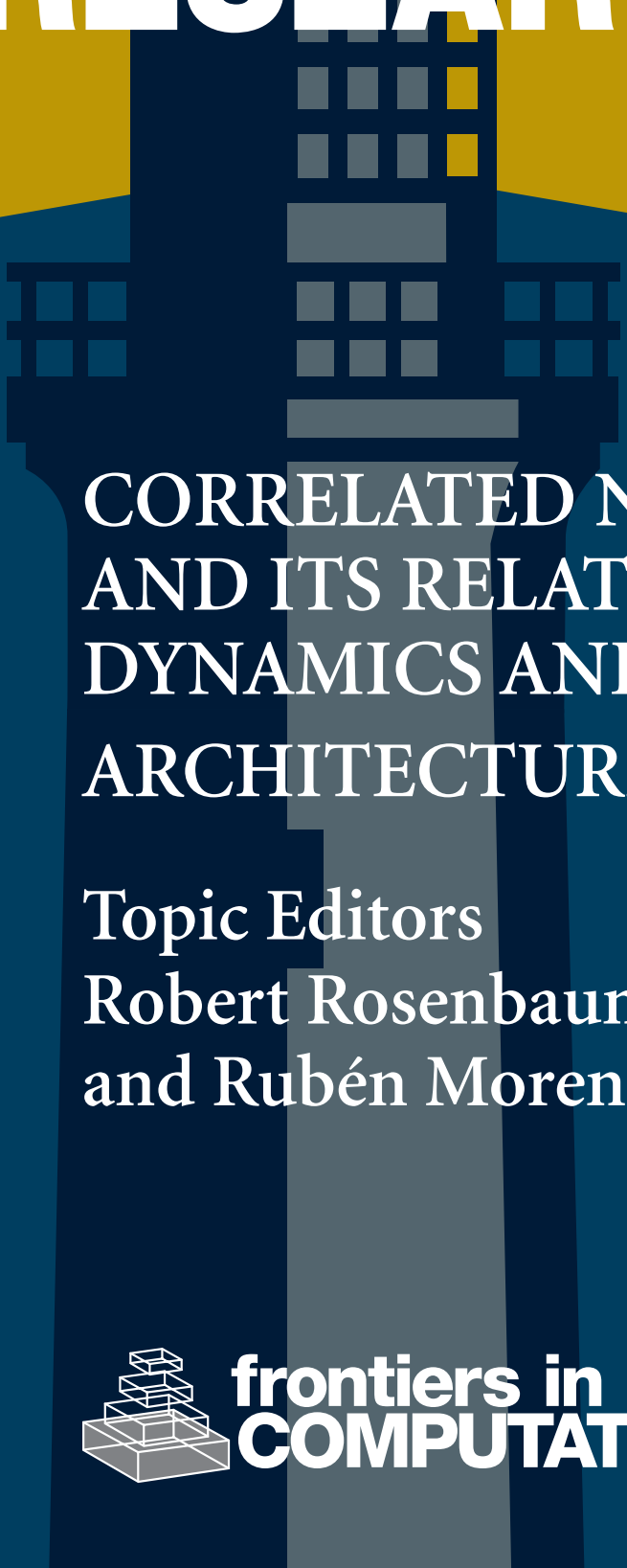
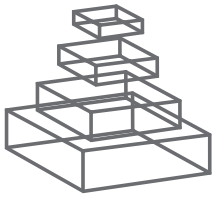
frontiers
RESEARCH TOPICS



CORRELATED NEURONAL ACTIVITY AND ITS RELATIONSHIP TO CODING, DYNAMICS AND NETWORK ARCHITECTURE

Topic Editors
Robert Rosenbaum, Tatjana Tchumatchenko and Rubén Moreno-Bote

frontiers in
COMPUTATIONAL NEUROSCIENCE



frontiers

FRONTIERS COPYRIGHT STATEMENT

© Copyright 2007-2014
Frontiers Media SA.
All rights reserved.

All content included on this site, such as text, graphics, logos, button icons, images, video/audio clips, downloads, data compilations and software, is the property of or is licensed to Frontiers Media SA ("Frontiers") or its licensees and/or subcontractors. The copyright in the text of individual articles is the property of their respective authors, subject to a license granted to Frontiers.

The compilation of articles constituting this e-book, wherever published, as well as the compilation of all other content on this site, is the exclusive property of Frontiers. For the conditions for downloading and copying of e-books from Frontiers' website, please see the Terms for Website Use. If purchasing Frontiers e-books from other websites or sources, the conditions of the website concerned apply.

Images and graphics not forming part of user-contributed materials may not be downloaded or copied without permission.

Individual articles may be downloaded and reproduced in accordance with the principles of the CC-BY licence subject to any copyright or other notices. They may not be re-sold as an e-book.

As author or other contributor you grant a CC-BY licence to others to reproduce your articles, including any graphics and third-party materials supplied by you, in accordance with the Conditions for Website Use and subject to any copyright notices which you include in connection with your articles and materials.

All copyright, and all rights therein, are protected by national and international copyright laws.

The above represents a summary only. For the full conditions see the Conditions for Authors and the Conditions for Website Use.

Cover image provided by Ibbl sarl, Lausanne CH

ISSN 1664-8714
ISBN 978-2-88919-357-8
DOI 10.3389/978-2-88919-357-8

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

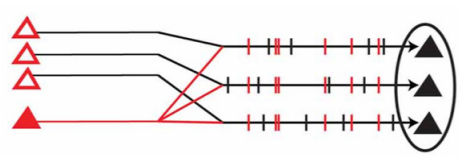
Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - 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

CORRELATED NEURONAL ACTIVITY AND ITS RELATIONSHIP TO CODING, DYNAMICS AND NETWORK ARCHITECTURE

Topic Editors:
Robert Rosenbaum, University of Notre Dame, USA
Tatjana Tchumatchenko, Max Planck Institute for Brain Research, Germany
Rubén Moreno-Bote, Parc Sanitari Sant Joan de Déu and Universitat de Barcelona, Spain
Centrode Investigación Biomédicaen Redde Salud Mental, Spain



A schematic of spike train correlations arising from shared inputs.

Correlated activity in populations of neurons has been observed in many brain regions and plays a central role in cortical coding, attention, and network dynamics. Accurately quantifying neuronal correlations presents several difficulties. For example, despite recent advances in multicellular recording techniques, the number of neurons from which spiking activity can be simultaneously recorded remains orders magnitude smaller than the size of local networks. In addition, there is a lack of consensus on the distribution of pairwise spike cross correlations obtained in extracellular multi-unit recordings. These challenges highlight the need for theoretical and computational approaches to understand how correlations emerge and to decipher their functional role in the brain.

Table of Contents

frontiers in
COMPUTATIONAL NEUROSCIENCE

EDITORIAL
published: 27 August 2014
doi: 10.3389/fncom.2014.00102

Correlated neuronal activity and its relationship to coding, dynamics and network architecture

***Robert Rosenbaum*[1,2]*, *Tatjana Tchumatchenko*[3] *and Rubén Moreno-Bote*[4,5]**

[1] *Department of Applied and Computational Mathematics and Statistics, University of Notre Dame, Notre Dame, IN, USA*
[2] *Center for the Neural Basis of Cognition, Pittsburgh, PA, USA*
[3] *Department Theory of Neural Dynamics, Max Planck Institute for Brain Research, Frankfurt am Main, Germany*
[4] *Research Unit, Parc Sanitari Sant Joan de Déu and Universitat de Barcelona, Barcelona, Spain*
[5] *Centro de Investigación Biomédica en Red de Salud Mental (CIBERSAM), Barcelona, Spain*
**Correspondence: robertr@pitt.edu*

Edited and reviewed by:
Misha Tsodyks, Weizmann Institute of Science, Israel

Keywords: neuronal correlations, neural synchrony, neural coding, spike train analysis, neuronal networks, noise correlation

Correlated and synchronous activity in populations of neurons has been observed in many brain regions and has been shown to play a crucial role in cortical coding, attention, and network dynamics (Singer and Gray, 1995; Salinas and Sejnowski, 2001). However, we still lack a detailed knowledge of the origin and function, if any, of neuronal correlations. In this Research Topic, new ideas about these long standing questions are put forward. One group of studies in this Research Topic investigates the interaction of neuronal correlations with cellular and circuit mechanisms at the level of single neurons and cell pairs. Bolhasani et al. (2013) study the interaction between direct synaptic coupling between two neurons with correlated stochastic input to the neurons. They find that excitatory synaptic coupling can alter the transfer of pairwise correlations from current input to spike output. Interestingly, there is an optimal value of synaptic coupling strength for which the sensitivity of output correlations to input correlations is maximized.

Bird and Richardson (2014) study the interaction between long term plasticity, synaptic vesicle depletion at multiple release sites and presynaptic spiking correlations. They find that there is an optimal number of release sites for driving postsynaptic spiking when synchrony is present in the presynaptic spike trains. Schwalger and Lindner (2013) investigated correlations between the interspike intervals of oscillator model neurons with adaptation. They reveal a fundamental connection between interval correlations and the phase response curve of the neuron model. They also show that when firing rates are high, negative interval correlations cause long-timescale variability of a model neuron's activity to be small.

A second group of studies in this Research Topic investigates neuronal correlations on the level of networks. The key questions that these studies addressing are: (1) How are pairwise and higher order correlations generated in networks and which of them are important for a given network? and (2) How should we uncover and interpret spike train correlations in a given dataset?

Four studies Zhou et al. (2013), Grytskyy et al. (2013), Barreiro et al. (2014), and Jahnke et al. (2013) have focused on the first question.

Zhou et al. (2013) investigated coupled pairs of neurons receiving temporally correlated input currents. They show that pairs of neurons may be more synchronized if they have some degree of heterogeneity in their intrinsic properties. Temporal correlations in the noise that these neurons receive may also promote synchrony.

Grytskyy et al. (2013) have addressed how recurrent neural networks can support the generation of pairwise correlations. The authors put forward a unified framework for the generation of pairwise correlations in recurrent networks and hypothesize that many different single model neurons, when coupled to a network, may generate the same pairwise correlation structures. Interestingly, the authors could show the equivalence of different single neuron models in a linear approximation to a model with fluctuating continuous variables. This could be a useful tool for assessing correlations across models and experiments.

In a complementary study, Barreiro et al. (2014) have focused on the emergence of pairwise and higher order correlations in retina models. The authors find that maximum entropy pairwise models capture surprisingly well the network spiking dynamics. What is surprising about these results is that higher-order correlations in this type of models can be constrained to be far lower than the statistically possible limits and that their strength depends more on the structure of the common input than on the synaptic connectivity profile.

Jahnke et al. (2013) focused on spike patterns rather than correlations and proposed a mechanism for precise spike time pattern generation and replay in neural networks that lack strong densely connected feed-forward structures. The authors put forward the hypothesis that a non-linearity in synaptic summation rules may explain the lack of observed strong feed-forward structures in live networks.

A team lead by Sonja Grün has tackled the second question, how spike correlations may be detected in a given data set. Torre et al. (2013) have extended our methodical toolbox and proposed a new method for the extraction of statistically overrepresented spike patterns that may be the functionally significant "cell assemblies" proposed by Abeles (1982). The challenge this study has taken on is to extract from large number of simultaneously recorded neurons candidate assemblies that are systematically co-activated. This search algorithm may help to reveal how

precise multi-neuron synchronization patterns that go beyond the standard pairwise analysis may relate to behavior.

In an opinion article, Zanin and Papo (2013) also address the second question. They suggest that one has to be cautious about interpreting neuronal correlations between neurons or brain areas, because typical measurements of effective connectivity might lead to false positives even when the neurons or the brain areas are indeed performing independent computations.

A third group of studies in this Research Topic addresses the computational advantages of neuronal correlations in the brain. Kilpatrick (2013) studied neuronal networks that sustain bump attractors, a well-established model for the maintenance of spatial cues in working memory tasks (Funahashi et al., 1989; Wimmer et al., 2014). In these models, the position of the bump undergoes a diffusion process, implying that the encoded memory degrades as the time progresses. Notably, Kilpatrick found that connecting several areas with similar bump attractors resulted in an increased stability of the stored memories because the variability within the areas could be averaged out. However, if the variability across areas was correlated, the diffusion of the bump attractor underwent larger variability. This study, therefore, suggests that correlated noise across neuronal areas can impoverish the precision of the encoding of spatial cues in working memory task.

In another study, Dipoppa and Gutkin (2013) found that correlations might have a positive role in working memory tasks by a mechanism that they named "correlation-induced gating." These authors and others have previously showed that correlations tend to destabilize the memory trace of an item stored in working memory. This result might suggest that correlations are deleterious for working memory, but Dipoppa and Gutkin argue that this is not the case: correlations in working memory circuits can be strongly beneficial to suppress the harmful interference of distractors, irrelevant items that do not need to be stored in memory to solve the ongoing task. This study, therefore, shows in an elegant way how changing correlations within specific neuronal population can allow for flexible gating of sensory information into working memory circuits.

Previous works have showed that synchronization between neuronal ensembles might play an important role in the binding of features belonging to a same object (Engel and Singer, 2001). In a theoretical work presented in this Research Topic, Finger and Koenig (2014) took an important step forward by showing that binding of features in natural images can be mediated by phase synchronization in a network of neural oscillators. The authors also found that the network, trained with natural images, developed small-world properties, and even allowed binding of features over long distances. This study strongly supports the idea that neuronal correlations in the brain might play an important computational role.

In a study where the LFP and single-cell activity were recorded in the hippocampal formation of epileptic patients, Alvarado-Rojas et al. (2013) found that activity of a sizable fraction of neurons preceded interictal epileptiform discharges, as measured by LFP activity.

These studies give conspicuous examples for the ambivalent nature of neuronal correlations: in some conditions correlations might be a signature of dynamic instability of the network, but in other conditions correlations might be used to perform complex and flexible computations, such as binding or information gating. Although these works have provided new clues about the role of neuronal correlations, there are yet many unsolved questions, such as how neuronal correlations are generated and propagated (Moreno et al., 2002; Moreno-Bote and Parga, 2006; de la Rocha et al., 2007; Ostojic et al., 2009; Renart et al., 2010; Rosenbaum et al., 2010, 2011; Tchumatchenko et al., 2010; Cohen and Kohn, 2011; Tchumatchenko and Wolf, 2011; Helias et al., 2014) and how correlations are shaped by limited information in sensory inputs and by neuronal computations. It is clear that the study of the impact of neuronal correlations on information transmission and brain computation, and vice versa, is still an arena for exciting new discoveries.

REFERENCES

Abeles, M. (1982). *Local Cortical Circuits: An Electrophysiological Study*. Berlin: Springer. doi: 10.1007/978-3-642-81708-3

Alvarado-Rojas, C., Lehongre, K., Bagdasaryan, J., Bragin, A., Staba, R., Engel, J., et al. (2013). Single-unit activities during epileptic discharges in the human hippocampal formation. *Front. Comput. Neurosci.* 7:140. doi: 10.3389/fncom.2013.00140

Barreiro, A., Gjorgjieva, J., Rieke, F., and Shea-Brown, E. (2014). When do microcircuits produce beyond-pairwise correlations? *Front. Comput. Neurosci.* 8:10. doi: 10.3389/fncom.2014.00010

Bird, A., and Richardson, M. (2014). Long-term plasticity determines the postsynaptic response to correlated afferents with multivesicular short-term synaptic depression. *Front. Comput. Neurosci.* 8:2. doi: 10.3389/fncom.2014.00002

Bolhasani, E., Azizi, Y., and Valizadeh, A. (2013). Direct connections assist neurons to detect correlation in small amplitude noises. *Front. Comput. Neurosci.* 7:108. doi: 10.3389/fncom.2013.00108

Cohen, M. R., and Kohn, A. (2011). Measuring and interpreting neuronal correlations. *Nat. Neurosci.* 14. 811–819. doi: 10.1038/nn.2842

de la Rocha, J., Doiron, B., Shea-Brown, E., Josić, K., and Reyes, A. (2007). Correlation between neural spike trains increases with firing rate. *Nature* 448, 802–806. doi: 10.1038/nature06028

Dipoppa, M., and Gutkin, B. S. (2013). Correlations in background activity control persistent state stability and allow execution of working memory tasks. *Front. Comput. Neurosci.* 7:108. doi: 10.3389/fncom.2013.00139

Engel, A. K., and Singer, W. (2001). Temporal binding and the neural correlates of sensory awareness. *Trends Cogn. Sci.* 5, 16–25. doi: 10.1016/S1364-6613(00)01568-0

Finger, H., and Koenig, P. (2014). Phase synchrony facilitates binding and segmentation of natural images in a coupled neural oscillator network. *Front. Comput. Neurosci.* 7:195. doi: 10.3389/fncom.2013.00195

Funahashi, S., Bruce, C. J., and Goldman-Rakic, P. S. (1989). Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *J. Neurophys.* 61, 331–349.

Grytskyy, D., Tetzlaff, T., Diesmann, M., and Helias, M. (2013). A unified view on weakly correlated recurrent networks. *Front. Comput. Neurosci.* 7:131. doi: 10.3389/fncom.2013.00131

Helias, M., Tetzlaff, T., and Diesmann, M. (2014). The correlation structure of local neuronal networks intrinsically results from recurrent dynamics. *PLoS Comput. Biol.* 10:e1003428. doi: 10.1371/journal.pcbi.1003428

Jahnke, S., Memmesheimer, R.-M., and Timme, M. (2013). Propagating synchrony in feed-forward networks. *Front. Comput. Neurosci.* 7:153. doi: 10.3389/fncom.2013.00153

Kilpatrick, Z. P. (2013). Interareal coupling reduces encoding variability in multi-area models of spatial working memory. *Front. Comput. Neurosci.* 7:82. doi: 10.3389/fncom.2013.00082

Moreno, R., de la Rocha, J., Renart, A., and Parga, N. (2002). Response of spiking neurons to correlated inputs. *Phys. Rev. Lett.* 89:288101. doi: 10.1103/PhysRevLett.89.288101

Moreno-Bote, R., and Parga, N. (2006). Auto- and cross-correlograms for the spike response of leaky integrate-and-fire neurons with slow synapses. *Phys. Rev. Lett.* 96:028101. doi: 10.1103/PhysRevLett.96.028101

Ostojic, S., Brunel, N., and Hakim, V. (2009). How connectivity, background activity, and synaptic properties shape the crosscorrelation between spike trains. *J. Neurosci.* 29, 10234–10253. doi: 10.1523/JNEUROSCI.1275-09.2009

Renart, A., de la Rocha, J., Bartho, P., Hollender, L., Parga, N., Reyes, A., et al. (2010). The asynchronous state in cortical circuits. *Science* 327, 587–590. doi: 10.1126/science.1179850

Rosenbaum, R., Trousdale, J., and Josić, K. (2011). The effects of pooling on spike train correlations. *Front. Neurosci.* 5:58. doi: 10.3389/fnins.2011.00058

Rosenbaum, R. J., Trousdale, J., and Josić, K. (2010). Pooling and correlated neural activity. *Front. Comput. Neurosci.* 4:9. doi: 10.3389/fncom.2010.00009

Salinas, E., and Sejnowski, T. (2001). Correlated neuronal activity and the flow of neural information. *Nat. Rev. Neurosci.* 2, 539–550. doi: 10.1038/35086012

Schwalger, T., and Lindner, B. (2013). Patterns of interval correlations in neural oscillators with adaptation. *Front. Comput. Neurosci.* 7:164. doi: 10.3389/fncom.2013.00164

Singer, W., and Gray, C. M. (1995). Visual feature integration and the temporal correlation hypothesis. *Annu. Rev. Neurosci.* 18, 555–586. doi: 10.1146/annurev.ne.18.030195.003011

Tchumatchenko, T., Malyshev, A., Geisel, T., Volgushev, M., and Wolf, F. (2010). Correlations and synchrony in threshold neuron models. *Phys. Rev. Lett.* 104, 058102. doi: 10.1103/PhysRevLett.104.058102

Tchumatchenko, T., and Wolf, F. (2011). Representation of dynamical stimuli in populations of threshold neurons. *PLoS Comput. Biol.* 7:e1002239. doi: 10.1371/journal.pcbi.1002239

Torre, E., Picado-Muino, D., Denker, M., Borgelt, C., and Grün, S. (2013). Statistical evaluation of synchronous spike patterns extracted by frequent item set mining. *Front. Comput. Neurosci.* 7:132. doi: 10.3389/fncom.2013.00132

Wimmer, K., Nykamp, D. Q., Constantinidis, C., and Compte, A. (2014). Bump attractor dynamics in prefrontal cortex explains behavioral precision in spatial working memory. *Nat. Neurosci.* 17, 431–439. doi: 10.1038/nn.3645

Zanin, M., and Papo, D. (2013). Efficient neural codes can lead to spurious synchronization. *Front. Comput. Neurosci.* 7, 125. doi: 10.3389/fncom.2013.00125

Zhou, P., Burton, S., Urban, N., and Ermentrout, G. (2013). Impact of neuronal heterogeneity on correlated colored noise-induced synchronization. *Front. Comput. Neurosci.* 7:113. doi: 10.3389/fncom.2013.00113

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.

Received: 15 July 2014; accepted: 07 August 2014; published online: 27 August 2014.
Citation: Rosenbaum R, Tchumatchenko T and Moreno-Bote R (2014) Correlated neuronal activity and its relationship to coding, dynamics and network architecture. Front. Comput. Neurosci. ***8****:102. doi: 10.3389/fncom.2014.00102*
This article was submitted to the journal Frontiers in Computational Neuroscience.
Copyright © 2014 Rosenbaum, Tchumatchenko and Moreno-Bote. 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) or licensor 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.

frontiers in
COMPUTATIONAL NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 06 February 2014
doi: 10.3389/fncom.2014.00010

When do microcircuits produce beyond-pairwise correlations?

Andrea K. Barreiro[1]*†, Julijana Gjorgjieva[2†], Fred Rieke[3] and Eric Shea-Brown[1,3]

[1] Department of Applied Mathematics, University of Washington, Seattle, WA, USA
[2] Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge, UK
[3] Department of Physiology and Biophysics, University of Washington, Seattle, WA, USA

Edited by:
Robert Rosenbaum, University of Pittsburgh, USA

Reviewed by:
Tim Gollisch, University Medical Center Göttingen, Germany
Tatjana Tchumatchenko, Max Planck Institute for Brain Research, Germany

***Correspondence:**
Andrea K. Barreiro, Department of Mathematics, Southern Methodist University, 3200 Dyer Street, PO Box 750156, Dallas, TX 75275-0156, USA
e-mail: abarreiro@smu.edu

†Present address:
Andrea K. Barreiro, Department of Mathematics, Southern Methodist University, Dallas, USA;
Julijana Gjorgjieva, Center for Brain Science, Harvard University, Cambridge, USA

Describing the collective activity of neural populations is a daunting task. Recent empirical studies in retina, however, suggest a vast simplification in how multi-neuron spiking occurs: the activity patterns of retinal ganglion cell (RGC) populations under some conditions are nearly completely captured by pairwise interactions among neurons. In other circumstances, higher-order statistics are required and appear to be shaped by input statistics and intrinsic circuit mechanisms. Here, we study the emergence of higher-order interactions in a model of the RGC circuit in which correlations are generated by common input. We quantify the impact of higher-order interactions by comparing the responses of mechanistic circuit models vs. "null" descriptions in which all higher-than-pairwise correlations have been accounted for by lower order statistics; these are known as pairwise maximum entropy (PME) models. We find that over a broad range of stimuli, output spiking patterns are surprisingly well captured by the pairwise model. To understand this finding, we study an analytically tractable simplification of the RGC model. We find that in the simplified model, bimodal input signals produce larger deviations from pairwise predictions than unimodal inputs. The characteristic light filtering properties of the upstream RGC circuitry suppress bimodality in light stimuli, thus removing a powerful source of higher-order interactions. This provides a novel explanation for the surprising empirical success of pairwise models.

Keywords: retinal ganglion cells, maximum entropy distribution, stimulus-driven, correlations, computational model

1. INTRODUCTION

Information in neural circuits is often encoded in the activity of large, highly interconnected neural populations. The combinatoric explosion of possible responses of such circuits poses major conceptual, experimental, and computational challenges. How much of this potential complexity is realized? What do statistical regularities in population responses tell us about circuit architecture? Can simple circuit models with limited interactions among cells capture the relevant information content? These questions are central to our understanding of neural coding and decoding.

Two developments have advanced studies of synchronous activity in recent years. First, new experimental techniques provide access to responses from the large groups of neurons necessary to adequately sample synchronous activity patterns (Baudry and Taketani, 2006). Second, maximum entropy approaches from statistical physics have provided a powerful approach to distinguish genuine higher-order synchrony (correlations) from that explainable by pairwise statistical interactions among neurons (Martignon et al., 2000; Amari, 2001; Schneidman et al., 2003). These approaches have produced diverse findings. In some instances, activity of neural populations is extremely well described by pairwise interactions alone, so that pairwise maximum entropy (PME) models provide a nearly complete description (Shlens et al., 2006, 2009). In other cases, while pairwise models bring major improvements over independent descriptions, it is not clear that they fully capture the data (Martignon et al., 2000; Schneidman et al., 2006; Tang et al., 2008; Yu et al., 2008; Montani et al., 2009; Ohiorhenuan et al., 2010; Santos et al., 2010). Empirical studies indicate that pairwise models can fail to explain the responses of spatially localized triplets of cells (Ohiorhenuan et al., 2010; Ganmor et al., 2011), as well as the activity of populations of ∼100 cells responding to natural stimuli (Ganmor et al., 2011). Overall, the diversity of empirical results highlights the need to understand the network and input features that control the statistical complexity of synchronous activity patterns.

Several themes have emerged from efforts to link the correlation structure of spiking activity to circuit mechanisms using both abstract (Amari et al., 2003; Krumin and Shoham, 2009; Macke et al., 2009; Roudi et al., 2009a) and biologically-based models (Bohte et al., 2000; Martignon et al., 2000; Roudi et al., 2009b); these models, however, do not provide a full description for why the PME models succeed or fail to capture neural circuit dynamics. First, thresholding non-linearities in circuits with Gaussian input signals can generate correlations that cannot be explained by pairwise statistics (Amari et al., 2003); the deviations from pairwise predictions are modest at moderate population sizes (Macke et al., 2009), but may become severe as population size grows large (Amari et al., 2003; Macke et al., 2011). The pairwise model also fails in networks of recurrent integrate-and-fire

units with adapting thresholds and refractory potassium currents (Bohte et al., 2000). The same is true for "Boltzmann-type" networks with hidden units (Koster et al., 2013). Finally, small groups of model neurons that perform logical operations can be shown to generate higher-order interactions by introducing noisy processes with synergistic effects (Schneidman et al., 2003), but it is unclear what neural mechanisms might produce similar distributions. These diverse findings point to the important role that circuit features and mechanisms—input statistics, input/output relationships, and circuit connectivity—can play in regulating higher-order interactions. Nevertheless, we lack a systematic understanding that links these features and their combinations to the success and failure of pairwise statistical models.

A second theme that has emerged is the use of perturbation approaches to explain why maximum entropy models with purely pairwise interactions capture circuit behavior in the limit in which the population firing rate is very low (i.e., the total number of firing events from all cells in the same small time window is small) (Cocco et al., 2009; Roudi et al., 2009a; Tkacik et al., 2009). Also in this regime, higher-order interactions cannot be introduced as an artifact of under-sampling the network (Tkacik et al., 2009), a concern at higher population firing rates. However, the low to moderate population firing rates observed in many studies permit *a priori* a fairly broad range in the quality of pairwise fits. What is left to explain then is why circuits operating outside the low population firing rate regime often produce fits consistent with the PME model.

We approach this issue here by systematically characterizing the ability of PME models to capture the responses of a class of circuit models with the following defining features. First, we consider relatively small circuits of 3–16 cells, each with identical intrinsic dynamics (i.e., spike-generating mechanism and level of excitability). Second, we assume a particular structure for inputs across the circuit. Each neuron receives the same global input which, for example, represents stimuli in the receptive fields of all modeled cells. Neurons also receive an independent, Gaussian-like noise term. Third, the circuit has either no reciprocal coupling, or has all-to-all excitatory or gap junction coupling. We begin with circuit models fully constrained by measured properties of primate ON parasol ganglion networks, receiving full-field and checkerboard light inputs. We then explore a simple thresholding model for which we exhaustively search over the entire parameter space.

We identify general principles that describe higher-order spike correlations in the circuits we study. First, in all cases we examined, the overall strength of higher-order correlations are constrained to be far lower than the statistically possible limits. Second, for the higher-order correlations that do occur, the primary factor that determines how significant they will be is the bimodal vs. unimodal profile of the common input signal. A secondary factor is the strength of recurrent coupling, which has a non-monotonic impact on higher-order correlations. Our findings provide insight into why some previously measured activity patterns are well captured by PME descriptions, and provide predictions for the mechanisms that allow for higher-order spike correlations to emerge.

2. RESULTS

2.1. QUANTIFYING HIGHER-ORDER CORRELATIONS IN NEURAL CIRCUITS

One strategy to identify higher-order interactions is to compare multi-neuron spike data against a description in which any higher-order interactions have been removed in a principled way—that is, a description in which all higher-order correlations are completely described by lower-order statistics. Such a description may be given by a maximum entropy model (Jaynes, 1957a,b; Amari, 2001), in which one identifies the most unstructured, or maximum entropy, distribution consistent with the constraints. Comparing the predicted and measured probabilities of different responses tests whether the constraints used are sufficient to explain observed network activity, or whether additional constraints need to be considered. Such constraints would produce additional structure in the predicted response distribution, and hence lower the entropy.

A common approach is to limit the constraints to a given statistical order—for example, to consider only the first and second moments of the distributions, which are determined by the mean and pairwise interactions. In the context of spiking neurons, we denote $\mu_i \equiv \mathbf{E}[x_i]$ as the firing rate of neuron i and $\hat{\rho}_{ij} \equiv \mathbf{E}[x_i x_j]$ as the joint probability that neurons i and j will fire. The distribution with the largest entropy for a given μ_i and $\hat{\rho}_{ij}$ is referred to as the *PME* model.

We use the Kullback–Leibler divergence, $D_{\mathrm{KL}}(P, \tilde{P})$, to quantify the accuracy of the PME approximation $\tilde{P}$ to a distribution P. This measure has a natural interpretation as the contribution of higher-order interactions to the response entropy $S(P)$ (Amari, 2001; Schneidman et al., 2003), and may in this context be written as the difference of entropies $S(\tilde{P}) - S(P)$. In addition, $D_{\mathrm{KL}}(P, \tilde{P})$ is approximately $-\log_2 L$, where L is the average likelihood (over different observations) that a sequence of data drawn from the distribution P was instead drawn from the model $\tilde{P}$ (Cover and Thomas, 1991; Shlens et al., 2006). For example, if $D_{\mathrm{KL}}(P, \tilde{P}) = 1$, the average likelihood that a single sample, i.e., a single network response, came from $\tilde{P}$ relative to the likelihood that it came from P is 2^{-1} (we use the base 2 logarithm in our definition of the Kullback–Leibler divergence, so all numerical values are in units of bits).

An alternative measure of the quality of the pairwise model comes from normalizing $D_{\mathrm{KL}}(P, \tilde{P})$ by the corresponding distance of the distribution P from an *independent maximum entropy* fit $D_{\mathrm{KL}}(P, P_1)$, where P_1 is the highest entropy distribution consistent with the mean firing rates of the cells (equivalently, the product of single-cell marginal firing probabilities) (Amari, 2001). Many studies (Schneidman et al., 2006; Shlens et al., 2006, 2009; Roudi et al., 2009a) use

$$\Delta = 1 - \frac{D_{\mathrm{KL}}\left(P, \tilde{P}\right)}{D_{\mathrm{KL}}\left(P, P_1\right)}; \tag{1}$$

a value of $\Delta = 1$ indicates that the pairwise model perfectly captures the additional information left out of the independent model, while a value of $\Delta = 0$ indicates that the pairwise model

gives no improvement over the independent model. To aid comparison with other studies, we report values of Δ in parallel with $D_{\mathrm{KL}}(P, \tilde{P})$ when appropriate.

We next explore and interpret the achievable range of $D_{\mathrm{KL}}(P, \tilde{P})$ values. The problem is made simpler if, following previous studies (Bohte et al., 2000; Amari, 2001; Macke et al., 2009; Montani et al., 2009), we consider only permutation-symmetric spiking patterns, in which the firing rate and correlation do not depend on the identity of the cells; i.e., $\mu_i = \mu$, $\hat{\rho}_{ij} = \hat{\rho}$ for $i \neq j$. We start with three cells having binary responses and assume that the response is stationary and uncorrelated in time. From symmetry, the possible network responses are

$$
\begin{aligned}
p_0 &= P[(0,0,0)] \\
p_1 &= P[(1,0,0)] = P[(0,1,0)] = P[(0,0,1)] \\
p_2 &= P[(1,1,0)] = P[(1,0,1)] = P[(0,1,1)] \\
p_3 &= P[(1,1,1)],
\end{aligned}
$$

where p_i denotes the probability that a particular set of i cells spike and the remaining $3 - i$ do not. Possible values of (p_0, p_1, p_2, p_3) are constrained by the fact that P is a probability distribution, so that the sum of p_i over all eight states is one.

To assess the numerical significance of $D_{\mathrm{KL}}(P, \tilde{P})$, we can compare it with the maximal achievable value for any symmetric distribution on three spiking cells. For three cells, the maximal value is $D_{\mathrm{KL}}(P, \tilde{P}) = 1$ (or 1/3 bits per neuron), achieved by the XOR operation (Schneidman et al., 2003). This distribution is illustrated in **Figure 1A** (right), together with two distributions produced by our mechanistic circuit models—illustrating observed deviations from PME fits for unimodal (left) and bimodal (middle) distributions of inputs (see below). The *KL*-divergence for these two patterns is 0.0013 and 0.091, respectively. As suggested by these bar plots (and explored in detail below), the distributions produced by a wide set of mechanistic circuit models are quite well captured by the PME approximation: to use the likelihood interpretation described above, an observer would need to draw many more samples from these distributions in order to distinguish between the true and model distributions: ≈1000 times and ≈10 times, respectively, in comparison to the XOR operator.

To further identify appropriate "benchmark" values of $D_{\mathrm{KL}}(P, \tilde{P})$ with which to compare our mechanistic circuit models, in **Figure 1B** we show plots of $D_{\mathrm{KL}}(P, \tilde{P})$ and Δ vs. firing rate produced by an exhaustive sampling of symmetric distributions on three cells. From this picture, we can see that it is possible to find symmetric, three-cell spiking distributions that are poorly fit by the pairwise model at a range of firing rates and pairwise correlations, with the largest values of $D_{\mathrm{KL}}(P, \tilde{P})$ found at low correlations (note that the XOR distribution has an average pairwise covariance of zero (i.e., $\mathbf{E}[X_1 X_2] = \mathbf{E}[X_1]\,\mathbf{E}[X_2]$)).

2.1.1. A condition for higher-order correlations

Possible solutions to the symmetric PME problem take the form of exponential functions characterized by two parameters, λ_1 and λ_2, which serve as Lagrange multipliers for the constraints:

$$
P[(x_1, x_2, x_3)] = \frac{1}{Z} \exp[\lambda_1 (x_1 + x_2 + x_3) + \lambda_2 (x_1 x_2 + x_2 x_3 + x_1 x_3)]. \qquad (2)
$$

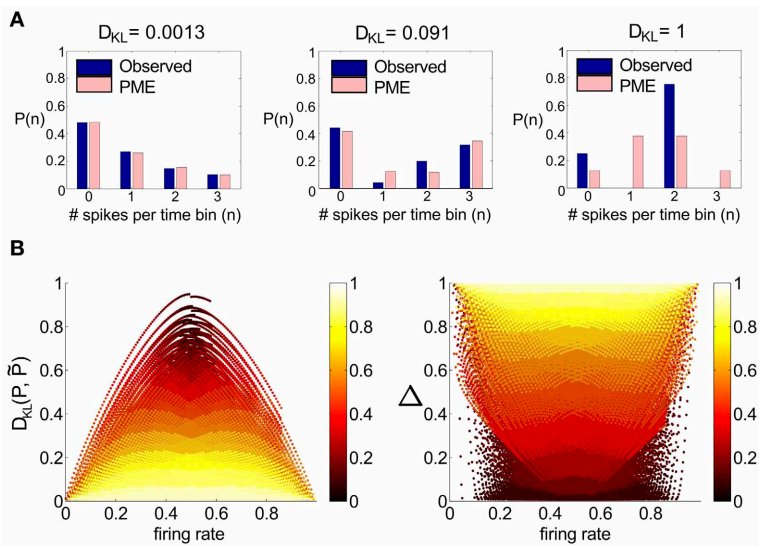


FIGURE 1 | A survey of the quality of the pairwise maximum entropy (PME) model for symmetric spiking distributions on three cells. (A) Probability distribution P (dark blue) and pairwise approximation $\tilde{P}$ (light pink) for three example distributions. From left to right: an example from the simple sum-and-threshold model receiving skewed common input; an example from the sum-and-threshold model receiving bimodal common input [specifically, the distribution with maximal $D_{\mathrm{KL}}(P, \tilde{P})$]; a specific probability distribution resulting from application of the XOR operator [for illustration of a "worst case" fit of the PME model (Schneidman et al., 2003)]. **(B)** $D_{\mathrm{KL}}(P, \tilde{P})$ vs. firing rate and Δ vs. firing rate, for a comprehensive survey of possible symmetric spiking distributions on three cells (see text for details). Firing rate is defined as the probability of a spike occurring per cell per random draw of the sum-and-threshold model, as defined in Equation (16). Color indicates output correlation coefficient ρ ranging from black for $\rho \in (0, 0.1)$, to white for $\rho \in (0.9, 1)$, as illustrated in the color bars.

The factor Z normalizes P to be a probability distribution.

By combining individual probabilities of events as given by Equation (2) the following relationship must be satisfied by any symmetric PME solution:

$$\frac{p_3}{p_0} = \left(\frac{p_2}{p_1}\right)^3. \tag{3}$$

This is equivalent to the condition that the *strain* measure of Ohiorhenuan and Victor (2010) be zero (in particular, the strain is negative whenever $p_3/p_0 - (p_2/p_1)^3 < 0$, a condition identified in Ohiorhenuan and Victor (2010) as corresponding to sparsity in the neural code).

For three-cell, symmetric networks, models that exactly satisfy Equation (3) will also be exactly described via PME. Moreover, note that probability models that meet this constraint fall on a surface in the space of (normalized) histograms, given by the probabilities p_j. One can verify by straightforward calculations (see Appendix) that—given fixed lower order moments—$D_{\mathrm{KL}}(P, \tilde{P})$ is a convex function of the probabilities p_j. This has interesting consequences for predicting when large vs. small values of $D_{\mathrm{KL}}(P, \tilde{P})$ will be found (see Appendix).

It is not necessary to assume permutation symmetry when deriving the PME fit $\tilde{P}$ to an observed distribution P, or in computing derived quantities such as $D_{\mathrm{KL}}(P, \tilde{P})$, and we do not do so in this study. However, most of the distributions we study are derived from mechanistic models that are themselves symmetric or near-symmetric. Therefore, we anticipate that the simplified calculations for permutation-symmetric distributions will yield analytical insight into our findings.

2.2. MECHANISMS THAT IMPACT BEYOND-PAIRWISE CORRELATIONS IN TRIPLETS OF ON-PARASOL RETINAL GANGLION CELLS

Having established the range of beyond-pairwise correlations that are possible statistically, we turn our focus to coding in retinal ganglion cell (RGC) populations, an area that has received a great deal of attention empirically. Specifically, PME approaches have been effective in capturing the activity of small RGC populations (Schneidman et al., 2006; Shlens et al., 2006, 2009). This success does not have an obvious anatomical correlate; there are multiple opportunities in the retinal circuitry for interactions among three or more ganglion cells. We explored circuits composed of three RGC cells with input statistics, recurrent connectivity and spike-generating mechanisms based directly on experiment. We based our model on ON parasol RGCs, one of the RGC types for which PME approaches have been applied extensively (Shlens et al., 2006, 2009). In addition, by examining how marginal input statistics are shaped by stimulus filtering, we also reveal the role that the specific filtering properties of ON parasol cells have in shaping higher-order interactions.

2.2.1. RGC model

We modeled a single ON parasol RGC in two stages (for details see section 4). First, we characterized the light-dependent excitatory and inhibitory synaptic inputs to cell k ($g_k^{\mathrm{exc}}(t)$, $g_k^{\mathrm{inh}}(t)$) in response to randomly fluctuating light inputs $s_k(t)$ via a linear-nonlinear model, e.g.,:

$$g_k^{\mathrm{exc}}(t) = N^{\mathrm{exc}}\left[L^{\mathrm{exc}} * s_k(t) + \eta_k^{\mathrm{exc}}\right], \tag{4}$$

where N^{exc} is a static non-linearity, L^{exc} is a linear filter, and η_k^{exc} is an effective input noise that captures variability in the response to repetitions of the same time-varying stimulus. These parameters were determined from fits to experimental data collected under conditions similar to those in which PME models have been tested empirically (Shlens et al., 2006, 2009; Trong and Rieke, 2008). The modeled excitatory and inhibitory conductances captured many of the statistical features of the real conductances, particularly the correlation time and skewness (data not shown).

Second, we used Equation (4) and an equivalent expression for $g_k^{\mathrm{inh}}(t)$ as inputs to an integrate-and-fire model incorporating a non-linear voltage and history-dependent term to account for refractory interactions between spikes (Badel et al., 2007, 2008). The voltage evolution equation was of the form

$$\frac{dV}{dt} = F(V, t - t_{\mathrm{last}}) + \frac{I_{\mathrm{input}}(t)}{C}, \tag{5}$$

where $F(V, t - t_{\mathrm{last}})$ was allowed to depend on the time of the last spike t_{last}. Briefly, we obtained data from a dynamic clamp experiment (Sharpe et al., 1993; Murphy and Rieke, 2006) in which currents corresponding to $g^{\mathrm{exc}}(t)$ and $g^{\mathrm{inh}}(t)$ were injected into a cell and the resulting voltage response measured. The input current I_{input} injected during one time step was determined by scaling the excitatory and inhibitory conductances by driving forces based on the measured voltage in the previous time step; that is,

$$I_{\mathrm{input}}(t) = -g^{\mathrm{exc}}(t)(V - V_E) - g^{\mathrm{inh}}(t)(V - V_I), \tag{6}$$

We used this data to determine F and C using the procedure described in Badel et al. (2007); details, including values of all fitted parameters, are described in section 4. Recurrent connections were implemented by adding an input current proportional to the voltage difference between the two coupled cells.

The prescription above provided a flexible model that allowed us to study the responses of three-cell RGC networks to a wide range of light inputs and circuit connectivities. Specifically, we simulated RGC responses to light stimuli that were (1) constant, (2) time-varying and spatially uniform, and (3) varying in both space and time. Correlations between cell inputs arose from shared stimuli, from shared noise originating in the retinal circuitry (Trong and Rieke, 2008), or from recurrent connections (Dacey and Brace, 1992; Trong and Rieke, 2008). Shared stimuli were described by correlations among the light inputs s_k. Shared noise arose via correlations in η_k^{exc} and η_k^{ink} as described in section 4. The recurrent connections were chosen to be consistent with observed gap-junctional coupling between ON parasol cells. We also investigated how stimulus filtering by L^{exc} and L^{inh} influenced network statistics. To compare our results with empirical studies, constant light, and spatially and temporally fluctuating checkerboard stimuli were used as in Shlens et al. (2006, 2009).

2.2.2. The feedforward RGC circuit is well-described by the PME model for full-field light stimuli

We start by considering networks without recurrent connectivity and with constant, full-field (i.e., spatially uniform) light stimuli. Thus, we set $s_k(t) = 0$ for $k = 1, 2, 3$, so that the cells received only Gaussian correlated noise η_k^{exc} and η_k^{inh} and constant excitatory and inhibitory conductances. Time-dependent conductances were generated and used as inputs to a simulation of three model RGCs. Simulation length was sufficient to ensure significance of all reported deviations from PME fits (see section 4). We found that the spiking distributions were strikingly well-modeled by a PME fit, as shown in the righthand panel of **Figure 2A**; $D_{\text{KL}}\left(P, \tilde{P}\right)$ is 2.90×10^{-5} bits. This result is consistent with the very good fits found experimentally in Shlens et al. (2006) under constant light stimulation.

Next, we introduce temporal modulation into the full-field light stimuli such that each cell received the same stimulus, $s_k(t) = s(t)$, where $s(t)$ refreshed every few milliseconds with an independently chosen value from one of several marginal distributions. For our initial set of experiments, the marginal distribution was either Gaussian (as in Ganmor et al., 2011) or binary (as used in Shlens et al., 2006). For both choices, we explored inputs with a range of standard deviations (1/16, 1/12, 1/8, 1/6, 1/4, 1/3, or 1/2 of a baseline light intensity) and refresh rates (8, 40, or 100 ms). The shared stimulus produced strong pairwise correlation between conductances of neighboring cells. However, values of $D_{\text{KL}}(P, \tilde{P})$ remained small, under 10^{-2} bits in all conditions tested.

2.2.3. Impact of stimulus spatial scale

We next asked whether PME models capture RGC responses to stimuli with varying spatial scales. We fixed stimulus dynamics to match the two cases that yielded the highest $D_{\text{KL}}(P, \tilde{P})$ under the full-field protocol: for both Gaussian and binary stimuli, we used 8 ms refresh rate and $\sigma = 1/2$. The stimulus was generated as a random checkerboard with squares of variable size; each square in the checkerboard, or *stixel*, was drawn independently from the appropriate marginal distribution and updated at the corresponding refresh rate. The conductance input to each RGC was then given by convolving the light stimulus with its receptive field, where the stimulus was positioned with a fixed rotation and translation relative to the receptive fields. This position was drawn randomly at the beginning of each simulation and held constant throughout (see insets of **Figures 3B,C** for examples, and section 4 for further details).

The RGC spike patterns remained very well described by PME models for the full range of spatial scales. **Figure 3A** shows this by plotting $D_{\text{KL}}(P, \tilde{P})$ vs. stixel size. Values of $D_{\text{KL}}(P, \tilde{P})$ increased with spatial scale, sharply rising beyond 128 μm, where a stixel had approximately the same size as a receptive field center, illustrating that introducing spatial scale via stixels produces even closer fits by PME models (the points at 512 μm correspond to the full-field simulations).

Values reported in **Figure 3A** are *averages* of $D_{\text{KL}}(P, \tilde{P})$ produced by five random stimulus positions. At stixel sizes of 128 μm and 256 μm, the resulting spiking distributions differed significantly from position to position; in **Figure 3B**, we show the probabilities of the distinct singlet [e.g., $P(1, 0, 0)$] and doublet [e.g., $P(1, 1, 0)$] spiking events produced at 256 μm. Each stimulus position created a "cloud" of dots (identified by color); large dots show the average over 20 sub-simulations. Each sub-simulation was identified by a small dot of the same color; because the simulations were very well-resolved, most of them were contained within the large dots (and hence not visible in the figure). Heterogeneity across stimulus positioning is indicated by the distinct positioning of differently colored dots. At smaller spatial scales, the process of averaging stimuli over the receptive fields resulted in spiking distributions that were largely unchanged with stimulus position, as shown in **Figure 3C**, where singlet and doublet spiking probabilities are plotted for 60 μm stixels. Thus, filtered light inputs were largely homogeneous from cell to cell, as each receptive field sampled a similar number of independent, statistically identical inputs; the inset of **Figure 3C** shows the projection of input stixels onto cell receptive fields from an example with 60 μm stixels. The resulting excitatory conductances and spiking patterns were very close to cell-symmetric (see **Figures S2B,C**).

By contrast, spiking patterns showed significant heterogeneity from cell to cell when the stixel size was large, as illustrated in **Figure 3B**. This arises because each cell in the population may be located differently with respect to stixel boundaries, and therefore receive a distinct pattern of input activity; this is illustrated by the inset of **Figure 3B**, which shows the projection of input stixels onto cell receptive fields from one such simulation. However, PME models gave excellent fits to data regardless of heterogeneity in RGC responses (see **Figures S2E,F**); as seen in **Figure 3A**, over all 20 sub-simulations, and over all individual stixel positions, we found a maximal $D_{\text{KL}}(P, \tilde{P})$ value of 0.00811.

2.2.4. Conductance profiles and impact of stimulus filtering

Intrigued by the consistent finding of low values of $D_{\text{KL}}(P, \tilde{P})$ from the RGC model circuit despite stimulation by a wide variety of highly correlated stimulus classes, we sought to further characterize the processing of light stimuli by this circuit. In particular, we examined the effects of different marginal statistics of light stimuli, standard deviation of full-field flicker, and refresh rate on the marginal distributions of excitatory conductances. We focused on excitatory conductances because they exhibit stronger correlations than inhibitory conductances in ON parasol RGCs (Trong and Rieke, 2008).

With constant light stimulation (no temporal modulation) the excitatory conductances were unimodal and broadly Gaussian (**Figure 2A**, middle panel). For a short refresh rate (8 ms) or small flicker size (standard deviation 1/6 or 1/4 of baseline light intensity), temporal averaging via the filter L^{exc} and the approximately linear form of N^{exc} over these light intensities produced a unimodal, modestly skewed distribution of excitatory conductances, regardless of whether the flicker was drawn from a Gaussian or binary distribution (see **Figures 2B,C**, center panels). For a slower refresh rate (100 ms) and large flicker size (s.d. 1/3 or 1/2 of baseline light intensity), excitatory conductances had multi-modal and skewed features, again regardless of whether the flicker was drawn from a Gaussian or binary distribution (**Figure 2D**). Other parameters being equal, binary light input

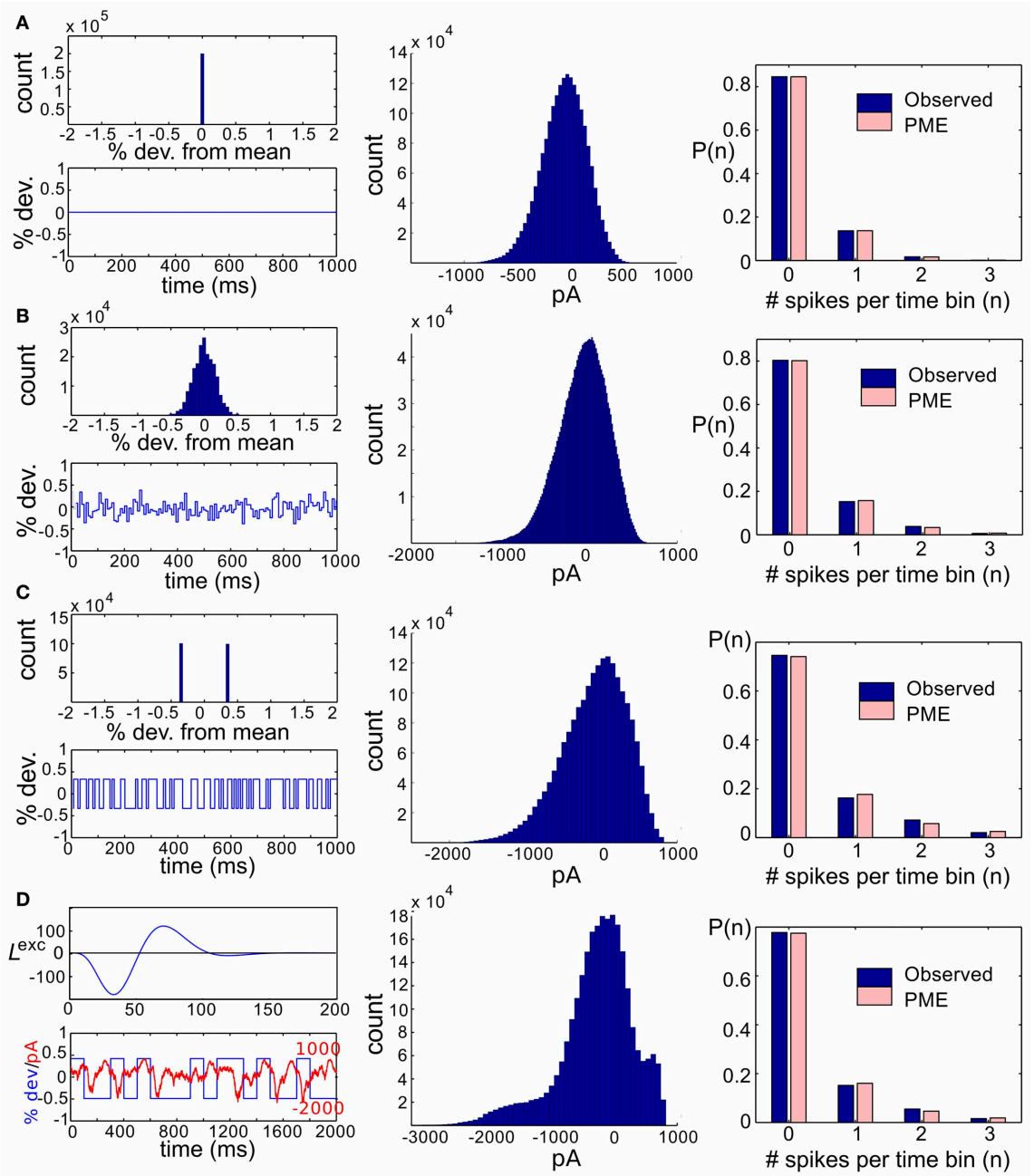


FIGURE 2 | Results for RGC simulations with constant light and full-field flicker. (A–C) (Left) A histogram and time series of stimulus, (center) a histogram of excitatory conductances and (right) the resulting distribution of spiking patterns. Stimuli are shown as deviations from a baseline intensity, expressed as a fraction of the baseline. Right panels show the probability distribution on spiking patterns P obtained from simulation ("Observed"; dark blue), and the corresponding pairwise approximation $\tilde{P}$ ("PME"; light pink). Each row gives these results for a different stimulus condition. **(A)** No stimulus (Gaussian noise only). **(B)** Gaussian input, standard deviation 1/6, refresh rate 8 ms. **(C)** Binary input, standard deviation 1/3, refresh rate 8 ms. **(D)** Binary input, standard deviation 1/3, refresh rate 100 ms. For panel **(D)**, the data in the left panel differs. (Left, top panel) The excitatory filter $L^{exc}(t)$ (Equation 7) is shown instead of a stimulus histogram; (Left, bottom panel) the normalized excitatory conductance, as a function of time (red dashed line), is superimposed on the stimulus (blue solid). (Center) The histogram of excitatory conductances and (right) the resulting distribution of spiking patterns. Both the form of the filter and the conductance trace illustrate that the LN model that processes light input acts as a (time-shifted) high pass filter.

produced more skewed conductances. While some conductance distributions had multiple local maxima, these were never well separated, with the envelope of the distribution still resembling a skewed distribution.

The mechanism that leads to unimodal distributions of conductances, even when light stimuli are binary, is high-pass filtering—a consequence of the differentiating linear filter in Equation (7) and illustrated in **Figure 2D**. To demonstrate this,

we constructed an alternative filter with a more monophasic shape [Equation (9), illustrated in **Figure S1**] and compared the excitatory conductance distributions side-by-side. We saw a striking difference in the response to long time scale, binary stimuli: the distributions produced by the monophasic filter reflected the bimodal shape of the input. Interestingly, the resulting simulation produced eight-times greater $D_{\mathrm{KL}}(P, \tilde{P})$ (**Figure 4**). This suggests that greater $D_{\mathrm{KL}}(P, \tilde{P})$ may occur when ganglion cell inputs are primarily characterized via monophasic filters, e.g., at low mean light levels for which the retinal circuit acts to primarily integrate, rather than differentiate over time.

In **Figure 4A**, we examine this effect over all full-field stimulus conditions by plotting $D_{\mathrm{KL}}(P, \tilde{P})$ from simulations with the monophasic filter, against $D_{\mathrm{KL}}(P, \tilde{P})$ from simulations in which the original filter was used with the same stimulus type. An increase in $D_{\mathrm{KL}}(P, \tilde{P})$ was observed across stimulus conditions, with a markedly larger effect for longer refresh rates. This consistent change could not be attributed to changes in lower order statistics; there was no consistent relationship between the change in pairwise model performance and either firing rate or pairwise correlations (data not shown). Instead, large effects in D_{KL} were accompanied by a striking increase in the bi- or multi-modality of excitatory conductances (see **Figure 4B**). In **Figure 4C**, we show an example stimulus and excitatory current trace taken from the simulation shown in **Figure 4B**: the monophasic filter allows the excitatory synaptic currents to track a long-timescale, bimodal stimulus with higher fidelity, transferring the bimodality of the stimulus into the synaptic currents. This finding was robust to specifics of the filtering process; we were able to reproduce the same results by designing integrating filters in different ways (data not shown).

2.2.5. Recurrent connectivity in the RGC circuit

We next considered the role of recurrence in shaping higher-order interactions by incorporating gap junction coupling into our simulations. We did this separately for each full-field stimulus

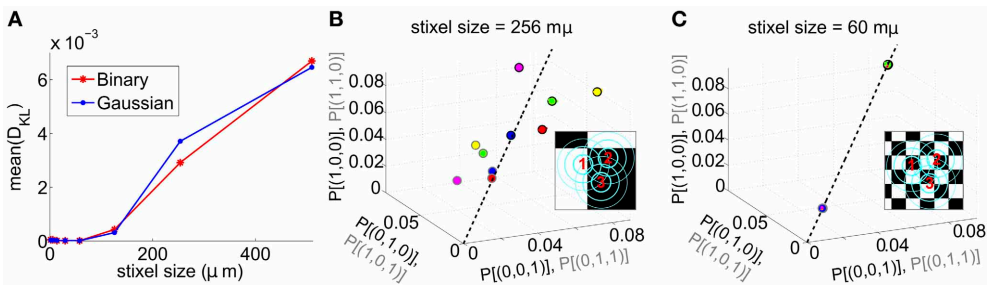


FIGURE 3 | Results for RGC simulations with light stimuli of varying spatial scale ("stixels"). **(A)** Average $D_{\mathrm{KL}}(P, \tilde{P})$ as a function of stixel size. Values were averaged over five stimulus positions, each with a different (random) stimulus rotation and translation; 512 μm corresponds to full-field stimuli. For the rest of the panels, data from the binary light distributions is shown; results from the Gaussian case are similar. **(B,C)** Probability of singlet and doublet spiking events, under stimulation by movies of 256 μm **(B)** and 60 μm **(C)** stixels. Event probabilities are plotted in 3-space, with the *x*, *y*, and *z* axes identifying the singlet (doublet) events 001 (011), 010 (101), and 100 (110), respectively. The black dashed line indicates perfect cell-to-cell homogeneity (e.g., $P[(1, 0, 0)] = P[(0, 1, 0)] = P[(0, 0, 1)]$). Both individual runs (dots) and averages over 20 runs (large circles) are shown, with averages outlined in black (singlet) and gray (doublet). Different colors indicate different stimulus positions. Insets: contour lines of the three receptive fields (at the 1 and 2 SD contour lines for the receptive field center; and at the zero contour line) superimposed on the stimulus checkerboard (for illustration, pictured in an alternating black/white pattern).

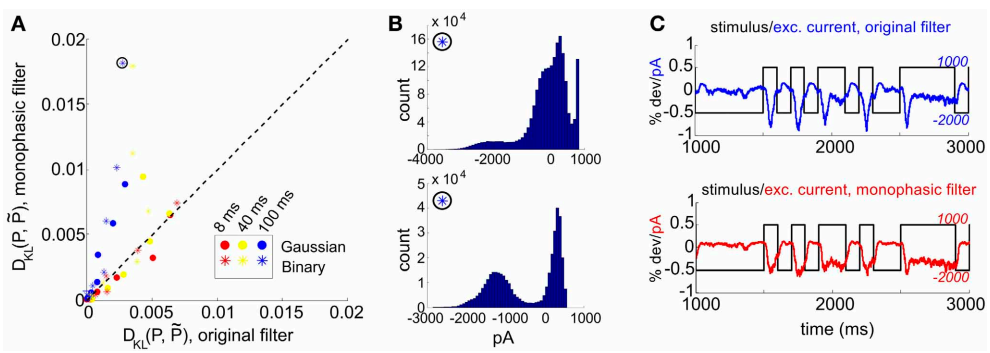


FIGURE 4 | Comparison of RGC simulations computed with the original ON parasol filter, vs. simulations using a more monophasic filter. **(A)** $D_{\mathrm{KL}}(P, \tilde{P})$ for original vs. monophasic filter. Data is organized by stimulus refresh rate (8, 40, and 100 ms) and marginal statistics (Gaussian vs. binary). **(B)** Histograms of excitatory conductances for an illustrative stimulus class, under original (top) and monophasic (bottom) filters. The marginal statistics and refresh rate are illustrated by icons inside black circles; here, binary stimuli with refresh rate 100 ms. The input standard deviation (expressed as a fraction of baseline light intensity) was 1/2. **(C)** Time course of stimulus and resulting excitatory conductances, from simulation shown in **(B)**: original (top) vs. monophasic (bottom) filters.

condition described earlier. In each case, we added gap junction coupling with strengths from 1 to 16 times an experimentally measured value (Trong and Rieke, 2008), and compared the resulting D_{KL} with that obtained without recurrent coupling (**Figure 5**).

At the experimentally measured coupling strength ($g^{\mathrm{gap}} = 1.1$ nS) itself, the fit of the pairwise model barely changed (**Figure 5A**) from the model without coupling. At twice the measured coupling strength ($g^{\mathrm{gap}} = 2.2$ nS), recurrent coupling had increased higher-order interactions, as measured by larger values of D_{KL} for all tested stimulus conditions. Higher order interactions could be further increased, particularly for long refresh rates (100 ms), by increasing the coupling strength to four or eight times its baseline level ($g^{\mathrm{gap}} = 4.4$ nS or $g^{\mathrm{gap}} = 8.8$ nS; see **Figures 5B,C**). Consistent with the intuition that very strong coupling leads to "all-or-none" spiking patterns, $D_{\mathrm{KL}}(P, \tilde{P})$ decreased as g^{gap} increased further, often to a level below what was seen in the absence of coupling (**Figure 5D**). In summary, the impact of coupling on D_{KL} is maximized at intermediate values of the coupling strength. However, the impact of recurrent coupling on the maximal values of D_{KL} evoked by visual stimuli is small overall, and almost negligible for experimentally measured coupling strengths.

2.2.6. Modeling heavy-tailed light stimuli in the RGC circuit

Finally, we repeated the full-field, recurrent, and alternate filter simulations previously described with light stimuli drawn from either Cauchy or heavy-tailed distributions: such distributions have been found to model the frequency of occurrence of luminance values in photographs of natural scenes (Ruderman and Bialek, 1994). In contrast to previous results with Gaussian and bimodal inputs, here we found very low $D_{\mathrm{KL}}(P, \tilde{P})$ over all stimulus conditions: the largest values found were more than an order of magnitude smaller than those obtained earlier. Specifically, for all conditions, we found $D_{\mathrm{KL}}(P, \tilde{P}) < 4.5 \times 10^{-4}$, over all 42 network realizations; for many simulations, this number did not meet a threshold for statistical significance (see section 4.1.7), indicating that P and $\tilde{P}$ were not statistically distinguishable. Using a more monophasic filter resulted in no apparent consistent change to $D_{\mathrm{KL}}(P, \tilde{P})$. When gap junction coupling was added, $D_{\mathrm{KL}}(P, \tilde{P})$ was maximized at an intermediate value; when $g^{\mathrm{gap}} = 8.8$, all simulations produced a statistically significant $D_{\mathrm{KL}}(P, \tilde{P}) \approx 3 - 4 \times 10^{-3}$. However, overall levels remained relatively low, roughly 1/2 the value achieved with Gaussian or binary stimuli.

To explain these findings, we examined the excitatory input currents: we found that over a broad range of refresh rates and stimulus variances, the marginal distributions of excitatory input conductances produced were remarkably unimodal in shape, and showed little skewness (**Figure 6A**). By examining the time evolution of the filtered stimuli (see **Figure 6B**), we see that heavy-tailed distributions allow rare, large events, but at the expense of medium-size events which explore the full range of the linear-nonlinear model used for stimulus processing (compare the blue with the red/green traces). When combined with the Gaussian background noise, this produces near-Gaussian excitatory conductances and, as may be expected from our original full-field simulations, very low D_{KL}.

We hypothesize that the methodology of averaging over the entire stimulus ensemble may not capture the significance of rare events that may individually be detected with high fidelity: D_{KL} was low even for full-field, high variance stimuli, which presumably caused (infrequent) global spiking events. Additionally, an important avenue for future work would be to test the ability of our RGC model, which was trained on Gaussian stimuli, to accurately model the response of a ganglion cell to stimuli whose variance is dominated by large events. Recent work examining the adaptation of retinal filtering properties to higher-order input statistics found little evidence of adaptation; however, the stimuli used in this work incorporated significant kurtosis but not heavy tails (Tkacik et al., 2012).

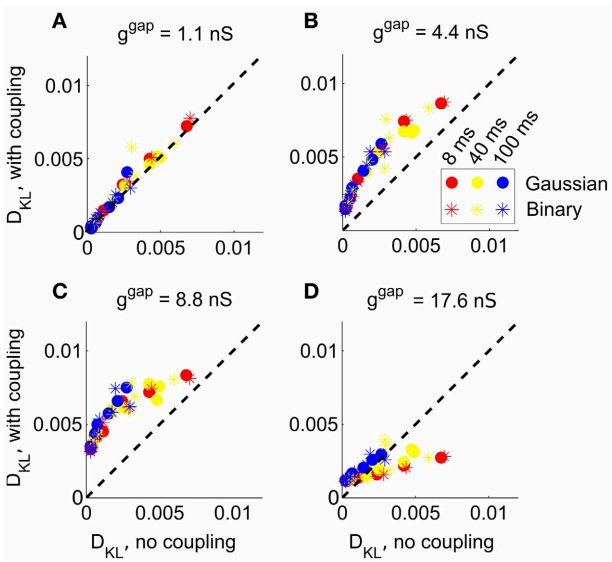


FIGURE 5 | The impact of recurrent coupling on RGC networks with full-field visual stimuli. The strength of gap junction connections was varied from a baseline level (relative magnitude $g = 1$, or absolute magnitude $g^{\mathrm{gap}} = 1.1$ nS) to an order of magnitude larger ($g = 16$, or $g^{\mathrm{gap}} = 17.6$ nS). In each panel, $D_{\mathrm{KL}}(P, \tilde{P})$ obtained with coupling is plotted vs. the value obtained for the same stimulus ensemble without coupling, for each of 42 different stimulus ensembles. **(A)** $g^{\mathrm{gap}} = 1.1$ nS (experimentally observed value); **(B)** $g^{\mathrm{gap}} = 4.4$ nS; **(C)** $g^{\mathrm{gap}} = 8.8$ nS; **(D)** $g^{\mathrm{gap}} = 17.6$ nS.

2.2.7. Summary of findings for RGC circuit

In summary, we probed the spiking response of a small array of RGC models to changes in light stimuli, gap junction coupling, and stimulus filtering properties, and identified two circumstances in which higher-order interactions were robustly generated in the spiking response. First, higher-order interactions were generated when excitatory currents had bimodal structure; we observed such structure when bimodal light stimuli was processed by a relatively monophasic filter. Secondly, higher-order interactions were maximized at an intermediate value of gap junction coupling; this value was, however, much larger (eight times) than the experimentally observed coupling strength.

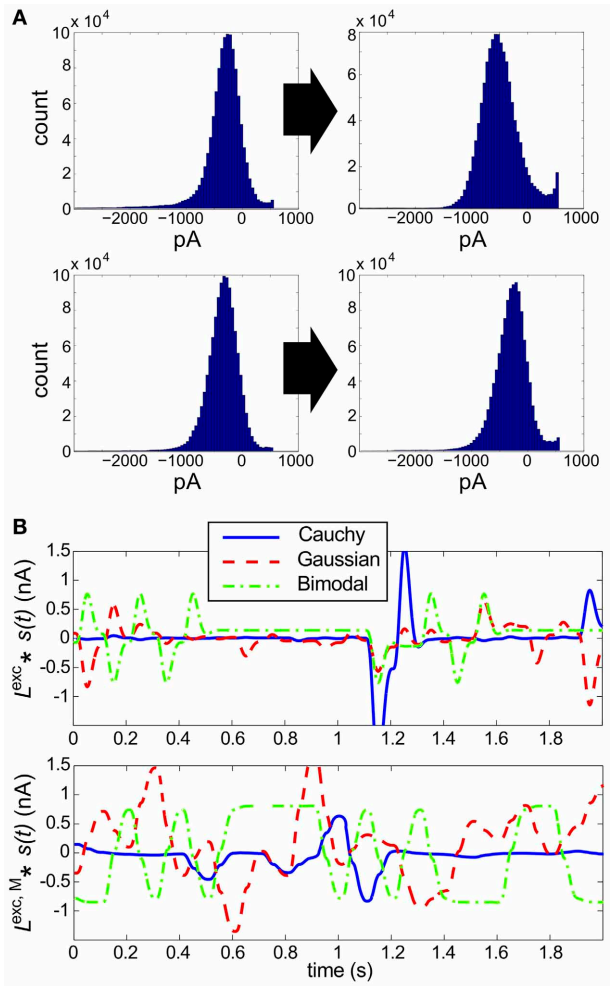


FIGURE 6 | Results for RGC simulations with heavy-tailed inputs. **(A)** Histograms of excitatory conductances, for the original (left) vs. monophasic (right) filter. The marginal statistics are heavy-tailed skew (top) and Cauchy (bottom) inputs, and refresh rate is 40 ms for both panels. The input standard deviation (expressed as a fraction of baseline light intensity) was 1/2 for both simulations. **(B)** Sample 100 ms stimuli, filtered by the original linear filter L_{exc} (top) and altered, monophasic filter $L_{\text{exc,M}}$(bottom). Cauchy (blue solid), Gaussian (red dashed), and bimodal (green dash-dotted) stimuli are shown.

2.3. A SIMPLIFIED CIRCUIT THAT EXPLAINS TRENDS IN RGC CELL MODEL

2.3.1. Setup and motivation

In the previous section, we developed results for a computational model tuned to a very specific cell type; we now ask whether these findings will hold for a more general class of neural circuits, or whether they are the consequence of system-specific features. To answer this question, we considered a simplified model of neural spiking: a feedforward circuit in which three spiking cells sum their inputs and spike according to whether or not they cross a threshold. Such highly idealized models of spiking have a long history in neuroscience (McCulloch and Pitts, 1943) and have been recently shown to predict the pairwise and higher-order activity of neural groups in both neural recordings and more complex dynamical spiking models (Nowotny and Huerta, 2003; Tchumatchenko et al., 2010; Yu et al., 2011; Leen and Shea-Brown, 2013).

In more detail, each cell j received an independent input I_j and a "triplet"—(global) input I_c that is shared among all three cells. Comparison of the total input $S_j = I_c + I_j$ with a threshold Θ determined whether or not the cell spiked in that random draw. An additional parameter, c, identified the fraction of the total input variance σ^2 originating from the global input; that is, $c \equiv \text{Var}[I_c]/\text{Var}[I_c + I_j]$. The global input was chosen from one of several marginal distributions, which included those used in the RGC model: Gaussian, bimodal, and heavy-tailed. The independent inputs I_j were, in all cases, chosen from a Gaussian distribution, consistent with our RGC model. When the common inputs are Gaussian, our model is equivalent to the Dichotomized Gaussian model previously studied by several groups (Amari et al., 2003; Macke et al., 2009, 2011; Yu et al., 2011), cf. (Tchumatchenko et al., 2010). For further details, see section 4.2.

In the RGC model large effects in D_{KL} were accompanied by a striking increase in the bi- or multi-modality of excitatory conductances. Why are bimodal inputs, shared across cells, able to produce spiking responses that deviate from the pairwise model? We use our simple thresholding model to provide some intuition for how bimodal common inputs to thresholding cells lead to spiking probabilities that violate the constraints (Equation 3) which must hold for the pairwise model. For example, suppose that the common input I_c can take on values that cluster around two separated values, $\mu_A < \mu_B$, but rarely in the interval between; that is, the distribution of I_c is *bimodal*. If μ_B is large enough to push the cells over threshold but μ_A is not, then we see that any contribution to the right-hand side of Equation (3), p_2/p_1, depends only on the distribution of the independent inputs I_j; if either one or two cells spike, then the common input must have been drawn from the cluster of values around μ_A, because otherwise all three cells would have spiked.

To be concrete, let $P[\mathbf{x}]$ refer to the probability of spiking event $\mathbf{x} = (x_1, x_2, x_3)$, and $P[\mathbf{x} \mid I_c \approx \mu_A]$ refer to the probability that $\mathbf{x}$ occurs, conditioned on the event $I_c \approx \mu_A$. Then

$$\begin{aligned} P[(1,0,0)] &= P[(1,0,0) \mid I_c \approx \mu_A]\, P[I_c \approx \mu_A] \\ &\quad + P[(1,0,0) \mid I_c \approx \mu_B]\, P[I_c \approx \mu_B] \\ &= P[(1,0,0) \mid I_c \approx \mu_A]\, P[I_c \approx \mu_A] \end{aligned}$$

because $P[(1,0,0) \mid I_c \approx \mu_B] = 0$: for the same reason,

$$P[(1,1,0)] = P[(1,1,0) \mid I_c \approx \mu_A]\, P[I_c \approx \mu_A]$$

therefore

$$\begin{aligned} \frac{p_2}{p_1} &= \frac{P[(1,1,0) \mid I_c \approx \mu_A]\, P[I_c \approx \mu_A]}{P[(1,0,0) \mid I_c \approx \mu_A]\, P[I_c \approx \mu_A]} \\ &= \frac{P[(1,1,0) \mid I_c \approx \mu_A]}{P[(1,0,0) \mid I_c \approx \mu_A]} \end{aligned}$$

On the other hand,

$$\frac{p_3}{p_0} = \frac{P\left[I_c \approx \mu_B\right] + P\left[(1,1,1) \mid I_c \approx \mu_A\right] P\left[I_c \approx \mu_A\right]}{P\left[(0,0,0) \mid I_c \approx \mu_A\right] P\left[I_c \approx \mu_A\right]}.$$

By changing the relative likelihood of drawing the common input from one cluster or the other, without changing the values of μ_A and μ_B themselves (that is, change $P[I_c \approx \mu_B]$ and $P[I_c \approx \mu_A]$ but leave the conditional probabilities (e.g., $P[(1,0,0) \mid I_c \approx \mu_A]$) fixed) one may change the ratio p_3/p_0 *without* changing the ratio p_2/p_1. Hence the constraint specifying those network responses exactly describable by PME models can be violated when the common input is bimodal.

In contrast, we may instead consider a *unimodal* common input, of which a Gaussian is a natural example. Here, the distribution of the common input I_c is completely described by its mean and variance; both parameters can impact the ratio p_3/p_0 (by altering the likelihood that the common input alone can trigger spikes) and the ratio p_2/p_1. Each value of I_c is consistent with both events p_1 and p_2, with the relative likelihood of each event depending on the specific value of I_c; it is no longer clear how to separate the two events. In the following sections, we will confirm this intuition by direct evaluation of the resulting departure from pairwise statistics.

2.3.2. Model input distributions

Motivated by our observations of excitatory currents that arose in the RGC model, we chose several input distributions that allow us to explore other salient features, such as symmetry and the probability of large events. A distribution is called *sub-Gaussian* if the probability of large events decays rapidly with event size, so that it can be bounded above by a scaled Gaussian distribution (see section 4). We considered two sub-Gaussian distributions; the Gaussian itself, and a skewed distribution with a sub-Gaussian tail (hereafter referred to as "skewed"). We also considered the two "heavy-tailed" distributions used as stimuli to the RGC model—the Cauchy distribution, and a skewed distribution with a Cauchy-like tail (hereafter referred to as "heavy-tailed skewed"). In these distributions, the probability of large events decays polynomially rather than exponentially.

For each choice of common input marginal, we varied the input parameters so as to explore a full range of firing rates and pairwise correlations: specifically, we varied the input correlation coefficient c in the range [0, 1], the *total* input standard deviation σ in the range [0, 4], and the threshold Θ in [0, 3]. In all cases the independent inputs I_j were chosen from a Gaussian distribution [of variance $(1-c)\sigma^2$]. For each choice of input parameters, we determine the resulting distribution on spiking states (as described in section 4) and compute the PME approximation.

2.3.3. Unimodal common inputs fail to produce significant higher-order interactions in three-cell feedforward circuits

We first considered common inputs chosen from a unimodal (e.g., Gaussian) distribution. If I_c is Gaussian, then the joint distribution of $\mathbf{S} = (S_1, S_2, S_3)$ is multivariate normal, and therefore characterized entirely by its means and covariances. Because the PME fit to a continuous distribution is precisely the multivariate normal that is consistent with the first and second moments, every such input distribution on $\mathbf{S}$ *exactly* coincides with its PME fit. However, even with Gaussian inputs, outputs (which are now in the binary state space $\{0,1\}^3$) will deviate from the PME fit (Amari et al., 2003; Macke et al., 2009). As shown below, non-Gaussian unimodal inputs can produce outputs with larger deviations. Nonetheless, these deviations are small for all cases in which inputs were chosen from a sub-Gaussian distribution, and PME models are quite accurate descriptions of circuits with a broad range of unimodal inputs.

We first considered circuits with either Gaussian or skewed common inputs. Over the full range of input parameters, distributions remained well fit by the pairwise model, with a maximum value of $D_{\mathrm{KL}}(P, \tilde{P})$ (of 0.0038 and 0.0035 for Gaussian and skewed inputs, respectively) achieved for high correlation values and σ comparable to threshold. In **Figure 7A** we illustrate these trends with a contour plot of $D_{\mathrm{KL}}(P, \tilde{P})$ for a fixed value of threshold (here, $\Theta = 1.5$) and Gaussian common inputs (the analogous plot for skewed inputs is qualitatively very similar, **Figure S3A**).

Clear patterns also emerged when we viewed $D_{\mathrm{KL}}(P, \tilde{P})$ as a function of *output* spiking statistics rather than *input* statistics (as in Macke et al., 2011). Non-linear spike generation can produce substantial differences between input and output correlations; this relationship can vary widely based on the specific non-linearity (Moreno et al., 2002; de la Rocha et al., 2007; Marella and Ermentrout, 2008; Shea-Brown et al., 2008; Vilela and Lindner, 2009; Barreiro et al., 2010, 2012; Tchumatchenko et al., 2010; Hong et al., 2012). **Figure 7B** shows $D_{\mathrm{KL}}(P, \tilde{P})$ and Δ for all threshold values (including the data shown in **Figure 7A**), but now plotted with respect to the output firing rate. The data were segregated according to the Pearson's correlation coefficient ρ between the responses of cell pairs ($\rho \equiv \frac{\mathrm{Cov}(x_i, x_j)}{\sqrt{\mathrm{Var}(x_i)\mathrm{Var}(x_j)}} = \frac{\hat{\rho} - \mu^2}{\mu(1-\mu)}$). For a fixed correlation, there was generally a one-to-one relationship between firing rate and $D_{\mathrm{KL}}(P, \tilde{P})$. For these distributions (**Figure 7B**, for Gaussian inputs; skewed inputs shown in **Figure S3B**), $D_{\mathrm{KL}}(P, \tilde{P})$ was maximized at an intermediate firing rate. Additionally, $D_{\mathrm{KL}}(P, \tilde{P})$ had a non-monotonic relationship with spike correlation: it increased from zero for low values of correlation, obtained a maximum for an intermediate value, and then decreased. These limiting behaviors agree with intuition: a spike pattern that is completely uncorrelated can be described by an independent distribution (a special case of PME model), and one that is perfectly correlated can be completely described via (perfect) pairwise interactions alone.

We next considered circuits in which inputs were drawn from one of two heavy-tailed distributions, the Cauchy distribution and a heavy-tailed skewed distribution, defined earlier. Here, distinctly different patterns emerge: for a fixed Θ, $D_{\mathrm{KL}}(P, \tilde{P})$ is maximized in regions of high input correlation and high input variance σ, but relatively high values of D_{KL} are achievable across a wide range of input values (see **Figure 7C** for Cauchy inputs; heavy-tailed skewed in **Figure S3C**). However, the maximum achievable values of D_{KL} were achieved at intermediate *output* correlations $\rho \approx 0.4$ (see **Figure 7D** for Cauchy inputs; heavy-tailed skewed shown in **Figure S3D**); this suggests that high input correlations do not result in high output correlations.

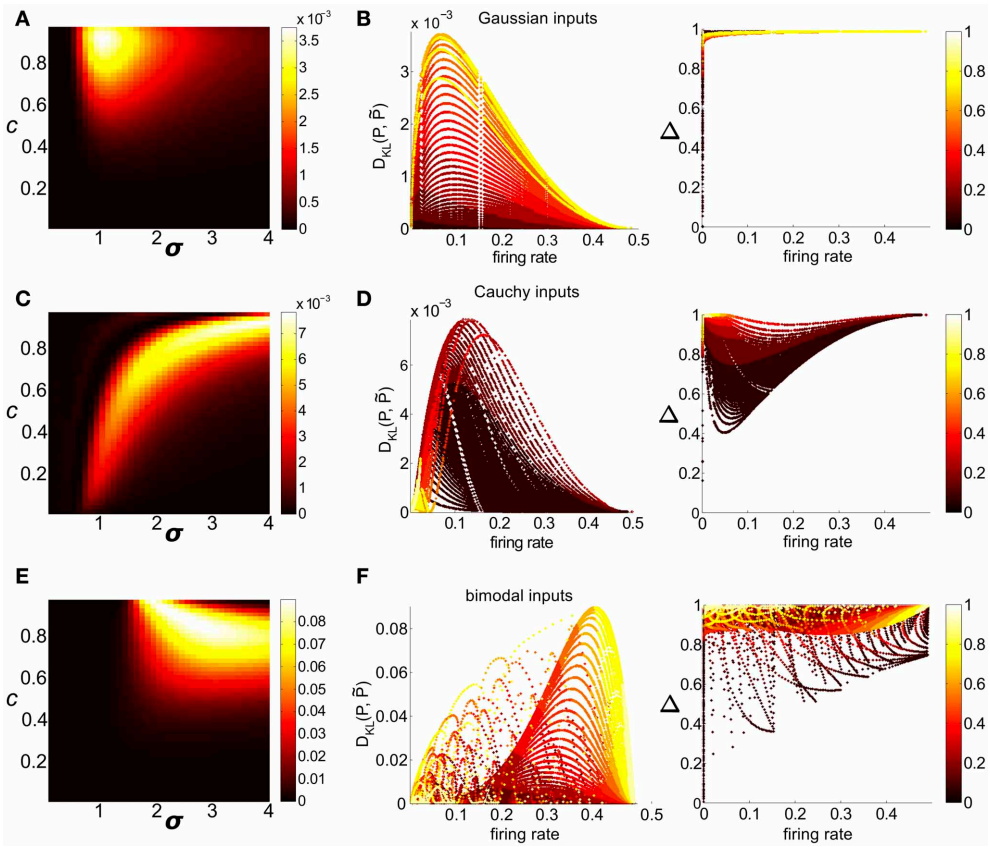


FIGURE 7 | Strength of higher-order interactions produced by the threshold model as input parameters vary, and the relationship of these higher-order interactions with other output firing statistics. **(A)** For Gaussian common inputs: $D_{\mathrm{KL}}(P, \tilde{P})$ as a function of input correlation c and input standard deviation σ, for a fixed threshold $\Theta = 1.5$. Color indicates $D_{\mathrm{KL}}(P, \tilde{P})$; see color bar for range. **(B)** For Gaussian common inputs: $D_{\mathrm{KL}}(P, \tilde{P})$ vs. firing rate (Left) and the fraction of multi-information (Δ) captured by the PME model vs. firing rate (Right). Each dot represents the value obtained from a single choice of the input parameters c, σ, and Θ; input parameters were varied over a broad range as described in section 2. Firing rate is defined as the probability of a spike occurring per cell per random draw of the sum-and-threshold model, as defined in Equation (16). Color indicates output correlation coefficient ρ ranging from black for $\rho \in (0, 0.1)$, to white for $\rho \in (0.9, 1)$, as illustrated in the color bars. **(C,D)**: as in **(A,B)**, but for Cauchy common inputs. **(E,F)**: as in **(A,B)**, but for bimodal common inputs.

This somewhat unintuitive finding may be explained by the structure of the PDF of a heavy-tailed common input, which favors (infrequent) large events at the expense of medium-size events. For instance, the probability that a Cauchy input is above a given threshold ($P[I_c > \Theta > \mathbf{E}[I_c]]$) is often much smaller than for a Gaussian distribution of the same variance. However, an input can trigger at best one single spiking event regardless of size: therefore a Cauchy common input generates fewer correlated spiking events with larger inputs, while a Gaussian common input triggers correlated spiking events with smaller, but more frequent, input values. As a result, heavy-tailed inputs are unable to explore the full range of output firing statistics: **Figure 7D** shows that high output correlations only occur at very low firing rates. Overall, $D_{\mathrm{KL}}(P, \tilde{P})$ reaches higher numerical values than for sub-Gaussian inputs, possibly reflecting the higher-order statistics in the input. However, the maximal $D_{\mathrm{KL}}(P, \tilde{P})$ attained still falls far short of exploring the full range of possible values (compare with **Figure 1B**).

Finally, we examine the behavior of the *strain*, which quantifies both the magnitude and sign of deviation from the pairwise model (see Ohiorhenuan and Victor, 2010). It has been previously observed that the strain is negative for the DG model (Macke et al., 2011), a condition that has been related to sparsity of the neural code and with which our results agree (data not shown). However, we found that any other choice of input marginal statistics, both positive and negative values are seen; for heavy-tailed common inputs, positive values predominated except at very low firing rates.

2.3.4. Bimodal triplet inputs can generate higher-order interactions in three-cell feedforward circuits

Having shown that a wide range of unimodal common inputs produced spike patterns that are well-approximated by PME fits, we next examined bimodal common inputs. Such inputs substantially increased departures from PME fits in the ganglion cell models described above. As in the previous section, we varied c,

σ, and Θ so as to explore a full range of firing rates and pairwise correlations.

As a function of input parameter values, $D_{\mathrm{KL}}(P, \tilde{P})$ is maximized for large input correlation and moderate input variance σ^2 [see **Figure 7E**, which illustrates $D_{\mathrm{KL}}(P, \tilde{P})$ for a fixed threshold $\Theta = 1.5$]. **Figure 7F** shows $D_{\mathrm{KL}}(P, \tilde{P})$ values as a function of the firing rate and pairwise correlation elicited by the full range of possible bimodal inputs. We see that $D_{\mathrm{KL}}(P, \tilde{P})$ is maximized at an intermediate (but relatively high: $\nu \approx 0.4$) firing rate, and for intermediate-to-large correlation values ($\rho \approx 0.6 - 0.8$).

We find distinctly different results when we view Δ (Equation 1), for these same simulations, as a function of output spiking statistics (right panels of **Figures 7B,D,F**). For unimodal, sub-Gaussian distributions (**Figure 7B**), Δ is very close to 1, with the few exceptions at extreme firing rates. For heavy-tailed and bimodal inputs (**Figures 7D,F**), Δ may be appreciably far from 1 (as small as 0.5) with the smallest numbers (suggesting a poor fit of the pairwise model) occurring for low correlation ρ. This highlights one interesting example where these two metrics for judging the quality of the pairwise model, $D_{\mathrm{KL}}(P, \tilde{P})$ and Δ, yield contrasting results.

Finally, we emphasize that while bimodal inputs can produce greater higher-order interactions than unimodal inputs, the values of $D_{\mathrm{KL}}(P, \tilde{P})$ accessible by feedforward circuits with global inputs remain far below their upper bounds at any given firing rate. The maximal values of $D_{\mathrm{KL}}(P, \tilde{P})$ reached by Cauchy and heavy-tailed skewed inputs were 0.0078 and 0.0153; bimodal common inputs reached a maximal value of 0.091. This is an order of magnitude smaller than possible departures among symmetric spike patterns (compare **Figure 1B**). The difference is illustrated in **Figure S4**, which compares the $D_{\mathrm{KL}}(P, \tilde{P})$ values obtained in the thresholding model and those obtained by direct exhaustive search at each firing rate by superposing the datapoints on a single axis.

2.3.5. Mathematical analysis of unimodal vs. bimodal effects

The central finding above is that circuits with bimodal inputs can generate significantly greater higher-order interactions than circuits with unimodal inputs. To probe this further, we investigated the behavior of $D_{\mathrm{KL}}(P, \tilde{P})$ for the feedforward threshold model with a perturbation expansion in the limit of small common input. We found that as the strength of common input signals increased, circuits with bimodal inputs diverged from the PME fit more rapidly than circuits with unimodal inputs; the full calculation is given in the Appendix. In brief, we determined the leading order behavior of $D_{\mathrm{KL}}(P, \tilde{P})$ in the strength c of (weak) common input. $D_{\mathrm{KL}}(P, \tilde{P})$ depended on c^3 for unimodal distributions, i.e., the low order terms in c dropped out; for symmetric unimodal distributions, such as a Gaussian, $D_{\mathrm{KL}}(P, \tilde{P})$ grew as c^4. For bimodal distributions, $D_{\mathrm{KL}}(P, \tilde{P})$ grew as c^2. Because of the c^2 dependence, rather than c^3 or c^4, as the strength of common input signals c increases, circuits with bimodal inputs are predicted to produce greater deviations from their PME fits.

2.3.6. Impact of recurrent coupling

We next modified our thresholding model to incorporate the effects of recurrent coupling among the spiking cells. To mimic gap junction coupling in the RGC circuit, we considered all-to-all, excitatory coupling, and assumed that this coupling occurs on a faster timescale compared with the timescale over which inputs arrive at the cells.

Our previous model was extended as follows: if the inputs arriving at each cell elicited any spikes, there was a second stage at which the input to each neuron receiving a connection from a spiking cell was increased by an amount g. This represented a rapid depolarizing current, assumed for simplicity to add linearly to the input currents. If the second stage resulted in additional spikes, the process was repeated: recipient cells received an additional current g, and their summed inputs were again thresholded. The sequence terminated when no new spikes occurred on a given stage; e.g., for $N = 3$, there were a maximum of three stages. The spike pattern recorded on a given trial was the total number of spikes generated across all stages.

We then explored the impact of varying g for a single representative value of σ and Θ, and several values of the correlation coefficient c. We found that as g increased $D_{\mathrm{KL}}(P, \tilde{P})$ varied smoothly, reflecting the underlying changes in the spike count distribution. For small c ($c = 0.02$ shown in **Figure 8A**), where the variance of common input is very small, the results varied little by input type: for all input types $D_{\mathrm{KL}}(P, \tilde{P})$ reached an interior maximum near $g \approx 1.7$. As c increases, the distinctions between inputs types become apparent (**Figures 8B,C** show $c = 0.2, 0.5$, respectively): for most input types and values of c, the value of $D_{\mathrm{KL}}(P, \tilde{P})$ reaches an interior maximum that exceeds its value without coupling (i.e., $g = 0$). However, overall values of $D_{\mathrm{KL}}(P, \tilde{P})$ remained modest, never exceeding 0.01 across the values explored here.

2.3.7. Summary of findings for simplified circuit model

We examined a highly idealized model of neural spiking, so as to explore the generality of our earlier findings in a small array of RGC models. We found that our main results from the RGC model—that higher-order interactions were most significant when inputs had bimodal structure, and that when fast excitatory recurrence was added to the circuit, higher-order interactions were maximized at an intermediate value of the recurrence strength—persisted in this simplified model. Moreover, we were able to show that the first of these findings is general, in that it holds over a complete exploration of parameter space.

2.4. SCALING OF HIGHER-ORDER INTERACTIONS WITH POPULATION SIZE

The results above suggest that unimodal, rather than bimodal, input statistics contribute to the success of PME models. Next, we examined whether this conclusion continues to hold when we increase network size. The permutation-symmetric architectures we have considered so far can be scaled up to more than three cells in several natural ways; for example, we can study N cells with a global common input.

We considered a sequence of models in which a set of N threshold spiking units received global input I_c [with mean 0 and variance $\sigma^2 c$] and an independent input I_j [with mean 0 and variance $\sigma^2(1 - c)$]. As for the three-cell network models considered previously, the output of each cell was determined by summing

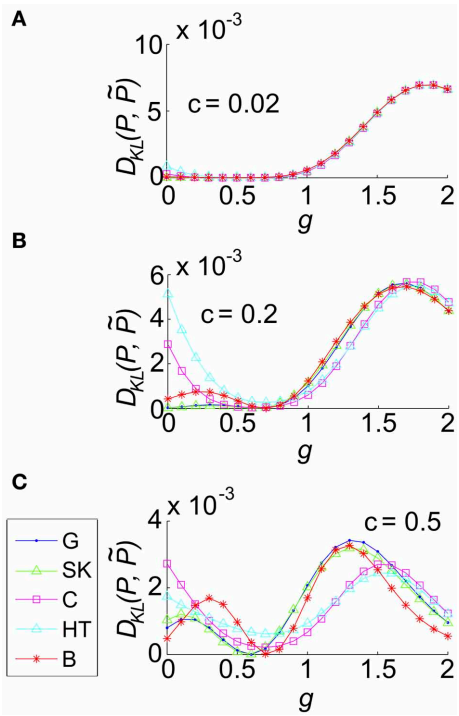


FIGURE 8 | The impact of recurrent coupling on the three-cell sum-and-threshold model. Each plot shows $D_{\mathrm{KL}}(P, \tilde{P})$ as a function of g, for a specific value of the correlation coefficient. In all panels, input standard deviation $\sigma = 1$, threshold $\Theta = 1.5$, $N = 3$ and symbols are as described in the legend for **(C)**. Abbreviations in the legend denote the marginal distribution of the common input: G, Gaussian; SK, skewed; C, Cauchy; HT, heavy-tailed skewed; B, bimodal. **(A)** For input correlation $c = 0.02$, **(B)** $c = 0.2$, and **(C)** $c = 0.5$.

and thresholding these inputs. Upon computing the probability distribution of network outputs (section 4), we fit a PME distribution. Again, we explored a range of σ, c, and Θ and recorded the maximum value of $D_{\mathrm{KL}}(P, \tilde{P})$ between the observed distribution P and its PME fit $\tilde{P}$. **Figure 9** shows this D_{KL}/N [i.e., entropy per cell (Macke et al., 2009)] for each class of marginal distributions.

We found that the maximum $D_{\mathrm{KL}}(P, \tilde{P})/N$ increased roughly linearly with N for Gaussian, skewed and Cauchy inputs; for heavy-tailed skew and bimodal inputs, $D_{\mathrm{KL}}(P, \tilde{P})/N$ appeared to saturate after an initial increase (**Figure 9**). The relative ordering for unimodal inputs shifted as N increased; as $N \to 16$, the maximal achievable $D_{\mathrm{KL}}(P, \tilde{P})$ for sub-Gaussian inputs overtook the values for heavy-tailed inputs. At all values of N, the values for Gaussian and skewed inputs tracked one another closely. Regardless, the values for all unimodal inputs remained substantially below the maximal value achievable for bimodal inputs. **Figure 9B** shows that the probability distributions produced by these inputs qualitatively agree with this trend: departures from PME were more visually pronounced for global bimodal inputs than for global unimodal inputs. In addition, the distributions for heavy-tailed and sub-Gaussian inputs differed qualitatively, offering a potential mechanism for different scaling behavior. Using the relationship between D_{KL} and likelihood ratios (described in section 2.1), at $N = 16$, the value $D_{\mathrm{KL}}/N \approx 0.1$ for bimodal global inputs corresponds to a likelihood ratio of 0.33 that a single draw from P (single network output) in fact came from the PME fit $\tilde{P}$ rather than from P; a likelihood <0.01 is reached for four draws.

We next extended our model with recurrent coupling to $N > 3$ cells. In addition to the parameters for the uncoupled network, we varied the coupling strength, g, for each type of input. As in the $N = 3$ network, coupling was all-to-all. As for the small networks explored in an earlier section, $D_{\mathrm{KL}}(P, \tilde{P})$ generally peaked at an intermediate value of the coupling strength g; however, the value of g decreased as population size N increased (illustrated in **Figure 10A**, for $c = 0.2$). This may be attributed to the increased potential impact of recurrence at larger population sizes; as N increases, the number of potential *additional* spikes that may be triggered increases; consequently the average recurrent excitation received by each cell increases, and therefore the probability that one or two spikes will trigger a cascade to N spikes. In **Figure 10B** we demonstrate that the impact of this effect may be captured by plotting $D_{\mathrm{KL}}(P, \tilde{P})$ as a function of an *effective* coupling parameter, $g^*N/3$. Here, we plot the curves for six population sizes ($N = 3$, 4, 6, 8, 10, and 12) and five common input types; each curve was scaled by normalizing $D_{\mathrm{KL}}(P, \tilde{P})$ by its maximum value. For many sets of parameter values, the resulting curves line up remarkably well, suggesting a universal scaling with the effective coupling parameter.

We also explored the overall possible impact of recurrence on higher-order interactions, by surveying a range of circuit parameters c, σ, Θ and g. The top panel of **Figure 10C** shows the maximal $D_{\mathrm{KL}}(P, \tilde{P})$ per neuron, for each type of input, up to population size $N = 8$. For unimodal inputs, recurrent coupling increased the available range of higher-order interactions modestly, compared with the range achieved with purely feedforward connections; however, these values remained significantly lower than those achieved for bimodal inputs.

Finally, we considered how higher-order interactions scale with population sampling size. The spike pattern distributions used to generate the last column of data points ($N = 8$) in the top panel of **Figure 10C** were reanalyzed by sub-sampling the spike pattern distributions on $k < 8$ cells. In each case, we chose our sub-population to be k nearest neighbors (for our setup, any subset of k cells is statistically identical). In the bottom panel of **Figure 10C**, we show the maximal value of $D_{\mathrm{KL}}(P, \tilde{P})$ per sub-sampled cell achieved over all input parameters (the curves for Gaussian, skewed and Cauchy inputs are so close together so as to be visually indistinguishable). This number increases or remains steady as k increases, indicating that sub-sampling a coupled network will depress the apparent higher-order interactions in the output spiking pattern.

To summarize, the greater impact of bimodal vs. unimodal input statistics on maximal values of $D_{\mathrm{KL}}(P, \tilde{P})$ persists in circuits with $N = 3$ cells up to $N = 16$ cells. Overall, for the circuit parameters producing maximal deviations from PME fits, it becomes easier to statistically distinguish between spiking distributions and their PME fits as the number of cells increases in feedforward networks.

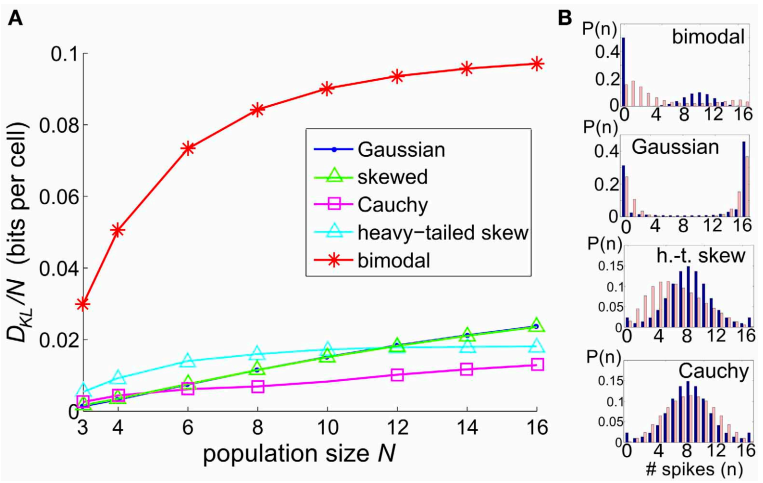


FIGURE 9 | The significance of higher-order interactions increases with network size. (A) Normalized maximal deviation, $D_{\mathrm{KL}}(P, \tilde{P})/N$, from the PME fit for the thresholding circuit model as network size N increases. For each N and common input distribution type, possible input parameters were in the following ranges: input correlation $c \in [0, 1]$, input standard deciation $\sigma \in [0, 4]$, and threshold $\Theta \in [0, 3]$. **(B)** Example sample distributions for different types of common input: from top, bimodal, Gaussian, heavy-tailed skew, and Cauchy common inputs. For each input type, the distribution that maximized $D_{\mathrm{KL}}(P, \tilde{P})$ for $N = 16$ is shown. Each distribution is illustrated with a bar plot contrasting the probabilities of spiking events in the true (dark blue) vs. pairwise maximum entropy (light pink) distributions.

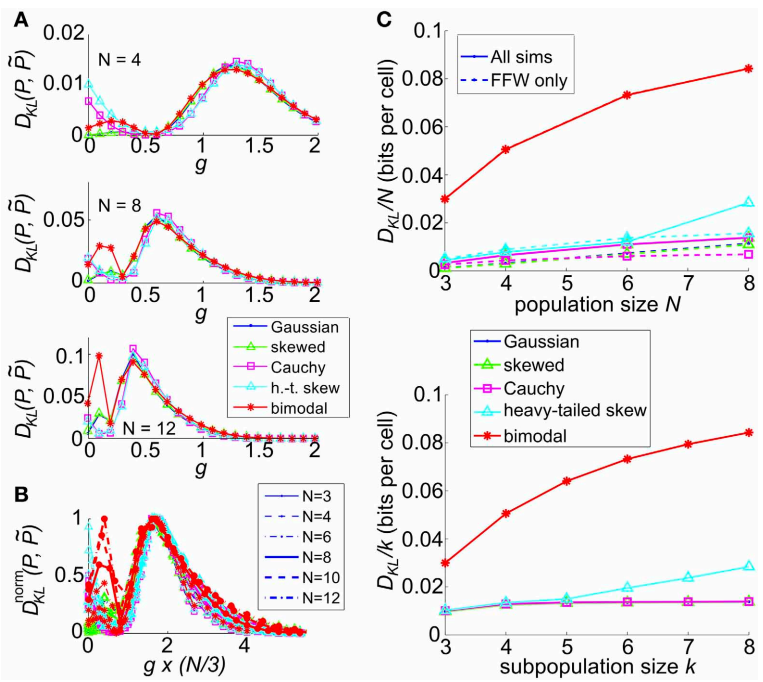


FIGURE 10 | The impact of recurrent coupling on the sum-and-threshold model, for increasing population size. (A) $D_{\mathrm{KL}}(P, \tilde{P})$ as a function of the coupling coefficient, g, for a specific value of population size N. In all plots, input standard deviation $\sigma = 1$, threshold $\Theta = 1.5$ and input correlation $c = 0.2$. From top: $N = 4$; $N = 8$; $N = 12$. **(B)** $D_{\mathrm{KL}}^{\mathrm{norm}}(P, \tilde{P})$ as a function of the coupling coefficient, g, for populations sizes $N = 3 - 12$. For each curve, $D_{\mathrm{KL}}(P, \tilde{P})$ was scaled by its maximal value and plotted as a function of the scaled coupling coefficient, $g^* N/3$, to illustrate a universal scaling with effective coupling strength. The line style of each curve indicates the population size N, as listed in the legend. The marker and line color indicate the common input marginal, as listed in the legend for **(A)**. **(C)** (Top) Maximal value of $D_{\mathrm{KL}}(P, \tilde{P})/N$, achieved over a survey of parameter values c, σ, Θ, and g, as a function of the population size N (solid lines). For each input marginal type, a second curve shows the maximal value obtained over only feed-forward simulations ($g = 0$; dashed lines). The marker and line color indicate the common input marginal, as listed in the legend for **(A)**. (Bottom) Maximal value of $D_{\mathrm{KL}}(P, \tilde{P})/k$, achieved over a survey of parameter values c, σ, Θ, and g, as a function of the *subsample* population size k. Data was subsampled from the $N = 8$ data shown in the top panel, by restricting analysis to k out of N cells.

3. DISCUSSION

We used mechanistic models to identify input patterns and circuit mechanisms which produce spike patterns with significant higher-order interactions—that is, with substantial deviations from predictions under a PME model. We focused on a tractable setting of small, symmetric circuits with common inputs. This revealed several general principles. First, we found that these circuits produced outputs that were much closer to PME predictions than required for a general spiking pattern. Second, bimodal input distributions produced stronger higher-order interactions than unimodal distributions. Third, recurrent excitatory or gap junction coupling could produce a further, moderate increase of higher-order correlations; the effect was greatest for coupling of intermediate strength.

These general results held for both an abstract threshold-and-spike model and for networks of non-linear integrate-and-fire units based on measured properties of one class of RGCs. Together with the facts that ON parasol cell filtering suppresses bimodality in light input, and that coupling among ON parasol cells is relatively weak, our findings provide an explanation for why their population activity is well captured by PME models.

3.1. COMPARISON WITH EMPIRICAL STUDIES

How do our maximum entropy fits compare with empirical studies? In terms of $D_{\mathrm{KL}}(P, \tilde{P})$—equivalently, the logarithm of the average relative likelihood that a sequence of data drawn from P was instead drawn from the model $\tilde{P}$—numbers obtained from our RGC models are very similar to those obtained by *in vitro* experiments on primate RGCs (Shlens et al., 2006, 2009). For example, in a survey of 20 numerical experiments under constant light conditions (each of length 100 ms, with spikes binned in 10 ms intervals), we find that $D_{\mathrm{KL}}(P, \tilde{P})$ ranges between 0 and 0.00029: similarly excellent fits were found by Shlens et al. (2006) (in which cell triplets were stimulated by constant light for 60 s with spikes binned at 10 ms), with one example given of 0.0008 (inferred from a reported likelihood ratio of 0.99944). These values can increase by an order of magnitude under full-field stimulation, as well as spatio-temporally varying stixel simulations (bounded above by 0.007). We can view the 60 μm stixel simulations as a model of the checkerboard experiments of Shlens et al. (2006), for which close fits by the PME distribution were also observed (likelihood numbers were not reported). Similarly, the values of Δ produced by our RGC model are close to those found by Schneidman et al. (2006); Shlens et al. (2006) under comparable stimulus conditions. We obtain $\Delta = 99.5\%$ (for cell group size $N = 3$) under constant illumination, which is near the range reported by Shlens et al. (2006) for the same bin size and stimulus conditions (98.6 ± 0.5, $N = 3 - 7$). For full-field stimuli we find a range of numbers from 95.7% to 99.3% ($N = 3$).

With regard to the circuit mechanisms behind these excellent fits by pairwise models, the findings that most directly address the experimental settings of Shlens et al. (2006, 2009), are (1) the finding that in the threshold model, unimodal inputs generate minimal higher-order interactions, compared to bimodal inputs, and (2) the particular stimulus filtering properties of parasol cells can suppress bimodality that may be present in an input stimulus, resulting in a unimodal distribution of input currents. First, we believe that unimodal inputs are consistent with the white-noise checkboard stimuli used in Shlens et al. (2006, 2009), where binary pixels were chosen to be small relative to the receptive field size; averaged over the spatial receptive field, they would be expected to yield a single Gaussian input by the central limit theorem. Second, temporal filtering may contribute to receipt of unimodal conductance inputs by cells for the full-field binary flicker stimuli that are delivered in Schneidman et al. (2006). With the 16.7 ms refresh rate used there, under the assumption that the filter time-scale of the cells studied in that paper is roughly similar to that of the ON parasol cell we consider, the filter would average a binary (and hence bimodal) stimulus into a unimodal shape (see **Figure 2C**, for example).

The simple threshold models that we have considered, meanwhile, give us a roadmap for how circuits could be driven in such a way as to lower Δ. The right columns of **Figures 7B,D,F** show Δ plotted as a function of firing rate for circuits of $N = 3$ cells receiving global common inputs; we observe that $\Delta \approx 1$ for Gaussian inputs over a broad range of firing rates and pairwise correlation coefficients, but that values of Δ can be depressed by 25–50% in the presence of a bimodal common input. Indeed, Shlens et al. (2006) showed that adding global bimodal inputs to a purely pairwise model can lead to a comparable departure in Δ. Our results are consistent with this finding, and explicitly demonstrate that the bimodality of the inputs—as well as their global projection—are characteristics that lead to this departure.

3.2. CONSEQUENCES FOR SPECIFIC NEURAL CIRCUITS

Our results make predictions about when neural circuits are likely to generate higher-order interactions. A comprehensive study of our simple thresholding model shows that bimodal inputs generate greater beyond-pairwise interactions than unimodal inputs. This result can be extended to other circuits where a clear input–output relationship exists, and be used to predict higher-order correlations by analyzing the impact of stimulus filtering on a statistically defined class of inputs. For example, the effect holds in our model of primate ON parasol cells, where a biphasic filter suppresses bimodality in a stimulus with a timescale matched to that of the filter. We can use these results to extrapolate to other classes of RGCs or other stimulus conditions in which filters are less biphasic (Victor, 1999). Indeed, when we process long time-scale bimodal inputs through a preliminary model of the midget cell circuit, stimulus bimodality is no longer suppressed and is associated with higher-order interactions (see **Figure 4**). We predict that greater higher-order interactions will be found for stimuli or RGC circuits that elicit bimodal activity that is thresholded when generating spikes—in comparison to the parasol circuits and stimuli studied in Shlens et al. (2006, 2009). We believe that this principle will be further applicable in other sensory systems.

We found that recurrent excitatory connections further increase higher-order interactions, which are maximized at an intermediate recurrence strength; in particular, when the strength of an excitatory recurrent input was comparable to the distance between rest and threshold (**Figure 8**). For the primate ON parasol cells we considered, the experimentally measured strength of gap junction coupling would lead to an estimated membrane

voltage jump of ≈1 mV in response to the firing of a neighboring RGC, while the voltage distance between the resting voltage and an approximate threshold is about 5–10 mV (Trong and Rieke, 2008). Consistent with this estimate, we found that in our ON parasol cell model, higher-order interactions were maximized when the strength of excitatory recurrence was eight times its experimentally measured value. The experimentally measured values of recurrence had little or no effect on higher-order interactions. We anticipate that this result may be used to predict whether recurrent coupling plays a role in generating higher-order interactions in other circuits where the average voltage jump produced by an electrical or synaptic connection can be measured.

To apply our findings to real circuits, we must also consider population size. A measurement from a neural circuit, in most cases, will be a subsample of a much larger, complete circuit. We addressed this question where it was computationally more tractable, for the thresholding model. Here, we found that the impact of higher-order interactions, as measured by entropy per cell unaccounted for by the pairwise model (D_{KL}/k), increases moderately as subsample size k increases. Since recurrent connectivity in our model is truly global, this is consistent with the suggestion of Roudi et al. (2009a) and others that the entropy can be expected to scale extensively with population size N, once N significantly exceeds the true spatial connectivity footprint: we may see different results with limited, local connectivity.

3.3. SCOPE AND OPEN QUESTIONS

There are many aspects of circuits left unexplored by our study. Prominent among these is heterogeneity. Only a few of our simulations produce heterogeneous inputs to model RGCs, and all of our studies apply to cells with identical response properties. This is in contrast to studies such as Schneidman et al. (2006), which examine correlation structures among multiple cell types. For larger networks, feedforward connections with variable spatial profiles also occur, between the extremes of independent and global input connections examined here. It is also possible that more complex input statistics could lead to greater higher-order interactions (Bethge and Berens, 2008). Finally, **Figure 9** indicates that some trends in $D_{\mathrm{KL}}(P, \tilde{P})$ vs. N appear to become non-linear for $N \gtrapprox 10$; for larger networks, our qualitative findings could change.

Our study also leaves largely open the role of different retinal filters in generating higher-order interactions. We have found that the specific filtering properties of ON parasol cells suppress bimodality in light inputs, suggesting that other classes of RGCs, such as midget cells, may produce more robust higher-order interactions (compare panels in **Figure 4B**). This predicts a specific mechanism for the development of higher-order interactions in preparations that include multiple classes of ganglion cells (Schneidman et al., 2006). For a complete picture, future studies will also need to account for the possible adaptation of stimulus filters in response to higher-order stimulus characteristics (Tkacik et al., 2012); we did not consider the latter effect here, where our filter was fit to the response of a cell to Gaussian stimuli with specific mean and variance. An allied possibility is that multiple filters will be required, as was found when fitting the responses of salamander retinal cells to LN models (Fairhall et al., 2006). Distinguishing the roles of linear filters vs. static non-linearities in determining which stimulus classes will give the greatest higher-order correlations is another important step. Finally, we considered circuits with a single step of inputs and simple excitatory or gap junction coupling; a plethora of other network features could also lead to higher-order interactions, including multi-layer feedforward structures, together with lateral and feedback coupling. We speculate that, in particular, such mechanisms could contribute to the higher-order interactions found in cortex (Tang et al., 2008; Montani et al., 2009; Ohiorhenuan et al., 2010; Oizumi et al., 2010; Koster et al., 2013).

A final outstanding area of research is to link tractable network mechanisms for higher-order interactions with their impact (or lack of impact) on information encoded in neural populations (Kuhn et al., 2003; Montani et al., 2009; Oizumi et al., 2010; Ganmor et al., 2011; Cain and Shea-Brown, 2013). A simple starting point is to consider rate-based population codes in which each stimulus produces a different "tuned" average spike count (see for e.g., chapter 3 of Dayan and Abbot, 2001). One can then ask whether spike responses can be more easily decoded to estimate stimuli for the full population response (i.e., P) to each stimulus or for its pairwise approximation ($\tilde{P}$). In our preliminary tests where higher-order correlations were created by inputs with bimodal distributions, we found examples where decoding of P vs. $\tilde{P}$ differed substantially. However, a more complete study would be required before general conclusions about trends and magnitudes of the effect could be made; such a study would include complementary approach in which the full spike responses P are themselves decoded via a "mismatched" decoder based on the pairwise model $\tilde{P}$ (Oizumi et al., 2010). Overall, we hope that the present paper, as one of the first that connects circuit mechanisms to higher-order statistics of spike patterns, will contribute to future research that takes these next steps.

4. MATERIALS AND METHODS

4.1. EXPERIMENTALLY-BASED MODEL OF A RGC CIRCUIT

We model the response of a individual RGC using data collected from a representative primate ON parasol cell, following methods in Murphy and Rieke (2006); Trong and Rieke (2008). Similar response properties were observed in recordings from 16 other cells. To measure the relationship between light stimuli and synaptic conductances, the retina was exposed to a full-field, white noise stimulus. The cell was voltage clamped at the excitatory (or inhibitory) reversal potential $V_E = 0$ mV ($V_I = -60$ mV), and the inhibitory (or excitatory) currents were measured in response to the stimulus. These currents were then turned into equivalent conductances by dividing by the driving force of ±60 mV; in other words

$$g^{\mathrm{exc}} = I^{\mathrm{exc}}/(V - V_E); \qquad V - V_E = -60 \text{ mV}$$
$$g^{\mathrm{inh}} = I^{\mathrm{inh}}/(V - V_I); \qquad V - V_I = 60 \text{ mV}$$

The time-dependent conductances g^{exc} and g^{inh} were now injected into a different cell using a dynamic clamp procedure

(i.e., the input current was varied rapidly to maintain the correct relationship between the conductance and the membrane voltage) and the voltage was measured at a resolution of 0.1 ms.

4.1.1. Stimulus filtering

To model the relationship between the light stimulus and synaptic conductances, the current measurements I^{exc} and I^{inh} were fit to a linear-nonlinear model:

$$g^{\text{exc}}(t) = N^{\text{exc}}\left[L^{\text{exc}} * s(t) + \eta^{\text{exc}}\right],$$
$$g^{\text{inh}}(t) = N^{\text{inh}}\left[L^{\text{inh}} * s(t) + \eta^{\text{inh}}\right]$$

where s is the stimulus, L^{exc} (L^{inh}) is a linear filter, N^{exc} (N^{inh}) is a non-linear function, and η^{exc} (η^{inh}) is a noise term. The linear filter was fit by the function

$$L^{\text{exc}}(t) = P_{\text{exc}}\,(t/\tau_{\text{exc}})^{n_{\text{exc}}} \exp\left(-t/\tau_{\text{exc}}\right) \sin\left(2\pi t/T_{\text{exc}}\right) \quad (7)$$

and the non-linear filter by the polynomial

$$N^{\text{exc}}(x) = A_{\text{exc}}x^2 + B_{\text{exc}}x + C_{\text{exc}}. \quad (8)$$

Fits minimized the mean-square distance between model and data. L^{inh} and N^{inh} were fit using the same parametrization.

The noise terms η_k^{exc}, η_k^{inh} were fit to reproduce the statistical characteristics of the residuals from this fitting. We simulated the noise terms η^{exc} and η^{inh} using Ornstein–Uhlenbeck processes with the appropriate parameters; these were entirely characterized by the mean, standard deviation, and time constant of autocorrelation $\tau_{\eta,\text{exc}}$ $\left(\tau_{\eta,\text{inh}}\right)$, as well as pairwise correlation coefficients for noise terms entering neighboring cells. The noise correlation coefficients were estimated from the dual recordings of Trong and Rieke (2008).

Linear filter parameters computed (also listed in **Table 1**) were $P_{\text{exc}} = -8 \times 10^4 \text{s}^{-1}$, $n_{\text{exc}} = 3.6$, $\tau_{\text{exc}} = 12$ ms, $T_{\text{exc}} = 105$ ms, and $P_{\text{inh}} = -1.8 \times 10^5\ \text{s}^{-1}$, $n_{\text{inh}} = 3.0$, $\tau_{\text{inh}} = 16$ ms, $T_{\text{inh}} = 120$ ms. Non-linearity parameters were $A_{\text{exc}} = -8.3 \times 10^{-7}$ nS, $B_{\text{exc}} = 7 \times 10^{-3}$ nS, $C_{\text{exc}} = -0.95$ nS, and $A_{\text{inh}} = 1.67 \times 10^{-6}$ nS, $B_{\text{inh}} = 6.2 \times 10^{-3}$ nS, $C_{\text{inh}} = 4.17$ nS. Noise parameters were measured to be $\text{mean}(\eta_k^{\text{exc}}) = 30$, $\text{std}(\eta_k^{\text{exc}}) = 500$, $\tau_{\eta,\text{exc}} = 22$ ms, and $\text{mean}(\eta_k^{\text{inh}}) = -1200$, $\text{std}(\eta_k^{\text{inh}}) = 780$, $\tau_{\eta,\text{inh}} = 33$ ms. In addition, excitatory (inhibitory) noise to different cells η_k^{exc}, η_j^{exc} (η_k^{inh}, η_j^{inh}) had a correlation coefficient of 0.3 (0.15).

For the filter demonstrated in **Figure 4**, we added a cosine component to the previous filter, i.e.,

$$L^{\text{exc,M}}(t) = P_{\text{exc,M}}\left(t/\tau_{\text{exc,M}}\right)^{n_{\text{exc,M}}} \exp\left(-t/\tau_{\text{exc,M}}\right) \times \left[\sin\left(2\pi t/T_{\text{exc,M,S}}\right) + R_{\text{exc,M}} \cos\left(2\pi t/T_{\text{exc,M,C}}\right)\right] \quad (9)$$

Here $P_{\text{exc,M}} = -3.2 \times 10^5\ \text{s}^{-1}$, $n_{\text{exc,M}} = 2$, $\tau_{\text{exc,M}} = 12$ ms, $T_{\text{exc,M,S}} = 120$ ms and $T_{\text{exc,M,C}} = 100$ ms, and $P_{\text{inh,M}} = -3.5 \times 10^5\ \text{s}^{-1}$, $n_{\text{inh,M}} = 2$, $\tau_{\text{inh,M}} = 13.2$ ms, $T_{\text{inh,M,S}} = 132$ ms and $T_{\text{inh,M,C}} = 110$ ms, while $R_{\text{exc,M}} = R_{\text{inh,M}} = 0.8$.

4.1.2. Voltage evolution

We create a model of the cell as a non-linear integrate-and-fire model using the method of Badel et al. (2007), in which the membrane voltage is assumed to respond as

$$\frac{dV}{dt} = F\left(V, t - t_{\text{last}}\right) + \frac{I_{\text{input}}(t)}{C} \quad (10)$$

where C is the cell capacitance, t_{last} is the time of the last spike before time t, and $I_{\text{input}}(t)$ is a time-dependent input current. We use the current-clamp data, which yields cell voltage in response to the input current $I_{\text{input}}(t) = -g^{\text{exc}}(t)(V - V_E) - g^{\text{inh}}(V - V_I)$, to fit a function $F(V, t)$. When voltage data is segregated according to the time since the last spike $t - t_{\text{last}}$, the $I - V$ curve is well fit by a function of the form

$$F\left(V, t - t_{\text{last}}\right) = \frac{1}{\tau_m}\left(E_L - V + \Delta_T e^{(V - V_T)/\Delta_T}\right) \quad (11)$$

where parameters are the membrane time constant τ_m, resting potential (E_L), spike width Δ_T and knee of the exponential curve V_T.

The values of these constants differed in each bin of voltage data; to estimate these constants, we first extracted their values from each mean $I - V$ curve. We found that these constants, as a function of $t - t_{\text{last}}$, were well fit by either a single exponential or a difference of two exponentials, with relaxation to a baseline rate (as in Badel et al., 2007, **Figure 3**). Specifically, we chose:

$$\frac{1}{\tau_m} = c_{\tau_m,1} + c_{\tau_m,2} e^{-(t - t_{\text{last}})/c_{\tau_m,3}}$$
$$E_L = c_{E_L,1} + c_{E_L,2}\left(e^{-(t - t_{\text{last}})/c_{E_L,3}} - e^{-(t - t_{\text{last}})/c_{E_L,4}}\right)$$
$$\Delta_T = c_{\Delta_T,1} + c_{\Delta_T,2}\left(e^{-(t - t_{\text{last}})/c_{\Delta_T,3}} - e^{-(t - t_{\text{last}})/c_{\Delta_T,4}}\right)$$
$$V_T = c_{V_T,1} + c_{V_T,2} e^{-(t - t_{\text{last}})/c_{V_T,3}} \quad (12)$$

We obtained the coefficients by least-squares fitting to the above functional forms: specifically, we found that (up to four digits): $\left(c_{\tau_m,1}, c_{\tau_m,2}, c_{\tau_m,3}\right) = (0.3719\ \text{ms}^{-1},\ 0.5412\ \text{ms}^{-1},\ 13.2726\ \text{ms})$, $\left(c_{E_L,1}, c_{E_L,2}, c_{E_L,3}, c_{E_L,4}\right) = (-59.4858\ \text{mV},\ 5.8966\ \text{mV},\ 8.3076\ \text{ms},\ 233.1114\ \text{ms})$, $\left(c_{\Delta_T,1}, c_{\Delta_T,2}, c_{\Delta_T,3}, c_{\Delta_T,4}\right) = (20.0487\ \text{ms},\ 19.0560\ \text{ms},\ 3.6280\ \text{ms},\ 2.4304\ \text{s})$, and $\left(c_{V_T,1}, c_{V_T,2}, c_{V_T,3}\right) = (-44.3323\ \text{mV},\ 25.1812\ \text{mV},\ 4.7653\ \text{ms})$. Coefficients are also listed in **Table 2**.

The capacitance was inferred from the voltage trace data by finding, at a voltage value where the voltage/membrane current relationship is approximately Ohmic, the value of C that minimizes error in the relation Equation (10) (Badel et al., 2007). The estimated value was $C = 28$ pF.

4.1.3. Spiking dynamics: feedforward network

For simulations without electronic coupling, our model neuron comprises Equations (10, 11) for $V < V_{\text{threshold}}$; a spike was detected when V reached $V_{\text{threshold}} = -30$ mV; voltage was then reset to $V_{\text{reset}} = -55$ mV. The cell was then unable to spike for an absolute refractory period of $\tau_{\text{abs}} = 3$ ms.

All simulations presented here were done in a three-cell network.

4.1.4. Spiking dynamics: recurrent network

Gap junction coupling was introduced as an additional current on the right-hand side of Equation (10):

$$\frac{I_{\text{gap},j}}{C} = -\frac{g^{\text{gap}}}{C}\sum_{k\neq j}\left(V_j - V_k\right) \tag{13}$$

The coupling strength g^{gap} was held constant during a simulation. When coupling was present (i.e., when $g^{\text{gap}} \neq 0$), g^{gap} was varied from the measured level (1.1 nS) (Trong and Rieke, 2008) to 16 times this value (17.6 nS) between simulations. When present, coupling was all-to-all.

As in the feedforward model, Equations (10, 11) were integrated for $V < V_{\text{threshold}}$, and a spike was detected when V reached $V_{\text{threshold}} = -30$ mV. To model the voltage trajectory immediately following a spike, an averaged spike waveform was extracted from voltage traces of the same ON parasol cell used to fit Equations (10, 11). This spike waveform was then used to replace 1 ms of the membrane voltage trajectory during and after a spike; at the end of the 1 ms, the voltage was released at approximately −58 mV. The cell was unable to spike for an absolute refractory period of $\tau_{\text{abs}} = 3$ ms. A relative refractory period was induced by introducing a declining threshold for the period of 3–6 ms following a spike, after which $V_{\text{threshold}}$ returns to −30 mV.

4.1.5. Cell receptive field and stimulation

We defined each cell's stimulus as the linear convolution of an image with its receptive field. The receptive fields include an ON center and an OFF surround, as in Chichilnisky and Kalmar (2002):

$$s_j(\vec{x}) = \exp\left(-\frac{1}{2}\left(\vec{x}-\vec{x}_j\right)^T \mathbf{Q}\left(\vec{x}-\vec{x}_j\right)\right) - k\exp\left(-\frac{1}{2}r\left(\vec{x}-\vec{x}_j\right)\mathbf{Q}r\left(\vec{x}-\vec{x}_j\right)\right) \tag{14}$$

where the parameters k and $1/r$ give the relative strength and size of the surround. $\mathbf{Q}$ specifies the shape of the center and was chosen to have a 1 standard deviation (SD) radius of 50 μm and to be perfectly circular. The receptive field locations $\vec{x}_1$, $\vec{x}_2$, and $\vec{x}_3$ were chosen so that the 1 SD outlines of the receptive field centers will tile the plane (i.e., they just touch). Other parameters used were $k = 0.3$, $r = 0.675$.

Stimulation images were defined on a 512 μm × 512 μm grid that overlapped all three receptive fields. For full-field stimuli, light intensity was chosen be spatially constant and refreshed every 8, 40, or 100 ms by choosing independently from the specified stimulus distribution (Gaussian, binary, Cauchy, or heavy-tailed skew). For spatially variable stimuli, a checkerboard pattern was imposed on the stimulation image: the intensity value in each checkerboard square was chosen independently and refreshed

Table 1 | Parameters used to model the transformation of stimuli into synaptic conductances for the RGC model, as described in Equations (7–9).

Model (MOD)	P_{MOD} (s^{-1})	τ_{MOD} (ms)	n_{MOD}	T_{MOD} (ms)	A_{MOD} (nS)	B_{MOD} (nS)	C_{MOD} (nS)
exc	-8×10^4	12	3.6	105	-8.3×10^{-7}	7×10^{-3}	−0.95
inh	-1.8×10^5	16	3.0	120	1.67×10^{-6}	6.2×10^{-3}	4.17
exc,M	-3.2×10^5	12	2	120*	-8.3×10^{-7}	7×10^{-3}	−0.95
inh,M	-3.5×10^5	13.2	2	132*	1.67×10^{-6}	6.2×10^{-3}	4.17

Additional parameters for monophasic filters

Model (MOD)	$T_{\text{MOD, S}}$ (ms)	$T_{\text{MOD, C}}$ (ms)	R_{MOD}
exc,M	120	100	0.8
inh,M	132	110	0.8

Asterisks (*) indicate parameters that are superceded by later rows; note that the monophasic filter equations contain two filtering timescales—for example $T_{exc,M,S}$ and $T_{exc,M,C}$, for the excitatory monophasic filter—and a relative weighting (e.g., $R_{exc,M}$).

Table 2 | Coefficients used to define refractory EIF model as specified in Equations (11, 12).

Parameter (PAR)	$c_{\text{PAR,1}}$	$c_{\text{PAR,2}}$	$c_{\text{PAR,3}}$ (ms)	$c_{\text{PAR,4}}$ (ms)
τ_m (actual fit: $1/\tau_m$)	0.3719 ms^{-1}	0.5412 ms^{-1}	13.2726	
V_T	−44.3323 mV	25.1812 mV	4.7653	
E_L	−59.4858 mV	5.8966 mV	8.3076	233.1114
Δ_T	20.0487 ms	19.0560 ms	3.6280	2430.4

The parameters $1/\tau_m$ and V_T were fit to single exponentials as functions of time, with three free parameters. The parameters E_L and Δ_T were fit to differences of exponentials and therefore have four parameters. Units in the first and second columns are as stated; coefficients in the third and fourth column are in units of milliseconds (ms).

at the appropriate interval. The checkerboard pattern was first given a random rotation and translation relative to the receptive fields: this was chosen at the outset of each batch of stixel simulations (for a total of five rotation/translation pairs per stixel size, refresh rate, and stimulus distribution). Two example placements are shown in **Figures S2A,D** for 256 μm and 60 μm pixels respectively.

4.1.6. Numerical methods

All simulations and data analysis were performed using MATLAB. Equations (10, 11) were integrated using the Euler method for $>10^5$ ms with a time step of 0.1 ms. The synaptic noise terms, η_k^{exc} and η_k^{inh}, as well as the light input, were generated independently for each simulation. In response to uniform light stimuli, firing rates were 11.51 ± 0.38 Hz (standard deviations given across a total of 60 cells; 3 cells each from 20 10^5 ms simulations); 10 ms bins were used to discretize the spiking output. Firing rates were higher for full-field stimuli, ranging from 12 to 43 Hz (firing rates increased with stimulus variance); therefore shorter (5 ms) bins were used to discretize spike output for all other simulations. With this range of firing rates and bin size, multiple spikes were very rare (occurring in <1% of occupied bins). Empirical spiking distributions were computed from the binned spike data.

For each stimulus condition, 20 simulations (or sub-simulations) were run, for a total integration time of $> 20 \times 10^5$ ms. These 20 sub-simulations were used to estimate standard errors in both the probability distribution over spiking events and $D_{\text{KL}}(P, \tilde{P})$. Numbers reported in section 2 are, unless specified otherwise, produced by collating the data from the 20 simulations.

To fit a maximum entropy model $\tilde{P}$ to an empirical probability distribution P, we used standard methods that have been explained elsewhere (Malouf, 2002). Briefly, we minimized the negative log-likelihood function:

$$L(\lambda) = -\sum_{\mathbf{x}} P(\mathbf{x}) \log \tilde{P}(\mathbf{x}, \lambda) \tag{15}$$

where

$$\tilde{P}(\mathbf{x}, \lambda) = Z_\lambda^{-1} \exp\left(\sum_k \lambda_k f_k(\mathbf{x})\right);$$

Z_λ is the partition function, $f_k, k = 1, \ldots, M$ is a set of functions or "features" of the spiking state, and λ is a vector of parameters, each of which serves as a Lagrange multiplier enforcing the constraint $\mathbf{E}_{\tilde{P}}[f_k]$. For the pairwise (PME) model on N cells, λ corresponds to N firing rates and $N(N-1)/2$ covariances, and the sum is over all possible spiking states of the system. For $N = 3$ there are six such parameters, and

$$\log \tilde{P}(\{x_1, x_2, x_3\}, \lambda) = \lambda_1 x_1 + \lambda_2 x_2 + \lambda_3 x_3 + \lambda_{1,2} x_1 x_2 + \lambda_{2,3} x_2 x_3 + \lambda_{1,3} x_1 x_3 - \log Z_\lambda.$$

The function in Equation (15) is a convex function of the parameters λ which will be minimized precisely (and uniquely) when $\tilde{P}$ matches the desired moments from P: e.g., $\mathbf{E}_P[x_1] = \mathbf{E}_{\tilde{P}}[x_1]$. Since $\tilde{P}$ is in log-linear form, the result will be the *maximum entropy* distribution that matches the desired moments (Malouf, 2002). In principle any unconstrained gradient descent method may be used; we used an implementation of the non-linear conjugate gradient method. The Kullback Leibler divergence $D_{\text{KL}}(P, \tilde{P})$ was computed using the identity $D_{\text{KL}}(P, \tilde{P}) = S(\tilde{P}) - S(P)$, where $S(P)$ is the entropy of P, i.e., $S(P) = -\sum_{\mathbf{x}} P(\mathbf{x}) \log P(\mathbf{x})$.

4.1.7. Convergence testing

To test our finding that the observed distributions were well-modeled by the PME fit, we also performed the PME analysis on each of the 20 simulations for each stimulus condition. While in general $D_{\text{KL}}(P, \tilde{P})$ can be quite sensitive to perturbations in P, the numbers remained small under this analysis. To confirm that our results for $D_{\text{KL}}(P, \tilde{P})$ are sufficiently resolved to remove bias from sampling, we performed an analysis in which we collect the 20 simulations in subgroups of 1, 2, 4, 5, 10, and 20, and plot the mean D_{KL} with estimated standard errors. As expected (e.g., Paninski, 2003), bias decreases as the length of subgroup increases and asymptotes at—or before—the full simulation length.

To provide a cross-validation test for the significance of our reported $D_{\text{KL}}(P, \tilde{P})$ values, we divided our data into halves (which we denote P_1 and P_2, each including data from 10 sub-simulations) and performed the PME analysis on one half (say P_1) to yield a model $\tilde{P}_1$. We then computed $D_{\text{KL}}(P_2, \tilde{P}_1)$ and $D_{\text{KL}}(P_2, P_1)$ (as in Yu et al., 2011), which we refer to the *cross-validated* and *empirical* likelihood, respectively. The former tests whether the PME fit is robust to over-fitting; the latter tests how well-resolved our "true" distribution is in the first place. Most cross-validated likelihoods fall on or near the identity line; most empirical likelihoods are close to zero [and importantly, significantly smaller than either $D_{\text{KL}}(P, \tilde{P})$ or $D_{\text{KL}}(P_2, \tilde{P}_1)$, indicating that $D_{\text{KL}}(P, \tilde{P})$ is accurately resolved]. We conclude that the deviations that we observe when these conditions are met can not be accounted for by the differences in testing and training data.

4.2. COMPUTATION OF SPIKING PATTERNS IN THE SIMPLIFIED MODEL

As a simplified model of a neural circuit, we consider a variant of the *Dichotomized Gaussian* (Amari et al., 2003; Macke et al., 2009, 2011), in which correlated inputs are thresholded to produce an output spike pattern. To be concrete, a set of N threshold spiking units is forced by a common input I_c [drawn from a probability distribution $P_C(y)$] and an independent input I_j [drawn from a probability distribution $P_I(y)$]. To relate these functions to the other free parameters in the model, $P_C(y)$ and $P_I(y)$ were always chosen so that I_j and I_c had mean 0 and variances $(1-c)\sigma^2$ and $c\sigma^2$, respectively (so that c yields the Pearson's correlation coefficient of the input to two cells). The output of each cell x_j is determined by summing and thresholding these inputs:

$$x_j = H\left(I_j + I_c - \Theta\right) \tag{16}$$

where H is the Heaviside function [$H(x) = 1$ if $x \geq 0$; $H(x) = 0$ otherwise]. Conditioned on I_c, the probability of each spike is

given by:

$$\begin{aligned}\textbf{Prob}\left[x_j = 1 \mid I_c = a\right] &= \textbf{Prob}\left[I_j + a - \Theta > 0\right] \\ &= \textbf{Prob}\left[I_j > \Theta - a\right] \\ &= \int_{\Theta-a}^{\infty} P_I(y)\, dy\end{aligned}$$

Similarly, we have the conditioned probability that $x_j = 0$:

$$\begin{aligned}\textbf{Prob}\left[x_j = 0 \mid I_c = a\right] &= \textbf{Prob}\left[I_j + a - \Theta < 0\right] \\ &= \textbf{Prob}\left[I_j < \Theta - a\right] \\ &= \int_{-\infty}^{\Theta-a} P_I(y)\, dy\end{aligned}$$

Because these are conditionally independent, the probability of any spiking event $(x_1, x_2, \ldots, x_N) = (A_1, A_2, \ldots, A_N)$ is given by the integral of the product of the conditioned probabilities against the density of the common input.

$$\textbf{Prob}\left[x_1 = A_1, \ldots, x_N = A_N\right] = \int_{-\infty}^{\infty} dy\, P_C(y) \prod_{j=1}^{N} \textbf{Prob}\left[x_j = A_j \mid I_c = y\right] \tag{17}$$

The integral in Equation (17) is numerically evaluated via an adaptive quadrature routine, such at Matlab's `quad` or `integral`.

Four distinct unimodal inputs were used; two with heavy tails (Cauchy and heavy-tailed with skew), and two with sub-Gaussian tails (Gaussian and skewed). A random variable X is *sub-Gaussian* if the probability of large events can be bounded above by a scaled Gaussian; that is, if there exist constants $C, c > 0$ such that

$$P\left(|X| > \lambda\right) \leq C \exp\left(-c\lambda^2\right)$$

for all λ (e.g., see Tao, 2012, p. 15).

Unimodal inputs I_j, I_c were chosen from different marginals with mean 0 and variances $(1 - c)\sigma^2$, $c\sigma^2$, respectively (for simplicity, we use σ^2 to refer to the variance of a generic probability distribution in the following three paragraphs). For Gaussian inputs with variance σ^2, $P(x) \propto e^{-x^2/2\sigma^2}$; for skewed inputs, $P(x) \propto (x + \mu)e^{-(x+\mu)^2/2a}$, for $x > -\mu$, where the parameter a sets the variance $2a(1 - \frac{\pi}{4})$ and shifting by $\mu = \sqrt{\frac{a\pi}{2}}$ ensures that the mean of $P(x)$ is zero.

The heavy-tailed unimodal inputs were chosen so that the rate of tail decay would mimic the I^{-2} luminance statistics found in natural scenes (Ruderman and Bialek, 1994):

$$P(x) \propto \frac{1}{x^2 + 1}, \qquad -X < x < X$$

$$P(x) \propto \frac{x}{\left(x^2 + 1\right)^{3/2}}, \qquad 0 \leq x < X$$

for the Cauchy and heavy-tailed with skew distributions, respectively. A finite support of X was necessary in order to ensure the distributions had finite moments; X was chosen to be 1000. Given X, the distributions were shifted and scaled to ensure mean 0 and variance σ^2.

Bimodal inputs with variance σ^2 were chosen in the following way: in all cases, $P(x)$ was chosen to be a discrete distribution with support on two values $\{0, X\}$ i.e., $P(X) = p$ and $P(0) = 1 - p$. If possible (i.e., if $\sigma^2 \leq 1/4$), X was chosen to be 1; otherwise, X was chosen so as to minimize the distance between 0 and X. Finally, $P(x)$ was shifted to have the desired mean value.

ACKNOWLEDGMENTS

This research was supported by NSF grants DMS-0817649 and 1056125 by a Career Award at the Scientific Interface from the Burroughs-Wellcome Fund (Eric Shea-Brown), by the Howard Hughes Medical Institute and by NIH grant EY-11850 (Fred Rieke), by a Trinity College Research Studentship (Julijana Gjorgjieva), and by an Early Career Award from the Mathematical Biosciences Institute (Andrea K. Barreiro).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://www.frontiersin.org/journal/10.3389/fncom.2014.00010/abstract

Figure S1 | Biphasic vs. monophasic filters used in simulations illustrated in Figure 4.

Figure S2 | Illustration of RGC simulations with light stimuli of varying spatial scale ("stixels"). (A–C) For stixel size 60 μm, results for one randomly chosen stimulus position. **(A)** Contour lines of the three receptive fields (at 0.5, 1, 1.5, and 2 SD; and at the zero contour line) superimposed on the stimulus checkerboard (for illustration, pictured in an alternating black/white pattern). The red scale bar indicates 100 μm. **(B)** Histograms of the excitatory conductances, for each cell. **(C)** Spike pattern distribution, as obtained from computational simulations of the RGC model ("Observed"; dark blue), and the corresponding pairwise fit ("PME"; light pink). All eight spike patterns are shown, to allow for the possibility of non-symmetric responses; the three different probabilities labeled p_1 correspond to $P[(1, 0, 0)]$, $P[(0, 1, 0)]$, and $P[(0, 0, 1)]$. **(D–F)** As in **(A–C)**, but for stixel size 256 μm. Panels **(E,F)** demonstrate that for this input, both excitatory inputs and spiking responses are heterogenous across the RGCs.

Figure S3 | Strength of higher-order interactions produced by the threshold model as input parameters vary; relationship with other output firing statistics. (A) For skewed common inputs: $D_{\mathrm{KL}}(P, \tilde{P})$ as a function of input correlation c and input standard deviation σ, for a fixed threshold $\Theta = 1.5$. Color indicates $D_{\mathrm{KL}}(P, \tilde{P})$; see color bar for range. **(B)** For skewed common inputs: $D_{\mathrm{KL}}(P, \tilde{P})$ vs. firing rate $\mathbf{E}[x_1]$ (Left) and the fraction of multi-information (Δ) captured by the PME model vs. firing rate $\mathbf{E}[x_1]$ (Right). In **(B)**, possible input parameters were varied over a broad range as described in section 2. Firing rate is defined as the probability of a spike occurring per cell per random draw of the sum-and-threshold model, as defined in Equation (16). Color indicates output correlation coefficient ρ ranging from black for $\rho \in (0, 0.1)$, to white for $\rho \in (0.9, 1)$, as illustrated in the color bars. **(C,D)**: as in **(A,B)**, but for heavy-tailed, skewed common inputs.

Figure S4 | The range of higher-order interactions produced by the threshold model varies across input type. Here, all values of $D_{\mathrm{KL}}(P, \tilde{P})$ produced by the three-cell threshold model (previously displayed in **Figures 7**, **S3**) are superimposed to show the contrast between different input distributions. By comparing these data with data from direct sampling of all symmetric spiking distributions on three cells (from **Figure 1** and shown here in yellow), one can see that only a limited set of output patterns are accessed by the feedforward thresholding model. Firing rate is defined as the probability of a spike occurring per cell per random draw of the sum-and-threshold model, as defined in Equation (16).

REFERENCES

Amari, S. (2001). Information geometry on hierarchy of probability distributions. *IEEE Trans. Inf. Theory* 47, 1701–1711. doi: 10.1109/18.930911

Amari, S., Nakahara, H., Wu, S., and Sakai, Y. (2003). Synchronous firing and higher-order interactions in neuron pool. *Neur. Comp.* 15, 127–142. doi: 10.1162/089976603321043720

Badel, L., Lefort, S., Berger, T. K., Petersen, C. C. H., Gerstner, W., and Richardson, M. J. E. (2008). Extracting non-linear integrate-and-fire models from experimental data using dynamic I–V curves. *Biol. Cybern.* 99, 361–370. doi: 10.1007/s00422-008-0259-4

Badel, L., Lefort, S., Brette, R., Petersen, C. C. H., Gerstner, W., and Richardson, M. J. E. (2007). Dynamic I-V curves are reliable predictors of naturalistic pyramidal-neuron voltage traces. *J. Neurophys.* 99, 656–666. doi: 10.1152/jn.01107.2007

Barreiro, A. K., Shea-Brown, E. T., and Thilo, E. L. (2010). Timescales of spike-train correlation for neural oscillators with common drive. *Phys. Rev. E* 81, 011916. doi: 10.1103/PhysRevE.81.011916

Barreiro, A. K., Thilo, E. L., and Shea-Brown, E. T. (2012). A-current and type I / type II transition determine collective spiking from common input. *J. Neurophysiol.* 108, 1631–1645. doi: 10.1152/jn.00928.2011

Baudry, M., and Taketani, M., (eds). (2006). *Advances in Network Electrophysiology Using Multi-Electrode Arrays*. New York, NY: Springer Press. doi: 10.1007/0-387-25858-2_15

Bethge, M., and Berens, P. (2008). Near-maximum entropy models for binary neural representations of natural images. *Adv. Neur. Inf. Proc. Syst.* 20, 97–104. doi: 10.1.1.68.3149

Bohte, S. M., Spekreijse, H., and Roelfsema, P. R. (2000). The effects of pair-wise and higher-order correlations on the firing rate of a postsynaptic neuron. *Neur. Comp.* 12, 153–179. doi: 10.1162/089976600300015934

Cain, N., and Shea-Brown, E. (2013). Impact of correlated neural activity on decision making performance. *Neur. Comp.* 25, 289–327. doi: 10.1162/NECO_a_00398

Chichilnisky, E. J., and Kalmar, R. S. (2002). Functional asymmetries in ON and OFF ganglion cells of primate retina. *J. Neurosci.* 22, 2737–2747. doi: 20026215

Cocco, S., Leibler, S., and Monasson, R. (2009). Neuronal couplings between retinal ganglion cells inferred by efficient inverse statistical physics methods. *Proc. Natl. Acad. Sci. U.S.A.* 106, 14058–14062. doi: 10.1073/pnas.0906705106

Cover, T. M., and Thomas, J. A. (1991). *Elements of Information Theory*. New York: Wiley. doi: 10.1002/0471200611

Dacey, D., and Brace, S. (1992). A coupled network for parasol but not midget ganglion cells in the primate retina. *Vis. Neurosci.* 9, 279–290. doi: 10.1017/S0952523800010695

Dayan, P., and Abbot, L. (2001). *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems*. Cambridge, MA: MIT Press. doi: 10.1016/S0306-4522(00)00552-2

de la Rocha, J., Doiron, B., Shea-Brown, E., Josic, K., and Reyes, A. (2007). Correlation between neural spike trains increases with firing rate. *Nature* 448, 802–806. doi: 10.1038/nature06028

Fairhall, A., Burlingame, C., Narasimhan, R., Harris, R., Puchalla, K., and Berry, M. (2006). Selectivity for multiple stimulus features in retinal ganglion cells. *J. Neurophys.* 96, 2724–2738. doi: 10.1152/jn.00995.2005

Ganmor, E., Segev, R., and Schneidman, E. (2011). Sparse low-order interaction network underlies a highly correlated and learnable population code. *Proc. Natl. Acad. Sci. U.S.A.* 108, 9679–9684. doi: 10.1073/pnas.1019641108

Hong, S., Ratte, S., Prescott, S., and De Schutter, E. (2012). Single neuron firing properties impact correlation-based population coding. *J. Neurosci.* 32, 1413–1428. doi: 10.1523/JNEUROSCI.3735-11.2012

Jaynes, E. T. (1957a). Information theory and statistical mechanics. *Physiol. Rev.* 106, 620–630. doi: 10.1103/PhysRev.106.620

Jaynes, E. T. (1957b). Information theory and statistical mechanics II. *Physiol. Rev.* 108, 171–190. doi: 10.1103/PhysRev.108.171

Koster, U., Sohl-Dickstein, J., Gray, C. M., and Olshausen, B. A. (2013). Higher order correlations within cortical layers dominate functional connectivity in microcolumns. ArXiv q-Bio/1301.0050.

Krumin, M., and Shoham, S. (2009). Generation of spike trains with controlled auto- and cross-correlation functions. *Neur. Comp.* 21, 1642–1664. doi: 10.1162/neco.2009.08-08-847

Kuhn, A., Aertsen, A., and Rotter, S. (2003). Higher-order statistics of input ensembles and the response of simple model neurons. *Neur. Comp.* 15, 67–101. doi: 10.1162/089976603321043702

Leen, D., and Shea-Brown, E. (2013). A simple mechanism for higher-order correlations in integrate-and-fire neurons. ArXiv q-Bio.NC/1306.5275.

Macke, J. H., Berens, P., Ecker, A. S., Tolias, A. S., and Bethge, M. (2009). Generating spike trains with specified correlation coefficients. *Neur. Comp.* 21, 397–423. doi: 10.1162/neco.2008.02-08-713

Macke, J. H., Opper, M., and Bethge, M. (2011). Common input explains higher-order correlations and entropy in a simple model of neural population activity. *Phys. Rev. Lett.* 106, 208102. doi: 10.1103/PhysRevLett.106.208102

Malouf, R. (2002). "A comparison of algorithms for maximum entropy parameter estimation," in *Proceedings of the Sixth Conference on Natural Language Learning* (Stroudsburg, PA), 49–55. doi: 10.3115/1118853.1118871

Marella, S., and Ermentrout, G. B. (2008). Class-II neurons display a higher degree of stochastic synchronization than class-I neurons. *Phys. Rev. E* 77, 041908. doi: 10.1103/PhysRevE.77.041918

Martignon, L., Deco, G., Laskey, K., Diamond, M., Freiwald, W., and Vaadia, E. (2000). Neural coding: higher-order temporal patterns in the neurostatistics of cell assemblies. *Neur. Comp.* 12, 2621–2653. doi: 10.1162/089976600300014872

McCulloch, W. S., and Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bull. Math. Biophys.* 5, 115–137. doi: 10.1007/BF02478259

Montani, F., Ince, R. A. A., Senatore, R., Arabzadeh, E., Diamond, M. E., and Panzeri, S. (2009). The impact of high-order interactions on the rate of synchronous discharge and information transmission in somatosensory cortex. *Phil. Trans. R. Soc. A* 367, 3297–3310. doi: 10.1098/rsta.2009.0082

Moreno, R., de la Rocha, J., Renart, A., and Parga, N. (2002). Response of spiking neurons to correlated inputs. *Phys. Rev. Lett.* 89, 288101. doi: 10.1103/PhysRevLett.89.288101

Murphy, G. J., and Rieke, F. (2006). Network variability limits stimulus-evoked spike timing precision in retinal ganglion cells. *Neuron* 52, 511–524. doi: 10.1016/j.neuron.2006.09.014

Nowotny, T., and Huerta, R. (2003). Explaining synchrony in feed-forward networks: are McCulloch-Pitts neurons good enough? *Biol. Cybern.* 89, 237–241. doi: 10.1007/s00422-003-0431-9

Ohiorhenuan, I. E., Mechler, F., Purpura, K. P., Schmid, A. M., Hu, Q., and Victor, J. D. (2010). Sparse coding and high-order correlations in fine-scale cortical networks. *Nature* 466, 617–621. doi: 10.1038/nature09178

Ohiorhenuan, I. E., and Victor, J. D. (2010). Information-geometric measure of 3-neuron firing patterns characterizes scale-dependence in cortical networks. *J. Comp. Neurosci.* 30, 125–141. doi: 10.1007/s10827-010-0257-0

Oizumi, M., Ishii, T., Ishibashi, K., and Okada, M. (2010). Mismatched decoding in the brain. *J. Neurosci.* 30, 4815–4826. doi: 10.1523/JNEUROSCI.4360-09.2010

Paninski, L. (2003). Estimation of entropy and mutual information. *Neur. Comp.* 15, 1191–1253. doi: 10.1162/089976603321780272

Roudi, Y., Nirenberg, S., and Latham, P. E. (2009a). Pairwise maximum entropy models for studying large biological systems: when they can work and when they can't. *PLoS Comp. Biol.* 5:e1000380. doi: 10.1371/journal.pcbi.1000380

Roudi, Y., Tyrcha, J., and Hertz, J. (2009b). Ising model for neural data: model quality and approximate methods for extracting functional connectivity. *Phys. Rev. E* 79, 051915. doi: 10.1103/PhysRevE.79.051915

Ruderman, D. L., and Bialek, W. (1994). Statistics of natural images: scaling in the woods. *Phys. Rev. Lett.* 73, 814–818. doi: 10.1103/PhysRevLett.73.814

Santos, G. S., Gireesh, E. D., Plenz, D., and Nakahara, H. (2010). Hierarchical interaction structure of neural activities in cortical slice cultures. *J. Neurosci.* 30, 8720–8733. doi: 10.1523/JNEUROSCI.6141-09.2010

Schneidman, E., Berry (II), M. J., Segev, R., and Bialek, W. (2006). Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature* 440, 1007–1012. doi: 10.1038/nature04701

Schneidman, E., Still, S., Berry (II), M. J., and Bialek, W. (2003). Network information and connected correlations. *Phys. Rev. Lett.* 91, 238701. doi: 10.1103/PhysRevLett.91.238701

Sharpe, L. T., Whittle, P., and Nordby, K. (1993). Spatial integration and sensitivity changes in the human rod visual system. *J. Physiol.* 461, 235–246.

Shea-Brown, E., Josić, K., Doiron, B., and de la Rocha, J. (2008). Correlation and synchrony transfer in integrate-and-fire neurons: basic properties and consequences for coding. *Phys. Rev. Lett.* 100, 108102. doi: 10.1103/PhysRevLett.100.108102

Shlens, J., Field, G. D., Gauthier, J. L., Greschner, M., Sher, A., Litke, A. M., et al. (2009). The structure of large-scale synchronized firing in primate retina. *J. Neurosci.* 29, 5022–5031. doi: 10.1523/JNEUROSCI.5187-08.2009

Shlens, J., Field, G. D., Gauthier, J. L., Grivich, M. I., Petrusca, D., Sher, A., et al. (2006). The structure of multi-neuron firing patterns in primate retina. *J. Neurosci.* 26, 8254–8266. doi: 10.1523/JNEUROSCI.1282-06.2006

Tang, A., Jackson, D., Hobbs, J., Smith, J. L., Patel, H., Prieto, A., et al. (2008). A maximum entropy model applied to spatial and temporal correlations from cortical networks *in vitro*. *J. Neurosci.* 28, 505–518. doi: 10.1523/JNEUROSCI.3359-07.2008

Tao, T. (2012). *Topics in Random Matrix Theory*. Providence, RI: American Mathematical Society. doi: 10.1142/S2010326311500018

Tchumatchenko, T., Malyshev, A., Geisel, T., and Wolf, F. (2010). Correlations and synchrony in threshold neuron models. *Phys. Rev. Lett.* 104, 058102. doi: 10.1103/PhysRevLett.104.058102

Tkacik, G., Ghosh, A., Schneidman, E., and Segev, R. (2012). Retinal adaptation and invariance in changes to higher-order stimulus statistics. arXiv:1201.3552.

Tkacik, G., Schneidman, E., Berry II, M. J., and Bialek, W. (2009). Spin glass models for a network of real neurons. arXiv:0912.5409.

Trong, P. K., and Rieke, F. (2008). Origin of correlated activity between parasol retinal ganglion cells. *Nat. Neurosci.* 11, 1343–1351. doi: 10.1038/nn.2199

Victor, J. (1999). Temporal aspects of neural coding in the retina and lateral geniculate. *Netw. Comput. Neur. Syst.* 10, 1–66. doi: 10.1088/0954-898X/10/4/201

Vilela, R. D., and Lindner, B. (2009). Comparative study of different integrate-and-fire neurons: spontaneous activity, dynamical response, and stimulus-induced correlation. *Phys. Rev. E* 80, 031909. doi: 10.1103/PhysRevE.80.031909

Yu, S., Huang, D., Singer, W., and Nikolić, D. (2008). A small world of neuronal synchrony. *Cereb. Cortex* 18, 2891–2901. doi: 10.1093/cercor/bhn047

Yu, S., Yang, H., Nakahara, H., Santos, G., Nikolić, D., and Plenz, D. (2011). Higher-order interactions characterized in cortical activity. *J. Neurosci.* 31, 17514–17526. doi: 10.1523/JNEUROSCI.3127-11.2011

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.

Received: 01 July 2013; accepted: 20 January 2014; published online: 06 February 2014.
Citation: Barreiro AK, Gjorgjieva J, Rieke F and Shea-Brown E (2014) When do microcircuits produce beyond-pairwise correlations? Front. Comput. Neurosci. ***8****:10. doi: 10.3389/fncom.2014.00010*
This article was submitted to the journal Frontiers in Computational Neuroscience.
Copyright © 2014 Barreiro, Gjorgjieva, Rieke and Shea-Brown. 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) or licensor 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.

APPENDIX

A.1 A MEASURE OF HIGHER-ORDER INTERACTIONS: $D_{\mathrm{KL}}(P, \tilde{P})$

We begin by observing that when $\tilde{P}$ is a maximum entropy distribution that approximates P (that is, it is log-linear, with coefficients chosen to enforce equality of a set of moments), then the KL-distance may be written as a difference of entropies (Cover and Thomas, 1991; Malouf, 2002):

$$D_{\mathrm{KL}}\left(P, \tilde{P}\right) = -S(P) + S\left(\tilde{P}\right)$$

Here, the entropy of a probability distribution P on $\{0, 1\}^3$ is given

$$S(P) = -p_0 \log\left(p_0\right) - 3p_1 \log\left(p_1\right) - 3p_2 \log\left(p_2\right) - p_3 \log\left(p_3\right) \tag{18}$$

if we use the fact that the distributions are permutation-symmetric [i.e., $p_1 \equiv P(1, 0, 0) = P(0, 1, 0) = P(0, 0, 1)$]. We take the logarithms in the definitions of the entropy S and KL-divergence D_{KL} to be base 2, so that any numerical values of these quantities are in units of bits. Using the fact that P must normalize to 1, we rewrite

$$S(P) = -\left(1 - 3p_1 - 3p_2 - p_3\right) \log\left(1 - 3p_1 - 3p_2 - p_3\right) - 3p_1 \log\left(p_1\right) - 3p_2 \log\left(p_2\right) - p_3 \log\left(p_3\right)$$

where the set of admissible distributions may now be described by the convex tetrahedron in $\mathbb{R}^3$, $\mathcal{C} = \{p_1, p_2, p_3 \geq 0;\ 3p_1 + 3p_2 + p_3 \leq 1\}$

We note that the set of distributions which satisfies a desired set of lower order moments is given by an affine subspace (in $\mathbb{R}^3$, a line) which intersects this tetrahedron:

$$\mu \equiv \mathbf{E}[X_i] = p_1 + 2p_2 + p_3$$
$$\hat{\rho} \equiv \mathbf{E}[X_i^2] = p_2 + p_3$$

Denoting this set by $\mathcal{C}_{\mu,\hat{\rho}}$, we note that $\mathcal{C}_{\mu,\hat{\rho}}$ is a convex set and that $S(\tilde{P})$ is constant on each $\mathcal{C}_{\mu,\hat{\rho}}$.

