique trajectory emanates from  $g_{Y,*2}$ . Along this trajectory,  $g_Y(k) = \text{const}$  until quantum-gravity contributions switch off below the Planck scale, where  $f_g$  quickly drops to tiny values and SM fields drive the flow. Unlike in the SM without the gravity-extension, the initial condition for the RG flow of  $g_Y$  is fixed at the Planck scale. Testing whether this results in an observationally viable value at the electroweak scale constitutes a strong observational test of the model. It also highlights that confronting quantum gravity with observations might be possible without reaching Planckian energies.

<sup>14</sup>Couplings that are asymptotically free, not asymptotically safe, already run at transplanckian scales.

The fixed-point structure underlying such “retrodictions” was found for the Abelian gauge coupling in Harst and Reuter (2011), further explored in Eichhorn and Versteegen (2018), cf. right panel in **Figure 12** and extended to a GUT setting in Eichhorn et al. (2018d).

In Zanusso et al. (2010); Vacca and Zanusso (2010); Oda and Yamada (2016); Eichhorn et al. (2016); Eichhorn and Held (2017); and Hamada and Yamada (2017), the gravity-contribution  $f_y$  to the Yukawa sector was calculated. Using beta functions of the form

$$\beta_{y_{t(b)}} = \frac{y_{t(b)}}{16\pi^2} \left( \frac{3y_{b(t)}^2}{2} + \frac{9y_{t(b)}^2}{2} - \frac{9}{4}g_2^2 - 8g_3^2 \right) - f_y y_{t(b)} - \frac{3y_{t(b)}}{16\pi^2} \left( \frac{1}{36} + Y_{t(b)}^2 \right) g_Y^2, \quad (58)$$

for the quarks of the third generation, with  $Y_t = 2/3$ ,  $Y_b = -1/3$ , supplemented by the assumption that the gauge sector of the SM is asymptotically free, and gravitational fixed-point values from a background-approximation results in a uniquely fixed top mass of about 170 GeV (Eichhorn and Held, 2018), cf. left panel of **Figure 12**.

Intriguingly, the SM beta functions with gravity in the approximation defined by Equations (57, 58) also admit an interacting fixed point such that the top Yukawa, bottom Yukawa and Abelian gauge coupling are fixed uniquely. They reach IR values in the vicinity of the observed ones, if the two gravity contributions  $f_g$  and  $f_y$  take appropriate values (Eichhorn and Held, 2018). In this scenario, the difference between top mass and bottom mass is generated through an interacting fixed point induced by gravity due to their different charges.

The fixed-point structure could be simpler in the scalar sector. Asymptotically safe quantum gravity flattens the Higgs potential: If all other SM couplings are asymptotically free, a fixed point at vanishing Higgs potential exists in line with intact shift-symmetry (Eichhorn and Held, 2017). It is IR attractive (Narain and Percacci, 2010; Percacci and Vacca, 2015; Labus et al., 2016; Oda and Yamada, 2016; Hamada and Yamada, 2017; Eichhorn, 2018). This extends to the Higgs portal coupling to scalar dark matter (Eichhorn, 2018). Taking the corresponding fixed-point values  $\lambda_{h*} = 0$  (for the Higgs quartic) and  $\lambda_{h\chi*} = 0$  (for the Higgs portal coupling) as initial conditions for the RG flow at the Planck scale, and setting all SM couplings to their observationally preferred Planck-scale values, one reaches a Higgs mass in the vicinity of the observed value, while the Higgs portal coupling remains zero at all scales. The first is a prediction (Shaposhnikov and Wetterich, 2010) put forward before the discovery of the Higgs at the LHC (Aad, 2012; Chatrchyan, 2012), see also Bezrukov et al. (2012). The second appears to be consistent with the non-detection of a scalar Higgs portal through direct searches (Athron et al., 2017; Aprile et al., 2018).

“Retrodictions” of SM couplings could be a much more generic consequence of quantum gravity than just of asymptotic safety as discussed in section 2.3.

## 5.2. Impact of Matter on Quantum Gravity – Backreaction Matters?

The impact of quantum fluctuations of matter on the gravitational fixed point has been studied in simple truncations. The corresponding theory space also contains non-minimal matter-curvature couplings (Narain and Percacci, 2010; Percacci and Vacca, 2015; Eichhorn and Lippoldt, 2017; Eichhorn et al., 2018e).

Matter fields deform the gravitational fixed point in truncations. Adding a small number of matter fields leads to the continued existence of a viable interacting fixed point. At larger number of matter fields, there are indications that further extensions of the truncation could be required (Meibohm et al., 2016; Eichhorn et al., 2018b).

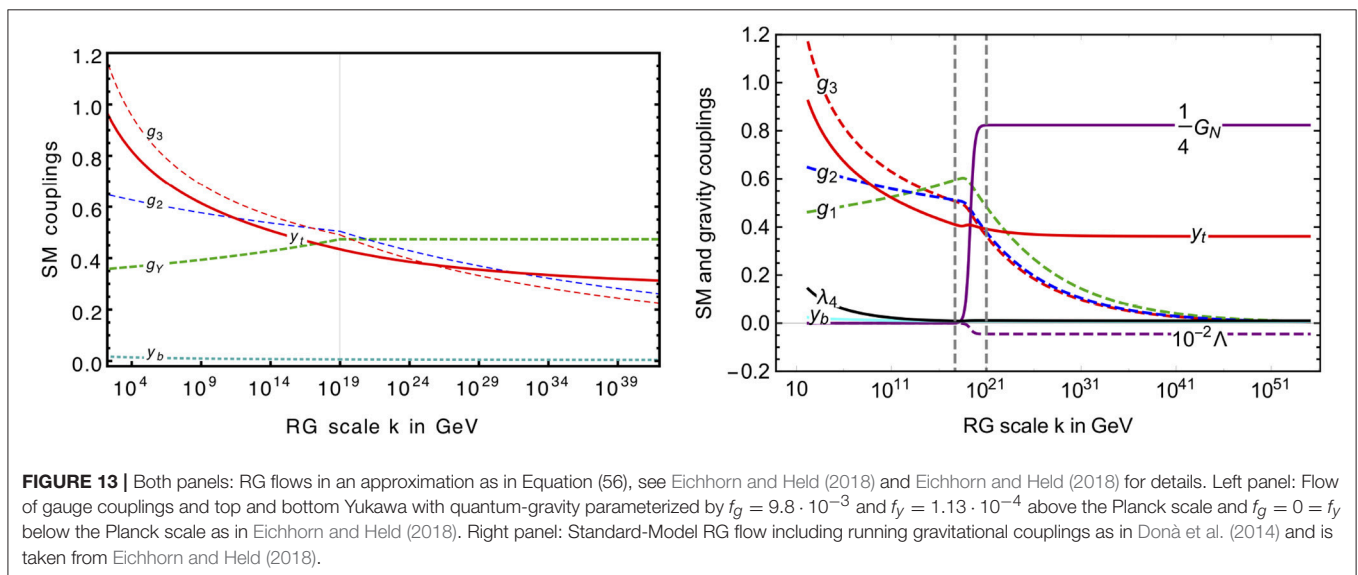
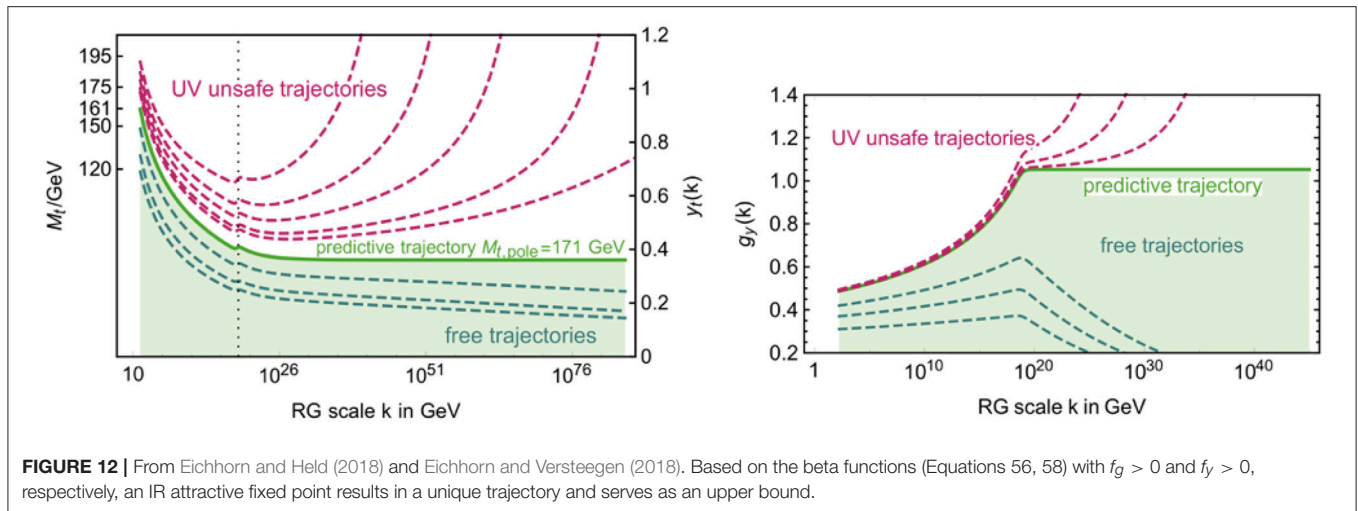
Assuming that asymptotic safety in gravity is driven by antiscreening metric fluctuations inducing a fixed point in the Newton coupling, the matter contribution to  $\beta_G$  is critical. Specifically,

$$\beta_G \Big|_{\text{matter}} = N_S G^2 a_S + N_D G^2 a_D + N_V G^2 a_V, \quad (59)$$

where  $a_S > 0$  (Donà et al., 2014; Percacci and Vacca, 2015; Meibohm et al., 2016; Labus et al., 2016; Donà et al., 2016; Biemans et al., 2017; Alkofer and Saueressig, 2018; Eichhorn et al., 2018b), agreeing with perturbative studies for  $d = 2 + \epsilon$  dimensions (Christensen and Duff, 1978) and studies of the one-loop effective action using heat-kernel techniques (Kabat, 1995; Larsen and Wilczek, 1996). Similarly, fermions screen the Newton coupling<sup>15</sup>  $a_D > 0$  (Donà et al., 2014; Meibohm et al., 2016; Eichhorn et al., 2018c), in agreement with perturbative studies (Kabat, 1995; Larsen and Wilczek, 1996). For vectors,  $a_V < 0$  (Donà et al., 2014; Christiansen et al., 2017; Biemans et al., 2017; Alkofer and Saueressig, 2018; Eichhorn et al., 2018c), also found with perturbative techniques (Kabat, 1995; Larsen and Wilczek, 1996). Background and fluctuation results are in agreement on this result (for fluctuation results, it is crucial to include the anomalous dimensions Donà et al., 2016; Eichhorn et al., 2018b).

A strong indication for (near-perturbative) asymptotic safety in matter-gravity systems comes from a comparison (Eichhorn et al., 2018b,c) of distinct “avatars” of the Newton coupling (Donà et al., 2016). It can be read off from the three-graviton vertex as well as gravity-matter vertices, just like the gauge coupling in gauge theories. For a dimensionless gauge coupling in the perturbative regime, two-loop universality equates the different avatars. Beyond perturbation theory, the Slavnov-Taylor-identities relating the avatars become nontrivial. Simply put, the stronger quantum effects are, the less trivial are the relation of classically equal couplings. Eichhorn et al. (2018b,c)

<sup>15</sup>For the background Newton coupling, this is more subtle: Choosing to impose the regulator on the spectrum of  $\nabla^2$ , or on  $\nabla^2 = \nabla^2 - R/4$  results in a different sign of the fermionic contribution to the running of  $G$  (Donà and Percacci, 2013 see also Alkofer, 2018; Alkofer and Saueressig, 2018). This highlights that the (unphysical) background-field dependence of the regulator can alter results in the background approximation in simple truncations, suggesting the need for a fluctuation calculation. The first choice agrees with the result from fluctuations calculations.



observe an effective universality of distinct avatars of the Newton coupling, which agree within an estimate of the systematic truncation error. This signals a near-perturbative nature of asymptotically safe gravity. Further, the delicate cancellations required between different contributions to the beta functions in order to achieve effective universality strongly point toward a physical fixed point instead of a truncation artifact.

## 6. OUTLOOK

Asymptotically safe models are of inherent theoretical interest when it comes to a comprehensive understanding of fundamental quantum field theories. Exciting progress in the last few years even hints at a possibility of asymptotically safe extensions of the Standard Model – with or without gravity. In quantum gravity, the idea of asymptotic safety resonates with a wider effort to analyze quantum spacetime from a Renormalization Group point of view. Hence, the many intriguing open questions that

remain to be answered in this area appear worth tackling, and the new (asymptotically safe) perspective on high-energy physics is exciting and potentially useful to explore.

## AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

## ACKNOWLEDGMENTS

I would like to thank all participants and speakers of the workshop Asymptotic safety in a dark universe at Perimeter Institute in June 2018, the conference Quantum spacetime and the Renormalization Group in Bad Honnef in June 2018 and the workshop on Quantum fields – from fundamental question to phenomenological applications at MITP in September 2018 for

stimulating discussions, and in particular acknowledge helpful discussions with L. Freidel, D. Litim, J. Pawłowski, M. Reuter, F. Sannino and C. Wetterich that have informed this review. It is a particular pleasure to thank all current and past members

of my research group for successful collaborations, inspiring discussions and for joining in the (only asymptotically safe) quest for quantum gravity. I acknowledge support by the DFG under grant no. Ei/1037-1.

## REFERENCES

- Aad, G. (2012). Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Phys. Lett. B* 2012:716. doi: 10.1016/j.physletb.2012.08.020
- Abel, S., and Sannino, F. (2017a). Radiative symmetry breaking from interacting UV fixed points. *Phys. Rev. D* 96:056028. doi: 10.1103/PhysRevD.96.056028
- Abel, S., and Sannino, F. (2017b). Framework for an asymptotically safe standard model via dynamical breaking. *Phys. Rev. D* 96:055021. doi: 10.1103/PhysRevD.96.055021
- Adeifeoba, A., Eichhorn, A., and Platania, A. (2018). Towards conditions for black-hole singularity-resolution in asymptotically safe quantum gravity. *arXiv:1808.03472 [gr-qc]*.
- Aida, T., and Kitazawa, Y. (1997). Two loop prediction for scaling exponents in (2+epsilon)-dimensional quantum gravity. *Nucl. Phys. B* 491:427. doi: 10.1016/S0550-3213(97)00091-6
- Aida, T., Kitazawa, Y., Kawai, H., and Ninomiya, M. (1994). Conformal invariance and renormalization group in quantum gravity near two-dimensions. *Nucl. Phys. B* 427:158. doi: 10.1016/0550-3213(94)90273-9
- Akerib, D. S., Alsum, S., Araújo, H. M., Bai, X., Bailey, A. J., Balajthy, J., et al. (2017). Results from a search for dark matter in the complete LUX exposure. *Phys. Rev. Lett.* 118:021303. doi: 10.1103/PhysRevLett.118.021303
- Alanne, T., and Blasi, S. (2018). The abelian gauge-Yukawa  $\beta$ -functions at large  $N_f$ . *arXiv:1808.03252 [hep-ph]*.
- Alanne, T., and Blasi, S. (2018). The  $\beta$ -function for Yukawa theory at large  $N_f$ . *JHEP* 1808:081.
- Alkofer, N. (2018). Asymptotically safe  $f(R)$ -gravity coupled to matter II: global solutions. *arXiv:1809.06162 [hep-th]*.
- Alkofer, N., and Saueressig, F. (2018). Asymptotically safe  $f(R)$ -gravity coupled to matter I: the polynomial case *arXiv:1802.00498 [hep-th]*
- Altarelli, G., and Isidori, G. (1994). Lower limit on the Higgs mass in the standard model: an Update. *Phys. Lett. B* 337:141. doi: 10.1016/0370-2693(94)91458-3
- Ambjorn, J., Coumbe, D., Gizbert-Studnicki, J., Gorlich, A., and Jurkiewicz, J. (2017). New higher-order transition in causal dynamical triangulations. *Phys. Rev. D* 95:124029. doi: 10.1103/PhysRevD.95.124029
- Ambjorn, J., Durhuus, B., and Jonsson, T. (1991). Three-dimensional simplicial quantum gravity and generalized matrix models. *Mod. Phys. Lett. A* 6:1133. doi: 10.1142/S0217732391001184
- Ambjorn, J., Görlich, A., Jurkiewicz, J., Kreienbuehl, A., and Loll, R. (2014). Renormalization group flow in CDT. *Class. Quant. Grav.* 31:165003. doi: 10.1088/0264-9381/31/16/165003
- Ambjorn, J., Jordan, S., Jurkiewicz, J., and Loll, R. (2011). A second-order phase transition in CDT. *Phys. Rev. Lett.* 107:211303. doi: 10.1103/PhysRevLett.107.211303
- Ambjorn, J., Jordan, v, Jurkiewicz, J., and Loll, R. (2012). Second- and first-order phase transitions in CDT. *Phys. Rev. D* 85:124044. doi: 10.1103/PhysRevD.85.124044
- Ambjorn, J., Jurkiewicz, J., and Loll, R. (2000). A nonperturbative lorentzian path integral for gravity. *Phys. Rev. Lett.* 85:924. doi: 10.1103/PhysRevLett.85.924
- Ambjorn, J., Jurkiewicz, J., and Loll, R. (2001). Dynamically triangulating lorentzian quantum gravity. *Nucl. Phys. B* 610:347. doi: 10.1016/S0550-3213(01)00297-8
- Anber, M. M., and Donoghue, J. F. (2012) On the running of the gravitational constant. *Phys. Rev. D* 85:104016. doi: 10.1103/PhysRevD.85.104016
- Anber, M. M., Donoghue, J. F., and El-Houssieny, M. (2011) Running couplings and operator mixing in the gravitational corrections to coupling constants. *Phys. Rev. D* 83:124003. doi: 10.1103/PhysRevD.83.124003
- Antipin, O., Dondi, N. A., Sannino, F., Thomsen, A. E., and Wang, Z. W. (2018a). Gauge-Yukawa theories: beta functions at large  $N_f$ . *arXiv:1803.09770 [hep-ph]*.
- Antipin, O., Dondi, N. A., Sannino, F., and Thomsen, A. E. (2018b). The  $a$ -theorem at large  $N_f$ . *arXiv:1808.00482 [hep-th]*.
- Antipin, O., Gillioz, M., Mølgaard, E., and Sannino, F. (2013). The  $a$  theorem for gauge-Yukawa theories beyond Banks-Zaks fixed point. *Phys. Rev. D* 87:125017. doi: 10.1103/PhysRevD.87.125017
- Antipin, O., and Sannino, F. (2017). Conformal window 2.0: the large  $N_f$  safe story. *arXiv:1709.02354 [hep-ph]*.
- Aprile, E., Aalbers, J., Agostini, F., Alfonsi, M., Althueser, L., Amaro, F. D., et al. (2018). Dark matter search results from a one ton-year exposure of XENON1T. *Phys. Rev. Lett.* 121:111302. doi: 10.1103/PhysRevLett.121.111302
- Arici, F., Becker, D., Ripken, C., Saueressig, F., and van Suijlekom, W. D. (2017). Reflection positivity in higher derivative scalar theories. *arXiv:1712.04308 [hep-th]*.
- Ashtekar, A., and Lewandowski, J. (1997). Quantum theory of geometry. I: Area operators. *Class. Q. Grav.* 14:A55. doi: 10.1088/0264-9381/14/1A/006
- Ashtekar, A., Reuter, M., and Rovelli, C. (2014). From general relativity to quantum gravity. *arXiv:1408.4336 [gr-qc]*.
- Athron, P., Balázs, C., Bringmann, T., Buckley, A., Chruszcz, M., Conrad, J., et al. (2017). Status of the scalar singlet dark matter model. *Eur. Phys. J. C* 77:568. doi: 10.1140/epjc/s10052-017-5113-1
- Avramidi, I. G., and Barvinsky, A. O. (1985). Asymptotic freedom in higher derivative quantum gravity. *Phys. Lett.* 159B:269. doi: 10.1016/0370-2693(85)90248-5
- Bahr, B., and Steinhaus, S. (2016). Numerical evidence for a phase transition in 4d spin foam quantum gravity. *Phys. Rev. Lett.* 117: 141302. doi: 10.1103/PhysRevLett.117.141302
- Bahr, B., and Steinhaus, S. (2017). Hypercuboidal renormalization in spin foam quantum gravity. *Phys. Rev. D* 95:126006. doi: 10.1103/PhysRevD.95.126006
- Bajc, B., Dondi, N. A., and Sannino, F. (2018). Safe sUSY. *JHEP* 1803:005. doi: 10.1007/JHEP03(2018)005
- Bajc, B., and Sannino, F. (2016). Asymptotically safe grand unification. *JHEP* 1612:141. doi: 10.1007/JHEP12(2016)141
- Banks, T., and Zaks, A. (1982). On the phase structure of vector-like gauge theories with massless fermions. *Nucl. Phys. B* 196:189. doi: 10.1016/0550-3213(82)90035-9
- Bardeen, W. A., Lee, B. W., and Shrock, R. E. (1976). Phase transition in the nonlinear  $\sigma$  Model in  $2 + \epsilon$  dimensional continuum. *Phys. Rev. D* 14:985. doi: 10.1103/PhysRevD.14.985
- Barducci, D., Fabbrichesi, M., Nieto, C. M., Percacci, R., and Skrinjar, V. (2018). In search of a UV completion of the standard model-78.000 models that don't work. *arXiv:1807.05584 [hep-ph]*. doi: 10.1007/JHEP11(2018)057
- Barnaby, N., and Kamran, N. (2008). Dynamics with infinitely many derivatives: the initial value problem. *JHEP* 0802:008. doi: 10.1088/1126-6708/2008/02/008
- Barvinsky, A. O., Blas, D., Herrero-Valea, M., Sibiryakov, S. M., and Steinwachs, C. F. (2017). Hořava gravity is asymptotically free (in 2+1 dimensions). *Phys. Rev. Lett.* 119:211301. doi: 10.1103/PhysRevLett.119.211301
- Becker, D., and Reuter, M. (2012). Running boundary actions, asymptotic safety, and black hole thermodynamics. *JHEP* 1207:172. doi: 10.1007/JHEP07(2012)172
- Becker, D., and Reuter, M. (2014). En route to background independence: broken split-symmetry, and how to restore it with bi-metric average actions. *Annals Phys.* 350:225. doi: 10.1016/j.aop.2014.07.023
- Becker, D., Ripken, C., and Saueressig, F. (2017). On avoiding Ostrogradski instabilities within Asymptotic Safety. *JHEP* 1712:121. doi: 10.1007/JHEP12(2017)121
- Ben Geloun, J., Kosłowski, T. A., Oriti, D., and Pereira, A. D. (2018). Functional renormalization group analysis of rank 3 tensorial group field

- theory: the full quartic invariant truncation. *Phys. Rev. D* 97:126018. doi: 10.1103/PhysRevD.97.126018
- Benedetti, D. (2012). Asymptotic safety goes on shell. *New J. Phys.* 14:015005. doi: 10.1088/1367-2630/14/1/015005
- Benedetti, D., Ben Geloun, J., and Oriti, D. (2015). Functional renormalisation group approach for tensorial group field theory: a rank-3 model. *JHEP* 1503:084. doi: 10.1007/JHEP03(2015)084
- Benedetti, D., and Caravelli, F. (2012). The local potential approximation in quantum gravity. *JHEP* 1206:017. doi: 10.1007/JHEP06(2012)017
- Benedetti, D., and Gurau, R. (2012). Phase transition in dually weighted colored tensor models. *Nucl. Phys. B* 855:420. doi: 10.1016/j.nuclphysb.2011.10.015
- Benedetti, D., Machado, P. F., and Saueressig, F. (2009). Asymptotic safety in higher-derivative gravity. *Mod. Phys. Lett. A* 24:2233. doi: 10.1142/S0217732309031521
- Berges, J., Tetradis, N., and Wetterich, C. (2002). Nonperturbative renormalization flow in quantum field theory and statistical physics. *Phys. Rept.* 363, 223–386. doi: 10.1016/S0370-1573(01)00098-9
- Bern, Z., Carrasco, J. J., Chen, W. M., Edison, A., Johansson, H., Parra-Martinez, J., et al. (2018). Ultraviolet properties of  $\mathcal{N} = 8$  supergravity at five Loops. *arXiv:1804.09311 [hep-th]*.
- Bern, Z., Carrasco, J. J. M., Chen, W. M., Johansson, H., Roiban, R., and Zeng, M. (2017). Five-loop four-point integrand of  $\mathcal{N} = 8$  supergravity as a generalized double copy. *Phys. Rev. D* 96:126012. doi: 10.1103/PhysRevD.96.126012
- Bezrukov, F., Kalmykov, M. Y., Kniehl, B. A., and Shaposhnikov, M. (2012). Higgs boson mass and new physics. *JHEP* 1210:140. doi: 10.1007/JHEP10(2012)140
- Bezrukov, F., and Shaposhnikov, M. (2015). Why should we care about the top quark Yukawa coupling? *J. Exp. Theor. Phys.* 120:335.
- Biemans, J., Platania, A., and Saueressig, F. (2017). Renormalization group fixed points of foliated gravity-matter systems. *JHEP* 1705:093. doi: 10.1007/JHEP05(2017)093
- Biemans, J., Platania, A., and Saueressig, F. (2017). Quantum gravity on foliated spacetimes: asymptotically safe and sound. *Phys. Rev. D* 95:086013. doi: 10.1103/PhysRevD.95.086013
- Bombelli, L., Lee, J., Meyer, D., and Sorkin, R. (1987). Space-time as a causal set. *Phys. Rev. Lett.* 59:521. doi: 10.1103/PhysRevLett.59.521
- Bonanno, A., Contillo, A., and Percacci, R. (2011a). Inflationary solutions in asymptotically safe  $f(R)$  theories. *Class. Quant. Grav.* 28:145026. doi: 10.1088/0264-9381/28/14/145026
- Bonanno, A., and Reuter, M. (1999). Quantum gravity effects near the null black hole singularity. *Phys. Rev. D* 60:084011. doi: 10.1103/PhysRevD.60.084011
- Bonanno, A., and Reuter, M. (2000). Renormalization group improved black hole space-times. *Phys. Rev. D* 62:043008. doi: 10.1103/PhysRevD.62.043008
- Bonanno, A., and Reuter, M. (2006). Spacetime structure of an evaporating black hole in quantum gravity. *Phys. Rev. D* 73:083005. doi: 10.1103/PhysRevD.73.083005
- Bonanno, A., and Reuter, M. (2008). Primordial entropy production and lambda-driven inflation from quantum Einstein gravity. *J. Phys. Conf. Ser.* 140:012008. doi: 10.1088/1742-6596/140/1/012008
- Bonanno, A., and Reuter, M. (2011b). Entropy production during asymptotically safe inflation. *Entropy* 13:274. doi: 10.3390/e13010274
- Bonanno, A., and Saueressig, F. (2017). Asymptotically safe cosmology—a status report. *Comptes Rendus Physique* 18:254. doi: 10.1016/j.crhy.2017.02.002
- Bond, A. D., Hiller, G., Kowalska, K., and Litim, D. F. (2017). Directions for model building from asymptotic safety. *JHEP* 1708:004. doi: 10.1007/JHEP08(2017)004
- Bond, A. D., and Litim, D. F. (2017). Theorems for asymptotic safety of Gauge theories. *Eur. Phys. J. C* 77:429. doi: 10.1140/epjc/s10052-017-4976-5
- Bond, A. D., and Litim, D. F. (2017). Asymptotic safety guaranteed in supersymmetry. *Phys. Rev. Lett.* 119:211601. doi: 10.1103/PhysRevLett.119.211601
- Bond, A. D., and Litim, D. F. (2018a). Price of asymptotic safety. *arXiv:1801.08527 [hep-th]*.
- Bond, A. D., and Litim, D. F. (2018b). More asymptotic safety guaranteed. *Phys. Rev. D* 97:085008. doi: 10.1103/PhysRevD.97.085008
- Bond, A. D., Litim, D. F., Medina Vazquez, G., and Steudtner, T. (2018). UV conformal window for asymptotic safety. *Phys. Rev. D* 97:036019. doi: 10.1103/PhysRevD.97.036019
- Bowman, P. O., Heller, U. M., Leinweber, D. B., Parappilly, M. B., Sternbeck, A., von Smekal, L., et al. (2007). Scaling behavior and positivity violation of the gluon propagator in full QCD. *Phys. Rev. D* 76:094505. doi: 10.1103/PhysRevD.76.094505
- Branchina, V., and Messina, E. (2013). Stability, higgs boson mass and new physics. *Phys. Rev. Lett.* 111:241801. doi: 10.1103/PhysRevLett.111.241801
- Braun, J. (2012). Fermion interactions and universal behavior in strongly interacting theories. *J. Phys. G* 39:033001. doi: 10.1088/0954-3899/39/3/033001
- Braun, J., Gies, H., and Scherer, D. D. (2011). Asymptotic safety: a simple example. *Phys. Rev. D* 83:085012. doi: 10.1103/PhysRevD.83.085012
- Brezin, E., and Zinn-Justin, J. (1992). Renormalization group approach to matrix models. *Phys. Lett. B* 288:54. doi: 10.1016/0370-2693(92)91953-7
- Bridle, I. H., Dietz, J. A., and Morris, T. R. (2017). The local potential approximation in the background field formalism. *JHEP* 1403:093. doi: 10.1007/JHEP03(2014)093
- Brown, L. S. (1977). Stress tensor trace anomaly in a gravitational metric: scalar fields. *Phys. Rev. D* 15:1469. doi: 10.1103/PhysRevD.15.1469
- Buttazzo, D., Degraffi, G., Giardino, P. P., Giudice, G. F., Sala, F., Salvio, A., et al. (2013). Investigating the near-criticality of the Higgs boson. *JHEP* 1312:089. doi: 10.1007/JHEP12(2013)089
- Buyukbese, T., and Litim, D. F. (2017). Asymptotic safety of gauge theories beyond marginal interactions. *PoS LATTICE 2016*:233. doi: 10.22323/1.256.0233
- Calcagni, G., Eichhorn, A., and Saueressig, F. (2013). Probing the quantum nature of spacetime by diffusion. *Phys. Rev. D* 87:124028. doi: 10.1103/PhysRevD.87.124028
- Campostrini, M., Pelissetto, A., Rossi, P., and Vicari, E. (1999). Improved high temperature expansion and critical equation of state of three-dimensional Ising-like systems. *Phys. Rev. E* 60:3526. doi: 10.1103/PhysRevE.60.3526
- Canet, L., Delamotte, B., Mouhanna, D., and Vidal, J. (2003). Nonperturbative renormalization group approach to the Ising model: a Derivative expansion at order partial\*\*4. *Phys. Rev. B* 68:064421. doi: 10.1103/PhysRevB.68.064421
- Cardy, J. L. (1988). Is there a c theorem in four-dimensions? *Phys. Lett. B* 215:749. doi: 10.1016/0370-2693(88)90054-8
- Carlip, S. (2017). Dimension and dimensional reduction in quantum gravity. *Class. Quant. Grav.* 34:193001. doi: 10.1088/1361-6382/aa8535
- Caswell, W. E. (1974). Asymptotic behavior of nonabelian gauge theories to two loop order. *Phys. Rev. Lett.* 33:244. doi: 10.1103/PhysRevLett.33.244
- Chatrchyan, S. (2012). Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Phys. Lett. B* 716:30. doi: 10.1016/j.physletb.2012.08.021
- Christensen, S. M., and Duff, M. J. (1978). Quantum gravity in two +  $\epsilon$  dimensions. *Phys. Lett.* 79B:213. doi: 10.1016/0370-2693(78)90225-3
- Christiansen, N., and Eichhorn, A. (2017). An asymptotically safe solution to the U(1) triviality problem. *Phys. Lett. B* 770:154. doi: 10.1016/j.physletb.2017.04.047
- Christiansen, N., Eichhorn, A., and Held, A. (2017). Is scale-invariance in gauge-Yukawa systems compatible with the graviton? *Phys. Rev. D* 96:084021. doi: 10.1103/PhysRevD.96.084021
- Christiansen, N., Falls, K., Pawłowski, J. M., and Reichert, M. (2018). Curvature dependence of quantum gravity. *Phys. Rev. D* 97:046007. doi: 10.1103/PhysRevD.97.046007
- Christiansen, N., Knorr, B., Pawłowski, J. M., and Reichert, M. (2015). Local Quantum Gravity. *Phys. Rev. D* 92:121501. doi: 10.1103/PhysRevD.92.121501
- Christiansen, N., Knorr, B., Pawłowski, J. M., and Rodigast, A. (2016). Global flows in quantum gravity. *Phys. Rev. D* 93:044036. doi: 10.1103/PhysRevD.93.044036
- Christiansen, N., Litim, D. F., Pawłowski, J. M., and Reichert, M. (2017). One force to rule them all: asymptotic safety of gravity with matter. *arXiv:1710.04669 [hep-th]*.
- Christiansen, N., Litim, D. F., Pawłowski, J. M., and Rodigast, A. (2017). Fixed points and infrared completion of quantum gravity. *Phys. Lett. B* 728:114. doi: 10.1016/j.physletb.2013.11.025
- Codello, A., Demmel, M., and Zanusso, O. (2014). Scheme dependence and universality in the functional renormalization group. *Phys. Rev. D* 90:027701. doi: 10.1103/PhysRevD.90.027701
- Codello, A., and D'Odorico, G. (2015). Scaling and renormalization in two dimensional quantum gravity. *Phys. Rev. D* 92:024026. doi: 10.1103/PhysRevD.92.024026

- Codello, A., D'Odorico, G., and Pagani, C. (2014). Consistent closure of renormalization group flow equations in quantum gravity. *Phys. Rev. D.* 89:081701. doi: 10.1103/PhysRevD.89.081701
- Codello, A., Langæble, K., Litim, D. F., and Sannino, F. (2016). Conformal Gauge-Yukawa theories away from four dimensions. *JHEP* 1607:118. doi: 10.1007/JHEP07(2016)118
- Codello, A., and Percacci, R. (2006). Fixed points of higher derivative gravity. *Phys. Rev. Lett.* 97:221301. doi: 10.1103/PhysRevLett.97.221301
- Codello, A., and Percacci, R. (2009). Fixed points of nonlinear sigma models in  $d > 2$ . *Phys. Lett. B* 672:280. doi: 10.1016/j.physletb.2009.01.032
- Codello, A., Percacci, R., and Rahmede, C. (2009). Investigating the ultraviolet properties of gravity with a wilsonian renormalization group equation. *Annals Phys.* 324:414. doi: 10.1016/j.aop.2008.08.008
- Cucchieri, A., Mendes, T., and Taurines, A. R. (2005). Positivity violation for the lattice Landau gluon propagator. *Phys. Rev. D* 71:051902. doi: 10.1103/PhysRevD.71.051902
- Daum, J.-E., and Reuter, M. (2012). Renormalization group flow of the holst action. *Phys. Lett. B* 710:215. doi: 10.1016/j.physletb.2012.01.046
- Daum, J. E., Harst, U., and Reuter, M. (2010). Running gauge coupling in asymptotically safe quantum gravity. *JHEP* 1001:084. doi: 10.1007/JHEP01(2010)084
- Daum, J. E., and Reuter, M. (2013). Einstein-cartan gravity, asymptotic safety, and the running immerzi parameter. *Annals Phys.* 334:351. doi: 10.1016/j.aop.2013.04.002
- David, F. (1988). Conformal field theories coupled to 2D gravity in the conformal gauge. *Mod. Phys. Lett. A* 3:1651. doi: 10.1142/S0217732388001975
- de Alwis, S. P. (2018). Exact RG flow equations and quantum gravity. *JHEP* 1803:118. doi: 10.1007/JHEP03(2018)118
- De Alwis, S. P. (2018). Higher derivative corrections to lower order RG flow equations. *arXiv:1809.04671 [hep-th]*.
- de Brito, G. P., Ohta, N., Pereira, A. D., Tomaz, A. A., and Yamada, M. (2018). Asymptotic safety and field parametrization dependence in the  $f(R)$  truncation. *arXiv:1805.09656 [hep-th]*.
- de Calan, C., Faria da Veiga, P. A., Magnen, J., and Seneor, R. (1991). Constructing the three-dimensional gross-neveu model with a large number of flavor components. *Phys. Rev. Lett.* 66:3233. doi: 10.1103/PhysRevLett.66.3233
- Delamotte, B. (2012). An introduction to the nonperturbative renormalization group. *Lect. Notes Phys.* 852:49. doi: 10.1007/978-3-642-27320-9\_2
- Demmel, M., Saueressig, F., and Zanusso, O. (2012). Fixed-functionals of three-dimensional quantum Einstein Gravity. *JHEP* 1211:131. doi: 10.1007/JHEP11(2012)131
- Demmel, M., Saueressig, F., and Zanusso, O. (2015). A proper fixed functional for four-dimensional Quantum Einstein Gravity. *JHEP* 1508:113. doi: 10.1007/JHEP08(2015)113
- Denz, T., Pawłowski, J. M., and Reichert, M. (2018). Towards apparent convergence in asymptotically safe quantum gravity. *Eur. Phys. J. C* 78:336. doi: 10.1140/epjc/s10052-018-5806-0
- Deser, S., and Nieuwenhuizen, P. V. (1974a). Nonrenormalizability of quantized fermion gravitation interactions. *Lett. Nuovo Cim.* 2:218.
- Deser, S., and Nieuwenhuizen, P. V. (1974b). Nonrenormalizability of the quantized einstein-maxwell system. *Phys. Rev. Lett.* 32:245.
- Di Francesco, P., Ginsparg, P. H., and Zinn-Justin, J. (1995). 2-D Gravity and random matrices. *Phys. Rept.* 254:1. doi: 10.1016/0370-1573(94)00084-G
- Dietz, J. A., and Morris, T. R. (2013a). Asymptotic safety in the  $f(R)$  approximation. *JHEP* 1301:108. doi: 10.1007/JHEP01(2013)108
- Dietz, J. A., and Morris, T. R. (2013b). Redundant operators in the exact renormalisation group and in the  $f(R)$  approximation to asymptotic safety. *JHEP* 1307:064. doi: 10.1007/JHEP07(2013)064
- Distler, J., and Kawai, H. (1989). Conformal field theory and 2D quantum gravity. *Nucl. Phys. B* 321:509. doi: 10.1016/0550-3213(89)90354-4
- Dittrich, B. (2017). "The continuum limit of loop quantum gravity-a framework for solving the theory," in *100 Years of General Relativity Loop Quantum Gravity*, eds A. Ashtekar and J. Pullin (Singapore: World Scientific Publishing), 153–179. doi: 10.1142/9789813220003\_0006
- Dittrich, B., Mizera S., and Steinhaus, S. (2016a). Decorated tensor network renormalization for lattice gauge theories and spin foam models. *New J. Phys.* 18:053009. doi: 10.1088/1367-2630/18/5/053009
- Dittrich, B., Schnetter, E., Seth, C. J., and Steinhaus, S. (2016b). Coarse graining flow of spin foam intertwiners. *Phys. Rev. D.* 94: 124050. doi: 10.1103/PhysRevD.94.124050
- D'Odorico, G., Saueressig, F., and Schutten, M. (2014). Asymptotic freedom in Hořava-lifshitz gravity. *Phys. Rev. Lett.* 113:171101. doi: 10.1103/PhysRevLett.113.171101
- Donà, P., Eichhorn, A., Labus, P., and Percacci, R. (2016). Asymptotic safety in an interacting system of gravity and scalar matter. *Phys. Rev. D.* 93:044049. doi: 10.1103/PhysRevD.93.044049
- Donà, P., Eichhorn, A., and Percacci, R. (2014). Matter matters in asymptotically safe quantum gravity. *Phys. Rev. D* 89:084035. doi: 10.1103/PhysRevD.89.084035
- Dona, P., and Percacci, R. (2013). Functional renormalization with fermions and tetrads. *Phys. Rev. D.* 87:045002. doi: 10.1103/PhysRevD.87.045002
- Dondi, N. A., Prochazka, V., and Sannino, V. (2018). Conformal Data of fundamental Gauge-Yukawa Theories. *Phys. Rev. D* 98:045002. doi: 10.1103/PhysRevD.98.045002
- Donoghue, J. F. (1994a). Leading quantum correction to the Newtonian potential. *Phys. Rev. Lett.* 72:2996. doi: 10.1103/PhysRevLett.72.2996
- Donoghue, J. F. (1994b). General relativity as an effective field theory: the leading quantum corrections. *Phys. Rev. D* 50:3874. doi: 10.1103/PhysRevD.50.3874
- Donoghue, J. F. (2012). The effective field theory treatment of quantum gravity. *AIP Conf. Proc.* 1483:73. doi: 10.1063/1.4756964
- Dou, D., and Percacci, R. (1998). The running gravitational couplings. *Class. Q. Grav.* 15:3449. doi: 10.1088/0264-9381/15/11/011
- Dowker, F. (2013). Introduction to causal sets and their phenomenology. *Gen. Rel. Grav.* 45:1651. doi: 10.1007/s10714-013-1569-y
- Ebert, D., Plefka, J., and Rodigast, A. (2008). Absence of gravitational contributions to the running Yang-Mills coupling. *Phys. Lett. B.* 660:579. doi: 10.1016/j.physletb.2008.01.037
- Ebert, D., Plefka, J., and Rodigast, A. (2009). Gravitational contributions to the running yang-mills coupling in large extra-dimensional brane worlds. *JHEP* 0902:028. doi: 10.1088/1126-6708/2009/02/028
- Eichhorn, A. (2012). Quantum-gravity-induced matter self-interactions in the asymptotic-safety scenario. *Phys. Rev. D.* 86:105021. doi: 10.1103/PhysRevD.86.105021
- Eichhorn, A. (2013). Faddeev-popov ghosts in quantum gravity beyond perturbation theory. *Phys. Rev. D* 87:124016. doi: 10.1103/PhysRevD.87.124016
- Eichhorn, A. (2015). The renormalization group flow of unimodular  $f(R)$  gravity. *JHEP* 1504:096. doi: 10.1007/JHEP04(2015)096
- Eichhorn, A. (2018a). Status of the asymptotic safety paradigm for quantum gravity and matter. *Found. Phys.* 48:1407. doi: 10.1007/s10701-018-0196-6
- Eichhorn, A. (2018b). Towards coarse graining of discrete Lorentzian quantum gravity. *Class. Quant. Grav.* 35:044001. doi: 10.1088/1361-6382/aaa0a3
- Eichhorn, A., and Gies, H. (2010). Ghost anomalous dimension in asymptotically safe quantum gravity. *Phys. Rev. D* 81:104010. doi: 10.1103/PhysRevD.81.104010
- Eichhorn, A., and Gies, H. (2011). Light fermions in quantum gravity. *New J. Phys.* 13:125012. doi: 10.1088/1367-2630/13/12/125012
- Eichhorn, A., Gies, H., Jaeckel, J., Plehn, T., Scherer, M. M., and Sondenheimer, R. (2015). The higgs mass and the scale of new physics. *JHEP* 1504:022. doi: 10.1007/JHEP04(2015)022
- Eichhorn, A., Hamada, Y., Lumma, J., and Yamada, M. (2018). Quantum gravity fluctuations flatten the Planck-scale higgs potential. *Phys. Rev. D.* 97:086004. doi: 10.1103/PhysRevD.97.086004
- Eichhorn, A., and Held, A. (2017). Viability of quantum-gravity induced ultraviolet completions for matter. *Phys. Rev. D.* 96:086025. doi: 10.1103/PhysRevD.96.086025
- Eichhorn, A., and Held, A. (2018). Mass difference for charged quarks from quantum gravity. *Phys. Rev. Lett.* 121:151302. doi: 10.1103/PhysRevLett.121.151302
- Eichhorn, A., and Held, A. (2018). Top mass from asymptotic safety. *Phys. Lett. B* 777:217. doi: 10.1016/j.physletb.2017.12.040

- Eichhorn, A., Held, A., and Griend, P. V. (2018a). Asymptotic safety in the dark. *arXiv:1802.08589 [hep-ph]*.
- Eichhorn, A., Held, A., and Pawłowski, J. M. (2016). Quantum-gravity effects on a Higgs-Yukawa model. *Phys. Rev. D* 94:104027. doi: 10.1103/PhysRevD.94.104027
- Eichhorn, A., Held, A., and Wetterich, C. (2018d). Quantum-gravity predictions for the fine-structure constant. *Phys. Lett. B* 782:198. doi: 10.1016/j.physletb.2018.05.016
- Eichhorn, A., Janssen, L., and Scherer, M. M. (2016). Critical  $O(N)$  models above four dimensions: small- $N$  solutions and stability. *Phys. Rev. D* 93:125021. doi: 10.1103/PhysRevD.93.125021
- Eichhorn, A., and Kosłowski, T. (2013). Continuum limit in matrix models for quantum gravity from the functional renormalization Group. *Phys. Rev. D* 88:084016. doi: 10.1103/PhysRevD.88.084016
- Eichhorn, A., and Kosłowski, T. (2018). Flowing to the continuum in discrete tensor models for quantum gravity. *Ann. Inst. H. Poincaré Comb. Phys. Interact.* 5:173. doi: 10.4171/AIHPD/52
- Eichhorn, A., Labus, P., Pawłowski, J. M., and Reichert, M. (2018b). Effective universality in quantum gravity. *arXiv:1804.00012 [hep-th]*.
- Eichhorn, A., and Lippoldt, S. (2017). Quantum gravity and standard-model-like fermions. *Phys. Lett. B* 767:142. doi: 10.1016/j.physletb.2017.01.064
- Eichhorn, A., Lippoldt, S., Pawłowski, J. M., Reichert, M., and Schiffer, M. (2018c). How perturbative is quantum gravity? *arXiv:1810.02828 [hep-th]*.
- Eichhorn, A., Lippoldt, S., and Skrinjar, V. (2018e). Nonminimal hints for asymptotic safety. *Phys. Rev. D* 97:026002. doi: 10.1103/PhysRevD.97.026002
- Eichhorn, A., and Versteegen, F. (2018). Upper bound on the abelian gauge coupling from asymptotic safety. *JHEP* 1801:030. doi: 10.1007/JHEP01(2018)030
- Elias-Miro, J., Espinosa, J. R., Giudice, G. F., Isidori, G., Riotto, A., and Strumia, A. (2012). Higgs mass implications on the stability of the electroweak vacuum. *Phys. Lett. B* 709:222. doi: 10.1016/j.physletb.2012.02.013
- Elizalde, E., Odintsov, S. D., and Romeo, A. (1996). Renormalization group properties of higher derivative quantum gravity with matter in (4-epsilon)-dimensions. *Nucl. Phys. B* 462:315. doi: 10.1016/0550-3213(95)00674-5
- Ellis, G. F. R., van Elst, H., Murugan, J., and Uzan, J. P. (2011). On the trace-free Einstein equations as a viable alternative to general relativity. *Class. Quant. Grav.* 28:225007. doi: 10.1088/0264-9381/28/22/225007
- Ellis, J., and Mavromatos, N. E. (2012). On the interpretation of gravitational corrections to gauge couplings. *Phys. Lett. B* 711:139. doi: 10.1016/j.physletb.2012.04.005
- Ellwanger, U. (1994). Flow equations for  $N$  point functions and bound states. *Z. Phys. C* 62:503. doi: 10.1007/BF01555911
- El-Showk, S., Paulos, M. F., Poland, D., Rychkov, S., Simmons-Duffin, D., and Vichi, A. (2012). Solving the 3D Ising model with the conformal bootstrap. *Phys. Rev. D* 86:025022. doi: 10.1103/PhysRevD.86.025022
- El-Showk, S., Paulos, M. F., Poland, D., Rychkov, S., Simmons-Duffin, D., and Vichi, A. (2014). Solving the 3d Ising model with the conformal bootstrap II. c-Minimization and precise critical exponents. *J. Stat. Phys.* 157:869. doi: 10.1007/s10955-014-1042-7
- Emir Gümrükçüoğlu, A., Saravani, M., and Sotiriou, T. P. (2018). Hořava gravity after GW170817. *Phys. Rev. D* 97:024032. doi: 10.1103/PhysRevD.97.024032
- Esbensen, J. K., Rytov, T. A., and Sannino, F. (2016). Quantum critical behavior of semisimple gauge theories. *Phys. Rev. D* 93:045009. doi: 10.1103/PhysRevD.93.045009
- Fabbrichesi, M., Percacci, R., Toner, A., and Zanusso, O. (2011). Asymptotic safety and the gauged  $SU(N)$  nonlinear  $\sigma$ -model. *Phys. Rev. D* 83:025016. doi: 10.1103/PhysRevD.83.025016
- Falkenberg, S., and Odintsov, S. D. (1998). Gauge dependence of the effective average action in Einstein gravity. *Int. J. Mod. Phys. A* 13:607. doi: 10.1142/S0217751X98000263
- Falls, K. (2015). Renormalization of Newton's constant. *Phys. Rev. D* 92:124057. doi: 10.1103/PhysRevD.92.124057
- Falls, K. (2017). Physical renormalization schemes and asymptotic safety in quantum gravity. *Phys. Rev. D* 96:126016. doi: 10.1103/PhysRevD.96.126016
- Falls, K., and Litim, D. F. (2014). Black hole thermodynamics under the microscope. *Phys. Rev. D* 89:084002. doi: 10.1103/PhysRevD.89.084002
- Falls, K., Litim, D. F., Nikolakopoulos, K., and Rahmede, C. (2013). A bootstrap towards asymptotic safety. *arXiv:1301.4191 [hep-th]*.
- Falls, K., Litim, D. F., Nikolakopoulos, K., and Rahmede, C. (2016). Further evidence for asymptotic safety of quantum gravity. *Phys. Rev. D* 93:104022. doi: 10.1103/PhysRevD.93.104022
- Falls, K., Litim, D. F., and Raghuraman, A. (2012). Black holes and asymptotically safe gravity. *Int. J. Mod. Phys. A* 27:1250019. doi: 10.1142/S0217751X12500194
- Fei, L., Giombi, S., and Klebanov, I. R. (2014). Critical  $O(N)$  models in  $6 - \epsilon$  dimensions. *Phys. Rev. D* 90:025018. doi: 10.1103/PhysRevD.90.025018
- Felipe, J. C. C., Brito, L. C. T., Sampaio, M., and Nemes, M. C. (2011). Quantum gravitational contributions to the beta function of quantum electrodynamics. *Phys. Lett. B* 700:86. doi: 10.1016/j.physletb.2011.04.061
- Finkelstein, D. R., Gaiutdinov, A. A., and Baugh, J. E. (2001). Unimodular relativity and cosmological constant. *J. Math. Phys.* 42:340. doi: 10.1063/1.1328077
- Fischler, M., and Oliensis, J. (1982). Two loop corrections to the evolution of the Higgs-Yukawa coupling constant. *Phys. Lett.* 119B:385. doi: 10.1016/0370-2693(82)90695-5
- Floerchinger, S. (2012). Analytic continuation of functional renormalization group equations. *JHEP* 1205:021. doi: 10.1007/JHEP05(2012)021
- Floresanini, R., and Percacci, R. (1995). The renormalization group flow of the Dilaton potential. *Phys. Rev. D* 52:896. doi: 10.1103/PhysRevD.52.896
- Folkerts, S., Litim, D. F., and Pawłowski, J. M. (2012). Asymptotic freedom of Yang-Mills theory with gravity. *Phys. Lett. B* 709:234. doi: 10.1016/j.physletb.2012.02.002
- Ford, C., Jack, I., and Jones, D. R. T. (1992). The standard model effective potential at two loops. *Nucl. Phys. B* 387:373. doi: 10.1016/0550-3213(92)90165-8
- Fradkin, E. S., and Tseytlin, A. A. (1982). Renormalizable asymptotically free quantum theory of gravity. *Nucl. Phys. B* 201:469. doi: 10.1016/0550-3213(82)90444-8
- Friedan, D. (1980). Nonlinear models in two epsilon dimensions. *Phys. Rev. Lett.* 45:1057. doi: 10.1103/PhysRevLett.45.1057
- Gastmans, R., Kallosh, R., and Truffin, C. (1978). Quantum gravity near two-dimensions. *Nucl. Phys. B* 133:417. doi: 10.1016/0550-3213(78)90234-1
- Gawedzki, K., and Kupiainen, A. (1985). Renormalizing the nonrenormalizable. *Phys. Rev. Lett.* 55:363. doi: 10.1103/PhysRevLett.55.363
- Gell-Mann, M., and Low, F. E. (1954). Quantum electrodynamics at small distances. *Phys. Rev.* 95:1300. doi: 10.1103/PhysRev.95.1300
- Gies, H. (2003). Renormalizability of gauge theories in extra dimensions. *Phys. Rev. D* 68:085015. doi: 10.1103/PhysRevD.68.085015
- Gies, H. (2012). Introduction to the functional RG and applications to gauge theories. *Lect. Notes Phys.* 852:287. doi: 10.1007/978-3-642-27320-9\_6
- Gies, H., Gneiting, C., and Sondenheimer, R. (2014). Higgs mass bounds from renormalization flow for a simple Yukawa model. *Phys. Rev. D* 89:045012. doi: 10.1103/PhysRevD.89.045012
- Gies, H., Knorr, B., and Lippoldt, S. (2015). Generalized parametrization dependence in quantum gravity. *Phys. Rev. D* 92:084020. doi: 10.1103/PhysRevD.92.084020
- Gies, H., Knorr, B., Lippoldt, S., and Saueressig, F. (2016). Gravitational two-loop counterterm is asymptotically safe. *Phys. Rev. Lett.* 116:211302. doi: 10.1103/PhysRevLett.116.211302
- Gies, H., and Martini, R. (2018). Curvature bound from gravitational catalysis. *Phys. Rev. D* 97:085017. doi: 10.1103/PhysRevD.97.085017
- Gies, H., Rechenberger, S., and Scherer, M. M. (2010). Towards an asymptotic-safety scenario for chiral Yukawa systems. *Eur. Phys. J. C* 66:403. doi: 10.1140/epjc/s10052-010-1257-y
- Gies, H., and Scherer, M. M. (2010). Asymptotic safety of simple Yukawa systems. *Eur. Phys. J. C* 66:387. doi: 10.1140/epjc/s10052-010-1256-z
- Gies, H., and Zambelli, L. (2017). Non-abelian Higgs models: paving the way for asymptotic freedom. *Phys. Rev. D* 96:025003. doi: 10.1103/PhysRevD.96.025003
- Glaser, L. (2018). The Ising model coupled to 2d orders. *Class. Quant. Grav.* 35:084001. doi: 10.1088/1361-6382/aab139
- Gliozzi, F., and Rago, A. (2014). Critical exponents of the 3d Ising and related models from Conformal Bootstrap. *JHEP* 1410:042. doi: 10.1007/JHEP10(2014)042

- Godfrey, N., and Gross, M. (1991). Simplicial quantum gravity in more than two-dimensions. *Phys. Rev. D* 43:1749. doi: 10.1103/PhysRevD.43.R1749
- Gonzalez-Martin, S., and Martin, C. P. (2017). Do the gravitational corrections to the beta functions of the quartic and Yukawa couplings have an intrinsic physical meaning? *Phys. Lett. B* 773:585. doi: 10.1016/j.physletb.2017.09.011
- Goroff, M. H., and Sagnotti, A. (1986). The ultraviolet behavior Of einstein gravity. *Nucl. Phys. B* 266:709.
- Gracey, J. A. (1996). The QCD beta function at  $O[1/N(f)]$ . *Phys. Lett. B* 373:178. doi: 10.1016/0370-2693(96)00105-0
- Groh, K., and Saueressig, F. (2010). Ghost wave-function renormalization in asymptotically safe quantum gravity. *J. Phys. A* 43:365403. doi: 10.1088/1751-8113/43/36/365403
- Gross, M. (1992). Tensor models and simplicial quantum gravity in  $>2$ -D. *Nucl. Phys. Proc. Suppl.* 25A:144. doi: 10.1016/S0920-5632(05)80015-5
- Guida, R., and Zinn-Justin, J. (1998). Critical exponents of the N vector model. *J. Phys. A* 31:8103. doi: 10.1088/0305-4470/31
- Gurau, R. (2011). The  $1/N$  expansion of colored tensor models. *Annales Henri Poincaré* 12:829. doi: 10.1007/s00023-011-0101-8
- Gurau, R. (2016). Invitation to random tensors. *SIGMA* 12:094. doi: 10.3842/SIGMA.2016.094
- Hamada, Y., and Yamada, M. (2017). Asymptotic safety of higher derivative quantum gravity non-minimally coupled with a matter system. *JHEP* 1708:070. doi: 10.1007/JHEP08(2017)070
- Hamber, H. W. (2009). Quantum gravity on the lattice. *Gen. Relat. Grav.* 41:817. doi: 10.1007/s10714-009-0769-y
- Hamber, H. W. (2015). Scaling exponents for lattice quantum gravity in four dimensions. *Phys. Rev. D* 92:064017. doi: 10.1103/PhysRevD.92.064017
- Hambye, T., and Riesselmann, K. (1997). Matching conditions and Higgs mass upper bounds revisited. *Phys. Rev. D* 55:7255. doi: 10.1103/PhysRevD.55.7255 [hep-ph/9610272].
- Hands, S., Kocic, A., and Kogut, J. B. (1993). Four fermi theories in fewer than four-dimensions. *Annals Phys.* 224:29. doi: 10.1006/aphy.1993.1039
- Harst, U., and Reuter, M. (2011). QED coupled to QEG. *JHEP* 1105, 119. doi: 10.1007/JHEP05(2011)119
- Harst, U., and Reuter, M. (2015). A new functional flow equation for Einstein–cartan quantum gravity. *Annals Phys.* 354:637. doi: 10.1016/j.aop.2015.01.006
- Harst, U., and Reuter, M. (2016). On selfdual spin-connections and asymptotic safety. *Phys. Lett. B* 753:395. doi: 10.1016/j.physletb.2015.12.016
- He, H. J., Kuang, Y. P., Wang, Q., and Yi, Y. P. (1992). Effective potential, renormalization, and nontrivial ultraviolet fixed point in D-dimensional four fermion theories ( $2 < D < 4$ ) to order  $1/N$  in  $1/N$  expansion. *Phys. Rev. D* 45:4610. doi: 10.1103/PhysRevD.45.4610
- Henson, J. (2006). “The causal set approach to quantum gravity,” in *Approaches to Quantum Gravity*, ed D. Oriti (Cambridge), 393–413.
- Higashijima, K., and Ito, E. (2003). Three-dimensional nonlinear sigma models in the Wilsonian renormalization method. *Prog. Theor. Phys.* 110:563. doi: 10.1143/PTP.110.563
- Holdom, B. (2011). Large N flavor beta-functions: a recap. *Phys. Lett. B* 694:74. doi: 10.1016/j.physletb.2010.09.037
- Holdom, B., and Ren, J. (2016a). QCD analogy for quantum gravity. *Phys. Rev. D* 93:124030. doi: 10.1103/PhysRevD.93.124030
- Holdom, B., and Ren, J. (2016b). Quadratic gravity: from weak to strong. *Int. J. Mod. Phys. D* 25:1643004. doi: 10.1142/S0218271816430045
- Holthausen, M., Kubo, J., Lim, K. S., and Lindner, M. (2013). Electroweak and conformal symmetry breaking by a strongly coupled hidden sector. *JHEP* 1312:076. doi: 10.1007/JHEP12(2013)076
- Horava, P. (2009). Quantum gravity at a lifshitz point. *Phys. Rev. D* 79:084008. doi: 10.1103/PhysRevD.79.084008
- Intriligator, K., and Sannino, F. (2015). Supersymmetric asymptotic safety is not guaranteed. *JHEP* 1511:023. doi: 10.1007/JHEP11(2015)023
- Jack, I., and Jones, D. R. T. (1991). The epsilon expansion of two-dimensional quantum gravity. *Nucl. Phys. B* 358:695. doi: 10.1016/0550-3213(91)90430-6
- Jack, I., and Osborn, H. (1984). Background field calculations in curved space-time. 1. General formalism and application to scalar fields. *Nucl. Phys. B* 234:331. doi: 10.1016/0550-3213(84)90067-1
- Jack, I., and Osborn, H. (1990). Analogs for the  $c$  Theorem for Four-dimensional renormalizable field theories. *Nucl. Phys. B* 343:647. doi: 10.1016/0550-3213(90)90584-Z
- Jack, I., and Osborn, H. (2014). Constraints on RG flow for four dimensional quantum field theories. *Nucl. Phys. B* 883:425. doi: 10.1016/j.nuclphysb.2014.03.018
- Jones, D. R. T. (1982). The two loop beta function for a  $G(1) \times G(2)$  Gauge Theory. *Phys. Rev. D* 25:581. doi: 10.1103/PhysRevD.25.581
- Jordan, S., and Loll, R. (2013). Causal dynamical triangulations without preferred foliation. *Phys. Lett. B* 724:155. doi: 10.1016/j.physletb.2013.06.007
- Kabat, D. N. (1995). Black hole entropy and entropy of entanglement. *Nucl. Phys. B* 453:281. doi: 10.1016/0550-3213(95)00443-V
- Kallosh, R., Linde, A. D., Linde, D. A., and Susskind, L. (1995). Gravity and global symmetries. *Phys. Rev. D* 52:912. doi: 10.1103/PhysRevD.52.912
- Kawai, H., Kitazawa, Y., and Ninomiya, M. (1993a). Scaling exponents in quantum gravity near two-dimensions. *Nucl. Phys. B* 393:280. doi: 10.1016/0550-3213(93)90246-L
- Kawai, H., Kitazawa, Y., and Ninomiya, M. (1993b). Ultraviolet stable fixed point and scaling relations in  $(2+\epsilon)$ -dimensional quantum gravity. *Nucl. Phys. B* 404:684. doi: 10.1016/0550-3213(93)90594-F
- Kawai, H., Kitazawa, Y., and Ninomiya, M. (1996). Renormalizability of quantum gravity near two-dimensions. *Nucl. Phys. B* 467:313. doi: 10.1016/0550-3213(96)00119-8
- Kawai, H., and Ninomiya, M. (1990). Renormalization group and quantum gravity. *Nucl. Phys. B* 336:115. doi: 10.1016/0550-3213(90)90345-E
- Khoze, V. V. (2013). Inflation and dark matter in the higgs portal of classically scale invariant standard model. *JHEP* 1311:215. doi: 10.1007/JHEP11(2013)215
- Kikukawa, Y., and Yamawaki, K. (1990). Ultraviolet fixed point structure of renormalizable four fermion theory in less than four-dimensions. *Phys. Lett. B* 234:497. doi: 10.1016/0370-2693(90)92046-L
- Knechtli, F., and Rinaldi, E. (2016). Extra-dimensional models on the lattice. *Int. J. Mod. Phys. A* 31:1643002. doi: 10.1142/S0217751X16430028
- Knorr, B., and Lippoldt, S. (2017). Correlation functions on a curved background. *Phys. Rev. D* 96:065020. doi: 10.1103/PhysRevD.96.065020
- Koch, B., and Saueressig, F. (2014a). Structural aspects of asymptotically safe black holes. *Class. Quant. Grav.* 31:015006. doi: 10.1088/0264-9381/31/1/015006
- Koch, B., and Saueressig, F. (2014b). Black holes within asymptotic safety. *Int. J. Mod. Phys. A* 29:1430011. doi: 10.1142/S0217751X14300117
- Kofinas, G., and Zarikas, V. (2015). Avoidance of singularities in asymptotically safe quantum Einstein gravity. *JCAP* 1510:069. doi: 10.1088/1475-7516/2015/10/069
- Kofinas, G., and Zarikas, V. (2016). Asymptotically safe gravity and non-singular inflationary big bang with vacuum birth. *Phys. Rev. D* 94:103514. doi: 10.1103/PhysRevD.94.103514
- Kopietz, P., Bartosch, L., and Schütz, F. (2010). Introduction to the functional renormalization group. *Lect. Notes Phys.* 798:1. doi: 10.1007/978-3-642-05094-7
- Kowalska, K., Bond, A., Hiller, G., and Litim, D. (2017). Towards an asymptotically safe completion of the standard model. *PoS EPS HEP2017:542*. doi: 10.22323/1.314.0542
- Kowalska, K., and Sessolo, E. M. (2018). Gauge contribution to the  $1/N_f$  expansion of the Yukawa coupling beta function. *JHEP* 1804, 027. doi: 10.1007/JHEP04(2018)027
- Krasnov, K., and Percacci, R. (2018). Gravity and unification: a review. *Class. Quant. Grav.* 35:143001. doi: 10.1088/1361-6382/aac58d
- Labus, P., Morris, T. R., and Slade, Z. H. (2016). Background independence in a background dependent renormalization group. *Phys. Rev. D* 94:024007. doi: 10.1103/PhysRevD.94.024007
- Labus, P., Percacci, R., and Vacca, G. P. (2016). Asymptotic safety in  $O(N)$  scalar models coupled to gravity. *Phys. Lett. B* 753:274. doi: 10.1016/j.physletb.2015.12.022
- Laiho, J., Bassler, S., Coumbe, D., Du, D., and Neelakanta, J. T. (2017). Lattice quantum gravity and asymptotic safety. *Phys. Rev. D* 96:064015. doi: 10.1103/PhysRevD.96.064015
- Laiho, J., and Coumbe, D. (2011). Evidence for asymptotic safety from lattice quantum gravity. *Phys. Rev. Lett.* 107:161301. doi: 10.1103/PhysRevLett.107.161301