By straightforward differentiation we can check that the Hessian of $-S(P)$ is positive definite, as long as the probabilities p_0, p_1, etc. are strictly greater than zero:

$$-D^2 S(P) = \begin{bmatrix} \frac{3}{p_1} & 0 & 0 \\ 0 & \frac{3}{p_2} & 0 \\ 0 & 0 & \frac{1}{p_3} \end{bmatrix} + \frac{1}{p_0} \begin{bmatrix} 9 & 9 & 3 \\ 9 & 9 & 3 \\ 3 & 3 & 1 \end{bmatrix}$$

Therefore $-S(P)$ is convex on $\mathcal{C}_{\mu,\hat{\rho}}$; since $S(\tilde{P})$ is constant, $D_{\mathrm{KL}}(P, \tilde{P})$ is likewise convex on $\mathcal{C}_{\mu,\hat{\rho}}$. As a consequence, if $D_{\mathrm{KL}}(P, \tilde{P})$ has a local minimum, then it is unique and a global minimum as well. Since $D_{\mathrm{KL}}(P, \tilde{P}) \geq 0$ with equality if and only if $P = \tilde{P}$, this minimum must be achieved occurs when $P = \tilde{P}$; the maximum is likewise achieved on the boundary of the admissible region $\mathcal{C}_{\mu,\hat{\rho}}$.

A.2 A MEASURE OF HIGHER-ORDER INTERACTIONS: STRAIN

We define the *strain*,

$$\gamma = \log\left(\frac{p_3 p_1^3}{p_0 p_2^3}\right) = \log p_3 - \log p_0 + 3\log p_1 - 3\log p_2 \tag{19}$$

a potential measure of the importance of higher-order interactions (Ohiorhenuan and Victor, 2010). By Equation (3), we can see that $\gamma = 0$ precisely for a pairwise maximum entropy (PME) distribution. We will show that as the distribution (p_0, p_1, p_2, p_3) is moved away from the constraint surface while fixing lower-order moments, the strain increases monotonically.

From the definition of lower-order moments,

$$\mu = \mathbf{E}[X_i] = p_1 + 2p_2 + p_3$$
$$\hat{\rho} = \mathbf{E}[X_i X_j] = p_2 + p_3$$

we can verify that in order to keep μ, $\hat{\rho}$ constant, if p_1 increases by z (i.e., $p_1 \to p_1 + z$), then we must also have $p_2 \to p_2 - z$ and $p_3 \to p_3 + z$. Then if each probability is strictly positive, then the derivative

$$\frac{\partial \gamma}{\partial z} = \frac{1}{p_3 + z} + \frac{1}{1 - p_3 - 3p_1 - 3p_2 - z} + \frac{3}{p_1 + z} + \frac{3}{p_2 - z}$$

is strictly positive as well. In particular, it is strictly positive at $z = 0$ and will remain positive until z reaches a value such that one of the denominators reaches 0. Therefore γ increases monotonically for $z > 0$ and decreases monotonically for $z < 0$.

A.3 AN ANALYTICAL EXPLANATION FOR UNIMODAL vs. BIMODAL EFFECTS

We consider an analytical argument to support the numerical results that bimodal inputs generate larger deviations from PME model fits than unimodal inputs. As a metric, we consider $D_{\mathrm{KL}}(P, \tilde{P})$—where P and $\tilde{P}$ are again the true and model distributions, respectively—when we perturb an independent spiking distribution by adding a common, global input of variance c. To simplify notation, the small parameter in the calculation will be denoted $\epsilon = \sqrt{c}$.

We now compute $S(P)$ and $S(\tilde{P})$ (defined in an earlier Appendix) by deriving a series expansion for each set of event probabilities. We can compute the true distribution P using the expressions derived in Equation (18); to recap, let the common input I_c have probability density $p(I_c)$, and the independent input to each cell, x, have density $p_s(x)$. Let Θ be the threshold for generating a spike (i.e., a "1" response). For each cell, a spike is generated if $x + I_c > \Theta$, i.e., with probability

$$d(I_c) = \int_{\Theta - I_c}^{\infty} p_s(x) dx.$$

Given I_c, this is conditionally independent for each cell. We can therefore write our probabilities by integrating over I_c as follows:

$$\begin{aligned} p_0 &= \int_{-\infty}^{\infty} p(I_c)(1-d(I_c))^3\, dI_c \\ p_1 &= \int_{-\infty}^{\infty} p(I_c)d(I_c)(1-d(I_c))^2\, dI_c \\ p_2 &= \int_{-\infty}^{\infty} p(I_c)d(I_c)^2(1-d(I_c))\, dI_c \\ p_3 &= \int_{-\infty}^{\infty} p(I_c)d(I_c)^3\, dI_c \end{aligned} \tag{20}$$

We develop a perturbation argument in the limit of very weak common input. That is, $p(I_c)$ is close to a delta function centered at $I_c = 0$. Take $p(I_c)$ to be a scaled function

$$p(I_c) = \frac{1}{\epsilon} f\left(\frac{I_c}{\epsilon}\right) \tag{21}$$

We place no constraints on $f(x)$, other than that it must be normalized ($\mathbf{E}[1] = 1$) and that its moments must be finite (so that $\mathbf{E}[I_c]$, $\mathbf{E}[I_c^2]$, and so forth will exist, where $\mathbf{E}[g(x)] \equiv \int_{-\infty}^{\infty} g(x)f(x)\, dx$).

For the moment, assume that the function $f(x)$ has a single maximum at $x = 0$. To evaluate the integrals above, we Taylor-expand $d(x)$ around $x = 0$. Anticipating a sixth-order term to survive, we keep all terms up to this order. This gives, for small x,

$$d(x) \approx d(0) + \sum_{k=1}^{6} a_k x^k + O(x^7)$$

where $a_1 = p_s(\Theta)$ (the other coefficients a_2-a_6 can be given similarly in terms of the independent input distribution at Θ). Substituting this into the expressions for p_0, etc., above, with $p(I_c)$ given as in Equation (21), gives us each event as a series in ϵ; for example,

$$p_3 = d_0^3 + \left(3a_1 d_0^2\, \mathbf{E}[x]\right)\epsilon + \left((3a_1^2 d_0 + 3a_2 d_0^2)\, \mathbf{E}[x^2]\right)\epsilon^2 + \ldots,$$

where expectations are, again, with respect to the unscaled PDF $f(x)$. The entropy $S(P)$ is now given by using these series expansions in Equation (18).

We note that our derivation does not rely on the fact that the distribution of common input is peaked at $I_c = 0$ in particular. For example, we could have a common input centered around μ. The common input distribution function would be of the form

$$p(I_c) = \frac{1}{\epsilon} f\left(\frac{I_c - \mu}{\epsilon}\right)$$

Changing ϵ regulates the variance, but doesn't change the mean or the peak (assuming, without loss of generality, that the peak of f occurs at zero). The peak of $p(I_c)$ now occurs at μ, and the appropriate Taylor expansion of $d(x)$ is

$$d(x) \approx d(\mu) + \sum_{k=1}^{6} b_k (x-\mu)^k + O(x^7),$$

where the coefficients b_k now depend on the local behavior of d around μ. The expectations that appear in the expansion of p_3, and so forth, are now centered moments taken around μ; the calculations are otherwise identical. In other words, the perturbation expansion requires the *variance* of the common input to be small, but not the mean.

For bimodal inputs, we consider a common input with a probability distribution of the following form:

$$p(I_c) = \left(1-\epsilon^2\right)\frac{1}{\epsilon} f\left(\frac{I_c}{\epsilon}\right) + \epsilon^2 \frac{1}{\epsilon} f\left(\frac{I_c - 1}{\epsilon}\right)$$

so that most of the probability distribution is peaked at zero, but there is a second peak of higher order (here taken at $I_c = 1$, without loss of generality). Again, we approximate the integrals given in Equation (20), and therefore the entropy $S(P)$, by Taylor expanding $d(x)$;

$$\begin{aligned} d(x) &\approx d(0) + \sum_{k=1}^{6} a_k x^k + O(x^7); \quad (x \approx 0) \\ &\approx d(1) + \sum_{k=1}^{6} b_k (x-1)^k + O\left((x-1)^7\right); \quad (x \approx 1) \end{aligned}$$

around the two peaks 0 and 1, respectively. For each integral we have the same contributions from the unimodal case, multiplied by $(1-\epsilon^2)$, as well as the corresponding contributions from the second peak multiplied by ϵ^2 (these weightings are chosen so that the common input has variance of order ϵ^2, as in the unimodal case). This makes clear at what order every term enters.

We now construct an expansion for the PME model $\tilde{P}$:

$$\begin{aligned} \tilde{P}(x_1, x_2, x_3) = \frac{1}{Z} \exp\big(&\lambda_1 (x_1 + x_2 + x_3) \\ &+ \lambda_2 (x_1 x_2 + x_2 x_3 + x_1 x_3)\big) \end{aligned}$$

We approach this problem by describing λ_1 and λ_2 as a series in ϵ. We match coefficients by forcing the first and second moments of $\tilde{P}$ to match those of P—as they must. Specifically, take

$$\lambda_1 = \tilde{\lambda} + \sum_{k=1}^{6} \epsilon^k u_k + O\left(\epsilon^7\right)$$

$$\lambda_2 = \sum_{k=1}^{6} \epsilon^k v_k + O\left(\epsilon^7\right)$$

where $\lambda_1 = \tilde{\lambda}$, $\lambda_2 = 0$ are the corresponding parameters from the independent case. The events $\tilde{p}_0$, $\tilde{p}_1$, $\tilde{p}_2$, and $\tilde{p}_3$ can be written as a series in ϵ. We then require that the mean and centered second

moments of $\tilde{P}$ match those of P; that is

$$p_1 + 2p_2 + p_3 = \tilde{p}_1 + 2\tilde{p}_2 + \tilde{p}_3$$
$$p_2 + p_3 - \left(p_1 + 2p_2 + p_3\right)^2 = \tilde{p}_2 + \tilde{p}_3 - \left(\tilde{p}_1 + 2\tilde{p}_2 + \tilde{p}_3\right)^2.$$

At each order k, this yields a system of two linear equations in u_k and v_k; we solve, inductively, up to the desired order; we now have $\tilde{P}$, and therefore $S(\tilde{P})$, as a series in ϵ.

Finally, we combine the two series to find that in the *unimodal* case,

$$D_{\mathrm{KL}}\left(P, \tilde{P}\right) = S\left(\tilde{P}\right) - S(P) = \epsilon^6 \left[\frac{a_1^6 \left(2\,\mathbf{E}[x]^3 - 3\,\mathbf{E}[x]\,\mathbf{E}[x^2] + \mathbf{E}[x^3]\right)^2}{2\left(1-d_0\right)^3 d_0^3}\right] + O\left(\epsilon^7\right) \quad (22)$$

If the first two odd moments of the distribution are zero (something we can expect for "symmetric" distributions, such as a Gaussian), then this sixth-order term is zero as well.

For the *bimodal* case

$$D_{\mathrm{KL}}\left(P, \tilde{P}\right) = S\left(\tilde{P}\right) - S(P) = \epsilon^4 \left[\frac{\left(d_1 - d_0\right)^6}{2\left(1-d_0\right)^3 d_0^3}\right] + O\left(\epsilon^5\right)$$

This last term depends on the distance $d_1 - d_0$, in other words, how much more likely the independent input is to push the cell over threshold when common input is "ON". We can also view this as depending on the ratio $\frac{d_1 - d_0}{1-d_0}$, which gives the fraction of previously non-spiking cells that now spike as a result of the common input.

The main point here, of course, is that $D_{KL}(P, \tilde{P})$ is of order ϵ^4 rather than ϵ^6. So, as the strength of a common binary vs. unimodal input increases, spiking distributions depart from the PME more rapidly.

frontiers in
COMPUTATIONAL NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 30 January 2014
doi: 10.3389/fncom.2014.00002

Long-term plasticity determines the postsynaptic response to correlated afferents with multivesicular short-term synaptic depression

Alex D. Bird*[1,2,3] *and* ***Magnus J. E. Richardson*[1]**

[1] *Warwick Systems Biology Centre, University of Warwick, Coventry, UK*
[2] *Warwick Systems Biology Doctoral Training Centre, University of Warwick, Coventry, UK*
[3] *School of Life Sciences, University of Warwick, Coventry, UK*

Edited by:
Tatjana Tchumatchenko, Max Planck Institute for Brain Research, Germany

Reviewed by:
Jean-Pascal Pfister, Cambridge University, UK (in collaboration with Simone Surace)
Michael Graupner, New York University, USA

****Correspondence:***
Alex D. Bird, Warwick Systems Biology Centre, Senate House, University of Warwick, CV4 7AL, Coventry, UK
e-mail: a.d.bird@warwick.ac.uk

Synchrony in a presynaptic population leads to correlations in vesicle occupancy at the active sites for neurotransmitter release. The number of independent release sites per presynaptic neuron, a synaptic parameter recently shown to be modified during long-term plasticity, will modulate these correlations and therefore have a significant effect on the firing rate of the postsynaptic neuron. To understand how correlations from synaptic dynamics and from presynaptic synchrony shape the postsynaptic response, we study a model of multiple release site short-term plasticity and derive exact results for the crosscorrelation function of vesicle occupancy and neurotransmitter release, as well as the postsynaptic voltage variance. Using approximate forms for the postsynaptic firing rate in the limits of low and high correlations, we demonstrate that short-term depression leads to a maximum response for an intermediate number of presynaptic release sites, and that this leads to a tuning-curve response peaked at an optimal presynaptic synchrony set by the number of neurotransmitter release sites per presynaptic neuron. These effects arise because, above a certain level of correlation, activity in the presynaptic population is overly strong resulting in wastage of the pool of releasable neurotransmitter. As the nervous system operates under constraints of efficient metabolism it is likely that this phenomenon provides an activity-dependent constraint on network architecture.

Keywords: long-term plasticity, short-term plasticity, synaptic depression, correlations and synchrony, voltage fluctuations

1. INTRODUCTION

Synapses play a key role in transmitting and processing information throughout the nervous system and long-term shifts in synaptic efficacy are believed to underpin learning and memory (Hebb, 2002; Markram et al., 2011). Synapses function through release of neurotransmitters that then bind to receptors on the postsynaptic cell and transiently alter the membrane conductance. Neurotransmitters in the presynaptic terminal are stored and transported in vesicles (Fox, 1988; Hu et al., 2008). A number of vesicles are positioned at active sites where they have a certain probability of being released when the presynaptic cell spikes. Empty release sites are restocked after a variable period, with an overall rate of a few Hz (Südhof, 2004). Both the number of contacts per presynaptic cell and the activity in the presynaptic network can generate correlations in the release of neurotransmitter at synapses onto a single neuron; we demonstrate that postsynaptic activity is governed by a balance between these two sources of correlation.

The usage of vesicles due to presynaptic firing and stochastic replenishment means that the number of vesicles available for release is a highly dynamic quantity that is dependent on the history of afferent activity. In the immature cortex, the relatively high release probability and limited availability of vesicles causes a progressive reduction in synaptic efficacy during a period of sustained neuronal activity (Reyes and Sakmann, 1999; Chen and Buonomano, 2012). This short-term reduction in synaptic strength is known as vesicle depletion depression: an unstocked active site cannot induce a postsynaptic response to any incident action potential (Abbot, 1997; Tsodyks and Markram, 1997; Zucker and Regehr, 2002). The phenomenon is believed to play a role in gain control (Abbot, 1997; Abbott and Regehr, 2004; Rothman et al., 2009), information transmission (Zador, 1998; Kilpatrick, 2012; Scott et al., 2012), and adaptation to sensory stimuli (Furukawa et al., 1982; Hallermann and Silver, 2012). The synaptic plasticity models introduced by Abbot (1997) and Tsodyks et al. (1998) capture short-term depression accurately; they match empirical data and allow a richness of network behavior (Tsodyks et al., 1998) to emerge beyond that predicted by static synapses. Such models consider the mean efficacy of the synapse, averaged across several presentations of the same presynaptic stimulus; the predicted postsynaptic response therefore varies continuously. Several recent studies have considered a quantal model of synaptic function incorporating short-term depression, with probabilistic vesicle release and replacement to reflect trial-to-trial variability (Fuhrmann et al., 2002; de la Rocha and Parga, 2005; Rosenbaum et al., 2012). The impact of stochastic vesicle dynamics is particularly marked when mean synaptic drive is insufficient to bring the postsynaptic neuron to threshold

and spiking activity is governed by fluctuations in the system (Gerstein and Mandelbrot, 1964; Kuhn, 2004). To induce postsynaptic firing in such a system it is necessary for the variable synaptic drive to exhibit coincidences; this occurs most regularly when that drive is correlated.

Correlations in neurotransmitter release between different sites can arise from two sources: from multiple contacts onto a postsynaptic neuron from the same presynaptic cell and from synchronous activity across the presynaptic population. The number of sites between a pair of neurons is fixed over short timescales, unlike the number of vesicles ready to release from the sites, but can vary widely over longer periods (Loebel et al., 2013) following potentiation or depression. Connections between neurons potentiate and depress in the long term chiefly through changes in this synaptic parameter—the number of independent release sites can be seen as a fundamental unit of memory. Synchronous firing in the presynaptic population emerges from the connectivity of neuronal networks (Aertsen et al., 1989) and has relevance for encoding sensory information (von der Malsburg, 1981; deCharms and Merzenich, 1996; Averbeck et al., 2006), motor control (Baker et al., 2001; Capaday, 2013) and decision making (Cohen and Newsome, 2008; Cain and Shea-Brown, 2013). Recent work suggests that modulation of correlations can be more significant for neuronal coding than alterations in the presynaptic firing rate (Seriès et al., 2004; Mitchell et al., 2009; Cohen and Kohn, 2011). Population synchronization is a transient phenomenon relative to the structural changes underlying long-term plasticity.

A detailed stochastic model of neurotransmitter dynamics at the presynaptic terminal is required to analyze the effects of presynaptic synchrony, particularly when long-term plasticity varies the structure of synapses through altering the number of release sites. It can be noted that multiple contacts between cells and transient correlations within a presynaptic population are likely to introduce considerable redundancy in the usage of vesicles: correlated events may lead to EPSPs many times larger than that required to reach threshold. However, evidence points to the nervous system operating under constraints of efficient metabolism (Levy and Baxter, 2002; Taschenberger et al., 2002; Savtchenko et al., 2012) suggesting such wastage would not commonly arise *in vivo*. It is therefore of interest to examine the effect on the postsynaptic cell of the interaction of partially synchronized afferent drive with multiple contacts per presynaptic cell. To this end, we analyze a model of a postsynaptic cell receiving input from a population of release sites distributed between different numbers of presynaptic neurons and with different levels of synchrony.

Following the basic model definitions, we first derive exact forms for the crosscorrelations of vesicle occupancies and release at multiple contacts from the same and different presynaptic cells. These correlations were previously derived by Rosenbaum et al. (2012) using a diffusion and additive-noise approximation, and our results show that this earlier method gave exact results for these quantities. We then go on to calculate the exact voltage mean and variance and, through comparison with the typical EPSP amplitude, argue that synaptic noise can become significantly non-Gaussian. We then derive two approximate limiting forms for the firing rate for low and high correlations and demonstrate that the postsynaptic response is optimal at intermediate levels of afferent correlations. We finally show that this effect is robust for neurons in which there is some level of synaptic homeostasis or soft limit on the total number of release sites.

2. METHODS

We consider a population of N presynaptic neurons synapsing onto a single postsynaptic neuron. A presynaptic neuron makes synapses with n vesicle occupancy sites from each of which neurotransmitter may be independently released with a probability p on the arrival of a presynaptic action potential, occurring at a constant Poissonian rate R_a. In between presynaptic action potentials, empty release sites are restocked independently at a constant Poissonain rate R_r. Initially, we consider that the total number of release sites onto the postsynaptic cell is fixed at $M = nN$ (example configurations are provided in **Figures 1A–C**). The number of independent release sites n was recently shown (Loebel et al., 2013) to be the synaptic parameter most closely correlated with the structural changes arising from long-term plasticity and so we will consider the effects of varying n (while initially keeping M constant) on the postsynaptic response. The binary variable x will be used to signify vesicle release-site occupancy: $x = 1$ if present or $x = 0$ if absent. The evolution of vesicle occupancy is given by the stochastic differential equation

$$\frac{dx}{dt} = (1 - x)\sum_m \delta(t - t_m) - \sum_k \varrho_k(x)\delta(t - t_k) \qquad (1)$$

where m counts the restock events occurring at a rate R_r and k counts the presynaptic action potentials occurring at a rate R_a. The binary random variable $\varrho_k(x)$ signifies whether a release was successful at the kth action potential: if $x = 1$ then $\varrho_k(x) = 1$ with probability p to model a successful release of neurotransmitter, and is 0 otherwise to model a failed release from a stocked site; if $x = 0$ then no release is possible and $\varrho_k(x) = 0$. The δs are

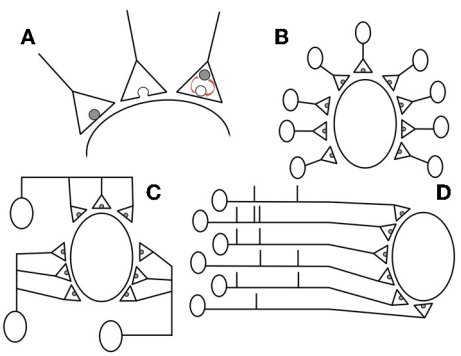


FIGURE 1 | We consider a population of *N* presynaptic neurons each featuring *n* independent release sites onto a single postsynaptic cell. **(A)** The stochastic dynamics are illustrated from left to right: if a vesicle is present it is released (with probability p) when an action potential arrives (Poissonian rate R_a); an empty release site; and restock of an empty release site (Poissonian rate R_r). **(B,C)** examples with $M = nN = 9$ with **(B)** $n = 1$, $N = 9$ and **(C)** $n = 3$, $N = 3$ contacts and presynaptic neurons, respectively. **(D)** Example spike trains for $M = N = 6$ correlated presynaptic neurons that feature $S = 3$ synchronous spikes.

Dirac delta functions and whenever a delta function multiplies a dynamic variable it is assumed that the value of the variable used is that immediately before the delta event occurs. In other words, the equations are non-anticipating and should be interpreted in an Itō sense (Gardiner, 2010).

2.1. CORRELATIONS FROM STRUCTURE

When a presynaptic neuron spikes, available vesicles at each of the n sites release their contents independently with probability p, and so the total number of release events is binomially distributed. Note that because these sites receive the same incoming action potentials correlations will arise despite the independent conditional release and restock events at each site. Globally, we first hold the total number of release sites, given by $M = nN$, constant so that the postsynaptic neuron receives a fixed overall excitatory drive. In this study we set $M = 5000$, which is of-the-order-of estimates by O'Kusky and Colonnier (1982), Megías et al. (2001), and Spruston (2008). This has the effect of maintaining the overall level of excitatory drive to the postsynaptic cell and in biological terms can be seen as a constraint of metabolic efficiency across the presynaptic population: as some contacts potentiate, others die out. The effects of relaxing this condition are discussed later. Recent analysis of long-term plasticity data has shown that changes in EPSP amplitude are captured by models in which the number of independent release sites n increases or decreases. Depending on the protocol, n can potentiate or depress by a factor of 5 or more (Loebel et al., 2013); a typical range for n is 5–50. However, contacts with a binomial n as low as 1 or as high as 100 sites have also been observed. Though the upper bound is unbiological, for completeness we vary n between 1 and 5000 in simulations.

2.2. CORRELATIONS FROM PRESYNAPTIC SYNCHRONY

The population of neurons driving a common target often displays substantial synchrony in spiking activity (Salinas and Sejnowski, 2000; Averbeck et al., 2006; Cohen and Kohn, 2011) (see **Figure 1D**). Here we model correlations in the presynaptic population by using a variation of the Multiple Interaction Process (MIP) introduced in Kuhn et al. (2003). We implement the process by considering a master spike train with a constant Poissonian rate NR_a/S. For each spike in the master train we pick S of the presynaptic neurons at random and assign a synchronous spike in their trains. If $S = 1$ this would imply no correlations in the presynaptic population and $S = N$ would be a fully synchronous presynaptic population. Note that the spiking of each presynaptic neuron is Poissonian at rate R_a as required and also that, given that one presynaptic neuron spikes, the probability that a particular other presynaptic neuron has a spike at the same time is $c = (S-1)/(N-1)$. In reality, shared spikes will not be entirely synchronous and so in later simulations (specifically, those leading to **Figures 6**, **7**) we add independent, normally distributed jitter to the spike times with mean 0 and standard deviation τ_j following de la Rocha and Parga (2005) and Cohen and Kohn (2011). Note that in **Figures 5**, **6A,B**, **7** the curves are truncated for increasing n because, for fixed S and fixed $M = nN$, it is invalid to have S greater than N. This is also the case for **Figures 6B,C** with increasing S.

2.3. POSTSYNAPTIC VOLTAGE

We treat the postsynaptic neuron as a leaky integrate-and-fire model with each neurotransmitter release event causing the voltage to jump by an amount a. The membrane voltage V has a resting value E and a spike threshold V_{th}. After a spike, V is reset to E and held there for a time τ_r to model the refractory period. If N presynaptic neurons each have n neurotransmitter release sites then the postsynaptic voltage is governed by

$$\tau \frac{dV}{dt} = E - V + a\tau \sum_{i=1}^{N} \sum_{j=1}^{n} \sum_{k} \varrho_k^{ij}(x_{ij}) \delta(t - t_k^i) \qquad (2)$$

where τ is the membrane time constant, x_{ij} is the occupancy variable for the ith presynaptic neuron's release site number j and k labels the order of incoming action potentials to release site with occupancy x_{ij}. Note that the spike times t_k^i are identical for all release sites with the same presynaptic neuron i and that some of the spike times will be common to release sites with distinct presynaptic neurons, depending on the level of synchrony given by the correlated MIP process parameterized by S. The values of other parameters used in simulations (unless otherwise stated) are given in (**Table 1**).

3. RESULTS

We first derive exact forms for the crosscorrelations of vesicle-occupancy and of neurotransmitter-release time series. The latter can then be used to calculate the exact membrane voltage variance. Two approximations of the postsynaptic firing rate then lead us to the main result of the paper: that long-term synaptic plasticity—through its alternation of the synaptic parameter n—sets the optimal postsynaptic response to a presynaptic population with correlated firing. Throughout this section the notation $\langle \phi \rangle$ denotes the steady-state expectation of the fluctuating quantity ϕ.

For the calculation of the crosscorrelations of objects separated by a time T, it is useful to consider how the steady-state expectation of the product of the occupancy x with some quantity ψ evaluated at an earlier time evolves with the separation time:

$$\frac{d}{dT} \langle x(T)\psi(0) \rangle = \langle (1 - x(T))\psi(0) \rangle R_r - \langle x(T)\psi(0) \rangle pR_a \qquad (3)$$

where the first term on the right-hand side is the rate that an empty site is filled and the second term is the rate that a full site releases its contents. This equation can be rearranged into the form

$$\tau_x \frac{d}{dT} \langle x(T)\psi(0) \rangle = \langle x \rangle \langle \psi \rangle - \langle x(T)\psi(0) \rangle \qquad (4)$$

where the time constant τ_x and steady-state occupancy $\langle x \rangle$ are

$$\tau_x = \frac{1}{R_r + pR_a} \text{ and } \langle x \rangle = \frac{R_r}{R_r + pR_a}. \qquad (5)$$

That the second quantity must be the steady-state occupancy $\langle x \rangle$ can be inferred by noting that in the limit $T \to \infty$ the expectation

Table 1 | Typical parameters used for the figures.

Parameter	Interpretation	Value
V	Postsynaptic membrane voltage	Varies
S	Number of presynaptic cells that fire together	Varies
n	Number of release sites per presynaptic neuron	Varies
N	Number of presynaptic neurons	Varies
M	Total number of vesicle release sites (nN)	5000
R_r	Rate at which empty vesicles are replaced at release sites	2 Hz
R_a	Rate of presynaptic spiking	2 Hz
p	Probability of spike arrival inducing neurotransmitter release at a site with a vesicle present	0.66
τ_j	Jitter standard deviation timescale	2 ms
E	Resting membrane voltage	−70 mV
V_{th}	Threshold at which action potentials are initiated	−55 mV
τ_r	Refractory period of a neuron after a spike	2 ms
τ	Membrane time constant	10 ms
a	EPSP amplitude induced by neurotransmitter released from a single vesicle	0.2 mV

$\langle x(T)\psi(0)\rangle$ in Equation (3) loses its T dependence and factorises into the product $\langle x\rangle\langle\psi\rangle$. Note that the exponential solution to the differential Equation (4) implies that all crosscorrelations that include the occupancy x take a simple exponential form

$$\text{Crosscorr}(x,\psi) = (\langle x\psi\rangle - \langle x\rangle\langle\psi\rangle)e^{-t/\tau_x} \tag{6}$$

where $\langle x\psi\rangle$ is the expectation evaluated in the limit $T \to 0$.

3.1. VESICLE OCCUPANCY CROSSCORRELATIONS

The autocorrelation of release-site occupancy can be calculated by making use of the fact that for the binary variable x we have $x^2 = x$ and so $\langle x^2\rangle = \langle x\rangle$. Putting $\psi = x$ in equation (6) gives

$$\text{Autocorr}(x) = \langle x\rangle(1-\langle x\rangle)e^{-|T|/\tau_x} = \frac{pR_aR_r}{(R_r+pR_a)^2}e^{-|T|/\tau_x} \tag{7}$$

where the extension of the exponential to negative times comes from a symmetry argument. For the crosscorrelation between different release sites, with occupancy variables x and x', we need to distinguish between cases where the release sites either share the same presynaptic neuron or have different presynaptic neurons when deriving $\langle xx'\rangle$. However, the derivation can be written in the same form by introducing a quantity γ that is the proportion of shared spikes: $\gamma = 1$ for release sites with the same presynaptic neuron or $\gamma = c = (S-1)/(N-1)$ for different presynaptic neurons. A steady-state equation for the zero-time expectation $\langle xx'\rangle$ can be found by considering the state where both sites are occupied and balancing the total rates into and out of this state

$$\langle x(1-x')\rangle R_r + \langle (1-x)x'\rangle R_r = \langle xx'\rangle(2R_ap - \gamma R_ap^2). \tag{8}$$

The terms on the left-hand side represent the total rate into the double occupancy state, whereas the terms on the right-hand side multiplying the expectation are the combined rates of individual vesicle release minus the coincidence term to prevent overcounting of events. We now combine terms to obtain the required expectation

$$\langle xx'\rangle_\gamma = \frac{2R_r\langle x\rangle}{2R_r + R_ap(2-\gamma p)} \tag{9}$$

where the γ subscript will be used later to distinguish the different cases. It can be inserted into Equation (6) with $\psi = x'$ to give

$$\text{Crosscorr}(x,x') = \frac{\gamma p^2 R_aR_r^2 e^{-|T|/\tau_x}}{(2R_r + pR_a(2-p\gamma))(R_r+pR_a)^2}. \tag{10}$$

Example plots of Equation (7), and Equation (10) for cases with $\gamma = 1$ and $\gamma = c$ are given in **Figures 2A,C,E**. It is interesting to note that our exact results are identical to those previously calculated in Rosenbaum et al. (2012) using a combined diffusion and additive-noise approximation, validating their method up to second-order statistics.

3.2. NEUROTRANSMITTER RELEASE CROSSCORRELATIONS

Though synchrony in the presynaptic population leads to positive correlations for release-site occupancy, we now show that the delayed restock following release leads to negative cross-correlations in the release events themselves. Let $\chi(t)$ and $\chi'(t)$ be trains of delta pulses representing neurotransmitter release from sites with occupancies defined by $x(t)$ and $x'(t)$, respectively, so that:

$$\chi(t) = \sum_k \varrho_k(x)\delta(t-t_k) \tag{11}$$

where k counts incoming action potentials at the contact with site occupancy x. In the steady state we have $\langle\chi\rangle = pR_a\langle x\rangle$ because the rate of release is equal to the release rate pR_a given vesicle occupancy multiplied by the occupancy probability $\langle x\rangle$. The auto and crosscorrelations can be straightforwardly calculated using the general result of Equation (6) by setting $\psi = \chi'$ and noting that $\langle\chi(T)\chi'(0)\rangle = pR_a\langle x(T)\chi'(0)\rangle$. However, some care needs to be taken when considering the case $T = 0$. The result of Equation (6) is valid in the limit $T \to 0$; but there is an additional delta function in the crosscorrelation when $T = 0$ with an amplitude equal to the rate of simultaneous events in χ and χ' that arises from the delta functions in Equation (11). The autocorrelation

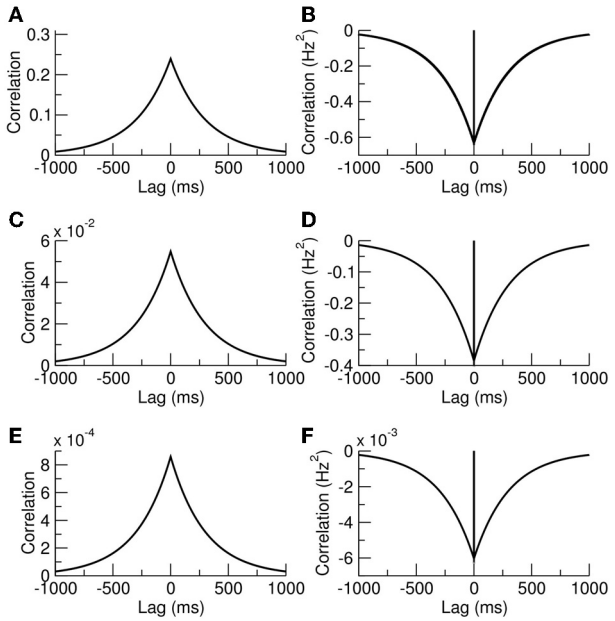


FIGURE 2 | Release-site occupancy is correlated, neurotransmitter-release events are anticorrelated. (A) Autocorrelation of a release-site occupancy and **(B)** autocorrelation in neurotransmitter release. **(C,D)** Crosscorrelations for distinct release sites sharing the same presynaptic cell. **(E,F)** Crosscorrelations for release sites with different presynaptic cells. The parameters were $N = 500$, $n = 10$, and $S = 10$ giving the probability of synchronous spikes $c = 0.018$.

function for χ therefore takes the form

$$\text{Autocorr}(\chi) = pR_a \langle x \rangle \delta(T) - (pR_a \langle x \rangle)^2 e^{-|T|/\tau_x} \quad (12)$$

where the rate of simultaneous events for the autocorrelation is just the mean release rate $pR_a \langle x \rangle$ and prefactor of the exponential is only $-\langle \chi \rangle^2$ because in the limit $T \to 0$ the expectation of $\langle \chi(T)\chi(0) \rangle$ is zero as there is no time for a restock. A similar consideration gives the result for the crosscorrelation

$$\text{Crosscorr}(\chi, \chi') = \gamma p^2 R_a \langle xx' \rangle_\gamma \delta(T) + R_a^2 p^2 ((1 - \gamma p) \langle xx' \rangle_\gamma - \langle x \rangle^2) e^{-|T|/\tau_x} \quad (13)$$

where we are treating cases for which the release is from distinct contacts sharing the same presynaptic neuron $\gamma = 1$ or from distinct presynaptic neurons where $\gamma = c$. In Equation (13) the prefactor of the delta function arises from the rate of simultaneous releases, which is equal to the arrival of simultaneous spikes γR_a multiplied by the probability that each contact releases a vesicle $p^2 \langle xx' \rangle_\gamma$. The prefactor of the exponential shares the same squared component $-\langle \chi \rangle^2 = -(pR_a \langle x \rangle)^2$ as the autocorrelation, but also has a non-zero contribution from $\langle \chi(T)\chi'(0) \rangle$ in the limit $T \to 0$. This quantity is equal to the probability that both sites are occupied $\langle xx' \rangle_\gamma$ multiplied by the probability of a release from site x' but no release from site x from a simultaneous presynaptic event, which is $R_a p(1 - \gamma p)$ multiplied by a subsequent release from site x just afterwards due to a second presynaptic spike, pR_a. This exact result is again identical to that derived previously using a diffusion and additive-noise approximation (Rosenbaum et al., 2012). Example autocorrelation and crosscorrelation functions are plotted in **Figures 2B,D,F**.

3.3. MEMBRANE VOLTAGE MEAN AND VARIANCE

The tonic component of the presynaptic drive can be characterized by the mean voltage, which is straightforward to calculate in the absence of a threshold. The dynamics of this quantity can be found by taking the expectation of Equation (2) to yield the steady-state result

$$\langle V \rangle = E + aM\tau pR_a \langle x \rangle = E + \frac{aM\tau pR_a R_r}{R_r + pR_a}. \quad (14)$$

Note that the mean voltage is independent of the synchrony S and is also independent of release-site number n when $M = nN$ is held fixed.

The effect of correlated synaptic fluctuations on the postsynaptic neuron can also be characterized by deriving the steady-state variance of the postsynaptic voltage (again in the absence of a threshold-reset mechanism). This quantity is derived in the Appendix using the auto and crosscorrelations of χ (Equations 12, 13) and takes the form

$$\text{Var}(V) = \frac{a^2 \tau Nnp R_a}{2} \left(\langle x \rangle + (n-1)p \langle xx' \rangle_1 + (N-1)ncp \langle xx' \rangle_c \right) + \frac{Nn(a\tau pR_a)^2}{1 + \tau R_r + p\tau R_a} \left((n-1)(1-p) \langle xx' \rangle_1 + (N-1)n(1-cp) \langle xx' \rangle_c - Nn \langle x \rangle^2 \right). \quad (15)$$

The first term arises from the δ-functions in Equations (12, 13) and the second term comes from the negative correlations in vesicle release due to short-term depression (the terms featuring exponentials in the same equations). For a related model (de la Rocha and Parga, 2005) it was demonstrated that on increasing the presynaptic rate a maximum can be seen in the conductance fluctuations. The exact result of Equation (15) allows for this effect of fluctuations in depressing synapses on the voltage itself to be analyzed. Example variances as a function of presynaptic rate are shown in **Figure 3** and, as expected from the previous analysis of conductance fluctuations (de la Rocha and Parga, 2005), the variance also shows a maximum at intermediate presynaptic rates.

Though the voltage variance measures one aspect of presynaptic fluctuations, it misses its increasing shot-noise nature as the correlations increase. Shot noise causes a non-Gaussian component in the tails of the membrane voltage distribution that, because they extend to the region of action-potential initiation, can significantly affect the post-synaptic firing rate (Richardson and Swarbrick, 2010). The mean EPSP amplitude can be used to see this effect: it is proportional to the mean of the vesicles released by a spike given the occupancy levels already computed, and so

$$\langle \text{EPSP} \rangle = apnS \langle x \rangle = \frac{apSnR_r}{R_r + pR_a}. \quad (16)$$

As correlations from increasing n or S become stronger, the mean EPSP amplitude increases. However, as noted above, the mean voltage (Equation 14) does not change under increasing n or S. Taken together, the implications are that in the limit of high correlations the synaptic drive becomes temporally sparse with large amplitude EPSPs generated from correlated events. This effect can be seen in simulations of the model with different parameter regimes (**Figure 4**). For parameters $N = 125$, $n = 1$, and $S = 1$ (no presynaptic synchrony) the presynaptic spikes (**Figure 4A**) and neurotransmitter release (**Figure 4D**) are uncorrelated, and in the full system with $M = 5000$ the EPSPs are relatively small (**Figures 4G,H**) and the resulting voltage distribution is close to Gaussian (**Figure 4I**). Increasing n (**Figure 4B**) or S (**Figure 4C**) to 25 leads to correlations in neurotransmitter release (**Figures 4E,F**), larger EPSPs (**Figures 4J,K,M,N**) and a more variable and skewed membrane voltage (**Figures 4L,O**). Note the right-hand tails from the skewed membrane voltages under conditions of presynaptic correlation that extend toward voltages where action potentials would be initiated.

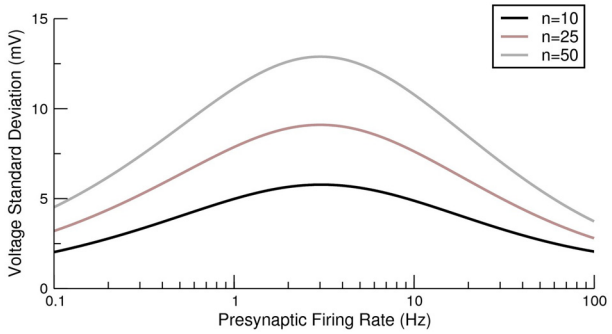


FIGURE 3 | Exact voltage variance for a postsynaptic neuron receiving multiple depressing synaptic contacts from a presynaptic population. Three examples are given with different numbers of neurotransmitter release sites per presynaptic neuron. For each case the synchrony was $S = 10$.

3.4. RELEASE SITE NUMBER AND POSTSYNAPTIC RATE

As the analyses of the previous section and examples in **Figure 4** demonstrate, for the case of few release sites and low synchrony the voltage distribution is close to Gaussian. However, for the case of many release sites the synchronous release events generate large EPSPs that are reminiscent of shot noise. With this in mind, approximations for the firing of the postsynaptic cell may be found for the cases of low n, when the voltage distribution is roughly Gaussian, and high n for which the EPSP amplitudes are of-the-order-of or larger than threshold.

3.4.1. Few release sites

For the low n approximation we rely on a recent observation (Alijani and Richardson, 2011) that the firing rate of integrate-and-fire neurons is relatively insensitive to temporal correlations as long as the subthreshold voltage mean and variance are matched. To this end we approximate the firing rate of the neuron by a white-noise equivalent that has a voltage mean μ equal to that

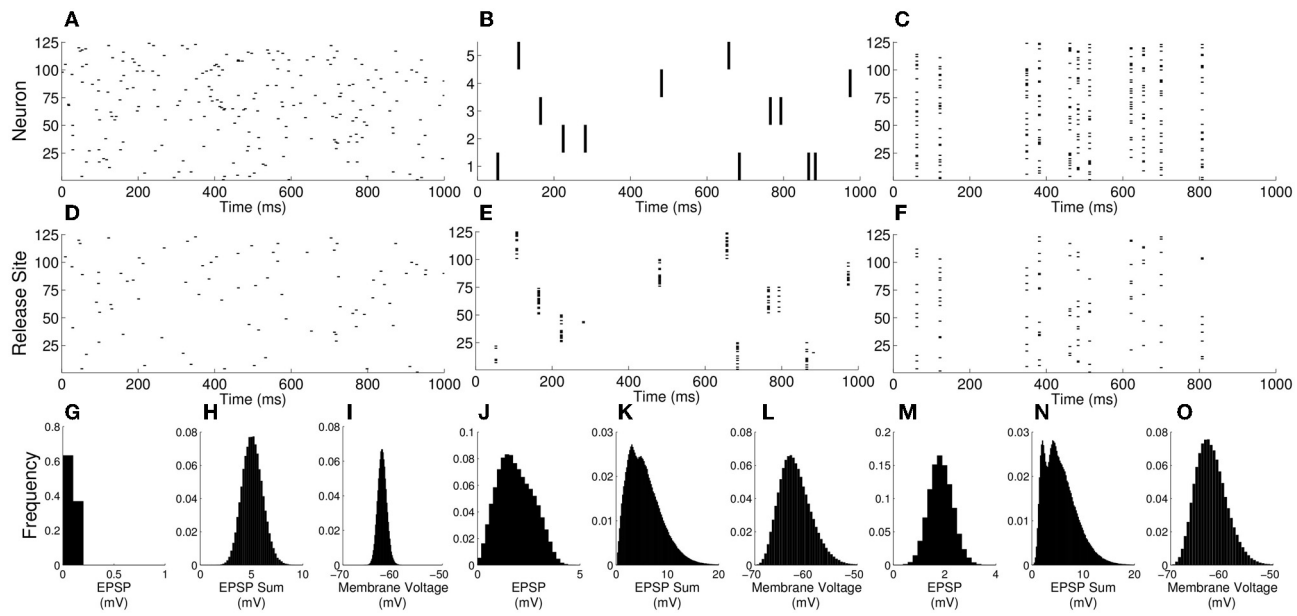


FIGURE 4 | Membrane voltage distributions become markedly non-Gaussian as correlations increase. (A–C) Rasters of presynaptic firing with: **(A)** $N = 125$, $n = 1$, and $S = 1$; **(B)** $N = 5$, $n = 25$, and $S = 1$; **(C)** $N = 125$, $n = 1$, and $S = 25$. **(D–F)** Rasters of neurotransmitter release for these firing patterns. **(G,J,M)** EPSP histograms for the above n and S values, but with N adjusted so that $M = nN = 5000$. **(H,K,N)** Histograms of the total synaptic drive over an interval of the membrane time constant for the same parameters. **(I,L,O)** Voltage histograms for the same parameters. Note that, whereas the voltage is close to Gaussian for the single release-site and no-presynaptic-synchrony case, it develops a tail to the right when correlations arise either from multiple release sites or presynaptic synchrony.

of Equation (14) and variance σ^2 equal to that of Equation (15). The firing rate of a leaky-integrate-and-fire neuron with these parameters is given (Brunel and Hakim, 1999) by the reciprocal of

$$\tau \int_0^\infty \frac{dz}{z} e^{-z^2/2} \left(e^{zz_{th}} - e^{zz_{re}} \right) \tag{17}$$

where $z_{th} = (V_{th} - E - \mu)/\sigma$ and in this case $z_{re} = -\mu/\sigma$.

3.4.2. Many release sites

For sufficiently large n the mean EPSPs are greater than that required to bring the neuron to threshold $apnS\langle x \rangle \gg V_{th} - E$, and so each synchronous presynaptic event is likely to cause the postsynaptic cell to spike. The postsynaptic cell receives input at a total rate of NR_a/S and so we can approximate the rate in the large n case by

$$r \sim \frac{NR_a}{S} = \frac{MR_a}{nS}. \tag{18}$$

Therefore, increasing the presynaptic synchrony S will reduce the postsynaptic response when n is large.

3.4.3. Optimal release-site number

Under conditions of a fixed number of release sites onto the postsynaptic cell $M = nN$, increasing n has no effect on the voltage mean (Equation 14), but increases the voltage variance (Equation 15). Therefore, as n increases from an initially small value, the approximation given by Equation (17) predicts that the postsynaptic cell will fire at an increasing rate. However, from Equation (18), which is valid for high n, we see that the postsynaptic firing rate decreases as n increases. Hence, there must be an intermediate n for which the response of the postsynaptic cell is optimized. This effect can be clearly seen in the examples given in **Figure 5** in which the postsynaptic rate is plotted as a function of n for fixed M. The intersections of the two approximations for each curve provide an estimate for the optimal n, which decreases as the presynaptic synchrony increases. It should be noted that this effect, which has a maximum as a function of release-site number at constant presynaptic rate, is a distinct phenomenon to the tuning curve as a function of presynaptic rate analyzed in de la Rocha and Parga (2005).

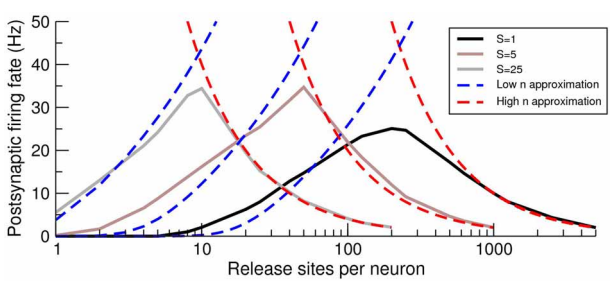


FIGURE 5 | The postsynaptic firing rate exhibits a maximum as a function of the number of pre-to-post release sites *n*. Firing-rate simulations (solid lines), low n approximation (Equation 17; blue dashed lines) and high n approximation (Equation 18; red-dashed lines) for various levels of presynaptic synchrony S as a function of the number of release sites n per presynaptic cell. The maximal postsynaptic response is close to the intersection of the approximate forms and the optimum n decreases with increasing synchrony S. Note that the curves are limited on their right because of the restriction $S \leq N$ (the maximal allowable synchrony is equal to the number of presynaptic neurons) so that the maximum n is $n = M/S$. This upper bound on n holds for similar curves in later figures.

3.5. LONG-TERM PLASTICITY AND RESPONSE TO SYNCHRONY

The post-synaptic firing rate is sensitive to correlations arising from multiple release sites, as discussed above, as well as to presynaptic synchrony (de la Rocha and Parga, 2005). In particular, the firing rate has a maximal response at an optimal n that is a function of the presynaptic synchrony as can be seen in **Figure 6**. When neurotransmitter release is too strongly correlated in the presynaptic population, the postsynaptic response weakens because the quantity of neurotransmitter released is in excess of that necessary to take the postsynaptic cell to threshold and therefore this limited resource is wasted. The reduction in response to over-strong correlations gives rise to the optimal responses in the space of n and S seen in **Figures 6A–C**. Note that the band of optimal postsynaptic response is linear with negative gradient in the n, S log–log plot and so the optimal synchrony in the presynaptic population has an inverse relation to the number of release sites n each presynaptic cell makes onto the postsynaptic target.

Analyses of long-term plasticity data (over a 12 h period) by Loebel et al. (2013) demonstrated that connections between thick-tufted layer-5 pyramidal cells in the rat somatosensory cortex alter their efficacy by changing the binomial parameter n, in preference to probability of release or quantal amplitude. Among the experiments analyzed certain connections potentiated four-fold, from an effective binomial n of ∼25 to ∼100. Assuming that the mean excitatory drive remains constant, this potentiation would lead to the postsynaptic cell becoming maximally responsive to signals encoded by weaker presynaptic synchrony (see **Figure 6C**). It would also cease to amplify strongly correlated stimuli as effectively. Other connections showed four-fold reductions in n from ∼40 to ∼10 under protocols that cause long-term depression. In this case the postsynaptic cell would now act as a better amplifier of stimuli encoded with larger correlations.

3.6. OPTIMAL RESPONSE AND SYNCHRONY JITTER

The effects of fluctuations in a synchronous presynaptic population can be modeled by adding a Gaussian-distributed jitter, of timescale τ_j, to the timing of each action potential. When the individual components of the synchronous MIP event are too dispersed temporally, i.e., when the jitter is greater than the membrane time constant $\tau_j > \tau$, the MIP event will fail to integrate in the postsynaptic cell. Under these circumstances the effect of correlations is diminished, as illustrated in **Figure 7**. When jitter is absent (**Figure 7A**), different values of presynaptic synchrony S produce distinct and clearly defined optimal response curves. With a physiological jitter timescale of $\tau_j = 2$ ms (**Figure 7B**) the curves for different synchronies shift upwards in n and the peak postsynaptic firing rate falls, particularly for larger synchrony. When $\tau_j = \tau$ (**Figure 7C**) only relatively strong synchrony values are significantly different from the independent case ($S = 1$).

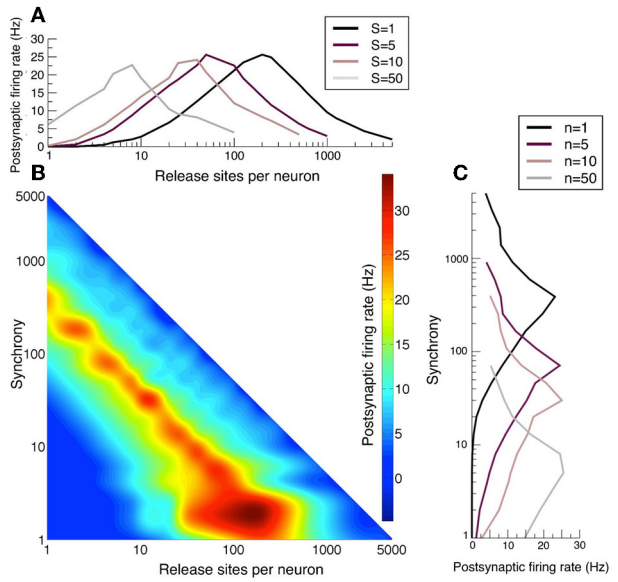


FIGURE 6 | Long-term plasticity that alters release-site number *n* sets the sensitivity to presynaptic synchrony. (A) Postsynatpic rate as a function of release-sites per presynaptic neuron n for different examples of presynaptic synchrony S. **(B)** Heat map of the postsynaptic rate as a function of presynaptic release-site number n and presynaptic synchrony S. **(C)** Postsynaptic rate as a function of presynaptic synchrony S for different examples of release-site number n. For these figures population spikes have been jittered with a standard deviation of 2 ms. Note that in panel **B** the optimal synchrony has an inverse dependency on the release-site number. Long-term potentiation makes the postsynaptic cell more sensitive to weak synchrony, whereas long-term depression sensitizes the cell to stronger synchrony.

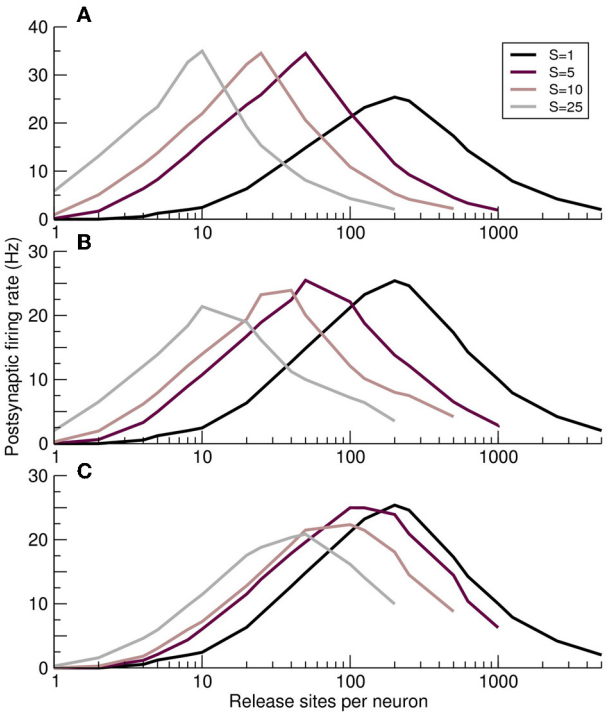


FIGURE 7 | Impact of synchrony jitter on the optimal response curves. (A–C) Postsynaptic firing rate as a function of the number of release sites per presynaptic neuron n for increasing jitter standard deviations τ_j. **(A)** No jitter $\tau_j = 0$. **(B)** Physiological levels of jitter $\tau_j = 2$ ms. **(C)** Response curves converge on the unsynchronized $S = 1$ case, as expected, when jitter is of the order of the postsynaptic membrane time constant $\tau_j = 10$ ms.

3.7. OPTIMAL-RESPONSE CURVES ARE A ROBUST FEATURE OF SYNAPTIC HOMEOSTASIS

Throughout much of the above analysis we held the total number of release sites $M = nN$ constant and demonstrated an optimal response curve in which the postsynaptic rate peaks at an intermediate n, which is dependent on the presynaptic synchrony S. The rationale for this choice is that, under conditions of homeostasis, synaptic potentiation (increasing n) amongst a subpopulation of presynaptic neurons will occur at the expense of pruning neurons that do not contribute to postsynaptic firing. This will lead to the postsynaptic neuron receiving afferent drive from fewer presynaptic neurons, though each of these will make more contacts (and vice-versa for long-term depression). The theoretical results and simulations are not predicated on the assumption of constant M and so it is interesting to investigate whether the optimal-response effect persists if this restriction is relaxed. Using the example $S = 10$ we plotted the postsynaptic rate as a function of the presynaptic neuron N and release site number n (see **Figure 8A**). As expected the postsynaptic rate increases with an increasing number of presynaptic neurons N or release sites n. Plotted on the same figure is the curve $N = M/n$ with $M = 5000$ that, because of its reciprocal relation will have low rates at either asymptotes, and an intermediate maximum (see **Figure 8B**). Also plotted is the curve $N = M_0$ where M_0 is a constant. This corresponds to a scenario in which the entire presynaptic population has either potentiated or depressed their contacts, thereby changing the number of release sites n a presynaptic neuron makes without altering the total number of presynaptic neurons N. For this case, which is arguably an extremum from the point-of-view of homeostatis, the intermediate maximum is lost: the postsynaptic rate increases monotonically and loses its n dependence when n is sufficiently large, as expected from the first form of Equation (18). However, for intermediate cases of homeostasis of the form $N = M_\kappa / n^\kappa$ with $\kappa = 3/4, 1/2, 1/4$ a maximal postsynaptic rate again occurs at some intermediate n (see **Figure 8B**). Given the dependence of the postsynaptic rate on n and N in **Figure 8A** it can be seen geometrically that any curve in which there is a reciprocal relation between N and n will likely feature a maximum at intermediate n and so the optimal-response curves are a robust feature of a postsynaptic neuron in which there is some degree of homeostatic restriction on the total number of afferent contacts.

4. DISCUSSION

We considered the effects of afferent correlations arising from multiple neurotransmitter release sites and a partially synchronized presynaptic population. We derived exact forms for the crosscorrelations of vesicle release site occupancy and vesicle release, and demonstrated that these are identical to those recently obtained from a diffusion and additive-noise approximation (Rosenbaum et al., 2012), validating that approach up

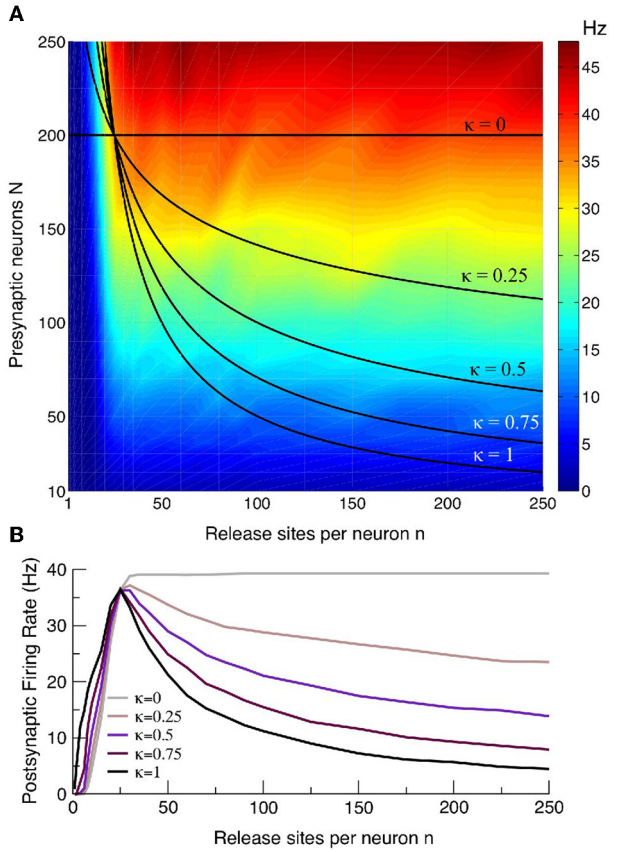


FIGURE 8 | Curves with a maximal postsynaptic rate at intermediate *n* are a robust feature. (A) Intensity plot of the postsynaptic rate as a function of presynaptic release site n and neuron number N for an example with $S = 10$. Also plotted are the relations $N = M_\kappa / n^\kappa$ for $\kappa = 0, 1/4, 1/2, 3/4, 1$, where $\kappa = 1$ corresponds to the homeostatic scenario principally considered in this paper for which there is a restriction $M = nN$ on the total number of afferent contacts. The case $\kappa = 0$ corresponds to a scenario with no such restriction, and the other values of κ are intermediate cases with varying degrees of homeostasis. **(B)** The postsynaptic rate as function of n for the curves in the upper panel. Cases for all values of κ, except $\kappa = 0$ in which there is no homeostatic restriction, show a maximal response at intermediate n. The example curves given have M_κ chosen so that they all pass through the point $n = 25$ and $N = 200$.

to second-order statistics and explaining their perfect agreement between theoretical and simulational results. We further calculated the exact variance of the membrane voltage, in absence of spike threshold. This quantity extends previous calculations (de la Rocha and Parga, 2005) of synaptic conductance fluctuations and allows for an estimation of the postsynaptic rate in the low-correlation Gaussian regime. For the high-correlation regime, due to multiple release sites n or strong synchrony S, we argued that the EPSPs become increasingly large, the nature of the synaptic fluctuations increasingly shot-noise like, and so the postsynaptic rate tends to the rate of synchronous presynaptic events. Combing these two results for the low and high correlation regimes, we demonstrated that the postsynaptic response is maximal for an intermediate number of release sites or synchrony. The system therefore exhibits a tuning-curve response to synchrony that can be modulated by long-term plasticity, which alters the number of release sites n.

Neurons respond maximally to specific stimuli when processing sensory input. A coordination of long-term plasticity, afferent synchrony and short-term depression therefore provides a potential tuning mechanism for cells to achieve this sensitivity. Efficient responsiveness would then depend on historical changes in synaptic connectivity (Taschenberger et al., 2002; Loebel et al., 2013) and the transient correlations evoked by a particular stimulus (Averbeck et al., 2006; Cohen and Kohn, 2011). More generally, neuronal networks balance fidelity of signal transmission with the metabolic costs associated with neurotransmitter recycling (Levy and Baxter, 2002; Savtchenko et al., 2012). Although a release of neurotransmitter beyond that necessary to induce a postsynaptic spike may have medium-term conductance implications or counteract strongly fluctuating inhibition, an efficient network would not be expected to exceed the degree of pairwise connectivity that maximizes response to common spike frequencies and correlations. On the other hand, signals encoded by small numbers of cells would require highly potentiated connections to transmit information with any degree of consistency. This implies that across a neuronal network the degree of clustering would be optimally balanced with individual synaptic weights.

To investigate maximal firing rate response to a defined excitatory drive, we have neglected the effects of synaptic inhibition. As *in vivo* network behaviors arise from a balance of excitation and inhibition, a development of the ideas presented here along the above lines would need to incorporate inhibitory effects on the total synaptic conductance. By altering the timescales on which excitatory inputs are integrated, inhibitory drive could allow a more finely-tuned response to afferent sub-populations with varying degrees of temporal dispersion. Another extension of this work would be to incorporate different forms of short-term synaptic plasticity into the model. This would be particularly appropriate when studying connections between specific cell-types where there is experimental evidence for other forms of synaptic dynamics. It is also likely that effects moderating synaptic depression, such as the increasing facilitation in the maturing neocortex (Reyes and Sakmann, 1999) would lead to qualitatively different behavior as cortical networks develop.

FUNDING

This research was supported by a Warwick Systems Biology Doctoral Training Centre fellowship to Alex D. Bird funded by the UK EPSRC and BBSRC funding agencies.

REFERENCES

Abbott, L. F. (1997). Synaptic depression and cortical gain control. *Science* 275, 221–224. doi: 10.1126/science.275.5297.221

Abbott, L. F., and Regehr, W. G. (2004). Synaptic computation. *Nature* 431, 796–803. doi: 10.1038/nature03010

Aertsen, A. M., Gerstein, G. L., Habib, M. K., and Palm, G. (1989). Dynamics of neuronal firing correlation: modulation of "effective connectivity". *J. Neurophysiol.* 61, 900–917.

Alijani, A. K., and Richardson, M. J. E., (2011). Rate response of neurons subject to fast or frozen noise: from stochastic and homogeneous to deterministic and heterogeneous populations. *Phys. Rev. E* 84, 011919. doi: 10.1103/PhysRevE.84.011919

Averbeck, B. B., Latham, P. E., and Pouget, A. (2006). Neural correlations, population coding and computation. *Nat. Rev. Neurosci.* 7, 358–366. doi: 10.1038/nrn1888

Baker, S. N., Spinks, R., Jackson, A., and Lemon, R. N. (2001). Synchronization in monkey motor cortex during a precision grip task. I. Task-dependent modulation in single-unit synchrony. *J. Neurophysiol.* 85, 869–885.

Brunel, N., and Hakim, V. (1999). Fast global oscillations in networks of integrate-and-fire neurons with low firing rates. *Neural Comput.* 11, 1621–1671. doi: 10.1162/089976699300016179

Cain, N., and Shea-Brown, E. (2013). Impact of correlated neural activity on decision-making performance. *Neural Comput.* 25, 289–327. doi: 10.1162/NECO_a_00398

Capaday, C., Ethier, C., Van Vreeswijk, C., and Darling, W. G. (2013). On the functional organization and operational principles of the motor cortex. *Front. Neural Circ.* 7:66. doi: 10.3389/fncir.2013.00066

Chen, W. X., and Buonomano, D. V. (2012). Developmental shift of short-term synaptic plasticity in cortical organotypic slices. *Neuroscience* 213, 38–46. doi: 10.1016/j.neuroscience.2012.04.018

Cohen, M. R., and Kohn, A. (2011). Measuring and interpreting neuronal correlations. *Nat. Neurosci.* 14, 811–819. doi: 10.1038/nn.2842

Cohen, M. R., and Newsome, W. T. (2008). Context-dependent changes in functional circuitry in visual area MT. *Neuron* 60, 162–173. doi: 10.1016/j.neuron.2008.08.007

de la Rocha, J., and Parga, N. (2005). Short-term synaptic depression causes a non-monotonic response to correlated stimuli. *J. Neurosci.* 25, 8416–8431. doi: 10.1523/JNEUROSCI.0631-05.2005

deCharms, R. C., and Merzenich, M. M. (1996). Primary cortical representation of sounds by the coordination of action-potential timing. *Nature* 381, 610–613. doi: 10.1038/381610a0

Fox, G. Q. (1988). A morphometric analysis of synaptic vesicle distributions. *Brain Res.* 475, 103–117. doi: 10.1016/0006-8993(88)90203-X

Fuhrmann, G., Segev, I., Markram, H., and Tsodyks, M. (2002). Coding of temporal information by activity-dependent synapses. *J. Neurophysiol.* 87, 140–148. doi: 10.1152/jn.00258.2001

Furukawa, T., Kuno, M., and Matsuura, S. (1982). Quantal analysis of a decremental response at hair cell-afferent fibre synapses in the goldfish sacculus. *J. Physiol.* 322, 181–195.

Gardiner, C. (2010). *Stochastic Methods: A Handbook for the Natural and Social Sciences*. Berlin: Springer.

Gerstein, G. L., and Mandelbrot, B. (1964). Random walk models for the spike activity of a single neuron. *Biophys. J.* 4, 41–68. doi: 10.1016/S0006-3495(64)86768-0

Hallermann, S., and Silver, R. A. (2012). Sustaining rapid vesicular release at active zones: potential roles for vesicle tethering. *Trends Neurosci.* 36, 1–10. doi: 10.1016/j.tins.2012.10.001

Hebb, D. O. (2002). *The Organization of Behavior: A Neuropsychological Theory*. Mahwah, NJ: L. Erlbaum Associates.

Hu, Y., Qu, L., and Schikorski, T. (2008). Mean synaptic vesicle size varies among individual excitatory hippocampal synapses. *Synapse* 62, 953–957. doi: 10.1002/syn.20567

Kilpatrick, Z. P. (2012). Short term synaptic depression improves information transfer in perceptual multistability. *Front. Comput. Neurosci.* 7:85. doi: 10.3389/fncom.2013.00085

Kuhn, A. (2004). Neuronal integration of synaptic input in the fluctuation-driven regime. *J. Neurosci.* 24, 2345–2356. doi: 10.1523/JNEUROSCI.3349-03.2004

Kuhn, A., Aertsen, A., and Rotter, S. (2003). Higher-order statistics of input ensembles and the response of simple model neurons. *Neural Comput.* 15, 67–101. doi: 10.1162/089976603321043702

Levy, W. B., and Baxter, R. A. (2002). Energy-efficient neuronal computation via quantal synaptic failures. *J. Neurosci.* 22, 4746–4755.

Loebel, A., Le Be, J. V., Richardson, M. J. E., Markram, H., and Herz, A. V. M. (2013). Matched pre- and post-synaptic changes underlie synaptic plasticity over long time scales. *J. Neurosci.* 33, 6257–6266. doi: 10.1523/JNEUROSCI.3740-12.2013

Markram, H., Gerstner, W., and Sjöström, P. J. (2011). A history of spike-timing-dependent plasticity. *Front. Synaptic Neurosci.* 3:4. doi: 10.3389/fnsyn.2011.00004

Megías, M., Emri, Z., Freund, T. F., and Gulyás, A. I. (2001). Total number and distribution of inhibitory and excitatory synapses on hippocampal CA1 pyramidal cells. *Neuroscience* 102, 527–540. doi: 10.1016/S0306-4522(00)00496-6

Mitchell, J. F., Sundberg, K. A., and Reynolds, J. H. (2009). Spatial attention decorrelates intrinsic activity fluctuations in Macaque area V4. *Neuron* 63, 879–888. doi: 10.1016/j.neuron.2009.09.013

O'Kusky, J., and Colonnier, M. (1982). A laminar analysis of the number of neurons, glia, and synapses in the adult cortex (area 17) of adult macaque monkeys. *J. Comp. Neurol.* 210, 278–290. doi: 10.1002/cne.902100308

Reyes, A., and Sakmann, B. (1999). Developmental switch in the short-term modification of unitary EPSPs evoked in layer 2/3 and layer 5 pyramidal neurons of rat neocortex. *J. Neurosci.* 19, 3827–3835.

Richardson, M. J. E., and Swarbrick, R. (2010). Firing-rate response of a neuron receiving excitatory and inhibitory synaptic shot noise. *Phys. Rev. Letts.* 105:178102. doi: 10.1103/PhysRevLett.105.178102

Rosenbaum, R., Rubin, J., and Doiron, B. (2012). Short term synaptic depression imposes a frequency dependent filter on synaptic information transfer. *PLoS Comput. Biol.* 8:e1002557. doi: 10.1371/journal.pcbi.1002557

Rothman, J. S., Cathala, L., Steuber, V., and Silver, R. A. (2009). Synaptic depression enables neuronal gain control. *Nature* 457, 1015–1018. doi: 10.1038/nature07604

Salinas, E., and Sejnowski, T. J. (2000). Impact of correlated synaptic input on output firing rate and variability in simple neuronal models. *J. Neurosci.* 20, 6193–6209.

Savtchenko, L. P., Sylantyev, S., and Rusakov, D. A. (2012). Central synapses release a resource-efficient amount of glutamate. *Nat. Neurosci.* 16, 10–12. doi: 10.1038/nn.3285

Scott, P., Cowan, A. I., and Stricker, C. (2012). Quantifying impacts of short-term plasticity on neuronal information transfer. *Phys. Rev. E* 85, 041921. doi: 10.1103/PhysRevE.85.041921

Seriès, P., Latham, P. E., and Pouget, A. (2004). Tuning curve sharpening for orientation selectivity: coding efficiency and the impact of correlations. *Nat. Neurosci.* 7, 1129–1135. doi: 10.1038/nn1321

Spruston, N. (2008). Pyramidal neurons: dendritic structure and synaptic integration. *Nat. Rev. Neurosci.* 9, 206–221. doi: 10.1038/nrn2286

Südhof, T. C. (2004). The synaptic vesicle cycle. *Annu. Rev. Neurosci.* 27, 509–547. doi: 10.1146/annurev.neuro.26.041002.131412

Taschenberger, H., Leão, R. M., Rowland, K. C., Spirou, G. A., and von Gersdorff, H. (2002). Optimizing synaptic architecture and efficiency for high-frequency transmission. *Neuron* 36, 1127–1143. doi: 10.1016/S0896-6273(02)01137-6

Tsodyks, M., Pawelzik, K., and Markram, H. (1998). Neural networks with dynamic synapses. *Neural Comput.* 10, 821–835. doi: 10.1162/089976698300017502

Tsodyks, M. V., and Markram, H. (1997). The neural code between neocortical pyramidal neurons depends on neurotransmitter release probability. *Proc. Natl. Acad. Sci. U.S.A.* 94, 719–723. doi: 10.1073/pnas.94.2.719

von der Malsburg, C. (1981). "Internal Report 81-2, Neurobiology, Max-Planck Institute for Biophysical Chemistry," in *The Correlation Theory of Brain Function*, (Gottingen).

Zador, A. A. (1998). Impact of synaptic unreliability on the information transmitted by spiking neurons. *J. Neurophysiol.* 79, 1219–1229.

Zucker, R.S., and Regehr, W. G. (2002). Short-term synaptic plasticity. *Annu. Rev. Physiol.* 64, 355–405. doi: 10.1146/annurev.physiol.64.092501.114547

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.

Received: 06 August 2013; accepted: 07 January 2014; published online: 30 January 2014.

Citation: Bird AD and Richardson MJE (2014) Long-term plasticity determines the postsynaptic response to correlated afferents with multivesicular short-term synaptic depression. Front. Comput. Neurosci. ***8****:2. doi: 10.3389/fncom.2014.00002*

This article was submitted to the journal Frontiers in Computational Neuroscience.

Copyright © 2014 Bird and Richardson. 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) or licensor 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.

APPENDIX

DERIVATION OF THE VOLTAGE VARIANCE

The voltage equation can be written in the form

$$\tau\frac{dV}{dt} = E - V + a\tau\zeta \tag{19}$$

where ζ is the summation of the release trains across the N presynaptic neurons and each of their n contacts

$$\zeta = \sum_{i=1}^{N}\sum_{j=1}^{n}\chi_{ij} \tag{20}$$

where χ_{ij} takes the form of Equation (11) for the ith presynaptic neuron's jth contact. The autocorrelation of ζ is therefore comprised of Nn autocorrelations of χ in the form of Equation (12), $Nn(n-1)$ crosscorrelations of χ for distinct release trains sharing the same presynaptic neuron given by Equation (13) with $\gamma = 1$ and $N(N-1)n^2$ crosscorrelations of χ for release trains with different presynaptic neurons given by Equation (13) with $\gamma = c$.

Taking expectations of both side of Equation (19) in the steady state gives

$$\langle V\rangle = E + a\tau\langle\zeta\rangle = E + aM\tau R_a p\langle x\rangle . \tag{21}$$

We can now solve Equation (19) to give

$$V - \langle V\rangle = a\int_{-\infty}^{t} dt' e^{-(t-t')/\tau}\left(\zeta(t') - \langle\zeta\rangle\right) \tag{22}$$

so that the voltage variance can be written as an integral over the autocorrelation of ζ, $\text{Autocorr}(\zeta) = \left\langle\left(\zeta(t') - \langle\zeta\rangle\right)\left(\zeta(t'') - \langle\zeta\rangle\right)\right\rangle$

$$(V - \langle V\rangle)^2 = a^2\int_{-\infty}^{t} dt'\int_{-\infty}^{t} dt'' e^{-(t-t')/\tau}e^{-(t-t'')/\tau}\text{Autocorr}(\zeta). \tag{23}$$

As discussed above, the autocorrelation of ζ is the sum of the various crosscorrelations of χ so that it must take the form

$$\text{Autocorr}(\zeta) = \alpha\delta(t' - t'') + \beta e^{-|t'-t''|/\tau_x} \tag{24}$$

where α and β are obtained from the prefactors of the terms in Equations (12, 13) multiplied by their respective contributions. Inserting Equation (24) into (23) and performing the integration gives

$$\text{Var}(V) = a^2\left(\frac{\alpha\tau}{2} + \frac{\beta\tau^2\tau_x}{\tau + \tau_x}\right). \tag{25}$$

On substituting the appropriate forms for α and β the result given in Equation (15) is obtained.

frontiers in
COMPUTATIONAL NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 27 January 2014
doi: 10.3389/fncom.2013.00195

Phase synchrony facilitates binding and segmentation of natural images in a coupled neural oscillator network

***Holger Finger*[1]* and Peter König*[1,2]**

[1] *Institute of Cognitive Science, University of Osnabrück, Osnabrück, Germany*
[2] *Institute of Neurophysiology and Pathophysiology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany*

Edited by:
Tatjana Tchumatchenko, Max Planck Institute for Brain Research, Germany

Reviewed by:
Abdelmalik Moujahid, University of the Basque Country, Spain
Andrea K. Barreiro, Southern Methodist University, USA

****Correspondence:***
Holger Finger, Institute of Cognitive Science, University of Osnabrück, Albrechtstraße 28, 49069 Osnabrück, Germany
e-mail: holger.finger@uos.de

Synchronization has been suggested as a mechanism of binding distributed feature representations facilitating segmentation of visual stimuli. Here we investigate this concept based on unsupervised learning using natural visual stimuli. We simulate dual-variable neural oscillators with separate activation and phase variables. The binding of a set of neurons is coded by synchronized phase variables. The network of tangential synchronizing connections learned from the induced activations exhibits small-world properties and allows binding even over larger distances. We evaluate the resulting dynamic phase maps using segmentation masks labeled by human experts. Our simulation results show a continuously increasing phase synchrony between neurons within the labeled segmentation masks. The evaluation of the network dynamics shows that the synchrony between network nodes establishes a relational coding of the natural image inputs. This demonstrates that the concept of binding by synchrony is applicable in the context of unsupervised learning using natural visual stimuli.

Keywords: oscillation, binding, synchronization, normative model, unsupervised learning, scene segmentation, object label, natural image statistics

1. INTRODUCTION

One of the central questions in neuroscience is how information about a given stimulus is processed in a distributed network of neurons such that it is perceived not only as a collection of unrelated features but as a unified single object. The concept of binding by synchrony has been proposed as a mechanism to coordinate the spatially distributed information processing in the cortex (Milner, 1974; Von Der Malsburg, 1981). Experiments in cat visual cortex have confirmed that inter-columnar synchronization indeed corresponds to a relational code that reflects global stimulus attributes (Gray et al., 1989; Singer, 1999; Engel and Singer, 2001). However, the physiological recordings in these early studies were based on the presentation of artificially designed stimuli. In a more recent study Onat et al. (2013) showed in experiments that long-range interactions in the visual cortex are compatible with Gestalt laws. This suggests that the concept of binding by synchrony is also feasible in the case of natural visual stimuli. It is still the center of a heated debate to what extend synchronized activity represents a neural code of binding and segmentation. Especially, how the neural system can learn this relational coding when it is exposed to new stimuli is still an open question. The most prominent possibility is that tangential cortico-cortical connections in the visual cortex lead to synchronized activity that implements Gestalt laws. Löwel and Singer (1992) showed in cats with artificially induced strabismus that selective stabilization of tangential connections occurs between cells that exhibit correlated activity induced by visual experience. Furthermore, König et al. (1993) found that the synchronization of cortical activity is impaired in these cats with artificial strabismus. These findings indicate that there is an important interplay between unsupervised learning of tangential connections on behavioral time scales and their role in synchronization phenomena on fast time scales.

The physiological experiments on binding by synchrony have been accompanied by theoretical studies early on. Sompolinsky et al. (1990) investigated how a model of coupled neural oscillators is able to process global stimulus properties in synchronization patterns using abstractly defined neuronal activation levels and predefined coupling strengths for the simulated network. These simulation results showed that the coupling of neural oscillators provides a viable mechanism implementing a coding of perceptual grouping. Such previous work includes studies ranging from networks build out of very simple elements to detailed simulations containing many compartments per unit.

To investigate the functional role of synchronization and its relation to coding, it is important to choose the right level of abstraction in the model. A simplification from detailed spiking neuron models to coupled phase oscillator models allows us to analyze neuronal synchronization in a broader context of a normative model involving unsupervised learning from natural stimuli. A review of these coupled neural oscillator models was done by Sturm and König (2001), where the authors show the derivation of simplified phase update equations from biologically measurable phase response curves. The simplifications in coupled phase oscillators are based on the assumption that neurons are close to their oscillatory limit-cycle and that a change in the phase of the neuronal inputs induces only a small perturbation to the neuronal phase. The phase update equation in our model is based on the Kuramoto model of coupled phase oscillators (Kuramoto, 1984) in the sense that our model also assumes

a very simple sinusoidal phase interaction function. This approximation of the phase interaction by a sinusoidal function allows us to use mathematical simplifications in the simulation of the model.

Very similar to the work of Sompolinsky et al. (1990), we extend the standard formulation of the Kuramoto model with a second variable per neuron to encode the activation of the oscillators. Therefore, in our model the state of a neuron is represented by 2 degrees of freedom, which are separated into activation and phase variables. This discrimination between coding of receptive field features by activation and coding of relationships by phase is a biologically motivated segregation of their different functional roles. Maye and Werning (2007) specifically compare the synchronization properties of these coupled phase oscillator models with mean-field oscillator models based on the Wilson-Cowan model (Wilson and Cowan, 1972). They state that the simplified coupled phase oscillators allow decoupling the simulation time constants of fast oscillatory time scales from slow rate coding time scales. Another advantage is an easier analysis of the synchronization patterns, because the direct encoding of the phase variables means that all contextual relationships are coded at the same time. Consequently, we use the dual variable phase model, because it is suitable to answer fundamental questions about the interactions between synchronization phenomena and contextual coding in neural systems.

In contrast to these phase oscillator models, most recent work on segmentation in networks of coupled neural oscillators is based on the so called "local excitatory global inhibitory oscillator network" (LEGION) or similar variants of this model, which was first proposed by Wang and Terman (1997). In LEGION the dynamics of each oscillatory period of individual units is simulated in detail by time-varying variables describing the internal states of each neuron. In contrast, in our model the oscillatory period is not simulated, but represented only implicitly in the phase variables. Nonetheless, several aspects which we analyze in this work were previously also investiged in LEGION. Namely, similar to Li and Li (2011) we use a small-world topology, to reduce the computational cost while still allowing binding by synchrony over large distances. We also use parallel computations to speed up the simulations, which was also previously done in LEGION by Bauer et al. (2012).

The above-mentioned previous theoretical studies mostly investigated the processing of artificial stimuli in close analogy to the physiological experiments. These stimuli are heavily dominated by artificial geometric patterns as bars and gratings. However, the concept of binding by synchrony makes much more general claims about grouping of sensory representations of natural stimuli. By now a fair number of databases with images considered to be natural is available. However, a problem with generic natural stimuli is that segmentation is not only difficult, but no general ground truth is available. The LabelMe database (Russell et al., 2008) is rather unique, as it contains a large collection of images together with human labeled annotations of image segments. In theoretical studies these labels may serve as a ground truth to evaluate how the relative phases between neurons are coding relational structures on natural stimuli.

The processing of natural stimuli in neural systems can be described as a normative approach in which the representation of the input is learned by an optimization of computational principles (Einhäuser and König, 2010). It has been successfully employed in modeling receptive field properties of simple and complex cells in primary visual cortex. Furthermore, response properties of neurons in higher areas and other modalities have been suggested to follow similar rules. This approach might be extended to include the computational principles that underlie tangential interactions that directly influence synchronization phenomena. This might answer the question whether the concept of binding by synchrony can work in principle with unsupervised learning and natural stimuli.

In this study we investigate whether the concept of binding by synchrony, as has been investigated using abstract stimuli, is viable for natural stimuli. The most important novelty of our approach is the combination of these different concepts described above into one single simulation model to allow the investigation of their interplay: Specifically, we combine normative model approaches of unsupervised learning from natural stimuli with the concept of binding by synchrony in a network of coupled phase oscillators. Importantly, the data driven approach, that utilizes general principles, minimizes the number of heuristics and free parameters. We present large-scale simulations of neural networks encoding real-world image scenes. In the first stage of our algorithm forward projections generate activation levels of neurons corresponding to the primary visual cortex. In the second stage these activation levels are used in a simulation of tangential coupled phase oscillators. We present results with forward projections based on designed Gabor filters that are a good approximation of receptive fields in the primary visual cortex. To allow later canonical generalization in higher network layers, we also present results with forward projections learned in a normative model approach with a sparse autoencoder using natural image statistics. In addition to these learned forward weights, the structural connectivity of the phase simulations is also learned unsupervised using the correlated activity induced by natural stimuli. Performance of the network is tested using images taken from the LabelMe database. Thereby we can investigate how synchronization phenomena might be utilized in sensory cortical areas to bind different attributes of the same stimulus and how it might be exploited for scene segmentation.

2. MATERIALS AND METHODS

The overall network architecture of our simulation model consists of two main parts: (1) Feedforward convolutional filters (red lines in **Figure 1**) are used to generate the activation levels for neurons in a layer corresponding to the primary visual cortex. On top of each pixel is a column of neurons which encode different features of a local patch in the input image (black bottom cuboid in **Figure 1**). Each feature type is described by a weight matrix which is applied using a 2-dimensional-convolutional operation on each rgb-color-channel of the input image. (2) The obtained activation levels in this 3-dimensional structure (black top cuboid in **Figure 1**) are subsequently used to simulate sparse connections (green lines in **Figure 1**) between coupled phase oscillators.

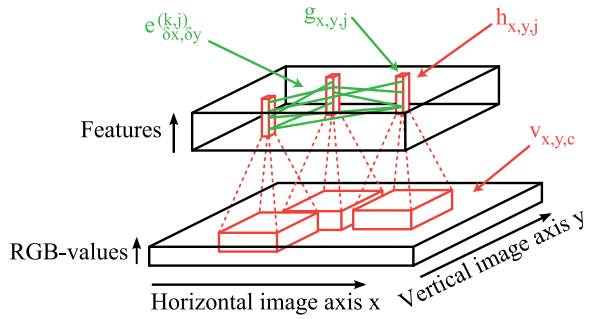


FIGURE 1 | Network structure. Feedforward convolutional filters (red) are applied on the input image $v_{x,y,c}$ (bottom) to generate the activation levels $h_{x,y,j}$ of feature columns (red blocks in the top). These activations $h_{x,y,j}$ are then transformed to activations of oscillators $g_{x,y,j}$ using simple local regularization steps. The intralayer connections $e^{k,j}_{\delta x,\delta y}$ (green) simulate these coupled phase oscillators which synchronize or desynchronize image features.

We start with the description of the stimulus material (section 2.1). This is followed by the description of the coupled phase oscillator model (section 2.2) and the sampling mechanism generating the horizontal sparse connections (section 2.3). Afterwards we describe the underlying mechanism of the feedforward generation of activation levels (section 2.4).

2.1. NATURAL STIMULUS MATERIAL

As stimulus material in our simulations we use images of suburban scenes from the LabelMe database (Russell et al., 2008). Due to computational time constrains we have to restrict the evaluations to a small subset of all available images in the database. In addition, the database is not fixed but new images and segmentation masks are often added. We use only the first 50 images in the folder *05june05_static_street_boston* so that we have a consistent and fixed dataset of well defined images.

These images have initially a resolution of 2560 × 1920 pixels. We first resize the images to 400 × 300 pixels to further reduce the computation time of the simulations. Subsequently we subtract the mean pixel values and apply a smoothed zero-phase (ZCA) whitening transformation (Bell and Sejnowski, 1997). For an input image X the whitened pixel values are given by $X_{\mathrm{ZCA}} = UDU^T X$, where U is a matrix containing the eigenvectors of the covariance matrix of the image statistics and D is a diagonal matrix with diagonal elements $\frac{1}{\sqrt{\lambda_i + 0.1}}$ where λ_i are the corresponding eigenvalues. This transformation applies local center-surround whitening filters that decrease the correlations in the input images. We implement this whitening transformation using a convolutional image filter.

The images in the LabelMe database come along with human labeled segmentation masks. These segmentation masks correspond to objects that are perceived as a unique concept with an associated abstract label like "tree," "car" or "house." We use these supervised segmentation masks for later evaluations of binding in the simulated phase maps. Please note that in our network simulations this segmentation information is not used at any moment in time. Instead, the network connectivity is based solely on unsupervised learning using the statistics of neuronal activations.

2.2. COUPLED PHASE OSCILLATOR MODEL

Our network of coupled phase oscillators is based on the oscillator model described by Sompolinsky et al. (1990). In the following, we use the same motivational derivation of the phase update equations. We model the probability of firing $P_{x,y,k}(t)$ per unit time of a neuron at image position (x, y) encoding feature type k at time t by an isochronous oscillator. In our simulations we represent the state of the neuronal oscillators by seperated activation variables $g_{x,y,k}$ and phase variables $\Phi_{x,y,k}$. These two variables are linked to the biological interpretation of firing probability by the equation

$$P_{x,y,k}(t) = g_{x,y,k}\left(1 + \lambda \cdot \cos(\Phi_{x,y,k}(t))\right), \tag{1}$$

where the parameter $0 < \lambda < 1$ controls the relative strength of the temporal oscillation in relation to the overall firing probability of the neuron. The phase progression is a periodic function $\Phi_{x,y,k}(t) = \Phi_{x,y,k}(t + 2\pi)$. In our work, the calculation of the activation levels $g_{x,y,k}$ significantly differs from the simple artificial tuning curves used in Sompolinsky et al. (1990). A detailed description of how these activation levels are obtained will be presented in section 2.4. The activation levels $g_{x,y,k}$ are normalized by dividing by the local sum of all activation levels at each image position such that $\sum_k g_{x,y,k} = 1 \forall x, y \in \mathbb{Z}$. In the simulations presented in this work the activation levels of each neuron are only computed once from the input image using feedforward projections (red lines in **Figure 1**) and are then kept constant during the simulation of the phase model. This simplification of constant activation levels is based on the assumption that the stimulus presentaion on behavioral timescales ($\approx$ seconds) remains constant during the phase synchronization which happens at very fast timescales (i.e., gamma frequency $\approx$ 40 Hz). Another argument to support this assumption is that the visual cortex seems to operate in a regime of self-sustained activity (Stimberg et al., 2009) and therefore we can assume constant activation levels during the phase simulation.

After these activation levels V are computed, we simulate the horizontal coupling between the phase oscillators. The phase connections in our network are described by a weighted graph $G = (H, E)$ where the neurons $g_{x,y,j} \in H$ are the vertices organized in a three dimensional block (**Figure 1**). An edge $e^{(j,k)}_{\delta x,\delta y} \in E$ describes synchronizing (positive) or desynchronizing (negative) connections from neurons $g_{x,y,j}$ to neurons $g_{x+\delta x, y+\delta y, k}$. The phase of each neuron is then modeled according to a differential equation describing weakly coupled phase oscillators (Kuramoto, 1984)

$$\frac{d\Phi_{x,y,k}(t)}{dt} = \omega - \frac{1}{\tau} \sum_{e^{(j,k)}_{\delta x,\delta y} \in E} g_{x,y,k} \cdot e^{(j,k)}_{\delta x,\delta y} \cdot g_{x-\delta x, y-\delta y, j} \cdot \sin\left(\Phi_{x,y,k}(t) - \Phi_{x-\delta x, y-\delta y, j}(t)\right), \tag{2}$$

where τ is the time scale of the phase interactions and ω is the natural frequency of the modeled neural oscillations. We assume that

all neurons have the same intrinsic natural frequency ω and the interaction strength $g_{x,y,k} \cdot e^{(j,k)}_{\delta x,\delta y} \cdot g_{x-\delta x,y-\delta y,j}$ is proportional to the activation levels of the pre- and post-sysnaptic neurons. Note that our model is in contrast to the more common formulation of the Kuramoto model with heterogeneous frequencies and fixed homogenous all-to-all interaction strengths.

A major difference to the phase update equation used in Sompolinsky et al. (1990) is that we neglect the noise term in the differential equation of each oscillator. The noise term in Sompolinsky et al. (1990) is used as the primary source of desynchronization in the network. In contrast, in our work, we use a normative model to learn not only synchronizing but also desynchronizing connections (see section 2.3). For an easier analysis and interpretation of the results, it is advantageous to have only a single source for the desynchronization in the network. Therefore, we decided to use a deterministic phase model, although it was previously shown that noise is an important factor to control the network coherence. In addition to a simpler interpretation it reduces the number of model parameters and is also more compatible to further applications of gradient descent learning to change the strength of the phase interactions.

We can further simplify the equation by using the fact that we model isochronous oscillators with homogeneous frequencies. In Equation 2 all phase variables $\Phi_{x,y,k}(t)$ have a constant phase progression with frequency ω. We can use a simple transformation to a new variable, which represents only the phase offsets between neurons:

$$\varphi_{x,y,k}(t) = \Phi_{x,y,k}(t) - \omega t. \tag{3}$$

This new phase variable $\varphi_{x,y,k}(t)$ describes the relative phase of neuron k to the global fixed network oscillation with frequency ω. Substitution into the equation above leads to a simplified phase update equation

$$\frac{d\varphi_{x,y,k}(t)}{dt} = -\frac{1}{\tau} \sum_{e^{(j,k)}_{\delta x,\delta y} \in E} g_{x,y,k} \cdot e^{(j,k)}_{\delta x,\delta y} \cdot g_{x-\delta x,y-\delta y,j} \cdot \sin\left(\varphi_{x,y,k}(t) - \varphi_{x-\delta x,y-\delta y,j}(t)\right). \tag{4}$$

In this equation it can be seen that the timescale τ of the phase interaction strength is decoupled from the oscillatory timescale $1/\omega$. Please also note, that a change of the parameter τ would not qualitatively change the results of our simulations. Instead it would just linearly change the units of the time axes. Therefore, we show the simulation results with the time axis measured in iterations, which could be linearly scaled to arbitrary time units to best fit to different biological measurements.

This phase update equation is used in our simulations to model the horizontal connections in the network. It allows directly specifying synchronizing interactions from neuron $g_{x,y,j}$ to neuron $g_{x+\delta x,y+\delta y,k}$ with a positive connection weight $e^{(j,k)}_{\delta x,\delta y}$ and desynchronizing interactions with a negative weight respectively. We simulate these coupled differential equations using a 4th-order Runge-Kutta method.

2.3. HORIZONTAL INTERACTION STRENGTHS

We use correlation statistics of the induced activation levels to set the intralayer connection strengths similar to a simple Hebbian learning rule. We write $\rho^{(k,m)}_{x,y}$ to denote the Pearson cross-correlation between the activations of feature type k at image position $(\tilde{x}, \tilde{y})$ and the activations of feature type m at the shifted image position $(\tilde{x}+x, \tilde{y}+y)$. Each correlation value in this tensor is calculated from the correlation statistics over approximately 1 million network activations induced by 50 natural images and presented at 236×86 image positions.

These horizontal connections make up the coupling between the neural oscillators. Instead of full connectivity, we use stochastically sampled sparse directed connections from the correlation matrix. To exclude noise in the correlation matrix, we use the Benjamini-Hochberg-Yekutieli procedure (Benjamini and Yekutieli, 2001) under arbitrary dependence assumptions with a false-discovery rate of 0.05.

The probability of a positive $(+1)$ or a negative connection (-1) in the connectivity graph $G = (H, E)$ is then given by

$$P\left(e^{(j,k)}_{x,y} = \pm 1\right) = \eta_{\pm} \cdot \frac{\max\left(0, \pm\rho^{(j,k)}_{x,y}\right)}{\sum_{\tilde{x},\tilde{y},m} \max\left(0, \pm\rho^{(m,k)}_{\tilde{x},\tilde{y}}\right)}, \tag{5}$$

where η_+ specifies the total number of afferent synchronizing connections and η_- the total number of afferent desynchronizing connections per neuron. Therefore, synchronizing connections exist only between naturally correlated features and desynchronizing connections between anti-correlated features.

We sample this sparse tangential connection pattern such that it is invariant to spatial shift transformations. The convolutional structure of the forward projections leads to activation and phase variables that are stored in a 3-dimensional block (top of **Figure 1**) with two dimensions given by the spatial extend of the image and one feature dimension. This convolutional structure can be exploited for the sparse horizontal connections to significantly speed up the computation. Therefore, we specify the properties of the coupled oscillator connections only for a generic feature column. These connections are then applied at each image position. Specifically, in our implementation each sampled tangential connection is specified by 5 variables: the horizontal and vertical connection length in image directions and the indices of the afferent and efferent feature maps and the connection weight. This has the advantage that the phase update equation can be implemented as a vectorized convolutional operation although the connection pattern is highly sparse.

2.4. FEEDFORWARD CONNECTIVITY

We compare the binding and segmentation performance of the coupled neural oscillator model using two different ways to generate the activation levels for the neurons. We first describe hand-crafted feedforward Gabor weights (section 2.4.1) and then the unsupervised learning of receptive fields using a convolutional autoencoder (section 2.4.2). Finally, activation functions are presented to further regularize the resulting feature representations (section 2.4.3).

2.4.1. Gabor filters

For reference we use a set of Gabor filters with specified orientation, frequency and color tuning to generate the activation levels for the phase simulation. Thereby we can analyze the phase oscillator network based on a regularly defined set of features that can be parameterized.

We generate linear convolutional weights (marked in red in **Figure 1**) using an approximate Gaussian derivative model, which was shown to be a good fit for the receptive fields of simple cells in the primate visual cortex (Young, 1987). We use only non-directional three-lobe monophasic receptive fields (Young and Lesperance, 2001) to reduce our model parameters. We implement the Gaussian derivative model using difference-of-offset-Gaussians with a slightly larger center compared to surround to code color offsets. The receptive fields that are used in our simulations have a size of 12x12 pixels and are defined by

$$W_{x,y} = g_{2\sigma}(y) \cdot \left(-5 \cdot g_{\sigma}(x + \sigma) + 10.1 \cdot g_{\sigma}(x) - 5 \cdot g_{\sigma}(x - \sigma)\right), \tag{6}$$

where $g_{\sigma}(x)$ is a one dimensional Gaussian distribution with standard deviations $\sigma = 1.5$ pixels (or $g_{2\sigma}(y)$ with standard deviation of $2\sigma = 3$ pixels) and the coordinates x and y are rotated giving a total of 8 orientations in steps of 22.5°. The convolutional filters are applied to the images with a stride of 2 pixels in both image dimensions and are followed by a sigmoidal activation function to scale the values to a reasonable interval between 0 and 1. We apply each orientation filter separately to all color channels (red, green, blue). Furthermore, we add features for the complementary color channels similar to the on-off discrimination in the visual pathway from the retina to the visual cortex. The direct linear dependency between these pairs of opponent-color channels is removed later with additional activation functions described in section 2.4.3. In summary, we have a total of 48 convolutional feature channels per image position: 8x orientations, 3x rgb-color channels, 2x opponent-color channels. This overcomplete neural representation of the input images is used to generate the activation levels for the phase simulations.

Cortical measurements show that the distribution of non-directional monophasic simple cells is roughly uniformly distributed between zero-, first- and second order Gaussian derivatives (Young and Lesperance, 2001). We performed the simulations presented here also with mixed receptive fields of zero-, first- and second-order Gaussian derivatives and obtained similar results. We present here only results with second order Gaussian derivatives, because this reduces the number of model parameters drastically.

2.4.2. Autoencoder filters

As a comparison to these regular hand-designed Gabor filters we analyze the oscillatory network based on activation levels generated by unsupervised learned autoencoder weights. A good overview of the concepts described in this section can be found in Le et al. (2011b), where the authors analyze different optimization methods for convolutional and sparse autoencoders. An autoencoder learns a higher level representation from the stimulus statistics such that the input stimuli can be reconstructed from the hidden representations. In addition, we optimize the sparsity of the activation levels in this representation, which was shown to learn connection weights which resemble receptive fields in the visual cortex (Olshausen and Field, 1996; Hinton, 2010; Le et al., 2011a).

A common trick in unsupervised learning in neural networks are shared connection weights to reduce the number of parameters that have to be learned, which can be accomplished by a convolutional feed-forward network in the case of images (LeCun et al., 1998; Lee et al., 2009). The structure of our convolutional autoencoder is shown in **Figure 2**. The feedforward projections that generate the activation of feature map j consist of convolutional filters $W_{x,y,c,j}$ (red lines in **Figure 2**) with input features $c \in \{1, 2, 3\}$ (rgb-colors) and a bias term b_j and is followed by a sigmoidal activation function. Therefore, the hidden layer activation map of feature $j \in \{1, 2, \ldots, J\}$ is described by

$$h_{x,y,j} = f\left(\sum_{c=1}^{3} W_{x,y,c,j} * v_{x,y,c} + b_j\right). \tag{7}$$

The hidden layer activation h of each input image sample is also a 3 dimensional block (horizontal and vertical image dimensions and the feature type). The weight matrix W is a 4 dimensional structure which describes the connection weights from a convolutional input block to one output column in the hidden layer. The convolutional image operations ($*$) are applied in the image directions x and y between all combinations of input feature maps c and all output feature maps j.

We use linear activation functions for the backward projections (blue lines in **Figure 2**) so that the output matches the scale of the input images (zero-mean). We use another set of weights $\hat{W}_{x,y,j,c}$ and bias terms $\hat{b}_c$ to describe these backward connections. Therefore, the activation in the reconstruction layer is given by

$$\hat{v}_{x,y,c} = \sum_{j=1}^{J} \hat{W}_{x,y,j,c} * h_{x,y,j} + \hat{b}_c, \tag{8}$$

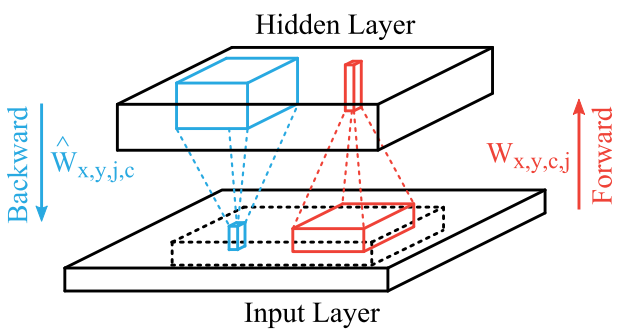


FIGURE 2 | Structure of the convolutional autoencoder. Convolutional forward weights (red) compute the hidden layer activation levels and convolutional backward weights (blue) generate the reconstruction of the given input. The reconstruction layer is compared to the centered part (dashed block) of the input layer.

where $J = 100$ is the number of different feature types. During the learning stage only the valid part (no zero padding) of the convolutions are used for the forward and backward projections to avoid edge effects of the image borders on the learned weights. Similar to the Gabor filters the convolutional filters have a size of 12x12 pixels and are applied using a stride of 2 pixels leading to a reduction in the resolution of the hidden layer.

We use the sum of 3 optimization functions to learn the forward and backward weights of the autoencoder. The first optimization term which is minimized is the reconstruction error averaged over all positions and training samples s and is given by

$$\Psi_1 = \left\langle \frac{1}{2} \left\| \hat{v}^{(s)}_{x,y,c} - v^{(s)}_{x,y,c} \right\|^2 \right\rangle_{x,y,s}. \quad (9)$$

The second term optimizes the sparseness of the hidden units as described by Hinton (2010) and Le et al. (2011a) with

$$\Psi_2 = \beta \cdot \sum_j \mathrm{KL}\left(\tilde{h} \middle\| \langle h^{(s)}_{x,y,j} \rangle_{x,y,s} \right), \quad (10)$$

where KL is the Kullback-Leibler-divergence between two Bernoulli distributions with expected values $\tilde{h}$ and $\langle h^{(s)}_{x,y,j} \rangle_{x,y,s}$. We set the desired average activation $\tilde{h} = 0.035$.

The third term is a weight decay (L2-norm) of all forward and backward weights and is given by

$$\Psi_3 = \frac{\lambda}{2} \cdot \left(\sum_{x,y,c,j} W^2_{x,y,c,j} + \sum_{x,y,j,c} \hat{W}^2_{x,y,j,c} \right). \quad (11)$$

This optimization term pushes all connection weights toward zero such that only the connections which help to extract useful features remain. Therefore, it provides a regularization mechanism during learning.

For the simulations presented in this paper we use a relative weighting between these optimization functions given by $\beta = 90$ and $\lambda = 0.3$. The gradients of the optimization functions are calculated using back propagation of error signals and were checked using numerical derivatives. The sum of the three terms described above is minimized with the limited memory Broyden-Fletcher-Goldfarb-Shanno algorithm (L-BFGS), which uses an approximation to the inverse Hessian matrix (Liu and Nocedal, 1989). We use the *minFunc* library of Mark Schmidt[1] with default parameters for line search with a strong Wolfe condition. We use L-BFGS because it converges much faster in comparison to standard gradient descent, especially in the case of autoencoders with sparseness constrains (Le et al., 2011b). Another advantage of L-BFGS is that extensive tuning of learning parameters as in standard gradient descent methods is not necessary.

The training data consists of 1000 color patches (60 × 60 pixels) sampled from the folder *05june05_static_street_boston* of the LabelMe database (Russell et al., 2008). This corresponds to 625.000 training samples per convolutional fragment where the forward weight matrix is applied. After 500 iterations the features are mostly oriented patches and sensitive to different colors.

2.4.3. Regularization of activation levels

Although the Gabor and autoencoder filters are both followed by a sigmoidal activation function, we further sparsify the activation levels $h_{x,y,k}$ with feature types $k \in \{1..K\}$ in a similar way to local cortical circuitry. We want to constrain the number of active neurons, rather than the mean activation levels. Therefore, we subtract at each image position the average local activation levels. Subsequently a half-wave rectification is applied to constrain the activation levels again to the positive domain with roughly half of the neurons inactivated:

$$\tilde{h}_{x,y,k} = \max\left(0, h_{x,y,k} - \sum\nolimits_{j=1}^{K} h_{x,y,j}\right). \quad (12)$$

Consequently the hard sparseness (Rehn and Sommer, 2007) is artificially increased and these inactivated neurons do not take part in the coupling of phase oscillations (see section 3.1). Thereby the number of possible interactions in the phase simulations is reduced.

As a last step we have to normalize the activation levels at every image position similar to local contrast adaptation in the visual system. We want to make sure that the overall local activation is uniform over the visual field such that an efficient coding of regions of high contrast and regions of low contrast is possible simultaneously. Therefore, we divide all activation levels by the sum of activations over all features at each image location:

$$g_{x,y,k} = \frac{\tilde{h}_{x,y,k}}{\sum_{j=1}^{K} \tilde{h}_{x,y,j}}. \quad (13)$$

As a result we have sparse activation maps with a large proportion of inactive neurons and the same average local activations at all image positions.

3. RESULTS

In a first step we analyse the properties of the activation patterns induced by the natural images (section 3.1). Subsequently we evaluate the correlation statistics of these induced feature activations (section 3.2) and the resulting sparse connectivity pattern (section 3.3). Based on this connectivity pattern we show simulations of the coupled phase oscillator model and the resulting dynamic phase maps (section 3.4). Finally, evaluations of these binding maps are presented based on human labeled segmentation masks (section 3.5).

3.1. SPARSENESS OF ACTIVATION

The simulation of the coupled phase oscillators is based on the activation levels that were generated from natural images. The phase coupling is highly dependent on the type of feature representation that is used to generate the activation levels. The first reason is that the connectivity is based on the correlation between features. The second reason is that also the actual strength of the dynamic coupling is proportional to the current activation levels. Therefore, the statistics of activation plays a crucial role in the formation of the dynamic binding maps.

[1] http://www.di.ens.fr/~mschmidt/Software/minFunc.html

Hand labeled photographs of suburban scenes from the LabelMe database (Russell et al., 2008) are used to generate feature representations with the linear convolutional forward weights followed by a sigmoidal function. The linear convolutional kernels of the Gabor receptive fields contain only one spatial frequency and equally spaced orientations (**Figure 3A**). In contrast, the learned weights of the sparse autoencoder (**Figure 3B**) cover a diverse set of spatial frequencies, colors and orientations.

We compare the activation levels of features obtained with the regular Gabor weights and the autoencoder weights. A very important characteristic of neuronal activations is the level of sparseness. A high level of activation sparseness means that the neuron is most of the time very silent and only rarely very active. This analysis of sparse coding should not be confused with the graph theoretic sparseness which will be analyzed in section 3.3. A qualitative comparison of the activation histograms (**Figure 4A**) shows that the autoencoder activations are sparser compared to the Gabor activations. The phase model is based on the assumption that the activation is restricted to the positive domain. Note that this is in contrast to many normative models of early visual processing which assume a feature code with a Gaussian distribution with zero mean. Furthermore, in our model we are mostly interested in the "hard sparseness" of the activation levels, meaning that the activation is most of the time exactly zero and only rarely very high (Rehn and Sommer, 2007). A comparison with a Gaussian distribution restricted to the positive domain with the same mean (dashed line in **Figure 4A**) reveals that the feature activations after the sigmoidal activation function are not necessarily sparse in the context of a positive distribution with this hard sparseness criteria.

The sigmoidal activation function is followed by the subtraction of mean, rectification and the division by the sum over all features. The resulting histograms of these activation levels (**Figure 4B**) show an increased hard sparseness for both types of receptive fields. These additional preprocessing steps are similar to local regulatory mechanisms in the cortex.

A quantitative evaluation of the sparseness of the activation levels is given by the kurtosis. We use the standard measure of excess kurtosis but without mean normalization because the phase model assumes a non-negative feature coding by activation. Therefore, we evaluate the hard sparseness of feature type j with activation levels $h^{(s)}_{x,y,j}$ by the kurtosis of a zero-centered distribution given by

$$\mathrm{kurt}_j = \frac{\left\langle \left(h^{(s)}_{x,y,j}\right)^4 \right\rangle_{x,y,s}}{\left(\left\langle \left(h^{(s)}_{x,y,j}\right)^2 \right\rangle_{x,y,s} \right)^2} - 3, \tag{14}$$

where $\langle . \rangle$ is the mean over all image positions (x, y) and image samples s from the labelMe database. The estimated median kurtosis over all receptive field types increases for the activations g after the normalization steps described above in comparison to the activations h before the normalizations (**Table 1**). A comparison with a Gaussian distribution, which has a kurtosis of 0, reveals that the additional activation functions indeed increase the sparseness and lead to a leptokurtic distribution of activations. Overall the activations generated by the autoencoder are more sparse in comparison with the hand designed Gabor filters.

The additional activation functions are crucial for the subsequent phase simulations. The mean subtraction and half-wave rectification increase the hard sparseness of activations. This reduction in the number of active neurons leads to a reduction in

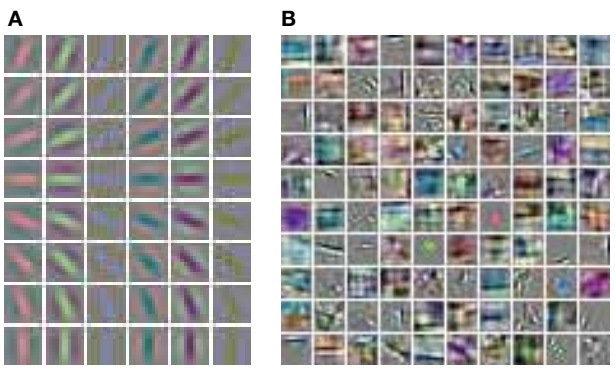


FIGURE 3 | Receptive fields of the feed-forward connections generating the activation levels for the phase simulations. (A) The regular Gabor filters are generated with 8 different orientations and 6 different color channels. **(B)** The convolutional autoencoder weights are learned by optimizing the reconstruction cost, sparseness and weight decay.

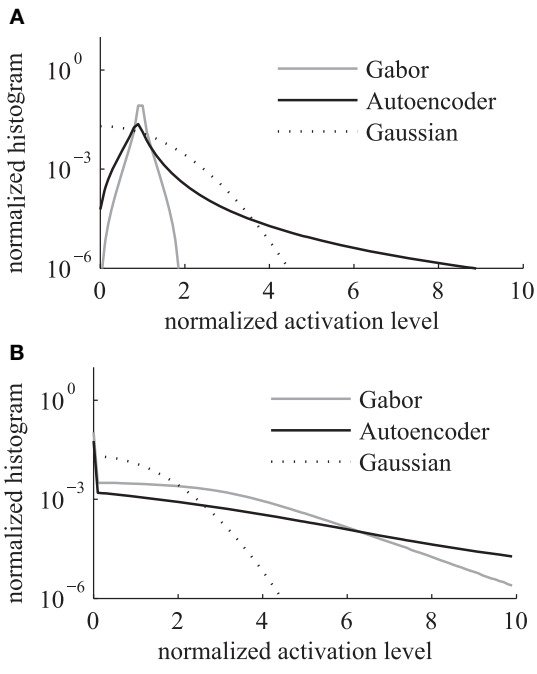


FIGURE 4 | Histogram of activation levels averaged over all feature types. The distributions of activation levels are compared to a Gabor distribution. **(A)** After sigmoidal activation function. **(B)** After mean subtraction, half-wave rectification and division by the sum.

Table 1 | Median kurtosis of feature activations.

	After sigmoid activations $h_{x,y,k}$	After normalizations[1] activations $g_{x,y,k}$
Gabor	−1.96	2.61
Autoencoder	−0.63	11.62

[1] *After the subtraction of mean, half-wave rectification and division by the local sum of the new activation levels.*

the number of active tangential phase connections. Therefore, the features in the input image do not only multiplicatively modulate the strength of the phase interaction but also deactivate many phase connections entirely leading to a completely new effective tangential connectivity pattern.

3.2. STATISTICS OF HORIZONTAL CROSS-CORRELATIONS

The horizontal connections between the coupled phase oscillators are sampled from the cross-correlations of induced activation levels as described in equation 5. Therefore, we describe the horizontal correlations in this section and evaluate the anisotropy of receptive field types. The 4 dimensional cross-correlation tensors $\rho_{x,y}^{(k,m)}$ as defined in section 2.3 are shown in **Figure 5** for 8 feature types. The Gabor receptive fields have a more regular correlation matrix (**Figure 5A**) compared to the learned autoencoder receptive fields (**Figure 5B**). The correlations between the activations of Gabor receptive fields are itself similar to high frequency Gabor functions. In contrast, the receptive fields learned by the autoencoder capture different spatial frequencies and a variety of different colors which is also reflected in the spatial cross-correlations. In both cases the horizontal cross-correlations extend over visual space up to three times the receptive field size. This suggests that the correlations indeed comprise higher-order correlation statistics of the natural images and not only interactions between overlapping receptive fields.

To analyze and compare the correlation tensor of the autoencoder and the Gabor filters, we calculate statistics for different correlation distances in visual space. The indices of the tensor are illustrated in the schematic in **Figure 6A**. For each distance r in visual space we calculate statistics over $\rho_j^{(k,m)}$ where

$$j \in R_r := \left\{ (x,y) \in \mathbb{Z}^2 \middle| r - \frac{1}{2} \leq \sqrt{x^2+y^2} < r + \frac{1}{2} \right\}. \quad (15)$$

The mean absolute value of the cross-correlations decreases for larger correlation distances r as shown in **Figure 6B**. The mean standard deviation of these absolute correlation values over different spatial directions also decreases but with a steeper slope (**Figure 6C**). To make a relative statement about the isotropy in the correlation tensor we also calculate the coefficient of variation over different directions. Therefore, we define the average anisotropy at radius r as

$$\text{anisotropy}(r) := \left\langle \frac{\text{std}_{j \in R_r}\left(\rho_j^{(k,m)}\right)}{\text{mean}_{j \in R_r}\left(\rho_j^{(k,m)}\right)} \right\rangle_{k,m} \quad (16)$$

This mean anisotropy averaged over all pairs of receptive field types has a local maximum at visual distances of around 8–10 pixels (**Figure 6D**). This suggests that the short range phase connections over this distance help more in the synchronization of fine structures. The anisotropy has a local minimum at distances around 15–16 pixels, where more long range phase connections are dominantly used to fill-in segment pixels with similar colors.

3.3. SPARSELY CONNECTED OSCILLATOR NETWORK

The correlation values are used to sample the sparse connections for the simulations of coupled phase oscillators. We restrict the sampled connectivity pattern in simulations of natural scenes to 200 synchronizing and 200 desynchronizing afferent connections per neuron if not stated otherwise. The phase simulations of natural image scenes are run in a network of $200 \times 150 \times 48$ neurons for Gabor features or $200 \times 150 \times 100$ for autoencoder features respectively. Therefore, the percentage of connections that are actually formed compared to all possible connections assuming full connectivity is approximately 0.014% in the case of Gabor features and 0.007% for autoencoder features. Thus, this procedure leads to a very sparse connectivity in comparison to a network of all-to-all interactions.

We evaluate the sampled connectivity based on natural image statistics using graph theoretic measures. The connectivity structure is represented as a graph $G = (H, E)$ as described in section 2.3. We compute the statistics not only over the graph of all connections E but also for the subgraph of synchronizing connections $E^+ := \left\{ e_{x,y}^{(j,k)} \in E \middle| e_{x,y}^{(j,k)} = +1 \right\}$ and the subgraph of desynchronizing connections $E^- := \left\{ e_{x,y}^{(j,k)} \in E \middle| e_{x,y}^{(j,k)} = -1 \right\}$ individually.

For a graph with edges E we calculate the fraction of intra-feature connections as

$$\mu = \frac{\left| \left\{ e_{\delta x,\delta y}^{k,m} \in E \middle| k = m \right\} \right|}{\left| \left\{ e_{\delta x,\delta y}^{k,m} \in E \middle| k \neq m \right\} \right|} \cdot 100\%. \quad (17)$$

The most obvious observation is that the fraction of intra-feature connections is larger for synchronizing connections in comparison to the desynchronizing connections (**Table 2**). The reason is that positive correlations, which are used to sample these synchronizing connections, are stronger between the same feature type shifted over visual space. In contrast negative correlations and thus desynchronizing connections are less likely to occur between the same feature type shifted over visual space. Another observation is that the fraction of intra-feature connections of the Gabor features is roughly twice as large as in the case of the autoencoder features. The reason is that we use 100 autoencoder features and only 48 Gabor features while the total number of sampled synchronizing and desynchronizing connections per feature remains constant.

A more elaborate evaluation of the sampled connectivity of our network can be done using the clustering coefficient and the small-world characteristics (Watts and Strogatz, 1998; Humphries et al., 2006), which are also shown in **Table 2**. To define the local clustering coefficient in an infinite graph $G = (H, E)$, we analyze

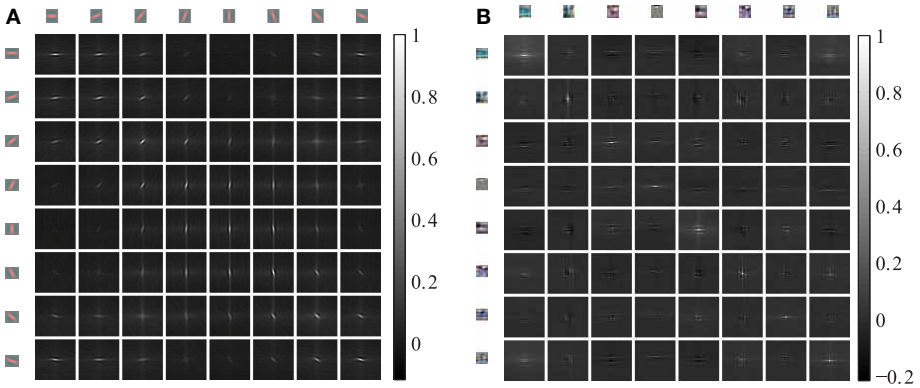


FIGURE 5 | Cross-correlations between different feature activations shifted in visual space. The shown cross-correlations are based on the activation levels induced by natural images. Only a subset of 8 features is shown. The patches on the top and left row show the forward weight matrix of the receptive fields. The other patches show the spatial correlation between these features. The feature weights are shown at the same spatial scale as the shifts in the cross-correlations. **(A)** The correlations between 8 oriented Gabor filters of one of the 6 color channels are shown. **(B)** The correlations between 8 randomly choosen autoencoder features are shown.

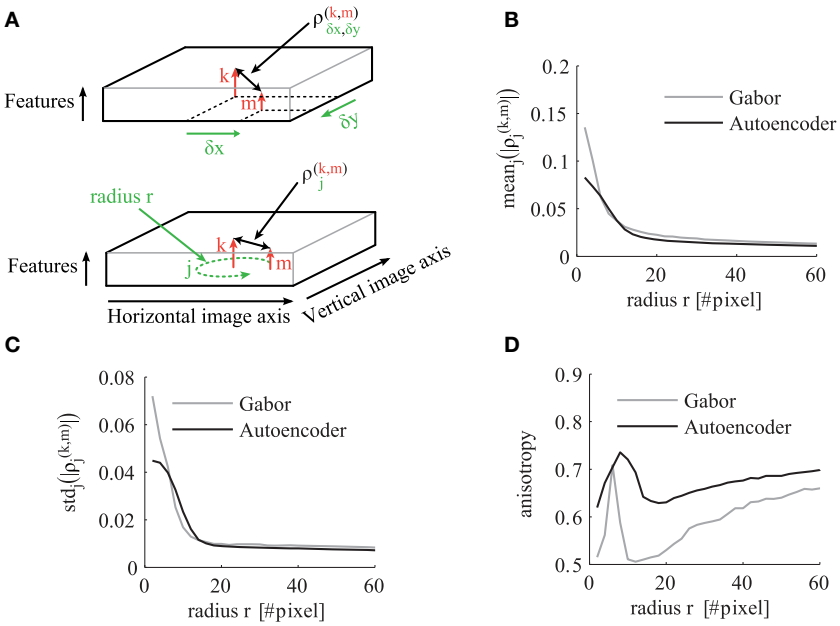


FIGURE 6 | The statistics of the correlation matrix evaluated for different distances *r* in visual space. (A) Schematic to illustrate the indices of the correlation tensor. In the top schematic the correlation tensor is indexed by horizontal (x) and vertical (y) offsets in visual space. In the bottom schematic the correlation tensor is indexed by $j \in R_r$ for a certain distance r in visual space. The other panels compare the correlation tensor of Gabor filters (gray) and autoencoder filters (black) for different distances r. All shown statistics are averaged over all pairs of receptive field types k and m. **(B)** Mean over all directions. **(C)** Standard deviation over different directions for a certain pair of feature types. **(D)** The anisotropy averaged over all pairs of receptive fields as described in the main text.

the connectivity of the neurons in a generic feature column at position $(x, y) = (0, 0)$. We define the neighbors of neuron $g_{0,0,k}$ coding feature type $k \in \{1..K\}$ as the set of all neurons which are directly connected in the graph as

$$N_k = \left\{ g_{x,y,m} \in H \middle| e^{k,m}_{x,y} \in E \vee e^{m,k}_{-x,-y} \in E \right\}, \qquad (18)$$

where we consider outbound ($e^{k,m}_{x,y}$) and inbound ($e^{m,k}_{-x,-y}$) connections of the neuron. Then we define the local clustering coefficient of a feature type k in our network as the fraction of the number of direct connections between neighbors to the number of pairs of neighbors:

$$\gamma_k = \frac{\left| \left\{ e^{m,n}_{x,y} \in E \middle| g_{\tilde{x}+x,\tilde{y}+x,m} \in N_k \wedge g_{\tilde{x},\tilde{y},n} \in N_k \right\} \right|}{|N_k| \cdot (|N_k| - 1)} \qquad (19)$$

We show the global clustering coefficients $\gamma =< \gamma_k >_k$ for our sampled networks comprising only the synchronizing, only

Table 2 | Graph theoretic statistics of the sparse connectivity pattern.

	Gabor			Autoencoder		
	All E	**Sync.** E^+	**Desync.** E^-	**All** E	**Sync.** E^+	**Desync.** E^-
Fraction of intra-feature connections μ	6.49%	12.87%	0.10%	3.12%	6.17%	0.07%
Global clustering coefficient γ (in 10^{-3})	3.21	3.53	0.64	3.18	2.49	1.27
Global clustering coefficient random γ_{random} (in 10^{-3})	2.35	1.11	1.25	2.18	1.03	1.18
Mean shortest path length λ	2.01	2.60	2.46	2.17	2.63	2.59
Mean shortest path length random λ_{random}	2.02	2.53	2.47	2.15	2.62	2.56
Small world index σ_{sw}	1.37	3.09	0.51	1.44	2.39	1.07

the desynchronizing or all connections in the second row of **Table 2**.

The evaluation of the graph comprising all connections shows that the mean clustering coefficient is roughly the same for the Gabor and the autoencoder features. But the evaluation of graphs individually reveals that the clustering coefficient of only the synchronizing graph is higher for the Gabor features in comparison to the autoencoder features. And reciprocally, the desynchronizing connections show a stronger clustering in the case of autoencoder features. An explanation for this difference is that the autoencoder learns a more diverse set of receptive fields by optimizing the reconstruction error. In comparison, the regular Gabor receptive fields cover only predefined colors, spatial frequencies and orientations, which are not optimized to cover a broad range of statistics in the input images. Therefore, the correlation structure in the Gabor activations shows stronger clustering. For comparison, we also show the corresponding clustering coefficients γ_{random} of the equivalent networks with the same connection lengths (measured in pixel distance) but rotated by random angles and connected to random features.

We can further use the small-world index to measure the capability of neurons in our network to reach other neurons via a small number of interaction steps. The small-world index is a quantitative definition of the presence of abundant clustering of connections combined with short average distances between neuronal elements, proposed by Humphries et al. (2006). It can characterize a large number of not fully connected network topologies. The connectivity within the 3-dimensional grid of our model is sampled such that it is invariant to shifts in the two image dimensions. Therefore, we have to slightly adapt the small-world index for our infinite horizontal sheet consisting of feature columns with identical connection patterns. We use the definition of the small-world index

$$\sigma_{\text{sw}} = \frac{\gamma/\gamma_{\text{random}}}{\lambda/\lambda_{\text{random}}}, \tag{20}$$

where the shortest path lengths λ and λ_{random} measure the number of network hops needed to connect two neurons within our sampled network and a random network respectively. We use the average over all shortest path lengths between all pairs of neurons within one feature column. A network graph must have a small-world index σ_{sw} larger than one to meet the small-world criteria. The evaluations show that the graph comprising the synchronizing connections exhibits small-world properties while the desynchronizing connections are closer to a random connectivity and do not exhibit small-world properties (**Table 2**). The small-world property might be helpful in the synchronization of distant neurons.

3.4. PHASE SIMULATIONS

The resulting connectivity pattern is used in the phase simulations. All shown simulations of the coupled phase oscillator networks are initialized with random phase variables. The activation levels are only set once in the beginning and remain the same throughout the phase simulations. During the simulations attractors are formed in the phase space and are localized in certain image regions.

A simulation of the coupled phase oscillator model with localized connectivity and with uniform activation levels shows that pinwheel structures will form in the phase map (**Figures 7A,B**). The connectivity length in the network determines the scale of the pinwheels. During the simulation these pinwheels attract each other and annihilate (Wolf and Geisel, 1998). The probability of the formation of pinwheels decreases for network connectivity patterns that are less locally dense but more sparse and spread out.

In the next simulations we use several feature types to encode different aspects of the input images. To visualize the resulting 3-dimensional structure of phase variables $\varphi_{x,y,k}$ we calculate the circular mean at each image position weighted by the corresponding activation levels:

$$\varphi_{x,y}^{\text{avg}} := \arg\left(\sum_k g_{x,y,k} e^{i\varphi_{x,y,k}}\right), \tag{21}$$

where arg is the complex argument. We show the average phase variables $\varphi_{x,y}^{\text{avg}}$ coded as color hue to visually represent the circular structure of the phase.

We use two simple artificial stimuli to demonstrate the basic function of the phase simulation in the presence of structure in the activation variables (**Figures 7C,D**). The stimuli of these simulations are artificially generated grayscale images containing bar segments and circle segments (insets in **Figures 7C,D**). The connectivity in both simulations is based on Gabor receptive fields with horizontal connectivity obtained from statistics of natural images. In the simulation of two collinear aligned bars the phase of the neurons coding the two bars are synchronizing although the

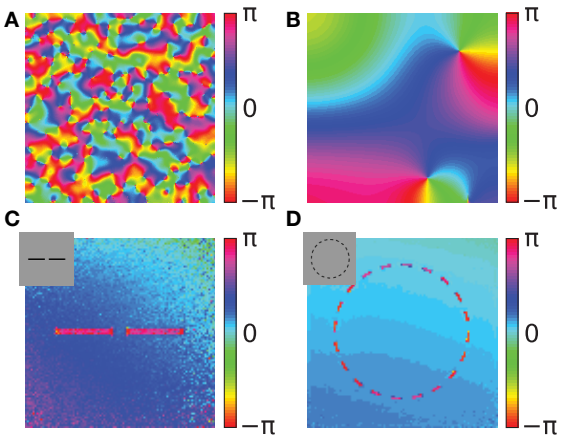


FIGURE 7 | Simulations of the coupled phase oscillator model. The phase variables are shown as color hue. **(A,B)** Phase simulations of a 2-dimensional grid of 200x200 neurons with uniform activation levels after 100 network iterations. A regular local connectivity is used with maximum length 3 **(A)** and 10 **(B)**. Simulations in **(C,D)** use the Gabor receptive fields and horizontal oscillator connectivity from the correlation statistics. The input images are artificial stimuli of two collinear aligned bars **(C)** and of a dashed circle **(D)** and are shown as an inset in the upper left corner. The shown phase maps in **(C,D)** are the circular mean of the phase variables weighted by the activation of the corresponding features. Panel **(C)** shows the average phase after 20 network iterations and panel **(D)** after 40 network iterations.

two bars are not directly connected in the image (**Figure 7C**). This suggests that the simulation can implement Gestalt laws of grouping, because neurons are grouped together by having the same phase value. Specifically, a human observer could interpret these two bars as one single continues line. Therefore, the simulation can be interpreted as implementing the Gestalt law of continuity because the neurons that are coding the two bars have the same phase. Please note, that in the simulation the gap between the two bars is not filled in because our model does not incorporate any feedback from the phase variables to the activation variables. In this study we focus on relational coding by phase variables and therefore neglect any recurrent dynamics in activation variables.

The other simulation uses a dashed black circle as input (**Figure 7D**). The phase map shows that all segments of the circle are synchronizing to the same phase value. The synchronized state of the circle means that the phase variables at different segments of the circle code the global attribute and bind the individual circle segments together. Similarly, humans usually perceive the circle segments all together as one single object. This indicates that the phase simulation can also implement the Gestalt law of closure. Depending on the initialization of random phase variables, cases exist where the circle does not synchronize to one coherent phase but forms a continuous phase progression one or multiple times from 0 to 2π. On one hand these simulations reproduce the previous studies demonstrating binding properties of coupled neural oscillators. On the other hand, in these simulations the connectivity is learned based on natural stimuli and not hand crafted. Hence, it demonstrates that these Gestalt properties are learned from the statistics of natural stimuli.

We next evaluate the concept of binding by synchrony also on natural visual scenes. All following simulations in this paper use color images from the LabelMe database (Russell et al., 2008) and either the Gabor filters or the autoencoder filters to generate the activation levels for the network. An example of a suburban scene is shown in **Figure 8A** with the corresponding human labeled segmentation masks in **Figure 8B**. We use the time constant $\tau = 1/3$ for the simulations based on Gabor filters and $\tau = 1/30$ for the simulations based on autoencoder filters. These values were chosen such that per iteration of the classical Runge-Kutta solver the phase of not more than 1% of all neurons changes more than $\pi/2$. The units of these time constants are arbitrary because our model of coupled phase oscillators describes the change in phase independent of the oscillation period. Examples of the resulting phase maps are shown in **Figure 8C** for Gabor activations and **Figure 8D** for autoencoder activations. The phase maps of simulations using autoencoder weights are blurred compared to the Gaborfilters because the peak of the receptive fields are not necessarily centered within the convolutional weight matrix, leading to shifts in visual space between different feature maps at segment boundaries. Yet in both examples an intuitive segmentation of the original can be recognized again in the distribution of phase values. We see a constantly increasing phase synchrony in labeled segments. This example suggests that high-level image objects are likely to synchronize to a coherent phase.

3.5. EVALUATION OF PHASE MAPS

We evaluate the simulated dynamic phase maps and compare them with human labeled binary segmentation masks of high level image objects from the LabelMe database. We begin with an evaluation of the resulting phase maps independently from the labeled image masks to show global properties of the coupled phase oscillator model and the influence of the number of horizontal connections (section 3.5.1). This is followed by an evaluation of the phase synchrony within labeled segments with respect to the surrounding of the segments (section 3.5.2). Finally a local evaluation of the phase maps at the boundaries of labeled segments is presented (section 3.5.3).

3.5.1. Phase synchrony

Segmentation and binding of neurons in the network can only be achieved if the phase variables are not random but also not completely synchronized. Therefore, we will first evaluate the local phase synchrony independent of segments in the image. We define the synchrony in a population M of neurons as

$$p_M = \left| \frac{\sum_{m \in M} g_m \cdot e^{i\varphi_m}}{\sum_{m \in M} g_m} \right|, \tag{22}$$

where M is defined as a set of 3-dimensional indices describing the position of the neurons.

In this section we analyze the simulation shown in **Figure 8** in more detail and evaluate how the number of synchronizing and desynchronizing connections effects the phase synchrony. We evaluate the local phase synchrony at image position (x, y) for a

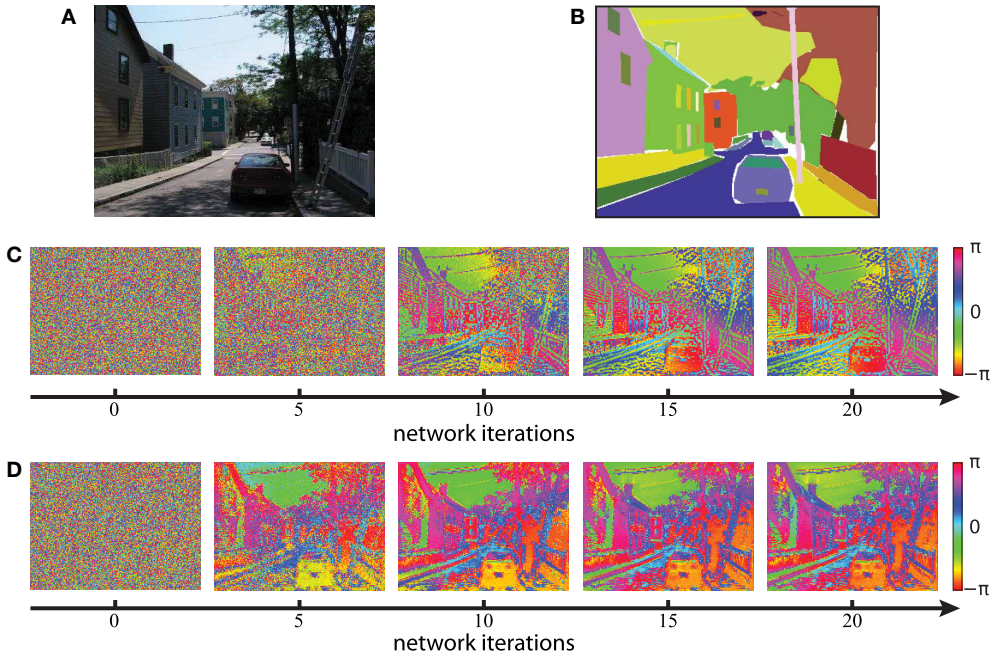


FIGURE 8 | Phase simulations of a natural image of a suburban scene. **(A)** A natural image from the LabelMe database is used as the input to generate neuronal activation maps. **(B)** The LabelMe images are accompanied by overlapping segmentation masks of labeled image regions. **(C,D)** The circular mean of the phase maps evaluated at different network iterations. Gabor filters were used in **(C)** and autoencoder filters in **(D)**.

certain radius r by calculating $p_{M_{x,y,r}}$ for neurons at positions

$$M_{x,y,r} = \left\{(\tilde{x},\tilde{y},k)\middle|(x-\tilde{x})^2+(y-\tilde{y})^2 < r^2,\right.$$
$$\left.(x,y) \in \mathbb{N}^2, k \in \{1..K\}\right\}, \quad (23)$$

where K is the number of feature maps. We average this quantity over all possible image positions (x, y). This mean local phase synchrony is shown in **Figure 9** for simulations using different number of connections, different iterations and for different radii r.

When the network has reached a steady state, the mean local phase synchrony depends on the number of synchronizing and desynchronizing connections (**Figures 9A,D**). The number of synchronizing connections increases the average local phase synchrony. In contrast, the number of desynchronizing connections can increase or decrease the average local phase synchrony depending on the number of synchronizing connections. At first sight, this may be counterintuitive. In the case of few synchronizing connections, the desynchronizing connections repel the associated phase variables from each other. This ultimately leads to a clustering in the circular phase space evoked by desynchronizing interactions. In the case of more synchronizing connections, the main force driving the network are attractor states and therefore desynchronizing connections decrease the overall phase synchrony.

The phase synchrony in the steady state condition increases with the ratio between synchronizing and desynchronizing connections up to a ratio of 16 times more synchronizing than desynchronizing connections (**Figures 9B,E**). Interestingly, the phase synchrony in the steady state condition decreases again in simulations with more than 800 synchronizing connections and very few desynchronizing connections. During the transient phase a very low or high ratio leads to a faster convergence to a more synchronized state. The slowest convergence is achieved at the cases with 4 times more desynchronizing connections or when the number of synchronizing and desynchronizing connections is balanced.

The phase simulations show synchronization behavior at a large variety of different spatial scales (**Figures 9C,F**). The level of synchrony at the steady state decreases for increasing radius of the phase synchrony evaluation. At all spatial scales the time to reach the steady state synchrony level is roughly the same. Only very localized regions over 1-2 pixel distances show a slightly faster convergence to the final phase synchrony level. When not otherwise stated we select in all simulations and evaluations an intermediate parameter range with balanced synchronizing and desynchronizing connections leading to rich dynamics. These standard parameters are marked with blue circles in **Figure 9**.

3.5.2. Segmentation index

The dynamic binding and segmentation of the simulated phase maps of natural images are evaluated using hand labeled segmentation masks. Here a baseline is necessary to accommodate for the higher probability of synchronization between neurons that are close by. Consequently we use the labeled image masks on the corresponding simulated phase maps and compare them to a baseline using the same image masks on simulations of different non-matching images.

The segmentation masks in the LabelMe database are specified as polygons on the images that are initially reduced in our

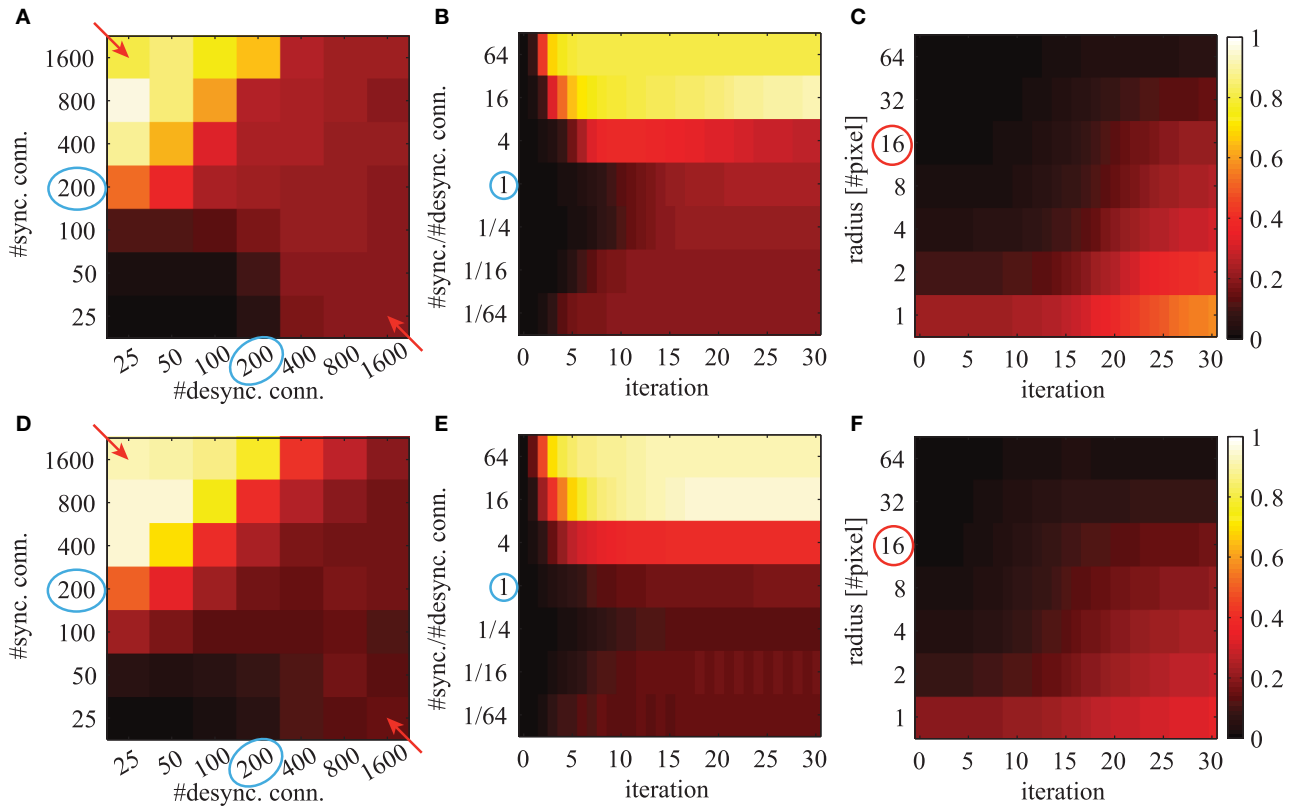


FIGURE 9 | The averaged local phase synchrony in circular image regions for different simulation parameters. All evaluations are based on the activation levels obtained from the image shown in **Figure 8A**. Simulations in the top row **(A–C)** are based on Gabor weights; simulations in the bottom row **(D–F)** are based on autoencoder weights. Blue circles indicate the standard parameters for subsequent evaluations. Colorbars of all panels are the same and shown on the right. The panels in the left column **(A,D)** show the phase synchrony after 20 iterations for different number of excitatory and inhibitory phase connections per neuron. The panels in the center column **(B,E)** show the phase synchrony for different ratios of excitatory to inhibitory connections as a function of network iterations. These ratios correspond to the diagonal elements marked with red arrows in panels **(A,D)**. And the shown time course of the average phase synchrony values are from the same simulations. In the right panels **(C,F)** the phase synchrony is shown for different sizes of the local circular region of the evaluations. The red circle indicates the radius which was used in the evaluations shown in the other panels.

simulation to a resolution of 400 × 300 pixels. The convolutional forward projections lead to a further reduction in the feature representation to a grid of 200 × 150 pixels. Therefore, we restrict the evaluations of the phase maps to segmentation masks which contain at least as many pixels as the specified patch size of the forward projections (6 × 6 neurons corresponding to 12 × 12 pixels in the input image). In addition, segments occupying more than half of the respective images are excluded to allow evaluations against a baseline synchrony of the surrounding regions. The range of labeled segments which is used in our evaluations is shown as a horizontal bar in **Figure 10**. Only in evaluations where the segment sizes are explicitly stated, we also evaluate these otherwise excluded very small and very large segments.

The number of labeled segments in the database decreases for larger segment sizes (**Figure 10A**). Yet the total area occupied by segments in the different bins increases for larger segment sizes (**Figure 10B**). Therefore, when applying labeled masks to non-matching images small segments are highly likely to fall into large segments where a large number of tangential connections is functionally active. Consequently the phase synchrony within labeled segments is not a sufficient baseline for an unbiased comparison with simulations of non-matching images. Therefore, we need a baseline to control for the unequal distribution of segment sizes and their occupied region in the images.

To accommodate for the statistics of segment sizes in the evaluation of the matching and non-matching natural scenes, we define a segmentation index (**Figure 11**) that sets the phase synchrony in segments into the context of the surrounding neurons. Concretely, the segmentation index evaluates how the phase of neurons inside of segments is more or less synchronized compared to the synchrony of random neurons inside and outside of the segment. The neighborhood N of a segment Q is generated using a diamond shaped grow operation on the segmentation mask repeatedly until the number of neurons in N is doubled compared to the original segment Q. Therefore, N is the union of the segment Q and the surrounding R of the segment (Q and R are annotated in the example shown in **Figure 11**).

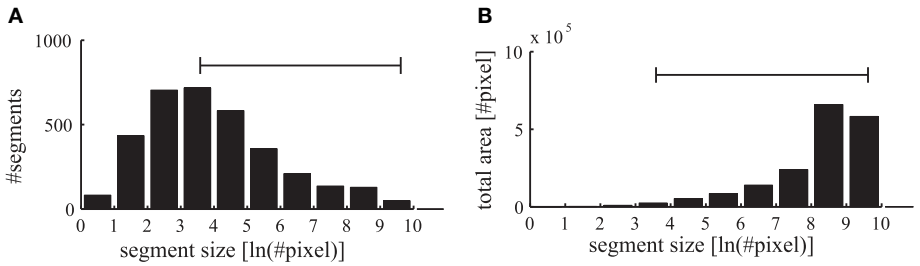


FIGURE 10 | Statistics of labeled image segments. (A) The histogram of evaluated segments from the LabelMe database for different segment sizes is shown. **(B)** The total area occupied by the segments in the corresponding bins. The range of segment sizes (36–15000 pixels) that are used for subsequent evaluations are marked with a horizontal bar.

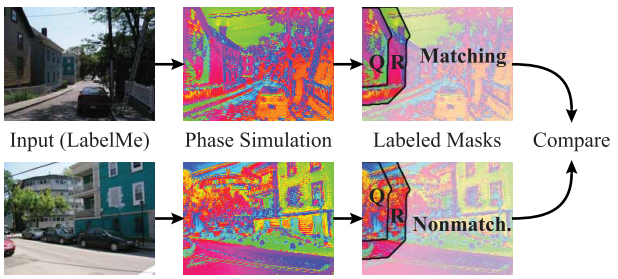


FIGURE 11 | Evaluation using hand labeled image masks. The evaluations compare the segmentation index of matching simulations and segmentation masks **(top row)** to a baseline of non-matching simulations and segmentation masks **(bottom row)**. The images from the LabelMe database (left column) are processed using the forward projections. The resulting features are used to simulate the phase of the coupled neural oscillators **(middle column)**. The segmentation index of these phase maps are then evaluated using the segmentations masks from the LabelMe database **(right column)**. The evaluation of the house in the top left is here shown as an example. The segmentation index compares the phase synchrony in the hand labeled region of the house (Q) to a baseline phase synchrony within the neighborhood ($Q \cup R$).

We calculate the phase synchrony values p_{Q_j} and p_{N_l} for random subsets $Q_j \subset Q$ and $N_l \subset N$ where $j, l \in \{1, \ldots, 100\}$ and $Q_j,\ N_l \in \mathbb{N}^{1000}$. We define the segmentation index of segment Q as the difference between the mean synchrony within the segment Q to the mean synchrony in the neighborhood $N = R \cup Q$:

$$\kappa(Q, N) = \langle p_{Q_j} \rangle_j - \langle p_{N_l} \rangle_l. \tag{24}$$

The segmentation index increases over simulation iterations for matching and non-matching masks and images (**Figure 12A**). The matching conditions have a steeper ascent and reach a higher segmentation index compared to the non-matching conditions. The difference between the matching segmentation index and the non-matching segmentation index increases for both simulations using Gabor weights and autoencoder weights (**Figure 12B**). The simulations using regular Gabor receptive fields show larger differences between matching and non-matching segmentation indices compared to the autoencoder weights. The ratio between matching and non-matching segmentation indices is roughly the same for both types of receptive fields. This demonstrates systematic binding in the phase maps of matching segments.

An evaluation for different segment sizes individually reveals more differences between the Gabor and autoencoder features. The evaluations of the matching conditions show that the segmentation index increases for larger segments in the case of the autoencoder features but decreases for larger segments in the case of the Gabor features (**Figure 12C**). An explanation is that the autoencoder contains more features with low spatial frequencies while the Gabor features are restricted to one specific spatial frequency.

The paired difference between matching and non-matching evaluations shows that the Gabor filter and the autoencoder have roughly the same performance for large segment sizes (**Figure 12D**). For small segment sizes the autoencoder has a decreased segmentation performance. One possible explanation might be that the receptive field weights are not centered (compare **Figure 3**) and therefore different feature neurons might be slightly misaligned relative to the hand labeled segmentation masks, which are defined as polygons with arbitrary precision on the image.

Overall the results show a significant difference between the matching and the non-matching segmentation indices for all evaluated segment sizes. The paired difference between the matching and the non-matching conditions increases as the simulation of the randomly initialized phase variables slowly converges to a state with clusters in the circular phase space. After about 20 network iterations the paired difference in the segmentation index reaches a high plateau. Therefore, the coupled phase oscillator model achieves a stable segmentation of the natural image scenes with a coding of binding by synchrony.

3.5.3. Segment boundaries

To evaluate how well the phase maps segment different labeled regions at their borders we calculate a metric at random locations of segment boundaries. We sample 50 random locations from all boundary lines of the segments in each simulated image from the LabelMe database. At these locations we use the angle of the segment boundary to divide a local region into two semicircles with a radius of 10 pixels such that one half lies approximately within the segment and the other half outside of the segment (**Figure 13A**). The mean phase difference between both semicircles decreases

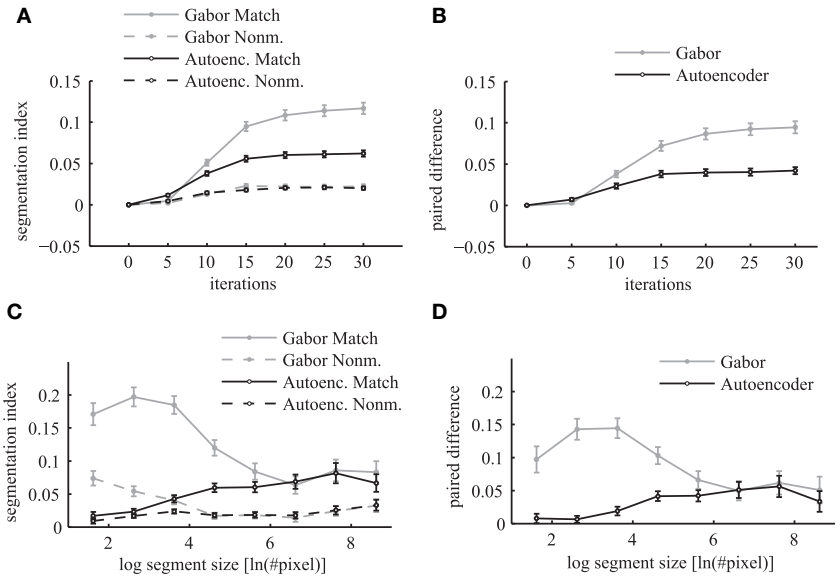


FIGURE 12 | Segmentation index. The mean segmentation index is shown as a function of network iterations averaged over all segments with more than 36 pixels in the top panels **(A,B)**. The segmentation index is shown as a function of different segment sizes after 20 network iterations in the bottom panels **(C,D)**. The panels on the left side **(A,C)** show the evaluations for matching images (solid lines) and non-matching images (dashed lines) individually. Panels on the right side **(B,D)** show the paired difference between matching and non-matching evaluations. In all panels the activation levels are obtained using Gabor filters (gray lines) and autoencoder filters (black lines). The errorbars in all panels are 95% confidence intervals.

over simulation time (**Figure 13B**). The paired difference between the phase difference in matching compared to non-matching images shows that the phase difference over matching segment boundaries is significantly larger (**Figure 13C**).

The evaluation of the phase difference as a function of the size of this circular region shows that the segmentation performance using autoencoder features decreases for very small regions (**Figures 13D,E**). This might be due to the above described misalignments between the learned receptive field centers. For very large evaluation regions the performance decreases for both receptive field types because the circular regions are likely to extend beyond the hand labeled segment regions.

It is possible to evaluate the segmentation performance of the dynamic binding maps without the need for a baseline on non-matching images if we use an unbiased performance estimator with a clearly defined chance level. Therefore, we measure how well the phase map can predict the angle of the borders of segmentation masks. We use the phase variables at randomly sampled locations on segment boundaries (**Figure 13A**) and compute the image direction with the largest change in the phase variables. We define the local variance in phase at image position (x, y) as

$$\vartheta_{x,y} = 1 - \frac{1}{5 \cdot K} \cdot \left| \sum_{k=1}^{K} e^{i\varphi_{x,y,k}} + e^{i\varphi_{x-1,y,k}} + e^{i\varphi_{x,y-1,k}} + e^{i\varphi_{x+1,y,k}} + e^{i\varphi_{x,y+1,k}} \right| \quad (25)$$

where the sum is over all $k \in \{1..K\}$ feature maps. We use the structure tensor of the local variance in phase to estimate the principal directions. To compute the structure tensor we use a Gaussian window function with a standard deviation of 3 pixels and the second order central finite difference of the local variance in phase. The eigenvector of the structure tensor gives an estimate of the border direction of the segmentation mask. The evaluation of the phase maps shows that the mean error in the estimation of the boundary angles decreases over simulation time (**Figure 13F**). A minimum is reached after around 20 network iterations with an error of approximately 28° in comparison to the chance level of 45°. This demonstrates that the phase gradient systematically aligns itself orthogonal to the segment boundaries.

4. DISCUSSION

Here we investigate the concept of binding by synchrony, as has been previously studied with abstract stimuli, in the context of unsupervised learning and natural stimuli. The model consists of coupled phase oscillators with a connectivity based on natural image statistics. Specifically, the correlation of neuronal activity governs the structure of local horizontal connections in the network. Hence the connections are not constructed according to a heuristic or intuition, but solely data driven. Therefore, we can expect it to generalize well to other cortical areas. We show that the sampled sparse connectivity based on positive correlations in induced activations by natural stimuli exhibits small-world properties. We hypothesize that the small world property is a signature of Gestalt laws in the form of regular local correlations (objects) that can be flexibly combined on a global scale. We show that these horizontal connections influence the dynamics of the phase variables such that an effective coding of contextual relationships between active neurons is implemented by phase

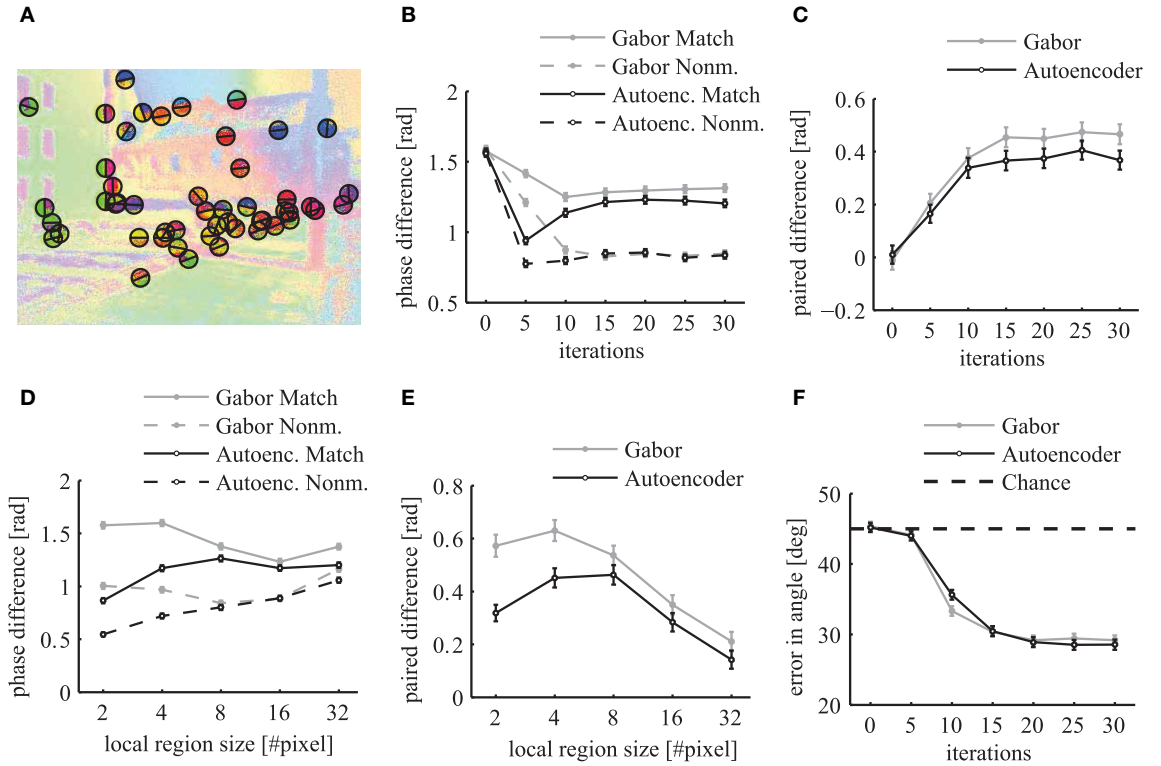


FIGURE 13 | Local evaluation of the phase segmentation. Results were obtained using Gabor filters (gray) and autoencoder filters (black). **(A)** Illustration of the randomly selected locations on segment borders and the corresponding semicircles as described in the main text. **(B)** The local phase difference at random segment border locations of matching images (solid lines) and non-matching images (dashed lines). **(C)** The paired difference between the local phase differences evaluated on matching and non-matching images. **(D)** The mean local phase difference as a function of different sizes of the local circular regions over which the phase is evaluated. **(E)** The paired difference between matching and nonmatching images. **(F)** The mean error in the estimated angle of segment boundaries. All errorbars are 95% confidence intervals.

synchronization. Therefore, our results reveal that the concept of binding by synchrony is viable for natural stimuli.

The evaluation of phase synchronization as a code for grouping and segmentation utilizes hand labeled image segments, corresponding to high level objects, as ground truth. The evaluations reveal that the phase maps are binding active neurons together if they encode different attributes of the same stimulus. It follows that the phase variables are coding global stimulus attributes in contrast to the coding of local stimulus attributes by the rate variables. The coding of these global contextual relationships is not directly influenced by the rate variables but only by their indirect modulation of the phase interactions. Furthermore, we illustrate that discontinuities are formed in the phase maps at the borders of segments and that these discontinuities can predict the orientation of segment boundaries. Therefore, our results suggest that the segmentation driven by bottom up dynamical processes using natural image statistics matches to a certain degree the top-down labeling of abstract image objects.

Our study connects three different subject areas: natural image statistics, dynamical models of neural networks and normative models of sensory processing. In the following we will discuss the motivations and implications of our study from each of these perspectives.

4.1. CHOICE OF NATURAL STIMULI

The choice of "natural" stimulus material is not as obvious as it might seem. A more natural choice from a biological perspective would be to use stimulus material generated by a moving agent. For example videos from a camera mounted to a cat's head were used previously to analyze the spatio-temporal structure of natural stimuli (Kayser et al., 2003). A similar setup from a human perspective is also possible (Açik et al., 2009). But time variant stimuli require more computational resources and the high number of horizontal connections in our simulations is computationally expensive although it is implemented as a vectorized operation. In addition, the analysis of the phase segmentation maps would be more difficult in the case of moving stimuli because of the unknown time lag between stimulus onsets and the resulting dynamic phase maps. Therefore, we decided to not use videos as stimulus material in the present study.