- Larsen, F., and Wilczek, F. (1996). Renormalization of black hole entropy and of the gravitational coupling constant. *Nucl. Phys. B* 458:249. doi: 10.1016/0550-3213(95)00548-X
- Lauscher, O., and Reuter, M. (2002). Ultraviolet fixed point and generalized flow equation of quantum gravity. *Phys. Rev. D* 65:025013. doi: 10.1103/PhysRevD.65.025013
- Lauscher, O., and Reuter, M. (2002). Flow equation of quantum Einstein gravity in a higher derivative truncation. *Phys. Rev. D* 66:025026. doi: 10.1103/PhysRevD.66.025026
- Lauscher, O., and Reuter, M. (2005). Fractal spacetime structure in asymptotically safe gravity. *JHEP* 0510:050. doi: 10.1088/1126-6708/2005/10/050
- Lewandowski, A., Meissner, K. A., and Nicolai, H. (2018). Conformal standard model, leptogenesis and dark matter. *Phys. Rev. D* 97:035024. doi: 10.1103/PhysRevD.97.035024
- Lindner, M., Schmidt, S., and Smirnov, J. (2014). Neutrino masses and conformal electro-weak symmetry breaking. *JHEP* 1410:177. doi: 10.1007/JHEP10(2014)177
- Litim, D. F. (2000). Optimization of the exact renormalization group. *Phys. Lett. B* 486:92. doi: 10.1016/S0370-2693(00)00748-6
- Litim, D. F. (2001). Optimized renormalization group flows. *Phys. Rev. D* 64:105007. doi: 10.1103/PhysRevD.64.105007
- Litim, D. F. (2004). Fixed points of quantum gravity. *Phys. Rev. Lett.* 92:201301. doi: 10.1103/PhysRevLett.92.201301
- Litim, D. F., Mojaza, M., and Sannino, F. (2016). Vacuum stability of asymptotically safe gauge-Yukawa theories. *JHEP* 1601:081. doi: 10.1007/JHEP01(2016)081
- Litim, D. F., and Pawłowski, J. M. (2002). Wilsonian flows and background fields. *Phys. Lett. B* 546:279. doi: 10.1016/S0370-2693(02)02693-X
- Litim, D. F., and Sannino, F. (2014). Asymptotic safety guaranteed. *JHEP* 1412:178. doi: 10.1007/JHEP12(2014)178
- Litim, D. F., and Zappala, D. (2011). Ising exponents from the functional renormalisation group. *Phys. Rev. D* 83:085009. doi: 10.1103/PhysRevD.83.085009
- Machacek, M. E., and Vaughn, M. T. (1983). Two loop renormalization group equations in a general quantum field theory. 1. Wave function renormalization. *Nucl. Phys. B* 222:83. doi: 10.1016/0550-3213(83)90610-7
- Machacek, M. E., and Vaughn, M. T. (1984). Two loop renormalization group equations in a general quantum field theory. 2. Yukawa Couplings. *Nucl. Phys. B* 236:221. doi: 10.1016/0550-3213(84)90533-9
- Machacek, M. E., and Vaughn, M. T. (1985). Two loop renormalization group equations in a general quantum field Theory. 3. Scalar quartic couplings. *Nucl. Phys. B* 249:70. doi: 10.1016/0550-3213(85)90040-9
- Machado, P. F., and Saueressig, F. (2008). On the renormalization group flow of  $f(R)$ -gravity. *Phys. Rev. D* 77:124045. doi: 10.1103/PhysRevD.77.124045
- Mackay, P. T., and Toms, D. J. (2010). Quantum gravity and scalar fields. *Phys. Lett. B* 684:251. doi: 10.1016/j.physletb.2009.12.032
- Mann, R., Meffe, J., Sannino, F., Steele, T., Wang, Z. W., and Zhang, C. (2017). Asymptotically safe standard model via vectorlike fermions. *Phys. Rev. Lett.* 119:261802. doi: 10.1103/PhysRevLett.119.261802
- Manrique, E., Rechenberger, S., and Saueressig, F. (2011a). Asymptotically safe lorentzian gravity. *Phys. Rev. Lett.* 106:251302. doi: 10.1103/PhysRevLett.106.251302
- Manrique, E., and Reuter, M. (2010). Bimetric truncations for quantum Einstein gravity and asymptotic safety. *Annals Phys.* 325:785. doi: 10.1016/j.aop.2009.11.009
- Manrique, E., Reuter, M., and Saueressig, F. (2011b). Matter induced bimetric actions for gravity. *Annals Phys.* 326:440. doi: 10.1016/j.aop.2010.11.003
- Manrique, E., Reuter, M., and Saueressig, F. (2011c). Bimetric renormalization group flows in quantum Einstein gravity. *Annals Phys.* 326: 463. doi:10.1016/j.aop.2010.11.006
- Markkanen, T., Rajantie, A., and Stopyra, S. (2018). Cosmological aspects of higgs vacuum metastability *arXiv:1809.06923 [astro-ph.CO]*.
- Meibohm, J., and Pawłowski, J. M. (2016). Chiral fermions in asymptotically safe quantum gravity. *Eur. Phys. J. C* 76, no. 5, 285. doi: 10.1140/epjc/s10052-016-4132-7
- Meibohm, J., Pawłowski, J. M., and Reichert, M. (2016). Asymptotic safety of gravity-matter systems. *Phys. Rev. D* 93:084035. doi: 10.1103/PhysRevD.93.084035
- Meissner, K. A., and Nicolai, H. (2007). Conformal symmetry and the standard model. *Phys. Lett. B* 648:312. doi: 10.1016/j.physletb.2007.03.023
- Metzner, W., Salmhofer, M., Honerkamp, C., Meden, V., and Schönhammer, K. (2012). Functional renormalization group approach to correlated fermion systems. *Rev. Mod. Phys.* 84:299. doi: 10.1103/RevModPhys.84.299
- Mølgaard, E., and Sannino, F. (2017). Asymptotically safe and free chiral theories with and without scalars. *Phys. Rev. D* 96:056004. doi: 10.1103/PhysRevD.96.056004
- Molinaro, E., Sannino, F., and Wang, Z. W. (2018). Safe pati-salam. *arXiv:1807.03669 [hep-ph]*.
- Morris, T. R. (1994). The exact renormalization group and approximate solutions. *Int. J. Mod. Phys. A* 9:2411.
- Morris, T. R. (2005). Renormalizable extra-dimensional models. *JHEP* 0501:2. doi: 10.1088/1126-6708/2005/01/002
- Morris, T. R. (2016). Large curvature and background scale independence in single-metric approximations to asymptotic safety. *JHEP* 1611:160. doi: 10.1007/JHEP11(2016)160
- Narain, G., and Percacci, R. (2010). Renormalization group flow in scalar-tensor theories. I. *Class. Quant. Grav.* 27:075001. doi: 10.1088/0264-9381/27/7/075001
- Narain, R., and Anishetty, R. (2013). Running couplings in quantum theory of gravity coupled with gauge fields. *JHEP* 1310:203. doi: 10.1007/JHEP10(2013)203
- Niedermaier, M. (2010). Gravitational fixed points and asymptotic safety from perturbation theory. *Nucl. Phys. B* 833:226. doi: 10.1016/j.nuclphysb.2010.01.016
- Niedermaier, M. R. (2009). Gravitational fixed points from perturbation theory. *Phys. Rev. Lett.* 103:101303. doi: 10.1103/PhysRevLett.103.101303
- Nielsen, N. G., Sannino, F., and Svendsen, O. (2015). Inflation from asymptotically safe theories. *Phys. Rev. D* 91:103521. doi: 10.1103/PhysRevD.91.103521
- Nieto, C. M., Percacci, R., and Skrinjar, V. (2017). Split weyl transformations in quantum gravity. *Phys. Rev. D* 96:106019. doi: 10.1103/PhysRevD.96.106019 [arXiv:1708.09760 (gr-qc)].
- Nink, A., and Reuter, M. (2013). On the physical mechanism underlying Asymptotic Safety. *JHEP* 1301:062. doi: 10.1007/JHEP01(2013)062
- Nink, A., and Reuter, M. (2016). The unitary conformal field theory behind 2D Asymptotic Safety. *JHEP* 1602:167. doi: 10.1007/JHEP02(2016)167
- Nink, A., Reuter, M., and Saueressig, F. (2013). Asymptotic safety in quantum gravity. *Scholarpedia* 8:31015. doi: 10.4249/scholarpedia.31015
- Nishimura, J., Tamura, S., and Tsuchiya, A. (1994).  $R^{**2}$  gravity in (2+epsilon)-dimensional quantum gravity. *Mod. Phys. Lett. A* 9:3565. doi: 10.1142/S0217732394003403
- Oda, K. y., and Yamada, M. (2016). Non-minimal coupling in higgs-yukawa model with asymptotically safe gravity. *Class. Quant. Grav.* 33:125011. doi: 10.1088/0264-9381/33/12/125011
- Ohta, N. (2017). Background scale independence in quantum gravity. *PTEP* 2017:033E02. doi: 10.1093/ptep/ptx020
- Ohta, N., Percacci, R., and Pereira, A. D. (2016). Gauges and functional measures in quantum gravity I: Einstein theory. *JHEP* 1606:115. doi: 10.1007/JHEP06(2016)115
- Ohta, N., Percacci, R., and Vacca, G. P. (2015). Flow equation for  $f(R)$  gravity and some of its exact solutions. *Phys. Rev. D* 92:061501. doi: 10.1103/PhysRevD.92.061501
- Ohta, N., Percacci, R., and Vacca, G. P. (2016). Renormalization group equation and scaling solutions for  $f(R)$  gravity in exponential parametrization. *Eur. Phys. J. C* 76:46. doi: 10.1140/epjc/s10052-016-3895-1
- Ostrogradsky, M. (1985). Mémoires sur les équations différentielles, relatives au problème des isopérimètres. *Mem. Acad. St. Petersburg* 6:385.
- Pagani, C., and Percacci, R. (2015). Quantum gravity with torsion and non-metricity. *Class. Quant. Grav.* 32:195019. doi: 10.1088/0264-9381/32/19/195019
- Pagani, C., and Reuter, M. (2018). Finite entanglement entropy in asymptotically safe quantum gravity. *JHEP* 1807:039. doi: 10.1007/JHEP07(2018)039
- Palanques-Mestre, A., and Pascual, P. (1984). The  $1/N^f$  expansion of the  $\gamma$  and beta functions in QED. *Commun. Math. Phys.* 95:277. doi: 10.1007/BF01212398
- Papenbrock, T., and Wetterich, C. (1995). Two loop results from one loop computations and nonperturbative solutions of exact evolution equations. *Z. Phys. C* 65:519. doi: 10.1007/BF01556140

- Pawlowski, J. M. (2007). Aspects of the functional renormalisation group. *Annals Phys.* 322:2831. doi: 10.1016/j.aop.2007.01.007
- Pawlowski, J. M., and Stock, D. (2018). Quantum improved schwarzschild-(A)dS and Kerr-(A)dS space-times. [*arXiv:1807.10512 hep-th*].
- Pawlowski, J. M., and Strodthoff, N. (2015). Real time correlation functions and the functional renormalization group. *Phys. Rev. D* 92:094009. doi: 10.1103/PhysRevD.92.094009
- Pelaggi, G. M., Plascencia, A. D., Salvio, A., Sannino, F., Smirnov, J., and Strumia, A. (2018). Asymptotically safe standard model extensions? *Phys. Rev. D* 97:095013. doi: 10.1103/PhysRevD.97.095013
- Pelissetto, A., and Vicari, E. (2002). Critical phenomena and renormalization group theory. *Phys. Rept.* 368:549. doi: 10.1016/S0370-1573(02)00219-3
- Percacci, R. (2017). "An introduction to covariant quantum gravity and asymptotic safety," in *100 years of General Relativity (Book 3)*, ed A. Ashtekar (Singapore: World Scientific Publishing).
- Percacci, R., and Vacca, G. P. (2014). Are there scaling solutions in the  $O(N)$ -models for large  $N$  in  $d > 4$  ? *Phys. Rev. D* 90:107702. doi: 10.1103/PhysRevD.90.107702
- Percacci, R., and Vacca, G. P. (2015). Search of scaling solutions in scalar-tensor gravity. *Eur. Phys. J. C* 75:188. doi: 10.1140/epjc/s10052-015-3410-0
- Percacci, R., and Vacca, G. P. (2017). The background scale ward identity in quantum gravity. *Eur. Phys. J. C* 77, 52. doi: 10.1140/epjc/s10052-017-4619-x
- Percacci, R., and Vacca, G. P. (2010). Asymptotic safety, emergence and minimal length. *Class. Q. Grav.* 27:245026 doi: 10.1088/0264-9381/27/24/245026
- Peskin, M. E. (1980). Critical point behavior Of the Wilson Loop. *Phys. Lett.* 94B:161. doi: 10.1016/0370-2693(80)90848-5
- Pica, C., and Sannino, F. (2011). UV and IR zeros of Gauge theories at the four loop order and beyond. *Phys. Rev. D* 83:035013. doi: 10.1103/PhysRevD.83.035013
- Pietrykowski, A. R. (2007). Gauge dependence of gravitational correction to running of gauge couplings. *Phys. Rev. Lett.* 98:061801. doi: 10.1103/PhysRevLett.98.061801
- Pietrykowski, A. R. (2013). Interacting scalar fields in the context of effective quantum gravity. *Phys. Rev. D* 87:024026. doi: 10.1103/PhysRevD.87.024026
- Platt, C., Hanke, W., and Thomale, R. (2013). Functional renormalization group for multi-orbital Fermi surface instabilities. *Adv. Phys.* 62, 453–562. doi: 10.1080/00018732.2013.862020
- Polonyi, J. (2003). Lectures on the functional renormalization group method. *Central Eur. J. Phys.* 1:1. doi: 10.2478/BF02475552
- Polyakov, A. M. (1975). Interaction of goldstone particles in two-dimensions. Applications to ferromagnets and massive yang-mills fields. *Phys. Lett.* 59B:79. doi: 10.1016/0370-2693(75)90161-6
- Rechenberger, S., and Saueressig, F. (2013). A functional renormalization group equation for foliated spacetimes. *JHEP* 1303:010. doi: 10.1007/JHEP03(2013)010
- Reuter, M. (1998). non-perturbative evolution equation for quantum gravity. *Phys. Rev. D* 57:971.
- Reuter, M., and Saueressig, F. (2002). Renormalization group flow of quantum gravity in the Einstein-Hilbert truncation. *Phys. Rev. D* 65:065016. doi: 10.1103/PhysRevD.65.065016
- Reuter, M., and Saueressig, F. (2011). Fractal space-times under the microscope: a renormalization group view on monte carlo data. *JHEP* 1112:012. doi: 10.1007/JHEP12(2011)012
- Reuter, M., and Saueressig, F. (2012). Quantum einstein gravity. *New J. Phys.* 14:055022. doi: 10.1088/1367-2630/14/5/055022
- Reuter, M., and Saueressig, F. (2013). Asymptotic safety, fractals, and cosmology. *Lect. Notes Phys.* 863:185. doi: 10.1007/978-3-642-33036-0\_8
- Reuter, M., and Schollmeyer, G. M. (2016). The metric on field space, functional renormalization, and metric-torsion quantum gravity. *Annals Phys.* 367:125. doi: 10.1016/j.aop.2015.12.004
- Reuter, M., and Schwindt, J. M. (2006) A Minimal length from the cutoff modes in asymptotically safe quantum gravity. *JHEP* 601:070. doi: 10.1088/1126-6708/2006/01/070
- Reuter, M., and Wetterich, C. (1994). Effective average action for gauge theories and exact evolution equations. *Nucl. Phys. B* 417, 181. doi: 10.1016/0550-3213(94)90543-6
- Reuter, M., and Wetterich, C. (1997). Gluon condensation in nonperturbative flow equations. *Phys. Rev. D* 56:7893. doi: 10.1103/PhysRevD.56.7893
- Robinson, S. P., and Wilczek, F. (2006). Gravitational correction to running of gauge couplings. *Phys. Rev. Lett.* 96:231601. doi: 10.1103/PhysRevLett.96.231601
- Rodigast, A., and Schuster, T. (2010). Gravitational corrections to Yukawa and  $\phi^4$  interactions. *Phys. Rev. Lett.* 104:081301. doi: 10.1103/PhysRevLett.104.081301
- Rosten, O. J. (2010). Fundamentals of the exact renormalization group. *arXiv:1003.1366 [hep-th]*.
- Rovelli, C., and Smolin, L. (1995). Discreteness of area and volume in quantum gravity. *Nucl. Phys. B* 442:593.
- Rytov, T. A., and Shrock, R. (2011). Higher-Loop corrections to the infrared evolution of a Gauge Theory with fermions. *Phys. Rev. D* 83:056011. doi: 10.1103/PhysRevD.83.056011
- Rytov, T. A., and Shrock, R. (2016). Infrared zero of  $\beta$  and value of  $\gamma_m$  for an  $SU(3)$  Gauge Theory at the five-loop level. *Phys. Rev. D* 94:105015. doi: 10.1103/PhysRevD.94.105015
- Safari, M. (2016). Splitting ward identity. *Eur. Phys. J. C* 76:201. doi: 10.1140/epjc/s10052-016-4036-6
- Salvio, A. (2018). Quadratic gravity. *arXiv:1804.09944 [hep-th]*.
- Sannino, F., and Shoemaker, I. M. (2015). Asymptotically safe dark matter. *Phys. Rev. D* 92:043518. doi: 10.1103/PhysRevD.92.043518
- Shaposhnikov, M., and Wetterich, C. (2010). Asymptotic safety of gravity and the higgs boson mass *Phys. Lett. B* 683:196. doi: 10.1016/j.physletb.2009.12.022
- Shaposhnikov, M., and Zenhausern, D. (2009a). Scale invariance, unimodular gravity and dark energy. *Phys. Lett. B* 671:187. doi: 10.1016/j.physletb.2008.11.054
- Shaposhnikov, M., and Zenhausern, D. (2009b). Quantum scale invariance, cosmological constant and hierarchy problem. *Phys. Lett. B* 671:162. doi: 10.1016/j.physletb.2008.11.041
- Simmons-Duffin, D. (2016). The conformal bootstrap. doi: 10.1142/9789813149441\_0001
- Sohnius, M. F. (1985). Introducing supersymmetry. *Phys. Rept.* 128:39. doi: 10.1016/0370-1573(85)90023-7
- Souma, W. (1999). Nontrivial ultraviolet fixed point in quantum gravity. *Prog. Theor. Phys.* 102:181. doi: 10.1143/PTP.102.181
- Stelle, K. S. (1977). Renormalization of higher derivative quantum gravity. *Phys. Rev. D* 16:953. doi: 10.1103/PhysRevD.16.953
- Surya, S. (2012). Evidence for a phase transition in 2D causal set quantum gravity. *Class. Quant. Grav.* 29:132001. doi: 10.1088/0264-9381/29/13/132001
- 't Hooft, G., and Veltman, M. J. G. (1974). One loop divergencies in the theory of gravitation. *Annales Poincare. Phys. Theor. A* 20:69.
- Tan, A., Xiao, M., Cui, X., Chen, X., Chen, Y., Fang, D., et al. (2016). Dark matter results from first 98.7 days of data from the PandaX-II experiment. *Phys. Rev. Lett.* 117:121303. doi: 10.1103/PhysRevLett.117.121303
- Tarasov, O. V., and Vladimirov, A. A. (1977). Two loop renormalization of the Yang-Mills Theory in an Arbitrary Gauge. *Sov. J. Nucl. Phys.* 25:585.
- Toms, D. J. (2007). Quantum gravity and charge renormalization. *Phys. Rev. D* 76:045015. doi: 10.1103/PhysRevD.76.045015
- Toms, D. J. (2008). Cosmological constant and quantum gravitational corrections to the running fine structure constant. *Phys. Rev. Lett.* 101:131301. doi: 10.1103/PhysRevLett.101.131301
- Toms, D. J. (2009). Quantum gravity, gauge coupling constants, and the cosmological constant. *Phys. Rev. D* 80:064040. doi: 10.1103/PhysRevD.80.064040
- Toms, D. J. (2010). Quantum gravitational contributions to quantum electrodynamics *Nature* 468:56. doi: 10.1038/nature09506
- Toms, D. J. (2011). Quadratic divergences and quantum gravitational contributions to gauge coupling constants. *Phys. Rev. D* 84:084016. doi: 10.1103/PhysRevD.84.084016
- Tsao, H. S. (1977). Conformal anomalies in a general background metric. *Phys. Lett.* 68B:79. doi: 10.1016/0370-2693(77)90039-9
- Unruh, W. G. (1989). A unimodular theory Of canonical quantum gravity. *Phys. Rev. D* 40:1048.

- Vacca, G. P., and Zambelli, L. (2015). Multimeson Yukawa interactions at criticality. *Phys. Rev. D* 91:125003. doi: 10.1103/PhysRevD.91.125003
- Vacca, G. P., and Zanusso, O. (2010). Asymptotic safety in Einstein gravity and scalar-fermion matter. *Phys. Rev. Lett.* 105:231601. doi: 10.1103/PhysRevLett.105.231601
- van de Ven, A. E. M. (1992). Two loop quantum gravity. *Nucl. Phys. B* 378, 309.
- Veneziano, G. (1979). U(1) without instantons. *Nucl. Phys. B* 159:213. doi: 10.1016/0550-3213(79)90332-8
- Weinberg, S. (1980). "UV divergences in quantum theories of gravitation," in *SPIRES Entry*, eds S. W., and Hawking, W. Israel (Cambridge: Cambridge University Press), 790–831. Available online at: <http://www.slac.stanford.edu/spires/find/hep/www?irn=784877>
- Wetterich, C. (1993). Exact evolution equation for the effective potential. *Phys. Lett. B* 301:90.
- Wilson, K. G., and Fisher, M. E. (1972). Critical exponents in 3.99 dimensions. *Phys. Rev. Lett.* 28:240. doi: 10.1103/PhysRevLett.28.240
- Woodard, R. P. (2015). Ostrogradsky's theorem on Hamiltonian instability. *Scholarpedia* 10:32243. doi: 10.4249/scholarpedia.32243
- Yagi, K., Blas, D., Barausse, E., and Yunes, N. (2014). Constraints on einstein-Æther theory and Hořava gravity from binary pulsar observations. *Phys. Rev. D* 89:084067.
- Zanusso, O., Zambelli, L., Vacca, G. P., and Percacci, R. (2010). Gravitational corrections to Yukawa systems. *Phys. Lett. B* 689:90. doi: 10.1016/j.physletb.2010.04.043
- Zinn-Justin, J. (2002). Quantum field theory and critical phenomena. *Int. Ser. Monogr. Phys.* 113:1. doi: 10.1093/acprof:oso/9780198509233.001.0001

**Conflict of Interest Statement:** The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Eichhorn. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Quadratic Gravity

Alberto Salvio\*

CERN, Theoretical Physics Department, Geneva, Switzerland

Adding the terms quadratic in the curvature to the Einstein-Hilbert action renders gravity renormalizable. This property is preserved in the presence of the most general renormalizable couplings with (and of) a generic quantum field theory (QFT). The price to pay is a massive ghost, which is due to the higher derivatives that the terms quadratic in the curvature imply. In this paper, the quadratic gravity scenario is reviewed including the recent progress on the related stability problem of higher derivative theories. The renormalization of the theory is also reviewed and the final form of the full renormalization group equations in the presence of a generic renormalizable QFT is presented. The theory can be extrapolated up to infinite energy through the renormalization group if all matter couplings flow to a fixed point (either trivial or interacting). Moreover, besides reviewing the above-mentioned topics, are some further insight on the ghost issue and the infinite energy extrapolation are provided. There is hope that in the future, this scenario might provide a phenomenologically viable and UV complete relativistic field theory of all interactions.

## OPEN ACCESS

### Edited by:

Oswaldo Cividanes,  
National University of La Plata,  
Argentina

### Reviewed by:

Marika Taylor,  
University of Southampton,  
United Kingdom  
Alan Stanley Cornell,  
University of the Witwatersrand,  
South Africa

### \*Correspondence:

Alberto Salvio  
alberto.salvio@cern.ch

### Specialty section:

This article was submitted to  
High-Energy and Astroparticle  
Physics,  
a section of the journal  
Frontiers in Physics

**Received:** 26 April 2018

**Accepted:** 29 June 2018

**Published:** 14 August 2018

### Citation:

Salvio A (2018) Quadratic Gravity.  
Front. Phys. 6:77.  
doi: 10.3389/fphy.2018.00077

**Keywords:** renormalization group, gravity, fixed point, relativity, field theory

## 1. INTRODUCTION AND SUMMARY

Relativistic field theories are the commonly accepted framework to describe particle physics and gravity, at least at currently accessible energies. An important question is whether such a framework could hold up to infinite energies and still agree with the experimental data. There are two serious difficulties that one has to overcome in order to give a positive answer to such a challenging question: the non-renormalizability of Einstein gravity [1, 2] and the presence of Landau poles in the Standard Model (SM).

Even if one does not quantize the gravitational field, it is known that quantum corrections due to any relativistic QFT generate terms that are not present in the Einstein-Hilbert action: specifically, local terms quadratic in the curvature tensor and with coefficients of dimension of non-negative powers of energy are generated [3], even if one sets them to zero at the classical level. Therefore, it is not possible to avoid them in a relativistic field theory. The resulting theory is commonly known as quadratic gravity<sup>1</sup> (QG). Starobinsky [4] exploited these unavoidable terms and noted that a non-singular solution that is initially in the de Sitter space can be obtained by taking them into account. This resulted in a pioneering model of inflation, one of the models favored by the Planck collaboration [5].

What happens if the quantum dynamics of the gravitational field is taken into account in QG? Weinberg [6] and Deser [7] suggested that QG is renormalizable (all physical quantities can be made finite by redefining the parameters and re-normalizing the fields) and few years later Stelle proved it rigorously [8].

<sup>1</sup>Other names sometimes used are “ $R^2$  gravity” and “higher derivative gravity,” as terms quadratic in the curvature have more than two derivatives of the gravitational field.

The presence of these local quadratic terms implies that classical QG belongs to the class of higher derivative theories analyzed a long time ago by Ostrogradsky [9], who proved that their Hamiltonian is unbounded from below. In QG, this manifests in the presence of a massive ghost, which is the price to pay to have a relativistic field theory of quantum gravity<sup>2</sup>. The importance of the quantum gravity problem has, however, encouraged several physicists to investigate whether QG can make sense and some recent progress in the ghost problem has been made. Most of the work done so far addressed the ghost problem within finite dimensional quantum mechanical models, and therefore, the case of a relativistic field theory (and in particular of QG) remains an important target for future research.

Another potential issue of QG is the clash between stability (understood as the absence of tachyons) and the absence of Landau poles [12, 13]: whenever the parameters were chosen to ensure stability, perturbation theory featured a Landau pole. Specifically, this Landau pole affected the parameter  $f_0$  appearing in the Lagrangian as  $\sqrt{-g}R^2/f_0^2$ , where  $g$  is the determinant of the spacetime metric  $g_{\mu\nu}$  and  $R$  is the Ricci scalar. Some progress has also been made in regards to this problem. In Salvio and Strumia [14], it was shown that QG coupled to a renormalizable QFT can hold up to infinite energies provided that all the couplings flow to a UV fixed point and the gravitational sector flows to conformal gravity (a version of gravity that is invariant under Weyl transformations,  $g_{\mu\nu}(x) \rightarrow e^{2\sigma(x)}g_{\mu\nu}(x)$ , where  $\sigma$  is a generic function of the spacetime point  $x$ ). The requirement that the QFT part enjoys a UV fixed point indicates the presence of several particles beyond the SM, which could be searched for with current and/or future particle experiments and could account for the strong evidence of new physics that we undoubtedly already have (such as neutrino oscillations and dark matter).

The aim of this work is to review what is known so far about QG (taking into account the coupling to a general renormalizable QFT). Other monographs and books on QG are present in the literature (see e.g., [15, 16], which focused on the renormalization of the theory). This review also includes the recent progress on the two problems mentioned above (the ghost and the Landau poles) and provides further insight on these issues. The article is structured as follows:

- In section 2, the action of QG coupled to a generic renormalizable QFT is discussed and the known physical degrees of freedom are identified with a new physically transparent method.
- Section 3 discusses the renormalizability of the theory; given that detailed proofs are present in the literature and, as mentioned above, books and reviews on this subject already

exist, we recall and elucidate a known intuitive argument in favor of renormalizability by providing more details than those currently available. In section 3, we also collect from the existing literature the full renormalization group equations (RGEs) for the dimensionless and dimensionful parameters of QG coupled to the most general renormalizable QFT.

- Section 4 is devoted to a pedagogical and detailed discussion of the ghost problem and the recent progress that has been made on this subject; most of the discussion, however, will be limited to simple finite dimensional quantum mechanical models and the extension to the full QG case remains an important target for future research.
- Section 5 reviews the issue of the Landau poles and how QG can flow to conformal gravity even in the presence of a generic QFT sector.

## 2. THE THEORY (INCLUDING A GENERAL MATTER SECTOR)

In this review, we do not consider only pure gravity but also its couplings to a general renormalizable matter sector.

### 2.1. Jordan-Frame Lagrangian

The full action in the so-called Jordan frame is,

$$S = \int d^4x \sqrt{-g} \mathcal{L}, \quad \mathcal{L} = \mathcal{L}_{\text{gravity}} + \mathcal{L}_{\text{matter}} + \mathcal{L}_{\text{non-minimal}}. \quad (2.1)$$

We describe in turn the three pieces—the pure gravitational Lagrangian,  $\mathcal{L}_{\text{gravity}}$ ; the matter Lagrangian,  $\mathcal{L}_{\text{matter}}$ ; the non-minimal couplings,  $\mathcal{L}_{\text{non-minimal}}$ .

#### The Pure Gravitational Lagrangian

$\mathcal{L}_{\text{gravity}}$  in quadratic gravity is obtained from the Einstein-Hilbert action by adding all possible local terms quadratic in the curvature, whose coefficients have the dimensionality of non-negative powers of energy:

$$\mathcal{L}_{\text{gravity}} = \alpha R^2 + \beta R_{\mu\nu} R^{\mu\nu} + \gamma R_{\mu\nu\rho\sigma} R^{\mu\nu\rho\sigma} - \frac{\bar{M}_P^2}{2} R - \Lambda, \quad (2.2)$$

where  $R_{\mu\nu\rho\sigma}$ ,  $R_{\mu\nu}$ , and  $R$  are the Riemann tensor, Ricci tensor, and Ricci scalar, respectively<sup>3</sup>, and the Greek indices are raised and lowered with  $g_{\mu\nu}$ . Furthermore,  $\alpha$ ,  $\beta$ , and  $\gamma$  are generic real coefficients. If the theory lives on a spacetime with boundaries, then one should also introduce in  $\mathcal{L}_{\text{gravity}}$  a term proportional to  $\square R$ , where  $\square$  is the covariant d'Alembertian, in order to preserve renormalizability [17–19]; in the applications described in this review such a term will not play any role and, therefore, will be neglected. Finally,  $\bar{M}_P$  and  $\Lambda$  are the reduced Planck mass and the cosmological constant, respectively.

<sup>3</sup>In this review we use the signature  $\eta_{\mu\nu} = \text{diag}(+1, -1, -1, -1)$  and define

$$R_{\mu\nu}{}^{\rho}{}_{\sigma} \equiv \partial_{\mu}\Gamma_{\nu\sigma}^{\rho} - \partial_{\nu}\Gamma_{\mu\sigma}^{\rho} + \Gamma_{\mu\tau}^{\rho}\Gamma_{\nu\sigma}^{\tau} - \Gamma_{\nu\tau}^{\rho}\Gamma_{\mu\sigma}^{\tau}, \quad \Gamma_{\mu\sigma}^{\rho} \equiv \frac{1}{2}g^{\rho\tau}(\partial_{\mu}g_{\sigma\tau} + \partial_{\sigma}g_{\mu\tau} - \partial_{\tau}g_{\mu\sigma}), \quad R_{\mu\nu} \equiv R_{\rho\mu}{}^{\rho}{}_{\nu}, \quad R \equiv g^{\mu\nu}R_{\mu\nu}.$$

<sup>2</sup>It should be noted that QG is distinct from the asymptotic safety proposal for quantum gravity made in Weinberg [10], where all the possible terms (including the non-renormalizable ones beyond the quadratic order) are included: in QG only renormalizable interactions are introduced so that only a finite number of parameters are present. This *guarantees* the predictivity of the theory. Possibly the ghost can be avoided by introducing an infinite series of higher-derivative terms [11], which can be viewed as non-local terms, but the resulting gravity theories contain infinite free parameters and are not known to be renormalizable.

One combination of the terms in (2.2) is a total (covariant) derivative, the topological Gauss-Bonnet term:

$$G \equiv R_{\mu\nu\rho\sigma} R^{\mu\nu\rho\sigma} - 4R_{\mu\nu} R^{\mu\nu} + R^2 = \frac{1}{4} \epsilon^{\mu\nu\rho\sigma} \epsilon_{\alpha\beta\gamma\delta} R^{\alpha\beta}_{\mu\nu} R^{\gamma\delta}_{\rho\sigma} = \text{divs}, \quad (2.3)$$

where  $\epsilon_{\mu\nu\rho\sigma}$  is the antisymmetric Levi-Civita tensor and “divs” represents the covariant divergence of some current. This total derivative does not contribute to the field equations and can be often ignored. It is therefore convenient to write (2.2) as,

$$\mathcal{L}_{\text{gravity}} = (\alpha - \gamma)R^2 + (\beta + 4\gamma)R_{\mu\nu}R^{\mu\nu} + \gamma G - \frac{\bar{M}_P^2}{2}R - \Lambda. \quad (2.4)$$

Furthermore, for reasons that will become apparent when the degrees of freedom will be identified in section 2.3, it is also convenient to express  $R_{\mu\nu}R^{\mu\nu}$  in terms of  $W^2 \equiv W_{\mu\nu\rho\sigma}W^{\mu\nu\rho\sigma}$ , where  $W_{\mu\nu\rho\sigma}$  is the Weyl tensor.

$$W_{\mu\nu\alpha\beta} \equiv R_{\mu\nu\alpha\beta} + \frac{1}{2}(g_{\mu\beta}R_{\nu\alpha} - g_{\mu\alpha}R_{\nu\beta} + g_{\nu\alpha}R_{\mu\beta} - g_{\nu\beta}R_{\mu\alpha}) + \frac{1}{6}(g_{\mu\alpha}g_{\nu\beta} - g_{\nu\alpha}g_{\mu\beta})R. \quad (2.5)$$

One has

$$\frac{1}{2}W_{\mu\nu\rho\sigma}W^{\mu\nu\rho\sigma} = \frac{1}{2}R_{\mu\nu\rho\sigma}R^{\mu\nu\rho\sigma} - R_{\mu\nu}R^{\mu\nu} + \frac{1}{6}R^2, \quad (2.6)$$

which, together with the definition of  $G$  in (2.3), gives

$$R_{\mu\nu}R^{\mu\nu} = \frac{W^2}{2} + \frac{R^2}{3} - \frac{G}{2}. \quad (2.7)$$

By inserting this expression of  $R_{\mu\nu}R^{\mu\nu}$  in (2.4), one finds,

$$\mathcal{L}_{\text{gravity}} = \frac{R^2}{6f_0^2} - \frac{W^2}{2f_2^2} - \epsilon G - \frac{\bar{M}_P^2}{2}R - \Lambda. \quad (2.8)$$

where

$$f_0^2 \equiv \frac{1}{2\beta + 2\gamma + 6\alpha}, \quad f_2^2 \equiv -\frac{1}{\beta + 4\gamma}, \quad \epsilon \equiv \frac{\beta}{2} + \gamma. \quad (2.9)$$

We have introduced the squares  $f_0^2$  and  $f_2^2$  because the absence of tachyonic instabilities requires  $f_0^2 > 0$  and  $f_2^2 > 0$ , as we will see in Sections 2.2, 2.3, and, in a more general context, in section 5.

## The Matter Lagrangian

The general matter content of a renormalizable theory includes real scalars  $\phi_a$ , Weyl fermions  $\psi_j$ , and vectors  $V_\mu^A$  (with field strength  $F_{\mu\nu}^A$ ) and its Lagrangian is,

$$\begin{aligned} \mathcal{L}_{\text{matter}} = & -\frac{1}{4}(F_{\mu\nu}^A)^2 + \frac{D_\mu\phi_a D^\mu\phi_a}{2} + \bar{\psi}_j i \not{D} \psi_j \\ & -\frac{1}{2}(Y_{ij}^a \psi_i \psi_j \phi_a + \text{h.c.}) - \mathcal{V}(\phi) - \frac{1}{2}(M_{ij} \psi_i \psi_j + \text{h.c.}), \end{aligned} \quad (2.10)$$

where

$$\mathcal{V}(\phi) = \frac{m_{ab}^2}{2} \phi_a \phi_b + \frac{A_{abc}}{3!} \phi_a \phi_b \phi_c + \frac{\lambda_{abcd}}{4!} \phi_a \phi_b \phi_c \phi_d, \quad (2.11)$$

where all the terms are contracted in a gauge-invariant way. The covariant derivatives are<sup>4</sup>:

$$D_\mu \phi_a = \partial_\mu \phi_a + i\theta_{ab}^A V_\mu^A \phi_b, \quad D_\mu \psi_j = \partial_\mu \psi_j + i t_{jk}^A V_\mu^A \psi_k + \frac{1}{2} \omega_{\mu}^{ab} \gamma_{ab} \psi_j$$

The gauge couplings are contained in the matrices  $\theta^A$  and  $t^A$ , which are the generators of the gauge group in the scalar and fermion representation, respectively, whereas  $Y_{ij}^a$  and  $\lambda_{abcd}$  are the Yukawa and quartic couplings, respectively. We have also added general renormalizable mass terms and cubic scalar interactions. Of course, for specific assignments of the gauge and global symmetries, some of these parameters can vanish, but here we use a general expression.

## The Non-minimal Couplings

$\mathcal{L}_{\text{non-minimal}}$  represents the non-minimal couplings between the scalar fields  $\phi_a$  and  $R$ :

$$\mathcal{L}_{\text{non-minimal}} = -\frac{1}{2} \xi_{ab} \phi_a \phi_b R, \quad (2.12)$$

where all terms are contracted in a gauge-invariant way. Non-minimal couplings are required by renormalizability, and if they are omitted at the classical level, quantum corrections generate them (as we will see in section 3.2.1).

## 2.2. Einstein Frame Lagrangian

The action in the Jordan frame is most suited to address the quantum aspects and to make contact with particle physics. However, when it comes to cosmological applications it is often better to express the gravitational part of the theory in a form closer to Einstein gravity [20, 21]. This will also help us in identifying the degrees of freedom in section 2.3. We now review how to obtain such a form of the theory and, in doing so, we shall neglect quantum corrections, which are anyway best studied in the Jordan frame.

The non-standard  $R^2$  term can be removed by adding to the Lagrangian the term

$$-\sqrt{-g} \frac{(R + 3f_0^2 \chi/2)^2}{6f_0^2}, \quad (2.13)$$

where  $\chi$  is an auxiliary field. This Lagrangian vanishes once the  $\chi$  EOM are used and we are therefore free to add it to the total Lagrangian. However, this has the effect of modifying the non-minimal couplings. The term linear in  $R$  in the Lagrangian now reads as:

$$-\frac{1}{2} \sqrt{-g} f(\chi, \phi) R, \quad f(\chi, \phi) \equiv \bar{M}_P^2 + \xi_{ab} \phi^a \phi^b + \chi. \quad (2.14)$$

<sup>4</sup>The spin-connection  $\omega_{\mu}^{ab}$  is defined as usual by  $\omega_{\mu}^{ab} = e^a_{\nu} \partial_{\mu} e^{b\nu} + e^a_{\rho} \Gamma_{\mu\sigma}^{\rho} e^{b\sigma}$  and  $\gamma_{ab} \equiv \frac{1}{4} [\gamma_a, \gamma_b]$ .

In order to get rid of this remaining non-standard term, we perform a Weyl transformation:

$$g_{\mu\nu} \rightarrow \frac{\bar{M}_P^2}{f} g_{\mu\nu}, \quad \phi^a \rightarrow \left(\frac{f}{\bar{M}_P^2}\right)^{1/2} \phi^a, \quad \psi_j \rightarrow \left(\frac{f}{\bar{M}_P^2}\right)^{3/4} \psi_j, \quad V_\mu^A \rightarrow V_\mu^A. \quad (2.15)$$

Now the Lagrangian can still be written as in (2.1), but with

$$\mathcal{L}_{\text{gravity}} = -\frac{W^2}{2f_0^2} - \frac{\bar{M}_P^2}{2} R + \text{divs.}, \quad \mathcal{L}_{\text{non-minimal}} = 0, \quad (2.16)$$

$$\begin{aligned} \mathcal{L}_{\text{matter}} = & -\frac{1}{4}(F_{\mu\nu}^A)^2 + \bar{\psi}_j i \not{D} \psi_j - \frac{1}{2}(Y_{ij}^a \psi_i \psi_j \phi_a + \text{h.c.}) \\ & - \frac{\sqrt{6}\bar{M}_P}{2\zeta}(M_{ij}\psi_i\psi_j + \text{h.c.}) + \frac{6\bar{M}_P^2}{\zeta^2} \frac{D_\mu \phi_a D^\mu \phi_a + \partial_\mu \zeta \partial^\mu \zeta}{2} \\ & - U(\zeta, \phi), \end{aligned} \quad (2.17)$$

where we defined<sup>5</sup>  $\zeta \equiv \sqrt{6f}$  and

$$U(\zeta, \phi) \equiv \frac{36\bar{M}_P^4}{\zeta^4} \left[ \mathcal{V}(\phi) + \Lambda + \frac{3f_0^2}{8} \left( \frac{\zeta^2}{6} - \bar{M}_P^2 - \xi_{ab}\phi_a\phi_b \right)^2 \right]. \quad (2.18)$$

In  $\mathcal{L}_{\text{gravity}}$ , we have not explicitly written the total derivatives as they typically do not play an important role in cosmology. These total derivatives emerge when the Weyl transformation is applied to the two terms proportional to  $\epsilon$  in (2.8).

The advantage of this form of the Lagrangian, known as the Einstein frame, is the absence of the non-minimal couplings and the  $R^2$  term. The latter has effectively been traded with the new scalar  $\zeta$ , which appears non-polynomially. The scalar kinetic terms are non-canonical and cannot be put in the canonical form with further field redefinitions given that the scalar field metric is not flat. Moreover, the Einstein frame potential  $U$  differs considerably from the Jordan-frame one,  $V + \Lambda$ . This result is a particular case of a more general theorem involving the generic functions  $f(R)$  of the Ricci scalar (for a review on  $f(R)$  theories see e.g., [22] and references therein). Also, note that the  $W^2$  term is present in the Einstein frame.

It is instructive to write the potential for  $\zeta$  when the other fields  $\phi_a$  are not present or are at the minimum of the potential and are not allowed to fluctuate (for example, because they have very large masses). In this case, one can make the kinetic term of  $\zeta$  canonical through the field redefinition  $\zeta = \sqrt{6}\bar{M}_P \exp(\omega/(\sqrt{6}\bar{M}_P))$ . The new field  $\omega$  feels a potential,

$$U(\omega) = \Lambda e^{-4\omega/\sqrt{6}\bar{M}_P} + \frac{3f_0^2 \bar{M}_P^4}{8} \left( 1 - e^{-2\omega/\sqrt{6}\bar{M}_P} \right)^2, \quad (2.19)$$

where we have neglected  $\mathcal{V}(\phi)$  and  $\xi_{ab}\phi_a\phi_b$  as they can be absorbed in  $\Lambda$  and  $\bar{M}_P^2$ , when the scalar fields  $\phi_a$  are absent or they are fixed to constant values. This is the potential of the famous Starobinsky's inflationary model [4]. There is a stationary point of  $U$  for

$$e^{-2\omega/\sqrt{6}\bar{M}_P} = \frac{3f_0^2 \bar{M}_P^4/8}{\Lambda + 3f_0^2 \bar{M}_P^4/8} \quad (2.20)$$

<sup>5</sup>Notice that in order for the metric redefinition in (2.15) to be regular one has to have  $f > 0$ , and thus we can safely take the square root of  $f$ .

whenever the right-hand side of the above equation is positive. For positive cosmological constant,  $\Lambda > 0$ , such a stationary point always exists for  $f_0^2 > 0$  when it is a point of minimum, but for  $f_0^2 < 0$  either the stationary point does not exist or it is a point of maximum, not minimum. This situation is illustrated in **Figure 1** and it is a special case of a more general result (valid when the other scalars  $\phi_a$  can fluctuate freely), which proves that a minimum of the potential exists only for  $f_0^2 > 0$  and will be presented in section 5.

### 2.3. The Degrees of Freedom of Quadratic Gravity

In section 2.2 we have seen that the  $R^2$  term is equivalent to a real scalar  $\zeta$ . We now complete the determination of the degrees of freedom of QG. We do so by working in the Einstein frame, where the gravity Lagrangian is the same as that in (2.16). The degrees of freedom associated with the matter Lagrangian can be identified with standard field theory methods and, therefore, we do not discuss them explicitly here.

The total derivatives ("divs") in (2.16) do not modify the degrees of freedom and for this reason will be neglected. Therefore, we focus on the following two terms in the gravity action:

$$S_W + S_{EH}, \quad (2.21)$$

where  $S_W$  is the part due to the unusual Weyl-squared term,

$$S_W = \int d^4x \sqrt{-g} \left[ -\frac{W^2}{2f_0^2} \right], \quad (2.22)$$

and  $S_{EH}$  is the usual Einstein-Hilbert part,

$$S_{EH} = \int d^4x \sqrt{-g} \left[ -\frac{\bar{M}_P^2}{2} R \right]. \quad (2.23)$$

We will use a 3 + 1 formalism (where space and time are treated separately). We do so because the identification of the degrees of freedom is particularly simple within that formalism.

In this section, however, we will expand the metric around the flat spacetime,  $ds_{\text{flat}}^2 = dt^2 - d\vec{x}^2$  as that is sufficient to determine the degrees of freedom perturbatively<sup>6</sup>. By choosing the Newtonian gauge, the metric describing the small linear fluctuations around the flat spacetime can be written as,

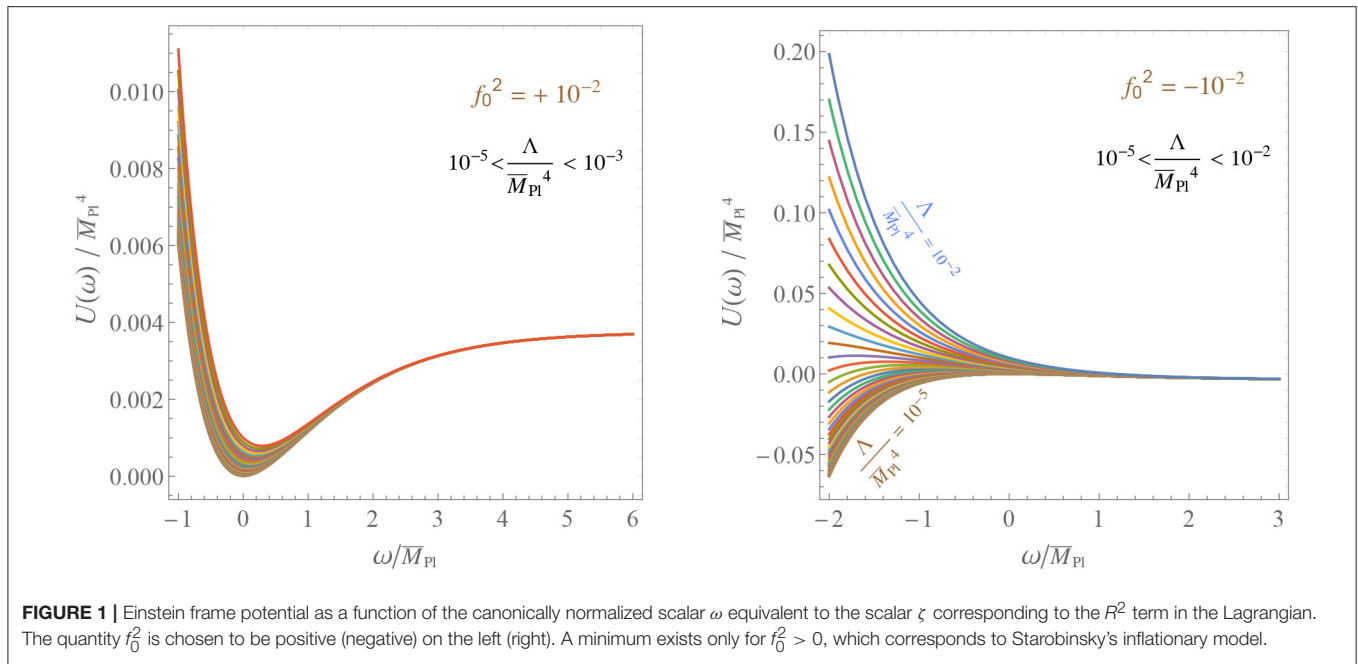
$$ds^2 = (1 + 2\Phi(t, \vec{x}))dt^2 - 2V_i(t, \vec{x})dt dx^i - [(1 - 2\Psi(t, \vec{x}))\delta_{ij} + h_{ij}(t, \vec{x})] dx^i dx^j. \quad (2.24)$$

By definition, the vector  $V_i$  (not to be confused with the spatial components of the gauge fields  $V_\mu^A$ ) and the tensor  $h_{ij}$  perturbations satisfy the following conditions:

$$\partial_i V_i = 0, \quad h_{ij} = h_{ji}, \quad h_{ii} = 0, \quad \partial_i h_{ij} = 0. \quad (2.25)$$

The Newtonian gauge is often used to study the small linear fluctuations around the Friedmann-Robertson-Walker (FRW)

<sup>6</sup>For a discussion of a possible form of the non-perturbative spectrum see Holdom and Ren [23, 24].



**TABLE 1** | Degrees of freedom in the gravitational sector.

Field	Spin	Mass
Graviton	2	0
Ghost	2	$M_2 \equiv f_2 \bar{M}_P / \sqrt{2}$
Scalar $\zeta$	0	$M_0 \equiv f_0 \bar{M}_P / \sqrt{2} + \dots$

The scalar  $\zeta$  is due to the  $R^2$  term in the Lagrangian; the dots in its mass  $M_0$  represent the possible contribution of other scalars mixing with  $\zeta$  (if any), which can be present in specific models.

cosmological metric (see e.g., [25] for a textbook treatment). Instead, here we study the fluctuations around the flat spacetime for simplicity. Also, sometimes the Newtonian gauge is defined for the scalar perturbations  $\Phi$  and  $\Psi$  only (see e.g., [25]). Here we consider a generalization, which also includes the non-scalar perturbations<sup>7</sup>. In Table 1, we provide the degrees of freedom of the gravitational sector (the part of the spectrum due to  $\mathcal{L}_{\text{gravity}}$ ). This includes the scalar  $\zeta$  found in section 2.2 and the ordinary graviton and a massive spin-2 ghost graviton, which will be identified in the following sections (2.3.1, 2.3.2, and 2.3.3).

### 2.3.1. Helicity-2 Sector

We start with the helicity-2 sector, whose quadratic action is denoted as  $S^{(2)}$ . Both  $S_{EH}$  and  $S_W$  contribute to this action. The helicity-2 quadratic action from  $S_{EH}$  and  $S_W$  are, respectively,

$$S_{ES}^{(2)} = \frac{\bar{M}_P^2}{8} \int d^4x \left( \dot{h}_{ij} \dot{h}_{ij} + h_{ij} \bar{\nabla}^2 h_{ij} \right),$$

$$S_W^{(2)} = -\frac{1}{4f_2^2} \int d^4x \left( \ddot{h}_{ij} \ddot{h}_{ij} + 2\dot{h}_{ij} \bar{\nabla}^2 \dot{h}_{ij} + h_{ij} \bar{\nabla}^4 h_{ij} \right), \quad (2.26)$$

<sup>7</sup>A possible gauge dependent divergence of  $h_{ij}$  has been set to zero by choosing the gauge appropriately.

where a dot denotes a derivative w.r.t. to time  $t$ ,  $\bar{\nabla}^4 \equiv (\bar{\nabla}^2)^2$  and  $\bar{\nabla}^2$  is the three-dimensional Laplacian. Therefore,

$$S^{(2)} = S_{EH} + S_W = \frac{\bar{M}_P^2}{8M_2^2} \int d^4x \left[ -\ddot{h}_{ij} \ddot{h}_{ij} - 2\dot{h}_{ij} \bar{\nabla}^2 \dot{h}_{ij} - h_{ij} \bar{\nabla}^4 h_{ij} + M_2^2 \left( \dot{h}_{ij} \dot{h}_{ij} + h_{ij} \bar{\nabla}^2 h_{ij} \right) \right], \quad (2.27)$$

where  $M_2^2 \equiv f_2^2 \bar{M}_P^2 / 2$ .

One can go to momentum space with a spatial Fourier transform,

$$h_{ij}(t, \vec{x}) = \int \frac{d^3p}{(2\pi)^{3/2}} e^{i\vec{p} \cdot \vec{x}} \sum_{\lambda=\pm 2} h_{\lambda}(t, \vec{p}) e_{ij}^{\lambda}(\hat{p}), \quad (2.28)$$

where  $e_{ij}^{\lambda}(\hat{p})$  are the usual polarization tensors for helicities  $\lambda = \pm 2$ . We recall that for  $\hat{p}$ , along the third axis, the polarization tensors satisfying (2.25) are given by,

$$e_{11}^{+2} = -e_{22}^{+2} = 1/2, \quad e_{12}^{+2} = e_{21}^{+2} = i/2, \quad e_{3i}^{+2} = e_{i3}^{+2} = 0,$$

$$e_{ij}^{-2} = (e_{ij}^{+2})^* \quad (2.29)$$

and for a generic momentum direction  $\hat{p}$  we can obtain  $e_{ij}^{\lambda}(\hat{p})$  by applying to (2.29) the rotation that connects the third axis with  $\hat{p}$ . The polarization tensors defined in this way also obey the orthonormality condition,

$$e_{ij}^{\lambda}(\hat{q})(e_{ij'}^{\lambda'}(\hat{q}))^* = \delta^{\lambda\lambda'}. \quad (2.30)$$

By using the Fourier expansion in (2.28), one obtains

$$\begin{aligned} S^{(2)} &= \frac{\bar{M}_P^2}{8M_2^2} \sum_{\lambda=\pm 2} \int dt d^3p \left[ -\dot{h}_\lambda^* \ddot{h}_\lambda + 2\dot{h}_\lambda^* \vec{p}^2 \dot{h}_\lambda - h_\lambda^* \vec{p}^4 h_\lambda \right. \\ &\quad \left. + M_2^2 \left( \dot{h}_\lambda^* \dot{h}_\lambda - h_\lambda^* \vec{p}^2 h_\lambda \right) \right] \\ &= \frac{\bar{M}_P^2}{8M_2^2} \sum_{\lambda=\pm 2} \int dt d^3p \left[ -\dot{h}_\lambda^* \ddot{h}_\lambda + (\omega_1^2 + \omega_2^2) |\dot{h}_\lambda|^2 \right. \\ &\quad \left. - \omega_1^2 \omega_2^2 |h_\lambda|^2 \right], \end{aligned} \quad (2.31)$$

where

$$\omega_1 \equiv \sqrt{\vec{p}^2 + M_2^2}, \quad \omega_2 \equiv |\vec{p}|. \quad (2.32)$$

The action  $S^{(2)}$  is the sum of the actions of Pais-Uhlenbeck oscillators, which will be studied in section 4.1.2. There we will see that this system is equivalent to a ghost d.o.f. with frequency  $\omega_1$  and a normal d.o.f. with frequency  $\omega_2$ . Therefore, the conclusion is that the helicity-2 sector features a massless field (the ordinary graviton) and a ghost field<sup>8</sup> with mass  $M_2 \equiv f_2 \bar{M}_P / \sqrt{2}$ . Therefore, as anticipated before, we see that  $f_2^2 > 0$  is required to avoid tachyonic instabilities. Lorentz invariance implies that the helicity-1 and helicity-0 components of the massive ghost should be present too. We will see how they emerge in the following sections 2.3.2 and 2.3.3. The derivation of the ghost field presented here simplifies and agrees with previous proofs based on the  $h_{\mu\nu}$  propagator [8, 27].