Differences in eye movements given different stimulus classes might also play a role in shaping the statistics in the visual input received by the primary visual cortex. There might be important interactions between saccadic eye movements and the dynamics of the horizontal connections in the visual cortex. One could simulate saccadic movements on static images using saliency maps and use the resulting images for the feedforward processing in

our model. But as with moving stimuli in general it would complicate the analysis and would not contribute directly to the understanding of the central questions of binding by synchrony.

The LabelMe database provides a large set of only static images. It has the advantage that the images are accompanied by labeled region masks of well defined objects. These high level labeled masks are often overlapping in the case of part-based segmentations of objects. The segmentation evaluation is tricky in the case of occluded objects. But the LabelMe database allows us to investigate the relationship between natural image statistics and the coding of high level image concepts. Therefore, we think it is a reasonable choice to use this database in our study.

4.2. BIOLOGICAL PLAUSIBILITY

As with most computational neural network models we have to ask ourselves in how far it is biologically plausible. To advance our knowledge about the underlying computation principles in the cortex, it is always a good choice to model only the level of detail which is necessary to explain the phenomena under investigation. Thereby we assure that the abstraction level of the model is as good as possible although it is very likely that some mechanisms below the level of detail modeled here play an important role in synchronization phenomena. We implement in our simulations the influence of correlated neuronal activity on large time scales to the network connectivity. Based on these connections we show how the dynamics on fast time scales can code for segmentation and binding. Therefore, we have to model the behavioral learning time scales (>days) to capture the natural image statistics and the dynamical network time scales (<seconds) simultaneously. Therefore, we consider the chosen network architecture of segregated rate and phase based coding suitable to investigate the role of correlated neuronal activity on the network dynamics and relational coding by synchronization.

The Kuramoto model restricts the dynamical interactions between coupled oscillators to a scalar phase variable. Breakspear et al. (2010) review this simplified model of coupled phase oscillators in the context of models of complex neurobiological systems. They find that it captures the core mechanisms of neuronal synchronization and a broad repertoire of rich, non-trivial cortical dynamics. Studies of the Kuramoto model mostly focus on regularly defined phase interactions without a separate network variable representing the activation levels of the oscillator neurons. This allows using mean-field approximations to further simplify the analysis of the Kuramoto model. In contrast, our study focuses on the simulation of heterogeneous connections which are modulated by heterogeneous activation levels induced by natural stimuli. Therefore, our simulation model is more similar to the diverse activations and connections found in biological neural systems but this comes with the drawback that a mean-field approximation is not warranted.

In principle two biological interpretations of the coupled phase oscillator model are possible. A conservative standpoint is an interpretation as a neural field model in which each network unit of our simulation represents a functional module, i.e., a cortical column, which is comprised of many biological neurons. In this case the phase variables would represent the average phase of a set of biological neurons, i.e., the phase of the local field potential. A second possible more fine-grained interpretation in which the phase oscillators represent individual biological neurons might seem far-fetched and oversimplified on first sight. Nonetheless the interpretation of the phase variables as spike timings might give further ideas about possible extensions of our proposed model. In this interpretation the oscillators represent the limit cycles of the dynamics of spike generation of biological neurons. The sinusoidal interaction function can then be related to an integral over the phase response function of a spiking neuron (Sturm and König, 2001). Furthermore, the spike interpretation could motivate the introduction of conduction delays in our model. This in turn might further allow studying spike-timing dependent plasticity in the context of a normative model.

Certainly, there are many phenomena that can only be modeled by more detailed spiking neuron models. For example spike-timing dependent plasticity could only be modeled with the phase oscillator model if we assume regular oscillatory firing but not in the case of irregular firing. For example, the ability of self-organizing recurrent networks (SORN) to learn spatio-temporal structures in the input depends on spike-timing dependent plasticity and irregular firing (Lazar et al., 2009). Similarly, Buonomano and Maass (2009) showed that spatiotemporal processing of natural stimuli can emerge from the dynamics of "hidden" neuronal states, such as short-term synaptic plasticity. Irregular firing is also needed for synfire chains of successively activated neural assemblies to explain the physiological measurements of spike patterns recurring with millisecond precision (Abeles, 1982). However, it might be possible to simulate some properties of synfire chains if we add more hierarchical layers and phase conduction in the feed forward projections in our model. Kumar et al. (2010) analyzed the coexistence of firing rate propagation and synchrony propagation in feed forward networks. Last but not least, self-organized criticality and cortical avalanches (Beggs and Plenz, 2003) can probably only be modeled with more detailed spike-based neuron models because the phenomenon requires a dynamical system of more complex coupled oscillators.

There are also other dynamical models of neural networks that were analyzed in the context of scene segmentation (Tononi et al., 1992). Wang and Terman (1997) described the local excitatory global inhibitory oscillator network (LEGION), which is comprised of units described by two differential equations that explicitly model a stable periodic orbit alternating between two phases with rapid transitions between them. This model has the advantage that fast synchronization of the coupled oscillators is possible. But it simulates each neuronal oscillation on a fast timescale and the synchronization of a population of neurons is only visible at certain simulated time points. In contrast, our phase model simplifies the phase plane to a continues phase variable averaged over many oscillatory periods, so that the phase relationships between all pairs of neurons is explicitly represented at all simulation time points. Another difference is that the implementation of LEGION involves many discontinuous operations to reduce the computation time. These discontinues operations prevent a normative model approach with optimizations using gradient descent. The full continuous dynamics in our model allows further optimizations of the horizontal connectivity using gradient descent methods.

In our model the forward connections are computed once and are then fixed during the phase simulation of horizontal connections. This is a very simplified model compared to the ongoing simultaneous processing of afferent and recurrent inputs in the cortex. But it is compatible with the fact that self sustained activity in the cortex can be measured also in the absence of stimulus inputs. Furthermore, computational models of cellular and network behavior support the conclusion that the cortical network operates in a recurrent rather than a purely feed-forward mode (Mariño et al., 2005). Therefore, it makes sense to simulate the lateral interactions decoupled from the time scale of forward projections that generate the activation levels.

We use the correlated neuronal activation levels as the probability to form horizontal intralayer connections. It was shown that the measured horizontal connectivity in the visual cortex of cats is indeed proportional to the correlation between receptive field wavelets in image statistics (Betsch et al., 2004). Our choice to use a sparse connectivity pattern instead of full connectivity with heterogenous connection strengths was initially intended as a computational shortcut to allow large-scale simulations. This sparse connectivity is in line with biological horizontal connectivity and reveals interesting properties that deserve further investigation. In the brain the binding of stimulus representations has to be distributed over many cortical areas. It was shown with graph theoretic measures that the sparse connectivity within the cortex is organized in hubs and shows properties of small-world networks (Sporns et al., 2004). One can speculate that this allows binding by temporal structure even between stimulus representations over distant cortical regions. Also in our network model the sampled sparse connection patterns generated from correlated neuronal activity were shown to have small-world properties in the case of synchronizing connections. Accordingly, we see in our network simulations fast synchronizations of distant neurons that are not directly connected. And in future studies our model could be extended to simulate even synchronizations between different cortical regions.

In the cortex a wide range of oscillatory frequencies at different spatial scales occur with cross-frequency couplings. This is highly prominent in different sleep stages (Belluscio et al., 2012) and plays an important role in memory encoding (Friese et al., 2012). Our model is highly simplified in the sense that all neurons are assumed to have the same oscillatory natural frequency. We simulate only horizontal connections between neurons with similar physiological properties which are operating in the same dynamical regime. In this context, the assumption that all active neurons are close to a similar dynamical limit cycle seems reasonable. In future work, several cortical rhythms could be implemented using several phase variables per neuron. One can conceive different algebraic structures which could efficiently represent cross-frequency couplings in the cortex. This would allow investigating fractal binding at different abstraction levels and segmentation at different scales.

In summary, the architecture of our model captures many important aspects of biological neural networks. In particular, it models the dynamical properties used for contextual coding and the unsupervised learning of statistics in natural stimuli. At the same time, our model keeps the simplicity required for the analysis of the network dynamics and allows relatively simple evaluations of the resulting phase relationships.

4.3. COMPARISON WITH OTHER NORMATIVE MODELS

In recent years the abstraction from complex differential equations describing biological neural networks to normative models of rate-based sensory processing improved our knowledge on the underlying computational principles of the cortex (Olshausen and Field, 2005). Unsupervised learning of the inherent statistics in the sensory input seems to be one of the main mechanisms governing the structural connectivity between neurons in low level sensory areas of the cortex (Olshausen and Field, 1996; Wiskott and Sejnowski, 2002; Körding et al., 2004). On the other hand relatively few studies have investigated the relationship between unsupervised learning using correlated neuronal activity and the coding of contextual relationships through binding by synchrony. In this section we describe differences and similarities between our model and other normative models of sensory processing in the brain.

Wyss et al. (2006) and Franzius et al. (2007) show that rate-coding neurons form a hierarchy of processing stages resembling the ventral visual pathway. These studies use optimization functions of optimal stability and decorrelation while exposing the network to natural stimuli. Although these models provide important insights into the information processing mechanisms in the cortex, they don't take into account the processing of contextual information and lack an implementation of relational coding between different features. In a similar way to these studies, we use the statistics of natural stimuli not only to learn feature representations but also to explain relational coding in the context of binding by synchrony. This approach could allow combining multi-scale image segmentation and object recognition into a hierarchical neuronal network model. A prerequisite for analyzing the segmentation by synchrony in a hierarchical network is an unsupervised learning of the feed-forward connections to generate the activation levels for higher network layers. We have shown that the proposed segmentation by synchrony works with receptive fields obtained from convolutional autoencoders, which can be stacked to obtain the forward and backward connections within a hierarchy. This allows a completely unsupervised learning of feed-forward, feed-back and intralayer connections using natural image statistics. Binding and extraction of features can be accomplished simultaneously within the hierarchy.

Biologically inspired autoencoder models were shown to be efficient for unsupervised learning of receptive fields by minimizing the reconstruction error of the input (Coates et al., 2010). Complex valued autoencoders have similar to our model 2 variables per network node (Baldi and Lu, 2012). To our knowledge the available publications investigating complex valued autoencoders focus mainly on the aspect of learning compressed representations of complex valued inputs. They do not directly address the biological motivation of binding by synchrony. They are usually strictly defined on the typical complex algebra and are not described by a differential equation which corresponds to coupled oscillators. The formalism of complex valued autoencoders might be adapted to allow further abstractions of our

model. This could support our understanding of the underlying computational principles of visual grouping and segmentation.

A very different and novel approach of coding contextual informations in autoencoder networks are mean-covariance restricted Boltzmann machines (Ranzato and Hinton, 2010). In these models latent hidden factors are used to efficiently represent the contextual information in the input in addition to the usual representation of pixel means in standard models of restricted Boltzmann machines. It was shown that the model can efficiently code pixel covariances in analogy to complex cells and pixel means in analogy to simple cells. However, the coding of contextual information in these models is limited to pair-wise interactions in the input layer. Therefore, this kind of generative model can capture only a linear combination of second order statistics so that contextual interactions between a large group of neurons is only possible through direct connections. In contrast, the grouping in our model is a dynamic process in which interactions between neurons are possible without a direct connection between them but through intermediate neurons. The reason is that our model uses a dynamical system approach with recurrent connections in contrast to probabilistic modeling of forward and backward connections.

Some mathematical theories of cortical processing mechanisms also take the contextual information into account. For example the free energy principle (Friston, 2010) and the theory of coherent infomax (Kay and Phillips, 2011) explicitly incorporate the context into single-variable local processors in the network. In contrast, the model presented in this paper takes the context into account in a separate phase variable, which codes relational properties similar to the dynamics on fast time scales in biological neural networks. Thereby our simulation allows to model higher order relational structures with a limited number of horizontal connections. In contrast, in the mathematical formalization of coherent infomax the contextual field input is assumed to be integrated into a single variable output of a local processor in the network. Thereby it doesn't allow implementing higher order relations between many local processors if the computational resources are limited. This limitation is of course only a matter of the used mathematical formalism and doesn't affect the general explanatory power of the free energy principle or the theory of coherent infomax. Therefore, in a broader sense our simulation model could be seen as an approximate implementation of these abstract concepts, although we use a biologically motivated architecture instead of a probabilistic derivation.

Our study combines aspects of these normative models of sensory processing and of detailed models of dynamical neural networks. We use only the statistics induced by natural images to learn unsupervised the forward and tangential phase connections. The supervised labeled segmentation masks are only used to evaluate how phase synchrony corresponds to a relational coding in the neural representation. Hence, the concept can be phrased completely in the form of a normative model. In future work, we plan to further formalize the model and conceive more complex learning rules for the phase interactions. These learning rules could replace the sampling of sparse connections from the correlation of activation by a more biologically motivated rule. For example, one could develop learning rules based on spike-timing dependent plasticity if phase delays are incorporated in the interactions of the network. This would additionally allow modeling phase locking between neurons and coding of syntactic relations in the network. These extensions to our model could provide new insights into the computational principles underlying higher order cognitive processes.

5. CONCLUSIONS

Our study revealed that the concept of binding by synchrony is viable in the context of unsupervised learning using natural stimuli. We show that the structural connectivity based on correlated activity leads to relational coding in a neural network model of coupled phase oscillators. The presented novel evaluation methodology for image segmentation revealed that the phase of neurons code global stimulus attributes. This strengthens the evidence that phase synchronization plays a key role to coordinate the spatially distributed information processing in the cortex. One could further speculate on how higher level coordination and binding between cortical areas might evolve from unsupervised learning based on correlated neuronal activity.

ACKNOWLEDGMENTS

The authors would like to thank Robert Märtin for his valuable comments and helpful suggestions.

FUNDING

This work was funded by the DFG through SFB 936 Multi-Site Communication in the Brain.

REFERENCES

Abeles, M. (1982). *Local Cortical Circuits: An Electrophysiological Study*, Vol. 6. New York, NY: Springer-Verlag. doi: 10.1007/978-3-642-81708-3

Açik, A., Onat, S., Schumann, F., Einhäuser, W., and König, P. (2009). Effects of luminance contrast and its modifications on fixation behavior during free viewing of images from different categories. *Vision Res.* 49:1541. doi: 10.1016/j.visres.2009.03.011

Baldi, P., and Lu, Z. (2012). Complex-valued autoencoders. *Neural Netw.* 33, 136–147. doi: 10.1016/j.neunet.2012.04.011

Bauer, M., Treichler, S., Slaughter, E., and Aiken, A. (2012). "Legion: expressing locality and independence with logical regions," in *High Performance Computing, Networking, Storage and Analysis (SC), 2012 International Conference* (Salt Lake City), 1–11.

Beggs, J. M., and Plenz, D. (2003). Neuronal avalanches in neocortical circuits. *J. Neurosci.* 23, 11167–11177. doi: 10.1523/jneurosci.0540-04.2004

Bell, A. J., and Sejnowski, T. J. (1997). The "independent components" of natural scenes are edge filters. *Vision Res.* 37:3327. doi: 10.1016/S0042-6989(97)00121-1

Belluscio, M. A., Mizuseki, K., Schmidt, R., Kempter, R., and Buzsáki, G. (2012). Cross-frequency phase-phase coupling between theta and gamma oscillations in the hippocampus. *J. Neurosci.* 32, 423–435. doi: 10.1523/JNEUROSCI.4122-11.2012

Benjamini, Y., and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Ann. Stat.* 29, 1165–1188.

Betsch, B. Y., Einhäuser, W., Körding, K. P., and König, P. (2004). The world from a cat's perspective–statistics of natural videos. *Biol. Cybern.* 90, 41–50. doi: 10.1007/s00422-003-0434-6

Breakspear, M., Heitmann, S., and Daffertshofer, A. (2010). Generative models of cortical oscillations: neurobiological implications of the Kuramoto model. *Front. Hum. Neurosci.* 4:190. doi: 10.3389/fnhum.2010.00190

Buonomano, D. V., and Maass, W. (2009). State-dependent computations: spatiotemporal processing in cortical networks. *Nat. Rev. Neurosci.* 10, 113–125. doi: 10.1038/nrn2558

Coates, A., Lee, H., and Ng, A. Y. (2010). An analysis of single-layer networks in unsupervised feature learning. *Ann. Arbor* 1001:48109.

Einhäuser, W., and König, P. (2010). Getting real—sensory processing of natural stimuli. *Curr. Opin. Neurobiol.* 20, 389–395. doi: 10.1016/j.conb.2010.03.010

Engel, A. K., and Singer, W. (2001). Temporal binding and the neural correlates of sensory awareness. *Trends Cogn. Sci.* 5, 16–25. doi: 10.1016/S1364-6613(00)01568-0

Franzius, M., Sprekeler, H., and Wiskott, L. (2007). Slowness and sparseness lead to place, head-direction, and spatial-view cells. *PLoS Comput. Biol.* 3:e166. doi: 10.1371/journal.pcbi.0030166

Friese, U., Köster, M., Hassler, U., Martens, U., Barreto, N. T., and Gruber, T. (2012). Successful memory encoding is associated with increased cross-frequency coupling between frontal theta and posterior gamma oscillations in human scalp-recorded EEG. *Neuroimage* 66, 642–647. doi: 10.1016/j.neuroimage.2012.11.002

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.* 11, 127–138. doi: 10.1038/nrn2787

Gray, C. M., König, P., Engel, A. K., and Singer, W. (1989). Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. *Nature* 338, 334–337. doi: 10.1038/338334a0

Hinton, G. (2010). A practical guide to training restricted Boltzmann machines. *Momentum* 9:1.

Humphries, M. D., Gurney, K., and Prescott, T. J. (2006). The brainstem reticular formation is a small-world, not scale-free, network. *Proc. R. Soc. B Biol. Sci.* 273, 503–511. doi: 10.1098/rspb.2005.3354

Kay, J. W., and Phillips, W. (2011). Coherent infomax as a computational goal for neural systems. *Bull. Math. Biol.* 73, 344–372. doi: 10.1007/s11538-010-9564-x

Kayser, C., Einhäuser, W., and König, P. (2003). Temporal correlations of orientations in natural scenes. *Neurocomputing* 52, 117–123. doi: 10.1016/S0925-2312(02)00789-0

König, P., Engel, A. K., Löwel, S., and Singer, W. (1993). Squint affects synchronization of oscillatory responses in cat visual cortex. *Eur. J. Neurosci.* 5, 501–508. doi: 10.1111/j.1460-9568.1993.tb00516.x

Körding, K. P., Kayser, C., Einhäuser, W., and König, P. (2004). How are complex cell properties adapted to the statistics of natural stimuli? *J. Neurophysiol.* 91, 206–212. doi: 10.1152/jn.00149.2003

Kumar, A., Rotter, S., and Aertsen, A. (2010). Spiking activity propagation in neuronal networks: reconciling different perspectives on neural coding. *Nat. Rev. Neurosci.* 11, 615–627. doi: 10.1038/nrn2886

Kuramoto, Y. (1984). *Chemical Oscillations, Waves, and Turbulence*. Berlin: Springer. doi: 10.1007/978-3-642-69689-3

Lazar, A., Pipa, G., and Triesch, J. (2009). SORN: a self-organizing recurrent neural network. *Front. Comput. Neurosci.* 3:23. doi: 10.3389/neuro.10.023.2009

Le, Q. V., Karpenko, A., Ngiam, J., and Ng, A. Y. (2011a). ICA with reconstruction cost for efficient overcomplete feature learning. *Adv. Neural Inf. Proc. Sys.* 24, 1017–1025.

Le, Q. V., Ngiam, J., Coates, A., Lahiri, A., Prochnow, B., and Ng, A. Y. (2011b). "On optimization methods for deep learning," in *Proceedings of the 28th International Conference on Machine Learning (ICML-11)* (Bellevue), 265–272.

LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proc. IEEE* 86, 2278–2324. doi: 10.1109/5.726791

Lee, H., Grosse, R., Ranganath, R., and Ng, A. Y. (2009). "Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations," in *Proceedings of the 26th Annual International Conference on Machine Learning* (Montreal), 609–616.

Li, C., and Li, Y. (2011). Fast and robust image segmentation by small-world neural oscillator networks. *Cogn. Neurodyn.* 5, 209–220. doi: 10.1007/s11571-011-9152-2

Liu, D. C., and Nocedal, J. (1989). On the limited memory BFGS method for large scale optimization. *Math. Program.* 45, 503–528. doi: 10.1007/BF01589116

Löwel, S., and Singer, W. (1992). Selection of intrinsic horizontal connections in the visual cortex by correlated neuronal activity. *Science* 255:209. doi: 10.1126/science.1372754

Mariño, J., Schummers, J., Lyon, D. C., Schwabe, L., Beck, O., Wiesing, P., et al. (2005). Invariant computations in local cortical networks with balanced excitation and inhibition. *Nat. Neurosci.* 8, 194–201. doi: 10.1038/nn1391

Maye, A., and Werning, M. (2007). Neuronal synchronization: from dynamic feature binding to object representations. *Chaos Complex. Lett.* 2, 315–325.

Milner, P. M. (1974). A model for visual shape recognition. *Psychol. Rev.* 81:521. doi: 10.1037/h0037149

Olshausen, B. A., and Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature* 381, 607–609. doi: 10.1038/381607a0

Olshausen, B. A., and Field, D. J. (2005). How close are we to understanding V1? *Neural Comput.* 17, 1665–1699. doi: 10.1162/0899766054026639

Onat, S., Jancke, D., and König, P. (2013). Cortical long-range interactions embed statistical knowledge of natural sensory input: a voltage-sensitive dye imaging study. *F1000Research* 2:51 doi: 10.3410/f1000research.2-51.v1

Ranzato, M., and Hinton, G. E. (2010). "Modeling pixel means and covariances using factorized third-order Boltzmann machines," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on* (San Francisco), 2551–2558.

Rehn, M., and Sommer, F. T. (2007). A network that uses few active neurones to code visual input predicts the diverse shapes of cortical receptive fields. *J. Comput. Neurosci.* 22, 135–146. doi: 10.1007/s10827-006-0003-9

Russell, B. C., Torralba, A., Murphy, K. P., and Freeman, W. T. (2008). LabelMe: a database and web-based tool for image annotation. *Int. J. Comput. Vis.* 77, 157–173. doi: 10.1007/s11263-007-0090-8

Singer, W. (1999). Neuronal synchrony: a versatile code review for the definition of relations? *Neuron* 24, 49–65. doi: 10.1016/S0896-6273(00)80821-1

Sompolinsky, H., Golomb, D., and Kleinfeld, D. (1990). Global processing of visual stimuli in a neural network of coupled oscillators. *Proc. Natl. Acad. Sci. U.S.A.* 87, 7200–7204. doi: 10.1073/pnas.87.18.7200

Sporns, O., Chialvo, D. R., Kaiser, M., and Hilgetag, C. C. (2004). Organization, development and function of complex brain networks. *Trends Cogn. Sci.* 8, 418–425. doi: 10.1016/j.tics.2004.07.008

Stimberg, M., Wimmer, K., Martin, R., Schwabe, L., Mariño, J., Schummers, J., et al. (2009). The operating regime of local computations in primary visual cortex. *Cereb. Cortex* 19, 2166–2180. doi: 10.1093/cercor/bhn240

Sturm, A., and König, P. (2001). Mechanisms to synchronize neuronal activity. *Biol. Cybern.* 84:153. doi: 10.1007/s004220000209

Tononi, G., Sporns, O., and Edelman, G. M. (1992). Reentry and the problem of integrating multiple cortical areas: simulation of dynamic integration in the visual system. *Cereb. Cortex* 2, 310–335. doi: 10.1093/cercor/2.4.310

Von Der Malsburg, C. (1981). "The correlation theory of brain function," in *Models of Neural Networks II: Temporal Aspects of Coding and Information Processing in Biological Systems* (Berlin), 95–119.

Wang, D., and Terman, D. (1997). Image segmentation based on oscillatory correlation. *Neural Comput.* 9, 805–836. doi: 10.1162/neco.1997.9.4.805

Watts, D. J., and Strogatz, S. H. (1998). Collective dynamics of 'small-world'networks. *Nature* 393, 440–442. doi: 10.1038/30918

Wilson, H. R., and Cowan, J. D. (1972). Excitatory and inhibitory interactions in localized populations of model neurons. *Biophys. J.* 12, 1–24. doi: 10.1016/S0006-3495(72)86068-5

Wiskott, L., and Sejnowski, T. J. (2002). Slow feature analysis: unsupervised learning of invariances. *Neural Comput.* 14, 715–770. doi: 10.1162/089976602317318938

Wolf, F., and Geisel, T. (1998). Spontaneous pinwheel annihilation during visual development. *Nature* 395, 73–78. doi: 10.1038/25736

Wyss, R., König, P., and Verschure, P. F. J. (2006). A model of the ventral visual system based on temporal stability and local memory. *PLoS Biol.* 4:e120. doi: 10.1371/journal.pbio.0040120

Young, R. A. (1987). The Gaussian derivative model for spatial vision: I. retinal mechanisms. *Spat. Vis.* 2, 273–293. doi: 10.1163/156856887X00222

Young, R. A., and Lesperance, R. M. (2001). The Gaussian derivative model for spatial-temporal vision: II. cortical data. *Spat. Vis.* 14, 3–4. doi: 10.1163/156856801753253591

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.

Received: 13 June 2013; accepted: 30 December 2013; published online: 27 January 2014.

*Citation: Finger H and König P (2014) Phase synchrony facilitates binding and segmentation of natural images in a coupled neural oscillator network. Front. Comput. Neurosci. **7**:195. doi: 10.3389/fncom.2013.00195*

This article was submitted to the journal Frontiers in Computational Neuroscience.

Copyright © 2014 Finger and König. 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) or licensor 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.

frontiers in
COMPUTATIONAL NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 29 November 2013
doi: 10.3389/fncom.2013.00164

Patterns of interval correlations in neural oscillators with adaptation

***Tilo Schwalger*[1,2]* and *Benjamin Lindner*[1,2]**

[1] *Bernstein Center for Computational Neuroscience, Berlin, Germany*
[2] *Department of Physics, Humboldt Universität zu Berlin, Berlin, Germany*

Edited by:
Tatjana Tchumatchenko, Max Planck Institute for Brain Research, Germany

Reviewed by:
Magnus Richardson, University of Warwick, UK
Richard Naud, Ecole Polytechnique Federale Lausanne, Switzerland

***Correspondence:**
Tilo Schwalger, Brain Mind Institute, École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland
e-mail: tilo@pks.mpg.de

Neural firing is often subject to negative feedback by adaptation currents. These currents can induce strong correlations among the time intervals between spikes. Here we study analytically the interval correlations of a broad class of noisy neural oscillators with spike-triggered adaptation of arbitrary strength and time scale. Our weak-noise theory provides a general relation between the correlations and the phase-response curve (PRC) of the oscillator, proves anti-correlations between neighboring intervals for adapting neurons with type I PRC and identifies a single order parameter that determines the qualitative pattern of correlations. Monotonically decaying or oscillating correlation structures can be related to qualitatively different voltage traces after spiking, which can be explained by the phase plane geometry. At high firing rates, the long-term variability of the spike train associated with the cumulative interval correlations becomes small, independent of model details. Our results are verified by comparison with stochastic simulations of the exponential, leaky, and generalized integrate-and-fire models with adaptation.

Keywords: spike-frequency adaptation, non-renewal process, serial correlation coefficient, phase-response curve, integrate-and-fire model, long-term variability

1. INTRODUCTION

The nerve cells of the brain are complex physical systems. They generate action potentials (spikes) by a nonlinear, adaptive, and noisy mechanism. In order to understand signal processing in single neurons, it is vital to analyze the sequence of the interspike intervals (ISIs) between adjacent action potentials. There is experimental evidence accumulating that the spiking in many cases is *not* a renewal process, i.e., a spike train with mutually independent ISIs, but that intervals are typically correlated over a few lags (Lowen and Teich, 1992; Ratnam and Nelson, 2000; Neiman and Russell, 2001; Nawrot et al., 2007; Engel et al., 2008) [further reports are reviewed in (Farkhooi et al., 2009; Avila-Akerberg and Chacron, 2011)]. These correlations are a basic statistics of any spike train with important implications for information transmission and signal detection in neural systems (Ratnam and Nelson, 2000; Chacron et al., 2001, 2004; Avila-Akerberg and Chacron, 2011) and man-made signal detectors (Nikitin et al., 2012). They are often characterized by the serial correlation coefficient (SCC)

$$\rho_k = \frac{\langle (T_i - \langle T_i \rangle)\,(T_{i+k} - \langle T_{i+k} \rangle) \rangle}{\langle (T_i - \langle T_i \rangle)^2 \rangle}, \qquad (1)$$

where T_i and T_{i+k} are two ISIs lagged by an integer k and $\langle \cdot \rangle$ denotes ensemble averaging. ISI correlations can be induced via correlated input to the neural dynamics, e.g. in the form of external colored noise (Middleton et al., 2003; Lindner, 2004), intrinsic noise from ion channels with slow kinetics (Fisch et al., 2012), or stochastic narrow-band input (Neiman and Russell, 2001, 2005; Bauermeister et al., 2013).

Another ubiquitous mechanism for ISI correlations are slow feedback processes mediating spike-frequency adaptation (Chacron et al., 2000; Liu and Wang, 2001; Benda et al., 2005)—a phenomenon describing the reduced neuronal response to slowly changing stimuli (Benda and Herz, 2003; Gabbiani and Krapp, 2006). In the stationary state, these adaptation mechanisms are typically associated with short-range correlations with a negative SCC at lag $k = 1$ and a reduced Fano factor as demonstrated by several numerical (Geisler and Goldberg, 1966; Wang, 1998; Liu and Wang, 2001; Benda et al., 2010) and analytical studies (Schwalger et al., 2010; Schwalger and Lindner, 2010; Farkhooi et al., 2011; Urdapilleta, 2011). The correlation structure of adapting neurons can show qualitatively different patterns, ranging from monotonically decaying correlations to damped oscillations when plotted as a function of the lag (Ratnam and Nelson, 2000). Because ISI correlations shape spectral measures (Chacron et al., 2004), they bear implications for neural computation in general. However, a simple theory that predicts and explains possible correlation patterns is still lacking.

In this article, we present a relation between the ISI correlation coefficient ρ_k and a basic characteristics of nonlinear neural dynamics, the *phase-response curve* (PRC). The PRC quantifies the advance (or delay) of the next spike caused by a small depolarizing current applied at a certain time after the last spike (Ermentrout, 1996). For neurons which integrate up their input (integrator neurons), the PRC is positive at all times (type I PRC) whereas neurons, which show subthreshold resonances (resonator neurons), possess a PRC that is partly negative (type II PRC) (Ermentrout, 1996; Izhikevich, 2005; Ermentrout and Terman,

2010). Below we show that resonator neurons possess a richer repertoire of correlation patterns than integrator neurons do.

2. RESULTS

2.1. MODEL

Spike frequency adaptation can be modeled by Hodgkin–Huxley type neurons with a depolarization-activated adaptation current (Wang, 1998; Ermentrout et al., 2001; Benda and Herz, 2003). However, the spiking of such conductance-based models can in many instances be approximated by simpler multi-dimensional integrate-fire (IF) models that are equipped with a spike-triggered adaptation current (Treves, 1993; Izhikevich, 2003; Brette and Gerstner, 2005); adapting IF models perform excellently in predicting spike times of real cells under noisy stimulation (Gerstner and Naud, 2009). Here, we consider a stochastic nonlinear multi-dimensional IF model for the membrane potential v, N auxiliary variables w_j $(j = 1, \ldots, N)$ and a spike-triggered adaption current $a(t)$:

$$\dot{v} = f_0(v, \mathbf{w}) + \mu - a + \xi(t), \tag{2a}$$

$$\dot{w}_j = f_j(v, \mathbf{w}), \tag{2b}$$

$$\tau_a \dot{a} = -a + \tau_a \Delta \sum_i \delta\,(t - t_i)\,. \tag{2c}$$

The membrane potential $v(t)$ is subject to weak Gaussian noise $\xi(t)$ with $\langle \xi(t)\xi(t')\rangle = 2D\delta(t - t')$ and noise intensity D. The dynamics is complemented by a spike-and-reset mechanism: whenever $v_{(\mathrm{T})}$ reaches a threshold $\upsilon(t)$, a spike is registered at time $t_i = t$ and $v(t)$ and $\mathbf{w}(t) = [w_1(t), \ldots, w_N(t)]^{\mathrm{T}}$ are reset to $v(t_i^+) = 0$ and $\mathbf{w}(t_i^+) = \mathbf{w}_r$ (where t_i^+ denotes the right-sided limit $t \to t_i + 0$). At the same time, $a(t)$ suffers a jump by $\Delta \geq 0$ as seen from Equation (2c), which resembles high-threshold adaptation currents (Wang, 1998; Liu and Wang, 2001). The constant input current μ is assumed to be sufficiently large to ensure ongoing spiking even in the absence of noise. Note that the model is non-dimensionalized by measuring time in units of the membrane time constant $\tau_{\mathrm{m}} \sim 10$ ms and voltage in units of the distance between reset and spike-initiating potential (a typical value is 15 mV). In particular, the adaptation time constant τ_a is measured relative to τ_{m} and the unit of the firing rate is $\tau_{\mathrm{m}}^{-1} \sim 100$ Hz.

An important special case, the adaptive exponential integrate-and-fire model (Brette and Gerstner, 2005) with purely spike-triggered adaptation and a white noise current with constant mean is illustrated in **Figure 1**. It assumes an exponential nonlinearity $f_0(v) = -\gamma v + \gamma\Delta_T \exp[(v - 1)/\Delta_T]$ (Fourcaud-Trocmé et al., 2003; Badel et al., 2008) and corresponds to $N = 0$. Time courses of $v(t)$ and $a(t)$ are shown in **Figures 1A1,B1** for two distinct correlation patterns possible in this model. The ISIs $T_i = t_i - t_{i-1}$ are obtained as differences between subsequent spiking times t_i. The sequence T_i, T_{i+1}, T_{i+2} displays patterns of *short-long-long* (**Figure 1A1**) and *short-long-short* (**Figure 1B1**), corresponding to a negative SCC, which decays monotonically with the lag k (**Figure 1A3**) or to an SCC oscillating with k (**Figure 1B3**). In the following, we develop a theory to analyze these and other correlation patterns possible in multi-dimensional adapting IF models.

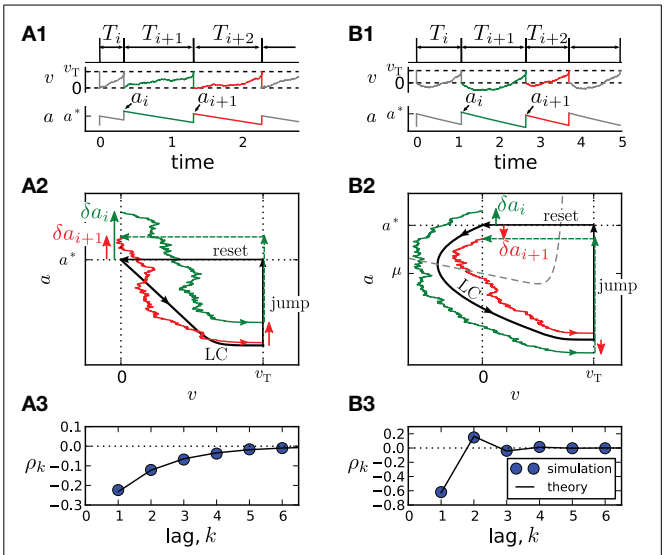


FIGURE 1 | Correlation patterns in the adaptive exponential IF model with $\tau_a = 10$, $\gamma = 1$, $\Delta_T = 0.1$, $\upsilon_T = 2$, $D = 0.1$. Adaptation is weak ($\Delta = 1$, $\mu = 15$) in **(A)** and strong ($\Delta = 10$, $\mu = 80$) in **(B)**. Membrane voltage $v(t)$ and adaptation variable $a(t)$ with ISI sequences $\{T_i\}$ and peak adaptation values $\{a_i\}$ are shown in **(A1,B1)**; time is in units of the membrane time constant τ_{m}. Colored pieces of trajectories in the phase plane (v, a) in **(A2,B2)** correspond to the respective colors in **(A1,B1)**. The deterministic limit cycle (LC), determined by the initial (post-spike) values $v = 0$, $a = a^*$, is indicated by a thick black line. For weak adaptation **(A2)** a short ISI T_i causes positive deviations $\delta a_i = a_i - a^*$ and $\delta a_{i+1} = a_{i+1} - a^*$ of peak values leading to long ISIs T_{i+1} and T_{i+2} and, hence, to a negative ISI correlation at all lags **(A3)**. Because of the qualitatively different limit cycle for strong adaptation **(B2)**, deviations δa_i and δa_{i+1} differ in sign, yielding an oscillatory correlation pattern **(B3)**.

2.2. GENERAL THEORY

In our model Equation (2), $a(t)$ is the only variable that keeps a memory of the previous spike times thereby inducing correlations between ISIs. Over one ISI the time course of adaptation is an exponential decay, relating two adjacent peak values $a_i = a(t_i^+)$ and $a_{i+1} = a(t_{i+1}^+)$ by

$$a_{i+1} = a_i e^{-T_{i+1}/\tau_a} + \Delta \tag{3}$$

(**Figures A1,B1**). We assume that in the deterministic case ($D = 0$) our model has a finite period T^* (i.e., the model operates in the tonically firing regime) and, hence, for $D = 0$ the map (3) has a stable fixed point

$$a^* = \Delta / \left[1 - \exp\left(-T^*/\tau_a\right)\right]. \tag{4}$$

The asymptotic deterministic dynamics can be interpreted as a limit-cycle like motion in the phase space from the reset point to the threshold and back by the instantaneous reset [cf. **Figures 1A2,B2**].

Weak noise will cause small deviations in the period $\delta T_i = T_i - T^* \approx T_i - \langle T_i\rangle$ that are mutually correlated with coefficient

$\rho_k = \langle \delta T_i \delta T_{i+k}\rangle / \langle \delta T_i^2\rangle$. The peak adaptation values, however, also fluctuate, $\delta a_i = a_i - a^*$, and both deviations are related by linearizing Equation (3):

$$\delta T_{i+1} = \frac{\tau_a}{a^*}\left(\delta a_i - e^{T^*/\tau_a}\delta a_{i+1}\right). \tag{5}$$

A second relation between the small deviations can be gained by considering how a small perturbation in the voltage dynamics affects the length of the period. This effect is captured by the infinitesimal phase response curve (PRC), $Z(t)$, $t \in (0, T^*)$ (Izhikevich, 2005; Ermentrout and Terman, 2010) (see Section 4 for the precise definition). During the interval T_{i+1}, the voltage dynamics in Equation (2a) can be written as $\dot v = f_0(v, \mathbf{w}) + \mu - (a^* + \delta a_i)e^{-(t-t_i)/\tau_a} + \xi(t)$. Compared to the deterministic limit cycle, the dynamics is perturbed by the weak noise and the small deviation in the adaptation $\delta a_i e^{-(t-t_i)/\tau_a}$ yielding in linear response

$$\delta T_{i+1} = \int_0^{T^*} \mathrm{d}t\, Z(t)\left(\delta a_i e^{-\frac{t}{\tau_a}} - \xi(t_i + t)\right). \tag{6}$$

Combining Equations (5), (6) we obtain the stochastic map

$$\delta a_{i+1} = \alpha\vartheta\,\delta a_i + \Xi_i, \tag{7}$$

where $\Xi_i = \frac{\alpha a^*}{\tau_a}\int_0^\infty \mathrm{d}t\, Z(t)\xi(t_i + t)$ are uncorrelated Gaussian random numbers and

$$\alpha = e^{-T^*/\tau_a}, \qquad \vartheta = 1 - \frac{a^*}{\tau_a}\int_0^{T^*} \mathrm{d}t\, Z(t)e^{-\frac{t}{\tau_a}}. \tag{8}$$

Note that local stability of the fixed point a^* requires that $|\alpha\vartheta| < 1$. The covariance $c_k = \langle \delta a_i \delta a_{i+k}\rangle$ of the auto-regressive process Equation (7) can be calculated by elementary means and using Equation (5) we obtain for $k \geq 1$:

$$\rho_k = -A(1-\vartheta)\,(\alpha\vartheta)^{k-1}, \qquad A = \frac{\alpha(1-\alpha^2\vartheta)}{1+\alpha^2-2\alpha^2\vartheta}. \tag{9}$$

In order to compute α and ϑ via Equation (8), we have to calculate T^* and $Z(t)$ (a^* then follows from Equation (4)), which can be done for simple systems analytically.

Our main result, Equations (8), (9), allows to draw a number of general conclusions. It shows that the SCC is always a geometric sequence with respect to the lag k that can generate qualitatively different correlation patterns depending on the value of ϑ and thus on PRC and adaptation current. Because $|\alpha\vartheta| < 1$ and $0 < \alpha < 1$, the prefactor A in Equation (9) is always positive. Consequently, ρ_1 is negative for $\vartheta < 1$ and positive for $\vartheta > 1$. Looking at Equation (8), we find that a positive PRC inevitably yields $\vartheta < 1$. This implies that adapting neurons with type I PRC possess negative correlations between adjacent ISIs. Intuitively, a short ISI causes in the following on average a higher inhibitory adaptation during the subsequent ISI. Such an inhibitory current always enlarges the ISI in type I neurons—hence, a short ISI is followed by a long ISI.

The sign of higher lags is determined by the base of the power: for $\vartheta > 0$ correlations decay monotonically, whereas for $\vartheta < 0$ the SCC oscillates. Two special cases are $\vartheta = 0$ with a negative correlation at lag 1 and vanishing correlations at all higher lags and $\vartheta = 1$ where all correlations vanish. Overall, we find five basic patterns corresponding to the cases $-\alpha^{-1} < \vartheta < 0$, $\vartheta = 0$, $0 < \vartheta < 1$, $\vartheta = 1$ and $1 < \vartheta < \alpha^{-1}$. These basic patterns cover all interval correlations discussed in previous theoretical studies (Schwalger and Lindner, 2010; Urdapilleta, 2011). Our geometric formula generalizes the theory for the perfect IF model with adaptation (Schwalger et al., 2010) to more realistic, nonlinear multi-dimensional IF models with adaptation.

The cumulative effect of the correlations can be described by the sum over all ρ_k, which determines the long-time limit of the Fano factor and the low-frequency limit of the spike train power spectrum (for a definition of these quantities, see Section 4.2). Evaluating the geometric series yields

$$\sum_{k=1}^{\infty}\rho_k = -\frac{A\,(1-\vartheta)}{1-\alpha\vartheta}. \tag{10}$$

This shows that adaptation in neurons with type I resetting ($\vartheta < 1$) leads to a negative summed correlation and hence a reduced long-term variability. Furthermore, at high firing rates achieved by a strong input current μ, the sum in Equation (10) can be approximated by

$$\sum_{k=1}^{\infty}\rho_k \simeq -\frac{1}{2} + \frac{1/2}{(1+\Delta\tau_a/\nu_\mathrm{T})^2}, \quad T^* \ll \tau_a. \tag{11}$$

In particular, for strong adaptation ($\Delta\tau_a \gg \nu_\mathrm{T}$) the sum is only slightly larger than $-1/2$. Note that by virtue of the fundamental relation $\lim_{t\to\infty} F(t) = C_\mathrm{V}^2\left(1 + 2\sum_{k=1}^\infty \rho_k\right)$ (Cox and Lewis, 1966) (see Section 4.2), the smallest possible value for the sum is $-1/2$ in order to ensure the non-negativity of the Fano factor $F(t)$. At this minimal value the long-term variability as expressed by the Fano factor vanishes even for a non-vanishing ISI variability as quantified by the coefficient of variation C_V. The latter quantity can also be estimated using the weak-noise theory: From Equation (7) one can calculate the variance of a_i and using Equation (5) an approximation for $C_\mathrm{V}^2 \approx \langle \delta T_i^2\rangle / T^{*2}$ can be obtained as follows:

$$C_\mathrm{V}^2 = 2D\frac{1+\alpha^2-2\alpha^2\vartheta}{[1-(\alpha\vartheta)^2]T^{*2}}\int_0^{T^*}\mathrm{d}t\,[Z(t)]^2. \tag{12}$$

2.3. ONE-DIMENSIONAL IF MODELS WITH ADAPTATION

In the simplest case ($N = 0, f_0(v, \mathbf{w}) = f(v)$) the PRC reads

$$Z(t) = Z(T^*)\exp\left[\int_t^{T^*}\mathrm{d}t' f'(v_0(t'))\right], \tag{13}$$

where $v_0(t)$ is the limit cycle solution and $Z(T^*) = [f(\nu_\mathrm{T}) + \mu - a^* + \Delta]^{-1}$ is the inverse of the velocity $\dot v_0(T^*)$ at the threshold, which is always positive. Thus, the PRC is positive for all $t \in (0, T^*)$, i.e., one-dimensional IF models show type I behavior. From our general considerations, this implies a negative SCC

at lag $k = 1$. The sign of the correlations at higher lags can be inferred from the sign of ϑ, for which one can show (Section 4) that

$$\vartheta = \big(f(0) + \mu - a^*\big)Z(0). \tag{14}$$

Because $Z(0) > 0$, the sign of ϑ is determined by the sign of $f(0) + \mu - a^*$. For weak adaptation such that $a^* < f(0) + \mu$ (achieved by a sufficiently small value of Δ or τ_a, **Figure 1A**), we will have $\vartheta > 0$ and a negative correlation at all lags (**Figure 1A3**). In this case, a short ISI occurring by fluctuation will cause a positive deviation δa_i (**Figure 1A2**, green arrow). Geometrically, it is plausible that such a positive deviation causes a likewise positive deviation δa_{i+1} in the subsequent cycle (**Figure 1A2**, red arrow). Because a positive deviation is associated with a long ISI, the initial short ISI is on average followed by longer ISIs.

In marked contrast, for strong adaptation such that $a^* > f(0) + \mu$ (achieved by a sufficiently large value of Δ or τ_a), ϑ becomes negative and hence the SCC's sign alternates with the lag. This alternation of the sign can be understood by means of the phase plane. Let us again consider a positive deviation δa_i due to a short preceding ISI (**Figure 1B2**, green arrow). Because $\dot{v}_0(0) = f(0) + \mu - a^* < 0$, the neuron is reset above the v-nullcline and hence hyperpolarizes at the beginning of the interval, i.e., the trajectory makes a detour into the region of negative voltage (corresponding to a "broad reset" in Naud et al. (2008)). A positive deviation δa_i leads to a larger detour (green trajectory) causing a sign inversion and hence a negative deviation δa_{i+1} (**Figure 1B2**, red arrow). Because a positive (negative) deviation corresponds on average to a long (short) ISI, the alternation of δa_i also entails an alternation of the ISI correlations. Thus, the distinction between monotonic and alternating patterns relates to a qualitative distinction of the voltage trace after resetting [cf. "sharp" vs. "broad" resets in Naud et al. (2008)].

As demonstrated in **Figures 1A3,B3**, our theory works well for the adapting exponential integrate-and-fire model. We next demonstrate the validity of our approach over a broad range of firing rates (**Figure 2**) for another important 1D model, the adapting leaky integrate-and-fire model (Treves, 1993) for which $f(v) = -\gamma v$ and

$$Z(t) = \exp\big[\gamma(t - T^*)\big]/\big(\mu - \gamma v_T - a^* + \Delta\big) \tag{15}$$

(here T^* has still to be determined from a transcendental equation). Changing the firing rate by varying the input current μ, we find a good agreement for the first two correlation coefficients and the sum of all ρ_k; the approximation of the CV shows deviations from simulation results when the input current μ becomes small (approaching the fluctuation-driven regime). In accordance with previous findings (Wang, 1998; Liu and Wang, 2001; Benda et al., 2010; Nesse et al., 2010; Schwalger et al., 2010; Schwalger and Lindner, 2010; Urdapilleta, 2011), the first correlation coefficient ρ_1 displays a minimum corresponding to strong anti-correlations between adjacent intervals. The correlations at lag 2 can be positive for a finite range of firing rates if the adaptation strength is sufficiently large (**Figure 2B**), whereas for moderate adaptation we find a negative ρ_2 at all firing rates

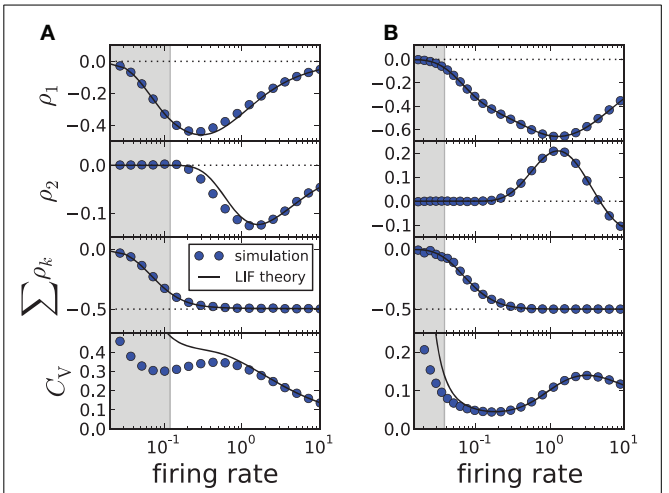


FIGURE 2 | ISI correlations and coefficient of variation (CV) of the adapting LIF model vs. firing rate $1/\langle T_i\rangle \approx 1/T^*$, where the rate is varied by increasing μ. The gray-shaded area corresponds to the fluctuation-driven regime ($\mu < \gamma v_T$), where the assumptions of the theory do not hold. The panels display (from top to bottom) ρ_1, ρ_2, the sum $\sum_{k=1}^{m} \rho_k$ and the CV for simulation (circles, $m = 100$) and theory (solid lines, $m \to \infty$). **(A)** Moderate adaptation: $\Delta = 1$, **(B)** strong adaptation: $\Delta = 10$. Both: $\gamma = 1$, $\tau_a = 10$, $D = 0.1$, $v_T = 1$. Note that the firing rate is given in units of the inverse membrane time constant τ_m^{-1}.

(**Figure 2A**). In both cases, however, the sum of SCCs approaches a value close to $-1/2$ for high firing rates as predicted by Equation (11) (**Figure 2**, bottom). This is strikingly similar to experimental data from weakly electric fish, in which some electro-receptors display a monotonically decaying SCC and some show an oscillatory SCC (Ratnam and Nelson, 2000) but all cells exhibit a sum close to $-1/2$ (Ratnam and Goense, 2004). Finally, **Figure 2** reveals a local maximum of the CV for some suprathreshold current μ—an effect that has been described by Nesse et al. (2008).

2.4. GENERALIZED INTEGRATE-AND-FIRE MODEL WITH ADAPTATION

Different correlation patterns become possible if we consider a type II PRC, which is by definition partly negative and can lead to a negative value of the integral in Equation (8), and hence to $\vartheta \geq 1$. This corresponds to a non-negative SCC at lag 1, which is infeasible in the one-dimensional case. To test the prediction $\rho_1 \geq 0$, we study the generalized integrate-and-fire (GIF) model (Brunel et al., 2003) with spike-triggered adaptation. This model is defined by $f_0(v, w) = -\gamma v - \beta w$ and $f_1(v, w) = (v - w)/\tau_w$. Using the method described in Section 4, the PRC is obtained as

$$Z(t) = \frac{e^{\frac{\nu}{2}(t-T^*)}\left[\cos(\Omega(t - T^*)) - \frac{1-\tau_w\gamma}{2\tau_w\Omega}\sin(\Omega(t - T^*))\right]}{\mu - \gamma v_T - \beta w_0(T^*) - a^* + \Delta} \tag{16}$$

where $\nu = \gamma + 1/\tau_w$, $\Omega = \sqrt{\frac{\beta+\gamma}{\tau_w} - \frac{\nu^2}{4}}$ and $w_0(t)$ is one component of the deterministic limit-cycle solution $[v_0(t), w_0(t), a_0(t)]$ that we calculated numerically.

In **Figure 3B** we demonstrate that all possible correlation patterns can be realized in the GIF model and that the predicted

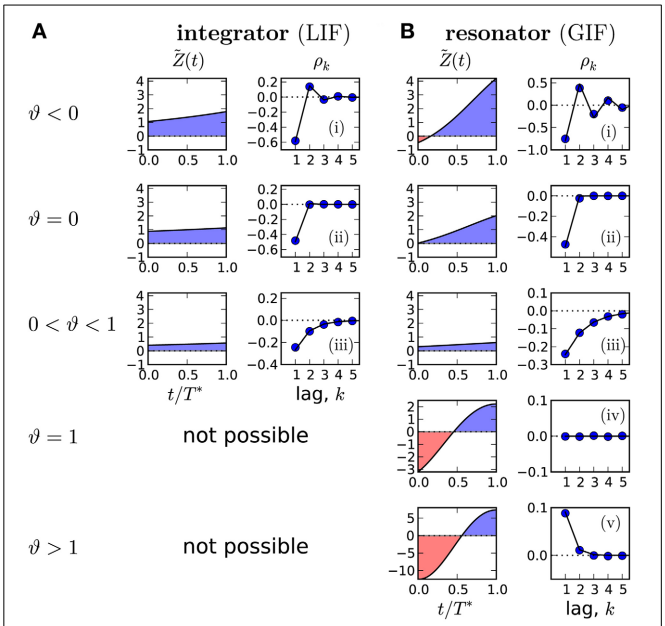


FIGURE 3 | Possible correlation patterns and corresponding PRCs (solid lines: theory, symbols: simulations of Equation (2)). For the adapting LIF model **(A)**, $\vartheta < 1$ and only three qualitative different cases are possible. The adapting GIF model **(B)** exhibits the full repertoire of correlation patterns because the PRC can be partly negative and ϑ can attain values from its entire physically meaningful interval $[-1/\alpha, 1/\alpha]$. The value of ϑ and hence the type of correlation pattern is set by the integral over the weighted PRC $\tilde{Z}(t) = Z(t)e^{-\frac{t}{\tau_a}}\frac{a^*T^*}{\tau_a}$, shown in left panels. LIF parameters: $D = 0.1$, $\tau_a = 2$, **(i)** $\mu = 20$, $\Delta = 10$, **(ii)** $\mu = 20$, $\Delta = 4.47$, (iii) $\mu = 5$, $\Delta = 1$. GIF parameters: **(i)** $\mu = 10$, $\beta = 3$, $\tau_a = 10$. **(ii)** $\mu = 11.75$, $\beta = 3$, $\tau_a = 10$. **(iii)** $\mu = 20$, $\beta = 1.5$, $\tau_a = 10$. **(iv)** $\mu = 2.12$, $\beta = 1.5$, $\tau_a = 1$, $\Delta = 10$. **(v)** $\mu = 1.5$, $\beta = 1.5$, $\tau_a = 1$, $\Delta = 9$ $D = 10^{-5}$. Unless stated otherwise, $\gamma = 1$, $\Delta = 1$, $\tau_w = 1.5$, $D = 10^{-4}$, $w_r = 0$.

SCCs agree quantitatively well in theory and model simulations (for comparison, see the SCC for the LIF in **Figure 3A**). To each distinct pattern belongs a range of ϑ (**Figure 3**, left), determined by the area under the weighted PRC $\tilde{Z}(t) = \frac{a^*}{\tau_a}e^{-\frac{t}{\tau_a}}Z(t)$. The function $\tilde{Z}(t)$ (left column in **Figures 3A,B**) illustrates, why an adapting GIF neuron can show vanishing (**Figure 3Biv**) or even *purely positive* ISI correlations (**Figure 3Bv**). In case of type II resetting, inhibitory input can *shorten* the ISI because of the negative part in the PRC; here inhibition acts like an excitatory input. Consequently, a short ISI will induce a stronger inhibition (adaptation) that now causes a likewise short interval and results thus in a positive correlation between adjacent ISIs. Also, the shortening effect of the adaption current in the early negative phase of the PRC can be exactly balanced by the delaying effect of the late positive phase of the PRC (pseudo-renewal case, in which the area under $\tilde{Z}$ is zero).

3. DISCUSSION

We have found a general relation between two experimentally accessible characteristics: the serial interval correlations and the phase response curve of a noisy neuron with spike-triggered adaptation. The theory predicts distinct correlation patterns like short-range negative and oscillatory correlations that have been observed in experiments (Ratnam and Nelson, 2000; Nawrot et al., 2007) and in simulation studies of adapting neurons (Chacron et al., 2000; Liu and Wang, 2001).

Beyond negative and oscillatory correlations, we have found, however, that resonator neurons with spike-frequency adaptation can exhibit purely positive ISI correlations or a pseudo-renewal process with uncorrelated intervals. Adaptation currents that are commonly associated with negative ISI correlations (Wang, 1998; Chacron et al., 2001; Liu and Wang, 2001; Chacron et al., 2003; Benda et al., 2010; Nesse et al., 2010) can thus induce a rich repertoire of correlation patterns. Despite the multitude of patterns, there is a universal limit for the cumulative correlations at high firing rates [cf. Equation (11)], which shows that the long-term variability of the spike train is in this limit always reduced in agreement with experimental studies (Ratnam and Goense, 2004).

Our analytical results apply to arbitrary adaptation strength and time scale but require that (1) the noise is weak and white, (2) the deterministic dynamics shows periodic firing with equal ISIs (i.e., a limit-cycle exists) and (3) the adaptation current is purely spike-triggered with (4) a single exponential decay time. Regarding the weak-noise assumption, we found from numerical simulations quantitative agreement with our theory for values of the coefficient of variation (CV) up to 0.4, which is, for instance, typical for neurons in the sensory periphery (Ratnam and Nelson, 2000; Neiman and Russell, 2004; Vogel et al., 2005). This holds even in the subthreshold regime at low CVs, where the deterministic system does not follow a limit cycle. In this case, T^* has to be replaced by the mean ISI. Moreover, we found qualitative agreement even for moderately strong noise with values of the CV up to 0.8, which is typical for cortical non-bursting neurons *in vivo* (e.g. **Figure 3** in Softky and Koch (1993)).

In the absence of a deterministic limit-cycle, i.e., in the fluctuation-driven regime at high CVs, different mathematical approaches have to be employed, such as those based on a hazard-function formalism (Muller et al., 2007; Nesse et al., 2010; Schwalger and Lindner, 2010; Farkhooi et al., 2011). Furthermore, for some parameter sets, we also observed repeat periods of the deterministic system that involved multiple ISIs corresponding to a periodic ISI sequence with $T_i = T_{i+n}$, where the smallest period is $n \geq 2$. Such cases can realize bursting (Naud et al., 2008), which we did not consider in the present study. However, we expect that these parameter regimes yield interesting correlation patterns because already in the noiseless case a periodic ISI sequence exhibits correlations between ISIs.

Regarding the last two assumptions, it seems that the analytical derivation cannot be easily extended to the cases of adaptation currents activated by the subthreshold membrane potential ("subthreshold adaptation" Ermentrout et al., 2001; Brette and Gerstner, 2005; Prescott and Sejnowski, 2008; Deemyad et al., 2012) and multiple-time-scale adaptation (Pozzorini et al., 2013). Ermentrout et al. (2001) have shown that the inclusion of subthreshold adaptation can lead to type II PRCs, which according to our theory could qualitatively change the correlation patterns. An adaptation dynamics depending on the subthreshold membrane potential also involves a fluctuating component because v is noisy. According to Schwalger et al. (2010), this stochasticity

could contribute positive correlations. The combined effect of spike-triggered, subthreshold and stochastic adaptation currents on the sign of the SCC is not clear.

The important cases of the fluctuation-driven regime and multiple-time-scale adaptation have been recently analyzed with respect to the first-order spiking statistics including the stationary firing rate as well as the mean response to time-dependent stimuli (Richardson, 2009; Naud and Gerstner, 2012). The second-order statistics, which describes the fluctuations of the spike train ("neural variability," cf. Section 4.2) and which limits the information transmission capabilities of neurons, is however still poorly understood theoretically in these cases. How adaptation shapes second-order statistics in the cases of multiple adaptation time scales, fluctuation-driven spiking and sub-threshold adaptation is an interesting topic for future investigations.

As an outlook we sketch, how our theory could be used to constrain unknown physiological parameters by measured SCCs and PRCs. For instance, from the mean ISI we can estimate $T^* = \langle T \rangle$. Furthermore, knowing $\rho_1 = -A(\alpha, \vartheta)(1-\vartheta)$ as well as the ratio $\rho_2/\rho_1 = \alpha\vartheta$ one can eliminate ϑ and solve for α. This allows to estimate the unknown adaptation time constant $\tau_a = -T^*/\ln\alpha$ and the amplitude of the adaptation current

$$a^* = \frac{\tau_a}{\alpha}\left(\alpha - \frac{\rho_2}{\rho_1}\right) \Bigg/ \int_0^{T^*} \mathrm{d}t\, Z(t) e^{-\frac{t}{\tau_a}} \,. \tag{17}$$

Although experimental PRCs are notoriously noisy (Izhikevich, 2005), the integral over $Z(t)$ determining our estimate of a^* is less error-prone. Combining our approach with advanced estimation methods for the PRC (Galán et al., 2005), may thus provide an alternative access to hidden physiological parameters using only spike time statistics.

4. MATERIALS AND METHODS

4.1. PHASE-RESPONSE CURVES OF ADAPTING IF MODELS

We use the phase-response curve $Z(t')$ to characterize the shift of the *next* spike following a small current pulse applied at a given "phase" $t' \in [0, T^*]$ of an ISI. More precisely, let us assume that the last spike occurred at time $t_0 = 0$. Then, the next spike time t_1 of the perturbed limit cycle dynamics $\dot{v} = f_0(v, \mathbf{w}) + \mu - a + \epsilon\delta(t - t')$, $v(0) = 0$, $\mathbf{w}(0) = \mathbf{w}_r$, $a(0) = a^*$, $0 < t' \leq T^*$ will be shifted by some amount $\delta T(t', \epsilon) = t_1 - T^*$. The infinitesimal PRC can be defined as the limit

$$Z(t') = -\lim_{\epsilon \to 0} \frac{\delta T(t', \epsilon)}{\epsilon}, \tag{18}$$

where the sign has been chosen such that a spike advance ($\delta T < 0$) due to a positive stimulation ($\epsilon > 0$) leads to a positive PRC. The definition of $Z(t)$ by the shift of the next spike differs from the PRC that describes the asymptotic spike shift but is equivalent to the so-called "first-order PRC," which is often measured in experiments (Netoff et al., 2012).

4.1.1. Adjoint equation and boundary conditions

The PRC can be computed using the adjoint method (see e.g. Ermentrout and Terman (2010)). To this end, the dynamics is linearized about the T^*-periodic limit cycle solution $\mathbf{y}_0(t) = [v_0(t), \mathbf{w}_0(t), a_0(t)]$. The linearized limit-cycle dynamics $\mathbf{y}(t) = \mathbf{y}_0(t) + \delta\mathbf{y}(t)$ corresponding to Equation (2) is given by

$$\dot{\delta\mathbf{y}} = A(t)\delta\mathbf{y} \tag{19}$$

with the Jacobian matrix

$$A(t) = \begin{pmatrix} \frac{\partial f_0}{\partial v} & \frac{\partial f_0}{\partial w_1} & \cdots & \frac{\partial f_0}{\partial w_N} & -1 \\ \tau_1^{-1}\frac{\partial f_1}{\partial v} & \tau_1^{-1}\frac{\partial f_1}{\partial w_1} & \cdots & \tau_1^{-1}\frac{\partial f_1}{\partial w_N} & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \tau_N^{-1}\frac{\partial f_N}{\partial v} & \tau_N^{-1}\frac{\partial f_N}{\partial w_1} & \cdots & \tau_N^{-1}\frac{\partial f_N}{\partial w_N} & 0 \\ 0 & \cdots & & 0 & -\tau_a^{-1} \end{pmatrix} \tag{20}$$

evaluated at $v = v_0(t), \mathbf{w} = \mathbf{w}_0(t)$. The linear response of the ISI to perturbations of the limit-cycle dynamics in an arbitrary direction is given by the vector $\mathbf{Z}(t) = [Z(t), Z_{w_1}(t), \ldots, Z_{w_N}(t), Z_a(t)]^{\mathrm{T}}$, where the first component is equal to the PRC defined above. This vector satisfies the adjoint equation $\dot{\mathbf{Z}} = -A^{\mathrm{T}}\mathbf{Z}$ (A^{T} denotes the transpose of A) with the normalization condition $\dot{v}_0(t)Z(t) + \dot{\mathbf{w}}_0(t)\mathbf{Z}_w(t) + \dot{a}_0(t)Z_a(t) = 1$. The remaining $N+1$ boundary conditions are obtained by the following consideration: On the limit cycle Γ, a phase $\phi : \Gamma \to [0, T^*]$ can be introduced in the usual way by inverting the map $t \mapsto \mathbf{y}_0(t)$ and setting $\phi = t$. Because we are interested in the shift of the *next* spike, it is useful to define the isochrons (sets of equal phase) as the sets of all points in phase space that will lead to the same first spike time. Put differently, phase points belonging to the same isochron will have their first threshold crossing in synchrony. As a consequence, the threshold hyperplane defined by the condition $v = v_{\mathrm{T}}$ is a special isochron corresponding to the phase $\phi = T^*$. Note that this definition of the phase implies that the reset line defined by the condition $v = 0, \mathbf{w} = \mathbf{w}_r$ does generally *not* correspond to $\phi = 0$ but to positive phases if $a < a^*$ and negative phases if $a > a^*$. Thus, off-limit-cycle trajectories suffer a phase jump upon reset. Close to the threshold, the isochrons are parallel to the threshold, and thus, a perturbation perpendicular to the v-direction does not change the phase. This insensitivity implies the boundary conditions $Z_{w_1}(T^*) = \ldots = Z_{w_N}(T^*) = Z_a(T^*) = 0$. Note that a definition of the PRC based on the asymptotic spike shift would require periodic boundary conditions (Ladenbauer et al., 2012).

From the above considerations, it becomes clear that the PRC $Z(t)$ can be computed for $t \in [0, T^*]$ by solving the system

$$\begin{pmatrix} \dot{Z} \\ \dot{Z}_{w_1} \\ \vdots \\ \dot{Z}_{w_N} \end{pmatrix} = -\begin{pmatrix} \frac{\partial f_0}{\partial v} & \tau_1^{-1}\frac{\partial f_1}{\partial v} & \cdots & \tau_N^{-1}\frac{\partial f_N}{\partial v} \\ \frac{\partial f_0}{\partial w_1} & \tau_1^{-1}\frac{\partial f_1}{\partial w_1} & \cdots & \tau_N^{-1}\frac{\partial f_N}{\partial w_1} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_0}{\partial w_N} & \tau_1^{-1}\frac{\partial f_1}{\partial w_1} & \cdots & \tau_N^{-1}\frac{\partial f_N}{\partial w_N} \end{pmatrix} \begin{pmatrix} Z \\ Z_{w_1} \\ \vdots \\ Z_{w_N} \end{pmatrix} \tag{21}$$

subject to the boundary conditions

$$Z(T^*) = \frac{1}{\dot{v}_0(T^*)} = \frac{1}{f_0(v_T, \mathbf{w}_0(T^*)) + \mu - a^* + \Delta}, \quad (22)$$

$$Z_{w_k}(T^*) = 0, \qquad k = 1, \ldots, N. \quad (23)$$

The PRC with respect to a is determined by

$$\dot{Z}_a = \frac{1}{\tau_a} Z_a + Z(t), \qquad Z_a(T^*) = 0. \quad (24)$$

The matrix in Equation (21) is again evaluated on the limit cycle at $v = v_0(t)$, $\mathbf{w} = \mathbf{w}_0(t)$ and is therefore time-dependent. An analytic solution of Equation (21) is possible for one-dimensional models with adaptation ($N = 0$) or general linear IF models although in most cases the deterministic period T^* still has to be computed numerically.

4.1.2. One-dimensional case

In the case $N = 0$, the PRC satisfies the equation $\dot{Z} = -f'(v_0)Z$ with boundary condition (22). The solution is given by Equation (13). In order to prove Equation (14), we compute $Z_a(t)$ from Equation (24) yielding

$$Z_a(t) = e^{\frac{t}{\tau_a}} \left(Z_a(0) + \int_0^t Z(t') e^{-\frac{t'}{\tau_a}} \, \mathrm{d}t' \right).$$

Evaluation of this expression for $t = T^*$ leads to $\vartheta = 1 + \frac{a^*}{\tau_a} Z_a(0)$. Finally, using the normalization condition $(f(0) + \mu - a^*)Z(0) - \frac{a^*}{\tau_a} Z_a(0) = 1$ yields Equation (14).

4.2. RELATION BETWEEN SECOND-ORDER STATISTICS OF SPIKE COUNT, SPIKE TRAIN AND INTERSPIKE INTERVALS

A stationary sequence of spike times $\{\ldots, t_{i-1}, t_i, t_{i+1}, \ldots\}$ is often characterized by the statistics of the spike train $x(t) = \sum_i \delta(t - t_i)$, the spike count $N(t) = \int_0^t \mathrm{d}t' \, x(t')$ or the sequence of ISIs $\{T_i = t_i - t_{i-1}\}$. In particular, neural variability can be quantified by the second-order statistics of these different descriptions as, for instance, the spike train power spectrum

$$S(f) = \int \mathrm{d}\tau \, e^{2\pi i f \tau} \langle x(t) x(t + \tau) \rangle, \quad (25)$$

the Fano factor

$$F(t) = \frac{\langle N(t)^2 \rangle - \langle N(t) \rangle^2}{\langle N(t) \rangle}, \quad (26)$$

and the coefficient of variation $C_V = \sqrt{\langle (T_i - \langle T_i \rangle)^2 \rangle} / \langle T_i \rangle$ and SCC ρ_k as defined in Equation (1). These statistics are connected by the fundamental relationship (Cox and Lewis, 1966) (see also (van Vreeswijk, 2010))

$$\lim_{t \to \infty} F(t) = \langle T_i \rangle \lim_{f \to 0} S(f) = C_V^2 \left(1 + 2 \sum_{k=1}^{\infty} \rho_k \right). \quad (27)$$

It shows that the summed SCC has a strong impact on the long-term variability of the spike train. In particular, a negative sum yields a more regular spike train on long time scales than a renewal process with the same C_V.

ACKNOWLEDGMENTS

This work was supported by Bundesministerium für Bildung und Forschung grant 01GQ1001A.

REFERENCES

Avila-Akerberg, O., and Chacron, M. J. (2011). Nonrenewal spike train statistics: causes and functional consequences on neural coding. *Exp. Brain Res.* 210, 353–371. doi: 10.1007/s00221-011-2553-y

Badel, L., Lefort, S., Brette, R., Petersen, C. C., Gerstner, W., and Richardson. M. J. (2008). Dynamic i-v curves are reliable predictors of naturalistic pyramidal-neuron voltage traces. *J. Neurophysiol.* 99, 656–666. doi: 10.1152/jn.01107.2007

Bauermeister, C., Schwalger, T., Russell, D., Neiman, A., and Lindner. B. (2013). Characteristic effects of stochastic oscillatory forcing on neural firing statistics: theory and application to paddlefish electroreceptor afferents. *PLoS Comput. Biol.* 9:e1003170. doi: 10.1371/journal.pcbi.1003170

Benda, J., and Herz, A. V. M. (2003). A universal model for spike-frequency adaptation. *Neural Comp.* 15, 2523. doi: 10.1162/089976603322385063

Benda, J., Longtin, A., and Maler, L. (2005). Spike-Frequency adaptation separates transient communication signals from background oscillations. *J. Neurosci.* 25, 2312–2321. doi: 10.1523/JNEUROSCI.4795-04.2005

Benda, J., Maler, L., and Longtin, A. (2010). Linear versus nonlinear signal transmission in neuron models with adaptation currents or dynamic thresholds. *J. Neurophysiol.* 104, 2806–2820. doi: 10.1152/jn.00240.2010

Brette, R., and Gerstner, W. (2005). Adaptive exponential integrate-and-fire model as an effective description of neuronal activity. *J. Neurophysiol.* 94, 3637–3642. doi: 10.1152/jn.00686.2005

Brunel, N., Hakim, V., and Richardson, M. J. (2003). Firing-rate resonance in a generalized integrate-and-fire neuron with subthreshold resonance. *Phys. Rev. E* 67(5 Pt 1), 051916–051916. doi: 10.1103/PhysRevE.67.051916

Chacron, M. J., Longtin, A., St-Hilaire, M., and Maler, L. (2000). Suprathreshold stochastic firing dynamics with memory in P-type electroreceptors. *Phys. Rev. Lett.* 85, 1576. doi: 10.1103/PhysRevLett.85.1576

Chacron, M. J., Lindner, B., and Longtin, A. (2004). Noise shaping by interval correlations increases neuronal information transfer. *Phys. Rev. Lett.* 92, 080601. doi: 10.1103/PhysRevLett.92.080601

Chacron, M. J., Longtin, A., and Maler, L. (2001). Negative interspike interval correlations increase the neuronal capacity for encoding time-dependent stimuli. *J. Neurosci.* 21, 5328.

Chacron, M. J., Pakdaman, K., and Longtin, A. (2003). Interspike interval correlations, memory, adaptation, and refractoriness in a leaky integrate-and-fire model with threshold fatigue. *Neural Comp.* 15, 253. doi: 10.1162/089976603762552915

Cox, D. R., and Lewis, P. A. W. (1966). *The Statistical Analysis of Series of Events*, chapter 4.6. (London: Chapman and Hall). doi: 10.1007/978-94-011-7801-3

Deemyad, T., Kroeger, J., and Chacron, M. J. (2012). Sub- and suprathreshold adaptation currents have opposite effects on frequency tuning. *J. Physiol.* 59(Pt 19), 4839–4858. doi: 10.1113/jphysiol.2012.234401

Engel, T. A., Schimansky-Geier, L., Herz, A., Schreiber, S., and Erchova, I. (2008). Subthreshold Membrane-Potential resonances shape Spike-Train patterns in the entorhinal cortex. *J. Neurophysiol.* 100, 1576. doi: 10.1152/jn.01282.2007

Ermentrout, B., Pascal, M., and Gutkin, B. (2001). The effects of spike frequency adaptation and negative feedback on the synchronization of neural oscillators. *Neural Comp.* 13, 1285–1310. doi: 10.1162/08997660152002861

Ermentrout, G. B. (1996). Type I membranes, phase resetting curves, and synchrony. *Neural Comp.* 8, 979. doi: 10.1162/neco.1996.8.5.979

Ermentrout, G. B., and Terman, D. H. (2010). *Mathematical Foundations of Neuroscience*. Springer. doi: 10.1007/978-0-387-87708-2

Farkhooi, F., Muller, E., and Nawrot, M. P. (2011). Adaptation reduces variability of the neuronal population code. *Phys. Rev. E* 83(5 Pt 1), 050905–050905. doi: 10.1103/PhysRevE.83.050905

Farkhooi, F., Strube-Bloss, M. F., and Nawrot, M. P. (2009). Serial correlation in neural spike trains: Experimental evidence, stochastic modeling, and single neuron variability. *Phys. Rev. E* 79, 021905–021910. doi: 10.1103/PhysRevE.79.021905

Fisch, K., Schwalger, T., Lindner, B., Herz, A., and Benda, J. (2012). Channel noise from both slow adaptation currents and fast currents is required to explain spike-response variability in a sensory neuron. *J. Neurosci.* 32, 17332–17344. doi: 10.1523/JNEUROSCI.6231-11.2012

Fourcaud-Trocmé, N., Hansel, D., van Vreeswijk, C., and Brunel, N. (2003). How spike generation mechanisms determine the neuronal response to fluctuating inputs. *J. Neurosci.* 23, 11628–11640.

Gabbiani, F., and Krapp, H. G. (2006). Spike-frequency adaptation and intrinsic properties of an identified, looming-sensitive neuron. *J. Neurophysiol.* 96, 2951–2962. doi: 10.1152/jn.00075.2006

Galán, R. F., Ermentrout, G. B., and Urban, N. N. (2005). Efficient estimation of phase-resetting curves in real neurons and its significance for neural-network modeling. *Phys. Rev. Lett.* 94, 158101. doi: 10.1103/PhysRevLett.94.158101

Geisler, C., and Goldberg, J. M. (1966). A stochastic model of the repetitive activity of neurons. *Biophys. J.* 6, 53–69. doi: 10.1016/S0006-3495(66)86639-0

Gerstner, W., and Naud, R. (2009). Neuroscience. how good are neuron models? *Science* 326, 379–380. doi: 10.1126/science.1181936

Izhikevich, E. M. (2003). Simple model of spiking neurons. *IEEE Trans. Neural. Netw.* 14, 1569–1572. doi: 10.1109/TNN.2003.820440

Izhikevich, E. M. (2005). *Dynamical Systems in Neuroscience: The Geometry of Excitability and Bursting*. (Cambridge, MA; London: MIT Press).

Ladenbauer, J., Augustin, M., Shiau, L., and Obermayer, K. (2012). Impact of adaptation currents on synchronization of coupled exponential integrate-and-fire neurons. *PLoS Comput. Biol.* 8:e1002478. doi: 10.1371/journal.pcbi.1002478

Lindner, B. (2004). Interspike interval statistics of neurons driven by colored noise. *Phys. Rev. E* 69, 022901. doi: 10.1103/PhysRevE.69.022901

Liu, Y.-H., and Wang, X.-J. (2001). Spike-frequency adaptation of a generalized leaky integrate-and-fire model neuron. *J. Comp. Neurosci.* 10, 25. doi: 10.1023/A:1008916026143

Lowen, S. B., and Teich, M. C. (1992). Auditory-nerve action potentials form a nonrenewal point process over short as well as long time scales. *J. Acoust. Soc. Am.* 92, 803. doi: 10.1121/1.403950

Middleton, J. W., Chacron, M. J., Lindner, B., and Longtin, A. (2003). Firing statistics of a neuron model driven by long-range correlated noise. *Phys. Rev. E* 68, 021920. doi: 10.1103/PhysRevE.68.021920

Muller, E., Buesing, L., Schemmel, J., and Meier, K. (2007). Spike-frequency adapting neural ensembles: Beyond mean adaptation and renewal theories. *Neural Comp.* 19, 2958–3110. doi: 10.1162/neco.2007.19.11.2958

Naud, R., and Gerstner, W. (2012). Coding and decoding with adapting neurons: a population approach to the peri-stimulus time histogram. *PLoS Comput. Biol.* 8:e1002711. doi: 10.1371/journal.pcbi.1002711

Naud, R., Marcille, N., Clopath, C., and Gerstner, W. (2008). Firing patterns in the adaptive exponential integrate-and-fire model. *Biol. Cybern.* 99, 335–347. doi: 10.1007/s00422-008-0264-7

Nawrot, M. P., Boucsein, C., Rodriguez-Molina, V., Aertsen, A., Grun, S., and Rotter, S. (2007). Serial interval statistics of spontaneous activity in cortical neurons *in vivo* and *in vitro*. *Neurocomputing* 70, 1717. doi: 10.1016/j.neucom.2006.10.101

Neiman, A., and Russell, D. F. (2001). Stochastic biperiodic oscillations in the electroreceptors of paddlefish. *Phys. Rev. Lett.* 86, 3443. doi: 10.1103/PhysRevLett.86.3443

Neiman, A., and Russell, D. F. (2004). Two distinct types of noisy oscillators in electroreceptors of paddlefish. *J. Neurophysiol.* 92, 492–509. doi: 10.1152/jn.00742.2003

Neiman, A., and Russell, D. F. (2005). Models of stochastic biperiodic oscillations and extended serial correlations in electroreceptors of paddlefish. *Phys. Rev. E* 71, 061915. doi: 10.1103/PhysRevE.71.061915

Nesse, W. H., Maler, L., and Longtin, A. (2010). Biophysical information representation in temporally correlated spike trains. *Proc. Natl. Acad. Sci. U.S.A.* 107, 21973–21978. doi: 10.1073/pnas.1008587107

Nesse, W. H., Negro, C. A., and Bressloff, P. C. (2008). Oscillation regularity in noise-driven excitable systems with multi-time-scale adaptation. *Phys. Rev. Lett.* 101, 088101–088101. doi: 10.1103/PhysRevLett.101.088101

Netoff, T., Schwemmer, M. A., and Lewis, T. J. (2012). "Experimentally estimating phase response curves of neurons: theoretical and practical issues." in *Phase Response Curves in Neuroscience: Theory, Experiment, and Analysis,* eds N. W. Schultheiss, A. A. Prinz, and R. J. Butera (Springer), 95–130.

Nikitin, A. P., Stocks, N. G., and Bulsara, A. R. (2012). Enhancing the resolution of a sensor via negative correlation: a biologically inspired approach. *Phys. Rev. Lett.* 109, 238103. doi: 10.1103/PhysRevLett.109.238103

Pozzorini, C., Naud, R., Mensi, S., and Gerstner, W. (2013). Temporal whitening by power-law adaptation in neocortical neurons. *Nat. Neurosci.* 16, 942–948. doi: 10.1038/nn.3431

Prescott, S. A., and Sejnowski, T. J. (2008). Spike-rate coding and spike-time coding are affected oppositely by different adaptation mechanisms. *J. Neurosci.* 28, 13649–13661. doi: 10.1523/JNEUROSCI.1792-08.2008

Ratnam, R., and Goense, J. B. M. (2004). "Variance stabilization of spike trains via non-renewal mechanisms - the impact on the speed and reliability of signal detection," in *Computational Neuroscience Meeting (CNS*2004)* (Baltimore).

Ratnam, R., and Nelson, M. E. (2000). Nonrenewal statistics of electrosensory afferent spike trains: implications for the detection of weak sensory signals. *J. Neurosci.* 20, 6672.

Richardson, M. J. E. (2009). Dynamics of populations and networks of neurons with voltage-activated and calcium-activated currents. *Phys. Rev. E* 80, 021928. doi: 10.1103/PhysRevE.80.021928

Schwalger, T., Fisch, K., Benda, J., and Lindner, B. (2010). How noisy adaptation of neurons shapes interspike interval histograms and correlations. *PLoS Comput. Biol.* 6:e1001026. doi: 10.1371/journal.pcbi.1001026

Schwalger, T., and Lindner, B. (2010). Theory for serial correlations of interevent intervals. *Eur. Phys. J. Spec. Top.* 187, 211–221. doi: 10.1140/epjst/e2010-01286-y

Softky, W. R., and Koch, C. (1993). The highly irregular firing of cortical cells is inconsistent with temporal integration of random EPSPs. *J. Neurosci.* 13, 334.

Treves, A. (1993). Mean-field analysis of neuronal spike dynamics. *Netw. Comput. Neural Syst.* 4, 259–284. doi: 10.1088/0954-898X/4/3/002

Urdapilleta, E. (2011). Onset of negative interspike interval correlations in adapting neurons. *Phys. Rev. E* 84, 041904. doi: 10.1103/PhysRevE.84.041904

van Vreeswijk, C. (2010). "Stochastic models of spike trains," in *Analysis of Parallel Spike Trains*, eds S. Grün and S. Rotter (Springer).

Vogel, A., Hennig, R. M., and Ronacher, B. (2005). Increase of neuronal response variability at higher processing levels as revealed by simultaneous recordings. *J. Neurophysiol.* 93, 3548. doi: 10.1152/jn.01288.2004

Wang, X. J. (1998). Calcium coding and adaptive temporal computation in cortical pyramidal neurons. *J. Neurophysiol.* 79, 1549–1566.

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.

Received: 01 August 2013; accepted: 26 October 2013; published online: 29 November 2013.

*Citation: Schwalger T and Lindner B (2013) Patterns of interval correlations in neural oscillators with adaptation. Front. Comput. Neurosci. **7**:164. doi: 10.3389/fncom.2013.00164*

This article was submitted to the journal Frontiers in Computational Neuroscience.

Copyright © 2013 Schwalger and Lindner. 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) or licensor 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.

frontiers in
COMPUTATIONAL NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 15 November 2013
doi: 10.3389/fncom.2013.00153

Propagating synchrony in feed-forward networks

***Sven Jahnke*[1,2,3]*, *Raoul-Martin Memmesheimer*[4] *and Marc Timme*[1,2,3]**

[1] *Network Dynamics, Max Planck Institute for Dynamics and Self-Organization (MPIDS), Göttingen, Germany*
[2] *Bernstein Center for Computational Neuroscience (BCCN), Göttingen, Germany*
[3] *Fakultät für Physik, Georg-August-Universität Göttingen, Göttingen, Germany*
[4] *Department for Neuroinformatics, Donders Institute, Radboud University, Nijmegen, Netherlands*

Edited by:
Tatjana Tchumatchenko, Max Planck Institute for Brain Research, Germany

Reviewed by:
Robert Rosenbaum, University of Pittsburgh, USA
Arvind Kumar, University of Freiburg, Germany
Raul C. Muresan, Romanian Institute of Science and Tehnology, Romania

****Correspondence:***
Sven Jahnke, Network Dynamics, Max Planck Institute for Dynamics and Self-Organization (MPIDS), Am Faßberg 17, 37077 Göttingen, Germany
e-mail: sjahnke@nld.ds.mpg.de

Coordinated patterns of precisely timed action potentials (spikes) emerge in a variety of neural circuits but their dynamical origin is still not well understood. One hypothesis states that synchronous activity propagating through feed-forward chains of groups of neurons (synfire chains) may dynamically generate such spike patterns. Additionally, synfire chains offer the possibility to enable reliable signal transmission. So far, mostly densely connected chains, often with all-to-all connectivity between groups, have been theoretically and computationally studied. Yet, such prominent feed-forward structures have not been observed experimentally. Here we analytically and numerically investigate under which conditions diluted feed-forward chains may exhibit synchrony propagation. In addition to conventional linear input summation, we study the impact of non-linear, non-additive summation accounting for the effect of fast dendritic spikes. The non-linearities promote synchronous inputs to generate precisely timed spikes. We identify how non-additive coupling relaxes the conditions on connectivity such that it enables synchrony propagation at connectivities substantially lower than required for linearly coupled chains. Although the analytical treatment is based on a simple leaky integrate-and-fire neuron model, we show how to generalize our methods to biologically more detailed neuron models and verify our results by numerical simulations with, e.g., Hodgkin Huxley type neurons.

Keywords: synchrony, networks, synfire chains, spike pattern, mathematical neuroscience, non-additive coupling, non-linear dendrites

1. SPIKE PATTERNS AND SIGNAL TRANSMISSION IN NEURONAL CIRCUITS

Reliable signal transmission is a core part of neuronal processing. A common hypothesis states that activity propagating along neuronal sub-populations that are connected in a feed-forward manner may support such signal transmission. Indeed, there is strong indication that activity propagation along feed-forward structures drives the generation of bird songs (Long et al., 2010) and experiments have shown propagation of synchronous and rate activity in feed-forward networks (FFNs) *in vitro* (Reyes, 2003; Feinerman et al., 2005; Feinerman and Moses, 2006). Sequential replay in the hippocampus and in neocortical networks also suggest underlying feed-forward mechanisms (August and Levy, 1999; Nadasdy et al., 1999; Lee and Wilson, 2002; Leibold and Kempter, 2006; Xu et al., 2011; Eagleman and Dragoi, 2012; Jahnke et al., 2012) and propagation of synchronous activity along feed-forward chains is a possible explanation for experimentally observed precise spike timing in the cortex (Riehle et al., 1997; Kilavik et al., 2009; Putrino et al., 2010). Further, the modular, hierarchical structure of many sensory and motor systems suggests propagation over sequences of areas in feed-forward manner, e.g., in bottom-up signal transfer (Felleman and Van Essen, 1991; Scannell et al., 1999; Bullmore and Sporns, 2009; Kumar et al., 2010).

Feed-forward structures which support the propagation of synchronous activity are termed synfire chains. The concept was introduced by Abeles (1982) as groups of neurons (layers) with dense anatomical connections between subsequent groups that are embedded in otherwise roughly randomly connected local neural circuits. Two major questions regarding the dynamical options for synfire activity include a) how synchrony may actively propagate and b) how such spatio-temporally coordinated spike timing may be robust against irregular background activity, because the synfire chains are part of a cortical network with dynamics defined by the so-called irregular balanced state (van Vreeswijk and Sompolinsky, 1996, 1998).

Addressing these points, theoretical studies have established conditions for stable propagation of synchrony in synfire chains (Diesmann et al., 1999; Gewaltig et al., 2001). Most synfire chain models assume functionally relevant FFNs that exhibit a very dense, often all-to-all connectivity between subsequent layers (Aviel et al., 2003; Mehring et al., 2003; Kumar et al., 2008) (see also a recent review on this topic Kumar et al., 2010). Such highly prominent feed-forward-structures, however, have not been found experimentally. Since cortical neural networks are overall sparse (e.g., Braitenberg and Schüz, 1998; Holmgren et al., 2003), we may also expect some level of dilution for embedded feed-forward chains. So far, computational model studies assumed that such chains created from existing connections in sparse recurrent networks exhibit strong synaptic efficiencies and specifically modified neuron properties to enable synchrony propagation (Vogels and Abbott, 2005).

Recently, we have shown that non-additive dendritic interactions promote propagation of synchrony (Jahnke et al., 2012). The non-additive dendritic interactions considered are mediated by fast dendritic spikes (Ariav et al., 2003; Gasparini et al., 2004; Polsky et al., 2004; Gasparini and Magee, 2006): upon stimulation within a time interval less than a few milliseconds, dendrites are capable of generating sodium spikes. These induce a strong, short and stereotypical depolarization in the soma. If this depolarization elicits a somatic spike, the spike occurs a fixed time interval after stimulation with sub-millisecond precision. This dendritic non-linearities relax the requirement of dense feed-forward anatomy and thereby allow for robust propagation of synchrony even in *diluted* FFNs with synapses of moderate strength within the biologically observed range.

In the present article, we analytically and numerically investigate in detail under which conditions synchronous activity may reliably propagate along the layers of an FFN where the inter-group connectivity is diluted, as may be expected when they are part of a sparse cortical network. An embedding network is mimicked by external, noisy input. We study the influence of the network setup, including the influence of the emulated embedding network, and of different types of standard linearly additive as well as non-additive dendritic interactions.

We derive analytical estimates for the critical connectivity—the minimal connectivity that allows robust propagation of synchrony. Some fundamental analytical results, in particular the ansatz for deriving a critical connectivity in the first place, have been briefly reported before (Jahnke et al., 2012). Here, we extend the approach and show how the bifurcation point, i.e., the transition point from the non-propagating to the propagating regime, can be estimated quantitatively from the neurons' ground state properties. We investigate the validity range of the analytical predictions and check them via direct numerical simulations. Furthermore, we discuss the applicability of our results to biologically more detailed neuron models and network setups. In particular, we argue that the assumptions underlying the analytical approach are met by a wide class of neuron models, including, e.g., conductance based leaky integrate-and-fire and Hodgkin–Huxley-type neurons.

The article is structured as follows: After introducing the neuron model and network setup in section 2, we study in the main part the propagation of synchrony in linearly coupled FFNs (section 3.1) and in FFNs incorporating dendritic non-linearities (section 3.2). In particular, we derive tools to study the system analytically, compare the results to computer simulations and elaborate differences of the dynamics of FFNs with and without non-additive dendritic interactions. In the final part (section 3.3), we discuss the application of our analytical results to biologically more detailed neuron models.

2. METHODS AND MODELS

2.1. NEURON MODEL

2.1.1. Linear model

Consider networks of leaky integrate-and-fire neurons that interact by sending and receiving spikes via directed connections. The state of neuron k at time t is described by its membrane potential $V_k(t)$ and its dynamics satisfy

$$\frac{dV_k(t)}{dt} = -\frac{V_k(t)}{\tau_k^{\mathrm{m}}} + I_k^{\mathrm{const}} + I_k^{\mathrm{net}}(t) + I_k^{\mathrm{ext}}(t), \tag{1}$$

where τ_k^{m} is the membrane time constant of neuron k, $I_k^{\mathrm{const}} := I_k^0/\tau_k^{\mathrm{m}}$ a constant input current, $I_k^{\mathrm{net}}(t)$ the input current caused by spikes within the network and $I_k^{\mathrm{ext}}(t)$ the input current arising from spikes from external sources. When the neuron's membrane potential reaches or exceeds the threshold Θ_k its membrane potential is reset to V_k^{reset} and a spike is sent to the postsynaptic neurons n, where it changes the postsynaptic potential after a delay τ_{nk}. After emitting a spike at $t = t_0$ the neuron becomes refractory for a time period t^{ref}, i.e., $V_k(t) = V_k^{\mathrm{reset}}$ for $t \in \left[t_0, t_0 + t^{\mathrm{ref}}\right]$.

To keep the model analytically tractable, we model the fast rise of the membrane potential upon the arrival of presynaptic spikes by instantaneous jumps of the membrane potential, such that the resulting input current reads

$$I_k^{\mathrm{net}}(t) = \sum_l \sum_m \epsilon_{kl} \delta\left(t - t_{lm}^f - \tau_{kl}\right). \tag{2}$$

Here ϵ_{kl} denotes the coupling strength from neuron l to neuron k, t_{lm}^f is the mth spike time of neuron l and τ_{kl} specifies the synaptic delay. In addition to spikes from the network each neuron receives excitatory and inhibitory random inputs that emulate an embedding network. These external inputs are modeled as random Poisson spike trains with rate ν^{exc} and ν^{inh}, respectively. The resulting input current is given by

$$I_k^{\mathrm{ext}}(t) = \sum_m \epsilon^{\mathrm{exc}} \delta\left(t - t_{km}^{\mathrm{ext,\,exc}}\right) + \sum_m \epsilon^{\mathrm{inh}} \delta\left(t - t_{km}^{\mathrm{ext,\,inh}}\right), \tag{3}$$

where $t_{km}^{\mathrm{ext,\,exc}}$ ($t_{km}^{\mathrm{ext,\,inh}}$) is the arrival time of the mth excitatory (inhibitory) spike to neuron k and $\epsilon^{\mathrm{exc}} > 0$ ($\epsilon^{\mathrm{inh}} < 0$) denote the corresponding coupling strength.

2.1.2. Non-linear model

In the above model all input currents are summed up linearly. To also investigate the effect of dendritic spikes we modulate the sum of synchronously arriving excitatory inputs by a non-linear dendritic modulation function $\sigma_{NL}(\cdot)$. This can be directly read off from experimental data (Ariav et al., 2003; Gasparini et al., 2004; Polsky et al., 2004; Gasparini and Magee, 2006): If the sum of excitatory inputs is below the dendritic threshold Θ_b, the single inputs are processed linearly ($\sigma_{NL}(\cdot)$ equals the identity). If the sum of inputs exceeds the dendritic threshold Θ_b, the depolarization is strongly non-linearly enhanced compared to that expected from linear summation. This is, in biological terms, due to a dendritic spike elicited. Larger inputs have been experimentally found to not further increase the somatic peak depolarization. The dendritic modulation function may then be modeled as

$$\sigma_{NL}(\epsilon) = \begin{cases} \epsilon & \text{for } \epsilon < \Theta_b \\ \kappa & \text{otherwise} \end{cases}. \tag{4}$$

The dendrites process synchronous inputs non-additively: inputs below the dendritic threshold are summed linear, inputs above this threshold are summed supra-linear and, due to the saturation, very large inputs are summed sub-linear.

If not stated otherwise, we consider only exactly simultaneous arriving spikes as sufficiently synchronous; to allow for exactly simultaneous arrivals, the synaptic delays are chosen as homogeneous $\tau_{kl} \equiv \tau$. The input currents caused by spikes that are received from the network are then given by

$$I_k^{\text{net}}(t) = \sum_{t^f} \left[\sigma_{NL} \left(\sum_{l \in M_{\text{exc}}(t^f)} \epsilon_{kl} \right) + \sum_{l \in M_{\text{inh}}(t^f)} \epsilon_{kl} \right] \delta\left(t - t^f - \tau\right). \quad (5)$$

Here, the sum over t^f denotes the sum over all times at which spike(s) are sent in the network, irrespective of which neuron(s) is (are) spiking. The sets $M_{\text{exc}}\left(t^f\right)$ and $M_{\text{inh}}\left(t^f\right)$ specify the set of neurons that send an excitatory or inhibitory spike at time t^f, respectively. (To describe a network with linear dendrites $\sigma_{NL}(\epsilon)$ is replaced by ϵ).

In section 3.3.1 we consider inhomogeneous delay distributions and finite dendritic integration window Δt (i.e., non-linear amplification of inputs received within finite time interval Δt) and discuss how the results achieved for homogeneous systems can be generalized.

2.2. NETWORK TOPOLOGY

We consider the propagation of synchrony in diluted Feed-Forward-Networks (FFNs, synfire-chains). They consist of a sequence of m layers, each composed of ω neurons. Neurons of one layer form excitatory projections to the neurons of the subsequent layer with probability p; the strength of an existing connection from neuron l to neuron k is denoted by ϵ_{kl}.

For simplicity of presentation, we consider homogeneous neuronal populations, i.e., all neurons have identical properties ($\tau_k^{\text{m}} = \tau^{\text{m}}$, $\Theta_k = \Theta$ and $V_k^{\text{reset}} = V^{\text{reset}}$ for all i), as well as homogeneous coupling strengths, i.e., $\epsilon_{kl} = \epsilon$ if a connection is realized, throughout this article. If not stated otherwise, we use $\tau^{\text{m}} = 14\,\text{ms}$ and $\Theta = 15\,\text{mV}$ as standard values for the membrane time constant and the neuron threshold.

2.3. GROUND STATE DYNAMICS

We consider networks, where the single neurons are placed in a "fluctuation driven regime," i.e., in the ground state the average input to each neuron is sub-threshold and spiking of neurons is caused by fluctuations of the inputs. This setup allows to emulate the dynamics of neurons which are part of a balanced network (van Vreeswijk and Sompolinsky, 1996, 1998). The neurons fire asynchronously and irregularly with low firing rate ν; the spike trains resemble Poissonian spike trains (Tuckwell, 1988; Brunel and Hakim, 1999; Brunel, 2000; Burkitt, 2006). Thus, the inputs to the neurons may be described by three Poissonian spike trains with rates ν^{exc} (external, excitatory), ν^{inh} (external, inhibitory) and $\nu^{\text{int}} = \nu p \omega$ (inputs from the preceding layer). Since the number of inputs N_T^X, $X \in \{\text{exc}, \text{inh}, \text{int}\}$, in a time interval T is Poisson distributed, the expected number of inputs $\left\langle N_T^X \right\rangle$ and the variance $\left\langle \left(N_T^X - \left\langle N_T^X \right\rangle\right)^2 \right\rangle$, equal $\nu^X T = \left\langle N_T^X \right\rangle = \left\langle \left(N_T^X - \left\langle N_T^X \right\rangle\right)^2 \right\rangle$. Then

$$\mu = I_0 + \tau^{\text{m}} \nu^{\text{exc}} \epsilon^{\text{exc}} + \tau^{\text{m}} \nu^{\text{inh}} \epsilon^{\text{inh}} + \tau^{\text{m}} p \omega \nu \epsilon \quad (6)$$

is the mean of the total input to the neurons in an interval of the size of the membrane time constant, $T = \tau^{\text{m}}$, and

$$\sigma^2 = \tau^{\text{m}} \nu^{\text{exc}} \left(\epsilon^{\text{exc}}\right)^2 + \tau^{\text{m}} \nu^{\text{inh}} \left(\epsilon^{\text{inh}}\right)^2 + \tau^{\text{m}} p \omega \nu \epsilon^2 \quad (7)$$

is its variance. In diffusion approximation, the distribution of membrane potentials $P_V(V)$ and the mean firing rate ν can be derived analytically (Brunel and Hakim, 1999; Brunel, 2000; Helias et al., 2010). In particular, for networks with low firing rates the probability density of membrane potentials (see, e.g., Tuckwell, 1988)

$$P_V(V) = \frac{1}{\sqrt{\pi \sigma^2}} \exp\left[-\left(\frac{V - \mu}{\sigma} \right)^2 \right] \quad (8)$$

is Gaussian and can be expressed in terms of the input current. In this approximation the average firing rate is

$$\nu = \frac{1}{\sqrt{\pi} \tau^{\text{m}}} \frac{\Theta - \mu}{\sigma} \exp\left[-\left(\frac{\Theta - \mu}{\sigma} \right)^2 \right] \quad (9)$$

and depends on μ and σ only via the quotient

$$\alpha := \frac{\Theta - \mu}{\sigma}, \quad (10)$$

which is the distance of the average input μ from the neurons' threshold Θ normalized by the standard deviation σ of the input. For the analytical derivations throughout this article we focus on the regime of low spiking rates $\left(\alpha \gtrapprox 2;\ \nu \lessapprox 1.5\text{Hz}\right)$.

In the absence of synchronous activity each neuron receives a large number of inputs from the external network and only a few inputs from the previous layer of the FFN, such that the ground state dynamics of the network is mainly established by the external inputs. To keep the input balanced we choose $\nu^{\text{exc}} = \nu^{\text{inh}} =: \nu^{\text{ext}}$ and $\epsilon^{\text{exc}} = -\epsilon^{\text{inh}} =: \epsilon^{\text{ext}}$ throughout the article.

2.4. PROPAGATION OF SYNCHRONY

To initiate propagating synchronous activity along the considered diluted FFN, we excite in the first layer a subgroup of $g_0 \leq \omega$ neurons to spike synchronously. This causes a synchronous input to the following layer after the synaptic delay τ and may therefore initiate synchronous spiking of a subgroup of neurons in that layer. These may again excite synchronous spiking in the next layer and so on. Depending on the ground state, i.e., the layout of the external network, on the layer size ω, and on the coupling strength ϵ, a synchronous pulse may or may not propagate along the FFN (cf. **Figures 1A,B,D,E**).

In addition to the triggered propagation, one might generally also expect the occurrence of spontaneous propagation of synchronous activity: Neurons of a particular layer share inputs from the previous layer and this causes correlations in their spiking

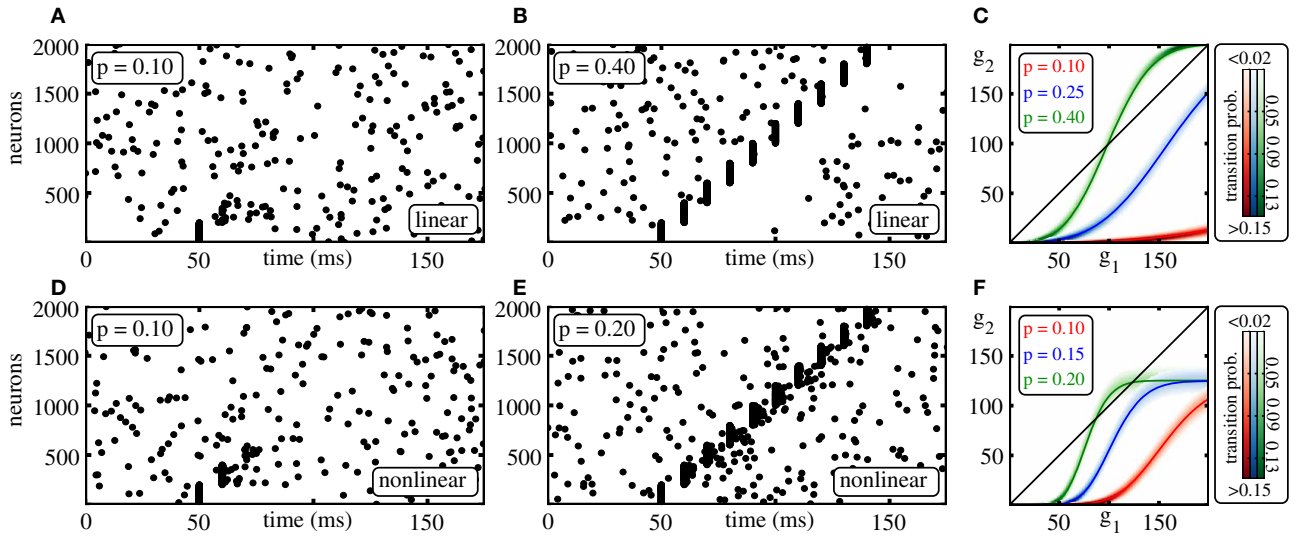


FIGURE 1 | Propagation of synchrony in diluted FFNs. (A,B,D,E) Raster plots of diluted feed-forward networks ($m = 10$, $\omega = 200$, $\epsilon = 0.25$ mV). With increasing connection probability p propagation of synchrony can be enabled **(A,B)** in networks with additive (linear) and **(D,E)** in networks with non-additive (non-linear) dendritic interactions ($\Theta_b = 4$ mV, $\kappa = 11$ mV). **(C,F)** Average number of synchronously active neurons in the second layer, g_2, vs. the number of synchronously active neurons in the initial layer, g_1; panel **(C)** linear, panel **(F)** non-additive dendritic interactions (average over $n = 10,000$ trials: solid line, transition probability: shading). Note that non-linear dendrites allow for sparser connectivity, **(E)** vs. **(B)** and for a sparser code, i.e., for smaller numbers of spiking neurons in an activated group, **(F)** vs. **(C)**.

activity. Over the layers these correlations can accumulate and lead to synchronous spiking (Aviel et al., 2003; Rosenbaum et al., 2010, 2011; Litvak et al., 2013). However, in the setups considered in this article, the effect is negligible due to two reasons: (1) each neuron receives a large number of external (uncorrelated) inputs and this background noise has a decorrelating effect, (2) we study the system near the critical point, i.e., for parameters where even synchronized spiking of all neurons of a particular layer is just sufficient to initiate a propagation of synchrony. Thus, spontaneous propagation of synchrony effectively does not occur.

We study the transition from the non-propagating to the propagating regime by means of a iterated map that yields the expectation value of the number of synchronously spiking neurons g_{i+1} in layer $i+1$ if g_i neurons are synchronously active in layer i. There is always one trivial fixed point, G_0, of this iterated map with $0 = G_0 = g_{i+1} = g_i$, which corresponds to absent activity. If $g_{i+1} < g_i$ for all $g_i > G_0$, synchronous activity will die out after a small number of layers. If $g_{i+1} \geq g_i$ for some substantial group size, $g_i > G_0$, a stable propagation of synchrony may be enabled (cf. **Figures 1C,F**). More precisely, we will show in this article that with increasing connectivity p the system undergoes a tangent bifurcation and two fixed points G_1 and $G_2 \geq G_1$ appear. If existing, G_1 is always unstable (the diagonal is crossed from below; the slope of the iterated map needs to be larger than one) and G_2 is always stable [all connections within the FFN are excitatory such that the iterated map is monotonically increasing (slope larger than zero, in particular larger than -1)]; further at G_2 there is an intersection with the diagonal from above thus the slope is smaller than one and stationary propagation with group sizes around G_2 is enabled.

In computer simulations, we determine for each given network setup by the following procedure whether a propagation is possible: after some initial time t^{init} we excite all neurons of the first layer to spike synchronously and measure the number of active neurons g_i in the ith layer at the expected spiking time $t_i^{\text{exp}} = t^{\text{init}} + i\tau$. If g_i is substantially larger than the number of active neurons arising from spontaneous activity in more than 50% of n trials (i.e., n repetition of the same simulation with different initial conditions), we denote the propagation of synchrony as successful. The critical connectivity p^*, that marks the transition from a regime where propagation of synchrony is not possible to a regime where propagation of synchrony is enabled, is found by determining the lowest connection probability p for which an initial synchronous pulse propagates successfully.

As the connections within the FFN are all excitatory, it is sufficient to check whether propagation of synchrony can be initiated by inducing synchronized spiking of all ω neurons of the first layer: Stationary propagation of synchrony can be enabled if there is a non-trivial stable fixed point (G_2) of the iterated map for the average group size. For purely excitatory connections the basin of attraction of this fixed point is bounded from the left by an unstable fixed point (G_1) and from the right by the maximum group size given by the layer size ω.

3. RESULTS AND DISCUSSION

Under which conditions can synchronous signals propagate robustly along diluted FFNs? To answer this question in detail, we first focus on networks with linear dendrites. Afterwards we study the propagation of synchrony in networks incorporating non-additive dendritic interactions and compare with the linear

case. Finally, we show that the derived results are directly applicable in biologically more detailed neuron models and network configurations.

3.1. FFNs WITH LINEAR DENDRITES

In this section, we consider linearly coupled FFNs. In the first part, we derive analytical estimates for the critical connectivity p_L^* that marks the transition from the non-propagating to the propagating regime; the initial steps follow the lines of Jahnke et al. (2012); Memmesheimer and Timme (2012). In the second part we investigate the influence of the external network on the propagation of synchrony and determine the parameter-region for which the analytical estimates are applicable. In particular, we show that the derived estimates are applicable in the biologically relevant parameter-region, where the spontaneous firing rate is low and the distribution of membrane potentials is sufficiently broad. Finally, we study how the properties of propagating synchronous pulses depend on different system parameters.

3.1.1. Analytical derivation of critical connectivity

To access the properties of propagation of synchrony we consider average numbers of active neurons in the different layers of an FFN: for this, we derive a iterated map which yields the expected number of neurons that will spike synchronously in one layer given that in the preceding layer a certain number of neurons was synchronously active.

If in the ith layer, g_i neurons spike synchronously, the number of synchronous inputs h a single neuron in layer $i+1$ receives follows a binomial distribution $h \sim B\left(g_i, p\right)$. We denote the spiking probability of a single neuron due to an input of strength x by $p_f\,(x)$. The average or expected spiking probability $p^{\mathrm{sp}}\left(g_i\right)$ of a single neuron in layer $i+1$ is then given by

$$p^{\mathrm{sp}}\left(g_i\right) = \mathrm{E}\left[p_f\left(h\epsilon\right)\middle|\, g_i\right] = \sum_{h=0}^{g_i} \binom{g_i}{h} p^h \left(1-p\right)^{g_i - h} p_f\left(h\epsilon\right). \quad (11)$$

Here and in the following we denote the expectation value of a function $f(X)$ of a random variable X by $\mathrm{E}[f(X)]$; conditional expectations are denoted by $\mathrm{E}[f(X)|Y]$. The expected number of spiking neurons in layer $i+1$ is then simply

$$\mathrm{E}\left[g_{i+1}\middle|\, g_i\right] = \omega p^{\mathrm{sp}}\left(g_i\right) \quad (12)$$

$$= \omega \sum_{h=0}^{g_i} \binom{g_i}{h} p^h \left(1-p\right)^{g_i - h} p_f\left(h\epsilon\right). \quad (13)$$

If the connection probability p is low and/or the connection strengths ϵ are small, the spontaneous spiking activity in the absence of synchrony is only weakly influenced by the spiking activity within the FFN. Thus as a starting point, we assume that the ground state is exclusively governed by external inputs (effectively setting $\epsilon_{ij} \equiv 0$). Then, the mean input to the neurons in an interval of length τ^{m} is $\mu = I_0$ with standard deviation $\sigma = \epsilon^{\mathrm{ext}}\sqrt{2\tau^{\mathrm{m}}\nu^{\mathrm{ext}}}$ (cf. section 2.3). Using the probability density (Equation 8), we calculate the spiking probability of a single neuron, $p_f\,(x)$, due to an input of strength x;

$$p_f\,(x) = \int_{\Theta - x}^{\Theta} P_V\,(V)\,dV \quad (14)$$

$$= \frac{1}{2}\left(\mathrm{Erf}\left[\frac{\Theta-\mu}{\sigma}\right] - \mathrm{Erf}\left[\frac{\Theta-\mu+x}{\sigma}\right]\right) \quad (15)$$

equals the probability of finding a neuron's membrane potential in the interval $[\Theta - x, \Theta]$. To derive a iterated map for the average number of active neurons (which maps $E[g_i] \to E[g_{i+1}]$), we interpolate $\mathrm{E}\left[g_{i+1}\middle|\, g_i\right]$ for continuous g_i and in the second step replace g_i by its expectation value $\mathrm{E}\left[g_i\right]$. The fixed points, $\mathrm{E}\left[g_{i+1}\middle|\, \mathrm{E}\left[g_i\right]\right] = \mathrm{E}\left[g_i\right]$, qualitatively determine the propagation properties of synchronous activity. In the rest of the manuscript we are dealing with the average number of active neurons in a given layer. Therefore, for simplicity we denote the expectation value of the average number of active neurons in a given layer i by g_i instead of $\mathrm{E}\left[g_i\right]$.

For sufficiently small connection probabilities p the map (Equation 12) has only one (trivial) fixed point $G_0 = g_{i+1} = g_i = 0$. Any initial synchronous pulse will die out after a small number of layers (see also **Figure 1**). With increasing connectivity two additional fixed points G_1 (unstable) and $G_2 \geq G_1$ (stable) appear via a tangent bifurcation.

For FFNs with purely excitatory couplings between the layers, the second fixed point G_2 (if it exists) is always stable: The spiking probability $p_f(x)$ is monotonically increasing with input x and thus also the iterated map (Equation 13) is monotonically increasing (i.e., the slope is larger than 0). Moreover, if G_2 exists the slope of the iterated map at this intersection point with the diagonal is smaller than 1. This implies that G_2 is stable and synchronous pulses of size $g_i \geq G_1$ typically initiate a propagation of synchrony with an average number of active neurons around G_2. The critical connectivity p_L^* at the bifurcation point marks the minimal connectivity that allows for stable propagation of synchrony.

Although the distribution of inputs from one layer to the subsequent one and the spiking probability of a single neuron $p_f(\cdot)$ are known, there is no analytic closed form solution to the fixed point equation $g_{i+1} = g_i = g_i^*$. In other words, we can compute the firing probability $p_f\,(x_0)$ for any x_0, and therefore also $\mathrm{E}\left[g_{i+1}\middle|\, g_i\right]$ for any g_i, but $g_i^* = \mathrm{E}\left[g_{i+1}\middle|\, g_i^*\right]$ is transcendental. We thus derive an approximate solution. We choose some expansion point g_i (see section 3.1.2 for details), and approximate the function $\mathrm{E}\left[g_{i+1}\middle|\, g_i^*\right]$ by a polynomial $g_i + S(g_i^* - g_i)$ in second order in $(g_i^* - g_i)$ near g_i. The arising quadratic fixed point equation $g_i^* = g_i + S(g_i^* - g_i)$ is then analytically solvable in g_i^*. This also allows to analytically compute the critical connectivity p_L^*: it is the parameter value at which the iterated map undergoes a tangent bifurcation, i.e., at which the two solutions of the fixed point equation become equal upon changing from complex-conjugate to real. Since the right hand side of Equation (13) does not offer itself for a direct series expansion in g_i^*, we derive $g_i + S(g_i^* - g_i)$ from an appropriate expansion of $p_f\,(h\epsilon)$ and a subsequent computation the arising expectation values.

In biologically relevant scenarios, the neurons usually receive a large number of synaptic inputs and thus the distribution of

membrane potentials $P_V(V)$ is broad, $P_V(V)$ changes slowly with V. Then, $P_V(V)$ around some $V = V_0$ can be approximated by considering a series expansion with a small order and it is possible to derive an approximation for the critical connectivity p_L^* based on an expansion of $p_f(\cdot)$. Expanding $p_f(x)$ into a Taylor series around some x_0 and using Equation (12) yields

$$g_{i+1} = \omega \mathrm{E}\left[\sum_{n=0}^{\infty} \frac{p_f^{(n)}(x_0)}{n!} (h\epsilon - x_0)^n \middle| g_i \right] \tag{16}$$

$$= \omega \sum_{n=0}^{\infty} \frac{p_f^{(n)}(x_0)}{n!} \mathrm{E}\left[(h\epsilon - x_0)^n \middle| g_i \right]. \tag{17}$$

Here and in the following we denote the nth derivative of a function $f(x)$ at $x = x_0$ by

$$f^{(n)}(x_0) = \left. \frac{d}{d^n x} f(x) \right|_{x=x_0}. \tag{18}$$

Replacing the derivatives of $p_f(\cdot)$ by the (one order lower) derivatives of probability density of membrane potentials $P_V(V)$ according to Equation (14) yields

$$g_{i+1} = \omega p_f(x_0) + \omega \sum_{n=1}^{\infty} \frac{P_V^{(n-1)}(V_0)}{(-1)^{n-1} n!} \mathrm{E}\left[(h\epsilon - x_0)^n \middle| g_i \right], \tag{19}$$

where we defined

$$V_0 := \Theta - x_0 \tag{20}$$

for better readability.

We have recently shown (Jahnke et al., 2012) that it is possible to derive a scaling law for the critical connectivity using

$$x_0 = g_i p \epsilon, \tag{21}$$

the (unknown) average input from one layer to the next during stationary synchrony propagation, as expansion point. For this choice the expectation value $\mathrm{E}\left[(h\epsilon - x_0)^n \middle| g_i \right]$ in Equation (19) simplifies to

$$\mathrm{E}\left[(h\epsilon - x_0)^n \middle| g_i \right] = \epsilon^n \mathrm{E}\left[(h - \mathrm{E}[h])^n \middle| g_i \right] = \epsilon^n m_n, \tag{22}$$

where we denote by m_n the nth central moment of the Binomial distribution $B(g_i, p)$, specifying the distribution of inputs to the $(i+1)$th layer. In the limit of large layer sizes ω and small coupling strengths ϵ keeping the maximal input $\epsilon\omega$ to each layer constant (to preserve the network state), all summands for $n \geq 2$ vanish, and Equation (19) simplifies to

$$g_{i+1} = \omega p_f\left(g_i p \epsilon\right). \tag{23}$$

Using the implicit function theorem one can show that this implies the scaling law

$$p_L^* = \frac{1}{\lambda \epsilon \omega} \tag{24}$$

where λ is a constant independent of ϵ and ω (Jahnke et al., 2012). We note that for the derivation of the scaling law (Equation 24) we did not use the actual functional form of the distribution of membrane potentials $P_V(V)$. Therefore this estimate holds if $P_V(V)$ is sufficiently slow changing with V such that the Taylor expansion (cf. Equation 16) is applicable, but its validity is not restricted to the low-rate approximation.

However, the dependence of the prefactor $1/\lambda$ on the layout of the external network remained unknown. Here, we present an approach that enables us to derive an approximate value for λ. We consider the expansion (Equation 19) around x_0 up to second order,

$$g_{i+1} \approx \omega p_f(x_0) + \omega P_V(V_0) \cdot \left(\epsilon g_i p - x_0\right) - \frac{\omega P_V^{(1)}(V_0)}{2} \left[\left(\epsilon g_i p - x_0\right)^2 + \epsilon^2 g_i p \left(1 - p\right) \right] \tag{25}$$

The truncated series (Equation 25) is quadratic in g_i such that the fixed points $g_{1/2}^* = g_{i+1} = g_i$ can be obtained analytically,

$$g_{1,2}^* = \gamma_L \pm \sqrt{\gamma_L^2 - \frac{x_0 \left(2P_V(V_0) + x_0 P_V^{(1)}(V_0)\right) - 2p_f(x_0)}{p^2 P_V^{(1)}(V_0) \epsilon^2}}, \tag{26}$$

where we defined

$$\gamma_L := \frac{p\epsilon\omega \left(2\left(P_V(V_0) + x_0 P_V^{(1)}(V_0)\right) + \left(p-1\right) P_V^{(1)}(V_0)\epsilon\right) - 2}{2p^2 P_V^{(1)}(V_0)\epsilon^2\omega}. \tag{27}$$

At the bifurcation point, the root in Equation (26) vanishes such that both fixed points agree ($g_1^* = g_2^*$) and $\gamma_L = g_1^* = g_2^*$ specifies the average size of a propagating synchronous pulse. Consequently, the critical connectivity is obtained by choosing p such that

$$\gamma_L^2 = \frac{x_0 \left(2P_V(V_0) + x_0 P_V^{(1)}(V_0)\right) - 2p_f(x_0)}{p^2 P_V^{(1)}(V_0)\epsilon^2} \tag{28}$$

which yields

$$p_L^* = \frac{1}{2} - \frac{1}{\epsilon} \left[\frac{\lambda^*}{P_V^{(1)}(V_0)} - \sqrt{\frac{2}{P_V^{(1)}(V_0)\omega} + \frac{\left(\epsilon P_V^{(1)}(V_0) - 2\lambda^*\right)^2}{4\left(P_V^{(1)}(V_0)\right)^2}} \right] \tag{29}$$

where we defined

$$\lambda^* := P_V(V_0) + x_0 P_V^{(1)}(V_0) - \sqrt{P_V^{(1)}(V_0)\left(x_0\left(2P_V(V_0) + x_0 P_V^{(1)}(V_0)\right) - 2p_f(x_0)\right)} \tag{30}$$

which is independent of the setup of the FFN and completely determined by the layout of the external network and the choice of the expansion point x_0.

As before we consider the limit of large layer sizes ω and small coupling strengths ϵ, i.e., we replace $\omega \to \frac{\text{const}}{\epsilon}$ and consider the leading terms of a series expansion of Equation (29). The expansion of the square bracket in Equation (29) yields

$$\frac{\lambda^*}{P_V^{(1)}(V_0)} - \sqrt{\frac{2}{P_V^{(1)}(V_0)}\frac{\epsilon}{\text{const}} + \frac{\left(\epsilon P_V^{(1)}(V_0) - 2\lambda^*\right)^2}{4\left(P_V^{(1)}(V_0)\right)^2}}$$

$$= \left[\frac{\lambda^*}{P_V^{(1)}(V_0)} - \frac{\lambda^*}{P_V^{(1)}(V_0)}\right] - \epsilon\left(\frac{1}{\lambda^* \cdot \text{const}} - \frac{1}{2}\right) + O\left(\epsilon^2\right), \quad (31)$$

such that the critical connectivity assumes the functional form given by Equation (24),

$$p_L^* \approx \frac{1}{\lambda^* \epsilon \omega}. \quad (32)$$

Thus $\lambda = \lambda^*$ defined by Equation (30) provides an approximation of the constant λ fully specifying the critical connectivity p_L^*.

3.1.2. Optimal expansion point

To derive Equation (30) we assumed that it is sufficient to consider the second order expansion of $p_f(x)$. It is thus necessary to choose an appropriate expansion point that results in fast convergence. In particular for the choice $x_0 = x_0^*$, that we will now derive, Equation (37) below, the bifurcation diagram near the bifurcation point is well approximated already for $k = 2$ (cf. **Figure 2**).

The size of a propagating group at the critical connectivity is γ_L (cf. Equation 27) and thus the resulting average input is $p_L^* \gamma_L \epsilon$. Our expansion point x_0 should lie near to this value, which is, of course, unknown prior to solving the fixed point equation. We will thus compute a range in which $p_L^* \gamma_L \epsilon$ has to lie and choose the expansion point appropriately within. We assume that ω is large and employ Equation (23) which allows an direct estimate of this range as we know the functional form explicitly. Equation (23) with $g_{i+1} = g_i$ is just another transcendental equation for the fixed points and it has zero, one, or two non-trivial fixed point solutions points g_1^* and g_2^*, which are then also solutions of Equation (19) with $g_{i+1} = g_i$. At the bifurcation point ($g^* = g_1^* = g_2^*$) where the diagonal is touched, the function $p_f(gp\epsilon)$ has to be concave and monotonic increasing with respect to g. The definition (Equation 14) of $p_f(x)$ implies that it is monotonic increasing for all $x \geq 0$. Moreover, it is concave for all $x \geq \Theta - \mu$,

$$p_f^{(1)}(x) = P_V(\Theta - x) \geq 0 \quad \text{for } x \geq 0 \quad (33)$$

$$p_f^{(2)}(x) = -P_V^{(1)}(\Theta - x) \leq 0 \quad \text{for } x \geq \Theta - \mu, \quad (34)$$

such that the bifurcation point satisfies

$$x_0 \geq \Theta - \mu. \quad (35)$$

The condition Equation (33) holds because $P_V(V) \geq 0$ is a probability density and Equation (34) is derived directly from differentiating Equation (8). To maximize the quality of the second order approximation Equation (25), we choose $x_0 = x_0^*$ such that the contribution to the expansion (Equation 19) of the $k = 3$rd order term equals zero. According to Equation (19), all 3rd order terms are proportional to $P_V^{(2)}(\Theta - x_0)$; so we determine the expansion point x_0^* as a deflection point of $P_V(\cdot)$, requiring that the second derivative of $P_V(V)$ vanishes for $V = \Theta - x_0^*$,

$$p_f^{(3)}(x_0^*) = \left.\frac{d^2 P_V(V)}{dV^2}\right|_{V=\Theta-x_0^*} \stackrel{!}{=} 0. \quad (36)$$

In the considered regime of low spiking rates, we find $x_0^* = \Theta - \mu \pm \frac{\sigma}{\sqrt{2}}$, cf. Equation (8). Due to Equation (35)

$$x_0^* = \Theta - \mu + \frac{\sigma}{\sqrt{2}}. \quad (37)$$

For $x_0 = x_0^*$ the bifurcation diagram near the bifurcation point is well approximated already for $k = 2$ (cf. **Figure 2**) and Equation (30) provides a good estimate of the critical connectivity p_L^* (cf. **Figure 3**).

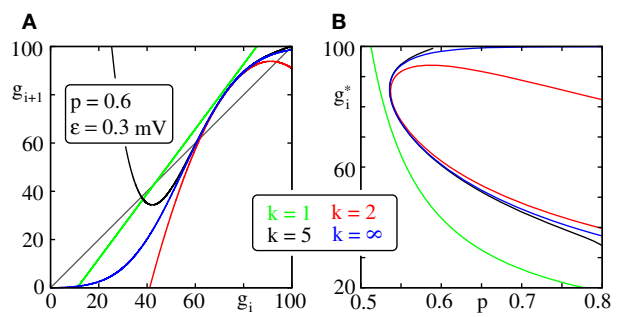


FIGURE 2 | Iterated map and bifurcation diagram for the average group size of a propagating synchronous pulse. (A) Iterated map (Equation 19) truncated after expansion order k (color code) with $x_0 = x_0^*$ (cf. Equation 37). **(B)** Fixed points of the iterated maps shown in **(A)**; with increasing connectivity two fixed points appear by a saddle node bifurcation. We note that already a second order expansion (red), i.e., the lowest order at which a saddle node bifurcation can occur, approximates the bifurcation diagram (blue) near the bifurcation point well.

3.1.3. Influence of external network

In the previous section we derived an iterated map for the average group size (cf. Equation 13) and an approximation for the critical connectivity p_L^* (cf. Equations 30 and 32) that marks the transition from FFNs which do not support propagation of synchrony to FFNs that do. In this section we focus on the robustness of our results. How does the critical connectivity change with the layout of the external network? For which parameter range does the estimate of the critical connectivity (given by Equations 30 and 32) yield reasonable results?

The derivation was based on the assumption that the ground state dynamics of the neurons of the FFN is completely determined by the external inputs. This assumption holds if the spontaneous firing rate ν of the neurons and/or the coupling strengths

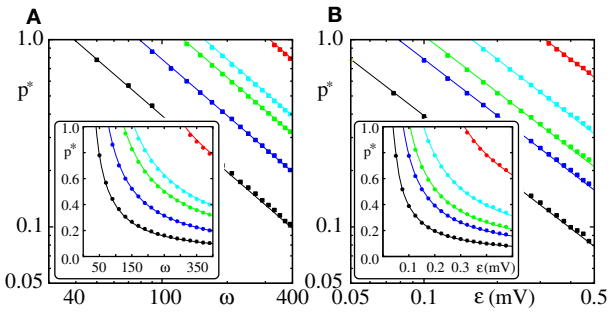


FIGURE 3 | Critical connectivity p_L^* in FFNs with linear dendrites decays algebraically with coupling strength ϵ and layer size ω. The parameters of the external inputs (emulated embedding network) are fixed ($I_0 = 5$ mV, $\nu^{\text{ext}} = 3$ kHz, $\epsilon^{\text{ext}} = 0.5$ mV). Panel **(A)** shows the critical connectivity p_L^* vs. the layer size ω for different coupling strengths ($\epsilon = \{0.05$ mV (red), 0.1 mV (cyan), 0.125 mV (green), 0.2 mV (blue), and 0.4 mV (black)$\}$) and panel **(B)** shows p_L^* vs. the coupling strength ϵ for different layer sizes ($\omega = \{50$ (red), 100 (cyan), 150 (green), 200 (blue), and 400 (black)$\}$). In the main panels we use a logarithmic scale, the insets have a linear scale. The squares indicate the connectivity above which a synchronous pulse propagates from the 1st to the 20th layer of a FFN in at least 50% of $n = 30$ trials. The critical connectivity given by Equation (32) (solid lines) with $x_0 = x_0^*$ (cf. Equation 37) is in good agreement with computer simulations. As predicted $p_L^* \propto (\epsilon\omega)^{-1}$ and the proportionality factor $1/\lambda$ is well approximated by the estimate $1/\lambda^*$ derived in Equation (30).

ϵ and/or the connectivity p are sufficiently small. We will generalize our approach and show how the impact of preceding layers on a layer's ground state can be taken into account. Thereafter we will compare the results with computer simulations, identify the regions in parameter space for which the derived approximations hold and discuss deviations between direct numerical simulations and analytics.

The first layer of an FFN receives inputs only from the external network and according to Equations (6, 7) the mean μ_1 and standard deviation σ_1 of its input is

$$\mu_1 = I_0 \tag{38}$$

$$\sigma_1 = \epsilon^{\text{ext}}\sqrt{2\tau^{\text{m}}\nu^{\text{ext}}}, \tag{39}$$

as assumed in the previous section. All following layers receive external inputs and spikes from their preceding layer(s). The mean μ_n and standard deviation σ_n of the input to neurons of the nth layer (with $n \geq 2$) reads (cf. Equations 6 and 7)

$$\mu_n = I_0 + \tau^{\text{m}} p\omega\nu_{n-1}\epsilon \tag{40}$$

$$\sigma_n = \sqrt{2\nu^{\text{ext}}\tau^{\text{m}}\,(\epsilon^{\text{ext}})^2 + p\omega\nu_{n-1}\tau^{\text{m}}\epsilon^2}. \tag{41}$$

Here we denote the spontaneous firing rate (in the absence of synchrony) of neurons of the $(n-1)$th layer by ν_{n-1}. It is given by Equation (9) as

$$\nu_{n-1} = \frac{1}{\sqrt{\pi}\tau^{\text{m}}}\frac{\Theta - \mu_{n-1}}{\sigma_{n-1}}\exp\left[-\left(\frac{\Theta - \mu_{n-1}}{\sigma_{n-1}}\right)^2\right]. \tag{42}$$

From layer to layer, the mean input, the standard deviation as well as the firing rate increase. For setups, where the ground state of the FFN is non-pathological, i.e., the firing rates of all layers are bounded, the additional corrections $\Delta X_n := X_n - X_{n-1}$ for $X \in \{\mu, \sigma, \nu\}$ decrease with n, and μ_n, σ_n and ν_n saturate for sufficiently large n. Thus, μ_∞ and σ_∞ describe the input to the neurons of an infinitely long FFN and the single neurons of such an FFN spike with an average rate ν_∞. Accordingly, replacing μ and σ by μ_∞ and σ_∞ in Equation (13) [where they appear as parameters of $p_f(\cdot)$] yields an iterated map for the average group size.

In **Figure 4**, we compare the critical connectivity found by numerically determining the bifurcation point of the iterated map (Equation 13) (i.e., we determined the connectivity p for which the iterated map touches the diagonal; solid lines) with computer simulations of propagating synchrony (markers). To also cover scenarios, where the input from the preceding layer is not negligible, we consider infinitely long FFNs (then, the distribution of membrane potentials is equal in all layers). In computer simulations this can be approximated by a sufficiently long FFN with periodic boundary conditions, i.e., an FFN where the last layer connects to the first layer. For moderate external inputs, i.e., moderate I_0 and ϵ^{ext}, already the analytical results neglecting the influence of the preceding layers (using μ_1 and σ_1) agree well with computer simulations (cf. **Figure 4A**, solid lines). However, for large external inputs, i.e., large I_0 and ϵ^{ext}, the critical connectivity is overestimated. Here, the assumption that the distribution of membrane potentials is not influenced by the connectivity of the FFN does not hold. The additional input shifts the membrane potentials to higher values and consequently a lower connectivity is required for a propagation of a synchronous pulse. The corrections given by Equations (38–42) account for these deviations to some extent (cf. **Figures 4B,C**; solid lines), in particular for setups where the spontaneous firing rate is low. However, for very large I_0 and ϵ^{ext}, the critical connectivity is under-estimated. Here, the spontaneous firing rate is too high and the low-rate approximation, Equations (8–9), is not adequate to describe the system; the firing rate and thus the mean input from the previous layer are over-estimated. This becomes particularly clear in **Figure 4C**, where we show the critical connectivity as a function of the strength of the external inputs ϵ^{ext}. For any given I_0 (different colors), the critical connectivity for small ϵ^{ext} is well approximated; with increasing ϵ^{ext} the firing rate increases [α decreases and thus ν increases; cf. Equations (9 and 10)] and when the coupling strengths ϵ^{ext} exceed a I_0-dependent threshold, the low-rate approximation becomes inapplicable.

Applying the methods in Brunel and Hakim (1999); Brunel (2000), the firing rate and the distribution of membrane potentials can be derived in diffusion approximation for states with higher spontaneous firing rates. Although most of the analytical considerations above are also applicable within this approximation, the determination of an optimal expansion point (cf. Equations 36 and 37) becomes more difficult and a closed form expression does not exist. However, the critical connectivity can be obtained by numerically determining the fixed points of the iterated map (Equation 13) and we find that it agrees with

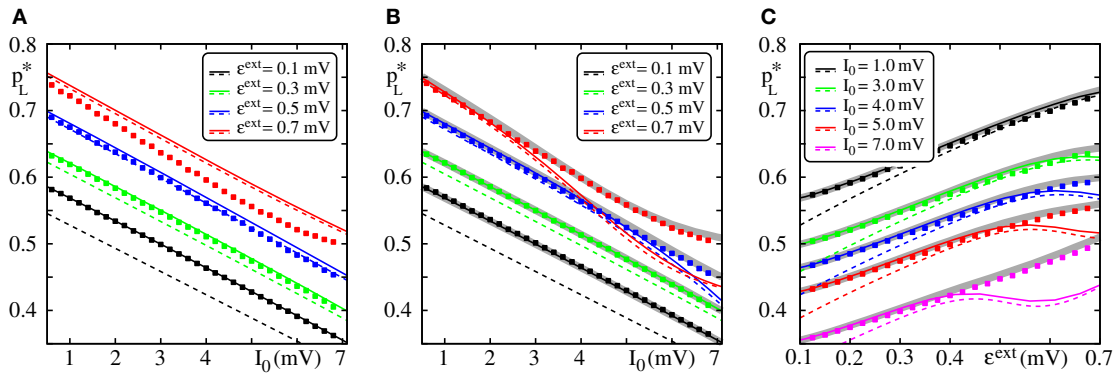


FIGURE 4 | Robustness of analytical estimates of the critical connectivity. (A–C) We consider the critical connectivity p_L^* of infinitely long FFNs, that are approximated by an FFN ($m = 20$, $\omega = 150$, $\epsilon = 0.2$ mV) with periodic boundary conditions in direct numerical simulations (markers), for different layouts of the external network. Panels **(A,B)** show p_L^* vs. I_0 for fixed ϵ^{ext} and panel **(C)** shows p_L^* vs. ϵ^{ext} for fixed I_0. The solid (colored) lines indicate the critical connectivity found by numerically determining the bifurcation point of the iterated map (Equation 13). In panel **(A)** we neglect the influence of previous layers on the ground state of a considered layer in the analytical computations [i.e., we use μ_1 and σ_1, cf. Equations (38) and (39)]. In **(B,C)** we employ corrections to account for their influence, cf. Equations (38–42). We show the third order correction, higher orders add only small modifications to the curves, but the numerical computations get more costly. The thick gray lines in **(B,C)** indicate the bifurcation point of the iterated map (Equation 13) with $P_V(V)$ derived from the diffusion approximation of leaky integrate-and-fire neuron dynamics with Poissonian input (Brunel and Hakim, 1999; Brunel, 2000). The dashed lines are the estimates of the critical connectivity given by Equations (30 and 32). Again, in panel **(A)** we neglect the influence of previous groups on the ground state, in panels **(B,C)** we use the third order correction. The estimates agree with the data from numerical simulations within the biologically relevant parameter range, where (1) the spontaneous spiking activity is low and (2) the distribution of membrane potentials is sufficiently broad. For further explanations see text (section 3.1.3).

computer simulations for the entire considered range of I_0 and ϵ^{ext}, (cf. **Figures 4B,C**; gray lines).

Analogous to the approach presented above, corrections for the influence of preceding layers can be taken into account for the analytical estimate of the critical connectivity derived in the previous section (Equations 30 and 32). Replacing the connectivity p by the approximation $p_L^* = (\lambda^* \epsilon \omega)^{-1}$ in Equations (40, 41) yields

$$\mu_n = I_0 + \tau^m / \lambda^*_{n-1} \nu_{n-1} \tag{43}$$

$$\sigma_n = \sqrt{2\nu^{ext}\tau^m (\epsilon^{ext})^2 + \epsilon \nu_{n-1} \tau^m / \lambda^*_{n-1}} \tag{44}$$

where $\lambda^*_{n-1} := \lambda^*(\mu_{n-1}, \sigma_{n-1})$ is given by Equation (30) and $\nu_{n-1} = \nu(\mu_{n-1}, \sigma_{n-1})$ is given by Equation (42). In **Figure 4** we show the estimate of the critical connectivity $p_L^* = \left(\lambda_n^* \epsilon \omega\right)^{-1}$ (cf. Equation 32) using λ_1^* (panel a; dashed line), i.e., neglecting the influence of the preceding layers, and using a higher correction order (panel b,c; dashed line: third order). For sufficiently large ϵ^{ext} the critical connectivity found by numerically determining the bifurcation point agrees with the analytical estimate given by Equation (32). As discussed above, the corrections Equations (43, 44) account for the deviations from the simulated data as long as the total spontaneous firing rate is sufficiently low. However, for small ϵ^{ext} the critical connectivity is under-estimated. Here, the standard deviation of the inputs (cf. Equation 7) is low, such that the distribution of membrane potentials $P_V(V)$ is narrow [for $\epsilon^{ext} \to 0$: $P_V(V) \to \delta(V - \mu)$; cf. Equation (8)], the spiking probability of one neuron, $p_f(\cdot)$, increases steeply in a small interval [for $\epsilon^{ext} \to 0$: $p_f(x) \to \Theta(x - \mu)$; cf. Equation (8)] and thus the approximation of $p_f(\cdot)$ by the leading terms of a Taylor expansion is not sufficiently accurate.

However, in the biologically plausible parameter regime, where the firing rates are small and the distribution of membrane potentials is broad, the critical connectivity is approximated well by Equation (32) together with Equation (30) (defining λ^*), Equation (37) (defining x_0^*) and the corrections that account for the influence of the preceding layers, Equations (43, 44).

3.1.4. Characteristics of propagating synchronous pulses

In the previous sections, we have shown that a synchronous pulse may propagate along a diluted FFN. In this section we study the characteristics and properties of a propagating synchronous signal. We consider them at the transition to stable propagation, p_L^*, because there they depend only weakly on the network setup. How large is the fraction of neurons that participate in propagating synchrony? How does this fraction depend on the network setup?

To answer such questions, we consider the effect of a propagation synchronous pulse on the single layers in the network, as a measure for the effective pulse size. In other words, we consider the mean input μ_L a neuron receives from the preceding layer if a synchronous pulse propagates along the FFN at the critical connectivity p_L^*. It is given by the product of the connection probability p_L^*, the connection strength ϵ and the average size of a propagating synchronous signal γ_L; using Equations (27) and (29) yields

$$\mu_L = \gamma_L p_L^* \epsilon = \frac{P_V(\Theta - x_0^*) + P_V^{(1)}(\Theta - x_0^*) x_0^* - \lambda^*}{P_V^{(1)}(\Theta - x_0^*)} \tag{45}$$

and after inserting λ^* as given by Equation (30),

$$\mu_L = \sqrt{\frac{x_0^*\left(2P_V(\Theta - x_0^*) + x_0^* P_V^{(1)}\left(\Theta - x_0^*\right)\right) - 2p_f\left(x_0^*\right)}{P_V^{(1)}\left(\Theta - x_0^*\right)}}. \quad (46)$$

According to Equation (46) the average input μ_L to the neurons due to a propagation of a synchronous pulse is independent of the layer size ω and coupling strength ϵ. For setups with moderate external inputs (i.e., inputs of the preceding layer influence the neurons' ground state only weakly; see also section 3.1.3) the distribution of membrane potentials $P_V(\cdot)$ (cf. Equation 8), the firing probability of single neurons $p_f(\cdot)$ (cf. Equation 14) as well as the expansion point (deflection point of $P_V(\cdot)$; cf. Equation 37)

$$x_0^* = \Theta - I_0 + \epsilon^{\text{ext}}\sqrt{\tau^{\text{m}}\nu^{\text{ext}}} \quad (47)$$

are fully determined by the external inputs (I_0, ν^{ext} and ϵ^{ext}). **Figures 5A,B** illustrates the dependence of μ_L on the layout of the external network and the FFN: as expected from our analytical considerations, the dependence on the layer size and coupling strength is weak when I_0 and ϵ^{ext} are kept fixed. With increasing mean of the external input (I_0) the distribution of membrane potentials $P_V(V)$ is shifted toward the threshold Θ, such that it is more likely to find the membrane potential of the neurons near the threshold and the critical connectivity decreases (cf. also **Figures 4A,B**). Naturally this implies a decreasing average input μ_L at p_L^*, which is shown in **Figure 5A** for different external couplings ϵ^{ext} and parameters of the FFN. Increasing the external coupling strength ϵ^{ext} (and with it the variance of the external input) causes a broadening of the distribution of membrane potentials; the membrane potentials of some neurons are shifted toward the threshold and the membrane potentials of other neurons are shifted away from it. If the fraction of neurons that participate in the propagation of the synchronous pulse is large, this implies an increasing critical connectivity (**Figure 5B**; cf. also **Figure 4C**).

The spiking probability of a single neuron due to the mean input μ_L equals the average fraction p^{frac} of neurons of one layer that participate in a propagating synchronous pulse,

$$p^{\text{frac}} = \frac{\gamma_L}{\omega} = p_f(\mu_L). \quad (48)$$

Interestingly, in the considered regime of low spiking rates and sufficiently broad distribution of membrane potentials, where the approximations given in section 3.1.1 are applicable, p^{frac} depends on the setup of the external inputs only via the quotient $\alpha = \frac{\Theta - \mu}{\sigma}$ (cf. Equation 10), or, equivalently, on the spontaneous firing rate ν of the neurons (cf. Equation 9). This can be shown by combining Equations (8, 37) and (Equation 46),

$$\mu_L = \sigma\left(\frac{e\pi}{2}\right)^{1/4}\left[\frac{\left(\sqrt{2}+2\alpha\right)\left(3+\sqrt{2}\alpha\right)}{2\sqrt{e\pi}} - \text{Erf}\left(\frac{1}{\sqrt{2}}\right) - \text{Erf}(\alpha)\right]^{1/2} \quad (49)$$

$$=: \sigma f_\mu(\alpha) \quad (50)$$

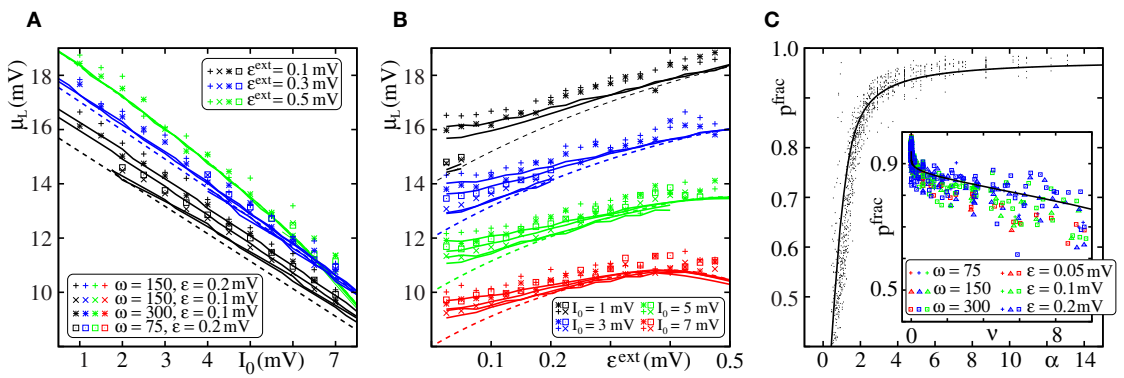


FIGURE 5 | Properties of propagating synchronous pulses at the transition from the no-propagation to the propagation regime. Panels **(A,B)** show the mean input μ_L that a layer receives due to a propagating synchronous pulse in the preceding layer. μ_L measures the effective pulse size (the impact of a propagating synchronous pulse) and is mainly determined by the external inputs rather than by the setup of the FFN. In **(A)** the variance of the external input (measured by ϵ^{ext}) is fixed and μ_L is plotted vs. I_0; in **(B)** the mean external input I_0 is fixed and μ_L is plotted vs. ϵ^{ext}. The markers indicate μ_L for FFNs of different sizes [ω and ϵ are given by the legend in **(A)**] obtained by numerical simulations of propagating synchrony. The dashed lines shows the approximation of μ_L given by Equation (46) (which is independent of ω and ϵ); the solid lines indicate $\mu_L = p_L^* G_2 \epsilon$; values of p_L^* and G_2 are found semi-analytically, by numerically identifying the bifurcation point of the analytically derived iterated map (Equation 13) for the different network setups (both analytical estimates are corrected for the influence of inputs from the preceding layer up to the first order). Panel **(C)** shows the fraction p^{frac} of neurons in a layer that participate in the propagation of a synchronous signal vs. α [(Equation 10); main panel] and vs. the spontaneous firing rate ν (inset). Data from different network setups are plotted without distinction as black dots in the main panel and with distinction by different colors and symbols in the inset (see legend); Simulations are repeated for different layouts of the external network ($I_0 \in \{1, 3, \ldots, 11\}$ mV; $\epsilon^{\text{ext}} \in \{0.1, 0.125, \ldots, 1.0\}$ mV). The solid lines indicate $p_f(\mu_L) = f_p(\alpha)$ as given by Equation (53). The layer size ω as well as the coupling strength ϵ influence p^{frac} only weakly. p^{frac} depends on the network setup mainly through α or, equivalently, through ν (cf. Equation 9): Measurement values from different network setups largely collapse to the graph of the function $p_f(\mu_L) = f_p(\alpha)$. For further explanations see text (section 3.1.4).

such that

$$p_f(\mu_L) = \frac{1}{2}\left[\mathrm{Erf}\left(\frac{\Theta-\mu}{\sigma}\right) + \mathrm{Erf}\left(\frac{\mu_L-\Theta+\mu}{\sigma}\right)\right] \quad (51)$$

$$= \frac{1}{2}\left[\mathrm{Erf}(\alpha) + \mathrm{Erf}\left(\frac{\mu_L}{\sigma}-\alpha\right)\right] \quad (52)$$

$$= \frac{1}{2}\left[\mathrm{Erf}(\alpha) + \mathrm{Erf}\left(f_\mu(\alpha)-\alpha\right)\right] =: f_p(\alpha). \quad (53)$$

In **Figure 5C** we compare the above predictions with direct numerical simulations: For different layer sizes ω, coupling strengths ϵ and layouts of the external networks (i.e., different values of I_0 and ϵ^{ext}), we detect whether propagation of a synchronous pulse is possible and if so, we numerically determine the average fraction of participating neurons as well as the spontaneous firing frequency. We find that indeed the size of the synchronous pulse is determined essentially by the quotient $\alpha = \frac{\Theta-\mu}{\sigma}$ and Equation (53) is a reasonable estimate of the average fraction of neurons spiking in each layer. With increasing α the fraction of participating neurons increases, it thus decreases with spontaneous firing rate ν see **Figure 5C**. For FFNs with low spontaneous spiking frequency almost all neurons of a layer participate in the propagation of a synchronous pulse.

3.2. FFNs WITH NON-LINEAR DENDRITES

In this section, we investigate propagation of synchrony mediated by dendritic non-linearities. Although the mechanism underlying the propagation is generally related to that in linear networks, the discontinuities introduced by non-additive dendritic interactions prevent a similar analytical approach. In the first part of this section, we thus derive analytical estimates for the critical connectivity p^*_{NL} in non-linearly coupled networks based on a self-consistency approach (see also Jahnke et al., 2012). In the second part, we study the transition from propagation of synchrony mediated by linear dendrites to propagation of synchrony mediated by non-additive dendritic interactions upon increasing the degree of non-linearity in the networks. In the last part, we evaluate the robustness of the analytical estimates with respect to the layout of the external network.

3.2.1. Analytical derivation of critical connectivity

Neurons with non-additive dendritic interaction process excitatory input by a non-linear dendritic modulation function σ_{NL} (see section 2.1), i.e., synchronous inputs that exceed the dendritic threshold Θ_b are amplified to an effective input of size κ (cf. Equation 4). Therefore the spiking probability of a single neuron due to a synchronous input of strength x, $p_f(\sigma_{NL}(x))$, is discontinuous and an approach based on expansion of $p_f(\cdot)$ is inappropriate. To derive an analytical expression for the critical connectivity p^*_{NL} in FFNs incorporating dendritic non-linearities, we consider the (average) fraction of neurons of one layer, p_γ, that receive an input x larger than the dendritic threshold, $x \geq \Theta_b$, due to the propagating synchronous pulse. If there is a stable (stationary) propagation of synchrony established, p_γ is constant throughout the layers, which allows us to formulate a self-consistency equation. The basic derivations have been published recently (Jahnke et al., 2012) and will be briefly reviewed in the following for the readers convenience.

For sufficiently small dendritic thresholds Θ_b and sufficiently large κ, the spiking probability of a neuron due to a sub-threshold input is small compared to the spiking probability of a supra-threshold input. Therefore, we approximate the spiking probability of a single neuron in response to a synchronous input of strength x by

$$p_f(\sigma_{NL}(x)) = \begin{cases} p_f(\kappa) & \text{if} \quad x \geq \Theta_b \\ 0 & \text{otherwise} \end{cases}, \quad (54)$$

i.e., we assume that somatic spikes due to the synchronous pulse are exclusively generated by dendritically enhanced inputs. We denote the fraction of neurons that receive a dendritic spike by p_γ. This may be considered as constant throughout the different layers if stable propagation of synchrony is enabled. Then the probability that a neuron receives exactly k inputs from the preceding layer follows a binomial distribution $k \sim B\left(\omega, p_\gamma p_f(\kappa)\, p\right)$, where $p_\gamma p_f(\kappa)\, p$ is the probability that (1) a neuron of the preceding layer receives a supra-threshold input (p_γ), (2) a somatic spike is elicited by that input $\left(p_f(\kappa)\right)$ and there is a connection from this spiking neuron to the considered neuron of the following layer (p). So we can formulate the self-consistency equation for p_γ,

$$p_\gamma = \sum_{k=\lceil\Theta_b/\epsilon\rceil}^{\omega} \binom{\omega}{k} \left(p_\gamma p_f(\kappa)\, p\right)^k \left(1 - p_\gamma p_f(\kappa)\, p\right)^{\omega-k}. \quad (55)$$

To solve Equation (55) we approximate the binomial distribution by a Gaussian distribution with mean $\delta := \omega p_\gamma p p_f(\kappa)$ and standard deviation $\sigma_\delta := \sqrt{\delta(1 - p_\gamma p p_f(\kappa))}$, which yields

$$p_\gamma = \frac{1}{2}\left[1 + \mathrm{Erf}\left(\frac{n}{\sqrt{2}}\right)\right], \quad (56)$$

where we defined

$$n := \frac{\delta - \Theta_b/\epsilon}{\sigma_\delta} \quad (57)$$

$$= \frac{\omega p_\gamma p p_f(\kappa) - \Theta_b/\epsilon}{\sqrt{\omega p_\gamma p p_f(\kappa)\left(1 - p_\gamma p p_f(\kappa)\right)}} \quad (58)$$

as the difference between the average number of inputs (δ) and the number of inputs needed to reach the dendritic threshold (Θ_b/ϵ) normalized by the standard deviation of the number of inputs (σ_δ). Solving definition (Equation 58) for p and replacing p_γ by Equation (56) yields

$$p_{NL} = \frac{n^2\epsilon + 2\Theta_b + n\sqrt{n^2\epsilon^2 + 4\Theta_b\left(\epsilon - \frac{\Theta_b}{\omega}\right)}}{p_f(\kappa)\epsilon(n^2+\omega)\left(1 + \mathrm{Erf}\left(\frac{n}{\sqrt{2}}\right)\right)}, \quad (59)$$

which is the connectivity p_{NL} where stable propagation of synchrony with some given n (or, equivalently, some given p_γ;

cf. Equation 56) is established. We note that a propagation of synchrony mediated by dendritic spikes requires

$$\epsilon\omega > \Theta_b \tag{60}$$

(otherwise even the input caused by a synchronized spiking of all neurons of a layer in a fully connected FFN ($p = 1$) is not sufficient to reach the dendritic threshold Θ_b).

For parameters fulfilling the inequality (Equation 60), $p_{NL}(n)$ has a global minimum (see Appendix) and the critical connectivity p^*_{NL}, again defined as the smallest connectivity that allows for a stable propagation of synchrony, matches that global minimum: any connectivity p_{NL} above the minimal connectivity p^*_{NL} has two preimages n_1 and n_2 corresponding to the both non-trivial fixed points G_1 and G_2 of the iterated map for the average group size (cf. **Figure 1** and section 2.4). However, there exists smaller connectivities for which a stationary propagation can be established. At the global minima p^*_{NL} both preimages n_1 and n_2 collapse to $n^* = n_1 = n_2$ and correspond to the fixed point $G = G_1 = G_2$ of the iterated map at the bifurcation point of the tangent bifurcation. Here the transition from the regime where no propagation of synchrony is possible to the regime where a propagation of synchrony is enabled takes place. For p_{NL} smaller than p^*_{NL} there are no preimages (i.e., a stationary propagation of synchrony mediated by non-additive dendritic interactions cannot be established); this scenario correspond to the absence of the non-trivial fixed points of the iterated map for connectivities below the tangent bifurcation.

In the following we will obtain the minima of p_{NL} (i.e., the critical connectivity p^*_{NL}) in the limit of large layer sizes ω and small coupling strength ϵ. We first derive an approximation of Equation (59) (cf. Equation 62), determine the validity range of this approximation (cf. Equation 69) and finally obtain an estimate for the critical connectivity (cf. Equation 71). As before, we fix the maximal input $\epsilon\omega$ to each neuron to preserve the network state and expand Equation (59) in a power series around $\epsilon \to 0$ and $\omega \to \infty$. Considering the leading terms yields

$$p_{NL} \approx p_{NL,\,a} := \frac{2\Theta_b}{p_f(\kappa)\,\epsilon\omega}\,\frac{1 + n\sqrt{\frac{\epsilon}{\Theta_b} - \frac{1}{\omega}}}{1 + \mathrm{Erf}\left(\frac{n}{\sqrt{2}}\right)}. \tag{61}$$

Further a propagation mediated by dendritic spikes (as introduced above) requires that the layer size ω and the coupling strength ϵ are sufficiently large such that a sufficiently large fraction of neurons of each layer receive a total input larger than the dendritic threshold Θ_b. In particular for diluted FFNs, this requirement translates to $\epsilon\omega \gg \Theta_b$ and Equation (61) simplifies further to

$$p_{NL,\,b} := \frac{2\Theta_b}{p_f(\kappa)\,\epsilon\omega}\,\frac{1 + n\sqrt{\frac{\epsilon}{\Theta_b}}}{1 + \mathrm{Erf}\left(\frac{n}{\sqrt{2}}\right)}. \tag{62}$$

Whereas p_{NL} has always a global minimum for $\epsilon\omega > \Theta_b$, this does not hold for the approximation $p_{NL,\,b}$, e.g., (cf. also **Figure 6C**)

$$\lim_{n\to-\infty}\left(p_{NL,\,b}\right) = -\infty. \tag{63}$$

However, we will now show that $p_{NL,\,b}$ has a (local) minimum if (and only if) $\epsilon \in \left(0, \frac{2\Theta_b}{\pi}\right]$ which approximates the global minimum of p_{NL} and therefore serve as an estimate for the critical connectivity. Starting with $\left.\frac{dp_{NL,\,b}(n)}{dn}\right|_{n=n^*} = 0$ yields

$$\sqrt{\frac{\Theta_b}{\epsilon}} = \sqrt{\frac{\pi}{2}}\exp\left(\frac{n^{*2}}{2}\right)\left(1 + \mathrm{Erf}\left(\frac{n^*}{\sqrt{2}}\right)\right) - n^* =: f\left(n^*\right), \tag{64}$$

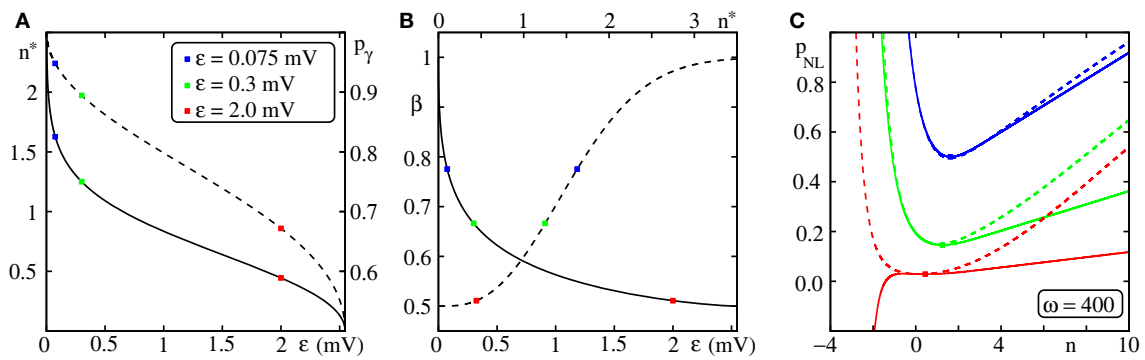


FIGURE 6 | Determining the critical connectivity in FFNs with non-additive dendritic interactions. (A) For a given setup, i.e., for a given dendritic threshold Θ_b and coupling strength $\epsilon < \frac{2\Theta_b}{\pi}$, the corresponding n^* (or equivalently p_γ; cf. Equation 56) is found by Equation (64). The solid line indicates n^* vs. ϵ (left vertical scale), the dashed line p_γ vs. ϵ (right vertical scale) and the markers $n^*(\epsilon)$ for $\epsilon = \{0.075, 0.3, 2.0\}$ mV (see legend). [Here, the dendritic threshold is $\Theta_b = 4$ mV, such that the estimate (Equation 64) is valid within the range $\epsilon \in (0, 2.55]$ mV; Equation (69)] **(B)** Knowing n^* allows to evaluate $\beta\left(\frac{\Theta_b}{\epsilon}\right) \in \left[\frac{1}{2}, 1\right]$ according to Equation (70). Panel **(B)** shows β (cf. Equation 70) vs. ϵ (solid line, lower horizontal axis) and β vs. n^* (dashed line, upper horizontal axis), respectively. **(C)** Finally, the critical connectivity p^*_{NL} is obtained by Equation (71) which depends on $\beta\left(\frac{\Theta_b}{\epsilon}\right)$. Panel **(C)** shows the connectivity p_{NL}[dashed; Equation (59)] and its approximation $p_{NL,\,b}$ [solid; Equation (62)] vs. n; for $\epsilon \in (0, \epsilon^{max}]$, $p_{NL,\,b}$ has a local minimum which agrees with the global minimum of p_{NL}. The markers indicate the critical connectivity p^*_{NL} obtained by the procedure described in **(A)** and **(B)**. For further explanations see text (section 3.2.1).

and n^* specifies the extremum of $p_{NL,\,b}(n)$. The second derivative of $p_{NL,\,b}(n)$ at the extremum n^* given by Equation (64) satisfies

$$\left.\frac{dp^2_{NL,\,b}}{dn^2}\right|_{n=n^*} = \frac{2n^*\sqrt{\frac{\Theta_b}{\epsilon}}}{p_f\,(\kappa)\,\omega\left(1+\mathrm{Erf}\left[\frac{n^*}{\sqrt{2}}\right]\right)} > 0 \tag{65}$$

if $n^* > 0$ such that the extremum actually is a minimum. Taken together, for a given setup, i.e., for given dendritic threshold Θ_b and coupling strength ϵ, the transcendent Equation (64) defines n^* which maximizes or minimize $p_{NL,\,b}(n)$ and if additionally $n^* > 0$ the extremum $p_{NL,\,b}\,(n^*)$ is a minimum.