### 2.3.2. Helicity-1 Sector

Next, we move to the helicity-1 sector, whose quadratic action is denoted here by  $S^{(1)}$ .  $S^{(1)}$  is given by the sum of the Einstein-Hilbert contribution,

$$S_{EH}^{(1)} = \frac{\bar{M}_P^2}{4} \int d^4x \left( \partial_i V_j \right)^2, \quad (2.33)$$

and the Weyl contribution,

$$S_W^{(1)} = -\frac{1}{2f_2^2} \int d^4x \left( \partial_i \dot{V}_j \partial_i \dot{V}_j - V_i \vec{\nabla}^4 V_i \right). \quad (2.34)$$

Therefore, the full quadratic action in the helicity-1 sector is,

$$S^{(1)} = \int d^4x \frac{\bar{M}_P^2}{4M_2^2} \left[ \dot{V}_j \vec{\nabla}^2 \dot{V}_j + V_i \vec{\nabla}^4 V_i - M_2^2 V_j \vec{\nabla}^2 V_j \right]. \quad (2.35)$$

Given that  $\vec{\nabla}^2$  is a negatively-defined operator, we see that  $V_i$  has a ghost kinetic term and a mass  $M_2$  and has, therefore, to be identified with the helicity-1 components of the massive spin-2 ghost.

<sup>8</sup>Starting from the initial action (2.8), it is possible to perform field redefinitions and use the auxiliary field method to make the ghost field explicitly appear in the Lagrangian [26]. This is equivalent to what has been done in section 2.2 to make the scalar field  $\zeta$  appear explicitly in the Lagrangian.

### 2.3.3. Helicity-0 Sector

We denote the helicity-0 action by  $S^{(0)}$ , which has one contribution from the Weyl-squared term and one from the Einstein-Hilbert term,  $S^{(0)} = S_W^{(0)} + S_{EH}^{(0)}$ . Expanding around the flat spacetime leads to the following helicity-0 action (modulo total derivatives)

$$S_W^{(0)} = -\frac{2}{3f_2^2} \int d^4x \left[ \vec{\nabla}^2 (\Phi + \Psi) \right]^2, \quad (2.36)$$

$$S_{ES}^{(S)} = \frac{\bar{M}_P^2}{2} \int d^4x \left[ -6\dot{\Psi}^2 + 4\Psi \vec{\nabla}^2 \Phi - 2\Psi \vec{\nabla}^2 \Psi \right]. \quad (2.37)$$

The variation of  $S^{(0)}$  with respect to  $\Phi$  gives

$$-\frac{4}{3f_2^2 \bar{M}_P^2} \vec{\nabla}^4 (\Phi + \Psi) + 2\vec{\nabla}^2 \Psi = 0. \quad (2.38)$$

We see that this equation does not depend on the time derivative of the fields and, therefore, has to be considered as a constraint. Solving for  $\Phi$ :

$$\Phi = -\Psi + 3M_2^2 \vec{\nabla}^{-2} \Psi. \quad (2.39)$$

In the expression above,  $\vec{\nabla}^{-2}$  denotes the inverse Laplacian, which can be defined by going to momentum space,  $\vec{p}$ , and identifying  $\vec{\nabla}^{-2} \rightarrow -1/\vec{p}^2$ . Inserting (2.39) into Equations (2.36) and (2.37) we get,

$$\begin{aligned} S^{(0)} &= \frac{\bar{M}_P^2}{2} \int d^4x \left[ -6\dot{\Psi}^2 - 6\Psi \vec{\nabla}^2 \Psi + 6M_2^2 \Psi^2 \right] \\ &= 3\bar{M}_P^2 \int d^4x \left[ -(\partial \Psi)^2 + M_2^2 \Psi^2 \right]. \end{aligned} \quad (2.40)$$

We see that the kinetic term of  $\Psi$  is of the ghost type and its mass is  $M_2$ . Therefore,  $\Psi$  represents the helicity-0 component of the ghost spin-2 field.

## 3. RENORMALIZATION

One of the main motivations for considering QG is its improved quantum behavior with respect to Einstein theory. Therefore, it seems appropriate to discuss the renormalization properties right after the definition of the theory.

### 3.1. Renormalizability

The renormalizability of QG is suggested by simple power counting arguments, general covariance, and dimensional analysis. It is therefore not surprising that some authors [6, 7] noted this property several decades ago. There are also formal proofs [8, 28] of the renormalizability of QG, but we do not reproduce them here because they are described in detail in the original articles<sup>9</sup>.

It is illuminating, however, to recall the main ingredients of the intuitive arguments in favor of renormalizability. Let us consider the expansion of QG around the flat spacetime,

<sup>9</sup>These formal derivations can also be extended to include the general renormalizable matter sector considered in section 2.

$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}$ , and a generic loop correction in momentum space. The vertices involving  $h_{\mu\nu}$  contain at most 4 powers of the momenta  $p$ , whereas the  $h_{\mu\nu}$ -propagator behaves as  $1/p^4$  for large momenta if an appropriate quantization is used [8] (see below). Therefore, in this case, the superficial degree of divergence should be four or less (see, for example, Chapter 12 of [29]). This conclusion holds good both in the pure QG and in the presence of the most general renormalizable QFT.

It is instructive to illustrate the quantization that leads to a propagator that behaves as  $1/p^4$  for large momenta. The presence of the ordinary graviton and the spin-2 ghost with mass  $M_2$  tells us that the  $h_{\mu\nu}$ -propagator should have two poles,

$$\frac{Z_{\text{graviton}}}{p^2 + i\epsilon}, \quad \frac{Z_{\text{ghost}}}{p^2 - M_2^2 + i\epsilon'}, \quad (3.1)$$

where  $Z_{\text{graviton}}$  and  $Z_{\text{ghost}}$  are the corresponding residues and we have allowed for two a priori different prescriptions,  $\epsilon$  and  $\epsilon'$ . Both the poles are proportional to the same tensor structure as they both have spin-2. The requirement that the  $h_{\mu\nu}$ -propagator behaves as  $p^4$  for large momenta leads to the condition  $Z_{\text{graviton}} = -Z_{\text{ghost}}$ . In this case, the  $h_{\mu\nu}$ -propagator is proportional to,

$$\frac{1}{p^2 + i\epsilon} - \frac{1}{p^2 - M_2^2 + i\epsilon'} = -\frac{M_2^2}{(p^2 + i\epsilon)(p^2 - M_2^2 + i\epsilon')} + \pi i \delta(p^2 - M_2^2)(\text{sign}(\epsilon') - \text{sign}(\epsilon)), \quad (3.2)$$

where we have used the formula,

$$\frac{1}{x \pm i\epsilon} = \mathcal{P} \frac{1}{x} \mp i\pi \delta(x) \quad (3.3)$$

with  $\mathcal{P}$  being the principal part. The second term on the right-hand side of Equation (3.3) corresponds to the fact that the poles are shifted in different directions in the complex energy plane for  $\text{sign}(\epsilon') \neq \text{sign}(\epsilon)$ . Therefore, one obtains a propagator that behaves as  $1/p^4$  only if<sup>10</sup>  $\text{sign}(\epsilon') = \text{sign}(\epsilon)$ . Given that the absolute values of  $\epsilon$  and  $\epsilon'$  are not important this final condition can be simplified to  $\epsilon = \epsilon'$ .

The condition  $\epsilon = \epsilon'$  implies that the ghost should be quantized by introducing an indefinite metric on the Hilbert space [8]. The easiest way to show this is by looking at the action  $S^{(0)}$  of the helicity-0 component of the ghost in (2.40). This allows us to avoid the complications due to spacetime indices. The corresponding Lagrangian is,

$$\mathcal{L}^{(0)} = \frac{1}{2} \left( -\dot{\Psi}^2 - \Psi \vec{\nabla}^2 \Psi + M_2^2 \Psi^2 \right), \quad (3.4)$$

where we have canonically normalized  $\Psi$  by rescaling  $\Psi \rightarrow \Psi/\sqrt{6M_P}$ . The conjugate variable is then,

$$\Pi_\Psi = \frac{\partial \mathcal{L}^{(0)}}{\partial \dot{\Psi}} = -\dot{\Psi} \quad (3.5)$$

<sup>10</sup>To convince ourselves of the correctness of this statement one could insert the propagator in (3.3) in a loop integral; the effect of the Dirac  $\delta$ -function is to drop one momentum integration and to add a power of momentum at the denominator, for a total of two (not four) momenta in the power counting.

and the canonical commutators are:

$$[\Psi(t, \vec{x}), \dot{\Psi}(t, \vec{y})] = -i\delta^{(3)}(\vec{x} - \vec{y}), \quad [\Psi(t, \vec{x}), \Psi(t, \vec{y})] = 0, \\ [\dot{\Psi}(t, \vec{x}), \dot{\Psi}(t, \vec{y})] = 0. \quad (3.6)$$

Performing a spatial Fourier transform and demanding  $\Psi$  to solve its EOM leads to,

$$\Psi(t, \vec{x}) = \int \frac{d^3p}{\sqrt{2(2\pi)^3\omega(\vec{p})}} \left( b_0(\vec{p}) e^{i\vec{p}\cdot\vec{x} - i\omega(\vec{p})t} + b_0(\vec{p})^\dagger e^{-i\vec{p}\cdot\vec{x} + i\omega(\vec{p})t} \right), \quad (3.7)$$

where  $\omega(\vec{p}) \equiv \sqrt{\vec{p}^2 + M_2^2}$ , and the commutation rules above imply the following:

$$[b_0(\vec{p}), b_0(\vec{q})^\dagger] = -\delta(\vec{p} - \vec{q}), \quad [b_0(\vec{p}), b_0(\vec{q})] = 0. \quad (3.8)$$

At this point we have a choice. We can either

- 1 interpret the  $b_0$  ( $b_0^\dagger$ ) as annihilation (creation) operators, or
- 2 interpret the  $b_0$  ( $b_0^\dagger$ ) as creation (annihilation) operators.

In Case 1, as we will see in section 4.2.1, one should introduce an indefinite metric on the Hilbert space. In Case 2, the indefinite metric can be avoided, but the energies are negative; this statement will be shown in section 4.2.1, but its correctness is intuitive because in that case one would interpret  $-\omega(\vec{p})$  (rather than  $+\omega(\vec{p})$ ) as the energy. Let us compute the propagator  $P(x)$  in the two cases. The definition is,

$$P(x) \equiv \langle 0 | T \Psi(t, \vec{x}) \Psi(0) | 0 \rangle = \theta(t) P_+(x) + \theta(-t) P_-(x), \quad (3.9)$$

where

$$P_+(x) \equiv \langle 0 | \Psi(t, \vec{x}) \Psi(0) | 0 \rangle, \quad P_-(x) \equiv \langle 0 | \Psi(0) \Psi(t, \vec{x}) | 0 \rangle \quad (3.10)$$

- 1 In Case 1, we have,

$$P_+(x) = - \int \frac{d^3p}{2(2\pi)^3 p_0} e^{-ipx}, \quad P_-(x) = P_+(-x) \quad (3.11)$$

where  $p_0 \equiv \omega(\vec{p})$ . The minus sign in (3.11) is due to the minus sign in the commutation relation (3.8). Therefore, by using a standard text-book derivation,

$$P(x) = - \int \frac{d^4p}{(2\pi)^4 (p^2 - M_2^2 + i\epsilon)} e^{-ipx}, \quad (3.12)$$

where  $\epsilon > 0$ . We see that this corresponds to  $Z_{\text{ghost}} = -Z_{\text{graviton}}$  and  $\epsilon' = \epsilon$ .

- 2 In Case 2, we still have,

$$P_+(x) = - \int \frac{d^3p}{2(2\pi)^3 p_0} e^{-ipx}, \quad P_-(x) = P_+(-x), \quad (3.13)$$

but now  $p_0 = -\omega(\vec{p})$  (the energies are negative) and one ends up with

$$P(x) = - \int \frac{d^4p}{(2\pi)^4 (p^2 - M_2^2 - i\epsilon)} e^{-ipx}. \quad (3.14)$$

Note that the overall minus sign has a different origin than in Case 1: here it is due to the negative energy condition  $p_0 = -\omega(\vec{p})$ , not to the commutators as the role of  $b_0$  and  $b_0^\dagger$  is switched. So, in this case, one still has  $Z_{\text{ghost}} = -Z_{\text{graviton}}$  but  $\epsilon' = -\epsilon$  and renormalizability does not occur.

Therefore, the conclusion is that renormalizability requires a quantization with an indefinite metric on the Hilbert space. In section 4.2.1, we will show that such a metric should be introduced also to ensure that the Hamiltonian is bounded from below. This raises an interpretational problem as in quantum mechanics the positivity of the metric is related to the positivity of probabilities. This problem will be addressed in section 4.2.6, where the state of the art of the related literature will be discussed.

## 3.2. RGEs

The renormalizability of the theory (including the gravitational sector) allows us to use the standard renormalization group machinery developed for field theories without gravity. The modified minimal subtraction (MS) scheme will be adopted in this review.

### 3.2.1. RGEs of the Dimensionless Parameters

The 1-loop RGEs of the dimensionless parameters are independent of the dimensionful quantities and it is therefore convenient to present them separately. Their expression for a general renormalizable matter sector is,

$$\frac{df_2^2}{d\tau} = -f_2^4 \left( \frac{133}{10} + \frac{N_V}{5} + \frac{N_F}{20} + \frac{N_S}{60} \right), \quad (3.15)$$

$$\frac{df_0^2}{d\tau} = \frac{5}{3}f_2^4 + 5f_2^2f_0^2 + \frac{5}{6}f_0^4 + \frac{f_0^4}{12}(\delta_{ab} + 6\xi_{ab})(\delta_{ab} + 6\xi_{ab}), \quad (3.16)$$

$$\frac{d\epsilon}{d\tau} = -\left[ \frac{196}{45} + \frac{1}{360} \left( 62N_V + \frac{11}{2}N_F + N_S \right) \right], \quad (3.17)$$

$$\begin{aligned} \frac{d\xi_{ab}}{d\tau} = & \frac{1}{6}\lambda_{abcd}(6\xi_{cd} + \delta_{cd}) + (6\xi_{ab} + \delta_{ab}) \sum_{k=a,b} \left[ \frac{Y_2^k}{6} - \frac{C_{2S}^k}{2} \right] + \\ & -\frac{5f_2^4}{3f_0^2}\xi_{ab} + f_0^2\xi_{ac} \left( \xi_{cd} + \frac{2}{3}\delta_{cd} \right) (6\xi_{db} + \delta_{db}), \end{aligned} \quad (3.18)$$

$$\begin{aligned} \frac{dY^a}{d\tau} = & \frac{1}{2}(Y^{\dagger b}Y^bY^a + Y^aY^{\dagger b}Y^b) + 2Y^bY^{\dagger a}Y^b + \\ & + Y^b\text{Tr}(Y^{\dagger b}Y^a) - 3\{C_{2F}, Y^a\} + \frac{15}{8}f_2^2Y^a, \end{aligned} \quad (3.19)$$

$$\begin{aligned} \frac{d\lambda_{abcd}}{d\tau} = & \sum_{\text{perms}} \left[ \frac{1}{8}\lambda_{abef}\lambda_{efcd} + \frac{3}{8}\{\theta^A, \theta^B\}_{ab}\{\theta^A, \theta^B\}_{cd} \right. \\ & - \text{Tr} Y^a Y^{\dagger b} Y^c Y^{\dagger d} + \frac{5}{8}f_2^4\xi_{ab}\xi_{cd} + \frac{f_0^4}{8}\xi_{ae}\xi_{cf}(\delta_{eb} \\ & + 6\xi_{eb})(\delta_{fd} + 6\xi_{fd}) + \frac{f_0^2}{4!}(\delta_{ae} + 6\xi_{ae})(\delta_{bf} + 6\xi_{bf})\lambda_{efcd} \left. \right] \\ & + \lambda_{abcd} \left[ \sum_{k=a,b,c,d} (Y_2^k - 3C_{2S}^k) + 5f_2^2 \right], \end{aligned} \quad (3.20)$$

where

$$\tau \equiv \ln(\mu/\mu_0)/(4\pi)^2, \quad (3.21)$$

$\mu$  is the  $\overline{\text{MS}}$  energy scale,  $\mu_0$  is a fixed energy, and  $N_V$ ,  $N_F$ , and  $N_S$  are the numbers of gauge fields, Weyl fermions, and real scalars, respectively. Also,  $Y_2^k$ ,  $C_{2S}^k$ , and  $C_{2F}$  are defined by

$$\text{Tr}(Y^{\dagger a}Y^b) = Y_2^a\delta^{ab}, \quad \theta_{ac}^A\theta_{cb}^A = C_{2S}^a\delta_{ab}, \quad C_{2F} = t^A t^A. \quad (3.22)$$

The sum over “perms” in the RGEs of the  $\lambda_{abcd}$  runs over the 4! permutations of  $abcd$ . We do not show the RGEs of the gauge couplings because they are not modified by the gravitational couplings (see [30–33]).

Some terms in the 2-loop RGEs have been determined [14]. For example, switching off all couplings, except  $f_0$ , one obtains the 2-loop RGE for  $f_0$  [14] as,

$$\frac{df_0^2}{d\tau} = \frac{5}{6}f_0^4 - \frac{1}{(4\pi)^2} \frac{5}{12}f_0^6. \quad (3.23)$$

However, a complete expression of the 2-loop RGEs for all couplings is not available yet.

Note that the coefficient  $\epsilon$  of the topological term  $G$  does not appear in the RGEs of the other parameters. Indeed,  $G$  vanishes when the spacetime is topologically equivalent to the flat spacetime, and the RGEs, being UV effects, are independent of the global spacetime properties.

The RGEs obtained as above are the result of several works. The first attempt to determine the RGEs of  $f_2$  and  $f_0$  was presented in Julve and Tonin [34]. The results of Julve and Tonin [34] are incomplete and contain some errors. An improved calculation was later provided by Fradkin and Tseytlin [30, 35], which, however, still contains an error in the RGE of  $f_0$ . The first correct calculation of the RGE of  $f_0$  in the pure gravity case appeared in Avramidi and Barvinsky [13]; indeed, the result of Avramidi and Barvinsky [13] was later checked by Salvio and Strumia [33] and Codello and Percacci [36] with completely different techniques. Salvio and Strumia [33] also extended the results of Avramidi and Barvinsky [13] to include the general couplings to renormalizable matter sectors. The RGE for  $\epsilon$  in the presence of general renormalizable matter fields can be found in Avramidi [16] (see also [37] for a more recent discussion). Also, Ohta and Percacci [38] checked the RGEs of  $f_2$ ,  $f_0$ , and  $\epsilon$  with functional renormalization group methods.

Equations (3.15) and (3.16) clearly show that even if the spacetime metric is not quantized and we do not introduce the terms quadratic in the curvature in the Lagrangian, such terms are anyhow generated by loops of matter fields, as originally shown in Utiyama and DeWitt [3].

### 3.2.2. RGEs of the Dimensionful Parameters

The 1-loop RGEs of the dimensionful parameters are,

$$\begin{aligned} \frac{d\tilde{M}_P^2}{d\tau} = & \frac{1}{3}m_{aa}^2 + \frac{1}{3}\text{Tr}(M^\dagger M) + 2\xi_{ab}m_{ab}^2 + \left( \frac{2f_2^2}{3} - \frac{5f_2^4}{3f_0^2} \right. \\ & \left. + 2X\right) \tilde{M}_P^2, \end{aligned} \quad (3.24)$$

$$\begin{aligned} \frac{d\Lambda}{d\tau} = & \frac{m_{ab}^2 m_{ab}^2}{2} - \text{Tr}[(MM^\dagger)^2] + \frac{5f_2^4 + f_0^4}{8}\tilde{M}_P^4 \\ & + (5f_2^2 + f_0^2)\Lambda + 4\Lambda X, \end{aligned} \quad (3.25)$$

$$\frac{dM}{d\tau} = \frac{1}{2}(Y^{\dagger b} Y^b M + M Y^{\dagger b} Y^b) + 2Y^b M^{\dagger} Y^b + Y^b \text{Tr}(Y^{\dagger b} M) - 3\{C_{2F}, M\} + \frac{15}{8}f_2^2 M + MX, \quad (3.26)$$

$$\begin{aligned} \frac{dm_{ab}^2}{d\tau} = & \lambda_{abef} m_{ef}^2 + A_{aef} A_{bef} - 2[\text{Tr}(Y^{\dagger a} Y^{\dagger b} M M^{\dagger}) + \\ & + \text{Tr}(Y^{\dagger a} Y^b M^{\dagger} M) + \text{Tr}(Y^a M^{\dagger} Y^b M^{\dagger}) \\ & + \text{Tr}(M Y^{\dagger a} M Y^{\dagger b})] + \\ & + \frac{5}{2}f_2^4 \xi_{ab} \bar{M}_P^2 + \frac{f_0^4}{2} (\xi_{ab} + 6\xi_{ae}\xi_{eb}) \bar{M}_P^2 + \\ & + f_0^2 (m_{ab}^2 + 3\xi_{bf} m_{af}^2 + 3\xi_{af} m_{bf}^2 + 6\xi_{ae}\xi_{bf} m_{ef}^2) + \\ & + m_{ab}^2 \left[ \sum_{k=a,b} (Y_2^k - 3C_{2S}^k) + 5f_2^2 + 2X \right], \quad (3.27) \end{aligned}$$

$$\begin{aligned} \frac{dA_{abc}}{d\tau} = & \lambda_{abef} A_{efc} + \lambda_{acef} A_{efb} + \lambda_{bcef} A_{efa} + \\ & - 2\text{Tr}(Y^{\dagger a} Y^{\dagger b} Y^c M^{\dagger}) - 2\text{Tr}(Y^{\dagger c} Y^a Y^{\dagger b} M) + \\ & + f_0^2 (A_{abc} + 3\xi_{af} A_{fbc} + 3\xi_{bf} A_{fac} + 3\xi_{cf} A_{fab}) + \\ & + 6f_0^2 (\xi_{ae}\xi_{bf} A_{efc} + \xi_{ae}\xi_{cf} A_{efb} + \xi_{be}\xi_{cf} A_{efa}) + \\ & + A_{abc} \left[ \sum_{k=a,b,c} (Y_2^k - 3C_{2S}^k) + 5f_2^2 + X \right], \quad (3.28) \end{aligned}$$

where the curly brackets represent the sum over the permutations of the corresponding indices, e.g.,  $Y^{\dagger a} Y^{\dagger b} = Y^a Y^{\dagger b} + Y^b Y^{\dagger a}$ . The symbol  $X$  represents a gauge-dependent quantity [14]. The RGEs of massive parameters are gauge dependent as the unit of mass is gauge dependent. Any dimensionless ratio of dimensionful parameters is physical and the corresponding RGE is indeed gauge-independent, as it can be easily checked from Equations (3.24) to (3.28).

The RGEs above for the most general renormalizable matter sector were obtained in Salvio and Strumia [14] and later checked in Anselmi and Piva [39]. However, before Salvio and Strumia [14] appeared, a number of articles computed the RGEs of some massive parameters in less general models. The RGE for  $\Lambda/\bar{M}_P^4$  in the pure gravity theory was determined in Avramidi and Barvinsky [13] and a detailed description of the methods used can be found in Avramidi [16]. The RGE of the ratio between the Higgs squared mass  $M_h^2$  and  $\bar{M}_P^2$  was computed in Salvio and Strumia [33] (where the matter sector was identified with the SM).

These general RGEs can be used to address issues related to the high-energy extrapolation, such as the UV-completeness or the vacuum stability of generic theories of the sort studied here.

## 4. GHOSTS

In this section, we discuss systems (such as quadratic gravity) featuring ghosts, recall the related problems, and present some possible solutions. We will mostly focus on finite dimensional systems but also discuss both the classical and quantum mechanical aspects.

### 4.1. Ghosts in Classical Mechanics

We consider a physical system described by a certain number of coordinates<sup>11</sup>  $q_i$  and restrict our attention to Lagrangians that depend on  $q_i$ ,  $\dot{q}_i$ ,  $\ddot{q}_i$  and, possibly, on time  $t$ ,

$$L(q, \dot{q}, \ddot{q}, t), \quad (4.1)$$

where the dot is the derivative w.r.t.  $t$  and, from now on, we understand the index  $i$ . This setup covers the case we are interested in: the Lagrangian of quadratic gravity depends both on the first and second derivatives of the field variables because of the extra terms quadratic in the curvature; moreover, an explicit dependence on time emerges, e.g., when a cosmological background is considered [21].

In the following paragraphs, we will first discuss the derivation of Euler-Lagrange equations of motion and then introduce the Hamiltonian approach. This discussion will be valid for QG as a particular case.

The least action principle in this context tells us that the variation  $\delta S$  of the action,  $S \equiv \int dt L$ , with respect to variations  $\delta q$  of the coordinates that vanish on the time boundaries (together with their first derivatives,  $\delta \dot{q}$ ) should be zero<sup>12</sup>:

$$0 = \delta S = \int dt \left( \frac{\partial L}{\partial q} \delta q + \frac{\partial L}{\partial \dot{q}} \delta \dot{q} + \frac{\partial L}{\partial \ddot{q}} \delta \ddot{q} \right). \quad (4.2)$$

Here, we should require that  $\delta \dot{q}$  also vanishes on the time boundaries because the values of  $q$  at two times are not sufficient to identify the motion as the equations involve derivatives higher than the second order. On performing integration by parts once on the second term in (4.2) and twice on the third term, we obtain the Euler-Lagrange equations of motion for four-derivative theories as follows,

$$\frac{d}{dt} \left( \frac{\partial L}{\partial \dot{q}} - \frac{d}{dt} \frac{\partial L}{\partial \ddot{q}} \right) = \frac{\partial L}{\partial q}. \quad (4.3)$$

We now move to the Hamiltonian approach. We start by defining two canonical coordinates,

$$q_1 \equiv q, \quad q_2 \equiv \dot{q}. \quad (4.4)$$

In this case, the conjugate momenta are defined by

$$p_l \equiv \frac{\delta L}{\delta \dot{q}_l} \equiv \frac{\partial L}{\partial \dot{q}_l} - \frac{d}{dt} \frac{\partial L}{\partial \ddot{q}_l}, \quad (4.5)$$

where the index  $l$  runs over  $\{1, 2\}$ . A motivation for this definition will be given below in section 4.1.1. For  $l = 1$  and  $l = 2$  separately, the conjugate momenta read

$$p_1 = \frac{\partial L}{\partial \dot{q}} - \frac{d}{dt} \frac{\partial L}{\partial \ddot{q}}, \quad p_2 = \frac{\partial L}{\partial \ddot{q}}. \quad (4.6)$$

Then as usual, one defines the Hamiltonian  $H$  as,

$$H = p_l \dot{q}_l - L(q, \dot{q}, \ddot{q}, t). \quad (4.7)$$

<sup>11</sup>Note that the case of fields can be obtained by interpreting the index  $i$  as a space coordinate  $\vec{x}$ .

<sup>12</sup>The summation on the index  $i$  is understood, for example,  $\frac{\partial L}{\partial \dot{q}} \delta q \equiv \sum_i \frac{\partial L}{\partial \dot{q}_i} \delta q_i$ .

#### 4.1.1. The Ostrogradsky Theorem

Under a non-degeneracy assumption, i.e., the fact that<sup>13</sup>  $\det(\partial^2 L / \partial \ddot{q}^2) \neq 0$ , it is possible to argue that the system is classically unstable<sup>14</sup>.

Indeed, this assumption allows us to express  $\ddot{q}$  as,

$$\ddot{q} = f(q, \dot{q}, p_2, t), \quad (4.8)$$

where  $f$  is the inverse of  $\partial L / \partial \ddot{q}$  viewed as a function of  $\ddot{q}$ . Once Equations (4.4) and (4.8) are used,  $H$  reads

$$H = p_1 q_2 + p_2 f(q_1, q_2, p_2, t) - L(q_1, q_2, f(q_1, q_2, p_2, t), t), \quad (4.9)$$

which is manifestly a function of the form,

$$H = H(q_1, p_1, t). \quad (4.10)$$

The form of  $H$  in (4.9) implies the celebrated *Ostrogradsky theorem* [9]: *the Hamiltonian obtained from a Lagrangian of the form  $L(q, \dot{q}, \ddot{q}, t)$ , which depends non-degenerately on  $\ddot{q}$  (i.e.,  $\det(\partial^2 L / \partial \ddot{q}^2) \neq 0$ ), is not bounded from below*. Indeed, the expression of  $H$  in (4.9) shows that  $H$  depends linearly on the momentum  $p_1$  and therefore goes to  $-\infty$  if  $p_1$  tends either to  $+\infty$  or  $-\infty$  (when  $q_2$  is non-vanishing). Note that this result is valid for QG as a particular case.

One may wonder why the conjugate momenta is defined as in (4.5). The reason is that the standard form of the Hamiltonian equations of motion follows in this case and, therefore, the Hamiltonian is a constant of motion if it does not depend explicitly on time. In order to see this, let us consider an infinitesimal variation of the Hamiltonian and compute it in two different ways, by using (4.7) and (4.10). Respectively we have,

$$dH = p_1 d\dot{q}_1 + \dot{q}_1 dp_1 - \frac{\partial L}{\partial q} dq - \frac{\partial L}{\partial \dot{q}} d\dot{q} - \frac{\partial L}{\partial \ddot{q}} d\ddot{q} - \frac{\partial L}{\partial t} dt, \quad (4.11)$$

$$dH = \frac{\partial H}{\partial q_1} dq_1 + \frac{\partial H}{\partial p_1} dp_1 + \frac{\partial H}{\partial t} dt. \quad (4.12)$$

By using the definition of the conjugate momenta in (4.6) and  $q_2 = \dot{q}$  in the first expression of  $dH$ , we obtain,

$$dH = \dot{q}_1 dp_1 - \frac{\partial L}{\partial q} dq - \frac{d}{dt} \left( \frac{\partial L}{\partial \ddot{q}} d\dot{q} - \frac{\partial L}{\partial t} dt \right) = \dot{q}_1 dp_1 - \frac{\partial L}{\partial q} dq - \dot{p}_2 d\dot{q} - \frac{\partial L}{\partial t} dt. \quad (4.13)$$

The Euler-Lagrange equations allow us to write the term  $\frac{\partial L}{\partial \ddot{q}} d\dot{q}$  as follows:

$$\frac{\partial L}{\partial \ddot{q}} d\dot{q} = \frac{d}{dt} \left( \frac{\partial L}{\partial \ddot{q}} - \frac{d}{dt} \frac{\partial L}{\partial \dot{q}} \right) dq = \dot{p}_1 dq \quad (4.14)$$

so,

$$dH = \dot{q}_1 dp_1 - \dot{p}_1 dq_1 - \frac{\partial L}{\partial t} dt. \quad (4.15)$$

<sup>13</sup> $\partial^2 L / \partial \ddot{q}^2$  denotes the Hessian matrix of  $L$ , whose elements are  $\partial^2 L / \partial \ddot{q}_i \partial \ddot{q}_j$ .

<sup>14</sup>Lagrangians that depend on even higher derivatives of  $q$  have been considered in the literature in the time-independent case [40], but these situations go beyond our scope as the quadratic gravity Lagrangian only depends on the derivative of  $q$  up to the second order.

Now, by comparing this expression with the one in (4.12) we obtain,

$$\dot{q}_1 = \frac{\partial H}{\partial p_1}, \quad \dot{p}_1 = -\frac{\partial H}{\partial q_1}, \quad \frac{\partial H}{\partial t} = -\frac{\partial L}{\partial t}. \quad (4.16)$$

Therefore, we see that in theories with a Lagrangian of the form  $L(q, \dot{q}, \ddot{q}, t)$ , which depends non-degenerately on  $\ddot{q}$  (i.e.,  $\det(\partial^2 L / \partial \ddot{q}^2) \neq 0$ ), the Hamiltonian equations have the standard form provided that the definition of the conjugate momenta are modified according to (4.5). By inserting the first two equations in (4.16) into (4.12), we obtain that the Hamiltonian is a constant of motion provided that  $\partial H / \partial t = 0$ .

#### (In)stabilities

If a system fulfills the hypothesis of the Ostrogradsky theorem, then it can develop instabilities. However, this theorem does not directly imply that all solutions of such a system are unstable. Here, by “stable solution” we mean a solution of the equations of motion such that for initial conditions close enough to the region of the phase space spanned by this solution the motion is bounded (it does not run away). There are several examples of systems of this type that feature bounded motions: the Pais-Uhlenbeck model [40] to be discussed in section 4.1.2 (in some cases even in the presence of interactions [41–46]) and quadratic gravity expanded at linear level around the flat or de Sitter spacetime [21, 47, 48].

#### 4.1.2. The Pais-Uhlenbeck Model

The Ostrogradsky theorem applies to a large class of higher derivative theories, but we have seen that it does not directly forbid the existence of stable solutions. To further understand the issues of higher derivative theories, it is convenient to analyze a simple system, which captures some of the essential characteristics of quadratic gravity. Therefore, in this section we focus on the Pais-Uhlenbeck model [40], whose Lagrangian is

$$L = -\frac{\ddot{q}^2}{2} + (\omega_1^2 + \omega_2^2) \frac{\dot{q}^2}{2} - \omega_1^2 \omega_2^2 \frac{q^2}{2} - V(q) = -\frac{1}{2} q \left( \frac{d^2}{dt^2} + \omega_1^2 \right) \left( \frac{d^2}{dt^2} + \omega_2^2 \right) q - V(q) + \text{total derivatives}. \quad (4.17)$$

Here,  $V$  is a function of  $q$  representing a possible interaction, and  $\omega_1$  and  $\omega_2$  are real parameters. As we will see,  $\omega_1$  and  $\omega_2$  represent the frequencies of two decoupled oscillators when  $V = 0$ . Apart from its simplicity, another reason for considering this model is that it closely resembles the helicity-2 sector of QG (see Equation 2.31). In QG  $\omega_1 \neq \omega_2$  at finite spatial momentum (see Equation 2.32); therefore, the unequal frequency case is particularly relevant.

#### Lagrangian Analysis

The Lagrangian equation of motion is,

$$\left( \frac{d^2}{dt^2} + \omega_1^2 \right) \left( \frac{d^2}{dt^2} + \omega_2^2 \right) q + V'(q) = \frac{d^4 q}{dt^4} + (\omega_1^2 + \omega_2^2) \frac{d^2 q}{dt^2} + \omega_1^2 \omega_2^2 q + V'(q) = 0. \quad (4.18)$$

Equation (4.18) makes it clear why one chooses  $\omega_1^2$  and  $\omega_2^2$  to be positive; otherwise the solutions of the equations of motion would feature tachyonic instabilities at least for vanishing  $V$ .

The corresponding classical solution, for given initial conditions  $q_0 \equiv q(0)$ ,  $\dot{q}_0 \equiv \dot{q}(0)$ ,  $\ddot{q}_0 \equiv \ddot{q}(0)$ ,  $\ddot{\ddot{q}}_0 \equiv \ddot{\ddot{q}}(0)$  at  $t = 0$ , is

$$q(t) = -\frac{\omega_2^2 q_0 + \ddot{q}_0}{\omega_1^2 - \omega_2^2} \cos(\omega_1 t) + \frac{\omega_1^2 q_0 + \ddot{q}_0}{\omega_1^2 - \omega_2^2} \cos(\omega_2 t) - \frac{\omega_2^2 \dot{q}_0 + \ddot{\ddot{q}}_0}{\omega_1(\omega_1^2 - \omega_2^2)} \sin(\omega_1 t) + \frac{\omega_1^2 \dot{q}_0 + \ddot{\ddot{q}}_0}{\omega_2(\omega_1^2 - \omega_2^2)} \sin(\omega_2 t). \quad (4.19)$$

This is a well-behaved system without run-away issues for unequal frequencies,  $\omega_1 \neq \omega_2$ . By taking the limit  $\omega_1 \rightarrow \omega_2 \equiv \omega$  in the expression above, one obtains

$$q(t) = \sin(\omega t) \left[ \frac{t(q_0 \omega^2 + \ddot{q}_0)}{2\omega} + \frac{3\dot{q}_0 \omega^2 + \ddot{\ddot{q}}_0}{2\omega^3} \right] + \cos(\omega t) \left[ q_0 - \frac{t(\dot{q}_0 \omega^2 + \ddot{\ddot{q}}_0)}{2\omega^2} \right]. \quad (4.20)$$

Note that the amplitudes of the sine and cosine functions above grow linearly with  $t$ .

Run-away (i.e., unstable) solutions can also appear for  $\omega_1 \neq \omega_2$  if a non-quadratic potential, i.e.,  $V \neq 0$ , is introduced. However, it has been found numerically that the system admits stable solutions regardless of the unboundedness of the Hamiltonian for some choices of  $V$ , such as  $V(q) \propto \sin(q)^4$  [43]. The situation for this potential is illustrated in **Figure 2**. In Pavšič [45], it was found that the solutions are unstable unless  $V$  is bounded from below *and* above. Of course, this can only be generically true for  $\omega_1 \neq \omega_2$  because, for equal frequencies, we have seen that the motion is unbounded even for  $V = 0$ , which is certainly bounded from below and above.

## Hamiltonian Analysis

We can now construct the Hamiltonian<sup>15</sup> by using the general formulae of section 4.1. Ostrogradsky's canonical variables defined in (4.5) and (4.4) in this case read

$$q_1 = q, \quad p_1 = \frac{\partial L}{\partial \dot{q}} - \frac{d}{dt} \frac{\partial L}{\partial \ddot{q}} = (\omega_1^2 + \omega_2^2) \dot{q} + \ddot{q}, \quad (4.21)$$

$$q_2 = \dot{q}, \quad p_2 = \frac{\partial L}{\partial \ddot{q}} = -\ddot{q}.$$

Note that the non-degeneracy hypothesis of the Ostrogradsky theorem is obviously satisfied in this case:  $\partial^2 L / \partial \ddot{q}^2 = -1 \neq 0$ . Indeed, by using the general formula in (4.9) we obtain (in the Pais-Uhlenbeck model  $f(q, \dot{q}, p_2, t) = -p_2$ ),

$$H = p_1 q_2 - \frac{1}{2} p_2^2 - \frac{\omega_1^2 + \omega_2^2}{2} q_2^2 + \frac{\omega_1^2 \omega_2^2}{2} q_1^2 + V(q_1), \quad (4.22)$$

which is obviously unbounded from below. From (4.16) the Hamiltonian equations of motion are,

$$\begin{cases} \dot{q}_1 = \frac{\partial H}{\partial p_1} = q_2, & \dot{p}_1 = -\frac{\partial H}{\partial q_1} = -\omega_1^2 \omega_2^2 q_1 - V'(q_1), \\ \dot{q}_2 = \frac{\partial H}{\partial p_2} = -p_2, & \dot{p}_2 = -\frac{\partial H}{\partial q_2} = -p_1 + (\omega_1^2 + \omega_2^2) q_2. \end{cases} \quad (4.23)$$

They imply the classical Euler-Lagrange equation of motion in (4.18).

When  $\omega_1 \neq \omega_2$ , the Hamiltonian in (4.22) can be brought in diagonal form (except for the effect of the interaction  $V$ )

$$H = -\frac{1}{2} (\tilde{p}_1^2 + \omega_1^2 \tilde{q}_1^2) + \frac{1}{2} (\tilde{p}_2^2 + \omega_2^2 \tilde{q}_2^2) + V(q_1) \quad (4.24)$$

through the canonical transformation,

$$\begin{aligned} q_1 &= \frac{\tilde{q}_2 - \tilde{p}_1/\omega_1}{\sqrt{\omega_1^2 - \omega_2^2}}, & q_2 &= \frac{\tilde{p}_2 - \omega_1 \tilde{q}_1}{\sqrt{\omega_1^2 - \omega_2^2}}, \\ p_1 &= \omega_1 \frac{\omega_1 \tilde{p}_2 - \omega_2^2 \tilde{q}_1}{\sqrt{\omega_1^2 - \omega_2^2}}, & p_2 &= \frac{\omega_2^2 \tilde{q}_2 - \omega_1 \tilde{p}_1}{\sqrt{\omega_1^2 - \omega_2^2}}. \end{aligned} \quad (4.25)$$

which satisfies  $q_1 p_1 - q_2 p_2 = \tilde{p}_2 \tilde{q}_2 - \tilde{p}_1 \tilde{q}_1$ . Its inverse is

$$\begin{aligned} \tilde{q}_1 &= \frac{p_1 - \omega_1^2 q_2}{\omega_1 \sqrt{\omega_1^2 - \omega_2^2}}, & \tilde{q}_2 &= \frac{\omega_1^2 q_1 - p_2}{\sqrt{\omega_1^2 - \omega_2^2}}, \\ \tilde{p}_1 &= \omega_1 \frac{\omega_2^2 q_1 - p_2}{\sqrt{\omega_1^2 - \omega_2^2}}, & \tilde{p}_2 &= \frac{p_1 - \omega_2^2 q_2}{\sqrt{\omega_1^2 - \omega_2^2}}. \end{aligned} \quad (4.26)$$

Note that, given the first equation in (4.25),  $V(q_1)$  introduces interactions between  $\tilde{q}_2$  and  $\tilde{p}_1$ . However, from (4.24) one can see that the system for  $V = 0$  is equivalent to two decoupled oscillators with frequencies  $\omega_1$  and  $\omega_2$ . Note that the first oscillator contributes negatively to the Hamiltonian; this is the manifestation of the Ostrogradsky theorem in this basis. Since the derivation of (4.24) is valid only for  $\omega_1 \neq \omega_2$  (because otherwise the transformation in (4.25) would be singular), one might hope to have a classical Hamiltonian that is bounded from below for  $\omega_1 = \omega_2$ . This is not the case as the Hamiltonian in the form given in (4.22) is valid for  $\omega_1 = \omega_2$  too and is not bounded from below.

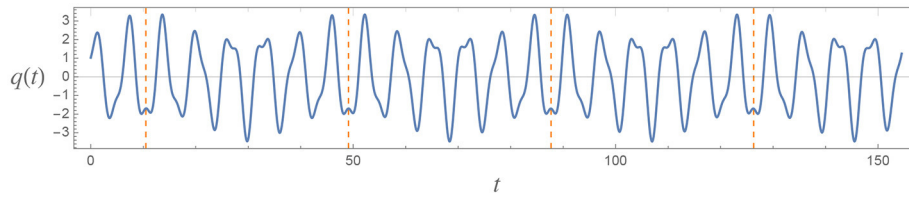
## 4.2. Quantum Mechanics With Ghosts

Before examining the peculiar features of the quantization with ghosts, let us spell out some basic assumptions of standard quantum mechanics, which will be made in the presence of ghosts too, including in the case of QG.

- Quantizing the theory consists in substituting the canonical coordinates  $q_j$  and conjugate momenta  $p_j$  with some operators acting on a vector space, whose elements are identified with the possible states of the system<sup>16</sup>.

<sup>15</sup>An analogous construction for QG was performed in Buchbinder et al. [15], Buchbinder and Lyakhovich [49, 50], and Kluson et al. [51].

<sup>16</sup>For simplicity, we will continue to use the same symbol to denote the quantum operators and the corresponding classical variables (when this does not create confusion).



**FIGURE 2 |** Solution to the equation of motion (4.18) of the Pais-Uhlenbeck model with  $V(q) = \lambda \sin(q)^4$ . The plot is presented in units of  $\omega_2$ . The other parameters are set as follows:  $\omega_1 = 2.1$ ,  $\lambda = 1.022$ . The motion appears to be bounded and periodic (the vertical dashed lines indicate the period).

- The Hamiltonian  $H$  in quantum mechanics is defined as a self-adjoint operator ( $H^\dagger = H$ ) with respect to some metric on the vector space of states.  $H$  generates the time evolution: the state  $|\psi_t\rangle$  at time  $t$  is given by,

$$|\psi_t\rangle = U(t)|\psi_0\rangle, \quad U(t) \equiv e^{-iHt}. \quad (4.27)$$

Moreover, the Hamiltonian is assumed to have the same expression in terms of  $q_j$  and  $p_j$  as in classical mechanics, Equation (4.10).

- The canonical coordinates  $q_j$  and their conjugate momenta  $p_j$  are promoted to operators by imposing the canonical commutators, i.e.,

$$[q_j, p_k] = i\delta_{jk}, \quad [q_j, q_k] = 0, \quad [p_j, p_k] = 0 \quad (4.28)$$

and requiring them to be self-adjoint:  $q_j^\dagger = q_j$  and  $p_j^\dagger = p_j$ .

The possible probabilistic interpretations of quantum theories with ghosts will be discussed in section 4.2.6.

Most of the efforts that have been done so far in quantizing theories with ghosts have focused on simple toy models, which isolate the main source of concern—the presence of four time-derivatives. The model that is typically studied is the quantum version of the Pais-Uhlenbeck construction given in section 4.1.2, which is perhaps the simplest four-derivative extension of an ordinary quantum mechanical model. Therefore, we will mostly focus on it. However, some of the results reviewed in this section can be applied to other models too.

#### 4.2.1. Trading Negative Energies With Negative Norms

The first thing one can prove is that some Hamiltonians that are not bounded from below can be quantized in a way that their quantum spectrum is instead bounded from below, but this is achieved by introducing an indefinite metric on the Hilbert space (as we will see, this is precisely the metric with respect to which  $H$ ,  $q_j$ , and  $p_j$  have been assumed to be self-adjoint). A classic example is the Pais-Uhlenbeck Hamiltonian<sup>17</sup> in Equation (4.24) for vanishing  $V$ , which we will now discuss in some detail.

<sup>17</sup>It is important to recall that Hamiltonian (4.24) is equivalent to the original Hamiltonian in (4.22) when  $\omega_1 \neq \omega_2$  a condition that is assumed to hold here (for the quantization of the equal frequency Pais-Uhlenbeck model see e.g., [46, 52–55]).

The part of the classical Hamiltonian that contributes negatively is

$$H_1 \equiv -\frac{1}{2}(\tilde{p}_1^2 + \omega_1^2 \tilde{q}_1^2), \quad (4.29)$$

and it is on this part that we shall focus as the other one  $H_2 \equiv \frac{1}{2}(\tilde{p}_2^2 + \omega_2^2 \tilde{q}_2^2)$ , being positive, can be quantized with standard methods. Note that the quadratic Hamiltonian of the ghost of QG can be written as the sum of Hamiltonians of the form (4.29), as is clear from Equations (2.40) and (2.35) and the fact that the Lagrangian (2.31) of the helicity-2 sector of QG is the sum of Pais-Uhlenbeck Lagrangians.

What allows us to trade the negative energy in Equation (4.29) with negative norm is the exchange of creation and annihilation operators: one defines the annihilation and creation operators, respectively, as

$$\tilde{a}_1 \equiv \sqrt{\frac{\omega_1}{2}} \left( \tilde{q}_1 - i \frac{\tilde{p}_1}{\omega_1} \right), \quad \tilde{a}_1^\dagger \equiv \sqrt{\frac{\omega_1}{2}} \left( \tilde{q}_1 + i \frac{\tilde{p}_1}{\omega_1} \right), \quad (4.30)$$

where we used  $\tilde{q}_1^\dagger = \tilde{q}_1$  and  $\tilde{p}_1^\dagger = \tilde{p}_1$ . The relative signs between  $\tilde{q}_1$  and  $\tilde{p}_1$  have been switched with respect to the standard case. Here, we keep the label 1 to recall that the oscillator with label 2 is subject to the usual definition of annihilation and creation operators:

$$\tilde{a}_2 \equiv \sqrt{\frac{\omega_2}{2}} \left( \tilde{q}_2 + i \frac{\tilde{p}_2}{\omega_2} \right), \quad \tilde{a}_2^\dagger \equiv \sqrt{\frac{\omega_2}{2}} \left( \tilde{q}_2 - i \frac{\tilde{p}_2}{\omega_2} \right). \quad (4.31)$$

From the canonical commutators (4.28) and by using the canonical transformation in (4.26) it follows

$$[\tilde{q}_j, \tilde{p}_k] = i\delta_{jk}, \quad [\tilde{q}_j, \tilde{q}_k] = 0, \quad [\tilde{p}_j, \tilde{p}_k] = 0, \quad (4.32)$$

which leads to

$$[\tilde{a}_j, \tilde{a}_k^\dagger] = \eta_{jk}, \quad [\tilde{a}_j, \tilde{a}_k] = 0, \quad [\tilde{a}_j^\dagger, \tilde{a}_k^\dagger] = 0, \quad (4.33)$$

where  $\eta_{11} = -1$ ,  $\eta_{22} = 1$ ,  $\eta_{12} = \eta_{21} = 0$ . One can now express  $\tilde{q}_1$  and  $\tilde{p}_1$  in terms of  $\tilde{a}_1$  and  $\tilde{a}_1^\dagger$  as usual and find

$$H_1 = -\omega_1 \tilde{a}_1^\dagger \tilde{a}_1 + \frac{\omega_1}{2} \equiv \omega_1 N_1 + \frac{\omega_1}{2}, \quad (4.34)$$

where we defined a number operator  $N_1 \equiv -\tilde{a}_1^\dagger \tilde{a}_1$  (see below) with an unusual minus sign. Indeed, with this definition  $N_1$ ,  $\tilde{a}_1$ , and  $\tilde{a}_1^\dagger$  satisfy the usual commutation relations

$$[N_1, \tilde{a}_1] = -\tilde{a}_1, \quad [N_1, \tilde{a}_1^\dagger] = \tilde{a}_1^\dagger, \quad (4.35)$$

which allows us to interpret  $\tilde{a}_1$  and  $\tilde{a}_1^\dagger$  as annihilation and creation operators, respectively: the eigenstates of  $N_1$ , i.e.,  $N_1|n_1\rangle = n_1|n_1\rangle$ , satisfy

$$\tilde{a}_1|n_1\rangle = c(n_1)|n_1-1\rangle, \quad \tilde{a}_1^\dagger|n_1\rangle = d(n_1)|n_1+1\rangle. \quad (4.36)$$

We can determine  $c$  and  $d$  up to an overall phase, once the normalization of  $|n_1\rangle$  is fixed. Here, for reasons that will become clear shortly, we allow some norms to be negative and we choose the normalizations<sup>18</sup>  $\langle n_1|n_1\rangle = v_{n_1}$ , where  $v_{n_1} = \pm 1$ . Notice now

$$-v_{n_1}n_1 = \langle n_1|\tilde{a}_1^\dagger \tilde{a}_1|n_1\rangle = |c(n_1)|^2 \langle n_1-1|n_1-1\rangle = |c(n_1)|^2 v_{n_1-1}, \quad (4.37)$$

which leads to

$$|c(n_1)|^2 = -\frac{v_{n_1}}{v_{n_1-1}} n_1. \quad (4.38)$$

If all the norms are positive, i.e., all  $v_{n_1} = 1$ , then it is possible to show with a standard textbook argument that the spectrum of  $N_1$  (and therefore, because of Equation (4.34), that of the Hamiltonian) is not bounded from below. This is because Equation (4.38) tells us that  $n_1 < 0$  and we can then reach an arbitrary large and negative value of  $n_1$  by acting with the annihilation operator.

The only way to avoid  $n_1 < 0$  is to take  $v_{n_1} = -v_{n_1-1}$ . Indeed, in this case (4.38) gives<sup>19</sup>

$$|c(n_1)|^2 = n_1, \quad (4.41)$$

which as usual implies that the spectrum of  $N_1$  is  $\{n_1\} = \{0, 1, 2, 3, \dots\}$  (and therefore  $N_1$  can appropriately be identified with a number operator) and the spectrum of the Hamiltonian is thus bounded from below. The state with  $n_1 = 0$  is interpreted as that without ghost quanta and so we require it to have a positive norm. Therefore,  $v_{n_1} = -v_{n_1-1}$  implies that the states with an even (odd) number of ghost quanta have positive (negative) norm.

A similar reasoning can be done in QG linearized around the flat spacetime: the energy becomes bounded from below

<sup>18</sup>More general assignments,  $v_{n_1} \neq \pm 1$  are equivalent because we can always re-normalize the states in a way that  $v_{n_1} = \pm 1$  as long as there are no zero norm states, which we assume here.

<sup>19</sup>In order to fix  $d(n_1)$  consider

$$-v_{n_1}(n_1+1) = \langle n_1|\tilde{a}_1^\dagger \tilde{a}_1|n_1+1\rangle = \langle n_1|a_1 \tilde{a}_1^\dagger|n_1+1\rangle = |d(n_1)|^2 \langle n_1+1|n_1+1\rangle = |d(n_1)|^2 v_{n_1+1}, \quad (4.39)$$

which gives

$$|d(n_1)|^2 = n_1 + 1. \quad (4.40)$$

if an indefinite metric on the Hilbert space is introduced (see section 3.1). Furthermore, we saw in section 3.1 that an indefinite metric should also be present in order for QG to be renormalizable. Therefore, insisting on having arbitrarily negative energies to preserve the positivity of the metric appears to have very little motivation.

As mentioned before, in this construction  $q_j$ ,  $p_j$ , and  $H$  are self-adjoint w.r.t. the indefinite metric. This leads to problems in the definition of probabilities, which we shall address in section 4.2.6.

#### 4.2.2. The Problem of the Wave-Function Normalization

So far we have given some features of the quantum theory, but we have not yet specified completely the quantization procedure. We still have to define the spectrum of the operators  $q_j$ .

Let us discuss this point in the Pais-Uhlenbeck model with  $\omega_1 \neq \omega_2$  for the sake of definiteness. One possibility would be to assume, as usual, that the spectrum is real for both  $q_1$  and  $q_2$ . However, this leads to non-normalizable wave functions [56, 57]. To see this, we consider the ground-state wave function  $\psi_0(q_1, q_2) \equiv \langle q_1, q_2|0\rangle$ , where  $|0\rangle$  is the vacuum, defined as  $\tilde{a}_1|0\rangle = 0$  and  $\tilde{a}_2|0\rangle = 0$ , while  $|q_1, q_2\rangle$  is an eigenstate of  $q_1$  and  $q_2$ . Using the standard representation for the conjugate momentum acting on the wave functions,  $p_i = -i\partial/\partial q_i$ , one obtains the ground-state wave function

$$\psi_0(q_1, q_2) \propto \exp\left(\frac{-q_1^2 \omega_1 \omega_2 + q_2^2}{2}(\omega_1 + \omega_2) - iq_1 q_2 \omega_1 \omega_2\right). \quad (4.42)$$

With this quantization,  $\psi_0(q_1, q_2)$  is non-normalizable along the  $q_2$ -direction. However,  $\psi_0(q_1, q_2)$  becomes normalizable when one performs the integral of  $|\psi_0(q_1, q_2)|^2$  on the imaginary  $q_2$ -axis.

This suggests that one could obtain a consistent quantization by requiring  $q_2$  to have a purely imaginary spectrum, while assuming a standard quantization (with real spectrum) for  $q_1$  [58].

#### 4.2.3. The Dirac-Pauli Quantization

The quantization with purely imaginary eigenvalues for a canonical variable  $\hat{x}$  was first discussed by Pauli [59] for Lagrangians with at most 2 time-derivatives, elaborating on a previous work by Dirac [60]. In the rest of this work, we will refer to this unusual quantization as the Dirac-Pauli quantization. To proceed, let us deduce some basic properties of the Dirac-Pauli quantization for a generic variable  $\hat{x}$ .

The defining property is that the spectrum of  $\hat{x}$  is purely imaginary:

$$\hat{x}|x\rangle = ix|x\rangle. \quad (4.43)$$

It follows  $\langle x'|\hat{x}|x\rangle = ix\langle x'|x\rangle$ , which, together with the self-adjointness of  $\hat{x}$ , i.e.,  $\langle x'|\hat{x}|x\rangle = \langle x|\hat{x}|x'\rangle^* = -ix'\langle x'|x\rangle^* = -ix'\langle x'|x\rangle$ , implies  $(x+x')\langle x'|x\rangle = 0$ . The general solution to this equation is  $\langle x'|x\rangle = \delta(x+x')h(x)$ , where  $h$  is a function that

we set to 1 without loss of generality: this can always be done by rescaling the states  $|x\rangle$ . Then, one obtains

$$\langle x'|x\rangle = \delta(x+x') \quad (4.44)$$

and the completeness<sup>20</sup> condition reads

$$\begin{aligned} \int dx |x\rangle \langle -x| &= 1, \quad \Longleftrightarrow \quad \int dx |x\rangle \langle x| = \eta, \\ \Longleftrightarrow \quad \int dx |x\rangle \langle x| \eta &= 1, \end{aligned} \quad (4.45)$$

where  $\eta$  is the operator defined by  $\eta|x\rangle = |-x\rangle$ .

It can be shown that the variable  $\hat{p}$  canonically conjugate to  $\hat{x}$  is also a Dirac-Pauli variable: i.e.,  $\hat{p}|p\rangle = ip|p\rangle$ , where  $p$  is a generic real number. To show this, we first notice that the operator  $\exp(\hat{p}a)$ , where  $a$  is a generic real number, generates translations in the coordinate space; for an infinitesimal  $a$  we have,

$$\hat{x}e^{\hat{p}a}|x\rangle = \hat{x}(1 + \hat{p}a)|x\rangle = i(x+a)e^{\hat{p}a}|x\rangle, \quad (4.46)$$

where, in the second step, we have used the canonical commutators in (4.28). This means

$$e^{\hat{p}a}|x\rangle = |x+a\rangle \quad (4.47)$$

(a possible overall factor  $k(a, x)$  in front of  $|x+a\rangle$  can be set to one by a suitable definition of  $\hat{p}$ ). From here we can construct the entire spectrum of  $\hat{p}$ . By applying  $e^{\hat{p}a}$  on  $\int dx |x\rangle$ , one discovers that this is an eigenstate with zero momentum, and by applying  $e^{-p\hat{x}}$  on it, where  $p$  is a generic real number, one generates all possible eigenstates  $|p\rangle$ :

$$\begin{aligned} |p\rangle &= \frac{1}{\sqrt{2\pi}} \int dx e^{-p\hat{x}} |x\rangle = \frac{1}{\sqrt{2\pi}} \int dx e^{-ipx} |x\rangle, \\ \Longleftrightarrow \quad \langle x|p\rangle &= \frac{1}{\sqrt{2\pi}} e^{ipx} \end{aligned} \quad (4.48)$$

where the factor  $1/\sqrt{2\pi}$  has been introduced to ensure the normalization condition

$$\langle p'|p\rangle = \delta(p+p'), \quad (4.49)$$

which, once again, leads to the completeness relation  $\int |p\rangle \langle p| \eta = 1$ . The states  $|p\rangle$  satisfy

$$\hat{p}|p\rangle = ip|p\rangle. \quad (4.50)$$

There are no other eigenstates as  $i\hat{p}$  is self-adjoint with respect to the positively defined metric  $\langle \cdot | \cdot \rangle_\eta \equiv \langle \cdot | \eta | \cdot \rangle$  and, therefore,  $\hat{p}$  can only have purely imaginary eigenvalues.

The Dirac-Pauli quantization may look strange at first sight, but it can be seen as a complex canonical transformation performed on variables quantized in the ordinary way:  $x \rightarrow ix$ ,  $p \rightarrow -ip$ .

In Table 2, the basic properties of a Dirac-Pauli variable are summarized.

<sup>20</sup>We require the completeness of the states  $|x\rangle$  as part of the definition of the vector space.

#### 4.2.4. Making the Wave Functions Normalizable

Let us now come back to our original problem, the non-normalizability of the wave functions. For the sake of definiteness, we again consider the Pais-Uhlenbeck model with  $\omega_1 \neq \omega_2$  and assume that  $q_2$  is a Dirac-Pauli variable, whereas  $q_1$  is an ordinary one. Then we obtain

$$\psi_0(q_1, q_2) \propto \exp\left(\frac{-q_1^2 \omega_1 \omega_2 - q_2^2}{2}(\omega_1 + \omega_2) + q_1 q_2 \omega_1 \omega_2\right), \quad (4.51)$$

which is now normalizable:

$$\begin{aligned} \langle 0|0\rangle &= \int dq_1 dq_2 \langle 0|q_1, -q_2\rangle \langle q_1, q_2|0\rangle \\ &= \int dq_1 dq_2 \psi_0(q_1, -q_2)^* \psi_0(q_1, q_2) < \infty, \end{aligned} \quad (4.52)$$

where we have used the decomposition of the identity in terms of eigenstates of the coordinate operators and we have taken into account Equation (4.45) for the Dirac-Pauli variable  $q_2$ . Moreover, recall that we have earlier required  $\langle 0|0\rangle$  to be positive; we fix  $\langle 0|0\rangle = 1$  by appropriately choosing the normalization constant. Then, by using (4.33), one can easily show that the state  $|n_1, n_2\rangle$ , where  $n_{1,2}$  are the occupation numbers of  $\tilde{a}_{1,2}$ , has norm  $(-1)^{n_1}$ . So, not only the ground state but also all the excited states are normalizable with this quantization.