Differentiating the right hand side of Equation (64),

$$\frac{df(n^*)}{dn^*} = n^* \cdot e^{\frac{n^{*2}}{2}}\sqrt{\frac{\pi}{2}}\left(1+\mathrm{Erf}\left[\frac{n^*}{\sqrt{2}}\right]\right) \tag{66}$$

$$\frac{d^2f(n^*)}{dn^{*2}} = n^* + \left(1+n^{*2}\right)e^{\frac{n^{*2}}{2}}\sqrt{\frac{\pi}{2}}\left(1+\mathrm{Erf}\left[\frac{n^*}{\sqrt{2}}\right]\right), \tag{67}$$

shows that $f\,(n^*)$ (as defined in Equation 64) is (1) minimal for $n^* = 0$ and (2) monotonically increasing for $n^* > 0$; according to Equation (64) the minimum $n^* = 0$ corresponds to

$$\epsilon^{\max} := \frac{\Theta_b}{\left[f(0)\right]^2} = \frac{2\Theta_b}{\pi} \approx 0.64\Theta_b. \tag{68}$$

The left hand side of Equation (64), i.e., $\sqrt{\Theta_b/\epsilon}$, is monotonically decreasing with ϵ from infinity to zero. Thus Equation (64) has a solution for any

$$\epsilon \in \left(0, \epsilon^{\max}\right] = \left(0, \frac{2\Theta_b}{\pi}\right] \tag{69}$$

and $p^*_{NL} := p^*_{NL,\,b}\,(n^*)$ is the (local) minimum of Equation (62) and provides an estimate for the critical connectivity, the (global) minimum of Equation (59).

For better readability we define the function $\beta(\cdot)$,

$$\beta\left(\frac{\Theta_b}{\epsilon}\right) := \frac{1}{2}\left(1+\mathrm{Erf}\left[\frac{n^*}{\sqrt{2}}\right]\right) - n^*\frac{e^{-\frac{n^{*2}}{2}}}{\sqrt{2\pi}}, \tag{70}$$

where $n^* = n^*\left(\frac{\Theta_b}{\epsilon}\right)$ as given by Equation (64). We note that $\beta\left(\frac{\Theta_b}{\epsilon}\right)$ can also be considered as a function of n^*. By combining Equations (62), (64), and (70) we obtain the critical connectivity

$$p^*_{NL} = \frac{\Theta_b}{p_f\,(\kappa)\,\epsilon\omega} \cdot \frac{1}{\beta\left(\frac{\Theta_b}{\epsilon}\right)}. \tag{71}$$

The function $\beta(\cdot)$ itself is monotonically decreasing with ϵ in the validity range $\epsilon \in (0, \epsilon^{\max}]$ of the above approximation: within this interval $n^* > 0$ and $\frac{d}{dn^*}f\,(n^*) > 0$ and thus the derivative

$$\frac{d\beta}{d\epsilon} = \frac{d\beta}{dn^*} \cdot \frac{dn^*}{d\sqrt{\Theta_b/\epsilon}} \cdot \frac{d\sqrt{\Theta_b/\epsilon}}{d\epsilon} \tag{72}$$

$$= -\frac{e^{-\frac{n^{*2}}{2}}n^{*2}}{\sqrt{2\pi}} \cdot \left(\frac{df\,(n^*)}{dn^*}\right)^{-1} \cdot \sqrt{\frac{\Theta_b}{4\epsilon^3}} \tag{73}$$

$$< 0. \tag{74}$$

Consequently β assumes its minimum

$$\beta^{\min} = \beta\left(n^* = 0\right) = \frac{1}{2} \tag{75}$$

for $\epsilon = \epsilon^{\max} = \frac{2\Theta_b}{\pi}$ and increases monotonically with decreasing ϵ against its asymptotic value

$$\beta^{\max} = \lim_{n^*\to\infty}\left[\frac{1}{2}\left(1+\mathrm{Erf}\left[\frac{n^*}{\sqrt{2}}\right]\right) - n^*\frac{e^{-\frac{n^{*2}}{2}}}{\sqrt{2\pi}}\right] = 1. \tag{76}$$

Thus the critical connectivity is bounded by

$$p^0 := \frac{\Theta_b}{p_f\,(\kappa)\,\epsilon\omega} \le p^*_{NL} \le 2 \cdot \frac{\Theta_b}{p_f\,(\kappa)\,\epsilon\omega} = 2 \cdot p^0 \tag{77}$$

and converges to the lower bound p^0 for small ϵ and to its upper bound $2p^0$ for large ϵ.

In **Figure 6** we visualize the determination of the critical connectivity (Equations 64, 70) and Equation (71). The critical connectivity obtained with the approach presented above agrees well with simulation data (cf. **Figure 7**).

3.2.2. Transition from linear to non-linear propagation

In the previous section we derived analytical estimates for the critical connectivity p^*_{NL} in FFNs with non-additive dendritic interactions; p^*_{NL} is determined by (1) the setup of the FFN (i.e., the layer size ω and coupling strength ϵ; cf. **Figure 7**), (2) the parameters of the non-linear modulation function (i.e., the dendritic threshold Θ_b and enhancement level κ) and (3) the layout of the external network (i.e., the mean external input I_0 and its variance, which is proportional to ϵ^{ext}). In this section, we discuss the influence of the parameters of the non-linear modulation function and study the transition from a regime where propagation of synchrony is mediated by dendritically enhanced inputs to a regime where the majority of inputs is processed linearly.

In general, with increasing threshold Θ_b more and more inputs are needed to reach this threshold and consequently the critical connectivity p^*_{NL} increases. If Θ_b exceeds μ_L, which is the average input to the neurons if a synchronous pulse propagates in linearly coupled FFNs (cf. Equation 45 and **Figure 5**), propagation mediated by linearly processed spikes is enabled for lower connectivities than propagation mediated by dendritic non-linearities. In this regime the linearly summed inputs (for $p = p^*_L$) are sufficient to maintain propagation of synchrony, but are not sufficient to cross the dendritic threshold. Increasing Θ_b even further has no

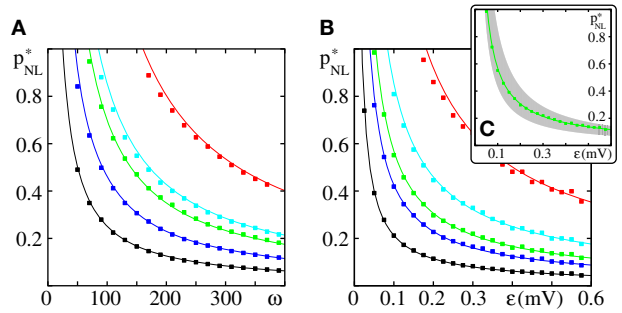


FIGURE 7 | Critical connectivity in FFNs with non-linear dendrites. The panels show **(A)** the critical connectivity p^*_{NL} vs. the layer size ω for different coupling strengths ($\epsilon = \{0.05\,\text{mV}$ (red), 0.1 mV (cyan), 0.125 mV (green), 0.2 mV (blue), and 0.4 mV (black)$\}$) and **(B)** p^*_{NL} vs. the coupling strength ϵ for different layer sizes ($\omega = \{50$ (red), 100 (cyan), 150 (green), 200 (blue), and 400 (black)$\}$). The points indicate the minimal connectivity for which a synchronous pulse propagates from the first to the last layer in an FFN with $m = 20$ layers in at least 50% of $n = 30$ trials. The critical connectivity given by Equation (71) (solid lines) is in good agreement with the computer simulations. **(C)** The critical connectivity is confined to the interval $p^*_{NL} \in [p^0, 2p^0]$ [indicated by the gray area for $\omega = 150$ (green), cf. Equation (77)] and approaches its lower bound for small ϵ and its upper bound for large ϵ. Like in linearly coupled networks the critical connectivity decays inversely proportional to layer size, $p^*_{NL} \propto \omega^{-1}$, (cf. also **Figure 3**), but the scaling with coupling strength is more complicated, $p^*_{NL} \propto \epsilon^{-1} \cdot 1/\beta\left(\frac{\Theta_b}{\epsilon}\right)$; the factor $\beta\left(\frac{\Theta_b}{\epsilon}\right) \in [0.5, 1]$ [cf. Equation (70) and **Figure 6**] measures the deviation from the algebraic decay (as found in linearly coupled networks). In this figure the parameters of the external network are fixed to $I_0 = 5\,\text{mV}$, $\nu^{\text{ext}} = 3\,\text{kHz}$, $\epsilon^{\text{ext}} = 0.5\,\text{mV}$.

influence on the critical connectivity p^*_{NL}, here a propagation of synchrony is possible for $p \geq p^*_L$ as discussed in section 3.1.

We illustrate this transition from non-linear to linear propagation in **Figure 8A**: We start with large $\Theta_b = \mu_L$ such that propagation is enabled for $p \approx p^*_L$ and also set $\kappa = \mu_L$. In fact, the linear critical connectivity p^*_L slightly under-estimates the observed critical connectivity p^*_{NL} as it does not account for the saturation of the non-linear modulation function, i.e., for the cutoff $\sigma_{NL}(x) = \kappa$ of inputs $x \geq \kappa$. With decreasing Θ_b the critical connectivity is substantially reduced and well approximated by Equation (71). Propagation of synchrony is now mainly mediated by dendritically enhanced inputs as described in section 3.2.1. The inset illustrates the impact of decreasing the dendritic threshold Θ_b on the iterated map. Initially, for $\Theta_b = \mu_L = \kappa$, the iterated map for linearly coupled and non-linearly coupled FFNs is similar; with decreasing Θ_b the jump like rise in the iterated map is shifted to lower group sizes and consequently the bifurcation point is shifted to lower connectivities.

The non-linear modulation function $\sigma_{NL}(\cdot)$ (cf. Equation 4) saturates for strong inputs, thus the enhancement level κ defines the maximal (effective) input to a neuron and $p_f(\kappa)$ is an upper bound for the spiking probability of any neuron in response to incoming inputs. This implies that in contrast to linearly coupled FFNs, the average size of a propagating synchronous pulse, γ_{NL}, given by the product of the probability of a neuron receiving sufficiently strong input to reach the dendritic threshold (p_γ; cf. Equation 56), the spiking probability due to that input [$p_f(\kappa)$] and the layer size ω, is bounded from above by

$$\gamma_{NL} = p_\gamma p_f(\kappa)\,\omega \leq \omega p_f(\kappa) =: \gamma^{\max}. \tag{78}$$

This bound decrease with decreasing κ as illustrated by **Figure 8B** (inset), where we compare the iterated maps for different values of κ. $p_f(\kappa)$ also influences the critical connectivity p^*_{NL} (cf. Equation 71): For small κ the spiking probability $p_f(\kappa)$ is low and thus p^*_{NL} is large (it may even exceed p^*_L). With increasing κ also $p_f(\kappa)$ increases and consequently the critical connectivity p^*_{NL} decreases; for very large κ the spiking probability $p_f(\kappa)$ approaches 1 (cf. Equation 14) and p^*_{NL} saturates (cf. **Figure 8B**).

In **Figure 8C** we show the critical connectivity for an additive enhancement by a constant Δ, i.e., inputs exceeding the dendritic threshold Θ_b are increased by the constant value $\Delta = \kappa - \Theta_b$. For small κ the critical connectivity p^*_{NL} is relatively large and may exceed p^*_L due to the low saturation level of the non-linear modulation function $\sigma_{NL}(\cdot)$ (cf. also **Figure 8B**). As mentioned above, with increasing κ, also $p_f(\kappa)$ increases and the critical connectivity p^*_{NL} decreases. However, for large κ and thus large dendritic threshold Θ_b propagation of synchrony mediated by linearly processed spikes is possible for lower connectivities than propagation mediated by dendritic non-linearities. Consequently, p^*_{NL} converges toward p^*_L (cf. also **Figure 8A**).

3.2.3. Influence of external network

In section 3.2.1 we derived an estimate of the critical connectivity p^*_{NL} for FFNs with non-additive dendritic interactions. So far we discussed the influence of the setup of the FFN (layer size ω and coupling strength ϵ) as well as the parameters of the non-linear modulation function σ_{NL} (dendritic threshold Θ_b and enhancement level κ). In the current section, we focus on the remaining determining factor, the layout of the external network. How does the critical connectivity change with the mean external input I_0 and external coupling strength ϵ^{ext} and how well are these changes covered by our analytics?

For the derivation of p^*_{NL} we assumed that somatic spikes are elicited exclusively by dendritically enhanced inputs (cf. Equation 54) and thus the critical connectivity depends on the layout of the external network only via $p_f(\kappa)$ (cf. also Equation 71), i.e., on the average spiking probability of a neuron receiving an input larger than the dendritic threshold $x \geq \Theta_b$. For sufficiently small $p_f(\kappa)$, $p^*_{NL} > 1$ and propagation of synchrony is not possible. With increasing $p_f(\kappa)$ the critical connectivity decreases and for $p_f(\kappa) \to 1$ it converges to $\Theta_b\,(\epsilon\omega\beta\,[\Theta_b/\epsilon])^{-1}$, independent of the external network.

In the regime of low spiking rates, changing the mean external input I_0 simply shifts the distribution of membrane potentials $P_V(V)$ (which is a Gaussian distribution centered at I_0; cf. Equation 8). Thus, with increasing I_0, $p_f(\kappa)$ increases and the critical connectivity p^*_{NL} decreases.

In **Figure 9A** we show the critical connectivity for different ϵ^{ext} [which determines the width of $P_V(V)$] vs. the mean external input I_0. For $I_0 = \Theta - \kappa$ (such that the sum of a dendritically enhanced input and the center of the distribution of membrane

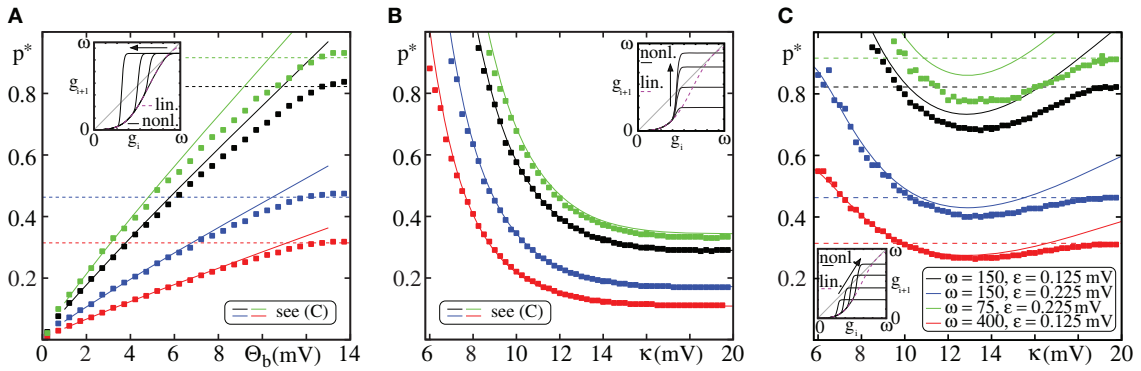


FIGURE 8 | Transition from linear to non-linear propagation. The figure shows the critical connectivity p^*_{NL} vs. the parameters of the non-linear modulation function σ_{NL} (cf. Equation 4) for different network setups (color code, see **(C)**). The lines are the theoretical predictions for p^*_{NL} [solid, Equation (71)] and p^*_L [dashed, Equation (32)]. The markers indicate the minimal connectivity for which a synchronous pulse propagates from the first to the last layer in an FFN ($I_0 = 5$ mV, $\nu^{ext} = 3$ kHz, $\epsilon^{ext} = 0.5$ mV) with $m = 20$ layers in at least 50% of $n = 30$ trials. The insets illustrate the effect of changing Θ_b and κ on the iterated map, cf. Equation (13), where connectivity is kept constant. **(A)** Critical connectivity vs. dendritic threshold Θ_b for constant enhancement level $\kappa = \mu_L \approx 13.7$ mV (cf. Equation 50). If the dendritic threshold Θ_b is sufficiently small such that $p_f(\Theta_b) \ll p_f(\kappa)$ (cf. Equation 54), the propagation of synchrony is mainly mediated by non-linear enhanced inputs and the critical connectivity can be estimated by Equation (71). For large Θ_b the probability that an input from the preceding layer exceeds the dendritic threshold is very low, propagation of synchrony is mainly mediated by linearly processed inputs and the critical connectivity is given by Equation (32). Between these scenarios (for moderate Θ_b) there is a "transition regime," where linear and non-linear propagation mix [similarly in **(C)**]. **(B)** Critical connectivity vs. enhancement level κ for constant threshold $\Theta_b = 4$ mV. For small enhancement levels κ the (maximal) spiking probability of a single neuron, $p_f(\kappa)$, is small and thus the critical connectivity p^*_{NL} is large. With increasing κ, $p_f(\kappa)$ increases and thus p^*_{NL} decreases; for large κ, $p_f(\kappa) \to 1$ (a neuron will almost surely spike upon the receipt of a non-linearly enhanced pre-synaptic input) and the critical connectivity saturates. **(C)** Critical connectivity vs. enhancement level κ for an additive enhancement by a constant $\Delta = \kappa - \Theta_b = 4$ mV. For further explanations see text (section 3.2.2).

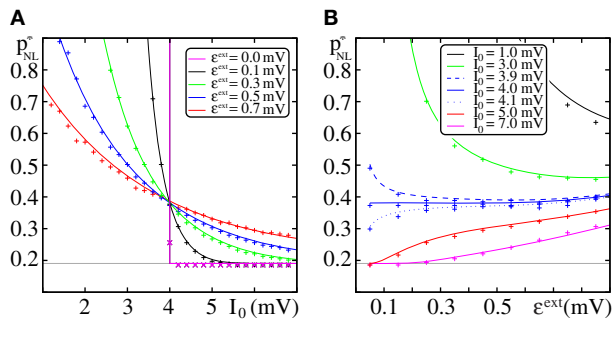


FIGURE 9 | Dependence of the critical connectivity p^*_{NL} on the layout of the external network. (A,B) The lines indicate the theoretical prediction for p^*_{NL} given by Equation (71) and agree well with the data from direct numerical simulations (markers; FFN with $\omega = 150$, $\epsilon = 0.2$ mV, $\Theta_b = 4$ mV, $\kappa = 11$ mV, $m = 20$). Panel **(A)** shows the critical connectivity vs. the mean external input I_0 for fixed ϵ^{ext} and panel **(B)** shows the critical connectivity vs. ϵ^{ext} for fixed mean external input I_0. The gray line indicates the minimal critical connectivity obtained for $p_f(\kappa) = 1$. With increasing mean (external) input I_0 the distribution of membrane potentials $P_V(V)$ is shifted toward the somatic threshold Θ, thus the spiking probability $p_f(\kappa)$ upon the reception of a non-linear enhanced input increases and the critical connectivity p^*_{NL} decreases. For $I_0 = \Theta - \kappa$, $p_f(\kappa) \approx 0.5$ (cf. Equation 80) and p^*_{NL} is largely independent of the layout of the external network [blue solid line in **(B)**; cf. also **(A)** where all curves coincide]. Further explanations see text (section 3.2.3).

potentials equals the somatic threshold Θ), $p_f(\kappa)$ simplifies to

$$p_f(\kappa) = \frac{1}{2}\left(\mathrm{Erf}\left[\frac{\Theta - I_0}{\sigma}\right] + \mathrm{Erf}\left[\frac{\kappa - \Theta + I_0}{\sigma}\right]\right) \quad (79)$$

$$= \frac{1}{2}\mathrm{Erf}\left(\frac{\Theta - I_0}{\sigma}\right) \quad (80)$$

and thus in the regime of low spiking rates, i.e., $(\Theta - I_0)/\sigma \gg 1$, $p_f(\kappa) \approx 0.5$ independent of the width of the distribution of membrane potentials. Consequently, all curves for different ϵ^{ext} coincide at this point. For $I_0 > \Theta - \kappa$ the majority of neurons (>50%) would spike upon receipt of a dendritically enhanced input. Thus $p_f(\kappa)$ increases and therewith the critical connectivity decreases upon decreasing ϵ^{ext}. In the limit of $\epsilon \to 0$, $P_V(V)$ converges toward a δ-distribution centered at I_0 and p_f becomes a step-function

$$p_f(\kappa) = \begin{cases} 0 & \kappa < \Theta - I_0 \\ 1 & \kappa \geq \Theta - I_0 \end{cases} \quad (81)$$

such that the critical connectivity is either constant and minimal for $I_0 \geq \Theta - \kappa$ or it diverges (no propagation possible) for $I_0 < \Theta - \kappa$ (cf. **Figure 9A**; magenta curve).

In **Figure 9B** we illustrate the effect of changing ϵ^{ext} on the critical connectivity for constant I_0. As discussed above for $I_0 = \Theta - \kappa$, $p_f(\kappa)$ and thus p^*_{NL} are rather independent of ϵ^{ext} and for $I_0 > \Theta - \kappa$ the critical connectivity increases with ϵ^{ext}. For $I_0 < \Theta - \kappa$ an increase of the width of the distribution of membrane potentials shifts the membrane potential of more and more neurons toward the relevant interval $[\Theta - \kappa, \Theta]$ and thus $p_f(\kappa)$ increases and the critical connectivity p^*_{NL} decreases.

For the derivation of p^*_{NL} we have assumed that the ground state dynamics is essentially not influenced by the spontaneous activity of the FFN itself (i.e., $\mu = I_0$ and $\sigma = \epsilon^{ext}\sqrt{2\tau^m \nu^{ext}}$). As

discussed in section 3.1.3, we can correct the results for such influences. However, since in non-linearly coupled FFNs the impact of (non-linearly enhanced) synchronous activity is much stronger than the impact of spontaneous activity (which is irregular and not amplified by non-additive dendritic interactions), we find that the deviations between the corrected and uncorrected version of p^*_{NL} is negligible.

Finally, we compare the critical connectivity for networks with and without non-additive dendritic interactions: The factor

$$c^{\text{rat}} := \frac{p^*_L}{p^*_{NL}} = \frac{p_f(\kappa)}{\lambda\Theta_b}\beta\left(\frac{\Theta_b}{\epsilon}\right) \tag{82}$$

measures how much the connectivity within the FFN can be reduced by introducing non-additive dendritic interactions. It is independent of the layer size ω and becomes maximal in the limit of small coupling strengths ϵ as $\beta(\Theta_b/\epsilon) \to \beta^{\max} = 1$ for $\epsilon \to 0$ (cf. Equation 76). It increases with decreasing Θ_b and increasing κ (see discussion in section 3.2.2). In **Figure 10** we show the influence of the external network. As discussed above, for small I_0, propagation of synchrony is not possible (the non-linear enhanced input is insufficient to elicit sufficiently many spikes in the layers of the FFN; white areas in **Figure 10**). With increasing I_0, p^*_{NL} decreases and c^{rat} increases.

3.3. GENERALIZATIONS

In the final section we discuss generalizations of the methods and results we derived. Compared to biological neurons, our models have simplifications which enable the analytical treatment, but might be suspected to be influential on the final result. These simplifications are the homogeneous delay distribution, the simplified initiation and impact of dendritic spikes, the limit of short synaptic currents and the sub-threshold leaky integrate-and-fire dynamics. Here, we verify that our results generalize to biologically more detailed neurons without these simplifications. In particular, we show that the estimates for the critical connectivity hold. Further, we consider a qualitatively different dendritic interaction function which assumes that the saturation is incomplete, i.e., beyond a region of saturation the impact of larger inputs increases. We show that the tools developed in the article are still applicable and reveal a new phenomenon, the coexistence of linear and non-linear propagation of synchrony.

In the first part (section 3.3.1), we discuss the influence of inhomogeneous delay distribution and finite dendritic integration windows. In the second part (section 3.3.2), we consider the non-linear modulation function with incomplete saturation. Finally, we consider biologically more detailed neuron models (section 3.3.3).

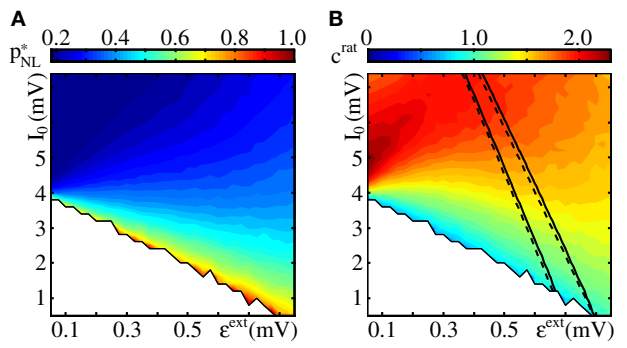


FIGURE 10 | Critical connectivity and reduction factor. Panel **(A)** shows the critical connectivity obtained from simulations of an FFN ($\omega = 150$, $\epsilon = 0.2$ mV, $m = 20$) incorporating non-additive dendritic interactions ($\Theta_b = 4$ mV, $\kappa = 11$ mV; see also **Figure 9**). Within the white area, propagation of synchrony is impossible because even for a fully coupled chain the input to the next layer (limited by the saturation of the non-linear modulation function and the layer size) is insufficient. Panel **(B)** shows the reduction factor c^{rat} (cf. Equation 82), the quotient between the critical connectivity in FFNs without and with non-additive dendritic interactions. The lines enclose the area for which the spontaneous firing is between $\nu \in [0.5, 1.5]$ Hz obtained from simulations (solid) and low-rate approximation (cf. Equation 9; dashed).

3.3.1. Heterogeneous delays

So far we considered FFNs with homogeneous delay distribution and dendritic modulation functions with integration window of zero length, i.e., only exactly synchronized inputs were possibly non-linearly amplified. Are these assumptions crucial for the obtained results? How does the critical connectivity change in the presence of heterogeneous delay distributions?

To answer this question, we consider synaptic delays τ_{kl} (specifying the synaptic delay between neuron l and k) uniformly drawn from

$$\tau_{kl} \in \left[\tau - \frac{\Delta T}{2}, \tau + \frac{\Delta T}{2}\right], \tag{83}$$

where τ is the mean delay. A direct consequence of heterogeneous delay distribution is that the spikes of the propagating synchronous signal are not simultaneous (i.e., exactly synchronized) anymore. To describe the system accurately one has to consider additionally to the size (g_i) also the temporal jitter (s_i) of the synchronous pulse in the ith layer and investigate the two-dimensional iterated map for (g_i, s_i) (e.g., Diesmann et al., 1999; Gewaltig et al., 2001; Goedeke and Diesmann, 2008). However, even if the synchronous pulse is blurred out to a pulse packet with finite width, for sufficiently large connectivity stable propagation still can be obtained (see e.g., Gewaltig et al., 2001).

For linearly coupled FFNs, with increasing width of the delay distribution, ΔT, the propagating pulse becomes broader and thus the critical connectivity p^*_L increases (cf. **Figures 11A,B**; squares). However, the scaling with layer size (cf. **Figure 11A**) and coupling strength (data not shown) is the same.

Under the assumption that the width of the pulse packet stays bounded, one can derive a lower bound for the critical connectivity. We assume that a pulse in layer i is perfectly synchronized and calculate the effective peak of the depolarization in the $(i+1)$th layer. Replacing the coupling strength ϵ by the effective depolarization ϵ' (derived below, cf. Equation 89) in the estimate of the critical connectivity (cf. Equation 32) one gains an estimate of the critical connectivity for systems with heterogeneous delays [Equation (90); shown in **Figure 11**]. Consider a perfectly synchronized pulse in layer i. Due to inhomogeneities in the delay, the inputs arriving at the $(i+1)$th layer are distributed uniformly in an interval of size ΔT (Equation 83). We assume that all inputs

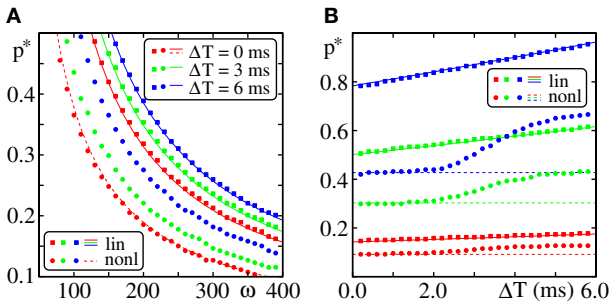


FIGURE 11 | Robustness against heterogeneities in the response delays. (A) Critical connectivity vs. layer size for FFNs ($m = 20$, $\epsilon = 0.25\,\text{mV}$, $I_0 = 5\,\text{mV}$, $\epsilon^{\text{ext}} = 0.5\,\text{mV}$) with additive (squares) and non-additive ($\Theta_b = 4\,\text{mV}$, $\kappa = 11\,\text{mV}$, $\Delta t = 2.5\,\text{ms}$; circles) dendritic interactions. Different colors indicate different widths of the delay distribution (cf. Equation 84). The solid lines indicate the critical connectivity p_L^* corrected for inhomogeneous delay distribution (cf. Equation 90), the dashed line p_{NL}^* for $\Delta T = 0\,\text{ms}$. **(B)** Critical connectivity vs. width of delay distribution ΔT. Different colors indicate different setups of the FFN (red: $\omega = 275$, $\epsilon = 0.4\,\text{mV}$; green: $\omega = 125$, $\epsilon = 0.25\,\text{mV}$; blue: $\omega = 200$, $\epsilon = 0.1\,\text{mV}$). Solid and dashed lines are p_L^* and p_{NL}^* as before.

arriving at a neuron of layer $i+1$ are equidistantly distributed over $[-\Delta T/2, \Delta T/2]$, i.e., the arrival time of the lth of a total number of k inputs is

$$t_l^{\text{arr}} = \tau - \frac{\Delta T}{2} + \frac{\Delta T}{k-1} \cdot (l-1)\,. \tag{84}$$

We consider the subthreshold dynamics only. Each single input depolarizes the neuron by an amount ϵ and afterwards the membrane potential $V(t)$ decays exponentially toward its asymptotic value (I_0) with the membrane time constant τ^{m} (cf. Equations 1, 2) until the next input arrives after a time interval $\frac{\Delta T}{k-1}$ (cf. Equation 84). Thus the total (effective) depolarization caused by the sum of these k inputs at the end of the considered time interval $\left(\tau + \frac{\Delta T}{2}\right)$ is

$$\Delta\epsilon_k = \sum_{l=1}^{k} \epsilon \exp\left(-\frac{1}{\tau^{\text{m}}} \frac{\Delta T}{k-1} (l-1)\right) \tag{85}$$

$$= \epsilon \frac{\exp\left(-\frac{\Delta T}{\tau^{\text{m}}} \frac{k}{k-1}\right) - 1}{\exp\left(-\frac{\Delta T}{\tau^{\text{m}}} \frac{1}{k-1}\right) - 1}. \tag{86}$$

We consider the effective depolarization per input, ϵ', in the limit of a large number of inputs k ($k \to \infty$),

$$\epsilon' = \lim_{k\to\infty} \left(\frac{\Delta\epsilon_k}{k}\right) \tag{87}$$

$$= \frac{\tau^{\text{m}}}{\Delta T} \left(1 - \exp\left[-\frac{\Delta T}{\tau^{\text{m}}}\right]\right) \epsilon \tag{88}$$

$$=: C(\Delta T)\,\epsilon. \tag{89}$$

Thus the correction factor $C(\Delta T) \leq 1$ defined in Equation (89) relates the coupling strength ϵ to the effective coupling strength ϵ' in the presence of inhomogeneous delays. The critical connectivity is then given by (cf. Equation 32)

$$p_L^* = \frac{1}{C(\Delta T)} \cdot \frac{1}{\lambda^* \epsilon \omega} \tag{90}$$

and this estimate agrees well with direct numerical simulations (cf. **Figure 11**).

For FFNs with dendritic non-linearities and inhomogeneous delays τ_{kl}, one has to consider a finite dendritic integration window Δt^{d}. Instead of amplifying only simultaneously received spikes (cf. Equation 5), the sum of spikes within the time interval Δt is considered. We denote the sum of inputs to a neuron within the time interval $[t - \Delta t, t]$ by

$$S_k^{\Delta t}(t) = \sum_l \sum_m \epsilon \chi_{[t-\Delta t,\, t]} \left(t_{lm}^f + \tau_{kl}\right), \tag{91}$$

where

$$\chi_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases} \tag{92}$$

is the indicator function and t_{lm}^f is the mth firing time of neuron l as before. If $S_k^{\Delta t}(t)$ exceeds the dendritic threshold Θ_b for some $t = t_0$, neuron k is depolarized additionally (to the depolarization arising from linear spike summation) by

$$\epsilon_\kappa^{\text{add}}(t_0) = \kappa - S_k^{\Delta t}(t_0) \tag{93}$$

such that the total (effective) depolarization caused by an input $x \geq \Theta_b$ equals κ, modeling the effect of a dendritic spike; cf. also section 3.3.3. After such an additional depolarization the dendrite becomes refractory for a time $t^{\text{ref,ds}}$ and does not transfer additional spikes within the interval $\left[t_0, t_0 + t^{\text{ref,ds}}\right]$. For $\Delta t = 0$ we recover the non-linear modulation function $\sigma_{NL}(\cdot)$ given by Equation (4). Due to the finite dendritic interaction window, a delay distribution with $\Delta T \leq \Delta t$ affects the critical connectivity only weakly (cf. **Figure 11B**). For $\Delta T > \Delta t$, some of the inputs received from the preceding layer upon a propagation of synchrony fall out of the dendritic interaction window ΔT and thus the critical connectivity increases. However, the scaling with layer size ω (cf. **Figure 11B**) and coupling strength ϵ (data not shown) is practically identical with the scenario $\Delta T = 0$.

Before we discuss propagation of synchrony in biologically more plausible neuron models in section 3.3.3, we consider generalization of the non-linear modulation function in the following section.

3.3.2. Coexistence of linear and non-linear propagation

In this article, we employed a non-linear modulation function $\sigma_{NL}(\epsilon)$ that is linear for dendritic stimulation smaller than the dendritic threshold, $\epsilon < \Theta_b$, and constant (i.e., saturates) for supra-threshold stimulation, $\epsilon \geq \Theta_b$ (cf. Equation 4).

Biologically, if the linear inputs are transmitted despite the dendritic sodium spike and are not shadowed by, e.g., an NMDA spike, they may lead to a second, later peak depolarization after the one generated by the sodium spike. Since our models replace depolarizations by jumps to the peak depolarization, we have to account for the later peak as soon as it exceeds the earlier one. In this part, we thus assume that if the synchronous input is so large that the depolarization it generates upon linear summation exceeds the depolarization κ generated by the dendritic spike, this former is considered as the effect of the input. In other words, we assume that the dendritic modulation function continues linearly beyond κ, i.e., we define

$$\sigma'_{NL}(\epsilon) = \begin{cases} \epsilon & \text{for} \quad \epsilon \leq \Theta_b \\ \kappa & \text{for} \quad \Theta_b \leq \epsilon \leq \kappa \\ \epsilon & \text{for} \quad \epsilon \geq \kappa \end{cases} \tag{94}$$

(cf. inset of **Figure 12A**).

The iterated map, mapping the number of active neurons in layer i to the average number of active neurons in layer $i+1$ may now have (depending on the system parameters) between one and five fixed points (cf. **Figure 12**). As before, $G_0 = 0$ is a trivial fixed point corresponding to the level of absent activity and the only fixed point of the iterated map for small connectivity p. With increasing connectivity p, two additional pairs of fixed points $G_1 \leq G_2$ and $G_3 \leq G_4$ appear via tangent bifurcations. The first pair of fixed points, G_1 and G_2, correspond to the propagation of synchrony mediated by non-additive dendritic interactions (as discussed in section 3.1), the second pair, G_3 and G_4, correspond to propagation of synchrony mediated by linearly processed inputs (as discussed in section 3.2). By further increasing the connectivity p, the fixed points G_2 and G_3 disappear via a tangent bifurcation (cf. **Figure 12A**). Within the region, where five fixed points exists, both types of propagation of synchrony coexists (illustrated in **Figures 12B–D**): Synchronized pulses of size $g_0 < G_1$ typically decay to zero after a small number of layers. Pulse sizes with $G_1 < g_0 < G_3$ typically initiate propagation of synchrony with an average pulse size around G_2 (where the propagation is mediated by non-additive dendritic interactions) and synchronous pulses of size $g_0 > G_3$ typically initiate propagation of synchrony with average pulse sizes around G_4 (linear propagation). For sufficiently large p, i.e., the fixed points G_2 and G_3 disappeared, a synchronized pulse of size $g_0 \geq G_1$ will initiate propagation of synchrony with pulse sizes around G_4; in this parameter region the non-additive dendritic interactions essentially increase the basin of attraction of G_4.

Within the framework of our analytical tractable model, we neglect, e.g., the initiation time of a dendritic spike (in our model non-linear amplifications are instantaneous) or the different shapes of potential deflections caused by linearly and non-linearly processed inputs. Therefore, propagating synchronous signals mediated either by linear or non-linear dendrites differ only in their size. In biological more detailed models (briefly discussed in section 3.3.3 below) both propagation types will be more distinct, e.g., the propagation frequency (speed) and the quality of synchrony of the propagating pulses are different (see also Jahnke et al., 2012).

3.3.3. Biological more detailed models

The model we mainly consider in this article has the advantage of being analytically tractable. Here we ask whether it over-simplifies the considered systems. More precisely, we study whether the results derived above, in particular the analytical estimates for the critical connectivity, generalize to biologically more detailed models.

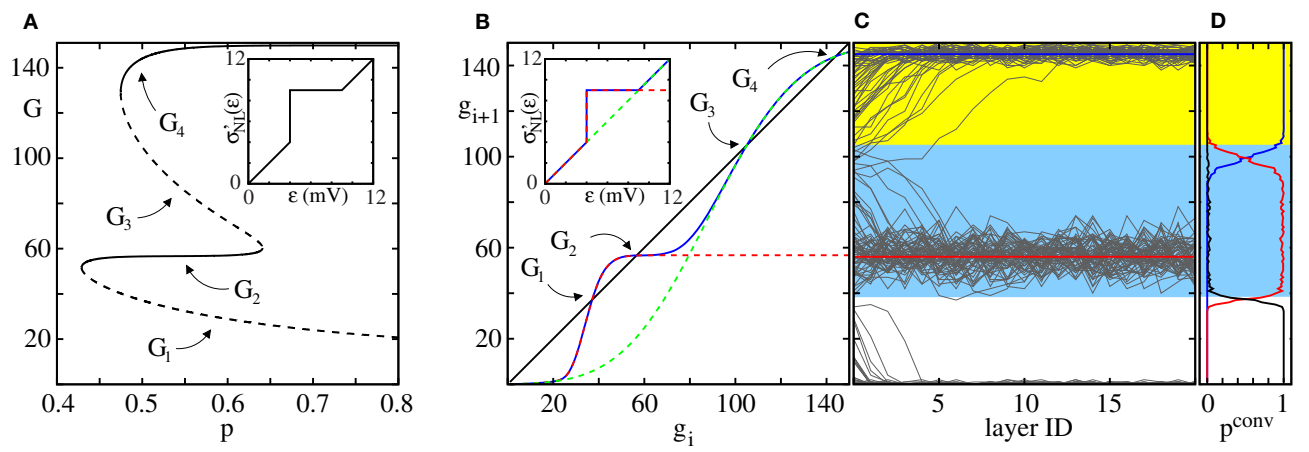


FIGURE 12 | Coexistence of linear and non-linear propagation. **(A)** Bifurcation diagram obtained from Equation (13) for an FFN ($\omega = 150$, $\epsilon = 0.225$ mV) with a non-linear modulation function σ'_{NL} with incomplete saturation [cf. Equation (94) and inset]. Panel **(B)** shows the iterated maps (Equation 13) for $p = 0.5$ with the different non-linear modulation functions considered in this article (linear coupling: green,dashed; non-linear coupling σ_{NL}: red, dashed; modified non-linear coupling σ'_{NL}: blue). Panel **(C)** depicts the development of the size of the synchronous pulse along the layers of the FFN (single trials). The blue and yellow regions are the basins of attraction of G_2 and G_4, respectively, derived from the data in panel **(B)**. Panel **(D)** shows the probability p^{conv} of converging to the linear propagation regime (yellow area, blue line) and the non-linear propagation regime (blue area, red line) after $m = 20$ layers (p^{conv} is obtained from $n = 150$ runs with different networks and initial conditions).

The main assumption underlying our analysis of linearly coupled networks is a very general one, namely that synchronous single inputs sum up linearly: we assumed that the spiking probability $p_f(\cdot)$ of a neuron due to the reception of x synchronous inputs of size ϵ equals the spiking probability due to the reception of one single input of size $y = x\epsilon$. Therefore, the results will hold also for more complex neuron models, as long as the effect of a synchronous input pulse is approximately the sum of the effects of single inputs. In particular, if the spiking probability due to an input of strength x, $p_f(x)$, is sufficiently slowly changing with x, according to Equation (24) the critical connectivity scales like $p_L^* \propto (\epsilon\omega)^{-1}$ for sufficiently large layer sizes and small coupling strengths. To fully compute the critical connectivity, the actual form of $p_f(\cdot)$ has to be known. Our leaky integrate-and-fire neuron with infinitesimally short current pulses approximates the behavior of a wide class of neuron models for which an analytical derivation of $p_f(\cdot)$ is impossible. Still even for more detailed models, $p_f(\cdot)$ is accessible for measurements in single neuron (computer) experiments.

In **Figure 13** we verify our predictions exemplary for two types of neuron models: We employ a model of conductance based leaky integrate-and-fire-type neurons with exponential input conductances (CB-type; see Appendix) and a Hodgkin-Huxley-type neuron model with alpha-function shaped input currents (HH-type; see Appendix). The post-synaptic potential induced by single excitatory inputs is shown in panels (a) and (b) and the scaling of the critical connectivity p_L^* with $\epsilon\omega$ in panel (c): the scaling of p_L^* is well described by $p_L^* \propto (\epsilon\omega)^{-1}$.

The main assumptions underlying our analysis of non-linearly coupled networks are (1) that the maximal spiking probability due to inputs which are subthreshold relative to the dendritic threshold, $p_f(\Theta_b)$, is significantly smaller than the spiking probability due to a suprathreshold input, $p_f(\kappa)$, and (2) that the temporal jitter of somatic spikes evoked by suprathreshold inputs is small such that synchronized inputs stay synchronized. Both conditions have been found to be satisfied in biological neurons (e.g., Ariav et al., 2003). Therefore, Equation (71) specifying the critical connectivity p_{NL}^* also holds for more detailed neuron models if these models incorporate biologically plausible features of fast dendritic spikes. To obtain a quantitative prediction of p_{NL}^*, it is sufficient to estimate (a) the number of inputs needed to elicit a dendritic spike, Θ_b/ϵ, (b) the layer size ω, and (c) the spiking probability due to the reception of a total input that is sufficiently strong to elicit a dendritic spike.

To investigate the scaling of the critical connectivity p_{NL}^* in direct numerical simulations, we account for the effects of dendritic spikes in the CB-type and HH-type: When the total excitatory input within the dendritic integration window exceeds the dendritic threshold level, a current pulse modeling the effect of a dendritic spike is initiated and causes an additional

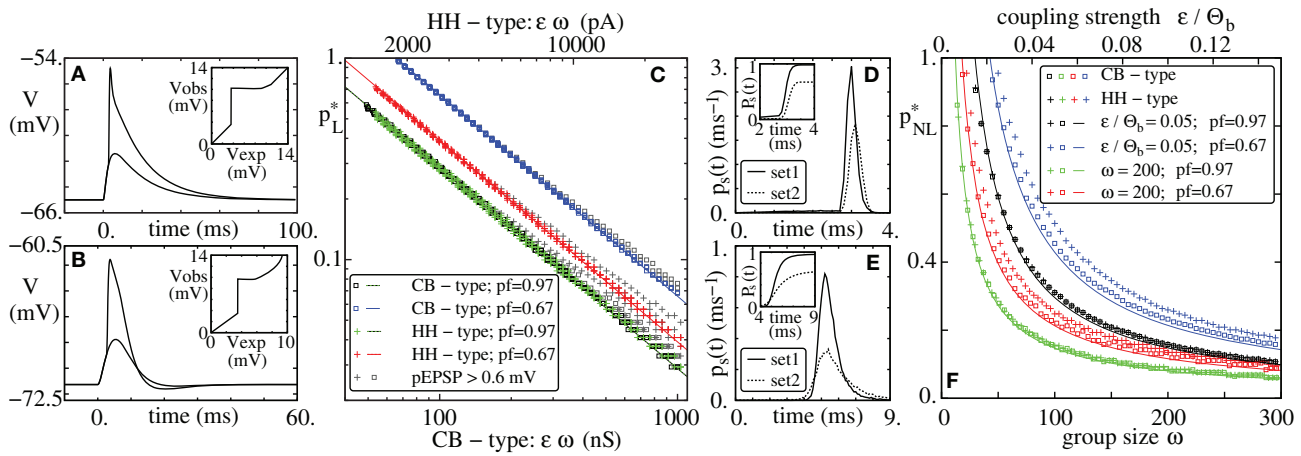


FIGURE 13 | Same scaling of propagating regime for networks of biologically more detailed neuron models. (A,B) Time course of the membrane potential of single neurons receiving inputs that are sufficiently strong to elicit a dendritic spike, with (non-linear model) and without (linear model) dendritic spike generation mechanism, for **(A)** a conductance based LIF-type neuron (henceforth: CB-type), and **(B)** a Hodgkin–Huxley-type neuron (HH-type). The insets show the observed peak of the induced postsynaptic potential (pEPSP) vs. the pEPSP expected from linear input summation (equivalent to the dendritic modulation function in the analytically tractable model). **(C)** Critical connectivity p_L^* vs. $\epsilon\omega$ in linearly coupled networks. For each value $\epsilon\omega$, we evaluated the critical connectivity for four different group sizes $\omega = 100, 300, 500, 700$ and four different coupling strengths $\epsilon = 0.3$, 0.6, 0.9, 1.2 nS (CB-type; squares; lower horizontal axis) and $\epsilon = 9$, 18, 27, 36 pA (HH-type; crosses; upper horizontal axis), respectively. The lines are fitted functions of the form $(\lambda\epsilon\omega)^{-1}$. The analytical estimate given by Equation (24) holds in the limit of large layer sizes ω and small couplings ϵ, therefore we exclude data points from the fitting where a single input yields an EPSP larger than 0.6 mV (CB-type: $\epsilon \geq 1.4$ nS; HH-type: $\epsilon \geq 46$ pA; these points are marked in gray). **(D,E)** Probability distribution of somatic spike times after stimulation of the neuron by an input which is sufficiently strong to generate a dendritic spike (**D**: CB-type, **E**: HH-type). We show exemplary two different configurations for the external inputs, which result in a total somatic spiking probability after dendritic spike generation of $p^f \approx 0.97$ (solid lines; set 1) and $p^f \approx 0.67$ (dashed lines; set 2). p^f equals the saturation level of the corresponding cumulative distribution function (shown in the insets). **(F)** Critical connectivity p_{NL}^* vs. group size ω (lower horizontal scale) and coupling strength ϵ normalized by threshold Θ_b (upper horizontal scale), respectively. The theoretical estimate of p_{NL}^* (cf. Equation 71) is a function of ω, Θ_b/ϵ and p^f, therefore the predictions agree for both models and the data from direct numerical simulations are consistent with the theoretical predictions. [All simulations of FFNs in this figure are obtained for inhomogeneous delay distribution with $\Delta T = 1$ ms (cf. Equation 83)].

depolarization of the soma of the post-synaptic neuron (see Appendix for details; cf. also section 3.3.1). In **Figure 13** we compare the results of direct numerical simulations with the estimate given by Equation (71). The post-synaptic potential induced by single excitatory inputs is shown in panels (A) and (B). Panel (D) and (E) shows the spiking probability of a single neuron (in the ground state of the FFN), p^f, due to an input exceeding the dendritic threshold level; as examples we present two different setups with $p^f = \{0.67, 0.97\}$. Panel (F) shows the scaling of p^*_{NL} with layer size and coupling strength and the good agreement of the analytical estimate with direct numerical simulations.

4. SUMMARY AND CONCLUSIONS

Propagation of synchrony in feed-forward sub-structures that are embedded in randomly connected recurrent networks has been a research topic for more than two decades now [see, e.g., review on this topic (Kumar et al., 2010)] and it is hypothesized that such propagation possibly explain the emergence of spatio-temporal spike patterns and information transmission.

In this article, we have analyzed diluted FFNs and investigated their capability to propagate synchrony. In addition to conventional additive (linear) input processing at single neurons, we considered non-additive dendritic interactions modeling the impact of fast dendritic spikes (Ariav et al., 2003; Gasparini et al., 2004; Polsky et al., 2004; Gasparini and Magee, 2006). We emulated the influence of the embedding recurrent network which establishes the irregular ground state in the FFN, by random Poissonian inputs (van Vreeswijk and Sompolinsky, 1996, 1998; Brunel, 2000). This approach does not account for back-reactions of activity within the FFN on the embedding network. It is justified as long as the connectivity and connection strength between the neurons of the FFN and the embedding network is low and weak compared to the feed-forward connectivity and connection strength. The back-reaction then influences the activity of the embedding network only weakly and a robust propagation of synchrony can be achieved (Vogels and Abbott, 2005; Kumar et al., 2008; Jahnke et al., 2012). Yet, if the condition is not met, synchronous activity within the FFN may spread out over the embedding network and potentially cause pathological activity ("synfire-explosions") (Mehring et al., 2003). For specifically structured networks also more complex interactions are possible, such as an enhancement of propagating synchrony (manuscript in preparation).

In the main part of the article, we studied the propagation of synchrony employing leaky integrate-and-fire neurons in the limit of temporally short synaptic inputs and homogeneous synaptic delays. Synchronous pulses consist of exactly synchronized (simultaneous) spikes. This allows to investigate propagation of synchrony by considering the size of a synchronized pulse only, so that the analysis becomes analytically tractable. Nevertheless, in the second part of our article we also consider systems with heterogeneous coupling delays and temporally extended interactions. In agreement with the literature (e.g., Diesmann et al., 1999; Gewaltig et al., 2001; Goedeke and Diesmann, 2008), we observe that pulse packets tend to synchronize along the layers of the FFN so that the results of our simplified description are directly applicable.

We derived scaling laws as well as quantitative estimates for the critical connectivity marking the bifurcation point between the regime where robust propagation of synchrony is possible and where it is not. In particular, based on a suitable series expansion we have shown that for linearly coupled FFNs the critical connectivity decays inversely proportional to layer size and coupling strength. Moreover, the proportionality factor can be estimated from the ground state properties of the single neurons. The estimate agrees with direct numerical simulations within the biologically relevant parameter regime where (a) the spontaneous firing rate of the neurons is low and (b) the distribution of membrane potentials is broad (each neuron receives a huge number of almost random presynaptic inputs). If a synchronous pulse propagates along the layers of a linearly coupled FFN, most of the neurons of each layer participate in the propagation of synchrony, independent of the actual layer size, coupling strength or layout of the external network.

For neurons incorporating non-additive dendritic interactions, the spiking probability as a function of the dendritic stimulation becomes discontinuous. Therefore, the analytical estimation of the critical connectivity in non-linearly coupled FFNs required a different approach than the treatment of linearly coupled FFNs. We have shown that the critical connectivity decays inversely proportional to the layer size (as in linearly coupled FFNs), and we have derived the dependence on the coupling strength which is more complicated. The critical connectivity is completely determined by layer size, spiking probability of the single neuron upon the reception of a non-linearly enhanced presynaptic input and the number of inputs required to reach the dendritic threshold. Our results indicate that in presence of non-linear dendrites, neurons process synchronous inputs similar to threshold units. Such units have been previously used as simplified rate neuron models to study activity propagation in discrete time, e.g., in Nowotny and Huerta (2003); Leibold and Kempter (2006); Cayco-Gajic and Shea-Brown (2013). Because the non-linear modulation function saturates, FFNs with non-additive dendritic interactions allow for a sparser coding, i.e., only a sub-fraction of each layer (the actual size depends on the non-linear enhancement level) participates in the propagation of synchrony. Whereas stable propagation of synchrony is possible in systems with and without dendritic non-linearities, it occurs in non-linearly coupled FFNs with substantially reduced feed-forward anatomy (reduced connectivity or reduced coupling strength) compared to linearly coupled FFNs.

The analytic derivation of the critical connectivity is based on rather general assumptions: (a) the effect of a synchronous input pulse is approximately the sum of the effects of single inputs and (b) for networks with non-additive dendritic interactions the spiking probability due to non-linearly enhanced input is substantially larger than due to a non-enhanced input. Therefore the predictions and estimates are directly applicable to networks of biologically more detailed neuron models.

In our article we have shown that even highly diluted feed-forward structures are suitable to reliably support the directed and constrained propagation of synchronous activity. Such structures occur naturally in sparse, random recurrent networks which are typical for the cortex. These structures might be enhanced

by simple synaptic plasticity to enable synchrony propagation. Fast dendritic spikes promote this propagation, as they selectively amplify synchronous inputs and are only weakly influenced by irregular background activity.

Indeed, important candidate regions for the generation of propagating synchrony such as the hippocampus and other, neocortical regions exhibiting replay of activity (Nadasdy et al., 1999; Lee and Wilson, 2002; Ji and Wilson, 2007; Xu et al., 2011; Eagleman and Dragoi, 2012) are sparse and show synaptic plasticity (Debanne et al., 1998; Kobayashi and Poo, 2004). Dendritic spikes as prominently found in, e.g., the hippocampus (Ariav et al., 2003; Gasparini et al., 2004; Polsky et al., 2004; Gasparini and Magee, 2006) trigger depolarizations and calcium influx sufficient to change synaptic strengths (Golding et al., 2002; Remy and Spruston, 2007) and the dendrites itself exhibit branch "strength potentiation," i.e., the strength of a dendritic spike on a dendritic branch exhibits experience- and activity-dependent plasticity (Losonczy et al., 2008; Makara et al., 2009; Müller et al., 2012).

Our work indicates that fast dendritic spikes reduce the required synaptic strength and connection density for replay of spike patterns. Moreover, their saturation and the resulting sparse coding might explain the observed variability during replay. Thus, in particular, our understanding of propagation along diluted feed-forward chains may now be combined with knowledge on synaptic plasticity and generation of activity accompanying replay (e.g., sharp wave/ripples) to gain an integrated mechanistic understanding for encoding, replay and memory transfer.

ACKNOWLEDGMENTS

This work was supported by the BMBF (Grant No. 01GQ1005B) [Sven Jahnke, Marc Timme], the DFG (Grant No. TI 629/3-1) [Sven Jahnke], the Swartz Foundation [Raoul-Martin Memmesheimer], and the Max Planck Society [Marc Timme]. Simulation results of networks with biologically more complex neuron models were obtained using the simulation software NEST (Gewaltig and Diesmann, 2007). Sven Jahnke thanks Harold Gutch, Elian Moritz, and Jonna Jahnke for stimulating discussions.

REFERENCES

Abeles, M. (1982). *Local Cortical Circuits: An Electrophysiological Study*. Berlin: Springer. doi: 10.1007/978-3-642-81708-3

Ariav, G., Polsky, A., and Schiller, J. (2003). Submillisecond precision of the input-output transformation function mediated by fast sodium dendritic spikes in basal dendrites of CA1 pyramidal neurons. *J. Neurosci.* 23, 7750–7758.

August, D. A., and Levy, W. B. (1999). Temporal sequence compression by an integrate-and-fire model of hippocampal area CA3. *J. Comput. Neurosci.* 6, 71–90. doi: 10.1023/A:1008861001091

Aviel, Y., Mehring, C., Abeles, M., and Horn, D. (2003). On embedding synfire chains in a balanced network. *Neural Comp.* 15, 1321–1340. doi: 10.1162/089976603321780290

Braitenberg, V., and Schüz, A. (1998). *Cortex: Statistics and Geometry of Neuronal Connectivity*. Berlin: Springer. doi: 10.1007/978-3-662-03733-1

Brunel, N. (2000). Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons. *J. Comp. Neurosci.* 8, 183–208. doi: 10.1023/A:1008925309027

Brunel, N., and Hakim, V. (1999). Fast global oscillations in networks of integrate-and-fire neurons with low firing rates. *Neural Comp.* 11, 1621–1671. doi: 10.1162/089976699300016179

Bullmore, E., and Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural and functional systems. *Nat. Rev. Neurosci.* 10, 186–198. doi: 10.1038/nrn2618

Burkitt, A. (2006). A review of the integrate-and-fire neuron model: I. Homogeneous synaptic input. *Biol. Cybern.* 95, 1–19. doi: 10.1007/s00422-006-0082-8

Cayco-Gajic, N. A., and Shea-Brown, E. (2013). Neutral stability, rate propagation, and critical branching in feedforward networks. *Neural Comput.* 25, 1768–1806. doi: 10.1162/NECO_a_00461

Dayan, P., and Abbott, L. (2001). *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems*. Cambridge: MIT Press.

Debanne, D., Gähwiler, B. H., and Thompson, S. M. (1998). Long-term synaptic plasticity between pairs of individual CA3 pyramidal cells in rat hippocampal slice cultures. *J. Physiol.* 507, 237–247. doi: 10.1111/j.1469-7793.1998.237bu.x

Diesmann, M., Gewaltig, M. O., and Aertsen, A. (1999). Stable propagation of synchronous spiking in cortical neural networks. *Nature* 402, 529–533. doi: 10.1038/990101

Eagleman, S. L., and Dragoi, V. (2012). Image sequence reactivation in awake V4 networks. *Proc. Natl. Acad. Sci. U.S.A.* 109, 19450–19455. doi: 10.1073/pnas.1212059109

Feinerman, O., and Moses, E. (2006). Transport of information along unidimensional layered networks of dissociated hippocampal neurons and implications for rate coding. *J. Neurosci.* 26, 4526–4534. doi: 10.1523/JNEUROSCI.4692-05.2006

Feinerman, O., Segal, M., and Moses, E. (2005). Signal propagation along unidimensional neuronal networks. *J. Neurophysiol.* 94, 3406–3416. doi: 10.1152/jn.00264.2005

Felleman, D. J., and Van Essen, D. V. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cereb. Cortex* 1, 1–47. doi: 10.1093/cercor/1.1.1

Gasparini, S., and Magee, J. C. (2006). State-dependent dendritic computation in hippocampal CA1 pyramidal neurons. *J. Neurosci.* 26, 2088–2100. doi: 10.1523/JNEUROSCI.4428-05.2006

Gasparini, S., Migliore, M., and Magee, J. C. (2004). On the initiation and propagation of dendritic spikes in CA1 pyramidal neurons. *J. Neurosci.* 24, 11046–11056. doi: 10.1523/JNEUROSCI.2520-04.2004

Gewaltig, M. O., and Diesmann, M. (2007). NEST (NEural Simulation Tool). *Scholarpedia* 2:1430. doi: 10.4249/scholarpedia.1430

Gewaltig, M. O., Diesmann, M., and Aertsen, A. (2001). Propagation of cortical synfire activity: survival probability in single trials and stability in the mean. *Neural Netw.* 14, 657–673. doi: 10.1016/S0893-6080(01)00070-3

Goedeke, S., and Diesmann, M. (2008). The mechanism of synchronization in feed-forward neuronal networks. *New J. Phys.* 10:015007. doi: 10.1088/1367-2630/10/1/015007

Golding, N. L., Staff, N. P., and Spruston, N. (2002). Dendritic spikes as a mechanism for cooperative long-term potentiation. *Nature* 418, 326–331. doi: 10.1038/nature00854

Helias, M., Deger, M., Rotter, S., and Diesmann, M. (2010). Instantaneous non-linear processing by pulse-coupled threshold units. *PLoS Comput. Biol.* 6:e1000929. doi: 10.1371/journal.pcbi.1000929

Holmgren, C., Harkany, T., Svennenfors, B., and Zilberter, Y. (2003). Pyramidal cell communication within local networks in layer 2/3 of rat neocortex. *J. Physiol.* 551, 139–153. doi: 10.1113/jphysiol.2003.044784

Jahnke, S., Timme, M., and Memmesheimer, R.-M. (2012). Guiding synchrony through random networks. *Phys. Rev. X* 2:041016. doi: 10.1103/PhysRevX.2.041016

Ji, D., and Wilson, M. A. (2007). Coordinated memory replay in the visual cortex and hippocampus during sleep. *Nat. Neurosci.* 10, 100–107. doi: 10.1038/nn1825

Kilavik, B. E., Roux, S., Ponce-Alvarez, A., Confais, J., Grün, S., and Riehle, A. (2009). Long-term modifications in motor cortical dynamics induced by intensive practice. *J. Neurosci.* 29, 12653–126563. doi: 10.1523/JNEUROSCI.1554-09.2009

Kobayashi, K., and Poo, M.-M. (2004). Spike train timing-dependent associative modification of Hippocampal CA3 recurrent synapses by mossy fibers. *Neuron* 41, 445–454. doi: 10.1016/S0896-6273(03)00873-0

Kumar, A., Rotter, S., and Aertsen, A. (2008). Conditions for propagating synchronous spiking and asynchronous firing rates in a cortical network model. *J. Neurosci.* 28, 5268–5280. doi: 10.1523/JNEUROSCI.2542-07.2008

Kumar, A., Rotter, S., and Aertsen, A. (2010). Spiking activity propagation in neuronal networks: reconciling differentperspectives on neural coding. *Nat. Rev. Neurosci.* 11, 615–627. doi: 10.1038/nrn2886

Lee, A. K., and Wilson, M. A. (2002). Memory of sequential experience in the Hippocampus during slow wave sleep. *Neuron* 36, 1183–1194. doi: 10.1016/S0896-6273(02)01096-6

Leibold, C., and Kempter, R. (2006). Memory capacity for sequences in a recurrent network with biological constraints. *Neural Compt.* 18, 904–941. doi: 10.1162/neco.2006.18.4.904

Litvak, V., Sompolinsky, H., Segev, I., and Abeles, M. (2013). On the transmission of rate code in long feedforward networks with excitatory-inhibitory balance. *J. Neurosci.* 23, 3006–3015.

Long, M. A., Jin, D. Z., and Fee, M. S. (2010). Support for a synaptic chain model of neuronal sequence generation. *Nature* 468, 394–399. doi: 10.1038/nature09514

Losonczy, A., Makara, J. K., and Magee, J. C. (2008). Compartmentalized dendritic plasticity and input feature storage in neurons. *Nature* 452, 436–441. doi: 10.1038/nature06725

Makara, J. K., Losonczy, A., Wen, Q., and Magee, J. C. (2009). Experience-dependent compartmentalized dendritic plasticity in rat hippocampal CA1 pyramidal neurons. *Nat. Neurosci.* 12, 1485–1487. doi: 10.1038/nn.2428

Mehring, C., Hehl, U., Kubo, M., Diesmann, M., and Aertsen, A. (2003). Activity dynamics and propagation of synchronous spiking in locally connected random networks. *Biol. Cybern.* 88, 395–408. doi: 10.1007/s00422-002-0384-4

Memmesheimer, R.-M. (2010). Quantitative prediction of intermittent high-frequency oscillations in neural networks with supralinear dendritic interactions. *Proc. Natl. Acad. Sci. U.S.A.* 107, 11092–11097. doi: 10.1073/pnas.0909615107

Memmesheimer, R.-M., and Timme, M. (2012). Non-additive coupling enables propagation of synchronous spiking activity in purely random networks. *PLoS Comput. Biol.* 8:e1002384. doi: 10.1371/journal.pcbi.1002384

Müller, C., Beck, H., Coulter, D., and Remy, S. (2012). Inhibitory control of linear and supralinear dendritic excitation in CA1 pyramidal neurons. *Neuron* 75, 851–864. doi: 10.1016/j.neuron.2012.06.025

Nadasdy, Z., Hirase, H., Czurko, A., Csicsvari, J., and Buzsaki, G. (1999). Replay and time compression of recurring spike sequences in the Hippocampus. *J. Neurosci.* 19, 9497–9507.

Nowotny, T., and Huerta, R. (2003). Explaining synchrony in feed-forward networks: are McCulloch-Pitts neurons good enough? *Biol. Cybern.* 89, 237–241. doi: 10.1007/s00422-003-0431-9

Polsky, A., Mel, B. W., and Schiller, J. (2004). Computational subunits in thin dendrites of pyramidal cells. *Nat. Neurosci.* 7, 621–627. doi: 10.1038/nn1253

Putrino, D., Brown, E. N., Mastaglia, F. L., and Ghosh, S. (2010). Differential involvement of excitatory and inhibitory neurons of cat motor cortex in coincident spike activity related to behavioral context. *J. Neurosci.* 30, 8048–8056. doi: 10.1523/JNEUROSCI.0770-10.2010

Remy, S., and Spruston, N. (2007). Dendritic spikes induce single-burst long-term potentiation. *Proc. Natl. Acad. Sci. U.S.A.* 104, 17192–17197. doi: 10.1073/pnas.0707919104

Reyes, A. D. (2003). Synchrony-dependent propagation of firing rate in iteratively constructed networks *in vitro*. *Nat. Neurosci.* 6, 593–599. doi: 10.1038/nn1056

Riehle, A., Grün, S., Diesmann, M., and Aertsen, A. (1997). Spike synchronization and rate modulation differentially involved in motor cortical function. *Science* 278, 1950–1953. doi: 10.1126/science.278.5345.1950

Rosenbaum, R. J., Trousdale, J., and Josic, K. (2010). Pooling and correlated neural activity. *Front. Comput. Neurosci.* 4:9. doi: 10.3389/fncom.2010.00009

Rosenbaum, R., Trousdale, J., and Josic, K. (2011). The effect of pooling on spike train correlations. *Front. Neurosci.* 5:58. doi: 10.3389/fnins.2011.00058

Scannell, J. W., Burns, G. A., Hilgetag, C. C., O'Neil, M. A., and Young, M. P. (1999). The connectional organization of the cortico-thalamic system of the cat. *Cereb. Cortex* 9, 277–299. doi: 10.1093/cercor/9.3.277

Tuckwell, H. (1988). *Introduction to Theoretical Neurobiology*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511623202

van Vreeswijk, C., and Sompolinsky, H. (1996). Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science* 274, 1724–1726. doi: 10.1126/science.274.5293.1724

van Vreeswijk, C., and Sompolinsky, H. (1998). Chaotic balanced state in a model of cortical circuits. *Neural Comp.* 10, 1321–1371. doi: 10.1162/089976698300017214

Vogels, T. P., and Abbott, L. F. (2005). Signal propagation and logic gating in networks of integrate-and-fire neurons. *J. Neurosci.* 25, 10786–10795. doi: 10.1523/JNEUROSCI.3508-05.2005

Xu, S., Jiang, W., Poo, M.-M., and Dan, Y. (2012). Activity recall in a visual cortical ensemble. *Nat. Neurosci.* 15, 449–455. doi: 10.1038/nn.3036

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.

Received: 23 June 2013; accepted: 11 October 2013; published online: 15 November 2013.

*Citation: Jahnke S, Memmesheimer R-M and Timme M (2013) Propagating synchrony in feed-forward networks. Front. Comput. Neurosci. **7**:153. doi: 10.3389/fncom.2013.00153*

This article was submitted to the journal Frontiers in Computational Neuroscience.

Copyright © 2013 Jahnke, Memmesheimer and Timme. 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) or licensor 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.

A. APPENDIX

A.1 PROOF OF EXISTENCE OF A GLOBAL MINIMUM OF $p_{NL}(n)$

We will show that $p_{NL}(n)$ as derived in Equation (59),

$$p_{NL}(n) = \frac{n^2\epsilon + 2\Theta_b + n\sqrt{n^2\epsilon^2 + 4\Theta_b\left(\epsilon - \frac{\Theta_b}{\omega}\right)}}{p_f(\kappa)\epsilon(n^2+\omega)\left(1+\mathrm{Erf}\left(\frac{n}{\sqrt{2}}\right)\right)} \tag{A.1}$$

$$= \frac{1}{p_f(\kappa)}\frac{2\Theta_b + n^2\epsilon\left(1+\sqrt{1+\frac{\alpha}{n^2}}\right)}{\left(1+\mathrm{Erf}\left(\frac{n}{\sqrt{2}}\right)\right)\left(n^2\epsilon+\omega\epsilon\right)}, \tag{A.2}$$

has a global minimum for $\epsilon\omega > \Theta_b$. In Equation (A.2) we defined

$$\alpha := \frac{4\Theta_b}{\epsilon}\left(1-\frac{\Theta_b}{\epsilon\omega}\right). \tag{A.3}$$

For $\epsilon\omega > \Theta_b$, p_{NL} is positive and continuous, and approaches

$$\lim_{n\to-\infty}\left(p_{NL}(n)\right) = \infty, \tag{A.4}$$

$$\lim_{n\to\infty}\left(p_{NL}(n)\right) = \frac{1}{p_f(\kappa)}, \tag{A.5}$$

in the limit of large/small n. Further, the derivative of p_{NL} can be written as

$$\frac{d}{dn}p_{NL}(n) = (2-h_1(n))\,h_2(n), \tag{A.6}$$

where we defined the functions

$$h_1(n) = \frac{1}{\epsilon\omega}\left(\frac{\sqrt{\frac{2}{\pi}}e^{-\frac{n^2}{2}}\left(n^2+\omega\right)\left(\frac{2\Theta_b}{\sqrt{\alpha+n^2}+n}+n\epsilon\right)}{1+\mathrm{Erf}\left(\frac{n}{\sqrt{2}}\right)} + \frac{4\Theta_b n}{\sqrt{\alpha+n^2}+n} + \frac{\alpha\epsilon\left(n^2+\omega\right)}{\alpha+n\left(\sqrt{\alpha+n^2}+n\right)}\right), \tag{A.7}$$

$$h_2(n) = \frac{\omega\left(\sqrt{\alpha+n^2}+n\right)}{p_f(\kappa)\left(\mathrm{Erf}\left(\frac{n}{\sqrt{2}}\right)+1\right)\left(n^2+\omega\right)^2}. \tag{A.8}$$

For $n > 0$ and $\epsilon\omega > \Theta_b$,

$$\alpha > 0, \tag{A.9}$$

$$h_1(n) > 0, \tag{A.10}$$

$$h_2(n) > 0, \tag{A.11}$$

and in the limit of large n,

$$\lim_{n\to\infty} h_1(n) = \frac{1}{\epsilon\omega}\left(0+2\Theta_b+\frac{\alpha\epsilon}{2}\right) \tag{A.12}$$

$$= 2\frac{2\Theta_b\omega\epsilon-\Theta_b^2}{\omega^2\epsilon^2} \tag{A.13}$$

$$\lim_{n\to\infty} h_2(n) = 0. \tag{A.14}$$

For $\epsilon\omega > \Theta_b$, $h_1(n)$ is smaller than two for sufficiently large n (cf. Equation A.13) and thus the derivative of $p_{NL}(n)$ becomes positive (cf. Equation A.6). Consequently p_{NL} approaches $1/p_f(\kappa)$ from below for large n (cf. also Equation A.5). This proves the existence of a global minimum of $p_{NL}(n)$, because $p_{NL} > 1/p_f(\kappa)$ for sufficiently small n (cf. Equation A.4).

A.2 BIOLOGICAL MORE DETAILED NEURON MODELS

In section 3.3.3 we consider biologically more detailed neuron models. In this appendix we present descriptions of these models including the parameters used for the numerical simulations in **Figure 13**. These simulations were done using NEST (Gewaltig and Diesmann, 2007), a simulator for spiking neural network models (available at http://www.nest-initiative.org). We implemented new model classes within the NEST framework to handle conductance-based leaky integrate-and-fire neurons with double exponential input conductances as well as non-linear dendritic interactions (source code available from Sven Jahnke).

A.2.1 CB-type model

The CB-type model is a leaky integrate-and-fire neuron with conductance based synapses, augmented with a mechanism for the generation of current pulses mimicking the effect of a dendritic spike (see also Memmesheimer, 2010; Jahnke et al., 2012). The subthreshold dynamics of the membrane potential V_l of neuron l obeys the differential equation

$$C_l^{\mathrm{m}}\frac{dV_l(t)}{dt} = g_l^L\left(V_l^{\mathrm{rest}}-V_l(t)\right) + g_l^A(t)\left(E^{\mathrm{Ex}}-V_l(t)\right) + g_l^G(t)\left(E^{\mathrm{In}}-V_l(t)\right) + I_l^{\mathrm{DS}}(t) + I_l^0. \tag{A.15}$$

Here, C_l^{m} is the membrane capacity, g_l^L is the resting conductance, V_l^{rest} is the resting membrane potential, E^{Ex} and E^{In} are the reversal potentials, and $g_l^A(t)$ and $g_l^G(t)$ are the conductances of excitatory and inhibitory synaptic populations, respectively. $I_l^{\mathrm{DS}}(t)$ models the current pulses caused by dendritic spikes and I_l^0 is a constant current gathering slow external and internal currents. The time course of single synaptic conductances contributing to $g_l^A(t)$ and $g_l^G(t)$ is given by the difference between two exponential functions (e.g., Dayan and Abbott, 2001) with time constants $\tau^{A,1}$ and $\tau^{A,2}$ for the excitatory and $\tau^{G,1}$ and $\tau^{G,2}$ for the inhibitory conductances. Whenever the membrane potential reaches the spike threshold Θ_l, the neuron sends a spike to its postsynaptic neurons, is reset to V_l^{reset} and becomes refractory for a period t_l^{ref}. Additionally to inputs from the preceding layer each neuron receives excitatory and inhibitory Poissonian input spike trains with rates ν^{ex} and ν^{in}; single inputs have coupling strength ϵ^{ex} and ϵ^{in}, respectively.

To account for dendritic spike generation, we consider the sum $g_{l,\Delta t}$ of excitatory input strengths (characterized by the coupling strengths), arriving at an excitatory neuron l within the time window Δt for non-linear dendritic interactions,

$$g_{l,\Delta t}(t) = \sum_j\sum_k \epsilon_{lj}\chi_{[t-\Delta t,\,t]}(t_{jk}^f+\tau), \tag{A.16}$$

where $\chi_{[t-\Delta t,\,t]}$ is the characteristic function of the interval $[t-\Delta t,\,t]$, t_{jk}^{f} is the kth firing time of neuron j and τ denotes the synaptic delay. We denote the peak conductance (coupling strength) for a connection from neuron j to neuron l by $g_{lj}^{\max}$. If $g_{l,\,\Delta t}$ exceeds a threshold g_{Θ}, a dendritic spike is initiated and the dendrite becomes refractory for a time window $t^{\mathrm{DS,ref}}$. The effect of the dendritic spike is incorporated into the model by the current pulse that reaches the soma a time τ^{DS} thereafter. This current pulse is modeled as the sum of three exponential functions,

$$I_l^{\mathrm{DS}}(t) = c(g_{\Delta t})\left[-Ae^{-\frac{t}{\tau^{\mathrm{DS},1}}} + Be^{-\frac{t}{\tau^{\mathrm{DS},2}}} - Ce^{-\frac{t}{\tau^{\mathrm{DS},3}}}\right], \quad (A.17)$$

with prefactors $A>0$, $B>0$, $C>0$, decay time constants $\tau^{\mathrm{DS},1}$, $\tau^{\mathrm{DS},2}$, $\tau^{\mathrm{DS},3}$ and a dimensionless correction factor $c\left(g_{\Delta t}\right)$, where $g_{\Delta t}$ is the summed excitatory input at the initiation time of the dendritic spike as given by Equation (A.16). The factor $c\left(g_{\Delta t}\right)$ modulates the pulse strength, ensuring that the peak of the excitatory postsynaptic potential (pEPSP) reaches the experimentally observed region of saturation. At very high excitatory inputs, the conventionally generated depolarization exceeds the level of saturation, $c\left(g_{\Delta t}\right)$ is zero and the pEPSP increases (cf. inset of **Figure 13A**).

Parameters for Figure 13

The single neuron parameters for the numerical simulations are $C_l^{\mathrm{m}} = C^{\mathrm{m}} = 400\,\mathrm{pF}$, $g_l^L = g^L = 25\,\mathrm{nS}$, $V_l^{\mathrm{rest}} = V^{\mathrm{rest}} = -65\,\mathrm{mV}$, $\Theta_l = \Theta = -50\,\mathrm{mV}$, $t_l^{\mathrm{ref}} = t^{\mathrm{ref}} = 3\,\mathrm{ms}$ and $V_l^{\mathrm{reset}} = V^{\mathrm{reset}} = -65\,\mathrm{mV}$. The reversal potentials are $E^{\mathrm{Ex}} = 0\,\mathrm{mV}$ and $E^{\mathrm{In}} = -75\,\mathrm{mV}$ and the time constants for the excitatory and inhibitory conductances are $\tau^{A,1} = \tau^{G,1} = 2.5\,\mathrm{ms}$ and $\tau^{A,2} = \tau^{G,2} = 0.5\,\mathrm{ms}$. The parameters of the dendritic spike current are $\Delta t = 2\,\mathrm{ms}$, $g^{\Theta} = 8.65\,\mathrm{nS}$, $\tau^{\mathrm{DS}} = 2.7\,\mathrm{ms}$, $A = 55\,\mathrm{nA}$, $B = 64\,\mathrm{nA}$, $C = 9\,\mathrm{nA}$, $\tau^{\mathrm{DS},1} = 0.2\,\mathrm{ms}$, $\tau^{\mathrm{DS},2} = 0.3\,\mathrm{ms}$, $\tau^{\mathrm{DS},3} = 0.7\,\mathrm{ms}$ and $t^{\mathrm{ref,\,DS}} = 5.2\,\mathrm{ms}$ and the dimensionless correction factor is given by $c(g) = \max\left\{1.5 - g\cdot 0.053\mathrm{nS}^{-1}, 0\right\}$. For the first setup ($p^f \approx 0.97$) we set $I_l^0 = I^0 = 250\,\mathrm{pA}$, $\nu^{\mathrm{ex}} = 2.4\,\mathrm{kHz}$, $\nu^{\mathrm{in}} = 0.6\,\mathrm{kHz}$, $\epsilon^{\mathrm{ex}} = 0.6\,\mathrm{nS}$ and $\epsilon^{\mathrm{in}} = 6.6\,\mathrm{nS}$; for the second setup ($p^f \approx 0.67$) we set $I_l^0 = I^0 = 0\,\mathrm{pA}$, $\nu^{\mathrm{ex}} = 20\,\mathrm{kHz}$, $\nu^{\mathrm{in}} = 5\,\mathrm{kHz}$, $\epsilon^{\mathrm{ex}} = 0.6\,\mathrm{nS}$ and $\epsilon^{\mathrm{in}} = -6.6\,\mathrm{nS}$.

A.2.2 HH-type model

We employ a standard model provided by NEST ("hh_psc_alpha"; Hodgkin–Huxley type neuron with alpha-function shaped postsynaptic currents) and incorporated a dendritic spike current as in the CB-Model. The membrane potential V_l of neuron l obeys the differential equation

$$C_l^{\mathrm{m}}\frac{dV_l(t)}{dt} = I_l^{\mathrm{Na}}(t) + I_l^{\mathrm{K}}(t) + I_l^{\mathrm{L}}(t) + I_l^0 + I_l^{\mathrm{ex}}(t) + I_l^{\mathrm{in}}(t) + I_l^{\mathrm{DS}}(t). \quad (A.18)$$

For clarity we drop the index l in the following; all quantities refer to some neuron l. In Equation (A.18),

$$I^{\mathrm{Na}}(t) = g^{\mathrm{Na}}m(t)^3h(t)\left[E^{\mathrm{Na}} - V(t)\right] \quad (A.19)$$

$$I^{\mathrm{K}}(t) = g^{\mathrm{K}}n(t)^4\left[E^{\mathrm{K}} - V(t)\right] \quad (A.20)$$

$$I^{\mathrm{L}}(t) = g^{\mathrm{L}}\left[E^{\mathrm{L}} - V(t)\right] \quad (A.21)$$

specify the Na^+current, the K^+ current and leak current. The dynamics of the gating variables m, n and h are governed by

$$\frac{dm(t)}{dt} = \alpha_{\mathrm{m}}(t)\,[1-m(t)] - \beta_{\mathrm{m}}(t)m(t) \quad (A.22)$$

$$\frac{dh(t)}{dt} = \alpha_{\mathrm{h}}(t)\,[1-h(t)] - \beta_{\mathrm{h}}(t)h(t) \quad (A.23)$$

$$\frac{dn(t)}{dt} = \alpha_{\mathrm{n}}(t)\,[1-n(t)] - \beta_{\mathrm{n}}(t)n(t), \quad (A.24)$$

where the voltage dependencies are given by

$$\alpha_n(t) = \frac{0.01\left[\tilde{V}(t)+55\right]}{1-\exp\left[-\frac{\tilde{V}(t)+55}{10}\right]} \quad (A.25)$$

$$\beta_n(t) = 0.125\cdot\exp\left[-\frac{\tilde{V}(t)\,+\,65}{80}\right] \quad (A.26)$$

$$\alpha_m(t) = \frac{0.1\left[\tilde{V}(t)+40\right]}{1-\exp\left[-\frac{V(t)+40}{10}\right]} \quad (A.27)$$

$$\beta_m(t) = 4\cdot\exp\left[-\frac{\tilde{V}(t)+65}{18}\right] \quad (A.28)$$

$$\alpha_h(t) = 0.07\cdot\exp\left[-\frac{\tilde{V}(t)+65}{20}\right] \quad (A.29)$$

$$\beta_h(t) = \left(1+\exp\left[-\frac{\tilde{V}(t)+35}{10}\right]\right)^{-1}. \quad (A.30)$$

In Equations (A.25–A.30) $\tilde{V}(t) := \frac{V(t)}{1\mathrm{mV}}$ is the value of membrane potential normalized by 1 mV. Spikes are detected by a combined threshold-and-local-maximum search, if there is a local maximum above a certain threshold of the membrane potential, $U_{\Theta} = 0\,\mathrm{mV}$, it is considered a spike (for more details see the NEST manual and the model implementation available at http://www.nest-initiative.org). After a synaptic delay time τ a spike initiates an alpha-function shaped current pulse at the postsynaptic neurons. The total excitatory and inhibitory input to neuron l is given by

$$I^{\mathrm{ex}}(t) = \sum_k \epsilon_k^{\mathrm{ex}}\frac{e}{\tau^{\mathrm{ex}}}\exp\left[-\frac{t}{\tau^{\mathrm{ex}}}\right]\Theta\left[t-t_k^{\mathrm{ex}}\right] \quad (A.31)$$

$$I^{\mathrm{in}}(t) = \sum_k \epsilon_k^{\mathrm{in}}\frac{e}{\tau^{\mathrm{in}}}\exp\left[-\frac{t}{\tau^{\mathrm{in}}}\right]\Theta\left[t-t_k^{\mathrm{in}}\right], \quad (A.32)$$

where $\epsilon_k^{\mathrm{ex}} > 0$ $\left(\epsilon_k^{\mathrm{in}} < 0\right)$ is the strength of the kth arriving excitatory (inhibitory) spike at neuron l, t_k^{ex} $\left(t_k^{\mathrm{in}}\right)$ denotes the reception time of that spike and e is the Euler constant [the currents $I^{\mathrm{ex}}(t)$ and $I^{\mathrm{in}}(t)$ are normalized such that an input of strength $\epsilon = 1\,\mathrm{pA}$

causes a peak current of 1 pA]. The time constants τ^{ex} and τ^{in} are the synaptic time constants. As before, we account for dendritic spike generation by considering the sum of excitatory input strengths received by neuron l within the time window Δt,

$$\epsilon_{\Delta t}(t) = \sum_k \epsilon_k^{ex} \chi_{[t-\Delta t,t]}(t_k^f + \tau). \quad (A.33)$$

If this sum exceeds the dendritic threshold I^{Θ}, a dendritic spike is initiated and we model its effect is by the current pulse

$$I^{DS}(t) = c(\epsilon_{\Delta t}) \left[-Ae^{-\frac{t}{\tau^{DS,1}}} + Be^{-\frac{t}{\tau^{DS,2}}} - Ce^{-\frac{t}{\tau^{DS,3}}} \right], \quad (A.34)$$

starting after a delay time τ^{DS} after the initiation time of the dendritic spike. The correction factor $c(\epsilon_{\Delta t})$ modulates the pulse strength such that the depolarization saturates for suprathreshold inputs until the effects of linearly summed input exceed the effects of the dendritic spike (cf. inset of **Figure 13B**).

A.2.3 Parameters for Figure 13

As before, we consider homogeneous neuronal properties. The single neuron parameters for the numerical simulations are $C^m = 200\,\text{pF}$, $E^K = -77\,\text{mV}$, $E^L = -70\,\text{mV}$, $E^{Na} = 50\,\text{mV}$, $g^K = 3600\,\text{nS}$, $g^L = 30\,\text{nS}$, $g^{Na} = 12000\,\text{nS}$, $\tau^{ex} = 2\,\text{ms}$ and $\tau^{in} = 2\,\text{ms}$. The parameters of the dendritic spike current are $\Delta t = 3.5\,\text{ms}$, $I^{\Theta} = 270\,\text{pA}$, $\tau^{DS} = 2.7\,\text{ms}$, $A = 27.5\,\text{nA}$, $B = 32\,\text{nA}$, $C = 4.5\,\text{nA}$, $\tau^{DS,1} = 0.2\,\text{ms}$, $\tau^{DS,2} = 0.3\,\text{ms}$, $\tau^{DS,3} = 0.7\,\text{ms}$ and $t^{ref,DS} = 5.2\,\text{ms}$ and the dimensionless correction factor is given by $c(\epsilon) = \max\left\{1.54 - \epsilon \cdot 0.002\,\text{pA}^{-1}, 0\right\}$. For the first setup ($p^f \approx 0.97$) we set $I^0 = 500\,\text{pA}$, $\nu^{ex} = 3\,\text{kHz}$, $\nu^{in} = 3\,\text{kHz}$, $\epsilon^{ex} = 20\,\text{pA}$ and $\epsilon^{in} = -20\,\text{pA}$; and for the second setup ($p^f \approx 0.67$) we set $I^0 = 250\,\text{pA}$, $\nu^{ex} = 10\,\text{kHz}$, $\nu^{in} = 10\,\text{kHz}$, $\epsilon^{ex} = 20\,\text{pA}$ and $\epsilon^{in} = -20\,\text{pA}$.

frontiers in
COMPUTATIONAL NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 07 November 2013
doi: 10.3389/fncom.2013.00156

Simultaneous stability and sensitivity in model cortical networks is achieved through anti-correlations between the in- and out-degree of connectivity

Juan C. Vasquez[1]*, Arthur R. Houweling[2] and Paul Tiesinga[1]*

[1] *Department of Neuroinformatics, Donders Institute for Brain, Cognition and Behavior, Radboud University Nijmegen, Nijmegen, Netherlands*
[2] *Department of Neuroscience, Erasmus Medical Center, Rotterdam, Netherlands*

Edited by:
Robert Rosenbaum, University of Pittsburgh, USA

Reviewed by:
Takuma Tanaka, Tokyo Institute of Technology, Japan
Xin Tian, Tianjin Medical University, China

****Correspondence:***
Juan C. Vasquez and Paul Tiesinga, Department of Neuroinformatics, Donders Institute for Brain, Cognition and Behavior, Radboud University Nijmegen, Heyendaalseweg 135, Postvak 66, 6525 AJ Nijmegen, Netherlands
e-mail: jc.vasquez@science.ru.nl; p.tiesinga@science.ru.nl

Neuronal networks in rodent barrel cortex are characterized by stable low baseline firing rates. However, they are sensitive to the action potentials of single neurons as suggested by recent single-cell stimulation experiments that reported quantifiable behavioral responses in response to short spike trains elicited in single neurons. Hence, these networks are stable against internally generated fluctuations in firing rate but at the same time remain sensitive to similarly-sized externally induced perturbations. We investigated stability and sensitivity in a simple recurrent network of stochastic binary neurons and determined numerically the effects of correlation between the number of afferent ("in-degree") and efferent ("out-degree") connections in neurons. The key advance reported in this work is that anti-correlation between in-/out-degree distributions increased the stability of the network in comparison to networks with no correlation or positive correlations, while being able to achieve the same level of sensitivity. The experimental characterization of degree distributions is difficult because all pre-synaptic and post-synaptic neurons have to be identified and counted. We explored whether the statistics of network motifs, which requires the characterization of connections between small subsets of neurons, could be used to detect evidence for degree anti-correlations. We find that the sample frequency of the 3-neuron "ring" motif ($1 \rightarrow 2 \rightarrow 3 \rightarrow 1$), can be used to detect degree anti-correlation for sub-networks of size 30 using about 50 samples, which is of significance because the necessary measurements are achievable experimentally in the near future. Taken together, we hypothesize that barrel cortex networks exhibit degree anti-correlations and specific network motif statistics.

Keywords: barrel cortex, detection threshold, nanostimulation, degree distribution, computational model, network motifs

INTRODUCTION

Rodents can be trained to use their whiskers to detect an object that predicts a reward and respond with licking to obtain this reward (Huber et al., 2012). The neural responses in barrel cortex to whisker stimulation are hypothesized to play an important role in performing this task (Petersen and Crochet, 2013). Animals can also be trained to detect electrical microstimulation (Butovas and Schwarz, 2007; Houweling and Brecht, 2008) or optogenetic stimulation (Huber et al., 2008) of barrel cortex. Microstimulation activates a large number of neurons that are spatially distributed within a few hundred microns around the stimulating electrode (Histed et al., 2009). An important question is how many neurons need to be activated for the subject to reliably detect the stimulation and whether some cell types are more sensitive than others. Answers to these questions may come from nanostimulation experiments in which a single neuron is activated through juxtacellular stimulation (Houweling and Brecht, 2008). These experiments show that adding trains of 10-15 action potentials in a single cortical neuron can indeed be detected, but the reliability of detection is low and reaction times are long compared to microstimulation.

The spontaneous firing rates in the barrel cortex are low, ranging from less than 1 Hz in the superficial layers to a few Hz in the deep layers (de Kock and Sakmann, 2009; Barth and Poulet, 2012), and whisker stimuli typically evoke a single spike (or none) in responsive neurons. The activity in the low firing rate state (LFS) is also stochastic, both in time as well as across cells, but the precise nature of sparse firing is still being quantified (Barth and Poulet, 2012). For a LFS a single spike could represent a significant perturbation, potentially yielding 28 additional spikes in postsynaptic neurons (London et al., 2010). The network state therefore needs to be stable against small fluctuations that may be amplified through recurrent connectivity. At the same time the aforementioned experiments show that the network is sensitive to small perturbations that are externally generated. Sensitivity and stability are connected and can in general not be optimized at the same time, as the increase in one causes a decrease in the other. Furthermore, stable LFS, in the sense of asynchronous and irregular activity, is difficult to achieve (Kumar et al., 2008).

We use two insights to find the optimal trade-off between stability and sensitivity. First, the external and internal generated firing rate fluctuations may have different statistics. The external perturbation is a train of action potentials [e.g., of 200 ms duration (Houweling and Brecht, 2008)] in a single neuron, thus correlated in time, whereas the internal fluctuations are likely to be of shorter duration and involve a more diverse set of neurons. Second, network structure may be such that these fluctuations have different stability properties (possibly through learning). Our guiding hypothesis is that simultaneous stability and sensitivity are achieved through an anti-correlation between the in- and out-degree of synaptic connectivity between neurons in barrel cortex. Thus, neurons with a low number of synaptic inputs have a high number of synaptic outputs and neurons with a high number of inputs have a low number of outputs. We further hypothesize that such an anti-correlation leads to a distribution of synaptic connectivity motifs that is different than for a random network (Milo et al., 2002). Experiments show that barrel cortical circuits have a motif distribution that is different from random (Song et al., 2005; Perin et al., 2011), whereas theoretical studies show that networks with non-random motif distribution have different synchronization properties (Roxin, 2011; Zhao et al., 2011; Litwin-Kumar and Doiron, 2012) (LaMar and Smith, 2010) and can emerge through synaptic plasticity during reward-based learning (Bourjaily and Miller, 2011a,b). Our work is the first that focuses on the effect on network dynamics of correlations between the in- and out-degree of the same neuron, rather than between in- and/or out-degrees of different neurons, which is referred to as assortativity (Newman, 2010).

Here we test these hypotheses in simplified networks of neurons. In order to focus on the effect of network structure, rather than the full dynamics of spiking neurons, we model neurons as binary units. The inputs to the binary units are determined through a connection matrix with a pre-specified degree distribution generated by a configuration model (Newman, 2010). We first describe how the networks are constructed and then determine (1) their stability in terms of the maximal coupling constant for which the LFS is still stable and (2) their sensitivity to single-cell perturbations using a receiver operating characteristic (ROC) analysis. Finally, we address the issue of how to detect evidence for anti-correlations in the degree distribution experimentally on the basis of sampling sub-networks.

Taken together, we find that anti-correlated networks are more stable than equivalent correlated and uncorrelated networks, but can still reach the same level of sensitivity, which represents a key theoretical advance in terms of a hypothesis for the experimentally observed sensitivity and stability of neuronal networks in the rodent barrel cortex. Furthermore, the hypothesis is of experimental significance, because our analysis shows that correlations in the degree distribution can be detected using sub-networks of sizes that are experimentally accessible in the near future.

METHODS

NETWORK DYNAMICS

The model network was composed of N binary excitatory neurons, whose state at time t is given by $x_i(\mathrm{t})$, a N-dimensional vector with ones for neurons that are active and zeros for ones that are not, here i is the index of the neuron. The new state $x_i(\mathrm{t}+1)$ is obtained in two steps. First, the probability $\nu_{i,t+1}$ of a neuron being active is calculated using Equation (1). Second, for each neuron the firing probability is compared to a random number that is uniformly distributed between 0 and 1. The neuron is set to 1 when the random number is less than or equal to the probability value

$$\nu_{i,\,t+1} = \frac{1}{1+\exp\left(h_0 - \frac{J}{Np_c}\sum_j w_{ij}\,x_{j,t}\right)} \tag{1}$$

The probability has a sigmoidal form, with the exponent consisting of a constant term h_0, which sets the probability of firing in the absence of inputs from other neurons, and a coupling term representing the network input. The coupling term contains the adjacency matrix w_{ij}, whose construction is described below, and in which $w_{ij} = 1$ if there is an input from neuron j to neuron i and $w_{ij} = 0$ otherwise. The overall probability of a connection is p_c. Hence the sum across rows of the adjacency matrix is on average Np_c and we normalize the coupling term by J/Np_c so that J then represents the overall coupling strength. The network activity is calculated in time bins that we consider to be 10 ms. The network has a high firing rate state (HFS), in which each neuron is active on each time step, to which the network will converge when enough neurons are active on a previous time step. We are primarily interested in the LFS, in which each neuron fires only in a fraction of the time bins, corresponding to a firing rate of approximately 1 Hz (Barth and Poulet, 2012). Alternatively, in a given time bin, only a fraction of neurons are active.

The network activity is represented by the mean probability of firing of a neuron during a time bin and is calculated as the total number of spikes divided by the number of neurons. When normalized by the bin width, it represents the mean firing rate of a network neuron in spikes per second (Hz).

NETWORK CONNECTIVITY

Our goal is to determine whether correlations in the in- and out-degree distribution are beneficial in that they increase sensitivity and/or stability relative to uncorrelated networks. Hence, we need a control network without degree correlations. Although the standard random network, Erdos-Renyi (ER) (Newman, 2010), does not have correlations in the degree distribution and is easy to generate samples of, it is not appropriate as a control because it has a sharp degree distribution (see below) and we instead need large variance degree distributions.

For ER networks with a connection probability p, the degree distribution (for both out- as well as in-degree) is given by a binomial distribution

$$p(k) = \binom{N-1}{k} p^k (1-p)^{N-1-k} \tag{2}$$

which has a mean of $(N-1)p$ and a variance of $(N-1)p(1-p)$, which in the limit of large N converges to Gaussian distribution

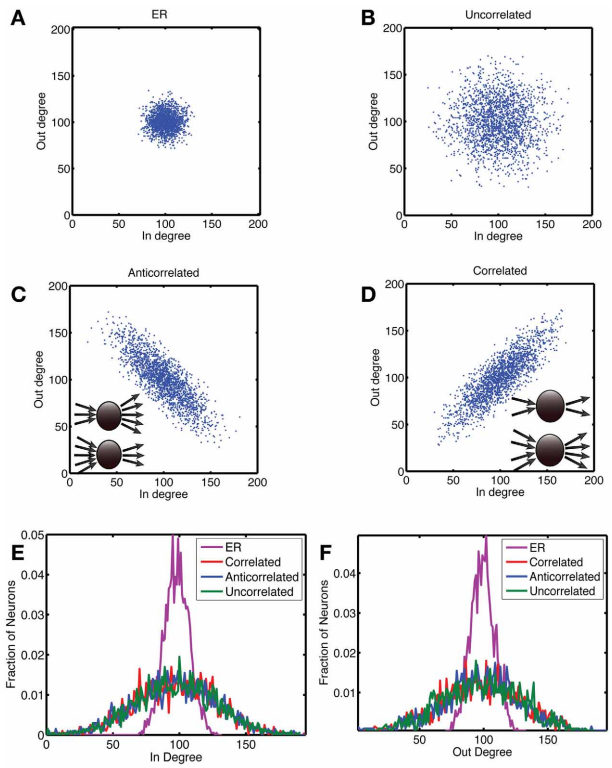


FIGURE 1 | Construction of networks with a correlation between out- and in-degree. In panels **(A)** to **(D)**, we show scatterplots of the out-degree vs. in-degree, whereas the corresponding marginal distributions are shown for **(E)** the in-degree and **(F)** the out-degree. We considered four types of networks, each with $N = 2000$ neurons and a connection probability of $p_c = 0.05$. **(A)** The Erdos-Renyi (ER) network in which each connection is chosen at random with a probability $p_c = 0.05$, for which there is no correlation [$\rho = 0.0034$ (standard deviation: 0.018)] and the relative variance of in- and out-degree across neurons is small for large networks. In order to examine networks with a higher variance of degree values, we first generated a degree distribution in the form of a truncated, bivariate Gaussian. In **(B)** the covariance matrix was diagonal, with equal variance for the out- and in-degrees, which yielded uncorrelated in- and out-degrees [$\rho = 0.0010$ (0.019)]. To generate correlations we started from a covariance matrix with unequal variances and rotated it by 45 degrees anticlockwise to obtain **(C)** anti-correlated [$\rho = 0.821$ (0.0085)] and by 45 degrees clockwise to obtain **(D)** correlated degree distributions [$\rho = 0.821$ (0.0085)]. In the anti-correlated case, nodes with a high out-degree had a low in-degree and vice versa, whereas in the correlated case, nodes with a high out-degree also had a high in-degree, as illustrated by the insets in **(C)** and **(D)**, respectively. **(E,F)** The networks were constructed so that the marginal distributions for the correlated (red), anti-correlated (blue) and uncorrelated (green) case were the same. The ER network (purple) had much tighter marginal distributions.

with a ratio of the standard deviation over the mean of

$$\sqrt{\frac{1-p}{N}} \tag{3}$$

This means that the distribution becomes very tight for large network sizes (**Figures 1A,E,F**).

Hence we generated networks from a truncated bivariate Gaussian for the joint in- and out-degree distribution as explained below (**Figures 1B–F**). We start from a bivariate Gaussian with a diagonal covariance matrix given by

$$p(x,y) = \frac{1}{\sqrt{4\pi^2}\sigma_x\sigma_y}\exp\left(-\frac{(x-\mu)^2}{2\sigma_x^2} - \frac{(y-\mu)^2}{2\sigma_y^2}\right) \tag{4}$$

which is rotated across 45 degrees clockwise or anticlockwise to obtain a distribution with positive and negative correlations, respectively. The resulting distribution is truncated below at 1 because the degree cannot be negative and we exclude the case of zero (since a zero degree neuron would not be considered part of the network) and above at twice the mean degree to make the distribution symmetric. The resulting distribution is normalized to make the integral over the positive quadrant equal to one. The short axis is represented by σ_x and the long axis is represented by σ_y. The mean degree μ was equal to Np_c, with a network size $N = 2000$ and connection probability $p_c = 0.05$ (Holmgren et al., 2003) this yields $\mu = 100$. The long axis was $\sigma_y = \mu/3$. The term dispersion refers to the ratio σ_x/σ_y, which was set to 0.3 for the standard parameter set.

Correlated degree distributions were obtained by sampling for each neuron i, the in- and out-degree from the above bivariate Gaussian, d_i^{in} and d_i^{out}. The simplest method for generating a realization of the corresponding network is the configuration method (Newman, 2010). A list with d_i^{out} stubs with value i, is made and concatenated into a list s_k^{out}. Likewise, a list with d_i^{in} stubs with value i, is made and concatenated into a list s_k^{in} and randomly permuted. From these two lists, pairs are picked from the same position, i.e., the kth stub on the out-list is matched to the kth stub on the in-list to make the connection s_k^{out} to s_k^{in}. This algorithm produces networks with two artifacts, there could be self-connections $s_k^{\text{out}} = s_k^{\text{in}}$, and a given connection could be sampled twice (or more), $s_k^{\text{out}} = s_l^{\text{out}}$ and $s_k^{\text{in}} = s_l^{\text{in}}$. For sparse networks the likelihood of self-edges is small (0.05%), but the probability for multi-edges was larger, around 2.7%. For the cases in which there were multi- or self-edges, we removed the corresponding links.

NETWORK STABILITY

Cortical networks with a low firing rate need to be stable in the sense that stochastic fluctuations should not lead to large increases in the firing rate that could be detected as a stimulation, resulting in a false positive. We characterized the network stability in three ways.

First, we simulated the network and determined the mean firing rate, averaged across neurons and across time bins, as a function of the coupling strength J for various levels of background activity h_0. To determine both the maximal stability and tease apart the contribution of neuronal heterogeneity and stochasticity to instability, we performed the simulations according to a number of different schemes. We considered the mean field limit, in which the network is taken to be so large that each neuron received the same number of inputs and that the resulting

mean firing rate of each neuron was the same. Equation 1 reduces in that case to

$$\nu_{t+1} = \frac{1}{1 + \exp(h_0 - J\nu_t)} \tag{5}$$

yielding the following equation for the fixed points

$$\nu = \frac{1}{1 + \exp(h_0 - J\nu)} \tag{6}$$

which correspond to the roots of the function

$$f(x) = x - \frac{1}{1 + \exp(h_0 - Jx)} \tag{7}$$

and can be obtained by iterating the fixed point equation Equation 6 or using Matlab's root finder fzero. The background field h_0 determines the baseline firing rate r_0, which is the rate obtained in the absence of coupling, $J = 0$:

$$r_0 = \frac{1}{\Delta t} \frac{1}{1 + \exp(h_0)} \tag{8}$$

where we have divided by the bin size Δt to obtain a firing rate in Hz.

There is always a high firing rate solution for sufficiently high coupling strength J, because when all neurons are active on a given time step, they will also all be active on the next time step. There can also be a low firing rate solution which depends on the coupling strength and the baseline firing rate. The coupling strength J_c at a given baseline firing rate below which the LFS exist is the upper limit of stability. The stochastic dynamics generates fluctuations, which could push the network away from the LFS, whereas a degree distribution with a large variance would cause a dispersion in the mean firing rate across neurons. These effects are characterized by performing the full simulations without stochasticity to determine the effect of firing rate dispersion,

$$\nu_{i,t+1} = \frac{1}{1 + \exp\left(h_0 - \frac{J}{Np_c} \sum_j w_{ij}\nu_{j,t}\right)} \tag{9}$$

and the stochastic version in Equation (1) to determine the effect of fluctuations.

Second, in the latter case, the state (LFS vs. HFS) reached is not deterministic, because a network can have a firing rate that fluctuates around the LFS or veers off to the HFS due to a somewhat larger fluctuation. We therefore performed the simulation multiple times and recorded how often (on what fraction of the trials) the network ended up at the HFS state as a function of the coupling constant. In this case we defined J_c to be the value of the coupling constant at which 50% of the states converged to the HFS within 400 time steps. The initial condition of the network was obtained by making a random set of neurons active in such a way that on average it had the same number of active neurons as expected based on the firing rate in the mean-field limit.

Third, when fluctuations stay in the basin of attraction (BOA) of the LFS, the network will not diverge, which means that the above fraction is an indirect measure of the BOA. We also determined a more direct measure by starting networks from different initial conditions, each with a different number of active neurons, and determining which fraction of trials goes to the HFS within 400 time steps. These initial states are characterized by the effective number N_{eff} of active neurons as is explained in the Results section and represented in Equation 11.

NETWORK SENSITIVITY

The sensitivity to a perturbation in experiment is tested in the model by activating a few selected neurons for a fixed duration. The stimulation was characterized by the number n_p of neurons stimulated (typically $n_p = 8$), the number of time bins, T_{stim}, the stimulation lasted (typically $T_{\text{stim}} = 6$) and the mean out-degree of the stimulated neurons represented by N_{eff}. For a fair comparison between different networks we randomly picked the stimulated neurons from the network and repeated the stimulation for 50 different realizations of the network. In order to estimate the effect of out-degree on the detection of the stimulation, we also ordered neurons based on their out-degree, with the highest out-degree first. This ordered set was divided into ten groups of equal size. We then randomly selected the stimulated neurons from a specific group and compared how the network response depended on which group was being stimulated.

ROC ANALYSIS

The ROC is obtained by picking a threshold and determining how often a firing rate response from the unstimulated network exceeds this threshold: the fraction of false positives. In addition, it is determined how often the firing rate of the stimulated network exceeds this threshold, this is the fraction of true positives. The ROC curve is traced out by plotting the true positives vs. the false positives for each possible threshold. When the distributions are exactly the same, the number of true positives equals the number of false positives, hence the ROC is the diagonal with an area under the curve (AUC) of 0.5. The deviation of the ROC curves from the diagonal, or equivalently deviation of the AUC from 0.5, is a measure for how different the distributions are and maps for Gaussian distributions on to d′, which is the difference in means of the distributions divided by the standard deviation (Kingdom and Prins, 2010). This also means that one can determine how many trials are needed to detect, given a particular ROC value, a difference between stimulation trials and unstimulated responses. The errors in the ROC curve and AUC value were determined by resampling of the simulated trials. Typically $N_r = 2000$ resamplings were used.

FUZZY CLUSTERING AND PERCEPTRON ANALYSIS

Fuzzy c-means (FCM) was used to cluster data points, such as a vector of network firing rates in consecutive time bins, or the motif distribution for a particular realization of a network, into groups with similar properties. FCM can be understood by first considering K-means clustering. In a K-means clustering, a number of clusters is chosen and the objects to be clustered are assigned on a random basis to each of the potential clusters (Duda et al., 2001). The name of the algorithm derives from

the convention that the number of clusters is denoted by K. Using these assignments, the mean of each cluster is found. Then, using these means, objects are re-assigned to each cluster based on which cluster center they are closest to. This process repeats until the cluster centers have converged onto stable values or a maximum iteration count is reached. This type of clustering minimizes the sum of the squared distances of the clustered objects from their cluster means. FCM functions in the same way, but rather than belonging to any particular cluster, each object i is assigned a set of normalized probabilities u_{ij} of belonging to cluster j (Bezdek, 1981). This is equivalent to minimizing a non-linear objective function of the distances of the objects from the cluster centers, characterized by the "fuzzifier" parameter, which is set to two. After the algorithm converges each data point is assigned to the cluster to which it is most likely to belong (maximizing the u_{ij} with respect to the cluster index j). A more complete description is given in (Fellous et al., 2004).

The perceptron algorithm is a method to classify responses x of the network (Duda et al., 2001). Here the vector $x = (r_t, r_{t+1})$ represents either a point in the firing rate return map or it represents the binary activity for each neuron during a particular time bin. The algorithm tries to find a weight vector w such that the sign w^Tx is positive when x belongs to group 1 (stimulated network) and negative when it belongs to group 2 (unstimulated).

ANALYSING MOTIF COUNT DISTRIBUTIONS

To investigate whether we could use motif statistics (restricting ourselves to 3-node motifs) of smaller parts of the complete network to distinguish between networks with different degree correlations, we generated $N_r = 1000$ realizations of each network type: correlated, anti-correlated and uncorrelated. We used smaller networks, $N = 200$, because these networks are adequate to represent sub-network statistics of size N_{sub} up to 200. We used standard parameters, $p_c = 0.05$, now yielding $\mu = 10$ and $\sigma_y = \mu/3 = 3.33$ and $\sigma_x = 0.3\sigma_y = 1.0$ for the smaller network. From each realization we sample sub-networks of N_{sub} from 4 to 24 in steps of 4 and from 30 to 200 in steps of 10. For each (sub) network we count the number of 3-node motifs using the explicit formulas given in Table III of Itzkovitz et al. (2003). Each motif is labeled by a number according to the convention also found in Itzkovitz et al. (2003). The counts in an ER network vary with powers of the expected number of edges per node k and network size N, $\lambda N^3 (k/N)^e$, where λ is a factor representing the symmetry of the pattern [see Table III in Itzkovitz et al. (2003)] and e is the number of edges in the pattern, which defines the complexity of the motif.

As a first step in the analysis we determined the mean and standard deviation of the motif count across the N_r realizations. To reduce the size of statistical fluctuations we also pooled motif counts by averaging them across N_{av} realizations. We either split the original N_r realizations into N_r/N_{av} groups, yielding a reduced number of data points or we randomly sampled with replacement $N_r N_{av}$ samples from the original N_r samples to keep the same number of pooled motif counts. The count distribution was often not Gaussian, which meant we could not use the t-test for the difference in mean count over the standard deviation. Hence, we utilized a ROC analysis. In order to obtain error estimates we created $N_b = 20$ different sets of $N_r = 500$ realizations, each of which were obtained by randomly sampling with replacement from amongst the $N_r = 1000$ original realizations.

We also wanted to determine whether incorporating counts of pairs of motifs would improve the ability to distinguish between networks with different degree correlations. We considered each realization, drawn from one or the other group of networks, as a two-dimensional data point and used FCM to find two clusters. FCM outputs the confidence (or probability) that the data point belongs to cluster 1. This value can be used as part of an ROC procedure. For a given threshold, the true positive corresponds the fraction of data points belonging to group 1 for which the confidence exceeds the threshold, whereas the false positive corresponds to the fraction exceeding threshold that belongs to the second group. We applied this procedure for each possible pair of motifs and for each sub-network size.

RESULTS

ANTI-CORRELATED NETWORKS ARE MOST STABLE IN THE ZERO-NOISE CASE

The mean-field limit, corresponding to an infinite network, is studied by considering the dynamics of a network where each neuron has the same firing rate, each neuron has the same number of synaptic inputs, i.e., in-degree, and there is no stochasticity. In this case the dynamical equations reduce to a self-consistent equation for the average firing rate ν (Equation 6 in Methods), which is solved according to the fixed point method. There are typically two stable solutions, one corresponding the HFS, in which the neuron is constantly firing (firing probability $\nu = 1$ or close to one) and one corresponding to the LFS at much lower rates, together with one unstable solution in between (**Figure 2A**). For high enough coupling constants only the HFS solution remains. We studied this by starting from an initial value of ν_t near zero and then iterating Equation 5 until convergence, if there is a LFS, it will converge to the LFS and if there is no LFS it will converge to the HFS, resulting in a sudden jump in firing rate as a function of J (**Figure 2B**). The coupling strength for which this jump occurs is denoted by J_c and depends on the baseline firing rate r_0 (defined in Equation 8, **Figures 2B,C**). The higher r_0 the less stable the network is. The firing rate of the LFS for J values just before it becomes unstable, referred to as r_c, is the maximum firing rate that the network can sustain, which varies approximately linear with the baseline firing rate (**Figure 2D**).

The effect of network size is studied by iterating Equation 9 for a vector of firing rate values, which ignores the effects of noise that are present in the full equations, Equation 1. In these finite size systems, the LFS is less stable, as reflected in the J_c values that are much below the mean-field limit (**Figure 2E**). There also is a difference between networks depending on their degree correlations, with the anti-correlated network being more stable than the ER network, correlated and uncorrelated networks. These differences become more pronounced for larger networks (**Figure 2E**). The difference also depends on the baseline firing rate, with the anti-correlated network again being the most stable (**Figure 2F**).

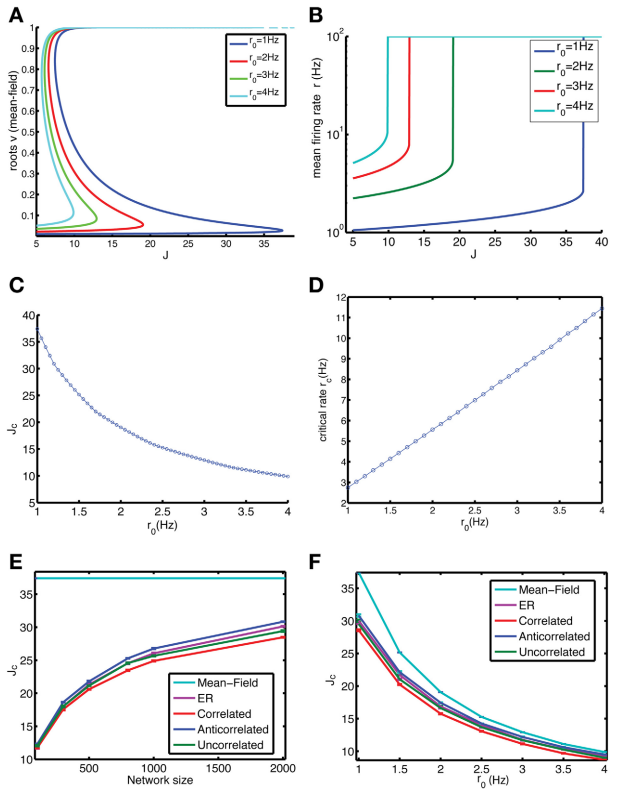


FIGURE 2 | Anti-correlations in the degree distribution improve the stability of the low firing rate state (LFS). We compared the stability of finite-size networks with different degree correlation structure by iterating Equation 9 (which is Equation 1 without taking into account stochastic spiking). **(A)** The mean-field solution, corresponding to an infinite-size network, is simulated by assuming that the firing rate of each neuron is equal, yielding Equation 6, of which all roots are shown in the graph. **(B)** Mean firing rate r vs. coupling constant J in the mean-field limit for different values for the baseline firing rate r_0. When the LFS loses stability, the only remaining solution is the HFS. As a result the plotted firing rate suddenly jumps to the maximum possible rate of 100 Hz (corresponding to 1 spike per bin). **(C)** The range of stable coupling constants, which are between 0 and J_c, decreases with increasing baseline firing rate. **(D)** The firing rate r_c of the LFS just before it turns unstable increases linearly with r_0. **(E)** The stability of the LFS depends on system size and approaches the mean-field limit (cyan) gradually as network size N increases (baseline rate $r_0 = 1$ Hz). The anti-correlated network (blue) is always more stable than the ER (purple), correlated (red), and uncorrelated (green) networks. **(F)** The difference between the mean field J_c and that of the finite-size networks decreases with baseline firing rate (network size $N = 2000$).

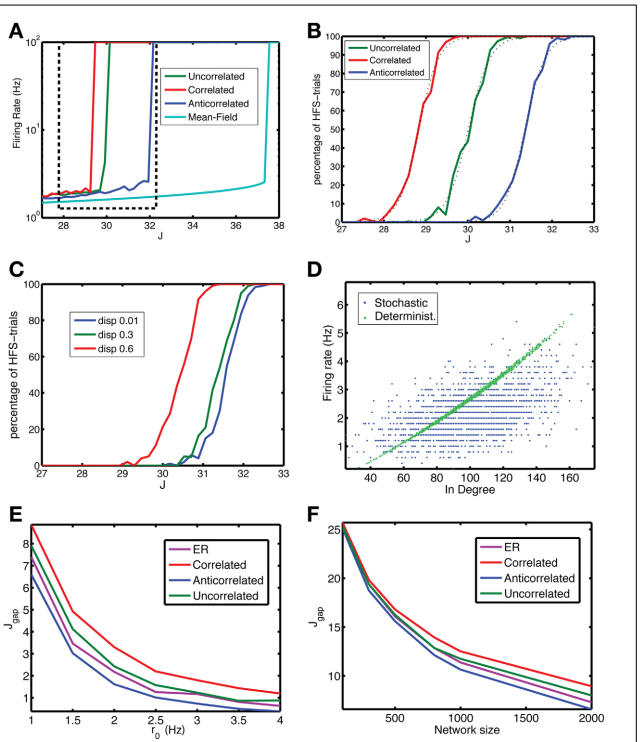


FIGURE 3 | The anti-correlated network is more stable against fluctuations. (A) The firing rate vs. coupling strength for the mean-field solution (cyan) and networks with uncorrelated (green), correlated (red) or anti-correlated (blue) degree distributions ($r_0 = 1$ Hz, $N = 2000$). The anti-correlated degree distribution leads to the most stable network. The dashed box approximately indicates the interval of coupling strengths highlighted in panels **(B)** and **(C)**. **(B)** Despite the existence of a stable LFS for a particular coupling strength, fluctuations in network activity may perturb the network away from it and the network ends up in the co-existing stable HFS state. The fraction of states that end up in the HFS state is close to zero far below J_c and increases to unity above J_c. The LFS state is more stable for the anti-correlated (blue) network than for the uncorrelated network (green), which in turn is more stable than the correlated network (red). The dashed lines are fits to the sigmoidal function in Equation 10. **(C)** The stability depends on the strength of the correlation. When the width (dispersion) corresponding to the small axis in the bivariate Gaussian degree distribution is increased, which means lower correlation, the stability is reduced. Data are for an anti-correlated network. **(D)** A neuron's firing rate is correlated with its in-degree, but the degree of correlation is reduced to 0.519 (0.014) due to jitter in this relation for Equation 1 (blue dots) from 0.997 (0.002) for Equation 9 (green dots). Data for anti-correlated network, $J = 30.96$. **(E,F)** The degree of stability can be qualified by J_{gap}, the distance of the J_c for the finite-size network from that for the mean-field network, shorter distances meaning more stable networks. J_{gap} decreases with the **(E)** baseline firing rate r_0 and with **(F)** network size. In both panels the anti-correlated network (blue line) corresponds to the lowest curve indicating higher stability compared to ER (purple), uncorrelated (green) and correlated (red), an advantage that increases with network size. The network had $N = 2000$ neurons, for each coupling strength $N_t = 100$ simulations were performed, with a length of 500 time steps, of which the first 100 were discarded as a transient.

ANTI-CORRELATED NETWORKS ARE MORE STABLE AGAINST FLUCTUATIONS

The dynamics of binary networks is stochastic because on each time step the expected firing rate is translated into a binary value. Hence, the firing rate, either averaged across network neurons during one time bin, or of one neuron averaged over a few time bins, will fluctuate. These fluctuations will alter the stability because these fluctuations could drive the network out of the BOA of the LFS toward that of the HFS state. The firing rate in the LFS state vs. coupling constant curve for the stochastic network (**Figure 3A**) looks similar to that for the zero noise case (not shown), but the fraction of trials on which the HFS state is reached displays a sigmoidal behavior (**Figure 3B**): with some networks switching to the HFS state close to, but below the critical coupling constant J_c, whereas most of the networks go to HFS for

coupling constants above J_c. In between there is a transition point where an equal number of networks go to the LFS and HFS state. The anti-correlated state is more stable, because this transition point lies to the right of the transition point for the other networks (**Figure 3B**). We have fitted the probability to the following expression,

$$\mathrm{p(J)} = 1/(1 + \exp(-(\mathrm{J} - \mathrm{J_h})/\sigma_\mathrm{J}) \quad (10)$$

where J_h is the transition point and σ_J represents the sharpness of the transition. The transition for correlated and anti-correlated networks is sharper than for uncorrelated networks, as indicated by the $\sigma_J = 0.424$ and 0.420, compared to 0.391, respectively, with R^2 values (fraction of explained variance) all approximately 0.999.

The in- and out-degrees are drawn from a bivariate Gaussian, which has a long axis, in the direction of the correlation, and a short axis perpendicular to that direction (Equation 4, Methods). Increasing the standard deviation along the short axis, termed dispersion, reduces the degree of correlation. In addition, it makes the anti-correlated network less stable (**Figure 3C**).

The stability properties of the finite-size networks are different from that in the mean-field limit (**Figure 2**), because the firing rate of a neuron depends on the number of inputs (in-degree), which varies across neurons in the network (**Figure 3D**, green dots). The correlation between the neuron's firing rate and its in-degree is almost perfect for the non-stochastic network, with squared Pearson correlation $R^2 = 0.997$ (0.002), but becomes jittered due to the stochastic spiking resulting in a squared Pearson correlation of 0.519 (0.014) (**Figure 3D**, blue dots).

The mean-field limit represents the highest level of stability, because both finite-size and noise effects reduce it. The reduction in stability can be captured into J_{gap}, which is the mean-field critical coupling minus the critical coupling value for the noisy, finite-size network. The smaller J_{gap} is, the more stable the system is. The gap decreases both with baseline firing rate (**Figure 3E**) and network size (**Figure 3F**). As the network size increases, the comparative stability advantage of anti-correlated networks increases.

The stability against fluctuations can be analyzed differently. Non-linear dynamical systems are characterized in terms of the basin of attraction (BOA). Consider a simple one-dimensional system with two stable fixed points (and an unstable one in between) (Strogatz, 1994). Depending on the initial condition of the one state variable, the system will converge to one or the other fixed point. The catchment area of the first fixed point, the range of initial conditions that converge toward it, is the BOA. There is a well-defined boundary between the two BOAs. Our goal is to characterize this boundary between LFS and HFS for the binary networks studied here, which is complicated because of the high dimensionality of the state space and the stochasticity, which means that a given initial condition near the boundary could converge to a LFS or HFS depending on the role of the dice. The first issue means we have to find a more effective and compact description of the initial state. Our initial choice was to use the number N_a of active neurons in the initial condition. However, when the N_a highest out-degree neurons are active, the network is more likely to converge to the HFS than when the N_a lowest out-degree neurons are active, even though the initial state has an equal number of active neurons. Hence, we used the so called effective number of active neurons, where each neuron's contribution is weighted by their out-degree:

$$N_{\mathrm{eff}} = N \frac{\sum_{i \in \mathrm{active}} d_i^{\mathrm{out}}}{\sum_i d_i^{\mathrm{out}}} \quad (11)$$

We started the simulations from a random initial state, characterized by a specific number of active neurons (range: between 0 and 200), and repeated this procedure enough times ($N_r = 4000$) to ensure sufficient coverage across the relevant N_{eff} values. For each N_{eff} value so sampled, a fraction converged to the LFS and the remainder went to the HFS state (**Figure 4A**). For small N_{eff} most states converge to LFS and for N_{eff} larger than a transition value $N_{\mathrm{eff},90}$ most converge to the HFS (**Figure 4B**). We choose as transition value the lowest N_{eff} value for which 90% or more states went to the HFS. The transition value $N_{\mathrm{eff},90}$ decreases with coupling strength J (**Figures 4C,D**) until its value comes close to the number of active neurons represented by the average firing rate of the mean-field network, at which point stability is lost. This is because the BOA of the LFS shrinks

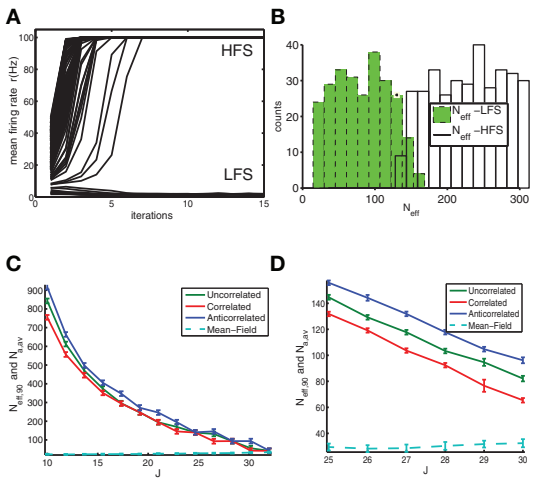


FIGURE 4 | The basin of attraction of the LFS is larger for anti-correlated networks indicating enhanced stability against fluctuations. (A) Simulations were started from initial states with a different number N_a of active neurons, which is translated into an N_{eff} value (see text) to allow for a fair comparison of initial conditions. We show the firing rate as a function of time (in units of iterations). For low N_a the anti-correlated network converged to the LFS, whereas for high N_a runs it converged to the HFS. **(B)** This was reflected in the histogram where green filled bars indicate the number of states with a particular N_{eff} that converged to the LFS and the open bars indicate the number of states that converged to the HFS. Data for anti-correlated network with $J = 25$. **(C)** $N_{\mathrm{eff},90}$ as a function of coupling constant J for uncorrelated (green), correlated (red) and anti-correlated (blue) networks together with the number $N_{a,av}$ of active neurons corresponding to the firing rate of the mean-field solution (cyan dashed line) as a reference. **(D)** Close-up of panel **(C)**. The data were obtained from a network of $N = 2000$ neurons, with a baseline rate of 1 Hz. For each coupling strength, and, each network type we used $N_r = 1000$ initial conditions and averaged across 4 realizations of the network.

to zero and most initial conditions go to the HFS. The anti-correlated network is more stable because it can sustain initial states with a higher number of active neurons and still return to the LFS as compared to other networks. Furthermore, for the anti-correlated networks the BOA is finite for larger values of the coupling constant compared to other networks. Overall, when a sufficient number of neurons are active in the initial condition, both the effective and unnormalized number of active neurons yield similar results for the size of the BOA (not shown).

THE SENSITIVITY OF THE NETWORK CAN BE CHARACTERIZED USING ROC ANALYSIS

During spontaneous (unstimulated) activity in the network, the firing rate will fluctuate from time bin to time bin, which can be considered random draws from a distribution. When the network is stimulated, the average firing rate will be altered, trivially because of the activated neurons, but non-trivially through the downstream effect of this stimulation on the other neurons. The stimulation is characterized by the number of cells n_p stimulated (and their out-degree, see below) and the duration of the stimulation T_{stim}. We used $n_p = 8$ and $T_{\text{stim}} = 6$. Its effect on the network can be detected when there is a systematic difference between the network states, quantified, for instance, in terms of the mean firing rate of the overall activity. An ROC analysis quantifies how different the distribution of firing rate is between the stimulated and unstimulated networks and how easy it is to detect this difference and can thus be compared to measured behavioral responses. In all of the following analyses we exclude the stimulated cells themselves. One reason is that the decision process would be based on downstream neurons, hence we should detect the difference in the downstream population.

The histogram of the simulated firing rates was shifted relative to that of the unstimulated network (**Figure 5A**). In **Figure 5B**, the ROC curve corresponding to the empirical distributions in panel a is shown. The evaluation of the corresponding AUC, as a function of time is shown in panel c. Before the stimulation at $t = 10$, the statistics of both networks are the same, yielding an AUC of close to 0.5, whereas after stimulation the AUC rises to the 0.75. The ability to detect a stimulation increases with the strength of the coupling constant (**Figure 5D**). This can be simply understood because a higher J increases the impact of presynaptic activity on the neuron's firing rate, hence it also increases the effect of stimulation. There is no difference in sensitivity due to the correlation structure of the network as long as neurons with similar out-degrees are stimulated, because the sensitivity only depends on the out-degree. The AUC also increases with baseline firing rate of the network (**Figure 5E**), which indicates that network state changes, such as those occurring during arousal or with attention in which the overall firing rate increases, could improve task performance. Also for this behavior there was no difference between networks with the different type of degree correlations. The derivative of the mean firing rate r with respect to J increases with baseline firing rate r_0, suggesting that the effect of a stimulation on the network firing rate increases with r_0, which is indeed borne out by the simulation results in **Figure 5E**.

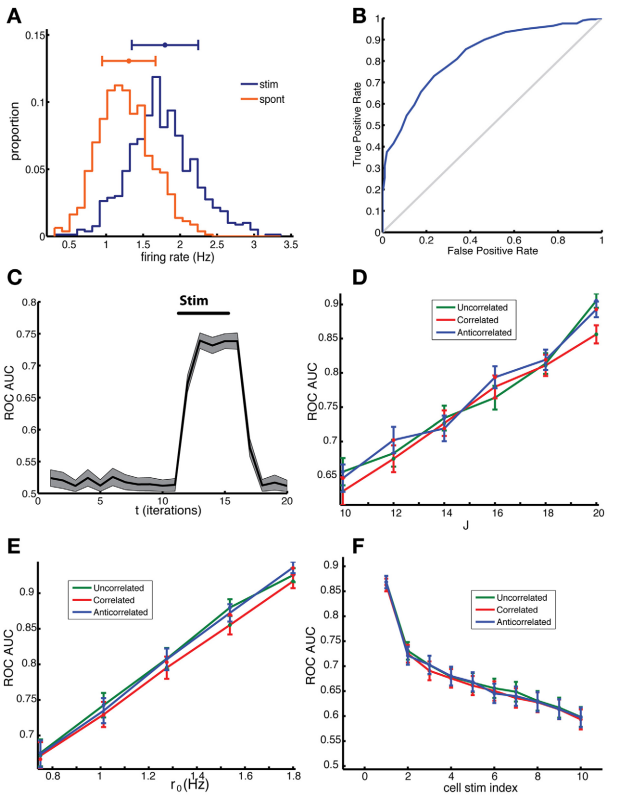


FIGURE 5 | Network sensitivity, when evaluated using an ROC analysis, depends only on the mean out-degree of the stimulated neurons and not on the degree correlations. (A) Distribution of firing rate across cells in a 10 ms time bin for spontaneous activity (red) and for the stimulated network (blue), in which 8 random cells were stimulated. Note that the stimulated cells were not included in this ROC analysis and we used the binary responses $x_i(t)$ to determine the firing rate. **(B)** The corresponding ROC curve (blue) quantifies the difference between the distributions, relative to the diagonal (gray), which represents distributions that cannot be distinguished. **(C)** The area under the curve (AUC) for the ROC curves calculated for different time bins. The AUC before stimulation was close to 0.5 because the distributions were the same apart from fluctuations due to sampling. After the stimulation, which started at $t = 10$ and ended at $t = 15$, the AUC rose to around 0.75. **(D)** The AUC increases with increasing coupling constant and **(E)** with increasing baseline firing rate. **(F)** The AUC depended on the mean out-degree of the stimulated neurons. Neurons were divided into ten groups according to their out-degree, with the first group having the highest out-degree. The group index is indicated on the x-axis. The results in **(D–F)** were not significantly different for correlated (red), anti-correlated (blue) or uncorrelated (green) networks, t-test, $p = 0.4479, 0.6279, 0.7421$, respectively. The network was comprised of $N = 2000$ neurons, of which $n_p = 8$ neurons were stimulated for the duration of $T_{\text{stim}} = 6$ time units starting on the 10^{th} bin. In panel **(A–C)** results for an uncorrelated network are shown. Parameters: **(A–C,F)** $J = 18$, $r_0 = 1$; for **(D)** $r_0 = 1$ and **(E)** $J = 20$.

Stimulus detection depended on which cells were stimulated, with their average out-degree being the most important factor. We chose n_p neurons to be stimulated randomly from 10 different groups with different mean out-degree, which were generated as follows. First all neurons were ordered according to their out-degree, with the highest out-degree neurons coming first, and then divided into ten equally-sized groups, labeled

1 to 10. Multiple stimulation trials were done with n_p neurons picked from one of the groups from which the group AUC was determined. The AUC for each group was then plotted as a function of the group label (**Figure 5F**). The AUC values for the first group were much higher than for the next groups demonstrating clearly that the group with the highest mean out-degree also had the highest AUC.

Taken together, these simulations show that the correlations in the degree distribution do not directly affect network sensitivity to stimulation. Rather, this sensitivity is determined by the out-degree of the stimulated neurons. Networks can display a higher sensitivity if they have a larger variability in the out-degree distribution and those cells with the highest out-degree are being stimulated. ER networks have a low variance in the out-degree, and will therefore have a reduced sensitivity compared to the networks here, compare the AUC of the first group to that of the fifth group which represents neurons with an out-degree closest to the mean.

The fluctuations in firing rate during spontaneous activity are expected to have different temporal correlations compared to those in the stimulated network, as an increase due to an external stimulation is going to persist across the time bins during which the stimulation takes place. Hence, the detection rate could improve by taking into account (spatio) temporal correlations. The first step is to consider the correlation in network firing rate r between two consecutive time bins. When r_{t+1} is plotted vs. r_t a return map would be obtained. However, because the firing rate values are restricted to $x/(N\Delta t)$, where x is an integer between 0 and N, and N the network size, the return map would have a non-informative appearance. Hence, we made a density representation, by replacing each sample by a two-dimensional Gaussian (kernel density estimate) with a standard deviation (bandwidth) optimally estimated from data following the Silverman's rule of thumb (Silverman, 1986). The hot spot in the return map density obtained for stimulated networks (**Figure 6B**, plus sign) is shifted along the diagonal in the positive direction (i.e., higher rates) in comparison to the return map for spontaneous activity (**Figure 6A**).

We determined whether such a two-dimensional representation would improve the detection rate. An equal number of samples from spontaneous activity and from stimulated activity were provided to a fuzzy clustering method (FCM) routine in Matlab in order to find two clusters (Fellous et al., 2004). The FCM returns for each data point i the probability u_{ij} that it belongs to cluster j. As the sum of probabilities needs to be unity, for two possible clusters we only need to consider u_{i1}. We thus obtain a distribution of u_{i1} values for data points from the spontaneous activity and a distribution for data points from the stimulated network. The difference between these distributions is a measure for how well stimulation can be detected and can thus be subjected to a ROC analysis. In this ROC analysis the u_{i1} values are treated in exactly the same way as the firing rates used to obtain the results in **Figure 5**. The resulting AUC values were 5% higher than based on the distribution of firing rates in one bin (t-test, $p = 0$).

In the firing-rate based detection procedure, each neuron (except the directly stimulated ones) carries equal weight. The cells that are not directly connected to the stimulated neurons would display firing rate fluctuations that are unrelated to the stimulation, hence act as noise that reduces probability of detection. The signal to noise of the firing rate fluctuations could be improved by weighing those neurons less. To explore this hypothesis we applied a perceptron procedure (see Methods) to learn the optimal weights for classifying the network state vectors (Duda et al., 2001). An equal number of network states for spontaneous activity and for stimulated networks were supplied to the perceptron routine together with the corresponding class labels. The output was a weight for each neuron. As before the activity of the directly stimulated neurons was not included in this analysis. To determine what features contributed to the weight we plotted the weight vs. feature value in a scatter plot and calculated the corresponding Pearson correlation. There was a small, but significant correlation between the weight and the in-degree

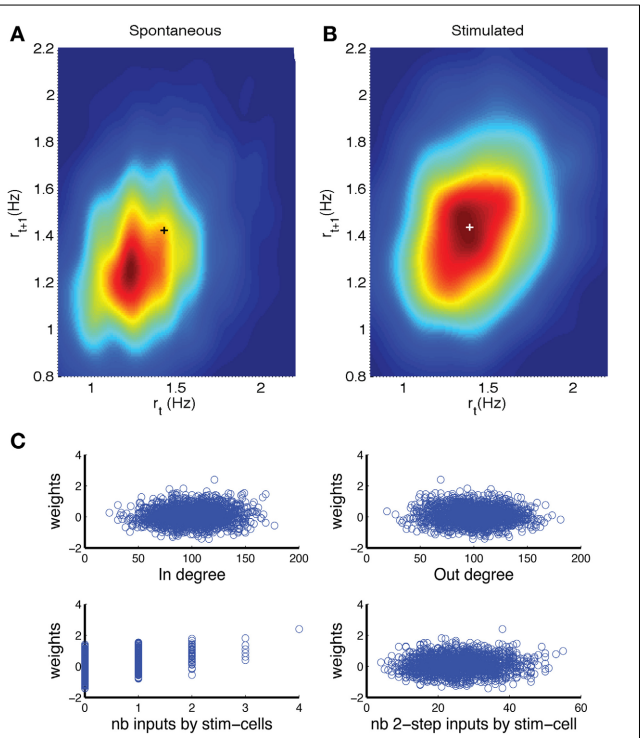


FIGURE 6 | Detection can be improved by including past activity and weighting neurons depending on how many inputs they receive from directly stimulated neurons. (A,B) Density representation of the firing rate return map, wherein the probability of obtaining consecutive rate values (r_t, r_{t+1}) is represented by a color scale, with red indicating the highest probability and blue indicating a near zero probability. The results are shown for **(A)** spontaneous activity and **(B)** stimulated activity. The plusses indicate the location of the peak in panel **(B)**. **(C)** Analysis of factors that contribute to a neuron's weight in detection decision that is outputted by the perceptron procedure. There was significant but small correlation between weight and (top left) in-degree or (top right) out-degree. There was a correlation between the weight and the number of direct inputs from stimulated neurons (bottom left), but only a weak correlation with the number of inputs from cells that received direct input from the stimulated cells (bottom right). The network was comprised of $N = 2000$ neurons, coupling constant $J = 18$, baseline firing rate $r_0 = 1$ Hz. In the stimulated network, $n_p = 8$ neurons were stimulated for a duration of $T_{\text{stim}} = 6$ time bins.

(**Figure 6C**, top left, correlation 0.073 ± 0.03, $p = 0.0011$) and with the out-degree (**Figure 6C**, top right, correlation -0.052 ± 0.027, $p = 0.018$). There was a strong correlation between the weight and the number of direct inputs the neuron received from stimulated neurons (**Figure 6C**, bottom left, correlation 0.410 ± 0.15, $p = 0.0$). The number of indirect inputs from stimulated neurons was less relevant (**Figure 6C**, bottom right, correlation 0.072 ± 0.032, $p = 0.012$). We calculated this by determining the number of inputs from cells that received direct inputs.

Taken together, these analyses show that our estimates for the detection of stimulation based on overall firing rate are underestimates and can be improved by taking into account network history and by selecting which neurons to listen to. The latter of which may be achieved through synaptic plasticity and the appropriate learning rules.

DETECTING ANTI-CORRELATION IN THE DEGREE DISTRIBUTION WITH LIMITED DATA

The results here establish that anti-correlation between in- and out-degrees results in more stable, but equally sensitive networks, compared to networks without correlations between in- and out-degree, or positive correlations between them. Hence, learning to detect a stimulation could proceed by altering the correlation between in- and out-degree. To demonstrate such a learning effect, the in- and out-degree of a number of neurons needs to be sampled. Classical tracing techniques are not appropriate because they involve the connections to or from multiple nearby neurons (Lanciego and Wouterlood, 2011). For instance, when the retrograde tracer horseradish peroxidase is injected, it is absorbed by multiple axon terminals and transported to their respective cell bodies. These axon terminals do not necessarily synapse on one and the same neuron near the injection site. Hence, the data cannot be used to determine the in-degree of a neuron near the injection site.

New viral-based techniques could help, because they work by infecting a few cells in the neighborhood where the virus is injected (Wickersham et al., 2007; Osakada et al., 2011). The virus will then retrogradely label the cells presynaptic to these cells by crossing one synapse and one synapse only. In the presynaptic cells the infection stops because the virus misses the proteins necessary to cross another synapse. The challenge with this method is to infect only one cell, with both an anterogradely and retrogradely crossing virus.

Currently, the gold standard is to simultaneously record multiple cells *in vitro* and assess connections by inducing action potentials in one neuron at a time and recording the post-synaptic responses in the other cells. The current record is 12 cells recorded simultaneously (Song et al., 2005; Perin et al., 2011). This means that the anti-correlation in the degree distribution will have to be assessed indirectly, by sampling from sub-networks.

Motifs represent patterns in the connectivity that occur more often than expected if the connections were made random (Milo et al., 2002). For instance, consider a network for which the average probability of a connection is p. For two neurons, if these connections are made randomly, the probability of having no connection is $(1-p)^2$, for having one connection $2p(1-p)$ and for having a bidirectional connection p^2. When it is found that bidirectional connections occur significantly more than the expected p^2 then there is additional, non-random, structure in the network (Song et al., 2005). Motifs most often refer to triplets of neurons and the patterns of connectivity between them that occur more often than expected in a random network (Milo et al., 2002). A motif distribution is the number of times each motif occurs in a network and a motif is considered present when it occurs more often than in a control network. Motif distributions are affected by many network properties such as, for instance, the degree distribution. The networks studied here, even when uncorrelated, have a different degree distribution than the ER network, which means that ER random networks are not a good control. Hence, we have to numerically generate the control distributions rather than having access to the analytical expression for the expected rate of each motif. In addition, in experimental settings we do not have access to the whole network from which to determine the motif distribution, we have to do with sub-networks. These sub-networks do not come from the same network, rather they come from networks sampled from an ensemble of networks with similar properties. To obtain estimates for how to observe evidence for anti-correlation in the degree distribution we need to deal with each of these issues.

The overall goal is to distinguish between pairs of networks with anti-correlated, uncorrelated and correlated degree distribution with the same marginal distribution for in- and out-degree.

We considered 13 different motifs that consisted of three connected neurons and gave each motif a numerical label as shown in **Figure 7A**. We determined the number of motifs in each realization of a network with correlated, anti-correlated or uncorrelated degree distribution and took the average. This was done for the full network (here reduced to $N = 200$) as well as for sub-networks (size N_{sub}). The complexity of a motif corresponds to the number of edges in the pattern, ranging from 2 to 6, which determines how often it is counted in a network. We normalized the counts such that they took values on the order of unity in order to better compare them across motifs. The mean count as a function of N_{sub} converged to a constant for network sizes between 50–100 neurons (**Figure 7B**), with more complex motifs requiring larger N_{sub}. The width of the count distribution, quantified as the standard deviation, decreased with N_{sub} as the-3/2 power (**Figure 7C**). Hence, for large enough networks the differences in mean counts across network type can be detected with certainty. This power law behavior is consistent with the results for a Binomial process with probability p and on the order of $n \sim N_{\text{sub}}^3$ trials, for which the mean is np and the variance is $np(1-p)$. In that case the normalized mean is p, and its variance $(1\text{-}p)/n$ (see also Equations 2, 3), leading to a standard deviation varying as $n^{-1/2} = N_{\text{sub}}^{-3/2}$.

We are looking for motifs whose counts are different between the analyzed network types. In **Figure 7D** we show the count as a function of motif for the three network types, with the highest difference occurring for the complex motifs 110 and 238. However, the counts for these motifs, which have the largest number of edges, are characterized by a large standard deviation. When we plot the motifs in an order based on the ratio of count difference over standard deviation, motif 98 comes up as winner instead (**Figure 7E**). **Figure 7D** shows that there are fewer motif 98 in

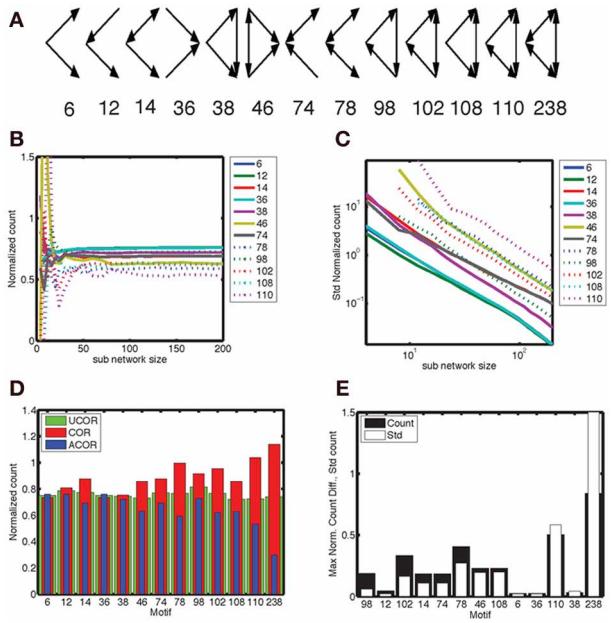


FIGURE 7 | Motif 98 is the most sensitive to degree correlations. (A) There are 13 motifs that involve 3 connected nodes. Below the graphical representation we plot the numbering used here, which follows Itzkovitz et al. (2003). The expected number of motifs depends on network size, hence we normalize the count by $N^3(k/N)^e$, with N the number of nodes, k the expected number of edges per node and e the number of edges in the motif. In addition, we include a numerical factor representing the equivalent permutations [listed in Table 3 in Itzkovitz et al. (2003)]. **(B)** The normalized counts, averaged across a thousand realizations, converge to constant values for sub-networks larger than 50–100 nodes, with the precise value depending on the complexity of the motif involved. **(C)** The standard deviation of the normalized counts fall off as $N^{-3/2}$. We illustrate the results for the anti-correlated network, which are typical for the correlated and uncorrelated network also. In addition, we omitted motif 238 because it occurs at such a low probability that it makes the statistics noisy. **(D)** The normalized counts for each motif for the (red) correlated, (blue) anti-correlated and (green) uncorrelated networks. We used the counts for the full network, rather than sub-networks. Network size in panel **(D)** and **(E)** was $N = 200$. We used a bivariate Gaussian degree distribution with a mean number of nodes equal to 10, a standard deviation along the long axis of $\sigma_y = 3.33$ and along the short axis of $\sigma_x = 1.0$. **(E)** The maximum difference in mean count between all three possible comparisons (black bars), relative to the mean standard deviation of these counts across the three network types. The motifs are ordered on the count over standard deviation ratio, starting with the largest. According to this analysis motif 98 should be used to best distinguish between different network correlation structures.

anti-correlated networks compared to correlated networks. This can be understood intuitively by noting that in "ring" motif 98 each neuron has the same number of inputs as outputs, namely 1, which is more representative for correlated networks (**Figure 1D**, inset) than for anti-correlated networks (**Figure 1C**, inset). Furthermore, this means there is a lower probability of closing the ring, because in an anti-correlated network a neuron with many inputs has fewer outputs to get to the next neuron in the ring.

The count distributions are not Gaussian for small sub-networks. **Figure 8A** shows the count distribution for motif 98 for networks with $N = 200$. Each network gives rise to a symmetric appearing distribution, with the peak at a different location depending on the network type. The distribution for the anti-correlated and correlated network were farthest apart, with that of the uncorrelated distribution situated in the middle. For $N_{\text{sub}} = 30$ (**Figure 8B**), the corresponding distributions fell on top of each other and are asymmetric because the counts are always positive. To compare the distributions we therefore performed an ROC analysis. As expected based on the reduced overlap between distributions, the AUC increases with sub-network size, and motif 98 comes out on top with the highest AUC (**Figure 8C**). Furthermore, given the lower overlap between the anti-correlated and correlated distribution (**Figure 8A**), the AUC values for the comparison between anti-correlated and correlated network is higher (**Figure 8C**) than for either the comparison between anti-correlated and uncorrelated (**Figure 8D**) or correlated with uncorrelated (not shown).

In experiments only relatively small networks can be mapped, up to 12 cells using paired recordings and a few tens to hundreds using population calcium imaging. For these numbers the degree correlations cannot be reliably distinguished based on a single measurement. We therefore pooled measurements to see if this improved discriminability for more experimentally accessible smaller sub-networks. This procedure (pooling motif counts across $N_{av} = 50$ network realizations) indeed reduced overlap between distributions (**Figure 8E**, compare to **Figure 8B**). The more motif counts were pooled, the higher the AUC was (**Figure 8F**). Furthermore, the value of unity, corresponding to perfect discriminability is reached for smaller sub-network sizes. For $N_{av} = 50$, $N_{\text{sub}} = 30$ networks are perfectly discriminable and the AUC transitions from values just above 0.7 to unity between $N_{av} = 30$ and 50 (**Figure 8F**). Taken together, sub-networks of a few tens of neurons could be used to test our hypothesis experimentally.

The question is whether this result can be improved by including counts for multiple different motifs (**Figure 9A**). Without pooling, motif 98 by itself outperforms any pair of motif counts, according to the AUC value (**Figure 9B**). To determine the AUC value for pairs of motif counts we used the FCM procedure as outlined in the methods section. When counts are pooled (**Figure 9C**), some motif pairs outperform motif 98 by a small margin. The pairs are highlighted in **Figure 9D**, and involve motif 98 itself. The more separated the cloud of points corresponding to different network types is, the better the FCM procedure classifies the networks, compare the plusses (correct discrimination) and dots (incorrect) in **Figure 9A**.

DISCUSSION

The overall firing rates in barrel cortex (de Kock et al., 2007; Greenberg et al., 2008; Barth and Poulet, 2012) are much lower than might be expected based on the classic experiments in macaque visual cortex (Hubel and Wiesel, 1968). Neural activity is also variable, which can be characterized as across trial reliability, or in terms of the coefficient of variation and Fano factor of spontaneous activity (Shadlen and Newsome, 1998). These measures reveal that the activity is similar to that of a Poisson process, in which the occurrence of a spike in a time bin is uncorrelated with whether or not a spike occurred in previous bins. The mean

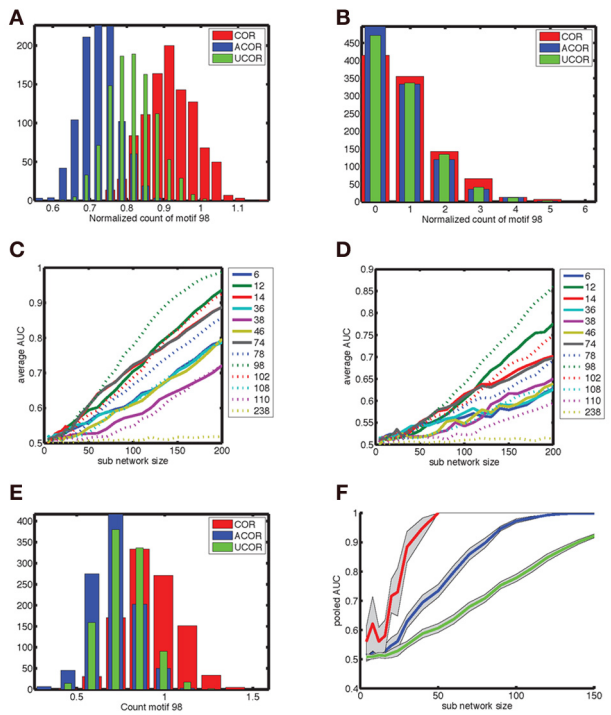


FIGURE 8 | Degree correlations can be distinguished by pooling fifty measurements of networks at least thirty neurons in size. (A) Motif count varies across network realizations, but degree correlations can be distinguished when the corresponding distribution show little overlap. We show the distribution of the normalized counts of motif 98 for (red) correlated, (blue) anti-correlated, and (green) uncorrelated networks with 200 neurons. **(B)** For smaller sub-networks ($N_{sub} = 30$), the distributions overlap. Furthermore, these distributions are not Gaussian as they are skewed because counts are always positive. Hence, a more general procedure, such as the ROC analysis needs to be used instead of looking at the differences in mean count relative to the standard deviation. **(C,D)** The area under the ROC curve (AUC) as a function of sub-network size N_{sub} for the comparison **(C)** between correlated and anti-correlated networks and **(D)** between anti-correlated and uncorrelated networks. Each motif is labeled with a line style and color as indicated in the legend. Motif 98 is most sensitive in both cases (as well as for the correlated vs. uncorrelated comparison that is not shown). It is more difficult to distinguish an anti-correlated network from an uncorrelated one than to distinguish it from a correlated network. The average AUC values were determined based on the AUC value for each of twenty different motif distributions of 500 network realizations, which were sampled randomly with replacement out of 1000 realizations. **(E)** The motif distribution for $N_{sub} = 30$ can be pooled across $N_{av} = 50$ network realizations in order to shrink the width of the distribution, so that the differences in mean counts become clearer [(compare to panel **(B)**]. **(F)** The AUC for larger N_{av} values reaches unity (distributions are perfectly distinguishable) for smaller sub-network sizes. We show (green) no pooling, (blue) pooling across $N_{av} = 5$ realizations and (red) pooling across $N_{av} = 50$ realizations. The AUC goes from 0.7 to 1.0 between $N_{sub} = 30$ and 50 when pooled across $N_{av} = 50$ realizations, indicating that networks of size 30 can be used to determine degree correlation structure.

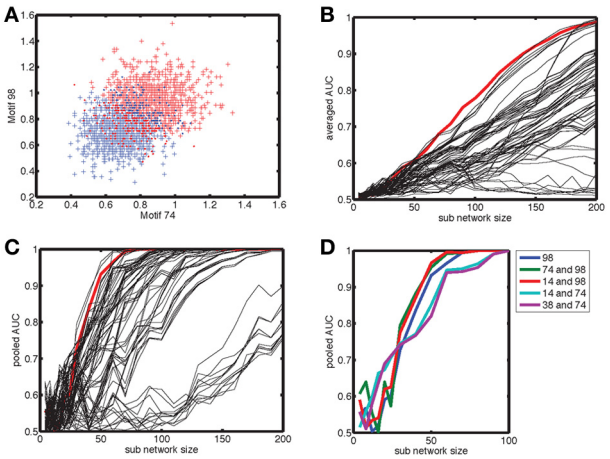


FIGURE 9 | The gain in discriminability by using the joint distribution of motif counts rather than the marginal distribution is limited. (A) The outcome of the FCM procedure to find two clusters. Red points indicate counts obtained from a correlated network and blue points are those for the anti-correlated network. The plusses indicate points correctly classified by FCM and the dots represent incorrectly classified network realizations. Each motif-pair ROC curve is obtained by applying the ROC analysis to the FCM generated probability of each network to belong to cluster one (see methods). **(B)** The resulting AUC as a function of sub-network size for motif 98 (red) is higher than for all possible pairs of motifs (black curves). The AUC shown is for comparing anti-correlated and correlated networks. **(C)** When the motif count is pooled across $N_{av} = 20$ realizations some two-motif curves exceed the single motif curve, which are shown separately in panel **(D)**. This suggests that for a specific size of the sub-network distinguishability can be improved by considering pairs of motif counts.

firing rate is maintained by the intrinsic excitability of neurons and their synaptic inputs, including recurrent excitation. High variability together with a low firing rate implies that the network dynamics should be stable against fluctuations in the mean activity in the sense that these fluctuations do not generate states with networks bursts in which all neurons in the network are active at the same time.

Experiments show that rodents can detect single-cell stimulation in barrel cortex, in which a single neuron is electrically stimulated to produce a high-frequency train of action potentials (Houweling and Brecht, 2008). This may mean that single-cell stimulation can cause an increase (or decrease) in the firing rate of the local network that is significantly different from that occurring during spontaneous activity. Taken together, this means that cortical networks with a low firing rate should at the same time be stable against fluctuations in firing rate and sensitive to weak stimulation. The overall goal of this paper was to a find a potential explanation for how the contrasting demands of sensitivity and stability can be realized. To achieve this we examined the dynamics of binary neural networks with correlation between the in- and out-degrees of neurons. In the following we summarize the main results with the aim of linking the detection performance of the network to experimentally obtained behavioral results, the mechanism by which sensitivity and stability can be achieved, and predicting the anatomical signatures of the hypothesized network. We also discuss the role of other biophysical factors, such as inhibition, not taken into account in the present study.

GENERATING THE NETWORK CONNECTIVITY UNDERLYING ENHANCED STABILITY

Our guiding hypothesis is that networks with an anti-correlation between the in- and out-degree of neurons are more stable and

equally sensitive as other networks with comparable marginal degree distributions. Network sensitivity is generated by neurons with a high out-degree, because these would amplify the effect of nanostimulation the most. This amplification would also destabilize the network, so these cells should not be activated during spontaneous activity. As the input to neurons is proportional to the mean firing rate and their in-degree, this can be achieved by making sure that high out-degree neurons have low in-degrees. To maintain the average degree, both in and out, there then also need to be neurons with a low out-degree and a high in-degree. We implemented this hypothesis as an anti-correlation in the in- and out-degree.