At this point it is good to mention that Hawking and Hertog [61] proposed a way to deal with four-derivative degrees of freedom, but they ended up with non-normalizable wave functions. They then suggested solving the problem by integrating out  $\dot{q}$ . As we have seen, this issue does not arise if the appropriate quantization described above is performed (treating  $q$  as an ordinary variable and  $\dot{q}$  as a Dirac-Pauli one)

Other consistent quantizations are possible [62, 63]. For example, one could quantize  $\tilde{q}_1$  à la Dirac-Pauli, by treating  $\tilde{q}_2$  as an ordinary variable (the variables with a tilde have been defined in Equation 4.26). We will address this point after having introduced the path-integral formulation of the theory.

A Dirac-Pauli quantization for the ghost of QG has not been studied yet and is a very interesting topic for future research. By analogy, with the results obtained in the Pais-Uhlenbeck model, one expects normalizable wave functions in the QG case too.

#### 4.2.5. Path-Integral Formulation

We now present the path-integral formulation of a theory with an arbitrary number of ordinary canonical variables  $q_1, \dots, q_n$  and Dirac-Pauli variables  $\tilde{q}_1, \dots, \tilde{q}_m$  [58, 64]. A state with definite canonical coordinates is denoted here as,

$$|q\rangle = |q_1, \dots, q_n, \tilde{q}_1, \dots, \tilde{q}_m\rangle. \quad (4.53)$$

We are interested in understanding whether the quantization presented above is consistent in the presence of interactions. Even in ordinary quantum theories the real-time path integral is only a formal object, whose consistency at the rigorous level is unclear. For this reason, we consider the imaginary-time path integral (what would be called the Euclidean path integral in a QFT).

**TABLE 2 |** Basic properties of a Dirac-Pauli variable (and its conjugate momentum) compared to the ordinary case.

Canonical variable	$\hat{x}$ on states	$\hat{p}$ on states	$\hat{x}$ on functions	$\hat{p}$ on functions
Dirac-Pauli variable	$\hat{x} \chi\rangle = i\chi \chi\rangle$	$\hat{p} \rho\rangle = i\rho \rho\rangle$	$\langle\chi \hat{x} \psi\rangle = -i\chi\langle\chi \psi\rangle$	$\langle\chi \hat{p} \psi\rangle = \frac{d}{d\chi}\langle\chi \psi\rangle$
	$\hat{x} \rho\rangle = -\frac{d}{d\rho} \rho\rangle$	$\hat{p} \chi\rangle = \frac{d}{d\chi} \chi\rangle$	$\langle\rho \hat{x} \psi\rangle = -\frac{d}{d\rho}\langle\rho \psi\rangle$	$\langle\rho \hat{p} \psi\rangle = -i\rho\langle\rho \psi\rangle$
Ordinary variable	$\hat{x} \chi\rangle = \chi \chi\rangle$	$\hat{p} \rho\rangle = \rho \rho\rangle$	$\langle\chi \hat{x} \psi\rangle = \chi\langle\chi \psi\rangle$	$\langle\chi \hat{p} \psi\rangle = -i\frac{d}{d\chi}\langle\chi \psi\rangle$
	$\hat{x} \rho\rangle = -i\frac{d}{d\rho} \rho\rangle$	$\hat{p} \chi\rangle = i\frac{d}{d\chi} \chi\rangle$	$\langle\rho \hat{x} \psi\rangle = i\frac{d}{d\rho}\langle\rho \psi\rangle$	$\langle\rho \hat{p} \psi\rangle = \rho\langle\rho \psi\rangle$

These properties are derived in the text or are simple extensions of the properties derived in the text.

In formulating a quantum theory with the path integral, one notices that the full information on the dynamics of the system is encoded in the object  $\langle q_f | \exp(-iHt) | q_i \rangle$ , where  $|q_i\rangle$  and  $|q_f\rangle$  are generic states with definite coordinates. Indeed, once this object is known we can determine how the wave function evolves in time. In the presence of some Dirac-Pauli variables, one can do something similar, but one inserts an operator  $\eta$  defined by

$$\eta |q_1, \dots, q_n, \bar{q}_1, \dots, \bar{q}_m\rangle \equiv |q_1, \dots, q_n, -\bar{q}_1, \dots, -\bar{q}_m\rangle. \quad (4.54)$$

Namely, instead of considering  $\langle q_f | \exp(-iHt) | q_i \rangle$ , one tries to evaluate  $\langle q_f | \eta \exp(-iHt) | q_i \rangle$ . This is convenient for reasons that will become apparent soon, but note that  $\langle q_f | \eta \exp(-iHt) | q_i \rangle$  encodes the full dynamical information just like  $\langle q_f | \exp(-iHt) | q_i \rangle$  as they both give the matrix elements of the time-evolution operators with respect to a complete basis.

Working with an imaginary time  $t \rightarrow -i\tau$ , one is thus interested in computing the matrix element  $\langle q_f | \eta \exp(-H\Delta\tau) | q_i \rangle$ , where  $\Delta\tau$  is some imaginary-time interval. This, as usual, can be done by decomposing  $\Delta\tau$  in the sum of a very large number  $N$  of very small intervals  $d\tau$ , i.e.,  $d\tau \equiv \Delta\tau/N$ . By writing  $\exp(-H\Delta\tau) = \prod_{j=1}^N \exp(-Hd\tau)$  and inserting  $N-1$  times the identity  $\int dq |q\rangle \langle q| \eta = 1$ , one ends up with

$$\langle q_f | \eta e^{-H\Delta\tau} | q_i \rangle = \int \prod_{j=1}^N \langle q_j | \eta e^{-Hd\tau} | q_{j-1} \rangle \prod_{k=1}^{N-1} dq_k, \quad (4.55)$$

where  $q_N \equiv q_f$  and  $q_1 \equiv q_0$ . To evaluate  $\langle q_j | \eta \exp(-Hd\tau) | q_{j-1} \rangle$ , we insert the identity in the form  $\int dp_{j-1} \eta |p_{j-1}\rangle \langle p_{j-1}| = 1$ :

$$\begin{aligned} \langle q_j | \eta \exp(-Hd\tau) | q_{j-1} \rangle &= \int dp_{j-1} \langle q_j | p_{j-1} \rangle \langle p_{j-1} | e^{-Hd\tau} | q_{j-1} \rangle \\ &= \int \frac{dp_{j-1}}{2\pi} e^{ip_{j-1}(q_j - q_{j-1}) - \bar{H}(q_{j-1}, p_{j-1})d\tau}, \end{aligned} \quad (4.56)$$

where we have used Equation (4.48) and defined

$$\bar{H}(q, p) \equiv \frac{\langle p | H | q \rangle}{\langle p | q \rangle}. \quad (4.57)$$

Here we use a compact notation where the indices and sums over the various degrees of  $q_1, \dots, q_n$  and  $\bar{q}_1, \dots, \bar{q}_m$  are understood.

By letting  $N \rightarrow \infty$ , one, thus, obtains the imaginary-time path integral

$$\langle q_f | \eta e^{-H\Delta\tau} | q_i \rangle = \int \delta q \delta p e^{\int d\tau (ipq' - \bar{H}(q, p))} \quad \text{where} \quad \delta q \delta p = \frac{dp_0}{2\pi} \lim_{N \rightarrow \infty} \prod_{j=1}^{N-1} \frac{dq_j dp_j}{2\pi}, \quad (4.58)$$

a prime denotes a derivative w.r.t.  $\tau$ , the integral over  $\tau$  is from an initial time  $\tau_i$  and a final time  $\tau_f$ , such that  $\Delta\tau = \tau_f - \tau_i$  and it is understood that the integral over  $\delta q$  is performed only over those configurations that satisfy  $q(\tau_i) = q_i$  and  $q(\tau_f) = q_f$ .

We see that, modulo the usual subtleties related to the integration over an infinite-dimensional functional space that are present in any quantum theory, the only requirement for the existence of the path integral is that the real part of  $\bar{H}(q, p)$  (not<sup>21</sup> the classical Hamiltonian  $H(q, p)$ ) be bounded from below and that  $\bar{H}(q, p)$  diverge fast enough when the canonical coordinates tend to infinity (so that the integrations over  $q$  and  $p$  converge).

These conditions are satisfied in the Pais-Uhlenbeck model where  $q_1$  is quantized in the ordinary way and  $q_2$  is quantized à la Dirac-Pauli, at least when the interaction term  $V$  is bounded from below<sup>22</sup> (the usual condition). Indeed, from the Hamiltonian (4.22) it follows

$$\bar{H}(q, p) = ip_1 q_2 + \frac{1}{2} p_2^2 + \frac{\omega_1^2 + \omega_2^2}{2} q_2^2 + \frac{\omega_1^2 \omega_2^2}{2} q_1^2 + V(q_1), \quad (4.59)$$

which has the required properties. For the Pais-Uhlenbeck model, the Euclidean path integral is

$$\langle q_f | \eta e^{-H\Delta\tau} | q_i \rangle = \int \delta q_1 \delta q_2 \delta p_1 \delta p_2 \exp \left[ \int d\tau (ip_1 q_1' + ip_2 q_2' - \bar{H}(q, p)) \right]. \quad (4.60)$$

This expression can be further simplified since some integrations can be explicitly performed. Given the first term in (4.59), the  $\delta p_1$  integral gives  $\delta(q_2 - q_1')$ , such that the  $\delta q_2$  path integral

<sup>21</sup>In ordinary quantum theories  $\bar{H}(q, p) = H(q, p)$ , but in the presence of Dirac-Pauli variables this is not generically the case because of the extra  $i$  appearing in the eigenvalues of the Dirac-Pauli coordinates and momenta.

<sup>22</sup>If one introduces a more complicated interaction that depends on the other coordinate and momenta  $V(q, p)$ , the condition is that  $\text{Re} \bar{V}(q, p)$  be bounded from below.

just fixes  $q_2 = q'_1$ . Next, the remaining terms in  $\tilde{H}$  are a sum of positive squares and  $V(q_1)$  so all other integrals are convergent assuming that  $V$  is bounded from below. Performing the remaining integrals, one finds the Lagrangian Euclidean path integral:

$$\langle q_f | \eta e^{-H\Delta\tau} | q_i \rangle \propto \int \delta q \exp \left[ - \int d\tau L_E(q) \right], \quad (4.61)$$

where the classical Euclidean Lagrangian is

$$L_E = \frac{1}{2} \left( \frac{d^2 q}{d\tau^2} \right)^2 + \frac{\omega_1^2 + \omega_2^2}{2} \left( \frac{dq}{d\tau} \right)^2 + \frac{\omega_1^2 \omega_2^2}{2} q^2 + V(q). \quad (4.62)$$

The Lagrangian path integral appears to be well-defined as  $L_E$  is bounded from below.

The expression in (4.62) also allows us to study the classical limit. Going back to real time one obtains precisely the Lagrangian we started from, Equation (4.17). As discussed in section 4.1.2, for some interactions  $V(q)$  (bounded from below and above) there are stable solutions. In a generic theory, one expects that the requirement of having stable solutions place stringent conditions on the possible interactions, which so far have not been fully classified. The path integral formulation tells us that, in the classical limit, the dynamics is dominated by the solution(s) with least Euclidean action. In the Pais-Uhlenbeck case, these correspond to time-independent solutions that minimize the full potential  $\frac{\omega_1^2 \omega_2^2}{2} q^2 + V(q)$ . All unbounded solutions, if any, should be negligible in the classical limit as the derivative terms always contribute positively to the Lagrangian in (4.62). As usual, perturbations around a given solution should be computed through the path integral and, given that the path integral appears to be well-defined, no pathologies are expected. Therefore, it is possible that the Dirac-Pauli quantization could solve the potential problems raised by the Ostrogradsky theorem.

The path integral (4.61) makes it clear that, if  $V(q)$  is chosen to be non-negative everywhere, no negative energies can be present. If they did, then we should observe a divergence of  $\langle q_f | \eta \exp(-H\Delta\tau) | q_i \rangle$  as  $\Delta\tau \rightarrow \infty$ , but the right-hand side of (4.61) does not diverge in that limit as the Lagrangian is a sum of positive terms.

Another issue is that in a theory where the Hamiltonian  $H$  is self-adjoint with respect to an indefinite norm (and nothing else is known) there is no theorem guaranteeing the reality of the energy spectrum. However, it is still possible that the spectrum is real, as we have seen in the case of the unequal-frequency Pais-Uhlenbeck model in section 4.2.1. Even if one introduces a non-trivial interaction term  $V \neq 0$  in the Pais-Uhlenbeck model with generic unequal frequencies, no complex energies can appear as long as  $V$  is small enough that perturbation theory can be trusted. Indeed, a complex energy would require a zero-norm state, but only positive and negative norm eigenstates of  $H$  with no degeneracies are found in section 4.2.1. In a theory where some of the eigenvalues of  $H$  turn out to be complex, one should find a sensible interpretation for them. A possible interpretation could be that those states are unstable and some of them (the ones with eigenvalues with positive imaginary parts) lead to a violation

of causality<sup>23</sup> [66, 67]. However, in Sotiriou and Faraoni [21] it was pointed out that there are some conditions to be fulfilled in order for this violation of causality to be observable and it is easy to engineer a model where these conditions are not met.

Let us come back to the path integral. What would have happened if we had used a different quantization? One could have quantized  $\tilde{q}_1$  à la Dirac-Pauli and  $\tilde{q}_2$  as an ordinary variable (the variables with a tilde have been defined in Equation (4.26) when  $\omega_1 \neq \omega_2$ ). Then, one would have obtained

$$\langle \tilde{q}_f | \eta e^{-H\Delta\tau} | \tilde{q}_i \rangle = \int \delta \tilde{q}_1 \delta \tilde{q}_2 \delta \tilde{p}_1 \delta \tilde{p}_2 \exp \left[ \int d\tau (i\tilde{p}_1 \dot{\tilde{q}}_1 + i\tilde{p}_2 \dot{\tilde{q}}_2 - \tilde{H}(\tilde{q}, \tilde{p})) \right], \quad (4.63)$$

where

$$\tilde{H}(\tilde{q}, \tilde{p}) = \frac{1}{2} (\tilde{p}_1^2 + \omega_1^2 \tilde{q}_1^2) + \frac{1}{2} (\tilde{p}_2^2 + \omega_2^2 \tilde{q}_2^2) + \tilde{V}(\tilde{q}_2, \tilde{p}_1) \quad (4.64)$$

and, according to Equation (4.25),

$$\tilde{V}(\tilde{q}_2, \tilde{p}_1) = V \left( \frac{\tilde{q}_2 - i\tilde{p}_1/\omega_1}{\sqrt{\omega_1^2 - \omega_2^2}} \right). \quad (4.65)$$

Given that  $V$  is computed in the complex quantity  $(\tilde{q}_2 - i\tilde{p}_1/\omega_1)/\sqrt{\omega_1^2 - \omega_2^2}$ , the requirement that  $\text{Re}\tilde{H}(\tilde{q}, \tilde{p})$  is bounded from below leads to very peculiar conditions on the function  $V$ , which seems very hard to be fulfilled for reasonable  $V$ , and thus very hard to be kept in generalizing these results to QG. Therefore, while other quantizations could still be consistent, dedicated studies of these alternative path-integral quantizations in the presence of interactions are not known.

The computation of the Lagrangian path integral has been carried out here within the Pais-Uhlenbeck model. We have used explicitly that some variables are quantized à la Dirac-Pauli. If a Dirac-Pauli quantization for QG will be provided, then one could also perform the same calculation in QG. One expects that the Lagrangian path-integral for QG is consistent if the classical Euclidean Lagrangian is bounded from below, which is the case for some choices of the parameters, but there is no substitute of a complete calculation to reach this conclusion. Such calculation would also provide a non-perturbative definition of quantum QG.

#### 4.2.6. Probabilities

We now turn to the possible definitions of probabilities in the presence of ghosts. We have learned in Sections 3.1 and 4.2.1 that both the renormalizability of QG and the requirement that the quantum Hamiltonian must be bounded from below lead to the presence of an indefinite metric. This raises problems in defining the probability that a certain event occurs. In quantum mechanics, the possible outcomes of the measurement of an observable  $A$  (a self-adjoint operator,  $A^\dagger = A$ ) are in one-to-one

<sup>23</sup>Nevertheless the commutators between any two field operators at points separated by a spacelike distance are zero [65], like in usual QFT. In QG, this property can be easily proved by using the expansion of the free ghost field in creation and annihilation operators introduced as in section 4.2.1 and then by applying the unitary operator that transforms the free ghost field in the interacting one.

correspondence with the eigenstates  $|a\rangle$  of  $A$  with probabilities given by the Born rule

$$P(\psi \rightarrow a) = \frac{|\langle a|\psi\rangle|^2}{\langle a|a\rangle\langle\psi|\psi\rangle}, \quad (4.66)$$

where  $|\psi\rangle$  is the state of the system before the measurement. If some of the states have negative norms, then the direct application of the Born rule in the presence of ghosts leads to some negative probabilities.

Since  $P(\psi \rightarrow a)$  can be negative only when the denominator  $\langle a|a\rangle\langle\psi|\psi\rangle$  is negative, a first idea could be to substitute (4.66) with the following modified Born rule:

$$P(\psi \rightarrow a) = \frac{|\langle a|\psi\rangle|^2}{|\langle a|a\rangle\langle\psi|\psi\rangle|}, \quad (4.67)$$

However, (4.67) does not generically satisfy another basic requirement, that the sum of  $P(\psi \rightarrow a)$  over all possible eigenvalues  $a$  is 1. This is because

$$\sum_a \frac{|\langle a|\psi\rangle|^2}{|\langle a|a\rangle\langle\psi|\psi\rangle|} = \sum_a \frac{\langle\psi|a\rangle\langle a|\psi\rangle}{|\langle a|a\rangle\langle\psi|\psi\rangle|} \quad (4.68)$$

and here generically we have

$$\sum_a \frac{|a\rangle\langle a|}{|\langle a|a\rangle|} \neq 1. \quad (4.69)$$

Indeed, if we assume the eigenstates  $|a\rangle$  to form a complete basis and decompose an arbitrary state  $|\alpha\rangle$  as  $|\alpha\rangle = \sum_{a'} \alpha_{a'} |a'\rangle$ , where  $\alpha_{a'}$  are complex numbers, we have

$$\sum_a \frac{|a\rangle\langle a|\psi\rangle}{|\langle a|a\rangle|} = \sum_{aa'} \frac{\alpha_{a'} |a\rangle\langle a|a'\rangle}{|\langle a|a\rangle|} \quad (4.70)$$

and in general  $\langle a|a'\rangle/|\langle a|a\rangle|$  is not equal to  $\delta_{aa'}$  because some of the states can have negative norm. This is what some people call the “unitarity problem” (we do not use this terminology here as the time evolution operator is unitary w.r.t. indefinite norm).

We now discuss the most popular ways to address this problem.

### Lee-Wick Idea

Lee and Wick [68] proposed that a theory with an indefinite metric can still have a unitary  $S$ -matrix provided that all stable states have positive norm. Since the  $S$ -matrix connects only asymptotic states that, by definition, are stable, one expects that under this hypothesis the transition probabilities between asymptotic states are positive and add up to one. The Lee-Wick idea has been studied in the context of QG in a number of papers [39, 69–76].

To understand this idea more in detail, let us denote with  $|\sigma\rangle$  and  $|\sigma'\rangle$  two generic stable states and consider the  $S$ -matrix elements

$$S_{\sigma'\sigma} \equiv \langle\sigma'|S|\sigma\rangle, \quad (4.71)$$

where we have normalized  $|\sigma\rangle$  and  $|\sigma'\rangle$  to 1 (the Lee-Wick hypothesis implies that the norm of stable states are positive and therefore can be normalized to 1). The operator  $S \equiv \lim_{\Delta t \rightarrow \infty} U(\Delta t)$  is unitary with respect to the indefinite norm by construction, but we are interested in proving the unitarity of the  $S$ -matrix in 4.71 because this is what would allow us to claim that the probabilities add up to one: indeed, using the standard Born rule (4.66) leads to

$$\sum_{\sigma'} P(\sigma \rightarrow \sigma') = \sum_{\sigma'} |\langle\sigma'|S|\sigma\rangle|^2 = \sum_{\sigma'} S_{\sigma'\sigma}^* S_{\sigma'\sigma}. \quad (4.72)$$

Now, one can rewrite

$$\sum_{\sigma'} |\langle\sigma'|S|\sigma\rangle|^2 = \sum_{\sigma'} \langle\sigma|S^\dagger|\sigma'\rangle\langle\sigma'|S|\sigma\rangle \quad (4.73)$$

and this expression would be equal to 1 in two cases:

1. if  $\sum_{\sigma'} |\sigma'\rangle\langle\sigma'| = 1$  or, more generally,
2. if  $S|\sigma\rangle$  can be written as a linear combination of the stable states only.

The first condition cannot be true because we know there are negative norm states, which can never be written as linear combinations of positive norm states only; indeed, in the presence of negative norm states  $\sum_{\sigma'} |\sigma'\rangle\langle\sigma'| = 1$  is replaced by

$$\sum_{\sigma'} |\sigma'\rangle\langle\sigma'| = 1 - \Pi_-, \quad (4.74)$$

where  $\Pi_-$  is the projector on the negative-norm subspace. So, one has to assume Condition 2, which, although plausible (as one expects  $S$  to connect stable states with stable states only), *has to be proved*. To see when the important probabilistic condition  $\sum_{\sigma'} |\langle\sigma'|S|\sigma\rangle|^2 = 1$  is satisfied, it is convenient to rewrite it in a form that can be more easily verified by an explicit calculation. To do so, we note that

$$\sum_{\sigma'} \langle\sigma|S^\dagger|\sigma'\rangle\langle\sigma'|S|\sigma\rangle = 1 - \langle\sigma|S^\dagger\Pi_-S|\sigma\rangle, \quad (4.75)$$

where we have used Equation (4.74). By writing as usual  $S \equiv 1 + iT$ , one has

$$\langle\sigma|S^\dagger\Pi_-S|\sigma\rangle = \langle\sigma|T^\dagger\Pi_-T|\sigma\rangle, \quad (4.76)$$

which follows from  $\Pi_-|\sigma\rangle = 0$ . The unitarity of  $S$  implies  $i(T^\dagger - T) = T^\dagger T$  and, by taking the diagonal matrix element  $T_{\sigma\sigma} \equiv \langle\sigma|T|\sigma\rangle$  and using once again Equation (4.74),

$$2\text{Im}T_{\sigma\sigma} = \sum_{\sigma'} \langle\sigma|T^\dagger|\sigma'\rangle\langle\sigma'|T|\sigma\rangle + \langle\sigma|T^\dagger\Pi_-T|\sigma\rangle \quad (4.77)$$

Given that  $\Pi_-$  can be written as  $\sum_g |g\rangle\langle g|$  where  $|g\rangle$  represents a complete basis on the negative-norm subspace, we see that the condition that the probabilities sum up to one is equivalent to the condition that the ghost states  $|g\rangle$  do not contribute to the imaginary part of the forward scattering amplitude, represented

here by  $T_{\sigma\sigma}$ . Anselmi [75] has recently found that this condition is satisfied if one modifies appropriately the prescription to determine the ghost propagator<sup>24</sup>.

One issue is that, in order to claim that the negative norm states are unstable, which is a basic assumption of the Lee-Wick proposal, one needs a consistent way of computing the probability of ghost decays; otherwise how do we tell if the ghost is unstable or not? Since there is one ghost field in QG, the use of the standard Born rule (4.66) to compute this probability leads to a negative number. This is not necessarily a non-sense as Lee and Wick proposed to consider as physical states only the asymptotic ones and regard the ghost just as a virtual state, which is not directly observable. In this case, it might be consistent to assign negative probabilities to such somewhat unobservable events, as pointed out by Feynman [79].

However, one can also argue that the Lee-Wick proposal might not address all potential problems because scattering theory (described by the S-matrix) is not the only application of quantum mechanics.

### Defining Positive Norms

Although renormalizability and the existence of a state of minimum energy lead to an indefinite metric, one can still try to define positively defined metrics with the desired property: positive probabilities that add up to one when used in the Born rule. This possibility was studied in a number of articles [58, 63, 80–85].

Let us consider an example of a positively defined metric. The path-integral formula (4.58) suggests to consider the  $\eta$ -metric  $\langle \cdot | \cdot \rangle_\eta \equiv \langle \cdot | \eta | \cdot \rangle$ , where  $\eta$  is defined for a generic theory in Equation (4.54). This metric is positively defined because

$$\langle q'_1, \dots, q'_n, \bar{q}'_1, \dots, \bar{q}'_m | \eta | q_1, \dots, q_n, \bar{q}_1, \dots, \bar{q}_m \rangle = \prod_{j=1}^n \delta(q_j - q'_j) \prod_{k=1}^m \delta(\bar{q}_k - \bar{q}'_k) \quad (4.78)$$

and  $|q_1, \dots, q_n, \bar{q}_1, \dots, \bar{q}_m\rangle$  is complete. In (4.78), we used (4.44) for the Dirac-Pauli variables  $\bar{q}_1, \dots, \bar{q}_m$  and the usual normalization  $\langle q_j | q'_j \rangle = \delta(q_j - q'_j)$  for the ordinary variables  $q_1, \dots, q_n$ . The  $\eta$ -metric can be used to compute the probabilities of measuring  $q_1, \dots, q_n, \bar{q}_1, \dots, \bar{q}_m$  and the corresponding conjugate momenta (in the case of Dirac-Pauli variables, the outcomes of an experiment can be identified with the imaginary parts of the eigenvalues). Below we will show that the probabilities add up to one.

Before doing so, we generalize this approach to other observables. First, we have to clarify the meaning of “observables” in this context. An observable  $A$  is represented by an operator with a complete set of eigenstates,  $|a\rangle$ . Indeed, in this case we can define a positively defined metric in the following way. Let us define an operator  $P_A$  through<sup>25</sup>

$$\langle a' | P_A | a \rangle \equiv \delta_{aa'}. \quad (4.79)$$

Note that  $P_A$  satisfies  $P_A^\dagger = P_A$  and depends in general on  $A$ . The new positively defined metric is defined by

$$\langle \psi_2 | \psi_1 \rangle_A \equiv \langle \psi_2 | P_A | \psi_1 \rangle, \quad (4.80)$$

where  $|\psi_{1,2}\rangle$  are generic states. By using this new metric, one can define the probabilities with the usual Born rule: the probability that the outcome of an experiment will measure  $a$  for an observable  $A$  given that the state before the measurement is  $|\psi\rangle$  is given by

$$P(\psi \rightarrow a) \equiv \frac{|\langle a | \psi \rangle_A|^2}{\langle a | a \rangle_A \langle \psi | \psi \rangle_A}. \quad (4.81)$$

These probabilities indeed satisfy the basic properties—they are positive and they add up to one,

$$\sum_a P(\psi \rightarrow a) = \sum_a \frac{\langle \psi | a \rangle_A \langle a | \psi \rangle_A}{\langle a | a \rangle_A \langle \psi | \psi \rangle_A} = \frac{\langle \psi | P_A | \psi \rangle}{\sqrt{\langle \psi | \psi \rangle_A}} \left( \sum_a \frac{|a\rangle \langle a| P_A}{\langle a | P_A | a \rangle} \right) \frac{|\psi\rangle}{\sqrt{\langle \psi | \psi \rangle_A}} = 1, \quad (4.82)$$

where we used

$$\sum_a \frac{|a\rangle \langle a| P_A}{\langle a | P_A | a \rangle} = 1, \quad (4.83)$$

which follows from the completeness of  $\{|a\rangle\}$  and the defining property of  $P_A$ , Equation (4.79). Note that this result also holds for time-dependent  $|\psi\rangle$  and, therefore, probability is conserved under time evolution. In the specific case when  $\langle a | a \rangle$  is either positive or negative (it never vanishes), an explicit expression for  $P_A$  is (after having normalized the state in a way that  $\langle a | a \rangle = \pm 1$ )

$$P_A \equiv \Pi_+^A - \Pi_-^A, \quad (4.84)$$

where  $\Pi_+^A$  and  $\Pi_-^A$  are the projectors on the positive norm and negative norm eigenstates of  $A$ , respectively.

### 4.3. Cosmology

In practice, the cosmological predictions of QG would be basically those of a standard QFT coupled to Einstein gravity if it were not for the  $W^2$  term. This term, as we have seen, corresponds to a spin-2 ghost with mass  $M_2 = f_2 \bar{M}_P / \sqrt{2}$ . Therefore, unless one takes  $f_2$  really tiny, the only significant effects of the ghost occur in an inflationary context. We will focus then on the inflationary behavior of the theory here.

The first step in studying the cosmological applications of the theory is to find an FRW metric that satisfies the classical equations. From the experience gained with the Pais-Uhlenbeck model in section 4.2.5, one expects that the classical limit provides precisely the classical action we started from, Equations (2.8), (2.10), and (2.12). This is what is assumed basically in the entire literature on the subject. The actual proof of this property would be a significant progress in the understanding of QG.

The FRW metric is

$$ds^2 = dt^2 - a(t)^2 \delta_{ij} dx^i dx^j, \quad (4.85)$$

<sup>24</sup>See also Abe et al. [77] and Donoghue and Menezes [78] for other discussions about unitarity.

<sup>25</sup>This defines  $P_A$  because an operator is defined once we give all matrix elements in a complete basis.

where  $a$  is the scale factor and we have neglected the spatial curvature parameter as during inflation the energy density is dominated by the scalar fields. The metric in (4.85) leads to standard Friedmann equations as the  $W^2$  term vanishes on conformally flat metrics and does not contribute to the equations of motion. When the hypothesis of homogeneity and isotropy is relaxed the  $W^2$  term contributes instead and its effect has been studied in a number of works [21, 47, 86–94] (see [21] for a general treatment), where the perturbations around the FRW metrics were considered. We do not reproduce the calculations here as they are performed in detail in the original articles. One of the most important results obtained so far is that all perturbations found by solving the linear equations around the FRW metric remain bounded as time passes by Peter et al. [19], Salvio [21], Ivanov and Tokareva [47], Tokareva [48], and Salles and Shapiro [95], contrary to what one would naively expect from the Ostrogradsky theorem. Moreover, by quantizing these linear perturbations with an indefinite metric (with an appropriate generalization of section 4.2.1) one obtains that the conserved Hamiltonian of the full system is bounded from below [21]. What happens beyond the linear order, however, has not been discussed in detail and is an important target for future research.

In QG, there are several possible inflaton candidates. First, QG gives a natural implementation of Starobinsky's inflationary model [4] as the  $R^2$  is mandatory in order to have renormalizability. Furthermore, other possible scalar fields can participate: at the very least the theory should contain the Higgs boson, which has been discovered at the Large Hadron Collider. A detailed analysis of the inflationary dynamics and observable predictions in some specific realizations of the QG scenario is provided in Kannike et al. [20], Salvio [21], and Salvio and Strumia [33].

#### 4.4. Black Holes

After the discovery of gravitational waves interpreted as the product of a binary black hole merger [96], the interest in black hole solutions have increased. Therefore, it is important to study the existence and properties of static spherically symmetric solutions in QG, where the metric is given in spherical coordinates  $\{r, \theta, \varphi\}$  by two functions  $f_1$  and  $f_2$  of  $r$ :

$$ds^2 = f_1(r)dt^2 - \frac{dr^2}{f_2(r)} - r^2(d\theta^2 + \sin^2\theta d\varphi^2). \quad (4.86)$$

This has been initiated in a number of articles. The first work was done by Stelle [12], who computed the correction to Newton's law due to the extra gravitational terms. A first, the observation is that the Schwarzschild solution of Einstein gravity in the vacuum ( $f_1(r) = f_2(r)$ ) is also a solution of the vacuum equations of QG (i.e., in the absence of matter) [12, 97, 98]. Also, Lu et al. [98], Holdom [99], Lü et al. [100, 103], Cai et al. [101], Lin et al. [102], Goldstein and Mashiyane [104], Kokkotas et al. [105], and Stelle [106] found numerically and studied new black hole solutions (not present in Einstein gravity) and Holdom and Ren [107] identified a new class of static

spherically symmetric solutions without horizon (called the 2-2-hole), which can, nevertheless, mimic the Schwarzschild solution outside the horizon, with interesting implications for the black hole information paradox.

Keeping in mind the Ostrogradsky theorem, an important question is whether a stable black hole (or pseudo black hole, such as the 2-2-hole) exists in the theory. Lü et al. [103] pointed out that the Schwarzschild solution is stable for large horizon radius  $r_h$ , but becomes unstable (see also [108]) when  $r_h$  is taken below a critical value set basically by the inverse ghost mass  $\sim 1/M_2$  (see also [106]). The endpoint of the instability is conjectured to be another black hole solution, which is not present in Einstein gravity and may be stable when  $r_h$  is small. Holdom and Ren [107] considered the creation of a static spherically symmetric solution generated by a thin spherically symmetric shell of matter; when the shell radius  $l \lesssim r_h$  the new 2-2-hole is found.

Once again in all these works the classical equations (valid as  $\hbar \rightarrow 0$ ) of QG are taken to be those generated by the starting action in (2.8), which is what we expect, but as pointed out in section 4.3, a proof is still missing in the literature.

#### 5. REACHING INFINITE ENERGY

Given that QG (coupled to a general renormalizable matter sector) is renormalizable, one can hope that the theory remains valid up to infinite energy. However, soon after the calculation of the gravitational  $\beta$ -functions of Avramidi and Barvinsky [13] it was realized to be a major obstacle to UV-completeness: the  $\beta$ -function of  $f_0^2$  in (3.16) is not negative for  $f_0^2 > 0$  and, therefore, the theory features a growth of  $f_0$  as the energy increases, until perturbation theory in  $f_0$  cannot be trusted anymore<sup>26</sup>.

Then, a number of authors [15, 109–113] explored the case  $f_0^2 < 0$  claiming that asymptotic freedom can be achieved for all couplings (both the gravitational and matter couplings) if the matter sector is chosen appropriately. Although such programme can lead to mathematically consistent asymptotically free theories, there is a big phenomenological problem when one chooses  $f_0^2 < 0$ .

Let us consider for simplicity the case where the scalar  $\zeta$  (corresponding to the  $R^2$  term and introduced in section 2.2) does not mix with other scalars (if any). Then, the squared mass of  $\zeta$  equals  $M_0^2 = f_0^2 \bar{M}_P^2/2$  (see Table 1), which clearly indicates that for  $f_0^2 < 0$  the scalar  $\zeta$  is tachyonic. One way to obtain  $M_0^2 = f_0^2 \bar{M}_P^2/2$  is to use the Einstein frame Lagrangian in (2.17) and (2.18) and compute its quadratic approximation for the small fluctuations around the flat spacetime. Another way is to calculate (directly in the Jordan frame) the propagator of  $h_{\mu\nu} \equiv g_{\mu\nu} - \eta_{\mu\nu}$ , a procedure that was originally performed in Stelle [8], which obtained precisely the masses given in Table 1. This confirms that  $f_0^2 < 0$  leads to a tachyonic instability<sup>27</sup>. Yet

<sup>26</sup>Different statements in the literature (even recent) appear because some results for the  $\beta$  of  $f_0$  (obtained before the correct results of [13]) contained wrong signs.

<sup>27</sup>Similarly,  $f_2^2 < 0$  leads to a tachyonic instability in the ghost sector and, therefore, this case is commonly avoided as not even consistent with asymptotic freedom (for a discussion of the tachyonic case see, however, [115–117]).

another way to see why  $f_0^2 < 0$  is phenomenologically problematic is to look at the Newtonian potential  $V_N(r)$  due to the Lagrangian (2.8) [12, 114],

$$V_N(r) = -\frac{G_N M}{r} \left( 1 - \frac{4}{3} e^{-M_2 r} + \frac{1}{3} e^{-M_0 r} \right), \quad (5.1)$$

where  $G_N$  is Newton's gravitational constant and  $M$  is the mass of the point particle generating the potential. As noted even in the original article [12] by Stelle, this expression only gives an acceptable Newtonian limit for real  $M_2$  and  $M_0$  (i.e., for positive  $f_2^2$  and  $f_0^2$ ): otherwise one would obtain oscillating  $1/r$  terms.

One could hope that a phenomenologically viable  $f_0^2 < 0$  is achieved by introducing more scalars (besides  $\zeta$ ). However, a general argument, which we now describe, indicates that this is not the case. Consider the Einstein frame potential  $U$  (defined in Equation 2.18) along the  $\zeta$ -direction, which can be conveniently parameterized as

$$U = \frac{1}{\zeta^4} [a_1 + a_2(\zeta^2 - a_3)^2] \quad (5.2)$$

where  $a_1, a_3$  are suitable coefficients, which depend on the other scalar fields, whereas  $a_2 = 3f_0^2 \bar{M}_P^4/8 < 0$  (having assumed  $f_0^2 < 0$  here). A necessary condition for the existence of a minimum of  $U$  is that

$$\frac{\partial U}{\partial \zeta} = 0, \quad \text{that is} \quad \zeta^2 = \frac{a_1 + a_2 a_3^2}{a_2 a_3}. \quad (5.3)$$

Notice that, if the solution for  $\zeta^2$  exists, that is  $(a_1 + a_2 a_3^2)/a_2 a_3 > 0$ , then it is unique. Moreover, note that  $a_2 < 0$  implies that  $U$  goes to a negative value as  $\zeta \rightarrow \infty$ . Therefore, there are only three possibilities:

- There is no acceptable solution to (5.3) (no solution with  $\zeta^2 > 0$ ).
- The solution to (5.3) is a maximum of the potential (or at most a saddle point once the other scalars are included).
- The solution to (5.3) is a point of minimum of  $U$ , but occurs for a negative value of  $U$  (in contradiction with the positive value of the observed cosmological constant). Indeed, if it corresponded to a positive value of  $U$  then there would also be a maximum (or a saddle point) given that  $U$  goes to a negative value for  $\zeta \rightarrow \infty$  and this would contradict the uniqueness of the solution in (5.3).

The conclusion is that a minimum of  $U$  (if any) must have  $U < 0$ . This argument generalizes the situation illustrated in **Figure 1**, where only the field  $\zeta$  was considered.

## 5.1. Conformal Gravity as the Infinite Energy Limit of Quadratic Gravity

Given that the experiments lead us to take  $f_0^2 > 0$ , what happens when  $f_0$  grows and leaves the domain of validity of perturbation theory? In Salvio and Strumia [14] (see also references therein), by using a perturbative expansion in  $1/f_0$ , it was shown that, when  $f_0$  grows up to infinity in the infinite energy limit, the

scalar due to the  $R^2$  term decouples from the rest of the theory and  $f_0$  does not hit any Landau pole, provided that all scalars have asymptotically Weyl-invariant couplings (see below) and all other couplings approach fixed points. Then, QG can flow to a Weyl-invariant theory, a.k.a. conformal gravity, at infinite energy. Given the importance of Weyl invariance for the high-energy limit of QG, let us give some more details on this topic. A Weyl transformation acts as follows on the various fields (the metric  $g_{\mu\nu}$ , the scalars  $\phi_a$ , the fermions  $\psi_i$ , and the vectors  $V_\mu^A$ ):

$$\begin{aligned} g_{\mu\nu}(x) &\rightarrow e^{2\sigma(x)} g_{\mu\nu}(x), & \phi_a(x) &\rightarrow e^{-\sigma(x)} \phi_a(x), \\ \psi_i(x) &\rightarrow e^{-3\sigma(x)/2} \psi_i(x), & V_\mu^A &\rightarrow V_\mu^A, \end{aligned} \quad (5.4)$$

where  $\sigma$  is a generic function of  $x$ . A scalar has Weyl-invariant couplings when all dimensionful parameters vanish and  $\xi_{ab} = -\delta_{ab}/6$ . This precise value of  $\xi_{ab}$  emerges because in this case the non-invariance of the kinetic term of the  $\phi_a$  precisely cancels the non-invariance of the non-minimal couplings, Equation (2.12).

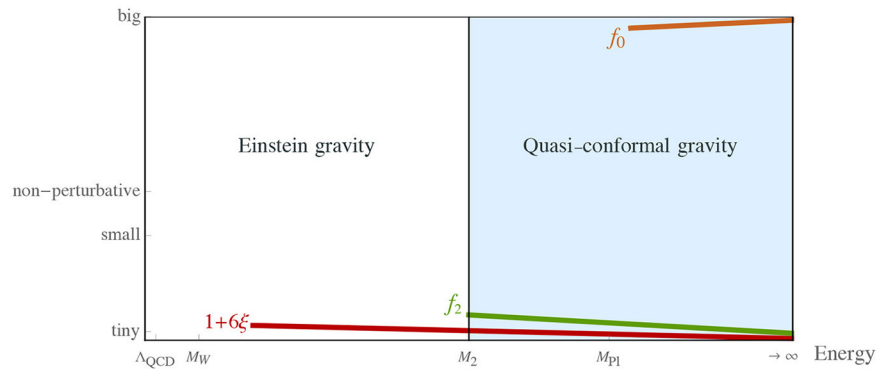
The idea that one can approach a Weyl-invariant theory at large energy has been investigated in a number of articles [78, 118–124]. We do not reproduce the proof of Salvio and Strumia [14] because it is described in detail there, but some remarks are in order regarding the implications of this result.

It is important to note that the condition to have a UV fixed point guarantees not only the UV-completeness of the QFT part<sup>28</sup> but also of the gravitational part of the theory (when all parameters flow to their conformal value). This opens the road to the construction and study of relativistic field theories of *all* interactions that are fundamental, i.e., hold up to infinite energy. This scenario leads to several extra fields (in addition to those present in the SM) as the study of the one-loop  $\beta$ -functions of the SM reveals the presence of Landau poles. These new fields can then be used to explain in an innovative way the current pieces of evidence for physics beyond the SM (such as neutrino oscillations, dark matter, and baryon asymmetry of the universe). This nearly unexplored field of research represents a very important target for future research.

## 5.2. RGEs for Conformal Gravity and Matter

Although flowing to conformal gravity at infinite energy can be consistent, at finite energy, conformal invariance is broken by the scale anomaly and the  $R^2$  term as well as a non-vanishing value of  $\delta_{ab} + 6\xi_{ab}$  are generated. However, this is a multiloop effect (see [14, 130–132] and references therein). The full set of one-loop

<sup>28</sup>Some SM extensions including gauge fields, fermions and scalars can feature a UV fixed point for all couplings and their corresponding phenomenology have been studied [125–129].



**FIGURE 3 |** Schematic behavior of the gravitational couplings as functions of the energy in a possible interesting scenario. At high energies the theory is approximately given by conformal gravity, with small corrections (which include the UV irrelevant Einstein-Hilbert and cosmological constant terms). Both  $1/f_0$  and  $\delta_{ab} + 6\xi_{ab}$  remain very small for the reasons given above. The coupling  $f_2$  associated with the  $W^2$  term is also chosen to be small both to maintain perturbativity and thus calculability and to provide interesting and potentially observable effects at the inflationary scales. The running of  $f_2$  is depicted only up to the mass of the corresponding degrees of freedom,  $M_2 = f_2 M_P / \sqrt{2}$ . A large coupling  $f_0$  influences physics only at energies much above the Planck mass as its role compared to the Einstein-Hilbert term is suppressed by  $E^2/(f_0^2 M_P^2)$ , where  $E$  is the typical energy of the process under study. Below  $M_2$  the gravitational theory resembles Einstein gravity plus small corrections. The energy flows from the scale below which strong interactions are non-perturbative,  $\Lambda_{QCD}$ , up to infinite energy (passing through the mass of the  $W$ -boson  $M_W$ , the ghost mass  $M_2$  and the Planck mass  $M_{Pl}$ ).

RGE in conformal gravity are given by,

$$\frac{df_2^2}{d\tau} = -f_2^4 \left( \frac{199}{15} + \frac{N_V}{5} + \frac{N_F}{20} + \frac{N_S}{60} \right) \quad (5.5)$$

$$\frac{dY^a}{d\tau} = \frac{1}{2} (Y^{\dagger b} Y^b Y^a + Y^a Y^{\dagger b} Y^b) + 2Y^b Y^{\dagger a} Y^b + Y^b \text{Tr}(Y^{\dagger b} Y^a) - 3\{C_{2F}, Y^a\} + \frac{15}{8} f_2^2 Y^a, \quad (5.6)$$

$$\begin{aligned} \frac{d\lambda_{abcd}}{d\tau} = & \sum_{\text{perms}} \left[ \frac{1}{8} \lambda_{abef} \lambda_{efcd} + \frac{3}{8} \{\theta^A, \theta^B\}_{ab} \{\theta^A, \theta^B\}_{cd} \right. \\ & \left. - \text{Tr} Y^a Y^{\dagger b} Y^c Y^{\dagger d} + \frac{5}{288} f_2^4 \delta_{ab} \delta_{cd} + \lambda_{abcd} \right. \\ & \left. \left[ \sum_{k=a,b,c,d} (Y_2^k - 3C_{2S}^k) + 5f_2^2 \right] \right] \quad (5.7) \end{aligned}$$

for  $f_0 \rightarrow \infty$  and  $\xi_{ab} \rightarrow -\frac{1}{6}\delta_{ab}$ . We do not show the RGE of the gauge couplings because they are not modified by the gravitational couplings (see [30–33]). The RGE of  $f_2$  was originally derived in Salvio and Strumia [14], Fradkin and Tseytlin [30], Shapiro and Zheksenaev [133], de Berredo-Peixoto and Shapiro [134], while those of  $Y^a$  and  $\lambda_{abcd}$  were obtained in Salvio and Strumia [14]. Also, Ohta and Percacci [135] checked the RGEs of  $f_2$  with functional renormalization group methods. This set of equations allows us to search for fundamental theories that enjoy total asymptotic freedom/safety: all couplings (including the gravitational ones) flow either to zero or to an interacting fixed point in the UV.

In **Figure 3**, a pictorial representation of a possible resulting gravitational scenario (described in the caption) is provided. That behavior suggests a new paradigm of inflation based on a quasi-conformal theory, a theory where  $f_0$  is large and  $\xi_{ab} \approx -\delta_{ab}/6$ , which so far has been left as a very interesting future development.

The general RGEs in (5.5)–(5.7) can be used to address high-energy issues in the scenario presented above, e.g., the actual verification of a UV fixed point and vacuum stability.

## 6. CONCLUDING REMARKS

QG, appropriately extended to include renormalizable couplings with and of a QFT, gives a renormalizable relativistic field theory of all interactions, which is predictive and computable. It has therefore attracted the interest of several researchers since decades and continues to be an important framework in the quest for a UV complete and phenomenologically viable relativistic field theory.

The price to pay is the presence of a ghost and consequently of an indefinite norm on the Hilbert space (which is implied both by renormalizability and the requirement of having a Hamiltonian that is bounded from below). Therefore, much of this review has been dedicated to illustrate some possible ways to address the ghost problem (such as the Dirac-Pauli quantization, the Lee-Wick approach and the possibility to introduce positively defined metrics on the Hilbert space) focusing on simple finite dimensional quantum mechanical models. The full extension of these techniques to the field theory case (and especially the QG case) has not been done yet and is an important goal for future research.

If QG is coupled to a QFT, which enjoys a UV fixed point, then the whole theory can hold up to infinite energy<sup>29</sup> and might still

<sup>29</sup> If a full solution of the ghost problem in quadratic gravity is found and the theory can be made UV complete (possibly with the inclusion of matter fields) one could also have a window on strongly coupled theories through the holographic principle [136, 137] (in particular the AdS/CFT correspondence [138]) by using QG as the higher dimensional theory on an asymptotically anti-de Sitter (AdS) space. Actually, several works in this direction already appear in the literature (see e.g., [139–142]).

be compatible with data. So far, potentially viable theories have only be found for  $f_0^2 > 0$ , given that  $f_0^2 < 0$  leads to a tachyonic instability (as is clear both in the Jordan and Einstein frame). The explicit construction of a QFT sector that satisfies all collider and cosmological bounds and explain the evidence for new physics has not been achieved yet and is an outstanding target for future research. The deep UV behavior of the theory may be the one of a Weyl invariant theory (conformal gravity): the gravitational coupling  $f_0$  and the non-minimal couplings of the scalar  $\xi_{ab}$  reach the Weyl invariant values  $f_0 \rightarrow \infty$  and  $\xi_{ab} \rightarrow -\delta_{ab}/6$ , whereas all other couplings approach a UV fixed point.

## REFERENCES

- Goroff MH, Sagnotti A. Quantum gravity at two loops. *Phys Lett.* (1985) **160B**:81.
- Goroff MH, Sagnotti A. The ultraviolet behavior of einstein gravity. *Nucl Phys B* (1986) **266**:709–36.
- Utiyama R, DeWitt BS. Renormalization of a classical gravitational field interacting with quantized matter fields. *J Math Phys.* (1962) **3**:608.
- Starobinsky AA. A new type of isotropic cosmological models without singularity. *Phys Lett.* (1980) **91B**:99–102. doi: 10.1016/0370-2693(80)90670-X
- Ade PAR, Aghanim N, Arnaud M, Arroja F, Ashdown M, Aumont J, et al. Planck 2015 results. XX. Constraints on inflation. *Astron Astrophys.* (2016) **594**:A20. doi: 10.1051/0004-6361/201525898
- Weinberg S. Problems in gauge field theories. In: Smith JR, editor. *The Proceedings of the XVII International Conference on High Energy Physics*. Chilton, Didcot: Rutherford Laboratory, III-59 (1974).
- Deser S. The state of quantum gravity. In: Arnowitt R, Nath P, editors. *The Proceedings of the Conference on Gauge Theories and Modern Field Theory*. Cambridge, MA: MIT Press (1975).
- Stelle KS. Renormalization of higher derivative quantum gravity. *Phys Rev D* (1977) **16**:953–69.
- Ostrogradsky M. *Memoires Sur les équations Différentielles Relatives au Problème des Isopérimètres*. Mem. Ac. St. Petersburg VI (1850), p. 385. Available online at: <https://babel.hathitrust.org/cgi/pt?id=mdp.39015038710128;view=1up;seq=405>
- Weinberg S. In *Understanding the Fundamental Constituents of Matter*. Zichichi, A, editor. New York, NY: Plenum Press (1977). [Weinberg S. In *General Relativity: An Einstein Centenary Survey*. Hawking SW, Israel W, editors. (Cambridge: University Press, 1980), p. 790–831.]
- Biswas T, Mazumdar A, Siegel W. Bouncing universes in string-inspired gravity. *J Cosmol Astropart Phys.* (2006) **0603**:9. doi: 10.1088/1475-7516/2006/03/009
- Stelle KS. Classical gravity with higher derivatives. *Gen Rel Grav.* (1978) **9**:353–71.
- Avramidi IG, Barvinsky AO. Asymptotic freedom in higher derivative quantum gravity. *Phys Lett.* (1985) **159B**:269–74. doi: 10.1016/0370-2693(85)90248-5
- Salvio A, Strumia A. Agravity up to infinite energy. *Eur Phys J C* (2018) **78**:124. doi: 10.1140/epjc/s10052-018-5588-4
- Buchbinder IL, Odintsov SD, Shapiro IL. *Effective Action in Quantum Gravity*. Bristol: IOP (1992).
- Avramidi IG. Covariant methods for the calculation of the effective action in quantum field theory and investigation of higher derivative quantum gravity. arXiv:hep-th/9510140.
- de Berredo-Peixoto G, Shapiro IL. Higher derivative quantum gravity with Gauss-Bonnet term. *Phys Rev D* (2005) **71**:064005. doi: 10.1103/PhysRevD.71.064005
- Shapiro IL. Effective action of vacuum: semiclassical approach. *Class Quant Grav.* (2008) **25**:103001. doi: 10.1088/0264-9381/25/10/103001
- Peter P, Salles FDO, Shapiro IL. On the ghost-induced instability on de Sitter background. arXiv:1801.00063. doi: 10.1103/PhysRevD.97.064044
- Kannike K, Hütsi G, Piza L, Racioppi A, Raidal M, Salvio A, et al. Dynamically induced planck scale and inflation. *J High Energy Phys.* (2015) **1505**:065. doi: 10.1007/JHEP05(2015)065
- Salvio A. Inflationary perturbations in no-Scale theories. *Eur Phys J C* (2017) **77**:267. doi: 10.1140/epjc/s10052-017-4825-6
- Sotiriou TP, Faraoni V.  $f(R)$  theories of gravity. *Rev Mod Phys.* (2010) **82**:451. doi: 10.1103/RevModPhys.82.451
- Holdom B, Ren J. QCD analogy for quantum gravity. *Phys Rev D* (2016) **93**:124030. doi: 10.1103/PhysRevD.93.124030
- Holdom B, Ren J. Quadratic gravity: from weak to strong. *Int J Mod Phys D* (2016) **25**:1643004. doi: 10.1142/S0218271816430045
- Weinberg S. *Cosmology*. Oxford: Oxford University Press (2008).
- Hindawi A, Ovrut BA, Waldram D. Consistent spin two coupling and quadratic gravitation. *Phys Rev D* (1996) **53**:5583.
- Johnston DA. Sedentary ghost poles in higher derivative gravity. *Nucl Phys B* (1988) **297**:721–32. doi: 10.1016/0550-3213(88)90555-X
- Barvinsky AO, Blas D, Herrero-Valea M, Sibiryakov SM, Steinwachs CF. Renormalization of gauge theories in the background-field approach. arXiv:1705.03480. doi: 10.1007/JHEP07(2018)035
- Weinberg S. *The Quantum Theory of Fields. Vol. 1: Foundations*. Cambridge: University press.
- Fradkin ES, Tseytlin AA. Renormalizable asymptotically free quantum theory of gravity. *Nucl Phys B* (1982) **201**:469–91. doi: 10.1016/0550-3213(82)90444-8
- Narain G, Anishetty R. Charge renormalization due to graviton loops. *J High Energy Phys.* (2013) **1307**:106. doi: 10.1007/JHEP07(2013)106
- Narain G, Anishetty R. Running couplings in quantum theory of gravity coupled with gauge fields. *J High Energy Phys.* (2013) **1310**:203. doi: 10.1007/JHEP10(2013)203
- Salvio A, Strumia A. Agravity. *J High Energy Phys.* (2014) **1406**:80. doi: 10.1007/JHEP06(2014)080
- Julve J, Tonin M. Quantum gravity with higher derivative terms. *Nuovo Cim B* (1978) **46**:137–52.
- Fradkin ES, Tseytlin AA. Renormalizable asymptotically free quantum theory of gravity. *Phys Lett.* (1981) **104B**:377.
- Codello A, Percacci R. Fixed points of higher derivative gravity. *Phys Rev Lett.* (2006) **97**:221301. doi: 10.1103/PhysRevLett.97.221301
- Einhorn MB, Jones DRT. Gauss-Bonnet coupling constant in classically scale-invariant gravity. *Phys Rev D.* (2015) **91**:084039. doi: 10.1103/PhysRevD.91.084039
- Ohta N, Percacci R. Higher derivative gravity and asymptotic safety in diverse dimensions. *Class Quant Grav.* (2014) **31**:015024. doi: 10.1088/0264-9381/31/1/015024
- Anselmi D, Piva M. The ultraviolet behavior of quantum gravity. arXiv:1803.07777.
- Pais A, Uhlenbeck GE. On field theories with nonlocalized action. *Phys Rev.* (1950) **79**:145.
- Pagani E, Tecchiolli G, Zerbini S. On the problem of stability for higher order derivatives: lagrangian systems. *Lett Math Phys.* (1987) **14**:311.
- Smilga AV. Benign versus malicious ghosts in higher-derivative theories. *Nucl Phys B* (2005) **706**:598. doi: 10.1016/j.nuclphysb.2004.10.037

## AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and approved it for publication.

## ACKNOWLEDGMENTS

I thank Matej Pavšič, Ilya Shapiro, Alessandro Strumia, and Hardi Veermäe for useful discussions and correspondence. This work was supported by the ERC grant NEO-NAT.

43. Pavšič M. Stable self-interacting Pais-Uhlenbeck oscillator. *Mod Phys Lett A* (2013) **28**:1350165. doi: 10.1142/S0217732313501654
44. Kaparulin DS, Lyakhovich SL, Sharapov AA. Classical and quantum stability of higher-derivative dynamics. *Eur Phys J C* (2014) **74**:3072. doi: 10.1140/epjc/s10052-014-3072-3
45. Pavšič M. Pais-Uhlenbeck oscillator and negative energies. *Int J Geom Meth Mod Phys*. (2016) **13**:1630015. doi: 10.1142/S0219887816300154
46. Smilga A. Classical and quantum dynamics of higher-derivative systems. *Int J Mod Phys A* (2017) **32**:1730025. doi: 10.1142/S0217751X17300253
47. Ivanov MM, Tokareva AA. Cosmology with a light ghost. *J Cosmol Astropart Phys*. (2016) **1612**:018. doi: 10.1088/1475-7516/2016/12/018
48. Tokareva A. Inflation with light Weyl ghost. *EPJ Web Conf.* (2016) **125**:03020. doi: 10.1051/epjconf/201612503020
49. Buchbinder IL, Lyakhovich SL. Canonical Quantization and Local Measure of  $R^2$  Gravity. *Class. Quant. Grav.* (1987) **4**:1487.
50. Buchbinder IL, Lyakhovich SL. Canonical quantization of theories with higher derivatives: quantization of  $R^2$  gravitation. *Theor Math Phys*. (1987) **72**:824.
51. Kluson J, Oksanen M, Tureanu A. Hamiltonian analysis of curvature-squared gravity with or without conformal invariance. *Phys Rev. D* (2014) **89**:064043. doi: 10.1103/PhysRevD.89.064043
52. Mannheim PD, Davidson A. Dirac quantization of the Pais-Uhlenbeck fourth order oscillator. *Phys Rev A* (2005) **71**:042110. doi: 10.1103/PhysRevA.71.042110
53. Smilga AV. Ghost-free higher-derivative theory. *Phys Lett B* (2006) **632**:433. doi: 10.1016/j.physletb.2005.10.014
54. Mannheim PD. Solution to the ghost problem in fourth order derivative theories. *Found Phys.* (2007) **37**:532. doi: 10.1007/s10701-007-9119-7
55. Bolonek K, Kosinski P. On double frequency limit of Pais-Uhlenbeck oscillator. arXiv:quant-ph/0612009.
56. Woodard RP. Avoiding dark energy with  $1/r$  modifications of gravity. *Lect Notes Phys.* (2007) **720**:403. doi: 10.1007/978-3-540-71013-4\_14
57. Woodard RP. Ostrogradsky's theorem on Hamiltonian instability. *Scholarpedia* (2015) **10**:32243. doi: 10.4249/scholarpedia.32243
58. Salvio A, Strumia A. Quantum mechanics of 4-derivative theories. *Eur Phys J C* (2016) **76**:227. doi: 10.1140/epjc/s10052-016-4079-8
59. Pauli W. On Dirac's new method of field quantization. *Rev Mod Phys.* (1943) **15**:175.
60. Dirac PAM. The physical interpretation of quantum mechanics. *Proc. R. Soc Lond. A* (1942) **180**:1.
61. Hawking SW, Hertog T. Living with ghosts. *Phys Rev D* (2002) **65**:103515. doi: 10.1103/PhysRevD.65.103515
62. Bender CM, Mannheim PD. Exactly solvable PT-symmetric Hamiltonian having no Hermitian counterpart. *Phys Rev D* (2008) **78**:025022. doi: 10.1103/PhysRevD.78.025022
63. Bender CM, Mannheim PD. No-ghost theorem for the fourth-order derivative Pais-Uhlenbeck oscillator model. *Phys Rev Lett.* (2008) **100**:110402. doi: 10.1103/PhysRevLett.100.110402
64. Boulware DG, Gross DJ. Lee-wick indefinite metric quantization: a functional integral approach. *Nucl Phys B* (1984) **233**:1.
65. Lee TD, Wick GC. Finite theory of quantum electrodynamics. *Phys Rev D* (1970) **2**:1033.
66. Coleman S. Acausality. In Zichichi A, editor. *Erice 1969, Ettore Majorana School On Subnuclear Phenomena*. New York, NY: Academic Press (1970), p. 282–327.
67. Grinstein B, O'Connell D, Wise MB. Causality as an emergent macroscopic phenomenon: the Lee-Wick O(N) model. *Phys Rev D* (2009) **79**:105019. doi: 10.1103/PhysRevD.79.105019
68. Lee TD, Wick GC. Negative metric and the unitarity of the S matrix. *Nucl Phys B* (1969) **9**:209.
69. Tomboulis E. Renormalizability and asymptotic freedom in quantum gravity. *Phys Lett.* (1980) **97B**:77.
70. Antoniadis I, Tomboulis ET. Gauge invariance and unitarity in higher derivative quantum gravity. *Phys Rev D* (1986) **33**:2756.
71. Hasslacher B, Mottola E. Asymptotically free quantum gravity and black holes. *Phys Lett.* (1981) **99B**:221.
72. Salvio A. Solving the standard model problems in softened gravity. *Phys Rev. D* (2016) **94**:096007. doi: 10.1103/PhysRevD.94.096007
73. Anselmi D, Piva M. A new formulation of Lee-Wick quantum field theory. *J High Energy Phys.* (2017) **1706**:066. doi: 10.1007/JHEP06(2017)066
74. Anselmi D, Piva M. Perturbative unitarity of Lee-Wick quantum field theory. *Phys Rev D* (2017) **96**:045009. doi: 10.1103/PhysRevD.96.045009
75. Anselmi D. On the quantum field theory of the gravitational interactions. *J High Energy Phys.* (2017) **1706**:086. doi: 10.1007/JHEP06(2017)086
76. Anselmi D. Fakeons and Lee-Wick models. (2018). arXiv:1801.00915. doi: 10.1007/JHEP02(2018)141
77. Abe Y, Inami T, Izumi K, Kitamura T. Matter scattering in  $R^2_{\mu\nu}$  gravity and unitarity. *PTEP* (2018) **2018**:031E01. doi: 10.1093/ptep/pty010
78. Donoghue JF, and Menezes G. Gauge assisted quadratic gravity: a framework for UV complete quantum gravity. (2018). arXiv:1804.04980. doi: 10.1103/PhysRevD.97.126005
79. Feynman RP. Negative probability. In: Hiley BJ, Peat FD, editors. *Quantum Implications: Essays in Honor of David Bohm*, Chap. 13. London: Routledge and Kegan Paul (1987), p. 235–48.
80. Bender CM, Brody DC, Jones HF. Complex extension of quantum mechanics. *Phys Rev Lett.* (2002) **89**:270401. Erratum: [*Phys. Rev. Lett.* (2004) **92**:119902].
81. Bender CM. Making sense of non-Hermitian Hamiltonians. *Rept Prog Phys.* (2007) **70**:947. doi: 10.1088/0034-4885/70/6/R03
82. Mannheim PD. Antilinearity rather than hermiticity as a guiding principle for quantum theory. arXiv:1512.04915.
83. Raidal M, Veermäe H. On the quantisation of complex higher derivative theories and avoiding the ostrogradsky ghost. *Nucl Phys B* (2017) **916**:607. doi: 10.1016/j.nuclphysb.2017.01.024
84. Mannheim PD. Appropriate inner product for PT-symmetric hamiltonians. *Phys Rev D* (2018) **97**:045001. doi: 10.1103/PhysRevD.97.045001
85. Strumia A. Interpretation of quantum mechanics with indefinite norm. arXiv:1709.04925
86. Berkin AL. Contribution of the Weyl tensor to  $R^2$  inflation. *Phys Rev D* (1991) **44**:1020.
87. Hamada KJ, Horata S, Yukawa T. Space-time evolution and CMB anisotropies from quantum gravity. *Phys Rev D* (2006) **74**:123502. doi: 10.1103/PhysRevD.74.123502
88. Clunan T, Sasaki M. Tensor ghosts in the inflationary cosmology. *Class Quant Grav.* (2010) **27**:165014. doi: 10.1088/0264-9381/27/16/165014
89. Weinberg S. Asymptotically safe inflation. *Phys Rev D* (2010) **81**:083535. doi: 10.1103/PhysRevD.81.083535
90. Nelson W. Restricting fourth order gravity via cosmology. *Phys Rev D* (2010) **82**:124044. doi: 10.1103/PhysRevD.82.124044
91. Deruelle N, Sasaki M, Sendouda Y, Youssef A. Inflation with a Weyl term, or ghosts at work. *J Cosmol Astropart Phys.* (2011) **1103**:040. doi: 10.1088/1475-7516/2011/03/040
92. Myung YS, Moon T. Primordial massive gravitational waves from Einstein-Chern-Simons-Weyl gravity. *J Cosmol Astropart Phys.* (2014) **1408**:061. doi: 10.1088/1475-7516/2014/08/061
93. Shapiro IL, Pelinson AM, Salles FO. Gravitational waves and perspectives for quantum gravity. *Mod Phys Lett A* (2014) **29**:1430034. doi: 10.1142/S0217732314300341
94. Myung YS, Moon T. Scale-invariant tensor spectrum from conformal gravity. *Mod Phys Lett A* (2015) **30**:1550172. doi: 10.1142/S0217732315501722
95. Salles FO, Shapiro IL. Do we have unitary and (super)renormalizable quantum gravity below the Planck scale? *Phys Rev. D* (2014) **89**:084054. Erratum: [*Phys Rev D* (2014) **90**:129903]. doi: 10.1103/PhysRevD.89.084054
96. Abbott BP, Abbott R, Abbott TD, Abernathy MR, Acernese F, Ackley K. [LIGO Scientific and Virgo Collaborations], Observation of gravitational waves from a binary black hole merger. *Phys Rev Lett.* (2016) **116**:061102. arXiv:1602.03837.
97. Nelson W. Static solutions for 4th order gravity. *Phys Rev D* (2010) **82**:104026. doi: 10.1103/PhysRevD.82.104026
98. Lu H, Perkins A, Pope CN, Stelle KS. Black holes in higher-derivative gravity. *Phys Rev Lett.* (2015) **114**:171601. doi: 10.1103/PhysRevLett.114.171601
99. Holdom B. On the fate of singularities and horizons in higher derivative gravity. *Phys Rev D* (2002) **66**:084010. doi: 10.1103/PhysRevD.66.084010

100. Lü H, Perkins A, Pope CN, Stelle KS. Spherically symmetric solutions in higher-derivative gravity. *Phys Rev D* (2015) **92**:124019. doi: 10.1103/PhysRevD.92.124019
101. Cai YF, Cheng G, Liu J, Wang M and Zhang H. Features and stability analysis of non-Schwarzschild black hole in quadratic gravity. *J High Energy Phys.* (2016) **1601**:108. doi: 10.1007/JHEP01(2016)108
102. Lin K, Qian WL, Pavan AB, Abdalla E. (Anti-) de sitter electrically charged black hole solutions in higher-derivative gravity. *EPL* (2016) **114**:60006. doi: 10.1209/0295-5075/114/60006
103. Lü H, Perkins A, Pope CN, Stelle KS. Lichnerowicz modes and black hole families in ricci quadratic gravity. *Phys Rev D* (2017) **96**:046006. doi: 10.1103/PhysRevD.96.046006
104. Goldstein K, Mashiyane JJ. Ineffective higher derivative black hole hair. *Phys Rev D* (2018) **97**:024015. doi: 10.1103/PhysRevD.97.024015
105. Kokkotas K, Konoplya RA, Zhidenko A. Non-Schwarzschild black-hole metric in four dimensional higher derivative gravity: analytical approximation. *Phys Rev D* (2017) **96**:064007. doi: 10.1103/PhysRevD.96.064007
106. Stelle KS. Abdus Salam and quadratic curvature gravity: vlassical solutions. *Int J Mod Phys A* (2017) **32**:1741012. doi: 10.1142/S0217751X17410123
107. Holdom B, Ren J. Not quite a black hole. *Phys Rev D* (2017) **95**:084034. doi: 10.1103/PhysRevD.95.084034
108. Myung YS. Stability of Schwarzschild black holes in fourth-order gravity revisited. *Phys Rev D* (2013) **88**:024039. doi: 10.1103/PhysRevD.88.024039
109. Buchbinder IL, Kalashnikov OK, Shapiro IL, Vologodsky VB, Wolfengaut JJ. The stability of asymptotic freedom in grand unified models coupled to  $R^2$  gravity. *Phys Lett B* (1989) **216**:127.
110. Tomboulis ET. Renormalization and unitarity in higher derivative and nonlocal gravity theories. *Mod Phys Lett A* (2015) **30**:1540005. doi: 10.1142/S0217732315400052
111. Einhorn MB, Jones DRT. Induced gravity I: real scalar field. *J High Energy Phys.* (2016) **1601**:019. doi: 10.1007/JHEP01(2016)019
112. Einhorn MB, Jones DRT. Induced gravity II: grand unification. *J High Energy Phys.* (2016) **1605**:185. doi: 10.1007/JHEP05(2016)185
113. Einhorn MB, Jones DRT. Renormalizable, asymptotically free gravity without ghosts or tachyons. *Phys Rev D* (2017) **96**:124025. doi: 10.1103/PhysRevD.96.124025
114. Alvarez-Gaume L, Kehagias A, Kounnas C, Lüst D, Riotto A. Aspects of quadratic gravity. *Fortsch. Phys.* (2016) **64**:176. doi: 10.1002/prop.201500100
115. Narain G, Anishetty R. Short distance freedom of quantum gravity. *Phys Lett B* (2012) **711**:128. doi: 10.1016/j.physletb.2012.03.070
116. Narain G. Exorcising ghosts in induced gravity. *Eur Phys J C* (2017) **77**:683. doi: 10.1140/epjc/s10052-017-5249-z
117. Narain G. Signs and stability in higher-derivative gravity. *Int J Mod Phys A* (2018) **33**:1850031. doi: 10.1142/S0217751X18500318
118. Fradkin ES, Vilkovisky GA. Conformal invariance and asymptotic freedom in quantum gravity. *Phys Lett.* (1978) **77B**:262.
119. Zee A Einstein gravity emerging from quantum Weyl gravity. *Ann Phys.* (1983) **151**:431.
120. Shapiro IL, Cognola G. Interaction of low-energy induced gravity with quantized matter and phase transition induced to curvature. *Phys Rev D* (1995) **51**:2775.
121. Hamada KJ. Resummation and higher order renormalization in 4-D quantum gravity. *Prog Theor Phys.* (2002) **108**:399. doi: 10.1143/PTP.108.399
122. Hamada KJ. Renormalizable 4D quantum gravity as a perturbed theory from CFT. *Found Phys.* (2009) **39**:1356. doi: 10.1007/s10701-009-9358-x
123. Donoghue JF. Conformal model of gravitons. *Phys Rev. D* (2017) **96**:044006. doi: 10.1103/PhysRevD.96.044006
124. Alvarez E, Anero J, Gonzalez-Martin S. Quadratic gravity in first order formalism. *J Cosmol Astropart Phys.* (2017) **1710**:008. doi: 10.1088/1475-7516/2017/10/008
125. Giudice GF, Isidori G, Salvio A, Strumia A. Softened Gravity and the Extension of the Standard Model up to Infinite Energy. *J High Energy Phys.* (2015) **1502**:137. doi: 10.1007/JHEP02(2015)137
126. Holdom B, Ren J, Zhang C. Stable asymptotically free extensions (SAFEs) of the standard model. *J High Energy Phys.* (2015) **1503**:028. doi: 10.1007/JHEP03(2015)028
127. Pelaggi GM, Strumia A, Vignali S. Totally asymptotically free trification. *J High Energy Phys.* (2015) **1508**:130. doi: 10.1007/JHEP08(2015)130
128. Mann R, Meffe J, Sannino F, Steele T, Wang ZW, Zhang C. Asymptotically safe standard model via vectorlike fermions. *Phys Rev Lett.* (2017) **119**:261802. doi: 10.1103/PhysRevLett.119.261802
129. Pelaggi GM, Plascencia AD, Salvio A, Sannino F, Smirnov J, Strumia A. Asymptotically safe standard model extensions? arXiv:1708.00437. doi: 10.1103/PhysRevD.97.095013
130. Hathrell SJ. Trace anomalies and  $\lambda\phi^4$  theory in curved space. *Ann Phys.* (1982) **139**:136.
131. Hathrell SJ. Trace anomalies and QED in curved space. *Anna Phys.* (1982) **142**:34.
132. Jack I, Osborn H. Analogs for the  $c$  theorem for four-dimensional renormalizable field theories. *Nucl Phys B* (1990) **343**:647.
133. hapiro IL, Zhukenshaev AG. Gauge dependence in higher derivative quantum gravity and the conformal anomaly problem. *Phys Lett B* (1994) **324**:286.
134. de Berredo-Peixoto G, Shapiro IL. Conformal quantum gravity with the Gauss-Bonnet term. *Phys Rev.* (2003) **D70**:044024. doi: 10.1103/PhysRevD.70.044024
135. Ohta N, Percacci R. Ultraviolet fixed points in conformal gravity and general quadratic theories. *Class Quant Grav.* (2016) **33**:035001. doi: 10.1088/0264-9381/33/3/035001
136. t Hooft G. Dimensional reduction in quantum gravity. *Conf Proc C* (1993) **930308**:284.
137. Susskind L. The World as a hologram. *J Math Phys.* (1995) **36**:6377.
138. Maldacena JM. The large N limit of superconformal field theories and supergravity. *Int J Theor Phys.* (1999) **38**:1113. [*Adv Theor Math Phys.* (1998) **2**:231].
139. Fukuma M, Matsuura S. Holographic renormalization group structure in higher derivative gravity. *Prog Theor Phys.* (2002) **107**:1085. doi: 10.1143/PTP.107.1085
140. Dong X. Holographic entanglement entropy for general higher derivative gravity. *J High Energy Phys.* (2014) **1401**:044. doi: 10.1007/JHEP01(2014)044
141. Erdmenger J, Flory M, Sleight C. Conditions on holographic entangling surfaces in higher curvature gravity. *J High Energy Phys.* (2014) **1406**:104. doi: 10.1007/JHEP06(2014)104
142. Bhattacharjee S, Sarkar S, Wall AC. Holographic entropy increases in quadratic curvature gravity. *Phys Rev D* (2015) **92**:064006. doi: 10.1103/PhysRevD.92.064006

**Conflict of Interest Statement:** The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Salvio. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Euclidean Wormholes, Baby Universes, and Their Impact on Particle Physics and Cosmology

Arthur Hebecker\*, Thomas Mikhail and Pablo Soler

*Institute for Theoretical Physics, University of Heidelberg, Heidelberg, Germany*

## OPEN ACCESS

### Edited by:

Alberto Salvio,  
European Organization for Nuclear  
Research (CERN), Switzerland

### Reviewed by:

Thomas Hertog,  
KU Leuven, Belgium  
Ashkbiz Danehkar,  
Harvard-Smithsonian Center for  
Astrophysics, United States

### \*Correspondence:

Arthur Hebecker  
a.hebecker@thphys.uni-heidelberg.de

### Specialty section:

This article was submitted to  
High-Energy and Astroparticle  
Physics,  
a section of the journal  
Frontiers in Astronomy and Space  
Sciences

**Received:** 29 June 2018

**Accepted:** 10 September 2018

**Published:** 08 October 2018

### Citation:

Hebecker A, Mikhail T and Soler P  
(2018) Euclidean Wormholes, Baby  
Universes, and Their Impact on  
Particle Physics and Cosmology.  
*Front. Astron. Space Sci.* 5:35.  
doi: 10.3389/fspas.2018.00035

The euclidean path integral remains, in spite of its familiar problems, an important approach to quantum gravity. One of its most striking and obscure features is the appearance of gravitational instantons or wormholes. These renormalize all terms in the Lagrangian and cause a number of puzzles or even deep inconsistencies, related to the possibility of nucleation of “baby universes.” In this review, we revisit the early controversies surrounding these issues as well as some of the more recent discussions of the phenomenological relevance of gravitational instantons. In particular, wormholes are expected to break the shift symmetries of axions or Goldstone bosons non-perturbatively. This can be relevant to large-field inflation and connects to arguments made on the basis of the Weak Gravity or Swampland conjectures. It can also affect Goldstone bosons which are of physical interest in the context of the strong CP problem or as dark matter.

**Keywords:** string theory, quantum gravity, euclidean wormhole, axions, particle physics -cosmology connection, inflation, weak gravity conjecture, gravitational instanton

## 1. INTRODUCTION

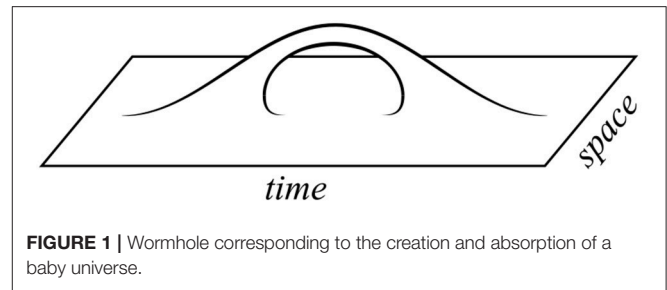
It is reasonable to think that a consistent theory of quantum gravity has to allow for topology change. Indeed, if the euclidean path integral has any relevance at all, then it appears unnatural to forbid 4-manifolds with non-trivial topology. After all, they are locally indistinguishable from  $\mathbb{R}^4$ . Further evidence in favor of topology change comes, for example, from string theory: String interactions and loops rely entirely on topology change in the worldsheet theory, the latter being a relatively well-understood examples of 2d quantum gravity. In addition, 10d supergravity theories with their stringy UV completion involve controlled examples of topology change. These occur if one dynamically moves through special loci in Calabi-Yau moduli space, e.g., through a conifold point.

However, our point of departure will be more simple minded, focusing on topology change in 4d effective quantum gravity. Consider the evolution of 3d spatial manifolds in time. It is natural to think that in the course of this evolution an  $\mathbb{R}^3$  can transit to an  $\mathbb{R}^3$  plus an  $S^3$  “baby universe,” which subsequently reunite becoming again an  $\mathbb{R}^3$  (cf. **Figure 1**). This can be viewed as a tunneling transition, which gains quantitative support from the existence of a corresponding euclidean solution—the Giddings-Strominger wormhole (Giddings and Strominger, 1988a). While topology change has been discussed before (Wheeler, 1955; Regge, 1961; Hawking, 1978, 1987, 1988; Ellis et al., 1984; Lavrelashvili et al., 1987), the Giddings-Strominger solution (Giddings and Strominger, 1988a) and especially the application to the cosmological constant problem suggested

by Coleman (1988c) led to an enormous spike of activity (Coleman, 1988a; Giddings and Strominger, 1988b, 1989b; Grinstein and Wise, 1988; Hawking and Laflamme, 1988; Lee, 1988; Rubakov, 1988; Abbott and Wise, 1989; Brown et al., 1989; Burgess and Kshirsagar, 1989; Choi and Holman, 1989; Coleman and Lee, 1989, 1990a,b; Duff, 1989; Fischler and Susskind, 1989; Gilbert, 1989; Grinstein, 1989; Grinstein and Hill, 1989; Klebanov et al., 1989; Nielsen and Ninomiya, 1989; Polchinski, 1989b; Preskill, 1989; Preskill et al., 1989; Rey, 1989; Tamvakis, 1989; Carlip and De Alwis, 1990; Grinstein and Maharana, 1990; Grinstein et al., 1990; Hawking, 1990a,b, 1991a,b; Hawking and Page, 1990; Tamvakis and Vayonakis, 1990; Lyons and Hawking, 1991; Linde, 1992; Twamley and Page, 1992; Kaplunovsky, unpublished; see Coleman et al., 1991 for an early overview).

As part of these investigations, severe problems in the resulting picture of a macroscopic spacetime surrounded by baby universes were uncovered (Fischler and Susskind, 1989; Kaplunovsky, unpublished; Polchinski, 1989b; Hawking, 1990b). While the interest has then subsided, important results have continued to appear over the years (Kallosh et al., 1995; Nirov and Rubakov, 1995; Barcelo et al., 1996; Gibbons et al., 1996; Rubakov and Shvedov, 1996a,b; Green and Gutperle, 1997; Rey, 1999; Gutperle and Sabra, 2002; Bergshoeff et al., 2004, 2005; Maldacena and Maoz, 2004; Collinucci, 2005; Bergshoeff et al., 2006; Dijkgraaf et al., 2006; Arkani-Hamed et al., 2007b; Bergman and Distler, 2007; Chiodaroli and Gutperle, 2009a,b; Cortes and Mohaupt, 2009; Mohaupt and Waite, 2011; Betzios et al., 2018). It has, however, neither been shown that wormholes and baby universe are unphysical nor has a satisfactory overall picture been developed. Thus, euclidean wormholes or gravitational instantons have remained a lurking fundamental issue in our understanding of quantum gravity. We emphasize that this issue is not easily dismissed as a problem of the UV completion. On the contrary, large wormholes tend to be as puzzling as small ones, such that the problems appear to be there even in the low-energy effective theory<sup>1</sup>.

More recently, the interest in wormholes has been renewed in the context of large-field inflation, axion-physics, and the widespread excitement (see e.g., Cheung and Remmen, 2014; Brown et al., 2015, 2016; de la Fuente et al., 2015; Hebecker et al., 2015; Heidenreich et al., 2015; Montero et al., 2015; Rudelius, 2015; Bachlechner et al., 2016; Choi and Kim, 2016; Junghans, 2016; Kaloper et al., 2016; Kappl et al., 2016; Kooner et al., 2016; Klaewer and Palti, 2017) about the Weak Gravity Conjecture and the Landscape/Swampland paradigm (Vafa, 2005; Arkani-Hamed et al., 2007a; Ooguri and Vafa, 2007, 2016; Brennan et al., 2017). This is natural since wormholes have the potential to break global symmetries, such as the shift symmetry of the axion. In addition, they may be considered the macroscopic, gravitational version of instantons in pretty much the same way as charged black



**FIGURE 1** | Wormhole corresponding to the creation and absorption of a baby universe.

holes are the macroscopic version of charged particles. Thus, the interest in the Weak Gravity Conjecture and its implications for phenomenology naturally lead to an enhanced interest in (euclidean) wormholes (Hebecker et al., 2015; Montero et al., 2015; Heidenreich et al., 2016; Harlow, 2016; Alonso and Urbano, 2017; Hebecker et al., 2017; Hertog et al., 2017; Ruggeri et al., 2018; Shiu and Staessens, 2018).

Our review is motivated in several ways: First, as just explained, it is timely to reconsider the wormhole issue in view of the growing interest in generic quantum gravity constraints on effective field theories. Second, the unsolved problems from the 90's are, in our opinion, as important as ever. Additionally, one of the main phenomenological targets in the otherwise rather theory-driven wormhole debate have always been axions<sup>2</sup>. Since axions are becoming more and more central in Beyond-the-Standard-Model research, scrutinizing their generic features is of particular importance. Finally, we believe that the post-90's theoretical developments of AdS/CFT, holography and (gravitational) entanglement have not yet been fully exploited in the context of euclidean wormholes. Thus, significant technical progress may be expected concerning the fundamental issues raised by those objects.

In the long run, we can think of two different outcomes: On the one hand, wormhole effects may turn out to be absent from certain theories, in particular from the 4d quantum gravity describing the real world. This would solve many puzzles. Advocates of this possibility have to address a number of questions. In particular, what is the specific mechanism behind this “wormhole censorship”? As we will argue, it appears difficult to imagine such a mechanism which would not also forbid topology change in general. This, of course, would be a radical step. Related to this: How can we forbid wormholes in 4d while maintaining their central role in the 2d quantum gravity known as string theory? Furthermore, if wormholes are forbidden, what is the generic gravitational effect responsible for the breaking of global shift symmetries of axions? On the other hand, if wormholes exist, they represent a radical departure from standard interpretations of effective field theories. As we will describe, the correct understanding of their effects requires solving numerous fundamental problems. In the hope that these questions can be successfully addressed in the near future, we

<sup>1</sup>In this review we focus on large wormholes. An interesting and closely related topic, which lies beyond the scope of this work, are topological fluctuations of spacetime at small scales (the Planck scale Wheeler, 1955; Hawking, 1978 or string scale Iqbal et al., 2008) as constituents of a microscopic description of quantum gravity.

<sup>2</sup>We will use the name axion for any shift-symmetric periodic scalar, even if unrelated to QCD.

consider it worthwhile summarizing the state of the art and describing the main puzzles and open issues posed by wormholes.

We start in section 2 by recalling how instantons (of either gauge-theoretic or stringy nature) generate a potential for any scalar to which they are minimally coupled. We then describe the famous Giddings-Strominger solution (Giddings and Strominger, 1988a), which corresponds to a throat with cross section  $S^3$ , connecting two points in  $\mathbb{R}^4$  (cf. **Figure 1**). The throat is supported by  $H_3$  flux, and the dual of  $H_3$  is the field strength of an axion. This axion then naturally couples to the two wormhole ends, which can locally be interpreted as instanton and anti-instanton. The axionic shift symmetry is potentially broken by a “dilute gas” of such wormholes. We also briefly comment on dilatonic instantons as they generically arise in string theory, emphasizing that it has by now been established that wormhole solutions do really arise in string-derived models (Tamvakis, 1989; Bergshoeff et al., 2005; Arkani-Hamed et al., 2007b; Hertog et al., 2017).

Next, in section 3, we discuss how the low-energy effective action is corrected by wormholes (of Giddings-Strominger type and, more generally, by any “spacetime handles” of the form displayed in **Figure 1**). We follow the pioneering work by Coleman (1988c) and Preskill (1989). Crucially, in contrast to instantons, wormholes induce a bilocal action, which has the potential to break locality or even quantum coherence. However, the bilocal correction can be turned into a local one by introducing appropriate auxiliary integration variables ( $\alpha$  parameters). Alternatively, this can be captured by thinking in terms of a “state of baby universes,” the absorption and emission of which is described by operators  $a^\dagger$  and  $a$ . In this language, the  $\alpha$  parameters are simply the eigenvalues of  $\hat{\alpha} = a + a^\dagger$ . If the (infinitely many)  $\alpha$  parameters take definite and not excessively large values, effective 4d locality and the dilute gas approximation are maintained. However, exact predictivity for Lagrangian parameters on the basis of some underlying microscopic theory is lost.

Section 4 is devoted to phenomenological applications. The early literature focuses on the indeterminacy of effective coupling constants. In particular, Coleman argued that the cosmological constant is statistically driven to zero value by the distribution of  $\alpha$  parameters and their interplay with large-scale 4d gravity (Coleman, 1988c). The violation of axionic shift symmetries and other global symmetries has also been studied from the beginning (see e.g., Rey, 1989). More recently, the shift symmetry of a large- $f$  axion has been discussed in the context of wormholes and their interplay with the Weak Gravity Conjecture (Hebecker et al., 2015; Montero et al., 2015; Heidenreich et al., 2016). We review some of this discussion, pointing out in particular difficulties in making strong, generic arguments against large-field axionic inflation (Hebecker et al., 2017). Additionally, we discuss possible wormhole effects on axions with  $f < M_P$  (including but not limited to the QCD axion) following (Alonso and Urbano, 2017). These may be relevant to ultralight dark matter, axion stars and black hole superradiance.

Open conceptual issues are the main subject of section 5. There are many of those, making the whole subject interesting but at the same time very difficult. We start with the

FKS catastrophe (Fischler and Susskind, 1989; Kaplunovsky, unpublished), which turns Coleman’s cosmological constant calculation into an argument for an overdensity of large wormholes. We go on to briefly discuss the generic problems of euclidean quantum gravity and, in particular, the negative-mode problems possibly affecting the Giddings-Strominger solution (Rubakov and Shvedov, 1996a; Alonso and Urbano, 2017; Hertog et al., 2017). Finally, we discuss the quantum cosmology involving macroscopic universes and a baby universe state. This can be relatively well understood in a 1d toy model, but becomes already rather complicated in 2d quantum gravity. The latter case has of course received particular attention since its “large universe” may be the worldsheet of a fundamental string, while the baby universe state is represented by the dynamical target space of string theory. Finally, we analyse the Wheeler-DeWitt perspective as well as issues arising in the AdS/CFT paradigm. We conclude in section 6.

## 2. FROM INSTANTONS TO WORMHOLES

In this section we describe the simplest wormhole configurations, extrema of the euclidean action of Einstein gravity coupled to axionic fields (and possibly dilatons). We start with a brief description of the related but much better understood case of flat spacetime, where instantons arise as euclidean saddle points of gauge theories.

### 2.1. Instantons

Let us start by recalling the familiar case of a 4d gauge theory with

$$\mathcal{L} = \frac{1}{2g^2} \text{tr} F_{\mu\nu} F^{\mu\nu}. \quad (1)$$

For simplicity the gauge group is taken to be  $SU(2)$ . The euclidean path integral necessarily involves certain finite action configurations (instantons) for which the field strength is non-zero in the vicinity of some point  $x_0 \in \mathbb{R}^4$  and falls off quickly as  $|x - x_0| \rightarrow \infty$ . Moreover, the value of

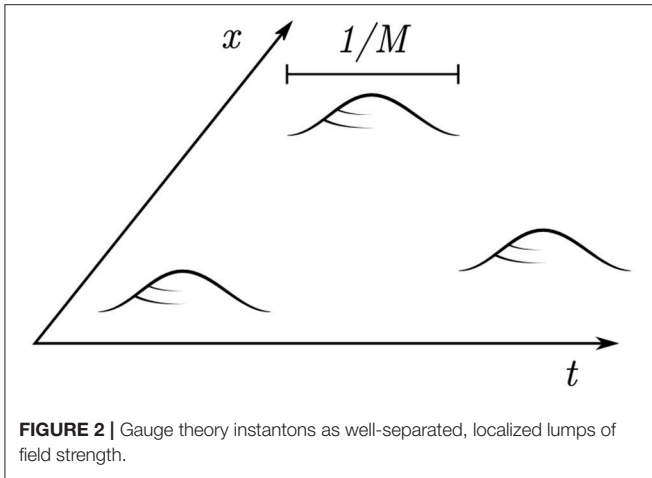
$$n = \frac{1}{8\pi^2} \int \text{tr}(F \wedge F) \quad (2)$$

is integer, with  $n = \pm 1$  characterizing a single instanton or anti-instanton (see e.g., Coleman, 1979; Vainshtein et al., 1982; Tong, 2005; Bianchi et al., 2008; Vandoren and van Nieuwenhuizen, 2008). The minimal action for such  $n = \pm 1$  configurations is

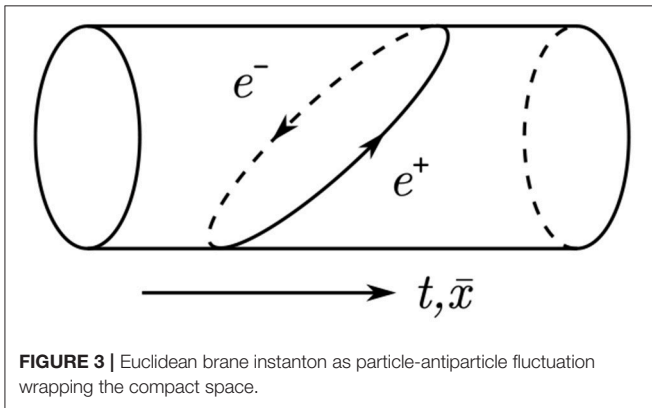
$$S = \frac{8\pi^2}{g^2}. \quad (3)$$

The underlying solutions have 8 moduli: the components of  $x_0$ , a size modulus, and three zero modes associated to global  $SU(2)$  transformations.

In calculating the partition function of the theory, one has to sum over any number of such instanton or anti-instanton configurations and integrate over all their moduli. This can be done very explicitly (see below) in the so-called dilute instanton gas approximation, i.e., assuming that the regions where  $F$  is



**FIGURE 2** | Gauge theory instantons as well-separated, localized lumps of field strength.



**FIGURE 3** | Euclidean brane instanton as particle-antiparticle fluctuation wrapping the compact space.

significantly non-zero are much smaller than their distance. Unfortunately, this clashes with the fact that a large contribution comes from very extended instantons, making the calculation e.g., in the practically interesting case of QCD non-trivial. A relevant toy model can however be obtained by Higgsing the gauge theory at  $M \gg \Lambda$ , with  $\Lambda$  the confinement scale. The largest instantons now have size  $\sim 1/M$  and the dilute gas approximation can be parametrically controlled (cf. **Figure 2**).

Another equally familiar case is that of stringy or exotic instantons. To recall this case, start with the toy model of a 5d gauge theory on  $\mathbb{R}^{1,3} \times S^1$ . Clearly, if charged particles exist, this theory has tunneling processes in which a particle-antiparticle pair emerges from the vacuum and annihilates after passing around the  $S^1$  in opposite directions (cf. **Figure 3**). In the euclidean theory, this corresponds to a 0-brane wrapped on the  $S^1$  at some point  $x_0 \in \mathbb{R}^4$ . The generalization to string compactifications with appropriate Dp-branes (or Ep-branes, with “E” for euclidean) wrapped on  $(p+1)$ -cycles of the compact space is obvious (for reviews see e.g., Akerblom et al., 2007; Blumenhagen et al., 2009; Ibanez and Uranga, 2012).

Crucially, in both of the above examples a shift symmetric, periodic scalar coupling to the instantons is naturally expected to be present. In the first case, it is the analog of the QCD axion,

coupling through

$$\mathcal{L} \supset \theta \operatorname{tr}(F \wedge F)/8\pi^2. \quad (4)$$

In the second case, it is the “Wilson-line” scalar descending from the 5d gauge field or, more generally, the 4d scalar descending from the Ramond-Ramond  $C_{p+1}$ -form field dimensionally reduced on the appropriate  $(p+1)$  cycle.

For us, the above prelude serves only to motivate the following model theory of *generic* (or *fundamental*) instantons: It is defined by the partition function

$$Z = \int D\phi D\theta e^{-S[\phi, \theta]} \sum_{n=1}^{\infty} \sum_{\vec{n}=1}^{\infty} \frac{1}{n! \vec{n}!} \prod_{i=1}^n \left( \int d^4 x_i M^4 e^{-S_I + i\theta(x_i)} \right) \prod_{\bar{i}=1}^{\vec{n}} \left( \int d^4 x_{\bar{i}} M^4 e^{-S_I - i\theta(x_{\bar{i}})} \right), \quad (5)$$

which can of course be extended to a prescription for calculating Greens functions in the usual way. In this theory, the instantons are fundamental, zero-dimensional objects coupling to the axion-like field (just axion from now on) in the mathematically natural way: The axion is interpreted as a zero-form gauge potential which simply has to be evaluated at the position of the charged object (in the stringy language a D(−1) brane)<sup>3</sup>. Furthermore,  $\phi$  stands for all other fields in the model and  $S_I$  is the instanton action. It arises (together with the typical instanton scale  $M$ ) as the tunneling suppression factor  $M^4 \exp(-S_I)$ , which can also be interpreted as the instanton density.

Famously, the instanton and anti-instanton sum exponentiate and the two exponents involving  $\theta$  combine to produce a cosine. This gives

$$Z = \int D\phi D\theta \exp \left( -S[\phi, \theta] + \int d^4 x 2M^4 e^{-S_I} \cos(\theta(x)) \right). \quad (6)$$

We emphasize that, apart from possible corrections to the dilute gas approximation, this is exact. Furthermore, it can be easily extended to situations in which the instantons couple, in addition to the necessary topological coupling to the zero-form  $\theta$ , to other fields. For example,  $S_I$  may depend on the background values of some of the degrees of freedom denoted by  $\phi$ .

## 2.2. Giddings-Strominger Solution

At the end of the previous section, we advertised the point of view that instantons coupled to axions are a limiting case of the general concept of a  $p$ -form gauge theory: In this case  $p = 0$  and the charged object is zero-dimensional. By analogy to the gauge theory, one then expects the existence of objects akin to black branes. In other words, there might exist purely gravitational solutions charged under the axion which represent

<sup>3</sup>Note that this coupling remains imaginary even in the euclidean formulation. A pragmatic way to see this is to recall that  $\theta$  is introduced as a periodic variable. A possibly deeper way is to think of instantons as tunneling events in the lorentzian theory and of  $\exp(i\theta)$  as a relative phase between initial and final state. The latter is of course not affected by Wick rotation.

the continuation of instantons into the high-mass (or high-tension) regime.

An object which fulfills such an expectation at least partially is the Giddings-Strominger wormhole (Giddings and Strominger, 1988a), sometimes also referred to as a gravitational instanton. It is based on the euclidean action ( $M_P = 1$ )

$$S = \int d^4x \sqrt{g} \left( -\frac{1}{2}R + \frac{f^2}{2} g^{\mu\nu} \partial_\mu \theta \partial_\nu \theta \right). \quad (7)$$

Equivalently, one can use the dual formulation in terms of a 2-form gauge theory with field strength  $H_3 = dB_2$ :

$$S = \int d^4x \sqrt{g} \left( -\frac{1}{2}R + \frac{1}{2f^2} \frac{1}{3!} H_{\mu\nu\kappa} H^{\mu\nu\kappa} \right). \quad (8)$$

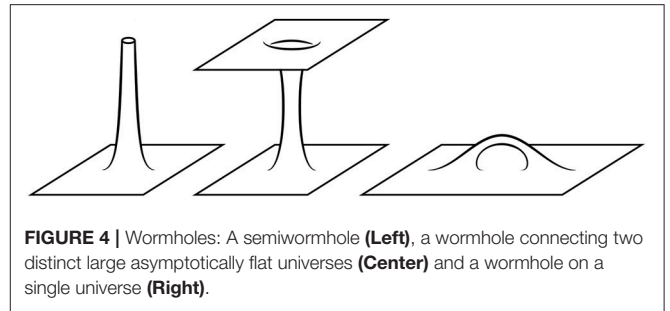
At the classical level, the duality relation is simply  $H = f^2 * d\theta$ . However, the equivalence of the two theories extends, of course, to the full quantum systems. To see this, the dualization must be done under the path integral and care must be taken to get the signs of the kinetic terms right. The outcome is that, both in the euclidean and in the lorentzian versions, the fields have standard (non-ghostlike) kinetic terms on both sides of the duality (see Burgess and Kshirsagar, 1989; Collinucci, 2005; Bergshoeff et al., 2006; Arkani-Hamed et al., 2007b; Hebecker et al., 2017 for details). The wormhole solution to be discussed momentarily exists only in the euclidean theory, but both in the 0-form and 2-form formulation. However, while the  $B_2/H_3$  fields are real, the corresponding values of  $\theta/d\theta$  are imaginary.

Now, the relevance of an “instanton-like” euclidean solution is, of course, that it defines a saddle point of the path integral and hence a very specific, easily quantifiable contribution to the partition function. For the  $B_2$  path integral, the Giddings-Strominger saddle point is then right in the standard integration domain, i.e., “on the real axis” of field space. By contrast, in the  $\theta$  path integral the corresponding saddle point is “on the imaginary axis,” requiring the deformation of the contour and raising the question whether such complex saddles contribute. Complex saddles are certainly known to contribute in certain cases (for a toy model relevant to the present setting see Arkani-Hamed et al., 2007b). Thus, while we favor the (real)  $B_2$  formulation for obvious reasons in what follows, there is nothing wrong in principle with the  $\theta$  formulation<sup>4</sup>.

After these preliminaries, let us describe the solution (Giddings and Strominger, 1988a). It can be motivated by starting from a field theory instanton and including gravitational backreaction: If an instanton couples to an axion  $\theta$ , the dual theory carries non-zero 3-form flux,

$$\int_{S^3} H = n, \quad n \in \mathbb{Z}, \quad (9)$$

<sup>4</sup>Occasionally, the impression is raised that the  $\theta$  formulation requires a wrong-sign kinetic term if one wants the wormhole solution to exist. While this perspective might technically be equivalent to what was said above, we find it conceptually misleading. In our reading, one studies a well-defined physical theory without ghost fields. It is only the desire to estimate the contribution from a certain complex saddle which leads one to work with imaginary  $\theta$  temporarily.



**FIGURE 4 |** Wormholes: A semiwormhole (Left), a wormhole connecting two distinct large asymptotically flat universes (Center) and a wormhole on a single universe (Right).

on any sphere containing  $n$  instantons (or an instanton of charge  $n$ ). Placing the instanton(s) at the origin and assuming spherical symmetry, it is immediately clear that one must have

$$H = \frac{n\epsilon}{2\pi^2}. \quad (10)$$

Here  $\epsilon$  is defined as the volume form of  $S^3$  in the description of  $\mathbb{R}^4$  as  $\mathbb{R}_+ \times S^3$ .

The above  $H$  automatically satisfies the Bianchi identity  $dH = 0$  and the equation of motion  $d * H = 0$  (for any spherically symmetric metric). It induces a non-zero energy momentum tensor and the corresponding Einstein equation is solved by

$$ds^2 = \left(1 + \frac{C}{r^4}\right)^{-1} dr^2 + r^2 d\Omega_3^2, \quad C = -\frac{n^2}{24\pi^4 f^2}. \quad (11)$$

Here  $d\Omega_3^2$  denotes the round metric on the unit sphere.

This geometry is asymptotically flat for  $r \rightarrow \infty$  and has a coordinate singularity at  $r = r_0 \equiv |C|^{1/4}$ . The space given by restricting  $r \in [r_0, \infty)$  forms what is often termed a semiwormhole (see Figure 4). Gluing two such solutions at the 3-spheres defined by  $r = r_0$ , one obtains a smooth wormhole connecting two flat universes (see Figure 4). A topologically distinct, approximate solution can be obtained if the two asymptotically flat regions of Figure 4 are interpreted as distant parts of the same universe—cf. Figure 4. One then has a wormhole joining two regions of the same large universe. This becomes exact in the limit that the two wormhole ends are infinitely far apart.

The wormhole action is particularly easy to compute using the trace of the Einstein equation:

$$S_w = \frac{1}{f^2} \int H \wedge *H = \frac{n^2}{2\pi^2 f^2} 2 \int_{r_0}^{\infty} \frac{dr}{r\sqrt{r^4 + C}} = \frac{\pi\sqrt{6}}{4} \cdot \frac{|n|}{f}. \quad (12)$$

Notice the factor 2 appearing because a wormhole consists of two solutions of the form of (11), each restricted to  $r > r_0$ .

The most straightforward interpretation of this is as follows: Suppressed by an overall factor  $\exp(-S_w)$ , the partition function includes processes in which an  $S^3$  baby universe supported by  $H_3$ -flux “bubbles off” at some space-time point  $x$  and is absorbed later on at  $y$  ( $x, y \in \mathbb{R}^4$ ). From the low-energy perspective, this is equivalent to an instanton (of charge  $n$  and action  $S_w/2 \sim |n|/f$ ) at  $x$  and a corresponding anti-instanton at  $y$ . Computational

control in semiclassical gravity requires  $r_0 \sim \sqrt{|n|/f} \gg 1$ . This should then give rise to a cosine potential for  $\theta$  and further instanton-induced operators. It has, however, been argued that, in contrast to the instantonic situation, no such potential is induced because of the unavoidable pairing of instantons and anti-instantons (Heidenreich et al., 2016). Counterarguments have been given (Hebecker et al., 2017), based essentially on the intuition that local physics is ignorant of the overall constraint on instantons vs. anti-instantons in a very large space-time (recall that the action stays finite as  $|x - y| \rightarrow \infty$ ). However, this debate is overshadowed by a much deeper issue which will permeate the rest of this review: Once one allows for wormholes, one has effectively allowed for baby-universes propagating between points  $x$  and  $y$ . But then such baby universes must also be allowed to be part of the initial and final states of any process. More generally, there exists a “baby-universe state” in addition to our space-time and any wormhole effects (such as the naive cosine potential) depend on it.

### 2.3. Dilatonic Wormholes

Before coming to the physical effects of wormholes and baby universes, we want to briefly comment on generalizations of the Giddings-Strominger solution which involve a dilaton (Giddings and Strominger, 1988a, 1989b; Bergshoeff et al., 2006; Heidenreich et al., 2016; Hebecker et al., 2017). This is important since such dilatons are always present in the simplest stringy models allowing for wormholes.

Consider an action in which the axionic kinetic term depends on a further massless scalar field  $\phi$ ,

$$S = \int d^4x \sqrt{g} \left( -\frac{1}{2}R + \frac{1}{2}\mathcal{K}(\phi)g^{\mu\nu}\partial_\mu\theta\partial_\nu\theta + \frac{1}{2}g^{\mu\nu}\partial_\mu\phi\partial_\nu\phi \right), \quad (13)$$

or equivalently

$$S = \int d^4x \sqrt{g} \left( -\frac{1}{2}R + \frac{1}{2}\mathcal{F}(\phi)H_{\mu\nu\kappa}H^{\mu\nu\kappa} + \frac{1}{2}g^{\mu\nu}\partial_\mu\phi\partial_\nu\phi \right), \quad (14)$$

with  $\mathcal{F} \equiv 1/(3!\mathcal{K})$ . As before, spherical symmetry ensures that the equation of motion for  $H$  is automatically satisfied. A new, non-trivial differential equation for the radial profile of  $\phi$  arises. Remarkably, the differential equation for  $g_{rr}$  (the only non-trivial part of the Einstein equation) decouples and the metric (11) remains a solution<sup>5</sup>. We will not discuss the solution  $\phi(r)$  in any detail. It is, however, interesting to note that, switching from  $H_3/B_2$  to  $d\theta/\theta$  for the moment, the common trajectory  $\{\phi(r), \theta(r)\}$  describes a geodesic in field space. This generalizes to the case of several axionic and several non-axionic scalars (cf. Footnote 5 and Arkani-Hamed et al., 2007b).

<sup>5</sup>In fact, the metric (11) solves the equations of motion of the more general action

$$S = \int d^4x \sqrt{g} \left( -\frac{1}{2}R + \frac{1}{2}G_{IJ}(\phi)g^{\mu\nu}\partial_\mu\phi^I\partial_\nu\phi^J \right), \quad (15)$$

where a set moduli  $\phi^I$  and a (non-positive-definite) metric  $G_{IJ}$  on moduli space have been introduced (Arkani-Hamed et al., 2007b).

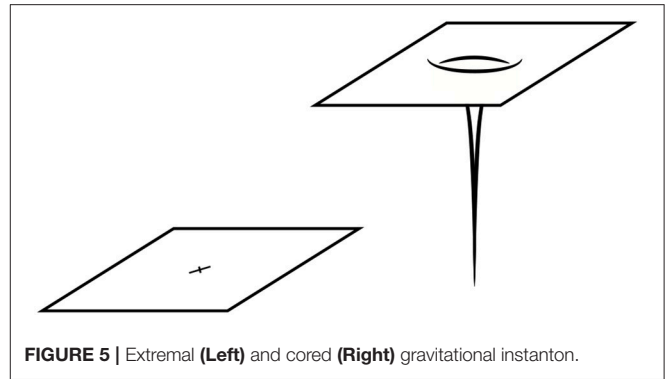


FIGURE 5 | Extremal (Left) and cored (Right) gravitational instanton.

Motivated by stringy and supergravity examples, we now restrict attention to the special case

$$\mathcal{F}(\phi) = \frac{1}{3!f^2} \exp(-\beta\phi). \quad (16)$$

Without loss of generality one can assume  $\beta \geq 0$ . Three different classes of solutions can be distinguished: First, as long as  $\beta < 2\sqrt{2/3}$ , the Giddings-Strominger wormhole continues to exist (metric of 11 with  $C < 0$ ). This is the case of our main interest. Second, there is the extremal gravitational instanton, corresponding to  $C = 0$ . The geometry is a flat space-time with the origin removed, but  $\phi$  diverges as one approaches  $r = 0$ . Third, there are “cored gravitational instantons,” corresponding to  $C < 0$ . In this case one has a curvature singularity at  $r = 0$  (cf. Figure 5). The last two cases have the significant drawback that they are not fully controlled within the low-energy effective theory and we will hence not discuss them further (see however Bergshoeff et al., 2006; Heidenreich et al., 2016; Hebecker et al., 2017).

In the simplest (usually highly supersymmetric) string compactifications, axions are always accompanied by a dilatonic scalar or saxion, as above. However, the simplest models do not allow for  $\beta < 2\sqrt{2/3}$ . Naively, one may then hope that wormholes do not arise in consistent theories of quantum gravity. But it turns out that the problem with the allowed  $\beta$  range can be overcome (Tamvakis, 1989; Bergshoeff et al., 2005; Arkani-Hamed et al., 2007b; Hertog et al., 2017). The underlying idea is simple: A wormhole can be charged under several axions, each with its own saxion with a certain  $\beta$ . The trajectory which the solution follows in the saxionic field space involves all the axions and can be characterized by a single effective  $\beta$ . The latter can be in the desired range even if the  $\beta$ -values of the ingredients were not<sup>6</sup>. Thus, one can by now be certain that Giddings-Strominger wormholes exist in the euclidean version of supergravity theories coming from string theory. This makes all the puzzles to be discussed below even more troubling<sup>7</sup>.

<sup>6</sup>The necessary condition for the existence of wormholes and the way in which multiple axions help to satisfy it can also be discussed in the language of time-like geodesics in the axion/saxion field space, cf. Footnote 5.

<sup>7</sup>A simpler but less rigorous argument that wormholes are “not in the swampland” can be given as follows: Surely somewhere in the string theory landscape there

### 3. THE EFFECT OF WORMHOLES

Two results of the previous section are essential for what follows. First, a dilute gas of instantons can be resummed (or “integrated out”) to obtain a correction to the effective action. Second, a very similar contribution to the path integral arises in gravitational theories with an axion. The objects to be summed over are wormholes or gravitational instantons. The main novelty is that they couple to the low-energy degrees of freedom (including the background metric) at two spacetime points rather than just at one. We now want to discuss the correction to the effective action arising in this second case following (Coleman, 1988; Preskill, 1989; Coleman and Lee, 1990b). We note that, while the specific Giddings-Strominger solution discussed above may be the simplest and best understood euclidean wormhole, the following analysis does not rely on any of its details. What matters is that the euclidean path integral includes contributions from topologies like that of **Figure 4** (on the right). All that we will use is that they are exponentially suppressed by a sufficiently large euclidean action and that the coupling to soft field modes occurs at two uncorrelated points (Hawking, 1988, 1990b).

#### 3.1. The Bilocal Action

We begin with a heuristic derivation of the bilocal action which captures the effect of wormholes at the semi-classical level. For this, we first recall the field theoretic partition function with instantons, Equation (5), and restrict it to the one-instanton sector for notational simplicity:

$$Z_1 = \int D\phi D\theta e^{-S[\phi, \theta]} \left( \int d^4x e^{-S_I + i\theta(x)} \right). \quad (17)$$

Here the prefactor  $M^4$  has been reabsorbed in the instanton action. (To be careful, one should then either work in Planck units or at least choose  $x$  dimensionless.)

The above is unnecessarily explicit in that  $\theta$  has been separated from all the other fields  $\phi$ . At the same time, it is oversimplified in that only the dependence of the instanton action on  $\theta$  has been kept:  $S_I[\theta] \equiv S_I + i\theta$ . A more general version, in which  $\theta$  is just one of the many fields denoted by  $\phi$ , reads

$$Z_1 = \int D\phi e^{-S[\phi]} \left( \int d^4x e^{-S_I[x, \phi]} \right). \quad (18)$$

Here  $S_I[x, \phi]$  is the single-instanton tunneling action for the space-time point  $x$  in a background field  $\phi$ . It is clear that obtaining this action in a concrete model is highly non-trivial: One would have to find the analog of the well-known instanton or wrapped-euclidean-brane solution in an, in general non-constant, background of all fields in the theory. However, we are satisfied with an approximation: the fields  $\phi$  are restricted to be

exists a low-energy effective theory containing an ungauged abelian Higgs model. Clearly, the global  $U(1)$  of this model will not be exact. The resulting effective axion will thus have a non-perturbatively generated cosine potential. This potential is in general exponentially suppressed and hence very small. The saxion, i.e., the radial direction of the complex Higgs scalar, is stabilized. Thus, wormholes based on this effective axion will exist.

soft relative to the instanton scale  $M$ . The action can then be expanded in terms of local operators:

$$S_I[x, \phi] = S_I + c_1\phi(x) + c_2\phi^2(x) + c_3(\partial\phi(x))^2 + \dots \quad (19)$$

Here  $S_I$  is the instanton action on the unperturbed background, say at  $\phi \equiv 0$ . With this, the transition to wormholes is simple.

Indeed, the wormhole analog of (18) is

$$Z_{1,w} = \int Dg D\phi e^{-S[g, \phi]} \left( \int d^4x \sqrt{g(x)} \int d^4y \sqrt{g(y)} e^{-S_w[x, y, g, \phi]} \right). \quad (20)$$

Here  $\int Dg$  stands for the integral over all soft (relative to the wormhole size) metrics on the topologically trivial background universe into which the wormhole is inserted. In addition,  $\phi$  stands for all further fields, including the axion or the dual 2-form<sup>8</sup>. As before, appealing to our restriction to soft fields and metric configurations, the wormhole action can be written as a series of local operators at  $x$  and  $y$ :

$$S_w[x, y, g, \phi] = S_w + c_1\phi(x) + c_1\phi(y) + c_3\phi(x)\phi(y) + \dots + c_4(\partial\phi(x))^2(\partial\phi(y))^2 + \dots \quad (21)$$

For simplicity terms depending on a non-trivial metric background have not been displayed. It is clear that such terms, involving various curvature invariants at  $x$  and at  $y$  as well as products thereof, will also be present. The crucial novelty compared to the instanton case is that one is dealing with a double functional Taylor expansion and that products of local operators involving all fields will in general arise. Thus, one generically has the bilocal expression

$$S_w[x, y, g, \phi] = S_w + \sum_{ij} \tilde{\Delta}_{ij} \mathcal{O}_i(x) \mathcal{O}_j(y), \quad (22)$$

or, equivalently,

$$e^{-S_w[x, y, g, \phi]} = \frac{1}{2} \sum_{ij} \Delta_{ij} \mathcal{O}_i(x) \mathcal{O}_j(y). \quad (23)$$

In the last expression, the exponential  $\exp\left(\sum_{ij} \tilde{\Delta}_{ij} \mathcal{O}_i \mathcal{O}_j\right)$  has been expanded and the suppression factor  $\exp(-S_w)$  has been absorbed in the new coefficients  $\Delta_{ij}$ :

$$\Delta_{ij} \sim e^{-S_w}. \quad (24)$$

Finally, one inserts (23) in (20) and writes down analogous expressions for any number of wormholes. In doing so, the dilute gas approximation is used, i.e., that typical distances between wormhole ends are much larger than the wormhole diameter. The sum exponentiates, exactly as in the instanton case, giving

$$Z_w = \int Dg D\phi e^{-S[g, \phi] + I} \quad (25)$$

<sup>8</sup>If one uses specifically the Giddings-Strominger solution, the value of the axion corresponds to the one far away from the wormhole. The fast change of the axion inside the throat is not part of what we want to call the background field.

with the bilocal action

$$I = \frac{1}{2} \int d^4x \sqrt{g} \int d^4y \sqrt{g} \sum_{ij} \Delta_{ij} \mathcal{O}_i(x) \mathcal{O}_j(y). \quad (26)$$

### 3.2. Local Action Involving $\alpha$ Parameters

Following Coleman (1988c) and Preskill (1989), one can give the action  $I$  a local form at the expense of introducing a set of auxiliary parameters  $\alpha_i$ . Up to some irrelevant normalization factor, one has

$$e^I = \prod_i \left( \int d\alpha_i \right) \exp \left( -\frac{1}{2} \sum_{ij} \alpha_i \Delta_{ij}^{-1} \alpha_j \right) \exp \left( \sum_i \alpha_i \int d^4x \sqrt{g} \mathcal{O}_i(x) \right). \quad (27)$$

It is natural to write the original action  $S$  of our physical system using the basis of local operators as in the wormhole action  $I$  above:

$$S[g; \lambda] = \sum_i \lambda_i \int d^4x \sqrt{g} \mathcal{O}_i(x). \quad (28)$$

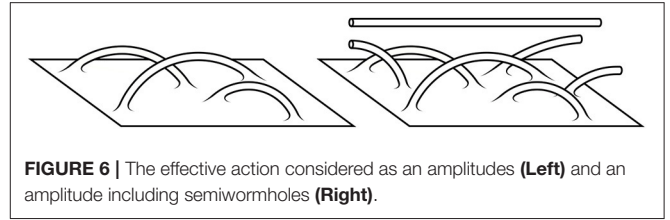
Here  $\lambda_i$  are the coupling constants. For example,  $\lambda_1$ ,  $\lambda_2$ , and  $\lambda_3$  could be the cosmological constant, the coefficient of the Einstein-Hilbert term, and of the  $R^2$ -term, respectively. To minimize the notational complexity, we suppress the dependence on the non-metric fields  $\phi$  here and below. Of course, all of the above holds with as many further fields as one needs.

Comparing (27) and (28), one sees that the effect of wormholes amounts to shifting the coupling constants of the original action:  $\lambda_i \rightarrow \lambda_i - \alpha_i$ . Put differently, one can use the “shifted” action  $S[g; \lambda - \alpha]$ , remembering of course to integrate over the  $\alpha$  parameters. The partition function with wormhole effects included (see Equation 25 and recall that we suppress  $\phi$ ) now reads

$$\begin{aligned} Z_w &= \int Dg e^{-S[g; \lambda] + I[g]} \\ &= \int Dg D\alpha G(\alpha) e^{-S[g; \lambda - \alpha]} \\ &= \int D\alpha G(\alpha) \left[ \int Dg e^{-S[g; \lambda - \alpha]} \right], \end{aligned} \quad (29)$$

with  $G(\alpha) = \exp \left( -\frac{1}{2} \sum_{ij} \alpha_i \Delta_{ij}^{-1} \alpha_j \right)$  the gaussian weighting factor. In the above, we also use the somewhat sloppy notation  $D\alpha$  for the integration over all  $\alpha_i$ , in spite of the fact that the index  $i$  is discrete.

In the last expression in (29), one recognizes the familiar partition function without wormholes inside the square brackets. The wormhole effect is reduced to shifting the coupling constants of that theory by  $\alpha_i$ . Since these  $\alpha$  parameters are constants in space and time, one can take the point of view that they simply have to be measured and no relevance should be ascribed to the gaussian weight factor governing their distribution. By contrast, one may argue that statistical predictions for their values are



**FIGURE 6** | The effective action considered as an amplitudes (Left) and an amplitude including semiwormholes (Right).

possible, which of course involves this weight factor. This is a multiverse-type situation, discovered (and discussed by many authors) long before the string theory multiverse entered the stage.

### 3.3. Baby Universes

The physics behind  $\alpha$  parameters becomes more lucid if one thinks of the wormholes in terms of  $S^3$  baby universes which are emitted and absorbed by our macroscopic space-time (left hand side of Figure 6). To derive the corresponding formulae, one considers the situation with a single operator and hence a single  $\alpha$  parameter for notational simplicity. Equation (27) then reads

$$e^I = \int \frac{d\alpha}{\sqrt{2\pi}} \exp \left( -\frac{1}{2} \alpha^2 + \alpha \sqrt{\Delta} \int d^4x \sqrt{g} \mathcal{O}(x) \right), \quad (30)$$

obtained after rescaling  $\alpha \rightarrow \alpha \sqrt{\Delta}$  and introducing the normalization factor  $1/\sqrt{2\pi}$  for later convenience.

Equation (30) can be viewed as a power series in  $\mathcal{O}(x)$  encoding the sum of process in which baby universes are created and annihilated at locations corresponding to the various values taken by  $x$ . All of this has of course to be inserted under the  $Dg$  integral over soft background metrics. To make this manifest, one defines baby universe creation and annihilation operators  $a^\dagger$ ,  $a$  satisfying the usual commutation relation  $[a, a^\dagger] = 1$ . The state with no baby universes  $|0\rangle$  is referred to as the baby universe vacuum. The normalized state with  $n$  baby universes is then given by

$$|n\rangle = \frac{(a^\dagger)^n}{\sqrt{n!}} |0\rangle. \quad (31)$$

The analogs of the conventional position operator of the harmonic oscillator and its eigenstates are defined as

$$\hat{a} = a + a^\dagger, \quad \hat{a}|\alpha\rangle = \alpha|\alpha\rangle. \quad (32)$$

Since the ground state obeys  $|0\rangle \sim \int d\alpha \exp(-\alpha^2/4)|\alpha\rangle$ , one immediately sees that

$$\langle 0 | (a + a^\dagger)^n | 0 \rangle = \int \frac{d\alpha}{\sqrt{2\pi}} \exp \left( -\frac{1}{2} \alpha^2 \right) \alpha^n. \quad (33)$$

This allows one to rewrite (30) according to

$$e^I = \int \frac{d\alpha}{\sqrt{2\pi}} \exp \left( -\frac{1}{2} \alpha^2 \right) \exp \left( \alpha \tilde{\mathcal{O}} \right) = \langle 0 | e^{(a+a^\dagger)\tilde{\mathcal{O}}} | 0 \rangle, \quad (34)$$

where  $\tilde{\mathcal{O}}$  is an abbreviation for  $\tilde{\mathcal{O}} = \sqrt{\Delta} \int d^4x \sqrt{g} \mathcal{O}(x)$ .