In the standard Erdos-Renyi networks, the relative variance in the degree distribution for large networks becomes too small to have out-degrees that are much larger than the mean degree, which is needed to reach the desired sensitivity. Hence, we needed to broaden the degree distribution artificially by using a truncated bivariate Gaussian distribution. Networks with this sampled degree distribution were generated via the configuration model (Newman, 2010). This configuration model generates networks with self-edges and multi-edges. Analytical calculations show that the probability for obtaining a network with one or more of these edges is close to one for the large mean degrees we consider (Blitzstein and Diaconis, 2006). Nevertheless, the number of these edges is low and their impact on the dynamics was limited.

There are a number of ways to address the multi and self-edge problem in a more principled approach that differ in computational efficiency and ease of implementation. First, one can use the configuration model procedure, but reject an invalid edge and find a valid replacement. This carries the risk that the algorithm stops when there are no valid edges available, which means that the whole procedure has to be restarted. Alternatively, as mentioned before, one can identify the invalid edges when the network construction has been completed and remove them or replace them by valid ones. See Blitzstein and Diaconis (2006) for a review. Second, one can find one graph that satisfied the degree distribution using the Havel-Hakim procedure (Viger and Latapy, 2005; Erdos et al., 2010; Chatterjee et al., 2011) and generate samples from the overall graph distribution by swapping links (Blitzstein and Diaconis, 2006). Swapping links refers to the procedure where randomly chosen existing links $i \rightarrow j$ and $k \rightarrow l$ are swapped into $i \rightarrow k$ and $j \rightarrow l$ when this yields a simple graph without self-edges and multi-edges. This requires careful calibration of the number of swaps and also introduces bias because these swaps do not change the number of triangles in the network (Roberts and Coolen, 2012). Third, a sequential method can be defined that produces all possible graphs, by randomly selecting amongst the allowed edges that keep the residual degree distribution graphical (Del Genio et al., 2010; Kim et al., 2012). A degree distribution is graphical when there exists a simple graph with that distribution, after each step the degree distribution is lowered to account for the connections realized, and this is referred to as the residual degree distribution. This method does not produce the graphs with the correct probability. Hence, averages based on these graphs have to be reweighted to take this into account. Furthermore, in our hands, an implementation of this method produces graphs with a correlation between the in- and/or out-degrees between different nodes, which is referred to as assortativity. This necessitates a number of link swaps to remove these correlations. Fourth, edges can be sampled according to a Boltzman function (Park and Newman, 2004), where the expectation value of the degree of a node is fixed through a Lagrange multiplier, for which the appropriate value has to be picked, which can be achieved, for instance, through a maximum likelihood approach or iterative rescaling (Chatterjee et al., 2011). Taken together, we opted to use the simplest method here, because these alternative methods for network generation were computationally more intensive and also suffered from aforementioned additional drawbacks, such as graphs that were not sampled according to a uniform probability (Del Genio et al., 2010) or other biases in the network statistics (Roberts and Coolen, 2012). Recently developed methods for generating networks with degree correlations, both in a single neuron as well as between pairs of neurons look very promising (Roberts and Coolen, 2012).

STABILITY IS ENHANCED WHEN THE IN- AND OUT-DEGREE ARE ANTI-CORRELATED

Our aim was to find stable networks, by which we mean that fluctuations do not cause a cascade of recurrent excitation resulting in all cells being active at the same time. One solution would be to have inhibitory neurons, but this does not affect the stability of the LFS, it just changes the ultimate level of activity reached (Avermann et al., 2012). Stability can be assessed in a number of different ways. First, stability in the nonlinear system sense: is the LFS a fixed point of a noise-less, infinite size system? We determined that there was a range of coupling strengths J, below J_c, for which such a LFS exists. The higher the baseline firing rate, the smaller that range is. Finite-size systems have a smaller range of stable coupling strength, because there is heterogeneity, not every neuron has the same in-degree. For instance, the uncorrelated network had a higher variance in the degree distribution than the ER network, and also had a smaller J_c. Interestingly, networks with a positive correlation between in- and out-degrees reduced stability even more, leading to a lower J_c, whereas for networks with a degree anti-correlation, J_c was higher, even exceeding the value for the ER network of the same size.

These calculations ignore the effects of fluctuations, which we subsequently introduced by making the dynamics stochastic. This did not alter the stability as determined before in terms of the existence of the LFS, but introduced other features. The LFS has a BOA with a fuzzy boundary due to the stochastic dynamics. A network can then be unstable when the fluctuations are large enough to leave the BOA when you wait long enough. This is primarily a concern for J values close to (and below) J_c. We determined the fraction of trials during which the network left the LFS BOA during the simulated time interval. As expected the anti-correlated network is more stable, because J_c is larger. For coupling constants away from J_c, this way of characterizing the BOA does not work. Hence, we started the network in states with many more neurons active than would be expected as a result of any normal fluctuation, and determined whether it converged to the LFS or HFS. This revealed that the BOA was larger for the anti-correlated network even away from J_c.

Taken together, these results clearly show that anti-correlated networks are more stable than uncorrelated ones, which means they can operate stably at higher coupling strengths and baseline firing rates, which confers advantages when the sensitivity is higher for higher coupling strengths and baselines rates. Furthermore, their sensitivity is enhanced compared to ER networks with the same connection probability, because of a subset of neurons with a high out-degree.

Recent experiments summarized in Barth and Poulet (2012) show that the average firing rate in sensory cortex is low, especially in superficial layers. This holds for spontaneous as well as evoked activity, and for both anesthetized animals and awake animals and is the basis for the parameter settings in the model. Nevertheless, there is a small subset of cells that display high firing rates. Cells that have recently been active express the immediate-early gene *c-fos*. When the *c-fos* promoter is used to express the fluorescent marker GFP, the recently active cells can be targeted for recording *in vivo* and *in vitro*. The so called fosGFP+ cells had a higher firing rate both *in vivo* and *in vitro* and received more excitatory inputs and less inhibitory inputs (Yassin et al., 2010). Furthermore, these cells are more likely to be connected amongst themselves. In the anti-correlated networks, there are neurons with a high in-degree but a low out-degree which make the network more stable, and neurons with high out-degree but low in-degree that make the network more sensitive. The fosGFP+ neurons could correspond to the former group, which form the backbone for the spontaneous activity. We did not explicitly build in assortativity in the network to preferentially connect high in-degree neurons to each other as suggested by Yassin et al. (2010). We take from this result that the prevailing homeostatic processes create networks with more strongly connected sub-networks and produce cell-to-cell heterogeneity in the balance between excitation and inhibition. Training to detect electrical stimulation should thus be able to induce similar changes in network structure.

THE SENSITIVITY ESTIMATED USING DIFFERENT MEASURES OF NETWORK ACTIVITY

Rodents were able to distinguish between patterns of neural activity during spontaneous activity and those caused by single-cell nanostimulation. Nevertheless, this distinction was small, given the effect size measured experimentally (Houweling and Brecht, 2008). One hypothesis is that the total amount of activity (firing rate) due to nanostimulation significantly exceeds that expected of a typical fluctuation. For a stationary network dynamics, this implies a fixed threshold above which a fluctuation is more likely caused by nanostimulation, whereas fluctuations below the threshold are more likely due to spontaneous activity. This can be quantified using a ROC curve, and the area under it, the AUC. The ROC is the curve traced out by varying this threshold and plotting the true positive rate (nanostimulation above threshold) vs. false positive (spontaneous fluctuations above threshold). When both distributions for the fluctuations are Gaussians, the AUC corresponds to the difference in means divided by the (common) standard deviation (Kingdom and Prins, 2010). Hence it is a measure of the difference in response relative to the size fluctuations around it. We found that the main determinant of the AUC is the out-degree of the stimulated neurons, independent of the correlation between in- and out-degree in the network. The AUC increases with coupling strength and baseline firing rate. The anti-correlated network has an advantage because it allows for a broader range of J and r_0 values. It thus has an increased stability at equal sensitivity.

The above represents an underestimate of the sensitivity, because it assumes that the activity of each neuron contributes equally to the detection (decision) and that the temporal signature of the firing rate fluctuation is not informative. Our further analysis shows that each of these factors would improve detection performance and makes it therefore likely that state-of-the-art classification approaches such as support vector machines would even further improve performance. Taken together this means that as a system the rodent brain could reach a much higher sensitivity than predicted here, when it could utilize all the information available in the network activity. Model simulations of spike pattern detection by cortical networks (Haeusler and Maass, 2007) suggests that laminar models with plastic synapses allow for more accurate estimates of the detection capability compared to neural networks that do not take into account the layered structure of cortex.

DETECTING SIGNATURES OF ANTI-CORRELATED DEGREE DISTRIBUTIONS

The model makes the prediction that anti-correlated networks would be more appropriate for the detection of nanostimulation in stable networks. To test this prediction we need to be able to distinguish correlations in the degree structure of the network without having access to all the inputs and all the outputs of a subset of neurons. We find that anti-correlations change the frequency of specific network motifs in a way that is independent of the network size, which means that it can be determined by averaging across many smaller sub-networks. A "ring" motif, number 98, which was a projection from neuron 1 to 2, from 2 to 3 and from 3 to 1, discriminated best between correlated and anti-correlated networks (**Figure 7**). Pairs of motif counts increased discriminability to a small extent, and only when the counts were pooled. This shows that these networks can be detected experimentally based on sampling sub-networks comprised of 30 neurons, when enough samples are available.

FUTURE STUDIES SHOULD INCORPORATE MULTIPLE TYPES OF INTERNEURONS

The model was highly simplified so that we could focus on the connectivity structure. Having established the advantages of anti-correlation, our goal is to study the effects in more realistic networks. There are many other biophysical features that could be included in the model that would change the results quantitatively or, in some cases, even qualitatively. Here we highlight a small selection of the most relevant ones.

The first issue is inhibition. Experimental evidence shows that two types of inhibitory neurons, those expressing parvalbumin (PV) and somatostatin (SOM), are relevant in determining the gain of the response of pyramidal cells to whisker stimulation, visual stimulation or current injection (Gentet et al., 2010; Kwan and Dan, 2012; Lee et al., 2012). Avermann and coworkers

(Avermann et al., 2012) constructed a model of L2/3 in barrel cortex constrained by *in vitro* measurements and studied the effect of stimulating varying amounts of pyramidal cells expressing channelrhodopsin by light pulses. In this model the strongest projection, in terms of the connection probability and synaptic strength, was from pyramidal cells to fast spiking (FS) interneurons (corresponding to PV neurons). Even when a relatively small fraction of the pyramidal cells were stimulated, almost all FS cells were recruited. For higher fractions of stimulated pyramidal cells, the non-fast spiking (NFS) interneurons (such as SOM interneurons) would become gradually activated. As a result the pyramidal cell activity remained low despite strong stimulation. The authors hypothesize that the strong inhibition is a mechanism to maintain sparse spiking in the pyramidal cells, with the NFS cells providing a back-up inhibitory mechanism. It is not clear how this computational model would be applicable to *in vivo* dynamics where FS cells are already spontaneously active. Furthermore, the level of activity in the different interneurons depends on brain state (Gentet et al., 2010). We have simulated binary networks with inhibitory neurons and find that anti-correlated degree distributions in the E–E sub-network improve stability (and yield the same sensitivity).

Detection could also take place by a state change in the network. The network has a LFS and a not too biologically plausible HFS, which in the context of a network with inhibition would perhaps correspond to something like an upstate. The true positive rate would correspond to how often single-cell stimulation would drive the network out of the BOA for the LFS, whereas the false positive rate would correspond to how often this would happen in the spontaneous state. The latter is given by the fraction of trials the system goes to the HFS state (**Figure 3**). The former can be tuned by changing the number of neurons and the duration of stimulation. A proper examination of this issue would require a network with a population of inhibitory neurons (Avermann et al., 2012).

A second issue is the effect of including spike timing. Synapses are sensitive through short-term depression and facilitation to the temporal patterns of stimulation (Abbott and Regehr, 2004), which could thereby affect the postsynaptic response in a nonlinear fashion, thereby preferentially activating specific populations of neurons. Dendritic nonlinearities also affect the impact of synaptic inputs based on their temporal coincidence and whether they arrive on the same part of the dendrite (Gasparini et al., 2004; Major et al., 2008; Polsky et al., 2009; Lavzin et al., 2012). Either of these effects could increase the sensitivity to external stimulation, while not appreciably changing the stability, thereby strengthening the results reported in this paper. However, to fully quantify these effects would require new and more extensive simulations that fall outside the scope of this paper.

PERCEPTUAL RELEVANCE OF ELECTRICAL OR OPTICAL STIMULATION IN EXPERIMENT

Our study explores a hypothesis for how to achieve detection of an electrical stimulation by quick recurrent excitation that escapes before being shut down by inhibition, without destabilizing the spontaneous state. We now review the relevant literature focusing on the difference between electrical and sensory stimulation and the role of inhibition.

The barrel cortex normally processes thalamic activity generated in response to whisker stimulation. According to the canonical cortical circuit (Douglas and Martin, 2004; Lefort et al., 2009; Petersen and Crochet, 2013) this activity arrives first in layer 4 (L4) of the barrel column representing the stimulated whisker and then goes to L2/3 and subsequently to L5. It stands to reason that when during a task an animal needs to make a decision based on whisker stimulation, this is based on activity in L2/3 or L5 that came there by way of L4. The path taken by activity induced by optical, micro- or nanostimulation does not necessarily directly involve L4 and improving detection could thus require altering the underlying cortical circuit.

When monkeys were trained to detect microstimulation at a location in the visual cortex corresponding to a specific retinotopic location, the stimulation threshold for detection was reduced from about 50 μA to 5 μA over a few thousands of trials (Ni and Maunsell, 2010). At the same time the contrast threshold needed to detect real visual stimuli at the same retinotopic location increased from 4–8% to 8–60%. When the monkeys were subsequently retrained on detecting visual stimuli, the sensitivity was recovered in another few thousand trials, but the sensitivity to electrical stimulation was reduced. One possible interpretation is that learning to detect electrical stimulation reorganizes the recurrent circuits in L2/3 to become more sensitive at the expense of the L4 to L2/3 feedforward connection.

The animal improves its performance when learning to detect microstimulation, which could also be the case for single-cell nanostimulation modeled here. This improvement could occur because of one or more of the following reasons. First, the network could become more anti-correlated by changing the in-degrees. This means that the stability of the network would improve over time and perhaps that the number of false positives would reduce. Second, the out-degree of the stimulated neurons could increase, so that the nanostimulation signal becomes louder, hence the true positives should increase. Third, the neurons involved in the detection process become more sensitive to neurons directly downstream of the stimulated cells.

The threshold for detecting microstimulation in monkey visual cortex matches the strength necessary to elicit action potentials in mouse and cat cortex in the neighborhood of the electrode, 5–10 μA (Histed et al., 2009) and in rat barrel cortex 2–5 μA (Houweling and Brecht, 2008). These numbers did not depend on whether metal or glass pipette electrodes were used. Stimulation close above this threshold activated a set of widely dispersed neurons within a few hundred microns from the electrode, through antidromic action potentials in axons that are close to the electrode. As a result, the spatial pattern of activation was very sensitive to small changes in the location of the electrode.

Similar stimulus strength, 10 μA for 0.1 to 0.5 ms, yielding charge transfers on the order of 1 nC, applied in the infragranular layers could be detected in rats (Butovas and Schwarz, 2007). The authors (Butovas and Schwarz, 2003) estimate that this corresponds to activating 80% of the pyramidal cells within 450 micron of the electrode, yielding an increase in their firing rate of 25% corresponding to about 0.5 excess spike per neuron. Interestingly,

trains of electrical stimulation were more effective, indicating that temporal correlation may be necessary to distinguish stimulation from spontaneous activity. Physiological measurements indicated that synapses of pyramidal cells on fast spiking interneurons depress more than the pyramidal to pyramidal synapses, which means that pulse trains could lead to more a prominent increase in activity than single stimuli (Holmgren et al., 2003).

Optogenetics was used to determine how many neurons in L2/3 would be required to generate a change in activity that would be detectable by a mouse (Huber et al., 2008). The authors' estimate of 300 neurons producing one action potential was based on a measured distribution of light intensity thresholds necessary to elicit an action potential, the number of neurons expressing the light-sensitive channelrhodopsin (ChR2) channels and the spatial fall off of the light intensity, and represents according to these authors an overestimate. The number of 300 neurons corresponds to about 5% of the approximately 6500 neurons present in a mouse barrel column (Lefort et al., 2009).

Nanostimulation refers to electrical activation of an individual neuron with a glass pipette in the juxtacellular configuration. Nanostimulation in rat barrel cortex must have led to behaviorally relevant changes in network activity, as the animal was able to detect nanostimulation, but the average effect size was rather small (Houweling and Brecht, 2008). The nature of this activity could not be assessed, but experiments in mouse visual cortex may shed some light on this. Single-cell stimulation led to spikes in the stimulated neuron and calcium transients in some of the surrounding neurons that could be detected using two-photon microscopy (Kwan and Dan, 2012). Such stimulation induced postsynaptic activity in very few other pyramidal cells, 20 out of 1152 measured. SOM interneurons [corresponding to the NFS of Avermann et al. (2012)] were most strongly activated, 5 out of 17 measured. PV expressing cells did not respond to this stimulation, but their calcium transients were most strongly correlated to the network activity produced by the rest of the measured cells. This indicates that in this state the SOM cells would be required to damp the increase in activity generated by the recurrently connected pyramidal cell network.

SUMMARY

Taken together, experimental results suggest that detection of single-cell stimulation requires a quick propagation of excitatory cell activity, before the various types of inhibition kick in. Our studies indicate that anti-correlated degree distributions could be an important strategy for increasing sensitivity while maintaining stability.

AUTHOR CONTRIBUTIONS

Paul Tiesinga and Arthur R. Houweling designed the research project, Paul Tiesinga and Juan C. Vasquez wrote the code, Juan C. Vasquez and Paul Tiesinga performed the simulations, Paul Tiesinga wrote the manuscript together with Arthur R. Houweling.

ACKNOWLEDGMENTS

This work was supported by the Netherlands Organization for Scientific Research (NWO), through a grant entitled "Reverse physiology of the cortical microcircuit," grant number 635.100.023.

REFERENCES

Abbott, L. F., and Regehr, W. G. (2004). Synaptic computation. *Nature* 431, 796–803. doi: 10.1038/nature03010

Avermann, M., Tomm, C., Mateo, C., Gerstner, W., and Petersen, C. C. (2012). Microcircuits of excitatory and inhibitory neurons in layer 2/3 of mouse barrel cortex. *J. Neurophysiol.* 107, 3116–3134. doi: 10.1152/jn.00917.2011

Barth, A. L., and Poulet, J. F. (2012). Experimental evidence for sparse firing in the neocortex. *Trends Neurosci.* 35, 345–355. doi: 10.1016/j.tins.2012.03.008

Bezdek, J. C. (1981). *Pattern Recognition with Fuzzy Objective Function Algorithms*. New York, London: Plenum. doi: 10.1007/978-1-4757-0450-1

Blitzstein, J. K., and Diaconis, P., (2006). *Algorithm for Graphs with Prescribed Degrees*. (Stanford, CA), preprint.

Bourjaily, M. A., and Miller, P. (2011a). Excitatory, inhibitory, and structural plasticity produce correlated connectivity in random networks trained to solve paired-stimulus tasks. *Front. Comput. Neurosci.* 5:37. doi: 10.3389/fncom.2011.00037

Bourjaily, M. A., and Miller, P. (2011b). Synaptic plasticity and connectivity requirements to produce stimulus-pair specific responses in recurrent networks of spiking neurons. *PLoS Comput. Biol.* 7:e1001091. doi: 10.1371/journal.pcbi.1001091

Butovas, S., and Schwarz, C. (2003). Spatiotemporal effects of microstimulation in rat neocortex: a parametric study using multielectrode recordings. *J. Neurophysiol.* 90, 3024–3039. doi: 10.1152/jn.00245.2003

Butovas, S., and Schwarz, C. (2007). Detection psychophysics of intracortical microstimulation in rat primary somatosensory cortex. *Eur. J. Neurosci.* 25, 2161–2169. doi: 10.1111/j.1460-9568.2007.05449.x

Chatterjee, S., Diaconis, P., and Sly, A. (2011). Random graphs with a given degree sequence. *Ann. Appl. Probab.* 21, 1400–1435. doi: 10.1214/10-AAP728

de Kock, C. P., Bruno, R. M., Spors, H., and Sakmann, B. (2007). Layer- and cell-type-specific suprathreshold stimulus representation in rat primary somatosensory cortex. *J. Physiol.* 581, 139–154. doi: 10.1113/jphysiol.2006.124321

de Kock, C. P., and Sakmann, B. (2009). Spiking in primary somatosensory cortex during natural whisking in awake head-restrained rats is cell-type specific. *Proc. Natl. Acad. Sci. U.S.A.* 106, 16446–16450. doi: 10.1073/pnas.0904143106

Del Genio, C. I., Kim, H., Toroczkai, Z., and Bassler, K. E. (2010). Efficient and exact sampling of simple graphs with given arbitrary degree sequence. *PLoS ONE* 5:e10012. doi: 10.1371/journal.pone.0010012

Douglas, R. J., and Martin, K. A. (2004). Neuronal circuits of the neocortex. *Annu. Rev. Neurosci.* 27, 419–451. doi: 10.1146/annurev.neuro.27.070203.144152

Duda, R. O., Hart, P. E., and Stork, D. G. (2001). *Pattern Classification*. New York, Chichester: Wiley.

Erdos, P. L., Miklos, I., and Toroczkai, Z. (2010). A simple Havel-Hakimi type algorithm to realize graphical degree sequences of directed graphs. *Electron. J. Combin.* 17, 1–10. Available online at: http://www.combinatorics.org/ojs/index.php/eljc/article/view/v17i1r66Article number #R66

Fellous, J. M., Tiesinga, P. H., Thomas, P. J., and Sejnowski, T. J. (2004). Discovering spike patterns in neuronal responses. *J. Neurosci.* 24, 2989–3001. doi: 10.1523/JNEUROSCI.4649-03.2004

Gasparini, S., Migliore, M., and Magee, J. C. (2004). On the initiation and propagation of dendritic spikes in CA1 pyramidal neurons. *J. Neurosci.* 24, 11046–11056. doi: 10.1523/JNEUROSCI.2520-04.2004

Gentet, L. J., Avermann, M., Matyas, F., Staiger, J. F., and Petersen, C. C. (2010). Membrane potential dynamics of GABAergic neurons in the barrel cortex of behaving mice. *Neuron* 65, 422–435. doi: 10.1016/j.neuron.2010.01.006

Greenberg, D. S., Houweling, A. R., and Kerr, J. N. (2008). Population imaging of ongoing neuronal activity in the visual cortex of awake rats. *Nat. Neurosci.* 11, 749–751. doi: 10.1038/nn.2140

Haeusler, S., and Maass, W. (2007). A statistical analysis of information-processing properties of lamina-specific cortical microcircuit models. *Cereb. Cortex* 17, 149–162. doi: 10.1093/cercor/bhj132

Histed, M. H., Bonin, V., and Reid, R. C. (2009). Direct activation of sparse, distributed populations of cortical neurons by electrical microstimulation. *Neuron* 63, 508–522. doi: 10.1016/j.neuron.2009.07.016

Holmgren, C., Harkany, T., Svennenfors, B., and Zilberter, Y. (2003). Pyramidal cell communication within local networks in layer 2/3 of rat neocortex. *J. Physiol.* 551, 139–153. doi: 10.1113/jphysiol.2003.044784

Houweling, A. R., and Brecht, M. (2008). Behavioural report of single neuron stimulation in somatosensory cortex. *Nature* 451, 65–68. doi: 10.1038/nature06447

Hubel, D. H., and Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *J. Physiol.* 195, 215–243.

Huber, D., Gutnisky, D. A., Peron, S., O'Connor, D. H., Wiegert, J. S., Tian, L., et al. (2012). Multiple dynamic representations in the motor cortex during sensorimotor learning. *Nature* 484, 473–478. doi: 10.1038/nature11039

Huber, D., Petreanu, L., Ghitani, N., Ranade, S., Hromadka, T., Mainen, Z., et al. (2008). Sparse optical microstimulation in barrel cortex drives learned behaviour in freely moving mice. *Nature* 451, 61–64. doi: 10.1038/nature06445

Itzkovitz, S., Milo, R., Kashtan, N., Ziv, G., and Alon, U. (2003). Subgraphs in random networks. *Phys. Rev. E. Stat. Nonlin. Soft Matter Phys.* 68, 026127. doi: 10.1103/PhysRevE.68.026127

Kim, H., Del Genio, C. I., Bassler, K. E., and Toroczkai, Z. (2012). Constructing and sampling directed graphs with given degree sequences. *New J. Phys.* 14, 1–23. doi: 10.1088/1367-2630/14/2/023012

Kingdom, F., and Prins, N. P. D. (2010). *Psychophysics: a Practical Introduction*. London: Academic.

Kumar, A., Schrader, S., Aertsen, A., and Rotter, S. (2008). The high-conductance state of cortical networks. *Neural Comput.* 20, 1–43. doi: 10.1162/neco.2008.20.1.1

Kwan, A. C., and Dan, Y. (2012). Dissection of cortical microcircuits by single-neuron stimulation *in vivo*. *Curr. Biol.* 22, 1459–1467. doi: 10.1016/j.cub.2012.06.007

LaMar, M. D., and Smith, G. D. (2010). Effect of node-degree correlation on synchronization of identical pulse-coupled oscillators. *Phys. Rev. E. Stat. Nonlin. Soft Matter Phys.* 81, 046206. doi: 10.1103/PhysRevE.81.046206

Lanciego, J. L., and Wouterlood, F. G. (2011). A half century of experimental neuroanatomical tracing. *J. Chem. Neuroanat.* 42, 157–183. doi: 10.1016/j.jchemneu.2011.07.001

Lavzin, M., Rapoport, S., Polsky, A., Garion, L., and Schiller, J. (2012). Nonlinear dendritic processing determines angular tuning of barrel cortex neurons *in vivo*. *Nature* 490, 397–401. doi: 10.1038/nature11451

Lee, S. H., Kwan, A. C., Zhang, S., Phoumthipphavong, V., Flannery, J. G., Masmanidis, S. C., et al. (2012). Activation of specific interneurons improves V1 feature selectivity and visual perception. *Nature* 488, 379–383. doi: 10.1038/nature11312

Lefort, S., Tomm, C., Floyd Sarria, J. C., and Petersen, C. C. (2009). The excitatory neuronal network of the C2 barrel column in mouse primary somatosensory cortex. *Neuron* 61, 301–316. doi: 10.1016/j.neuron.2008.12.020

Litwin-Kumar, A., and Doiron, B. (2012). Slow dynamics and high variability in balanced cortical networks with clustered connections. *Nat. Neurosci.* 15, 1498–1505. doi: 10.1038/nn.3220

London, M., Roth, A., Beeren, L., Hausser, M., and Latham, P. E. (2010). Sensitivity to perturbations *in vivo* implies high noise and suggests rate coding in cortex. *Nature* 466, 123–127. doi: 10.1038/nature09086

Major, G., Polsky, A., Denk, W., Schiller, J., and Tank, D. W. (2008). Spatiotemporally graded NMDA spike/plateau potentials in basal dendrites of neocortical pyramidal neurons. *J. Neurophysiol.* 99, 2584–2601. doi: 10.1152/jn.00011.2008

Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., and Alon, U. (2002). Network motifs: simple building blocks of complex networks. *Science* 298, 824–827. doi: 10.1126/science.298.5594.824

Newman, M. E. J. (2010). *Networks: an Introduction*. Oxford: Oxford University Press.

Ni, A. M., and Maunsell, J. H. (2010). Microstimulation reveals limits in detecting different signals from a local cortical region. *Curr. Biol.* 20, 824–828. doi: 10.1016/j.cub.2010.02.065

Osakada, F., Mori, T., Cetin, A. H., Marshel, J. H., Virgen, B., and Callaway, E. M. (2011). New rabies virus variants for monitoring and manipulating activity and gene expression in defined neural circuits. *Neuron* 71, 617–631. doi: 10.1016/j.neuron.2011.07.005

Park, J., and Newman, M. E. (2004). Statistical mechanics of networks. *Phys. Rev. E. Stat. Nonlin. Soft Matter Phys.* 70, 066117. doi: 10.1103/PhysRevE.70.066117

Perin, R., Berger, T. K., and Markram, H. (2011). A synaptic organizing principle for cortical neuronal groups. *Proc. Natl. Acad. Sci. U.S.A.* 108, 5419–5424. doi: 10.1073/pnas.1016051108

Petersen, C. C., and Crochet, S. (2013). Synaptic computation and sensory processing in neocortical layer 2/3. *Neuron* 78, 28–48. doi: 10.1016/j.neuron.2013.03.020

Polsky, A., Mel, B., and Schiller, J. (2009). Encoding and decoding bursts by NMDA spikes in basal dendrites of layer 5 pyramidal neurons. *J. Neurosci.* 29, 11891–11903. doi: 10.1523/JNEUROSCI.5250-08.2009

Roberts, E. S., and Coolen, A. C. (2012). Unbiased degree-preserving randomization of directed binary networks. *Phys. Rev. E. Stat. Nonlin. Soft Matter Phys.* 85, 046103. doi: 10.1103/PhysRevE.85.046103

Roxin, A. (2011). The role of degree distribution in shaping the dynamics in networks of sparsely connected spiking neurons. *Front. Comput. Neurosci.* 5:8. doi: 10.3389/fncom.2011.00008

Shadlen, M. N., and Newsome, W. T. (1998). The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. *J. Neurosci.* 18, 3870–3896.

Silverman, B. W., (1986). *Density Estimation for Statistics and Data Analysis*. London: Chapman and Hall.

Song, S., Sjostrom, P. J., Reigl, M., Nelson, S., and Chklovskii, D. B. (2005). Highly nonrandom features of synaptic connectivity in local cortical circuits. *PLoS Biol* 3:e68. doi: 10.1371/journal.pbio.0030068

Strogatz, S. H., (1994). *Nonlinear Dynamics and Chaos: With Applications in Physics, Biology, Chemistry, and Engineering.* Reading, Wokingham: Addison-Wesley Pub.

Viger, F., and Latapy, M. (2005). "Efficient and simple generation of random simple connected graphs with prescribed degree sequence," in *Computing and Combinatorics, Proceedings,* ed L. H. Wang (Berlin: Springer-Verlag Berlin), 440–449. doi: 10.1007/11533719_45

Wickersham, I. R., Lyon, D. C., Barnard, R. J., Mori, T., Finke, S., Conzelmann, K. K., et al. (2007). Monosynaptic restriction of transsynaptic tracing from single, genetically targeted neurons. *Neuron* 53, 639–647. doi: 10.1016/j.neuron.2007.01.033

Yassin, L., Benedetti, B. L., Jouhanneau, J. S., Wen, J. A., Poulet, J. F., and Barth, A. L. (2010). An embedded subnetwork of highly active neurons in the neocortex. *Neuron* 68, 1043–1050. doi: 10.1016/j.neuron.2010.11.029

Zhao, L., Beverlin, B. 2nd., Netoff, T., and Nykamp, D. Q. (2011). Synchronization from second order network connectivity statistics. *Front. Comput. Neurosci.* 5:28. doi: 10.3389/fncom.2011.00028

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.

Received: 10 July 2013; accepted: 17 October 2013; published online: 07 November 2013.

Citation: Vasquez JC, Houweling AR and Tiesinga P (2013) Simultaneous stability and sensitivity in model cortical networks is achieved through anti-correlations between the in- and out-degree of connectivity. Front. Comput. Neurosci. 7:156. doi: 10.3389/fncom.2013.00156

This article was submitted to the journal Frontiers in Computational Neuroscience.

Copyright © 2013 Vasquez, Houweling and Tiesinga. 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) or licensor 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.

frontiers in
COMPUTATIONAL NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 23 October 2013
doi: 10.3389/fncom.2013.00132

Statistical evaluation of synchronous spike patterns extracted by frequent item set mining

Emiliano Torre[1]*, David Picado-Muiño[2], Michael Denker[1], Christian Borgelt[2] and Sonja Grün[1,3]

[1] *Institute of Neuroscience and Medicine (INM-6) and Institute for Advanced Simulation (IAS-6), Jülich Research Centre and JARA, Jülich, Germany*
[2] *European Centre for Soft Computing, Mieres, Spain*
[3] *Theoretical Systems Neurobiology, RWTH Aachen University, Aachen, Germany*

Edited by:
Ruben Moreno-Bote, Foundation Sant Joan de Deu, Spain

Reviewed by:
Shigeru Shinomoto, Kyoto University, Japan
Srdjan Ostojic, Ecole Normale Superieure, France

***Correspondence:**
Emiliano Torre, Institute of Neuroscience and Medicine (INM-6) and Institute for Advanced Simulation (IAS-6), Jülich Research Centre, Wilhelm-Johnen-Strasse, 52425 Jülich, Germany
e-mail: e.torre@fz-juelich.de

We recently proposed frequent itemset mining (FIM) as a method to perform an optimized search for patterns of synchronous spikes (*item sets)* in massively parallel spike trains. This search outputs the occurrence count (*support*) of individual patterns that are not trivially explained by the counts of any superset (*closed frequent item sets*). The number of patterns found by FIM makes direct statistical tests infeasible due to severe multiple testing. To overcome this issue, we proposed to test the significance not of individual patterns, but instead of their signatures, defined as the pairs of pattern size z and support c. Here, we derive in detail a statistical test for the significance of the signatures under the null hypothesis of full independence (*pattern spectrum filtering*, PSF) by means of surrogate data. As a result, injected spike patterns that mimic assembly activity are well detected, yielding a low false negative rate. However, this approach is prone to additionally classify patterns resulting from chance overlap of real assembly activity and background spiking as significant. These patterns represent false positives with respect to the null hypothesis of having one assembly of given signature embedded in otherwise independent spiking activity. We propose the additional method of *pattern set reduction* (PSR) to remove these false positives by conditional filtering. By employing stochastic simulations of parallel spike trains with correlated activity in form of injected spike synchrony in subsets of the neurons, we demonstrate for a range of parameter settings that the analysis scheme composed of FIM, PSF and PSR allows to reliably detect active assemblies in massively parallel spike trains.

Keywords: higher-order correlations, neuronal cell assemblies, spike patterns, spike synchrony, multiple testing, data mining

1. INTRODUCTION

The cortex is comprised of a highly interconnected network of neurons and thus one may speculate that information processing in the brain may only be understood on the basis of the concerted activity of the neuronal population. Hebb (1949) suggested that neurons coordinate their activities by organizing in functional groups, termed cell assemblies. Synchronous spike input to receiving neurons is known to be more effective in generating output spikes (Abeles, 1982; König et al., 1996), which leads to the hypothesis that temporal coordination of spiking activity or correlational processing is the defining expression of an active cell assembly (Singer et al., 1997; Harris, 2005). As excitatory postsynaptic potentials are small in amplitude compared to the gap between the resting potential and the neuronal firing threshold, it is expected that a cell assembly is composed of many neurons firing in a correlated fashion. This observation is the basis for the assumption that higher-order synchronous spiking activity serves as a signature expression of an active assembly (Riehle et al., 1997; Berger et al., 2010; Staude et al., 2010b; Shimazaki et al., 2012).

In order to observe and detect such signatures in the brain, the spiking activities of many neurons must be recorded simultaneously. Fortunately, in recent years considerable progress has been made in the development of multi-electrode recording techniques [e.g., Nicolelis, 1998; Buzsaki, 2004; Hatsopoulos et al., 2007; Riehle et al., 2013], which enable to record the activity of hundred(s) of neurons. Such massively parallel spike train data pose statistical challenges due to the inherent complexity of the required multivariate approaches. Most notably, increasing the number of observed neurons leads to a combinatorial explosion of the number of potential spike patterns that need to be detected and tested. Based on pairwise correlation analyses only, the existence and functional relevance of neuronal correlations could be demonstrated in various cortical systems and behavioral paradigms [e.g., Gerstein and Aertsen, 1985; Riehle et al., 1997; Kohn and Smith, 2005; Berger et al., 2007; Fujisawa et al., 2008; Feldt et al., 2009; Humphries, 2011; Masud and Borisyuk, 2011]. Nevertheless, a correlation analysis considering the complete set of simultaneously recorded spike trains is required to uncover also higher-order correlations among neurons. In recent years several such approaches were developed, each of which focuses on different aspects: (i) methods to determine the presence of higher-order spike correlations with a minimum order without explicitly identifying the participating neurons [e.g., Louis et al., 2010a; Staude et al., 2010a,b]; (ii) methods that test whether

individual neurons participate in synchronous spiking activity without identifying the groups of correlated neurons [e.g., Berger et al., 2010]; (iii) methods that test for the presence of correlation as predicted by a specific correlation model such as a synfire chain (Abeles, 1991), that is, spatio-temporal spike patterns or propagation of synchronous spiking activity [e.g., Abeles and Gerstein, 1988; Schrader et al., 2008; Gerstein et al., 2012; Gansel and Singer, 2012]; (iv) methods that directly identify the members of cell assemblies on the basis of the patterns of synchronous spiking activity [e.g., Gerstein et al., 1978; Pipa et al., 2008; Feldt et al., 2009; Gansel and Singer, 2012; Shimazaki et al., 2012; Picado-Muiño et al., 2013].

In Picado-Muiño et al. (2013) we presented the basic approach and relevant statistics to employ frequent item set mining (FIM) to identify significant patterns of spike synchrony in massively parallel spike trains. FIM enables fast and efficient counting of synchronous spike patterns by pruning the tree of all possible patterns. To address the problem of multiple testing, statistics are not computed for individual patterns, but on the pattern spectrum that collects the number of observed patterns based on their signature. A signature is defined as the pair (z, c) of pattern size z (i.e., number of participating neurons) and *support* c (i.e., number of pattern occurrences). In *pattern spectrum filtering* (PSF) those identified sets of neurons for which patterns with the same signature (z, c) occur also in appropriate surrogate data are then marked as chance patterns and discarded.

Here, we extend the approach of Picado-Muiño et al. (2013) in three ways that will enable the application of the method to biological data. First, we refine the statistical test employed in pattern spectrum filtering for reporting significant patterns of a given signature (Section 2). Then, we introduce a subsequent analysis step, termed *pattern set reduction* (PSR), to additionally filter out those patterns that are detected as significant, but are compositions of chance spikes or patterns and the actual cell assembly pattern (Section 3). Finally, we report on the performance of our method related to features describing the data (e.g., coincidence rate, assembly pattern size, firing rate heterogeneity or non-stationarity) and analysis parameters (Section 4). The discussion (Section 5) includes a step-by-step instruction on how to utilize the proposed method in the context of massively parallel spike trains obtained from electrophysiological recordings.

2. SPIKE PATTERN DETECTION AND STATISTICAL TESTING

In this section we introduce our approach to detect frequent synchronous spike patterns in massively parallel spike trains (MPST). We first briefly review frequent item set mining (FIM) and related terminology and definitions as proposed in Picado-Muiño et al. (2013) as a tool to efficiently detect and count synchronous spike patterns in MPST. Then we derive a modified version of the FIM-based statistics proposed in Picado-Muiño et al. (2013) for assessing pattern significance.

2.1. FREQUENT ITEMSET MINING

Given N parallel spike trains with neuron ids $1, 2, \ldots, N$, observed in the time window $[0, T)$, we partition $[0, T)$ into b exclusive bins $\{b_i\}_{i=1}^{b}$ of identical width $w = T/b$ (typically chosen as a few ms): $b_i = [(i-1) \cdot w, i \cdot w)$. If one or more spikes of one neuron fall into a bin, we consider the bin occupied and reduce the entry to 1 (*clipping*), so that each time bin contains at most one spike per neuron. Spikes from different neurons falling into the same time bin are defined as *synchronous* (see **Figure 1A**). Borrowing terminology from FIM, we define each neuron id as an *item*, the set T_i of all items spiking in b_i as the i-th *transaction* in the binned data, and $\{T_i\}_{i=1}^{b}$ as the *transaction list*. Given a *minimum pattern size* z_0, each set of $z \geq z_0$ items in T_i constitutes a *pattern of synchronous spikes*, or *item set* (see **Figure 1B**). Here we

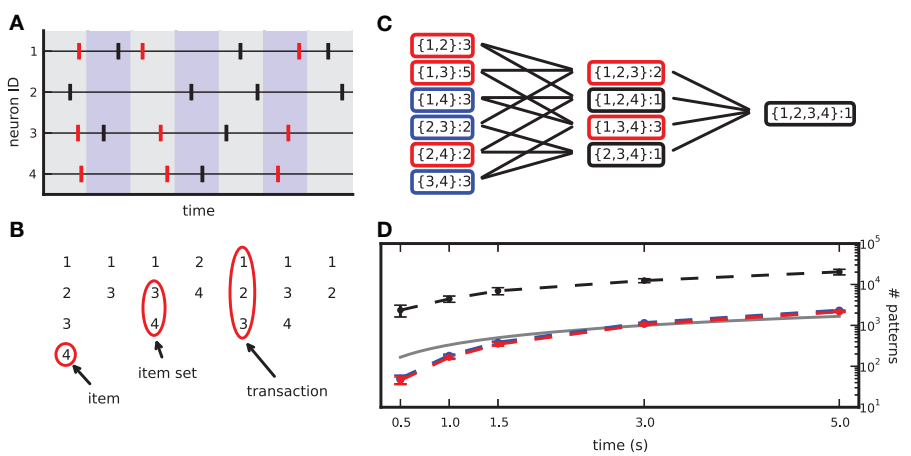


FIGURE 1 | From spike data to closed frequent itemsets. (A) Sketch of a raster plot of 4 neurons firing in parallel. Shaded colors separate adjacent bins. Red spikes mark the occurrences of the synchronous pattern composed of neurons 1, 3, 4. **(B)** Transaction list derived from the spike data in **(A)** after binning. **(C)** List of item sets obtained from **(B)**, together with their occurrence counts. Black boxes mark non-frequent item sets (support set to 2), blue boxes mark non-closed frequent item sets, red boxes mark CFISs. **(D)** Average number of item sets (dashed black line), frequent item sets (dashed blue line) and CFISs (dashed red line) obtained from 100 simulations of 100 parallel independent spike trains with a firing rate of 20 Hz, as a function of the simulation time. Other parameters are bin width $w = 3$ ms and minimum pattern size $z_0 = 2$. Bars mark ± 1 std. dev. The solid line indicates the number of time bins (and thus transactions) as a function of the simulation time.

set z_0 to 2. Due to clipping, each item set occurs at most once per transaction. The number of occurrences of an item set in the transaction list is the *support* of that item set.

A transaction that contains K items yields $2^K - K - 1$ different (but possibly overlapping) item sets of size $z \geq 2$, that is, all 2^K possible subsets without the empty set and the K singletons. The total number of different item sets in a transaction list can thus largely exceed the number of transactions (i.e., time bins). This number grows with the duration of the data set (see **Figure 1D**) and with the number of parallel spike trains (not shown).

In order to limit the data to potentially interesting and non-trivial item sets, we select only item sets whose support c is larger than or equal to a *minimum support* c_0 ($c_0 \geq 1$) as introduced by Picado-Muiño et al. (2013). Here we set c_0 to 2. An item set whose support equals or exceeds the minimum support is called *frequent item set*. For $c_0 > 1$, frequent item sets are usually a small fraction of all item sets (**Figure 1D**, compare black dashed line to blue dashed line). Furthermore, we discard any frequent item set occurring as many times as any of its supersets. These patterns are trivially explained by the occurrences of their supersets, which are more significant due to the larger number of neurons involved. Non-trivial frequent item sets are called *closed frequent item sets* (CFISs; see **Figure 1C**). Discarding non-closed frequent item sets does not yield any loss of information. Indeed, the set $\mathcal{F}$ of all frequent item sets can be reconstructed from the set $\mathcal{C}$ of CFISs by

$$\mathcal{F} = \bigcup_{I \in \mathcal{C}} \bigcup_{J \subset I,\, |J| \geq z_0} J.$$

The support $s(I)$ of a non-closed frequent item set $I \in \mathcal{F}$ can be computed as $s(I) = \max_{J \in C,\, J \supset I} s(J)$.

If A and B are two CFISs such that $B \subsetneq A$, and c_A, c_B their respective supports, it follows from the definition of CFISs that $c_B > c_A$ (*a priori* property). We refer to the (non-empty) set $A \backslash B$ as the *excess items* of A with respect to B, and to the difference $c_B - c_A$ as the *excess occurrences* of B with respect to A.

Following Picado-Muiño et al. (2013), we make use of frequent itemset mining [FIM; for a review, see Goethals (2010), Borgelt (2012)] to extract CFISs and their support from an MPST transaction list. FIM performs a non-redundant search for spike patterns, starting from those of size z_0 and then moving on to supersets of increasing size. Starting at lowest-size patterns, the search is organized in a search tree in layers of increasing pattern size. A branch connects two patterns if one is a subset of the other. Each pattern is visited at most once. FIM exploits the *apriori* property to stop the search at infrequent patterns, as no supersets of an infrequent item set can be frequent. The output of FIM is a list of all CFISs with their support (**Figure 1C**).

2.2. PATTERN SPECTRUM FILTERING

Direct statistical tests of all individual patterns occurring in MPST are not suitable, as they cause a severe multiple testing problem yielding large occurrences of false positives (FPs), or enhanced levels of false negatives (FNs) after statistical corrections. Therefore Picado-Muiño et al. (2013) proposed to pool CFISs according to their size z (number of neurons involved) and their support c (number of occurrences) in a two-dimensional histogram (*pattern spectrum*) and to evaluate patterns of the same signature (z, c) for significance by a Monte-Carlo approach using surrogate data. Here we present a refinement of this original approach, named *pattern spectrum filtering* (PSF), that bases the test for a specific signature (z, c) also on patterns of higher size and support than specified by the signature.

In order to implement the null hypothesis $\mathcal{H}_0$ of independent spiking, and to approximate the p-values of the signatures (z, c), from the original data (**Figure 2A**) we repeatedly generate surrogate data (**Figure 2B**), collect from each one its CFISs through FIM as done for the original data, and compute the corresponding surrogate pattern spectrum (**Figure 2C**). The surrogates are generated from the original data by intentionally destroying correlations while keeping other features, such as firing rates, intact [e.g., by spike randomization or spike dithering, Louis et al. (2010b)].

Let $\succeq$ be the partial ordering on the real plane, that is, $(x_*, y_*) \succeq (x, y)$ if $x_* \geq x$ and $y_* \geq y$, where $\succ$ holds if at least one inequality is strict. From each surrogate pattern spectrum we compute a binary spectrum which takes value 1 at each signature (z, c) such that at least one signature $(z_*, c_*) \succeq (z, c)$ is occupied, and value 0 otherwise [in contrast to Picado-Muiño et al. (2013) where only the occupation of signature (z, c) is checked]. Formally, we define the *signature operator* $\mathrm{sgt}(\cdot)$ such that, given a CFIS A with size $z_A = |A|$ and occurrence count c_A, $\mathrm{sgt}(A) := (z_A, c_A)$. For each list $\mathcal{S}_i$ of CFISs from one surrogate data set, let $\hat{P}_i$ be the *binary pattern spectrum*, defined for each $z, c \geq 2$ by:

$$\hat{P}_i(z, c) := \begin{cases} 1 \text{ if } \exists A \in \mathcal{S}_i : \mathrm{sgt}(A) \succeq (z, c) \\ 0 \text{ otherwise} \end{cases}.$$

Averaging the binary spectra at each signature, we get the *p-value spectrum* $\hat{P}$:

$$\hat{P}(z, c) := \frac{1}{K} \#\left(\mathcal{S}_i : \exists A \in \mathcal{S}_i : \mathrm{sgt}(A) \succeq (z, c)\right).$$

$\hat{P}(z, c)$ yields an estimate of the probability to observe (one or more) patterns with signature $(z_*, c_*) \succeq (z, c)$ under $\mathcal{H}_0$ (see **Figure 2D**).

We then classify any signature (z, c) whose p-value is lower than the significance level α_* as significant. Given the desired overall significance level α for PSF, we derive α_* from α by Bonferroni correction for the number m of tests, i.e., the number of signatures in the data to test for: $\alpha_* = \alpha/m$. Any signature (z, c) for which $\hat{P}(z, c) < \alpha_*$ is classified as significant. Formally, we introduce the *significance spectrum* $\hat{S}$ defined at each (z, c) by

$$\hat{S}(z, c) := \begin{cases} 1 & \text{if } (z, c) \text{ is significant} \\ 0 & \text{otherwise} \end{cases}.$$

In **Figure 2E** $\hat{S}(z, c) = 1$ is marked in white, $\hat{S}(z, c) = 0$ in gray. The border between the two is the *detection border*, on the left of which signatures in the original data are classified as

not significant and rejected. Signatures to its right ($\hat{S}(z, c) = 1$) are considered as significant (marked in red in **Figure 2E**). The corresponding patterns and their supports are listed in **Figure 2F**.

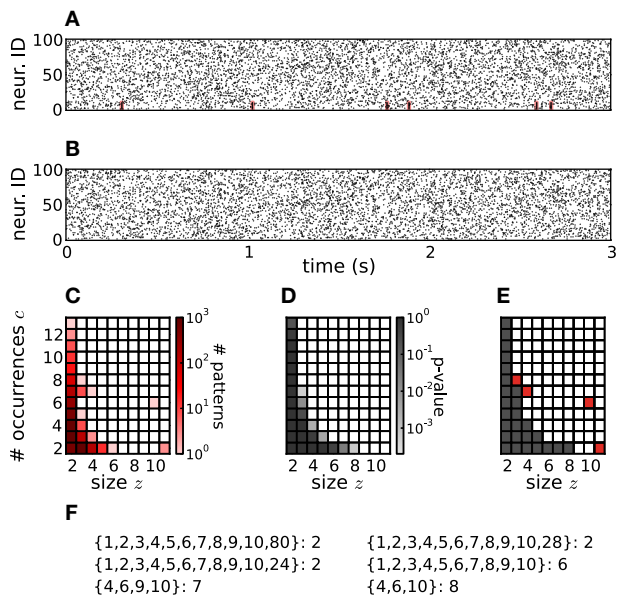


FIGURE 2 | PSF on artificial data. (A) Raster plot of 100 parallel simulated spike trains consisting of independent Poisson activity plus 6 injections of one pattern of synchronous spikes (highlighted in red) from neurons 1 to 10, occurring at random times (see Section 4 for details). The total firing rate of each neuron is 20 Hz, the simulation time is 3 s. **(B)** Same as in **(A)**, but without injection of synchronous patterns. The spike trains are therefore completely independent. **(C)** Pattern spectrum of CFISs extracted from the data in **(A)** by FIM ($z_0 = 2$, $c_0 = 2$, $w = 5$ ms). Counts are color-coded (logarithmic scale). **(D)** *P*-value spectrum drawn from 5000 surrogate, independent data sets of the type shown in **(B)**. *P*-values are color-coded (logarithmic scale). **(E)** Significance spectrum (overall significance $\alpha = 0.01$, Bonferroni-corrected for $m = 50$ tests yielding $\alpha_* = 2 \cdot 10^{-4}$). Gray squares indicate signatures that are not significant, white squares mark potentially significant signatures. Red squares mark significant signatures of the pattern spectrum shown in **(C)**, i.e., which fall into white squares of the significance spectrum. **(F)** List of patterns detected by PSF. Besides the injected pattern $A = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$, PSF also classifies additional patterns as significant, all being subsets or supersets of *A*.

3. PATTERN SET REDUCTION

PSF tests the significance of patterns under the null hypothesis $\mathcal{H}_0$ of fully uncorrelated spike trains. However, PSF might fail in rejecting patterns that result from combinations of chance spikes or chance patterns with the assembly pattern (see list of detected patterns in **Figure 2F** besides the injected one). These patterns are a specific kind of false positive, not resulting from merely independent data. They may be subsets or supersets of the assembly pattern, or patterns that partially overlap with it (**Figures 3A–C**). In this section we define the type of FPs that may occur, investigate why PSF is prone to return such FPs, and propose an additional statistical analysis, termed *pattern set reduction* (PSR), to remove them.

3.1. TYPES OF FPs

3.1.1. Chance subsets

If a CFIS *A* repeats c_A times and a subset *B* of *A* (with $|B| \geq z_0$) has *c* additional chance occurrences, *B* represents a CFIS repeating $c_B = c_A + c$ total times. We call *B* a *chance subset* of *A*, having *c* excess occurrences (**Figure 3A**). PSF is designed to test the significance of signature ($|B|$, c_B) under $\mathcal{H}_0$ (complete independence), thus disregarding the fact that c_A occurrences are due to pattern *A*. As a result it classifies *B* as a significant pattern, thus yielding an FP outcome. This is illustrated in **Figure 2F**, where e.g., pattern {4, 6, 10} occurs twice by chance plus 6 times as a subset of pattern {1, 2, 3, 4, 5, 6, 7, 8, 9, 10}. The corresponding signature (3, 8) is significant compared to the surrogates (**Figure 2E**), so that PSF does not reject it.

3.1.2. Chance supersets

If a CFIS *B* occurs c_B times and another set *C* of neurons fires by chance synchronously with *B* in *c* of those c_B transactions (with $c \geq c_0$), then the pattern $A = B \cup C$ represents a CFIS repeating $c_A = c$ times. We call *A* a *chance superset* of *B*, with $|C|$ excess neurons (**Figure 3B**). PSF tests the significance of signature ($|A|$, c_A) under $\mathcal{H}_0$, disregarding the fact that $|B|$ of the $|A|$ neurons of *A* are due to the presence of pattern *B*. The test is therefore prone to classify *A* as significant. This is the case for patterns $\{1, 2, \ldots, 10, 80\}$, $\{1, 2, \ldots, 10, 28\}$ and $\{1, 2, \ldots, 10, 24\}$ in **Figure 2F**, each of which occurs twice as a superset of $\{1, 2, \ldots, 10\}$. The corresponding signature (11, 2) is significant compared to the surrogates (**Figure 2E**), so that PSF classifies these patterns as significant.

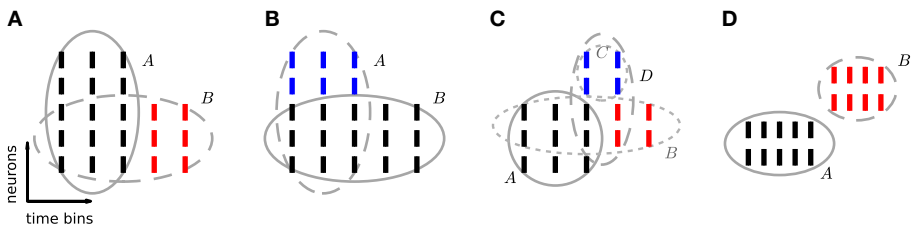


FIGURE 3 | Excess occurrences and excess items. Sketch of the possible relationship between a reference pattern and patterns sharing neuron identities and/or time occurrences with it. In each panel, ticks represent individual spikes. Rows correspond to neurons and columns to transactions, i.e., time bins. Spikes forming a pattern are grouped by an ellipse. The reference pattern of each panel is shown by black ticks and is indicated by a solid ellipse. **(A)** *B* is a subset of *A* with excess occurrences (red). **(B)** *A* is a superset of *B* with excess items (blue). **(C)** *B* is a subset of *A* with excess occurrences (red). Neurons in *C* (blue) additionally fire synchronously to *A* and to excess occurrences of *B*. Thus pattern $D = B \cup C$ forms a CFIS, which partially overlaps with *A*. **(D)** Patterns *A* and *B* are disjoint: they are composed of different neuron identities and occur at different time bins.

3.1.3. Chance overlapping sets

The simultaneous presence of excess items and excess occurrences can yield yet another type of FP outcome, namely patterns that overlap with the actual assembly. Given an assembly A, assume that a subset B of A has additional chance occurrences. If an additional set C of neurons disjoint from A fires synchronously to A *and* to an excess occurrence of B for a total of $c \geq c_0$ chance times, then the set $D = B \cup C$ represents a CFIS which partially overlaps with A (**Figure 3C**). PSF is prone to classify D as significant.

3.1.4. Disjoint patterns

Two patterns which do not have items in common are *disjoint* (**Figure 3D**). In contrast to the previous classes of chance patterns, the presence of an active assembly does not enhance chance patterns disjoint from it. PSF therefore correctly estimates their significance and manages to filter out almost all of them, as shown in 4.

3.2. PSR STATISTICS

Let $\mathcal{P}$ be the class of CFISs reported as significant by PSF. Given a pair $(A, B) \in \mathcal{P} \times \mathcal{P}$ such that $B \subset A$ (therefore $c_B > c_A$ by definition of CFIS, and $|B| < |A|$), we propose statistical tests to assess the conditional significance of either A *given* B ($A|B$) or B *given* A ($B|A$), i.e., of one pattern given that the other represents an assembly pattern. These tests can be applied, using different strategies, to the class of all such (A, B) pairs, reducing $\mathcal{P}$ to a subclass $\mathcal{Q}$ of patterns which are mutually significant given each other.

3.2.1. Subset filtering

This procedure aims at rejecting FPs that are chance subsets of other CFISs. For each pair $(A, B) \in \mathcal{P} \times \mathcal{P}$ such that $B \subset A$ (so that $c_B > c_A$), B has $c_B - c_A$ excess occurrences with respect to A. Subset filtering tests $B|A$, i.e., the null hypothesis $\mathcal{H}_0^{B|A}$ that B is a chance subset of the actual assembly A, by assessing the significance of the excess occurrences of B. Equivalently, $\mathcal{H}_0^{B|A}$ states that the pattern B' defined by the same items as B but its excess occurrences only (red spikes in **Figure 3A**) is a chance pattern. If $\mathcal{H}_0^{B|A}$ is rejected, B is kept and A discarded, otherwise A is kept and B discarded. Thus, the procedure keeps either A or B and discards the other (*exclusive*). We present two alternatives to test $\mathcal{H}_0^{B|A}$.

3.2.1.1. Exact test. This test computes the p-value of the signature $(|B|, c_B - c_A)$ of B'. If $c_B - c_A < c_0$, B is classified as a chance subset of A. Otherwise, let T'_A be the transaction list obtained from T by discarding the transactions where A occurred, and keeping in the remaining transactions only the items composing A. All the excess occurrences of subsets of A must be contained in T'_A. B' itself is a CFIS in this transaction list: it is an item set because $|B'| = |B| \geq z_0$, it is frequent because $c_B - c_A \geq c_0$, it is closed because otherwise B itself would be non-closed. To test the significance of B', one can therefore run FIM and PSF on surrogates of T'_A to estimate the significance of its signature $(|B|, c_B - c_A)$**.** If $(|B|, c_B - c_A)$ is significant, B' is significant in T'_A and B is classified as significant in T (given A). Otherwise, B is classified as non-significant.

3.2.1.2. Approximate test. This test approximates the p-value of the signature $(|B|, c_B - c_A)$ in T'_A by the p-value of the signature $(|B|, c_B - c_A + h)$, $h \geq 1$, in T, already obtained when performing PSF. In contrast to T'_A, T is composed of more neurons than those which can actually form chance subsets of A (because it does not contain the items of A only), and more transactions than those where such subsets could actually display excess occurrences (because it also contains the transaction where A is already present). Therefore, the p-value of $(|B|, c_B - c_A)$ would be underestimated if computed over T instead of T'_A. Parameter h heuristically corrects for this by substituting it with the p-value of a signature with the same size but higher support. The lower h, the higher the probability to reject B. If $h \geq c_A$, then $(|B|, c_B - c_A + h) \succeq (|B|, c_B)$ and B is necessarily reported as significant. This test avoids to run FIM and PSF on T'_A and is therefore computationally more efficient.

3.2.2. Superset filtering

This procedure aims at rejecting FPs that are chance supersets of other CFISs. For each pair $(A, B) \in \mathcal{P} \times \mathcal{P}$ such that $B \subset A$ (so that $|B| < |A|$), A has $|A| - |B|$ excess items with respect to B. Subset filtering tests $A|B$, i.e., the null hypothesis $\mathcal{H}_0^{A|B}$ that A is a chance superset of the actual assembly B, by assessing the significance of the excess items of A. Equivalently, $\mathcal{H}_0^{A|B}$ states that the pattern A' defined by the same transactions as A but containing its excess items only (blue spikes in **Figure 3B**), is a chance pattern. If $\mathcal{H}_0^{A|B}$ is rejected, A is kept and B discarded from $\mathcal{P}$, otherwise B is kept and A discarded from $\mathcal{P}$. Thus, the procedure keeps either A or B and discards the other (*exclusive*). We present two alternatives to test $\mathcal{H}_0^{A|B}$.

3.2.2.1. Exact test. This test computes the significance of the signature $(|A| - |B|, c_A)$ of A'. If $|A| - |B| < z_0$, A is classified as a chance superset of B. Otherwise, let $T'_{\bar{B}}$ be the transaction list obtained from T by keeping only the transaction where B occurred, and discarding from them the items constituting B. All groups of excess items of B (i.e., neurons that fire synchronously to B) must be contained in $T'_{\bar{B}}$. A' itself is a CFIS of this transaction list: it is an item set because $|A'| = |A| - |B| \geq z_0$, it is frequent because $c_A \geq c_0$, it is closed because otherwise A itself would be non-closed. To test the significance of A', one can therefore run FIM and PSF on surrogates of $T'_{\bar{B}}$ to estimate the p-value of its signature $(|A| - |B|, c_A)$**.** If $(|A| - |B|, c_A)$ is significant, A' is significant in $T'_{\bar{B}}$ and A is classified as significant in T (given B). Otherwise, A is classified as non-significant.

3.2.2.2. Approximate test. This test approximates the p-value of the signature of A' in $T'_{\bar{B}}$ by the p-value of signature $(|A| - |B| + k, c_A)$, $k \geq 1$, in T, already obtained when performing PSF. In contrast to $T'_{\bar{B}}$, T is composed of more neurons than those that can actually form excess items of B (because it contains the items of B, too), and more transactions than those where supersets of B could actually occur (because it contains also transactions where B does not occur). Therefore, the p-value of $(|A| - |B|, c_A)$ would be underestimated if computed over T instead of $T'_{\bar{B}}$. Parameter k heuristically corrects for this by substituting it with the p-value of a signature with the same support

but higher size. The lower k, the higher the probability to reject A. Note that if $k \geq |B|$ then $(|A| - |B| + k,\ c_A) \succeq (|A|,\ c_A)$ and A is necessarily reported as significant. This test allows to avoid running FIM and PSF on $T'_{\tilde{B}}$ for each B.

3.2.3. Covered-spikes criterion

This simple selection strategy consists of taking all pairs $(A, B) \in \mathcal{P} \times \mathcal{P}$ for which $B \subset A$, and keeping for each pair the pattern covering the largest number of spikes, while rejecting the other. Specifically, the criterion prefers A to B if $z_A \cdot c_A \geq z_B \cdot c_B$, B to A otherwise. It does not involve significance tests, but is based on the observation that, given the probability p for a neuron to spike in a time bin, the probability for z neurons to fire synchronously in a bin is approximately p^z, so that the probability that this pattern occurrs c times is binomially distributed and approximately proportional to $p^{z \cdot c}$. The larger the $z \cdot c$ score, the less likely a pattern of such size and support. This matches the finding that the detection border separating non-significant signatures (marked gray in **Figure 2E**) from significant ones (marked white in **Figure 2E**) in the significance spectrum exhibits a hyperbolic shape. The criterion thus keeps the less likely of the two patterns.

A variant consists in keeping the pattern with the largest $(z - 1) \cdot c$ score. This choice is motivated by the observation that a pattern of size z and support c can be seen as $z - 1$ spike trains which synchronize their spikes to another train c times. Thus, $(z - 1) \cdot c$ spikes are coincident to spikes in another spike train. Keeping the pattern with the largest $(z - 1) \cdot c$ score amounts to keeping the pattern which covers more coincident spikes. Geometrically, penalizing the pattern size corrects for the fact that the hyperbolic shape of the detection border in **Figure 2E** is elongated toward the pattern support (y-axis) rather than being equilateral.

3.2.4. Combined filtering

Subset filtering, superset filtering and covered-spikes criterion can be combined into a filtering procedure which tests for both excess coincidences and excess items. Combined filtering tests for each pair $(A, B) \in \mathcal{P} \times \mathcal{P}$ both the null hypotheses $\mathcal{H}_0^{B|A}$ (i.e., that B is a chance subset of A) *and* $\mathcal{H}_0^{A|B}$ (i.e., that A is a chance superset of B). If one of the null hypotheses is rejected, the corresponding pattern is retained as significant. Thus, if both hypotheses are rejected, both patterns are retained (*inclusive*). Accepting one null hypothesis does not necessarily lead to the rejection of the corresponding pattern (in contrast to subset or superset filtering): the pattern is rejected only if the other pattern is accepted, i.e., if the other null hypothesis is rejected. If both $\mathcal{H}_0^{B|A}$ and $\mathcal{H}_0^{A|B}$ are accepted, one of the two patterns is kept based on the covered-spikes criterion.

4. CALIBRATION ON ARTIFICIAL DATA

In this section we compare the performance (in terms of FPs and FNs) of PSF to PSF followed by PSR to illustrate the advantages yielded by the latter. For the sake of computational efficiency we employ the approximate versions of the tests for the subset and superset filtering with parameters $h = 1$ and $k = 2$, respectively. We test different types of artificial data that involve typical features of experimental data. After studying the general behavior of the analysis method for stationary, homogeneous data, we study data sets with heterogeneous firing rates across neurons, and with non-stationary firing rates in time.

4.1. CORRELATED DATA

As a model for data containing assembly activity, we generate correlated spike trains by a modified version of the single-interaction-process [SIP; Kuhn et al. (2003); Berger et al. (2010)], which we keep calling SIP for convenience. First, we simulate $N = 100$ parallel independent Poisson spike trains as background activity. Then we model assembly activity by inserting synchronous spike events in a subset of z of the N neurons (the *SIP neurons*, with ids 1 to z). This is done by generating a hidden Poisson process with the desired number c of pattern occurrences, from which spikes are copied into each of the z spike trains of the SIP neurons. Thus, as compared to the model proposed by Kuhn et al. (2003) we insert correlated firing only in a specific subset of the parallel processes. Before insertion of the synchronous patterns, the background firing rate of the SIP neurons is reduced by the rate of the hidden process to ensure the same firing rate for all neurons. In the simplest scenario, the firing rates and the pattern occurrence rate are stationary over time and homogeneous across neurons. More complicated cases will include either non-stationarity or heterogeneity of rates. The purpose of the analysis of such data is to test under controlled conditions if the simulated assembly is indeed detected and can be distinguished from background activity.

4.2. INDEPENDENT DATA

To implement the null-hypothesis $\mathcal{H}_0$ of complete independence needed to derive the significance of signatures of the correlated data, we generate independent Poisson processes of the same rates as the data to be tested, thus keeping the same marginal statistics. This is one way of implementing the null-hypothesis. However, in the context of analyzing real experimental data, one may want to keep more statistical features of the experimental data (e.g., non-stationary and heterogenous firing rates, deviation from Poisson, and so on). This can be realized by the use of more complex surrogates derived by manipulation of the original data, e.g., spike dithering (Grün, 2009; Louis et al., 2010b).

4.3. ASSESSING SIGNIFICANCE

We evaluate the performance of our analysis in terms of the average number of FPs and FNs obtained with PSF and PSR in $R = 1000$ iterations on the same model of correlated data (SIP of size z in $N = 100$ parallel spike trains). To study the performance of our analysis, we investigate 243 models differing in the size of the injected assembly $z = 2, \ldots, 10$, its injection count $c = 2, \ldots, 10$, and the firing rates $r = 5$, 10 or 20 Hz (here: homogeneous for all neurons). We analyse each model with a bin width $w = 3$ ms and $w = 5$ ms for the detection of synchronous spike patterns. See **Table 1** for an overview of the parameter combinations. For the significance estimation we generate surrogate data, i.e., independent Poisson processes with the same firing rates as the correlated data, and analyse them with FIM as done for the correlated data. This procedure is repeated for $K = 5000$ times to derive the p-value spectrum and then the significance

spectrum by employing an overall significance level of $\alpha = 0.01$, Bonferroni-corrected for the number of signatures tested. The latter is given by the number of signatures existent in the correlated data, which never exceeded $m = 50$. In order to have the same corrected significance level for each of the 1000 iterations of each SIP model, we always correct for $m = 50$ tests, instead of correcting for the individual number $m' < m$ of signatures found in each data set. This yields the corrected significance level $\alpha_* = 2 \cdot 10^{-4}$, which is typically more conservative than correcting individually for m' tests. This procedure allows us to use a single significance spectrum for all 81 SIP models with the same firing rates, differing by parameters z and c only, and for all 1000 realizations of each model. To obtain the p-values with precision α_* we generate $K = 1/\alpha_* = 5.000$ surrogates, compute their binary spectra and average them to draw the p-value spectrum (see Section 2.2).

Figure 4 shows significance spectra obtained from surrogate data for models differing by the firing rate r (5, 10 or 20 Hz) analysed with different bin widths w (dark gray for $w = 3$ ms, light gray for $w = 5$ ms; $\alpha_* = 2 \cdot 10^{-4}$). The set of non-significant signatures shows a hyperbolic shape, which grows with both r and w to higher z and higher c. Both factors, higher firing rates and larger bin width, cause more spikes per bin, and therefore larger and more frequent chance patterns.

Table 1 | Parameters for calibration of the method.

Simulation parameters		Analysis parameters	
Background activity	**SIP**	**FIM**	**Statistical tests**
$N = 100$	$z = 2, \ldots, 10$	$w = 3, 5$ ms	$\alpha_* = 2 \cdot 10^{-4}$
$r = 5, 10, 20$ Hz	$c = 2, \ldots, 10$	$c_0 = 2$	$K = 5000$
$T = 3$ sec		$z_0 = 2$	$R = 1000$

Parameters for the background activity: N: number of neurons, r: firing rate, T: simulation time. Parameters for correlated data: z: number of neurons in correlated activity (size of SIP), c: SIP occurrences. Analysis parameters: w: bin width, c_0: minimum item set support, z_0: minimum item set size. Statistical parameters: α_*: Bonferroni-corrected significance level (for m = 50 tests), K: number of surrogates, R: number of simulation runs per SIP model.

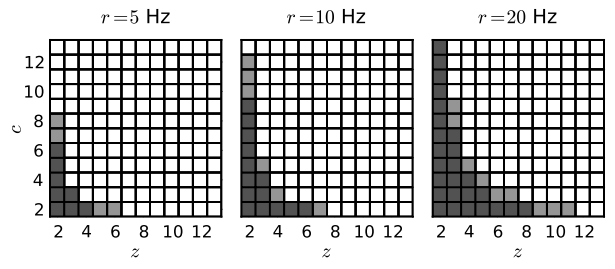


FIGURE 4 | Significance spectra for different parameter sets. Independent Poisson spike trains ($N = 100$; $T = 3$ s) of different firing rates ($r = 5$, 10 or 20 Hz) serve as surrogates for the computation of three significance spectra (from left to right). Each square represents a (z, c) signature. Dark-shaded gray squares mark non-significant signatures obtained with $w = 3$ ms. Light-shaded squares represent further non-significant signatures for $w = 5$ ms. White squares indicate significant signatures for both choices of the bin width. Other parameters: $z_0 = 2$, $c_0 = 2$, $\alpha_* = 2 \cdot 10^{-4}$, $K = 5000$.

4.4. PERFORMANCE, HOMOGENEOUS FIRING RATES

For each SIP parameter set we simulate the corresponding model $R = 1000$ times, and evaluate FPs and FNs of each realization. Their averages measure the performance of the analysis for each parameter constellation.

As previously discussed (Section 3), in the presence of correlations PSF tends to classify chance subsets, supersets or overlapping sets as significant, thus yielding FPs. **Figure 5**, top row, shows this effect on simulations of SIP models differing by SIP size (x-axis of each panel) and injection count (y-axis). For each model, the FP level is computed as an average over 1000 stochastic simulations. The total amount of FPs increases as the SIP size and/or the number of injections get larger. The contribution of FP supersets (green) and FP subsets (blue) is about the same, while in comparison FP overlapping sets (yellow) occur only at higher values for z and c, and FP disjoint patterns (purple) are almost never observed. As shown in **Figure 5**, bottom row, PSR (here, combined filtering) largely reduces the amount of FPs. Although the PSR statistical tests apply to chance subsets (blue) and supersets (green) only (Section 3.2), they successfully remove most of the overlapping patterns (yellow) as well. The reason is that, if there is a CFIS D overlapping with the actual assembly A by z_0 or more items, their intersection B is a CFIS as well (**Figure 3C**). In most cases PSF classifies B as significant together with A and D. If so, PSR likely rejects D when testing $H_0^{D|B}$, and rejects B when testing $H_0^{B|A}$.

A reduction of the amount of FPs typically comes at the expense of enhanced FNs. In particular, FNs may occur if the real pattern is rejected in favor of one of its subsets or supersets. **Figure 6** shows, for a range of combinations of SIP size and injection count, the resulting level of FPs, FNs, and the maximum of the two (as a measure of overall performance) after performing each of the proposed PSR strategies. The significance spectrum used to determine significance for all realizations of the SIP models is the one for $w = 3$ ms shown in **Figure 4** (top right, dark-shaded entries). For the FPs shown in **Figure 6**, top row, the color-coded level refers to the fraction of simulations (out of 1000) containing one or more FPs. This measure takes values between 0 and 1, unlike the average FP counts shown in **Figure 5**. This representation simplifies the comparison with the average FN level, which ranges between 0 to 1 since here only a single spike pattern is injected in every simulation. To aid the comparison between the performances of PSF and PSR, gray dots mark those squares that correspond to models where the error rates exceed 5%. PSF on its own never performs well in terms of FNs and FPs simultaneously, while all PSR strategies yield a range of models for which both quantities are low. In summary, the relative improvement of PSR versus PSF shows that any PSR strategy reduces the FP rate considerably, while causing only a minor increase in the FN rate.

4.5. PERFORMANCE, HETEROGENEOUS FIRING RATES

If neurons have the same spiking statistics, the spike pattern statistics depends on the pattern size only. Thus, the p-value of

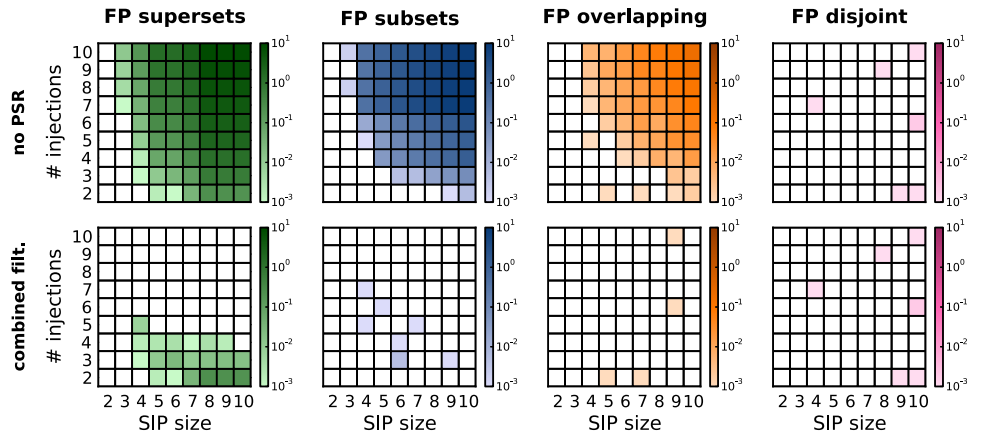


FIGURE 5 | Average number of FPs, distinguished by type, after PSF and PSR. Average number of FPs obtained for different SIP models on $R = 1000$ model simulations. FPs are shown after performing PSF (top) and then PSR with combined filtering (bottom), and are distinguished by type (columns from left to right: FP supersets, FP subsets, FP overlapping, FP disjoint patterns). Each panel shows the average number of FPs obtained for different SIP models, each corresponding to a square in the grid: the models differ by the SIP size (from 2 to 10; *x*-axis) and its injection count (from 2 to 10; *y*-axis). Other parameters (same for all simulations): $N = 100$, $T = 3\,\mathrm{s}$, $r = 20\,\mathrm{Hz}$, $w = 3\,\mathrm{ms}$, $K = 5000$, $\alpha_* = 2 \cdot 10^{-4}$.

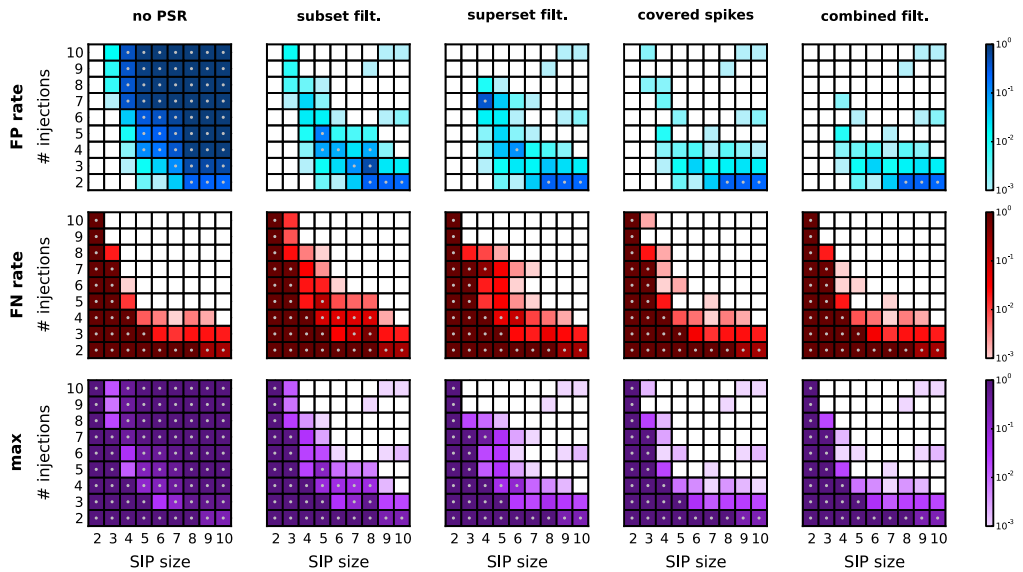


FIGURE 6 | Performance of PSR with homogeneous, stationary firing rates. Performance of PSR with different filtering methods, measured as the fraction of $R = 1000$ simulations where FPs (top row) and FNs (second row) are detected (thus the fraction represents a rate). The maximum of the two (third row) indicates the combined error rate. Each matrix shows the performance for 81 different SIP models varying by SIP size (from 2 to 10, *x*-axis) and number of SIP injections (from 2 to 10, *y*-axis), of stationary and homogeneous neuronal firing rates ($r = 20\,\mathrm{Hz}$). The performance value is color-coded (see color bar, logarithmic scale). White squares mark SIP models where no simulations led to false outcomes. Gray dots mark entries where the error rate is above 5%. Each column corresponds to a different PSR strategy applied after PSF, from left to right: no filtering, subset filtering, superset filtering, covered-spikes criterion, combined filtering. Other parameters (same for all panels): $N = 100$, $T = 3\,\mathrm{s}$, $w = 3\,\mathrm{ms}$, $K = 5.000$, $\alpha_* = 2 \cdot 10^{-4}$.

each pattern is fully determined by the pattern signature. This does not hold when neurons have different spiking statistics, and in particular different firing rates. Here we discuss the case of heterogeneous firing rates across neurons, which are often present in electrophysiological data. Higher firing rates lead to a higher spiking probability per time bin. Patterns composed of neurons with higher firing rate are more likely to occur by chance, and are therefore less significant than patterns composed of neurons with lower rates. Thus, the *p*-values of patterns with the same signature (z, c) differ for different compositions of the firing rates. Pooling patterns by size and support in the pattern spectrum does not take into account the heterogeneity of firing rates across neurons and thus may lead to a biased statistics.

To investigate the robustness of our method against firing rate heterogeneity, we first simulate independent data consisting of 100 neurons, with a small population of neurons (2 to 10) firing at a higher rate (20 Hz) than the rest of the neurons (5 Hz). We simulate 1000 data sets of this type, and evaluate FPs in each of them by means of FIM and PSF ($K = 5000$ surrogates). In none of the simulations we detect significant signatures, i.e., FPs. The opposite scenario, where 2 to 10 neurons fire at 5 Hz and the others at 20 Hz, does not yield FPs as well. Thus, employing rate-preserving surrogates allows PSF to correctly estimate the significance of signatures under $\mathcal{H}_0$, also when rates are heterogeneous across neurons.

Next we study correlated data characterized by heterogeneous background firing rates. We investigate two cases based on a SIP model. In scenario S1, a pattern is injected in a set of neurons firing with lower firing rate ($r_S = 5$ Hz) than the independent neurons firing at rate $r_I = 20$ Hz (**Figure 7**, left column). In contrast, in scenario S2 the pattern is injected in neurons with higher firing rates ($r_S = 20$ Hz, $r_I = 5$ Hz; **Figure 7**, right column). In comparison to the homogeneous case where all neurons fire at 5 Hz (data not shown), the overall performance drops significantly, but does not so compared to the 20 Hz homogeneous case (see **Figure 6**, right column). This is consistent with the previous finding that higher rates worsen the performance by shifting the detection border in the significance spectrum to the right (**Figure 4**, left vs. right). This also explains why FP and FN rates in scenario S1 are higher than in scenario S2: the average firing rate in the former ranges (depending on the SIP model) from 18.5 to 19.7 Hz, in the latter from 5.3 to 7 Hz. Our choice of using PSR with combined filtering leads to a better performance in this scenario than the covered spikes criterion (not shown). Taken together, these results indicate that the method can deal well with heterogeneity of firing rates without severe performance loss.

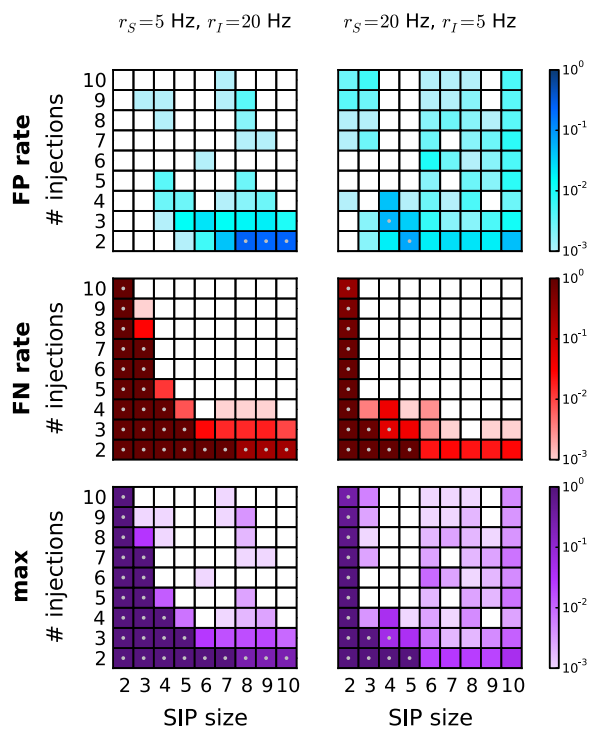


FIGURE 7 | Performance of PSR with heterogeneous firing rates. Performance of PSR (combined filtering with parameters $h = 1$, $k = 2$) in terms of FP rates (*top* row), FN rates (middle row) and the combined error rates (maximum of FP and FN rates) (*bottom* row) of data with heterogeneous rates. Left column: SIP neurons fire at $r_S = 5$ Hz, independent neurons fire at $r_I = 20$ Hz. Gray dots mark entries where the error rate is above 5%. Right column: SIP neurons fire at $r_S = 20$ Hz and independent neurons at $r_I = 5$ Hz. Other parameters (same for all panels): $N = 100$, $T = 3$ s, $w = 3$ ms, $K = 5.000$, $\alpha_* = 2 \cdot 10^{-4}$.

4.6. PERFORMANCE, NON-STATIONARY FIRING RATES

Now we want to consider the case when the firing rates of the neurons are not stationary in time. To explore the sensitivity of our method to non-stationarities we employ simulated data, again consisting of 100 parallel spike trains, which fire in two consecutive epochs of length T_1 and T_2 (the total simulation time $T = T_1 + T_2$ is 3 s, as in the data previously analysed) at different rates ($r_1 = 5$ Hz and $r_2 = 20$ Hz; or vice versa), homogeneously across the neurons in both epochs. In the first epoch, correlated activity is inserted by the SIP model. SIP of size 2 to 10, injected 2 to 7 times, amount to a coincidence rate of 1.33 to 4.66 Hz in the first epoch. The background rate is reduced correspondingly. For comparison, we also study the stationary case, where all neurons fire at $r = 10$ Hz. The performance for the three scenarios is shown in **Figure 8** (first column: $r_1 = 5$ Hz, $r_2 = 20$ Hz; second column: $r_1 = 20$ Hz, $r_2 = 5$ Hz; third column: $r_{1,2} = 10$ Hz). Although our analysis performs better (detection border more to the left) in the stationary case ($r = 10$ Hz; third column), it can still recover SIP activity with no FPs in a large portion of the parameter space, provided that rate-preserving surrogates are employed. As in the heterogeneous case, FPs increase when the SIP neurons have higher firing rates and thus more FP subsets occur. As apparent from **Figure 8**, bottom row, the method can correctly detect significant patterns in a wide range of models also in the presence of non-stationary rates. To study whether short transients in the firing rates tend to generate FPs, we repeated the analysis for $T_1 = 0.5$ s, $T_2 = 2.5$ s, setting first $r_1 = 5$ Hz, $r_2 = 20$ Hz and then $r_1 = 20$ Hz, $r_2 = 5$ Hz. In all cases we do not find enhanced FPs (data not shown), indicating that employing rate-preserving surrogates suffices to correct for rate non-stationarity in independent data.

5. DISCUSSION

In this study we have presented a method to detect significant patterns of synchronous spiking in a subset of massively parallel spike trains in the presence of background activity. Our work is rooted in Picado-Muiño et al. (2013), where we demonstrated how to efficiently detect spike patterns in such data, and assess their significance under the null hypothesis of independent firing. Here we refined this significance test, which evaluates the significance of patterns using PSF on the basis of the pattern signature (size and support). PSF is prone to report FP patterns that arise due to the activation of an actual assembly mixed with chance synchrony because of background activity. To identify and remove these FP

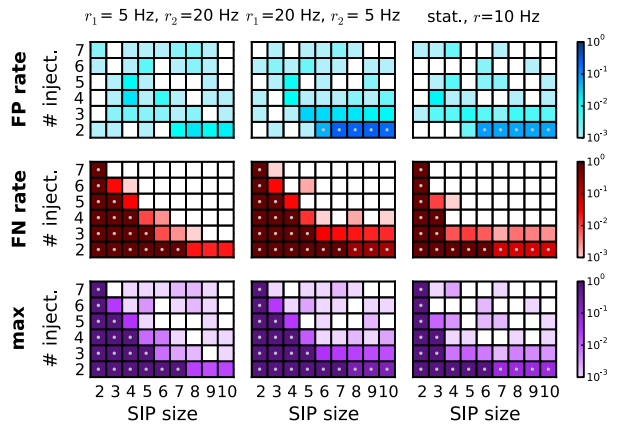


FIGURE 8 | Performance of PSR with non-stationary firing rates. Performance of PSR (combined filtering) in terms of FP rate (top row), FN rate (middle row), and maximum of the two (bottom row), computed over $R = 1000$ simulations per SIP model. Each panel shows the performance for 54 different models varying by SIP size (from 2 to 10, *x*-axis) and SIP injections (from 2 to 7, *y*-axis). In the first two columns, the simulations consist of two epochs. The first epoch of duration $T_1 = 1.5$ s is composed of a stationary, homogeneous SIP model (with firing rate $r = r_1$), followed by a second epoch of $T_2 = 1.5$ s with independent spiking at a rate of $r_2 \neq r_1$. In the first column, the rate compositions are $r_1 = 5$ Hz and $r_2 = 20$ Hz, and in the second column $r_1 = 20$ Hz and $r_2 = 5$ Hz. For comparison, the third column shows the performance for the stationary case with all neurons firing at rate $r = 10$ Hz, and a duration of $T = 3$ s. Other parameters (same for all panels): $N = 100$, $w = 3$ ms, $K = 5.000$, $\alpha_* = 2 \cdot 10^{-4}$. Gray dots mark entries where the error rate is above 5%.

detections, we introduced here PSR as an additional statistical testing step. As shown in **Figure 6** (second to last columns), PSR succeeds in eliminating FPs for a wide range of parameters, at the expense of a minor increase in FNs. A series of calibrations demonstrates the effectiveness of our approach under conditions of heterogeneous and non-stationary firing rates.

The relevance of higher-order correlations for information processing in the nervous system is hotly debated. Approaches based on maximum entropy models, such as Schneidman et al. (2006), suggest that higher-oder correlations contribute by a negligible fraction to the total network correlation, which appears to be dominated by pairwise correlations. However, it is important to stress that for correlations of a specific order, maximum entropy models estimate the overall magnitude of that correlation order, and are not sensitive to individual correlation structures of that order. Thus, the presence of a single group of correlated neurons with a certain size in the data is not enough for maximum entropy models to report significant correlation of the corresponding order. The study by Shlens et al. (2006) addresses this point, discussing that maximum entropy models may miss higher-order correlations because they overall contribute only by a negligible fraction to the total correlation. Besides, Roudi et al. (2009) showed that the statistical power of maximum entropy models describing spike correlations in heavily undersampled biological systems (such as parallel recordings with electrode arrays) is low. Despite these challenges, Ohiorhenuan et al. (2010) have shown using a maximum entropy model approach that in visual cortex local microcircuits exhibit evidence of higher-order interactions, whereas correlation statistics across long-range connections are explained on the basis of pair-wise interactions. However, methods designed to investigate individual spike patterns are needed to investigate the detailed structure of correlation in groups of spiking neurons.

A majority of current methods for spike correlation analysis limit themselves to fully synchronous patterns or to patterns with a specific size of typically low order [e.g., Grün et al., 2002a,b; Berger et al., 2007, 2010; Shimazaki et al., 2012]. Other approaches, such as CuBIC (Staude et al., 2010b), conclude on the presence of higher order correlations based on the statistics of the population activity without identifying the specific units engaged in such correlations. While Gansel and Singer (2012) presented a method for the detection of higher-order patterns, they identify pattern subsets by a purely heuristic procedure that is not accessible by analytic treatment, and that tests patterns directly, which requires a number of statistical corrections to avoid FPs (at the expense of FNs). Our proposed method instead first tests the significance of pattern signatures. PSF eliminates non-significant signatures based on surrogate data through the significance spectrum (see **Figure 4**), and determines the class $\mathcal{P}$ of associated significant patterns. Testing patterns on the basis of their signature rather than testing individual patterns reduces the number of required statistical tests to the number of signatures found in the data. We have shown that the composition of assembly and background spikes typically leads to the identification of additional significant patterns (i.e., FPs). In order to remove this type of FPs, we introduced here the PSR procedure that is based on conditional pairwise tests.

We have tested the performance of our analysis on artificial data where we embedded groups of synchronously spiking neurons in background activity of independent Poisson spike trains [SIP, cf. Kuhn et al. (2003)]. We studied the rate of FP and FN detections for occurrence rates of the synchronous pattern varying from 0.66 to 3.33 Hz, which reflect plausible values for the activation frequency of the assumed assemblies (Grün et al., 1999; Denker et al., 2010). The analysis shows in particular that by introducing PSR, assembly detection becomes possible with near perfect reliability and precision for a large range of SIP parameters. The transition shifts toward higher support and assembly size as the bin width or the firing rates increase (cf., **Figure 4**). Nevertheless, for physiologically realistic parameters only for very small or very infrequent SIP injections these patterns cannot be distinguished from chance synchrony. Moreover, evaluating patterns obtained from a larger set of simultaneously recorded neurons will have only minor impact on our findings due to a slight increase in the average size of observed patterns.

Non-stationarities of the firing rate in time or across neurons are a common concern faced by correlation analysis methods. The effect of non-stationary firing rates on PSF is two-fold. First, the surrogates used to calculate the significance estimates on pattern signatures should adequately reproduce the experimental rate profiles. Even if the underlying rate profile is not known, a variety of suitable approaches for surrogate generation is available for this task (Grün, 2009; Louis et al., 2010b). However, the sensitivity of

detecting assembly activations is further affected by where these occur with respect to the rate non-stationarity. In this respect we tested the performance of PSF and PSR in a scenario of step-wise non-stationary firing rates where spike patterns were injected at selected rate levels only. Compared to the stationary case, the method retains a high performance for large parameter regimes (**Figure 8**), and shows only a slight increase in the number of FNs. For very large rate non-stationarities, a time-resolved analysis may be used to additionally aid the detection, as done, e.g., in the Unitary Events analysis (Grün et al., 2002b). In a similar framework, we found that also heterogeneous firing rates across neurons (**Figure 7**) exhibit a performance similar to the stationary case. While we see minor increases in the number of FPs, we remark that to a large extent these are indeed supersets of the injected pattern due to the high probability of gaining an additional coincident spike by chance from the set of neurons spiking at high rates.

In this study we assumed that assemblies occur at the time resolution of the data, i.e., that spike times of the assemblies are not jittered in time. In electrophysiological data this is a rare scenario, and instead spike synchrony typically occurs with a temporal jitter of up to several milliseconds [Grün et al., 1999; Pazienti et al., 2008]. In order to capture such slightly imprecise synchrony, exclusive binning is typically applied (Grün et al., 1999), where the bin width is chosen large enough to capture the jittered spike pattern. However, the spikes of the pattern may be split into adjacent bins with a probability that depends on the jitter, bin size, and pattern size. Therefore, the original synchronous events are destroyed, leading to increased FN rates (Grün et al., 1999). In **Figure 9** we show how this effect can have a substantial impact on the performance of the method. We applied PSF followed by PSR (combined filtering) on data where synchronous patterns are injected with a jitter of ± 1 ms, and analysed with a bin width of $w = 3$ ms (left column) and $w = 5$ ms (right column). The performance drops considerably due to an increase of the FP rate for higher z and c, and an overall increase of the FN rate. The performance is slightly better for a bin width of 5 ms. Consistent with these findings, Grün et al. (1999) showed that for two parallel spike trains about 60% of the synchronous events are lost if the bin width corresponds to the jitter width. An earlier modification of exclusive time binning [multiple shift method, Grün et al., 1999] that avoids the splitting of jittered synchrony was not trivially applicable to large numbers of parallel spike trains. In Picado-Muiño et al. (submitted) we demonstrate how to implement a method for pattern detection based on the inter-spike distances rather than discrete time binning. This approach successfully detects jittered spike patterns and therefore trivially exhibits a performance in the context of PSF that is similar to that achieved in the absence of jitter (see Picado-Muiño et al., submitted, for details). Thus, it also complements the PSR framework presented in this study. Therefore, we suggest to detect jittered synchrony by the continuous detection method and perform the analysis by the proposed sequence of FIM, PSF, and PSR.

A further scenario that remains to be addressed in the future is unreliability in spiking activity that causes neurons to selectively skip participation in assembly activations. This scenario

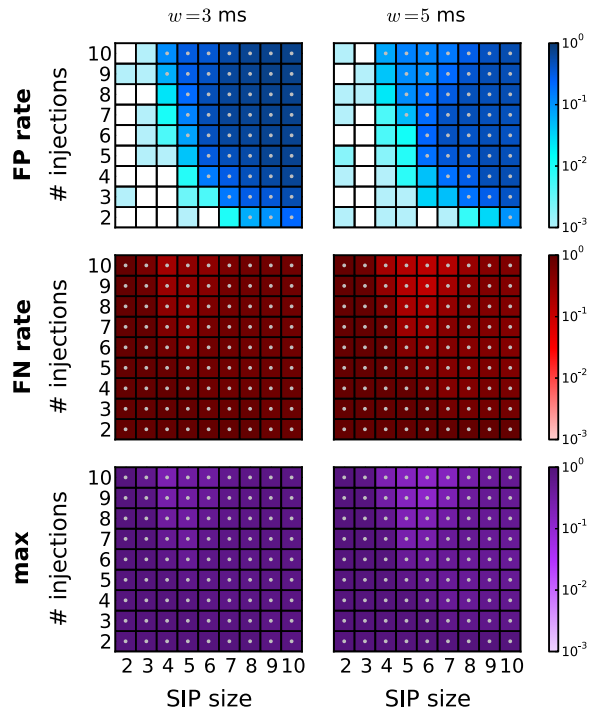


FIGURE 9 | Performance of PSR under jittered synchrony. Performance of PSR (combined filtering with $h = 1$, $k = 2$) on data from SIP models with jittered synchrony. The spikes of SIP events are randomly jittered up to ± 1 ms around the original occurrence time. The performance is shown in terms of FP rates (first row), FN rates (middle row) and maximum of the two (bottom row) for different bin widths: $w = 3$ ms (left column) and $w = 5$ ms (right column).

was discussed in the context of the synfire chain model, where it was shown that stable propagation of synchronous spike packages through the network happens reliably although the probability that individual neurons participate in each activation of the synfire chain is lower than 1 (Diesmann et al., 1999). Selective participation may arise as a consequence of synaptic failure. The multiple interaction process [MIP; Kuhn et al., 2003] was proposed as a stochastic model implementing such a behavior. Our method would interpret the variable composition of spikes in a single MIP event as occurrences of multiple SIP events of lower support.

We conclude with a discussion of the practical implementation of the proposed analysis on data from electrophysiological recordings. Given a set of parallel spike recordings obtained at a resolution (i.e., binning) w, we choose the minimum pattern size z_0 and the minimum pattern support c_0 of the analysis. First, the spike data is binned and, using FIM, the CFISs and the corresponding pattern signatures are obtained from the transaction list. While this approach is feasible for the experimental data available today, with several hundreds of parallel recordings the computational effort may become too large. In this scenario, we suggest to pre-filter the data entering the analysis as suggested by Berger et al. (2010) before applying FIM on the reduced set of neurons. To monitor dynamic changes in the correlation structure

of the activity, e.g., if assemblies are time locked to a particular behavioral event, one may choose to additionally perform the analysis in sliding windows.

Next, the significance of the observed patterns is evaluated by PSF under the null-hypothesis of full independence implemented by uncorrelated surrogate data. For experimental data, several techniques for surrogate generation based on stochastic sampling have been proposed in the past [for a review, see Grün, 2009]. Surrogates that preserve the firing rate profiles, such as spike dithering, seem most appropriate since PSF determines pattern significance based on the firing rates. Given the significance level α and m detected pattern signatures, a minimum of $K = \lceil m/\alpha \rceil$ surrogates are required to achieve the Bonferroni-corrected significance level $\alpha_* = \alpha/m$. Once the surrogates have been generated, we follow the procedure described for the simulated data. CFISs, pattern signatures and the resulting binary pattern spectrum are obtained for each surrogate run. Next, the p-value spectrum is obtained as an average of the binary spectra (see Section 2.2). The signatures whose p-values do not exceed the Bonferroni-corrected significance level α_* are marked as significant, and the CFISs of significant signatures are collected into the class $\mathcal{P}$ of potential assemblies. Finally, PSR with combined filtering is performed to reduce $\mathcal{P}$ to a subclass $\mathcal{Q}$ of patterns which are mutually significant with respect to each other.

In summary, the use of FIM combined with the statistical tests described in this study and in Picado-Muiño et al. (submitted) represents a powerful tool to extract candidate assemblies from experimental data. The method is statistically rigid, computationally feasible, robust against heterogeneity in the data, and powerful enough to deal with the limited amount of data typically available from electrophysiological experiments. We expect that our approach will help to reveal how precise spike synchronization observed by pairwise analysis in relation to behavior (Riehle et al., 1997) is manifested at the level of neuronal populations.

ACKNOWLEDGMENTS

We thank Günter Palm for stimulating discussions. This work was partially supported by the Helmholtz Alliance on Systems Biology, BrainScales (EU Grant 269912), the Helmholtz Portfolio Supercomputing and Modelling for the Human Brain (SMHB), and the Spanish Ministry for Economy and Competitiveness (MINECO Grant TIN2012-31372).

SOFTWARE AND SUPPLEMENTAL MATERIAL

The FIM library underlying the Python scripts with which we carried out our experiments is available at http://www.borgelt.net/pyfim.html. Python and shell scripts for related experiments as well as more extensive result diagrams are available at http://www.borgelt.net/accfim.html and http://www.borgelt.net/cocofim.html. Please also consult http://www.spiketrain-analysis.org for these codes and further information on the analysis of parallel spike trains.

REFERENCES

Abeles, M. (1982). Role of cortical neuron: integrator or coincidence detector? *Israel J. Med. Sci.* 18, 83–92.

Abeles, M. (1991). *Corticonics: Neural Circuits of the Cerebral Cortex* (1st Edn). Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511574566

Abeles, M., and Gerstein, G. L. (1988). Detecting spatiotemporal firing patterns among simultaneously recorded single neurons. *J. Neurophysiol.* 60, 909–924.

Berger, D., Borgelt, C., Louis, S., Morrison, A., and Grün, S. (2010). Efficient identification of assembly neurons within massively parallel spike trains. *Comput. Intell. Neurosci.* 2010:439648. doi: 10.1155/2010/439648

Berger, D., Warren, D., Normann, R., Arieli, A., and Grün, S. (2007). Spatially organized spike correlation in cat visual cortex. *Neurocomputing* 70, 2112–2116. doi: 10.1016/j.neucom.2006.10.141

Borgelt, C. (2012). "Frequent item set mining," in *Wiley Interdisciplinary Reviews (WIREs): Data Mining and Knowledge Discovery*, Vol. 2, J. (Chichester: Wiley and Sons), 437–456. doi:10.1002/widm.1074

Buzsaki, G. (2004). Large-scale recording of neuronal ensembles. *Nat. Neurosci.* 7, 446–451. doi: 10.1038/nn1233

Denker, M., Riehle, A., Diesmann, M., and Grün, S. (2010). Estimating the contribution of assembly activity to cortical dynamics from spike and population measures. *J. Comput. Neurosci.* 29, 599–613. doi: 10.1007/s10827-010-0241-8

Diesmann, M., Gewaltig, M.-O., and Aertsen, A. (1999). Stable propagation of synchronous spiking in cortical neural networks. *Nature* 402, 529–533. doi: 10.1038/990101

Feldt, S., Waddell, J., Hetrick, V. L., Berke, J. D., and Zochowski, M. (2009). Functional clustering algorithm for the analysis of dynamic network data. *Phys. Rev. E* 79:056104. doi: 10.1103/PhysRevE.79.056104

Fujisawa, S., Amarasingham, A., Harrison, M. T., and Buzsáki, G. (2008). Behavior-dependent short-term assembly dynamics in the medial prefrontal cortex. *Nat. Neurosci.* 11, 823–833. doi: 10.1038/nn.2134

Gansel, K. S., and Singer, W. (2012). Detecting multineuronal temporal patterns in parallel spike trains. *Front. Neuroinformatics* 6:18. doi: 10.3389/fninf.2012.00018

Gerstein, G. L., and Aertsen, A. M. H. J. (1985). Representation of cooperative firing activity among simultaneously recorded neurons. *J. Neurophysiol.* 54, 1513–1528.

Gerstein, G. L., Perkel, D. H., and Subramanian, K. N. (1978). Identification of functionally related neural assemblies. *Brain Res.* 140, 43–62. doi: 10.1016/0006-8993(78)90237-8

Gerstein, G. L., Williams, E. R., Diesmann, M., Grün, S., and Trengove, C. (2012). Detecting synfire chains in parallel spike data. *J. Neurosci. Methods* 206, 54–64. doi: 10.1016/j.jneumeth.2012.02.003

Goethals, M. D. (2010). *Data Mining and Knowledge Discovery Handbook* (2nd Edn.) Chapter Frequent Set Mining. Berlin: Springer Verlag, 321–338 doi: 10.1007/978-0-387-09823-4_16

Grün, S. (2009). Data-driven significance estimation of precise spike correlation. *J. Neurophysiol.* 101, 1126–1140. doi: 10.1152/jn.00093.2008

Grün, S., Diesmann, M., and Aertsen, A. (2002a). 'Unitary Events' in multiple single-neuron spiking activity. I. Detection and significance. *Neural Comput.* 14, 43–80. doi: 10.1162/089976602753284455

Grün, S., Diesmann, M., and Aertsen, A. (2002b). 'Unitary Events' in multiple single-neuron spiking activity. II. Non-Stationary data. *Neural Comput.* 14, 81–119. doi: 10.1162/089976602753284464

Grün, S., Diesmann, M., Grammont, F., Riehle, A., and Aertsen, A. (1999). Detecting unitary events without discretization of time. *J. Neurosci. Methods* 94, 67–79. doi: 10.1016/S0165-0270(99)00126-0

Harris, K. (2005). Neural signatures of cell assembly organization. *Nat. Rev. Neurosci.* 5, 339–407.

Hatsopoulos, N. G., Xu, Q., and Amit, Y. (2007). Encoding of movement fragments in the motor cortex. *J. Neurosci.* 27, 5105–5114. doi: 10.1523/JNEUROSCI.3570-06.2007

Hebb, D. O. (1949). *The Organization of Behavior: A Neuropsychological Theory*. New York, NY: John Wiley and Sons.

Humphries, M. D. (2011). Spike-train communities: finding groups of similar spike trains. *J. Neurosci.* 31, 2321–2336. doi: 10.1523/JNEUROSCI.2853-10.2011

Kohn, A., and Smith, M. A. (2005). Stimulus dependence of neuronal correlations in primary visual cortex of the Macaque. *J. Neurosci.* 25, 3661–3673. doi: 10.1523/JNEUROSCI.5106-04.2005

König, P., Engel, A. K., and Singer, W. (1996). Integrator or coincidence detector? The role of the cortical neuron revisited. *TINS* 19, 130–137. doi: 10.1016/S0166-2236(96)80019-1

Kuhn, A., Aertsen, A., and Rotter, S. (2003). Higher-order statistics of input ensembles and the response of simple model neurons. *Neural Comput.* 15, 67–101. doi: 10.1162/089976603321043702

Louis, S., Borgelt, C., and Grün, S. (2010a). Complexity distribution as a measure for assembly size and temporal precision. *Neural Netw.* 23, 705–712. doi: 10.1016/j.neunet.2010.05.004

Louis, S., Gerstein, G. L., Grün, S., and Diesmann, M. (2010b). Surrogate spike train generation through dithering in operational time. *Front. Comput. Neurosci.* 4:127. doi: 10.3389/fncom.2010.00127

Masud, M., and Borisyuk, R. (2011). Statistical technique for analysing functional connectivity of multiple spike trains. *J. Neurosci. Methods* 196, 201–219. doi: 10.1016/j.jneumeth.2011.01.003

Nicolelis, M. A. L. (Ed.) (1998). *Methods for Neural Ensemble Recordings*. Boca Raton, FL: CRC Press. doi: 10.1201/9781420048254

Ohiorhenuan, I. E., Mechler, F., Purpura, K. P., Schmid, A. M., Hu, Q., and Victor, J. D. (2010). Sparse coding and high-order correlations in fine-scale cortical networks. *Nature* 466, 617–621. doi: 10.1038/nature09178

Pazienti, A., Maldonado, P. E., Diesmann, M., and Grün, S. (2008). Effectiveness of systematic spike dithering depends on the precision of cortical synchronization. *Brain Res.* 1225, 39–46. doi: 10.1016/j.brainres.2008.04.073

Picado-Muiño, D., Borgelt, C., Berger, D., Gerstein, G. L., and Grün, S. (2013). Finding neural assemblies with frequent item set mining. *Front. Neuroinform.* 7:9. doi: 10.3389/fninf.2013.00009

Pipa, G., Wheeler, D. W., Singer, W., and Nikolic, D. (2008). Neuroxidence: reliable and efficient analysis of an excess or deficiency of joint-spike events. *J. Neurosci. Methods* 25, 64–88.

Riehle, A., Grün, S., Diesmann, M., and Aertsen, A. (1997). Spike synchronization and rate modulation differentially involved in motor cortical function. *Science* 278, 1950–1953. doi: 10.1126/science.278.5345.1950

Riehle, A., Wirtssohn, S., Grün, S., and Brochier, T. (2013). Mapping the spatio-temporal structure of motor cortical lfp and spiking activities during reach-to-grasp movements. *Front. Neural Circuits* 7:48. doi: 10.3389/fncir.2013.00048

Roudi, Y., Nirenberg, S., and Latham, P. E. (2009). Pairwise maximum entropy models for studying large biological systems: When they can work and when they canâĂŹt. *PLos Comput. Biol.* 5:e1000380. doi: 10.1371/journal.pcbi.1000380

Schneidman, E., Berry, M. J., Segev, R., and Bialek, W. (2006). Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature* 440, 1007–1012. doi: 10.1038/nature04701

Schrader, S., Grün, S., Diesmann, M., and Gerstein, G. (2008). Detecting synfire chain activity using massively parallel spike train recording. *J. Neurophysiol.* 100, 2165–2176. doi: 10.1152/jn.01245.2007

Shimazaki, H., Amari, S.-i., Brown, E. N. B., and Grün, S. (2012). State-space analysis of time-varying higher-order spike correlation for multiple neural spike train data. *PLoS Comput. Biol.* 8:e1002385. doi: 10.1371/journal.pcbi.1002385

Shlens, J., Field, G. D., Gauthier, J. L., Grivich, M. I., Petrusca, D., Sher, A., et al. (2006). The structure of multi-neuron firing patterns in primate retina. *J. Neurosci.* 26, 8254–8266. doi: 10.1523/JNEUROSCI.1282-06.2006

Singer, W., Engel, A. K., Kreiter, A. K., Munk, M. H. J., Neuenschwander, S., and Roelfsema, P. R. (1997). Neuronal assemblies: necessity, signature and detectability. *Trends Cogn. Sci.* 1, 252–261. doi: 10.1016/S1364-6613(97)01079-6

Staude, B., Grün, S., and Rotter, S. (2010a). Higher-order correlations in non-stationary parallel spike trains: statistical modeling and inference. *Front. Comput. Neurosci.* 4:16. doi: 10.3389/fncom.2010.00016

Staude, B., Rotter, S., and Grün, S. (2010b). Cubic: cumulant based inference of higher-order correlations in massively parallel spike trains. *J. Comput. Neurosci.* 29, 327–350. doi: 10.1007/s10827-009-0195-x

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.

Received: 31 May 2013; accepted: 11 September 2013; published online: 23 October 2013.

Citation: Torre E, Picado-Muiño D, Denker M, Borgelt C and Grün S (2013) Statistical evaluation of synchronous spike patterns extracted by frequent item set mining. Front. Comput. Neurosci. 7:132. doi: 10.3389/fncom.2013.00132

This article was submitted to the journal Frontiers in Computational Neuroscience.

Copyright © 2013 Torre, Picado-Muiño, Denker, Borgelt and Grün. 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) or licensor 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.

frontiers in
COMPUTATIONAL NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 21 October 2013
doi: 10.3389/fncom.2013.00139

Correlations in background activity control persistent state stability and allow execution of working memory tasks

Mario Dipoppa [1,2]*† ***and Boris S. Gutkin*** [1,3]*

[1] *Departement d'Etudes Cognitives, Ecole Normale Superieure, Group for Neural Theory, Laboratoire des Neurosciences Cognitives INSERM U960, Paris, France*
[2] *Ecole Doctorale Cerveau Cognition Comportement, Université Pierre et Marie Curie, Paris, France*
[3] *Centre national de la recherche scientifique, Paris, France*

Edited by:
Robert Rosenbaum, University of Pittsburgh, USA

Reviewed by:
Carson C. Chow, National Institutes of Health, USA
Zachary P. Kilpatrick, University of Houston, USA

***Correspondence:**
Mario Dipoppa and Boris S. Gutkin, Departement d'Etudes Cognitives, Ecole Normale Superieure, Group for Neural Theory, Laboratoire des Neurosciences Cognitives INSERM U960, 29 rue d'Ulm, 75005 Paris, France
e-mail: m.dipoppa@ucl.ac.uk; boris.gutkin@ens.fr

†Present address:
Mario Dipoppa, University College London, 21 University Street, London WC1E 6DE, UK

Working memory (WM) requires selective information gating, active information maintenance, and rapid active updating. Hence performing a WM task needs rapid and controlled transitions between neural persistent activity and the resting state. We propose that changes in correlations in neural activity provides a mechanism for the required WM operations. As a proof of principle, we implement sustained activity and WM in recurrently coupled spiking networks with neurons receiving excitatory random background activity where background correlations are induced by a common noise source. We first characterize how the level of background correlations controls the stability of the persistent state. With sufficiently high correlations, the sustained state becomes practically unstable, so it cannot be initiated by a transient stimulus. We exploit this in WM models implementing the delay match to sample task by modulating flexibly in time the correlation level at different phases of the task. The modulation sets the network in different working regimes: more prompt to gate in a signal or clear the memory. We examine how the correlations affect the ability of the network to perform the task when distractors are present. We show that in a winner-take-all version of the model, where two populations cross-inhibit, correlations make the distractor blocking robust. In a version of the mode where no cross inhibition is present, we show that appropriate modulation of correlation levels is sufficient to also block the distractor access while leaving the relevant memory trace in tact. The findings presented in this manuscript can form the basis for a new paradigm about how correlations are flexibly controlled by the cortical circuits to execute WM operations.

Keywords: correlations, background activity, working memory, spiking neural network, persistent activity

INTRODUCTION

Working memory (WM), defined as short term storage of information that is actively used on-line to carry out actions and decisions and drive learning, is one of the key processes that underpins our cognitive abilities. WM is characterized by an information bottleneck with resources restricting its "on-line" capacity to a relatively limited number of items at high levels of performance (Miller, 1956; Luck and Vogel, 1997; Cowan, 2001; Vogel et al., 2001) and a rapid decrease in performance with item number due to limited resource allocation (Wilken and Ma, 2004; Bays and Husain, 2008; van den Berg et al., 2012) as suggested by the recent experiments. Furthermore by its very nature, WM is characterized by the need to operate on the stored information rapidly. Such limitations and rapid operations of WM create the need for selective gating and rapid updating as well as active information maintenance to enable its immediate use (Frank et al., 2001). One of the central unresolved issues is how the multiple requirements for WM are carried out by the brain circuits: whether the maintenance, read-in, gating, and read-out are implemented by separated systems (e.g., as suggested by Baddeley, 2003) or by operations within the same neural circuit (e.g., as recently put forward by Machens et al., 2005).

Electrophysiological data from primate performing delayed-response tasks show that persistent neuronal activity in prefrontal cortex (PFC) underlies the maintenance of WM: during the delay period between the stimulus presentation and the read-out, neurons selective to the memorized stimulus fire spikes at an elevated rate with respect to the resting state (Fuster and Alexander, 1971; Fuster and Jervey, 1981; Funahashi et al., 1989; Miller et al., 1996; Romo et al., 1999).

In order to highlight the unique requirements of the WM as a neural process let us focus on the DMS task with distractors as a prototypical example (Miller et al., 1996). In this task the subject must remember the identity of an item briefly shown (the sample) and respond correctly only when the item is shown again (match) all the while ignoring other items flashed (distractors). To execute correctly this task, the neural circuitry needs to perform three operations (**Figure 1A**): first, encode and maintain in memory the sensory stimulus during the delay period; second, robustly maintain the memory face to distractors presentation; third, erase the memory trace at task completion to make the store available again, given the limited WM capacity. These operations are translated in terms of neural activity as follows: item-related activity is turned on rapidly and selectively by the sample-stimulus, is

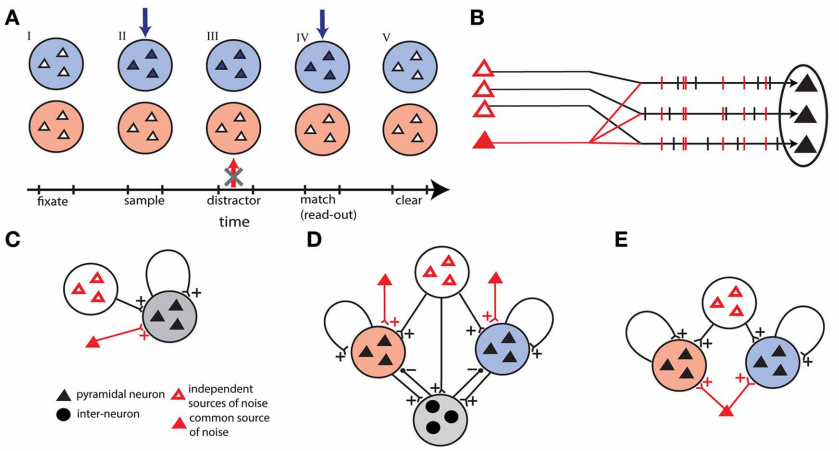


FIGURE 1 | Outline of the models. (A) Time sequence of the delay match-to-sample task for the working memory network. Active neurons are represented in full colors. Successively: (I) both populations are in a quiescent state, (II) sample stimulus (blue arrow) activates blue population, (III) the network prevent a distracting stimulus (red arrow) to activate the red population, (IV) match stimulus allows the read-out of the encoded memory in the blue population, and (V) persistent activity is erased in the blue population. **(B)** Correlations in external background activity generated by a common source of noise, in addition to independent sources of noise. **(C)** Single unit network receiving shared and independent sources of noise. **(D)** Winner-take-all network with two competing excitatory populations coupled through one inhibitory population. In addition to independent sources the excitatory populations receive background activity by two different common noise sources. **(E)** Two-unit network with two excitatory populations receiving shared noise.

protected from distractors during the delay period, and is rapidly turned off on response by the match.

A number of spiking network models have been conceived to describe the neural substrate for WM where persistent activity is maintained by recurrent connections that allow for co-existing attractor memory states and a ground non-memory state (Amit and Brunel, 1997; Compte et al., 2000; Brunel and Wang, 2001; Gutkin et al., 2001; Laing and Chow, 2001; Machens et al., 2005; Miller and Wang, 2006; Ardid et al., 2010). In some of these models, protection from distractors and memory clearance are performed through the recruitment of inhibition (Compte et al., 2000; Brunel and Wang, 2001; Machens et al., 2005). As an alternative to the erasing-by-inhibition paradigm, it has been shown, in a spatial WM model, that a transient excitatory stimulus matching the memory trace "location" on the network extinguishes the persistent state by transiently synchronizing the spike-times of the neurons (Gutkin et al., 2001; Laing and Chow, 2001). This work, along with Machens et al. (2005) showed how the read-out and clear-out can be merged into a single operation. However, in these alternative frameworks, protection from distractors, or selective gating, was not addressed. Here we propose that the gating is obtained by flexibly controlling the spike-time structure of the WM network activity. In support of this idea, it has been shown that spike-time synchronization is modulated in association with cognitive processing (Abeles et al., 1993; Riehle et al., 1997; Funahashi and Inoue, 2000) and in particular in WM (Sakurai and Takahashi, 2006; Pipa and Munk, 2011).

Critically, WM trace appears in the context of on-going background activity. While background activity is not related to task parameters, this is not without structure. Correlations have been found broadly in spontaneous neural activity in the cortex (Tsodyks et al., 1999). In particular, it has been shown that nearby neurons receive common inputs from afferent neurons making their voltages correlated (Lampl et al., 1999). Effects of correlations have been widely studied for their effect on population code (Salinas and Sejnowski, 2001), to measure network connectivity (Aertsen et al., 1989; Cocco et al., 2009), on neural dynamics for coupled neurons (Ly and Ermentrout, 2009), and for multiple independent neurons (Galán et al., 2006; Moreno-Bote et al., 2008).

In computational models of WM the background activity has been largely seen as problematic for memory maintenance. For example one of the more sensitive technical issues addressed by several computational proposals is how to stabilize the WM trace in face of random background activity (Compte et al., 2000). The benefits of external input correlations on persistent activity in recurrent networks have only recently started to be addressed theoretically (Buice et al., 2010; Polk et al., 2012). For the specific case of line-attractor networks (modeling parametric WM) Polk et al. (2012) showed in a detailed analysis how properly tuned input noise correlations can promote stability of the persistent firing rate. This was further noted in Lim and Goldman (2012) who also showed that the correlation structure of background noise can suggest the optimal architecture of neural networks for short term memory performance.

Finally, in this article we examined the influence of input correlations on recurrent spiking networks, finding that the correlation level in fact may destabilize the persistent activity state, rendering it a slow transient state. Buice et al. (2010) used a path integral approach to integrate the effects of correlations and synchronization into a rate model of recurrent networks and examined the stability of the persistent state. For a bistable firing-rate network they noted that transient increases in input

correlations (synchronizing noise input) can lead to a turn-off of the persistent activity. This approach may in fact provide an analytical framework of the observations we make in the present manuscript for recurrent spiking networks and the correlation-based control of the persistent state lifetime. In this manuscript we also go beyond noting that input correlations defined the lifetime of persistent activity; we show that input correlations can effectively control the access to WM by disallowing transient stimuli to initiate persistent activity. The functional consequences of these two effects are the central topic of this work.

To demonstrate that, by controlling the correlation-driven synchronization of the background activity it is possible to control the lifetime of the persistent state and to manipulate selectively the transitions in sustained activity and consequently to perform the required operations of the WM task, we first consider a minimal recurrent network. In this recurrent network the neurons receive an excitatory random background noise, and background correlations are induced by a common noise source. Then we implement a discrete item WM model where the modulation of the background correlation level sets the network into different regimes allowing for loading of memory, protection from distractors and memory persistence. In addition we show the possibility to merge the read-out and the clearance in a single operation since the presentation of the match stimulus can directly quench the persistent activity.

MATERIALS AND METHODS

NEURAL MODELS

In this work we study recurrent spiking networks that show bistability between a ground state and an active persistent spiking state. Our goal is to construct and analyze a minimal network capable of showing the required bistability. Hence we consider networks of recurrently connected excitatory pyramidal neurons. The elements of the network are represented by non-linear "point" neurons that are sparsely connected by instantaneous excitatory recurrent synapses. The dynamics of a neuron's membrane potential v is described by the Quadratic Integrate and Fire (QIF) equation, which represents the normal form of type 1 spike generating dynamics (Ermentrout, 1996):

$$\tau \frac{dv}{dt} = v^2 - b^2 + I_{\text{syn}}(t) \tag{1}$$

$$v(t) = V_t \rightarrow V_r \tag{2}$$

where τ represents the membrane time constant, $-b$ is the resting potential, $I(t)$ the input current, V_t a spike threshold, and V_r the reset membrane potential. The voltage of the neuron is scaled such that v is a non-dimensional variable. When the membrane potential neuron attains the threshold value $v = V_t$, a spike is emitted and a post-synaptic current (PSC) is transmitted to an output neuron. We set the parameters as follows: $V_r = -20$, $V_t = 20$, $b = 1$ and $\tau = 20$ ms.

The input current to a given cell in the network is decomposed into three different components:

$$I_{\text{syn}}(t) = I_r(t) + I_s(t) + I_{\text{ba}}(t) \tag{3}$$

where $I_r(t)$ represents the recurrent input due to other neurons in the network, $I_s(t)$ represents the input from external stimuli directed to the network, and $I_{\text{ba}}(t)$ represents a non-specific background activity. Each of the three currents corresponds to a sum of PSCs originating from synaptic inputs generated by the presynaptic neurons at times t_n. The PSCs are modeled with delta pulses:

$$I(t) = \sum_a \sum_{\{t_n\}} J_a \tau \delta(t - t_n) \tag{4}$$

where J_a represents the synaptic strength for a given connection and could be positive (corresponding to an AMPA synapse) or negative (corresponding to a GABA synapse).

BACKGROUND ACTIVITY AND CORRELATIONS MEASURES

Ample data shows that cortical neurons receive a large amount of non-specific cortical and subcortical inputs whose structure is not directly related to the specific task and stimulus [e.g., see Shadlen and Newsome (1994) and summary of data in Amit and Brunel (1997)]. We refer to this type of input as an external background activity. It is taken to be composed of sequences of excitatory PSCs of synaptic strength J_0 and with the synaptic times generated by a Poisson process. The synaptic currents are depolarizing in accordance with the notion that cortical neurons receive inputs from long-range excitatory glutamatergic projections.

In our model, this background activity can be either unstructured (uncorrelated) or structured (correlated). The correlation level, between two spike trains $S_i(t)$ and $S_j(t)$ is given by:

$$\lambda_{ij} = \frac{1}{\langle S_i(t) \rangle} \int \text{CCVF}_{ij}(s) ds \tag{5}$$

where CCVF corresponds to the cross-covariance function (Brette, 2009). This function is normalized to zero if $S_i(t)$ and $S_j(t)$ are generated by independent Poisson processes.

We consider two ways for constructing the background activity:

Uncorrelated background activity

All N neurons receive spike trains generated by N independent channels with rate ν_0. This leads to $\text{CCVF}(s) = 0$ and thus the correlation level is $\lambda_{ij} = 0$.

Correlations induced by a common source of noise (Figure 1B)

All the N neurons receive inputs both from independent channels, with frequency $(1 - \lambda)\nu_0$, and from a common channel, with frequency $\lambda\nu_0$ and $0 \leq \lambda \leq 1$. Each channel generates a spike train with Poisson statistics. The average background input rate is ν_0 for each neuron. The cross-covariance function is then $\text{CCVF}(s) = \lambda \langle S_i(t) \rangle \delta(s)$ and the correlation level is $\lambda_{ij} = \lambda$. This gives purely spatial correlations.

We measure the correlation level of the synaptic input among cells in the network with the mean Pearson correlation coefficient. We first compute a running mean (averaged over a time window of 5 ms) of the synaptic input $I_a^i(t)$ for each cell during a

certain interval of time. Then we compute the Pearson correlation between the synaptic input of two cells:

$$\rho_{ij} = \frac{\mathrm{cov}(I_a^i, I_a^j)}{\sigma(I_a^i)\sigma(I_a^j)} \quad (6)$$

Finally we compute the average over all the cell pairs of the network $\rho = [2/N(N-1)] \sum N_{i=1} \sum_{j=i+1}^{N} \rho_{ij}$. In particular, in **Figure 4**, we performed this measure for the recurrent input $a = r$ and background input $a = ba$.

FUNCTIONAL NETWORK STRUCTURES IMPLEMENTING WM TASKS

In this work we study three different networks. We start out by studying a homogeneous network of recurrently coupled excitatory neurons. This network can be also thought of as a encoding a single item of WM: a "single-unit network". The second model consists of two homogeneous excitatory networks coupled together through a population of inhibitory neurons: a "winner-take-all network" of two discrete competing short-term memory items. The third model is made up of two recurrent excitatory populations without mutual connections: a "two-unit network".

Single-unit network

A homogeneous network with $N = 100$ identical sparsely coupled neurons is represented in **Figure 1C**. Each neuron in the network receives synaptic inputs from cN other excitatory neurons, where $c = 0.2$ is the probability of connection, and $J = 0.26$ is the recurrent synaptic strength [described in Equation (4)]. Neurons receive excitatory inputs also from external background activity, with synaptic strength $J_0 = 0.151$ and firing rate $\nu_0 = 106\,\mathrm{Hz}$. Neurons also receive an excitatory input from external sensory stimuli with synaptic strength $J_1 = 1.5$ and firing rate $\nu_1 = 56\,\mathrm{Hz}$ for a duration of 50 ms, as will be described hereafter. Parameters of the network are chosen such that the network sustains a quiescent state, with low firing rate ($f < 5\,\mathrm{Hz}$), and a persistent state, with high firing rate ($\approx$20 Hz).

Winner-take-all network

The second model is a reduced version of the network proposed by Amit and Brunel (1997) (**Figure 1D**). The network is composed of two excitatory populations and one inhibitory neural population. Each of the two excitatory populations has $N_E = 40$ neurons, and the third population is made up of $N_I = 20$ inhibitory neurons. An excitatory neuron receives synaptic inputs from $c_{\mathrm{EE}}N_E$ ($c_{\mathrm{EE}} = 0.45$) neurons of the same population, with synaptic strength $J_{\mathrm{EE}} = 0.3$, and from $c_{\mathrm{EI}}N_I$ ($c_{\mathrm{EI}} = 0.35$) inhibitory neurons with synaptic strength $J_{\mathrm{EI}} = -0.25$. An inhibitory neuron receives synaptic inputs from $c_{\mathrm{IE}}N_E$ ($c_{\mathrm{IE}} = 0.34$) excitatory neurons with synaptic strength $J_{\mathrm{IE}} = 0.05$ from each excitatory population. Other parameters of the network are: $J_0 = 0.4$, $J_1 = 1.5$, $\nu_0 = 60\,\mathrm{Hz}$, and $\nu_1 = 17\,\mathrm{Hz}$. In addition we augment the mutual inhibition network with the added feature to control the amount of correlated noise in each excitatory population. More precisely each excitatory population receives background activity by common noise sources in addition to independent sources. In such a way the correlation level λ is regulated independently in each excitatory population.

Two-unit network

We devised a third version of our network models that is made of two excitatory independent populations, each one making recurrent connections with itself. Both networks share a common excitatory noise source projecting simultaneously to all excitatory neurons in addition to the independent uncorrelated background noise (**Figure 1E**). Since the common noise source is shared between the two populations the correlation level λ varies equally in the two excitatory populations. The parameters of each excitatory population are those given for the single-unit network, with the only difference that we used here a larger network ($N = 1000$) and we scaled accordingly the recurrent synaptic strength ($J = 0.026$).

DELAYED MATCH-TO-SAMPLE TASK

We study the spike-timing based mechanisms able to implement the DMS task (**Figure 1A**). The sequence of operations and the neural dynamics aim to reproduce the experimental results of Miller et al. (1996). For illustrative purposes, the discrete items can be viewed as corresponding to colors. The activation of one excitatory population encodes color blue, that we define population B, while the other encodes color red, that we define population R. If the populations are both in a quiescent state, the state of the network represents the absence of color information. For simplicity we represent the spontaneous state as the quiescent state (average firing rate $\approx$0 Hz).

During the task, the animal has to maintain a memory of an item (a color) during the delay period. In terms of neural activity, the corresponding excitatory population should be activated and maintained in a persistent sate. Additionally the model should protect the memory from the presentation of a distractor stimulus. At task completion after the decision, the system should erase rapidly the memory, i.e., the persistent activity should be deactivated to its quiescent state.

To establish that a network performs a WM task correctly we require it to perform all the operations of the task. The first operation, *load*, corresponds to loading the memory by the sample signal, and corresponds to B that is activated in a persistent state while R is in a quiescent state. The second operation, *protect*, corresponds to the maintenance of the blue item memory in the face of the distractor presentation. In terms of activity it corresponds to B maintained in the persistent state and R that is not activated to the persistent state even when the red stimulus is presented during the delay period. In networks (**Figure 1E**) where population B and population R are not connected, the operation *protect* can be separated in two independent sub-operations: *maintain* (maintain item memory in population B) and *block* (prevent activation of population R). The third operation, *clear* corresponds to the clearance of the memory encoded in the network. This is equivalent to the erasing of the persistent activity in the network. Note that in this work we do not focus explicitly on the read-out mechanism following the presentation of the match stimulus.

In particular in the winner-take-all network (resp. the two-unit network), operation *load* is executed with success if the sample stimulus activates population B. This is measured before distractor presentation during 350–450 ms (resp. 350–450 ms): $\nu_B > 5$ Hz and $\nu_R < 5$ Hz, where $\nu_B < 5$ and ν_R denote the average population firing rates of populations blue and red, respectively. Operation *protect* is executed with success if population B maintains the persistent state and population R is not activated. This is measured before match presentation during 750–850 ms (resp. 700–800): $\nu_B > 5$ Hz and $\nu_R < 5$ Hz. Operation *clear* is executed with success if population B is deactivated at task completion. This is measured during an interval after match presentation, during 1150–1250 ms (resp. 1050–1150 ms): $\nu_B < 5$ Hz and $\nu_R < 5$ Hz.

NUMERICAL ANALYSIS

All the numerical results are obtained by algorithms run in Python. The differential equations are integrated with Euler steps of $dt = 0.1$ ms. The mean population firing rate f is computed over population average in 10 ms.

Data points for networks and associated error bars are computed by averaging over simulated individual network realizations. We generated random connectivity matrices such that every neuron receives the same number of input connections. For a fixed network connectivity matrix, we computed the average over 100 realizations of background activity and stimuli for each of 30 random realizations of the network connectivity matrix when not otherwise stated.

RESULTS

EFFECTS OF CORRELATIONS ON PERSISTENT ACTIVITY STATE IN THE SINGLE-UNIT NETWORK: ERASING AND BLOCKING THE MEMORY TRACE

We examine how correlations in the background activity control selective persistent activity in WM networks. Hence we start out by analyzing how background correlations affect the transitions between the quiescent and self-sustained states in our network model.

Correlations in background activity are generated by the addition of a common noise source to independent stochastic channels (see **Figure 1B**). By changing the relative firing rate of the common source with respect to the independent channels we control the correlation level λ. We set two different protocols represented in **Figure 2A**. In the first protocol, the correlation level is increased instantaneously from $\lambda = 0$ to some value $\lambda > 0$ at 500 ms. Therefore given that the stimulus activates the persistent state, this protocol allows us to test the effects of the correlations on the probability that the active state is erased and we refer to it as the erasing protocol. In the second protocol the correlation level is set $\lambda > 0$ for all the time, before the transient stimulus appears. In this way it is possible to see the effect of correlations on blocking the ability of the stimulus, presented during 50–100 ms, to initiate the persistent state and we refer to it as the blocking protocol.

We first demonstrate the prevalent effect of correlations: control of active memory state and control of access to the memory. In an example of the erasing protocol the excitatory stimulus activates the network into a persistent state; at 500 ms correlations are increased and the persistent state is disrupted (**Figure 2A**). In an example of the blocking protocol the excitatory stimulus is not able to activate the persistent state (see **Figure 2B**). In order to understand how these effects depend on the activity parameters we ran a large number of simulations where we injected background activity with different correlation levels for $0 \leq \lambda \leq 1$ to networks with different connection probability c to measure how this effect spread thanks to the network architecture (**Figure 2C**). We compared networks with the same scaled synaptic strength J such that $cJN = \text{const.} = 5.2$. In the erasing protocol we estimated the erasing probability $P_e(c, \lambda)$ defined as the probability for the network to have the firing rate $\nu < 5$ Hz in the interval 800–900 ms. We discarded trials where the network is not in a persistent state ($\nu > 5$ Hz during 400–500 ms). In the blocking protocol we estimated the probability that the correlations block the stimulus; the blocking probability $P_b(c, \lambda)$ defined as the probability for the network to have the firing rate $\nu < 5$ Hz in the interval 400–500 ms. This could also be seen as a gating of the persistent activity. We observe that for both protocols, the increase of both c and λ disrupts the persistent state: in the first case by erasing it and in the second case by blocking its activation.

We hence wanted to assess how the network size influences the stability of the persistent state under the various background activity regimes (**Figure 2D**). We compared networks with different size N with an equal average synaptic input $cJN = \text{const.} = 5.2$. We measured both erasing probability $P_e(N, \lambda)$ and blocking probability $P_b(N, \lambda)$ as a function of λ. We observe that both the erasing probability and the blocking probability (P_e and P_b) increase with the network size N. We observe that both for fixed c and for fixed N $P_b > P_e$. Finally we studied the probabilities P_e and P_b as function of N, fixing both $J = 0.26$ and the number of inputs that each neuron receives, i.e., $cN = \text{cost} = 20$. We computed these probabilities averaging over 500 trials. We found that with such a scaling both P_e and P_b are approximately constant (**Figure 2E**).

In order to determine whether the optimal stimulus parameters to load of a memory (or activation of a persistent state) depend on the correlation strength we measured the loading probability $(1 - P_b)$ as function of the stimulus strength (ν_1) and for different values of λ (**Figure 3A**), we computed the probabilities of **Figure 3** averaging over 300 trials. Different values of λ change the amplitude of $(1 - P_b)$ but do not shift the tuning with respect to ν_1. We also found that there are two peaks of $(1 - P_b)$: one at about $\nu_1 \approx 20$ Hz and another at about $\nu_1 \approx 50$ Hz. To test whether the positions of the two peaks depend on the recurrent network properties we measured the loading probability as function of ν_1 and for different values of the recurrent synaptic strength J (**Figure 3B**). Similarly to the previous results, different values of J change the amplitude of $(1 - P_b)$ but do not shift the peaks of the curves with respect to ν_1. In summary this indicates that indeed the strength of the stimulus required to active the persistent state with a set probability is dependent on the background correlations, and yet the tuning is rather broad.

To further investigate the effect of correlation on the stability of the persistent state we determined the lifetime of the sustained activity and the level of correlations prior to the erasing time. We defined the end of the persistent state t_{stop} (magenta vertical line,

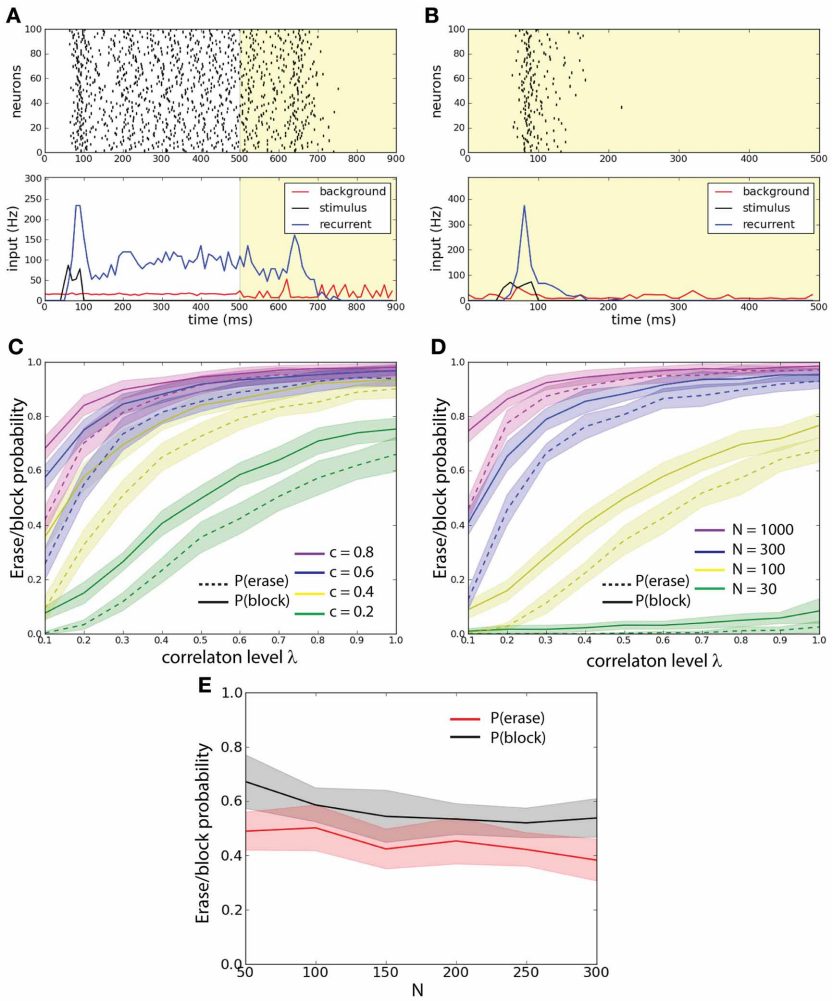


FIGURE 2 | External background correlations destabilize the persistent state in a single-unit network. (A) Erasing persistent state with correlations. Examples of firing rate (top), and average synaptic input for one trial. $\lambda = 0$ until 500 ms and $\lambda = 0.8$ after 500 ms (yellow shaded areas). The population is activated by a stimulus during 50–100 ms. Correlations in background activity erase the persistent state. **(B)** Correlations gate the activation of the persistent state. $\lambda = 0.8$ all the time (yellow shaded areas). The network receives an excitatory stimulus during 50–100 ms. The stimulus fails to activate the persistent state in presence of background correlations. **(C)** Erasing probability P_e (continuous lines) and blocking probability P_b (dashed lines) as function of c and λ. Synaptic strength is scaled such that $cJN = 0.52$. Both P_e and P_b increase with increasing λ and c. **(D)** P_e (continuous lines) and P_b (dashed lines) as function of N and λ. Synaptic strength is scaled such that $cJN = 0.52$. Both P_e and P_b increase with increasing λ and N. **(E)** P_e and P_b as function of N with fixed $\lambda = 0.6$ and $J = 0.26$ and with c scaled such that $cN = const = 20$. P_e and P_b remain approximately constant.

Figure 4A) as the first period of 10 ms (after the correlation onset) during which the firing rate of the network falls below 5 Hz. Noticing that in most of the trails a peak of activity was preceding the persistent state erasing, we defined the time of such a peak t_{peak} (black vertical line, **Figure 4A**) as the last period of 10 ms before t_{stop} that the firing rate attains a local maximum (in time) and that is beyond 20 Hz. We determined for each trial where the persistent state was not erased before the onset of correlations $t_{\text{corr}} = 800$ ms (red vertical line, **Figure 4A**), the interval $\Delta t_{\text{c.p.}} = t_{\text{peak}} - t_{\text{corr}}$. We determined the interval $\Delta t_{\text{p.s.}} = t_{\text{stop}} - t_{\text{peak}}$. We performed this protocol for three different values of the correlation level: $\lambda = \{0.3, 0.6, 0.9\}$ (**Figure 4B**). We found that the distribution of $\Delta t_{\text{c.p.}}$ decreases with time for all the values of correlations. When the level of correlations is larger (**Figure 4B**, top) the probability of reaching the peak earlier in time slightly increases with λ. Furthermore we found that the interval between the peak of activity and the erasing of the activity in the network is narrowly distributed in time. Finally this interval is independent of the correlation level, meaning that the correlations do not have a strong effect on this timing (**Figure 4B**, bottom). We computed these distributions averaging over 500 trials.

The mean Pearson correlation coefficient ρ (see Materials and Methods) of the synaptic input in the network during the interval with uncorrelated background activity was compared with that during the interval with correlated background activity just preceding the peak. Only trials where $t_{\text{peak}} - t_{\text{corr}} >$

100 ms were considered. The interval with uncorrelated background activity is defined as the 100 ms preceding t_{corr} (gray shaded area, **Figure 4A**). The interval with correlated background activity is defined as the 100 ms preceding t_{peak} (red shaded area, **Figure 4A**). We computed ρ for the background input (red lines, **Figure 4C**) and for the recurrent input (black lines, **Figure 4C**) both for the uncorrelated interval (dashed lines) and for the correlated input (continuous lines) when $\lambda = 0.6$. We found that ρ during the correlated input is smaller in the recurrent input with respect than in the background input. However, ρ of the recurrent input is larger during the correlated interval than during the uncorrelated interval. Interestingly we found that during the correlated interval, while the correlation coefficient of the background input increases with λ (**Figure 4D**, bottom), the correlation coefficient of the recurrent input instead remains approximately equally distributed when λ is changed (**Figure 4D**, top). This suggests that the network has reached the maximal amount of sustainable correlations before turning off.

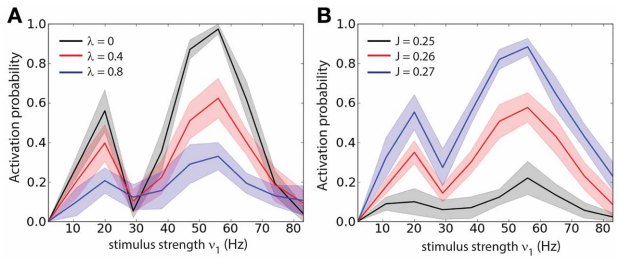


FIGURE 3 | Stimulus strength tuning is independent of the correlation level. (A) Probability of activation ($1 - P_b$) as function of the stimulus strength (ν_1) for different values of the correlation level (λ). **(B)** Probability of activation as function of the stimulus strength (ν_1) for different values of the recurrent synaptic strength (J).

To understand whether the persistent activity deactivation is caused by an increase of spike synchrony we tracked the synchrony of the spike times using the multivariate SPIKE-distance measure S (Kreuz et al., 2013) (**Figure 5A**). The spike synchrony is given by $1 - S$ spanning the values between 0 (no synchrony) and 1 (perfect synchrony). We compared the average spike synchrony during two intervals, similarly to **Figure 4**: the first interval corresponds to the 100 ms preceding the start of correlated background activity and the second interval corresponds to the 100 ms preceding the last peak of activity before the deactivation of the persistent activity (provided that the onset of this last interval does not precede the start of the correlated background activity). The distribution of the average value of $1 - S$ (computed

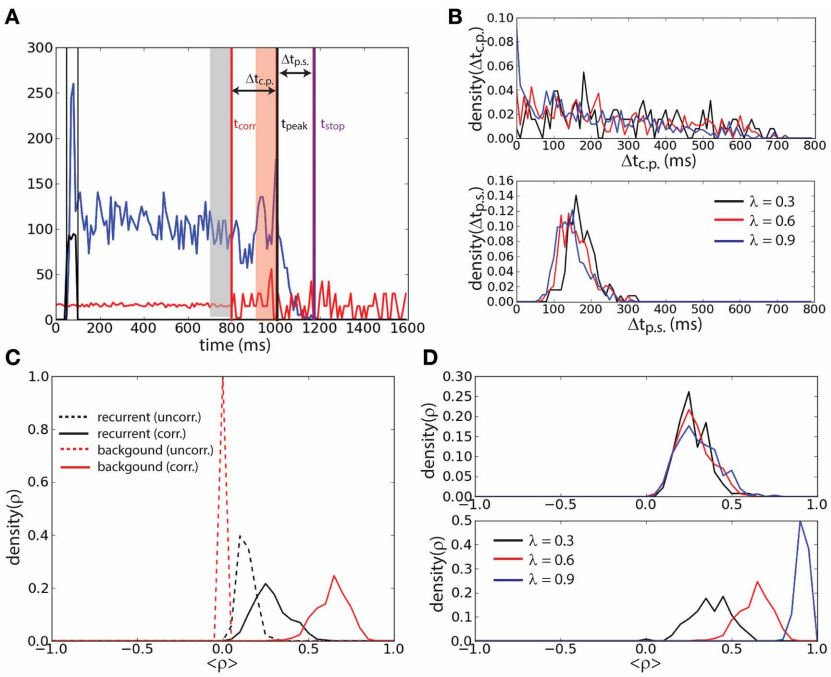


FIGURE 4 | Persistent state suppression is preceded by an increase of recurrent correlation. (A) Timing outline of erasing persistent activity. Mean recurrent input (blue trace), mean background input (red trace), and external input (black trace) are represented together with the time at which activity is erased (t_{stop}, magenta vertical line), time of the last peak of activity before erasing (t_{peak}, black vertical line), and time of the onset of the correlations in background activity (t_{corr}, red vertical line). The interval of uncorrelated activity and the interval of correlated background activity during which the correlation coefficient is measured in panels **(C)** and **(D)** are represented with gray and red shaded areas, respectively. **(B)** Distribution of the interval $\Delta t_{c.p.} = t_{peak} - t_{corr}$ (top) and distribution of the interval $\Delta t_{p.s.} = t_{stop} - t_{peak}$ (bottom) for different values of λ. **(C)** Mean Pearson correlation coefficient (ρ) of the recurrent input (black traces) and of the background input (red traces), computed during the uncorrelated interval (dashed traces) and the correlated interval (continuous traces). **(D)** Coefficient ρ of the recurrent input computed during the correlated interval for different values of λ (top). Same analysis for the background input (bottom).

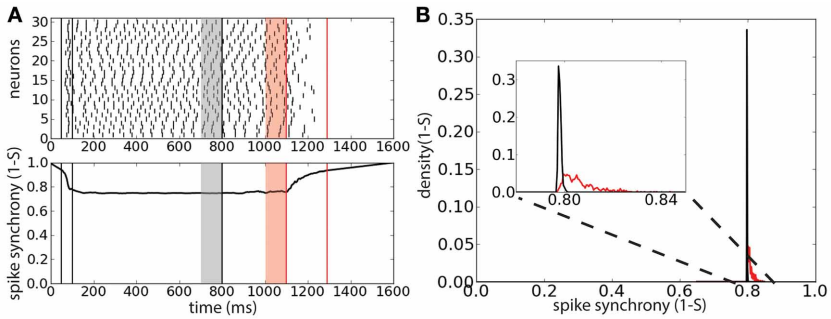


FIGURE 5 | Persistent state suppression is preceded by a weak increase of spike synchronization. (A) Raster plot of 30 representative neurons of the network (top) and spike synchrony $1 - S$ (bottom). Same protocol of that described in **Figure 4**. Gray shaded area corresponds to the 100 ms interval preceding the background correlation onset ($\lambda = 0.6$); red shaded area corresponds to the 100 ms preceding the last peak of activity before sustained activity suppression. **(B)** Distribution of the average spike synchrony ($1 - S$) during the two interval described previously. Inset corresponds to a magnification of the relevant interval of spike synchrony values.

over 2000 trials) during these two intervals shows that there is a weak increase of spike synchrony preceding the persistent activity deactivation with respect to the case of uncorrelated background activity (**Figure 5B**). This weak increase could indicate that few spike coincidences might be the cause of the persistent activity turning off.

EFFECTS OF BACKGROUND ACTIVITY CORRELATIONS IN A WINNER-TAKE-ALL NETWORK

We show here that modulating appropriately in space and time the correlation level of the background activity in a network performing a WM task significantly improves correct execution of all the required operations: *load*, *protect*, and *clear*.

We compared two different versions of the winner-take-all network; each made of two excitatory populations *B* and *R*, representing respectively colors blue and red. The two populations interact via a third population of inhibitory neurons that creates a winner-take-all mechanism. The two versions differ in that the first one receives only uncorrelated background activity while in the second each excitatory population receives also background activity from a different common noise source (**Figure 1D**).

We fixed the stimuli sequence as follows: during 50–150 ms a sample blue stimulus excites population *B*; during 450–550 ms a distractor red stimulus excites population *R*; during 850–950 a match blue stimulus excites again population *B*. The operations that the network has to do are to load the blue item in memory, to protect the memory at red item presentation, and to clear the memory after the match presentation.

Brunel and Wang (2001) pointed out that in order to perform the DMS task correctly, the distractor stimulus strength needs to be controlled with care: above a certain strength persistent memory-trace is perturbed by the distractor. For our case we suppose that it is reasonable to assume that all sensory stimuli in the task are of the same strength. As a preliminary test we want to confirm that in absence of background correlations the network without common noise source does not perform efficiently when the stimuli are too strong, as was already stated in the reference network described by Brunel and Wang (2001). In the example shown in **Figure 6A** the distractor activates *R* and via the inhibitory population the persistent state in *B* is deactivated leading to a failure of the operation *protect*.

We then consider the network represented in **Figure 1D** that allows the modulation of the correlation level λ in each excitatory population independently. The network initially receives uncorrelated background activity to $\lambda = 0.9$. After the first item has been loaded, the system increases the correlation level in the background activity of the other non-activated population *R*. After the match stimulus has been presented, the correlation level is increased also in population *B* to $\lambda = 0.9$.

We show an example of the network executing the WM task where the correlation level is modulated independently in the excitatory populations (**Figure 6B**). In this particular example we illustrate a trial where the network performs the required operations of the WM task (compare with **Figure 1D** and see below for statistics across trials). The distractor excites *R* only transiently such that excitation does not last enough to disrupt the activity in *B*, in addition as shown below this happens also for strong distractor stimuli. Therefore the operation *protect* has been executed with success and the memory is maintained. At the end of the match stimulus the persistent activity is disrupted also in population *B* caused by the increase of λ in that population too. Therefore the operation *clear* is executed with success and the memory is erased in the network. This example illustrates that the success of the operations *protect* and *clear* in the network with correlations are not due to the presence of inhibition as was set in the model of Brunel and Wang (2001).

To get quantitative measures of performance for these two networks (with and without correlations in background activity), we analyzed the statistics of *load* and *protect* performance, as a function of the stimulus intensity ν_1 (and thus its strength) (**Figure 7A**). We consistently find higher *protect* performance for correlated background activity than for uncorrelated background activity throughout the whole range for ν_1. In fact the success of the *protect* operation depends only gradually on

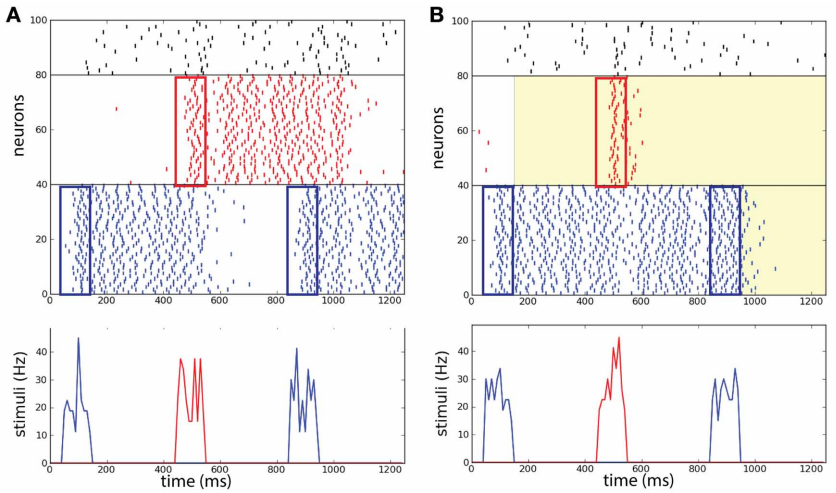


FIGURE 6 | Selective correlations implemented in a working memory task. Two competing populations network, with two item-selective excitatory populations (blue and red) and one inhibitory non-selective population (black). **(A)** Without background correlations, the distracting stimulus activates population *R* and population *B* is deactivated. **(B)** After the activation of *B* at 150 ms, a common source of noise increases the correlations in background activity ($\lambda = 0.9$) in *R*. The correlations block the activation of *R* and maintain the persistent state *B*. After the completion of the task at 950 ms the correlations erase persistent activity in *B*. (Top) Raster plot of the neural activity in the task. (Bottom) Successively: sample stimulus to *B* (50–150 ms), distracting stimulus to *R* (450–550 ms), and match stimulus to *B* (850–950 ms).

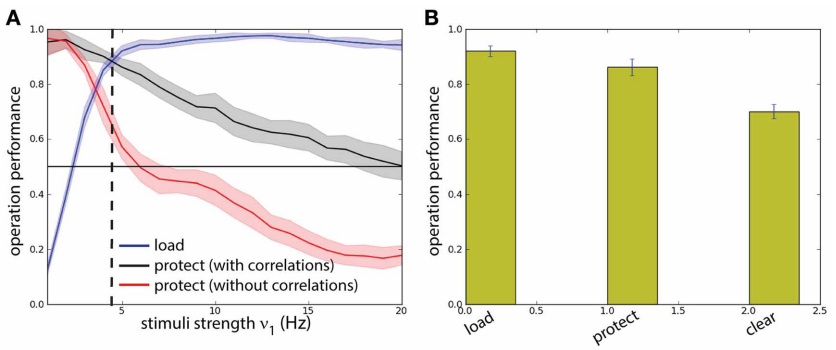


FIGURE 7 | Background correlations increase working memory performance. Protocol reported in **Figure 6**. **(A)** Comparison between a two competing populations network with and without correlations. Dashed line: optimal value for the network with correlations. **(B)** The performance of the network is measured on four different operations for the optimal value of $\nu_1 = 4.8$ Hz: probability of activating *B* by the sample stimulus (load), probability of preventing memory disruption by a distracting stimulus in the protocol with correlations (protect), probability of erasing of the memory at the end of the task (clear). All probabilities have a high value showing that the network has good task performance, above chance.

the distractor strength. On the other hand in order to perform operation *protect* above chance level in the network with uncorrelated background activity distractors should be carefully adjusted to have intensity $\nu_1 < 5$ Hz. However, in this range the operation *load* is suboptimal. Hence the uncorrelated model fails in the task. This fact illustrates a recurrent problem in the protect-by-inhibition paradigm: it needs fine-tuning and achieves only low performance if the stimuli are too strong. Instead, using correlations as a mechanism to protect the activity does not need precise fine-tuning as can been seen in the large range in which both *load* and *protect* are well above chance level. We found the value $\nu_1 = 4.8$ Hz maximizes the joint probability of executing with success *load* and *protect* (**Figure 7A**, vertical dashed line). We show in **Figure 7B** probability of success of the three operations *load*, *protect*, and *clear* finding that all of them score a value higher than chance level.

IMPLEMENTING WORKING MEMORY TASK BY FLEXIBLE CORRELATIONS MODULATION

We now go on to show that mutual inhibition is not a required mechanism for implementing the WM task. We show here that modulating appropriately the background activity correlations in time in a network without inhibitory population allows correct execution of all the required WM operations: *load*, *maintain*, *block*, and *clear* (Please note that since the network studied

here is made of two separated excitatory populations the component *maintain* and *block* of the operation *protect* can be treated separately).

Network operating regimes

In order to characterize the network performance statistics during the task we need to track three probabilities. The first probability $P_{\text{g.o.}} = P_e P_b$ corresponds to the joint probability of deactivating by correlations the network that is in the persistent state (erase) and to block activation of the network that is in the quiescent state and is excited by a stimulus. When $P_{\text{g.o.}}$ dominates over the other probabilities the system is in a gate-out regime, i.e., memory cannot neither be loaded nor maintained in the network. The second probability $P_{\text{g.i.}} = (1 - P_e)(1 - P_b)$ corresponds to the joint probability that, despite the correlations, the network maintains the persistent state, if previously activated, and that the stimulus activates the persistent state when the network is in the quiescent state. When $P_{\text{g.i.}}$ dominates, the system is in a gate-in regime, i.e., the memory can be loaded and maintained in the network. Finally the third probability $P_{\text{s.g.}} = (1 - P_e)P_b$ corresponds to the probability of maintaining the persistent activity in the presence of correlations while blocking the activation of a persistent state with correlated background activity when the system is in a quiescent state and is excited by a stimulus. When $P_{\text{s.g.}}$ dominates the system is in a selective-gate regime, i.e., the memory is maintained but cannot be loaded. In a sense we want to show that correlations in the background activity can selectively switch the network from the gate-in regime at the outset of the task, to the selective-gate regime during the memory period. We do not consider the probability $P_e(1 - P_b)$.