Equation (34) can be considered a convenient formal expression for a power series in  $\tilde{\mathcal{O}}$ . But it is much more than that: It formalizes the interpretation of the partition function and of the process depicted on the left hand side of **Figure 6** in terms of a baby universe Hilbert space. Equation (34) calculates the amplitude relating two spatial slices of the parent universe, allowing for any number of wormholes to be inserted between initial and final time.

The most important point here is that, in this approach, it is both easy and obviously necessary to allow for more general initial and final states: There is simply no reason to treat those as baby universe vacua. For example, one can also consider the transition amplitude between states with  $n_1$  and  $n_2$  baby universes:

$$\langle n_2 | e^{(a+a^\dagger)\tilde{\mathcal{O}}} | n_1 \rangle. \quad (35)$$

In fact, arbitrary states  $\psi_1$  and  $\psi_2$  can be considered, another relevant case being that of so-called  $\alpha$ -vacua:

$$\langle \alpha | e^{(a+a^\dagger)\tilde{\mathcal{O}}} | \alpha \rangle = e^{\alpha \tilde{\mathcal{O}}}. \quad (36)$$

Here we ignore the divergent prefactor related to the  $\delta$ -function normalization of “momentum eigenstates.”

It is easy to see that, for an arbitrary number of operators and arbitrary initial and final states, the above amplitude generalizes to

$$\langle \psi_2 | \exp \left( \sum_i \sqrt{\Delta_{ii}} \int d^4x \sqrt{g} \mathcal{O}_i(x) (a_i + a_i^\dagger) \right) | \psi_1 \rangle. \quad (37)$$

Here, a basis of local operators has been chosen such that the matrix  $\Delta_{ij}$  is diagonal. The  $a_i^\dagger$  and  $a_i$  carry the same index as the local operators and create or annihilate baby universes of type  $i$ . If everything is based on the Giddings-Strominger solution of lowest charge, one may think of these baby universes as of transverse spheres  $S^3$  in a perturbed wormhole geometry (or some appropriate quantum superposition thereof).

The Hamiltonian  $(a + a^\dagger)\tilde{\mathcal{O}}$  was first derived by Coleman (1988a) by summing explicitly over all possible wormhole and semiwormhole configurations. For completeness, we now briefly explain this computation, for the case of a single type of wormhole for simplicity. Consider a 4-manifold  $M$  of the type shown in the right hand side of **Figure 6**. The initial boundary consists of a large 3-manifold parent universe and  $n_1$  incoming baby universes. Of those,  $n_1 - r$  later on merge with  $M$ . The final boundary consists again of a large 3-manifold and  $n_2$  outgoing baby universes,  $n_2 - r$  of which emerged from  $M$ . Thus,  $r$  baby universes simply travel from the initial to the final boundary without interacting with the parent universe. Furthermore  $m$  baby universes form complete wormholes on  $M$ . The path integral sums over all such configuration:

$$\sum_{r,m} e^{-S} |_{n_1, n_2}. \quad (38)$$

As before, one assumes that each semiwormhole attached to the parent universe contributes a factor  $\tilde{\mathcal{O}} = \sqrt{\Delta} \int d^4x \sqrt{g} \mathcal{O}(x)$ .

Taking into account the combinatorics and carrying out the summation over  $m$  yields

$$\begin{aligned} \sum_{r,m} e^{-S} |_{n_1, n_2} &= \sqrt{n_1!} \sqrt{n_2!} e^{\tilde{\mathcal{O}}^2/2} \sum_{r=0}^{\min(n_1, n_2)} \frac{\tilde{\mathcal{O}}^{n_1+n_2-r}}{(n_1-r)! (n_2-r)! r!} \\ &= \langle n_2 | e^{(a^\dagger+a)\tilde{\mathcal{O}}} | n_1 \rangle. \end{aligned} \quad (39)$$

Here the second equality follows by applying Baker-Campbell-Hausdorff in the form  $\exp[(a + a^\dagger)\tilde{\mathcal{O}}] = \exp(a^\dagger\tilde{\mathcal{O}}) \exp(a\tilde{\mathcal{O}}) \exp(\tilde{\mathcal{O}}^2/2)$  and inserting the identity operator written as a sum over  $|r\rangle\langle r|$ . Thus, the language of  $a$  and  $a^\dagger$  introduced earlier is nothing but a convenient way of counting wormhole topologies.

### 3.4. The Perspective of $\alpha$ -vacua and the Wormhole Density

It is clear that the appearance of  $\alpha$  parameters in the path integral has the potential to change physics dramatically: Since these parameters are space-time independent, the whole universe (including its time evolution) can be thought of as a superposition of independent universes, each with a specific set of fixed  $\alpha$  parameters.

This has become even more apparent in the last subsection, when the baby universe state characterized by the  $\alpha$  parameters was introduced. Since all effective operator coefficients or couplings are shifted according to  $\lambda_i \rightarrow \lambda_i - \alpha_i$ , the baby universe state determines the 4d low-energy effective field theory. A whole landscape of such theories, equivalent to the space of  $\alpha$ -vacua, exists. At this level, every hope of predicting coupling constants from some fundamental theoretical principle appears to be lost.

The situation might not be, however, quite as bad: for transitions among baby universe vacua an integral over the  $\alpha$  parameters with a very specific measure arises. This makes sense in a compact euclidean universe, for example for a large 4-sphere (or a set of large 4-spheres), where no initial or final baby universe state is required. Specifically a 4-sphere geometry is reminiscent of the Hartle-Hawking definition (Hartle and Hawking, 1983) of the Wheeler-DeWitt wave function of the universe (DeWitt, 1967; Wheeler, 1967). Thus, one may think of the integral over  $\alpha$  parameters (with the concrete measure derived earlier) as of a preferred wave function of the baby universe state. This point of view allows for at least a statistical prediction of effective coupling constants.

It is essential that the  $\alpha$ -parameters are eigenvalues of the Hamiltonian governing the interaction of our large-scale 4d world with the baby-universe state. This was derived above and it can also be seen intuitively: one can not distinguish in principle whether a wormhole attached at a given position  $x$  corresponds to a baby universe being absorbed or being created. Hence one always encounters the combination  $(a + a^\dagger) \mathcal{O}(x) \equiv \hat{\alpha} \mathcal{O}(x)$  in the effective Hamiltonian. When an operator coefficient is measured, one is projected to a subsector of the theory belonging to a certain eigenvalue  $\alpha$ . Further dynamical evolution can not change this value.

As a consistency check one can estimate the density of wormhole ends following Preskill (Preskill, 1989). This is crucial

to understand the validity of the dilute gas approximation. Returning to the perspective of the bilocal action, section 3.1, one can focus on a single operator, the cosmological constant. According to (26), the effect of wormhole insertions is then encoded in

$$e^I \sim e^{\frac{1}{2} V_4^2 \Delta} = \sum_{n=0}^{\infty} \frac{1}{n!} \left( \frac{1}{2} V_4^2 \Delta \right)^n, \quad (40)$$

where  $V_4 = \int d^4x \sqrt{g(x)}$  and  $\Delta \sim e^{-S_w}$ . Here the  $n$ -th order term corresponds to  $n$  wormholes. The dominant contribution to the sum comes from terms with  $n \sim N_w \equiv \frac{1}{2} V_4^2 \Delta$ , such that the wormhole density in typical configurations is

$$\frac{N_w}{V_4} = \frac{1}{2} V_4 \Delta. \quad (41)$$

One arrives at the disturbing conclusion that this density grows with the volume  $V_4$ .

Fortunately, the result changes if one considers physics at fixed  $\alpha$ . According to (36) the wormhole sum is now encoded in

$$e^{\alpha V_4 \sqrt{\Delta}} = \sum_{n=0}^{\infty} \frac{1}{n!} (\alpha V_4 \sqrt{\Delta})^n. \quad (42)$$

This sum is dominated by terms with  $n \sim N_{w,\alpha} \equiv \alpha V_4 \sqrt{\Delta}$ . A non-divergent density of wormhole ends in spacetime follows:

$$\frac{N_{w,\alpha}}{V_4} = \alpha \sqrt{\Delta}. \quad (43)$$

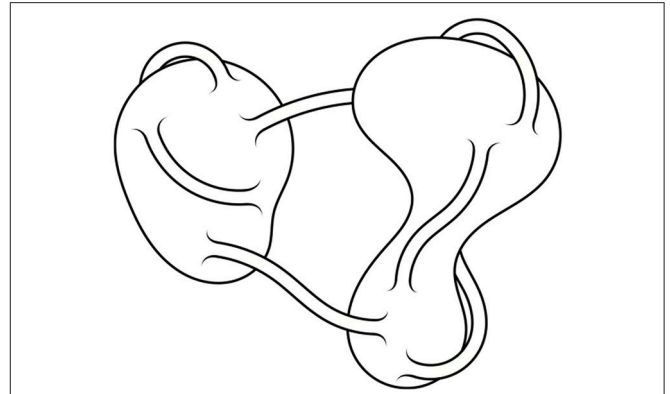
Thanks to the suppression factor  $\sqrt{\Delta} \sim e^{-S_w/2}$ , this density is expected to be very small for large wormholes with a correspondingly large euclidean action. The problem encountered above in the vacuum-to-vacuum amplitude,  $|0\rangle \rightarrow |0\rangle$ , appears to have been resolved. Technically, the reason is that the sum has been re-organized by combining events where a wormhole is absorbed and created at the same point:  $a, a^\dagger \rightarrow (a + a^\dagger)$ . However, together with the suppression factor  $e^{-S_w/2}$  comes, of course, the unknown parameter  $\alpha$ . In the integration over  $\alpha$ , the problem of an overdensity of wormhole ends can in principle reappear. This is the subject of section 5.1

### 3.5. Multiple Large Universes

Only the case of one large parent universe with many small-scale wormholes attached has been considered so far. It is, however, completely logical to allow for multiple large universes. Wormholes can connect one large universe to itself or to another large universe, cf. **Figure 7**. When all wormholes are integrated out, the large universes become disconnected.

Following Fischler and Susskind (1989) and Preskill (1989), one can single out one particular large universe and consider the expectation value of an observable  $\mathcal{A}(x)$  in that universe. Keeping the values of the  $\alpha$  parameters (which are common to all large universes) fixed for the moment, one has

$$\langle \mathcal{A}(x) \rangle_\alpha = \int Dg_d e^{-S[g_d; \lambda - \alpha]} \mathcal{A}(x). \quad (44)$$



**FIGURE 7** | Large universes connected by wormholes—figure adapted from Fischler and Susskind (1989).

Here  $Dg_d$  (with “d” for disconnected) stands for the integration over all large-scale metrics, including a summation over manifolds with many components. Making this summation over the number of disconnected components explicit,

$$\begin{aligned} \langle \mathcal{A}(x) \rangle_\alpha &= \sum_N \frac{1}{N!} \prod_{n=0}^N \left( \int Dg_n e^{-S[g_n; \lambda - \alpha]} \right) \int Dg e^{-S[g; \lambda - \alpha]} \mathcal{A}(x) \\ &= \exp \left( \int Dg' e^{-S[g'; \lambda - \alpha]} \right) \int Dg e^{-S[g; \lambda - \alpha]} \mathcal{A}(x). \end{aligned} \quad (45)$$

Here, in the first line,  $g$  is the metric on the distinguished large universe and  $g_n$  are the metrics on the other disconnected components. The second line used the fact that the sum over disconnected geometries exponentiates, introducing the variable  $g'$  for the metric on a generic such component.

Reinstating the  $\alpha$ -integration gives

$$\langle \mathcal{A}(x) \rangle = \int Dg D\alpha G(\alpha) P(\alpha) e^{-S[g; \lambda - \alpha]} \mathcal{A}(x), \quad (46)$$

with the probability distribution

$$P(\alpha) = \exp \left( \int Dg e^{-S[g; \lambda - \alpha]} \right). \quad (47)$$

In the calculation of the partition function, i.e., without the insertion of a local operator, no connected component is singled out. The sum over topologies then exponentiates without the need to split of one of the factors:

$$Z = \int D\alpha G(\alpha) P(\alpha). \quad (48)$$

As discussed later, the double exponential  $P(\alpha)$  is responsible both for the initial excitement in wormhole physics (Coleman’s solution to the cosmological constant problem Coleman, 1988c) as well as for a particularly serious conceptual problem (the FKS

catastrophe Fischler and Susskind, 1989; Kaplunovsky, unpublished).

## 4. PHENOMENOLOGICAL APPLICATIONS

### 4.1. Random Values of Couplings and the Cosmological Constant Problem

If one accepts that euclidean wormholes contribute to the path integral, one may clearly be concerned that all of the familiar local physics will break down. The reason is that the action of the wormhole contributions does not grow with the separation of the two points where they attach to our macroscopic spacetime. The possible loss of quantum coherence has also been initially discussed in this context. However, it has quickly been established (Coleman, 1988a; Giddings and Strominger, 1988b) that a local effective field theory description can be recovered by introducing  $\alpha$  parameters in the path integral or, equivalently,  $\alpha$  vacua in the canonical approach (cf. the discussion in the last chapter).

The implications of this are nevertheless quite dramatic: All coupling constants depend on the  $\alpha$  vacuum, i.e., on the a priori unknown baby universe state. This state is an unavoidable additional piece of information which has to come on top of the quantum-field-theoretic initial conditions given on a Cauchy surface of our spacetime manifold. By measuring couplings one is effectively determining some of the infinitely many  $\alpha$  parameters. There seems to be no hope of predicting these couplings on the basis of a unique theory of everything, even if the latter was known to us. From a modern point of view, this is of course very similar to the situation which has anyway been widely accepted after the advent of the string theory landscape (Bousso and Polchinski, 2000; Giddings et al., 2002; Kachru et al., 2003; Susskind, 2003; Denef and Douglas, 2004). In fact, both ways of randomizing coupling constants may be at work simultaneously. The familiar deep issue of the measure problem of eternal inflation (the leading candidate mechanism for populating the landscape) has a cousin in the form of the measure on or the dynamics of the baby universe state.

The above situation may be viewed as the generic phenomenological implication of euclidean wormholes or gravitational instantons. For the initial popularity of this paradigm, it was crucial that an apparently very successful attempt was made early on to derive a statistical prediction for one of the couplings—the cosmological constant (Coleman, 1988c) (for early applications of wormholes to other phenomenologically relevant couplings see Grinstein and Wise, 1988; Choi and Holman, 1989; Gilbert, 1989; Nielsen and Ninomiya, 1989; Preskill et al., 1989). In fact, a distribution infinitely peaked at zero was found, making the prediction exact. Subsequently many caveats were discovered such that the “cosmological constant prediction” is not viewed as a central motivation for wormhole physics at present. Nevertheless, because of its intrinsic interest and its immense historical importance we review the argument in the remainder of this subsection (for reviews discussing this as well as other early

approaches to the cosmological constant problem see Weinberg, 1989; Carroll et al., 1992).

The argument is due to Coleman (1988c) and can be given using just the leading terms of the bare gravitational action:

$$S[g] = \int d^4x \sqrt{g} \left( \Lambda - \frac{M_P^2}{2} R + \dots \right) = \int d^4x \sqrt{g} \sum_i \lambda_i \mathcal{O}_i. \quad (49)$$

Here  $\lambda_1 = \Lambda$  and  $\lambda_2 = -M_P^2/2$  characterize the cosmological constant and the Planck scale. As discussed before, including the effects of wormholes and allowing for multiple large parent universes (as in Figure 7) leads to the partition function (cf. Equation 48)

$$Z = \int D\alpha \exp \left( -\frac{1}{2} \sum_{i,j} \alpha_i \Delta_{ij}^{-1} \alpha_j \right) \exp \left( \int Dg \exp \left( -\int d^4x \sqrt{g} \sum_i (\lambda_i - \alpha_i) \mathcal{O}_i \right) \right). \quad (50)$$

Since wormholes have been integrated out, the relevant metric in the above expression refers to a single parent universe. As argued in Coleman (1988c), this expression is dominated by values of  $\alpha$  which correspond to  $\Lambda_{\text{eff}} = \lambda_1 - \alpha_1 > 0$ . Furthermore, the sum over topologies is dominated by spheres. The path integral over metrics can then be estimated in the saddle point approximation:

$$Z = \int D\alpha \exp \left( -\frac{1}{2} \sum_{i,j} \alpha_i \Delta_{ij}^{-1} \alpha_j \right) \exp \left( \exp \left( -\int d^4x \sqrt{g} \left( \Lambda_{\text{eff}} - \frac{M_{P,\text{eff}}^2}{2} R \right) \right) \right), \quad (51)$$

where  $M_{P,\text{eff}}^2 = M_P^2 + 2\alpha_2$  and the sum is restricted to  $i, j = 1, 2$ . Thus, all one needs is the action of a 4-sphere solution with the above effective Planck scale and cosmological constant. Given that a 4-sphere of radius  $r$  has volume  $V_4 = (8/3)\pi^2 r^4$  and scalar curvature  $R = 12/r^2$ , this action is

$$S_{\text{sphere}} = -\frac{24\pi^2 M_{P,\text{eff}}^4}{\Lambda_{\text{eff}}}. \quad (52)$$

This gives

$$Z = \int D\alpha \exp \left( -\frac{1}{2} \sum_{i,j} \alpha_i \Delta_{ij}^{-1} \alpha_j \right) \exp \left[ \exp \left( 96\pi^2 \frac{(M_P^2/2 + \alpha_2)^2}{(\Lambda + \alpha_1)} \right) \right], \quad (53)$$

where  $\alpha_1$  was redefined  $\alpha_1 \rightarrow -\alpha_1$ .

The key point of this result is the double exponential enhancement of the measure governing the  $\alpha$ -parameter integration at the point where the effective cosmological constant vanishes.

As already emphasized, serious caveats exist and the above logic is nowadays generally not considered a valid solution to the cosmological constant problem. First, the measured value of the cosmological constant is not any more consistent with zero. Second, firm evidence exists for cosmological inflation, and it is not clear how such an early quasi-de-Sitter period fits in the argument for vanishing  $\Lambda$ . Finally, as we will discuss in detail in sections 5.1 and 5.2, the above argument may run into problems with an overdensity of wormholes (the FKS catastrophe) and the sign or negative-mode problem of euclidean quantum gravity.

Nevertheless, reinterpretations of Coleman's mechanism have recently been explored in the context of the cosmological constant and other fine tuning problems of the Standard Model (Kawai and Okada, 2011, 2012; Hamada et al., 2014, 2015). The authors take a Lorentzian approach to the dynamics of multiple large universes connected by wormholes. Such a real-time formulation avoids the problems of the euclidean path integral of gravity, but at the same time modifies the conclusions obtained by Coleman. The analysis is based on the Wheeler-DeWitt wave function for a system of multiple large universes emerged in an evolving baby universe gas. By tracing out the unobserved large and baby universes, a density matrix  $\rho$  describing our large universe is derived. The dependence of  $\rho$  on the universe volume  $z$  and on the couplings of the effective action, i.e., on the wormhole-induced  $\alpha$  parameters, is studied. A problematic feature is the divergence of integrals over universe volumes  $z$  arising in the density matrix calculation. It is treated by an IR cutoff  $z_{IR}$  corresponding to a maximum universe size. Under these assumptions, it is argued that the density matrix  $\rho$  peaks at vanishing cosmological constant, as in Coleman's mechanism, albeit with a much milder power-law dependence. This is interpreted as a prediction for a vanishing effective cosmological constant at asymptotically large times.

## 4.2. Axion Potentials From Wormholes

The main current phenomenological interest in wormholes lies in their interplay with axions. Axions have been an important ingredient in models of particle physics and cosmology since they were first proposed as solutions to the strong CP problem, and have found much wider applications ever since, e.g., as components of dark matter or as inflaton candidates. From a top-down perspective, axions are among the most generic outcomes of string compactifications, and are hence extremely well motivated (see e.g., Arvanitaki et al., 2010).

Axions enjoy a global shift symmetry  $\phi \rightarrow \phi + \epsilon$  that prevents the appearance of a potential at the perturbative level. It is only non-perturbative effects such as charged instantons and wormholes that can break these symmetries and give axions a mass. In fact, the explicit example of the Giddings-Strominger wormhole arises in the presence of axions and carries a corresponding charge given by (9). This is precisely the type of object required to generate an axion potential, as we review next following Rey (1989) (see also Alonso and Urbano, 2017).

Recall the discussion of section 3.3 on the wormhole correction  $I[g, \phi]$  to the low energy effective action of a large parent universe propagating in a plasma of baby universes. It is

given by the effective Hamiltonian (37), which can be written in the form Rey (1989)

$$e^I = \langle \psi_2 | \exp \left[ \sum_{n \in \mathbb{Z}} e^{-S_w(n)/2} K_n \int d^4x \sqrt{g(x)} \mathcal{O}_n(x) (a_n + a_{-n}^\dagger) \right] | \psi_1 \rangle. \quad (54)$$

Here, the exponential factor  $e^{-S_w(n)/2}$  has been extracted from the matrix  $\Delta_{mn}$ , making the dependence on it explicit. The remainder is denoted by  $K_n$ . The states  $|\psi_i\rangle$  live in the Fock space of baby universes on which the parent universe propagates. Correspondingly,  $a_n$  and  $a_n^\dagger$  represent baby universe annihilation and creation operators.

Baby universes associated to Giddings-Strominger wormholes carry an axionic charge given by (9). That is, they satisfy  $[Q, a_n] = -na_n$  and  $[Q, a_n^\dagger] = na_n^\dagger$ , where  $Q$  generates the  $U(1)$  axionic shift symmetry. This charge is the reason why the combination  $(a_n + a_{-n}^\dagger)$  appears in (54), generalizing Equation (37): it is impossible for an observer in the parent universe to distinguish between the annihilation of a baby universe of charge  $n$ , and the creation of a baby universe of charge  $-n$ . These two processes hence generate the same local perturbation  $\mathcal{O}_n$ . Total charge conservation implies that the effective operators  $\mathcal{O}_n(x)$  must be charged as well  $[Q, \mathcal{O}_n] = n\mathcal{O}_n$ , i.e., they transform as  $\mathcal{O}_n(x) \rightarrow e^{in\epsilon} \mathcal{O}_n(x)$  under the axionic shift  $\phi \rightarrow \phi + \epsilon f$ . From this one can deduce that the local operators must be of the form  $\mathcal{O}_n(x) = e^{in\phi/f} \mathcal{O}_S(x)$ , where  $\mathcal{O}_S(x)$  is a singlet.

One can explicitly evaluate (54) by choosing the baby universes to be in an  $\alpha$ -eigenstate (introduced in section 3.3), i.e.,  $|\psi_1\rangle = |\psi_2\rangle = |\alpha\rangle$ , with  $(a_n + a_{-n}^\dagger)|\alpha\rangle = \alpha_n|\alpha\rangle$ .<sup>9</sup> The correction to the low energy action of a large parent universe propagating in such a background is hence given by

$$\begin{aligned} I &= \sum_{n \in \mathbb{Z}} e^{-S_w(n)/2} K_n \int d^4x \sqrt{g(x)} \mathcal{O}_S(x) |\alpha_n| \exp \left( \frac{in\phi}{f} + i\delta_n \right) \\ &= \sum_{n \in \mathbb{N}_0} e^{-S_w(n)/2} K_n \int d^4x \sqrt{g(x)} \mathcal{O}_S(x) |\alpha_n| \cos \left( \frac{n\phi}{f} + \delta_n \right) \end{aligned} \quad (55)$$

where  $\alpha_n = |\alpha_n| e^{i\delta_n}$ . Of course, it is easy to consider propagation between more general baby universe states. For example,  $|\psi_1\rangle = |\psi_2\rangle = |0\rangle$  would lead to an integral of (55) over  $\alpha_n$  with a Gaussian measure analogous to (34).

The operator  $\mathcal{O}_S$  can be expanded in a set of singlet operators, e.g.,  $\mathcal{O} = 1 + a\mathcal{R} + \dots$ . Of these, the most interesting one is the unit operator, which leads to a potential for the axion. Taking

<sup>9</sup>Since the operator  $A_n := a_n + a_{-n}^\dagger$  is not Hermitian, one may worry that no basis of eigenvectors exists. To show its existence notice that the operators  $C_n := A_n + A_n^\dagger, \bar{C}_n := i(A_n - A_n^\dagger)$  are Hermitian. A quick calculation shows that  $[C_n, \bar{C}_m] = 0$ , thus  $C_n$  and  $\bar{C}_n$  can be diagonalized simultaneously with an orthonormal basis. Since  $2A_n = C_n - i\bar{C}_n$  these basis elements are also eigenvectors of  $A_n$ . This also shows that the eigenvalues of  $A_n$ , which are precisely the  $\alpha$ -parameters, will generically be complex.

into account only wormholes with charge  $n = \pm 1$  the induced potential is of the form

$$V_w(\phi) \sim \frac{|\alpha_1|}{r_0^4} e^{-\frac{S_w}{2}} \cos\left(\frac{\phi}{f} + \delta\right). \quad (56)$$

The coefficient of the potential is hard to calculate in general. In most cases (in particular for  $f < M_P$ ) its precise value is not very relevant due to the dominant exponential suppression  $e^{-S_w/2}$ . In the following, whenever an explicit estimate is needed, we will follow (Alonso and Urbano, 2017) and use the wormhole neck radius  $r_0^4 = C = (24\pi^4 f^2 M_P^2)^{-1}$  as in (56). At this stage, if no other potential exists, the phase  $\delta \equiv \delta_1$  is unphysical and could be absorbed by a shift in the axion field. Generically, in the presence of other terms in the potential, there is no reason why  $\delta_1$  should not appear. The dimensionless parameter  $|\alpha_1|$  depends on the baby universe state and is not predicted by the theory. For explicit evaluations one can assume that  $|\alpha_1|$  is an order one parameter (Alonso and Urbano, 2017). A possible justification could be the expectation value  $\int d|\alpha| e^{-|\alpha|^2} |\alpha|^2 = \sqrt{\pi}/4$  of order one. It is not unreasonable to use the Gaussian distribution since the latter appears when considering propagation between baby universe  $|0\rangle$ -vacua. In principle, however, the  $\alpha$ -parameters could take any value.

### 4.3. Superplanckian Axions

#### 4.3.1. Large Field Excursions and Inflation

One of the most interesting applications of axions is inflation (see e.g., Baumann, 2011; Baumann and McAllister, 2015; Westphal, 2015 for reviews with emphasis on stringy contexts). The perturbative shift symmetry that axions enjoy makes them ideal inflaton candidates in models of large field inflation. In these constructions, the inflaton traverses distances in field space larger than the Planck scale,  $\Delta\phi \gtrsim M_P$ . Generically, such large field displacements imply a high UV sensitivity of the model since higher-order terms in the potential,  $\Delta V(\phi) \sim \phi^{n+4}/\mu_{UV}^n$ , become relevant (here  $\mu_{UV} \lesssim M_P$  is a UV cutoff scale). This may clash with the slow-roll requirement of a smooth potential. Successful models of large field inflation hence demand a fine control of UV corrections, as it is indeed provided by axions.

One of the main reasons for the current interest in large field models is their prediction of observable primordial tensor modes in the CMB. These are parametrized by the tensor-to-scalar ratio  $r$ . Under mild assumptions, the Lyth bound (Lyth, 1997) relates  $r$  to the inflaton field displacement,

$$\Delta\phi \gtrsim \left(\frac{r}{0.01}\right)^{1/2} M_P. \quad (57)$$

The current experimental bounds (Ade et al., 2015, 2016b,a) are  $r < 0.07$  (95% confidence level, Planck, BICEP2/Keck-Array combined), with near future experiments expected to strengthen this bound significantly. The combination of these searches with the Lyth bound and the UV sensitivity of large field inflation provides an ideal playground for testing UV features of effective field theories and possibly quantum gravity.

As already emphasized, the main challenge facing large field models is the control of UV corrections. Symmetries are required

to avoid a drastic tuning of higher dimension operators. This is naturally realized by axions since, due to the shift symmetry  $\phi \rightarrow \phi + \epsilon$ , the axion potential vanishes automatically. This symmetry is mildly broken by non-perturbative effects, such as instantons and wormholes, which generate a periodic potential of the form

$$\begin{aligned} V(\phi) &= \Lambda^4 \sum_n e^{-S_n} \cos\left(\frac{n\phi}{f} + \delta_n\right) \\ &= \Lambda^4 e^{-S_1} \cos\left(\frac{\phi}{f} + \delta_1\right) + \dots \end{aligned} \quad (58)$$

Here  $\Lambda$  is a typical UV scale and  $n$ -dependent non-exponential prefactors have been suppressed. As discussed previously, gauge instantons and axionic wormholes induce such potentials (Equations (6) and (56), respectively). Different harmonics correspond to instantons/wormholes of different instanton number/axionic charge  $n$ .

The idea of *natural inflation* (Freese and Kinney, 2015) is to use the  $n = 1$  term in (58). Neglecting higher harmonics is justified in many cases due to the expectation that  $e^{-S_n} \ll e^{-S_1}$  for  $n > 1$ . Slow roll inflation then requires  $f > M_P$  (notice that the maximum field displacement of the canonically normalized axion is  $2\pi f$ ). In this simplest form, models of natural inflation are disfavored by Planck (Ade et al., 2015, 2016a,b). However, this can be remedied by small corrections, e.g., from higher harmonics. More importantly, natural inflation continues to play the role of a “benchmark model” exemplifying in the simplest way the interplay between UV theory constraints and observations. Our considerations also have applications in models of axion monodromy (Silverstein and Westphal, 2008; McAllister et al., 2010).

One might try to use wormholes to generate the inflationary potential, but one immediately runs into difficulties. Semiclassically, the charge- $n$  wormhole action is  $S_n \sim nM_P/f$ . In the regime of interest,  $f \gtrsim M_P$ , higher harmonics are not sufficiently suppressed,  $e^{-S_{n+1}} \ll e^{-S_n}$ , at least for terms with  $n \lesssim f/M_P$ . Thus, there is no hierarchy between the first few terms in the series (58) and hence no perturbative control.

A closely related and more profound issue is the fact that the lowest charge instantons are microscopic and subject to strong corrections. The spectrum of microscopic instantons does not need to resemble the classical spectrum of macroscopic wormholes (just like the spectrum of charged elementary particles does not resemble the spectrum of Reissner-Nordstrom black holes). This suggests that the dominant axion potential will be generated by some microscopic non-perturbative effect, over which one has little control, and macroscopic wormholes will only induce higher corrections. The ideal situation for inflation would then take the form

$$V(\phi) = \Lambda_{inf}^4 \cos\left(\frac{\phi}{f}\right) + \sum_{n > n_c} \Lambda_w^4 e^{-S_w(n)/2} \cos\left(\frac{n\phi}{f} + \delta_n\right). \quad (59)$$

Here  $\Lambda_{inf}$  is the scale of the inflationary potential, generated by a microscopic instanton. The sum is only over macroscopic

wormholes, i.e., those whose radius of curvature is larger than the cutoff length.

Given this setup, one may ask how important the wormhole contribution is Montero et al. (2015) and Hebecker et al. (2015). To have a successful model of inflation, it should be subdominant:

$$\frac{\Lambda_w^4 e^{-S_w(n)/2}}{\Lambda_{inf}^4} \ll 1. \quad (60)$$

Because of the exponential dependence, this constraint is highly sensitive to the wormhole action  $S_w(n) = (\pi\sqrt{6}/4) |n| M_P / f$ . The dependence on the prefactor  $\Lambda_w$  is much milder and one can, as in (56), write  $\Lambda_w^4 \sim r_n^{-4} = 24\pi^4 f^2 M_P^2 / n^2$ , where  $r_n$  is the radius of the  $S^3$  at the neck of the wormhole. The constraint (60) becomes

$$\frac{\Lambda_w^4 e^{-S_w(n)/2}}{\Lambda_{inf}^4} \sim \frac{e^{-(3\pi^3/2) r_n^2 M_P^2}}{r_n^4 \Lambda_{inf}^4} \ll 1. \quad (61)$$

This bound takes its tightest form for the wormholes with lowest charge. One should, however, only consider those which are controlled in effective field theory, i.e., whose neck radius  $r_n$  is larger than a UV cutoff scale  $r_n \gtrsim \mu_{UV}^{-1}$ . This condition defines  $n_c$  in (59). The constraint can now be further rewritten in terms of the cutoff (Hebecker et al., 2015, 2017)

$$\frac{e^{-(3\pi^3/2) M_P^2 / \mu_{UV}^2}}{\Lambda_{inf}^4 / \mu_{UV}^4} \ll 1. \quad (62)$$

One sees that, parametrically, inflation is in trouble in theories with a high cutoff,  $\mu_{UV} \sim M_P$ . The reason is that one has an  $\mathcal{O}(1)$  number in the exponent, hence an  $\mathcal{O}(1)$  numerator, and a parametrically small denominator. However, taking into account the surprisingly large numerical prefactor  $3\pi^3/2$  and the value  $\Lambda_{inf}^4 \sim 10^{-8} M_P^4$  relevant for phenomenological large field inflation, the conclusion changes dramatically. One finds that the inequality (62) is saturated at  $\mu_{UV} \simeq 2.5 M_P$  (corresponding to  $r_n \simeq 0.4 M_P$ ). Thus, even the smallest controlled wormhole solutions appear to be harmless (Hebecker et al., 2017).<sup>10</sup>

#### 4.3.2. The Weak Gravity Conjecture (WGC)

The inflationary potential (59) is perfectly acceptable from a (bottom-up) effective field theory perspective. As just discussed, macroscopic wormholes do not affect this potential significantly. However, one may be concerned that the contribution from smaller wormholes was removed by hand, and this is, at least naively, the dominant one. To argue for generic constraints

coming from this regime, where one loses semiclassical control, one has to resort to ideas about the quantum gravity *swampland*.

The concept of the swampland (Vafa, 2005) refers to the set of apparently consistent low-energy effective field theories which are, nevertheless, inconsistent with a UV completion in quantum gravity. It arises most naturally in string theory, where it represents the complement, in the space of effective field theories, of the vast *landscape* of string compactifications. Effective theories in the swampland are those that cannot arise from a UV-complete fundamental theory, and in particular from string theory.

Several criteria have been conjectured to discern whether a given theory belongs to the swampland. Most of them refer to properties of the spectra of operators charged under gauge symmetries. The simplest and perhaps most solidly founded of the swampland conjectures are the statements that every symmetry must be local and that the whole lattice of corresponding gauge charges consistent with charge quantization must be populated (see e.g., Banks and Seiberg, 2011). That is, for every symmetry there must exist a gauge potential ( $A_\mu$  in the case of a one-form), and there must exist states carrying every possible set of charges (every integer charge for a single  $U(1)$  with an appropriate normalization).

A more stringent, albeit more speculative conjecture is the WGC (Arkani-Hamed et al., 2007a): It states that at least some of the charged states must be super-extremal, that is, their charge-to-mass ratio must be larger than that of the corresponding extremal gravitational solution:

$$z_{WGC} \equiv \left(\frac{q}{m}\right)_{WGC} \geq \left(\frac{Q}{M}\right)_{ext}. \quad (63)$$

This is generally described as the statement that “gravity is always the weakest force,” since when (63) is satisfied, the gauge repulsion of two distant equal-charge objects dominates their gravitational attraction. In case of a single  $U(1)$ , the extremal object corresponds to an extremal Reissner-Nordstrom black hole, which in appropriate units satisfies  $z_{ext} = M_P^{-1}$ .<sup>11</sup> Since macroscopic gravitational solutions cannot be super-extremal (super-extremal black holes contain a naked singularity, violating cosmic censorship), one expects (63) to be satisfied by microscopic objects. For such states, quantum corrections can become relevant, pushing them away from extremality.

Now, what does all of this have to do with axions, wormholes and inflation? In general, abelian gauge theories arise from  $p$ -form gauge fields under which  $p$ -dimensional objects (i.e., whose world-volumes are  $p$ -dimensional) are charged. The swampland conjectures, and in particular the WGC, are expected to hold for all possible  $p$  (Arkani-Hamed et al., 2007a). The case of particles

<sup>10</sup>The exponentials in (61) and (62) are highly sensitive to the precise definition of the cutoff or, equivalently, the critical radius  $r_c$  above which wormholes are considered “macroscopic.” This was analyzed in Hebecker et al. (2017) using string compactifications with  $g_s = 1$  and self-dual compactification radius  $R$ . Equating the (appropriate power) of the wormhole  $S^3$  volume with the volume of the compact torus,  $(2\pi^2 r_c^2)^2 \equiv (2\pi R)^6$ , one obtains a suppression of  $e^{-S/2} \sim 10^{-68}$ . Imposing instead that the great circle of the  $S^3$  be equal to the torus  $S^1$ s,  $2\pi r_c \equiv 2\pi R$ , the suppression becomes  $e^{-S/2} \sim 10^{-13}$ . In neither of these cases are macroscopic wormholes able to affect inflation. Nevertheless, minor modifications of the definition of  $r_c$  could change this conclusion.

<sup>11</sup>The WGC has been motivated by the requirement that no stable bound states with arbitrary charge should exist. Super-extremal objects implement this requirement by permitting otherwise stable extremal black holes to decay through Schwinger pair production. It remains to be rigorously proven, however, that this requirement arises from fundamental consistency conditions. Unfortunately, the exciting and active field of the WGC lies outside the scope of this review and we limit ourselves to the axionic version and the consequences for natural inflation since this directly relates to our main subject.

with electric charge corresponds to  $p = 1$ , strings coupled to a two-form field to  $p = 2$  and, most relevant for our interests, axions can be understood as  $p = 0$  gauge fields, to which “zero-dimensional” instantons/wormholes couple. This interpretation can be made manifest by considering a standard one-form gauge field in 5d, reduced on a circle to 4d. The component of the gauge field along the circle, the Wilson line, becomes a periodic axion in 4d, whose periodicity reflects the higher dimensional gauge symmetry. In this way, one can relate the mass  $m$  and charge  $q$  of a 5d particle to the action  $S_n$  and axionic charge  $n$  of a 4d instanton, respectively. The WGC (63), when applied to axions is hence expected to read

$$\left(\frac{n}{S_n}\right)_{\text{WGC}} \geq \left(\frac{N}{S_N}\right)_{\text{ext}}. \quad (64)$$

Just like extremal black holes satisfy  $M/Q \sim eM_P$  with  $e$  the gauge coupling, one generally expects *extremal instantons* to satisfy  $S_N/N \sim M_P/f$ . If the WGC (64) is satisfied by the instanton of lowest charge  $n = 1$ , this means that  $S_1 f \lesssim M_P$ . This is incompatible with the basic requirements of large field natural inflation ( $f \gtrsim M_P$ ) in regimes of perturbative control,  $S_{\text{inst}} \gtrsim 1$ .<sup>12</sup> Setups in which the instanton that satisfies the WGC is not the one of lowest charge have been proposed as a loophole to this strong constraint (Rudelius, 2015; Brown et al., 2015, 2016) and are being actively investigated (Hebecker et al., 2015)<sup>13</sup>.

The main difficulty in making the requirement (64) more precise is to properly identify what one means by an extremal instanton/wormhole. In setups where the axion arises from a 5d gauge field, one can see that the higher dimensional extremal black holes correspond to the extremal instantons introduced in section 2.3. However, these setups always involve a dilaton field (the radius of the compactification circle) for which the coupling parameter  $\beta$  of Equation (16) is  $\beta = 2\sqrt{2/3}$ . Recall from section 2.3 that wormhole solutions only exist for  $\beta < 2\sqrt{2/3}$ . Since the main interest (at least for inflation) is in the case where the dilaton has been stabilized and disappears from the low energy theory, i.e.,  $\beta = 0$ , the relation to higher dimensional black holes is lost, along with a rigorous notion of an extremal instanton/wormhole.

Hence, with our current understanding, some amount of guesswork is required to properly interpret (64) in a pure axion-Einstein theory. Following Hebecker et al. (2017), we will assume that, on the right hand side of the bound, one needs to use the classical action of a macroscopic wormhole. With this interpretation, the WGC states that some

microscopic “wormhole” has a charge-to-action ratio larger than its macroscopic counterpart, i.e., that  $S_n \leq (\pi\sqrt{6}/4) |n| M_P/f$ .

Finally, we return to the effective model of natural inflation (59). As discussed before, the sum over macroscopic wormholes is generically suppressed strongly enough to be ignored. Ideally, one could hope that the uncontrolled microscopic wormholes somehow disappear from the low energy theory. However, the WGC implies<sup>14</sup> quite the opposite, namely that (at least some) microscopic wormholes/instantons are less suppressed than their macroscopic counterparts and inflation is strongly affected.

A possible caveat to this conclusion is the implicit assumption that all instantons enter the potential with  $\mathcal{O}(1)$  prefactors. This, however, is not in principle required by the WGC. The smoothness of the inflationary potential may be preserved if the coefficients of dangerous corrections vanish or are highly suppressed, i.e., if  $\Delta V \ll \Lambda_{\text{inf}}^4$  (see de la Fuente et al., 2015 for a model potentially realizing this possibility).

To discuss this point more generally, one can split the correction to the potential according to

$$\begin{aligned} \Delta V &= \Delta V_1 + \Delta V_2 \\ \Delta V_1(\phi) &\sim \sum_n r_n^{-4} e^{-S_n} && \text{for } r_n \gg \mu_{UV}^{-1} \\ \text{with} &&& \\ \Delta V_2(\phi) &\sim \sum_n r_n^{-4} (\mu_{UV} r_n)^\alpha e^{-S_n} && \text{for } r_n \lesssim \mu_{UV}^{-1}, \end{aligned} \quad (65)$$

with  $\alpha > 0$ . Here  $\Delta V_1$  comes from macroscopic instantons or wormholes and is harmless, as explained above. By contrast,  $\Delta V_2$  comes from their microscopic counterparts and is dangerous according to the WGC. The reason is that small, low-charge instantons are not exponentially suppressed,  $e^{-S_n} \sim \mathcal{O}(1)$  for  $n \sim \mathcal{O}(1)$ . However, the prefactor of those instantons can be smaller than the naively expected  $r_n^{-4}$ . This has been parametrized by including a factor  $(\mu_{UV} r_n)^\alpha$ .

As an example, let the microscopic instantons be gauge instantons of some non-abelian 4d gauge theory. The presence of charged fermions of mass  $m$  does not affect the contribution of large instantons (relative to  $1/m$ ). By contrast, the contribution of small instantons is suppressed by  $(mr_n)^\alpha$  with  $\alpha$  proportional to the number of fermions (as in the lower line in (65)). An analogous suppression can arise in the case of brane-instantons due to the presence of fermionic zero-modes. These are generally lifted by the SUSY breaking required for inflation. The formula (65) is grossly oversimplified in that just a single threshold,  $\mu_{UV}$ , occurs. It is only intended to illustrate how the smallness of corrections could in principle come about. Indeed, one sees that  $\Delta V \ll V_{\text{inf}} \sim H^2 M_P^2$  may be satisfied (for appropriate  $\alpha$ ) together with  $H \lesssim \mu_{UV} \ll M_P$ . Finding specific implementations of such a mechanism remains challenging.

To summarize: effects of macroscopic wormholes in the low energy Einstein-axion theory are in general not strong enough to constrain models of natural inflation. Nevertheless, expectations based on the WGC place potentially strong bounds on such

<sup>12</sup>The loss of perturbative control has a particularly nice interpretation in string theory compactifications, where one generically finds that trans-planckian axions  $f \gtrsim M_P$  arise only in regimes where either the string coupling becomes strong, or Kaluza-Klein/winding modes become light (Banks et al., 2003).

<sup>13</sup>More generally, current approaches to large field axion inflation can be roughly divided into multi-axion (Kim et al., 2005; Dimopoulos et al., 2008) and monodromy (Silverstein and Westphal, 2008; McAllister et al., 2010; Blumenhagen and Plauschinn, 2014; Hebecker et al., 2014; Marchesano et al., 2014) models (also useful in the relaxation mechanism Graham et al., 2015). The WGC and related swampland ideas can be generalized to such setups, and lead to interesting phenomenological features and constraints. The strength of these depends on subtleties in the precise formulation of the WGC and is being intensely debated.

<sup>14</sup>More precisely, this requires one of the stronger forms of the conjecture. For example, one may demand that the instanton satisfying the bound  $S_n/n < M_P/f$  has  $n = 1$  or at least  $n \sim \mathcal{O}(1)$ .

models. In particular, trans-Planckian axion decay constants are expected to arise only in regimes where perturbative control is lost, e.g., where microscopic wormholes/instanton spoil slow-roll conditions. Several possible ways around such constraints exist and are being actively studied. Whether such loopholes can be implemented in specific (string theory) setups is the subject of ongoing research.

#### 4.4. Subplanckian Axions and Goldstone Bosons

Sub-Planckian axions  $f < M_P$  are not suitable to accommodate inflation but they are extremely interesting in other phenomenological setups. Again, it is their shift symmetry and the resulting exponential suppression of their masses that makes them stand out among the plethora of fields relevant at low energy.

Let us repeat here for convenience the wormhole induced axion potential (56):

$$V_w(\phi) \approx \frac{|\alpha_1|}{r_0^4} e^{-S_w/2} \cos\left(\frac{\phi}{f} + \delta\right). \quad (66)$$

The mass induced by this potential is given by

$$m^2 = 24\pi^4 M_P^2 |\alpha_1| e^{-S_w/2} = 24\pi^4 M_P^2 |\alpha_1| e^{-\frac{\pi\sqrt{6}}{8} \frac{M_P}{f}}. \quad (67)$$

In contrast to trans-Planckian axions, for a decay constant smaller than the Planck scale the wormhole contribution is strongly suppressed through the exponential  $e^{-S_w/2}$ . This ensures that the axion is very light, making it suitable for many phenomenological applications. The exponential dependence implies that small changes in  $f$  drastically change the value of  $m$ , allowing for a wide range of values for the mass. This observation will be a recurring theme in the following.

Another feature specific to sub-Planckian axions is that even wormholes of unit charge are macroscopic, in the sense that the size of their throat  $r_0$  is larger than the Planck scale. This is a rather peculiar property, but it is necessary for (66) to be trustworthy. More in general, in an effective theory with UV cutoff  $\mu_{UV}$ , the validity of (66) requires  $r_0^{-1} \sim \sqrt{f M_P} < \mu_{UV}$ . If the cutoff scale becomes too low, one may expect sizeable corrections to the action of the wormholes with lowest charge.<sup>15</sup> The results described in this section assume the validity of (66), with  $S_w$  taking its classical value. The important caveat just mentioned should nevertheless be taken into account when interpreting these results.

We proceed now to review potential phenomenological applications of axions with an induced wormhole potential of the form (66). Significant parts of our discussion follows (Alonso and Urbano, 2017)<sup>16</sup>.

<sup>15</sup>The scale  $\mu_{UV} \sim \sqrt{f M_P}$  itself arises in the context of the (magnetic) WGC as an intrinsic UV cutoff. The unit charge wormhole lies precisely at this scale, and is hence potentially subject to relatively sizeable corrections to its action.

<sup>16</sup>Mild discrepancies with the results of Alonso and Urbano (2017) arise from the inclusion of a (Gibbons-Hawking-York) boundary contribution to the action of a semi-wormhole in Alonso and Urbano (2017). Our perspective is that of a summation over full wormholes, where no such contribution arises. The semi-wormhole factor  $e^{-S_w/2}$  appears only effectively through a rewriting.

##### 4.4.1. Black Hole Superradiance and Bosenovas

As just explained, axions play a special role in testing quantum gravity, especially wormhole or baby universe effects. The reason is their extremely suppressed potential. Furthermore, assuming that the relevant  $\alpha$  parameters take their natural  $\mathcal{O}(1)$  value, the potential and hence the mass are predicted in terms of the decay constant.

However, a generic (in particular non-QCD) axion is hard to observe. One classical possibility is black hole superradiance (Damour et al., 1976; Zouros and Eardley, 1979; Detweiler, 1980). This term characterizes the energy deposition by a spinning black hole into a light scalar field, not-necessarily an axion, of suitable mass. The relevance for the discovery of axions has been emphasized in the context of the “string axiverse” (Arvanitaki et al., 2010) and continues to receive much attention (see e.g., Arvanitaki et al., 2015, 2017; Brito et al., 2017a,b; Cardoso et al., 2018). A recent discussion in the wormhole context appears in Alonso and Urbano (2017).

The dependence of superradiance on the most important physics parameters are easily explained. Consider a spinning black hole with mass  $M$ , angular momentum  $J$  and typical radius  $R \equiv M/8\pi M_P^2$ . One generally uses the spin parameter  $a = J/M$  to characterize its rotation, with  $a = R$  corresponding to extremality.<sup>17</sup> Superradiance is a classical instability which draws energy from the black hole and deposits it in the field oscillations of a light scalar, localized in a spherical region outside the horizon. Very roughly, one may think in terms of (classical analogs of) electron shells of an atom being populated by this scalar. The effect relies on the black hole being near extremality and on the Compton wavelength of the axion being comparable to the black hole radius,  $R \sim 1/m$ .

It is instructive to consider what happens if this latter condition is not met (Arvanitaki et al., 2010): For an extremal black hole and  $R \gg 1/m$ , the instability time scale is given by Zouros and Eardley (1979)

$$\tau \simeq (10^7 R) \exp(1.84 mR). \quad (68)$$

In this regime, the Compton wave length is small and only modes with a large angular excitation superradiate. But such modes experience a high and thick centrifugal barrier, leading to an exponential suppression of the rate  $1/\tau$ . For subcritical  $a$  the exponential suppression is even stronger. In the opposite regime  $R \ll 1/m$ , one has (Detweiler, 1980)

$$\tau = 24 R (R/a) (mR)^{-9}. \quad (69)$$

In this limit, low modes are available for superradiance. However, the potential well is now very wide and the modes spread out. One may say that the scalar’s Compton wavelength is too large such that the small overlap with the black holes induces a suppression.

Efficient superradiance hence requires a relation between the black hole mass and the axion Compton wavelength. Stellar black holes ( $2 - 100 M_\odot$ ) correspond to axion masses of  $10^{-13} -$

<sup>17</sup>Intuitively,  $a$  is the radius which a shell with mass  $M$ , rotating at the speed of light, would need to have to generate  $J$ . It can not exceed the Schwarzschild radius corresponding to  $M$ .

$10^{-10}$  eV and supermassive black holes ( $10^6 - 10^8 M_\odot$ ) to  $10^{-19} - 10^{-16}$  eV. The crucial signal for an axion in one of these regions would be gaps in the spectrum of rotating black holes. At present, spin and mass observations of stellar black holes exclude the range (Arvanitaki et al., 2015)

$$6 \times 10^{-13} \text{ eV} \lesssim m \lesssim 2 \times 10^{-11} \text{ eV}. \quad (70)$$

Note that a detection of axion-induced superradiance is also possible through gravitational waves. The gravitational wave signals from, e.g., axion annihilation or axion transitions may be detected by future experiments at LIGO, VIRGO and at LISA (Arvanitaki et al., 2015, 2017; Brito et al., 2017a,b; Cardoso et al., 2018).

In our context, i.e., for a pure-quantum-gravity potential, the relation (67) between mass and decay constant may in principle provide information beyond the generic axion case. Of course, the mass is subject to the uncertainties from the  $\alpha$  parameter and fluctuation determinant. However, as can be seen by solving (67) for  $f$ ,

$$f = \frac{M_P (\pi \sqrt{6}/8)}{\ln(24\pi^4 M_P^2 |\alpha_1|/m^2)}, \quad (71)$$

the sensitivity to these uncertainties is extremely weak. Indeed, for  $|\alpha_1| = 1$  the above excluded mass window corresponds to the surprisingly narrow range  $1.23 \times 10^{16} \text{ GeV} \lesssim f \lesssim 1.28 \times 10^{16} \text{ GeV}$ . Thus, under the above assumptions, an axion discovery at the edge of the present mass window would imply a very precise determination of  $f$ . Similarly, the mass window  $10^{-19} \text{ eV} \lesssim m \lesssim 10^{-16} \text{ eV}$  accessible via supermassive black holes translates to  $1.06 \times 10^{16} \text{ GeV} \lesssim f \lesssim 1.13 \times 10^{16} \text{ GeV}$ . However, the key question is then whether an independent measurement of  $f$  for such a “quantum gravity axion” is conceivable.

It turns out that the answer to this question is positive: To measure the mass, it suffices to study superradiance at linear order. However, to get independent information about  $f$ , higher-order terms in the  $\cos(\phi/f)$  potential have to be probed. This is possible, for example, in the context of the so-called bosonova. The term derives from analogous condensed matter phenomena (Donley Elizabeth et al., 2001). In a bosonova, the self-interactions of the growing axion cloud around the black hole lead to a dynamical collapse: Part of the extracted energy is ejected and the rest returned to the black hole. This may happen multiple times until enough spin has been extracted from the black hole and superradiance (at least for the given level) is lost (Yoshino and Kodama, 2012).

Among the observable effects are a continuous gravitational wave signal as well as bursts of gravitational waves. For the continuous case, an analysis based on a possible axion cloud of the stellar black hole Cygnus X-1 was reported in Yoshino and Kodama (2015a). Assuming that the LIGO upper limit is similar to that for gravitational waves from rotating distorted neutron stars, an expected exclusion range was derived. For  $1.1 \times 10^{-12} \text{ eV} < m < 2.5 \times 10^{-12} \text{ eV}$ , it restricts  $f$  to lie below  $10^{15} - 10^{16} \text{ GeV}$ . This can be understood intuitively since, as  $f$  grows, the bosonova cuts off the superradiance instability

at higher axion densities, leading to larger signals. Notice that the bound on  $f$  is in the range relevant for wormhole induced potentials as discussed above. Realistic detection prospects exist also for gravitational wave bursts (Yoshino and Kodama, 2015b). Present limitations of the theoretical analysis are related to the need for including backreaction and extending certain parts of the numerics from the Schwarzschild to the Kerr solution (for details see e.g., Yoshino and Kodama, 2015b).

#### 4.4.2. QCD Axion

For the QCD axion an interesting observation can be made (Alonso and Urbano, 2017). The total potential, including the contribution from the usual QCD instantons, is given by

$$V(\phi) = -\Lambda_{QCD}^4 \cos\left(\frac{\phi}{f}\right) - \frac{1}{r_0^4} e^{-\frac{S_w}{2}} \cos\left(\frac{\phi}{f} + \delta\right), \quad (72)$$

where the axion  $\phi$  is defined such that the QCD instanton induced potential is minimized at  $\phi = 0$ . The phase  $\delta$  is redefined accordingly and is generically non-zero since there is no obvious reason for the two terms in the potential to have the same minimum. Furthermore, the  $|\alpha_1|$  parameter has been set to one.

The dependence of the axion mass on the decay constant is interesting. With increasing  $f$ , the QCD contribution decreases while the wormhole one grows. Hence, the axion mass features a minimum as a function of  $f$ . It is not unreasonable to expect, on theoretical grounds, that gravitational effects are subdominant with respect to gauge contributions. This requires that  $f \lesssim 1.4 \times 10^{16} \text{ GeV}$ . This bound can also be derived from phenomenological considerations. The phase of the wormhole contribution implies a shift of the minimum of the potential and hence a non-zero QCD  $\theta$ -parameter  $\theta_{\text{eff}}$ . The experimental bound  $\theta_{\text{eff}} \lesssim 10^{-10}$  coming from the neutron electric dipole moment constrains the wormhole contribution. Specifically, assuming  $\sin(\delta) \sim \mathcal{O}(1)$ , one finds a bound on the decay constant  $f \lesssim 1.2 \times 10^{16} \text{ GeV}$ . One might have suspected that the tight requirement  $\theta_{\text{eff}} \lesssim 10^{-10}$  would lead to a stronger bound on  $f$ . This is not the case, however, due to the strong exponential dependence of the wormhole contribution.

In the regime of small wormhole corrections, one can expand the potential (72) and obtain the axion mass and effective  $\theta_{\text{eff}}$  parameter as functions of  $f$ ,  $\Lambda_{QCD}$  and  $M_P$ :

$$m^2 \approx \frac{\Lambda_{QCD}^4}{f^2} + 24\pi^4 M_P^2 \cos(\delta) \exp\left(-\frac{\pi\sqrt{6}}{8} \frac{M_P}{f}\right) \quad (73)$$

$$\theta_{\text{eff}} \approx 24\pi^4 \sin(\delta) \frac{f^2 M_P^2}{\Lambda_{QCD}^4} \exp\left(-\frac{\pi\sqrt{6}}{8} \frac{M_P}{f}\right). \quad (74)$$

The minimal mass obtained from (73) is  $m \gtrsim 4 \times 10^{-9} \text{ eV}$ . Notice that the bound coming from superradiance described above is irrelevant in this case.

#### 4.4.3. Axions as Dark Matter

Despite its success on scales larger than 10 kpc, the scale of stellar distributions in typical galaxies, it is not clear yet if the cosmological  $\Lambda$ CDM model is consistent with observations at smaller distances (Weinberg et al., 2015). The tension arises

from the cuspy halo cores and an abundance of satellite galaxies predicted by numerical simulations but incompatible with observations. Using an extremely light scalar field with mass  $10^{-22} - 10^{-21}$  eV, it is possible to construct a model of dark matter with the same large scale predictions as CDM, in which, however, these problems are absent. The key idea here is that the large Compton wavelength of a light particle can suppress the formation of structures on sufficiently small scales. This dark matter model goes by the name of Fuzzy Dark Matter (FDM) (Hu et al., 2000).

Because of its extreme lightness, an axion with the induced potential (66) is an ideal candidate for FDM. Information about the possible values of  $f$  can be obtained by reproducing the observed relic abundance via the misalignment mechanism (Abbott and Sikivie, 1983; Preskill et al., 1983; Dine and Fischler, 1983). Assuming an initial misalignment angle of order one  $\theta_i = \phi_i/f \approx 1$ , the axion contribution to today's energy density (normalized by the critical energy density) is given by Arvanitaki et al. (2010) and Kim and Marsh (2016) (see also Hui et al., 2017)

$$\Omega_a h^2 \approx 0.1 \left( \frac{f}{10^{17} \text{ GeV}} \right)^2 \left( \frac{m}{10^{-22} \text{ eV}} \right)^{\frac{1}{2}} \quad (75)$$

where  $h = 0.678$  is the dimensionless Hubble parameter. Requiring that the axion accounts for (a large fraction of) the measured dark matter energy density, i.e., that  $\Omega_a h^2 \approx 0.1$ , implies a relation between the axion mass and its decay constant. For the FDM range of masses  $10^{-22} \lesssim m \lesssim 10^{-21}$  eV the axion decay constant must lie in the range  $5.6 \times 10^{16} \lesssim f \lesssim 10^{17}$  GeV.

It is interesting to compare these relations to those predicted by a wormhole induced mass (again using  $|\alpha_1| \sim \mathcal{O}(1)$  as a benchmark) (Alonso and Urbano, 2017). Plugging (67) into (75), one obtains that the correct relic density is obtained when  $f \approx 10^{16}$  GeV, which corresponds to an axion mass  $m \approx 7 \times 10^{-19}$  eV. While still valid as a candidate for dark matter, this mass is above the appropriate regime for the FDM scenario  $m \lesssim 10^{-21}$  eV. For the FDM setup, the exponential suppression  $e^{-S_w/2}$  is too strong to obtain the full dark matter relic abundance.

This conclusion is rather general and relates to the WGC described in section 4.3.2: Consider a generic non-perturbative axion mass  $m^2 = M_p^2 e^{-S_i}$ . A mass in the FDM range  $m \lesssim 10^{-21}$  eV requires  $S_i \gtrsim 220$ . At the same time, obtaining the right relic abundance through Equation (75) requires a rather large decay constant  $f \gtrsim 5.6 \times 10^{16}$  GeV. These two estimations combined lead to the interesting but potentially troublesome relation  $f S_i \gtrsim 5 M_p$ . The situation is similar to that of natural inflation described in section 4.3.2: demanding the production of enough FDM pushes instanton effects into the sub-extremal range  $f S_i \gtrsim M_p$ . This regime conflicts with the WGC which requires the presence of super-extremal (and hence naively dominant) instantons.

Of course, this conclusion is subject to several caveats. First, the exponential dependence on the instanton action makes the constraint highly sensitive to the precise extremality bound that enters the WGC. As previously discussed, the WGC for wormhole generated potential suggests  $f S_i \leq \sqrt{6\pi} M_p/8 \approx 0.96 M_p$ . Other setups (e.g., axio-dilaton instantons) provide

slightly different numerical bounds, but none of them seem to prevent the conflict. Second, the dark matter abundance expressed by Equation (75) assumes an initial angle of axion misalignment  $\theta_i = \phi_i/f \approx 1$ . Larger initial displacements can lead to an enhanced axion density. Specifically, for generic  $-\pi < \theta_i < \pi$ , Equation (75) should read

$$\Omega_a h^2 \approx 0.1 \left( \frac{f}{10^{17} \text{ GeV}} \right)^2 \left( \frac{m}{10^{-22} \text{ eV}} \right)^{\frac{1}{2}} \theta_i^2 f(\theta_i). \quad (76)$$

The function  $f(\theta_i)$  accounts for anharmonicities of the potential when the initial value of the axion is close to maximum  $\theta_i \rightarrow \pi$ , where it diverges. Using the approximate analytic expression for  $f(\theta_i)$  given in Visinelli and Gondolo (2009), one can estimate that an initial tuning  $\theta_i \approx 0.91\pi$  leads to the correct relic abundance for  $m \approx 10^{-21}$  eV and  $f S_i \approx M_p$ .