To obtain the network performance on the DMS task we considered the statistical results presented in **Figure 2** for the single excitatory population ($N = 1000$ and $c = 0.2$). We note that there is a difference between the erasing probability $P_e(\lambda)$ and the blocking probability $P_b(\lambda)$ in function of the correlation level. In **Figure 8** we present the results for network we consider in this manuscript. We see that when $P_{\text{g.i.}}(\lambda)$ dominates ($\lambda < 0.04$), the system is in a gate-in regime, i.e., the memory can be loaded and maintained in the network (**Figure 8**). When $P_{\text{s.g.}}(\lambda)$ dominates ($0.04 < \lambda < 0.11$) the system is in a selective-gate regime. We do not consider here the gate-out regime that corresponds to $P_{\text{g.o.}}(\lambda)$ dominating over the other probabilities ($\lambda > 0.11$). We set the gate-in regime at $\lambda = 0$ and the selective-gate regime at $\lambda = 0.07$.

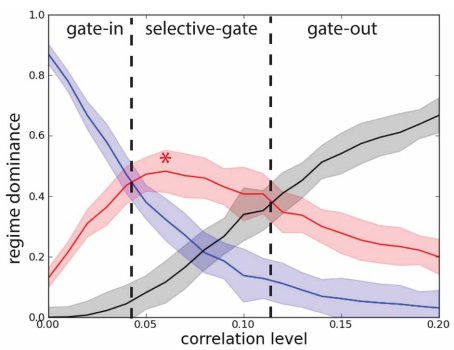


FIGURE 8 | Correlations set the network working regime. Probability that the network operates in different regimes with network size $N = 1000$. Joint probability of (not-)erase and (not-)block: $P_{\text{g.o.}}(\lambda)$ (black curve), $P_{\text{g.i.}}(\lambda)$ (blue curve), and $P_{\text{s.g.}}(\lambda)$ (red curve). Gate-out regime corresponds to domination of $P_{\text{g.o.}}(\lambda)$ and falls in the range $\lambda > 0.11$. Gate-in regime corresponds to dominance of $P_{\text{g.i.}}(\lambda)$ and falls in the range $\lambda < 0.04$. Selective-gate corresponds to dominance of $P_{\text{s.g.}}(\lambda)$ and falls in the range $0.04 < \lambda < 0.11$, with maximal value at $\lambda = 0.07$ (red star).

Modulation of correlation level in time

We now show that correlations induced by a global common noise source to the whole network executes the DMS task efficiently by modulating the correlation level λ during the different phases of the task. We note that in the mutual inhibition model, at task completion, increase in the correlation level induces the gate-out regime and erases the memory. We show here, in a two-unit model (**Figure 1E**), how the presentation of the match stimulus directly can erase the memory thereby implementing a direct match-based suppression without requiring inhibition. In this model each of the two excitatory populations receives background activity from sources independent to each neuron and from a noise source common to all neurons.

An example of the network performing the DMS task is represented in **Figure 9A**. The stimuli are presented in the following sequence: sample stimulus to population B at time during 100–150 ms, distractor stimulus to population R during 450–500 ms, and match stimulus to population B at time 800–850. In the beginning the system is in the gate-in regime ($\lambda = 0$): the sample stimulus activates B. From 300 ms the network is set in a selective-gate regime ($\lambda = 0.07$): the distractor stimulus activates transiently population R while persistent activity is maintained in population B. At the end of the task the match stimulus, first, increases the activity in B and, then, destroys it.

We can then compute the task performance of the network, corresponding to the success rate that the operations *load*, *maintain*, *block*, and *clear* are executed successfully (**Figure 9B**). These measure are all above chance level. Notice the high level of performance of the *clear* operation.

DISCUSSION

RESULTS AND DATA DISCUSSION

In this work we present a novel paradigm explaining how the persistent activity can be modulated on-line by the mean of both information-related signal and background activity. This paradigm is based on our result showing that background correlations influence the transition between the persistent state and the quiescent state in a bistable recurrent neural network. We call this phenomenon correlation-induced gating.

In order to implement a multi-unit network performing a WM task, we began by establishing the basis of the correlation-induced gating on a single-unit network. We show that background correlations block and erase a persistent state in a homogeneous recurrent neural network representing a single unit. We found that the transition rate from the persistent state to the quiescent state increases, with the network size and with the connection probability. For all situations the probabilities increase

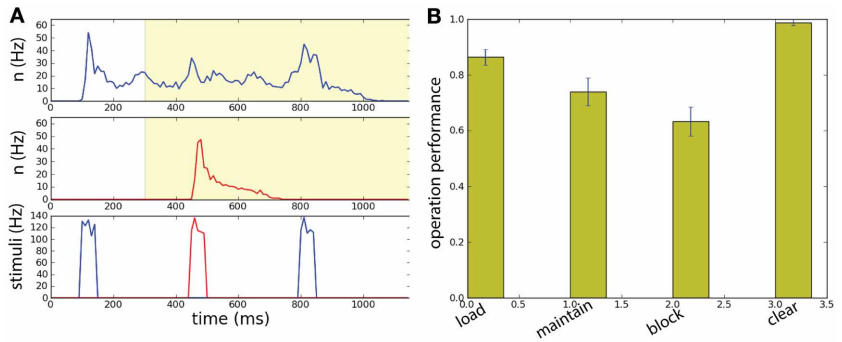


FIGURE 9 | Spatially uniform correlations in a two unit network performing a working memory task. (A) The network has size $N = 1000$. Memory is loaded in population B by the sample stimulus. Background correlations block the activation of persistent population R by a distractor while maintaining the memory in the population B. The match stimulus erases the memory in population B, playing both the role of read-out and clear. (Top) Raster plot of the neural activity for populations B and R. (Bottom) Input current of the: sample stimulus, distracting stimulus, and match stimulus. The network has size $N = 1000$. **(B)** With intermediate correlation level ($\lambda = 0.07$) in background activity the network can execute successfully the working memory task. Task performance of the network for the four operations: load, block, maintain, clear.

with the correlation level. Increasing the network size, while fixing the connections probability and renormalizing the synaptic inputs to keep the average input strength constant scales up the probabilities. In other words, in larger size network with weaker but more numerous synapses, correlations appear to have a stronger effect. On the other hand, when we fix the total number of synapses each neurons receives, growing network size does not appear to have much effect on the correlation driven probabilities. These effects could be related to the fact that the amount of correlation between neurons sharing common input is mainly determined by pooling (Rosenbaum et al., 2010, 2011).

We implemented a winner-take-all network composed of two excitatory populations and one inhibitory population. Each of the excitatory populations receives background input from independent noise sources and a noise source common to the neurons of such population. The amount of correlation could be changed independently in the two excitatory populations. By increasing the level of correlations in the populations encoding an irrelevant information we prevented a distractor from loading a memory item in such population. In particular we showed that this model allows to prevent stronger distractors with respect to a model inspired by Brunel and Wang (2001) where the distractor is blocked only by the mutual inhibition. Our model could therefore explain how the response to the distractor stimulus in a WM task could be as strong as for the sample stimulus (Miller et al., 1996). This effect would be in fact not compatible with a model where a distractor is prevented by mutual inhibition.

We implemented a WM network differing by the same previous one in the construction of background correlations that are induced by a shared source. We show that modulating the correlation level in background activity we can set the system in the different regimes. This time instead of modulating the correlations level "in space" we modulate it in time. Depending on the strength of the correlations the system is set in different operating points, namely the gate-in, selective-gate, and gate-out regimes. The gate-in regime allows to load a memory in the WM store and to maintain it subsequently. The selective-gate regime maintains a previously loaded memory but blocks the load of any new memory. The gate-out regime blocks the network both to load and to maintain a memory. We can switch instantaneously from one dynamic regime to the other by tuning the strength of the background activity correlations. We further show that the projection of a strong match stimulus can be sufficient to clear the memory at task completion, thereby suggesting that correlations also play a role in match-suppression.

We must also note that in this work we considered spatial correlations and their effect on the persistent activity and WM task executions. In a companion paper we have shown that temporal structure also has an important effect: the gating modes are modulated by the oscillatory frequency content of the background activity (Dipoppa and Gutkin, 2013). While in the companion paper the block and erase probabilities, and thus the gating modes of the network, are modulated by the oscillation frequency, in this work they are modulated by the correlation level. As opposed to the non-monotone relationship between the oscillation frequency and the block and erase probabilities [Figure 3A of Dipoppa and Gutkin (2013)], there is a monotone relationship between the correlation level and the same measures (**Figures 2C,D**). Hence control of the WM through spatial correlations could be implemented by a simple increase or decrease of activity within a neural population furnishing connections common to the WM store, while the oscillatory control would require more complex task-dependent shifting between the frequency bands. We would like to speculate that the two mechanisms could represent two independent modes of control over the WM networks. Furthermore, in this work we uniquely examine the role of mutual inhibition and show that the spatial correlation structure alleviates the network sensitivity to stimulus strength.

Although we do not propose a mechanism for the read-out of the memory information we note that the mechanism

proposed by Brunel and Wang (2001) for their network would be compatible with our model. This mechanism corresponds to the fact that a match stimulus will elicit a stronger response with respect to a distractor stimulus in the first few tenths of milliseconds since the first will excite a network that is already in the persistent state. Hence we might speculate that a complementary population of neurons sensitive to rapid transients in the activity might be a way to signal read-out differentially.

MODEL PREDICTIONS AND OPEN QUESTIONS

The novel paradigm that we present here allows to manipulate persistent activity through background correlations. An advantage of the correlation-induced gating with respect to the inhibition-induced gating is that the gate can be rapidly and flexibly opened or closed depending on the correlation level, instead of being fixed by the network connectivity structure.

The effects that we find for the flexible changes in the correlation levels is the major prediction of the model. We predict that an examination of multi-unit electrophysiological recordings of animals performing a WM task will show the following modulation of correlation level: low level during loading and intermediate level during maintenance (as in the two-unit model of **Figure 1E**) or alternatively high level of correlations in the population of neuron selective for a non-memorized item during the delay period (as in the winner-take-all model of **Figure 1D**). To our knowledge experiments specifically analyzing how correlations change in the PFC as the delay-response task unfolds are still lacking.

At the same time, there are several lines of indirect evidence that lead us to believe that task dependent correlation modulation is indeed possible. First, it has been found that there is a modulation of spike coincidences during different phases of a motor task (Riehle et al., 1997). Riehle et al. (1997) found that at times during the delay when the animal was expecting to generate a response there were transients of synchronized spikes. Furthermore for successful trials there were more synchronized spikes during the delay period than for failure trails. This indeed suggests that spike coincidence is modulated in a functional way. The increase of excess synchrony at response (or expected response) times is compatible with the correlation based memory clearance discussed in this manuscript. Furthermore, it has been found that a change in representation during the delay-response task leads to an increase of synchronization (Sakamoto et al., 2008). Pipa and Munk (2011) analyzed multi-unit activity during the delay period of a match-to-sample task and found that on correct vs. incorrect trials there is a modulation of spike synchronization and further, synchronous spike events are more prevalent at match presentation. This last point again suggests that increased correlations may be involved in erasing the memory trace.

In fact there is ample literature relating changes in oscillatory synchrony, coherence and frequency during WM tasks (Tallon-Baudry et al., 1998; Pesaran et al., 2002; Lee et al., 2005; Pipa et al., 2009). For example Pesaran et al. (2002) found that gamma-band spiking coherence is increased during the delay period in the lateral intraparietal cortex (LIP) in primates performing a delayed response task. Given that LIP is coupled to the PFC and is also involved in WM trace (Chafee and Goldman-Rakic, 1998), this is suggesting of increased input correlations to the PFC during the WM task. In the context of irregular poisson firing, oscillatory coherence is nothing other than correlations organized both in time (the frequency) and space. Oscillatory effects are beyond the scope of this paper and are a subject of the companion manuscript (Dipoppa and Gutkin, 2013).

The data reviewed above does show that there is a modulation of activity correlations during the WM task, yet it does not provide the mechanism. Here we propose that the mechanism is in the background input correlations generated by a common source. One hence might ask where such inputs may be coming from. As hinted above, one source could be coherent firing activity in the cortical regions coupled to the PFC and involved in WM processing (e.g., LIP). In addition, we propose that the source of shared background input generating spatial correlations can reside in the striatum, a subcortical area thought to be involved in WM. In fact the structure of the cortico-striatal loops as been longly seen as a disadvantage for the WM capacity if the memory store is located also in the striatum. Since the number of striatal neurons is much lower than the number of pyramidal neurons (Lange et al., 1976) and the loop is based on divergence (resp. convergence) in the striato-cortical (resp. cortico-striatal) direction, then the striatum could not have the same memory capacity of the cortex. It has been suggested that instead that divergent/convergent structure could be useful since the basal ganglia do not encode the individual information of WM but they control the gate of other region and decide when they can be updated (Frank et al., 2001). We also suggest that striatum plays a gating role since it could be the source of the common noise that creates the different regimes.

The correlation-induced gating is a robust effect to parameters variation. We propose the following explanation for this phenomenon: background correlations induce spike-times synchronization in the recurrent network, as was found similarly for independent neurons by Galán et al. (2006), and this leads to persistent activity erasing and block because of the refractory period of the neurons, as was found by Laing and Chow (2001) and Gutkin et al. (2001). Providing a proof of this assumption and a mathematical explanation of the correlation-induced gating will be the subject of future research.

ACKNOWLEDGMENTS

The authors want to thank Ole Jensen, Romain Brette, Christian Machens, and Thomas Kreuz for constructive discussions. Mario Dipoppa was partially supported by MESR (France) (for Mario Dipoppa). Boris S. Gutkin was partially supported by CNRS, ANR-Blanc Grant Dopanic, CNRS Neuro IC grant, Neuropole Ile de France, Ecole de Neuroscience de Paris collaborative grant, and LABEX Institut des Etudes Cognitives, INSERM, and ENS.

REFERENCES

Abeles, M., Bergman, H., Margalit, E., and Vaadia, E. (1993). Spatiotemporal firing patterns in the frontal cortex of behaving monkeys. *J. Neurophysiol.* 70, 1629–1638.

Aertsen, A. M., Gerstein, G. L., Habib, M. K., and Palm, G. (1989). Dynamics of neuronal firing correlation: modulation of "effective connectivity". *J. Neurophysiol.* 61, 900–917.

Amit, D. J., and Brunel, N. (1997). Model of global spontaneous activity and local structured activity during delay periods in the cerebral cortex. *Cereb. Cortex* 7, 237–252. doi: 10.1093/cercor/7.3.237

Ardid, S., Wang, X.-J., Gomez-Cabrero, D., and Compte, A. (2010). Reconciling coherent oscillation with modulation of irregular spiking activity in selective attention:gamma-range synchronization between sensoryand executive cortical areas. *J. Neurosci.* 30, 2856–2870. doi: 10.1523/JNEUROSCI.4222-09.2010

Baddeley, A. (2003). Working memory: looking back and looking forward. *Nat. Rev. Neurosci.* 4, 829–839. doi: 10.1038/nrn1201

Bays, P. M., and Husain, M. (2008). Dynamic shifts of limited working memory resources in human vision. *Science* 321, 851–854. doi: 10.1126/science.1158023

Brette, R. (2009). Generation of correlated spike trains. *Neural Comput.* 21, 188–215. doi: 10.1162/neco.2009.12-07-657

Brunel, N., and Wang, X.-J. (2001). Effects of neuromodulation in a cortical network model of object working memory dominated by recurrent inhibition. *J. Comput. Neurosci.* 11, 63–85. doi: 10.1023/A:1011204814320

Buice, M. A., Cowan, J. D., and Chow, C. C. (2010). Systematic fluctuation expansion for neural network activity equations. *Neural Comput.* 22, 377–426. doi: 10.1162/neco.2009.02-09-960

Chafee, M. V., and Goldman-Rakic, P. S. (1998). Matching patterns of activity in primate prefrontal area 8a and parietal area 7ip neurons during a spatial working memory task. *J. Neurophysiol.* 79, 2919–2940.

Cocco, S., Leibler, S., and Monasson, R. (2009). Neuronal couplings between retinal ganglion cells inferred by efficient inverse statistical physics methods. *Proc. Natl. Acad. Sci. U.S.A.* 106, 14058–14062. doi: 10.1073/pnas.0906705106

Compte, A., Brunel, N., Goldman-Rakic, P. S., and Wang, X.-J. (2000). Synaptic mechanisms and network dynamics underlying spatial working memory in a cortical network model. *Cereb. Cortex* 10, 910–923. doi: 10.1093/cercor/10.9.910

Cowan, N. (2001). The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *Behav. Brain Sci.* 24, 87–114. doi: 10.1017/S0140525X01003922

Dipoppa, M., and Gutkin, B. S. (2013). Flexible frequency control of cortical oscillations enables computations required for working memory. *Proc. Natl. Acad. Sci. U.S.A.* 110, 12828–12833. doi: 10.1073/pnas.1303270110

Ermentrout, G. B. (1996). Type I membranes, phase resetting curves, and synchrony. *Neural Comput.* 8, 979–1001. doi: 10.1162/neco.1996.8.5.979

Frank, M., Loughry, B., and O' Reilly, R. (2001). Interactions between frontal cortex and basal ganglia in working memory: a computational model. *Cogn. Affect. Behav. Neurosci.* 1, 137–160. doi: 10.3758/CABN.1.2.137

Funahashi, S., Bruce, C. J., and Goldman-Rakic, P. S. (1989). Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *J. Neurophysiol.* 61, 331–349.

Funahashi, S., and Inoue, M. (2000). Neuronal interactions related to working memory processes in the primate prefrontal cortex revealed by cross-correlation analysis. *Cereb. Cortex* 10, 535–551. doi: 10.1093/cercor/10.6.535

Fuster, J. M., and Alexander, G. E. (1971). Neuron activity related to short-term memory. *Science* 173, 652–654. doi: 10.1126/science.173.3997.652

Fuster, J. M., and Jervey, J. P. (1981). Inferotemporal neurons distinguish and retain behaviorally relevant features of visual stimuli. *Science* 212, 952–955. doi: 10.1126/science.7233192

Galán, R. F., Fourcaud-Trocmé, N., Ermentrout, G. B., and Urban, N. N. (2006). Correlation-induced synchronization of oscillations in olfactory bulb neurons. *J. Neurosci.* 26, 3646–3655. doi: 10.1523/JNEUROSCI.4605-05.2006

Gutkin, B. S., Laing, C. R., Colby, C. L., Chow, C. C., and Ermentrout, B. G. (2001). Turning on and off with excitation: the role of spike-timing asynchrony and synchrony in sustained neural activity. *J. Comput. Neurosci.* 11, 121–134. doi: 10.1023/A:1012837415096

Kreuz, T., Chicharro, D., Houghton, C., Andrzejak, R. G., and Mormann, F. (2013). Monitoring spike train synchrony. *J. Neurophysiol.* 109, 1457–1472. doi: 10.1152/jn.00873.2012

Laing, C. R., and Chow, C. C. (2001). Stationary bumps in networks of spiking neurons. *Neural Comput.* 13, 1473–1494. doi: 10.1162/089976601750264974

Lampl, I., Reichova, I., and Ferster, D. (1999). Synchronous membrane potential fluctuations in neurons of the cat visual cortex. *Neuron* 22, 361–374. doi: 10.1016/S0896-6273(00)81096-X

Lange, H., Thorner, G., and Hopf, A. (1976). Morphometric-statistical structure analysis of human striatum, pallidum, and nucleus subthalamicus: III. nucleus subthalamicus. *J. für Hirnforsch. (J. Hirnforsch.)* 17, 31–41.

Lee, H., Simpson, G. V., Logothetis, N. K., and Rainer, G. (2005). Phase locking of single neuron activity to theta oscillations during working memory in monkey extrastriate visual cortex. *Neuron* 45, 147–156. doi: 10.1016/j.neuron.2004.12.025

Lim, S., and Goldman, M. S. (2012). Noise tolerance of attractor and feedforward memory models. *Neural Comput.* 24, 332–390. doi: 10.1162/NECO_a_00234

Luck, S. J., and Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature* 390, 279–281. doi: 10.1038/36846

Ly, C., and Ermentrout, G. (2009). Synchronization dynamics of two coupled neural oscillators receiving shared and unshared noisy stimuli. *J. Comput. Neurosci.* 26, 425–443. doi: 10.1007/s10827-008-0120-8

Machens, C. K., Romo, R., and Brody, C. D. (2005). Flexible control of mutual inhibition: a neural model of two-interval discrimination. *Science* 307, 1121–1124. doi: 10.1126/science.1104171

Miller, E. K., Erickson, C. A., and Desimone, R. (1996). Neural mechanisms of visual working memory in prefrontal cortex of the macaque. *J. Neurosci.* 16, 5154–5167.

Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychol. Rev.* 63, 81–97. doi: 10.1037/h0043158

Miller, P., and Wang, X.-J. (2006). Inhibitory control by an integral feedback signal in prefrontal cortex: a model of discrimination between sequential stimuli. *Proc. Natl. Acad. Sci. U.S.A.* 103, 201–206. doi: 10.1073/pnas.0508072103

Moreno-Bote, R., Renart, A., and Nestor Parga, N. (2008). Theory of input spike auto- and cross-correlations and their effect on the response of spiking neurons. *Neural Comput.* 7, 1651–1705. doi: 10.1162/neco.2008.03-07-497

Pesaran, B., Pezaris, J. S., Sahani, M., Mitra, P. P., and Andersen, R. A. (2002). Temporal structure in neuronal activity during working memory in macaque parietal cortex. *Nat. Neurosci.* 5, 805–811. doi: 10.1038/nn890

Pipa, G., and Munk, M. H. J. (2011). Higher order spike synchrony in prefrontal cortex during visual memory. *Front. Comput. Neurosci.* 5, 1–13. doi: 10.3389/fncom.2011.00023

Pipa, G., Städtler, E. S., Rodriguez, E. F., Waltz, J. A., Muckli, L., Singer, W., et al. (2009). Performance- and stimulus-dependent oscillations in monkey prefrontal cortex during short-term memory. *Front. Integr. Neurosci.* 3:25. doi: 10.3389/neuro.07.025.2009

Polk, A., Litwin-Kumar, A., and Doiron, B. (2012). Correlated neural variability in persistent state networks. *Proc. Natl. Acad. Sci. U.S.A.* 109, 6295–6300. doi: 10.1073/pnas.1121274109

Riehle, A., Grün, S., Diesmann, M., and Aertsen, A. (1997). Spike synchronization and rate modulation differentially involved in motor cortical function. *Science* 278, 1950–1953. doi: 10.1126/science.278.5345.1950

Romo, R., Brody, C. D., Hernández, A., and Lemus, L. (1999). Neuronal correlates of parametric working memory in the prefrontal cortex. *Nature* 399, 470–473. doi: 10.1038/20939

Rosenbaum, R., Trousdale, J., and Josic, K. (2010). Pooling and correlated neural activity. *Front. Comput. Neurosci.* 4:9. doi: 10.3389/fncom.2010.00009

Rosenbaum, R., Trousdale, J., and Josic, K. (2011). The effects of pooling on correlated neural variability. *Front. Neurosci.* 5:58. doi: 10.3389/fnins.2011.00058

Sakamoto, K., Mushiake, H., Saito, N., Aihara, K., Yano, M., and Tanji, J. (2008). Discharge synchrony during the transition of behavioral goal representations encoded by discharge rates of prefrontal neurons. *Cereb. Cortex* 18, 2036–2045. doi: 10.1093/cercor/bhm234

Sakurai, Y., and Takahashi, S. (2006). Dynamic synchrony of firing in the monkey prefrontal cortex during

working-memory tasks. *J. Neurosci.* 26, 10141–10153. doi: 10.1523/JNEUROSCI.2423-06.2006

Salinas, E., and Sejnowski, T. J. (2001). Correlated neuronal activity and the flow of neural information. *Nat. Rev. Neurosci.* 2, 539–550. doi: 10.1038/35086012

Shadlen, M. N., and Newsome, W. T. (1994). Noise, neural codes and cortical organization. *Curr. Opin. Neurobiol.* 4, 569–579. doi: 10.1016/0959-4388(94)90059-0

Tallon-Baudry, C., Bertrand, O., Peronnet, F., and Pernier, J. (1998). Induced γ-band activity during the delay of a visual short-term memory task in humans. *J. Neurosci.* 18, 4244–4254.

Tsodyks, M., Kenet, T., Grinvald, A., and Arieli, A. (1999). Linking spontaneous activity of single cortical neurons and the underlying functional architecture. *Science* 286, 1943–1946. doi: 10.1126/science.286.5446.1943

van den Berg, R., Shin, H., Chou, W.-C., George, R., and Ma, W. J. (2012). Variability in encoding precision accounts for visual short-term memory limitations. *Proc. Natl. Acad. Sci. U.S.A.* 109, 8780–8785. doi: 10.1073/pnas.1117465109

Vogel, E. K., Woodman, G. F., and Luck, S. J. (2001). Storage of features, conjunctions, and objects in visual working memory. *J. Exp. Psychol. Hum. Percept. Perfor.* 27, 92–114. doi: 10.1037/0096-1523.27.1.92

Wilken, P., and Ma, W. J. (2004). A detection theory account of change detection. *J. Vis.* 4, 1120–1135. doi: 10.1167/4.12.11

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.

Received: 10 July 2013; accepted: 25 September 2013; published online: 21 October 2013.

Citation: Dipoppa M and Gutkin BS (2013) Correlations in background activity control persistent state stability and allow execution of working memory tasks. Front. Comput. Neurosci. 7:139. doi: 10.3389/fncom.2013.00139

This article was submitted to the journal Frontiers in Computational Neuroscience.

Copyright © 2013 Dipoppa and Gutkin. 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) or licensor 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.

frontiers in
COMPUTATIONAL NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 18 October 2013
doi: 10.3389/fncom.2013.00140

Single-unit activities during epileptic discharges in the human hippocampal formation

Catalina Alvarado-Rojas[1,2], Katia Lehongre[1,2], Juliana Bagdasaryan[1,2], Anatol Bragin[3], Richard Staba[3], Jerome Engel Jr.[3], Vincent Navarro[1,2,4] and Michel Le Van Quyen[1,2]*

[1] Centre de Recherche de l'Institut du Cerveau et de la Moelle Epinière, INSERM UMRS 975 - CNRS UMR 7225, Hôpital de la Pitié-Salpêtrière, Paris, France
[2] Université Pierre et Marie Curie, Paris, France
[3] Department of Neurology, David Geffen School of Medicine at UCLA, Los Angeles, CA, USA
[4] Epilepsy Unit, Groupe Hospitalier Pitié-Salpêtrière, Paris, France

Edited by:
Ruben Moreno-Bote, Foundation Sant Joan de Deu, Spain

Reviewed by:
Emili Balaguer-Ballester, Bournemouth University, UK
Abdelmalik Moujahid, University of the Basque Country UPV/EHU, Spain

***Correspondence:**
Michel Le Van Quyen, Centre de Recherche de l'Institut du Cerveau et de la Moelle épinière, INSERM UMRS 975 - CNRS UMR 7225, Hôpital de la Pitié-Salpêtrière, 47 Bd de l'Hôpital, 75651 Paris, Cedex 13, France
e-mail: quyen@ t-online.de

Between seizures the brain of patients with epilepsy generates pathological patterns of synchronous activity, designated as interictal epileptiform discharges (ID). Using microelectrodes in the hippocampal formations of 8 patients with drug-resistant temporal lobe epilepsy, we studied ID by simultaneously analyzing action potentials from individual neurons and the local field potentials (LFPs) generated by the surrounding neuronal network. We found that ∼30% of the units increased their firing rate during ID and 40% showed a decrease during the post-ID period. Surprisingly, 30% of units showed either an increase or decrease in firing rates several hundred of milliseconds before the ID. In 4 patients, this pre-ID neuronal firing was correlated with field high-frequency oscillations at 40–120 Hz. Finally, we observed that only a very small subset of cells showed significant coincident firing before or during ID. Taken together, we suggested that, in contrast to traditional views, ID are generated by a sparse neuronal network and followed a heterogeneous synchronization process initiated over several hundreds of milliseconds before the paroxysmal discharges.

Keywords: interictal epileptiform discharges, microelectrode recordings, multiunit activity, temporal lobe epilepsy, spike synchronization

INTRODUCTION

Synchronization of local and distributed neuronal assemblies is thought to underlie fundamental brain processes such as perception, learning, and cognition (Varela et al., 2001). In neurological diseases, neuronal synchrony can be altered and in epilepsy may play an important role in enhanced cellular excitability (Jasper and Penfield, 1954). Besides ictal events or seizures, interictal discharges (ID) are a typical signature of abnormal neuronal synchronization, seen spontaneously between seizures in scalp and intracranial EEG. They are used as a clinical indicator for the location of the epileptogenic zone, the region that generates seizures. Furthermore, it is believed that this region contains both, the seizure onset zone and the surrounding "irritative zone," which generates ID and limits with normal tissue (Talairach and Bancaud, 1966). These transient epileptic synchronization events are characterized by a large-amplitude, rapid component lasting 50–100 ms that is usually followed by a slow wave of 200–500 ms duration (de Curtis and Avanzini, 2001). In some cases, they are associated with an oscillation in the high frequency range greater than 40 Hz (Bragin et al., 1999; Jacobs et al., 2011; Le Van Quyen, 2012). Despite their fundamental importance in diagnosing and treating epilepsy, little is known about the neurophysiological mechanisms generating these events in the human brain. Experimental work on animals and human tissue propose the paroxysmal depolarization shift (PDS) as the cellular correlate of ID (Prince and Wong, 1981; Avoli and Williamson, 1996). This event is defined as a burst of action potentials on a large depolarization, followed by a longer hyperpolarization. However, *in vivo* human evidence is scarce, because of the limited opportunities to study the behavior of single neurons in human subjects. To overcome this difficulty, epilepsy patients suitable for surgical treatment are sometimes studied with intracranial depth electrodes in order to record EEG activity from deep cortical structures and accurately identify the regions originating seizures. Using depth electrodes specially adapted with microelectrodes (Fried et al., 1997; **Figure 1A**), ID can be studied by simultaneously recording action potentials from individual neurons and the local field potentials (LFP). Studies using microelectrode technology, have reported a variable and complex relation between ID and the activity of individual neurons, more heterogeneous than simple PDS (Babb et al., 1973; Wyler et al., 1982; Ulbert et al., 2004; Keller et al., 2010; Alarcon et al., 2012). In particular, a large diversity of neuronal response were found including an increase or decrease in their firing rates or even changes in firing that precede the defining interictal discharge itself. Most of these studies were performed on patients with neocortical epilepsy that exhibit a wide range of heterogeneity. In the present work, we recorded ID in the hippocampal formation of 8 patients with drug-resistant mesial temporal lobe epilepsy. Our objective is to describe firing patterns and neuronal synchronization of single-unit activities during spontaneous IDs.

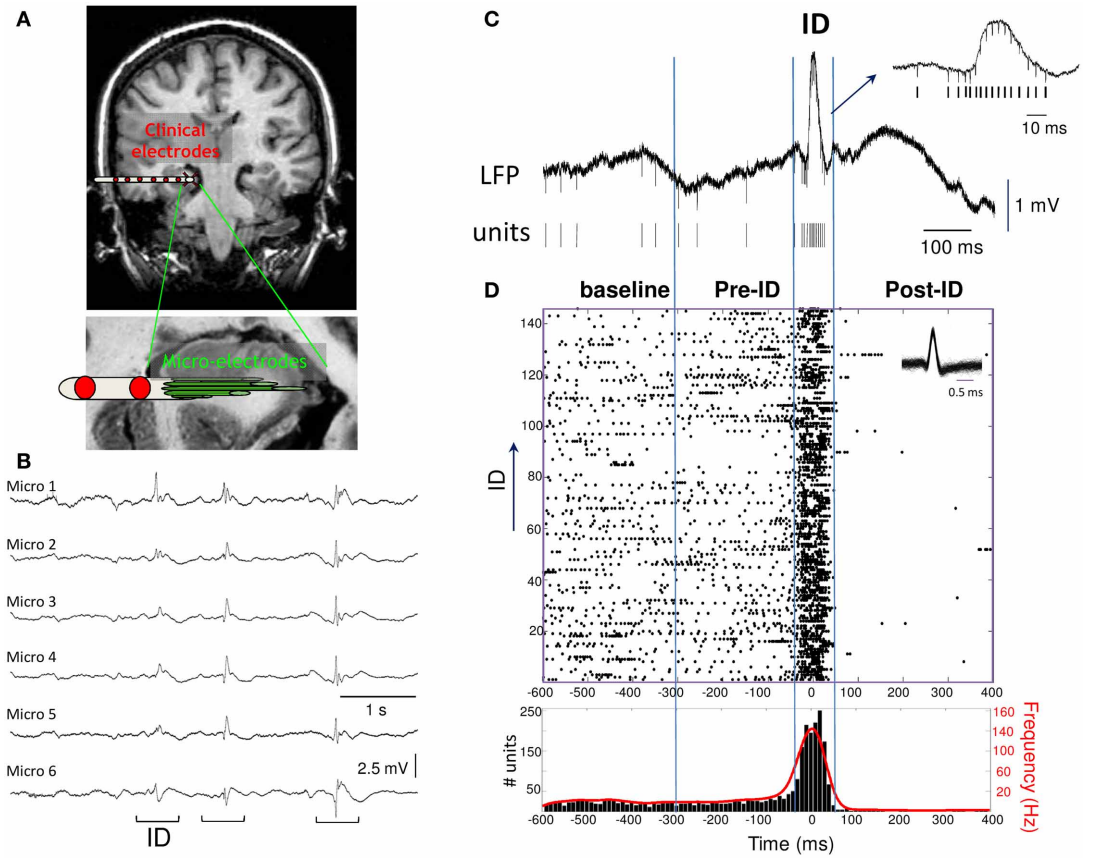


FIGURE 1 | (A) Macro- and micro-electrodes superimposed on a magnetic resonance imaging scan. Nine microwires (40 μm diameter) extend beyond the tip of each macro-electrode and record the hippocampal formation. **(B)** Interictal discharges (ID) recorded with microelectrode local field potentials from adjacent electrodes in the hippocampus of a patient. **(C)** Example of wide-band recording of an ID event with the corresponding extracted single unit activities. **(D)** Raster plot and peri-event histogram (bin size, 10 ms) of the single unit activity shown above. Note the strong changes in the firing rate and instantaneous frequency (red) during the ID.

MATERIALS AND METHODS

DATABASE

Subjects were 8 patients [two female, mean age ± standard deviation (SD) 36.3 ± 10.5 years] with pharmacologically intractable temporal lobe epilepsy who were implanted with 8–14 intracranial depth electrodes in order to localize epileptogenic regions for possible resection. The placement of the electrodes was determined exclusively by clinical criteria (Fried et al., 1999). Extending beyond the tip of each electrode were nine Pt-Ir microwires (40 μm diameter) with inter-tip spacing of 500 μm, eight active recording channels and one reference. Each microwire was sampled at 28 kHz (Cheetah recording system, Neuralynx Inc., Tucson, AZ). Spatial localizations were determined on the basis of postimplant computed tomography scans coregistered with preimplant 1.5T MRI scans. Our results are based on microelectrode recordings located in the anterior hippocampus (n = 40 channels in 5 patients) and entorhinal cortex (n = 24 channels in 3 patients). The recording states were quiet wakefulness and slow waves sleep (stages 1–4). All studies conformed to the guidelines of the Medical Institutional Review Board at University of California, Los Angeles.

SPIKE SORTING

In order to detect single-units, all channels were high-pass filtered at 300 Hz and were visually examined for the presence of unit activities. In those microwires with clear unit activities, we performed spike detection (>4:1 signal to noise ratio) to obtain multi-unit activities (MUA). Single-unit activities were extracted with spike sorting using KlustaKwik 1.7 program (Software: http://klustakwik.sourceforge.net/; Harris et al., 2000) which employs the 10 principal components of the spike shape and an unsupervised Conditional Expectation Maximization (CEM) clustering algorithm (Hazan et al., 2006). After automatic clustering, the clusters containing non-spike waveforms were visually deleted and then the units were further isolated using a manual cluster cutting method. Only units with clear boundaries and less than 0.5% of spike intervals within a 1 ms refractory period are included in the present analysis. Typically we isolated 1 or 2 distinct neurons from each microwire, but in several cases we observed up to 4 distinct neurons from a single microwire. The instantaneous spike frequency was measured by convolving the timing of each unit with a Gaussian function of standard deviation of 20 ms (Ts = 1 ms), set close to the modal interspike

interval (Le Van Quyen et al., 2008, 2010). This operation leads to an analog trace of the instantaneous firing rate (Paulin, 1996).

OSCILLATION ANALYSIS

LFP are complementary to action potential information and have shown prominent oscillatory activity within the high-frequency frequency range from 40 to 300 Hz (Worrell et al., 2012). A wavelet time-frequency analysis was used to determine precisely the mean frequency, maximum amplitude and onset and offset of these LFP oscillations. The advantage of the wavelet analysis lies in the fact that the time resolution is variable with frequency, so that high frequencies have a sharper time resolution (Le Van Quyen and Bragin, 2007). The Complex Morlet wavelet was here applied that uses a wave-like scalable function that is well-localized in both time and frequency:

$$\Psi_{\tau,f}(u) = \sqrt{f}\exp(j2\pi f(u-\tau))\exp\left(-\frac{(u-\tau)^2}{2\sigma^2}\right).$$

This wavelet represents the product of a sinusoidal wave at frequency f, with a Gaussian function centered at time τ, with a standard deviation σ proportional to the inverse of f. The wavelet coefficients of a signal $x(t)$ as a function of time (τ) and frequency (f) are defined as: $W(\tau, f) = \int_{-\infty}^{+\infty} x(u)\Psi_{\tau,f}(u)du$. It depends solely on σ, which sets the number of cycles of the wavelet: $nco = 6f\sigma$. The value nco determines the frequency resolution of the analysis by setting the width of the frequency interval for which phase are measured. Here, we chose $nco = 5$. For baseline correction, the average and SD of power were first computed at each frequency of the baseline period. Then, the average baseline power was subtracted from all time windows at each frequency, and the result scaled by 1/SD, yielding baseline-adjusted Z scores. Significant increases with respect to baseline activity showed up as positive Z-values and tabulated probability values indicate that, for absolute values of $Z > 3.09$, we have $P < 0.001$. The Kolmogorov–Smirnov test was applied to assess the distribution normality of the wavelet coefficients, using a 0.05 probability level.

SPIKE SYNCHRONIZATION

Different measures exist to detect and quantify synchronization between spike trains (Brown et al., 2004; Kreuz et al., 2007). In this study, we used two complementary techniques: (1) Cross-correlation analysis was performed for cell pairs (Perkel et al., 1967; Amarasingham et al., 2012). To evaluate the significance of the correlation, we used a boot-strap method that accounts for the firing rate changes of the neurons (Hatsopoulos et al., 2003; Grün, 2009). Since the widths of the peaks in the original cross-correlograms were typically in the range of 5–30 ms (Krüger and Mayer, 1990), the spikes were jittered by adding a random value from a normal distribution with a 50-ms SD and 0 mean to the spike times. For each cell pair, 1000 jittered spike trains were created, and the expected cross-correlogram (and 99% confidence interval) was estimated on 1 ms time bins. For any given cell pair where at least one bin in the [1.5 ms, 30 ms] interval exceeded the 99% confidence interval, the interaction was considered significant. (2) A method for identifying statistically conspicuous spike coincidences was implemented to detect the number of quasi-simultaneous appearances of spikes over small coincidence windows, here of 5 ms (Gütig et al., 2002; Quian Quiroga et al., 2002). Their occurrence was then studied in relation to surrogate data generated by dithering the individual, original spike times within a given time interval. Here, each spike in the original data set was randomly and independently jittered on a uniform interval of $[-5, +5]$ ms to form a surrogate dataset. By repeating the procedure 1000 times, the 99.9% confidence interval for each bin ($p = 0.001$) was calculated.

RESULTS

Microelectrode recordings were selected by an expert electroencephalographer to have very abundant and persistent ID in the hippocampus (5 patients) or entorhinal cortex (3 patients) during quiet wakefulness or slow-wave sleep (recording durations from 10 to 118 min; total recording time: 6 h). All ID were recorded in the epileptic zone and appeared as spatially synchronous patterns emerging at about the same time on the same bundle of microelectrodes (**Figure 1B**). A standard, threshold-based ID detector was performed to automatically detect, from the LFP, events showing a pointed peak with a large amplitude, large slope and duration of 20–100 ms, appearing at a frequency of 0.07 ± 0.30 Hz (range: 0.01–0.21 Hz). After expert visual confirmation, 862 ID were identified showing a large pattern of morphological characteristics typical for sharp waves, spikes and spike-wave discharges (Niedermeyer, 2005). Events were aligned by the sharpest peak of the discharge (**Figure 1C**). In order to analyze the patterns of neuronal activity around the discharge, we defined a baseline period (–600 to –300 ms), pre-ID period (–300 to –50 ms), the interictal discharge (–50 to 50 ms), the post-ID period (50–400 ms). The activities of different neurons per microelectrode were identified with a spike sorting algorithm and a total of 75 single units were selected for analysis. To visualize the discharge-related activity of single neurons, peri-stimulus raster plots and timing histograms were constructed for the period of 1 s before and after each event (**Figure 1D**).

During the ID period (–50 to 50 ms), we found that around 40% of the recorded units showed some change in firing, whereas 60% remain unchanged. About 32% increased their firing rate more than 2 times during ID relative to baseline epochs [**Figure 2B**; right-tail t-test: $T_{(23)} = 1.78$; $p = 0.04$; an example of a cell can be seen in **Figure 1D**]. The firing rate of these cells showed a considerable degree of variability (range from 1.4 to 99 Hz) with a mean of 9.4 ± 19.7 Hz during ID (baseline: 2.7 ± 3.1 Hz). During the post-ictal period, 40% of units decreased firing by half [50–400 ms, mean firing rate: 1.8 ± 2.7 Hz and baseline: 7.0 ± 2.7 Hz, left-tail t-test: $T_{(29)} = -3.73$; $p = 4.1 \cdot 10^{-4}$, **Figure 2C**]. In addition to this modulated single unit activity during ID, many units showed a significant change in firing preceding the interictal discharge. From 30% of single units that significantly changed during the pre-ID period, 12% increased [mean firing rate: 10.0 ± 13.5 Hz and baseline: 4.2 ± 5.8 Hz; $T_{(8)} = -3.45$; $p = 0.004$] and 18% decreased [mean firing rate: 2.3 ± 7.0 Hz and baseline: 5.2 ± 11.6 Hz; $T_{(13)} = -1.64$; $p = 0.06$] their firing rate (–300 to –50 ms, **Figure 2A**; examples are given in **Figure 2D**). On the corresponding channels, we were interested in the relationship between these pre-ID firing changes and

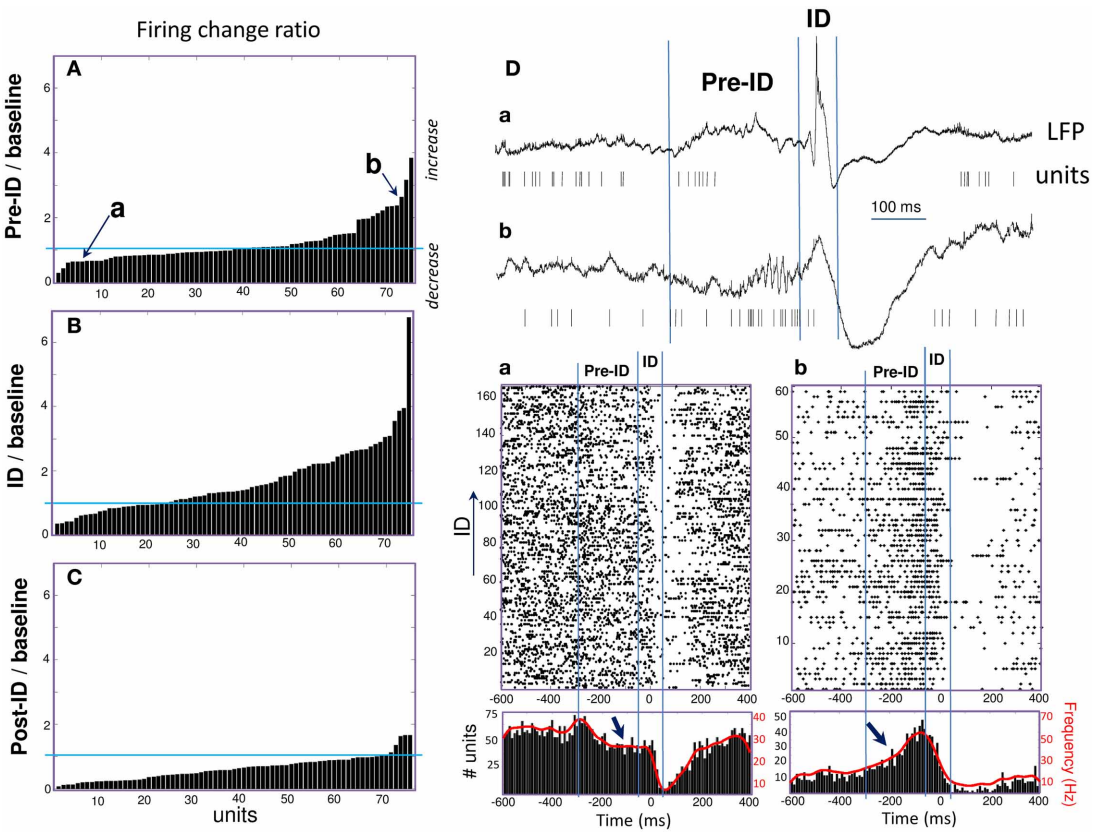


FIGURE 2 | (A–C) Peri-ID firing changes of single units defined as the ratio of changes in discharge probability between the baseline and pre-ID, ID or post-ID. **(D)** Examples of two units (Top: raw data; Bottom: Raster plots and peri-event histograms) recorded in different patients and showing significant decrease **(a)** or increase **(b)** during the pre-ID period (see arrows).

LFP (<300 Hz). Spectral power was performed by using Morlet wavelet analysis (20–300 Hz) and pre-ID changes in LFP were tested for significant increases/decreases from baseline of specific frequency bands ($p < 0.001$). In 4 subjects we observed that pre-ID neuronal firing pattern was correlated with an increase in high-frequency oscillations between 40 and 120 Hz (mean peak from baseline SD: $Z = 6.1$, range from 4.3 to 9.1). **Figure 3** shows average time-frequency representations around the ID for the two patients of **Figure 2D**. Main changes in spectral power can be seen in the LFP preceding the interictal discharge and correlate closely with the increase or decrease in neuronal firing.

Finally, we analyzed unit synchronization during ID between pairs of units simultaneously recorded in two different microelectrodes. Because of the inter-tip spacing of 500 μm, the units are assumed to reflect adjacent but different neuronal populations. Two complementary methods have been used to address the synchrony between spike trains. First, analysis of cross-correlograms between pairs of units was performed for each cell pairs that showed a sufficient number of spikes (>100) during ID. The significance of the correlation was obtained by jittering each pair of spike trains and by computing the 99% confidence interval. Of the 120 cross-correlograms constructed, only 5 cross-correlograms (about 4%) had a significant peak that occurred within ±25 ms around the origin, indicating that these neuronal pairs were discharging in a correlated way. **Figure 4** (top) illustrates examples of significant peaks in cross-correlograms of two units. In addition to cross-correlation analysis, we also analyzed the overall level of synchronicity from the number of quasi simultaneous appearances of spikes. In order to not overestimate the number of random synchronous spikes due to the elevated firing rate, we used jitter techniques to infer millisecond-precise temporal synchrony (Hatsopoulos et al., 2003). Here, spikes of one of the pairs of neurons were time jittered by ±5 ms to generate jittered peri-stimulus raster plots of unit coincidence that could be used to assess the statistical significance of bin fluctuations in the non-jittered spike series. Because the jittered data sets preserve firing rates on timescales much broader than that of the jitter interval (in this case, 5 ms), the overall effect of the analysis is to identify those pairs that showed excessive co-firing at short latencies that cannot be accounted for by firing rates varying at timescales of tens of milliseconds. Despite the strong increase in about 30% of the recorded units during ID, only a very small subset of cells (18 of 120 analyzed pairs, about 15%) showed significant coincident firing before or during ID. For two patients, **Figure 4** (bottom) illustrates pairs of units that showed significant coincident firing ($p = 0.001$) during ID (A) and the pre-ID period (B).

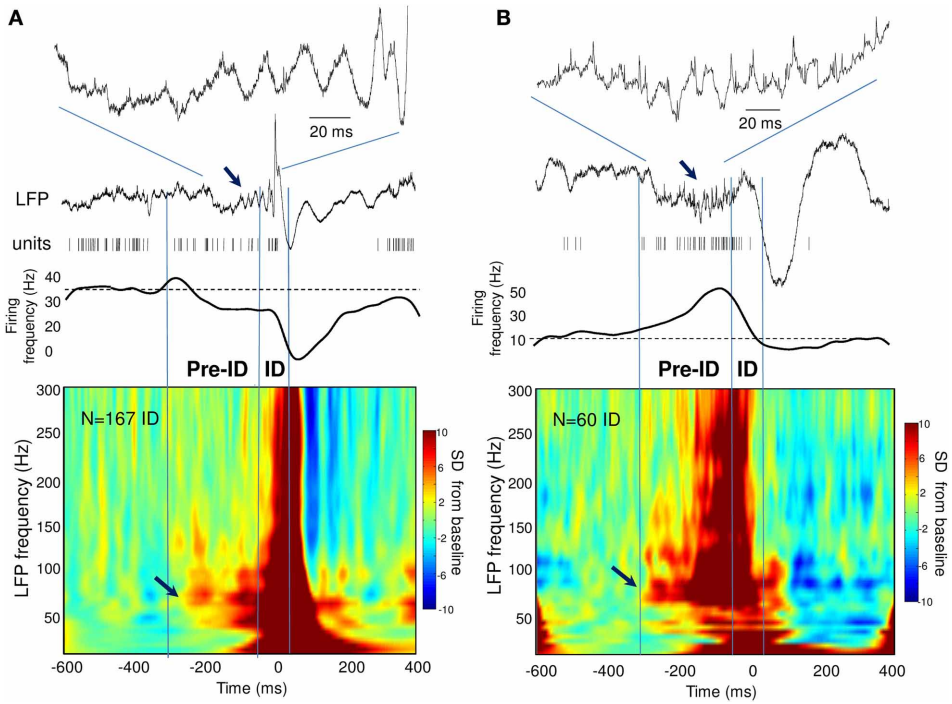


FIGURE 3 | (A,B) Time-frequency representations of the LFP around the ID for two patients showing pre-ID changes in neuronal firing. Note the increase in LFP high-frequency oscillations between 40 and 120 Hz preceding the ID and closely correlated with the decrease **(A)** or increase **(B)** in neuronal firing (see arrows).

DISCUSSION

We found that a large subset of the recorded units showed significant changes in firing in or around ID in the hippocampal formation of patients with mesial temporal epilepsy. Around 30% of the unit increased their firing rate during ID while 40% showed a decrease during the post-ID period. This percentage of modulated neurons agrees with that described by Wyler et al. (1982), who found that 44% of recorded neurons showed primarily an increase in firing rate near the interictal discharge peak. Surprisingly, a subset of 30% of units showed significant firing rate variations several hundred of milliseconds before the ID. In a few patients, we observed that this neuronal firing pattern was related with elevated LFP oscillations at 40–120 Hz. Finally, based on two statistical methods that identify spike synchronization, we found that only a very small subset of cells showed significant coincident firing before or during ID.

Our observations of neuronal firing during the interictal discharge are consistent with the paroxysmal depolarizing shift (PDS) mechanism—a large depolarization phase followed by a long hyperpolarization—that have been studied in animal models of epilepsy (Matsumoto and Marsan, 1964; Prince, 1968). The first part of the depolarization phase is believed to be generated by intrinsic membrane conductance (de Curtis et al., 1999), and the later from feedback recurrent synaptic excitation mediated by AMPA and NMDA receptor subtypes, and glutamate receptor-coupled calcium conductances (Traub et al., 1993). Thus, PDS has been shown to be the result of giant excitatory postsynaptic potentials. The PDS is usually followed by a hyperpolarization, which represents GABA-mediated recurrent inhibition, as well as Ca^{2+}-dependent K^{+} currents. Interestingly, consistent with *in vitro* studies on hippocampal slices from human patients with temporal lobe epilepsy (Cohen et al., 2002; Wittner et al., 2009), the presence of a similar suppression of unit activities in our *in vivo* data suggests that IDs can occur in cortical regions maintaining substantial inhibitory function.

However, in contrast to simple models of PDS and in line with other observations in human epileptic neocortex (Keller et al., 2010), we found that ID, rather than requiring a large synchronization of neurons, can occur with relatively sparse single neuron participation (estimated at about 30% of the cells). Furthermore, a small subset of the units significantly increased or decreased their firing well before ID. Concomitant with changes in firing rate for certain neurons, at least in some patients, high-frequency oscillations at 40–120 Hz can be seen in the LFP preceding the ID and correlate closely with the changes in neuronal firing. Because interneurons are involved in the generation of high frequency oscillations through mechanisms of post-inhibition resetting of neuronal firing (Cobb et al., 1995; Ylinen et al., 1995; Le Van Quyen et al., 2008; Le Van Quyen, 2012), it is here tempting to speculate that GABA-mediated events may contribute to enhance synchronization of local epileptic networks before ID. Interestingly, emerging evidence indicates that GABA promotes epileptiform synchronization (Pavlov et al., 2013). For instance, GABA receptor-mediated inhibition can facilitate thalamocortical processes leading to the occurrence of generalized spike and wave discharges that occur during absence seizures (Danober et al., 1998). Following a similar mechanism, ID may be caused by a rebound synchronization of cells that may start firing

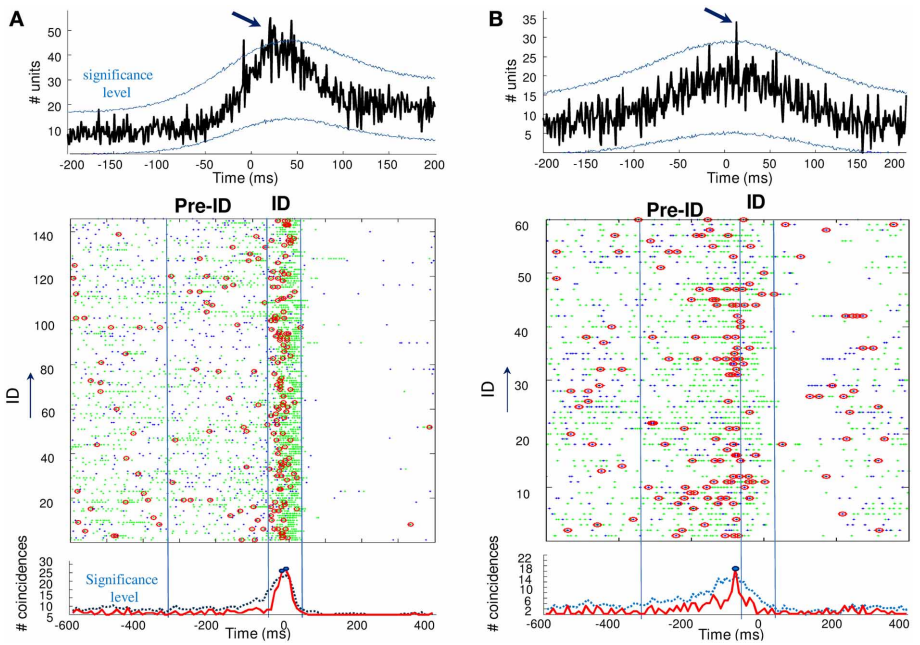


FIGURE 4 | Top: Cross-correlograms between pairs of units during ID in two patients (A,B). The blue lines are the significance levels computed from 1000 jittered spike trains. In both cases, the center peak exceeds the significance level (arrows) and the pairs of units are considered to be significantly correlated. Bottom: Unit synchronizations (red circles) were defined as coincidences between the two units (green and blue points) occurring over a 5-ms interval. Note the significant increase in coincidences during ID **(A)** and the pre-ID period **(B)**, over the statistical threshold defined by a random jitter of the original data.

synchronously shortly after inhibition ceases and permit the fast component of the ID. Moreover, intense synaptic activation of $GABA_A$ receptors in the hippocampus can lead to a shift in GABAergic neurotransmission from inhibitory to excitatory, contributing to epileptic discharges (Kohling et al., 2000; Cohen et al., 2002). Interestingly, pre-event changes have also been seen in advance of seizures in an animal model of temporal lobe epilepsy (Bower and Buckmaster, 2008) and around seizure onset in human epilepsy (Babb and Crandall, 1976; Truccolo et al., 2011), suggesting a possible similar mechanism before seizures.

Taken together, our data suggest that ID in patients with temporal lobe epilepsy is not a simple paroxysm of hypersynchronous excitatory activity, but rather represents a heterogeneous synchronization process possibly initiated by GABAergic responses in small subsets of cells and emerging over hundreds of milliseconds before the paroxysmal discharges.

ACKNOWLEDGMENTS

Catalina Alvarado-Rojas was supported by the Administrative Department for Science, Technology and Innovation (COLCIENCIAS), Colombia. Vincent Navarro was supported by a Contrat Inferface INSERM. This work was supported by funding from the program "Investissements d'avenir" ANR-10-IAIHU-06 and from the ICM and OCIRP.

REFERENCES

Alarcon, G., Martinez, J., Kerai, S. V., Lacruz, M. E., Quiroga, R. Q., Selway, R. P., et al. (2012). *In vivo* neuronal firing patterns during human epileptiform discharges replicated by electrical stimulation. *Clin. Neurophysiol.* 123, 1736–1744. doi: 10.1016/j.clinph.2012.02.062

Amarasingham, A., Harrison, M. T., Hatsopoulos, N. G., and Geman, S. (2012). Conditional modeling and the jitter method of spike resampling. *J. Neurophysiol.* 107, 517–531. doi: 10.1152/jn.00633.2011

Avoli, M., and Williamson, A. (1996). Functional and pharmacological properties of human neocortical neurons maintained *in vitro*. *Prog. Neurobiol.* 48, 519–554. doi: 10.1016/0301-0082(95)00050-X

Babb, T. L., Carr, E., and Crandall, P. H. (1973). Analysis of extracellular firing patterns of deep temporal lobe structures in man. *Electroencephalogr. Clin. Neurophysiol.* 34, 247–257. doi: 10.1016/0013-4694(73)90252-6

Babb, T. L., and Crandall, P. H. (1976). Epileptogenesis of human limbic neurons in psychomotor epileptics. *Electroencephalogr. Clin. Neurophysiol.* 40, 225–243. doi: 10.1016/0013-4694(76)90147-4

Bower, M. R., and Buckmaster, P. S. (2008). Changes in granule cell firing rates precede locally recorded spontaneous seizures by minutes in an animal model of temporal lobe epilepsy. *J. Neurophysiol.* 99, 2431–2442. doi: 10.1152/jn.01369.2007

Bragin, A., Engel, J. Jr., Wilson, C. L., Fried, I., and Mathern, G. W. (1999). Hippocampal and entorhinal cortex high frequency oscillations (100–500 Hz) in human epileptic brain and in kainic acid-treated rats with chronic seizures. *Epilepsia* 40, 127–137. doi: 10.1111/j.1528-1157.1999.tb02065.x

Brown, E. N., Mitra, P. P., and Kass, R. E. (2004). Multiple neural spike train data analysis: State-of-the-art and future challenges. *Nat. Neurosci.* 7, 456–461. doi: 10.1038/nn1228

Cobb, S. R., Buhl, E. H., Halasy, K., Paulsen, O., and Somogyi, P. (1995). Synchronization of neuronal activity in hippocampus by individual GABAergic interneurons. *Nature* 378, 75–78. doi: 10.1038/378075a0

Cohen, I., Navarro, V., Clemenceau, S., Baulac, M., and Miles, R. (2002). On the origin of interictal activity in human temporal lobe epilepsy *in vitro*. *Science* 298, 1418–1421. doi: 10.1126/science.1076510

Danober, L., Deransart, C., Deapulis, A., Vergnes, M., and Marescaux, C. (1998). Pathophysiological mechanisms of genetic absence epilepsy in the rat. *Prog. Neurobiol.* 55, 27–57. doi: 10.1016/S0301-0082(97)00091-9

de Curtis, M., and Avanzini, G. (2001). Interictal spikes in focal epileptogenesis. *Prog. Neurobiol.* 63, 541–567. doi: 10.1016/S0301-0082(00)00026-5

de Curtis, M., Radici, C., and Forti, M., (1999). Cellular mechanisms underlying spontaneous interictal spikes in a model of focal cortical epileptogenesis. *Neuroscience* 88, 107–117. doi: 10.1016/S0306-4522(98)00201-2

Fried, I., MacDonald, K. A., and Wilson, C. L. (1997). Single neuron activity in human hippocampus and amygdala during recognition of faces and objects. *Neuron* 18, 753–765. doi: 10.1016/S0896-6273(00)80315-3

Fried, I., Wilson, C. L., Maidment, N. T., Engel, J., Behnke, E., Fields, T. A., et al. (1999). Cerebral microdialysis combined with single-neuron and electroencephalographic recording in neurosurgical patients. *J. Neurosurg.* 91, 697–705. doi: 10.3171/jns.1999.91.4.0697

Grün, S. (2009). Data-driven significance estimation for precise spike correlation. *J. Neurophysiol.* 101, 1126–1140. doi: 10.1152/jn.00093.2008

Gütig, R., Aertsen, A., and Rotter, S. (2002). Statistical significance of coincident spikes: count-based versus rate-based statistics. *Neural Comput.* 14, 121–153. doi: 10.1162/089976602753284473

Harris, K. D., Henze, D. A., Csicsvari, J., Hirase, H., and Buzsáki, G. (2000). Accuracy of tetrode spike separation as determined by simultaneous intracellular and extracellular measurements. *J. Neurophysiol.* 84, 401–414.

Hatsopoulos, N., Geman, S., Amarasingham, A., and Bienenstock, E. (2003). At what time scale does the nervous system operate? *Neurocomputing* 52–54, 25–29. doi: 10.1016/S0925-2312(02)00773-7

Hazan, L., Zugaro, M., and Buzsáki, G. (2006). Klusters, NeuroScope, NDManager: A free software suite for neurophysiological data processing and visualization. *J. Neurosci. Methods* 155, 207–216. doi: 10.1016/j.jneumeth.2006.01.017

Jacobs, J., Kobayashi, K., and Gotman, J. (2011). High-frequency changes during interictal spikes detected by time-frequency analysis. *Clin. Neurophysiol.* 122, 32–42. doi: 10.1016/j.clinph.2010.05.033

Jasper, H., and Penfield, W. (1954). *Epilepsy and the Functional Anatomy of the Human Brain*. New York, NY: Little, Brown and Co.

Keller, C. J., Truccolo, W., Gale, J. T., Eskandar, E., Thesen, T., Carlson, C., et al. (2010). Heterogeneous neuronal firing patterns during interictal epileptiform discharges in the human cortex. *Brain* 133, 1668–1681. doi: 10.1093/brain/awq112

Kohling, R., Vreugdenhil, M., Bracci, E., and Jefferys, J. G. (2000). Ictal epileptiform activity is facilitated by hippocampal GABAA receptor-mediated oscillations. *J. Neurosci.* 20, 6820–6829.

Kreuz, T., Haas, J. S., Morelli, A., Abarbanel, H. D. I., and Politi, A. (2007). Measuring spike train synchrony. *J. Neurosci. Methods* 165, 151–161. doi: 10.1016/j.jneumeth.2007.05.031

Krüger, J., and Mayer, M. (1990). Two types of neuronal synchrony *in* monkey striate cortex. *Biol. Cybern.* 64, 135–140. doi: 10.1007/BF02331342

Le Van Quyen, M. (2012). Editorial "Special issue on High-frequency oscillations in cognition and epilepsy". *Prog. Neurobiol.* 98, 239–318. doi: 10.1016/j.pneurobio.2012.06.009

Le Van Quyen, M., and Bragin, A. (2007). Analysis of dynamic brain oscillations: methodological advances. *Trends Neurosci.* 30, 365–373. doi: 10.1016/j.tins.2007.05.006

Le Van Quyen, M., Bragin, A., Staba, R., Crepon, B., Wilson, C. L., and Engel, J. Jr. (2008). Cell type-specific firing during ripple oscillations in the hippocampal formation of humans. *J. Neurosci.* 28, 6104–6110. doi: 10.1523/JNEUROSCI.0437-08.2008

Le Van Quyen, M., Staba, R., Bragin, A., Dickson, C., Valderrama, M., Fried, I., et al. (2010). Large-scale microelectrode recordings of high frequency gamma oscillations in human cortex during sleep. *J. Neurosci.* 30, 7770–7782. doi: 10.1523/JNEUROSCI.5049-09.2010

Matsumoto, H., and Marsan, C. A. (1964). Cortical cellular phenomena in experimental epilepsy: interictal manifestations. *Exp. Neurol.* 80, 286–304. doi: 10.1016/0014-4886(64)90025-1

Niedermeyer, E. (2005). "Abnormal EEG patterns: epileptic and paroxysmal," in *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*, eds E. Niedermeyer and F. Lopes da Silva (Philadelphia, PA: Lippincott Williams and Wilkins).

Paulin, M. G. (1996). "System identification of spiking sensory neurons using realistically constrained nonlinear time series models," in *Advances in Processing and Pattern Analysis of Biological Signals,* eds I. Gath and G. Inbar (New York, NY: Plenum), 183–194. doi: 10.1007/978-1-4757-9098-6_13

Pavlov, I., Kaila, K., Kullmann, D. M., and Miles, R. (2013). Cortical inhibition, pH and cell excitability in epilepsy: what are optimal targets for antiepileptic interventions? *J. Physiol.* 591, 765–774. doi: 10.1113/jphysiol.2012.237958

Perkel, D., Gerstein, G., and Moore, G. (1967). Neuronal spike trains and stochastic point processes. II. Simultaneous spike trains, *Biophys. J.* 7, 419–440. doi: 10.1016/S0006-3495(67)86597-4

Prince, D. (1968). Inhibition in 'Epileptic' neurons. *Exp. Neurol.* 21, 307–321. doi: 10.1016/0014-4886(68)90043-5

Prince, D. A., and Wong, R. K. (1981). Human epileptic neurons studied *in vitro*. *Brain Res.* 210, 323–333. doi: 10.1016/0006-8993(81)90905-7

Quian Quiroga, R., Kreuz, T., and Grassberger, P. (2002). Event synchronization: a simple and fast method to measure synchronicity and time delay patterns. *Phys. Rev. E.* 66, 041904. doi: 10.1103/PhysRevE.66.041904

Talairach, J., and Bancaud, J. (1966). Lesion, "irritative" zone and epileptogenic focus. *Confin. Neurol.* 27, 91–94. doi: 10.1159/000103937

Traub, R. D., Miles, R., and Jefferys, J. G. R., (1993). Synaptic and intrinsic conductances shape picrotoxininduced synchronized afterdischarges in the guinea pig hippocampal slice. *J. Physiol.* 461, 525–547.

Truccolo, W., Donoghue, J. A., Hochberg, L. R., Eskandar, E. N., Madsen, J. R., Anderson, W. S., et al. (2011). Single-neuron dynamics in human focal epilepsy. *Nat. Neurosci.* 14, 635–641. doi: 10.1038/nn.2782

Ulbert, I., Heit, G., Madsen, J., Karmos, G., and Halgren, E. (2004). Laminar analysis of human neocortical interictal spike generation and propagation: current source density and multiunit analysis *in vivo*. *Epilepsia* 45, 48–56. doi: 10.1111/j.0013-9580.2004.04011.x

Varela, F., Lachaux, J. P., Rodriguez, E., and Martinerie, J. (2001). The brainweb: phase synchronization and large-scale integration. *Nat. Rev. Neurosci.* 2, 229–239. doi: 10.1038/35067550

Wittner, L., Huberfeld, G., Clemenceau, S., Eross, L., Dezamis, E., Entz, L., et al. (2009). The epileptic human hippocampal cornu ammonis 2 region generates spontaneous interictal like activity *in vitro*. *Brain* 132, 3032–3046. doi: 10.1093/brain/awp238

Worrell, G. A., Jerbi, K., Kobayashi, K., Lina, J. M., Zelmann, R., and Le Van Quyen, M. (2012). Recording and analysis techniques for high-frequency oscillations. *Prog. Neurobiol.* 98, 265–278. doi: 10.1016/j.pneurobio.2012.02.006

Wyler, A. R., Ojemann, G. A., and Ward, A. A. (1982). Neurons in human epileptic cortex: correlation between unit and EEG activity. *Ann. Neurol.* 11, 301–308. doi: 10.1002/ana.410110311

Ylinen, A., Bragin, A., Nadasdy, Z., Jando, G., Szabo, I., Sik, A., et al. (1995). Sharp wave-associated high-frequency oscillation (200 Hz) in the intact hippocampus: network and intracellular mechanisms. *J. Neurosci.* 15, 30–46.

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.

Received: 31 May 2013; accepted: 27 September 2013; published online: 18 October 2013.

Citation: Alvarado-Rojas C, Lehongre K, Bagdasaryan J, Bragin A, Staba R, Engel Jr J, Navarro V and Le Van Quyen M (2013) Single-unit activities during epileptic discharges in the human hippocampal formation. Front. Comput. Neurosci. 7:140. doi: 10.3389/fncom.2013.00140

This article was submitted to the journal Frontiers in Computational Neuroscience.

Copyright © 2013 Alvarado-Rojas, Lehongre, Bagdasaryan, Bragin, Staba, Engel, Navarro and Le Van Quyen. 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) or licensor 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.

frontiers in
COMPUTATIONAL NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 18 October 2013
doi: 10.3389/fncom.2013.00131

A unified view on weakly correlated recurrent networks

***Dmytro Grytskyy*[1]*, *Tom Tetzlaff*[1], *Markus Diesmann*[1,2] *and Moritz Helias*[1]**

[1] *Institute of Neuroscience and Medicine (INM-6) and Institute for Advanced Simulation (IAS-6), Jülich Research Centre and JARA, Jülich, Germany*
[2] *Medical Faculty, RWTH Aachen University, Germany*

Edited by:
Ruben Moreno-Bote, Foundation Sant Joan de Deu, Spain

Reviewed by:
Brent Doiron, University of Pittsburgh, USA
Shigeru Shinomoto, Kyoto University, Japan

***Correspondence:**
Dmytro Grytskyy, Jülich Research Centre and JARA, Institute of Neuroscience and Medicine (INM-6) and Institute for Advanced Simulation (IAS-6), Building 15.22, 52425 Jülich, Germany
e-mail: d.grytskyy@fz-juelich.de

The diversity of neuron models used in contemporary theoretical neuroscience to investigate specific properties of covariances in the spiking activity raises the question how these models relate to each other. In particular it is hard to distinguish between generic properties of covariances and peculiarities due to the abstracted model. Here we present a unified view on pairwise covariances in recurrent networks in the irregular regime. We consider the binary neuron model, the leaky integrate-and-fire (LIF) model, and the Hawkes process. We show that linear approximation maps each of these models to either of two classes of linear rate models (LRM), including the Ornstein–Uhlenbeck process (OUP) as a special case. The distinction between both classes is the location of additive noise in the rate dynamics, which is located on the output side for spiking models and on the input side for the binary model. Both classes allow closed form solutions for the covariance. For output noise it separates into an echo term and a term due to correlated input. The unified framework enables us to transfer results between models. For example, we generalize the binary model and the Hawkes process to the situation with synaptic conduction delays and simplify derivations for established results. Our approach is applicable to general network structures and suitable for the calculation of population averages. The derived averages are exact for fixed out-degree network architectures and approximate for fixed in-degree. We demonstrate how taking into account fluctuations in the linearization procedure increases the accuracy of the effective theory and we explain the class dependent differences between covariances in the time and the frequency domain. Finally we show that the oscillatory instability emerging in networks of LIF models with delayed inhibitory feedback is a model-invariant feature: the same structure of poles in the complex frequency plane determines the population power spectra.

Keywords: correlations, linear response, Hawkes process, leaky integrate-and-fire model, binary neuron, linear rate model, Ornstein–Uhlenbeck process

1. INTRODUCTION

The meaning of correlated neural activity for the processing and representation of information in cortical networks is still not understood, but evidence for a pivotal role of correlations increases (recently reviewed in Cohen and Kohn, 2011). Different studies have shown that correlations can either decrease (Zohary et al., 1994) or increase (Sompolinsky et al., 2001) the signal to noise ratio of population signals, depending on the readout mechanism. The architecture of cortical networks is dominated by convergent and divergent connections among the neurons (Braitenberg and Schüz, 1991) causing correlated neuronal activity by common input from shared afferent neurons in addition to direct connections between pairs of neurons and common external signals. It has been shown that correlated activity can faithfully propagate through convergent-divergent feed forward structures, such as synfire chains (Abeles, 1991; Diesmann et al., 1999), a potential mechanism to convey signals in the brain. Correlated firing was also proposed as a key to the solution of the binding problem (von der Malsburg, 1981; Bienenstock, 1995; Singer, 1999), an idea that has been discussed controversially (Shadlen and Movshon, 1999). Independent of a direct functional role of correlations in cortical processing, the covariance function between the spiking activity of a pair of neurons contains the information about time intervals between spikes. Changes of synaptic coupling, mediated by spike-timing dependent synaptic plasticity (STDP, Markram et al., 1997; Bi and Poo, 1999), are hence sensitive to correlations. Understanding covariances in spiking networks is thus a prerequisite to investigate the evolution of synapses in plastic networks (Burkitt et al., 2007; Gilson et al., 2009, 2010).

On the other side, there is ubiquitous experimental evidence of correlated spike events in biological neural networks, going back to early reports on multi-unit recordings in cat auditory cortex (Perkel et al., 1967; Gerstein and Perkel, 1969), the observation of closely time-locked spikes appearing at behaviorally relevant points in time (Kilavik et al., 2009; Ito et al., 2011) and collective oscillations in cortex [recently reviewed in Buzsáki and Wang (2012)].

The existing theories explaining correlated activity use a multitude of different neuron models. Hawkes (1971) developed the theory of covariances for linear spiking Poisson neurons (Hawkes processes). Ginzburg and Sompolinsky (1994) presented the approach of linearization to treat fluctuations around the point of stationary activity and to obtain the covariances for

networks of non-linear binary neurons. The formal concept of linearization allowed Brunel and Hakim (1999) and Brunel (2000) to explain fast collective gamma oscillations in networks of spiking leaky integrate-and-fire (LIF) neurons. Correlations in feed-forward networks of LIF models are studied in Moreno-Bote and Parga (2006), exact analytical solutions for such network architectures are given in Rosenbaum and Josic (2011) for the case of stochastic random walk models, and threshold crossing neuron models are considered in Tchumatchenko et al. (2010) and Burak et al. (2009). Covariances in structured networks are investigated for Hawkes processes (Pernice et al., 2011), and in linear approximation for LIF (Pernice et al., 2012) and exponential integrate-and-fire neurons (Trousdale et al., 2012). The latter three works employ an expansion of the propagator (time evolution operator) in terms of the order of interaction. Finally Buice et al. (2009) investigate higher order cumulants of the joint activity in networks of binary model neurons.

Analytical insight into a neuroscientific phenomenon based on correlated neuronal activity often requires a careful choice of the neuron model to arrive at a solvable problem. Hence a diversity of models has been proposed and is in use. This raises the question which features of covariances are generic properties of recurrent networks and which are specific to a certain model. Only if this question can be answered one can be sure that a particular result is not an artifact of oversimplified neuronal dynamics. Currently it is unclear how different neuron models relate to each other and whether and how results obtained with one model carry over to another. In this work we present a unified theoretical view on pairwise correlations in recurrent networks in the asynchronous and collective-oscillatory regime, approximating the response of different models to linear order. The joint treatment allows us to answer the question of genericness and moreover naturally leads to a classification of the considered models into only two categories, as illustrated in **Figure 1**. The classification in addition enables us to extend existing theoretical results to biologically relevant parameters, such as synaptic delays and the presence of inhibition, and to derive explicit expressions for the time-dependent covariance functions, in quantitative agreement with direct simulations, which can serve as a starting point for further work.

The remainder of this article is organized as follows. In the first part of our results in "Covariance structure of noisy rate models" we investigate the activity and the structure of covariance functions for two versions of linear rate models (LRM); one with input the other with output noise. If the activity relaxes exponentially after application of a short perturbation, both models coincide with the OUP. We mainly consider the latter case, although most results hold for arbitrary kernel functions. We extend the analytical solutions for the covariances in networks of OUP (Risken, 1996) to the neuroscientifically important case of synaptic conduction delays. Solutions are derived first for general forms of connectivity in "Solution of the convolution equation with input noise" for input noise and in "Solution of convolution equation with output noise" for output noise. After analyzing the spectral properties of the dynamics in the frequency domain in "Spectrum of the dynamics," identifying poles of the propagators

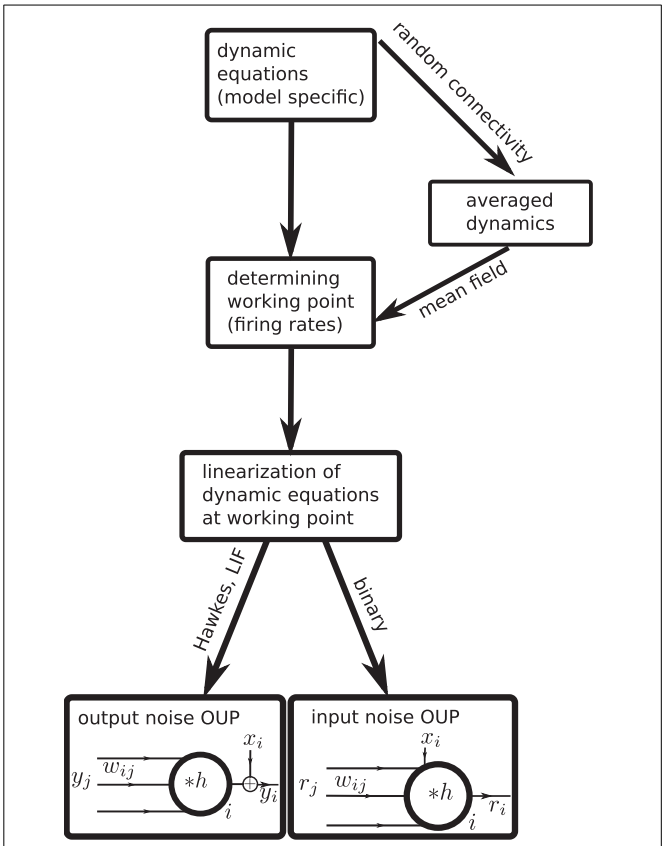


FIGURE 1 | Mapping different descriptions of neuronal dynamics to linear rate models (LRM). The arrows indicate analytical methods which enable a mapping from the original spiking (LIF model, Hawkes model) or binary neuron dynamics to the analytically more tractable linear rate models. Depending on the original dynamics (spiking or binary) the resulting LRM contains an additive noise component x either on the output side (left) or on the input side (right).

and their relation to collective oscillations in neuronal networks, we show in "Population-averaged covariances"' how to obtain pairwise averaged covariances in homogeneous Erdös-Rényi random networks. We explain in detail the use of the residue theorem to perform the Fourier back-transformation of covariance functions to the time domain in "Fourier back transformation" for general connectivity and in "Explicit expression for the population averaged cross covariance in the time domain" for averaged covariance functions in random networks, which allows us to obtain explicit results and to discuss class dependent features of covariance functions.

In the second part of our results in "Binary neurons," "Hawkes processes," and "Leaky integrate-and-fire neurons" we consider the mapping of different neuronal dynamics on either of the two flavors of the linear rate models discussed in the first part. The mapping procedure is qualitatively the same for all dynamics as illustrated in **Figure 1**: Starting from the dynamic equations of the respective model, we first determine the working point described in terms of the mean activity in the network. For unstructured homogeneous random networks this

amounts to a mean-field description in terms of the population averaged activity (i.e., firing rate in spiking models). In the next step, a linearization of the dynamical equations is performed around this working point. We explain how fluctuations can be considered in the linearization procedure to improve its accuracy and we show how the effective linear dynamics maps to the LRM. We illustrate the results throughout by a quantitative comparison of the analytical results to direct numerical simulations of the original non-linear dynamics. The appendices "Implementation of noisy rate models," "Implementation of binary neurons in a spiking simulator code," and "Implementation of Hawkes neurons in a spiking simulator code." describe the model implementations and are modules of our long-term collaborative project to provide the technology for neural systems simulations (Gewaltig and Diesmann, 2007).

2. COVARIANCE STRUCTURE OF NOISY RATE MODELS

2.1. DEFINITION OF MODELS

Let us consider a network of linear model neurons, each characterized by a continuous fluctuating rate r and connections from neuron j to neuron i given by the element w_{ij} of the connectivity matrix $\mathbf{w}$. We assume that the response of neuron i to input can be described by a linear kernel h so that the activity in the network fulfills

$$\mathbf{r}(t) = h(\circ) * [\mathbf{w}\mathbf{r}(\circ - d) + \boldsymbol{b}\mathbf{x}(\circ)](t), \tag{1}$$

where $f(\circ - d)$ denotes the function f shifted by the delay d, $\mathbf{x}$ is an uncorrelated noise with

$$\langle x_i(t)\rangle = 0, \qquad \langle x_i(s)x_j(t)\rangle = \delta_{ij}\delta(s-t)\rho^2\,, \tag{2}$$

e.g., a Gaussian white noise and $(f * g)(t) = \int_{-\infty}^{t} f(t-t')\,g(t')\,dt'$ is the convolution. With the particular choice $\boldsymbol{b} = \mathbf{w}\delta(\circ - d)*$ we obtain

$$\mathbf{r}(t) = [h(\circ) * \mathbf{w}(\mathbf{r}(\circ - d) + \mathbf{x}(\circ - d))](t). \tag{3}$$

We call the dynamics (3) the linear noisy rate model (LRM) with noise applied to output, as the sum $r + x$ appears on the right hand side. Alternatively, choosing $\boldsymbol{b} = \mathbf{1}$ we define the model with input noise as

$$\mathbf{r}(t) = h(\circ) * [\mathbf{w}\mathbf{r}(\circ - d) + \mathbf{x}(\circ)](t). \tag{4}$$

Hence, Equations (3) and (4) are special cases of (1). In the following we consider the particular case of an exponential kernel

$$h(s) = \frac{1}{\tau}\theta(s)\,e^{-s/\tau}, \tag{5}$$

where θ denotes the Heaviside function, $\theta(t) = 1$ for $t > 0$, 0 else. Applying to (1) the operator $O = \tau\frac{d}{ds} + 1$ which has h as a Green's function (i.e., $Oh = \delta$) we get

$$\tau\frac{d}{dt}\mathbf{r}(t) + \mathbf{r}(t) = \mathbf{w}\mathbf{r}(t-d) + \mathbf{b}\mathbf{x}(t), \tag{6}$$

which is the equation describing a set of delay coupled Ornstein-Uhlenbeck-processes (OUP) with input or output noise for $\boldsymbol{b} = \mathbf{1}$ or $\boldsymbol{b} = \mathbf{w}\delta(\circ - d)*$, respectively. We use this representation in "Binary neurons" to show the correspondence to networks of binary neurons.

2.2. SOLUTION OF THE CONVOLUTION EQUATION WITH INPUT NOISE

The solution for the system with input noise obtained from the definition (4) after Fourier transformation is

$$\mathbf{R} = H_d\mathbf{w}\mathbf{R} + H\mathbf{X}, \tag{7}$$

where the delay is consumed by the kernel function $h_d(s) = \frac{1}{\tau}\theta(s-d)e^{-(s-d)/\tau}$. We use capital letters throughout the text to denote objects in the Fourier domain and lower case letters for objects in the time domain. Solved for $\mathbf{R} = (1 - H_d\mathbf{w})^{-1}H\mathbf{X}$ the covariance function of $\mathbf{r}$ in the Fourier domain is found with the Wiener–Khinchin theorem (Gardiner, 2004) as $\langle\mathbf{R}(\omega)\mathbf{R}^T(-\omega)\rangle$, also called the cross spectrum

$$\begin{aligned}\mathbf{C}(\omega) &= \langle\mathbf{R}(\omega)\mathbf{R}^T(-\omega)\rangle \\ &= (1 - H_d(\omega)\mathbf{w})^{-1}H(\omega)\langle\mathbf{X}(\omega)\mathbf{X}^T(-\omega)\rangle \\ &\quad H(-\omega)(1 - H_d(-\omega)\mathbf{w}^T)^{-1} \\ &= (H_d(\omega)^{-1} - \mathbf{w})^{-1}\mathbf{D}(H_d(-\omega)^{-1} - \mathbf{w}^T)^{-1},\end{aligned} \tag{8}$$

where we introduced the matrix $\mathbf{D} = \langle\mathbf{X}(\omega)\mathbf{X}^T(-\omega)\rangle$. From the second to the third line we used the fact that the non-delayed kernels $H(\omega)$ can be replaced by delayed kernels $H_d(\omega)$ and that the corresponding phase factors $e^{i\omega d}$ and $e^{-i\omega d}$ cancel each other. If $\mathbf{x}$ is a vector of pairwise uncorrelated noise, $\mathbf{D}$ is a diagonal matrix and needs to be chosen accordingly in order for the cross spectrum (8) to coincide (neglecting non-linear effects) with the cross spectrum of a network of binary neurons, as described in "Equivalence of binary neurons and Ornstein–Uhlenbeck processes".

2.3. SOLUTION OF CONVOLUTION EQUATION WITH OUTPUT NOISE

For the system with output noise we consider the quantity $y_i = r_i + x_i$ as the dynamic variable representing the activity of neuron i and aim to determine pairwise correlations. It is easy to get from (3) after Fourier transformation

$$\mathbf{R} = H_d\mathbf{w}(\mathbf{R} + \mathbf{X}), \tag{9}$$

which can be solved for $\mathbf{R} = (1 - H_d\mathbf{w})^{-1}H_d\mathbf{w}\mathbf{X}$ in order to determine the Fourier transform of $\mathbf{Y}$ as

$$\mathbf{Y} = \mathbf{R} + \mathbf{X} = (1 - H_d\mathbf{w})^{-1}\mathbf{X}. \tag{10}$$

The cross spectrum hence follows as

$$\begin{aligned}\mathbf{C}(\omega) &= \langle\mathbf{Y}(\omega)\mathbf{Y}^T(-\omega)\rangle \\ &= (1 - H_d(\omega)\mathbf{w})^{-1}\langle\mathbf{X}(\omega)\mathbf{X}^T(-\omega)\rangle(1 - H_d(-\omega)\mathbf{w}^T)^{-1} \\ &= (1 - H_d(\omega)\mathbf{w})^{-1}\mathbf{D}(1 - H_d(-\omega)\mathbf{w}^T)^{-1},\end{aligned} \tag{11}$$

with $\mathbf{D} = \langle \mathbf{X}(\omega)\mathbf{X}^T(-\omega)\rangle$. $\mathbf{D}$ is a diagonal matrix with the i-th diagonal entry ρ_i^2. For the correspondence to spiking models $\mathbf{D}$ must be chosen appropriately, as discussed in "Hawkes processes" and "Leaky integrate-and-fire neurons" for Hawkes processes and LIF neurons, respectively.

2.4. SPECTRUM OF THE DYNAMICS

For both linear rate dynamics, with output and with input noise, the cross spectrum $\mathbf{C}(\omega)$ has poles at certain frequencies ω in the complex plane. These poles are defined by the zeros of $\det(H_d(\omega)^{-1} - \mathbf{w})$ and the corresponding term with the opposite sign of ω. The zeros of $\det(H_d(\omega)^{-1} - \mathbf{w})$ are solutions of the equation

$$H_d(\omega)^{-1} = (1 + i\omega\tau)e^{i\omega d} = L_j$$

where L_j is the j-th eigenvalue of $\mathbf{w}$. The same set of poles arises from (1) when solving for $\mathbf{R}$. For $d > 0$ and the exponential kernel (5), the poles can be expressed as

$$z_k(L_j) = \frac{i}{\tau} - \frac{i}{d} W_k\left(L_j \frac{d}{\tau} e^{\frac{d}{\tau}}\right), \tag{12}$$

where W_k is the k-th of the infinitely many branches of the Lambert-W function (Corless et al., 1996). For vanishing synaptic delay $d = 0$ there is obviously only one solution for every L_j given by $z = \frac{-i}{\tau}(L_j - 1)$.

Given the same parameters d, $\mathbf{w}$, τ, the pole structures of the cross spectra of both systems (8) and (11) are identical, since the former can be obtained from the latter by multiplication with $(H_d(\omega)H_d(-\omega))^{-1} = (H(\omega)H(-\omega))^{-1}$, which has no poles. The only exception causing a different pole structure for the two models is the existence of an eigenvalue $L_j = 0$ of the connectivity matrix $\mathbf{w}$, corresponding to a pole $z(0) = \frac{i}{\tau}$. However, this pole corresponds to an exponential decay of the covariance for input noise in the time domain and hence does not contribute to oscillations. For output noise, the multiplication with the term $(H(\omega)H(-\omega))^{-1}$, vanishing at $\omega = \frac{i}{\tau}$, cancels this pole in the covariance. Consequently both dynamics exhibit similar oscillations. A typical spectrum of poles for a negative eigenvalue $L_j < 0$ is shown in **Figures 2B,D**.

2.5. POPULATION-AVERAGED COVARIANCES

Often it is desirable to consider not the whole covariance matrix but averages over subpopulations of pairs of neurons. For instance the average over the whole network would result in a single scalar value. Separately averaging pairs, distinguishing excitatory and inhibitory neuron populations, yields a 2 by 2 matrix of covariances. For these simpler objects closed form solutions can be obtained, which already preserve some useful information and show important features of the network. Averaged covariances are also useful for comparison with simulations and experimental results.

In the following we consider a recurrent random network of $N_e = N$ excitatory and $N_i = \gamma N$ inhibitory neurons with synaptic weight w for excitatory and $-gw$ for inhibitory synapses. The probability p determines the existence of a connection between

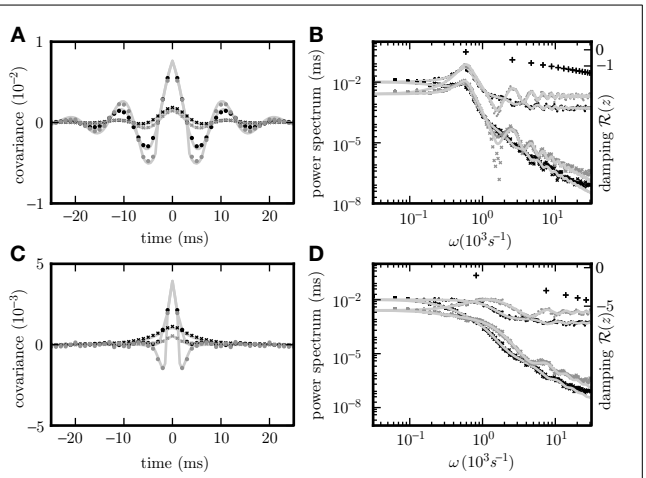


FIGURE 2 | Pole structure determines dynamics. Autocovariance of the population activity **(A,C)** measured in ρ^2/τ and its Fourier transform called power spectrum **(B,D)** of the rate models with output noise (dots) (3) and input noise (diagonal crosses) (4) for delays $d = 3$ ms **(A,B)**, and $d = 1$ ms **(C,D)**. Black symbols show averages over the excitatory population activity and gray symbols over the inhibitory activity obtained by direct simulation. Light gray curves show theoretical predictions for the spectrum (20) and the covariance (21) for output noise and the spectrum (17) and the covariance (18) for input noise. Black crosses (12) in **(B,D)** denote the locations of the poles of the cross spectra - with the real parts corresponding to the damping (vertical axis), and the imaginary parts to oscillation frequencies (horizontal axis). The detailed parameters for this and following figures are given in "Parameters of simulations".

two randomly chosen neurons. We study the dynamics averaged over the two subpopulations by introducing the quantities $r_a = \frac{1}{N_a}\sum_{j\in a} r_j$ and noise terms $x_a = \frac{1}{N_a}\sum_{j\in a} x_j$ for $a \in \{\mathcal{E}, \mathcal{I}\}$; indices $\mathcal{I}$ and $\mathcal{E}$ stand for inhibitory and excitatory neurons and corresponding quantities. Calculating the average local input $N_a^{-1}\sum_{j\in a} w_{jk} r_k$ to a neuron of type a, we obtain

$$\begin{aligned} N_a^{-1}\sum_{j\in a}\sum_k w_{jk} r_k &= N_a^{-1}\left(\sum_{j\in a}\sum_{k\in\mathcal{E}} w_{jk} r_k + \sum_{j\in a}\sum_{k\in\mathcal{I}} w_{jk} r_k\right) \\ &= N_a^{-1}\left(pN_a w\sum_{k\in\mathcal{E}} r_k - pN_a gw\sum_{k\in\mathcal{I}} r_k\right) \\ &= pwN(r_\mathcal{E} - \gamma g r_\mathcal{I}), \end{aligned} \tag{13}$$

where, from the second to the third line we used the fact that in expectation a given neuron k has pN_a targets in the population a. The reduction to the averaged system in (13) is exact if in every column k in $\mathbf{w}_{jk}$ there are exactly K non-zero elements for $j \in \mathcal{E}$ and γK for $j \in \mathcal{I}$, which is the case for networks with fixed out-degree (number of outgoing connections of a neuron to the neurons of a particular type is kept constant), as noted earlier (Tetzlaff et al., 2012). For fixed in-degree (number of connections to a neuron coming in from the neurons of a particular type is kept constant) the substitution of $r_{j\in a}$ by r_a is an additional approximation, which could be considered as an average over possible realizations of the random connectivity. In both cases the

effective population-averaged connectivity matrix **M** turns out to be

$$\mathbf{M} = Kw \begin{pmatrix} 1 & -\gamma g \\ 1 & -\gamma g \end{pmatrix}, \tag{14}$$

with $K = pN$. So the averaged activities fulfill the same Equations (3) and (4) with the non-averaged quantities **r**, **x**, and **w** replaced by their averaged counterparts $\bar{\mathbf{r}} = (r_{\mathcal{E}}, r_{\mathcal{I}})^T$, $\bar{\mathbf{x}} = (x_{\mathcal{E}}, x_{\mathcal{I}})^T$, and **M**. The population averaged activities r_a are directly related to the block-wise averaged covariance matrix $\bar{\mathbf{c}} = \begin{pmatrix} c_{\mathcal{E}\mathcal{E}} & c_{\mathcal{E}\mathcal{I}} \\ c_{\mathcal{I}\mathcal{E}} & c_{\mathcal{I}\mathcal{I}} \end{pmatrix}$, with $c_{ab} = N_a^{-1} N_b^{-1} \sum_{i \in a} \sum_{j \in b} c_{ij}$. With

$$\begin{aligned} \bar{D}_{ab} &= N_a^{-1} N_b^{-1} \left\langle \sum_{i \in a} x_i \sum_{j \in b} x_j \right\rangle \\ &= N_a^{-1} N_b^{-1} \sum_{i \in a} \sum_{j \in b} D_{ij} \\ &= \delta_{ab} N_a / N_a^2 \rho^2 = \delta_{ab} N_a^{-1} \rho^2 \end{aligned} \tag{15}$$

we replace **D** by $\bar{\mathbf{D}} = \rho^2 \begin{pmatrix} N^{-1} & 0 \\ 0 & (\gamma N)^{-1} \end{pmatrix}$ and **c** by $\bar{\mathbf{c}}$ so that the same Equations (11) and (8) and their general solutions also hold for the block-wise averaged covariance matrices.

The covariance matrices separately averaged over pairs of excitatory, inhibitory or mixed pairs are shown in **Figure 2** for both linear rate dynamics (3) and (4). (Parameters for all simulations presented in this article are collected in "Parameters of simulations," the implementation of LRM is described in "Implementation of noisy rate models"). The poles of both models shown in **Figure 2B** are given by (12) and coincide with the peaks in the cross spectra (8) and (11) for output and input noise, respectively. The results of direct simulation and the theoretical prediction are shown for two different delays, with the longer delay leading to stronger oscillations.

Figure 3C shows the distribution of eigenvalues in the complex plane for two random connectivity matrices with different synaptic amplitudes w. The model exhibits a bifurcation, if at least one eigenvalue assumes a zero real part. For fixed out-degree the averaging procedure (13) is exact, reflected by the precise agreement of theory and simulation in **Figure 3D**. For fixed in-degree, the averaging procedure (13) is an approximation, which is good only for parameters far from the bifurcation. Even in this regime still small deviations of the theory from the simulation results are visible in **Figure 3B**. On the stable side close to a bifurcation, the appearance of long living modes causes large fluctuations. These weakly damped modes appearing in one particular realization of the connectivity matrix are not represented after the replacement of the full matrix **w** by the average **M** over matrix realizations. The eigenvalue spectrum of the connectivity matrix provides an alternative way to understand the deviations. By the averaging the set of N eigenvalues of the connectivity matrix is replaced withby the two eigenvalues of the reduced matrix **M**, one of which is zero due to identical rows of **M**. The eigenvalue spectrum of the full matrix

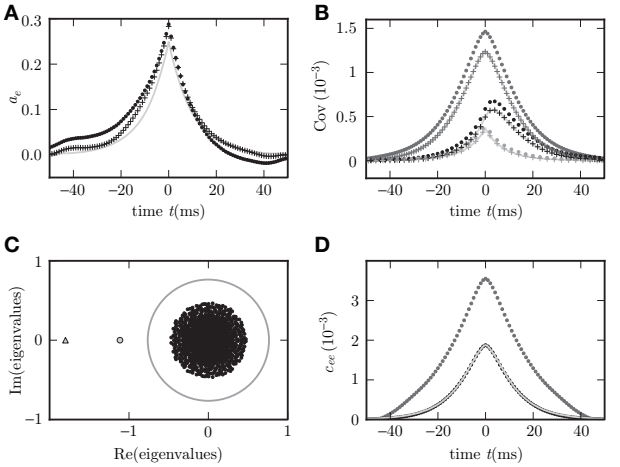


FIGURE 3 | Limits of the theory for fixed in-degree and fixed out-degree. Autocovariance **(A)** and covariance **(B)** in random networks with fixed in-degree (dots) and fixed out-degree (crosses). Simulation results for $c_{\mathcal{E}\mathcal{E}}$, $c_{\mathcal{E}\mathcal{I}}$, and $c_{\mathcal{I}\mathcal{I}}$ are shown in dark gray, black and light gray, respectively for synaptic weight $w = 0.011$ far from bifurcation. For larger synaptic weight $w = 0.018$ close to bifurcation (see text at the end of "Population-averaged covariances"), $c_{\mathcal{E}\mathcal{E}}$ is also shown in **(D)** for fixed in-degree (dark gray dots) and for fixed out-degree (black dots). Corresponding theoretical predictions for the autocovariance (34) **(A)** and the covariance (18) **(B,D)** are plotted as light gray curves throughout. The set of eigenvalues is shown as black dots in panel **(C)** for the smaller weight. The gray circle denotes the spectral radius $w\sqrt{Np(1-p)(1+\gamma g^2)}$ (Rajan and Abbott, 2006; Kriener et al., 2008) confining the set of eigenvalues for the larger weight. The small filled gray circle and the triangle show the effective eigenvalues L of the averaged systems for small and large weight, respectively.

is illustrated in **Figure 3C**. Even if the eigenvalue(s) $L^{\mathbf{M}}$ of **M** are far in the stable region (corresponding to $\Im(z(L^{\mathbf{M}})) > 0$) some eigenvalues $L^{\mathbf{w}}$ of the full connectivity matrix in the vicinity of the bifurcation region may still have an imaginary part becoming negative and the system can feel their influence, shown in **Figure 3D**.

2.6. FOURIER BACK TRANSFORMATION

Although the cross spectral matrices (8) and (11) for both dynamics look similar in the Fourier domain, the procedures for back transformation differ in detail. In both cases, the Fourier integral along the real ω-axis can be extended to a closed integration contour by a semi-circle with infinite radius centered at 0 in the appropriately chosen half-plane. The half-plane needs to be selected such that the contribution of the integration along the semi-circle vanishes. By employing the residue theorem (Bronstein et al., 1999) the integral can be replaced by a sum over residua of the poles encircled by the contour. For a general covariance matrix we only need to calculate $\mathbf{c}(t)$ for $t \geq 0$, as for $t < 0$ the solution can be found by symmetry $\mathbf{c}(t) = \mathbf{c}^T(-t)$.

For input noise it is possible to close the contour in the upper half-plane where the integrand $\mathbf{C}(\omega)\, e^{i\omega t}$ vanishes for $|\omega| \to \infty$ for all $t > 0$, as $|C_{ij}(\omega)|$ decays as $|\omega|^{-2}$. This can be seen from (8), because the highest order of $H_d^{-1} \propto \omega$ appearing in $\det(H_d^{-1} - \mathbf{w})$ is equal to the dimensionality N of **w** ($N = 2$ for **M**), and in

det(adjugate matrix ij of $H_d^{-1} - \mathbf{w}$) it is $N-1$ ($i=j$) or $N-2$ ($i \neq j$). So $|(H_d^{-1} - \mathbf{w})^{-1}|$ is proportional to $|\omega|^{-1}|e^{-i\omega d}|$ and $|\mathbf{C}(\omega)| \propto |\omega|^{-2}$ for large $|\omega|$.

For the case of output noise (11) $\mathbf{C}(\omega)$ can be obtained from the $\mathbf{C}(\omega)$ for input noise (8) multiplied with $(H_d(\omega)H_d(-\omega))^{-1} \sim |\omega|^2$ for large $|\omega|$. The multiplication with this factor changes the asymptotic behavior of the integrand, which therefore contains terms converging to a constant value and terms decaying like $|\omega|^{-1}$ for $|\omega| \to \infty$. These terms result in non-vanishing integrals over the semicircle in the upper half-plane and have to be considered separately. To this end we rewrite (11) as

$$\begin{aligned}\mathbf{C}(\omega) &= ((1 - H_d(\omega)\mathbf{w})^{-1}H_d(\omega)\mathbf{w} + 1)\\ &\quad \mathbf{D}(\mathbf{w}^T H_d(-\omega)(1 - H_d(-\omega)\mathbf{w}^T)^{-1} + 1) \qquad (16)\\ &= (1 - H_d(\omega)\mathbf{w})^{-1}H_d(\omega)\mathbf{w}\mathbf{D}\mathbf{w}^T H_d(-\omega)(1 - H_d(-\omega)\mathbf{w}^T)^{-1}\\ &\quad + (1 - H_d(\omega)\mathbf{w})^{-1}H_d(\omega)\mathbf{w}\mathbf{D}\\ &\quad + \mathbf{D}\mathbf{w}^T H_d(-\omega)(1 - H_d(-\omega)\mathbf{w}^T)^{-1}\\ &\quad + \mathbf{D},\end{aligned}$$

and find the constant term $\mathbf{D}$ which turns into a δ-function in the time domain. The first term in the second line of (16) decays like $|\omega|^{-2}$ and can be transformed just as $\mathbf{C}(\omega)$ for input noise closing the contour in the upper half-plane. The second and third term are the transposed complex conjugates of each other, because of the dependence of H on $-\omega$ instead of ω, and require a special consideration. Multiplied by $e^{i\omega t}$ under the Fourier integral, the first term is proportional to $H_d e^{i\omega t} \sim \omega^{-1} e^{i\omega(t-d)}$ and vanishes faster than $|\omega|^{-1}$ for large $|\omega|$ in the upper half-plane for $t > d$ and in the lower half plane for $t < d$. For the second term the half planes are interchanged. The application of the residue theorem requires closing the integration contour in the half-plane where the integral over the semi-circle vanishes faster than $|\omega|^{-1}$. For $\mathbf{w} = \mathbf{M}$ and in the general case of a stable dynamics all poles of the first term are in the upper half-plane $\Im(z_k(L_j)) > 0$, and have no contribution to $\mathbf{c}(t)$ for $t < d$. For the second term the same is true for $t > -d$; these terms correspond to the jumps of $\mathbf{c}(t)$ after one delay, caused by the effect of the sending neuron arriving at the other neurons in the system after one synaptic delay. These terms correspond to the response of the system to the impulse of the sending neuron – hence we call them "echo terms" in the following (Helias et al., 2013). The presence of such discontinuous jumps at time points d and $-d$ in the case of output noise is reflected in the convolution of $h\mathbf{w}$ with $\mathbf{D}$ in the time domain in (37). For input noise the absence of discontinuities can be inferred from the absence of such terms in (33), where the derivative of the correlation function is equal to the sum of finite terms. The first summand in (16) corresponds to the covariance evoked by fluctuations propagating through the system originating from the same neuron and we call it "correlated input term". In the system with input noise a similar separation into effective echo and correlated input terms can be performed. We obtain the correlated input term as the covariance in an auxiliary population without outgoing connections and echo terms as the difference between the full covariance between neurons within the network and the correlated input term.

2.7. EXPLICIT EXPRESSION FOR THE POPULATION AVERAGED CROSS COVARIANCE IN THE TIME DOMAIN

We obtain the population averaged cross spectrum in a recurrent random network of Ornstein–Uhlenbeck processes (OUP) with input noise by inserting the averaged connectivity matrix $\mathbf{w} = \mathbf{M}$ (14) into (8). The explicit expression for the covariance function follows by taking into account all (both) eigenvalues of $\mathbf{M}$ with values 0 and $L = Kw(1 - \gamma g)$. The detailed derivation of the results presented in this section are documented in "Calculation of the Population Averaged Cross Covariance in Time Domain". The expression for the cross spectrum (8) takes the form

$$\begin{aligned}\mathbf{C}(\omega) = f(\omega)f(-\omega) &\left(\mathbf{1} + Kw\begin{pmatrix}\gamma g & -\gamma g\\ 1 & -1\end{pmatrix}H_d(\omega)\right)\\ &\mathbf{D}\left(\mathbf{1} + Kw\begin{pmatrix}\gamma g & 1\\ -\gamma g & -1\end{pmatrix}H_d(-\omega)\right), \qquad (17)\end{aligned}$$

where we introduced $f(\omega) = (H_d(\omega)^{-1} - L)^{-1}$ as a short hand. Sorting the terms by their dependence on ω, introducing the functions $\Phi_1(\omega), \ldots, \Phi_4(\omega)$ for this dependence, and $\varphi_1(t), \ldots, \varphi_4(t)$ for the corresponding functions in the time domain, the covariance in the time domain $\mathbf{c}(t) = \frac{1}{2\pi}\int_{-\infty}^{+\infty}\mathbf{C}(\omega)e^{i\omega t}d\omega$ takes the form

$$\begin{aligned}\mathbf{c}(t) &= \mathbf{D}\varphi_1(t)\\ &\quad + Kw\left(\begin{pmatrix}\gamma g & -\gamma g\\ 1 & -1\end{pmatrix}\mathbf{D}\varphi_2(t) + \mathbf{D}\begin{pmatrix}\gamma g & 1\\ -\gamma g & -1\end{pmatrix}\varphi_3(t)\right)\\ &\quad + K^2w^2\begin{pmatrix}\gamma g & -\gamma g\\ 1 & -1\end{pmatrix}\mathbf{D}\begin{pmatrix}\gamma g & 1\\ -\gamma g & -1\end{pmatrix}\varphi_4(t).\end{aligned}$$

The previous expression is valid for arbitrary $\mathbf{D}$. In simulations presented in this article we consider identical marginal input statistics for all neurons. In this case the averaged activities for excitatory and inhibitory neurons are the same, so we can insert the special form of $\mathbf{D}$ given in (15), which results in

$$\begin{aligned}\mathbf{c}(t) &= \frac{\rho^2}{N}\begin{pmatrix}1 & 0\\ 0 & \gamma^{-1}\end{pmatrix}\varphi_1(t) \qquad (18)\\ &\quad + \frac{\rho^2}{N}Kw\begin{pmatrix}\gamma g & -g\\ 1 & -\gamma^{-1}\end{pmatrix}\varphi_2(t) + \frac{\rho^2}{N}Kw\begin{pmatrix}\gamma g & 1\\ -g & -\gamma^{-1}\end{pmatrix}\varphi_3(t)\\ &\quad + \frac{\rho^2}{N}(\gamma+1)K^2w^2\begin{pmatrix}\gamma g^2 & g\\ g & \gamma^{-1}\end{pmatrix}\varphi_4(t).\end{aligned}$$

The time-dependent functions $\varphi_1, \ldots, \varphi_4$ are the same in both cases. Using the residue theorem $\varphi_i(t) = \frac{1}{2\pi}\int_{-\infty}^{+\infty}\Phi_i(\omega)e^{i\omega t}d\omega = i\sum_{z\in\text{poles of }\Phi_i}\text{Res}(\Phi_i, z)\, e^{izt}$ for $t \geqq 0$ they can be expressed as a sum over the poles $z_k(L)$ given by (12) and the pole $z = \frac{i}{\tau}$ of $H_d(\omega)$. At $\omega = z_k(L)$ the residue of $f(\omega)$ is $\text{Res}(f, \omega = z_k(L)) = \left(idL + i\tau e^{i\omega d}\right)^{-1}$, the residue of $H_d(\omega)$ at

$z = \frac{i}{\tau}$ is $-\frac{i}{\tau}e^{d/\tau}$, so that the explicit forms of $\varphi_1, \ldots, \varphi_4$ follow as

$$\begin{aligned}
\varphi_1(t) &= \sum_{\omega=z_k(L)} i\mathrm{Res}(f,\omega)f(-\omega)e^{i\omega t} \\
\varphi_2(t) &= \sum_{\omega=z_k(L)} i\mathrm{Res}(f,\omega)f(-\omega)H_d(\omega)e^{i\omega t} \\
&\quad + \frac{e^{(d-t)/\tau}}{\tau} f\left(\frac{i}{\tau}\right) f\left(-\frac{i}{\tau}\right) \\
\varphi_3(t) &= \sum_{\omega=z_k(L)} i\mathrm{Res}(f,\omega)f(-\omega)H_d(-\omega)e^{i\omega t} \\
\varphi_4(t) &= \sum_{\omega=z_k(L)} i\mathrm{Res}(f,\omega)f(-\omega)H_d(\omega)H_d(-\omega)e^{i\omega t} \\
&\quad + \frac{e^{-t/\tau}}{2\tau} f\left(\frac{i}{\tau}\right) f\left(-\frac{i}{\tau}\right).
\end{aligned} \tag{19}$$

The corresponding expression for $\mathbf{C}(\omega)$ for output noise is obtained by multiplying (17) with $H_d^{-1}(\omega)H_d^{-1}(-\omega) = (1+\omega^2\tau^2)$

$$\begin{aligned}
\mathbf{C}(\omega) &= H_d^{-1}(\omega)H_d^{-1}(-\omega)f(\omega)f(-\omega) \\
&\quad \times (\mathbf{1} + Kw\begin{pmatrix} \gamma g & -\gamma g \\ 1 & -1 \end{pmatrix} H_d(\omega))\mathbf{D}(\mathbf{1} + Kw\begin{pmatrix} \gamma g & 1 \\ -\gamma g & -1 \end{pmatrix} H_d(-\omega)),
\end{aligned} \tag{20}$$

which, after Fourier transform, provides the expression for $\mathbf{c}(t)$ in the time domain for $t \geqq 0$

$$\begin{aligned}
\mathbf{c}(t) &= \mathbf{M}\mathbf{D}\mathbf{M}^T\varphi_1(t) + \mathbf{M}\mathbf{D}\varphi_0(t) + \mathbf{D}\delta(t) \\
&= K^2w^2\frac{\rho^2}{N}(1+\gamma g^2)\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}\varphi_1(t) + Kw\frac{\rho^2}{N}\begin{pmatrix} 1 & -g \\ 1 & -g \end{pmatrix}\varphi_0(t) \\
&\quad + \frac{\rho^2}{N}\begin{pmatrix} 1 & 0 \\ 0 & \gamma^{-1} \end{pmatrix}\delta(t).
\end{aligned} \tag{21}$$

As in (18), the first line holds for arbitrary $\mathbf{D}$, and the second for $\mathbf{D}$ given by (15), valid if the firing rates are homogeneous. φ_1 is defined as before, and

$$\varphi_0(t) = \theta(t-d)\sum_{\omega=z_k(L)} \left(dL + \tau e^{i\omega d}\right)^{-1} e^{i\omega t} \tag{22}$$

vanishes for $t < d$. All matrix elements of the first term in (21) are identical. Therefore all elements of $\mathbf{c}(t)$ are equal for $0 < |t| < d$. Both rows of the matrix in front of φ_0 are identical, so for $t > 0$ the off diagonal term $c_{\mathcal{I}\mathcal{E}}$ coincides with $c_{\mathcal{E}\mathcal{E}}$ and $c_{\mathcal{E}\mathcal{I}}$ with $c_{\mathcal{I}\mathcal{I}}$ and vice versa for $t < 0$.

As an illustration we show the functions $\varphi_0, \ldots, \varphi_4$ for one set of parameters in **Figure 4**. The left panels (**A,C**) correspond to contributions to the covariance caused by common input to a pair of neurons, the right panels (**B,D**) to terms due to the effect of one of the neurons' activities on the remaining network (echo terms). The upper panels (**A,B**) belong to the model with input noise, the lower panel (**C,D**) to the one with output noise.

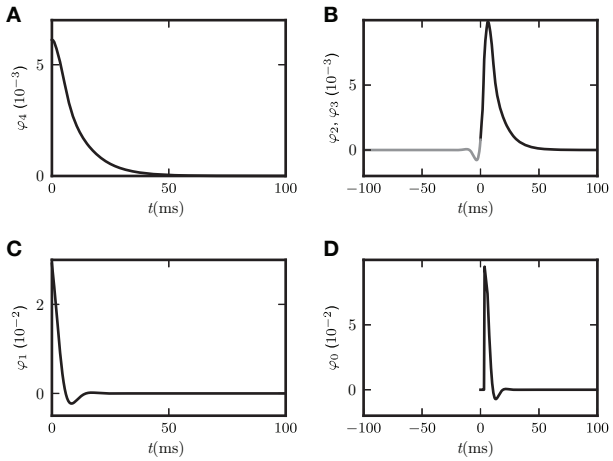


FIGURE 4 | Functions φ_0 (D), φ_1 (C), φ_2, φ_3 (B), φ_4 (A) introduced in (19) and (22) for decomposition of covariance $\mathbf{c}(t)$. In **(B)** $\varphi_3(-t)$ is shown in gray and $\varphi_2(t)$ in black. The two functions are continuations of each other, joint at $t = 0$. Both functions appear in the echo term for input noise. The function φ_0 in **(D)** describing the corresponding echo term in the case of output noise is shifted to be aligned with the function in **(B)** to facilitate the comparison of **(B,D)**. Parameters in all panels are $d = 3$ ms, $\tau = 10$ ms, $L = -1.72$.

For the rate dynamics with output noise, the term with φ_1 in (21) (shown in **Figure 4C**) is symmetric and describes the common input covariance and the term with φ_0 (shown in **Figure 4D**) is the echo part of the covariance. For the rate dynamics with input noise (18) the term containing φ_4 (shown in **Figure 4A**) is caused by common input and is hence also symmetric, the terms with φ_2 and φ_3 (shown in **Figure 4B**) correspond to the echo part and have hence their peak outside the origin. The second echo term in (18) is equal to the first one transposed and with opposite sign of the time argument, so we show $\varphi_2(t)$ and $\varphi_3(-t)$ together in one panel in **Figure 4B**. Note that for input noise, the term with φ_1 describes the autocovariance, which corresponds to the term with the δ-function in case of output noise.

The solution (18) is visualized in **Figure 6**, the solution (21) in **Figure 7**, and the decomposition into common input and echo parts is also shown and compared to direct simulations in **Figure 8**.

3. BINARY NEURONS

In the following sections we study, in turn, the binary neuron model, the Hawkes model and the LIF model and show how they can be mapped to one of the two OUPs; either the one with input or the one with output noise, so that the explicit solutions (18) and (21) for the covariances derived in the previous section can be applied. In the present section, we start with the binary neuron model (Ginzburg and Sompolinsky, 1994; Buice et al., 2009).

Following Ginzburg and Sompolinsky (1994) the state of the network of N binary model neurons is described by a binary vector $\mathbf{n} \in \{0,1\}^N$ and each neuron is updated at independently drawn time points with exponentially distributed intervals of mean duration τ. This stochastic update constitutes a source

of noise in the system. Given the i-th neuron is updated, the probability to end in the up-state ($n_i = 1$) is determined by the gain function $F_i(\mathbf{n})$ which depends on the activity $\mathbf{n}$ of all other neurons. The probability to end in the down state ($n_i = 0$) is $1 - F_i(\mathbf{n})$. Here we implemented the binary model in the NEST simulator (Gewaltig and Diesmann, 2007) as described in "Implementation of Binary Neurons in a Spiking Simulator Code". Such systems have been considered earlier (Ginzburg and Sompolinsky, 1994; Buice et al., 2009), and here we follow the notation employed in the latter work. In the following we collect results that have been derived in these works and refer the reader to these publications for the details of the derivations. The zero-time lag covariance function is defined as $c_{ij}(t) = \langle n_i(t)n_j(t)\rangle - a_i(t)a_j(t)$, with the expectation value $\langle\rangle$ taken over different realizations of the stochastic dynamics. Here $\mathbf{a}(t) = (a_1(t), \ldots, a_N(t))^T$ is the vector of mean activities $a_i(t) = \langle n_i(t)\rangle$. $c_{ij}(t)$ fulfills the differential equation

$$\tau\frac{d}{dt}c_{ij}(t) = -2c_{ij}(t) + \langle (n_j(t) - a_j(t))F_i(\mathbf{n})\rangle + \langle (n_i(t) - a_i(t))F_j(\mathbf{n})\rangle.$$

In the stationary state, the covariance therefore fulfills

$$c_{ij} = \frac{1}{2}\langle (n_j - a_j)F_i(\mathbf{n})\rangle + \frac{1}{2}\langle (n_i - a_i)F_j(\mathbf{n})\rangle. \quad (23)$$

The time lagged covariance $c_{ij}(t, s) = \langle n_i(t)n_j(s)\rangle - a_i(t)a_j(s)$ fulfills for $t > s$ the differential equation

$$\tau\frac{d}{dt}c_{ij}(t, s) = -c_{ij}(t, s) + \langle F_i(\mathbf{n}, t)(n_j(s) - a_j(s))\rangle. \quad (24)$$

This equation is also true for $i = j$, the autocovariance. The term $\langle F_i(\mathbf{n}, t)(n_j(s) - a_j(s))\rangle$ has a simple interpretation: it measures the influence of a fluctuation of neuron j at time s around its mean value on the gain of neuron i at time t (Ginzburg and Sompolinsky, 1994). We now assume a particular form for the coupling between neurons

$$F_i(\mathbf{n}, t) = \phi(\mathbf{J}_i\mathbf{n}(t - d)) = \phi\left(\sum_{k=1}^{N} J_{ik}n_k(t - d)\right), \quad (25)$$

where $\mathbf{J}_i$ is the vector of incoming synaptic weights into neuron i and ϕ is a non-linear gain function. Assuming that the fluctuations of the total input $\mathbf{J}_i\mathbf{n}$ into the i-th neuron are sufficiently small to allow a linearization of the gain function ϕ, we obtain the Taylor expansion

$$F_i(\mathbf{n}, t) = F_i(\mathbf{a}) + \phi'(\mathbf{J}_i\mathbf{a})\,\mathbf{J}_i(\mathbf{n}(t - d) - \mathbf{a}(t - d)),$$

where

$$\phi'(\mathbf{J}_i\mathbf{a}) \quad (26)$$

is the slope of the gain function at the point of mean input.

Up to this point the treatment of the system is identical to the work of Ginzburg and Sompolinsky (1994). Now we present an alternative approach for the linearization which takes into account the effect of fluctuations in the input. For sufficiently asynchronous network states, the fluctuations in the input $\mathbf{J}_i\mathbf{n}(t - d)$ to neuron i can be approximated by a Gaussian distribution $\mathcal{N}(\mu, \sigma)$. In the following we consider a homogeneous random network with fixed in-degree as described in "Population-averaged covariances". As each neuron receives the same number K of excitatory and γK inhibitory synapses, the marginal statistics of the summed input to each neuron is identical. The mean input to a neuron then is $\mu = KJ(1 - \gamma g)a$, where a is the mean activity of a neuron in the network. If correlations are small, the variance of this input signal distribution can be approximated as the sum of the variances of the individual contributions from the incoming signals, resulting in $\sigma^2 = KJ^2(1 + \gamma g^2)\,a(1 - a)$, where we used the fact that the variance of a binary variable with mean a is $a(1 - a)$. This results from a direct calculation: since $n \in \{0, 1\}$, $n^2 = n$, so that the variance is $\langle n^2\rangle - \langle n\rangle^2 = \langle n\rangle - \langle n\rangle^2 = a(1 - a)$. Averaging the slope ϕ' of the gain function over the distribution of the input variable results in the averaged slope

$$\langle\phi'\rangle \simeq \int_{-\infty}^{\infty} \mathcal{N}(\mu, \sigma, x)\,\phi'(x)\,dx \quad (27)$$

$$\text{with } \mathcal{N}(\mu, \sigma, x) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right).$$

The two alternative methods of linearization of ϕ are illustrated in **Figure 5**. In the given example, the linearization procedure taking into account the fluctuations of the input signal results in a smaller effective slope $\langle\phi'\rangle$ than taking the slope $\phi'(a)$ at the mean activity a near its maximum. Averaging the slope $\langle\phi'\rangle$ over this distribution fits simulation results better than $\phi'(a)$ calculated at the mean of a, as shown in **Figure 6**.

The finite slope of the non-linear gain function can be understood as resulting from the combination of a hard threshold with an intrinsic local source of noise. The inverse strength of this noise

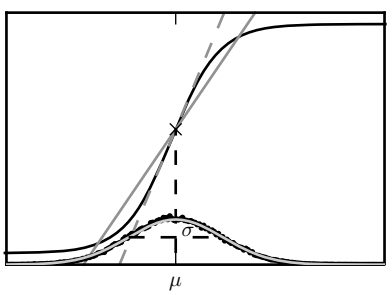


FIGURE 5 | Alternative linearizations of the binary neuron model. The black curve represents the non-linear gain function $\phi(x) = \frac{1}{2} + \frac{1}{2}\tanh(\beta x)$. The dashed gray line is its tangent at the mean input value (denoted by the diagonal cross). The solid curve is the slope $\langle\phi'\rangle$ averaged over the distribution of the fluctuating input (27). This distribution estimated from direct simulation is presented by black dots, the corresponding theoretical prediction of a normal distribution $\mathcal{N}(\mu, \sigma)$ (27) is shown as the light gray curve.

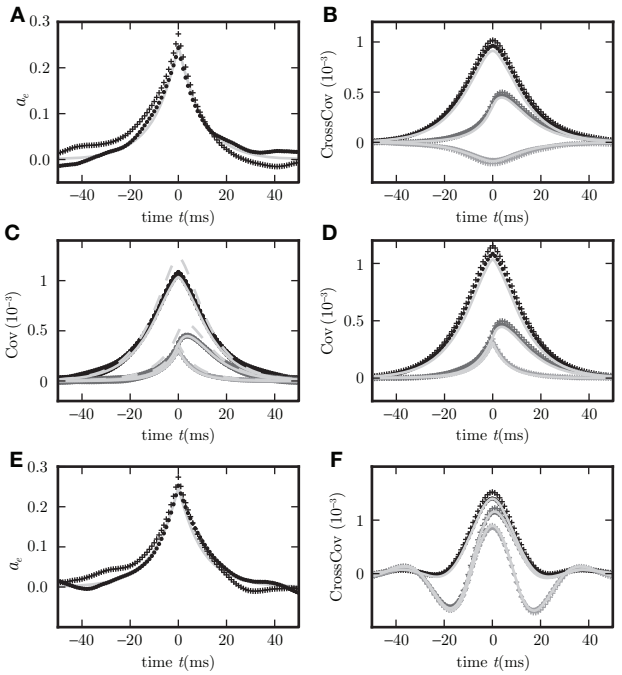


FIGURE 6 | Binary model neuron corresponds to OUP model with input noise. Autocovariance **(A)**, crosscovarince **(B)**, and autocovariance of population averaged activity **(C,D)**, for binary neurons (dots) and rate model with input noise (crosses). $c_{\mathcal{EE}}$, $c_{\mathcal{EI}}$ and $c_{\mathcal{II}}$ are shown in black, gray, and light gray. Corresponding theoretical predictions (18) in **(C,D)**, (34) in **(A)**, their difference in **(C)** are plotted as light gray curves throughout. Dashed curve in **(C)** represents the theoretical prediction using the linearization with the slope at the mean activity (26), the solid curve shows the results for the slope averaged over Gaussian distributed input fluctuations (27). The spread of the simulation results for binary neurons in **(C)** is due to different realizations of the random connectivity. **(E,F)** are the same as **(A,B)** but for the presence of a synaptic delay $d = 10$ ms instead of $d = 0.1$ ms.

determines the slope parameter β (Ginzburg and Sompolinsky, 1994). In this sense, the network model contains two sources of noise, the explicit local noise, quantified by β and the fluctuating synaptic input interpreted as self-generated noise on the network level, quantified by σ. Even in the absence of local noise ($\beta \to \infty$), the above mentioned linearization is applicable and yields a finite effective slope $\langle \phi' \rangle$ (27). In the latter case the resulting effective synaptic weight is independent of the original synapse strength (Grytskyy et al., 2013).

We now extend the classical treatment of covariances in binary networks (Ginzburg and Sompolinsky, 1994) by synaptic conduction delays. In (25) $F_i(\mathbf{n}, t)$ must therefore be understood as a functional acting on the function $\mathbf{n}(t')$ for $t' \in [-\infty, t]$, so that also synaptic connections with time delay d can be realized. We define an effective weight vector to absorb the gain factor as $\mathbf{w}_i = \beta_i \mathbf{J}_i$, with either $\beta_i = \phi'(\mu)$ or $\beta_i = \langle \phi' \rangle$ depending on the linearization procedure, and expand the right hand side of (24) to obtain

$$\langle F_i(\mathbf{n}, t)(n_j(s) - a_j(s)) \rangle = \sum_{k=1}^{N} w_{ik} c_{kj}(t - d, s).$$

Thus the cross-covariance fulfills the matrix delay differential equation

$$\tau \frac{d}{dt} \mathbf{c}(t, s) + \mathbf{c}(t, s) = \mathbf{w}\mathbf{c}(t - d, s). \tag{28}$$

This differential equation is valid for $t > s$. For the stationary solution, the differential equation only depends on the relative timing $u = t - s$

$$\tau \frac{d}{du} \mathbf{c}(u) + \mathbf{c}(u) = \mathbf{w}\mathbf{c}(u - d). \tag{29}$$

The same linearization applied to (23) results in the boundary condition for the solution of the previous equation

$$2\mathbf{c}(0) = \mathbf{w}\mathbf{c}(-d) + (\mathbf{w}\mathbf{c}(-d))^T \tag{30}$$

or, if we split $\mathbf{c}$ into its diagonal and its off-diagonal parts $\mathbf{c}_a$ and $\mathbf{c}_{\neq}$

$$2\mathbf{c}_{\neq}(0) = \mathbf{w}\mathbf{c}_{\neq}(-d) + (\mathbf{w}\mathbf{c}_{\neq}(-d))^T + \mathbf{O} \tag{31}$$
$$\text{with } \mathbf{O} = \mathbf{w}\mathbf{c}_a(-d) + (\mathbf{w}\mathbf{c}_a(-d))^T.$$

In the following section we use this representation to demonstrate the equivalence of the covariance structure of binary networks to the solution for OUP with input noise.

3.1. EQUIVALENCE OF BINARY NEURONS AND ORNSTEIN–UHLENBECK PROCESSES

In the following subsection we show that the same Equations (29) and (31) for binary neurons also hold for the Ornstein-Uhlenbeck process (OUP) with input noise. In doing so here we also extend the existing framework of OUP (Risken, 1996) to synaptic conduction delays d. A network of such processes is described by

$$\tau \frac{d}{dt} \mathbf{r}(t) + \mathbf{r}(t) = \mathbf{w}\mathbf{r}(t - d) + \mathbf{x}(t), \tag{32}$$

where $\mathbf{x}$ is a vector of pairwise uncorrelated white noise with $\langle \mathbf{x}(t) \rangle_x = 0$ and $\langle x_i(t) x_j(t + t') \rangle_x = \delta_{ij} \delta(t') \rho^2$. With the help of the Green's function G satisfying $(\tau \frac{d}{dt} + 1)\, G(t) = \delta(t)$, namely $G(t) = \frac{1}{\tau}\, \theta(t)\, e^{-t/\tau}$, we obtain the solution of Equation (32) as

$$\mathbf{r}(t) = \tau G(t) \mathbf{r}(0) + \int_0^t G(t - t')(\mathbf{w}\mathbf{r}(t' - d) + \mathbf{x}(t'))\, dt'.$$

The equation for the fluctuations $\delta\mathbf{r}(t) = \mathbf{r}(t) - \langle \mathbf{r}(t) \rangle_x$ around the expectation value

$$\delta\mathbf{r}(t) = \int_0^t G(t - t')(\mathbf{w}\delta\mathbf{r}(t' - d) + \mathbf{x}(t'))\, dt'$$

coincides with the noisy rate model with input noise (4) with delay d and convolution kernel $h = G$. In the next step we investigate the covariance matrix $c_{ij}(t, s) = \langle \delta r_i(t + s) \delta r_j(t) \rangle_x$ to show

for which choice of parameters the covariance matrices for the binary model and the OUP with input noise coincide. To this end we derive the differential equation with respect to the time lag s for positive lags $s > 0$

$$\begin{aligned}\tau\frac{d}{ds}\mathbf{c}(t,s) &= \left\langle \tau\frac{d}{ds}\delta\mathbf{r}(t+s)\delta\mathbf{r}^T(t)\right\rangle_x \qquad (33)\\ &= \langle(\mathbf{w}\delta\mathbf{r}(t+s-d) - \delta\mathbf{r}(t+s) + \mathbf{x}(t+s))\delta\mathbf{r}^T(t)\rangle_x\\ &= \mathbf{w}\mathbf{c}(t,s-d) - \mathbf{c}(t,s),\end{aligned}$$

where we used $\langle\mathbf{x}(t+s))\delta\mathbf{r}(t)\rangle_x = 0$, because the noise is realized independently for each time step and the system is causal. Equation (33) is identical to the differential equation satisfied by the covariance matrix (28) for binary neurons (Ginzburg and Sompolinsky, 1994). To determine the initial condition of (33) we need to take the limit $\mathbf{c}(t,0) = \lim_{s\to+0}\mathbf{c}(t,s)$. This initial condition can be obtained as the stationary solution of the following differential equation

$$\begin{aligned}\tau\frac{d}{dt}\mathbf{c}(t,0) &= \lim_{s\to+0}\left(\left\langle\tau\frac{d}{dt}\delta\mathbf{r}(t+s)\delta\mathbf{r}^T(t)\right\rangle_x + \left\langle\delta\mathbf{r}(t+s)\tau\frac{d}{dt}\delta\mathbf{r}^T(t)\right\rangle_x\right)\\ &= \lim_{s\to+0}\Big(\langle(\mathbf{w}\delta\mathbf{r}(t+s-d) - \delta\mathbf{r}(t+s) + \mathbf{x}(t+s))\delta\mathbf{r}^T(t)\rangle_x\\ &\quad + \langle\delta\mathbf{r}(t+s)(\delta\mathbf{r}^T(t-d)\mathbf{w}^T - \delta\mathbf{r}^T(t) + \mathbf{x}^T(t))\rangle_x\Big)\\ &= -2\mathbf{c}(t,0) + \mathbf{w}\mathbf{c}(t,-d) + \mathbf{c}(t-d,d)\mathbf{w}^T + \mathbf{D}.\end{aligned}$$

Here we used that $\langle\mathbf{x}(t+s)\delta\mathbf{r}^T(t)\rangle$ vanishes due to independent noise realizations and causality and

$$\begin{aligned}\mathbf{D} &= \lim_{s\to+0}\langle\delta\mathbf{r}(t+s)\mathbf{x}^T(t)\rangle_x\\ &= \lim_{s\to+0,\ s<d}\int_0^{t+s} G(t+s-t')(\mathbf{w}\underbrace{\langle\delta\mathbf{r}(t'-d)\mathbf{x}^T(t)\rangle_x}_{=0\text{ causality}} + \underbrace{\langle\mathbf{x}(t')\mathbf{x}^T(t)\rangle_x}_{=\mathbf{1}\delta(t-t')\rho^2})dt'\\ &= \lim_{s\to+0,\ s<d}\int_0^{t+s} G(t+s-t')\mathbf{1}\delta(t-t')\rho^2 dt'\\ &= \lim_{s\to+0,\ s<d} G(s)\mathbf{1}\rho^2 = \frac{1}{\tau}\mathbf{1}\rho^2.\end{aligned}$$

In the stationary state, $\mathbf{c}$ only depends on the time lag s and is independent of the first time argument t, which, with the symmetry $\mathbf{c}(-d)^T = \mathbf{c}(d)$ yields the additional condition for the solution of (33)

$$2\mathbf{c}(0) = \mathbf{w}\mathbf{c}(-d) + (\mathbf{w}\mathbf{c}(-d))^T + \mathbf{D}$$

or, if $\mathbf{c}$ is split in diagonal and off-diagonal parts $\mathbf{c}_a$ and $\mathbf{c}_{\neq}$, respectively,

$$\begin{aligned}2\mathbf{c}_{\neq}(0) &= \mathbf{w}\mathbf{c}_{\neq}(-d) + (\mathbf{w}\mathbf{c}_{\neq}(-d))^T + \mathbf{O}\\ 2\mathbf{c}_a(0) &= \mathbf{w}\mathbf{c}_{\neq}(-d) + (\mathbf{w}\mathbf{c}_{\neq}(-d))^T + \mathbf{D}\end{aligned}$$

with $\mathbf{O} = \mathbf{w}\mathbf{c}_a(-d) + (\mathbf{w}\mathbf{c}_a(-d))^T$. In the equation for the autocovariance $\mathbf{c}_a$ the first two terms are contributions due to the cross covariance. In the state of asynchronous network activity with $c_{ij} \sim N^{-1}$ for $i \neq j$ these terms are typically negligible in comparison to the third term because $\sum_k w_{ik}c_{ki} \sim wKN^{-1} = pw$, which is typically smaller than 1 for small effective weights $w < 1$ and small connection probabilities $p \ll 1$. In this approximation with (33) the temporal shape of the autocovariance function is exponentially decaying with time constant τ. With $\mathbf{c}_a(0) \approx \mathbf{D}/2$ the approximate solution for the autocovariance is

$$\mathbf{c}_a(t) = \frac{\mathbf{D}}{2}\exp\left(-\frac{|t|}{\tau}\right). \qquad (34)$$

The cross covariance then satisfies the initial condition

$$\begin{aligned}2\mathbf{c}_{\neq}(0) &= \mathbf{w}\mathbf{c}_{\neq}(-d) + (\mathbf{w}\mathbf{c}_{\neq}(-d))^T + \mathbf{O}\\ \mathbf{O} &= \mathbf{w}\mathbf{D}/2 + (\mathbf{w}\mathbf{D}/2)^T,\end{aligned}$$

which coincides with (31) for binary neurons if the diagonal matrix containing the zero time autocorrelations $\mathbf{c}_a(0)$ for binary neurons is equal to $\mathbf{D}/2$, i.e., if the amplitude of the input noise $\rho^2 = 2\tau a(1-a)$ and the effective linear coupling satisfies $\mathbf{w}_i = \beta_i\mathbf{J}_i$. **Figure 6** shows simulation results for population averaged covariance functions in binary networks and in networks of OUPs with input noise where the parameters of the OUP network are chosen according to the requirements derived above. The theoretical results (18) agree well with the direct simulations of both systems. For comparison, both methods of linearization, as explained above, are shown. The linearization procedure which takes into account the noise on the input side of the non-linear gain function results in a more accurate prediction. Moreover, the results derived here extend the classical theory (Ginzburg and Sompolinsky, 1994) by considering synaptic conduction delays. **Figure 8** shows the decomposition of the covariance structure for a non-zero delay $d = 3$ ms. For details of the implementation see "Implementation of binary neurons in a spiking simulator code". The explicit effect of introducing delays into the system, such as the appearance of oscillations in the time dependent covariance, is presented in **(E,F)** of **Figure 6**, differing from **(A,B)** of this figure, respectively, only in the delay ($d = 10$ ms for **(E,F)**, $d = 0.1$ ms for **(A,B)**).

4. HAWKES PROCESSES

In the following section we show that to linear order the covariance functions in networks of Hawkes processes (Hawkes, 1971) are equivalent to those in the linear rate network with output noise. Hawkes processes generate spikes randomly with a time density given by $\mathbf{r}(t)$, where neuron i generates spikes at a rate $r_i(t)$, realized independently within each infinitesimal time step. Arriving spike trains $\mathbf{s}$ influence $\mathbf{r}$ according to

$$\mathbf{r}(t) = \nu + (h_d * \mathbf{J}\mathbf{s})(t), \qquad (35)$$

with the connectivity matrix $\mathbf{J}$ and the kernel function h_d including the delay. Here ν is a constant base rate of spike emission assumed to be equal for each neuron. Here we employ the implementation of the Hawkes model in the NEST simulator (Gewaltig and Diesmann, 2007). The implementation is

described in "Implementation of Hawkes neurons in a spiking simulator code".

Given neuron j spiked at time $u \leq t$, the probability of a spike in the interval $[t, t+\delta t)$ for neuron i is 1 if $i = j,\ u = t$ (the neuron spikes synchronously with itself) and $r_i(t)\delta t + o(\delta t^2)$ otherwise. Considering the system in the stationary state with the time averaged activity $\bar{\mathbf{r}} = \langle \mathbf{s}(t) \rangle$ we obtain a convolution equation for time lags $\tau \geq 0$ for the covariance matrix with the entry $c_{ij}(\tau)$ for the covariance between spike trains of neurons i and j

$$\begin{aligned}
\mathbf{c}(\tau) &= \langle \mathbf{s}(t+\tau)\mathbf{s}^T(t)\rangle - \langle \mathbf{s}(t+\tau)\rangle\langle \mathbf{s}^T(t)\rangle \qquad (36)\\
&= \langle (\delta(\tau)\mathbf{1} + \mathbf{r}(t+\tau))\mathbf{s}^T(t)\rangle - \bar{\mathbf{r}}\bar{\mathbf{r}}^T\\
&= \langle \mathbf{r}(t+\tau)(\mathbf{s}^T(t) - \bar{\mathbf{r}}^T)\rangle + \mathbf{D}_{\bar{\mathbf{r}}}\\
&= \langle (\nu + (h_d * \mathbf{J}\mathbf{s})(t+\tau))(\mathbf{s}^T(t) - \bar{\mathbf{r}}^T)\rangle + \mathbf{D}_{\bar{\mathbf{r}}}\\
&= h_d * \mathbf{J}\langle \mathbf{s}(t+\tau)(\mathbf{s}^T(t) - \bar{\mathbf{r}}^T)\rangle + \mathbf{D}_{\bar{\mathbf{r}}}\\
&= (h_d * \mathbf{J}\mathbf{c})(\tau) + \mathbf{D}_{\bar{\mathbf{r}}},
\end{aligned}$$

with the diagonal matrix $\mathbf{D}_{\bar{\mathbf{r}}} = \delta(\tau)\mathrm{diag}(\bar{\mathbf{r}})$, which has been derived earlier (Hawkes, 1971). If the rates of all neurons are equal, $\bar{\mathbf{r}}_i = \bar{r}$, all entries in the diagonal matrix are the same, $\mathbf{D}_{\bar{\mathbf{r}}} = \delta(\tau)\mathbf{1}\bar{r}$. In the subsequent section we demonstrate that the same convolution Equation (36) holds for the linear rate with output noise.

4.1. CONVOLUTION EQUATION FOR LINEAR NOISY RATE NEURONS

For the linear rate model with output noise we use Equation (3) for time lags $\tau > 0$ to obtain a convolution equation for the covariance matrix of the output signal vector $\mathbf{y} = \mathbf{r} + \mathbf{x}$ as

$$\begin{aligned}
\mathbf{c}(\tau) &= \langle \mathbf{y}(t+\tau)(\mathbf{y}^T(t) - \bar{\mathbf{r}}^T)\rangle \qquad (37)\\
&= \langle (h_d * \mathbf{w}\mathbf{y} + \mathbf{x})(t+\tau)(\mathbf{y}^T(t) - \bar{\mathbf{r}}^T)\rangle\\
&= (h_d * \mathbf{w}\mathbf{c})(\tau) + \langle \mathbf{x}(t+\tau)(\mathbf{r}^T(t) - \bar{\mathbf{r}}^T)\rangle + \langle \mathbf{x}(t+\tau)\mathbf{x}^T(t)\rangle\\
&= (h_d * \mathbf{w}\mathbf{c})(\tau) + \mathbf{D},
\end{aligned}$$

where we utilized that due to causality the random noise signal generated at $t+\tau$ has no influence on $\mathbf{r}(t)$, so the respective correlation vanishes. $\mathbf{D}$ is the covariance of the noise as in (11), $D_{ij}(\tau) = \langle x_i(t)x_j(t+\tau)\rangle = \delta_{ij}\delta(\tau)\rho^2$. If ρ is chosen such that ρ^2 coincides with the averaged activity $\bar{r}$ in a network of Hawkes neurons and the connection matrix $\mathbf{w}$ is identical to $\mathbf{J}$ of the Hawkes network, the Equations (36) and (37) are identical. Therefore the cross spectrum of both systems is given by (11).

4.2. NON-LINEAR SELF-CONSISTENT RATE IN RECTIFYING HAWKES NETWORKS

The convolution Equation (36) for the covariance matrix of Hawkes neurons is exact if no element of $\mathbf{r}$ is negative, which is particularly the case for a network of only excitatory neurons. Especially in networks including inhibitory couplings, the intensity r_i of neuron i may assume negative values. A neuron with $r_i < 0$ does not emit spikes, so the instantaneous rate is given by $\lambda_i = [r_i(t)]_+ = \theta(r_i(t))\, r_i(t)$, with the Heaviside function θ. We now take into account this effective nonlinearity –the rectification of the Hawkes model neuron– in a similar manner as we already used to linearize binary neurons. If the network is in the regime of low spike rates, the fluctuations in the input of each neuron due to the Poissonian arrival of spikes are large compared to the fluctuations due to the time varying intensities $\mathbf{r}(t)$. Considering the same homogeneous network structure as described in "Population-averaged covariances," the input statistics is identical for each cell i, so the mean activity $\lambda_0 = \langle \lambda_i \rangle$ is the same for all neurons i. The superposition of the synaptic inputs to neuron i cause an instantaneous intensity r_i that follows approximately a Gaussian distribution $\mathcal{N}(\mu, \sigma, r_i)$ with mean $\mu = \langle r \rangle = \nu + \lambda_0 KJ(1 - g\gamma)$ and standard deviation $\sigma = \sqrt{\langle r^2 \rangle - \langle r \rangle^2} = J\sqrt{\frac{\lambda_0}{2\tau}K(1 + g^2\gamma)}$. These expressions hold for the exponential kernel (5) due to Campbell's theorem (Papoulis and Pillai, 2002), because of the stochastic Poisson-like arrival of incoming spikes, where the standard deviation of the spike count is proportional to the square root of the intensity λ_0. The rate λ_0 is accessible by explicit integration over the Gaussian probability density as

$$\begin{aligned}
\lambda_0 &= \int_{-\infty}^{\infty} \mathcal{N}(\mu, \sigma, r)\, r\, \theta(r)\, dr\\
&= \frac{1}{\sqrt{2\pi}\sigma}\int_0^{\infty} \exp\left(-\frac{(r-\mu)^2}{2\sigma^2}\right) r\, dr\\
&= \frac{-\sigma}{\sqrt{2\pi}}\int_0^{\infty} \exp\left(-\frac{(r-\mu)^2}{2\sigma^2}\right)\frac{-(r-\mu)}{\sigma^2}\, dr\\
&\quad + \frac{\mu}{\sqrt{2\pi}\sigma}\int_0^{\infty} \exp\left(-\frac{(r-\mu)^2}{2\sigma^2}\right) dr\\
&= \frac{\sigma}{\sqrt{2\pi}}\exp\left(-\frac{\mu^2}{2\sigma^2}\right) + \frac{\mu}{2}\left(1 - \mathrm{erf}\left(-\frac{\mu}{\sqrt{2}\sigma}\right)\right).
\end{aligned}$$

This equation needs to be solved self-consistently (numerically or graphically) to determine the rate in the network, as the right hand side depends on the rate λ_0 itself through μ and σ. Rewritten as

$$\begin{aligned}
\lambda_0 &= \frac{\sigma}{\sqrt{2\pi}}\exp\left(-\frac{\mu^2}{2\sigma^2}\right) + \mu P_{\mu,\sigma}(r > 0)\\
P_{\mu,\sigma}(r > 0) &= \frac{1}{2} - \frac{1}{2}\mathrm{erf}\left(-\frac{\mu}{\sqrt{2}\sigma}\right), \qquad (38)
\end{aligned}$$

$P_{\mu,\sigma}(r > 0)$ is the probability that the intensity of a neuron is above threshold and therefore contributes to the transmission of a small fluctuation in the input. A neuron for which $r < 0$ acts as if it was absent. Hence we can treat the network with rectifying neurons completely analogous to the case of linear Hawkes processes, but multiply the synaptic weight J or $-gJ$ of each neuron with $P_{\mu,\sigma}(r > 0)$, i.e., the linearized connectivity matrix is

$$\mathbf{w} = P_{\mu,\sigma}(r > 0)\mathbf{J}. \qquad (39)$$

Figure 7 shows the agreement of the covariance functions obtained from direct simulation of the network of Hawkes processes and the analytical solution (21) with average firing rate

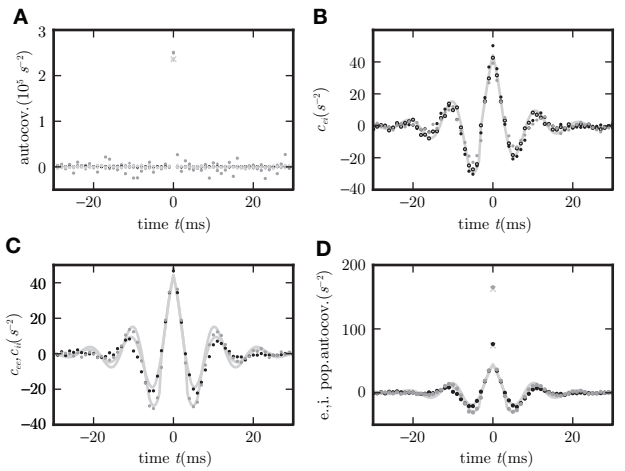


FIGURE 7 | Covariance structure in spiking networks corresponds to OUP with output noise. (A) Autocovariance obtained by direct simulation of the LIF (black), Hawkes (gray), and OUP (light gray) models for excitatory (dots) and inhibitory neurons (crosses). **(B)** Covariance $c_{\mathcal{EI}}$ averaged over disjoint pairs of neurons for LIF (black dots), Hawkes (gray dots), and OUP with output noise (empty circles). **(C)** Covariance averaged over disjoint pairs of neurons of the same type. **(D)** Autocovariance of the population averaged activity. Averages in **(C,D)** over excitatory neurons as black dots, over inhibitory neurons as gray dots. Corresponding theoretical predictions (21) are plotted as light gray curves in all panels except **(A)**. Light gray diagonal crosses in **(A,D)** denote theoretical peak positions determined by the firing rate $\bar{r}$ as $\bar{r}\Delta t$ (where $\Delta t = 0.1$ ms is the time resolution of the histogram).

λ_0 determined by (38), setting the effective strength of the noise $\rho^2 = \lambda_0$, and the linearized coupling as described above. The detailed procedure for choosing the parameters in the direct simulation is described together with the implementation of the Hawkes model in "Implementation of Hawkes neurons in a spiking simulator code".

5. LEAKY INTEGRATE-AND-FIRE NEURONS

In this section we consider a network of LIF model neurons with exponentially decaying postsynaptic currents and show its equivalence to the network of OUP with output noise, valid in the asynchronous irregular regime. A spike sent by neuron j at time t arrives at the target neuron i after the synaptic delay d, elicits a synaptic current I_i that decays with time constant τ_s and causes a response in the membrane potential V_i proportional to the synaptic efficacy J_{ij}. With the time constant τ_m of the membrane potential, the coupled set of differential equations governing the subthreshold dynamics of a single neuron i is (Fourcaud and Brunel, 2002)

$$\begin{aligned}\tau_m \frac{dV_i}{dt} &= -V_i + I_i(t)\\ \tau_s \frac{dI_i}{dt} &= -I_i + \tau_m \sum_{j=1,j}^{N} J_{ij} s_j(t-d),\end{aligned} \tag{40}$$

where the membrane resistance was absorbed into the definitions of J_{ij} and I_i. If V_i reaches the threshold V_θ at time point t_k^i the neuron emits an action potential and the membrane potential is reset to V_r, where it is clamped for the refractory time τ_r. The spiking activity of neuron i is described by this sequence of action potentials, the spike train $s_i(t) = \sum_k \delta(t - t_k^i)$. The dynamics of a single neuron is deterministic, but in network states of asynchronous, irregular activity and in the presence of external Poisson inputs to the network, the summed input to each cell can well be approximated as white noise (Brunel, 2000) with first moment $\mu_i = \tau_m \sum_j J_{ij} r_j$ and second moment $\sigma_i^2 = \tau_m \sum_j J_{ij}^2 r_j$, where r_j is the stationary firing rate of neuron j. The stationary firing rate of neuron i is then given by Fourcaud and Brunel (2002)

$$r_i^{-1} = \tau_r + \tau_m \sqrt{\pi}\left(F(y_\theta) - F(y_r)\right) \tag{41}$$

$$f(y) = e^{y^2}(1 + \mathrm{erf}(y)) \quad F(y) = \int^y f(y)\,dy$$

$$\text{with } y_{\theta,r} = \frac{V_{\theta,r} - \mu_i}{\sigma_i} + \frac{\alpha}{2}\sqrt{\frac{\tau_s}{\tau_m}} \quad \alpha = \sqrt{2}\left|\zeta\left(\frac{1}{2}\right)\right|,$$

with Riemann's zeta function ζ. The response of the LIF neuron to the injection of an additional spike into afferent j determines the impulse response $w_{ij}h(t)$ of the system. The time integral $w_{ij} = w_{ij}\int_0^\infty h(t)\,dt$ is the DC-susceptibility, which can formally be written as the derivative of the stationary firing rate by the rate of the afferent r_j, which, evaluated by help of (41), yields (Helias et al., 2013, Results and App. A)

$$w_{ij} = \frac{\partial r_i}{\partial r_j} = \alpha J_{ij} + \beta J_{ij}^2 \tag{42}$$

$$\text{with } \alpha = \sqrt{\pi}(\tau_m r_i)^2 \frac{1}{\sigma_i}\left(f(y_\theta) - f(y_r)\right)$$

$$\text{and } \beta = \sqrt{\pi}(\tau_m r_i)^2 \frac{1}{2\sigma_i^2}\left(f(y_\theta)\frac{V_\theta - \mu_i}{\sigma_i} - f(y_r)\frac{V_r - \mu_i}{\sigma_i}\right).$$

In the strongly fluctuation-driven regime, the temporal behavior of the kernel h is dominated by a single exponential decay, whose time constant can be determined empirically. In a homogeneous random network the firing rates of all neurons are identical $r_i = \bar{r}$ and follow from the numerical solution of the self-consistency Equation (41). Approximating the autocovariance function of a single spike train by a δ-peak scaled by the rate $\bar{r}\delta(t)$, one obtains for the covariance function $\mathbf{c}$ between pairs of spike trains the same convolution Equation (36) as for Hawkes neurons (Helias et al., 2013, cf. equation 5). As shown in "Convolution equation for linear noisy rate neurons" this convolution equation coincides with that of a linear rate model with output noise (37), where the diagonal elements of $\mathbf{D}$ are chosen to agree to the average spike rate $\rho^2 = \bar{r}$. The good agreement of the analytical cross covariance functions (21) for the OUP with output noise and direct simulation results for LIF are shown in **Figure 7**.

6. DISCUSSION

In this work we describe the path to a unified theoretical view on pairwise correlations in recurrent networks. We consider binary

neuron models, LIF models, and linear point process models. These models containing a non-linearity (spiking threshold in spiking models, non-linear sigmoidal gain function in binary neurons, strictly positive rates in Hawkes processes) are linearized, taking into account the distribution of the fluctuating input.

The work presents results for several neuron models: We derive analytical expressions for delay-coupled OUP with input and with output noise, we extend the analytical treatment for stochastic binary neurons to the presence of synaptic delays, present a method that takes into account network-generated noise to determine the effective gain function, extend the theory of Hawkes processes to the existence of delays and inhibition, and present in Equation (12) a condition for the onset of global oscillations caused by delayed feedback, generalized to feedback pathways through different eigenvalues of the connectivity.

Some results qualitatively extend the existing theory (delays, inhibition), others improve the accuracy of existing theories (linearization including fluctuations). More importantly, our approach enables us to demonstrate the equivalence of each of these models after linear approximation to a linear model with fluctuating continuous variables. The fact that linear perturbation theory leads to effective linear equations is of course not surprising, but the analytical procedure firstly enables a mapping between models that conserves quantitative results and secondly allows us to uncover common structures underlying the emergence of correlated activity in recurrent networks. For the commonly appearing exponentially decaying response kernel function, these rate models coincide with the OUP (OUP, Uhlenbeck and Ornstein, 1930; Risken, 1996). We find that the considered models form two groups, which, in linear approximation merely differ by a matrix valued factor scaling the noise and in the choice of variables interpreted as neural activity. The difference between these two groups corresponds to the location of the noise: spiking models—LIF models and Hawkes models—belong to the class with noise on the output side, added to the activity of each neuron. The non-spiking binary neuron model corresponds to an OUP where the noise is added on the input side of each neuron. The closed solution for the correlation structure of OUP holds for both classes.

We identify different contributions to correlations in recurrent networks: the solution for output noise is split into three terms corresponding to the δ-peak in the autocovariance, the covariance caused by shared input, and the direct synaptic influence of stochastic fluctuations of one neuron on another–the latter echo terms are equal to propagators acting with delays (Helias et al., 2013). A similar splitting into echo and correlated input terms for the case of input noise is shown in **Figure 8**. For increasing network size $N \to \infty$, keeping the connection probability p fixed, so that $K = pN$, and with rescaled synaptic amplitudes $J \sim 1/\sqrt{N}$ (van Vreeswijk and Sompolinsky, 1996; Renart et al., 2010) the echo terms vanish fastest. Formally this can be seen from (18): the multiplicative factor of the common covariance term φ_4 does not change with N while the other coefficients decrease. So ultimately all four entries of the matrix $\mathbf{c}$ have the same time dependence determined by the common covariance term φ_4. In particular the covariance between excitation and inhibition $c_{\mathcal{E}\mathcal{I}}$ becomes

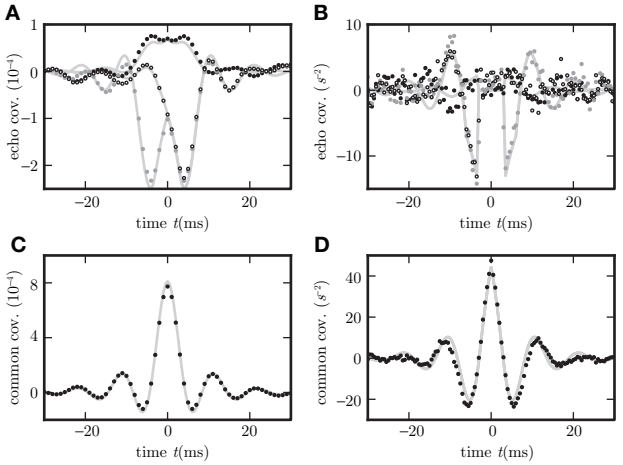


FIGURE 8 | Different echo terms for spiking and non-spiking neurons. Binary non-spiking neurons shown in **(A,C)** and LIF in **(B,D)**. **(A)**,**(B)** Echo terms by direct influence of the neuron's output on the network in dependence of neuron types (in **A**,**B** $c_{\mathcal{E}\mathcal{E}}$,$c_{\mathcal{E}\mathcal{I}}$, and $c_{\mathcal{I}\mathcal{I}}$ are plotted as black, gray dots and circles). **(C,D)**, Contributions to the covariance evoked by correlated and common input (black dots) measured with help of auxiliary model neurons which do not provide feedback to the network. Corresponding theoretical predictions (16) are plotted as light gray curves throughout.

symmetric in this limit. This finally provides a quantitative explanation of the observation made in (Renart et al., 2010) that the time-lag between excitation and inhibition vanishes in the limit of infinitely large networks. For a different synaptic rescaling $J \sim N^{-1}$ while keeping ρ^2 constant by appropriate additional input to each neuron (see Helias et al., 2013 applied to the LIF model), all multiplicative factors decrease $\sim N^{-1}$ and so does the amplitude of all covariances. Hence the asymmetry of $c_{\mathcal{E}\mathcal{I}}$ does not vanish in this limit. The same results hold for the case of output noise where the term with φ_1 describes the common input part of the covariance. In this case and for finite network size, $c_{\mathcal{I}\mathcal{E}}$ coincides with $c_{\mathcal{E}\mathcal{E}}$ and $c_{\mathcal{E}\mathcal{I}}$ with $c_{\mathcal{I}\mathcal{I}}$ for $t > 0$, having a discontinuous jump at the time of the synaptic delay $t = d$. For time lags smaller than the delay all four covariances coincide. This is due to causality, as the second neuron cannot feel the influence of a fluctuation that happened in the first neuron less than one synaptic delay before. The covariance functions for systems corresponding to an OUP with input noise contain neither discontinuities nor sharp peaks at $t = d$, but $c_{\mathcal{E}\mathcal{I}}$ and $c_{\mathcal{I}\mathcal{E}}$ have maxima and minima near this location. This observation can be interpreted as a result of the stochastic nature of the binary model where changes in the input influence the state of the neuron only with a certain probability. So, the entries of $\mathbf{c}$ in this case take different values for $|t| < d$ but show the tendency to approach each other with increasing $|t| \gg d$. This tendency increases with network size. Our analytical solutions (18) for input noise and (21) for output noise hence explain the model-class dependent differences in the shape of covariance functions.

The two above mentioned synaptic scaling procedures are commonly termed "strong coupling" ($J \sim 1/\sqrt{N}$) and "weak

coupling" ($J \sim 1/N$), respectively. The results shown in **Figure 6** were obtained for $J = 2/\sqrt{N}$ and $\beta = 0.5$, so the number of synapses required to cause a notable effect on the gain function is $1/(\beta J) = \sqrt{N}$, which is small compared to the number of incoming synapses pN. Hence the network is in the strong coupling regime. Also note that for infinite slope of the gain function, $\beta \to \infty$, the magnitude of the covariance becomes independent of the synaptic amplitude J, in agreement with the linear theory presented here. This finding can readily be understood by the linearization procedure, presented in the current work, that takes into account the network- generated fluctuations of the total input. The amplitude σ of these fluctuations scales linearly in J and the effective susceptibility depends on J/σ in the case $\beta \to \infty$, explaining the invariance (Grytskyy et al., 2013). In the current manuscript we generalized this procedure to finite slopes β and to other models than the binary neuron model.

Our approach enables us to map results obtained for one neuron model to another, in particular we extend the theory of all considered models to capture synaptic conduction delays, and devise a simpler way to obtain solutions for systems considered earlier (Ginzburg and Sompolinsky, 1994). Our derivation of covariances in spiking networks does not rely on the advanced Wiener-Hopf method (Hazewinkel, 2002), as earlier derivations (Hawkes, 1971; Helias et al., 2013) do, but only employs elementary methods. Our results are applicable for general connectivity matrices, and for the purpose of comparison with simulations we explicitly derive population averaged results. The averages of the dynamics of the linear rate model equations are exact for random network architectures with fixed out-degree, and approximate for fixed in-degree. Still, for non-linear models the linearization for fixed in-degree networks are simpler, because the homogeneous input statistics results in an identical linear response kernel for all cells. Finally we show that the oscillatory properties of networks of integrate-and-fire models (Brunel, 2000; Helias et al., 2013) are model-invariant features of all of the studied dynamics, given inhibition acts with a synaptic delay. We relate the collective oscillations to the pole structure of the cross spectrum, which also determines the power spectra of population signals such as EEG, ECoG, and the LFP.

The presented results provide a further step to understand the shape and to unify the description of correlations in recurrent networks. We hope that our analytical results will be useful to constrain the inverse problem of determining the synaptic connectivity given the correlation structure of neurophysiological activity measurements. Moreover the explicit expressions for covariance functions in the time domain are a necessary prerequisite to understand the evolution of synaptic amplitudes in systems with spike-timing dependent plasticity and extend the existing methods (Burkitt et al., 2007; Gilson et al., 2009, 2010) to networks including inhibitory neurons and synaptic conduction delays.

ACKNOWLEDGMENTS

We gratefully appreciate ongoing technical support by our colleagues in the NEST Initiative, especially Moritz Deger for the implementation of the Hawkes model. Binary and spiking network simulations performed with NEST (www.nest-initiative.org). Partially supported by the Helmholtz Association: HASB and portfolio theme SMHB, the Jülich Aachen Research Alliance (JARA), the Next-Generation Supercomputer Project of MEXT, and EU Grant 269921 (BrainScaleS).

REFERENCES

Abeles, M. (1991). *Corticonics: Neural Circuits of the Cerebral Cortex*. 1st Edn. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511574566

Ascher, D., Dubois, P. F., Hinsen, K., Hugunin, J., and Oliphant, T. (2001). *An Open Source Project: Numerical Python*. Technical Report UCRL-MA-128569, Livermore, CA: Lawrence Livermore National Laboratory.

Bi, G.-Q., and Poo, M.-M. (1999). Distributed synaptic modification in neural networks induced by patterned stimulation. *Nature* 401, 792–796. doi: 10.1038/44573

Bienenstock, E. (1995). A model of neocortex. *Network* 6, 179–224. doi: 10.1088/0954-898X/6/2/004

Braitenberg, V., and Schüz, A. (1991). *Anatomy of the Cortex: Statistics and Geometry*. Berlin; Heidelberg; New York: Springer-Verlag.

Bronstein, I. N., Semendjajew, K. A., Musiol, G., and Mühlig, H. (1999). *Taschenbuch der Mathematik*. 4th Edn. Frankfurt am Main: Verlag Harri Deutsch.

Brunel, N. (2000). Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons. *J. Comput. Neurosci.* 8, 183–208. doi: 10.1023/A:1008925309027

Brunel, N., and Hakim, V. (1999). Fast global oscillations in networks of integrate-and-fire neurons with low firing rates. *Neural Comput.* 11, 1621–1671. doi: 10.1162/089976699300016179

Buice, M. A., Cowan, J. D., and Chow, C. C. (2009). Systematic fluctuation expansion for neural network activity equations. *Neural Comput.* 22, 377–426. doi: 10.1162/neco.2009.02-09-960

Burak, Y., Lewallen, S., and Sompolinsky, H. (2009). Stimulus-dependent correlations in threshold-crossing spiking neurons. *Neural Comput.* 21, 2269–2308. doi: 10.1162/neco.2009.07-08-830

Burkitt, A. N., Gilson, M., and van Hemmen, J. (2007). Spike-timing-dependent plasticity for neurons with recurrent connections. *Biol. Cybern.* 96, 533–546. doi: 10.1007/s00422-007-0148-2

Buzsáki, G., and Wang, X. J. (2012). Mechanisms of gamma oscillations. *Annu. Rev. Neurosci.* 35, 203–225. doi: 10.1146/annurev-neuro-062111-150444

Cohen, M. R., and Kohn, A. (2011). Measuring and interpreting neuronal correlations. *Nat. Rev. Neurosci.* 14, 811–819. doi:10.1038/nn.2842.

Corless, R. M., Gonnet, G. H., Hare, D. E. G., Jeffrey, D. J., and Knuth, D. E. (1996). On the lambert w function. *Adv. Comput. Math.* 5, 329–359. doi: 10.1007/BF02124750

Diesmann, M., Gewaltig, M.-O., and Aertsen, A. (1999). Stable propagation of synchronous spiking in cortical neural networks. *Nature* 402, 529–533. doi: 10.1038/990101

Fourcaud, N., and Brunel, N. (2002). Dynamics of the firing probability of noisy integrate-and-fire neurons. *Neural Comput.* 14, 2057–2110. doi: 10.1162/089976602320264015

Galassi, M., Davies, J., Theiler, J., Gough, B., Jungman, G., Booth, M., and Rossi, F. (2006). *GNU Scientific Library Reference Manual 2nd Edn*. Bristol: Network Theory Limited.

Gardiner, C. W. (2004). *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences*, 3rd Edn. Springer Series in Synergetics. Berlin: Springer.

Gerstein, G. L., and Perkel, D. H. (1969). Simultaneously recorded trains of action potentials: analysis and functional interpretation. *Science* 881, 828–830. doi: 10.1126/science.164.3881.828

Gewaltig, M.-O., and Diesmann, M. (2007). NEST (NEural Simulation Tool). *Scholarpedia* 2, 1430. doi: 10.4249/scholarpedia.1430

Gilson, M., Burkitt, A. N., Grayden, D. B., Thomas, D. A., and van Hemmen, J. L. (2009). Emergence of network structure due to spike-timing-dependent plasticity in recurrent neuronal networks. I. Input selectivity - strengthening correlated input pathways. *Biol. Cybern.* 101, 81–102. doi: 10.1007/s00422-009-0319-4

Gilson, M., Burkitt, A. N., and van Hemmen, J. L. (2010). STDP in recurrent neuronal networks. *Front. Comput. Neurosci.* 4:23. doi: 10.3389/fncom.2010.00023

Ginzburg, I., and Sompolinsky, H. (1994). Theory of correlations in stochastic neural networks. *Phys. Rev. E* 50, 3171–3191. doi: 10.1103/PhysRevE.50.3171

Grytskyy, D., Tetzlaff, T., Diesmann, M., and Helias, M. (2013). Invariance of covariances arises out of noise. *AIP Conf. Proc.* 1510, 258–262. doi: 10.1063/1.4776531

Hawkes, A. (1971). Point spectra of some mutually exciting point process. *J. R. Statist. Soc. Ser. B* 33, 438–443.

Hazewinkel, M. (Ed.) (2002). *Encyclopaedia of Mathematics*. Dordrecht: Kluwer Academic Publishers.

Helias, M., Tetzlaff, T., and Diesmann, M. (2013). Echoes in correlated neural systems. *New J. Phys.* 15:023002. doi: 10.1088/1367-2630/15/2/023002

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci. U.S.A.* 79, 2554–2558. doi: 10.1073/pnas.79.8.2554

Ito, J., Maldonado, P., Singer, W., and Grün, S. (2011). Saccade-related modulations of neuronal excitability support synchrony of visually elicited spikes. *Cereb. Cortex* 21, 2482–2497. doi: 10.1093/cercor/bhr020

Jones, E., Oliphant, T., Peterson, P., et al. (2001). *SciPy: Open Source Scientific Tools for Python*. Available online at: http://www.scipy.org/

Kilavik, B. E., Roux, S., Ponce-Alvarez, A., Confais, J., Gruen, S., and Riehle, A. (2009). Long-term modifications in motor cortical dynamics induced by intensive practice. *J. Neurosci.* 29, 12653–12663. doi: 10.1523/JNEUROSCI.1554-09.2009

Kriener, B., Tetzlaff, T., Aertsen, A., Diesmann, M., and Rotter, S. (2008). Correlations and population dynamics in cortical networks. *Neural Comput.* 20, 2185–2226. doi: 10.1162/neco.2008.02-07-474

Markram, H., Lübke, J., Frotscher, M., and Sakmann, B. (1997). Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs. *Science* 275, 213–215. doi: 10.1126/science.275.5297.213

Moreno-Bote, R., and Parga, N. (2006). Auto- and crosscorrelograms for the spike response of leaky integrate-and-fire neurons with slow synapses. *Phys. Rev. Lett.* 96:028101. doi: 10.1103/PhysRevLett.96.028101

Morrison, A., and Diesmann, M. (2008). "Maintaining causality in discrete time neuronal network simulations," in *Lectures in Supercomputational Neuroscience: Dynamics in Complex Brain Networks*, eds P. beim Graben, C. Zhou, M. Thiel, and J. Kurths (Understanding Complex Systems, Springer), 267–278.

Morrison, A., Mehring, C., Geisel, T., Aertsen, A., and Diesmann, M. (2005). Advancing the boundaries of high connectivity network simulation with distributed computing. *Neural Comput.* 17, 1776–1801. doi: 10.1162/0899766054026648

Papoulis, A., and Pillai, S. U. (2002). *Probability, Random Variables, and Stochastic Processes*, 4th Edn. Boston, MA: McGraw-Hill.

Perkel, D. H., Gerstein, G. L., and Moore, G. P. (1967). Neuronal spike trains and stochastic point processes. II. Simultaneous spike trains. *Biophys. J.* 7, 419–440. doi: 10.1016/S0006-3495(67)86597-4

Pernice, V., Staude, B., Cardanobile, S., and Rotter, S. (2011). How structure determines correlations in neuronal networks. *PLoS Comput. Biol.* 7:e1002059. doi: 10.1371/journal.pcbi.1002059

Pernice, V., Staude, B., Cardanobile, S., and Rotter, S. (2012). Recurrent interactions in spiking networks with arbitrary topology. *Phys. Rev. E* 85:031916. doi: 10.1103/PhysRevE.85.031916

Plesser, H. E., Eppler, J. M., Morrison, A., Diesmann, M., and Gewaltig, M.-O. (2007). "Efficient parallel simulation of large-scale neuronal networks on clusters of multiprocessor computers," in *Euro-Par 2007: Parallel Processing*, Vol. 4641 of *Lecture Notes in Computer Science*, eds A.-M. Kermarrec, L. Bougé, and T. Priol (Berlin: Springer-Verlag), 672–681.

Python Software Foundation. (2008). *The Python Programming Language*. Available online at: http://www.python.org.

Rajan, K., and Abbott, L. (2006). Eigenvalue spectra of random matrices for neural networks. *Phys. Rev. Lett.* 97:188104. doi: 10.1103/PhysRevLett.97.188104

Renart, A., De La Rocha, J., Bartho, P., Hollender, L., Parga, N., Reyes, A., and Harris, K. D. (2010). The asynchronous state in cortical cicuits. *Science* 327, 587–590. doi: 10.1126/science.1179850

Risken, H. (1996). *The Fokker-Planck Equation*. Verlag; Berlin; Heidelberg: Springer.

Rosenbaum, R., and Josic, K. (2011). Mechanisms that modulate the transfer of spiking correlations. *Neural Comput.* 23, 1261–1305. doi: 10.1162/NECO_a_00116

Rotter, S., and Diesmann, M. (1999). Exact digital simulation of time-invariant linear systems with applications to neuronal modeling. *Biol. Cybern.* 81, 381–402. doi: 10.1007/s004220050570

Rumelhart, D. E., McClelland, J. L., and the PDP Research Group (1986). *Parallel Distributed Processing, Explorations in the Microstructure of Cognition: Foundations*, Vol. 1. Cambridge, MA: MIT Press.

Shadlen, M. N., and Movshon, A. J. (1999). Synchrony unbound: A critical evaluation of the temporal binding hypothesis. *Neuron* 24, 67–77. doi: 10.1016/S0896-6273(00)80822-3

Singer, W. (1999). Neuronal synchrony: a versatile code for the definition of relations? *Neuron* 24, 49–65. doi: 10.1016/S0896-6273(00)80821-1

Sompolinsky, H., Yoon, H., Kang, K., and Shamir, M. (2001). Population coding in neuronal systems with correlated noise. *Phys. Rev. E* 64:51904. doi: 10.1103/PhysRevE.64.051904

Tchumatchenko, T., Malyshev, A., Geisel, T., Volgushev, M., and Wolf, F. (2010). Correlations and synchrony in threshold neuron models. *Phys. Rev. Lett.* 104, 058102. doi: 10.1103/PhysRevLett.104.058102

Tetzlaff, T., Helias, M., Einevoll, G., and Diesmann, M. (2012). Decorrelation of neural-network activity by inhibitory feedback. *PLoS Comput. Biol.* 8:e1002596. doi: 10.1371/journal.pcbi.1002596

Trousdale, J., Hu, Y., Shea-Brown, E., and Josic, K. (2012). Impact of network structure and cellular response on spike time correlations. *PLoS Comput. Biol.* 8:e1002408. doi: 10.1371/journal.pcbi.1002408

Uhlenbeck, G. E., and Ornstein, L. S. (1930). On the theory of the brownian motion. *Phys. Rev.* 36, 823–841. reprinted in Wax (1954). doi: 10.1103/PhysRev.36.823

van Vreeswijk, C., and Sompolinsky, H. (1996). Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science* 274, 1724–1726. doi: 10.1126/science.274.5293.1724

van Vreeswijk, C., and Sompolinsky, H. (1998). Chaotic balanced state in a model of cortical circuits. *Neural Comput.* 10, 1321–1371. doi: 10.1162/089976698300017214

von der Malsburg, C. (1981). *The Correlation Theory of Brain Function*. Internal report 81-2, Department of Neurobiology, Max-Planck-Institute for Biophysical Chemistry, Göttingen, Germany.

Wax, N. (eds.). (1954). *Selected Papers on Noise and Stochastic Processes*. New York, NY: Dover Publications.

Zohary, E., Shadlen, M. N., and Newsome, W. T. (1994). Correlated neuronal discharge rate and its implications for psychophysical performance. *Nature* 370, 140–143. doi: 10.1038/370140a0

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.

Received: 22 April 2013; accepted: 10 September 2013; published online: 18 October 2013.

Citation: Grytskyy D, Tetzlaff T, Diesmann M and Helias M (2013) A unified view on weakly correlated recurrent networks. Front. Comput. Neurosci. 7:131. doi: 10.3389/fncom.2013.00131

This article was submitted to the journal Frontiers in Computational Neuroscience.

Copyright © 2013 Grytskyy, Tetzlaff, Diesmann and Helias. 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) or licensor 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.

APPENDIX

CALCULATION OF THE POPULATION AVERAGED CROSS COVARIANCE IN TIME DOMAIN

We obtain the population averaged cross spectrum for the Ornstein-Uhlenbeck process with input noise by inserting the averaged connectivity matrix $\mathbf{w} = \mathbf{M}$ (14) into (8). The two eigenvalues of $\mathbf{M}$ are 0 and $L = Kw(1 - \gamma g)$. Taking these into account, we first rewrite the term

$$
\begin{aligned}
&(H_d(\omega)^{-1} - \mathbf{M})^{-1} \\
&\quad = \det(H_d(\omega)^{-1} - \mathbf{M})^{-1} \begin{pmatrix} H_d(\omega)^{-1} + Kw\gamma g & -Kw\gamma g \\ Kw & H_d(\omega)^{-1} - Kw \end{pmatrix} \\
&\quad = ((H_d(\omega)^{-1} - 0)(H_d(\omega)^{-1} - L))^{-1} \left(H_d(\omega)^{-1}\mathbf{1} + Kw \begin{pmatrix} \gamma g & -\gamma g \\ 1 & -1 \end{pmatrix} \right) \\
&\quad = f(\omega) \left(\mathbf{1} + Kw \begin{pmatrix} \gamma g & -\gamma g \\ 1 & -1 \end{pmatrix} H_d(\omega) \right),
\end{aligned}
$$

where we introduced $f(\omega) = (H_d(\omega)^{-1} - L)^{-1}$. The corresponding transposed and conjugate complex term follows analogously. Hence we obtain the expression for the cross spectrum (17). The residue of $f(\omega)$ at $\omega = z_k(L)$ is

$$
\begin{aligned}
\mathrm{Res}(f, \omega = z_k(L)) &= \lim_{\omega_1 \to \omega} \frac{\omega_1 - \omega}{f^{-1}(\omega_1)} \\
\overset{\text{l'Hopital}}{=} \lim_{\omega_1 \to \omega} \frac{1}{(f^{-1})'(\omega_1)} &= \left(\frac{d(e^{i\omega d}(1 + i\omega\tau))}{d\omega} \right)^{-1} \\
&= \left(id e^{i\omega d}(1 + i\omega\tau) + i\tau e^{i\omega d} \right)^{-1} \\
&= \left(idL + i\tau e^{i\omega d} \right)^{-1},
\end{aligned}
$$

where in the last step we used the condition for a pole $H_d(z_k)^{-1} = e^{iz_k d}(1 + iz_k\tau) = L$ (see "Spectrum of the dynamics"). The residue of $H_d(\omega)$ at $z(0) = \frac{i}{\tau}$ is $-\frac{i}{\tau}e^{d/\tau}$. Using the residue theorem, we need to sum over all poles within the integration contour $\{z_k(L)|k \in \mathbb{N}\} \cup \frac{i}{\tau}$ to get the expression for $\mathbf{c}(t) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \mathbf{C}(\omega)e^{i\omega t}d\omega = i\sum_{z \in \{z_k(L)|k\in\mathbb{N}\}\cup\frac{i}{\tau}} \mathrm{Res}(\mathbf{C}(z), z)e^{izt}$ for $t \geqq 0$. Sorting (17) to obtain four matrix prefactors and remainders with different frequency dependence, $\Phi_1(\omega) = f(\omega)f(-\omega)$, $\Phi_2(\omega) = f(\omega)f(-\omega)H_d(\omega)$, $\Phi_3(\omega) = \Phi_2(-\omega)$, and $\Phi_4(\omega) = f(\omega)f(-\omega)H_d(\omega)H_d(-\omega)$, we get (18). $\mathbf{C}(\omega)$ for output noise (20) is obtained by multiplying the expression for $\mathbf{C}(\omega)$ for input noise with $H_d^{-1}(\omega)H_d^{-1}(-\omega) = (1 + \omega^2\tau^2)$. In order to perform the back Fourier transformation one first needs to rewrite the cross spectrum in order to isolate the frequency independent term and the two terms that vanish for either $t < d$ or $t > d$, as described in "Fourier back transformation,"

$$
\begin{aligned}
\mathbf{C}(\omega) &= f(\omega)(\mathbf{1} + Kw \begin{pmatrix} \gamma g & -\gamma g \\ 1 & -1 \end{pmatrix} H_d(\omega))\mathbf{M}\mathbf{D}^T f(-\omega) \\
&\quad (\mathbf{1} + Kw \begin{pmatrix} \gamma g & 1 \\ -\gamma g & -1 \end{pmatrix} H_d(-\omega)) \\
&\quad + f(\omega)(\mathbf{1} + Kw \begin{pmatrix} \gamma g & -\gamma g \\ 1 & -1 \end{pmatrix} H_d(\omega))\mathbf{M}\mathbf{D} \\
&\quad + \mathbf{D}\mathbf{M}^T f(-\omega)(\mathbf{1} + Kw \begin{pmatrix} \gamma g & 1 \\ -\gamma g & -1 \end{pmatrix} H_d(-\omega)) + \mathbf{D} \\
&= f(\omega)\mathbf{M}\mathbf{D}\mathbf{M}^T f(-\omega) + f(\omega)\mathbf{M}\mathbf{D} + \mathbf{D}\mathbf{M}^T f(-\omega) + \mathbf{D},
\end{aligned}
$$

where in the last step we used $\begin{pmatrix} \gamma g & -\gamma g \\ 1 & -1 \end{pmatrix}\mathbf{M} = 0$, because $\mathbf{M}$ is symmetric, obtaining (21). For each of the first three terms in the last expression the right integration contour needs to be chosen as described in "Fourier back transformation" on the example of the general expression (16).

IMPLEMENTATION OF NOISY RATE MODELS

The dynamics is propagated in time steps of duration Δt (note that in other works we use h as a symbol for the computation step size, which here is used as the symbol for the kernel). The product of the connectivity matrix with the vector of output variables at the end of the previous step $i - 1$ is the vector $\mathbf{I}(t_i)$ of inputs at the current step i. The intrinsic time scale of the system is determined by the time constant τ. For sufficiently small time steps $\Delta t \ll \tau$ these inputs can be assumed to be time independent within one step. So we can use (3) or (4) and analytically convolve the kernel function h assuming the input to be constant over the time interval Δt. This corresponds to the method of exponential integration (Rotter and Diesmann, 1999, see App. C.6) requiring only local knowledge of the connectivity matrix $\mathbf{w}$. Note that this procedure becomes exact for $\Delta t \to 0$ and for finite Δt is an approximation. The propagation of the initial value $r_j(t_{i-1})$ until the end of the time interval takes the form $r_j(t_{i-1})\, e^{-\Delta t/\tau}$ because $h(t_i) = h(t_{i-1})\, e^{-\Delta t/\tau}$, so we obtain the expression $r_j(t_i)$ at the end of the step as

$$
r_j(t_i) = e^{-\Delta t/\tau}\, r_j(t_{i-1}) + (1 - e^{-\Delta t/\tau})\, I_j(t_i), \tag{43}
$$

where I_j denotes the input to the neuron j. For output noise the output variable of neuron j is $y_j = r_j + x_j$, with the locally generated additive noise x_j and hence the input is $I_j(t_i) = (\mathbf{w}\mathbf{y}(t_i))_j$. In the case of input noise the output variable is r_j and the additional noise is added to the input variable, $I_j(t_i) = (\mathbf{w}\mathbf{r}(t_i))_j + x_j(t_i)$. In both cases x_j is implemented as a binary noise: in each time step, x_j is independently and randomly chosen to be 1 or -1 with probability 0.5 multiplied with $\rho/\sqrt{\Delta t}$ to satisfy (2) for discretized time. Here the δ-function is replaced by a "rectangle" function that is constant on the interval of length Δt, vanishes elsewhere, and has unit integral. The factor Δt^{-1} in the expression for x^2 ensures the integral to be unity. So far, the implementation assumes the synaptic delay to be zero. To implement a non-zero synaptic delay d, each object representing a neuron contains an array b of length $l_d = d/\Delta t$ acting as a ring buffer. The input $I_j(t_i)$ used to calculate the output rate at step i according to (43) is then taken from position i mod l_d of this array and after that replaced by the input presently received from the network, so that the new input will be used only after one delay has passed. This sequence

of buffer handling can be represented as

$$I_j(t_i) \leftarrow b[i \bmod l_d]$$
$$b[i \bmod l_d] \leftarrow \begin{cases} (\mathbf{w}\,\mathbf{r})_j + x_j & \text{for input noise} \\ (\mathbf{w}\,\mathbf{y})_j & \text{for output noise} \end{cases}.$$

The model is implemented in Python version 2.7 (Python Software Foundation, 2008) using numpy 1.6.1 (Ascher et al., 2001) and scipy 0.9.0 (Jones et al., 2001).

IMPLEMENTATION OF BINARY NEURONS IN A SPIKING SIMULATOR CODE

The binary neuron model is implemented in the NEST simulator, version 2.2.1 (Gewaltig and Diesmann, 2007), which allows distributed simulation on parallel machines and handles synaptic delays in the established framework for spiking neurons (Morrison et al., 2005). The name of the model is "`ginzburg_neuron`". In NEST information is transmitted in form of point events, which in case of binary neurons are sent if the state of the neuron changes: one spike is sent for a down-transition and two spikes at the same time for an up-transition, so the multiplicity reflects the type of event. The logic to decode the original transitions is implemented in the function handle shown in Alg. 2. If a single spike is received, the synaptic weight w is subtracted from the input buffer at the position determined by the time point of the transition and the synaptic delay. In distributed simulations a single spike with multiplicity 2 sent to another machine is handled on the receiving side as two separate events with multiplicity 1 each. In order to decode this case on the receiving machine we memorize the time (t_{last}) and origin (global id gid_{last} of the sending neuron) of the last arrived spike. If both coincide to the spike under consideration, the sending neuron has performed an up transition $0 \to 1$. We hence add twice the synaptic weight $2w$ to the input buffer of the target neuron, one that reflects the real change of the system state and another that compensates the subtraction of w after reception of the first spike of a pair. The algorithm relies on the fact that within NEST two spikes that are generated by one neuron at the same time point are delivered sequentially to the target neurons. This is assured, because neurons are updated one by one: The update propagates each neuron by a time step equal to the minimal delay $d_{\min}$ in the network. All spikes generated within one update step are written sequentially into the communication buffers, and finally the buffers are shipped to the other processors (Morrison et al., 2005). Hence a pair of spikes generated by one neuron within a single update step will be delivered consecutively and will not be interspersed by spikes from other neurons with the same time stamp.

The model exhibits stochastic transitions (at random points in time) between two states. The transitions are governed by probabilities $\phi(h)$. Using asynchronous update (Rumelhart et al., 1986), in each infinitesimal interval $[t, t + \delta t)$ each neuron in the network has the probability $\frac{1}{\tau}\delta t$ to be chosen for update (Hopfield, 1982). A mathematically equivalent formulation draws the time points of update independently for all neurons. For a particular neuron, the sequence of update points has exponentially distributed intervals with mean duration τ, i.e., it forms a Poisson process with rate τ^{-1}. We employ the latter formulation to incorporate binary neuron models in the globally time-driven spiking simulator NEST (Gewaltig and Diesmann, 2007) and constrain the points of transition to a discrete time grid $\Delta t = 0.1$ ms covering the interval $d_{\min} \geq \Delta t$. This neuron state update is implemented by the algorithm shown in Alg. 1. Note that the field h is updated in steps of Δt while the activity state is updated only when the current time exceeds the next potential transition point. As the last step of the activity update we draw an exponentially distributed time interval to determine the new potential transition time. The potential transition time is represented with a higher resolution (on the order of microseconds) than Δt to avoid a systematic bias of the mean inter-update-interval. This update scheme is identical to the one used in (Hopfield, 1982). Note that the implementation is different from the classical asynchronous update scheme (van Vreeswijk and Sompolinsky, 1998), where in each discrete time step Δt exactly one neuron is picked at random. The mean inter-update-interval (time constant τ in Alg. 1) in the latter scheme is determined by $\tau = \Delta t N$, with N the number of neurons in the network. For small time steps both schemes converge so that update times follow a Poisson process.

At each update time point the neuron state becomes 1 with the probability given by the function ϕ applied to the input at that time according to (25) and 0 with probability $1 - \phi$. The input is a function of the whole system state and is constant between spikes which indicate state changes. Each neuron therefore maintains a state variable h at each point in time holding the summed input and being updated by adding and subtracting the input read from the ring buffer b at the point readpos(t) corresponding to the current time (see Morrison et al., 2005, for the implementation of the ring buffer, i.p. Fig 6). The ring buffer enables us to implement synaptic delays. For technical

Algorithm 1 | Update function of a binary neuron embedded in the spiking network simulator NEST.

```
y ← 0 // initially neuron is inactive
t_next ← −τ log(rand()) // next time point of update

for each time step t:

    h ← h + b[readpos(t)]

    if t > t_next:
        // up−state with probability given by
        // gain function φ depending on input h(t)

        if φ(h) > rand():
            y_new ← 1
        else:
            y_new ← 0

        if y_new ≠ y:
            // down transition: send single spike
            // up transition: send two spikes
            send (y_new + 1 spikes)

            y ← y_new

        // add an exponentially distributed time interval
        t_next ← t_next − τ log(rand())
```

The function readpos(t) *returns a position in the ring buffer b corresponding to the current time point.*

Algorithm 2 | Input spike handler of a binary neuron embedded in the spiking network simulator NEST.

```
1  handle(t_spike, d, gid, m):
2
3      if m = 1:
4
5      // multiplicity = 1, either a single 1→0 event
6      // or the first or second of a pair of 0→1 events
7
8          if gid = gid_lastspike and t_spike = t_lastspike:
9
10             // received twice the same event, so transition 0→1
11             // add 2w to compensate for subtraction after reception
12             // of first event
13             b[pos(t_spike, d, t)] ← b[pos(t_spike, d, t)] + 2w
14
15         else:
16             // count this event negatively,
17             // assuming it comes as single event
18             // transition 1→0
19             b[pos(t_spike, d, t)] ← b[pos(t_spike, d, t)] − w
20     else:
21
22     // multiplicity != 1
23
24         if m = 2:
25
26             // count this event positively, transition 0→1
27             b[pos(t_spike, d, t)] ← b[pos(t_spike, d, t)] + w
28             gid_lastspike ← gid
29             t_lastspike ← t_spike
```

The simulation kernel calls the handle function for each spike event to be delivered to the neuron. A spike event is characterized by the time point of occurrence t_{spike}*, the synaptic delay* d *after which the event should reach the target, the global id* gid *identifying the sending neuron, and the multiplicity* $m \geq 1$*, indicating the reception of multiple spike events. The function* $\text{pos}(t_{\text{spike}}, d, t)$ *returns the position in the ring buffer* b *to which the spike is added so that it will be read at time* $t + d$ *by the update function of the neuron, see Alg. 1.*

reasons this implementation requires a minimal delay of a single simulation time step (Morrison and Diesmann, 2008). The gain function ϕ applied to the input h has the form

$$\phi(h) = c_1 h + c_2 \frac{1}{2}\left(1 + \tanh(c_3(h - \theta))\right), \tag{44}$$

where throughout this manuscript we used $c_1 = 0$, $c_2 = 1$, and $c_3 = \beta$, as defined in "Parameters of simulations".

IMPLEMENTATION OF HAWKES NEURONS IN A SPIKING SIMULATOR CODE

Hawkes neurons (Hawkes, 1971) were introduced in the NEST simulator in version 2.2.0 (Gewaltig and Diesmann, 2007). The name of the model is "`pp_psc_delta`". In the following we describe the implemented neuron model in general and mention the particular choices of parameter and correspondences to the theory presented in "Hawkes processes". The dynamics of the quasi-membrane potential u is integrated exactly within a time step Δt of the simulation (Rotter and Diesmann, 1999), expressing the voltage $u(t_i)$ at the end of time step i by the membrane potential at the end of the previous time step $u(t_{i-1})$ as

$$u(t_i) = e^{-\Delta t/\tau}\, u(t_{i-1}) + (1 - e^{-\Delta t/\tau})\, R_m I_e + b(t_i), \tag{45}$$

where I_e is a time-step wise constant input current (equal to 0 in all simulations presented in this article) and $R_m = \tau_m / C_m$ is the membrane resistance. The buffer $b(t_i)$ contains the summed contributions of incoming spikes, multiplied by their respective synaptic weight, which have arrived at the neuron within the interval $(t_{i-1}, t_i]$. b is implemented as a ring-buffer in order to handle the synaptic delay, logically similar as in "Implementation of noisy rate models," described in detail in Morrison et al. (2005). The instantaneous spike emission rate is $\lambda = [c_1 u + c_2 e^{c_3 u}]_+$, where we use $c_3 = 0$ in all simulations presented here. The quantities in the theory "Hawkes processes," in particular in (35), are related to the parameters of the simulated model in the following way. The quantity r relates to the membrane potential u as $r = c_1 u + c_2$ and the background rate ν agrees to $c_2 = \nu$. Hence the synaptic weight J_{ij} corresponds to the synaptic weight in the simulation multiplied by c_1. For the correspondence of the Hawkes model to the OUP with output noise of variance ρ^2 we use (38) to adjust the background rate ν in order to obtain the desired rate $\lambda_0 = \rho^2$ and we choose the synaptic weight J of the Hawkes model so that the linear coupling strength w of the OUP agrees to the effective linear weight given by (39). These two constraints can be fulfilled simultaneously by solving (38) and (39) by numerical iteration. The spike emission of the model is realized either with or without dead time. In this article we only used the latter. In the presence of a dead time, which is constrained to be larger than the simulation time step, at most one spike can be generated within a time step. A spike is hence emitted with the probability $p_{\geq 1} = 1 - e^{\lambda \Delta t}$, where $e^{\lambda \Delta t}$ is the probability of the complementary event (emitting 0 spikes), implemented by comparing a uniformly distributed random number to $p_{\geq 1}$. The refractory period is handled as described in Morrison et al. (2005). Without refractoriness, the number of emitted spikes is drawn from a Poisson distribution with parameter $\lambda \Delta t$, implemented in the GNU Scientific Library (Galassi et al., 2006). Reproducibility of the random sequences for different numbers of processes and threads is ensured by the concept of random number generators assigned to virtual processes, as described in (Plesser et al., 2007).

PARAMETERS OF SIMULATIONS

For all simulations we used $\gamma = 0.25$ corresponding to the biologically realistic fraction of inhibitory neurons, a connectivity probability $p = 0.1$, and a simulation time step of $\Delta t = 0.1$ ms. For binary neurons we measured the covariance functions with a resolution of 1 ms, for all other models the resolution is 0.1 ms. Simulation time is 10, 000 ms for linear rate and for LIF neurons, 50, 000 ms for Hawkes, and 100, 000 ms for binary neurons. The covariance is obtained for a time window of ± 100 ms.

The parameters for simulations of the LIF model presented in **Figure 7** and **Figure 8** are $J = 0.1$ mV, $\tau = 20$ ms, $\tau_s = 2$ ms, $\tau_r = 2$ ms, $V_\theta = 15$ mV, $V_r = 0$, $g = 6$, $d = 3$ ms, $N = 8000$. The number of neurons in the corresponding networks of other models is the same. Cross covariances are measured between the summed spike trains of two disjoint populations of $N_{\text{rec}} = 1000$ neurons each. The single neuron autocovariances a_α are averaged over a subpopulation of 100 neurons. The autocovariances of the population averaged activity $\frac{1}{N_\alpha} a_\alpha + C_{\alpha\alpha}$ for population $\alpha \in \{\mathcal{E}, \mathcal{I}\}$ (shown in **Figure 7**) are constructed from the estimated single neuron population averaged autocovariances a_α and cross covariances $C_{\alpha\alpha}$. This enables us to estimate a_α and $C_{\alpha\alpha}$ from the activity of a small subpopulation and still assigns the

correct relative weights to both contributions. The corresponding effective parameters describing the system dynamics are $\mu = 15$ mV, $\sigma = 10$ mV, $r = 23.6$ Hz (see (40) and the following text for details).

The parameters of the Hawkes model and of the noisy rate model with output noise yielding quantitatively agreeing covariance functions are:

- For simulations of the noisy rate model with output noise presented in **Figure 7** and **Figure 2** the parameters are $w = 0.0043$, $g \approx 5.93$, $\tau = 4.07$ ms, $\rho^2 = 23.6$ Hz, $d = 3$ ms (see (3), (4)). In **Figure 2** also results for $d = 1$ ms and for input noise are shown. Signals are measured from $N_{\text{rec}} = 500$ neurons in each population to obtain $c_{\mathcal{EI}}$, $c_{\mathcal{IE}}$ and from the whole population to determine $c_{\mathcal{EE}}$ and $c_{\mathcal{II}}$. The cross covariances $C_{\mathcal{EE}}$ and $C_{\mathcal{II}}$ are estimated from two disjoint subpopulations each comprising half of the neurons of the respective population.
- For the network of Hawkes neurons presented in **Figure 7** we used $\lambda_0 \approx 22.54$ Hz (see (38)), $J = 0.0055$ mV, $d = 3$ ms, and the same g and τ as for the noisy rate model. We measured the cross covariances in the same way as for the LIF model, but using the spike trains from sub-populations of $N_{\text{rec}} = 2000$ neurons. The autocovariances of the population averaged activity were estimated from the whole populations.

The network of binary neurons shown in **Figure 8** uses $\theta = -3.89$ mV, $\beta = 0.5$ mV^{-1}, $J = 0.02$ mV, $d = 3$ ms (see (25), (44)), and the same g and τ as the noisy rate model. Covariances are measured using the signals from all neurons.

The simulation results for the network of binary neurons presented in **Figure 6** uses $\theta = -2.5$ mV, $\tau = 10$ ms, $\beta = 0.5$ mV^{-1}, $g = 6$, $J \approx 0.0447$ mV, $N = 2000$ and the smallest possible value of synaptic delay is $d = 0.1$ ms equal to time resolution (the same set of parameters only with modified $\beta = 1$ mV^{-1} was used to create **Figure 5**). The cross covariances $C_{\mathcal{EE}}$ and $C_{\mathcal{II}}$ are estimated from two disjoint subpopulations each comprising half of the neurons of the respective population, $c_{\mathcal{EI}}$ is measured between two such subpopulations. For $c_{\mathcal{EE}}$ and $c_{\mathcal{II}}$ we used the full populations.

The parameters required for a quantitative agreement with the rate model with input noise are $w \approx 0.011$, $\rho \approx 2.23\sqrt{\text{ms}}$. We used the same parameters in **Figure 3**, where additionally results for $w = 0.018$ are shown. The population sizes are the same as for the binary network. The covariances are estimated in the same way as for the rate model with output noise. Note that the definition of noisy rate models has no limitation for units of ρ^2. These can be arbitrary and are chosen differently as required by the correspondence with either spiking or binary neurons.

frontiers in
COMPUTATIONAL NEUROSCIENCE

OPINION ARTICLE
published: 10 September 2013
doi: 10.3389/fncom.2013.00125

Efficient neural codes can lead to spurious synchronization

Massimiliano Zanin[1,2]* and David Papo[3]

[1] Departamento de Engenharia Electrotécnica, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Lisboa, Portugal
[2] Innaxis Foundation and Research Institute, Madrid, Spain
[3] Center for Biomedical Technology, Technical University of Madrid, Madrid, Spain
*Correspondence: mzanin@mzanin.com

Edited by:
Ruben Moreno-Bote, Foundation Sant Joan de Deu, Spain

Reviewed by:
Klaus Wimmer, Universitat de Barcelona, Spain

Keywords: synchronization, neural models, boolean code, EEG/MEG, stimuli

Experimental and computational evidence shows that cognitive function requires an optimal balance between global integrative and local functionally specialized processes (Tononi et al., 1998). This balance can be described in terms of transient short-lived episodes of synchronized activity between different parts of the brain (Friston, 2000; Breakspear, 2002). Synchronization over multiple frequency bands is thought to subserve fundamental operations of cortical computation (Varela et al., 2001; Fries, 2009), and to be one of the mechanisms mediating the large-scale coordination of scattered functionally specialized brain regions. For instance, transient synchronization of neuronal oscillatory activity in the 30–80 Hz range has been proposed to act as an integrative mechanism, binding together spatially distributed neural populations in parallel networks during sensory perception and information processing (Singer, 1995; Miltner et al., 1999; Rodriguez et al., 1999). More generally, synchrony may subserve an integrative function in cognitive functions as diverse as motor planning, working or associative memory, or emotional regulation (Varela, 1995).

Over the past 15 years, cognitive neuroscientists have tried to capture and quantify neural synchronies across distant brain regions both during spontaneous brain activity and in association with the execution of a wide range of cognitive tasks, using neuroimaging techniques such as functional resonance imaging, electro- or magneto-encephalography. Theoretical advances in various fields including non-linear dynamical systems theory have allowed the study of various types of synchronization from time series (Pereda et al., 2005), and to address important issues such as determining whether observed couplings do not reflect a mere correlation between activities recorded at two different brain regions but rather a causal relationship (Granger, 1969) whereby a brain region would cause the activity of the other one.

However, not all measured synchrony may in fact represent neurophysiologically and cognitively relevant computations: various confounding effects may mislead into identifying functional connectivity, defined as the temporal correlations between spatially remote neurophysiological events, with effective connectivity, i.e., the influence one neuronal system exerts over another (Friston, 1994). For instance, measured synchrony may stem from common thalamo-cortical afferents or neuromodulatory input from ascending neurotransmitter systems, or may be the visible part of indirect effective connectivity. Other technique-specific artifactual sources of synchrony, for instance induced by volume conduction, are also well-known to cognitive neuroscientists (Stam et al., 2007).

Here, we address a further (extracranial) confounding source: the appearance of simultaneous, yet uncorrelated stimuli. We show how the activity of two groups of binary neurons, whose output code is optimized to represent rare events with short codes, can exhibit a synchronization when such rare events appear, even in the absence of shared information or common computational activities.

1. THE MODEL

We suppose that a neuron codifies an external stimulus with a set of spikes, to transmit information about the event to other regions of the neural system. For the sake of simplicity, let's also suppose that all stimuli are drawn from a finite set of events $\mathrm{E} = \{e_1, \ldots, e_N\}$, N being the total number of events. Each event i is characterized by two strongly related features: the frequency of appearance f_i and the importance factor m_i. Clearly, rare events are also the most important ones. For instance, the image of a group of trees is quite common for an animal, and should not attract his attention. On the other hand, a predator appearing behind such trees is far less frequent, and the importance of a fast response to the event, high. Therefore, for each event i, the relation $m_i = 1/f_i$ is defined.

Each neuron optimizes its code to represent such an environment, i.e., it assigns a symbol s_i drawn from an alphabet $\mathcal{S}$ to each input event i. As the neuron natural language is composed of spikes, each symbol s_i is defined as a sequence of spikes and silences; this is represented by a sequence of 0's and 1's, of arbitrary length, forming a Boolean code. In other words, and from an information science perspective, each symbol s_i is a number in its Boolean representation.

In the creation of the code, the neurons use all their available knowledge concerning their environment, given by f_i and m_i, trying to fulfilling two conditions. First, the cost associated with the transmission of information should be minimized, thus as few spikes as possible should be generated; this favors large symbols with few 1's and a large proportion of 0's. This condition is energy saving, but increases the neuron's response time. Therefore, a second condition ensures that the neuron minimizes symbol length, particularly those associated with events or

items of great importance, i.e., with low f_i and high m_i.

A *cost* given by:

$$C = \sum_i \left[\alpha \frac{b_i f_i}{l_i} + (1-\alpha)\, l_i m_i \right] \quad (1)$$

accounts for the trade-off between these conditions is associated to each code, and minimized by the neuron in a training phase representing a natural selection process. The contribution of each symbol i to the total is given by two terms—see Equation 1. The first, involving the number of spikes in the symbol (b_i), its expected frequency of appearance (f_i) and its length (l_i), expresses the probability of having the neuron spiking , at a given time, and thus the expected energetic cost of the code. The second term penalizes the appearance of long symbols codifying important messages. Finally, the parameter α defines the balance between both contributions to the total cost: for $\alpha \approx 0$ ($\alpha \approx 1$) the total cost is dominated by the length of important symbols (by the energetic cost).

Two additional requirements are added. First, for different events no to be confused, all symbols should be different, i.e., $s_i \neq s_j$. Second, all symbols should start with a spike (a 1) and have at least one zero, in order to be recognizable and to avoid codes composed only of silences or spikes.

Due to the computational cost of optimizing such codes when multiple events are considered, the process is performed by means of a *greedy algorithm* Cormen et al. (2001), that is, by starting with an empty set, and adding one symbol at the time, making the locally optimal choice at each iteration.

2. RESULTS

We now explore how a spurious synchronization between different neurons (or groups of them) can be achieved even in the absence of any information transfer.

Neurons are supposed to work independently, that is, they receive independent inputs from the environment and create their optimal code to process and transmit such information. For instance, two groups of neurons may receive two different and uncorrelated stimuli, corresponding to the image of a predator and the sound of a thunder.

Following this idea, a large number of neurons are modeled and their codes created. Each neuron has its independent set of stimuli, half of them highly probable (and therefore, less important), and half of them with low probability of appearance.

Using this information, all codes are generated, and a time series for each neuron is created, by presenting sequences of stimuli at random, and recording the neuron's corresponding activity. Time series are divided into two parts of equal length. During the first half, neurons are stimulated by high-probability events; the opposite occurs during the second half. Following the previous example, we suppose that the organism is resting quietly at the beginning, and then spots a predator and hears a thunder. Furthermore, we suppose that neurons do not respond with the same velocity to the external stimuli: each neuron receives its inputs with a delay drawn from a uniform distribution defined between 0 and 400 time steps.

Figure 1 Left depicts the evolution of the time series generated by two groups of neurons, each one composed of 500 neurons, for $\alpha = 0.1$, 40 stimuli, and a transition interval of 400. Each series is clearly divided in two epochs, the first one corresponding to the time window [0, 5000], in which no relevant event appears, and a second window [5000, 10000] in which neurons respond to rare external stimuli. As previously described, an efficient code requires important stimuli to be codified with short symbols, which, in turns, are associated with high spike densities. This effect is clearly shown in **Figure 1** Left, where the proportion of spiking neurons after time 5000 is roughly increased by 0.05.

As neural codes are independently generated for the 1000 neurons considered, with different probability distributions, and external stimuli are also triggered in an independent way, no synchronization is expected between both time series. Indeed, if one computes the Pearson's correlation coefficient between both series within the time window [0, 5000], the result is in the order of 10^{-4}. Nonetheless, an interesting result is obtained when the correlation is calculated by means of a sliding window; in other words, a time-varying correlation is obtained, whose value at time t represents the dynamics of both neural groups in the interval $[t-200, t+200]$. Intuitively, when analyzing the series near time 5000, both series share the same trend, i.e., an upward dynamics, thus leading to a positive synchronization. Such

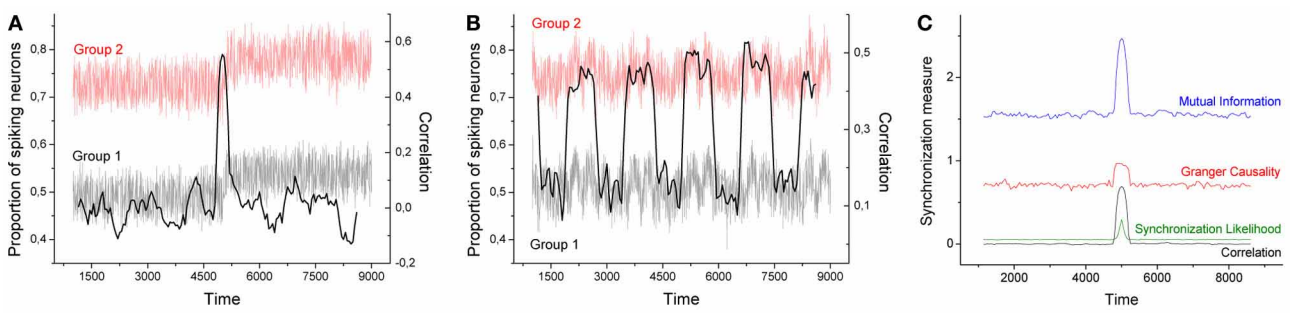


FIGURE 1 | (A,B) Time series of the proportion of spiking neurons generated by two groups of 500 neurons (gray and red lines). In panel **A** (panel **B**), the probability of finding rare events is changed at time 5000 (is continuously changed). Time series of group 2 is represented with an offset of 0.25. The black solid line represents the evolution of the Pearson's correlation coefficient between both groups, calculated with a sliding window of size 400. **(C)** Average values for the four synchronization metrics, using the same event sets of panel **(A)**. All neural codes are optimized for $\alpha = 0.1$.

effect is shown in **Figure 1** Left, black line and right scale: around time 5000 the Pearson's correlation coefficient jumps to 0.6.

To confirm this result, **Figure 1** Right reports the average synchronization level obtained in 100 realizations of the previously described process, as obtained by 4 commonly used metrics for the assessment of synchronization in brain activity:

- Correlation: Pearson's linear correlation between the two time series.
- Granger causality: following the original definition in Wiener (1956), a time series is said to cause a second one if one can improve the prediction of the evolution of the latter by incorporating information about the past dynamics of the former. Such relationship is tested by means of bivariate autoregressive models (AR). The value here reported is the value of $1 - \alpha^*$, α^* being the critical level of significance for which the first time series can be considered causal to the second one.
- Mutual information: assesses the quantity of information, measured in *bits*, that two time series share. In other words, it measures how much knowing one of these time series reduces uncertainty about the other.
- Synchronization Likelihood: arguably one of the most popular index for assessing the presence of generalized synchronization, returns a normalized estimate of the dynamical interdependencies between two or more time series (Stam and Van Dijk, 2002). It relies on the detection of simultaneously occurring patterns, even when they are different in the two signals.

As can be seen in **Figure 1** Right, all four metrics present a peak around time 5000, indicating that they all detect this spurious synchronization between the two groups of neurons.

This spurious synchronization is caused by the optimization of the neural code, in which the length of important events is minimized, thus increasing the proportion of spiking neurons when rare events are presented to the system.

The example proposed in **Figure 1** Left is not very ecological as the set of events presented in the two halfs of the considered period only included frequent ([0, 5000]) and infrequent ([5000, 10000]) events. **Figure 1** Center presents a more realistic example, in which the probability of finding rare events is continuously varied between two intermediate values. The resulting time series (gray and light red lines) are highly noisy, while it is still possible to detect some trends. The black solid line represents the evolution of the Pearson's correlation coefficient calculated over a sliding window of size 400. Even in this noisy configuration, it is possible to detect regions in which the correlation between the two time series is strongly increased - similar results were obtained with the three other considered metrics.

3. DISCUSSION

In conclusion, we showed that synchronization can appear when the response of two groups of binary neurons is modulated by the simultaneous appearance of uncommon stimuli, even if both groups do not share information and are not performing a common computation. This is due to the way neural codes are constructed, i.e., to the preference of short symbols, with high spiking rates, representing uncommon events. The present toy model is not intended to mirror actual neural functioning, but rather to draw attention to a possible source of spurious synchronization occurring at the system level of description of neural activity typical of standard neuroimaging techniques. In particular, our results show that even a measure such as the Granger causality can be fooled into signaling causal relationships in the presence of mere coincidences corresponding to no underlying computation. This confirms that claims of causality from (multiple) bivariate time series should always be taken with caution (Pereda et al., 2005), as true causality can only be assessed if the set of two time series contains all possible relevant information and sources of activities for the problem (Granger, 1980), a condition that a neurophysiological experiment can only rarely comply with. Finally, it is important to remark that our model's main suggestion that some of the correlations one would observe in neural activity would not correspond to genuine computation holds true even for resting brain activity, which is operationally defined by the absence of exogenous stimulation. This is explained by the fact that resting brain activity is characterized by unobservable, endogenous activity stemming from numerous simultaneous sources rendering spurious coincidences a plausible occurrence.

ACKNOWLEDGMENTS

Authors acknowledge the usage of the resources, technical expertise and assistance provided by supercomputing facility CRESCO of ENEA in Portici (Italy).

REFERENCES

Breakspear, M. (2002). Nonlinear phase desynchronization in human electroencephalografic data. *Hum. Brain Mapp.* 15, 175–198. doi: 10.1002/hbm.10011

Cormen, T. H., Leiserson, C. E., Rivest, R. L., and Stein, C. (2001). *Introduction to Algorithms*, Cambridge, MA: MIT press.

Fries, P. (2009). Neuronal gamma-band synchronization as a fundamental process in cortical computation. *Annu. Rev. Neurosci.* 32, 209–224. doi: 10.1146/annurev.neuro.051508.135603

Friston, K. J. (1994). Functional and effective connectivity in neuroimaging: a Synthesis. *Hum. Brain Mapp.* 2, 56–78. doi: 10.1002/hbm.460020107

Friston, K. J. (2000). The labile brain. I. Neuronal transients and nonlinear coupling. *Philos. Trans. R. Soc. Lond.* 355, 215–236. doi: 10.1098/rstb.2000.0560

Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* 37, 424–438. doi: 10.2307/1912791

Granger, C. W. J. (1980). Testing for causality: a personal viewpoint. *J. Econ. Dyn. Cont.* 2, 329–352. doi: 10.1016/0165-1889(80)90069-X

Miltner, W. H. R., Braun, C., Arnold, M., Witte, H., and Taub, E. (1999). Coherence of gamma-band EEG activity as a basis of associative learning. *Nature* 397, 434–436. doi: 10.1038/17126

Pereda, E., Quian Quiroga, R., and Bhattacharya J. (2005). Nonlinear multivariate analysis of neurophysiological signals. *Prog. Neurobiol.* 77, 1–37. doi: 10.1016/j.pneurobio.2005.10.003

Rodriguez, E., George, N., Lachaux, J.-P., Martinerie, J., Renault, B., and Varela, F. J. (1999). Perception's shadow: Long distance synchronization of human brain activity. *Nature* 397, 430–433. doi: 10.1038/17120

Singer, W. (1995). "Putative functions of temporal correlations in neocortical processing," in *Large-Scale Neuronal Theories of the Brain*. eds C. Koch and J. L. Davis (Cambridge, MA: MIT Press), 201–237.

Stam, C. J., Nolte, G., and Daffertshofer, A. (2007). Phase lag index: assessment of functional connectivity from multi channel EEG and MEG with diminished bias from common sources. *Hum. Brain Mapp.* 28, 1178–1193. doi: 10.1002/hbm.20346

Stam, C. J., and Van Dijk, B. W. (2002). Synchronization likelihood: an unbiased measure of generalized synchronization in multivariate data sets. *Physica D* 163, 236–251. doi: 10.1016/S0167-2789(01)00386-4

Tononi, G., Edelman, G. M., and Sporns, O. (1998). Complexity and coherency: integrating information in the brain. *Trends Cogn. Sci.* 2, 474–484. doi: 10.1016/S1364-6613(98)01259-5

Varela, F., Lachaux, J. P., Rodriguez, E., and Martinerie, J. (2001). The brainweb: phase synchronization and large-scale integration. *Nat. Rev. Neurosci.* 2, 229–239. doi: 10.1038/35067550

Varela, F. J. (1995). Resonant cell assemblies: a new approach to cognitive functions and neuronal synchrony. *Biol. Res.* 28, 81–95.

Wiener, N. (1956). "The theory of prediction," in *Modern Mathematics for Engineers*, eds E. F. Beckenbach (New York, McGraw-Hill), 165–190.

Received: 01 June 2013; accepted: 21 August 2013; published online: 10 September 2013.

*Citation: Zanin M and Papo D (2013) Efficient neural codes can lead to spurious synchronization. Front. Comput. Neurosci. **7**:125. doi: 10.3389/fncom.2013.00125*

This article was submitted to the journal Frontiers in Computational Neuroscience.

Copyright © 2013 Zanin and Papo. 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) or licensor 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.

frontiers in
COMPUTATIONAL NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 21 August 2013
doi: 10.3389/fncom.2013.00113

Impact of neuronal heterogeneity on correlated colored noise-induced synchronization

Pengcheng Zhou[1,2], Shawn D. Burton[2,3], Nathaniel N. Urban[2,3] and G. Bard Ermentrout[2,4]*

[1] *Program in Neural Computation, Carnegie Mellon University, Pittsburgh, PA, USA*
[2] *Center for the Neural Basis of Cognition, Pittsburgh, PA, USA*
[3] *Department of Biology, Carnegie Mellon University, Pittsburgh, PA, USA*
[4] *Department of Mathematics, University of Pittsburgh, Pittsburgh, PA, USA*

Edited by:
Robert Rosenbaum, University of Pittsburgh, USA

Reviewed by:
Tatjana Tchumatchenko, Max Planck Institute for Brain Research, Germany
Eric Shea-Brown, University of Washington, USA

****Correspondence:***
G. Bard Ermentrout, Department of Mathematics, University of Pittsburgh, 139 University Place, Pittsburgh, PA 15260, USA
e-mail: bard@pitt.edu

Synchronization plays an important role in neural signal processing and transmission. Many hypotheses have been proposed to explain the origin of neural synchronization. In recent years, correlated noise-induced synchronization has received support from many theoretical and experimental studies. However, many of these prior studies have assumed that neurons have identical biophysical properties and that their inputs are well modeled by white noise. In this context, we use colored noise to induce synchronization between oscillators with heterogeneity in both phase-response curves and frequencies. In the low noise limit, we derive novel analytical theory showing that the time constant of colored noise influences correlated noise-induced synchronization and that oscillator heterogeneity can limit synchronization. Surprisingly, however, heterogeneous oscillators may synchronize better than homogeneous oscillators given low input correlations. We also find resonance of oscillator synchronization to colored noise inputs when firing frequencies diverge. Collectively, these results prove robust for both relatively high noise regimes and when applied to biophysically realistic spiking neuron models, and further match experimental recordings from acute brain slices.

Keywords: synchrony, correlation, colored noise, heterogeneity, neural oscillators, phase-response curve

1. INTRODUCTION

Synchronization of neural oscillators is thought to play a critical role in sensory, motor, and cognitive processes (Sanes and Donoghue, 1993; Fries et al., 2001; Wang, 2010). In many networks, synchronization is achieved via direct coupling such as through gap junctions and chemical synapses. However, there are several systems (notably, the mammalian olfactory bulb) where the mode of coupling is less clear and neural synchrony is hypothesized to arise from partially correlated presynaptic inputs (Galán et al., 2006; Marella and Ermentrout, 2010). Indeed, in non-oscillatory networks of neurons, such correlated input is largely responsible for the output correlations of the neurons (de la Rocha et al., 2007). Thus, a natural question is: how do the properties of neurons and networks alter output correlations for a given degree of input correlation? At small input correlations, output and input correlations can be regarded as linearly proportional; this ratio is called the *susceptibility* (Shea-Brown et al., 2008). For example, (de la Rocha et al., 2007) showed that the susceptibility depends on the background firing rate of the neuron. For some model systems, this susceptibility can be computed using linear response theory (which assumes small perturbations around the stationary state).

When neurons fire regularly, they can be regarded as noisy nonlinear oscillators and, as such, there are many mathematical techniques available for their analysis. In particular, the *phase-response curve* (PRC) provides a compact and useful characterization of the responses of a nonlinear oscillator to external perturbations. The PRC describes the shift in timing of, say, an action potential as a function of the timing of the input relative to the last action potential. In several studies, we have described the theoretical relationship between the shape of the PRC and the ability of *identical* neurons to transfer partially synchronized activity (Marella and Ermentrout, 2008; Abouzeid and Ermentrout, 2009). In these studies, the only source of heterogeneity considered between neural oscillators was their unshared (uncorrelated) inputs, which consisted of white noise. Recently, we extended these methods to cases in the low noise limit in which the oscillators were not identical and showed how heterogeneity in intrinsic properties could significantly degrade the output correlation in pairs receiving common inputs (Burton et al., 2012).

In this study, we extend this theory to include colored noise inputs and, further, report some surprising effects of heterogeneity. First, we derive a set of equations for the distribution of phase differences for pairs of heterogeneous oscillators driven by a partially correlated Ornstein-Uhlenbeck (OU) process (low-pass filtered noise). We next show that the theory developed for phase reduced models works well with a conductance-based biophysical model. We then show that, quite surprisingly, at low input correlations, heterogeneity can sometimes produce higher output correlations than the homogeneous case. That is, consider two distinct oscillators, A and B, such that the AA pair has a small susceptibility and the BB pair a larger susceptibility. Then, at low correlations, the susceptibility of the AB pair can sometimes

exceed that of the AA pair. We confirm this somewhat counter-intuitive prediction with recordings from regularly firing mitral cells of the main olfactory bulb. In addition to heterogeneity in response properties, neurons can fire at different frequencies, and such frequency differences can significantly impact correlated-noise induced synchronization (Markowitz et al., 2008; Burton et al., 2012). Here, we find that for some frequency differences between oscillators, there is an optimal time scale of correlated noise that will maximally synchronize the oscillators. We do not see this effect when the oscillators have the same frequency.

2. MATERIALS AND METHODS

2.1. PHASE REDUCTION MODEL

In Appendix, we provide a brief overview of how to reduce a general weakly perturbed limit cycle to a single differential equation for the phase of the cycle. If we assume that the original limit cycle represents repetitive firing of a single compartment neuron model that is driven by a noisy current, $I(t)$, then we obtain:

$$\frac{d\theta}{dt} = 1 + \epsilon\Delta(\theta)I(t)/C_m \tag{1}$$

where C_m is the membrane capacitance, θ is the phase (or, typically, the time since the last spike), and $\Delta(\theta)$ is the PRC of the neuron. The PRC describes the phase-dependent shift in the spike times of an oscillator receiving small perturbations. It is readily measured in neurons and other biological oscillators (Torben-Nielsen et al., 2010) and provides a compact representation of the effects of stimuli on the timing of action potentials. $\Delta(\theta)$ has dimensions of milliseconds per millivolt; that is, the shift in timing of the next action potential per millivolt perturbation of the potential. Mathematically, for a given model, $\Delta(\theta)$ is found by solving a certain differential equation (see Appendix). It is a periodic function of phase and, with no loss in generality, we can normalize the period to be 2π for simplicity.

2.2. STATIONARY DENSITY

Given the reduced model (Equation 1), we can now turn to the main question at hand, which is: how do oscillating heterogeneous neurons transfer correlations? We will consider two types of heterogeneity: differences in the PRC shapes and differences in natural frequencies. We drive the oscillators with correlated filtered noise. After reduction to phase variables, we obtain:

$$\theta_1' = 1 + \epsilon\Delta_1(\theta_1)x \tag{2}$$

$$\theta_2' = 1 + \epsilon\Delta_2(\theta_2)y + \epsilon^2\omega \tag{3}$$

$$x' = -x/\tau + \xi_x/\sqrt{\tau} \tag{4}$$

$$y' = -y/\tau + \xi_y/\sqrt{\tau} \tag{5}$$

θ_1 and θ_2 are the phases of two oscillators, and $\Delta_1(\theta)$ and $\Delta_2(\theta)$ are PRCs of the two oscillators. Without loss of generality, we set the natural frequency of one oscillator to 1. The parameter ω then determines the magnitude of the difference in natural frequencies between the two oscillators. $\epsilon \ll 1$, thus the noise is weak and the frequency difference is small. The processes x and y are generated by an OU process with the same time constant τ. ξ_x and ξ_y are two correlated white noise processes, i.e., $\langle\xi_x(t)\xi_x(t')\rangle = \delta(t-t')$, $\langle\xi_y(t)\xi_y(t')\rangle = \delta(t-t')$, $\langle\xi_x(t)\xi_y(t')\rangle = c\delta(t-t')$, where c is the degree of correlation.

We remark that the allowable frequency difference is $O(\epsilon^2)$, which seems considerably smaller than the magnitude of the noise, which is ϵ. However, as the noise has zero mean, what matters is the variance of the noise, which has magnitude ϵ^2. Thus, the scales of both the frequency difference and the synchronizing inputs (correlations in the noise) are similar. If the frequency differences are larger, then no synchronization is possible.

Our goal is to compute the stationary distribution of the phase difference between two neurons since this will enable us to compute various measures of correlation and synchrony. Thus, some variable substitution will be helpful: $\theta = \theta_1$, $\phi = \theta_2 - \theta_1$. Therefore, ϕ is the phase difference between the two oscillators. With this change of variables, the equations are:

$$\theta' = 1 + \epsilon\Delta_1(\theta)x \tag{6}$$

$$\phi' = \epsilon[\Delta_2(\theta+\phi)y - \Delta_1(\theta)x] + \epsilon^2\omega \tag{7}$$

and x, y are as above. Let $\rho(x, y, \theta, \phi, t)$ represent the probability density function at time t:

$$Pr(X(t) \in (x, x+dx), Y(t) \in (y, y+dy), \Theta(t) \in (\theta, \theta+d\theta),$$
$$\Phi(t) \in (\phi, \phi+d\phi)) = \rho(x, y, \theta, \phi)dxdyd\theta d\phi \tag{8}$$

We denote the stationary density (long-time behavior as $t \to \infty$) as $\rho_{ss}(x, y, \theta, \phi)$.

Our goal is to compute the probability density of the phase difference between the two oscillators, $R(\phi)$, which is:

$$R(\phi) := \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{0}^{2\pi} \rho_{ss}(x, y, \theta, \phi)\, dxdyd\theta \tag{9}$$

If the oscillators are perfectly synchronized, then $R(\phi)$ will be a delta function centered at $\phi = 0$. If the oscillators are completely independent, then $R(\phi) = 1/(2\pi)$. In Appendix, we show that $R(\phi)$ satisfies a simple first order boundary value problem (BVP). We present the exact equation for this in Results.

2.3. ORDER PARAMETER

Once we get the distribution of phase differences, $R(\phi)$, we need a number to estimate the synchronization, which means the sharpness of this distribution. In this study, we use an order parameter (OP) to do this. We define:

$$\mathrm{OP} = \sqrt{C^2 + S^2} \tag{10}$$

$$C = \int_0^{2\pi} R(\phi)\cos(\phi)d\phi$$

$$S = \int_0^{2\pi} R(\phi)\sin(\phi)d\phi$$

$$\theta = \mathrm{atan2}(C, S)$$

OP is a representation of sharpness and θ is the estimation of the peak position. For certain types of heterogeneity, $R(\phi)$ is peaked at $\phi = 0$; in this case, we can show that the cross correlation of the spike times is $(R(0) - 1/(2\pi))/(2\pi)$ (Burton et al., 2012). However, OP provides a better global measure of the synchrony and is not dependent on the peak being centered at 0; we will therefore use OP in our current results.

2.4. MORRIS-LECAR MODEL

The Morris-Lecar (ML) model (Rinzel and Ermentrout, 1989) is a simplified two-dimensional system membrane model that we use to compare the phase models with a full biophysical model:

$$C\frac{dV_1}{dt} = I_1 - g_L(V_1 - V_L) - g_K w_1(V_1 - V_K) - g_{Ca} m_\infty(V_1)(V_1 - V_{Ca}) + \sigma x \quad (11)$$

$$\frac{dw_1}{dt} = \phi\frac{w_\infty(V_1) - w_1}{\tau_w(V_1)} \quad (12)$$

$$C\frac{dV_2}{dt} = I_2 - g_L(V_2 - V_L) - g_K w_2(V_2 - V_K) - g_{Ca} m_\infty(V_2)(V_2 - V_{Ca}) + \sigma x \quad (13)$$

$$\frac{dw_2}{dt} = \phi\frac{w_\infty(V_2) - w_2}{\tau_w(V_2)} \quad (14)$$

$$x' = -x/\tau + \xi_x/\sqrt{\tau} \quad (15)$$

$$y' = -y/\tau + \xi_y/\sqrt{\tau} \quad (16)$$

with $\langle \xi_1(t),\ \xi_1(t')\rangle = \delta(t - t')$, $\langle \xi_2(t),\ \xi_2(t')\rangle = \delta(t - t')$, and $\langle \xi_1(t), \xi_2(t')\rangle = c\delta(t - t'), c \in [0, 1]$. The auxiliary functions are:

$$m_\infty(V) = 0.5 \cdot (1 + \tanh((V - V_a)/V_b)) \quad (17)$$

$$w_\infty(V) = 0.5 \cdot (1 + \tanh((V - V_c)/V_d)) \quad (18)$$

$$\tau_w(V) = \frac{1}{\cosh((V - V_c)/(2V_d))} \quad (19)$$

The parameters used in this paper are: $V_K = -84\,mV$, $V_L = -60\,mV$, $V_{Ca} = 120\,mV$, $g_K = 8\frac{mS}{cm^2}$, $g_L = 2\frac{mS}{cm^2}$, $g_{Ca} = 4\frac{mS}{cm^2}$, $C = 20\frac{\mu F}{cm^2}$, $V_a = -1.2\,mV$, $V_b = 18\,mV$, $V_c = 2\,mV$, and $V_d = 30\,mV$. I_1, I_2 and ϕ_1, ϕ_2 vary for each figure.

To get the phase from the noisy voltage signal generated by the ML model, we first apply the Hilbert transform (HT) to $V(t)$ which allows us to get a phase. However, the phase is not uniform as it is not a temporal phase. We then map the HT phase to a temporal phase on the noise-free limit cycle which gives a uniform phase-distribution as required by the theory. This allows us to estimate $R(\phi)$ for the biophysical model, where ϕ here is the phase difference between two ML model neurons that are driven with partially correlated noise.

In some of the figures, we simulate the phase-reduced dynamics for the ML model. To do this, we must compute the infinitesimal PRC, $\Delta_{ML}(\theta)$. As described in Appendix, the PRC for the model is the voltage component of the solution to the adjoint equation (Equation 32). The software package XPP (Ermentrout, 2002) includes an algorithm for computing the adjoint solution for an exponentially stable limit cycle, so we simply compute various limit cycles (say with very different parameters but similar periods) and then compute $\Delta_{ML}(\theta)$ for those specific parameters. We save the result as a lookup table and then numerically solve the phase equation.

2.5. NUMERICS

To get solutions to the stochastic phase and membrane equations, we use the Euler-Murayama method. We solve the BVP for the stationary phase difference density using a custom BVP solver written in MATLAB. All codes are available by request.

3. RESULTS

3.1. APPROXIMATION OF THE PHASE DIFFERENCE DENSITY

Oscillators driven with a correlated fluctuating signal will exhibit a degree of synchronization that depends on the size of the signal, the strength of correlation, and the similarity of the two oscillators. Thus, for example, identical oscillators driven by small enough identical white noise will synchronize perfectly (Pikovsky et al., 1997; Teramae and Tanaka, 2004). The rate at which these identical oscillators synchronize depends on the properties of the noise - in particular, its autocorrelation (Nakao et al., 2007; Goldobin et al., 2010). In general, and especially in biological systems, there will be a great deal of heterogeneity in any pair of oscillators. For example, for neurons, there is always some source of independent noise so that the input correlation is always less than 1. The neurons may also be firing at slightly different frequencies. Finally, even if the neurons are adjusted to fire at the same frequency, their distribution of ion channels can be very different and, thus, their response to correlated signals can be quite different (Burton et al., 2012). If the fluctuating inputs are sufficiently small, then any stable limit cycle oscillator can be reduced to a so-called phase model where the dynamics are characterized by a single variable, the phase, such that the firing is considered to occur at a phase of 0 and the time between spikes is mapped onto an angle between zero and 2π. Here, we consider driven pairs of heterogeneous oscillators that receive partially correlated filtered noise. As our main examples come from neuroscience, we assume that the external inputs are implemented as currents, in which case the phase model for the pair of neural oscillators has the form:

$$\theta_1' = 1 + \epsilon\Delta_1(\theta_1)x(t)$$

$$\theta_2' = 1 + \epsilon^2\omega + \epsilon\Delta_2(\theta_2)y(t)$$

where $x(t)$ and $y(t)$ are OU processes with the same time constant, τ, and with correlation c; $\Delta_{1,2}(\theta)$ are the PRCs for the two oscillators; ϵ is a small positive parameter (characterizing the magnitude of the fluctuations); and ω accounts for the frequency difference in the unperturbed oscillators (see Materials and Methods, Equations 2–5). We are primarily interested in the distribution of the phase difference, $\phi := \theta_2 - \theta_1$. In the Appendix (Equation 62), we show that $R(\phi)$, the

probability density function for the phase difference, satisfies a simple BVP:

$$\frac{d}{d\phi}\{[c \cdot g(\phi) - C_1]R(\phi)\} + (4\pi\omega - C_2)R(\phi) = K$$

$$R(\phi) = R(\phi + 2\pi)$$

$$g(\phi) = g(\phi + 2\pi)$$

$$\int_{-\pi}^{\pi} R(\phi)d\phi = 1$$

$$K = 2\omega - \frac{C_2}{2\pi}$$

The $2\pi-$periodic function $g(\phi)$ and the constants, $C_{1,2}$, depend in a complicated way on the forms of the PRCs and the time constant of the noise, τ (see Appendix). However, all quantities can be found by integrating elementary functions. If the oscillators have the same PRC, then $C_2 = 0$ and $g(\phi)$ is even symmetric. If the oscillators have the same frequency, then $\omega = 0$. When both C_2 and ω vanish, we can immediately solve the BVP, yielding $R(\phi) = N/(C_1 - cg(\phi))$, where N is a normalization constant so that the integral is 1. This is the result found in Marella and Ermentrout (2008) for white noise, but is clearly also true for colored noise. When the oscillators are identical and there is no difference in frequencies, the phase difference density is symmetric and always peaks at 0. However, when $\omega - C_2$ is nonzero, the peak of the phase difference density will generally be offset. We note that ϵ does not appear in the expression for $R(\phi)$, which says that the phase difference density is, to a first approximation, independent of the amplitude of the noise. **Figure 1** shows typical results comparing the perturbation calculation and the simulation of the Langevin equation. In **Figure 1A**, at fairly high noise $\epsilon = 1$, there is some distortion at the peak of the distribution, but as predicted from the theory, the distribution magnitude is largely independent of the magnitude of the fluctuations. **Figure 1B** shows a similar simulation, but the correlation of noise is lower ($c = 0.5$ vs. $c = 0.8$), the noise is faster ($\tau = 0.25$ vs. $\tau = 1$), and the PRCs are identical. In this case, even the higher noise simulations match the theory. We once again emphasize that the perturbation expansion requires a small value of ϵ, but clearly, the simulations show that ϵ can be nearly 1 and still yield good agreement.

We note that the density of the phase differences can be related to more conventional measures of correlation. In Burton et al. (2012), we showed that the spike time cross-correlation (CC) between a pair of weakly noisy oscillators is:

$$\mathrm{CC}(\tau) = \frac{1}{2\pi}\left[R(-\tau) - \frac{1}{2\pi}\right] \tag{20}$$

For example, if the oscillators are asynchronous, then they have a uniform phase difference density and the cross-correlation will be 0. This calculation confirms ones intuition that *different* neurons that receive correlated noise will have spike time cross-correlations that peak off-center.

Figure 2 shows that we can apply the theory through two levels of simplification. The ML system is a simple, biophysically realistic model for a spiking neuron (Rinzel and Ermentrout, 1989). With different choices of parameters, the onset to oscillatory behavior can be either through a Hopf bifurcation (HB) or a saddle-node on an invariant circle (SNIC) bifurcation. The PRCs that result from these distinct bifurcations are often quite different (Brown et al., 2004; Izhikevich, 2007) and thus have quite different synchronization properties. In **Figure 2**, we tune the ML model so that each cell has the same frequency but the parameters are quite different and so the PRCs are different (see **Figure 2B**).

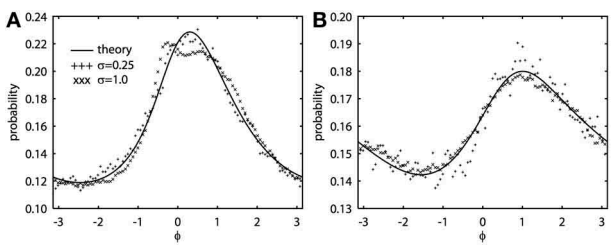


FIGURE 1 | Novel analytical theory of correlated colored noise-induced synchronization of heterogeneous oscillators matches Monte Carlo simulations for low to moderate levels of noise. Stationary phase difference density is shown as computed from the solution of the BVP and through Monte Carlo simulation from $t = 1000$ to $t = 201000$ in steps of 0.05. Monte Carlo data binned into 100 bins between $-\pi$ and π. There is a frequency difference of $\epsilon^2/2$ where ϵ is the magnitude of the noise. Here $\Delta_j(\theta_j) = \sin(a_j) - \sin(\theta_j + a_j) + b_j \sin(2\theta_j)$, where $j = 1, 2$ for two oscillators. **(A)** $\tau = 1$, $a_1 = 0.1$, $a_2 = 0.6$, $b_1 = 0.32$, $b_2 = 0.3$, and $c = 0.8$. **(B)** $\tau = 0.25$, $a_1 = a_2 = 0.5$, $b_1 = b_2 = 0.3$, and $c = 0.5$.

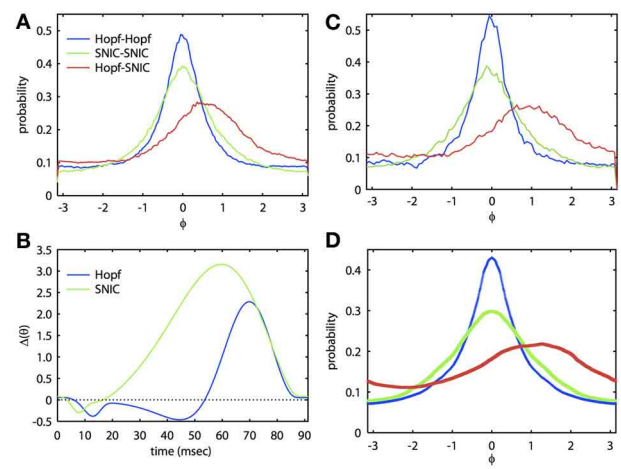


FIGURE 2 | Analytical theory accurately predicts synchronization of biophysically realistic spiking neuron models. (A) Invariant phase difference density computed from the reconstructed phase of two ML model neurons receiving partially correlated colored noise (period is 91.25 ms, $\tau = 5$ ms, $c = 0.8$). Three cases are illustrated with either identical (homogeneous) or mixed (heterogeneous) PRCs. The "Hopf" case corresponds to a set of parameters where the oscillatory activity arises via a HB and the "SNIC" case through a SNIC bifurcation (Rinzel and Ermentrout, 1989); see Appendix for parameters. **(B)** PRCs for the two cases. **(C)** Same as **(A)**, but using simulations of the phase reduced equations. **(D)** Solutions to the BVP using the PRCs from **(B)**.

In **Figure 2C**, we show the results of a Monte Carlo simulation in which the biophysical model is driven by correlated noise. Phase is reconstructed from the voltage traces using a Hilbert transform and from these, we obtain phase difference histograms. In this figure, the correlation c is 0.8, $\tau = 5$ ms, and the natural period of the oscillation is 91.25 ms. For the same degree of correlation, two HB oscillators are much better at synchronizing than are two SNIC oscillators. This result is consistent with the theory developed in Marella and Ermentrout (2008) for white noise and also for spike time correlations over fast timescales (i.e., spike synchrony) (Barreiro et al., 2010). At this high correlation, the heterogeneous HB-SNIC pair shows greatly reduced synchrony from either of the homogeneous cases and a shift in the peak *even though there is no frequency difference.* **Figure 2B** shows the two PRCs that were determined using the adjoint method. We then used these PRCs to compute the invariant densities for the corresponding phase reduced models. The invariant density is a function that describes the distribution of phase differences of the two neurons over some time interval consisting of many cycles. Thus, the peak of the invariant density indicates the most likely phase difference, and a large peak at zero phase difference would indicate that the two neurons are well synchronized. Comparison between **Figures 2A,C** shows excellent agreement. Finally, we substituted the numerically computed PRCs into our BVP and computed the invariant density. The result of this calculation is shown in **Figure 2D**. There are small differences in the amplitude, but the shapes and the shift of the densities in the heterogeneous case are almost identical. Thus, through two levels of reduction (first, from the full model to the phase model, and second, from the Langevin phase model to the approximate invariant density), we see that our analytical method works very well at estimating the invariant density of phase differences between neural oscillators.

3.2. PRC HETEROGENEITY

Our approximation of the invariant density, while requiring that we solve a BVP, allows us to explore the effects of heterogeneity much faster than simulating the appropriate Monte Carlo system. Thus, we will use this method to explore the effects of PRC heterogeneity, frequency differences, and the color of the noise on the ability of oscillators to synchronize. One simple global measure of synchrony/correlation for systems whose natural dynamics are periodic is the circular variance, $\sigma_{circle} = 1 - \mathrm{OP}$, where we define an order parameter (OP) (see Materials and Methods, Equation 10):

$$\mathrm{OP} = \left[\left(\int_{-\pi}^{\pi} R(\phi) \cos\phi \, d\phi \right)^2 + \left(\int_{-\pi}^{\pi} R(\phi) \sin\phi \, d\phi \right)^2 \right]^{\frac{1}{2}}$$

For a flat distribution, OP $= 0$ ($\sigma_{\mathrm{circle}} = 1$) and for a delta function distribution, OP $= 1$ ($\sigma_{\mathrm{circle}} = 0$). The OP is a commonly used measure for the degree of synchronization between two oscillators (Kuramoto, 2003).

In general, one expects that the synchrony between two oscillators forced with correlated noise would be greatest if the oscillators are homogeneous. Certainly, if the inputs are identical (i.e., no independent or unshared noise), then identical oscillators will synchronize perfectly, while heterogeneous oscillators will not synchronize perfectly. That is, the phase density will not be a delta function. [See Burton et al. (2012) for a proof]. However, surprisingly, at low input correlations, it is possible for a heterogeneous pair of oscillators to produce greater synchrony than one (but not both) of the respective homogeneous pairs of oscillators. **Figure 3** illustrates the behavior of two separate homogeneous pairs of oscillators (blue and green lines, respectively) as the input correlation varies from 0 to 1. A third, heterogeneous pair comprised of an oscillator from each homogeneous pair is shown in red. **Figure 3A** shows the two different PRCs; pairs of oscillators with the green PRC ("PRC 1-PRC 1") produce weaker synchrony than pairs of oscillators with the blue PRC ("PRC 2-PRC 2"). This is demonstrated in **Figure 3B**, where the correlation is set to 0.8. Note that the phase difference density for PRC 2-PRC 2 pair is more peaked than that for PRC 1-PRC 1 pair, while both densities are more peaked than the heterogeneous "PRC 1-PRC 2" pair. As noted above, the peak of the heterogeneous pair is not at the origin but rather, is shifted to the left. In order to get a global measure of synchrony, we plot OP as a function of the input correlation (**Figure 3C**). As $c \to 1$, both homogeneous pairs approach OP $= 1$ (i.e., perfect synchrony) while the heterogeneous pair never exceeds OP $= 0.4$. However, at low correlations, the heterogeneous pair can actually synchronize better than the PRC 1-PRC 1 pair (compare red to green lines in inset). That is, in the presence of low correlations, a "good synchronizer" paired with a "bad synchronizer" performs better than the homogeneous pair of bad synchronizers. This effect is not just due

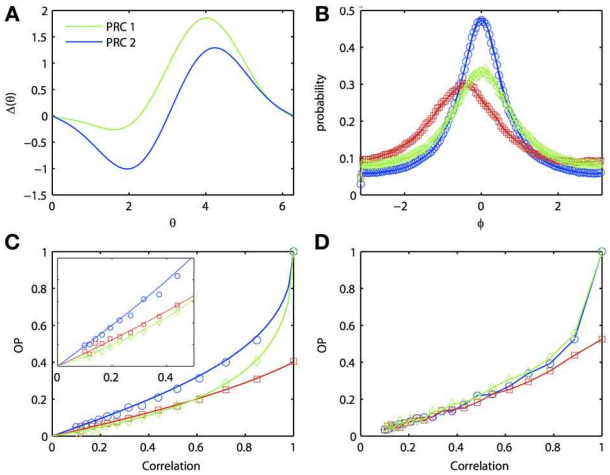


FIGURE 3 | Oscillator heterogeneity can enhance correlated noise-induced synchronization at low input correlations. (A) Two PRCs with the form $\Delta_j(\theta_j) = \sin(a_j) - \sin(a_j + \theta_j) + b_j \sin(2\theta_j)$, $a_1 = 0.1$, $b_1 = 0.32$, $a_2 = 0.6$, and $b_6 = 0.3$. **(B)** $R(\phi)$ with different combinations of PRCs. Blue: PRC 2-PRC 2; red: PRC 1-PRC 2; green: PRC 1-PRC 1. Solid lines are theoretical predictions from the solution to the BVP and open symbols are Monte Carlo simulation results (same notation applies in following figures). Parameters used: $\tau = 1$ and $c = 0.8$. **(C)** Synchronization as input correlation varies from 0 to 1; inset shows magnification when $c < 0.5$. **(D)** Same as **(C)**, but using the ML model. Parameters used: $I_1 = 110$, $\phi_1 = 0.04616$, $I_2 = 120$, and $\phi_2 = 0.04$.

to our approximate expansion as the Monte Carlo simulations show the same phenomenon. **Figure 3D** further hints that we can also see the effect in the full ML model, although the results are not as clear.

3.2.1. Experimental evidence

Could this subtle difference in the ability of neural oscillators to transfer correlation be seen in experiments? To answer this, we re-examined data from a previous study (Burton et al., 2012). Mitral cells from the mouse main olfactory bulb were injected with constant current overlaid with frozen noise to evoke noisy periodic firing. PRCs were then experimentally estimated using our previously described method using the spike-triggered average (Ermentrout et al., 2007). [Complete methods are provided in Burton et al. (2012)]. In this dataset, we found several examples where injecting partially correlated noise produced greater synchrony between two different mitral cells firing at the same rate than for one of the mitral cells across different trials (experimentally simulating a homogeneous pair of mitral cells). **Figure 4** illustrates an example. In **Figure 4A**, we show the voltage traces (top) of two mitral cells receiving correlated inputs, and the spike times (middle) and phase (bottom) as determined by a simple linear interpolation between spikes. **Figure 4B** shows the PRCs from each of these two cells along with their fit to the exponential-sine PRC model (see Appendix, Equation 64). In **Figure 4C**, we show the phase difference density as constructed from the linear phase interpolation of the two cells. In this example, the currents delivered through the electrodes are perfectly correlated. However,

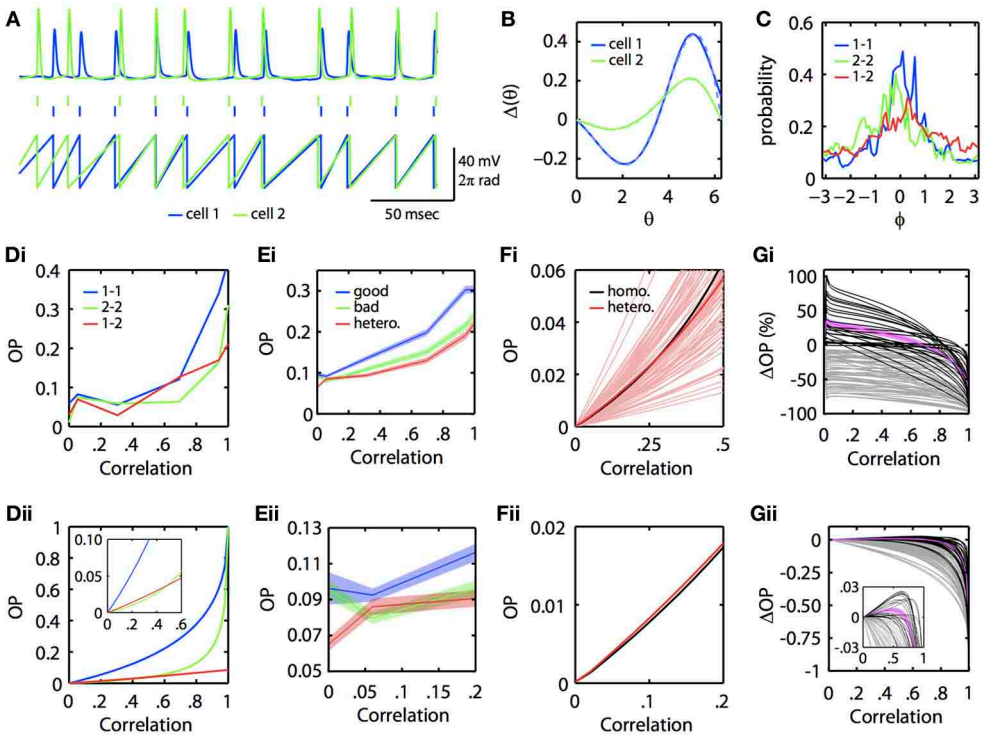


FIGURE 4 | Physiological neuronal heterogeneity can enhance correlated noise-induced synchronization at low input correlations. (A) Example linear interpolation of phase between recorded spike times of two mitral cells injected with perfectly correlated colored noise. Top: experimentally recorded membrane potentials. Middle: raster plot of spike times. Bottom: phase. **(B)** Experimentally estimated PRCs for the two cells shown in **(A)**. Dashed lines are fits of the exponential-sine PRC model to the estimated PRCs. **(C)** Phase difference densities of the two cells during injection of perfectly correlated currents. Densities were calculated from pairs of 5 sec recordings. Blue and green curves show densities for homogeneous pairs of cell 1 and cell 2, respectively. The red curve shows the density for the heterogeneous pair of cell 1 with cell 2. **(Di)** Experimental and **(Dii)** theoretical OP vs. input correlation for homogeneous and heterogeneous pairs of the two mitral cells. Theoretical curves calculated by solving the BVP with the model PRC fits and $\tau = 5$. [Note that the same results were obtained in separate calculations for $\tau = 3$, the time scale of the noise used in Burton et al. (2012)]. **(Ei–Eii)** Mean OP (±SEM) vs. input correlation across 85 pairs of mitral cells (formed from 27 separate mitral cell recordings described in Burton et al. (2012)). For each pair, the cell with the greatest area under its homogeneous OP vs. correlation curve was classified as the "good synchronizer" of the pair. **(Fi)** Theoretical OP vs. input correlation (with $\tau = 3$) for each of the 85 heterogeneous pairs from **(E)** (light red lines), plotted against the theoretical OP vs. input correlation of a homogeneous pair formed from the average mitral cell PRC. Note that many (but not all) of the heterogeneous pairs exceed the homogeneous pair in the low correlation range shown. On average (dark red line), physiological heterogeneity enhances synchrony for input correlations up to ∼0.27. **(Fii)** Magnification of the homogeneous and average heterogeneous lines from **Fi** for low input correlations. **(Gi)** Percent and **(Gii)** absolute change in theoretical OP for heterogeneous vs. homogeneous bad pairs of mitral cells. Black lines plot OP changes for pairs in which heterogeneity increased synchrony at low input correlations; magenta line plots mean OP enhancement (±SEM) for these pairs. Grey lines plot OP changes for pairs in which heterogeneity did not increase synchrony. Note that heterogeneity mediates the greatest percent increase in OP at low (<0.1) input correlations, similar to experimental results shown in **(E)**. Inset shows magnification when $|\Delta OP| < 0.03$.

unlike the simulations, the neurons themselves are intrinsically noisy, so there is a substantial component of "private" noise. Nevertheless, one can see that cell 1 (blue) synchronizes better across trials than does cell 2 (green) across trials. **Figures 4Di,Dii** show the OP as reconstructed from the experimental data and as obtained by using the computed PRCs, respectively. This shows that at low correlations, the heterogeneous pair ("1–2") can synchronize better than the "2–2" homogeneous pair (but not the "1–1" homogeneous pair). The inset in 4Dii magnifies the low c region.

Are the results presented in **Figures 4A–D** for a single pair of mitral cells consistent across a larger population of mitral cells? To answer this, we examined recordings from 27 regularly firing mitral cells, from which we were able to form 85 different pairs of mitral cells with highly similar ($\leq$5 Hz difference) firing rates. For each pair of mitral cells, we computed the OP across varying input correlations for both homogeneous combinations and the heterogeneous combination. We automatically classified the mitral cell with the greatest homogeneous OP across all levels of input correlation as the "good synchronizer" of the mitral cell pair. **Figure 4E** shows the mean OP vs. correlation across the 85 good, bad, and heterogeneous mitral cell pairs. Note that, even with this relatively insensitive classification of good vs. bad synchronizers, there is a region at low input correlations where, on average, heterogeneous pairs synchronize better than the bad homogeneous pairs. This phenomenon is seen more clearly when we use the experimentally estimated PRCs and the BVP to compute the OP vs. input correlation. **Figure 4Fi** plots OP vs. c for all heterogeneous pairs (light red lines), the mean of the heterogeneous pairs (dark red line), and the OP for a single homogeneous pair whose PRC is the mean of all the PRCs (black line). For many cases (but not all), heterogeneity increases the OP above that achieved by a uniform population of neural oscillators with the mean PRC. **Figure 4Fii** magnifies the mean OP vs. c curves at low correlation; the red curve is clearly higher than the black curve.

We then quantified the degree to which physiological levels of heterogeneity [as experimentally measured in mitral cells (Burton et al., 2012)] can enhance synchrony between neural oscillators. Using the BVP and our experimentally estimated mitral cell PRCs, we calculated the percent and absolute change in OP for all 85 heterogenous vs. homogeneous bad mitral cell pairs. That is, for the example pair in **Figure 4Dii**, we subtracted the green from the red line to calculate the absolute change in OP, and divided this difference by the green line to calculate the percent change in OP. **Figures 4Gi,Gii** plot the results of this analysis for all 85 pairs. In 26 of these pairs (plotted in black), heterogeneity enhanced synchrony at low input correlations, with a mean increase in OP (plotted in magenta; $\pm$SEM) of up to 36%. Thus, in relative terms, physiological levels of heterogeneity can significantly enhance correlated noise-induced synchrony at low input correlations. While this relative enhancement in synchrony corresponds to an admittedly low absolute increase in OP of up to 0.01 on average (**Figure 4Gii**), we nevertheless expect this phenomenon to significantly contribute to patterns of oscillatory synchrony in the olfactory bulb and potentially other brain regions (see Discussion).

3.2.2. Good vs. bad synchronizers

When is a neuron a good vs. bad synchronizer? Here, the BVP is much simpler since we just have to compare homogeneous pairs. In this case, the probability density function can be written as:

$$R(\phi) = \frac{N}{1 - c\frac{g(\phi)}{g(0)}} \tag{21}$$

where N is a normalization and $g(\phi)$ is defined above by setting $n = m$. For low values of c, we get:

$$R(\phi) \approx N\left[1 + c\frac{g(\phi)}{g(0)}\right] \tag{22}$$

and integrating, we can find N:

$$\frac{1}{N} \approx 2\pi\left[1 + c\frac{1}{2\pi}\int_0^{2\pi} g(\phi)/g(0)\, d\phi\right] \tag{23}$$

Since the two neurons are identical, the peak of $R(\phi)$ occurs at $\phi = 0$ and, so, we can estimate the zero lag cross-correlation as $[R(0) - 1/(2\pi)]/(2\pi)$. Using the approximations above, we see that:

$$\mathrm{CC} \approx \frac{c}{2\pi}\left(1 - \frac{1}{2\pi}\int_0^{2\pi}\frac{g(\phi)}{g(0)}\, d\phi\right) := cS \tag{24}$$

That is, the cross-correlation is linearly proportional to the input correlation (for small c) and this factor [called the susceptibility (de la Rocha et al., 2007)], is a simple function of $g(\phi)$. We can maximize S if we can make the integral as small as possible. Note that $g(\phi)$ is periodic, and the integral over a period is proportional to the constant Fourier coefficient. Recall that $g(\phi)$ is a low-pass filtered version of $h(\phi)$, which is the autocorrelation function of the PRC. Thus, $h(0)$ is positive and so is $g(0)$. The integral of $g(\phi)$ is proportional to the integral of $h(\phi)$, which is just $2\pi a_0^2$ where a_0 is the DC component of the PRC. Hence, we can minimize the integral and maximize the correlation transfer (susceptibility) if we mimimize the DC component of the PRC. This fact generalizes the conclusions in Marella and Ermentrout (2008) and Abouzeid and Ermentrout (2009) that state that more sinusoidal PRCs are the best synchronizers. For example, with a PRC of the form $(\sin(a) - \sin(x + a))\exp(C(x - 2\pi))$, we obtain the best synchrony when $a = -\arctan C$.

Can we determine when a pair of oscillators will have the property that a good-bad heterogeneous pair is better than a bad-bad homogeneous pair? Since the effect is only seen at low correlations, this suggests a perturbation expansion for small c. We write $R(\phi) = \sum c^n R_n(\phi)$ and find that R_0 is constant and so to order 1, $R(\phi) = R_0 + cR_1(\phi)$. From this, we can compute OP, $\mathrm{OP} = c\int_0^{2\pi}\cos\phi R_1(\phi)\, d\phi$. The formulas for this are not terribly useful, but we can illustrate the results with a simple example. Let $\Delta_j(\phi) = \sin(a_j) - \sin(\phi + a_j)$, where $j = 1, 2$ for two oscillators. Then:

$$\mathrm{OP}_{jk} = c\frac{K}{1 + \left(\tau^2 + 1\right)\left[\sin^2 a_j + \sin^2 a_k\right]} \tag{25}$$

Thus, for $0 \leq a_1 < a_2 \leq \pi/2$, we always have $OP_{11} > OP_{12} > OP_{22}$ for all τ and sufficiently small values of c. This provides a simple and surprising illustration that heterogeneity will improve synchrony at low correlations for very simple PRCs. We remark, however, that this phenomenon does not always hold. Pairs of PRCs can be found such that OP is always bigger for both of the homogeneous oscillator pairs than for the heterogeneous oscillator pair, as can be seen from **Figures 4F,G**.

3.2.3. PRC heterogeneity tunes the sharpness and peak position of the phase difference density

If two neurons are identical but driven with partially correlated noise, then the peak of the phase difference density will be centered at $\phi = 0$, which means that the two oscillators will tend to have the same phase. However, with heterogeneity, the peak will shift depending on the degree of heterogeneity, just as two coupled oscillators will shift if they have different intrinsic frequencies. **Figure 5** shows how the peak of the phase difference density is shifted by oscillator heterogeneity. Using the two term double sinusoidal form PRC (Equation 63), we keep PRC 1 constant ($a_1 = 0.1, b_1 = 0.32$) as we vary PRC 2 ($b_2 = 0.3$ is constant and a_2 varies from $-\pi$ to π). From the results shown in **Figure 5**, we can conclude that heterogeneity can tune oscillator synchronization in both the sharpness and peak position of the phase difference density, which might be useful in neural signal coding. We also note that OP is minimized when the peak is at $\pm\pi/2$ and that "changing the sign" of the PRC (e.g., setting $a_2 = \pi$) shifts the peak but has very little effect on the OP.

3.3. FREQUENCY DIFFERENCES HIGHLY LIMIT SYNCHRONIZATION

In the above results, we assume that all oscillators have the same natural frequency, which means $\omega = 0$. This is a somewhat unreasonable assumption for real neurons. Thus, we now study how synchronization is dependent on the frequency differences between oscillators. **Figure 6** shows the effects of frequency differences on a pair of oscillators that have different PRCs (of the two term double sinusoidal form, Equation 63) and are driven by partially correlated noise. With no frequency difference, the heterogeneity in oscillator PRCs yields a shift in the peak position (**Figure 5**), consistent with previous measurements of synchrony between irregularly firing neurons (Tchumatchenko et al., 2010). This means that, if frequency differences can shift the peak in the opposite direction [e.g., see Figure 1C of Burton et al. (2012)], then changes in frequency could "cancel" the effects of PRC heterogeneity so that the peak of the phase difference density is at 0. This cancellation can be seen in **Figure 6** near $\omega = 0.2$. However, this cancellation comes at a loss to precision, as seen by the decrease in OP. While not shown, we remark that the drop in OP is symmetric about $\omega = 0$; thus, a negative frequency difference will not result in a larger OP. While it remains to be proven, we conjecture that the OP is always maximal when there is no frequency difference. This differs from the case that we looked at in the previous section where heterogeneity (in PRCs) can sometimes lead to a larger OP than homogeneity.

3.4. CORRELATED NOISE-INDUCED SYNCHRONIZATION IS DEPENDENT ON THE TIME CONSTANT OF THE NOISE

Because of the natural decay times of synapses, broadband inputs into neurons have some temporal correlations. Thus, we now explore how the temporal properties of noise interact with heterogeneities in the PRCs. **Figure 7** shows that synchronization decreases monotonically as τ increases for $\omega = 0$, while there exists an optimal value of τ that achieves the greatest synchronization for $\omega = 0.5$. This means synchronization of two oscillators with different frequencies (i.e., $\omega \neq 0$) can have a resonance with τ. Furthermore, as seen in **Figures 7B,D** the peak of the phase difference density depends on τ only when there is a frequency difference between the two oscillators. In **Figure 8**, we explore this resonance in more detail where $R(\phi)$ is plotted as τ varies. The left panels (showing the solution to the BVP and the results of Monte Carlo simulation) show that when $\omega = 0$, the peak position of $R(\phi)$ is largely unchanged and the magnitude decreases monotonically with τ. There is a sharp drop off in $R(\phi)$ at $\tau \approx 2$. A different result emerges in the right panels, where a frequency difference exists ($\omega = 0.5$). At low and high values of τ, $R(\phi)$ is almost flat with a distinct resonance when $\tau \approx 1$. We see the same resonance in the biophysical ML model when the neurons have different frequencies and different PRCs (**Figure 9**).

We can see why the frequency differences are needed for resonance by considering the simplest example of identical PRCs of the form $\Delta(\phi) = \sin a - \sin(\phi + a)$. In this case, we solve the BVP:

$$\frac{G(\phi, \tau)R(\phi)}{d\phi} = \alpha \frac{1 + \tau^2}{\tau}(R(\phi) - 1) \tag{26}$$

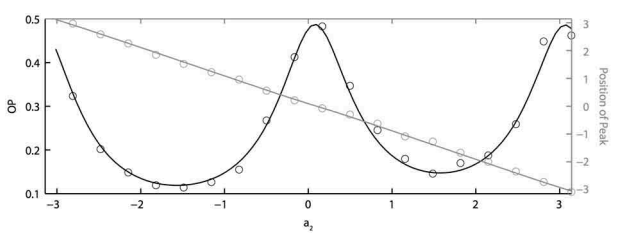


FIGURE 5 | PRC heterogeneity tunes the sharpness and peak position of the phase difference density. OP vs. a_2 (black axis) and the peak position of the phase difference density vs. a_2 (grey axis). Parameters used: $a_1 = 0.1$, $b_1 = 0.32$, $b_2 = 0.3$, $\tau = 1$, and $c = 0.8$.

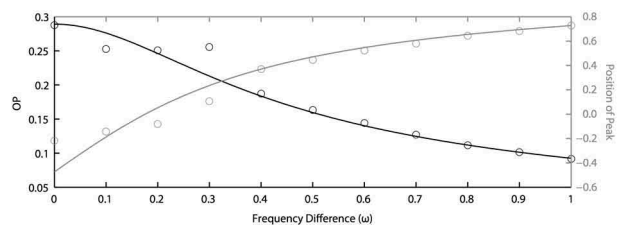


FIGURE 6 | Frequency differences limit noise-induced synchronization. OP decreases quickly as frequency differences increase (black axis). The peak position of the phase difference density is shifted by changing frequency differences between two oscillators (grey axis). Parameters used here: $a_1 = 0.1$, $b_1 = 0.32$, $a_2 = 0.6$, $b_2 = 0.3$, $\tau = 1$, and $c = 0.8$.

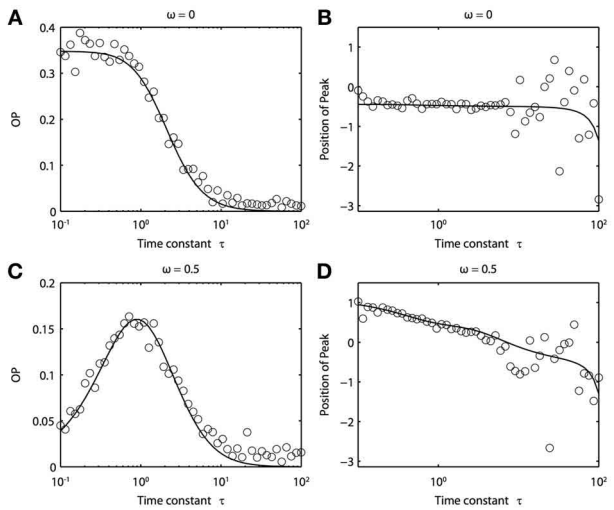


FIGURE 7 | Frequency differences between oscillators change the dependence of synchrony on the time constant of the correlated noise. **(A)** OP and **(B)** the peak position of the phase difference density for oscillators with no frequency difference ($\omega = 0$). **(C)** OP and **(D)** the peak position of the phase difference density for oscillators with a frequency difference ($\omega = 0.5$). Parameters used: $a_1 = 0.1$, $b_1 = 0.32$, $a_2 = 0.6$, $b_2 = 0.3$, and $c = 0.8$.

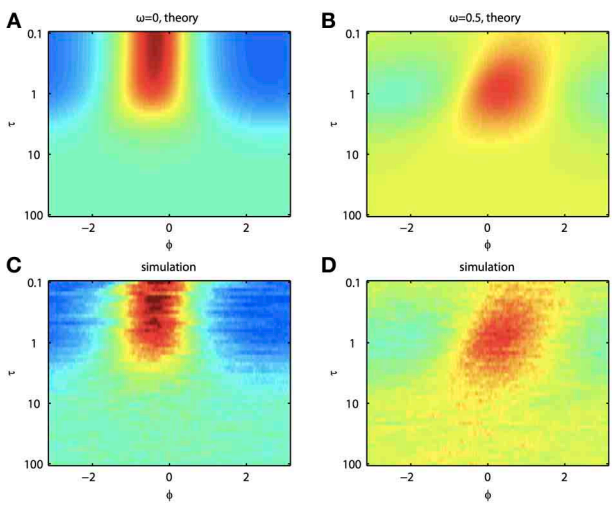


FIGURE 8 | Synchronization of oscillators with different frequencies resonates with the time constant of the correlated noise. **(A)** BVP solution and **(C)** results of Monte Carlo simulation for $R(\phi)$ between two oscillators with no frequency difference ($\omega = 0$) as τ varies. X-axis shows the phase difference, ϕ, and higher $R(\phi)$ is plotted as hotter colors in the heat map. **(B)** and **(D)**, same as **(A)** and **(C)** for two oscillators with a frequency difference ($\omega = 0.5$). Parameters used: $a_1 = 0.1$, $b_1 = 0.32$, $a_2 = 0.6$, $b_2 = 0.3$, and $c = 0.8$.

where $G(\phi, \tau) = (1 + \tau^2)\sin(a)^2 + 1 - c\cos\phi$. Here, α is proportional to the frequency difference. In particular, note that when $\omega = 0$, G is independent of τ and otherwise, τ acts to weaken the correlated noise-induced synchronization as it increases the part of G that is not phase dependent. However, the right side of this equation shows that the effect of the frequency difference is minimized when $\tau = 1$, and thus we expect resonance in the OP. This effect disappears when $\alpha = 0$.

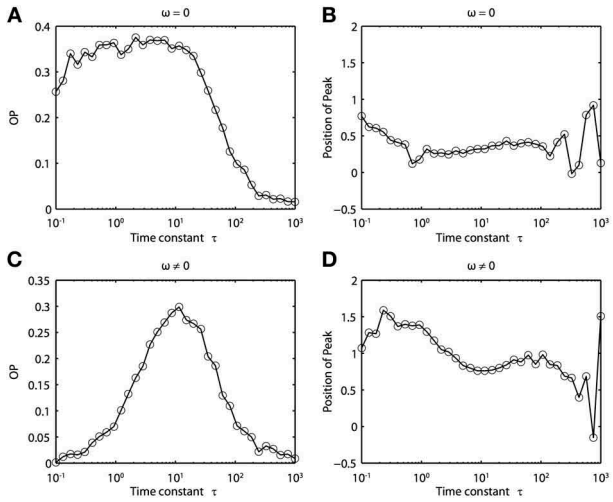


FIGURE 9 | Frequency differences between biophysically realistic spiking neuron models change the dependence of synchrony on the time constant of the correlated noise. **(A)** OP and **(B)** peak position for model neurons with different PRCs but the same frequency ($I_1 = 120$, $\phi_1 = 0.04$, $I_2 = 110$, and $\phi_2 = 0.04616$). **(C)** OP and **(D)** peak position for model neurons with different PRCs (same as above) and slightly different frequencies ($I_1 = 120$, $\phi_1 = 0.041$, $I_2 = 110$, and $\phi_2 = 0.04616$).

4. DISCUSSION

In our current study, we have extended a number of previous results describing the ability of neural oscillators to synchronize in the presence of correlated noise. Our methods are similar to those in Burton et al. (2012), with the additional aspect that we now use colored noise (OU process). The properties of the noise show up only through a convolution of the autocorrelation function of the noise with the phase functions $h_{nm}(\phi)$ that, in turn, depend only on the PRCs (see Appendix, Equation 56). Thus, we could easily generalize this work to noise with an arbitrary autocorrelation function. In addition, we have now included many more examples of the theory and shown that the conclusions from the perturbation theory continue to be valid for full biophysical models (cf. **Figure 2**). Further, we have shown that for low input correlations, heterogeneity can actually improve synchrony both pairwise and in large populations. We demonstrated that this theoretical effect can be seen in experimental recordings of regularly firing olfactory bulb mitral cells. Thus, we have significantly extended the findings presented in Burton et al. (2012), and our results on colored noise further suggest some experimentally testable phenomena, such as the resonance seen in slightly detuned oscillators (**Figures 7–9**). These novel findings and their biological implications are discussed in more detail below.

4.1. HETEROGENEITY CAN IMPROVE SYNCHRONY

We found that correlated noise can synchronize a heterogeneous pair of oscillators (comprised of a "bad synchronizer" and a

"good synchronizer") better than a homogeneous pair of bad synchronizers at low levels of input correlation and verified this experimentally. We showed that good (bad) synchronizers are characterized by having a relatively low (high) DC component in their PRC. Consistent with this, oscillators with the generic "type II" PRC (i.e., $\sin\phi$) are better synchronizers than oscillators with the generic "type I" PRC (i.e., $1-\cos\phi$).

Several authors have previously studied the effects of heterogeneity on the transfer of correlation. As we noted in Introduction, at low correlations, the output correlation is linearly proportional to the input correlation through a factor, S, called the susceptibility (de la Rocha et al., 2007; Shea-Brown et al., 2008). If we let $S(A, B)$ denote the susceptibility for two neurons, A, B, then what we have found in our current study is that in some cases, $S(A, A) > S(A, B) > S(B, B)$. Note that in our study, we are looking at output correlation related to spike-to-spike synchronization, whereas in many other studies of output correlation, the interest is in *spike count* correlation. We can regard our measure of synchrony as the same as spike count correlation, but over a time window that is of the order of the mean interspike interval. In a recent paper, (Shea-Brown et al., 2008) showed that for spike count correlation, $S(A, B) = \sqrt{S(A, A)S(B, B)}$ and thus, trivially, we can obtain $S(A, A) > S(A, B) > S(B, B)$ when A is "better" than B at transferring correlation. We want to emphasize that their result is for long time windows (that is, the window length tends to infinity). Which neurons are better than others at the transfer of correlation depends very strongly on the window of time through which you measure the correlation. Indeed, Barreiro et al. (2010) and Abouzeid and Ermentrout (2011) showed that type II PRCs have larger susceptibilities than type I for short time windows (i.e., spike synchrony) but the trend is reversed for large time windows (i.e., rate correlation).

Interestingly, the efficiency of correlated-noise induced synchronization is also modulated by firing rate in the low input correlation regime (de la Rocha et al., 2007; Tchumatchenko et al., 2010). Given that changes in firing rate can modulate PRC shape in biophysically realistic neuron models and in real neurons (Gutkin et al., 2005; Marella and Ermentrout, 2008; Stiefel et al., 2008, 2009; Schultheiss et al., 2010; Fink et al., 2011; Burton et al., 2012), whether or not (and the degree to which) PRC heterogeneity will enhance synchrony may depend in part on the firing rate. However, in the simplest cases (such as models like the leaky-integrate and fire model and the theta model), the only effect of the firing rate on the shape of the PRC is to change its amplitude. Since amplitude (but not shape) changes can be absorbed into the size of the noise, and our theory shows that the phase difference density is independent of the size of the noise (at least, if it is small enough), changes in firing rate will have no effect on the synchronization of pairs of neurons firing at the same or nearly the same rates.

The ability of cellular heterogeneity to regulate which oscillators synchronize best as a function of input correlation likely contributes to coding in many neural systems. In the olfactory bulb, where oscillatory synchrony appears to be critical to olfactory coding [for review, see Bathellier et al. (2010)], tens of "sister" mitral cells are linked to each glomerulus where they receive highly correlated afferent input (Carlson et al., 2000; Schoppa and Westbrook, 2001). Each sister mitral cell of a glomerulus may also participate in independent (i.e., unshared) lateral inhibitory circuits with non-sister mitral cells of surrounding glomeruli, mediated by local inhibitory granule cells (Dhawale et al., 2010; Tan et al., 2010). On average, sister mitral cells are thus subject to high input correlations while non-sister mitral cells are subject to low (though non-zero) input correlations (Dhawale et al., 2010). Further, we and others have demonstrated that mitral cells exhibit substantial cell-to-cell heterogeneity (Padmanabhan and Urban, 2010; Angelo and Margrie, 2011; Angelo et al., 2012; Burton et al., 2012). Based on our current results, this heterogeneity will thus act to reduce output synchrony of sister mitral cells but *enhance* output synchrony of non-sister mitral cells. Thus, in the context of the olfactory system, heterogeneity will promote encoding of combinatorial sensory information (i.e., activation of non-sister mitral cells by odor combinations).

Our results suggest that heterogeneity can only enhance correlation-induced synchronization by a moderate amount between two neural oscillators (up to 36% in BVP solutions using mitral cell PRCs). Two properties of the olfactory bulb nevertheless suggest that even this moderate effect can significantly influence patterns of oscillatory synchrony in the olfactory system. First, the reciprocal dendrodendritic connectivity between mitral cells and granule cells enables activity-dependent regulation of granule cell recruitment (Arevian et al., 2008), which can lead to amplification of granule cell-mediated correlated noise-induced synchronization (Marella and Ermentrout, 2010). Second, mitral cells separated by up to ∼2 mm can engage in lateral inhibitory interactions (Egger and Urban, 2006), thus multiplying the synchrony-enhancing effect of cellular heterogeneity across a potentially large fraction of the ∼40,000 total mitral cells per mouse olfactory bulb (Benson et al., 1984). Whether neural oscillator heterogeneity exists in, and significantly enhances, correlated-noise induced synchrony in other brain regions remains a promising topic of future research.

4.2. RESONANCE

In addition to the above findings, we found that there exists some resonance of correlated noise-induced synchronization with respect to the time scale of the noise. That is, we found a local maximum in OP as the time scale of the correlated noise varied. Surprisingly, this only occurs when there is a difference in the frequencies between the two oscillators. The requirement for the frequency difference would seem to contradict earlier work (Galán et al., 2008), where it was found that the Liapunov exponent was most negative when the noise has a particular time scale. However, when the noise is only partially correlated, the uncorrelated part of the noise causes a drift in the phase difference. The degree of this drift is also dependent on the time scale of the noise, and thus the two effects cancel. A frequency difference breaks this symmetry by adding an additional drift term, which prevents one from factoring out the resonance. A frequency difference thus leads to a dependence of OP on the time scale of the noise. We have not yet tested this idea experimentally, but it seems to be quite robust, having been found in both the simple

phase models (**Figures 7**, **8**) and in the biophysical ML model (**Figure 9**).

4.3. LIMITATIONS OF THE THEORY

The analysis that we have done in this paper and in our earlier papers requires that the neurons fire almost periodically. This means that the activity of the neurons is *mean driven* rather than *fluctuation driven* so that the coefficient of variation of the inter-spike intervals should be small. While this may not be the case in all areas of the brain, there are many regions, such as the olfactory bulb, where the firing rate can be quite regular and synchronous as indicated by the large rhythmic local field potentials. Assuming that the neurons are firing at a fairly regular rate, it is also reasonable to ask how well the PRC describes such noisy neurons. An extensive review of the caveats of PRC theory for real neurons can be found in Smeal et al. (2010). Another issue is the actual estimation of the PRCs in the presence of noise. Several studies have shown that background synaptic activity and other forms of noise do not significantly affect the shape of the PRC (Ermentrout et al., 2011; Netoff et al., 2012).

In conclusion, we have extended our previous work to demonstrate that oscillator heterogeneity and frequency differences interact with the time scale of input noise to regulate how correlated noise synchronizes uncoupled oscillators.

4.4. DATA SHARING

All codes are available by request from the authors. They include Matlab and XPPaut files.

ACKNOWLEDGMENTS

This work was supported by National Institute on Drug Abuse Predoctoral Training Grant R90DA023426 and a R.K. Mellon Foundation Presidential Fellowship in the Life Sciences (Pengcheng Zhou), an Achievement Rewards for College Scientists Foundation Fellowship (Shawn D. Burton), National Institute on Deafness and Other Communication Disorders Grant 5R01DC011184-07 (Nathaniel N. Urban and G. Bard Ermentrout), and National Science Foundation grant DMS1219754 (G. Bard Ermentrout).

REFERENCES

Abouzeid, A., and Ermentrout, B. (2009). Type-II phase resetting curve is optimal for stochastic synchrony. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 80:011911. doi: 10.1103/PhysRevE.80.011911

Abouzeid, A., and Ermentrout, B. (2011). Correlation transfer in stochastically driven neural oscillators over long and short time scales. *Phys. Rev. E* 84:061914. doi: 10.1103/PhysRevE.84.061914

Angelo, K., and Margrie, T. W. (2011). Population diversity and function of hyperpolarization-activated current in olfactory bulb mitral cells. *Sci. Rep.* 1:50. doi: 10.1038/srep00050

Angelo, K., Rancz, E. A., Pimentel, D., Hundahl, C., Hannibal, J., Fleischmann, A. et al. (2012). A biophysical signature of network affiliation and sensory processing in mitral cells. *Nature* 488, 375–378. doi: 10.1038/nature11291

Arevian, A. C., Kapoor, V., and Urban, N. N. (2008). Activity-dependent gating of lateral inhibition in the mouse olfactory bulb. *Nat. Neurosci.* 11, 80–87. doi: 10.1038/nn2030

Barreiro, A. K., Shea-Brown, E., and Thilo, E. L. (2010). Time scales of spike-train correlation for neural oscillators with common drive. *Phys. Rev. E* 81:011916. doi: 10.1103/PhysRevE.81.011916

Bathellier, B., Gschwend, O., and Carleton, A. (2010). "Temporal coding in olfaction," in *The Neurobiology of Olfaction,* ed A. Menini (Boca Raton, FL: CRC Press), 329–348.

Benson, T. E., Ryugo, D. K., and Hinds, J. W. (1984). Effects of sensory deprivation on the developing mouse olfactory system: a light and electron microscopic, morphometric analysis. *J. Neurosci.* 4, 638–653.

Brown, E., Moehlis, J., and Holmes, P. (2004). On the phase reduction and response dynamics of neural oscillator populations. *Neural Comput.* 16, 673–715. doi: 10.1162/089976604322860668

Burton, S. D., Ermentrout, G. B., and Urban, N. N. (2012). Intrinsic heterogeneity in oscillatory dynamics limits correlation-induced neural synchronization. *J. Neurophysiol.* 108 2115–2133. doi: 10.1152/jn.00362.2012

Carlson, G. C., Shipley, M. T., and Keller, A. (2000). Long-lasting depolarizations in mitral cells of the rat olfactory bulb. *J. Neurosci.* 20, 2011–2021.

de la Rocha, J., Doiron, B., Shea-Brown, E., Josić, K., and Reyes, A. (2007). Correlation between neural spike trains increases with firing rate. *Nature* 448 802–806. doi: 10.1038/nature06028

Dhawale, A. K., Hagiwara, A., Bhalla, U. S., Murthy, V. N., and Albeanu, D. F. (2010). Non-redundant odor coding by sister mitral cells revealed by light addressable glomeruli in the mouse. *Nat. Neurosci.* 13, 1404–1412. doi: 10.1038/nn.2673

Ermentrout, B. (2002). *Simulating, Analyzing, and Animating Dynamical Systems: A Guide to XPPAUT for Researchers and Students.* Philadelphia, PA: SIAM. doi: 10.1137/1.9780898718195

Fries, P., Reynolds, J. H., Rorie, A. E., and Desimone, R. (2001). Modulation of oscillatory neuronal synchronization by selective visual attention. *Science* 291, 1560–1563. doi: 10.1126/science.1055465

Egger, V., and Urban, N. N. (2006). Dynamic connectivity in the mitral cell-granule cell microcircuit. *Semin Cell Dev. Biol.* 17, 424–432. doi: 10.1016/j.semcdb.2006.04.006

Ermentrout, G. B., Beverlin, B., Troyer, T., and Netoff, T. I. (2011). The variance of phase-resetting curves. *J. Comput. Neurosci.* 31, 185–197. doi: 10.1007/s10827-010-0305-9

Ermentrout, G. B., Galán, R. F., and Urban, N. N. (2007). Relating neural dynamics to neural coding. *Phys. Rev. Lett.* 99:248103. doi: 10.1103/PhysRevLett.99.248103

Fink, C. G., Booth, V., and Zochowski, M. (2011). Cellularly-driven differences in network synchronization propensity are differentially modulated by firing frequency. *PLoS Comput. Biol.* 7:e1002062.doi: 10.1371/journal.pcbi.1002062

Galán, R. F., Ermentrout, G. B., and Urban, N. N. (2008). Optimal time scale for spike-time reliability: theory, simulations, and experiments. *J. Neurophysiol.* 99, 277–283. doi: 10.1152/jn.00563.2007

Galán, R. F., Fourcaud-Trocmé, N., Ermentrout, G. B., and Urban, N. N. (2006). Correlation-induced synchronization of oscillations in olfactory bulb neurons. *J. Neurosci.* 26, 3646–3655. doi: 10.1523/JNEUROSCI.4605-05.2006

Goldobin, D. S., Teramae, J. N., Nakao, H., and Ermentrout, G. B. (2010). Dynamics of limit-cycle oscillators subject to general noise. *Phys. Rev. Lett.* 105:154101. doi: 10.1103/PhysRevLett.105.154101

Gutkin, B. S., Ermentrout, G. B., and Reyes, A. D. (2005). Phase-response curves give the responses of neurons to transient inputs. *J. Neurophysiol.* 94, 1623–1635. doi: 10.1152/jn.00359.2004

Izhikevich, E. M. (2007). *Dynamical Systems in Neuroscience: The Geometry of Excitability and Bursting.* Cambridge, MA: MIT Press.

Kuramoto, Y. (2003). *Chemical Oscillations, Waves, and Turbulence.* Mineola, NY: Dover.

Marella, S., and Ermentrout, B. (2010). Amplification of asynchronous inhibition-mediated synchronization by feedback in recurrent networks. *PLoS Comput. Biol.* 6:e1000679. doi: 10.1371/journal.pcbi.1000679

Marella, S., and Ermentrout, G. B. (2008). Class-II neurons display a higher degree of stochastic synchronization than class-I neurons. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 77:041918. doi: 10.1103/PhysRevE.77.041918

Markowitz, D. A., Collman, F., Brody, C. D., Hopfield, J. J., and Tank, D. W. (2008). Rate-specific synchrony: using noisy oscillations to detect equally active neurons. *Proc. Natl. Acad. Sci. U.S.A.* 105, 8422–8427. doi: 10.1073/pnas.0803183105

Nakao, H., Arai, K., and Kawamura, Y. (2007). Noise-induced synchronization and clustering in ensembles of uncoupled limit-cycle oscillators.

Phys. Rev. Lett. 98:184101. doi: 10.1103/PhysRevLett.98.184101

Netoff, T., Schwemmer, M. A., and Lewis, T. J. (2012). "Experimentally estimating phase response curves of neurons: theoretical and practical issues," in *Phase Response Curves in Neuroscience: Theory, Experiment, and Analysis,* eds N. W. Schltheiss, A. A. Prinz, R. J. (New York, NY: Springer), 95–129.

Padmanabhan, K., and Urban, N. N. (2010). Intrinsic biophysical diversity decorrelates neuronal firing while increasing information content. *Nat. Neurosci.* 13, 1276–1282. doi: 10.1038/nn.2630

Pfeuty, B., Mato, G., Golomb, D., and Hansel, D. (2003). Electrical synapses and synchrony: the role of intrinsic currents. *J. Neurosci.* 23, 6280–6294.

Pfeuty, B., Mato, G., Golomb, D., and Hansel, D. (2005). The combined effects of inhibitory and electrical synapses in synchrony. *Neural Comput.* 17, 633–670. doi: 10.1162/0899766053019917

Pikovsky, A. S., Rosenblum, M. G., Osipov, G. V., and Kurths, J. (1997). Phase synchronization of chaotic oscillators by external driving. *Phys. Nonlinear. Phenom.* 104, 219–238.

Risken, H. (1984). *Fokker-Planck Equation, Springer Series in Synergetics,* Vol. 18. Berlin: Springer. doi: 10.1007/978-3-642-96807-5

Rinzel, J., and Ermentrout, G. B. (1989). "Analysis of neural excitability and oscillations," in *Methods in Neuronal Modeling*, eds C. Koch and I. Segev (Cambridge, MA: MIT Press), 135–169.

Sanes, J. N., and Donoghue, J. P. (1993). Oscillations in local field potentials of the primate motor cortex during voluntary movement. *Proc. Natl. Acad. Sci. U.S.A.* 90, 4470–4474. doi: 10.1073/pnas.90.10.4470

Schoppa, N. E., and Westbrook, G. L. (2001). Glomerulus-specific synchronization of mitral cells in the olfactory bulb. *Neuron* 31, 639–651. doi: 10.1016/S0896-6273(01)00389-0

Schultheiss, N. W., Edgerton, J. R., and Jaeger, D. (2010). Phase response curve analysis of a full morphological globus pallidus neuron model reveals distinct perisomatic and dendritic modes of synaptic integration. *J. Neurosci.* 30, 2767–2782. doi: 10.1523/JNEUROSCI.3959-09.2010

Shea-Brown, E, Josić, K., de la Rocha, J., and Doiron, B. (2008). Correlation and synchrony transfer in integrate-and-fire neurons: basic properties and consequences for coding. *Phys. Rev. Lett.* 100:108102. doi: 10.1103/PhysRevLett.100.108102

Smeal, R. M., Ermentrout, G. B., and White, J. A. (2010). Phase-response curves and synchronized neural networks. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 365, 2407–2422. doi: 10.1098/rstb.2009.0292

Stiefel, K. M., Gutkin, B. S., and Sejnowski, T. J. (2008). Cholinergic neuromodulation changes phase response curve shape and type in cortical pyramidal neurons. *PLoS ONE* 3:e3947. doi: 10.1371/journal.pone.0003947

Stiefel, K. M., Gutkin, B. S., and Sejnowski, T. J. (2009). The effects of cholinergic neuromodulation on neuronal phase-response curves of modeled cortical neurons. *J. Comput. Neurosci.* 26, 289–301. doi: 10.1007/s10827-008-0111-9

Tan, J., Savigner, A., Ma, M., and Luo, M. (2010). Odor information processing by the olfactory bulb analyzed in gene-targeted mice. *Neuron* 65, 912–926. doi: 10.1016/j.neuron.2010.02.011

Tchumatchenko, T., Malyshev, A., Geisel, T., Volgushev, M., and Wolf, F. (2010). Correlations and synchrony in threshold neuron models. *Phys. Rev. Lett.* 104:058102. doi: 10.1103/PhysRevLett.104.058102

Teramae, J. N., and Tanaka, D. (2004). Robustness of the noise-induced phase synchronization in a general class of limit cycle oscillators. *Phys. Rev. Lett.* 93:204103. doi: 10.1103/PhysRevLett.93.204103

Torben-Nielsen, B., Uusisaari, M., and Stiefel, K. M. (2010). A comparison of methods to determine neuronal phase-response curves. *Front. Neuroinform.* 4:6. doi: 10.3389/fninf.2010.00006

Tsubo, Y., Teramae, J. N., and Fukai, T. (2007). Synchronization of excitatory neurons with strongly heterogeneous phase responses. *Phys. Rev. Lett.* 99:228101. doi: 10.1103/PhysRevLett.99.228101

Wang, X. J. (2010). Neurophysiological and computational principles of cortical rhythms in cognition. *Physiol. Rev.* 90, 1195–1268. doi: 10.1152/physrev.00035.2008

White, J. A., Chow, C. C., Rit, J., Soto-Treviño, C., and Kopell, N. (1998). Synchronization and oscillatory dynamics in heterogeneous, mutually inhibited neurons. *J. Comput. Neurosci.* 5, 5–16. doi: 10.1023/A:1008841325921

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.

Received: 25 May 2013; accepted: 25 July 2013; published online: 21 August 2013.
Citation: Zhou P, Burton SD, Urban NN and Ermentrout GB (2013) Impact of neuronal heterogeneity on correlated colored noise-induced synchronization. Front. Comput. Neurosci. 7:113. doi: 10.3389/fncom.2013.00113
This article was submitted to the journal Frontiers in Computational Neuroscience.
Copyright © 2013 Zhou, Burton, Urban and Ermentrout. 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) or licensor 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.

APPENDIX

REDUCTION TO A PHASE MODEL

Consider a general oscillator receiving a possibly noisy time-dependent signal:

$$\frac{dX}{dt} = F(X) + \epsilon N(X, t) \tag{27}$$

Here $N(X, t)$ is the external or imposed inputs into the system. For single compartment neural models, N will typically only affect the membrane potential, e.g., as an injected or synaptic current. We assume that $X' = F(X)$ has as a solution an exponentially stable limit cycle, $U(t + T) = U(t)$ and that ϵ is a small positive parameter characterizing the magnitude of the input. We are interested in how the phase of the limit cycle evolves in time in the presence of small inputs. The phase of a limit cycle is easy to define when a point lies *on* the limit cycle. For example, for neurons, the phase is just the time since the last spike of the cell. However, if the limit cycle is attracting, then it is also possible to define the phase of a point that is near, but not directly on the limit cycle. Specifically, there is a function $\Theta(X)$ that maps a point near the limit cycle, X, to the phase that it will eventually reach as it is attracted to the limit cycle (asymptotic phase). Clearly $\Theta(U(t)) = t$. Define the phase to be $\theta(t) = \Theta(X(t))$, so that by the chain rule:

$$\frac{d\theta}{dt} = \nabla_X \Theta(X) \cdot \frac{dX}{dt} \tag{28}$$

$$= \nabla_X \Theta(X) \cdot F(X) + \epsilon \nabla_X \Theta(X) \cdot N(X(t), t) \tag{29}$$

$$= 1 + \epsilon \nabla_X \Theta(X) \cdot N(X(t), t) \tag{30}$$

Thus, in the absence of inputs, the phase moves around the circle at constant velocity. This expression is exact, but not very helpful since it requires knowledge of the solution $X(t)$. Kuramoto's approximation (which is valid for small ϵ) is to replace $X(t)$ in the right-hand side by $U(\theta(t))$, where U is the unperturbed limit cycle (Kuramoto, 2003). This closes the system yielding:

$$\frac{d\theta}{dt} = 1 + \epsilon Z(\theta) \cdot N(U(\theta), t) \tag{31}$$

where we have defined $Z(\theta) := \nabla_X \Theta(U(\theta))$. The function, $Z(\theta)$ is the so-called adjoint function satisfying the linear equation:

$$Z' = -\left(D_X F(U(t))\right)^T Z(t) \tag{32}$$

with $Z(t) \cdot U'(t) = 1$. Here $D_X F(U(t))$ means the linearization of $F(X)$ evaluated along the limit cycle.

In single compartment neuron models, inputs appear only in the voltage component of the neural oscillator in the form of external currents so that the dot product in Equation 31 becomes scalar multiplication:

$$\frac{d\theta}{dt} = 1 + \epsilon \Delta(\theta) I(U(\theta), t)/C \tag{33}$$

where I is the input current, C is the capacitance, and $\Delta(\theta)$ is the voltage component of the vector Z. The quantity, $\Delta(\theta)$, is sometimes called the infinitesimal PRC and, for small perturbations, is proportional to the PRC.

DERIVATION OF THE STATIONARY DENSITY OF PHASE DIFFERENCES

The Langevin equations that drive the phase models (Equations 4–6) correspond to a forward Fokker-Planck (FP) equation that can be written as (Risken, 1984):

$$\begin{aligned}\frac{\partial \rho}{\partial t} = &\frac{1}{\tau}\left\{\frac{\partial}{\partial x}\left(\frac{1}{2}\frac{\partial}{\partial x} + x\right) + \frac{\partial}{\partial y}\left(\frac{1}{2}\frac{\partial}{\partial y} + y\right) + c\frac{\partial^2}{\partial x \partial y}\right\}\rho - \frac{\partial}{\partial \theta}\rho \\ &-\epsilon\left\{\frac{\partial}{\partial \theta}[\Delta_1(\theta) x \rho] + \frac{\partial}{\partial \phi}[(\Delta_2(\theta + \phi) y - \Delta_1(\theta) x)\rho]\right\} \\ &-\epsilon^2 \omega \frac{\partial}{\partial \phi}\rho\end{aligned} \tag{34}$$

When the distribution of phase differences is stationary, $\frac{\partial \rho}{\partial t} = 0$. Our goal is to exploit the smallness of ϵ to compute this stationary density with which we can compute the marginal distribution of the phase difference.

ANALYTICAL SOLUTION

We expand the steady state ρ in ϵ:

$$\rho(x, y, \theta, \phi) = \rho_0(x, y, \theta, \phi) + \epsilon\rho_1(x, y, \theta, \phi) + \epsilon^2\rho_2(x, y, \theta, \phi) \tag{35}$$
$$\iiiint \rho_0(x, y, \theta, \phi)dxdyd\theta d\phi = 1$$
$$\iiiint \rho_n(x, y, \theta, \phi)dxdyd\theta d\phi = 0, \quad n = 1, 2$$

We define the operator:

$$L_0 = \frac{1}{\tau}\left\{\frac{\partial}{\partial x}\left(\frac{1}{2}\frac{\partial}{\partial x} + x\right) + \frac{\partial}{\partial y}\left(\frac{1}{2}\frac{\partial}{\partial y} + y\right) + c\frac{\partial^2}{\partial x\partial y}\right\} + \frac{\partial}{\partial\theta} \tag{36}$$

At steady state condition ($\frac{\partial\rho}{\partial t} = 0$), we substitute the above expansion into the FP equation and collect the powers of ϵ. We need to go to ϵ^2:

$$0 = L_0\rho_0 \tag{37}$$
$$0 = L_0\rho_1 - \left\{\frac{\partial}{\partial\theta}[\Delta_1(\theta)x\rho_0] + \frac{\partial}{\partial\phi}[(\Delta_2(\theta + \phi)y - \Delta_1(\theta)x)\rho_0]\right\} \tag{38}$$
$$0 = L_0\rho_2 - \left\{\frac{\partial}{\partial\theta}[\Delta_1(\theta)x\rho_1] + \frac{\partial}{\partial\phi}[(\Delta_2(\theta + \phi)y - \Delta_1(\theta)x)\rho_1]\right\} - a\frac{\partial}{\partial\phi}\rho_0 \tag{39}$$

Solving Equation 37

Equation 37 is just a linear separable equation, independent of ϕ, so, by inspection:

$$\rho_0(x, y, \theta, \phi) = \frac{1}{2\pi}G(x, y)R(\phi) \tag{40}$$

where:

$$G(x, y) = \frac{1}{\sqrt{1 - c^2}\pi}e^{-\frac{1}{1-c^2}(x^2+y^2-2cxy)} \tag{41}$$

and $R(\phi)$ remains to be determined. Note that $G(x, y)$ is just the standard stationary solution to the multivariate OU equation. At this juncture, we remark that our main goal is to find $R(\phi)$, which is the marginal density of the phase differences between the two oscillators.

Solving Equation 38

Both Equations 38 and 39 have the form $L_0\rho = b(x, y, \theta)$. L_0 operates on the space of functions defined on $R^2 \times S^1$ that are twice continuously differentiable in x, y and continuously differentiable in θ. In this space, L_0 has a one-dimensional nullspace spanned by $G(x, y)$ (constant in θ) and so L_0 is not invertible. However, $L_0\rho(x, y, \theta) = b(x, y, \theta)$ does have a solution provided that $b(x, y, \theta)$ is orthogonal to the null space of L_0^*, which is the adjoint linear operator of L_0. Since L_0 is a standard probability operator, its adjoint is always 1 (i.e., the function that is 1 everywhere).

Since:

$$b_1(x, y, \theta) = \frac{xG(x, y)}{2\pi}[\Delta_1'(\theta)R(\phi) - \Delta_1(\theta)R'(\phi)] + yG(x, y)[\Delta_2'(\theta + \phi)R(\phi) + \Delta_2(\theta + \phi)R'(\phi)] \tag{42}$$

we see that $\iiint b_1(x, y, \theta)dxdyd\theta = 0$. Thus, $L_0\rho_1 = b_1$ has a solution. Since:

$$L_0[xG(x, y)] = -xG(x, y)/\tau \tag{43}$$
$$L_0[yG(x, y)] = -yG(x, y)/\tau \tag{44}$$

we look for a solution of the form:

$$\rho_1(x, y, \theta, \phi) = \frac{w_1(\theta, \phi)xG(x, y) + w_2(\theta, \phi)yG(x, y)}{2\pi} \tag{45}$$

Inserting this into Equation 38, we find that $w_j(\theta, \phi)$ must satisfy:

$$\frac{\partial}{\partial\theta}w_1(\theta,\phi)+\frac{w_1(\theta,\phi)}{\tau}=-\Delta_1'(\theta)R(\phi)+\Delta_1(\theta)R'(\phi) \tag{46}$$

$$\frac{\partial}{\partial\theta}w_2(\theta,\phi)+\frac{w_2(\theta,\phi)}{\tau}=-\Delta_2'(\theta+\phi)R(\phi)-\Delta_2(\theta+\phi)R'(\phi) \tag{47}$$

w_j must be periodic functions of θ; we defer their exact solution to later, but note that there is always a unique periodic solution to each of these equations.

Solving Equation 39

We now have:

$$b_2(x,y,\theta)=\frac{\partial}{\partial\theta}[\Delta_1(\theta)x\rho_1]+\frac{\partial}{\partial\phi}[(\Delta_2(\theta+\phi)y-\Delta_1(\theta)x)\rho_1]+a\frac{\partial}{\partial\phi}\rho_0 \tag{48}$$

In order to solve Equation 39, for $\rho_2(x, y, \theta, \phi)$, we must have:

$$\begin{aligned}
0&=\iint_{-\infty}^{\infty}\int_0^{2\pi}b_2(x,y,\theta)dxdyd\theta\\
&=0+\frac{\partial}{\partial\phi}\left\{\iint_{-\infty}^{\infty}\int_0^{2\pi}\left\{\frac{\Delta_2(\theta+\phi)}{2\pi}[w_2(\theta,\phi)y^2+w_1(\theta,\phi)xy]G(x,y)\right.\right.\\
&\quad\left.\left.-\frac{\Delta_1(\theta)}{2\pi}[w_1(\theta,\phi)x^2+w_2(\theta,\phi)xy]G(x,y)+\frac{a}{2\pi}R(\phi)G(x,y)\right\}dxdyd\theta\right\}\\
&=\frac{1}{4\pi}\frac{\partial}{\partial\phi}\left\{\int_0^{2\pi}\{\Delta_2(\theta+\phi)[w_2(\theta,\phi)+c\cdot w_1(\theta,\phi)]-\Delta_1(\theta)[w_1(\theta,\phi)+c\cdot w_2(\theta,\phi)]\}d\theta+4\pi\omega R(\phi)\right\}\\
&=\frac{1}{4\pi}\frac{\partial}{\partial\phi}[f(\phi)+4\pi aR(\phi)]
\end{aligned} \tag{49}$$

where:

$$\begin{aligned}
f(\phi)&=\int_0^{2\pi}[\Delta_2(\theta+\phi)v_2(\theta,\phi)-\Delta_1(\theta)v_1(\theta,\phi)]d\theta\\
v_1(\theta,\phi)&=w_1(\theta,\phi)+cw_2(\theta,\phi)\\
v_2(\theta,\phi)&=w_2(\theta,\phi)+cw_1(\theta,\phi)
\end{aligned} \tag{50}$$

Given Equations 46 and 47, we see that $v_1(\theta)$ and $v_2(\theta)$ satisfy:

$$\begin{aligned}
v_1'(\theta)+\frac{v_1(\theta)}{\tau}&=-[\Delta_1'(\theta)+c\cdot\Delta_2'(\theta+\phi)]R(\phi)+[\Delta_1(\theta)-c\cdot\Delta_2(\theta+\phi)]R'(\phi)\\
&:=q_1(\theta)
\end{aligned} \tag{51}$$

$$\begin{aligned}
v_2'(\theta)+\frac{v_2(\theta)}{\tau}&=-[c\cdot\Delta_1'(\theta)+\Delta_2'(\theta+\phi)]R(\phi)+[c\cdot\Delta_1(\theta)-\Delta_2(\theta+\phi)]R'(\phi)\\
&:=q_2(\theta)
\end{aligned} \tag{52}$$

For Equations 51 and 52, we can write down the solution of $v_1(\theta)$ and $v_2(\theta)$ in terms of $q_1(\theta)$ and $q_2(\theta)$ (see Appendix):

$$v_n(\theta)=\int_0^{\infty}e^{-\frac{s}{\tau}}q_n(\theta-s)ds,\quad n=1,2 \tag{53}$$

We substitute $v_n(\phi)$ into $f(\phi)$,

$$\begin{aligned} f(\phi) &= \int_0^{2\pi} [\Delta_2(\theta+\phi)v_2(\theta) - \Delta_1(\theta)v_1(\theta)]d\theta \\ &= \int_0^{\infty} e^{-\frac{s}{\tau}} ds \int_0^{2\pi} [\Delta_2(\theta+\phi)q_2(\theta-s) - \Delta_1(\theta)q_1(\theta-s)]d\theta \\ &= \int_0^{\infty} e^{-\frac{s}{\tau}} ds \int_0^{2\pi} h(\theta,\phi,s)d\theta \end{aligned} \tag{54}$$

where:

$$\begin{aligned} h(\theta,\phi,s) &= \Delta_2(\theta+\phi)q_2(\theta-s) - \Delta_1(\theta)q_1(\theta-s) \\ &= \Delta_1(\theta-s)[c\cdot\Delta_2(\theta+\phi) - \Delta_1(\theta)]R'(\phi) \\ &\quad + \Delta_2(\theta+\phi-s)[-\Delta_2(\theta+\phi) + c\cdot\Delta_1(\theta)]R'(\phi) \\ &\quad + \Delta_1'(\theta-s)[-c\cdot\Delta_2(\theta+\phi) + \Delta_1(\theta)]R(\phi) \\ &\quad + \Delta_2'(\theta+\phi-s)[-\Delta_2(\theta+\phi) + c\cdot\Delta_1(\theta)]R(\phi) \end{aligned} \tag{55}$$

Define:

$$\begin{aligned} g_{mn}(\phi) &= \int_0^{\infty} h_{mn}(s+\phi)e^{-\frac{s}{\tau}} ds \\ h_{mn}(s) &= \int_0^{2\pi} \Delta_m(\theta)\Delta_n(\theta+s)d\theta \end{aligned} \tag{56}$$

Since $\Delta_1(\theta)$ and $\Delta_2(\theta)$ are periodic functions,

$$\begin{aligned} \int_0^{2\pi} h(\theta,\phi,s)d\theta &= \int_0^{2\pi} \{\Delta_1(\theta-s)\{[c\Delta_2(\theta+\phi) - \Delta_1(\theta)]R'(\phi) + [c\Delta_2'(\theta+\phi) - \Delta_1'(\theta)]R(\phi)\} \\ &\quad + \Delta_2(\theta+\phi-s)\{[c\Delta_1(\theta) - \Delta_2(\theta+\phi)]R'(\phi) - [c\Delta_1'(\theta) - \Delta_2'(\theta+\phi)]R(\phi)\}\}d\theta \\ &= \{c[h_{12}(s+\phi) + h_{21}(s-\phi)] - [h_{11}(s) + h_{22}(s)]\}R'(\phi) \\ &\quad + \left\{c\frac{d}{d\phi}[h_{12}(s+\phi) + h_{21}(s-\phi)] - \frac{d}{d\phi}[h_{11}(s+\phi) - h_{22}(s+\phi)]\big|_{\phi=0}\right\}R(\phi) \end{aligned} \tag{57}$$

$$\begin{aligned} f(\phi) &= \int_0^{\infty} e^{-\frac{s}{\tau}} ds \int_0^{2\pi} h(\theta,\phi,s)d\theta \\ &= \{c[g_{12}(\phi) + g_{21}(-\phi)] - [g_{11}(0) + g_{22}(0)]\}R'(\phi) \\ &\quad + \left\{c\cdot\frac{d}{d\phi}[g_{12}(\phi) + g_{21}(-\phi)] - [g_{11}'(0) - g_{22}'(0)]\right\}R(\phi) \\ &= \frac{d}{d\phi}\{[c\cdot g(\phi) - C_1]R(\phi)\} - C_2R(\phi) \end{aligned} \tag{58}$$

where:

$$g(\phi) = g_{12}(\phi) + g_{21}(-\phi) \tag{59}$$

$$C_1 = g_{11}(0) + g_{22}(0) \tag{60}$$

$$C_2 = g_{11}'(0) - g_{22}'(0) \tag{61}$$

Combined with Equations 49–58, we have a boundary value problem (BVP):

$$\frac{d}{d\phi}\{[c \cdot g(\phi) - C_1]R(\phi)\} + (4\pi\omega - C_2)R(\phi) = K \tag{62}$$

$$R(\phi) = R(\phi + 2\pi)$$

$$g(\phi) = g(\phi + 2\pi)$$

$$\int_{-\pi}^{\pi} R(\phi)d\phi = 1$$

$$K = 2\omega - \frac{C_2}{2\pi}$$

To solve this BVP, we need to compute $g(\phi)$ for a given PRC. We use two forms of the PRC in this paper:

$$\Delta(\theta) = \sin(a) - \sin(\theta + a) + b\sin(2\theta) \tag{63}$$

and

$$\Delta(\theta) = A[\sin(B) - \sin(\theta + B)]e^{C(\theta - 2\pi)} \quad (\theta \in (0, 2\pi),\ \Delta(\theta) = \Delta(\theta + 2\pi)) \tag{64}$$

The required integrals can be computed for both PRC forms. More generally, all PRCs can be written in Fourier form and, again, the integrals are readily computed to obtain $g(\phi)$ (see below).

***Small correlation approximation for R*(ϕ)**

We use a BVP solver to get the numerical solution for $R(\phi)$, but we would like to better understand the form of $R(\phi)$ at low values of correlation, c, so we expand R as a series in c. As K is dependent on c, we must also expand K in c. Finally, we need to keep the normalization condition for $R(\phi)$, hence:

$$R(\phi) = R_0(\phi) + cR_1(\phi) + \ldots \qquad K = K_0 + cK_1 + \ldots \tag{65}$$

$$R_0(\phi) = R(\phi)\big|_{c=0}, \quad \int_{-\pi}^{\pi} R_0(\phi)d\phi = 1$$

$$\int_{-\pi}^{\pi} R_n(\phi) = 0,\ n \geq 1$$

We substitute these expressions into the BVP, Equation 62 and find after equating powers of c:

$$-C_1R_0'(\phi) + (4\pi\omega - C_2)R_0(\phi) = K_0 \tag{66}$$

$$-C_1R_1'(\phi) + (4\pi a - C_2)R_1(\phi) + [g(\phi)R_0(\phi)]' = K_1 \tag{67}$$

We can integrate both left and right side over $[-\pi, \pi]$, then use the assumptions and periodicity requirements above to get:

$$K_0 = \frac{4\pi\omega - C_2}{2\pi}, \quad K_1 = 0 \tag{68}$$

Rewriting these equations,

$$R_0'(\phi) + DR_0(\phi) = Q_0 \tag{69}$$

$$D = \frac{C_2 - 4\pi\omega}{C_1}, \quad Q_0 = -\frac{K_0}{C_1}$$

$$R_1'(\phi) + DR_1(\phi) = Q_1(\phi) \tag{70}$$

$$Q_1(\phi) = \frac{[g(\phi)R_0(\phi)]'}{C_1}$$

we see:

$$R_0(\phi) = \frac{1}{2\pi}$$

$$[R_1(\phi)e^{D\phi}]' = Q_1(\phi)e^{D\phi}$$

We can use numerical methods to get the solution of $R_1(\phi)$ given $R_0(\phi)$ and for some choices of PRCs we can get exact expressions. For example, for the two term double sinusoidal form PRC (Equation 63) we get:

$$R_1(\phi) = \frac{1}{C_1}\left\{\frac{\tau}{\tau^2+1}\frac{\cos(\phi+a_2-a_1)-D\sin(\phi+a_2-a_1)}{1+D^2}+\frac{2b_1b_2\tau}{4\tau^2+1}\frac{2\cos(2\phi)-D\sin(2\phi)}{4+D^2}\right\} \tag{71}$$

DIFFERENT PRCS

Double sines form PRC

For the PRC:

$$\Delta_m(\theta) = \sin(a_m) - \sin(\theta + a_m) + b_m \sin(2\theta)$$

We have:

$$\begin{aligned} h_{mn}(s) &= \int_0^{2\pi} \Delta_m(\theta)\Delta_n(\theta+s)d\theta \\ &= 2\pi\sin(a_m)\sin(a_n) + \pi\cos(s-a_m+a_n) + b_m b_n \pi\cos(2s) \end{aligned} \tag{72}$$

$$\begin{aligned} g_{mn}(\phi) &= \int_0^{\infty} h_{mn}(s+\phi)e^{-\frac{s}{\tau}}ds \\ &= 2\pi\tau\sin(a_m)\sin(a_n) + \pi\tau\frac{\cos(\phi+a_n-a_m)-\tau\sin(\phi+a_n-a_m)}{\tau^2+1} \\ &\quad + b_m b_n \pi\tau\frac{\cos(2\phi)-2\tau\sin(2\phi)}{4\tau^2+1} \end{aligned} \tag{73}$$

$$g'_{mn}(\phi) = -\pi\tau\frac{\tau\cos(\phi+a_n-a_m)+\sin(\phi+a_n-a_m)}{\tau^2+1} - 2b_m b_n\pi\tau\frac{2\tau\cos(2\phi)+\sin(2\phi)}{4\tau^2+1} \tag{74}$$

$$\begin{aligned} g(\phi) &= g_{12}(\phi) + g_{21}(-\phi) \\ &= 4\pi\tau\sin(a_1)\sin(a_2) + 2\pi\tau\frac{\cos(\phi+a_2-a_1)}{\tau^2+1} + 2b_1b_2\pi\tau\frac{\cos(2\phi)}{4\tau^2+1} \end{aligned} \tag{75}$$

$$\begin{aligned} g'(\phi) &= g'_{12}(\phi) - g'_{21}(-\phi) \\ &= -2\pi\tau\frac{\sin(\phi+a_2-a_1)}{\tau^2+1} - 4b_1b_2\pi\tau\frac{\sin(2\phi)}{4\tau^2+1} \end{aligned} \tag{76}$$

$$C_1 = g_{11}(0) + g_{22}(0) = 2\pi\tau(\sin^2(a_1)+\sin^2(a_2)) + \frac{2\pi\tau}{\tau^2+1} + \frac{(b_1^2+b_2^2)\pi\tau}{4\tau^2+1} \tag{77}$$

$$C_2 = g'_{11}(0) - g'_{22}(0) = \frac{4\pi\tau^2(b_2^2-b_1^2)}{4\tau^2+1} \tag{78}$$

Exponential-sine form PRC

For empirical PRCs:

$$\begin{aligned} &\Delta_1(\theta) = a_1[\sin(b_1) - \sin(b_1+\theta)]e^{c_1(\theta-2\pi)} \\ &\Delta_2(\theta) = a_2[\sin(b_2) - \sin(b_2+\theta)]e^{c_2(\theta-2\pi)} \\ &\theta \in (0, 2\pi) \end{aligned}$$

We have:

$$h_{mn}(s) = B_1 \cdot e^{-c_m s}(hc_1, hv_1(s)) + B_0 \cdot e^{c_n s}(hc_0, hv_0(s)) \tag{79}$$

$$g_{mn}(\phi) = C_0 \cdot e^{\frac{\phi}{\tau}} - e^{-c_m\phi}(gc_1, gv_1(\phi)) - e^{c_n\phi}(gc_0, gv_0(\phi)) \tag{80}$$

Where $s \in [0, 2\pi)$ and $\phi \in [0, 2\pi)$, both have the period of 2π. Note also that (,) means the inner product of two vectors. The above quantities are defined as:

$$B_1 = a_m \cdot a_n \cdot e^{-2\pi c_n} \cdot [1 - e^{-2\pi c_m}]$$

$$B_0 = a_m \cdot a_n \cdot [1 - e^{-2\pi c_n}]$$

$$D_1 = c_m + \frac{1}{\tau}$$

$$D_0 = c_n - \frac{1}{\tau}$$

$$k_0 = \frac{\sin(b_m) \cdot \sin(b_n)}{c_m + c_n} - \frac{\sin(b_m)}{1 + (c_m + c_n)^2}[(c_m + c_n) \cdot \sin(b_n) - \cos(b_n)]$$

$$k_1 = \frac{1}{2(c_m + c_n)}, \quad k_2 = \frac{1}{4 + (c_m + c_n)^2}, \quad k_3 = -\frac{c_m + c_n}{2[4 + (c_m + c_n)^2]}$$

$$k_4 = \frac{(c_m + c_n) \cdot \sin(b_n)}{1 + (c_m + c_n)^2}, \quad k_5 = \frac{\sin(b_n)}{1 + (c_m + c_n)^2}$$

$$j_0 = \frac{\sin(b_m) \cdot \sin(b_n)}{c_m + c_n} - \frac{\sin(b_n)}{1 + (c_m + c_n)^2}[(c_m + c_n) \cdot \sin(b_m) - \cos(b_m)]$$

$$j_1 = \frac{1}{2(c_m + c_n)}, \quad j_2 = -\frac{1}{4 + (c_m + c_n)^2}, \quad j_3 = -\frac{c_m + c_n}{2[4 + (c_m + c_n)^2]}$$

$$j_4 = -\frac{(c_m + c_n) \cdot \sin(b_m)}{1 + (c_m + c_n)^2}, \quad j_5 = \frac{\sin(b_m)}{1 + (c_m + c_n)^2}$$

$$hc_1 = [k_0, k_1, k_2, k_3, k_4, k_5]$$

$$hc_0 = [j_0, j_1, j_2, j_3, j_4, j_5]$$

$$hv_1 = [1, \cos(s + b_n - b_m), \sin(s - b_m - b_n), \cos(s - b_m - b_n), \sin(s - b_m), \cos(s - b_m)]$$

$$hv_0 = [1, \cos(s + b_n - b_m), \sin(s + b_m + b_n), \cos(s + b_m + b_n), \sin(s + b_n), \cos(s + b_n)]$$

$$gc_1 = \frac{B_1}{1 + D_1^2}[-\frac{1 + D_1^2}{D_1}k_0, -D_1 k_1, k_1, k3 - D_1 k_2, -k_2 - D_1 k3, k5 - D_1 k_4, -k4 - D_1 k_5]$$

$$gc_0 = \frac{B_0}{1 + D_0^2}[\frac{1 + D_0^2}{D_0}j_0, D_0 j_1, j_1, j3 + D_0 j_2, -j_2 + D_0 j3, j5 + D_0 j_4, -j4 + D_0 j_5]$$

$$gv_1(\phi) = [1, \cos(\phi + b_n - b_m), \sin(\phi + b_n - b_m), \sin(\phi - b_m - b_n),$$
$$\cos(\phi - b_m - b_n), \sin(\phi - b_m), \cos(s - b_m)]$$

$$gv_0(\phi) = [1, \cos(\phi + b_n - b_m), \sin(\phi + b_n - b_m), \sin(\phi + b_m + b_n),$$
$$\cos(\phi + b_m + b_n), \sin(\phi + b_n), \cos(s + b_n)]$$

$$C_0 = \frac{e^{-2\pi\tau}}{1 - e^{-\frac{2\pi}{\tau}}}[(e^{-2\pi c_m} - 1) \cdot (gc_1, gv_1(0)) + (e^{2\pi c_n} - 1) \cdot (gc_0, gv_0(0))]$$

Fourier form PRC

For the Fourier form of the PRC:

$$\Delta_m(\theta) = \sum_{k=-\infty}^{\infty} a_{m,k} e^{ik\theta}$$

$$h_{mn}(s) = \int_0^{2\pi} \Delta_m(\theta)\Delta_n(\theta) ds$$
$$= \sum_{k_1, k_2} a_{m, k_1} a_{n, k_2} \int_0^{2\pi} e^{ik_1\theta} e^{ik_2(\theta + s)} d\theta$$

$$= 2\pi \sum_{k=-\infty}^{\infty} a_{m,k} a_{n,-k} e^{-iks} \tag{81}$$

$$g_{mn}(\phi) = \int_0^{\infty} h_{mn}(s+\phi) e^{-\frac{s}{\tau}} ds$$

$$= 2\pi \sum_{k=-\infty}^{\infty} a_{m,k} a_{n,-k} \int_0^{\infty} e^{-ik(s+\phi)} e^{-\frac{s}{\tau}} ds$$

$$= 2\pi \sum_{k=-\infty}^{\infty} \frac{a_{m,k} a_{n,-k}}{k^2 + \frac{1}{\tau^2}} (\frac{1}{\tau} - ik) e^{-ik\phi} \tag{82}$$

frontiers in
COMPUTATIONAL NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 14 August 2013
doi: 10.3389/fncom.2013.00108

Direct connections assist neurons to detect correlation in small amplitude noises

E. Bolhasani, Y. Azizi and A. Valizadeh*

Department of Physics, Institute for Advanced Studies in Basic Sciences, Zanjan, Iran

Edited by:
Ruben Moreno-Bote, Foundation Sant Joan de Deu, Spain

Reviewed by:
Germán Mato, Centro Atomico Bariloche, Argentina
Zachary P. Kilpatrick, University of Houston, USA

****Correspondence:***
A. Valizadeh, Department of Physics, Institute for Advanced Studies in Basic Sciences, Gava zang, PO Box 45195-1159, Zanjan, Iran
e-mail: valizade@iasbs.ac.ir

We address a question on the effect of common stochastic inputs on the correlation of the spike trains of two neurons when they are coupled through direct connections. We show that the change in the correlation of small amplitude stochastic inputs can be better detected when the neurons are connected by direct excitatory couplings. Depending on whether intrinsic firing rate of the neurons is identical or slightly different, symmetric or asymmetric connections can increase the sensitivity of the system to the input correlation by changing the mean slope of the correlation transfer function over a given range of input correlation. In either case, there is also an optimum value for synaptic strength which maximizes the sensitivity of the system to the changes in input correlation.