A third caveat is the fact that, as discussed around Equation (65), the energy scale at which instantons generate a mass may be lowered if a UV cutoff  $\mu_{UV}$  exists below the Planck scale, e.g., due to the presence of fermionic modes. Consider an axion mass of the form  $m^2 = \mu_{UV}^2 e^{-S_i}$ , and assume that the instantons saturate the bound  $f S_i \approx M_p$ . It is easy to see that the linear dependence of  $m$  on  $\mu_{UV}$  (as opposed to its exponential dependence on  $S_i$ ) requires an extremely low instanton scale. In fact, for generic initial misalignment  $\theta_i = 1$ , the cutoff scale should be  $\mu_{UV} \approx 10^{-12}$  eV. When the potential is generated by wormholes, a similar suppression could be in principle achieved by tuning the  $|\alpha_1|$ -parameter.

Finally, as in its applications to large field inflation, mild formulations of the WGC allow for loopholes in which sub-extremal instantons dominate the potential, and hence avoid the above constraints. In particular, systems with multiple axion are being actively investigated in this respect (see e.g., Bachlechner et al., 2018).

The above mechanisms can quite possibly reconcile axions as candidates of FDM with the WGC. It is nevertheless very interesting that such models, motivated mainly by their phenomenological applications, are probing quantum gravity constraints.

Whether fuzzy or not, axions and their induced gravitational potentials provide well-motivated dark matter candidates. Further phenomenological features, such as the formation and stability of substructures (e.g., axion stars or oscillons) also depend on the ratio of  $f$  and  $m$ . These will hopefully be experimentally probed in the near future. Moreover, direct detection experiments such as CASPER (Jackson Kimball et al., 2017) and HeXenia (Crespo Urrutia et al., unpublished) can also be expected to test the regime of extremely small (QCD-) axion masses in the foreseeable future (see also Alonso-Álvarez and Jaeckel, 2017).

In summary, there exists by now a whole set of promising phenomenological directions probing very light scalars, especially axions, which relate in a non-trivial way to quantum gravity and gravitational instantons.

## 5. CONCEPTUAL ISSUES

The discussion of wormholes presented so far has glossed over some fundamental questions which may change our perspectives on, or even invalidate, several of the results described in previous sections. Some of these issues were recognized and thoroughly discussed immediately after the first wormhole solutions were constructed, while others have been raised more recently, when wormholes have been considered in string and holographic setups. It is fair to say, nevertheless, that none of them has been fully understood yet. It is possible that the correct interpretation of wormholes and topology change will remain obscure until a controllable non-perturbative description of quantum gravity is found. It may well be, on the other hand, that the puzzles posed by wormholes can guide us in the pursuit of such a theory.

### 5.1. FKS Catastrophe

Following Coleman's intriguing proposal for a wormhole-based solution to the cosmological constant problem (Coleman, 1988c), Fischler, Kaplunovsky and Susskind have argued that an inconsistency may be hidden in the underlying calculation (Fischler and Susskind, 1989; Kaplunovsk, unpublished). Concretely, they extended Coleman's argument by including an  $R^2$  term in the action and by allowing for Wilsonian renormalization group (RG) running. As a result, they found an overdensity of wormholes, even of those with large radius.

The analysis of Fischler and Susskind (1989) follows that of Coleman (cf. section 4.1) very closely: The starting point is the action

$$\begin{aligned} S[g] &= \int d^4x \sqrt{g} \left( \Lambda - \frac{M_p^2}{2} R + \gamma R^2 + \dots \right) \\ &= \int d^4x \sqrt{g} \sum_i \lambda_i \mathcal{O}_i \end{aligned} \quad (77)$$

with  $\lambda_1 = \Lambda$ ,  $\lambda_2 = -M_p^2/2$  and  $\lambda_3 = \gamma$ . The partition function, including wormholes and multiple large universes, reads

$$\begin{aligned} Z &= \int D\alpha \exp \left( -\frac{1}{2} \sum_{i,j} \alpha_i \Delta_{ij}^{-1} \alpha_j \right) \\ &\exp \left( \int Dg \exp \left( -\int d^4x \sqrt{g} \sum_i (\lambda_i - \alpha_i) \mathcal{O}_i \right) \right). \end{aligned} \quad (78)$$

As before, the integral over metrics is performed in the saddle point approximation. This amounts to evaluating the action of (77) on a sphere of radius  $r$  and extremizing in  $r$ . But, on dimensional grounds, the  $R^2$  part of the action, evaluated on a sphere, gives an  $r$ -independent contribution. Hence the euclidean de-Sitter solution of Section 4.1 remains entirely unchanged. The resulting partition function is an exact copy of (53), except that the  $R^2$  part has to be added to the saddle-point action in the

double exponent:

$$\begin{aligned} Z &= \int D\alpha \exp \left( -\frac{1}{2} \sum_{i,j} \alpha_i \Delta_{ij}^{-1} \alpha_j \right) \\ &\exp \left[ \exp \left( 96\pi^2 \frac{(M_p^2/2 + \alpha_2)^2}{(\Lambda + \alpha_1)} + (\gamma + \alpha_3) \right) \right]. \end{aligned} \quad (79)$$

Again,  $\alpha_1$  has been redefined  $\alpha_1 \rightarrow -\alpha_1$ , and  $\gamma$  and  $\alpha_3$  have been rescaled to avoid numerical prefactors.

Now, by the same mechanism that drives  $\alpha_1$  to  $-\Lambda$  (Coleman's solution of the cosmological constant problem), the parameter  $\alpha_3$  is driven to infinity. This will turn out to be problematic. To explain the issue, the Wilsonian RG perspective is useful.

Start with the effective action at some UV length scale  $\rho_{UV}$ . The wormholes to be integrated out come in all sizes  $\rho > \rho_{UV}$ . Indeed, even in the simple Giddings-Strominger case with a single axion, the different wormhole charges give rise to a discrete set of wormholes of different radii. Thus, one can think of going down in energy in a renormalization-group-like way: One first integrates out wormholes of sizes  $\rho \in [\rho_{UV}, \rho_1]$ , then those with  $\rho \in [\rho_1, \rho_2]$ , and so on (with  $\rho_{UV} < \rho_1 < \rho_2 < \dots$ ). Very schematically, the previous formulae can be adjusted to this perspective by

$$\int D\alpha \rightarrow \prod_{\rho} \int D\alpha(\rho) \quad (80)$$

and

$$\lambda_i - \alpha_i \rightarrow \lambda_i - \sum_{\rho} \alpha_i(\rho). \quad (81)$$

The above wormhole-induced change of the couplings follows from iterating the basic step

$$\lambda_i(\rho + \Delta\rho) = \lambda_i(\rho) - \alpha_i(\rho). \quad (82)$$

In addition to (and intertwined with) this stepwise renormalization by wormholes, the usual RG running takes place. According to Fischler and Susskind (1989), the combined effect may be described by a set of modified RG equations,

$$\frac{d\tilde{\lambda}_i(\rho)}{d\ln(\rho)} = -\beta(\tilde{\lambda}_i(\rho)) - \tilde{\alpha}_i(\rho), \quad (83)$$

where  $\tilde{\lambda}_i = \lambda_i \rho^{\dim(\lambda_i)}$  and  $\tilde{\alpha}_i = \alpha_i(\rho) \rho^{\dim(\alpha_i)}$  are the dimensionless coupling constants and  $\alpha$  parameters, respectively. The first part of (83) is just the standard general form of an RG equation, the additional  $\tilde{\alpha}_i$  terms encode the wormhole effect. A redefinition of the  $\alpha_i$  is necessary when deriving this from the above stepwise procedure (i.e., when taking the continuum limit  $\Delta\rho \rightarrow 0$ ). This is left implicit here.

It will be useful for what follows to spell out the leading terms in the  $\beta$  function for the simple three-operator model considered:

$$\frac{d\tilde{\Lambda}}{d\ln(\rho)} = 4\tilde{\Lambda} + c_1 + \tilde{\Lambda}/\tilde{M}_P^2 + \gamma/\tilde{M}_P^2 + \cdots + \tilde{\alpha}_1(\rho) \quad (84)$$

$$\frac{d\tilde{M}_P^2}{d\ln(\rho)} = 2\tilde{M}_P^2 + c_2 + \tilde{\Lambda}/\tilde{M}_P^2 + \gamma/\tilde{M}_P^2 + \cdots + \tilde{\alpha}_2(\rho) \quad (85)$$

$$\frac{d\tilde{\gamma}}{d\ln(\rho)} = c_3 + \tilde{\Lambda}/\tilde{M}_P^2 + \gamma/\tilde{M}_P^2 + \cdots + \tilde{\alpha}_3(\rho). \quad (86)$$

Here the two leading terms  $4\tilde{\Lambda}$  and  $2\tilde{M}_P^2$  correspond to the naive scaling dimension of the operators. The terms  $c_i$  arise through the quartic, quadratic and logarithmic divergence of the three operator coefficients in question. The numerical prefactors of all other terms have been suppressed for brevity<sup>18</sup>.

As the above discussion shows, the Wilsonian procedure of integrating out high-scale perturbative fluctuations and wormholes induces a dependence of each effective coupling constant  $\lambda_i$  on *all* the  $\alpha$  parameters. The relevant distribution function, e.g., in (79), hence becomes

$$P(\alpha) = \exp\left(24\pi^2 \frac{M_P(\alpha)^4}{\Lambda(\alpha)} + \gamma(\alpha)\right), \quad (87)$$

where in the FKS truncation  $\alpha \equiv \{\alpha_1, \alpha_2, \alpha_3\}$  and the analysis is restricted to the single-universe-case for simplicity (hence no double-exponent). We suppress the further complication that one needs a different  $\alpha_i$  for each  $\rho$ , for the whole range of  $\rho$ . It is sufficient to consider higher scales as having been integrated out, such that  $P(\alpha)$  is interpreted as governing the physics at some low effective scale  $1/\rho$ , with a single set of  $\alpha$  parameters,  $\alpha_i = \alpha_i(\rho)$ , all belonging to that scale.

The next key point is to understand how the  $\alpha$  parameters are related to the wormhole density. To see this, return to the simple toy model with only one wormhole type and thus one  $\alpha$ -parameter. Consider the Taylor-expansion of the  $\alpha$  distribution:

$$P(\alpha) = \sum_{n=0}^{\infty} c_n \alpha^n. \quad (88)$$

Under the integral over the  $\alpha$ -parameters, this can be rewritten using baby universe operators, cf. (33). Thus, the  $n$ th term in the expansion corresponds to an amplitude with the insertion of  $n$  baby universe operators. It represents a configuration with  $n$  wormhole ends. The average number of such wormhole ends is then given by

$$\bar{N} = \frac{1}{P(\alpha)} \sum_{n=0}^{\infty} n c_n \alpha^n = \frac{1}{P(\alpha)} \alpha \frac{\partial P(\alpha)}{\partial \alpha}, \quad (89)$$

<sup>18</sup>A very naive way to derive, for example, the first equation is to write the loop corrected cosmological constant in the schematic form  $\Lambda = \Lambda_0 + c_1 \mu^4 + \gamma \mu^6/M_P^2 + \Lambda \mu^2/M_P^2$ . Here the correction terms correspond to the usual one-loop quartic divergence and the leading one-loop tadpole diagrams involving  $\gamma$  and  $\Lambda$  itself. The expression  $(\partial/\partial \ln \mu)(\Lambda/\mu^4)$  gives our desired perturbative  $\beta$ -function if one identifies the regulator scale  $\mu$  with  $1/\rho$ . Explicit formulae for such  $\beta$ -functions have more recently appeared in the context of “asymptotic safety,” see e.g., Reuter (1998), Litim (2004), and Falls et al. (2016).

where  $P(\alpha)$  appears in the denominator for normalization. Utilizing (87) now gives

$$\bar{N} \sim -\frac{M_P(\Lambda)^4}{\Lambda(\alpha)^2} \alpha \frac{\partial \Lambda(\alpha)}{\partial \alpha} + \frac{M_P(\Lambda)^2}{\Lambda(\alpha)} \alpha \frac{\partial M_P(\Lambda)^2}{\partial \alpha} + \alpha \frac{\partial \gamma(\alpha)}{\partial \alpha}, \quad (90)$$

where several unwieldy numerical prefactors were suppressed.

The curvature-squared of the classical 4-sphere solution is  $\sim \Lambda/M_P^2$ . Dividing by the corresponding volume,  $V_4 \sim M_P^4/\Lambda^2$ , gives the wormhole density

$$\nu = \frac{\bar{N}}{V_4} \sim -\alpha \frac{\partial \Lambda(\alpha)}{\partial \alpha}, \quad (91)$$

where only the leading term in the limit of small  $\Lambda$  have been kept. It is easy to see that the above logic goes through also in the case of multiple  $\alpha$  parameters, giving

$$\nu \sim -\sum_i \alpha_i \frac{\partial \Lambda(\alpha)}{\partial \alpha_i}. \quad (92)$$

Since  $\alpha_3$  is driven to infinity, the third term will dominate this expression. The relevant  $\alpha_3$  dependence of  $\Lambda$  follows from the  $\gamma$  term on the r.h. side of (84). This is clear since  $\gamma$  involves an additive  $\alpha_3$  contribution according to (86). Thus,

$$\nu \sim \alpha_3 \frac{\partial}{\partial \alpha_3} \Lambda \sim \alpha_3 \frac{\partial}{\partial \gamma} \left( \frac{\gamma}{M_P^2 \rho^2} \cdot \frac{1}{\rho^4} \right) \sim \frac{\alpha_3}{M_P^2 \rho^6}. \quad (93)$$

However, the maximum attainable density of wormholes of size  $\rho$  is given by  $\nu \sim \rho^{-4}$ . Since  $\alpha_3$  is driven to large values, (93) will saturate this bound, corresponding to the maximal  $\alpha_3$  value

$$\alpha_3 \sim M_P^2 \rho^2. \quad (94)$$

It follows that the path integral is dominated by close packing configurations. Moreover, this effect persists as  $\rho$  increases, contrary to the expectation that large wormholes should be suppressed.

Arguments against this so-called FKS or giant wormhole catastrophe were raised in Preskill (1989) and Coleman and Lee (1989), but both proposed resolutions were criticized by Polchinski in Polchinski (1989b). According to Preskill (1989), small wormholes can, when they are packed sufficiently densely, “crowd out” larger wormholes. This “excluded volume” resolution has been criticized in Polchinski (1989b) on the grounds that it violates the Wilsonian RG perspective: The effect of small wormholes should not be more drastic than to change the parameters of the effective action at lower energies. Moreover, an explicit toy model calculation was presented to demonstrate that the proposed excluded volume mechanism fails to suppress large wormholes.

The argument of Coleman and Lee (1989) is related but different at the technical level. Here, it is suggested that small wormholes induce charge violating interactions which are sufficiently strong to destabilize larger wormholes. From a microscopic perspective, small wormholes “bleed off” the

charge stabilizing the large ones. While this mechanism can be consistent with a Wilsonian RG perspective, it is clearly peculiar to the Giddings-Strominger and related wormhole solutions for which charge (or 3-form flux) is essential. Polchinski (1989b) argues against this resolution on the grounds that our focus on classical saddle points is merely due to our technical inability to treat more general topology-changing transitions (e.g., euclidean wormholes which do not solve the classical equations of motion). If included, such more general wormholes will not fall victim to the destabilization effect of Coleman and Lee (1989), reinstating the FKS catastrophe.

Finally, as emphasized in Hawking (1990b), the divergence of the measure  $P(\alpha)$  in certain regions of the  $\alpha$  parameter space calls for regularization. Depending on the cutoff procedure, different preferred values for the  $\alpha$  parameters and hence the effective couplings may be obtained. This can affect both the original argument for vanishing  $\Lambda$  as well as the “infinite force” driving  $\alpha_3$  to infinity and leading to the giant wormhole problem.

## 5.2. Euclidean Quantum Gravity and Negative Modes

The most immediate suspicion that wormholes should give rise to is that they are based on a very poorly understood sum over four-geometries and topologies, described by the euclidean path integral of quantum gravity. As is well known, this formulation suffers from serious technical and interpretational pitfalls.

Of course, the non-renormalizability of quantum gravity implies that the effective description in terms of the Einstein-Hilbert action will break down at some UV scale (e.g., the string scale) at which new degrees of freedom (excited string modes) will become important. This should not, however, pose an obstacle as long as considerations are restricted to wormholes whose size is much larger than the UV scale, i.e.,  $\rho \gg \ell_{UV}$ .

Much more worrisome is the fact that the euclidean version of the Einstein-Hilbert action is unbounded from below. Consider a conformal transformation  $g_{\mu\nu} \rightarrow \Omega^2 g_{\mu\nu}$ , under which

$$S = -\frac{1}{2} \int d^4x \sqrt{g} R \rightarrow S = -\frac{1}{2} \int d^4x \sqrt{g} \Omega^2 R - 3 \int d^4x \sqrt{g} g^{\mu\nu} \nabla_\mu \Omega \nabla_\nu \Omega \quad (95)$$

By choosing a rapidly varying conformal factor  $\Omega$ , one can make the action arbitrarily large and negative, even when the original metric  $g_{\mu\nu}$  satisfies the equations of motion ( $R_{\mu\nu} = 0$  in the absence of a cosmological constant). As a consequence, saddle points of the action, including the Giddings Strominger wormhole, necessarily possess negative modes.

This infamous conformal factor problem has been the subject of much debate, and several prescriptions have been given in order to avoid it. The most common approach, that of Gibbons, Hawking and Perry (GHP) (Gibbons et al., 1978), amounts to a rotation in the path integral contour such that the conformal factor of the metric takes imaginary values (see Schleich, 1987; Hartle and Schleich, 1988; Mazur and Mottola, 1990 for further discussions). This prescription provides us with a satisfactory action which is bounded from below,

but has dramatic consequences for Coleman’s argument for a vanishing cosmological constant described in section 4.1 (and perhaps more generally for Baum and Hawking’s mechanism Baum, 1983; Hawking, 1984, of which Coleman’s is a particular implementation). The vanishing of the cosmological constant arises from divergent probability amplitudes of the form  $P(\alpha) \sim \exp \left[ \exp \left( \frac{1}{4\Lambda(\alpha)\kappa^4(\alpha)} \right) \right]$ , whose ultimate origin is the conformal factor problem. A complex contour of integration leads to a better defined euclidean quantum gravity, but it results in a crucial change of sign  $P(\alpha) \sim \exp \left[ \exp \left( -\frac{1}{4\Lambda(\alpha)\kappa^4(\alpha)} \right) \right]$  or  $P(\alpha) \sim \exp \left[ -\exp \left( \frac{1}{4\Lambda(\alpha)\kappa^4(\alpha)} \right) \right]$ , depending on how the contour rotation is precisely implemented. Either way, these amplitudes give no explanation of the smallness of the cosmological constant (Fischler et al., 1989; Polchinski, 1989c).

The conformal factor problem also obscures the correct interpretation of wormholes in a different respect. It is well known that, in non-gravitational theories, minima of the euclidean action give a real contribution to the ground state energy, breaking the degeneracy of classically equivalent vacua, e.g., by inducing non-perturbative potentials for axions. In the presence of a negative mode the corresponding contribution to the energy becomes pure imaginary (from the one-loop determinant contribution), signaling an instability of a classical vacuum against tunneling (Coleman, 1979, 1988b). These statements, however, do not generalize straightforwardly to gravitational theories where there is no direct correlation between the euclidean path integral and the WKB prescription, and so the correct interpretation of negative modes remains unclear in this case (Lavrelashvili et al., 1985; Tanaka and Sasaki, 1992; Lavrelashvili, 1998; Tanaka, 1999; Khvedelidze et al., 2000; Lavrelashvili, 2000; Gratton and Turok, 2001; Hackworth and Weinberg, 2005; Dunne and Wang, 2006; Lavrelashvili, 2006; Brown and Weinberg, 2007; Battarra et al., 2013; Yang, 2013; Lee and Weinberg, 2014; Lee, 2014; Koehn et al., 2015).

The conformal factor problem would naively suggest that there is an infinite number of negative modes around wormhole solutions. The gravitational action, however, is largely redundant due to its invariance under diffeomorphisms. In order to properly count the number of the negative modes, one should carefully fix the gauge and take constraints into account to identify the physical degrees of freedom of the theory. The negative modes in the conformal sector of the metric are expected to be removed in the process, possibly by the GHP or similar prescriptions. There is an important caveat, however, when one tries to apply this procedure to wormholes. The gauge constraints can only be properly taken into account in the real-time theory, around solutions of the lorentzian equations of motion. Topologically non-trivial manifolds such as wormholes do not admit non-degenerate metrics, and hence cannot represent such real solutions<sup>19</sup>.

<sup>19</sup>In two dimensions this is known as the Anderson-DeWitt problem (Anderson and DeWitt, 1986) (see also Strominger, 1994), but it is generic to higher dimensions as well (Horowitz, 1991).

These subtleties in the interplay between gauge redundancies and constraints, and the transition to euclidean space, have led to contradictory statements regarding the role of negative modes around wormhole solutions. Rubakov and Shvedov have argued in Rubakov and Shvedov (1996a) that, after implementing the GHP rotation, one physical negative-action deformation of the Giddings Strominger wormholes exists. This was interpreted as an instability of large parent universes against decay by emission of baby universes. It has been argued (Barvinsky, 1998), however, that such a negative deformation, being in the conformal sector of the metric, should correspond to a gauge degree of freedom and hence disappear from the spectrum. In fact, an alternative computation in which physical modes were identified in the Lorentzian theory (where the wormhole solution is complex), has more recently found no negative modes (Alonso and Urbano, 2017). The issue becomes even more obscure in the presence of extra scalar fields (such as the dilatons of section 2.3), where scalar and metric deformations are intertwined (Kim et al., 1997; Khvedelidze et al., 2000), or in the presence of a cosmological constant. The appropriate interpretation of negative modes around wormhole solutions is hence still an open question.

The above considerations make it clear that the path integral approach to quantum gravity and the role played by gravitational instantons are still obscure. Our degree of understanding of different issues is quite disparate. While still mysterious in many aspects, the euclidean path integral has illuminated important setups of quantum gravity, several of which involve non trivial topologies (including the description of thermodynamic properties of black holes Gibbons and Hawking, 1977, the instability of hot flat space against black hole nucleation Gross et al., 1982, or the instability of the Kaluza-Klein vacuum Witten, 1982). It seems hence quite likely that topology change through euclidean wormholes is unavoidable and, following the arguments of section 3, will induce corrections in the low energy effective action. The interpretation of the resulting path integral is however still much open to debate.

Alternative formulations of effective quantum gravity will ultimately be necessary to illuminate these issues. Recently, an approach to the Lorentzian path integral based on Picard-Lefschetz theory has been used in Feldbrugge et al. (2017a), Feldbrugge et al. (2017b), Diaz Dorronsoro et al. (2017), Feldbrugge et al. (2018), and Diaz Dorronsoro et al. (2018) to explore certain aspects of quantum gravity. In this approach, wormholes would correspond to complex extrema of the Einstein-Hilbert-axion action. Picard-Lefschetz theory would then determine how the contour in the path integral is to be deformed into the complexified field space, and which saddles contribute to the path integral. It would be interesting to understand in this framework what the role played by gravitational instantons and wormholes is.

In order to shed some light on the conceptual problems raised by wormholes, we describe in the following sections toy models in setups where topology change is better understood, namely, theories of gravity in lower dimensions. Although some of the simplifications that arise in such theories surely hide crucial aspects of quantum gravity in four and higher dimensions, they

allow us to understand some fundamental aspects of wormholes in relatively controlled settings.

### 5.3. One-Dimensional Universes: Feynman Diagrams

In the next four subsections (Sections 5.3–5.6), we discuss the dynamics of the baby universe state and its interplay with the dynamics of our large universe. More precisely, almost all of this discussion will be in the context of toy models, the most developed and complex of which rely on 2d quantum gravity (Polchinski, 1989a; Banks and Lykken, 1990; Banks and O’Loughlin, 1991; Cooper et al., 1991; Hawking, 1991b; Lyons and Hawking, 1991). Such baby-universe and quantum-cosmology toy model calculations have been performed in the context of non-critical string theory (Polchinski, 1989a; Banks and O’Loughlin, 1991; Cooper et al., 1991) and will be the subject of section 5.6. However, to prepare the stage, we will start with 1d quantum gravity in the present section (Strominger, 1988; Hawking, 1991b), and its Wheeler-DeWitt formulation with baby universes (Banks, 1988; Strominger, 1988; Fischler et al., 1989; Giddings and Strominger, 1989a; Cooper et al., 1991) in Sect. 5.4. Two-dimensional quantum gravity corresponding to critical string theory (Hawking, 1991b; Lyons and Hawking, 1991) will be described in section 5.5.

As promised, we now start with the simplest case following (Strominger, 1988; Hawking, 1991b). Consider the one-dimensional diffeomorphism invariant theory with action

$$S[X, e] = \int d\tau \left( e^{-1} g_{\mu\nu} \dot{X}^\mu \dot{X}^\nu - em^2 \right). \quad (96)$$

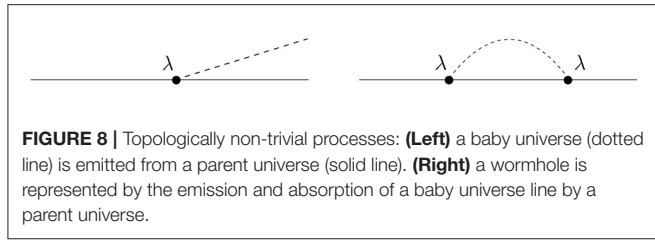
This obviously describes a free particle moving in a target space of  $D$  dimensions with metric  $g_{\mu\nu}$ . Upon quantization of the fields  $X^\mu$ , one is dealing the quantum mechanics of that particle.

The interest here, however, is in interpreting this as a theory of gravity in one dimension. In this sense, one can refer to the particle as the universe, with euclidean worldline element given by  $ds^2 = e^2 d\tau^2$  and  $D$  matter fields  $X^\mu$ . The parameter  $m^2$  hence corresponds to a one dimensional cosmological constant. Of course, such a toy model lacks many interesting features that arise in higher dimensions (to begin with, the Ricci scalar vanishes identically in one dimension, and there is no corresponding Einstein-Hilbert term in Equation 96). However, studying topology change in the one dimensional model can illuminate some points that are obscure in higher dimensions.

The theory described by Equation (96) is gauge invariant under local time reparametrizations. One can conveniently fix the gauge such that  $e = N$ , where  $N$  is constant<sup>20</sup>. It measures the proper length of the worldline and hence, in a path integral approach, it must be integrated over together with the matter fields:

$$\begin{aligned} \langle X_f | X_i \rangle_0 &= \int_0^\infty dN \int_{X_i}^{X_f} \mathcal{D}X \\ &\exp \left[ - \int_0^1 d\tau \left( N^{-1} g_{\mu\nu} \dot{X}^\mu \dot{X}^\nu + Nm^2 \right) \right]. \quad (97) \end{aligned}$$

<sup>20</sup>In one dimension the vielbein  $e(t)$  coincides with the lapse function  $N(t)$ . The gauge is fixed such that this becomes a constant  $N(t) \equiv N$ .



Here (euclidean) time was chosen to run from  $\tau_i = 0$  to  $\tau_f = 1$ . The subscript zero indicates that this corresponds to a path integral of a single-component universe, i.e., a single line in the absence of wormholes or baby universes.

Just as in higher dimensions, the action in (97) is unbounded from below if the target space metric  $g_{\mu\nu}$  has Minkowskian signature. The negative mode arises in this case from the matter field  $X^0$  associated to the target-space time direction. The solution is clear here: one needs to Wick rotate the target spacetime metric (i.e.,  $X^0 \rightarrow iX^0$ ). That is, one considers the propagation of the euclidean one-dimensional universe (particle) through a euclidean  $D$ -dimensional target spacetime. From now on, hence,  $g_{\mu\nu}$  is considered to have euclidean signature.

The path integral in (97) can be carried out explicitly, yielding (Strominger, 1988; Hawking, 1991b)

$$\langle X_f | X_i \rangle_0 = \int d^D P \frac{e^{iP(X_f - X_i)}}{P^2 + m^2} \quad (98)$$

where scalar products are taken with the target space metric  $g_{\mu\nu}$ . This is of course nothing but the euclidean propagator of a free (as indicated by the subscript) scalar of mass  $m$  in  $D$  dimension.

In order to discuss topology change and the emission of baby universes, one can introduce in the path integral (the sum over one-geometries) processes such as those shown in the **Figure 8**. To reflect as closely as possible the higher dimensional case, one would like to implement topology change as a process in which small baby universes are nucleated from large ones. Unfortunately, one dimensional universes are pointlike and there is no notion of big or small. One can, however, introduce baby universes as a different species of particles (universes) with much smaller mass than the parent universe. For concreteness, introduce a single type of baby universe with zero mass:  $m_b = 0$ .

The effect of a single wormhole on a parent universe propagator (right diagram of **Figure 8**) is given by

$$\begin{aligned} \langle X_f | X_i \rangle_1 &= \int_0^\infty dN \int_{X_i}^{X_f} \mathcal{D}X e^{-S[X,N]} \\ &\quad \left( -\lambda^2 N^2 \int_0^1 d\tau_1 \int_0^1 d\tau_2 \langle X(\tau_2) | X(\tau_1) \rangle_{0,b} \right) \\ &= \int_0^\infty dN \int_{X_i}^{X_f} \mathcal{D}X e^{-S[X,N]} \\ &\quad \left( -\lambda^2 N^2 \int_0^1 d\tau_1 \int_0^1 d\tau_2 \int d^D P \frac{e^{iP[X(\tau_2) - X(\tau_1)]}}{P^2} \right) \end{aligned} \quad (99)$$

where  $\lambda$  controls the coupling between parent and baby universes. Upon summing over arbitrary numbers of wormholes, their contribution exponentiates in a standard fashion to yield

$$\langle X_f | X_i \rangle = \sum_{n=0}^\infty \frac{1}{n!} \langle X_f | X_i \rangle_n = \int_0^\infty dN \int_{X_i}^{X_f} \mathcal{D}X e^{-S[X,N] - I[X,N]} \quad (100)$$

where  $S$  is given by (96), and  $I$  is the bilocal wormhole contribution:

$$I[X,N] = \lambda^2 \int \frac{d^D P}{P^2} \left( \int d\tau_2 N e^{iP X(\tau_2)} \right) \left( \int d\tau_1 N e^{-iP X(\tau_1)} \right). \quad (101)$$

As in previous discussions of bilocal operators, one can introduce (complex)  $\alpha$  parameters to make the action local, at the expense of having variable coupling constants:

$$\begin{aligned} e^{-I[X,N]} &= \int \mathcal{D}\alpha(P) e^{-\int d^D P P^2 |\alpha(P)|^2} \\ &\quad \exp \left[ \lambda N \int_0^1 d\tau \int d^D P (\alpha(P) e^{iP X} + \text{c.c.}) \right]. \end{aligned} \quad (102)$$

All this discussion resembles closely the description of wormholes in four dimensions. The one-dimensional theory has split into super selection sectors, labeled by  $\alpha(P)$ , which determine an infinite set of new couplings on the worldline. The target space momentum  $P$  labels the different species of wormholes, in analogy to the index  $i$  of the generic wormhole parameters  $\alpha_i$  in previous sections. Following the analogy with higher dimensional wormholes, one would affirm that the couplings  $\alpha(P)$  have a probability distribution  $e^{-|\alpha|^2}$ .

One can also use the advantageous perspective of a parent universe as a particle propagating in  $D$  dimensions. As mentioned before, the sum over parent universe one-geometries, in the absence of baby universes, is nothing but the propagator of a free scalar field  $\Phi(X)$  in  $D$  dimensions. The sum over non-trivial one-geometries is represented naturally by the sum over (connected) Feynman diagrams, where the field  $\Phi(X)$  has a cubic coupling to a light baby-universe scalar field  $\phi(X)$ . That is, all the results described previously can be derived from a quantum field theory in  $D$ -dimensional target space, with action:

$$\begin{aligned} S[\Phi, \phi] &= \frac{1}{2} \int d^D X \sqrt{g} (g^{\mu\nu} \partial_\mu \Phi \partial_\nu \Phi \\ &\quad + g^{\mu\nu} \partial_\mu \phi \partial_\nu \phi + m^2 \Phi^2 + \lambda \Phi^2 \phi). \end{aligned} \quad (103)$$

It is in fact easy to check that equation (100) is reproduced by

$$\langle X_f | X_i \rangle = \int \mathcal{D}\Phi \mathcal{D}\phi \Phi(X_i) \Phi(X_f) e^{-S[\Phi, \phi]}. \quad (104)$$

One can furthermore see that the  $\alpha(P)$  parameters induced in the worldline effective action by wormholes are nothing but the Fourier modes of the baby universe field

$$\phi(X) = \int d^D P \alpha(P) e^{-iP X}. \quad (105)$$

The  $D$ -dimensional space on which this  $(\Phi, \phi)$ -theory lives is nothing but the superspace (in Wheeler's sense) of the one-dimensional universe.

## 5.4. The Wheeler-DeWitt Perspective

It is also instructive to take a canonical rather than path integral approach to wormholes. The basic ingredient in the canonical treatment of quantum gravity is the Wheeler-DeWitt (WDW) equation, which imposes time reparametrization invariance as a constraint on the wave function of the universe. We follow the discussion of Strominger (1988).

In a one dimensional theory with action given by (96), a single universe is described by a wave function on superspace  $\Phi(X)$ . After fixing the gauge  $e(\tau) = N$ , the action is invariant under time translations  $\tau \rightarrow \tau + \text{const}$ , which are generated by the Hamiltonian

$$H = g^{\mu\nu} P_\mu P_\nu / 4 + m^2. \quad (106)$$

Here,  $P_\mu$  are the canonical momenta for the matter fields  $P_\mu = \frac{2}{N} g_{\mu\nu} \dot{X}^\nu$ . Invariance of the quantum theory under these transformations is imposed by the WDW equation

$$H \Phi(X) = 0 \quad (107)$$

where  $P_\mu = -i \nabla_\mu$ . Equation (107) should describe the dynamics of a one-dimensional (pointlike) universe, i.e., its propagation in target spacetime or superspace. It is not, however, a Schrödinger-type equation, but rather a Klein-Gordon equation in a (possibly curved)  $D$ -dimensional background. This, together with our goal of describing a system of an arbitrary number of universes, naturally suggests that  $\Phi(X)$  should be treated like a quantum field in superspace rather than as the wave function of a single universe.

With this interpretation, the linear WDW equation describes the dynamics of a free quantum field, which acts on the Fock space of an arbitrary number of universes propagating in superspace<sup>21</sup>. One expects that (107) only represents the leading approximation to a theory of interacting universes. In fact, such a theory was already introduced in the previous section. The superspace action  $S[\Phi, \phi]$  of equation (103) describes the dynamics of a parent universe field  $\Phi(X)$ , interacting with a baby universe field  $\phi(X)$  through a  $\lambda \phi \Phi^2$  coupling. The resulting equation of motion for  $\Phi$ ,

$$(\nabla^2 - m^2) \Phi = \lambda \Phi \phi, \quad (108)$$

indeed generalizes the WDW equation to the case of interacting universes.

A meta-observer capable of measuring different multi-universe states would straightforwardly interpret this theory as a quantum field theory of point-like particles propagating in  $D$ -dimensional spacetime. However, the interpretation is much more subtle for an observer living on the worldline of a single parent universe propagating in a background of baby

universes. Such an interpretation was described in section 5.3: The sum over one-geometries (Feynman diagrams) derived from the superspace action (103) is reproduced by the worldline action modified by an infinite set of  $\alpha$ -parameters, representing the baby universe field  $\phi(X)$ . In order for the single parent universe approximation to be valid, one has to make sure that the background metric in superspace is adiabatic, and that interactions among universes are small.

In the classical limit of the superspace theory, one can consider the baby universe to be in an eigenstate that satisfies the baby universe equations of motion. That is, one can replace  $\phi(X)$  by a solution  $\alpha(X)$  of the equation

$$\nabla^2 \alpha = 0 \quad (109)$$

where the backreaction of parent universes has been neglected<sup>22</sup>. The gravitational worldline theory of the parent universe in such a classical baby universe background is given by the action:

$$S = \int d\tau \left( \frac{1}{N} \dot{X}^2 - Nm - N\lambda \alpha(X) \right) \quad (110)$$

The worldline observer would measure a potential given by  $\alpha(X)$ , which in turn is determined by the superspace equations of motion (109).

Of course, when quantum fluctuations of the baby universe field are taken into account, the effective coupling constants induced on the parent universe worldline theory are no longer deterministic and are subject to the superspace quantum uncertainty principle. One would conclude, for example, that there is an intrinsic indeterminacy in the worldline potential  $\alpha(X)$  once its first derivative has been measured to a finite accuracy. The interpretation of these quantum uncertainties in the worldline couplings is still somewhat obscure.

The above logic should generalize to higher dimensional theories. In the two-dimensional case, which will be considered in more detail in the following sections, the setup is just string theory. The WDW operator implementing time-reparametrization invariance corresponds to the worldsheet Hamiltonian  $H^{23}$ . The WDW equation is then  $H\Phi(X) = 0$ , where  $\Phi(X)$  is the wave function of a single universe, a function on superspace. In order to discuss multiple universes and topology change, one promotes  $\Phi$  to a quantum field, and interprets the WDW equation as the linearized equation of motion of the corresponding superspace theory (a string field theory). This step is sometimes referred to as "third quantization." Topology change arises when one introduces interactions between string fields, leading to a non-linear generalization of the WDW equation. A two dimensional observer would interpret this theory as a gravitational theory on a genus zero worldsheet, with couplings determined by the background configuration of baby-strings. These would

<sup>21</sup>Despite being a free theory, interesting dynamics, such as universe production, can arise if the target spacetime metric  $g_{\mu\nu}$  is curved (Fischler et al., 1989).

<sup>22</sup>Of course, this equation would in general be modified by self interaction terms coming from a baby universe potential  $V(\phi)$ .

<sup>23</sup>More generally, the BRST operator which implements full reparametrization invariance.

represent, in turn, target space fields such as the metric, two-form and the dilaton. In the classical limit of the superspace theory, this background could be in a classical state satisfying the equations of motion, which would in turn lead to a determination of the worldsheet couplings. However, just as in one dimension, quantum fluctuation in superspace would lead to an intrinsic uncertainty in such couplings.

One can try to understand four dimensional wormholes in analogy to the previous discussions. The main idea is to promote the wave function of the universe to a field in superspace, and to interpret the WDW equation as the linearized equation of motion of the corresponding quantum field theory. Non-linearities arise when interactions are introduced. These would represent the effects of wormholes in the multi-universe theory. The qualitative picture should be similar to the lower dimensional cases. Four dimensional observers measure coupling constants that are determined by a background of baby universes that propagate in superspace. In the classical limit of superspace, these couplings are determined by the corresponding equations of motion, but in the quantum theory they are subject to the uncertainty principle.

Unfortunately, the infinite dimensional superspace of four dimensional universes is too complicated for this approach to be tractable in practice. One can drastically simplify the problem by reducing superspace to a finite number of dimensions, e.g., in mini-superspace approximations. An analysis of such setups, with emphasis of phenomenological implications, has been performed in Fischler et al. (1989) (see also Giddings and Strominger, 1989a). Baby universes are modeled as small spheres, interacting with large universes through non-linear terms in the WDW equation. The main phenomenological focus is on the cosmological constant problem, for which the outcome appears to be negative: While a variant of the Baum-Hawking-Coleman enhancement at  $\Lambda_{\text{eff}} \rightarrow 0$  is recovered, it occurs for empty and cold universes rather than for inflationary or big-bang cosmologies. A way beyond this negative result would require a non-standard re-interpretation of boundary conditions in the WDW equations in the multi-universe setting.

## 5.5. The Two-Dimensional Case: Critical Strings

The one dimensional theory described above is useful in many respects to understand wormhole properties in higher dimensions. It still lacks, however, important ingredients, some of which appear in the much richer context of two dimensional quantum gravity.

The way non trivial one-topologies in the path integral were introduced was rather *ad hoc*. In two dimensions, on the contrary, the sum over non-trivial topologies arises naturally. It is the basis of (perturbative) string theory. Furthermore, the superspace of one dimensional theories of gravity is finite dimensional, in contrast to the infinite dimensional superspace of worldsheet and higher dimensional theories.

Hence, one would like to discuss string theories as two dimensional models of quantum gravity. As is well known, in critical string theory the two-dimensional metric can be (locally) gauged away, and the resulting theory contains only the matter

fields  $X^\mu$ , with  $\mu = 1, \dots, D$ , as physical degrees of freedom.<sup>24</sup> In conformal gauge, the two-dimensional action is given by

$$S_P[X] = \frac{1}{2\pi\alpha'} \int d^2z (\partial X^\mu \bar{\partial} X_\mu + R\Phi_0). \quad (111)$$

For simplicity, the  $D$ -dimensional background on which the string propagates has a flat metric and constant dilaton, and all other fields (tachyon, two-form, etc) vanish.

A single spherical universe with  $g$  handles (=wormholes) attached corresponds to worldsheets of genus  $g > 0$ . Just like in the general treatment of previous sections, one can take a dilute wormhole approximation and replace these wormholes with local operators at each endpoints. One can show that these operators are nothing else than the standard vertex operators of string theory (Lyons and Hawking, 1991).

These vertex operators are in one to one correspondence with target spacetime fields. In the dilute gas approximation, only the lowest string modes contribute significantly. These are, other than possible tachyons, the target space metric, two-form and dilaton fields. They correspond to the traceless symmetric, anti-symmetric, and trace parts of the local vertex operators

$$V^{\mu\nu}(K; z) = \partial X^\mu \bar{\partial} X^\nu e^{iKX} \quad (112)$$

where  $X(z)$  are the embedding functions of the worldsheet into target space, and  $K$  is a target spacetime momentum.

As usual, upon summing over wormhole contributions with such vertex operators attached to each end, one gets a bilocal contribution to the two-dimensional effective action,

$$\Delta S = \int d^D K \left[ \left( \int d^2 z_1 V^{\mu\nu}(K; z_1) \right) \Delta_{\mu\nu\rho\sigma}(K) \left( \int d^2 z_2 V^{\rho\sigma}(K; z_2) \right) \right], \quad (113)$$

where  $\Delta_{\mu\nu\rho\sigma}(K)$  is the wormhole action, which is nothing but the  $D$ -dimensional target space propagator of massless gravitons, two-forms and dilatons.

Once again, one can introduce a set of  $\alpha_{\mu\nu}(K)$  parameters to turn this into a local contribution to the worldsheet action. The resulting path integral is:

$$Z = \int \mathcal{D}X e^{-S_P[X] - I[X]} \quad (114)$$

where  $S_P$  is the original Polyakov action, and the wormhole contribution is given by a path integral

$$e^{-I[X]} = \int \mathcal{D}\alpha_{\mu\nu}(K) \exp \left[ \int d^D K \alpha \Delta^{-1} \alpha^* \right] \exp \left[ \int d^D K \alpha_{\mu\nu} \partial X^\mu \bar{\partial} X^\nu e^{iKX} + \text{c.c.} \right]. \quad (115)$$

It is important to notice that, since wormholes have been integrated out, the path integral (114) is only over a sphere,

<sup>24</sup>We are only considering bosonic degrees of freedom, e.g., by restricting attention to bosonic string theory with  $D = 26$ .

which is to be interpreted as the parent universe. The effects of worldsheets with higher genus are encoded in the wormhole contribution  $I[X]$  via  $\alpha$ -parameters.

From the parent worldsheet point of view, wormholes have introduced a randomness in the coupling constants. Of course, this has a natural interpretation in target space: The  $\alpha$ -parameters, which can be conveniently denoted  $\{\alpha_{\mu\nu}(K)\} = \{G_{\mu\nu}(K), B_{\mu\nu}(K), D(K)\}$ , simply describe the background of metric, two-form and dilaton fields on which the string propagates.

So far only the dominant wormhole contributions, coming from massless string modes, have been considered. Massive modes will of course contribute to terms of higher dimension in the effective action, introducing an infinite set of  $\alpha(K)$ -parameters. Their quantization, i.e., the path integral over this infinite set of target space fields, should lead to string field theory (this interesting relation goes beyond the scope of this review).

## 5.6. Two Dimensional Quantum Cosmology

In this section we would like to consider two dimensional quantum cosmology in baby universe backgrounds as a toy model of the four-dimensional case. String theory in critical dimensions is not ideal for this purpose since the worldsheet metric can be gauged away. One can nevertheless follow (Polchinski, 1989a; Banks and Lykken, 1990; Banks and O'Loughlin, 1991; Cooper et al., 1991; Carneiro da Cunha and Martinec, 2003) and can consider a generally covariant theory with scalar matter fields  $X^i$ , with  $i = 1, \dots, D$  and general target space dimension  $D$ :

$$S = \frac{1}{8\pi} \int d^2\sigma \sqrt{\gamma} \left[ \gamma^{ab} \partial_a X \cdot \partial_b X + \omega R + \lambda \right]. \quad (116)$$

Here  $\gamma_{ab}$  is the worldsheet metric,  $\lambda$  is the cosmological constant, and the topological  $\omega R$ -term counts the genus of the worldsheet. The signature of the  $D$ -dimensional  $X$ -space is taken to be euclidean. It is useful to fix the gauge such that the metric becomes  $\gamma_{ab} = e^\phi \hat{\gamma}_{ab}$ , where  $\hat{\gamma}_{ab}$  is an arbitrary fiducial metric. The path integral over worldsheet metrics reduces to that over the Liouville field  $\phi$ , with an action determined by the conformal anomaly (Polyakov, 1981):

$$S = \frac{1}{8\pi} \int d^2\sigma \sqrt{\hat{\gamma}} \left\{ \hat{\gamma}^{ab} \partial_a X \cdot \partial_b X + \lambda e^\phi + \frac{26-D}{12} \left[ \hat{\gamma}^{ab} \partial_a \phi \partial_b \phi + 2\hat{R}\phi \right] \right\}. \quad (117)$$

Here  $\omega$  has been reabsorbed by a shift of  $\phi$  and a rescaling of  $\lambda$ . The equations of motion for  $\phi$  are solved by metrics of constant curvature  $R(\gamma) = R(e^\phi \hat{\gamma}) \sim \lambda$ , supporting the interpretation of  $\lambda$  as a two-dimensional cosmological constant.

Notice that in (117) the action for the metric degree of freedom  $\phi$  takes the same form as that for the matter fields  $X^i$ . One can naturally interpret  $\{\phi, X^i\}$  as parametrising a  $D+1$ -dimensional target space on which the string propagates. Interestingly, the target spacetime has euclidean signature for  $D < 26$ , and lorentzian for  $D \geq 26^{25}$ . In the latter case, the

Liouville mode  $\phi$  plays the role of a time-like coordinate in target space. It is this situation that most closely resembles gravitational theories in four dimensions (Polchinski, 1989a).

With this interpretation, Equation (117) corresponds to a subset of a more general class of 2d gravitational theories, where all  $D+1$  scalars enter on equal footing,

$$S = \frac{1}{8\pi} \int d^2\sigma \sqrt{\hat{\gamma}} \left[ T(X) + \left( \hat{\gamma}^{ab} G_{\mu\nu}(X) + i\epsilon^{ab} B_{\mu\nu}(X) \right) \partial_a X^\mu \partial_b X^\nu + 2\hat{R}\Phi(X) + \dots \right] \quad (118)$$

with  $X^0$  corresponding to the Liouville mode  $\phi$ . The function  $T(X)$  plays the role of a cosmological constant. Preserving two-dimensional diffeomorphism invariance at the quantum level is equivalent to conformal invariance and imposes strong constraints on the couplings  $\{T, G_{\mu\nu}, \Phi, B_{\mu\nu}, \dots\}$ , namely the vanishing of their  $\beta$ -functions. These constraints correspond, in  $(D+1)$ -dimensional target space, to the equations of motion of a tachyon, the metric, and the dilaton fields (setting  $B_{\mu\nu} = 0$  for simplicity):

$$\nabla^2 T - \nabla\Phi \cdot \nabla T = V'(T), \quad (119)$$

$$\nabla^2 \Phi - 2(\nabla\Phi)^2 = \frac{1}{6}(D-25) + V(T), \quad (120)$$

$$R_{\mu\nu} - \frac{1}{2}G_{\mu\nu}R = -2\nabla_\mu \nabla_\nu \Phi + G_{\mu\nu} \nabla^2 \Phi + \nabla_\mu T \nabla_\nu T - \frac{1}{2}G_{\mu\nu}(\nabla T)^2 \quad (121)$$

where  $V(T) = -T^2 + \dots$  is the target space tachyon potential.

These equations describe the dynamics of the background on which the string propagates. In our context, this background is the “baby universe state” surrounding our spacetime. It is a condensate of baby universes in the same sense that the string target space is a condensate of string states. Since the background equations of motion arise from the requirement of diffeomorphism invariance of the worldsheet, they should contain the 2d WDW equation. Non-linearities in these equations go beyond the standard WDW framework and reflect baby universe interactions. In other words, they come from topology change.

A solution is given by the linear dilaton background

$$T(X) = 0, \quad G_{\mu\nu} = \eta_{\mu\nu}, \quad \Phi(X) = -\sqrt{\frac{D-25}{12}} X^0. \quad (122)$$

Notice that the dilaton controls the string coupling  $g_s \sim e^\Phi$ . The semiclassical regime is realized in the limit  $D \rightarrow \infty$  for positive  $X^0$ . At early time  $X^0$ , the theory is strongly coupled. In the solution (122), the tachyon is balanced on top of its potential. This vanishing of the two-dimensional cosmological constant is obviously unstable against condensation of tachyons. In the

<sup>25</sup>For the Weyl invariant case of the critical string  $D = 26$  the Liouville mode  $\phi$  is a gauge degree of freedom and disappears from the spectrum. Of course this is the

best studied case. Lower central charges  $D \leq 1$  have also received much attention in the context of matrix models (see e.g., Klebanov, 1991; Ginsparg and Moore, 1993; Klebanov and Hashimoto, 1995; Martinec, 2004).

linearized approximation  $V(T) = -T^2$ , this happens with a homogeneous profile (in the limit  $D \rightarrow \infty$ )

$$T(X^0) = \lambda e^{\sqrt{\frac{12}{D}} X^0}. \quad (123)$$

This solution is valid for small values of  $T(X)$ . Higher order terms in the tachyon potential  $V(T)$  soon become relevant as the tachyon rolls down, but are hard to compute. It is conceivable that these terms produce a minimum away from zero, leading to a stable solution with constant  $T$ . It has been argued (Cooper et al., 1991) that this stability, i.e., the absence of growing modes in the WDW equation, will be interpreted by the worldsheet observer as the vanishing of the cosmological constant.

## 5.7. Wormholes in AdS/CFT

In the last few sections we have discussed the interpretation and effects of wormholes in low dimensional theories, where they are relatively well understood. However, given the simplicity of these models, in particular of their gravitational sectors, one should be very cautious when trying to extrapolate conclusions to four dimensional setups. In order to properly tackle the puzzles of wormholes, one needs to study them directly in theories of quantum gravity in higher dimensions. For this, one of the main tools presently at our disposal is the AdS/CFT correspondence.

Superstring theories in asymptotically AdS spacetimes are dual to conformal field theories living on the boundary (Gubser et al., 1998; Witten, 1998; Maldacena, 1999; Aharony et al., 2000). The partition function of the CFT should be encoded in a sum over all geometries with the correct asymptotics, possibly including topologically non-trivial ones. If wormhole configurations can be embedded in the low energy supergravity theories that arise in string theory AdS compactifications, one should arguably be able to interpret their effects, and in particular the  $\alpha$ -parameters they induce, on the field theory side.

This, however, poses severe problems (Bergshoeff et al., 2006; Arkani-Hamed et al., 2007b; Hertog et al., 2017; Ruggeri et al., 2018). It has been argued in Arkani-Hamed et al. (2007b) that AdS wormholes clash with locality of the boundary field theory. The cluster decomposition principle implies that for boundary operators  $\mathcal{O}_1$  and  $\mathcal{O}_2$  separated by a large (Euclidean) time  $T$ , the CFT correlator can be decomposed as

$$\langle \mathcal{O}_1 \mathcal{O}_2 \rangle = \langle \mathcal{O}_1 \rangle \langle \mathcal{O}_2 \rangle + \mathcal{O}(e^{-ET}) \quad (124)$$

where  $E$  is non-zero if the vacuum of the theory is unique. (The argument can also be extended to cases with a finite set of vacua.) Using the AdS/CFT dictionary, the correlators in (124) should be reproduced on the gravity side by a path integral over geometries. If these include wormholes,  $\alpha$ -parameters correct the effective couplings. Hence, the two point function on the left hand side of (124) should be given by

$$\begin{aligned} \langle \mathcal{O}_1 \mathcal{O}_2 \rangle &= \int D\alpha e^{-\alpha \Delta^{-1} \alpha} \langle \mathcal{O}_1 \mathcal{O}_2 \rangle_\alpha \\ &= \int D\alpha e^{-\alpha \Delta^{-1} \alpha} \langle \mathcal{O}_1 \rangle_\alpha \langle \mathcal{O}_2 \rangle_\alpha + \mathcal{O}(e^{-E_\alpha T}), \end{aligned} \quad (125)$$

where the correlators in the integrand are to be computed in the AdS gravitational theory with  $\alpha$ -shifted couplings. The second equality assumes the factorization (at large  $T$  and for fixed  $\alpha$ ) on the AdS side of the duality. We expect this not to be problematic, at least in the classical limit<sup>26</sup>. One can similarly compute the expectation values on the right hand side of (124):

$$\langle \mathcal{O}_1 \rangle \langle \mathcal{O}_2 \rangle = \int D\alpha_1 e^{-\alpha_1 \Delta^{-1} \alpha_1} \langle \mathcal{O}_1 \rangle_{\alpha_1} \int D\alpha_2 e^{-\alpha_2 \Delta^{-1} \alpha_2} \langle \mathcal{O}_2 \rangle_{\alpha_2}. \quad (126)$$

Equations (125) and (126) are inequivalent in general, in contradiction with the locality requirement stated in (124). To see this explicitly, assume that  $\mathcal{O}_1$  and  $\mathcal{O}_2$  are actually the same operator, just inserted at different times  $t_1$  and  $t_2$ . Then (125) gives the expectation value of  $\langle \mathcal{O} \rangle_\alpha^2$ , interpreted as a function of  $\alpha$  and using a Gaussian probability distribution  $P(\alpha) = e^{-\alpha \Delta^{-1} \alpha}$ . By contrast, (126) corresponds to the square of the expectation value of  $\langle \mathcal{O} \rangle_\alpha$ , with the same  $\alpha$ -distribution. These are equal only if  $\langle \mathcal{O} \rangle_\alpha$  is independent of  $\alpha$ , i.e., if the expectation values computed in AdS are independent of the couplings affected by wormholes.

Another problem is that the presence of wormholes in AdS can result in a violation of the BPS bound on the boundary super Yang-Mills theory (Bergshoeff et al., 2006). Bulk axions source the  $F \wedge F$  operator on the boundary, while the accompanying dilaton (always present in supersymmetric string compactifications) sources the gauge kinetic operator  $F \wedge \tilde{F}$ . It can be shown (Bergshoeff et al., 2006) that wormholes correspond, on the CFT side, to configurations that violate the BPS bound, namely, for which  $\langle |F - \tilde{F}|^2 \rangle < 0$ . These are obviously inconsistent, and pose a problem to the correct interpretation of wormholes in the holographic framework.

One might hope that string theory prevents the presence of wormholes in holographic setups where these paradoxes arise. In fact axions are always accompanied by dilatons in superstring compactifications and, as discussed in section 2.3, the existence of regular wormholes solutions depends crucially on their coupling. While the first wormhole constructions in AdS string compactifications indeed were singular (Rey, 1999; Maldacena and Maoz, 2004), regular solutions have been obtained more recently (Arkani-Hamed et al., 2007b; Hertog et al., 2017; Ruggeri et al., 2018). These analysis suggest that wormholes do exist in controlled holographic setups and hence represent a sharp paradox in AdS/CFT.

The correct resolution of this paradox is still not understood. One possibility is that some mechanism in string theory prevents topology change in holographic setups. One would need to understand in this case how such a mechanism is implemented and if it applies more generally to every string compactification. It could also be that wormholes exist but their effect on the effective action is not given in terms of  $\alpha$ -parameters (e.g., because of issues with negative modes discussed in section 5.2). Finally, another possibility is that the holographic dictionary, or

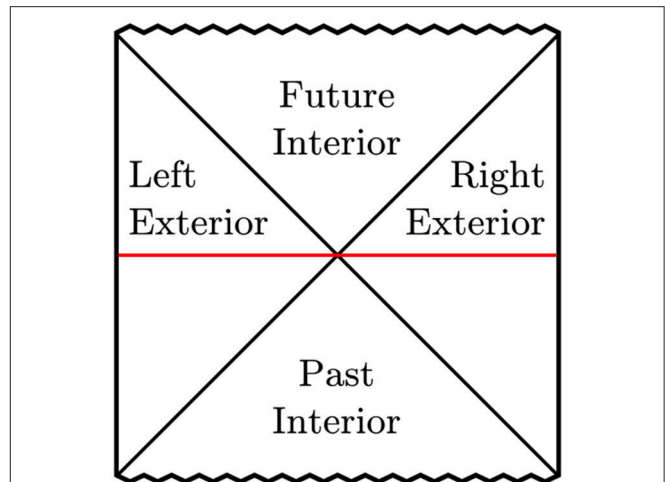
<sup>26</sup>Notice that, for some values of  $\alpha$ , massless modes could arise, for which  $E_\alpha = 0$ , and the corrections in (125) would not be exponentially suppressed. This caveat may affect the above argument, although it is not likely that it could reconcile the different structures of (125) and (126).

the correct understanding of the strongly coupled CFT, would encode the  $\alpha$ -parameters in a so far unknown manner. It is conceivable that the CFT could develop a vacuum degeneracy in its strong coupling regime which is not directly seen and which is only accessible through the  $\alpha$  parameters of the gravity dual. Alternatively, one would recover the correct factorization of two-point functions (124) if one considered the AdS theory to be in an  $\alpha$  eigenstate. It is however unknown how the CFT would encode the appropriate value of  $\alpha$ , or if there is a preferred  $\alpha$  in string compactifications.

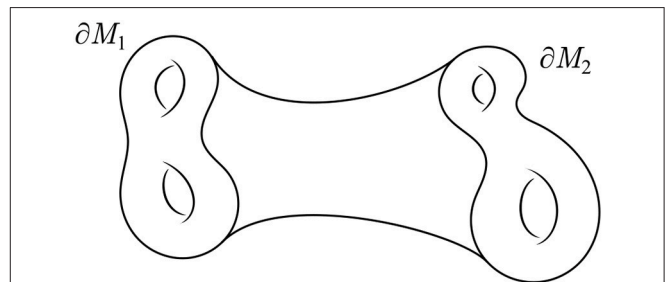
Let us now turn to a related apparent puzzle that arises when a wormhole connects two different AdS spacetimes rather than two distant regions of one AdS space. Such a geometry contains two boundaries and is hence dual to a pair of CFTs. Since the boundaries are disconnected, one naively expects CFT correlation functions of the type  $\langle \mathcal{O}_1(x_1) \mathcal{O}_2(x_2) \rangle_{\text{CFT}}$ , where  $x_1$  and  $x_2$  belong to different boundaries, to factorize as  $\langle \mathcal{O}_1(x_1) \rangle_{\text{CFT}_1} \langle \mathcal{O}_2(x_2) \rangle_{\text{CFT}_2}$ . But this contradicts the gravity side computation: Here, the presence of the wormhole, which connects the two dual AdS spaces, leads to non-trivial correlators between operators on the different boundaries. This problem is similar to the one described above, around (124)–(126).

In lorentzian signature the resolution of this puzzle is well known (Maldacena, 2003): AdS geometries with multiple boundaries always contain horizons that separate the different boundaries (Galloway et al., 2001)<sup>27</sup>. The prototypical example is an extended AdS-black hole which has two asymptotic AdS regions connected by a non-traversable wormhole or Einstein-Rosen bridge (see Figure 9). This geometry is dual to a pair of CFTs in an *entangled* state, the correlators of which hence do not factorize. Furthermore, the entanglement entropy of each boundary CFT is related to the entropy of the horizon that separates the boundaries. This can be explicitly checked with the Ryu-Takayanagi (RT) or the covariant Hubeny-Rangamani-Takayanagi prescription (Ryu and Takayanagi, 2006a,b; Hubeny et al., 2007): The entanglement entropy of a spacelike region  $A$  in the CFT is computed in the bulk by the area of a co-dimension-two minimal surface with boundary anchored on  $\partial A$ . As an example one can take  $A$  to be one of the boundaries of an AdS-black hole geometry. Since each CFT lives on a sphere one has  $\partial A = 0$ . The surface measuring the entanglement entropy of  $A$  then detaches from the boundary and moves into the bulk, becoming precisely the black hole horizon and hence measuring its area. This relation between Einstein-Rosen bridges and entanglement entropy has led to the remarkable conjecture, known as ER=EPR (Maldacena and Susskind, 2013), which says that entangled states (even microscopic ones) are generally described by wormholes.

While lorentzian wormholes, including their description in AdS/CFT, are a fascinating subject (see e.g., Visser, 1995; Maldacena and Qi, 2018), the focus of the present review is different: We are interested in euclidean wormholes, their interpretation as tunneling events, and the resulting contribution to the effective actions of gravitational theories. Unfortunately, it



**FIGURE 9** | Extended AdS-black hole with two boundaries connected by a wormhole or Einstein-Rosen bridge (horizontal line).



**FIGURE 10** | Illustration of a 3d euclidean AdS wormhole geometry with the two boundary components  $\partial M_1$  and  $\partial M_2$  corresponding to Riemann surfaces (Maldacena and Maoz, 2004; Hubeny et al., 2007).

is not immediately clear how to carry over the above discussion, especially the elegant resolution of the paradox, to this setting.

A promising way forward may be to consider euclidean rather than lorentzian wormholes which connect AdS spaces with two disconnected boundary components. The latter correspond to two euclidean CFTs (Maldacena and Maoz, 2004). Euclidean wormholes are different from their lorentzian counterparts in that they do not possess a horizon separating the two boundaries. In fact, the simplest examples are obtained by starting with global AdS and modding out a discrete symmetry. In this case there are no matter fields supporting the wormhole throat and one can not think of the wormhole ends as being localized at arbitrary points inside AdS spaces. Rather, the whole AdS space is the wormhole (cf. Figure 10).

Due in particular to the absence of horizons, the relation of these wormholes to entanglement entropy is not immediately clear (Maldacena and Maoz, 2004). However, Hubeny et al. (2007) have made a very intriguing suggestion (conceptually related to their time-dependent generalization of RT) for interpreting such euclidean wormholes in terms of entangled CFT states. The idea is to focus on a CFT state corresponding,

<sup>27</sup>see however Fujita et al. (2011) and Arias et al. (2011).

for example, to a 1-cycle  $A$  in  $\partial M_1$  in **Figure 10**. In this case the entangling surface is the minimal-length 1-cycle in the bulk to which  $A$  can be deformed. This relates to the minimal width of the wormhole at its waist. Interesting extensions include those to multiboundary wormholes (Balasubramanian et al., 2014), to situations with inflating wormhole interiors (Fischetti et al., 2015), and many others (see e.g., Mandal et al., 2015; Maxfield, 2015). In our context, the crucial question is whether such an entanglement interpretation of euclidean wormholes holds the key to resolving the problems described above. In particular, it is tantalizing to think that some generalization of ER=EPR can be applied to Euclidean wormholes, perhaps giving them a description in terms of “entangled instantons.” Can one hope to obtain a satisfactory interpretation of Coleman’s  $\alpha$ -parameters in holographic setups in this way? At a more modest level, the mere existence of well-established entanglement entropy interpretations of euclidean AdS wormholes strengthens the case of this review by making it less likely that such exotic objects can be dismissed altogether. Summarizing, it seems clear that AdS/CFT correspondence and holographic entanglement entropy suggest promising avenues to resolving the long-standing puzzles posed by wormholes.

## 6. CONCLUSIONS

We have reviewed a number of issues, both theoretical and phenomenological, arising in the context of gravitational instantons, euclidean wormholes and baby universes. The more recent interest in this old subject is related to the weak gravity conjecture, which is further connected to the interplay between charged microscopic objects and charged black branes. In the special case of the axion or 0-form gauge field, this is the interplay between microscopic instantons and gravitational instantons or wormholes.

The latter case is, however, very special. Indeed, if one insists that the macroscopic charged objects are not UV-sensitive, then cored (i.e., singular) gravitational instantons are excluded and the objects to be considered are the Giddings-Strominger wormholes. Those can be interpreted as tunneling processes in which an  $S^3$  baby universe is emitted in some region of 4d non-compact space-time and re-absorbed at an arbitrarily distant location. Allowing such processes unavoidably introduces a baby-universe state, characterized by so-called  $\alpha$  parameters, into our description of reality. This is a form of indeterminacy reminiscent of that induced by the string theory landscape. It is, however, of very different conceptual origin and potentially more severe in that parameters are scanned in a continuous way. While the axionic euclidean wormhole solution of Giddings and Strominger played a prominent role in the inception of this picture, it is really not that central: All one needs is some form of topology change.

Crucially, not only the instanton-induced axionic cosine potential is affected, but all coupling constants of the 4d effective theory. Historically, Coleman’s suggested solution to the cosmological constant problem played a crucial role in this discussion. This was based on the attempt to integrate over

the  $\alpha$  parameters together with the 4d geometry, producing a probabilistic distribution of  $\Lambda$ -values infinitely peaked at zero. However, this has become less believable due to severe technical problems and the fact that arguably a cold and empty universe is predicted.

The more modest recent discussions of phenomenology have mostly been limited to the axionic cosine potential, under the assumption that the relevant  $\alpha$  parameters take their arguably natural  $\mathcal{O}(1)$  value. For (effective) axions with  $f > M_p$ , this is relevant in the context of large-field inflation, where wormholes could in principle have a sizeable impact in the inflaton potential (Montero et al., 2015; Hebecker et al., 2017). However, it turns out that in this regime only wormholes with large 3-form flux are semiclassically controlled. As the UV cutoff is lowered, the required charge grows together with the wormhole action, and the induced potential falls exponentially. Thus, bounds independent of microscopic instantons and the weak gravity conjecture are hard to obtain. The potentially strong constraints on large-field inflation arising from the weak gravity conjecture are being intensely studied, and are one of the main reasons for the current interest in wormhole physics (Montero et al., 2015; Heidenreich et al., 2016; Hebecker et al., 2017).

By contrast, for small- $f$  axions, even minimally charged wormholes have radii above  $M_p^{-1}$  and are semiclassically controlled. This leads to interesting limits on, or even predictions of, axion masses for axions without (or with highly suppressed) microscopic instantons (Alonso and Urbano, 2017). Such bounds have immediate phenomenological relevance for black hole superradiance and light or ultralight dark matter. In the specific case of the QCD axion, the wormhole-induced potential starts to compete with the QCD-instanton effects at  $f \sim 10^{16}$  GeV, potentially spoiling the solution of the strong CP-problem at such relatively large decay constants.

While the above phenomenological considerations are intriguing and deserve further development, it is important to emphasize that deep conceptual issues remain unresolved. First, the Giddings-Strominger wormhole is a solution of euclidean quantum gravity and the status of the latter is unclear. This is in particular due to the negative modes associated with the conformal factor. Also, the question of whether the Giddings-Strominger solution has negative modes beyond those generically present in euclidean gravity, and how they should be interpreted, is being controversially discussed. However, we consider it unlikely that arguments along those lines can be strong enough to entirely forbid wormhole-type tunneling events. Indeed, in quantum mechanics, as a rule of thumb “anything that can happen will happen,” even without a stable euclidean saddle point. In this case one needs to understand which role, if any, is played by topology change, and what are the resulting effects on the low-energy effective field theory (e.g., whether  $\alpha$  parameters arise).

More drastically, one could argue that topology change may be strictly forbidden. Indeed, in lorentzian signature no smooth and everywhere defined metric can exist on a space-time “with a handle.” Thus, if one wants to think of the corresponding tunneling trajectory directly in the lorentzian theory, one is forced to deal with (mildly) singular points. We can not rule out

that this will probe unknown UV-features of the theory which will rule out the desired transitions. However, it must also be said string theory as our best candidate for a theory of quantum gravity is built on topology change in 2d and includes many examples of well-understood and controlled topology change in 10d. Thus, we find a generic censorship of topology change unlikely.

Accepting wormholes as a feature of the theory, the problems do unfortunately not end: One has to deal with the  $\alpha$  parameters and simply integrating over them together with the metric may lead to problems. One extreme instance of this is known as the Fischler-Kaplanovsky-Susskind catastrophe, which states that under reasonable assumptions the density even of large wormhole ends in 4d space becomes large and the dilute-gas approximation breaks down.

Stepping back and considering the role of  $\alpha$  parameters from a more fundamental perspective, one discovers that simply integrating over them is too simplistic. Indeed, the proper approach is the Wheeler-DeWitt equation describing the full dynamics of a superposition of many large universes interacting with a wormhole baby universe “gas” surrounding them. A helpful 1-dimensional analogy which we described is that of a heavy particle (electron) emitting and absorbing light particles (photons), the cloud of which represents a background field. The latter corresponds to the  $\alpha$  parameters, which hence have their own quantum dynamics. The Wheeler-DeWitt equation in this case encodes a standard quantum field theory. A more elaborate toy model takes the point of view of an observer living on the worldsheet of a string that propagates through target space. To this 2d observer, the sum over worldsheet topologies of string theory represents a sum over wormholes, and his  $\alpha$  parameters correspond to target space fields (metric, dilaton, etc.). Thus, understanding the values of  $\alpha$  parameters amounts to studying string field theory. Very interesting investigations of this setting have been undertaken in the context of “2d quantum cosmology” (Cooper et al., 1991). In particular, in the context of non-critical strings, insights into issues such as the emergence of time or the evolution of cosmological parameters (in particular the cosmological constant) and their interplay with wormholes appear to be within reach.

Unfortunately, even in these toy models, firm conclusions are hard to come by. Furthermore, the deep differences between one- or two-dimensional theories of gravity and higher-dimensional ones make the extrapolation of results highly speculative (Fischler et al., 1989; Giddings and Strominger, 1989a). It is conceivable that some mechanism forbids wormholes in four dimensions while allowing them in two dimensions. However,

we are not aware of such a constraint. It is hence crucial to obtain insight directly in higher dimensions. A powerful tool we have at hand is the AdS/CFT correspondence. In this context, wormholes pose a new type of puzzles. It has been argued that, while wormholes can be embedded in AdS string compactifications, their interpretation in terms of  $\alpha$  parameters lead to problems in the boundary field theory, such as non-localities or violations of the BPS bound (Bergshoeff et al., 2006; Arkani-Hamed et al., 2007b; Hertog et al., 2017; Ruggeri et al., 2018). The resolution of this conflict remains to be understood.

To summarize: the existence and effects of wormholes in theories of gravity remains, after almost 40 years, an important but enigmatic subject with both deep fundamental issues and potential phenomenological applications to be explored. Despite new insights into quantum gravity and string theory, progress in our understanding of wormholes has been slow. Our picture remains rather incomplete. Whether topology change (at low energy) is required or forbidden in four and higher dimensions remains to be conclusively settled. Either possibility opens new questions to be addressed. If wormholes exist, their effects lead, as we have discussed, to several puzzles to be resolved. If wormholes are absent altogether, the censorship mechanism at work needs to be understood. Furthermore, in this case one should also ask what are the model- and UV-independent objects (gravitational instantons) that break global axionic shift symmetry. We believe that these questions deserve further investigation.

## AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

## ACKNOWLEDGMENTS

We would like to thank W. Cottrell, P. Henkenjohann, J. Jaeckel, E. Kiritsis, P. Mangat, M. Montero, F. Rompineve, G. Shiu, M. Sloth, S. Theisen, T. Van Riet, and L. Witkowski for useful discussions. We would also like to thank one of the referees for drawing our attention to the interesting issue of the entanglement interpretation of euclidean wormholes in AdS. This work was supported by the DFG Transregional Collaborative Research Centre TRR 33 The Dark Universe. We acknowledge financial support by Deutsche Forschungsgemeinschaft within the funding programme Open Access Publishing, by the Baden-Württemberg Ministry of Science, Research and the Arts and by Ruprecht-Karls-Universität Heidelberg.

## REFERENCES

- Abbott, L. F., and Sikivie, P. (1983). A cosmological bound on the invisible axion. *Phys. Lett. B* 120, 133–136. doi: 10.1016/0370-2693(83)90638-X
- Abbott, L. F., and Wise, M. B. (1989). Wormholes and global symmetries. *Nucl. Phys. B* 325, 687–704. doi: 10.1016/0550-3213(89)90503-8

- Ade, P. A., Aghanim, N., Ahmed, Z., Aikin, R. W., Alexander, K. D., Arnaud, M., et al. (2015). Joint Analysis of BICEP2/Keck Array and Planck Data. *Phys. Rev. Lett.* 114:101301. doi: 10.1103/PhysRevLett.114.101301
- Ade, P. A., Ahmed, Z., Aikin, R. W., Alexander, K. D., Barkats, D., Benton, S. J., et al. (2016a). Improved constraints on cosmology and foregrounds from BICEP2 and Keck array cosmic microwave background

- data with inclusion of 95 GHz band. *Phys. Rev. Lett.* 116:031302. doi: 10.1103/PhysRevLett.116.031302
- Ade, P. A. R., Aghanim, N., Arnaud, M., Arroja, F., Ashdown, M., Aumont, J., et al. (2016b). Planck 2015 results. XX. Constraints on inflation. *Astron. Astrophys.* 594:A20. doi: 10.1051/0004-6361/201525898
- Aharony, O., Gubser, S. S., Maldacena, J. M., Ooguri, H., and Oz, Y. (2000). Large N field theories, string theory and gravity. *Phys. Rept.* 323, 183–386. doi: 10.1016/S0370-1573(99)00083-6
- Akerblom, N., Blumenhagen, R., Lust, D., and Schmidt-Sommerfeld, M. (2007). Instantons and holomorphic couplings in intersecting D-brane models. *J. High Energy Phys.* 8:44. doi: 10.1088/1126-6708/2007/08/044
- Alonso, R., and Urbano, A. (2017). Wormholes and masses for Goldstone bosons. *arXiv:1706.0741*.
- Alonso-Álvarez, G., and Jaeckel, J. (2018). Exploring ALPs beyond the canonical. *Phys. Rev. D* 98:023539. doi: 10.1103/PhysRevD.98.023539
- Anderson, A., and DeWitt, B. S. (1986). Does the topology of space fluctuate? *Found. Phys.* 16, 91–105. doi: 10.1007/BF01889374
- Arias, R. E., Botta Cantcheff, M., and Silva, G. A. (2011). Lorentzian AdS, wormholes and holography. *Phys. Rev. D* 83:066015. doi: 10.1103/PhysRevD.83.066015
- Arkani-Hamed, N., Motl, L., Nicolis, A., and Vafa, C. (2007a). The string landscape, black holes and gravity as the weakest force. *J. High Energy Phys.* 6:60. doi: 10.1088/1126-6708/2007/06/060
- Arkani-Hamed, N., Orgera, J., and Polchinski, J. (2007b). Euclidean wormholes in string theory. *J. High Energy Phys.* 12:18. doi: 10.1088/1126-6708/2007/12/018
- Arvanitaki, A., Baryakhtar, M., Dimopoulos, S., Dubovsky, S., and Lasenby, R. (2017). Black hole mergers and the QCD axion at advanced LIGO. *Phys. Rev. D* 95:043001. doi: 10.1103/PhysRevD.95.043001
- Arvanitaki, A., Baryakhtar, M., and Huang, X. (2015). Discovering the QCD axion with black holes and gravitational waves. *Phys. Rev. D* 91:084011. doi: 10.1103/PhysRevD.91.084011
- Arvanitaki, A., Dimopoulos, S., Dubovsky, S., Kaloper, N., and March-Russell, J. (2010). String axiverse. *Phys. Rev. D* 81:123530. doi: 10.1103/PhysRevD.81.123530
- Bachlechner, T. C., Eckerle, K., Janssen, O., and Kleban, M. (2018). Axions of evil. *Phys. Rev. D* 98:061301. doi: 10.1103/PhysRevD.98.061301
- Bachlechner, T. C., Long, C., and McAllister, L. (2016). Planckian axions and the weak gravity conjecture. *J. High Energy Phys.* 1:91. doi: 10.1007/JHEP01(2016)091
- Balasubramanian, V., Hayden, P., Maloney, A., Marolf, D., and Ross, S. F. (2014). Multiboundary wormholes and holographic entanglement. *Class. Quant. Grav.* 31:185015. doi: 10.1088/0264-9381/31/18/185015
- Banks, T. (1988). Prolegomena to a theory of bifurcating universes: a nonlocal solution to the cosmological constant problem or little lambda goes back to the future. *Nucl. Phys. B* 309, 493–512. doi: 10.1016/0550-3213(88)90455-5
- Banks, T., Dine, M., Fox, P. J., and Gorbato, E. (2003). On the possibility of large axion decay constants. *J. Cosmol. Astropart. Phys.* 306:1. doi: 10.1088/1475-7516/2003/06/001
- Banks, T., and Lykken, J. D. (1990). String theory and two-dimensional quantum gravity. *Nucl. Phys. B* 331, 173–180. doi: 10.1016/0550-3213(90)90021-5
- Banks, T., and O’Loughlin, M. (1991). Two-dimensional quantum gravity in Minkowski space. *Nucl. Phys. B* 362, 649–664. doi: 10.1016/0550-3213(91)90547-B
- Banks, T., and Seiberg, N. (2011). Symmetries and strings in field theory and gravity. *Phys. Rev. D* 83:084019. doi: 10.1103/PhysRevD.83.084019
- Barcelo, C., Garay, L. J., Gonzalez-Diaz, P. F., and Mena Marugan, G. A. (1996). Asymptotically anti-de Sitter wormholes. *Phys. Rev. D* 53, 3162–3171. doi: 10.1103/PhysRevD.53.3162
- Barvinsky, A. O. (1998). Solution of quantum Dirac constraints via path integral. *Nucl. Phys. B* 520, 533–560. doi: 10.1016/S0550-3213(98)00172-2
- Battarra, L., Lavrelashvili, G., and Lehnert, J.-L. (2013). Zoology of instanton solutions in flat potential barriers. *Phys. Rev. D* 88:104012. doi: 10.1103/PhysRevD.88.104012
- Baum, E. (1983). Zero cosmological constant from minimum action. *Phys. Lett.* 133B, 185–186. doi: 10.1016/0370-2693(83)90556-7
- Baumann, D. (2011). “Inflation,” in *Physics of the large and the small, TASI 09, proceedings of the Theoretical Advanced Study Institute in Elementary Particle Physics*, (Boulder, CO), 523–686.
- Baumann, D., and McAllister, L. (2015). *Inflation and String Theory*. Cambridge, UK: Cambridge University Press.
- Bergman, A., and Distler, J. (2007). Wormholes in Maximal Supergravity. *arXiv:0707.3168*.
- Bergshoeff, E., Collinucci, A., Gran, U., Roest, D., and Vandoren, S. (2004). Non-extremal D-instantons. *J. High Energy Phys.* 10:31. doi: 10.1088/1126-6708/2004/10/031
- Bergshoeff, E., Collinucci, A., Gran, U., Roest, D., and Vandoren, S. (2005). Non-extremal instantons and wormholes in string theory. *Fortsch. Phys.* 53, 990–996. doi: 10.1002/prop.200410227
- Bergshoeff, E., Collinucci, A., Ploegh, A., Vandoren, S., and Van Riet, T. (2006). Non-extremal D-instantons and the AdS/CFT correspondence. *J. High Energy Phys.* 1:61. doi: 10.1088/1126-6708/2006/01/061
- Betzios, P., Gaddam, N., and Papadoulaki, O. (2018). Antipodal correlation on the meron wormhole and a bang-crunch universe. *Phys. Rev. D* 97:126006. doi: 10.1103/PhysRevD.97.126006
- Bianchi, M., Kovacs, S., and Rossi, G. (2008). Instantons and supersymmetry. *Lect. Notes Phys.* 737, 303–470. doi: 10.1007/978-3-540-74233-3\_14
- Blumenhagen, R., Cvetic, M., Kachru, S., and Weigand, T. (2009). D-brane instantons in type II orientifolds. *Ann. Rev. Nucl. Part. Sci.* 59, 269–296. doi: 10.1146/annurev.nucl.010909.083113
- Blumenhagen, R., and Plauschinn, E. (2014). Towards universal axion inflation and reheating in string theory. *Phys. Lett. B* 736, 482–487. doi: 10.1016/j.physletb.2014.08.007
- Bousso, R., and Polchinski, J. (2000). Quantization of four form fluxes and dynamical neutralization of the cosmological constant. *J. High Energy Phys.* 6:6. doi: 10.1088/1126-6708/2000/06/006
- Brennan, T. D., Carta, F., and Vafa, C. (2017). The string landscape, the swampland, and the missing corner. *arXiv:1711.00864*.
- Brito, R., Ghosh, S., Barausse, E., Berti, E., Cardoso, V., Dvorkin, I., et al. (2017a). Gravitational wave searches for ultralight bosons with LIGO and LISA. *Phys. Rev. D* 96:064050. doi: 10.1103/PhysRevD.96.064050
- Brito, R., Ghosh, S., Barausse, E., Berti, E., Cardoso, V., Dvorkin, I., et al. (2017b). Stochastic and resolvable gravitational waves from ultralight bosons. *Phys. Rev. Lett.* 119:131101. doi: 10.1103/PhysRevLett.119.131101
- Brown, A. R., and Weinberg, E. J. (2007). Thermal derivation of the Coleman-De Luccia tunneling prescription. *Phys. Rev. D* 76:064003. doi: 10.1103/PhysRevD.76.064003
- Brown, J., Cottrell, W., Shiu, G., and Soler, P. (2015). Fencing in the Swampland: quantum Gravity Constraints on Large Field Inflation. *J. High Energy Phys.* 2015:23. doi: 10.1007/JHEP10(2015)023
- Brown, J., Cottrell, W., Shiu, G., and Soler, P. (2016). On axionic field ranges, loopholes and the weak gravity conjecture. *J. High Energy Phys.* 4:17. doi: 10.1007/JHEP04(2016)017
- Brown, J. D., Burgess, C. P., Kshirsagar, A., Whiting, B. F., and York, Jr., J. W. (1989). Scalar field wormholes. *Nucl. Phys. B* 328, 213–222. doi: 10.1016/0550-3213(89)90101-6
- Burgess, C. P., and Kshirsagar, A. (1989). Wormholes and duality. *Nucl. Phys. B* 324, 157–166. doi: 10.1016/0550-3213(89)90186-7
- Cardoso, V., Dias, O. J. C., Hartnett, G. S., Middleton, M., Pani, P., and Santos, J. E. (2018). Constraining the mass of dark photons and axion-like particles through black-hole superradiance. *arXiv:1801.01420*.
- Carlip, S., and De Alwis, S. P. (1990). Wormholes in (2+1)-dimensions. *Nucl. Phys. B* 337, 681–694. doi: 10.1016/0550-3213(90)90511-B
- Carneiro da Cunha, B., and Martinec, E. J. (2003). Closed string tachyon condensation and world sheet inflation. *Phys. Rev. D* 68:063502. doi: 10.1103/PhysRevD.68.063502
- Carroll, S. M., Press, W. H., and Turner, E. L. (1992). The cosmological constant. *Ann. Rev. Astron. Astrophys.* 30, 499–542. doi: 10.1146/annurev.aa.30.090192.002435
- Cheung, C., and Remmen, G. N. (2014). Naturalness and the weak gravity conjecture. *Phys. Rev. Lett.* 113:051601. doi: 10.1103/PhysRevLett.113.051601
- Chiodaroli, M., and Gutperle, M. (2009a). Instantons and wormholes for the universal hypermultiplet. *Nucl. Phys. B* 807, 138–154. doi: 10.1016/j.nuclphysb.2008.08.017

- Chiodaroli, M., and Gutperle, M. (2009b). Instantons and wormholes in  $N=2$  supergravity. *Phys. Rev. D* 79:085023. doi: 10.1103/PhysRevD.79.085023
- Choi, K., and Holman, R. (1989). A wormhole solution to the strong CP problem. *Phys. Rev. Lett.* 62:2575. doi: 10.1103/PhysRevLett.62.2575
- Choi, K., and Kim, H. (2016). Aligned natural inflation with modulations. *Phys. Lett. B* 759, 520–527. doi: 10.1016/j.physletb.2016.05.097
- Coleman, S. R. (1979). The uses of instantons. *Subnucl. Ser.* 15:805.
- Coleman, S. R. (1988a). Black holes as red herrings: topological fluctuations and the loss of quantum coherence. *Nucl. Phys. B* 307, 867–882. doi: 10.1016/0550-3213(88)90110-1
- Coleman, S. R. (1988b). Quantum tunneling and negative eigenvalues. *Nucl. Phys. B* 298, 178–186. doi: 10.1016/0550-3213(88)90308-2
- Coleman, S. R. (1988c). Why there is nothing rather than something: a theory of the cosmological constant. *Nucl. Phys. B* 310, 643–668. doi: 10.1016/0550-3213(88)90097-1
- Coleman, S. R., Hartle, J. B., Piran, T., and Weinberg, S. (eds.). (1991). “Quantum cosmology and baby universes,” in *Proceedings, 7th Winter School for Theoretical Physics* (Jerusalem; Singapore: World Scientific).
- Coleman, S. R., and Lee, K.-M. (1989). Escape from the menace of the giant wormholes. *Phys. Lett. B* 221, 242–249. doi: 10.1016/0370-2693(89)91705-X
- Coleman, S. R., and Lee, K.-M. (1990a). Big wormholes and little interactions. *Nucl. Phys. B* 341, 101–118. doi: 10.1016/0550-3213(90)90263-D
- Coleman, S. R., and Lee, K.-M. (1990b). Wormholes made without massless matter fields. *Nucl. Phys. B* 329, 387–409. doi: 10.1016/0550-3213(90)90149-8
- Collinucci, A. (2005). *Instantons and Cosmologies in String Theory*. Ph.D. thesis, University of Groningen.
- Cooper, A. R., Susskind, L., and Thorlacius, L. (1991). Two-dimensional quantum cosmology. *Nucl. Phys. B* 363, 132–162. doi: 10.1016/0550-3213(91)90238-S
- Cortes, V., and Mohaupt, T. (2009). Special geometry of euclidean supersymmetry III: the local r-map, instantons and black holes. *J. High Energy Phys.* 7:66. doi: 10.1088/1126-6708/2009/07/066
- Damour, T., Deruelle, N., and Ruffini, R. (1976). On quantum resonances in stationary geometries. *Lett. Nuovo Cim.* 15, 257–262. doi: 10.1007/BF02725534
- de la Fuente, A., Saraswat, P., and Sundrum, R. (2015). Natural inflation and quantum gravity. *Phys. Rev. Lett.* 114:151303. doi: 10.1103/PhysRevLett.114.151303
- Denef, F., and Douglas, M. R. (2004). Distributions of flux vacua. *J. High Energy Phys.* 5:72. doi: 10.1088/1126-6708/2004/05/072
- Detweiler, S. L. (1980). Klein-Gordon equation and rotating black holes. *Phys. Rev. D* 22, 2323–2326. doi: 10.1103/PhysRevD.22.2323
- DeWitt, B. S. (1967). Quantum theory of gravity. 1. The canonical theory. *Phys. Rev.* 160, 1113–1148. doi: 10.1103/PhysRev.160.1113
- Diaz Dorronsoro, J., Halliwell, J. J., Hartle, J. B., Hertog, T., and Janssen, O. (2017). Real no-boundary wave function in Lorentzian quantum cosmology. *Phys. Rev. D* 96:043505. doi: 10.1103/PhysRevD.96.043505
- Diaz Dorronsoro, J., Halliwell, J. J., Hartle, J. B., Hertog, T., Janssen, O., and Vreys, Y. (2018). Damped perturbations in the no-boundary state. *Phys. Rev. Lett.* 121:081302. doi: 10.1103/PhysRevLett.121.081302
- Dijkgraaf, R., Gopakumar, R., Ooguri, H., and Vafa, C. (2006). Baby universes in string theory. *Phys. Rev. D* 73:066002. doi: 10.1103/PhysRevD.73.066002
- Dimopoulos, S., Kachru, S., McGreevy, J., and Wacker, J. G. (2008). N-flation. *J. Cosmol. Astropart. Phys.* 808:3. doi: 10.1088/1475-7516/2008/08/003
- Dine, M., and Fischler, W. (1983). The not so harmless axion. *Phys. Lett. B* 120, 137–141. doi: 10.1016/0370-2693(83)90639-1
- Donley E. A., Claussen N. R., Cornish S. L., Roberts J. L., Cornell E. A., and Wieman C. E. (2001). Dynamics of collapsing and exploding Bose–Einstein condensates. *Nature* 412:295. doi: 10.1038/35085500
- Duff, M. J. (1989). The cosmological constant is possibly zero, but the proof is probably wrong. *Phys. Lett. B* 226:36. doi: 10.1016/0370-2693(89)90284-0
- Dunne, G. V., and Wang, Q.-h. (2006). Fluctuations about cosmological instantons. *Phys. Rev. D* 74:024018. doi: 10.1103/PhysRevD.74.024018
- Ellis, J. R., Hagelin, J. S., Nanopoulos, D. V., and Srednicki, M. (1984). Search for violations of quantum mechanics. *Nucl. Phys. B* 241:381. doi: 10.1016/0550-3213(84)90053-1
- Falls, K., Litim, D. F., Nikolakopoulos, K., and Rahmede, C. (2016). Further evidence for asymptotic safety of quantum gravity. *Phys. Rev. D* 93:104022. doi: 10.1103/PhysRevD.93.104022
- Feldbrugge, J., Lehnert, J.-L., and Turok, N. (2017a). Lorentzian quantum cosmology. *Phys. Rev. D* 95:103508. doi: 10.1103/PhysRevD.95.103508
- Feldbrugge, J., Lehnert, J.-L., and Turok, N. (2018). No rescue for the no boundary proposal: pointers to the future of quantum cosmology. *Phys. Rev. D* 97:023509. doi: 10.1103/PhysRevD.97.023509
- Feldbrugge, J., Lehnert, J. L., and Turok, N. (2017b). No smooth beginning for spacetime. *Phys. Rev. Lett.* 119:171301. doi: 10.1103/PhysRevLett.119.171301
- Fischetti, S., Marolf, D., and Wall, A. C. (2015). A paucity of bulk entangling surfaces: AdS wormholes with de Sitter interiors. *Class. Quant. Grav.* 32:065011. doi: 10.1088/0264-9381/32/6/065011
- Fischler, W., Klebanov, I. R., Polchinski, J., and Susskind, L. (1989). Quantum mechanics of the googolplexus. *Nucl. Phys. B* 327, 157–177. doi: 10.1016/0550-3213(89)90290-3
- Fischler, W., and Susskind, L. (1989). A wormhole catastrophe. *Phys. Lett. B* 217, 48–54. doi: 10.1016/0370-2693(89)91514-1
- Freese, K., and Kinney, W. H. (2015). Natural inflation: consistency with cosmic microwave background observations of planck and BICEP2. *J. Cosmol. Astropart. Phys.* 1503:044. doi: 10.1088/1475-7516/2015/03/044
- Fujita, M., Hatsuda, Y., and Takayanagi, T. (2011). Probing AdS Wormholes by Entanglement Entropy. *J. High Energy Phys.* 6:141. doi: 10.1007/JHEP06(2011)141
- Galloway, G. J., Schleich, K., Witt, D., and Woolgar, E. (2001). The AdS/CFT correspondence conjecture and topological censorship. *Phys. Lett. B* 505, 255–262. doi: 10.1016/S0370-2693(01)00335-5
- Gibbons, G. W., Green, M. B., and Perry, M. J. (1996). Instantons and seven-branes in type IIB superstring theory. *Phys. Lett. B* 370, 37–44. doi: 10.1016/0370-2693(95)01565-5
- Gibbons, G. W., and Hawking, S. W. (1977). Action integrals and partition functions in quantum gravity. *Phys. Rev. D* 15, 2752–2756. doi: 10.1103/PhysRevD.15.2752
- Gibbons, G. W., Hawking, S. W., and Perry, M. J. (1978). Path integrals and the indefiniteness of the gravitational action. *Nucl. Phys. B* 138, 141–150. doi: 10.1016/0550-3213(78)90161-X
- Giddings, S. B., Kachru, S., and Polchinski, J. (2002). Hierarchies from fluxes in string compactifications. *Phys. Rev. D* 66:106006. doi: 10.1103/PhysRevD.66.106006
- Giddings, S. B., and Strominger, A. (1988a). Axion induced topology change in quantum gravity and string theory. *Nucl. Phys. B* 306, 890–907. doi: 10.1016/0550-3213(88)90446-4
- Giddings, S. B., and Strominger, A. (1988b). Loss of incoherence and determination of coupling constants in quantum gravity. *Nucl. Phys. B* 307, 854–866. doi: 10.1016/0550-3213(88)90109-5
- Giddings, S. B., and Strominger, A. (1989a). Baby universes, third quantization and the cosmological constant. *Nucl. Phys. B* 321, 481–508. doi: 10.1016/0550-3213(89)90353-2
- Giddings, S. B., and Strominger, A. (1989b). String wormholes. *Phys. Lett. B* 230, 46–51. doi: 10.1016/0370-2693(89)91651-1
- Gilbert, G. (1989). Wormhole induced proton decay. *Nucl. Phys. B* 328, 159–170. doi: 10.1016/0550-3213(89)90097-7
- Ginsparg, P. H., and Moore, G. W. (1993). “Lectures on 2-D gravity and 2-D string theory,” in *Proceedings, Theoretical Advanced Study Institute (TASI 92): From Black Holes and Strings to Particles* (Boulder, CO), 277–469.
- Graham, P. W., Kaplan, D. E., and Rajendran, S. (2015). Cosmological relaxation of the electroweak scale. *Phys. Rev. Lett.* 115:221801. doi: 10.1103/PhysRevLett.115.221801
- Gratton, S., and Turok, N. (2001). Homogeneous modes of cosmological instantons. *Phys. Rev. D* 63:123514. doi: 10.1103/PhysRevD.63.123514
- Green, M. B., and Gutperle, M. (1997). Effects of D instantons. *Nucl. Phys. B* 498, 195–227. doi: 10.1016/S0550-3213(97)00269-1
- Grinstein, B. (1989). Charge quantization of wormholes and the finiteness of Newton’s constant. *Nucl. Phys. B* 321, 439–464. doi: 10.1016/0550-3213(89)90351-9
- Grinstein, B., and Hill, C. T. (1989). The trace anomaly and low-energy phenomenological implications of wormholes. *Phys. Lett. B* 220, 520–526. doi: 10.1016/0370-2693(89)90779-X
- Grinstein, B., and Maharana, J. (1990). Vertex operators for axionic wormholes. *Nucl. Phys. B* 333, 160–172. doi: 10.1016/0550-3213(90)90226-4

- Grinstein, B., Maharana, J., and Sudarsky, D. (1990). All orders vertex operators for axionic wormholes. *Nucl. Phys. B* 345, 231–247. doi: 10.1016/0550-3213(90)90616-L
- Grinstein, B., and Wise, M. B. (1988). Light scalars in quantum gravity. *Phys. Lett. B* 212, 407–410. doi: 10.1016/0370-2693(88)91788-1
- Gross, D. J., Perry, M. J., and Yaffe, L. G. (1982). Instability of flat space at finite temperature. *Phys. Rev. D* 25, 330–355. doi: 10.1103/PhysRevD.25.330
- Gubser, S. S., Klebanov, I. R., and Polyakov, A. M. (1998). Gauge theory correlators from noncritical string theory. *Phys. Lett. B* 428, 105–114. doi: 10.1016/S0370-2693(98)00377-3
- Gutperle, M., and Sabra, W. (2002). Instantons and wormholes in Minkowski and (A)dS spaces. *Nucl. Phys. B* 647, 344–356. doi: 10.1016/S0550-3213(02)00942-2
- Hackworth, J. C., and Weinberg, E. J. (2005). Oscillating bounce solutions and vacuum tunneling in de Sitter spacetime. *Phys. Rev. D* 71:044014. doi: 10.1103/PhysRevD.71.044014
- Hamada, Y., Kawai, H., and Kawana, K. (2014). Evidence of the big fix. *Int. J. Mod. Phys. A* 29:1450099. doi: 10.1142/S0217751X14500997
- Hamada, Y., Kawai, H., and Kawana, K. (2015). Weak scale from the maximum entropy principle. *PTEP* 2015:033B06. doi: 10.1093/ptep/ptv011
- Harlow, D. (2016). Wormholes, emergent gauge fields, and the weak gravity conjecture. *J. High Energy Phys.* 1:122. doi: 10.1007/JHEP01(2016)122
- Hartle, J. B., and Hawking, S. W. (1983). Wave function of the universe. *Phys. Rev. D* 28, 2960–2975. doi: 10.1103/PhysRevD.28.2960
- Hartle, J. B., and Schleich, K. (1988). “The conformal rotation in linearized gravity,” in *Quantum Field Theory and Quantum Statistics*, Vol. 2, eds I. A. Batalin, C. J. Isham, and G. A. Vilkovitsky (Bristol, UK: Hilger), 67–87.
- Hawking, S. W. (1978). Space-time foam. *Nucl. Phys. B* 144, 349–362. doi: 10.1016/0550-3213(78)90375-9
- Hawking, S. W. (1984). The cosmological constant is probably zero. *Phys. Lett.* 134B:403. doi: 10.1016/0370-2693(84)91370-4
- Hawking, S. W. (1987). Quantum coherence down the wormhole. *Phys. Lett. B* 195:337. doi: 10.1016/0370-2693(87)90028-1
- Hawking, S. W. (1988). Wormholes in space-time. *Phys. Rev. D* 37, 904–910. doi: 10.1103/PhysRevD.37.904
- Hawking, S. W. (1990a). Baby universes. 2. *Mod. Phys. Lett. A* 5, 453–466. doi: 10.1142/S0217732390000524
- Hawking, S. W. (1990b). Do wormholes fix the constants of nature? *Nucl. Phys. B* 335, 155–165. doi: 10.1016/0550-3213(90)90175-D
- Hawking, S. W. (1991a). The alpha parameters of wormholes. *Phys. Scripta T* 36, 222–227. doi: 10.1088/0031-8949/1991/T36/023
- Hawking, S. W. (1991b). The effective action for wormholes. *Nucl. Phys. B* 363, 117–131. doi: 10.1016/0550-3213(91)90237-R
- Hawking, S. W., and Laflamme, R. (1988). Baby universes and the nonrenormalizability of gravity. *Phys. Lett. B* 209, 39–41. doi: 10.1016/0370-2693(88)91825-4
- Hawking, S. W., and Page, D. N. (1990). The spectrum of wormholes. *Phys. Rev. D* 42, 2655–2663. doi: 10.1103/PhysRevD.42.2655
- Hebecker, A., Kraus, S. C., and Witkowski, L. T. (2014). D7-brane chaotic inflation. *Phys. Lett. B* 737, 16–22. doi: 10.1016/j.physletb.2014.08.028
- Hebecker, A., Mangat, P., Rompineve, F., and Witkowski, L. T. (2015). Winding out of the swamp: evading the weak gravity conjecture with F-term winding inflation? *Phys. Lett. B* 748, 455–462. doi: 10.1016/j.physletb.2015.07.026
- Hebecker, A., Mangat, P., Theisen, S., and Witkowski, L. T. (2017). Can gravitational instantons really constrain axion inflation? *J. High Energy Phys.* 2:97. doi: 10.1007/JHEP02(2017)097
- Heidenreich, B., Reece, M., and Rudelius, T. (2015). Weak gravity strongly constrains large-field axion inflation. *J. High Energy Phys.* 2015, 1–41. doi: 10.1007/JHEP12(2015)108
- Heidenreich, B., Reece, M., and Rudelius, T. (2016). Sharpening the weak gravity conjecture with dimensional reduction. *J. High Energy Phys.* 2:140. doi: 10.1007/JHEP02(2016)140
- Hertog, T., Trigiante, M., and Van Riet, T. (2017). Axion wormholes in AdS compactifications. *J. High Energy Phys.* 6:67. doi: 10.1007/JHEP06(2017)067
- Horowitz, G. T. (1991). Topology change in classical and quantum gravity. *Class. Quant. Grav.* 8, 587–602. doi: 10.1088/0264-9381/8/4/007
- Hu, W., Barkana, R., and Gruzinov, A. (2000). Cold and fuzzy dark matter. *Phys. Rev. Lett.* 85, 1158–1161. doi: 10.1103/PhysRevLett.85.1158
- Hubeny, V. E., Rangamani, M., and Takayanagi, T. (2007). A covariant holographic entanglement entropy proposal. *J. High Energy Phys.* 7:62. doi: 10.1088/1126-6708/2007/07/062
- Hui, L., Ostriker, J. P., Tremaine, S., and Witten, E. (2017). Ultralight scalars as cosmological dark matter. *Phys. Rev. D* 95:043541. doi: 10.1103/PhysRevD.95.043541
- Ibanez, L. E., and Uranga, A. M. (2012). *String Theory and Particle Physics: An Introduction to String Phenomenology*. Cambridge, UK: Cambridge University Press.
- Iqbal, A., Nekrasov, N., Okounkov, A., and Vafa, C. (2008). Quantum foam and topological strings. *J. High Energy Phys.* 4:11. doi: 10.1088/1126-6708/2008/04/011
- Jackson Kimball, D. F., Afach, S., Aybas, D., Blanchard, J. W., Budker, D., et al. (2017). Overview of the cosmic axion spin precession experiment (CASPER). *arXiv:1707.05312*.
- Junghans, D. (2016). Large-field inflation with multiple axions and the weak gravity conjecture. *J. High Energy Phys.* 2:128. doi: 10.1007/JHEP02(2016)128
- Kachru, S., Kallosh, R., Linde, A. D., and Trivedi, S. P. (2003). De Sitter vacua in string theory. *Phys. Rev. D* 68:046005. doi: 10.2172/812628
- Kallosh, R., Linde, A., Linde, D., and Susskind, L. (1995). Gravity and global symmetries. *Phys. Rev. D* 52, 912–935. doi: 10.1103/PhysRevD.52.912
- Kaloper, N., Kleban, M., Lawrence, A., and Sloth, M. S. (2016). Large field inflation and gravitational entropy. *Phys. Rev. D* 93:043510. doi: 10.1103/PhysRevD.93.043510
- Kappl, R., Nilles, H. P., and Winkler, M. W. (2016). Modulated natural inflation. *Phys. Lett. B* 753, 653–659. doi: 10.1016/j.physletb.2015.12.073
- Kawai, H., and Okada, T. (2011). Asymptotically vanishing cosmological constant in the multiverse. *Int. J. Mod. Phys. A* 26, 3107–3120. doi: 10.1142/S0217751X11053730
- Kawai, H., and Okada, T. (2012). Solving the naturalness problem by baby universes in the lorentzian multiverse. *Prog. Theor. Phys.* 127, 689–721. doi: 10.1143/PTP.127.689
- Khvedelidze, A., Lavrelashvili, G. V., and Tanaka, T. (2000). On cosmological perturbations in closed FRW model with scalar field and false vacuum decay. *Phys. Rev. D* 62:083501. doi: 10.1103/PhysRevD.62.083501
- Kim, J. E., and Marsh, D. J. E. (2016). An ultralight pseudoscalar boson. *Phys. Rev. D* 93:025027. doi: 10.1103/PhysRevD.93.025027
- Kim, J. E., Nilles, H. P., and Peloso, M. (2005). Completing natural inflation. *J. Cosmol. Astropart. Phys.* 501:5. doi: 10.1088/1475-7516/2005/01/005
- Kim, J. Y., Lee, H. W., and Myung, Y. S. (1997). Negative modes in the four-dimensional stringy wormholes. *Phys. Rev. D* 56, 6684–6687. doi: 10.1103/PhysRevD.56.6684
- Klaewer, D., and Palti, E. (2017). Super-planckian spatial field variations and quantum gravity. *J. High Energy Phys.* 1:88. doi: 10.1007/JHEP01(2017)088
- Klebanov, I. R. (1991). “String theory in two-dimensions,” in *Spring School on String Theory and Quantum Gravity (to be followed by Workshop)* (Trieste), 30–101.
- Klebanov, I. R., and Hashimoto, A. (1995). Nonperturbative solution of matrix models modified by trace squared terms. *Nucl. Phys. B* 434, 264–282. doi: 10.1016/0550-3213(94)00518-J
- Klebanov, I. R., Susskind, L., and Banks, T. (1989). Wormholes and the cosmological constant. *Nucl. Phys. B* 317, 665–692. doi: 10.1016/0550-3213(89)90538-5
- Koehn, M., Lavrelashvili, G., and Lehnert, J.-L. (2015). Towards a solution of the negative mode problem in quantum tunnelling with gravity. *Phys. Rev. D* 92:023506. doi: 10.1103/PhysRevD.92.023506
- Kooner, K., Parameswaran, S., and Zavala, I. (2016). Warping the weak gravity conjecture. *Phys. Lett. B* 759, 402–409. doi: 10.1016/j.physletb.2016.05.082
- Lavrelashvili, G. (2006). The Number of negative modes of the oscillating bounces. *Phys. Rev. D* 73:083513. doi: 10.1103/PhysRevD.73.083513
- Lavrelashvili, G. V. (1998). On the quadratic action of the Hawking-Turok instanton. *Phys. Rev. D* 58:063505. doi: 10.1103/PhysRevD.58.063505
- Lavrelashvili, G. V. (2000). Negative mode problem in false vacuum decay with gravity. *Nucl. Phys. Proc. Suppl.* 88, 75–82. doi: 10.1016/S0920-5632(00)00756-8
- Lavrelashvili, G. V., Rubakov, V. A., and Tinyakov, P. G. (1985). Tunneling transitions with gravitation: breaking of the quasiclassical approximation. *Phys. Lett.* 161B, 280–284. doi: 10.1016/0370-2693(85)90761-0

- Lavrelashvili, G. V., Rubakov, V. A., and Tinyakov, P. G. (1987). Disruption of quantum coherence upon a change in spatial topology in quantum gravity. *JETP Lett.* 46, 167–169.
- Lee, H., and Weinberg, E. J. (2014). Negative modes of Coleman-De Luccia bounces. *Phys. Rev. D* 90:124002. doi: 10.1103/PhysRevD.90.124002
- Lee, H. J. (2014). *Negative Modes in Vacuum Decay*. Ph.D. thesis, Columbia University.
- Lee, K. (1988). Wormholes and goldstone bosons. *Phys. Rev. Lett.* 61, 263–266. doi: 10.1103/PhysRevLett.61.263
- Linde, A. D. (1992). Hard art of the universe creation (stochastic approach to tunneling and baby universe formation). *Nucl. Phys. B* 372, 421–442. doi: 10.1016/0550-3213(92)90326-7
- Litim, D. F. (2004). Fixed points of quantum gravity. *Phys. Rev. Lett.* 92:201301. doi: 10.1103/PhysRevLett.92.201301
- Lyons, A., and Hawking, S. W. (1991). Wormholes in string theory. *Phys. Rev. D* 44, 3802–3818. doi: 10.1103/PhysRevD.44.3802
- Lyth, D. H. (1997). What would we learn by detecting a gravitational wave signal in the cosmic microwave background anisotropy? *Phys. Rev. Lett.* 78, 1861–1863. doi: 10.1103/PhysRevLett.78.1861
- Maldacena, J., and Qi, X.-L. (2018). Eternal traversable wormhole. *arXiv:1804.00491*.
- Maldacena, J., and Susskind, L. (2013). Cool horizons for entangled black holes. *Fortsch. Phys.* 61, 781–811. doi: 10.1002/prop.201300020
- Maldacena, J. M. (1999). The Large N limit of superconformal field theories and supergravity. *Int. J. Theor. Phys.* 38, 1113–1133. doi: 10.1023/A:1026654312961
- Maldacena, J. M. (2003). Eternal black holes in anti-de Sitter. *J. High Energy Phys.* 4:21. doi: 10.1088/1126-6708/2003/04/021
- Maldacena, J. M., and Maoz, L. (2004). Wormholes in AdS. *J. High Energy Phys.* 2:53. doi: 10.1088/1126-6708/2004/02/053
- Mandal, G., Sinha, R., and Sorokhaibam, N. (2015). The inside outs of  $\text{AdS}_3/\text{CFT}_2$ : exact AdS wormholes with entangled CFT duals. *J. High Energy Phys.* 1:36. doi: 10.1007/JHEP01(2015)036
- Marchesano, F., Shiu, G., and Uranga, A. M. (2014). F-term axion monodromy inflation. *J. High Energy Phys.* 9:184. doi: 10.1007/JHEP09(2014)184
- Martinec, E. J. (2004). “Matrix models and 2D string theory,” in *9th Frontiers of Mathematical Physics Summer School on Strings, Gravity and Cosmology* (Vancouver, BC), 403–457.
- Maxfield, H. (2015). Entanglement entropy in three dimensional gravity. *J. High Energy Phys.* 4:31. doi: 10.1007/JHEP04(2015)031
- Mazur, P. O., and Mottola, E. (1990). The gravitational measure, solution of the conformal factor problem and stability of the ground state of quantum gravity. *Nucl. Phys. B* 341, 187–212. doi: 10.1016/0550-3213(90)90268-1
- McAllister, L., Silverstein, E., and Westphal, A. (2010). Gravity waves and linear inflation from axion monodromy. *Phys. Rev. D* 82:046003. doi: 10.1103/PhysRevD.82.046003
- Mohaupt, T., and Waite, K. (2011). Euclidean actions, instantons, solitons and supersymmetry. *J. Phys. A* 44:175403. doi: 10.1088/1751-8113/44/17/175403
- Montero, M., Uranga, A. M., and Valenzuela, I. (2015). Transplanckian axions? *J. High Energy Phys.* 8:32. doi: 10.1007/JHEP08(2015)032
- Nielsen, H. B., and Ninomiya, M. (1989). A solution of strong CP problem in baby universe theory. *Phys. Rev. Lett.* 62, 1429–1432. doi: 10.1103/PhysRevLett.62.1429
- Nirov, K. S., and Rubakov, V. A. (1995). Strings as a model for parent and baby universes: total splitting rates. *Mod. Phys. Lett. A* 10, 3059–3068. doi: 10.1142/S0217732395003197
- Ooguri, H., and Vafa, C. (2007). On the geometry of the string landscape and the swampland. *Nucl. Phys. B* 766, 21–33. doi: 10.1016/j.nuclphysb.2006.10.033
- Ooguri, H., and Vafa, C. (2016). Non-supersymmetric AdS and the Swampland. *Adv. Theor. Math. Phys.* 21, 1787–1801.
- Polchinski, J. (1989a). A two-dimensional model for quantum gravity. *Nucl. Phys. B* 324, 123–140. doi: 10.1016/0550-3213(89)90184-3
- Polchinski, J. (1989b). Decoupling versus excluded volume or return of the giant wormholes. *Nucl. Phys. B* 325, 619–630. doi: 10.1016/0550-3213(89)90499-9
- Polchinski, J. (1989c). The phase of the sum over spheres. *Phys. Lett. B* 219, 251–257. doi: 10.1016/0370-2693(89)90387-0
- Polyakov, A. M. (1981). Quantum geometry of bosonic strings. *Phys. Lett. B* 103, 207–210. doi: 10.1016/0370-2693(81)90743-7
- Preskill, J. (1989). Wormholes in space-time and the constants of nature. *Nucl. Phys. B* 323, 141–186. doi: 10.1016/0550-3213(89)90592-0
- Preskill, J., Trivedi, S. P., and Wise, M. B. (1989). Wormholes in space-time and  $\theta$  (QCD). *Phys. Lett. B* 223, 26–31. doi: 10.1016/0370-2693(89)90913-1
- Preskill, J., Wise, M. B., and Wilczek, F. (1983). Cosmology of the invisible axion. *Phys. Lett. B* 120, 127–132. doi: 10.1016/0370-2693(83)90637-8
- Regge, T. (1961). General relativity without coordinates. *Nuovo Cim.* 19, 558–571. doi: 10.1007/BF02733251
- Reuter, M. (1998). Nonperturbative evolution equation for quantum gravity. *Phys. Rev. D* 57, 971–985. doi: 10.1103/PhysRevD.57.971
- Rey, S.-J. (1989). The axion dynamics in wormhole background. *Phys. Rev. D* 39:3185. doi: 10.1103/PhysRevD.39.3185
- Rey, S.-J. (1999). Holographic principle and topology change in string theory. *Class. Quant. Grav.* 16, L37–L43. doi: 10.1088/0264-9381/16/7/102
- Rubakov, V. A. (1988). On the third quantization and the cosmological constant. *Phys. Lett. B* 214, 503–507. doi: 10.1016/0370-2693(88)90108-6
- Rubakov, V. A., and Shvedov, O. Yu. (1996a). A negative mode about Euclidean wormhole. *Phys. Lett. B* 383, 258–261. doi: 10.1016/0370-2693(96)00766-6
- Rubakov, V. A., and Shvedov, O. Yu. (1996b). “Instability of space-time due to Euclidean wormholes,” in *Quarks’96. Proceedings, 9th International Seminar, Vol. 1, 2* (Yaroslavl), 206–214.
- Rudelius, T. (2015). Constraints on axion inflation from the weak gravity conjecture. *J. Cosmol. Astropart. Phys.* 1509:20. doi: 10.1088/1475-7516/2015/9/020
- Ruggeri, D., Trigiante, M., and Van Riet, T. (2018). Instantons from geodesics in AdS moduli spaces. *J. High Energy Phys.* 3:91. doi: 10.1007/JHEP03(2018)091
- Ryu, S., and Takayanagi, T. (2006a). Aspects of holographic entanglement entropy. *J. High Energy Phys.* 8:45. doi: 10.1088/1126-6708/2006/08/045
- Ryu, S., and Takayanagi, T. (2006b). Holographic derivation of entanglement entropy from AdS/CFT. *Phys. Rev. Lett.* 96:181602. doi: 10.1103/PhysRevLett.96.181602
- Schleich, K. (1987). Conformal rotation in perturbative gravity. *Phys. Rev. D* 36:2342–2363. doi: 10.1103/PhysRevD.36.2342
- Shiu, G., and Staessens, W. (2018). Strong dynamics and natural inflation. *arXiv:1807.00620*.
- Silverstein, E., and Westphal, A. (2008). Monodromy in the CMB: gravity waves and string inflation. *Phys. Rev. D* 78:106003. doi: 10.1103/PhysRevD.78.106003
- Strominger, A. (1988). “Baby universes,” in *Theoretical Advanced Study Institute in Elementary Particle Physics: Particles, Strings and Supernovae (TASI 88)* (Providence), 315–391.
- Strominger, A. (1994). “Les Houches lectures on black holes,” in *NATO Advanced Study Institute: Les Houches Summer School, Session 62: Fluctuating Geometries in Statistical Mechanics and Field Theory* (Les Houches).
- Susskind, L. (2003). “The anthropic landscape of string theory,” in *Universe or Multiverse?* ed B. Carr (Cambridge University Press), 247–266.
- Tamvakis, K. (1989). Two axion string wormholes. *Phys. Lett. B* 233, 107–111. doi: 10.1016/0370-2693(89)90624-2
- Tamvakis, K., and Vayonakis, C. E. (1990). Axionic instantons and the creation of expanding unicerses. *Nucl. Phys. B* 329, 519–532. doi: 10.1016/0550-3213(90)90154-6
- Tanaka, T. (1999). The No - negative mode theorem in false vacuum decay with gravity. *Nucl. Phys. B* 556, 373–396. doi: 10.1016/S0550-3213(99)00369-7
- Tanaka, T., and Sasaki, M. (1992). False vacuum decay with gravity: negative mode problem. *Prog. Theor. Phys.* 88, 503–528. doi: 10.1143/ptp/88.3.503
- Tong, D. (2005). “TASI lectures on solitons: instantons, monopoles, vortices and kinks,” in *Theoretical Advanced Study Institute in Elementary Particle Physics: Many Dimensions of String Theory (TASI 2005)* (Boulder, CO).
- Twamley, J., and Page, D. N. (1992). Imaginary wormholes. *Nucl. Phys. B* 378, 247–287. doi: 10.1016/0550-3213(92)90009-Z
- Vafa, C. (2005). The string landscape and the swampland. *arXiv:hep-th/0509212*.
- Vainshtein, A. I., Zakharov, V. I., Novikov, V. A., and Shifman, M. A. (1982). ABC’s of instantons. *Sov. Phys. Usp.* 25:195. doi: 10.1070/PU1982v025n04ABEH004533
- Vandoren, S., and van Nieuwenhuizen, P. (2008). Lectures on instantons. *arXiv:0802.1862*.
- Visinelli, L., and Gondolo, P. (2009). Dark matter axions revisited. *Phys. Rev. D* 80:035024. doi: 10.1103/PhysRevD.80.035024

- Visser, M. (1995). *Lorentzian Wormholes: From Einstein to Hawking*. New York, NY; Woodbury, MN: Springer Verlag.
- Weinberg, D. H., Bullock, J. S., Governato, F., Kuzio de Naray, R., and Peter, A. H. (2015). Cold dark matter: controversies on small scales. *Proc. Nat. Acad. Sci. U.S.A.* 112, 12249–12255. doi: 10.1073/pnas.1308716112
- Weinberg, S. (1989). The cosmological constant problem. *Rev. Mod. Phys.* 61, 1–23. doi: 10.1103/RevModPhys.61.1
- Westphal, A. (2015). String cosmology — Large-field inflation in string theory. *Int. J. Mod. Phys. A* 30:1530024. doi: 10.1142/S0217751X15300240
- Wheeler, J. A. (1955). Geons. *Phys. Rev.* 97:511. doi: 10.1103/PhysRev.97.511
- Wheeler, J. A. (1967). “Superspace and the nature of quantum geometrodynamics,” in *Battelle Rencontres - 1967 Lectures in Mathematics and Physics* (Seattle), ed S. C. DeWitt and J. A. Wheeler (New York, NY: Benjamin), 242–307.
- Witten, E. (1982). Instability of the Kaluza-Klein Vacuum. *Nucl. Phys. B* 195, 481–492. doi: 10.1016/0550-3213(82)90007-4
- Witten, E. (1998). Anti-de Sitter space and holography. *Adv. Theor. Math. Phys.* 2, 253–291. doi: 10.4310/ATMP.1998.v2.n2.a2
- Yang, I.-S. (2013). Recovering the negative mode for type B Coleman–de Luccia instantons. *Phys. Rev. D* 87:084026. doi: 10.1103/PhysRevD.87.084026
- Yoshino, H., and Kodama, H. (2012). Bosenova collapse of axion cloud around a rotating black hole. *Prog. Theor. Phys.* 128, 153–190. doi: 10.1143/PTP.128.153
- Yoshino, H., and Kodama, H. (2015b). The bosenova and axiverse. *Class. Quant. Grav.* 32:214001. doi: 10.1088/0264-9381/32/21/214001
- Yoshino, H., and Kodama, H. (2015a). Probing the string axiverse by gravitational waves from Cygnus X-1. *PTEP* 2015:061E01. doi: 10.1093/ptep/ptv067
- Zouros, T. J. M., and Eardley, D. M. (1979). Instabilities of massive scalar perturbations of a rotating black hole. *Ann. Phys.* 118, 139–155. doi: 10.1016/0003-4916(79)90237-9

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Hebecker, Mikhail and Soler. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Advantages of publishing in Frontiers



## OPEN ACCESS

Articles are free to read  
for greatest visibility  
and readership



## FAST PUBLICATION

Around 90 days  
from submission  
to decision



## HIGH QUALITY PEER-REVIEW

Rigorous, collaborative,  
and constructive  
peer-review



## TRANSPARENT PEER-REVIEW

Editors and reviewers  
acknowledged by name  
on published articles

## Frontiers

Avenue du Tribunal-Fédéral 34  
1005 Lausanne | Switzerland

**Visit us:** [www.frontiersin.org](http://www.frontiersin.org)

**Contact us:** [info@frontiersin.org](mailto:info@frontiersin.org) | +41 21 510 17 00



## REPRODUCIBILITY OF RESEARCH

Support open data  
and methods to enhance  
research reproducibility



## DIGITAL PUBLISHING

Articles designed  
for optimal readership  
across devices



## FOLLOW US

[@frontiersin](https://twitter.com/frontiersin)



## IMPACT METRICS

Advanced article metrics  
track visibility across  
digital media



## EXTENSIVE PROMOTION

Marketing  
and promotion  
of impactful research



## LOOP RESEARCH NETWORK

Our network  
increases your  
article's readership