Keywords: correlation, correlation transfer, coupling, inhomogeneity, synchrony

1. INTRODUCTION

The recent advent of novel recording techniques has made it easier to simultaneously record from a large number of neurons and has provided new possibilities to relate population activity to coding and information processing in the brain (Greenberg et al., 2008; Cohen and Kohn, 2011). Many researchers suggest that studying the correlated activity of neurons in a population is essential for understanding how information is coded in the brain (Zohary et al., 1994; Abbott and Dayan, 1999; Nirenberg and Latham, 2003; Averbeck et al., 2006; Biederlack et al., 2006; Schneidman et al., 2006; Pillow et al., 2008). Correlated spiking of neurons contributes in several cognitive functions such as attention (Steinmetz et al., 2000), sensory coding (Christopher deCharms and Merzenich, 1996; Bair et al., 2001; Doiron et al., 2004; Galán et al., 2006; Schoppa, 2006) and discrimination (Stopfer et al., 1997; Kenyon et al., 2004), motor behavior (Maynard et al., 1999) and population coding (Sompolinsky et al., 2001; Averbeck et al., 2006; Josic et al., 2009). In addition to the functional effects of such correlations between populations of neurons on neural coding, understanding how different parameters such as biological, network or stimulus parameters tune them is eventually being revealed (Shadlen and Newsome, 1998; Binder and Powers, 2001; Moreno et al., 2002; Moreno-Bote and Parga, 2006; Tchumatchenko et al., 2010b; Rosenbaum and Josić, 2011b). Correlation between neuronal activities is measured frequently by pairwise correlation coefficients and spike count correlations, and the ability of a neuronal system to transfer correlation can be quantified by the correlation transfer function (CTF), which determines the relation between the output correlation of a system under stimulus and a specific input correlation (Doiron et al., 2006; Shea-Brown et al., 2008; Rosenbaum and Josić, 2011b).

A periodic common input on two (or more) uncoupled oscillators can cause coherent behavior when both oscillators lock to the external force (Pikovsky et al., 2003). A very common example is the control of circadian rhythms of humans/animals by the light-dark stimulation (Roberts, 2005). In case of noisy inputs the counterpart of the phenomena appears as stochastic synchronization (SS) which is a general topic that addresses the phenomenon of irregular phase locking between two noisy non-linear oscillators (Neiman et al., 1999). In nervous systems, cross-correlations can arise either from the presence of direct synaptic connections (Csicsvari et al., 1998; Barthó et al., 2004) or from shared inputs from the surrounding network or sensory layers (Binder and Powers, 2001; Türker and Powers, 2001, 2004). Effect of direct synaptic connections and common inputs have been widely studied, but these two sources of correlation can be present concurrently in many physical and biological systems and their interplay can result in quite interesting phenomena. Couplings can regulate the activity of noisy oscillators and less variability in neuronal dynamics emerges through synchronization in networks of coupled noisy oscillators (Ly and Ermentrout, 2010; Tabareau et al., 2010; Zilli and Hasselmo, 2010). Studies on the correlation of spike trains have reported increase and decrease of correlation due to the presence of excitatory and inhibitory synapses, respectively (Rosenbaum and Josić, 2011a; Ly et al., 2012). When delay in communication and type of excitabilty of neurons are taken into account, the generality of these results can be debated since both excitatory and inhibitory synapses can be sources of synchrony and may increase correlation in different parameter ranges (Vreeswijk et al., 1994; Wang et al., 2012; Sadeghi and Valizadeh, 2013). Regarding the type of excitability and categorizing couplings as synchronizing and desynchronizing, it has been shown that shared inputs and direct couplings can show cooperative or disruptive effects on the correlation of noisy coupled oscillators (Ly and Ermentrout, 2009).

Possible differences between intrinsic parameters of neurons causes the message from the environment to the system to be decoded differently by the system components. Another aim of the current study is to investigate how the correlation is transferred by two neurons when the neurons are not identical. In such

a heterogeneous system, the temporal symmetry of spike correlation is lost (Tchumatchenko et al., 2010b). We will show that with small amplitude stochastic inputs, even a slight inhomogeneity in the intrinsic parameters can lead to a large reduction of the pairwise correlation coefficient in the case of uncoupled neurons. As expected, the results depend on the time bins over which the correlation is calculated: spike count correlations over long time bins are less affected by the heterogeneity but synchrony—alignment of the action potential in small time bins—is tightly dependent on the homogeneity of the system.

We have shown that correlated inputs and direct connections can either show cooperative or disruptive effects in different ranges of parameters. For uncoupled neurons, correlation susceptibility increases by increasing the amplitude of noise for mildly correlated inputs (De La Rocha et al., 2007; Shea-Brown et al., 2008; Tchumatchenko et al., 2010b). We show that when direct connections are present between non-identical neurons, the mean susceptibility is not a monotonic function of the amplitude of the correlated noisy input anymore. Reminiscence of stochastic resonance phenomena, an intermediate noise amplitude in this case, leads to larger a sensitivity of the system to the changes in input correlation. We have also shown that with monosynaptic connections between two neurons, presence of inhomogeneity in the intrinsic firing rate of the neurons can enhance correlation of spike trains while for symmetric couplings, maximum correlation is seen for homogeneous system. Changing mismatch and synaptic strengths between two neurons, it is possible to change the functional form of the correlation transfer function to optimize the mean correlation susceptibility which is an indicator of the sensitivity of the system to the change of input correlation in different ranges. In this way, as the most important result of current study, we will show that with direct couplings it is possible to detect correlation in small amplitude noises by increasing the sensitivity of the system to the change of correlation in small amplitude noisy inputs.

2. MATERIALS AND METHODS

The system under investigation consists of two coupled leaky integrate and fire (LIF) neurons (Knight, 1972), subjected to correlated stochastic inputs (see **Figure 1**). Subthreshold dynamics of the LIF neuron obeys the following first order equation:

$$\tau_m \frac{dv_i}{dt} = V_{\text{rest}} - v_i + I_i + I_{ij}, \qquad (1)$$

in which v_i is a voltage-like variable for each neuron labeled by $i = 1, 2$ with $\tau_m = 20\,\text{ms}$ and $V_{\text{rest}} = -70\,\text{mV}$. A severe nonlinearity is imposed on the model by considering a threshold value $v_{th} = -54\,\text{mV}$. Whenever this value is reached, the neuron *spikes* and the voltage resets to $v_{\text{reset}} = -60\,\text{mV}$. [Parameters taken from Troyer and Miller (1997)]. The spikes of the neurons are recorded as $x_i(t) = \sum_m \delta(t - t_i^m)$ where t_i^m is the time of m th spike of the neuron i, and $\delta(x)$ is the Dirac delta function.

Each model neuron receives a synaptic current through the direct connection from the other neuron I_{ij}, and an external current I_i representing the sensory input or the effect of the surrounding networks. In the model equations, external current to the neuron i comprises a constant (dc) and a stochastic component with amplitude σ. The stochastic inputs are sum of a common component $\xi_c(t)$ and an individual component $\xi_i(t)$:

$$I_i(t) = (1 \pm \delta)I + \sigma\left[\sqrt{1-c}\xi_i(t) + \sqrt{c}\xi_c(t)\right], \qquad (2)$$

where $\xi_c(t)$ and $\xi_i(t)$ are mutually independent Gaussian stochastic processes with zero mean and unit variance $\langle \xi_i(t)\xi_j(t')\rangle = \delta_{ij}\delta(t - t')$. The parameter $c \in [0, 1]$ determines correlation of external currents which will be referred to as the input correlation. With the minimal model we used, inhomogeneity in the intrinsic activity rates is imposed by different constant currents which are chosen as $I_1 = (1 + \delta)I$ and $I_2 = (1 - \delta)I$, where δ is referred to as the parameter of inhomogeneity. With non-zero δ the neurons 1 and 2 will be the high frequency (fast) and low frequency (slow) neurons, respectively. The currents are chosen suprathreshold (>14 mV) such that the neurons fire periodically at vanishing noise. Note that in this mean driven regime presence of small amplitude noise results in small jitters in firing times and a narrow distribution of interspike intervals.

Neurons are pulse coupled. The neuron i receives a pulse by the strength Δ_{ij} every time the neuron j fires, so the synaptic current in Equation 1 can be written as $I_{ij} = \Delta_{ij}x_j(t)$ where the synaptic strength Δ_{ij} can be positive (excitatory) or negative (inhibitory). For convenience, we call the connections 21 and 12, the forward and backward connections, respectively. Although the external and synaptic inputs appear as currents, they are actually measured in units of the membrane potential (mV) since a factor of the membrane resistance has been absorbed into their definition.

Co-fluctuations in the activity of neurons are measured over a range of timescales (for a review see Cohen and Kohn, 2011). Spike count correlation is usually measured over the time scales from tens of milliseconds to seconds, while synchrony, that is almost precise alignment of the spikes, is measured over the time scale of the typical width of an action potential. It has been shown that spike count correlation over the small bins, bins of the order of one millisecond, can be largely determined by

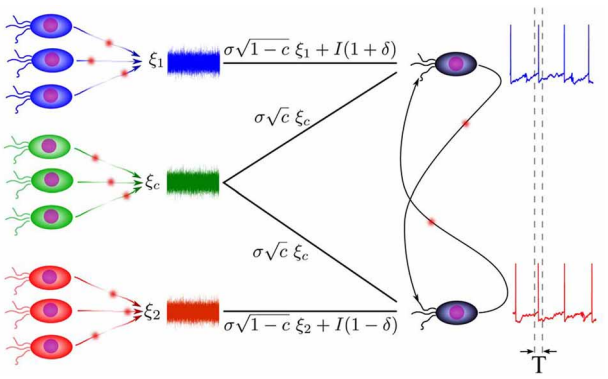


FIGURE 1 | Schematic representation of the model. Two neurons stimulated by common and independent components, are possibly connected together by direct excitatory synaptic connections. Correlation of spike trains is then calculated over time bins much smaller than the mean inter-spike intervals.

zero-lag conditional firing rate which quantifies exact synchrony (Tchumatchenko et al., 2010a). In this study we focus on synchrony, by describing spike counts and correlation coefficients in discrete bins of duration $T = 0.5$ ms. Correlation coefficient of spike counts $n_i(t) = \int_t^{t+T} x_i(s)ds$, is defined as the zero lag cross-correlation between n_1 and n_2:

$$\rho_T = \frac{\langle n_1(t)n_2(t)\rangle - \langle n_1(t)\rangle\langle n_2(t)\rangle}{\sqrt{\langle n_1(t)^2\rangle - \langle n_1(t)\rangle^2}\sqrt{\langle n_2(t)^2\rangle - \langle n_2(t)\rangle^2}}. \quad (3)$$

Dependence of the output correlation to the input correlation shows how correlation is transferred along neuronal layers in the nervous system (Rosenbaum and Josić, 2011a). With varying input correlation while other parameters are fixed, we compute $\rho_T(c)$, correlation of spike trains as a function of input correlation. To study sensitivity of correlation of output spike trains to the change of input correlation, we use *mean correlation susceptibility* (MCS), the mean slope of $\rho_T(c)$ in a given range of $c \in [c_1, c_2]$:

$$S_T(c_1, c_2) = \frac{\Delta\rho_T}{\Delta c}. \quad (4)$$

which shows ratio of the change of correlation of spike trains $\Delta\rho_T = \rho_T(c_2) - \rho_T(c_1)$ to the change of input correlation $\Delta c = c_2 - c_1$. For two identical neurons with no direct connection, this value is equal to one when it is evaluated over the full range of input correlation [0, 1].

3. RESULTS

We first present the results for two uncoupled neurons. In **Figure 2A** we have shown the cross-correlation coefficient as a function of the mismatch between intrinsic firing rates of neurons for low noise amplitude and different values of the input correlation. When there is no direct connection between the neurons, highly correlated inputs lead to a large output correlation in case of identical neurons. Even a small mismatch decreases the output correlation considerably if the noise is small amplitude. In this case, even common noises lead to a relatively low output correlation in the presence of a slight inhomogeneity (e.g., $\delta = 0.01$ in **Figure 2A**). For larger noise amplitudes, the output correlation is less sensitive to inhomogeneity (**Figure 2B**). The system is also less sensitive to inhomogeneity when the inputs are weakly correlated where both homogeneous and inhomogeneous systems show a small output correlation. In **Figures 2C,D** we have shown the correlation transfer function. It can be seen that while the slope of the correlation transfer function decreases with mismatch for all the values of input correlation, this dependence is only noticeable when inputs are highly (completely) correlated. Increasing the noise amplitude (while decreasing the constant input to avoid a change in the mean firing rate as explained below) makes the output correlation less sensitive to inhomogeneity, yet the maximum sensitivity to mismatch is observed for highly correlated inputs (**Figure 2D**).

To show how sensitive are the correlation of spike trains to the input correlation, in **Figure 2E** we have plotted MCS (mean slope

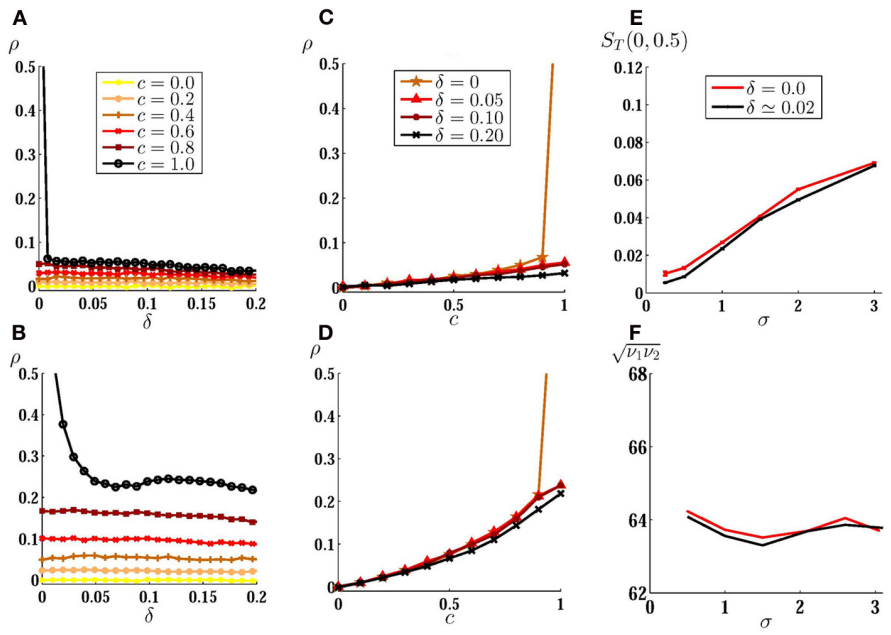


FIGURE 2 | Correlation of spike trains for two uncoupled neurons. (A) Correlation coefficient is plotted against inhomogeneity, the mismatch between input current of neurons, for different values of input correlation and low noise amplitude $\sigma = 1$ mV. In **(B)** the same results are shown for larger value of noise amplitude $\sigma = 5$ mV with the same mean firing rate as **(A)** (see materials and methods). **(C,D)** Correlation transfer function, which shows the dependence of correlation of spike trains to the input correlation, is plotted for different values of inhomogeneity for the same noise amplitudes as **(A)** and **(B)**. **(E)** Mean correlation susceptibility (MCS) is plotted for homogeneous and slightly inhomogeneous systems, as a function of noise amplitude, which shows the mean sensitivity of the output correlation to the change of input correlation over the range [0, 0.5]. In **(F)** the geometric mean of the firing rate of the two neurons is shown when σ is varied.

of $\rho_T(c)$ as described in materials and methods) as a function of the amplitude of the stochastic input for two uncoupled neurons over the range $c \in [0 - 0.5]$ for homogeneous ($\delta = 0$) and slightly inhomogeneous ($\delta = 0.02$) systems. The system shows low sensitivity to the change in input correlation for small amplitude noises and the sensitivity smoothly increases with noise amplitude. Also, the presence of inhomogeneity has negligible effect on the mean correlation suceptibility: as noted above, for uncoupled neurons effect of inhomogeneity is only significant when inputs are highly correlated and while MCS is calculated over a range of weakly correlated inputs, it is almost insensitive to small inhomogeneity. While increasing the amplitude of the fluctuations, we have decreased mean value of the input currents to keep the mean firing rate almost constant (∼64 Hz) as is shown in **Figure 2F**. In such a way the results observed in **Figure 2E** can not be attributed to the increase in firing rate which is known to increase the spike train correlation (De La Rocha et al., 2007; Shea-Brown et al., 2008). These results show that the correlation in small amplitude noises can not be suitably detected by a system of uncoupled neurons, whether the neurons have equal firing rates or their firing rates are different. To investigate the effect of direct couplings we have first considered a two neurons motif with just one unidirectional excitatory synapse. In many cases this configuration is favored when the synapses change through spike timing-dependent plasticity (Song et al., 2000). We considered an excitatory forward coupling from the high frequency neuron (as the presynaptic) to low frequency neuron (as the postsynaptic). In the absence of noise, any finite value of the forward coupling strength can lead to a zone of 1:1 synchrony, in which the dissimilar neurons fire in a causal master-slave fashion (Takahashi et al., 2009; Bayati and Valizadeh, 2012). In such causal limit the postsynaptic neuron fires immediately after receiving presynaptic stimulation (Woodman and Canavier, 2011; Wang et al., 2012). In our model delays in communication have been ignored, so in the causal 1:1 synchrony zones the postsynaptic neuron fires just one simulation time step after the firing of presynaptic neuron. Since the time bin on which the correlation is calculated contains several time steps (see materials and methods), such a causal master-slave firing leads to $\rho = 1$ (gray curves in **Figure 3**).

Stochastic inputs have non-trivial effects on the correlation of the spike trains of these two neurons. The output correlation is not a monotonically decreasing function of mismatch anymore, and in the presence of noise a small mismatch can increase the output correlation (**Figure 3A**). With zero mismatch, in the presence of one excitatory connection from neuron 1 to neuron 2 and in the absence of noise, the only stable state is the phase locked state in which neuron 2 fires one time step after neruon 1 (Bayati and Valizadeh, 2012). In the presence of noise this state looses stability as follows: because of the initial phase difference between the two neurons after master-slave firing (even though the phase difference is very small, just one time step), they respond slightly differently even to common noises. The different responses of the two neurons lead to a cumulative phase difference and if this phase difference results in the firing of neuron 2 before neuron 1

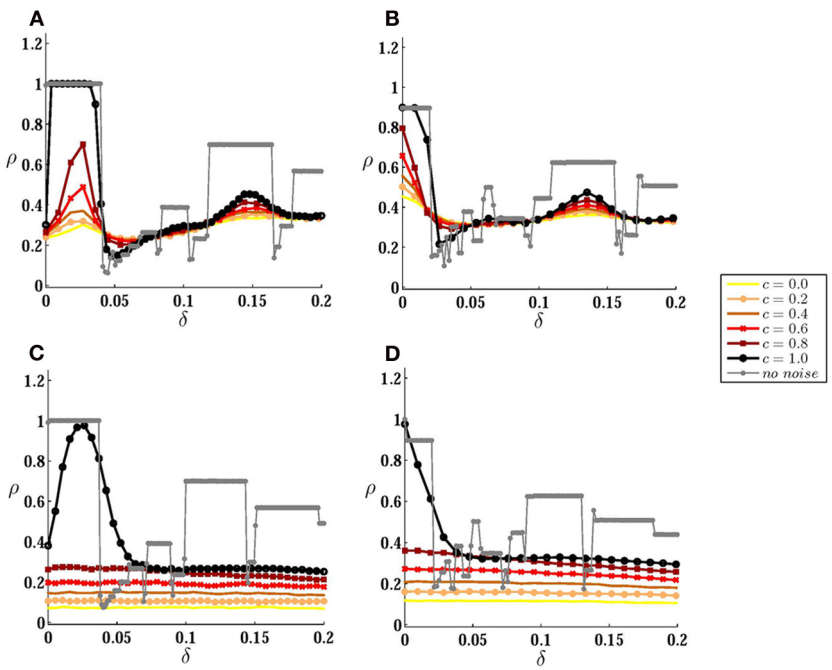


FIGURE 3 | Correlation of spike trains for coupled neurons. (A) Correlation coefficient is plotted against inhomogeneity for different values of input correlation, when the neurons are connected by a forward excitatory connection (from the high-frequency to the low-frequency neuron) of the strength $\Delta_{21} = 1$. **(B)** The same results are shown when the neurons are bidirectionally coupled by symmetric connections. In **(C)** and **(D)** the results are presented for larger noise amplitude $\sigma = 5$ mV. Noise amplitude in **(A)** and **(B)** is $\sigma = 1$ mV. The gray curves correspond to autonomous case when no stochastic input is present.

reaches threshold, the excitatory pulse from neuron 1 would be desynchronizing and makes the next firing of the two neurons further apart. The probability of the advancement of the phase of neuron 2 decreases in the presence of inhomogeneity (with $I_1 > I_2$), and with larger inhomogeneity it is less likely that the firing of neuron 2 (low frequency neuron) exceeds the firing of neuron 1 (high frequency neuron). When neuron 2 fires, before neuron 1 has reached the threshold, the excitatory pulse to the low frequency neuron will be synchronizing and if the voltage of neuron 2 is in the range $[v_{th} - \Delta_{21}, v_{th}]$ at the time of the firing of neuron 1, the neurons maintain causal master-slave firing. Further increasing the inhomogeneity lowers the probability of the voltage of the low frequency neuron reaching the range $[v_{th} - \Delta_{21}, v_{th}]$ at the time of the firing of the high frequency neuron, which results in the reduction of the spike trains correlation. A similar argument can explain the other notable rise and fall of the correlation which is seen in 1:2 locking zone of the noiseless system.

With symmetric bidirectional couplings, maximum correlation is obtained when the neurons are of the same firing rate (**Figure 3B**). When the neurons have equal firing rates (with $I_1 = I_2$) and in the absence of noise, each of the neurons can play the role of the master in a causal master-slave firing: in this case the connection from the master is synchronizing and the other connection has a desynchronizing effect (Bayati and Valizadeh, 2012). In the presence of small amplitude noise, the system can maintain causal locking by interchanging the role of two connections as synchronizing and desynchronizing. Suppose the firing of neuron 1 (master) is followed by the firing of neuron 2 (slave). Firing of neuron 2 exerts an excitatory pulse on neuron 1 but the phase advance of neuron 1 is relatively small because of the weak response of the LIF neuron at the beginning of its cycle (Mirollo and Strogatz, 1990). So it is probable that neuron 2 fires before neuron 1 reaches the threshold, then the excitatory pulse to neuron 1 would be synchronizing and neuron 1 fires immediately at the time it receives the pulse if its voltage is within the range $[v_{th} - \Delta_{12}, v_{th}]$ (note that the argument holds also in the presence of an absolute refractory period where the desynchonizing pulse from the slave neuron is ineffective). In the presence of inhomogeneity, it is the high frequency neuron that more probably plays the role of the master in a locked causal firing in the absence of the noise. In this case, in the presence of noise, inhomogeneity increases the probability that the voltage of low frequency neuron takes a value outside the range $[v_{th} - \Delta_{21}, v_{th}]$ at the time of the firing of the high frequency neuron, which reduces the correlation of spike trains as can be seen in **Figure 3B** for small values of inhomogeneity. For larger values of inhomogeneity, a bump can be seen again which belongs to the other main locking zone of the system in the absence of noise.

Intuitively, the relative amplitudes of noise and recurrent stimulations determine the behavior of the system and the most notable results can be expected when these two sources are of the same order, i.e., when neither the external noises nor recurrent stimulations are dominant. The results of **Figures 3A,B** are produced in this regime. For larger values of the noise amplitude, qualitative behavior of the system becomes more similar to the uncoupled system as shown in **Figures 3C,D**. For all partially correlated inputs, correlation of the spike trains is independent of the inhomogeneity and no signature of the locking zones is observed in the presence of large amplitude noises. It is only for common noise ($\gamma = 1$) that the effect of the unidirectional direct connection can be seen in the presence of strong noise in the region of the main locking zone.

In **Figure 4** we have plotted correlation of spike trains as a function of input correlation to inspect the effect of changing the correlation of the stochastic inputs on the correlation of the spike trains for a fixed value of the synaptic strength. When the noise amplitude is not large, depending on the mismatch, different dependencies of the output correlation to the input correlation can be observed (**Figures 4A,B**. Notably with changing mismatch it is possible to generate, for example, a system with higher sensitivity to the input correlation in different ranges of input correlation, or a negative slope $\rho_T(c)$. Comparing with the results of **Figure 3** it can be deduced that high sensitivities on the input correlation is seen on the main locking zone (where the neurons are causally locked in 1:1 zone in the absence of the noise), and a negative slope is seen between two main locking zones. Again, as can be seen in **Figures 4C,D**, strong noises wash the signature of the direct couplings, and $\rho_T(c)$ for large amplitude noises is qualitatively similar to the uncoupled neurons.

Impact of direct connections on the detection of the input correlation of low amplitude noisy inputs is more apparent in a plot of MCS. In **Figure 5A** we have plotted $S_T(0, 0.5)$ as a function of noise amplitude for several values of synaptic strength, for unidirectionally coupled neurons and in the presence of a small mismatch in the intrinsic firing rates. As shown in **Figure 5A**, a forward monosynaptic connection (from high frequency to low frequency neuron) can considerably change the performance of the heterogeneous system in detecting variable input correlation.

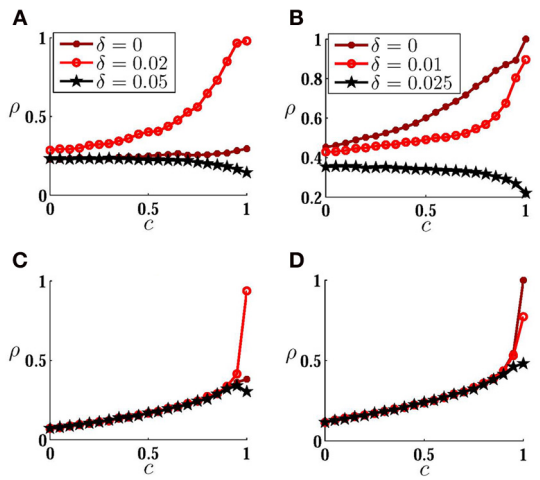


FIGURE 4 | Correlation transfer for coupled neurons. (A,B) Correlation of spike trains ρ_T is plotted versus input correlation c for different values of inhomogeneity, when the neurons are connected by a forward excitatory connection **(A)** and by symmetric bidirectional couplings **(B)** . In **(C)** and **(D)** the results are presented for larger noise amplitudes. All the parameters are the same as those in **Figure 3**.

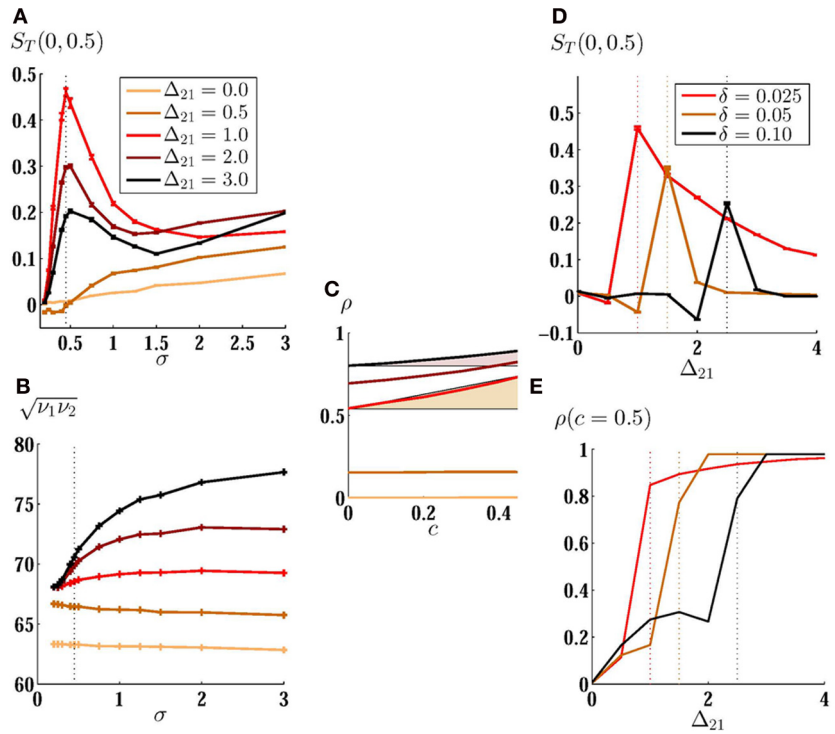


FIGURE 5 | Mean correlation susceptibility for coupled neurons. (A) MCS is plotted versus noise amplitude for two unidirectionally coupled non-identical neurons ($\delta = 0.02$). The results are shown for different values of synaptic strength. Maximum value of sensitivity to low amplitude noises can be obtained by $\Delta_{21} = 1$. In **(B)** and **(C)** the firing rate of the neurons and $\rho_T(c)$ are shown for the corresponding curves in **(A)**, respectively. Shadings in **(C)** are guide to eye for a comparison of the mean slope of the $\rho_T(c)$ for two different values of synaptic strength. **(D)** MCS is shown as a function of synaptic strength for different value of mismatch. The optimum value for synaptic strength grows for larger mismatch. Correlation of the spike trains for $c = 0.5$ is shown in **(E)**. It can be seen that the correlation saturates when coupling constant is increased. Vertical dotted lines are plotted to show where the mean sensitivity is maximized.

In an intermediate synaptic strength ($\Delta_{21} = 1$) MCS shows a faster growth and a higher maximum in relatively small amplitude noise. Further increasing of the synaptic strength or the noise amplitude reduces the performance of the system in the detection of the input correlation. With very large noise amplitudes, the effect of the direct connections is washed out and all the curves, including that of the uncoupled neurons, merge together and the MCS smoothly increases with noise amplitude.

Overall increase of the correlation of the spike trains is an intuitive expectation when direct excitatory couplings are present in the systems (although this can be dependent on the type of excitability of the neurons). But how can direct connections increase the sensitivity to the changes in input correlation? In **Figure 5B** we have shown the geometric mean of the firing rate of the two neurons $\sqrt{\nu_1 \nu_2}$ for the curves plotted in **Figure 5A**. Note that ν_2 may be different from the intrinsic firing rate of neuron 2 because of the presence of an excitatory afferent synapse. The results show that the increase in the mean correlation susceptibility cannot be attributed to the increase of the mean firing rate of neurons, since then, larger coupling constants would lead to more sensitivity as they increase the mean firing rate of the system. A simple explanation can be found in **Figure 5C**: the degree of amplification of the output correlation depends on the input correlation. A suitable choice of the synaptic strength would result in more amplification for higher input correlations and would increase the slope of $\rho_T(c)$. Increasing the synaptic strength further, decreases the sensitivity due to the saturation of the correlation of the spike trains for the upper bound of the input correlation. In calculating MCS we have considered the range [0, 0.5] for the input correlation. Reducing the upper bound of this range increases the synaptic strength which saturates the correlation of the spike trains, so the synaptic strength which gives the maximum sensitivity increases with decreasing the range over which the mean sensitivity is calculated.

The *best* synaptic strength, which maximizes sensitivity, depends also on the mismatch between the intrinsic firing rate of the neurons as can be implicitly deduced from the results shown in **Figures 3A,B**. In **Figure 5D** we have shown MCS as a function of the strength of the forward unidirectional coupling for three values of mismatch. Optimum value of synaptic strength is larger when the intrinsic firing rate of the neurons are more different. Plots of the spike train correlation ρ for upper limiting value of the input correlation $c = 0.5$ again shows that the maximum mean sensitivity in this range is obtained when the spike train correlation is not saturated for the upper bound of the range of c (**Figure 5E**).

All the results presented in this study have focused on the degree of zero-lag synchrony which is measured by the zero-lag cross correlation of the binned spike trains with small bin size. In the presence of an inhomogeneity and with asymmetric direct connections, it is possible that the maximum correlation of the spike trains appears in non-zero lag. In **Figure 6** we have shown the cross-correlation coefficient of spike trains as a function of the time lag for three values of noise strength and two values of the input correlation ($c = 0$ and $c = 0.5$). It can be seen that the maximum cross correlation for all the values appears in zero time lag (more precisely at a time lag equal to one simulation time step). Presence of other maximums is an indicator of almost periodic firing of the neurons which arises from the suprathreshold mean and the small amplitude stochastic fluctuations of the input current. Results in **Figure 6** are presented for one forward unidirectional coupling and sample values of inhomogeneity and synaptic strength. The results for other parameters are similar while the system is in the main locking zone in the absence of noise. This result shows a drawback of the simplified models we have used: LIF neurons with pulsatile instantaneous couplings can be synchronized with zero phase lag even in the presence of frequency mismatch, which is revealed as a maximum in correlation at zero lag (one simulation time step) when a small amplitude noise is added. Both mismatch and delay (synaptic and axonal) can be source of phase lag, when the neurons are modeled by limit cycle oscillators and more realistic models are used for synaptic currents. Our results are still valid when such phase lags are small, of the order of the time bins in the calculation of the correlation.

Above results were obtained for bidirectional symmetric couplings or for one unidirectional coupling. To find the *best* configuration through which direct couplings can improve the performance of the system in the detection of a variable input correlation, we have tested mutual couplings with different ratios of forward Δ_{21} and backward Δ_{12} connections. While the synaptic cost (sum of two synaptic strengths) is kept constant, different configurations can be designed by changing the ratio of the coupling constants $r = \Delta_{21}/\Delta_{12}$ (**Figures 7A,B**). In the absence of mismatch, the best configuration is that which preserves symmetry, i.e., the best performance results with equal forward and backward couplings. On the other hand, in the presence of mismatch, an asymmetric arrangement of couplings in which the forward coupling (from the high frequency neuron) is larger, improves the performance of the system. Interestingly, asymmetric excitatory couplings in favor of backward coupling (from the low frequency neuron), significantly decreases the sensitivity of the system since it plays the role of desynchronizing coupling as discussed above.

4. DISCUSSION

Both direct connections and common inputs can be sources of the correlated activity of neurons in the nervous system. Effect of direct connections is widely studied as a general problem in dynamical systems and in particalur in nervous systems (Kuramoto, 1991; Strogatz and Mirollo, 1991; Abbott and van Vreeswijk, 1993). Stochastic inputs are usually a source of temporal disorder but spatial order can be induced in a neuronal pool when the neurons share stochastic inputs from common sources (Binder and Powers, 2001; Türker and Powers, 2001, 2004). Because of the possible cooperative/competitive effects of common inputs and direct connections, interesting results can be

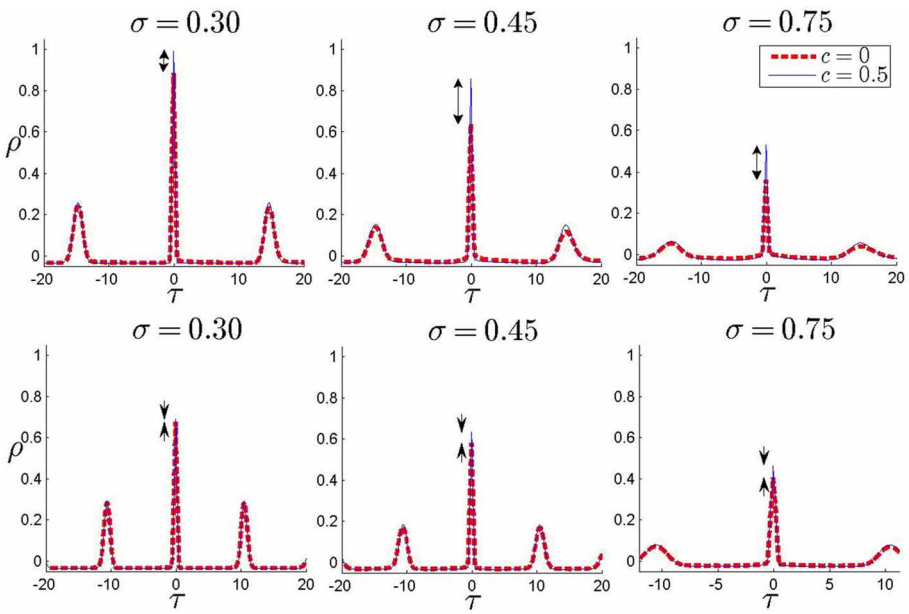


FIGURE 6 | Correlation coefficient for non-zero time lags. In each panel, correlation coefficient is plotted against time lag for two values of input correlation $c = 0$ and $c = 0.5$. The results are shown for three different noise amplitudes (shown above each panel) and two different values of inhomogeneity parameter, $\delta = 0.02$ in upper panels and $\delta = 0.15$ in lower panels. One unidirectional connection of strength $\Delta_{21} = 1$ is present from the high-frequency to the low-frequency neuron. The difference between two curves at zero lag gives MCS which has been shown in the plots by arrows.

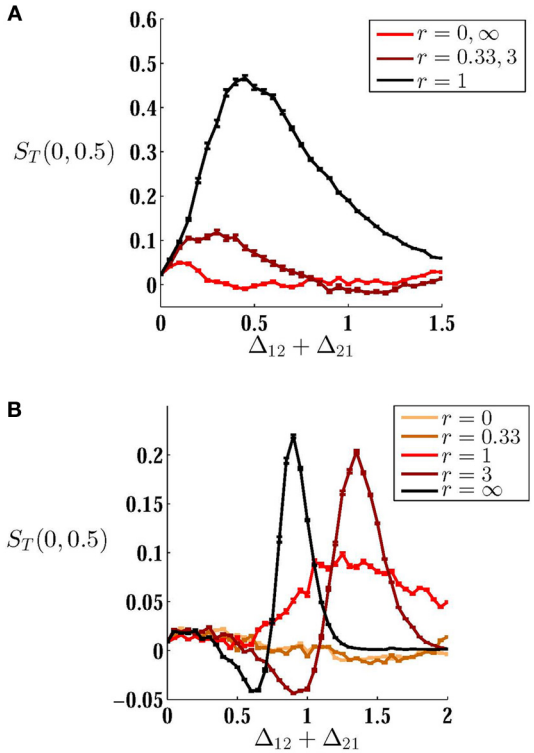


FIGURE 7 | How the configuration of connections affects sensitivity. **(A,B)** MCS is plotted against sum of synaptic weights for **(A)** homogeneous system $\delta = 0$, and **(B)** inhomogenous system $\delta = 0.1$. Different curves are plotted for different ratios r of forward and backward couplings indicated in the legends. When the neurons have equal intrinsic firing rates, symmetric configuration $r = 1$ shows the best performance with a suitable choice of synaptic strengths. For inhomogeneous case when the imbalance of couplings is in favor of forward coupling (from the high frequency to low frequency neuron) the sensitivity is considerably improved. When the backward coupling is larger $r < 1$, the system performance is quite poor. As is shown in axes labels, MCS is calculated over the range [0, 0.5] of input correlation.

expected when they are concurrently present in a system (Ostojic et al., 2009; Ly and Ermentrout, 2010; Tabareau et al., 2010; Zilli and Hasselmo, 2010; Rosenbaum and Josić, 2011a; Ly et al., 2012). In this study we have numerically inspected the effect of correlated stochastic inputs on the correlation of spike trains of two coupled LIF neurons. We have mainly focused on the correlation of spike trains when correlated small amplitude noises were imposed on a system of two coupled neurons, and the neurons were regularly and synchronously firing in the absence of noise. We have shown that such a system shows high sensitivity to the changes of input correlation, and therefore can be a suitable detector of the correlation in small amplitude noises. To study the system in a more general framework, we have considered neurons with different intrinsic firing rates. We have assumed neurons have equal membrane time constants, and inhomogeneity is imposed on the system by feeding the neurons with unequal suprathreshold constant currents. The inhomogeneity, determined by the difference in the mean input currents, along with synaptic strengths are the key-parameters that specify the response of the system to stochastic inputs.

While for uncoupled neurons the output correlation is a monotonically decreasing function of inhomogeneity, for coupled neurons with low noise amplitudes, spike trains correlation can be increased by increasing inhomogeneity in some ranges. This result holds for sufficiently small noise amplitudes and the system inherits this property from *n*:*m* locking zones for the autonomous system when there is no stochastic input present. This introduces inhomogeneity as an important parameter with non-trivial impact on the correlation of spike trains in coupled systems.

Another feature of the system is that the two sources of correlation, correlated inputs and direct excitatory connections, do not necessarily cooperate in the formation of correlated spike trains. For uncoupled neurons output correlation is a monotonically increasing function of input correlation and for weakly correlated inputs, the slope decreases with lowering noise amplitude (De La Rocha et al., 2007; Shea-Brown et al., 2008) and with increasing mismatch. With different choices of the synaptic strengths and the inhomogeneity, it is possible to change functional form of correlation transfer (dependence of output correlation to the input correlation) and design a system with different sensitivity to the input correlation. In particular, it is possible to design a system with negative mean slope of correlation transfer, showing a case with destructive effect of common noises on the correlation of spike trains, or a system with maximum sensitivity to the changes in input correlation in a given range by maximizing the slope of correlation transfer. The latter proposes that direct connections can increase the sensitivity of the system to the correlation of the neuron's stochastic inputs, especially when the noises are small amplitude. We have further shown that for a homogeneous system (where the neurons have equal intrinsic firing rates), the best configuration of the couplings which maximizes the mean sensitivity of the system in a given range, is a symmetric configuration with equal coupling constants. On the other hand, in the presence of inhomogeneity, an asymmetric configuration in which the synaptic constant from the high frequency neuron to the low frequency neuron is larger, improves the sensitivity. In either case, there is an optimum value of the synaptic constant which maximizes the sensitivity.

Competitive learning through conventional spike timing-dependent plasticity (STDP) in feed-forward networks leads to the potentiation of the synapses which convey correlated data and depression of those with uncorrelated activity (Babadi and Abbott, 2010). How does STDP change the lateral connections transverse to the path of data flow? It has been shown that in the recurrent networks, asymmetric connections arise through STDP and in the presence of inhomogeneity, such an asymmetric change is in favor of the connection from the high frequency to the low frequency neuron (Takahashi et al., 2009; Bayati and Valizadeh, 2012). Our results show that asymmetric connections can enhance the performance of inhomogeneous systems in the detection of input correlation, and interestingly such an optimum configuration of connections emerges through STDP (with asymmetric profile) in inhomogeneous neuronal pools (Bayati and Valizadeh, 2012).

Type of neuronal excitability can also affect the correlation transfer in neuronal pools (Galán et al., 2008; Abouzeid and Ermentrout, 2009; Barreiro et al., 2010). Phase resetting curve characterizes how small perturbations influence the oscillator's subsequent timing or phase. It has been recently shown that uncoupled type-II neurons with both negative and positive regions in their PRC transfer correlations more faithfully when the correlation is calculated over short time bins (Abouzeid and Ermentrout, 2011). Since the phase of a LIF neuron always advances in response to the external pulses, the results for LIF neurons are likely to apply for type-I neurons.

Correlation of spike trains over such small time bins that we have used $T = 0.5$ ms, is a measure of (almost) precise alignment of the action potentials. Similar results were obtained when we repeated the experiments with $T = 1$ ms but we expect qualitatively different results when the correlation of the spike counts is measured over the time scales comparable, or larger than the mean inter-spike interval. Less sensitivity to the inhomogeneity is expected when the correlation is evaluated over large time bins, but the effect of direct couplings warrants further studies to find out if correlation in small amplitude stochastic inputs can be revealed in co-variation of spike trains of coupled neurons over large time scales.

ACKNOWLEDGMENTS

Authors gratefully acknowledge Bahman Farnoudi for proof reading of the manuscript, and the reviewers Germán Mato and Zachary P. Kilpatrick, for careful reading of the manuscript and giving valuable comments.

REFERENCES

Abbott, L., and Dayan, P. (1999). The effect of correlated variability on the accuracy of a population code. *Neural Comput.* 11, 91–101. doi: 10.1162/089976699300016827

Abbott, L., and van Vreeswijk, C. (1993). Asynchronous states in networks of pulse-coupled oscillators. *Phys. Rev. E* 48:1483. doi: 10.1103/PhysRevE.48.1483

Abouzeid, A., and Ermentrout, B. (2009). Type-ii phase resetting curve is optimal for stochastic synchrony. *Phys. Rev. E* 80:011911. doi: 10.1103/PhysRevE.80.011911

Abouzeid, A., and Ermentrout, B. (2011). Correlation transfer in stochastically driven neural oscillators over long and short time scales. *Phys. Rev. E* 84:061914. doi: 10.1103/PhysRevE.84.061914

Averbeck, B. B., Latham, P. E., and Pouget, A. (2006). Neural correlations, population coding and computation. *Nat. Rev. Neurosci.* 7, 358–366. doi: 10.1038/nrn1888

Babadi, B., and Abbott, L. F. (2010). Intrinsic stability of temporally shifted spike-timing dependent plasticity. *PLoS Comput. Biol.* 6:e1000961. doi: 10.1371/journal.pcbi.1000961

Bair, W., Zohary, E., and Newsome, W. T. (2001). Correlated firing in macaque visual area mt: time scales and relationship to behavior. *J. Neurosci.* 21, 1676–1697.

Barreiro, A. K., Shea-Brown, E., and Thilo, E. L. (2010). Time scales of spike-train correlation for neural oscillators with common drive. *Phys. Rev. E* 81:011916. doi: 10.1103/PhysRevE.81.011916

Barthó, P., Hirase, H., Monconduit, L., Zugaro, M., Harris, K. D., and Buzsáki, G. (2004). Characterization of neocortical principal cells and interneurons by network interactions and extracellular features. *J. Neurophysiol.* 92, 600–608. doi: 10.1152/jn.01170.2003

Bayati, M., and Valizadeh, A. (2012). Effect of synaptic plasticity on the structure and dynamics of disordered networks of coupled neurons. *Phys. Rev. E* 86:011925. doi: 10.1103/PhysRevE.86.011925

Biederlack, J., Castelo-Branco, M., Neuenschwander, S., Wheeler, D. W., Singer, W., and Nikolić, D. (2006). Brightness induction: rate enhancement and neuronal synchronization as complementary codes. *Neuron* 52, 1073–1083. doi: 10.1016/j.neuron.2006.11.012

Binder, M. D., and Powers, R. K. (2001). Relationship between simulated common synaptic input and discharge synchrony in cat spinal motoneurons. *J. Neurophysiol.* 86, 2266–2275.

Christopher deCharms, R., and Merzenich, M. M. (1996). Primary cortical representation of sounds by the coordination of action-potential timing. *Nature* 381, 13.

Cohen, M. R., and Kohn, A. (2011). Measuring and interpreting neuronal correlations. *Nat. Neurosci.* 14, 811–819. doi: 10.1038/nn.2842

Csicsvari, J., Hirase, H., Czurko, A., and Buzsáki, G. (1998). Reliability and state dependence of pyramidal cell–interneuron synapses in the hippocampus: an ensemble approach in the behaving rat. *Neuron* 21, 179–189. doi: 10.1016/S0896-6273(00)80525-5

De La Rocha, J., Doiron, B., Eric Shea-Brown, K. J., and Reyes, A. (2007). Correlation between neural spike trains increases with firing rate. *Nature* 448, 802–806. doi: 10.1038/nature06028

Doiron, B., Lindner, B., Longtin, A., Maler, L., and Bastian, J. (2004). Oscillatory activity in electrosensory neurons increases with the spatial correlation of the stochastic input stimulus. *Phys. Rev. Lett.* 93:48101. doi: 10.1103/PhysRevLett.93.048101

Doiron, B., Rinzel, J., and Reyes, A. (2006). Stochastic synchronization in finite size spiking networks. *Phys. Rev. E* 74:030903. doi: 10.1103/PhysRevE.74.030903

Galán, R. F., Ermentrout, G. B., and Urban, N. N. (2008). Optimal time scale for spike-time reliability: theory, simulations, and experiments. *J. Neurophysiol.* 99, 277–283. doi: 10.1152/jn.00563.2007

Galán, R. F., Fourcaud-Trocmé, N., Ermentrout, G. B., and Urban, N. N. (2006). Correlation-induced synchronization of oscillations in olfactory bulb neurons. *J. Neurosci.* 26, 3646–3655. doi: 10.1523/JNEUROSCI.4605-05.2006

Greenberg, D. S., Houweling, A. R., and Kerr, J. N. (2008). Population imaging of ongoing neuronal activity in the visual cortex of awake rats. *Nat. Neurosci.* 11, 749–751. doi: 10.1038/nn.2140

Josic, K., Shea-Brown, E., Doiron, B., and De La Rocha, J. (2009). Stimulus-dependent correlations and population codes. *Neural Comput.* 21, 2774–2804. doi: 10.1162/neco.2009.10-08-879

Kenyon, G. T., Theiler, J., George, J. S., Travis, B. J., and Marshak, D. W. (2004). Correlated firing improves stimulus discrimination in a retinal model. *Neural Comput.* 16, 2261–2291. doi: 10.1162/0899766041941916

Knight, B. W. (1972). The relationship between the firing rate of a single neuron and the level of activity in a population of neurons experimental evidence for resonant enhancement in the population response. *J. Gen. Physiol.* 59, 767–778. doi: 10.1085/jgp.59.6.767

Kuramoto, Y. (1991). Collective synchronization of pulse-coupled oscillators and excitable units. *Phys. D Nonlin. Phenom.* 50, 15–30. doi: 10.1016/0167-2789(91)90075-K

Ly, C., and Ermentrout, G. B. (2009). Synchronization dynamics of two coupled neural oscillators receiving shared and unshared noisy stimuli. *J. Comput. Neurosci.* 26, 425–443. doi: 10.1007/s10827-008-0120-8

Ly, C., and Ermentrout, G. B. (2010). Coupling regularizes individual units in noisy populations. *Phys. Rev. E* 81:011911. doi: 10.1103/PhysRevE.81.011911

Ly, C., Middleton, J. W., and Doiron, B. (2012). Cellular and circuit mechanisms maintain low spike co-variability and enhance population coding in somatosensory cortex. *Front. Comput. Neurosci.* 6:7. doi: 10.3389/fncom.2012.00007

Maynard, E., Hatsopoulos, N., Ojakangas, C., Acuna, B., Sanes, J., Normann, R., and Donoghue, J. (1999). Neuronal interactions improve cortical population coding of movement direction. *J. Neurosci.* 19, 8083–8093.

Mirollo, R. E., and Strogatz, S. H. (1990). Synchronization of pulse-coupled biological oscillators. *SIAM J. Appl. Math.* 50, 1645–1662. doi: 10.1137/0150098

Moreno, R., de La Rocha, J., Renart, A., and Parga, N. (2002). Response of spiking neurons to correlated inputs. *Phys. Rev. Lett.* 89:288101. doi: 10.1103/PhysRevLett.89.288101

Moreno-Bote, R., and Parga, N. (2006). Auto-and crosscorrelograms for the spike response of leaky integrate-and-fire neurons with slow synapses. *Phys. Rev. Lett.* 96:28101. doi: 10.1103/PhysRevLett.96.028101

Neiman, A., Schimansky-Geier, L., Moss, F., Shulgin, B., and Collins,

J. J. (1999). Synchronization of noisy systems by stochastic signals. *Phys. Rev. E* 60:284. doi: 10.1103/PhysRevE.60.284

Nirenberg, S., and Latham, P. E. (2003). Decoding neuronal spike trains: How important are correlations? *Proc. Natl. Acad. Sci. U.S.A.* 100, 7348–7353. doi: 10.1073/pnas.1131895100

Ostojic, S., Brunel, N., and Hakim, V. (2009). How connectivity, background activity, and synaptic properties shape the cross-correlation between spike trains. *J. Neurosci.* 29, 10234–10253. doi: 10.1523/JNEUROSCI.1275-09.2009

Pikovsky, A., Rosenblum, M., and Kurths, J. (2003). *Synchronization: A Universal Concept in Nonlinear Sciences*, Vol. 12. Cambridge: Cambridge university press.

Pillow, J. W., Shlens, J., Paninski, L., Sher, A., Litke, A. M., Chichilnisky, E., et al. (2008). Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature* 454, 995–999. doi: 10.1038/nature07140

Roberts, J. E. (2005). Update on the positive effects of light in humansâĂă. *Photochem. Photobiol.* 81, 490–492. doi: 10.1562/2004-12-02-IR-391.1

Rosenbaum, R., and Josić, K. (2011a). Mechanisms that modulate the transfer of spiking correlations. *Neural Comput.* 23, 1261–1305. doi: 10.1162/NECO_a_00116

Rosenbaum, R., and Josić, K. (2011b). Membrane potential and spike train statistics depend distinctly on input statistics. *Phys. Rev. E* 84:051902. doi: 10.1103/PhysRevE.84.051902

Sadeghi, S., and Valizadeh, A. (2013). Synchronization of delayed coupled neurons in presence of inhomogeneity. *J. Comput. Neurosci.* 35, 1–12.

Schneidman, E., Berry, M. J., Segev, R., and Bialek, W. (2006). Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature* 440, 1007–1012. doi: 10.1038/nature04701

Schoppa, N. E. (2006). Synchronization of olfactory bulb mitral cells by precisely timed inhibitory inputs. *Neuron* 49, 271–283. doi: 10.1016/j.neuron.2005.11.038

Shadlen, M. N. and Newsome, W. T. (1998). The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. *J. Neurosci.* 18, 3870–3896.

Shea-Brown, E., Josić, K., de La Rocha, J., and Doiron, B. (2008). Correlation and synchrony transfer in integrate-and-fire neurons: basic properties and consequences for coding. *Phys. Rev. Lett.* 100:108102. doi: 10.1103/PhysRevLett.100.108102

Sompolinsky, H., Yoon, H., Kang, K., and Shamir, M. (2001). Population coding in neuronal systems with correlated noise. *Phys. Rev. E* 64:051904. doi: 10.1103/PhysRevE.64.051904

Song, S., Miller, K. D., and Abbott, L. F. (2000). Competitive hebbian learning through spike-timing-dependent synaptic plasticity. *Nat. Neurosci.* 3, 919–926. doi: 10.1038/78829

Steinmetz, P. N., Roy, A., Fitzgerald, P., Hsiao, S., Johnson, K., and Niebur, E. (2000). Attention modulates synchronized neuronal firing in primate somatosensory cortex. *Nature* 404, 131–133. doi: 10.1038/35004588

Stopfer, M., Bhagavan, S., Smith, B. H., and Laurent, G. (1997). Impaired odour discrimination on desynchronization of odour-encoding neural assemblies. *Nature* 390, 70–73. doi: 10.1038/36335

Strogatz, S. H. and Mirollo, R. E. (1991). Stability of incoherence in a population of coupled oscillators. *J. Statis. Phys.* 63, 613–635. doi: 10.1007/BF01029202

Tabareau, N., Slotine, J.-J., and Pham, Q.-C. (2010). How synchronization protects from noise. *PLoS Comput. Biol.* 6:e1000637. doi: 10.1371/journal.pcbi.1000637

Takahashi, Y. K., Kori, H., and Masuda, N. (2009). Self-organization of feed-forward structure and entrainment in excitatory neural networks with spike-timing-dependent plasticity. *Phys. Rev. E* 79:051904. doi: 10.1103/PhysRevE.79.051904

Tchumatchenko, T., Geisel, T., Volgushev, M., and Wolf, F. (2010a). Signatures of synchrony in pairwise count correlations. *Front. Comput. Neurosci.* 4:1. doi: 10.3389/neuro.10.001.2010

Tchumatchenko, T., Malyshev, A., Geisel, T., Volgushev, M., and Wolf, F. (2010b). Correlations and synchrony in threshold neuron models. *Phys. Rev. Lett.* 104:58102. doi: 10.1103/PhysRevLett.104.058102

Troyer, T. W., and Miller, K. D. (1997). Physiological gain leads to high isi variability in a simple model of a cortical regular spiking cell. *Neural Comput.* 9, 971–983. doi: 10.1162/neco.1997.9.5.971

Türker, K., and Powers, R. (2001). Effects of common excitatory and inhibitory inputs on motoneuron synchronization. *J. Neurophysiol.* 86, 2807–2822.

Türker, K., and Powers, R. (2004). The effects of common input characteristics and discharge rate on synchronization in rat hypoglossal motoneurones. *J. Physiol.* 541, 245–260. doi: 10.1113/jphysiol.2001.013097

Vreeswijk, C., Abbott, L. F., and Bard Ermentrout, G. (1994). When inhibition not excitation synchronizes neural firing. *J. Comput. Neurosci.* 1, 313–321. doi: 10.1007/BF00961879

Wang, S., Chandrasekaran, L., Fernandez, F. R., White, J. A., and Canavier, C. C. (2012). Short conduction delays cause inhibition rather than excitation to favor synchrony in hybrid neuronal networks of the entorhinal cortex. *PLoS Compu. Biol.* 8:e1002306. doi: 10.1371/journal.pcbi.1002306

Woodman, M. M., and Canavier, C. C. (2011). Effects of conduction delays on the existence and stability of one to one phase locking between two pulse-coupled oscillators. *J. Comput. Neurosci.* 31, 401–418. doi: 10.1007/s10827-011-0315-2

Zilli, E. A., and Hasselmo, M. E. (2010). Coupled noisy spiking neurons as velocity-controlled oscillators in a model of grid cell spatial firing. *J. Neurosci.* 30, 13850–13860. doi: 10.1523/JNEUROSCI.0547-10.2010

Zohary, E., Shadlen, M. N., and Newsome, W. T. (1994). Correlated neuronal discharge rate and its implications for psychophysical performance. *Nature* 370, 140–143. doi: 10.1038/370140a0

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.

Received: 29 April 2013; accepted: 24 July 2013; published online: 14 August 2013.
Citation: Bolhasani E, Azizi Y and Valizadeh A (2013) Direct connections assist neurons to detect correlation in small amplitude noises. Front. Comput. Neurosci. 7:108. doi: 10.3389/fncom.2013.00108

Copyright © 2013 Bolhasani, Azizi and Valizadeh. 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) or licensor 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.

frontiers in
COMPUTATIONAL NEUROSCIENCE

METHODS ARTICLE
published: 17 July 2013
doi: 10.3389/fncom.2013.00084

A generative spike train model with time-structured higher order correlations

James Trousdale[1], Yu Hu[2], Eric Shea-Brown[2,3] and Krešimir Josić[1,4]*

[1] Department of Mathematics, University of Houston, Houston, TX, USA
[2] Department of Applied Mathematics, University of Washington, Seattle, WA, USA
[3] Program in Neurobiology and Behavior, University of Washington, Seattle, WA, USA
[4] Department of Biology and Biochemistry, University of Houston, Houston, TX, USA

Edited by:
Robert Rosenbaum, University of Pittsburgh, USA

Reviewed by:
John A. Hertz, Niels Bohr Institute, Denmark
Sonja Grün, Research Center Juelich, Germany

****Correspondence:***
James Trousdale, Department of Mathematics, University of Houston, 641 PGH Building, Houston, TX 77204-3008, USA
e-mail: jrtrousd@math.uh.edu

Emerging technologies are revealing the spiking activity in ever larger neural ensembles. Frequently, this spiking is far from independent, with correlations in the spike times of different cells. Understanding how such correlations impact the dynamics and function of neural ensembles remains an important open problem. Here we describe a new, generative model for correlated spike trains that can exhibit many of the features observed in data. Extending prior work in mathematical finance, this *generalized thinning and shift* (GTaS) model creates marginally Poisson spike trains with diverse temporal correlation structures. We give several examples which highlight the model's flexibility and utility. For instance, we use it to examine how a neural network responds to highly structured patterns of inputs. We then show that the GTaS model is analytically tractable, and derive cumulant densities of all orders in terms of model parameters. The GTaS framework can therefore be an important tool in the experimental and theoretical exploration of neural dynamics.

Keywords: correlations, spiking neurons, neuronal networks, cumulant, neuronal modeling, neuronal network model, point processes

1. INTRODUCTION

Recordings across the brain suggest that neural populations spike collectively—the statistics of their activity as a group are distinct from that expected in assembling the spikes from one cell at a time (Bair et al., 2001; Salinas and Sejnowski, 2001; Harris, 2005; Averbeck et al., 2006; Schneidman et al., 2006; Shlens et al., 2006; Pillow et al., 2008; Ganmor et al., 2011; Bathellier et al., 2012; Hansen et al., 2012; Luczak et al., 2013). Advances in electrode and imaging technology allow us to explore the dynamics of neural populations by simultaneously recording the activity of hundreds of cells. This is revealing patterns of collective spiking that extend across multiple cells. The underlying structure is intriguing: For example, higher-order interactions among cell groups have been observed widely (Amari et al., 2003; Schneidman et al., 2006; Shlens et al., 2006, 2009; Ohiorhenuan et al., 2010; Ganmor et al., 2011; Vasquez et al., 2012; Luczak et al., 2013). A number of recent studies point to mechanisms that generate such higher-order correlations from common input processes, including unobserved neurons. This suggests that, in a given recording or given set of neurons projecting downstream, higher-order correlations may be quite ubiquitous (Barreiro et al., 2010; Macke et al., 2011; Yu et al., 2011; Köster et al., 2013). Moreover, these *higher-order correlations* may impact the firing statistics of downstream neurons (Kuhn et al., 2003), the information capacity of their output (Ganmor et al., 2011; Cain and Shea-Brown, 2013; Montani et al., 2013), and could be essential in learning through spike-time dependent synaptic plasticity (Pfister and Gerstner, 2006; Gjorgjieva et al., 2011).

What exactly is the impact of such collective spiking on the encoding and transmission of information in the brain? This question has been studied extensively, but much remains unknown. Results to date show that the answers will be varied and rich. Patterned spiking can impact responses at the level of single cells (Salinas and Sejnowski, 2001; Kuhn et al., 2003; Xu et al., 2012) and neural populations (Amjad et al., 1997; Tetzlaff et al., 2003; Rosenbaum et al., 2010, 2011). Neurons with even the simplest of non-linearities can be highly sensitive to correlations in their inputs. Moreover, such non-linearities are sufficient to accurately decode signals from the input to correlated neural populations (Shamir and Sompolinsky, 2004).

An essential tool in understanding the impact of collective spiking is the ability to generate artificial spike trains with a predetermined structure across cells and across time (Brette, 2009; Gutnisky and Josić, 2009; Krumin and Shoham, 2009; Macke et al., 2009). Such synthetic spike trains are the grist for testing hypotheses about spatiotemporal patterns in coding and dynamics. In experimental studies, such spike trains can be used to provide structured stimulation of single cells across their dendritic trees via glutamate uncaging (Gasparini and Magee, 2006; Reddy et al., 2008; Branco et al., 2010; Branco and Häusser, 2011). In addition, entire populations of neurons can be activated via optical stimulation of microbial opsins (Han and Boyden, 2007; Chow et al., 2010). Computationally, they are used to examine the response of non-linear models of downstream cells (Carr et al., 1998; Salinas and Sejnowski, 2001; Kuhn et al., 2003).

Therefore, much effort has been devoted to developing statistical models of population activity. A number of flexible, yet tractable probabilistic models of joint neuronal activity have been proposed. Pairwise correlations are the most common type of interactions obtained from multi-unit recordings. Therefore, many earlier models were designed to generate samples of neural activity patterns with predetermined first and second order statistics (Brette, 2009; Gutnisky and Josić, 2009; Krumin and Shoham, 2009; Macke et al., 2009). In these models, higher-order correlations are not explicitly and separately controlled.

A number of different models have been used to analyze higher-order interactions. However, most of these models assume that interactions between different cells are instantaneous (or near-instantaneous) (Kuhn et al., 2003; Johnson and Goodman, 2009; Staude et al., 2010; Shimazaki et al., 2012). A notable exception is the work of Bäuerle and Grübel (2005), which developed such methods for use in financial applications. In these previous efforts, correlations at all orders were characterized by the increase, or decrease, in the probability that groups of cells spike together at the same time, or have a common temporal correlation structure regardless of the group.

The aim of the present work is to provide a statistical method for generating spike trains with more general correlation structures across cells and time. Specifically, we allow distinct temporal structure for correlations at pairwise, triplet, and all higher orders, and do so separately for different groups of cells in the neural population. Our aim to describe a model that can be applied in neuroscience, and can potentially be fit to emerging datasets.

A sample realization of a multivariate generalized thinning and shift (GTaS) process is shown in **Figure 1**. The multivariate spike train consists of six marginally Poisson processes. Each event was either uncorrelated with all other events across the population, or correlated in time with an event in all other spike trains. This model was configured to exhibit activity that cascades through a sequence of neurons. Specifically, neurons with larger index tend to fire later in a population wide event (this is similar to a synfire chain (Abeles, 1991), but with variable timing of spikes within the cascade). In **Figure 1B**, we plot the "population cross-cumulant density" for three chosen neurons—the summed activity of the population triggered by a spike in a chosen cell. The center of mass of this function measures the average latency by which spikes of the neuron in question precede those of the rest of the population (Luczak et al., 2013). Finally, **Figure 1C** shows the third-order cross-cumulant density for the three neurons. The triangular support of this function is a reflection of a synfire-like cascade structure of the spiking shown in the raster plot of panel **(A)**: when firing events are correlated between trains, they tend to proceed in order of increasing index. We demonstrate the impact of such structured activity on a downstream network in section 2.2.3.

2. RESULTS

Our aim is to describe a flexible multivariate point process capable of generating a range of high order correlation structures. To do so, we extend the *TaS* (thinning and shift) model of temporally- and spatially-correlated, marginally Poisson counting processes (Bäuerle and Grübel, 2005). The TaS model itself generalizes the SIP and MIP models (Kuhn et al., 2003) which have been used in theoretical neuroscience (Tetzlaff et al., 2008; Rosenbaum et al., 2010; Cain and Shea-Brown, 2013). However, the TaS model has not been used as widely. The original TaS model is too rigid to generate a number of interesting activity patterns observed in multi-unit recordings (Ikegaya et al., 2004; Luczak et al., 2007, 2013). We therefore developed the *GTaS* which allows for a more diverse temporal correlation structure.

We begin by describing the algorithm for sampling from the GTaS model. This constructive approach provides an intuitive understanding of the model's properties. We then present a pair of examples, the first of which highlights the utility of the

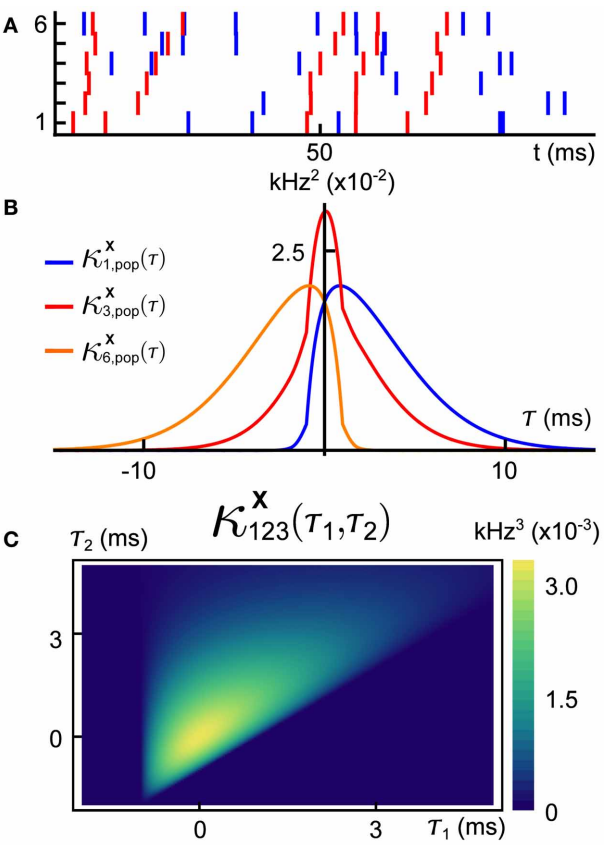


FIGURE 1 | (A) Raster plot of event times for an example multivariate Poisson process $\mathbf{X} = (X_1, \ldots, X_6)$ generated using the methods presented below. This model exhibits independent marginal events (blue) and population-level, chain-like events (red). **(B)** Some second order population cumulant densities (i.e., second order correlation between individual unit activities and population activity) for this model (Luczak et al., 2013). Greater mass to the right (resp. left) of $\tau = 0$ indicates that the cell tends to lead (resp. follow) in pairwise-correlated events. **(C)** Third-order cross-cumulant density for processes X_1, X_2, X_3. The quantity $\kappa^{\mathbf{X}}_{123}(\tau_1, \tau_2)$ yields the probability of observing spikes in cells 2 and 3 at an offset τ_1, τ_2 from a spike in cell 1, respectively, in excess of what would be predicted from the first and second order cumulant structure. All quantities are precisely defined in the Methods. Note: system parameters necessary to reproduce results are given in the Appendix for all figures.

GTaS framework. The second example demonstrates how sample point processes from the GTaS model can be used to study population dynamics. Next, we present the analysis which yields the explicit forms for the cross-cumulant densities derived in the context of the examples. We do so by first establishing a useful distributional representation for the GTaS process, paralleling Bäuerle and Grübel (2005). Using this representation, we derive cross-cumulants of a GTaS counting process, as well as explicit expressions for the cross-cumulant densities. After explaining the derivation at lower orders, we present a theorem which describes cross-cumulant densities at all orders.

2.1. GTaS MODEL SIMULATION

The GTaS model is parameterized first by a rate λ which determines the intensity of a "mother process"—a Poisson process on $\mathbb{R}$. The events of the mother process are marked, and the markings determine how each event is distributed among a collection of N daughter processes. The daughter processes are indexed by the set $\mathbb{D} = \{1, \ldots, N\}$, and the set of possible markings is the power set $2^{\mathbb{D}}$, the set of all subsets of $\mathbb{D}$. We define a probability distribution $p = (p_D)_{D \subset \mathbb{D}}$, assigning a probability to each possible marking, D. As we will see, p_D determines the probability of a joint event in all daughter processes with indices in the set D. Finally, to each marking, D, we assign a probability distribution Q_D, giving a family of shift (jitter) distributions $(Q_D)_{D \subset \mathbb{D}}$. Each (Q_D) is a distribution over $\mathbb{R}^N$.

The rate λ, the distribution p over the markings, and the family of jitter distributions $(Q_D)_{D \subset \mathbb{D}}$, define a vector $\mathbf{X} = (X_1, \ldots, X_N)$ of dependent daughter Poisson processes described by the following algorithm, which yields a single realization (see **Figure 2**):

1. Simulate the mother Poisson process of rate λ on $\mathbb{R}$, generating a sequence of event times $\{t^j\}$. (**Figure 2A**).
2. With probability p_{D^j} assign the subset $D^j \subset \mathbb{D}$ to the event of the mother process at time t^j. This event will be assigned only to processes with indices in D^j. (**Figure 2B**).
3. Sample a vector $(Y_1^j, \ldots, Y_N^j) = \mathbf{Y}^j$ from the distribution Q_{D^j}. For each $i \in D$, the time $t^j + Y_i^j$ is set as an event time for the marginal counting process X_i. (**Figure 2C**).

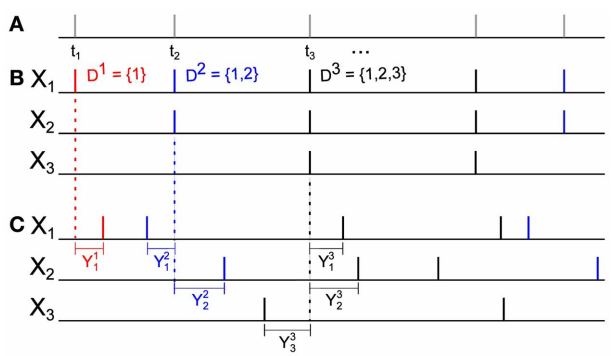


FIGURE 2 | An illustration of a GTaS simulation. (A) Step 1: Simulate the mother process—a time-homogeneous Poisson process with event times $\{t^j\}$. **(B)** Step 2: For each t^j in step 1, select a set $D^j \subset \mathbb{D}$ according to the distribution p_D, and project the event at time t^j to the subsets with indices in D^j. The legend indicates the colors assigned to three possible markings in this example. **(C)** Step 3: For each pair (t^j, D^j) generated in the previous two steps, draw $\mathbf{Y}^j$ from Q_{D^j}, and shift the event times in the daughter processes by the corresponding values Y_i^j.

Hence copies of each point of the mother process are placed into daughter processes after a shift in time. A primary difference between the GTaS model and the TaS model presented in Bäuerle and Grübel (2005) is the dependence of the shift distributions Q_D on the chosen marking. This allows for greater flexibility in setting the temporal cumulant structure.

2.2. EXAMPLES

2.2.1. Relation to SIP/MIP processes

Two simple models of correlated, jointly Poisson processes were defined in Kuhn et al. (2003). The resulting spike trains exhibit spatial correlations, but only instantaneous temporal dependencies. Each model was constructed by starting with independent Poisson processes, and applying one of two elementary point process operations: superposition and thinning (Cox and Isham, 1980). We show that both models are special cases of the GTaS model.

In the *single interaction process* (SIP), each marginal process X_i is obtained by merging an independent Poisson process with a common, global Poisson process. That is,

$$X_i(\cdot) = Z_i(\cdot) + Z_c(\cdot), \quad i = 1, \ldots, N,$$

where Z_c and each Z_i are independent Poisson counting processes on $\mathbb{R}$ with rates λ_c, λ_i, respectively. An SIP model is equivalent to a GTaS model with mother process rate $\lambda = \lambda_c + \sum_{i=1}^N \lambda_i$, and marking probabilities

$$p_D = \begin{cases} \frac{\lambda_i}{\lambda} & D = \{i\} \\ \frac{\lambda_c}{\lambda} & D = \mathbb{D} \\ 0 & \text{otherwise} \end{cases}.$$

Note that if $\lambda_c = 0$, each spike will be assigned to a different process X_i, resulting in N independent Poisson processes. Lastly, each shift distribution is equal to a delta distribution at zero in every coordinate (i.e., $q_D(y_1, \ldots, y_N) \equiv \prod_{i=1}^N \delta(y_i)$ for every $D \subset \mathbb{D}$). Thus, all joint cumulants (among distinct marginal processes) of orders 2 through N are delta functions of equal magnitude, $\lambda p_{\mathbb{D}}$.

The *multiple interaction process* (MIP) consists of N Poisson processes obtained from a common mother process with rate λ_m by *thinning* (Cox and Isham, 1980). The ith daughter process is formed by independent (across coordinates and events) deletion of events from the mother process with probability $p = (1 - \epsilon)$. Hence, an event is common to k daughter processes with probability ϵ^k. Therefore, if we take the perspective of retaining, rather than deleting events, the MIP model is equivalent to a GTaS process with $\lambda = \lambda_m$, and $p_D = \epsilon^{|D|}(1 - \epsilon)^{d - |D|}$. As in the SIP case, the shift distributions are singular in every coordinate. Below, we present a general result (Theorem 1.1) which immediately yields

as a corollary that the MIP model has cross-cumulant functions which are δ functions in all dimensions, scaled by ϵ^k, where k is the order of the cross-cumulant.

2.2.2. Generation of synfire-like cascade activity

The GTaS framework provides a simple, tractable way of generating cascading activity where cells fire in a preferred order across the population—as in a synfire chain, but (in general) with variable timing of spikes (Abeles, 1991; Abeles and Prut, 1996; Aertsen et al., 1996; Aviel et al., 2002; Ikegaya et al., 2004). More generally, it can be used to simulate the activity of *cell assemblies* (Hebb, 1949; Harris, 2005; Buzsáki, 2010; Bathellier et al., 2012), in which the firing of groups of neurons is likely to follow a particular order.

In the Introduction, we briefly presented one example in which the GTaS framework was used to generate synfire-like cascade activity (see **Figure 1**), and we present another in **Figure 3**. In what follows, we will present the explicit definition of this second model, and then derive explicit expressions for its cumulant structure. Our aim is to illustrate the diverse range of possible correlation structures that can be generated using the GTaS model.

Consider an N-dimensional counting process $\mathbf{X} = (X_1, \ldots, X_N)$ of GTaS type, where $N \geq 4$. We restrict the marking distribution so that $p_D \equiv 0$ unless $|D| \leq 2$ or $D = \mathbb{D}$. That is, events are either assigned to a single, a pair, or all daughter processes. For sets D with $|D| = 2$, we set $Q_D \sim \mathcal{N}(0, \Sigma)$—a Gaussian distributions of zero mean and some specified covariance. The choice of the precise pairwise shift distributions is not important. Shifts of events attributed to a single process have no effect on the statistics of the multivariate process—this will become clear in section 2.3, when we exhibit that a GTaS process is equivalent in distribution to a sum of independent Poisson processes. In that context, the shifts of marginal events are applied to the event times of only one of these Poisson processes, which does not impact its statistics.

It remains to define the jitter distribution for events common to the entire population of daughter processes, i.e., events marked by $\mathbb{D}$. We will show that we can generate cascading activity, and analytically describe the resulting correlation structure. We will say that a random variable T follows the exponential distribution Exp(α) if it has probability density

$$f(t|\alpha) = \alpha e^{-\alpha t}\Theta(t),$$

where $\Theta(t)$ is the Heaviside step function. We generate random vectors $\mathbf{Y} \sim Q_{\mathbb{D}}$ according to the following rule, for each $i = 1, \ldots, N$:

1. Generate independent random variables $T_i \sim \text{Exp}(\alpha_i)$ where $\alpha_i > 0$.
2. Set $Y_i = \sum_{j=1}^{i} T_j$.

In particular, note that these shift times satisfy $Y_N \geq \ldots \geq Y_2 \geq Y_1 \geq 0$, indicating the chain-like structure of these joint events.

From the definition of the model and our general result (Theorem 1.1) below, we immediately have that $\kappa_{ij}^{\mathbf{X}}(\tau)$, the second

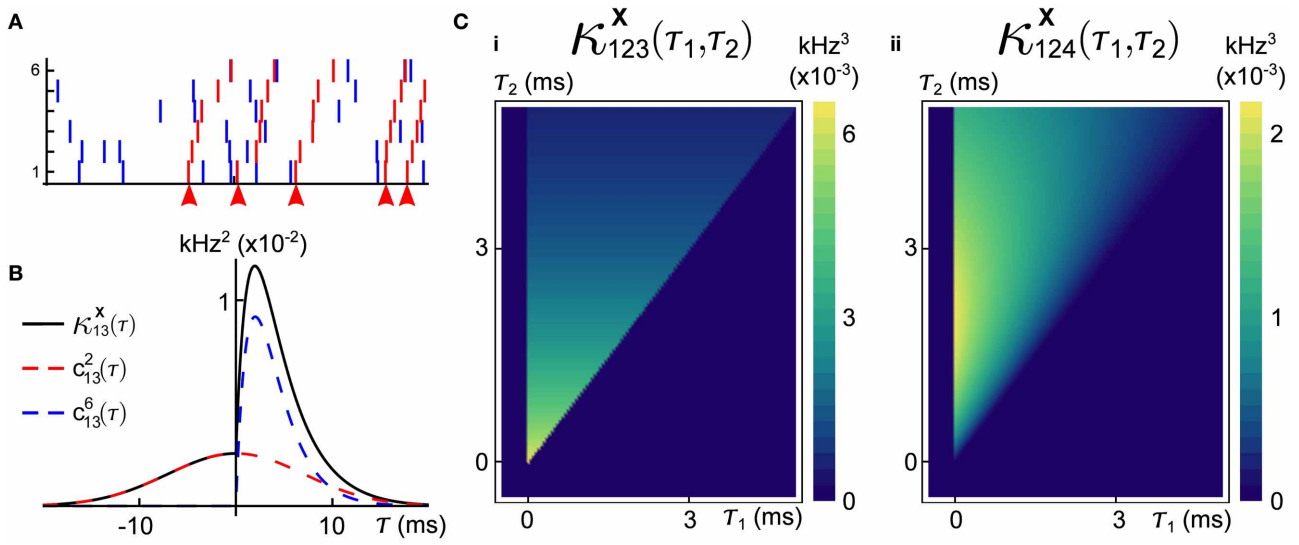


FIGURE 3 | An example of a six dimensional GTaS model exhibiting synfire-like cascading firing patterns. (A) A raster-plot of spiking activity over a 100 ms window. Blue spikes indicate either marginal or pairwise events (i.e., events corresponding to markings for sets $D \subset \mathbb{D}$ with $|D| \leq 2$). Red spikes indicate population-wide events which have shift-times given by cumulative sums of independent exponentials, as described in the text. Arrows indicate the location of the first spike in the cascade. **(B)** A second-order cross-cumulant $\kappa^{\mathbf{X}}_{13}$ (black line) of this model is composed of contributions from two sources: correlations due to second-order markings, which have Gaussian shifts (c^2_{13}—dashed red line), and correlations due to the occurrence of population wide events (c^N_{13}—dashed blue line). **(C)** Density plots of the third-order cross-cumulant density for triplets **(i)** (1, 2, 3) and **(ii)** (1, 2, 4)—the latter is given explicitly in Equation (6). System parameters are given in the Appendix.

order cross-cumulant density for the process (i, j), is given by

$$\kappa^{\mathbf{X}}_{ij}(\tau) = c^2_{ij}(\tau) + c^N_{ij}(\tau), \tag{1}$$

where

$$c^2_{ij}(\tau) = \lambda p_{\{i,j\}} \int q^{\{i,j\}}_{\{i,j\}}(t, t+\tau)dt,$$
$$c^N_{ij}(\tau) = \lambda p_{\mathbb{D}} \int q^{\{i,j\}}_{\mathbb{D}}(t, t+\tau)dt \tag{2}$$

define the contributions to the second order cross-cumulant density by the second-order, Gaussian-jittered events and the population-level events, respectively. Therefore, correlations between spike trains in this case reflect distinct contributions from second order and higher order events. The functions $q^{D'}_{D}$ indicate the densities associated with the distribution Q_D, projected to the dimensions of D'. All statistical quantities are precisely defined in the methods.

By exploiting the hierarchical construction of the shift times, we can find an expression for the joint density $q_{\mathbb{D}}$, necessary to explicitly evaluate Equation (1). For a general N-dimensional distribution,

$$f(y_1, \ldots, y_N) = f(y_N|y_1, \ldots, y_{N-1})f(y_{N-1}|y_1, \ldots, y_{N-2}) \cdots \cdot f(y_2|y_1)f(y_1). \tag{3}$$

Since $Y_1 \sim \text{Exp}(\alpha_1)$, we have $f(y_1) = \exp\left[-\alpha_1 y_1\right]\Theta(y_1)$, where $\Theta(y)$ is the Heaviside step function. Further, as $(Y_i - Y_{i-1})|(Y_1, \ldots, Y_{i-1}) \sim \text{Exp}(\alpha_i)$ for $i \geq 2$, the conditional densities of the y_i's take the form

$$f(y_i|y_1, \ldots, y_{i-1}) = f(y_i|y_{i-1}) = \alpha_i \exp\left[-\alpha_i(y_i - y_{i-1})\right] \cdot \Theta\,(y_i - y_{i-1}), \quad i \geq 2.$$

Substituting this in to the identity Equation (3), we have

$$q_{\mathbb{D}}(y_1, \ldots, y_N) = \begin{cases} \alpha_1 \exp\left[-\alpha_1 y_1\right] \prod_{i=2}^{N} \alpha_i & y_N \geq \ldots \geq \\ \quad \cdot \exp\left[-\alpha_i(y_i - y_{i-1})\right] & y_2 \geq y_1 \geq 0\,. \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

Using Theorem 1.1 (Equation A8) we obtain the Nth order cross-cumulant density (see the Methods),

$$\kappa^{\mathbf{X}}_{1\ldots N}(\tau_1, \ldots, \tau_{N-1}) = \lambda p_{\mathbb{D}} \int q_{\mathbb{D}}(t, t+\tau_1, \ldots, t+\tau_{N-1})dt \tag{5}$$
$$= \lambda p_{\mathbb{D}} \cdot \begin{cases} \prod_{i=1}^{N-1} \alpha_{i+1} & \tau_i \geq \tau_{i-1} \\ \quad \cdot \exp\left[-\alpha_{i+1}(\tau_i - \tau_{i-1})\right] & i = 1, \ldots, N-1, \\ 0 & \text{otherwise} \end{cases}$$

where, for notational convenience, we define $\tau_0 = 0$. A raster plot of a realization of this model is shown in **Figure 3A**. We note that the cross-cumulant densities of arbitrary subcollections of the counting processes $\mathbf{X}$ can be obtained by finding the appropriate marginalization of $q_{\mathbb{D}}$ via integration of Equation (4). In the case that common distributions are used to define the shifts, symbolic calculation environments (i.e., Mathematica) can quickly yield explicit formulas for cross-cumulant densities. Mathematica notebooks for **Figure 1** available upon request.

As a particular example, we consider the cross-cumulant density of the marginal processes X_1, X_3. Using Equations (2, 4), we find

$$c^N_{13}(\tau) = \lambda p_{\mathbb{D}} \Theta(\tau) \cdot \begin{cases} \frac{\alpha_2\alpha_3}{\alpha_3 - \alpha_2}\left\{\exp\left[-\alpha_2\tau\right] - \exp\left[-\alpha_3\tau\right]\right\} & \alpha_2 \neq \alpha_3 \\ \alpha_2\alpha_3\tau \exp\left[-\alpha_2\tau\right] & \alpha_2 = \alpha_3 \end{cases}.$$

An expression for $c^2_{13}(\tau)$ may be obtained similarly using Equation (2) and recalling that $Q_{\{i,j\}} \equiv \mathcal{N}(0, \Sigma)$ for all i, j. In **Figure 3B**, we plot these contributions, as well as the full covariance density.

Similar calculations at third order yield, as an example,

$$\kappa^{\mathbf{X}}_{124}(\tau_1, \tau_2) = \lambda p_{\mathbb{D}} \cdot \begin{cases} \frac{\alpha_2\alpha_3\alpha_4}{\alpha_4 - \alpha_3} \exp\left[-\alpha_2\tau_1\right]\left\{\exp\left[-\alpha_3(\tau_2 - \tau_1)\right]\right. & \\ \quad \left. - \exp\left[-\alpha_4(\tau_2 - \tau_1)\right]\right\} & \alpha_3 \neq \alpha_4, \\ \alpha_2\alpha_3\alpha_4(\tau_2 - \tau_1)\exp\left[-\alpha_2\tau_1 - \alpha_3(\tau_2 - \tau_1)\right] & \alpha_3 = \alpha_4 \end{cases} \tag{6}$$

where the cross-cumulant density $\kappa^{\mathbf{X}}_{124}(\tau_1, \tau_2)$ is supported only on $\tau_2 \geq \tau_1 \geq 0$. Plots of the third-order cross-cumulants for triplets $(1, 2, 3)$ and $(1, 2, 4)$ in this model are shown in **Figure 3C**. Note that, for the specified parameters, the conditional distribution of Y_4—the shift applied to the events of X_4 in a joint population event—given Y_2 follows a gamma distribution, whereas $Y_3|Y_2$ follows an exponential distribution, explaining the differences in the shapes of these two cross-cumulant densities.

General cross-cumulant densities of at least third order for the cascading model will have a form similar to that given in Equation (6), and will contain no signature of the correlation of strictly second order events. This highlights a key benefit of cumulants as a measure of dependence: although they agree with central moments up to third order, we know from Equation (23) below [or Equation (22) in the general case] that central moments necessarily exhibit a dependence on lower order statistics. On the other hand, cumulants are "pure" and quantify only dependencies at the given order which cannot be inferred from lower order statistics (Grün and Rotter, 2010).

One useful statistic for analyzing population activity through correlations is the *population cumulant density* (Luczak et al., 2013). The second order population cumulant density for cell i is defined by (see the Methods)

$$\kappa^{\mathbf{X}}_{i,\,\text{pop}}(\tau) = \sum_{j \neq i} \kappa^{\mathbf{X}}_{ij}(\tau).$$

This function is linearly related to the spike-triggered average of the population activity conditioned on that of cell i. In **Figure 4**

we show three different second-order population-cumulant functions for the cascading GTaS model of **Figure 3A**. When the second order population cumulant for a neuron is skewed to the right of $\tau = 0$ (as is $\kappa^{\mathbf{X}}_{1,\,\mathrm{pop}}$—blue line), a neuron tends to precede other cells in the population in pairwise spiking events. Similarly, skewness to the left of $\tau = 0$ ($\kappa^{\mathbf{X}}_{6,\,\mathrm{pop}}$—orange line) indicates a neuron which tends to trail other cells in the population in such events. A symmetric population cumulant density indicates a neuron is a follower *and* a leader. Taken together, these three second order population cumulants hint at the chain structure of the process.

Greater understanding of the joint temporal statistics in a multivariate counting process can be obtained by considering higher-order population cumulant densities. We define the third-order population cumulant density for the pair (i, j) to be

$$\kappa^{\mathbf{X}}_{ij,\,\mathrm{pop}}(\tau_1, \tau_2) = \sum_{k \neq i,\,j} \kappa^{\mathbf{X}}_{ijk}(\tau_1, \tau_2).$$

The third-order population cumulant density is linearly related to the spike-triggered population activity, conditioned on spikes in cells i and j separated by a delay τ_1. In **Figures 4B–D**, we present three distinct third-order population cumulant densities. Examining $\kappa^{\mathbf{X}}_{12,\,\mathrm{pop}}(\tau_1, \tau_2)$ (panel **B**), we see only contributions in the region $\tau_2 > \tau_1 > 0$, indicating that the pairwise event $1 \to 2$ often precedes a third spike elsewhere in the population (i.e., with a probability above chance). The population cumulant $\kappa^{\mathbf{X}}_{34,\,\mathrm{pop}}(\tau_1, \tau_2)$ has contributions in two sections of the plane (panel **C**). Contributions in the region $\tau_2 > \tau_1 > 0$ can be understood following the preceding example, while contributions in the region $\tau_2 < 0 < \tau_1$ imply that the firing of other neurons tends to precede the joint firing event $3 \to 4$. Lastly, contributions to $\kappa^{\mathbf{X}}_{16,\,\mathrm{pop}}(\tau_1, \tau_2)$ (panel **D**) are limited to $0 < \tau_2 < \tau_1$, indicating an above chance probability of joint firing events of the form $1 \to i \to 6$, where i indicates a distinct neuron within the population.

A distinct advantage of the study of population cumulant densities as opposed to individual cross-cumulant functions in

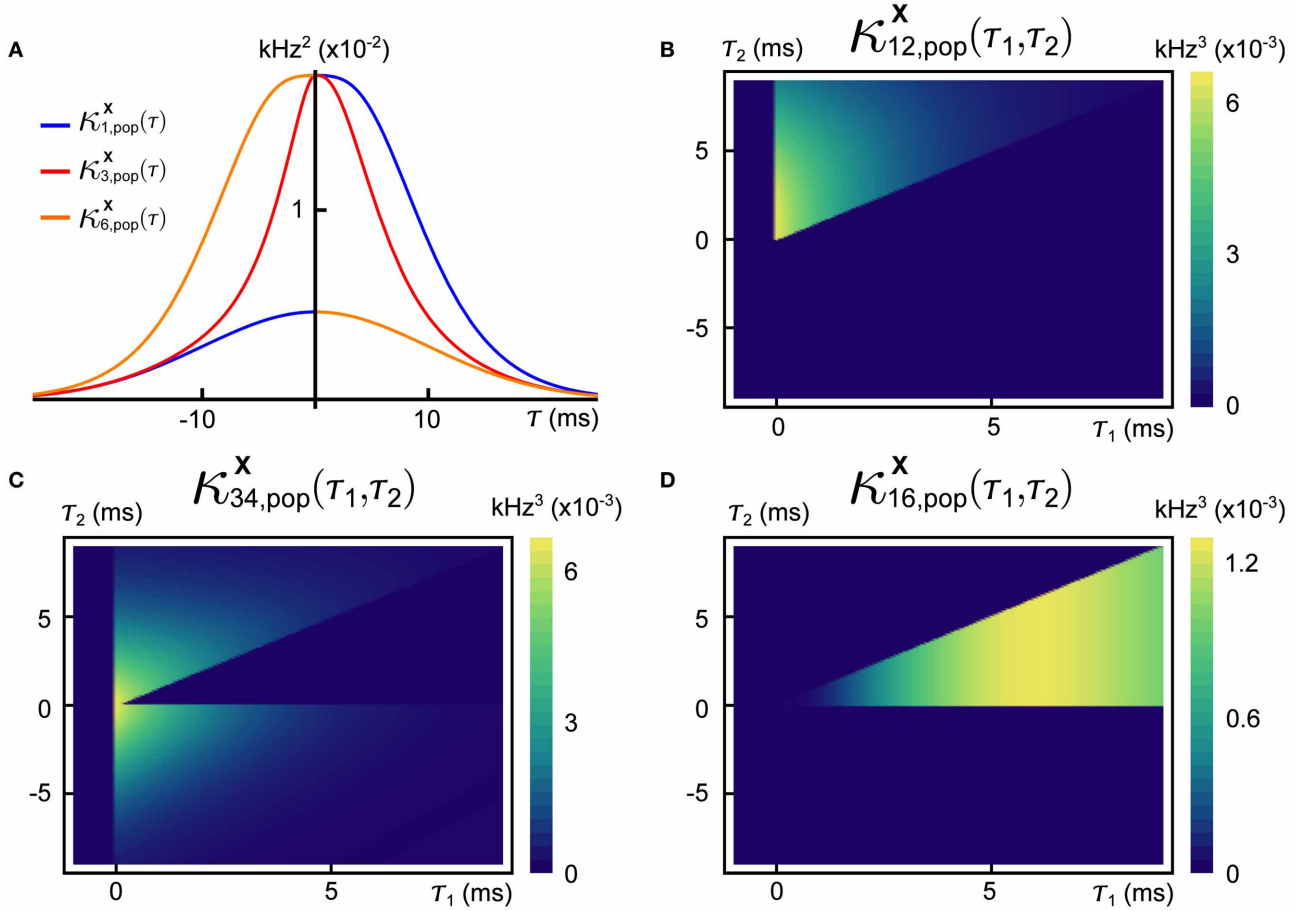


FIGURE 4 | Population cumulants for the synfire-like cascading GTaS process of Figure 3. See Equation (25) for the definition of population cumulants. **(A)** Second order population cumulant densities for processes 1, 3, and 6. Greater mass to the right (resp. left) of $\tau = 0$ indicates that a cell tends to lead (resp. follow) in pairwise-correlated events. **(B)** Third order population cumulant for processes X_1, X_2 in the cascading GTaS process. Concentration of the mass in different regions of the plane indicates temporal structure of events correlated between X_1, X_2 relative to the remainder of the population (see the text). **(C)** Same as **(B)**, but for processes X_3, X_4. **(D)** Same as **(B)**, but for processes X_1, X_6. System parameters are given in the Appendix.

practical applications is related to data (i.e., sample size) limitations. In many practical applications, where the temporal structure of a collection of observed point processes is of interest, we often deal with a small, noisy samples. It may therefore be difficult to estimate third- or higher-order cumulants. Population cumulants partially circumvent this issue by *pooling* (Tetzlaff et al., 2003; Rosenbaum et al., 2010, 2011) (or summing) responses, to amplify existing correlations and average out the noise in measurements.

We conclude this section by noting that even cascading GTaS examples can be much more general. For instance, we can include more complex shift patterns, overlapping subassemblies within the population, different temporal processions of the cascade, and more.

2.2.3. Timing-selective network

The responses of single neurons and neuronal networks in experimental (Meister and Berry II, 1999; Singer, 1999; Bathellier et al., 2012) and theoretical studies (Jeffress, 1948; Hopfield, 1995; Joris et al., 1998; Thorpe et al., 2001; Gütig and Sompolinsky, 2006) can reflect the temporal structure of their inputs. Here, we present a simple example that shows how a network can be selective to fine temporal features of its input, and how the GTaS model can be used to explore such examples.

As a general network model, we consider N leaky integrate-and-fire (LIF) neurons with membrane potentials V_i obeying

$$\frac{dV_i}{dt} = -V_i + \sum_{j=1}^{N} w_{ij}(F * z_j)(t) + w^{\text{in}} x_i(t), \quad i = 1, \ldots, N. \tag{7}$$

When the membrane potential of cell i reaches a threshold Vth, an output spike is recorded and the membrane potential is reset to zero, after which evolution of V_i resumes the dynamics in Equation (7). Here w_{ij} is the synaptic weight of the connection from cell j to i, w^{in} is the input weight, and we assume time to be measured in units of membrane time constants. The function $F = {\tau_{\text{syn}}}^{-1} e^{-(t-\tau_{\text{d}})/\tau_{\text{syn}}} \Theta(t - \tau_{\text{d}})$ is a delayed, unit-area exponential synaptic kernel with time-constant τ_{syn} and delay τ_{d}. The output of the ith neuron is

$$z_i(t) = \sum_j \delta(t - t_i^j),$$

where t_i^j is the time of the jth spike of neuron i. In addition, the input $\{x_i\}_{i=1}^N$ is

$$x_i(t) = \sum_j \delta(t - s_i^j),$$

where the event times $\{s_i^j\}$ correspond to those of a GTaS counting process $\mathbf{X}$. Thus, each input spike results in a jump in the membrane potential of the corresponding LIF neuron of amplitude w^{in}. The particular network we consider will have a ring topology (nearest neighbor-only connectivity)—specifically, for $i, j = 1, \ldots, N$, we let

$$w_{ij} = \begin{cases} w^{\text{syn}} & i - j \mod N \equiv 1 \text{ or } N-1 \\ 0 & \text{otherwise} \end{cases}.$$

We further assume that all neurons are *excitatory*, so that $w^{\text{syn}} > 0$.

A network of LIF neurons with synaptic delay is a minimal model which can exhibit fine-scale discrimination of temporal patterns of inputs without precise tuning (Izhikevich, 2006) (that is, without being carefully designed to do so, with great sensitivity to modification of network parameters). To exhibit this dependence we generate inputs from two GTaS processes. The first (the *cascading model*) was described in the preceding example. To independently control the mean and variance of relative shifts we replace the sum of exponential shifts with sums of gamma variates. We also consider a model featuring population-level events without shifts (the *synchronous model*), where the distribution $Q_{\mathbb{D}}$ is a δ distribution at zero in all coordinates.

The only difference between the two input models is in the temporal structure of joint events. In particular, the rates, and all long timescale spike count cross-cumulants (equivalent to the total "area" under the cross-cumulant density, see the Methods) of order two and higher are identical for the two processes. We focus on the sensitivity of the network to the temporal cumulant structure of its inputs.

In **Figures 5A,B**, we present two example rasters of the nearest-neighbor LIF network receiving synchronous (left) and cascading (right) input. In the second case, there is an obvious pattern in the outputs, but the firing rate is also increased. This is quantified in **Figure 5C**, where we compare the number of output spikes fired by a network receiving synchronous input (horizontal axis) with the same for a network receiving cascading input (vertical axis), over a large number of trials. On average, the cascading input increases the output rate by a factor of 1.5 over the synchronous inputs—we refer to this quantity as the *cascade amplification factor* (CAF).

Finally, in **Figure 5D**, we illustrate how the cascade amplification factor depends on the parameters that define the timing of spikes for the cascading inputs. First, we study the dependence on the standard deviation σ_{shift} of the gamma variates determining the shift distribution. We note that amplification factors above 1.5 hold robustly (i.e., for a range of shift σ_{shift} values). The amplification factors decrease with shift variance. In the inset to panel (**D**), we show how the gain depends on the mean of the shift distribution μ_{shift}. On an individual trial, the response intensity will depend strongly on the total number of input spikes. Thus, in order to enforce a fair comparison, the mother process and markings used were identical in each trial of every panel of **Figure 5**. We note that network properties, such as the membrane properties of individual cells or synaptic timescales, may have an equally large impact on the cascade amplification factor—indeed, as we explain below, the observed behavior of the CAF is a result of synergy between the timescales of input and interactions within the network.

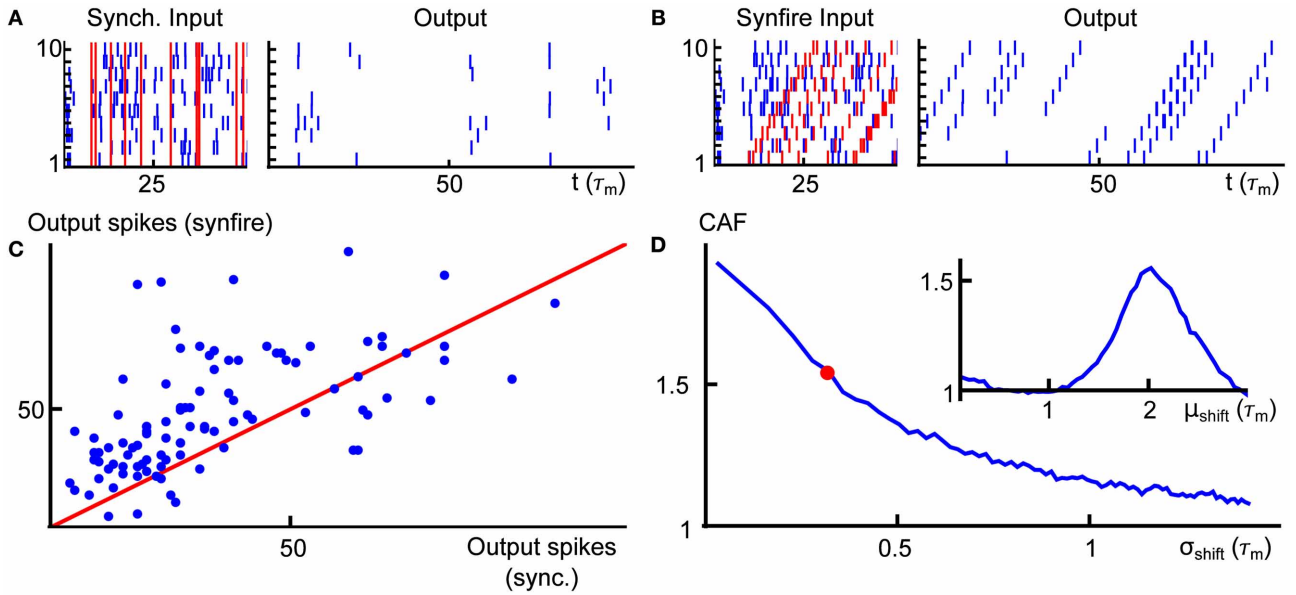


FIGURE 5 | (A) Example input (left) and output (right) for the nearest neighbor LIF network receiving input with synchronous input. **(B)** Same as **(A)**, but for cascading input. **(C)** Scatter plot of the output spike count of the network receiving synchronous (horizontal axis) and cascading input (vertical axis) with $\mu_{shift} = 2$, $\sigma_{shift} = 0.3$. The red line is the diagonal. **(D)** Average gain (rate in response to cascading input divided by rate in response to synchronous input) as a function of the standard deviation of the gamma variates which compose the shift vectors for population-level events (μ_{shift} was fixed at 2). The red dot indicates the value of σ_{shift} used in panel **(C)**. Inset shows the same gain as panel **(D)**, but for varying the mean of the shift distribution ($\sigma_{shift} = 0.3$). Spike counts in panels **(C,D)** were obtained for trials of length $T = 100$. Other system parameters are given in the Appendix.

These observations have simple explanations in terms of the network dynamics and input statistics. Neglecting, for a moment, population-level events, the network is configured so that correlations in activity decrease with topographic distance. Accordingly, the probability of finding neurons that are simultaneously close to threshold also decreases with distance. Under the synchronous input model, a population-level event results in a simultaneous increase of the membrane potentials of all neurons by an amount w^{in}, but unless the input is very strong (in which case every, or almost every, neuron will fire regardless of fine-scale input structure), the set of neurons sufficiently close to threshold to "capitalize" on the input and fire will typically be restricted to a topographically adjacent subset. Neurons which do not fire almost immediately will soon have forgotten about this population-level input. As a result, the output does not significantly reflect the chain-like structure of the inputs (**Figure 5A**, right).

On the other hand, in the case of the cascading input, the temporal structure of the input and the timescale of synapses can operate synergistically. Consider a pair of adjacent neurons in the ring network, called cells 1 and 2, arranged so that cell 2 is downstream from cell 1 in the direction of the population-level chain events. When cell 1 spikes, it is likely that cell 2 will also have an elevated membrane potential. The potential is further elevated by the delayed synaptic input from cell 1. If cell 1 spikes in response to a population-level chain event, then cell 2 imminently receives an input spike as well. If the synaptic filter and time-shift of the input spikes to each cell align, then the firing probability of cell 2 will be large relative to chance. This reasoning can be carried on across the network. Hence synergy between the temporal structure of inputs and network architecture allows the network to selectively respond to the temporal structure of the inputs (**Figure 5B**, right).

In Kuhn et al. (2003), the effect of higher order correlations on the firing rate gain of an integrate-and-fire neuron was studied by driving single cells using sums of SIP or MIP processes with equivalent firing rates (first order cumulants) and pairwise correlations (second order cumulants). In contrast, in the preceding example, the two inputs have equal long time spike count cumulants, and differ only in temporal correlation structure. An increase in firing rate was due to network interactions, and is therefore a population level effect. We return to this comparison in the Discussion.

These examples demonstrate how the GTaS model can be used to explore the impact of spatio-temporal structure in population activity on network dynamics. We next proceed with a formal derivation of the cumulant structure for a general GTaS process.

2.3. CUMULANT STRUCTURE OF A GTaS PROCESS

The GTaS model defines an N-dimensional counting process. Following the standard description for a counting process, $\mathbf{X} = (X_1, \ldots, X_N)$ on $\mathbb{R}^N$, given a collection of Borel subsets $A_i \in \mathcal{B}(\mathbb{R}), i = 1, \ldots, N$, then $\mathbf{X}(A_1 \times \cdots \times A_N) = (X_1(A_1), \ldots, X_N(A_N)) \in \mathbb{N}^N$ is a random vector where the value of each coordinate i indicates the (random) number of points which fall inside the set A_i. Note that the GTaS model defines processes that are marginally Poisson. All GTaS model parameters and related quantities are defined in **Table 1**.

Table 1 | Common notation used in the text.

$\mathbb{D}$	$\mathbb{D} = \{1, 2, \ldots, N\}$ where N is the system size of the GTaS process under consideration
$(p_D)_{D \subset \mathbb{D}}$	Marking probabilities of a GTaS process
$(Q_D)_{D \subset \mathbb{D}}$	Family of shift distributions on $\mathbb{R}^N$ for a GTaS process
$\mathcal{B}(\mathbb{R})$	Borel subsets of the real line $\mathbb{R}$
$\xi(D; A_1, \ldots, A_N)$	Independent Poisson variables which count points which, after shifting, lie in the sets A_i only along the dimensions corresponding to the indices of D. These counts consist of contributions from subsets marked for $D' \supset D$, but indices in $D' \backslash D$ end up outside the corresponding A_i. Defined in the statement of Theorem 0
$\zeta_D(A_1, \ldots, A_N)$	Independent Poisson variables which are context-dependent resummations of the variables $\xi(D; A_1, \ldots, A_N)$. Defined below Equation (10)
$\kappa(X_1, \ldots, X_N)$	Cross-cumulant of the random variables $X_1, \ldots, X_N$ defined in the Methods
$\kappa^{\mathbf{X}}_{i_1 \cdots i_k}(\tau_1, \ldots, \tau_{k-1})$	Cross-cumulant density defined in Equation (24)
$\kappa^{\mathbf{X}}_{i_1 \cdots i_{k-1}, \mathrm{pop}}(\tau_1, \ldots, \tau_{k-1})$	Population cumulant density defined in Equation (25)

For each $D \subset \mathbb{D} = \{1, \ldots, N\}$, define the tail probability $\bar{p}_D$ by

$$\bar{p}_D = \sum_{D \subset D' \subset \mathbb{D}} p_{D'}. \tag{8}$$

Since p_D is the probability that exactly the processes in D are marked, $\bar{p}_D$ is the probability that all processes in D, as well as possibly other processes, are marked. An event from the mother process is assigned to daughter process X_i with probability $\bar{p}_{\{i\}}$. As noted above, an event attributed to process i following a marking $D \ni i$ will be marginally shifted by a random amount determined by the distribution $Q_D^{\{i\}}$ which represents the projection of Q_D onto dimension i. Thus, the events in the marginal process X_i are shifted in an independent and identically distributed (IID) manner according to the mixture distribution Q_i given by

$$Q_i = \frac{\sum_{D \ni i} p_D Q_D^{\{i\}}}{\sum_{D \ni i} p_D}.$$

Note that IID shifting of the event times of a Poisson process generates another Poisson process of identical rate. Thus, the process X_i is marginally Poisson with rate $\lambda \bar{p}_{\{i\}}$ (Ross, 1995).

In deriving the statistics of the GTaS counting process $\mathbf{X}$, it will be useful to express the distribution of $\mathbf{X}$ as

$$\begin{pmatrix} X_1(A_1) \\ \vdots \\ X_N(A_N) \end{pmatrix} =_{\mathrm{distr}} \begin{pmatrix} \sum_{D \ni 1} \xi(D; A_1, \ldots, A_N) \\ \vdots \\ \sum_{D \ni N} \xi(D; A_1, \ldots, A_N) \end{pmatrix}. \tag{9}$$

Here, each $\xi(D; A_1, \ldots, A_N)$ is an independent Poisson process, and the notation $=_{\mathrm{distr}}$ indicates that the two random vectors are equal in distribution. This process counts the number of points which are marked by a set $D' \supset D$, but (after shifting) only the points with indices $i \in D$ lie in the corresponding set A_i. Precise definitions of the processes ξ and a proof of Equation (9) may be found in the Appendix. We emphasize that the Poisson processes $\xi(D)$ do not directly count points marked for the set D, but instead points which are marked for a set containing D that, after shifting, only have their D-components lying in the "relevant" sets A_i.

Suppose we are interested in calculating dependencies among a subset of daughter processes, $\{X_{i_j}\}_{i_j \in \bar{D}}$ for some set $\bar{D} \subset \mathbb{D}$, consisting of $|\bar{D}| = k$ distinct members of the collection of counting processes $\mathbf{X}$. Then the following alternative representation will be useful:

$$\begin{pmatrix} X_{i_1}(A_{i_1}) \\ \vdots \\ X_{i_k}(A_{i_k}) \end{pmatrix} =_{\mathrm{distr}} \begin{pmatrix} \sum_{i_1 \in D \subset \bar{D}} \zeta_D(A_1, \ldots, A_N) \\ \vdots \\ \sum_{i_k \in D \subset \bar{D}} \zeta_D(A_1, \ldots, A_N) \end{pmatrix} \tag{10}$$

where

$$\zeta_D(A_1, \ldots, A_N) = \sum_{\substack{D' \supset D \\ (\bar{D} \backslash D) \cap D' = \emptyset}} \xi(D'; A_1, \ldots, A_N).$$

We illustrate this decomposition in the cases $k = 2, 3$ in **Figure 6**. The sums in Equation (10) run over all sets $D \subset \mathbb{D}$ containing the indicated indices i_j and contained within $\bar{D}$. The processes ζ_D are comprised of a sum of all of the processes $\xi(D')$ (defined below Equation 9) such that D' contains all of the indices D, but no other indices which are part of the subset $\bar{D}$ under consideration. These sums are non-overlapping, implying that the ζ_D are also independent and Poisson.

The following examples elucidate the meaning and significance of Equation (10). We emphasize that the GTaS process is a completely characterized, joint Poisson process, and we use Equation (10) to calculate cumulants of a GTaS process. In principle, any other statistics can be obtained similarly.

2.3.1. Second order cumulants (covariance)

We first generalize a well-known result about the dependence structure of temporally jittered pairs of Poisson processes, X_1, X_2. Assume that events from a mother process with rate λ, are

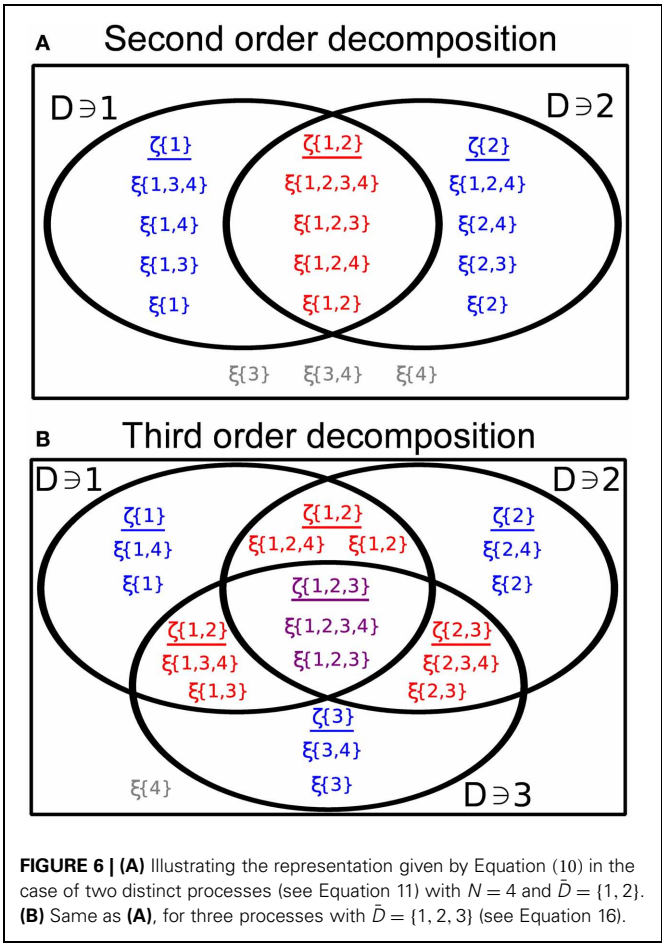


FIGURE 6 | (A) Illustrating the representation given by Equation (10) in the case of two distinct processes (see Equation 11) with $N = 4$ and $\bar{D} = \{1, 2\}$. **(B)** Same as **(A)**, for three processes with $\bar{D} = \{1, 2, 3\}$ (see Equation 16).

assigned to two daughter processes with probability p. Each event time is subsequently shifted independently according to a univariate distribution f. The cross-cumulant density (or cross-covariance function; see the Methods for cumulant definitions) then has the form (Brette, 2009)

$$\kappa_{12}^{\mathbf{X}}(\tau) = \lambda p \int f(t) f(t+\tau) dt = \lambda p (f \times f)(\tau).$$

We generalize this result within the GTaS framework. At second order, Equation (10) has a particularly nice form. Following Bäuerle and Grübel (2005) we write for $i \neq j$ (see **Figure 6A**)

$$\begin{pmatrix} X_i(A_i) \\ X_j(A_j) \end{pmatrix} =_{\text{distr}} \begin{pmatrix} \zeta_{\{i,j\}}(A_i, A_j) + \zeta_{\{i\}}(A_i) \\ \zeta_{\{i,j\}}(A_i, A_j) + \zeta_{\{j\}}(A_j) \end{pmatrix}. \tag{11}$$

The process $\zeta_{\{i,j\}}$ sums all $\xi(D')$ for which $\{1, 2\} \subset D'$, while the process $\zeta_{\{i\}}$ sums all $\xi(D')$ such that $i \in D', j \notin D'$, and $\zeta_{\{j\}}$ is defined likewise.

Using the representation in Equation (11), we can derive the second order cumulant (covariance) structure of a GTaS process. First, we have

$$\begin{aligned} \mathbf{cov}\big[X_i(A_i), X_j(A_j)\big] &= \kappa[X_i(A_i), X_j(A_j)] \\ &= \kappa[\zeta_{\{i,j\}}(A_i, A_j), \zeta_{\{i,j\}}(A_i, A_j)] \\ &\quad + \kappa[\zeta_{\{i\}}(A_i), \zeta_{\{i,j\}}(A_i, A_j)] \\ &\quad + \kappa[\zeta_{\{i,j\}}(A_i, A_j), \zeta_{\{j\}}(A_j)] \\ &\quad + \kappa[\zeta_{\{i\}}(A_i), \zeta_{\{j\}}(A_j)] \\ &= \kappa_2[\zeta_{\{i,j\}}(A_i, A_j)] + 0 \\ &= \mathbf{E}\big[\zeta_{\{i,j\}}(A_i, A_j)\big]. \end{aligned}$$

The third equality follows from the construction of the processes ζ_D: if $D \neq D'$, then the processes $\zeta_D, \zeta_{D'}$ are independent. The final equality follows from the observation that every cumulant of a Poisson random variable equals its mean.

The covariance may be further expressed in terms of model parameters (see Theorem 1.1 for a generalization of this result to arbitrary cumulant orders):

$$\begin{aligned} &\mathbf{cov}\big[X_i(A_i), X_j(A_j)\big] \\ &\quad = \lambda \sum_{D' \supset \{i,j\}} p_{D'} \int P\left(t + Y_i \in A_i, t + Y_j \in A_j \mid \mathbf{Y} \sim Q_{D'}\right) dt. \end{aligned} \tag{12}$$

In other words, the covariance of the counting processes is given by the weighted sum of the probabilities that the (i, j) marginal of the shift distributions yield values in the appropriate sets. The weights are the intensities of each corresponding component processes $\xi(D)$ which contribute events to both of the processes i and j.

In the case that $Q_D \equiv Q$, Equation (12) reduces to the solution given in Bäuerle and Grübel (2005). Using the tail probabilities defined in Equation (8), if $Q_D \equiv Q$ for all D, the integral in Equation (12) no longer depends on the subset D', and the equation may be written as

$$\begin{aligned} &\mathbf{cov}\big[X_i(A_i), X_j(A_j)\big] \\ &\quad = \lambda \bar{p}_{\{i,j\}} \int P\left(t + Y_i \in A_i, t + Y_j \in A_j \mid \mathbf{Y} \sim Q\right) dt. \end{aligned}$$

Using Equation (12), we may also compute the second cross-cumulant density (also called the *covariance density*) of the processes. From the definition of the cross-cumulant density [Equation (24) in the Methods], this is given by

$$\begin{aligned} \kappa_{ij}^{\mathbf{X}}(\tau) &= \lim_{\Delta t \to 0} \frac{\mathbf{cov}\big[X_i([0, \Delta t)), X_j([\tau, \tau + \Delta t))\big]}{\Delta t^2} \\ &= \lambda \sum_{D' \supset \{i,j\}} p_{D'} \\ &\quad \int \lim_{\Delta t \to 0} \frac{P\left(t + Y_i \in [0, \Delta t), t + Y_j \in [\tau, \tau + \Delta t) \mid \mathbf{Y} \sim Q_{D'}\right)}{\Delta t^2} dt. \end{aligned} \tag{13}$$

Before continuing, we note that given a random vector $\mathbf{Y} = (Y_1, \ldots, Y_N) \sim Q$, where Q has density $q(y_1, \ldots, y_N)$, the vector $\mathbf{Z} = (Y_2 - Y_1, \ldots, Y_N - Y_1)$ has density q_Z given by

$$q_Z(\tau_1, \ldots, \tau_{N-1}) = \int q(t, t+\tau_1, \ldots, t+\tau_{N-1})dt. \quad (14)$$

Assuming that the distributions $Q_{D'}$ have densities $q_{D'}$, and denoting by $q_{D'}^{\{i,j\}}$ the bivariate marginal density of the variables Y_i, Y_j under $Q_{D'}$, we have that

$$\kappa_{ij}^X(\tau) = \lambda \sum_{D' \supset \{i,j\}} p_{D'} \int q_{D'}^{\{i,j\}}(t, t+\tau)dt. \quad (15)$$

According to Equation (14), the integrals present in Equation (15) are simply the densities of the variables $Y_j - Y_i$, where $\mathbf{Y} \sim Q_{D'}$.

Thus $\kappa_{ij}^{\mathbf{X}}(\tau)$, which captures the additional probability for events in the marginal processes X_i and X_j separated by τ units of time beyond what can be predicted from lower order statistics is given by a weighted sum (in this case, the lower order statistics are marginal intensities—see the discussion around Equation (24) of the Methods). The weights are the "marking rates" $\lambda p_{D'}$ for markings contributing events to both component processes, while the summands are the probabilities that the corresponding shift distributions yield a pair of shifts in the proper arrangement—specifically, the shift applied to the event as attributed to X_i precedes that applied to the event mapped to X_j by τ units of time. This interpretation of the cross-cumulant density is quite natural, and will carry over to higher order cross-cumulants of a GTaS process. However, as we show next, this extension is not trivial at higher cumulant orders.

2.3.2. Third order cumulants

To determine the higher order cumulants for a GTaS process, one can again use the representation given in Equation (10). The distribution of a subset of three processes may be expressed in the form (see **Figure 6B**)

$$\begin{pmatrix} X_i(A_i) \\ X_j(A_j) \\ X_k(A_k) \end{pmatrix} =_{\text{distr}} \begin{pmatrix} \zeta_{\{i,j,k\}} + \zeta_{\{i,j\}} + \zeta_{\{i,k\}} + \zeta_{\{i\}} \\ \zeta_{\{i,j,k\}} + \zeta_{\{i,j\}} + \zeta_{\{j,k\}} + \zeta_{\{j\}} \\ \zeta_{\{i,j,k\}} + \zeta_{\{i,k\}} + \zeta_{\{j,k\}} + \zeta_{\{k\}}, \end{pmatrix}, \quad (16)$$

where, for simplicity, we suppressed the arguments of the different ζ_D on the right hand side. Again, the processes in the representation are independent and Poisson distributed. The variable $\zeta_{\{i,j,k\}}$ is the sum of all random variables $\xi(D)$ (see Equation 9) with $D \supset \{i,j,k\}$, while the variable $\zeta_{\{i,j\}}$ is now the sum of all $\xi(D)$ with $D \supset \{i,j\}$, but $k \notin D$. The rest of the variables are defined likewise. Using properties (C1) and (C2) of cumulants given in the Methods, and assuming that i, j, k are distinct indices, we have

$$\kappa(X_i(A_i), X_j(A_j), X_k(A_k)) = \kappa_3(\zeta_{\{i,j,k\}}) = \mathbf{E}\left[\zeta_{\{i,j,k\}}\right].$$

The second equality follows from the fact that all cumulants of a Poisson distributed random variable equal its mean. Similar to Equation (12), we may write

$$\kappa(X_i(A_i), X_j(A_j), X_k(A_k)) = \lambda \sum_{D' \supset \{i,j,k\}} p_{D'} \int P\left(t + Y_i \in A_i, \; t + Y_j \in A_j, t + Y_k \in A_k \,|\, \mathbf{Y} \sim Q_{D'}\right)dt.$$

The third cross-cumulant density is then given similarly to the second order function by

$$\kappa_{ijk}^{\mathbf{X}}(\tau_1, \tau_2) = \lambda \sum_{D' \supset \{i,j,k\}} p_{D'} \int q_{D'}^{\{i,j,k\}}(t, t+\tau_1, t+\tau_2)dt.$$

Here, we have again assumed the existence of densities $q_{D'}$, and denoted by $q_{D'}^{\{i,j,k\}}$ the joint marginal density of the variables Y_i, Y_j, Y_k under $q_{D'}$. The integrals appearing in the expression for the third order cross-cumulant density are the probability densities of the vectors $(Y_j - Y_i, Y_k - Y_i)$, where $\mathbf{Y} \sim Q_{D'}$.

2.3.3. General cumulants

Finally, consider a general subset of k distinct members of the vector counting process $\mathbf{X}$ as in Equation (10). The following theorem provides expressions for the cross-cumulants of the counting processes, as well as the cross-cumulant densities, in terms of model parameters in this general case. The proof of Theorem 1.1 is given in the Appendix.

Theorem 1.1. *Let $\mathbf{X}$ be a joint counting process of GTaS type with total intensity λ, marking distribution $(p_D)_{D \subset \mathbb{D}}$, and family of shift distributions $(Q_D)_{D \subset \mathbb{D}}$. Let $A_1, \ldots, A_k$ be arbitrary sets in $\mathcal{B}(\mathbb{R})$, and $\bar{D} = \{i_1, \ldots, i_k\} \subset \mathbb{D}$ with $|\bar{D}| = k$. The cross-cumulant of the counting processes may be written*

$$\kappa(X_{i_1}(A_1), \ldots, X_{i_k}(A_k)) = \lambda \sum_{D' \supset \bar{D}} p_{D'} \int P(t\mathbf{1} + \mathbf{Y}^{\bar{D}} \in A_1 \times \cdots \times A_k | \mathbf{Y} \sim Q_{D'})dt \quad (17)$$

where $\mathbf{Y}^{\bar{D}}$ represents the projection of the random vector $\mathbf{Y}$ onto the dimensions indicated by the members of the set $\bar{D}$. Furthermore, assuming that the shift distributions possess densities $(q_D)_{D \subset \mathbb{D}}$, the cross-cumulant density is given by

$$\kappa_{i_1 \cdots i_k}^X(\tau_1, \ldots, \tau_{k-1}) = \lambda \sum_{D' \supset \bar{D}} p_{D'} \int q_{D'}^{\bar{D}}(t, t+\tau_1, \cdots, t+\tau_{k-1})dt, \quad (18)$$

where $q_{D'}^{\bar{D}}$ indicates the kth order joint marginal density of $q_{D'}$ in the dimensions of $\bar{D}$.

An immediate corollary of Theorem 1.1 is a simple expression for the infinite-time-window cumulants, obtained by integrating

the cumulant density across all time lags τ_i. From Equation (A8), we have

$$\gamma^{\mathbf{X}}_{i_1\cdots i_k}(\infty) = \int\cdots\int \kappa^{X}_{i_1\cdots i_k}(\tau_1,\ldots,\tau_{k-1})d\tau_{k-1}\cdots d\tau_1$$
$$= \lambda \sum_{D'\supset \bar{D}} p_{D'}\cdot 1 = \lambda \bar{p}_{\bar{D}}. \quad (19)$$

This shows that the infinite time window cumulants for a GTaS process are non-increasing with respect to the ordering of sets, i.e.,

$$\gamma^{\mathbf{X}}_{i_1\cdots i_k}(\infty) \geq \gamma^{\mathbf{X}}_{i_1\cdots i_k i_{k+1}}(\infty).$$

We conclude this section with a short technical remark: Until this point, we have considered only the cumulant structure of sets of *unique* processes. However occasionally, one may wish to calculate a cumulant for a set of processes including repeats. Take, for example, a cumulant $\kappa(X_1(A_1), X_1(A_2), X_3(A_3))$. Owing to the marginally Poisson nature of the GTaS process, we would have (referring to the Methods for cumulant definitions)

$$\kappa(X_1(A_1), X_1(A_2), X_3(A_3))$$
$$= \kappa_{(2,1)}(X_1(A_1\cap A_2), X_3(A_3)) \quad \text{if} \quad \mathbf{X}\sim \text{GTaS}. \quad (20)$$

For a general counting process $\mathbf{X}$, it may be shown that

$$\kappa^{\mathbf{X}}_{113}(\tau_1,\tau_2) = \delta(\tau_1)\kappa^{\mathbf{X}}_{13}(\tau_2) + \text{ “non-singular contributions”}. \quad (21)$$

In addition, the second order auto-cumulant density may be written (Cox and Isham, 1980)

$$\kappa^{\mathbf{X}}_{ii}(\tau) = r_i\delta(\tau) + \text{ “non-singular contributions”},$$

where r_i is the stationary rate. The singular contribution shown in Equation (21) at third order is in analogy to the delta contribution proportional to the firing rate which appears in the second-order auto-cumulant density. For a GTaS process, the non-singular contributions in Equation (21) are identically zero, following directly from Equation (20). Expressions similar to Equations (20, 21) hold for general cases.

3. DISCUSSION

We have introduced a general method of generating spike trains with flexible spatiotemporal structure. The GTaS model is completely analytically tractable: all statistics of interest can be obtained directly from the distributions used to define it. It is based on an intuitive method of selecting and shifting point processes from a “mother” train. Moreover, the GTaS model can be used to easily generate partially synchronous states, cluster firing, cascading chains, and other spatiotemporal patterns of neural activity.

Processes generated by the GTaS model are naturally described by cumulant densities of pairwise and higher orders. This raises the question of whether such statistics are readily computable from data, so that realistic classes of GTaS models can be defined in the first place. One approach is to fit mechanistic models to data, and to use the higher order structure that is generated by the underlying mechanisms (Yu et al., 2011). A synergistic blend of other methods with the GTaS framework may also be fruitful—for example, the CuBIC framework of Staude et al. (2010) could be used to determine relevant marking orders, and the parametrically-described GTaS process could then be fit to allow generation of surrogate data after selection of appropriate classes of shift distributions. When it is necessary to infer higher order structure in the face of data limitations, population cumulants are an option to increase statistical power (albeit at the cost of spatial resolution; see **Figure 4**).

While the GTaS model has flexible higher order structure, it is always marginally Poisson. While throughout the cortex spiking is significantly irregular (Holt et al., 1996; Shadlen and Newsome, 1998), the level of variability differs across cells, with Fano factors ranging from below 0.5 to above 1.5—in comparison with the Poisson value of 1 (Churchland et al., 2010). Changes in variability may reflect cortical states and computation (Litwin-Kumar and Doiron, 2012; White et al., 2012). A model that would allow flexible marginal variability would therefore be very useful. Unfortunately, the tractability of the GTaS model is closely related to the fact that the marginal processes are Poisson. Therefore, an immediate generalization does not seem possible.

A number of other models have been used to describe population activity. Maximum entropy (ME) approaches also result in models with varied spatial activity; these are defined based on moments or other averaged features of multivariate spiking activity (Schneidman et al., 2006; Roudi et al., 2009). Such models are often used to fit purely spatial patterns of activity, though (Tang et al, 2008; Marre et al., 2009) have extended the techniques to treat temporal correlations as well. Generalized linear models (GLMs) have been used successfully to describe spatiotemporal patterns at second (Pillow et al., 2008), and third order (Ohiorhenuan et al., 2010). In comparison to the present GTaS method, both GLMs and ME models are more flexible. They feature well-defined approaches for fitting to data, including likelihood-based methods with well-behaved convexity properties. What the GTaS method contributes is an explicit way to generate population activity with explicitly specified high order spatio-temporal structure. Moreover, the lower order cumulant structure of a GTaS process can be modified independently of the higher order structure, though the reverse is not true.

There are a number of possible implications of such spatiotemporal structure for communication within neural networks. In section 2.2.3, we showed that these temporal correlations can play a role similar to that of spatial correlations established in Kuhn et al. (2003) for determining network input-output transfer. Our model allowed us to examine that impact of such temporal correlations on the network-level gain of a downstream population (cascade amplification factor). Even in a very simple network it was clear that the strength of the response is determined jointly by the temporal structure of the input to the network, and the connectivity within the network. Kuhn et al.

examined the effect of higher order structure on the firing rate gain of an integrate-and-fire neuron by driving it with a mixture of SIP or MIP processes (Kuhn et al., 2003). However, in these studies, only the spatial structure of higher order activity was varied. The GTaS model allows us to concurrently change the temporal structure of correlations. In addition, the precise control of the cumulants allows us to derive models which are equivalent up to a certain cross-cumulant order, when the configuration of marking probabilities and shift distributions allow it (as for the SIP and MIP processes of Kuhn et al. (2003), which are equivalent at second order).

Such patterns of activity may be useful when experimentally probing dendritic information processing (Gasparini and Magee, 2006), synaptic plasticity (Pfister and Gerstner, 2006; Gjorgjieva et al., 2011), or investigating the response of neuronal networks to complex patterns of input (Kahn et al., 2013). Spatiotemporal patterns may also be generated by cell assemblies (Bathellier et al., 2012). The firing in such assemblies can be spatially structured, and this structure may not be reflected in the activity of participating cells. Assemblies can exhibit persistent patterns of firing, sometimes with millisecond precision (Harris et al., 2002). The GTaS framework is well suited to describe exactly such activity patterns. The examples we presented can be easily extended to generate more complex patterns of activity with overlapping cell assemblies, different cells leading the activity, and other variations.

Understanding impact of spatiotemporal patterns on neural computations remains an open and exciting problem. Progress will require coordination of computational, theoretical, and experimental work—the latter taking advantage of novel stimulation techniques. We hope that the GTaS model, as a practical and flexible method for generating high-dimensional, correlated spike trains, will play a significant role along the way.

4. METHODS

4.1. CUMULANTS AS A MEASURE OF DEPENDENCE

We first define *cross-cumulants* (also called *joint cumulants*) (Stratonovich and Silverman, 1967; Kendall et al., 1969; Gardiner, 2009) and review some important properties of these quantities. Define the cumulant generating function g of a random vector $\mathbf{X} = (X_1, \ldots, X_N)$ by

$$g(t_1, \ldots, t_N) = \log\left(\mathbf{E}\left[\exp\left(\sum_{j=1}^{N} t_j X_j\right)\right]\right).$$

The $\mathbf{r}$-cross-cumulant of the vector $\mathbf{X}$ is given by

$$\kappa_{\mathbf{r}}(\mathbf{X}) = \frac{\partial^{|\mathbf{r}|}}{\partial t_1^{r_1} \cdots \partial t_N^{r_N}} g(t_1, \ldots, t_N)\Big|_{t_1 = \cdots = t_N = 0}.$$

where $\mathbf{r} = (r_1, \ldots, r_N)$ is a N-vector of positive integers, and $|\mathbf{r}| = \sum_{i=1}^{N} r_i$. We will generally deal with cumulants where all variables are considered at first order, without excluding the possibility that some variables are duplicated. In this case, we define the cross-cumulant $\kappa(\mathbf{X})$, of the variables in the random vector $\mathbf{X} = (X_1, \ldots, X_N)$ as

$$\kappa(\mathbf{X}) := \kappa_{\mathbf{1}}(\mathbf{X}) = \frac{\partial^N}{\partial t_1 \cdots \partial t_N} g(t_1, \ldots, t_N)\Big|_{t_1 = \cdots = t_N = 0}$$

where $\mathbf{1} = (1, \ldots, 1)$.

This relationship may be expressed in combinatorial form:

$$\kappa(X_1, \ldots, X_N) = \sum_{\pi} (|\pi| - 1)!(-1)^{|\pi|-1} \prod_{B \in \pi} \mathbf{E}\left[\prod_{i \in B} X_i\right] \tag{22}$$

where π runs through all partitions of $\mathbb{D} = \{1, \ldots, N\}$, and B runs over all blocks in a partition π. More generally, the $\mathbf{r}$-cross-cumulant may be expressed in terms of moments by expanding the cumulant generating function as a Taylor series, noting that

$$g(t_1, \ldots, t_N) = \sum_{\mathbf{r}} \frac{\kappa_{\mathbf{r}}(X_1, \ldots, X_N)}{\mathbf{r}!} x_1^{r_1} \cdots x_d^{r_N} \quad \text{with}$$

$$\mathbf{r}! = \prod_{i=1}^{N} r_i!,$$

similarly expanding the moment generating function $M(t) = e^{g(t)}$, and matching the polynomial coefficients. Note that the nth cumulant κ_n of a random variable X may be expressed as a joint cumulant via

$$\kappa_n(X) = \kappa(\underbrace{X, \ldots, X}_{\text{n copies of } X}).$$

We will utilize the following two principal properties of cumulants (Brillinger, 1965; Stratonovich and Silverman, 1967; Mendel, 1991; Staude et al., 2010):

(C1) Multilinearity - for any random variables X, Y, $\{Z_i\}_{i=2}^{N}$, we have

$$\kappa(aX + bY, Z_2, \ldots, Z_N) = a\kappa(X, Z_2, \ldots, Z_N) + b\kappa(Y, Z_2, \ldots, Z_N).$$

This holds regardless of dependencies amongst the random variables.

(C2) If any subset of the random variables in the cumulant argument is independent from the remaining, the cross-cumulant is zero—i.e., if $\{X_1, \ldots, X_{N_1}\}$ and $\{Y_1, \ldots, Y_{N_2}\}$ are sets of random variables such that each X_i is independent from each Y_j, then

$$\kappa_{(\mathbf{r}_X, \mathbf{r}_Y)}(X_1, \ldots, X_{N_1}, Y_1, \ldots, Y_{N_2}) = 0 \quad \text{for all } \mathbf{r}_X \in \mathbb{N}_+^{N_1}, \mathbf{r}_Y \in \mathbb{N}_+^{N_2}.$$

To exhibit another key property of cumulants, consider a 4-vector $\mathbf{X} = (X_1, X_2, X_3, X_4)$ with non-zero fourth cumulant and a random variable Z independent of each X_i. Define $\mathbf{Y} = (X_1 +$

$Z, X_2 + Z, X_3 + Z, X_4$). Using properties (C1), (C2) above, it follows that

$$\kappa(Y_1, Y_2, Y_3) = \kappa(X_1, X_2, X_3) + \kappa_3(Z).$$

On the other hand, it is also true that

$$\kappa(\mathbf{Y}) = \kappa(\mathbf{X}),$$

that is, adding the variable Z to only a subset of the variables in $\mathbf{X}$ results in changes to cumulants involving only that subset, but *not* to the joint cumulant of the entire vector. In this sense, an rth order cross-cumulant of a collection of random variables captures exclusively dependencies amongst the collection which cannot be described by cumulants of lower order. In the example above, only the joint statistical properties of a subset of $\mathbf{X}$ were changed. As a result, the total cumulant $\kappa(\mathbf{X})$ remained fixed.

From Equation (22), it is apparent that $\kappa(X_i) = \mathbf{E}[X_i]$, and $\kappa(X_i, X_j) = \mathbf{cov}[X_i, X_j]$. In addition, the third cumulant, like the second, is equal to the corresponding central moment:

$$\kappa(X_i, X_j, X_k) = \mathbf{E}\left[(X_i - \mathbf{E}[X_i])(X_j - \mathbf{E}[X_j])(X_k - \mathbf{E}[X_k])\right].$$

As cumulants and central moments agree up to third order, central moments up to third order inherit the properties discussed above at these orders. On the other hand, the fourth cumulant is *not* equal to the fourth central moment. Rather:

$$\begin{aligned}&\kappa(X_i, X_j, X_k, X_l)\\ &\quad = E[(X_i - \mathbf{E}[X_i])(X_j - \mathbf{E}[X_j])(X_k - \mathbf{E}[X_k])(X_l - \mathbf{E}[X_l])]\\ &-\mathbf{cov}[X_i, X_j]\,\mathbf{cov}[X_k, X_l] - \mathbf{cov}[X_i, X_k]\,\mathbf{cov}[X_j, X_l]\\ &\quad -\mathbf{cov}[X_i, X_l]\,\mathbf{cov}[X_j, X_k].\end{aligned} \tag{23}$$

Higher cumulants have similar (but more complicated) expansions in terms of central moments. Accordingly, central moments of fourth and higher order do not inherit properties (C1), (C2).

4.2. TEMPORAL STATISTICS OF POINT PROCESSES

In the Results, we present an extension of previous work (Bäuerle and Grübel, 2005) in which we construct and analyze multivariate counting processes $\mathbf{X} = (X_1, \ldots, X_N)$ where each X_i is marginally Poisson.

Formally, a counting process $\mathbf{X}$ is an integer-valued random measure on $\mathcal{B}(\mathbb{R}^N)$. Evaluated on subset $A_1 \times \cdots \times A_N$ of $\mathcal{B}(\mathbb{R}^N)$, the random vector $(X_1(A_1), \ldots, X_N(A_N))$ counts events in d distinct categories whose times of occurrence fall in to the sets A_i. A good general reference on the properties of counting processes (marginally Poisson and otherwise) is Daley and Vere-Jones (2002).

The assumption of Poisson marginals implies that for a set $A_i \in \mathcal{B}(\mathbb{R})$, the random variable $X_i(A_i)$ follows a Poisson distribution with mean $\lambda_i \ell(A_i)$, where ℓ is the Lebesgue measure on $\mathbb{R}$, and λ_i is the (constant) rate for the ith process. The processes under consideration will further satisfy a joint stationarity condition, namely that the distribution of the vector $(X_1(A_1 + t), \ldots, X_N(A_N + t))$ does not depend on t, where $A_i + t$ denotes the translated set $\{a + t : a \in A_i\}$.

We now consider some common measures of temporal dependence for jointly stationary vector counting processes. We will refer to the quantity $X_i[0, T]$ as the *spike count* of process i over $[0, T]$. The quantity $\gamma^{\mathbf{X}}_{i_1 \cdots i_k}(T)$ (which we will refer to as a *spike count cumulant*) is given by

$$\gamma^{\mathbf{X}}_{i_1 \cdots i_k}(T) = \frac{1}{T}\kappa[X_{i_1}[0, T], \ldots, X_{i_k}[0, T]]$$

measures kth order correlations amongst spike counts for the listed processes which occur over windows of length T. At second order, $\gamma^{\mathbf{X}}_{ij}(T)$ measures the covariance of the spike counts of processes i, j over a common window of length T. The infinite window spike count cumulant quantifies dependencies in the spike counts of point processes over arbitrarily long windows, and is given by

$$\gamma^{\mathbf{X}}_{i_1 \cdots i_k}(\infty) = \lim_{T \to \infty} \gamma^{\mathbf{X}}_{i_1 \cdots i_k}(T).$$

A related measure is the kth order cross-cumulant density $\kappa^{X}_{i_1, \ldots, i_k}(\tau_1, \ldots, \tau_{k-1})$, defined by

$$\begin{aligned}\kappa^{\mathbf{X}}_{i_1 \cdots i_k}(\tau_1, \ldots, \tau_{k-1}) = \lim_{\Delta t \to 0} \frac{1}{\Delta t^k}\kappa[X_{i_1}[0, \Delta t],\\ X_{i_2}[\tau_1, \tau_1 + \Delta t], \ldots, X_{i_k}[\tau_{k-1}, \tau_{k-1} + \Delta t]].\end{aligned} \tag{24}$$

The cross-cumulant density should be interpreted as a measure of the likelihood—above what may be expected from knowledge of the lower order cumulant structure—of seeing events in processes $i_2, \ldots, i_k$ at times $\tau_1 + t, \ldots, \tau_{k-1} + t$, conditioned on event in process i_1 at time t. The infinite window spike count cumulant is equal to the total integral under the cross-cumulant density,

$$\gamma^{\mathbf{X}}_{i_1 \cdots i_k}(\infty) = \int \cdots \int \kappa^{\mathbf{X}}_{i_1 \cdots i_k}(\tau_1, \ldots, \tau_{k-1}) d\tau_{k-1} \cdots d\tau_1.$$

As an example, we again consider the familiar second-order cross-cumulant density $\kappa^{X}_{ij}(\tau)$—often referred to as the *cross-covariance density* or *cross-correlation function*. Defining the conditional intensity $h_{ij}(\tau)$ of process j, conditioned on process i to be

$$h^{X}_{ij}(\tau) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} P(X_j[\tau, \tau + \Delta t] > 0 | X_i[0, \Delta t] > 0),$$

that is, the intensity of j conditioned on an event in process i which occurred τ units of time in the past, then it is not difficult to show that

$$\kappa^{X}_{ij}(\tau) = \lambda_i h_{ij}(\tau) - \lambda_i \lambda_j.$$

That is, the second order cross-cumulant density supplies the probability of chance of observing an event attributed to process i, followed by one attributed to process j, τ units of time later, above what would be expected from knowledge of first order statistics (given by the product of the marginal intensities, $\lambda_i\lambda_j$). More generally, at higher orders, the cross-cumulant density should be interpreted as a measure of the likelihood (above what may be expected from knowledge of the lower order correlation structure) of seeing events attribute to processes $i_2, \ldots, i_k$ at times $\tau_1 + t, \ldots, \tau_{k-1} + t$, conditioned on an event in process i_1 at time t.

Another statistic useful in the study of a correlated vector counting process $\mathbf{X}$ is the *population cumulant density*. At second-order, the population cumulant density for X_i takes the form (Luczak et al., 2013)

$$\kappa^{\mathbf{X}}_{i,\,\mathrm{pop}}(\tau) = \sum_{j \neq i} \kappa^{\mathbf{X}}_{ij}(\tau).$$

More generally, the kth order population cumulant density corresponding to the processes $X_{i_1}, \ldots, X_{i_{k-1}}$ is given by

$$\kappa^{\mathbf{X}}_{i_1 \cdots i_{k-1},\mathrm{pop}}(\tau_1, \ldots, \tau_{k-1}) = \sum_{j \neq i_1,\ldots,i_k} \kappa^{\mathbf{X}}_{i_1 \cdots i_{k-1} j}(\tau_1, \ldots, \tau_{k-1}). \quad (25)$$

FUNDING

This work was supported by NSF grants DMS-0817649, DMS-1122094, a Texas ARP/ATP award to Krešimir Josić, and by a Career Award at the Scientific Interface from the Burroughs Wellcome Fund and NSF Grant DMS-1122106 to Eric Shea-Brown.

REFERENCES

Abeles, M. (1991). *Corticonics: Neural Circuits of the Cerebral Cortex*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511574566

Abeles, M., and Prut, Y. (1996). Spatio-temporal firing patterns in the frontal cortex of behaving monkeys. *J. Physiol. Paris* 90, 249–250. doi: 10.1016/S0928-4257(97)81433-7

Aertsen, A., Diesmann, M., and Gewaltig, M. O. (1996). Propagation of synchronous spiking activity in feedforward neural networks. *J. Physiol. Paris* 90, 243–247. doi: 10.1016/S0928-4257(97)81432-5

Amari, S., Nakahara, H., Wu, S., and Sakai, Y. (2003). Synchronous firing and higher-order interactions in neuron pool. *Neural Comput.* 15, 127–142. doi: 10.1162/089976603321043720

Amjad, A. M., Halliday, D. M., Rosenberg, J. R., and Conway, B. A. (1997). An extended difference of coherence test for comparing and combining several independent coherence estimates: theory and application to the study of motor units and physiological tremor. *J. Neurosci. Meth.* 73, 69–79. doi: 10.1016/S0165-0270(96)02214-5

Averbeck, B. B., Latham, P. E., and Pouget, A. (2006). Neural correlations, population coding and computation. *Nat. Rev. Neurosci.* 7, 358–366. doi: 10.1038/nrn1888

Aviel, Y., Pavlov, E., Abeles, M., and Horn, D. (2002). Synfire chain in a balanced network. *Neurocomputing* 44, 285–292. doi: 10.1016/S0925-2312(02)00352-1

Bair, W., Zohary, E., and Newsome, W. (2001). Correlated firing in macaque visual area mt: time scales and relationship to behavior. *J. Neurosci.* 21, 1676–1697.

Barreiro, A. K., Shea-Brown, E., and Thilo, E. L. (2010). Time scales of spike-train correlation for neural oscillators with common drive. *Phys. Rev. E* 81:011916. doi: 10.1103/PhysRevE.81.011916

Bathellier, B., Ushakova, L., and Rumpel, S. (2012). Discrete neocortical dynamics predict behavioral categorization of sounds. *Neuron* 76, 435–449. doi: 10.1016/j.neuron.2012.07.008

Bäuerle, N., and Grübel, R. (2005). Multivariate counting processes: copulas and beyond. *Astin Bull.* 35, 379. doi: 10.2143/AST.35.2.2003459

Branco, T., Clark, B. A., and Hausser, M. (2010). Dendritic discrimination of temporal input sequences in cortical neurons. *Sci. Signal.* 329, 1671. doi: 10.1126/science.1189664

Branco, T., and Häusser, M. (2011). Synaptic integration gradients in single cortical pyramidal cell dendrites. *Neuron* 69, 885–892. doi: 10.1016/j.neuron.2011.02.006

Brette, R. (2009). Generation of correlated spike trains. *Neural Comput.* 21, 188–215. doi: 10.1162/neco.2009.12-07-657

Brillinger, D. R. (1965). An introduction to polyspectra. *Ann. Math. Stat.* 36, 1351–1374. doi: 10.1214/aoms/1177699896

Buzsáki, G. (2010). Neural syntax: cell assemblies, synapsembles, and readers. *Neuron* 68, 362–385. doi: 10.1016/j.neuron.2010.09.023

Cain, N., and Shea-Brown, E. (2013). Impact of correlated neural activity on decision-making performance. *Neural Comput.* 25, 289–327. doi: 10.1162/NECO_a_00398

Carr, C. E., Agmon-Snir, H., and Rinzel, J. (1998). The role of dendrites in auditory coincidence detection. *Nature* 393, 268–272. doi: 10.1038/30505

Chow, B. Y., Han, X., Dobry, A. S., Qian, X., Chuong, A. S., Li, M., et al. (2010). High-performance genetically targetable optical neural silencing by light-driven proton pumps. *Nature* 463, 98–102. doi: 10.1038/nature08652

Churchland, M. M., Yu, B. M., Cunningham, J. P., Sugrue, L. P., Cohen, M. R., Corrado, G. S., et al. (2010). Stimulus onset quenches neural variability: a widespread cortical phenomenon. *Nat. Neurosci.* 13, 369–378. doi: 10.1038/nn.2501

Cox, D. R., and Isham, V. (1980). *Point Processes*. Vol. 12. London: Chapman and Hall/CRC.

Daley, D. J., and Vere-Jones, D. (2002). *An Introduction to the Theory of Point Processes: Volume I: Elementary Theory and Methods*. Vol. 1. New York, NY: Springer.

Ganmor, E., Segev, R., and Schneidman, E. (2011). Sparse low-order interaction network underlies a highly correlated and learnable neural population code. *Proc. Natl. Acad. Sci. U.S.A.* 108, 9679–9684. doi: 10.1073/pnas.1019641108

Gardiner, C. W. (2009). *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences*. Berlin: Springer-Verlag.

Gasparini, S., and Magee, J. C. (2006). State-dependent dendritic computation in hippocampal CA1 pyramidal neurons. *J. Neurosci.* 26, 2088–2100. doi: 10.1523/JNEUROSCI.4428-05.2006

Gjorgjieva, J., Clopath, C., Audet, J., and Pfister, J.-P. (2011). A triplet spike-timing–dependent plasticity model generalizes the bienenstock–cooper–munro rule to higher-order spatiotemporal correlations. *Proc. Natl. Acad. Sci. U.S.A.* 108, 19383–19388. doi: 10.1073/pnas.1105933108

Grün, S., and Rotter, S. (2010). *Analysis of Parallel Spike Trains*. New York, NY: Springer. doi: 10.1007/978-1-4419-5675-0

Gütig, R., and Sompolinsky, H. (2006). The tempotron: a neuron that learns spike timing–based decisions. *Nat. Neurosci.* 9, 420–428. doi: 10.1038/nn1643

Gutnisky, D. A., and Josić, K. (2009). Generation of spatiotemporally correlated spike-trains and local-field potentials using a multivariate autoregressive process. *J. Neurophysiol.* 103, 2912–2930. doi: 10.1152/jn.00518.2009

Han, X., and Boyden, E. S. (2007). Multiple-color optical activation, silencing, and desynchronization of neural activity, with single-spike temporal resolution. *PLoS ONE* 2:e299. doi: 10.1371/journal.pone.0000299

Hansen, B. J., Chelaru, M. I., and Dragoi, V. (2012). Correlated variability in laminar cortical circuits. *Neuron* 76, 590–602. doi: 10.1016/j.neuron.2012.08.029

Harris, K. D. (2005). Neural signatures of cell assembly organization. *Nat. Rev. Neurosci.* 6, 399–407. doi: 10.1038/nrn1669

Harris, K. D., Henze, D. A., Hirase, H., Leinekugel, X., Dragoi, G., Czurkó,

A., et al. (2002). Spike train dynamics predicts theta-related phase precession in hippocampal pyramidal cells. *Nature* 417, 738–741. doi: 10.1038/nature00808

Hebb, D. O. (1949). *The Organization of Behavior: A Neuropsychological Theory*. New York, NY: Psychology Press.

Holt, G. R., Softky, W. R., Koch, C., and Douglas, R. J. (1996). Comparison of discharge variability *in vitro* and *in vivo* in cat visual cortex neurons. *J. Neurophysiol.* 75, 1806–1814.

Hopfield, J. J. (1995). Pattern recognition computation using action potential timing for stimulus representation. *Nature* 376, 33–36. doi: 10.1038/376033a0

Ikegaya, Y., Aaron, G., Cossart, R., Aronov, D., Lampl, I., Ferster, D., et al. (2004). Synfire chains and cortical songs: temporal modules of cortical activity. *Sci. Signal.* 304, 559. doi: 10.1126/science.1093173

Izhikevich, E. M. (2006). Polychronization: computation with spikes. *Neural Comput.* 18, 245–282. doi: 10.1162/089976606775093882

Jeffress, L. A. (1948). A place theory of sound localization. *J. Comp. Physiol. Psychol.* 41, 35–39. doi: 10.1037/h0061495

Johnson, D. H., and Goodman, I. N. (2009). Jointly Poisson processes. *arXiv preprint* arXiv:0911.2524.

Joris, P. X., Smith, P. H., and Yin, T. C. T. (1998). Coincidence detection in the auditory system: 50 years after Jeffress. *Neuron* 21, 1235–1238. doi: 10.1016/S0896-6273(00)80643-1

Kahn, I., Knoblich, U., Desai, M., Bernstein, J., Graybiel, A. M., Boyden, E. S., et al. (2013). Optogenetic drive of neocortical pyramidal neurons generates fMRI signals that are correlated with spiking activity. *Brain Res.* 1511, 33–45. doi: 10.1016/j.brainres.2013.03.011

Kendall, M. G., Stuart, A., and Ord, J. K. (1969). *The Advanced Theory of Statistics*. Vol. 1, 3rd Edn. London: Griffin.

Köster, U., Sohl-Dickstein, J., Gray, C. M., and Olshausen, B. A. (2013). Higher order correlations within cortical layers dominate functional connectivity in microcolumns. *arXiv preprint* arXiv:1301.0050.

Krumin, M., and Shoham, S. (2009). Generation of spike trains with controlled auto- and cross-correlation functions. *Neural Comput.* 21, 1642–1664. doi: 10.1162/neco.2009.08-08-847

Kuhn, A., Aertsen, A., and Rotter, S. (2003). Higher-order statistics of input ensembles and the response of simple model neurons. *Neural Comput.* 15, 67–101. doi: 10.1162/089976603321043702

Litwin-Kumar, A., and Doiron, B. (2012). Slow dynamics and high variability in balanced cortical networks with clustered connections. *Nat. Neurosci.* 15, 1498–1505. doi: 10.1038/nn.3220

Luczak, A., Bartho, P., and Harris, K. D. (2013). Gating of sensory input by spontaneous cortical activity. *J. Neurosci.* 33, 1684–1695. doi: 10.1523/JNEUROSCI.2928-12.2013

Luczak, A., Barthó, P., Marguet, S. L., Buzsáki, G., and Harris, K. D. (2007). Sequential structure of neocortical spontaneous activity *in vivo*. *Proc. Natl. Acad. Sci. U.S.A.* 104, 347–352. doi: 10.1073/pnas.0605643104

Macke, J. H., Berens, P., Ecker, A. S., Tolias, A. S., and Bethge, M. (2009). Generating spike trains with specified correlation coefficients. *Neural Comput.* 21, 397–423. doi: 10.1162/neco.2008.02-08-713

Macke, J. H., Opper, M., and Bethge, M. (2011). Common input explains higher-order correlations and entropy in a simple model of neural population activity. *Phys. Rev. Lett.* 106:208102. doi: 10.1103/PhysRevLett.106.208102

Marre, O., El Boustani, S., Frégnac, Y., and Destexhe, A. (2009). Prediction of spatiotemporal patterns of neural activity from pairwise correlations. *Phys. Rev. Lett.* 102:138101. doi: 10.1103/PhysRevLett.102.138101

Meister, M., and Berry II, M. J. (1999). The neural code of the retina. *Neuron* 22, 435. doi: 10.1016/S0896-6273(00)80700-X

Mendel, J. M. (1991). Tutorial on higher-order statistics (spectra) in signal processing and system theory: theoretical results and some applications. *Proc. IEEE* 79, 278–305. doi: 10.1109/5.75086

Montani, F., Phoka, E., Portesi, M., and Schultz, S. R. (2013). Statistical modelling of higher-order correlations in pools of neural activity. *Physica A* 392, 3066–3086. doi: 10.1016/j.physa.2013.03.012

Ohiorhenuan, I. E., Mechler, F., Purpura, K. P., Schmid, A. M., Hu, Q., and Victor, J. D. (2010). Sparse coding and high-order correlations in fine-scale cortical networks. *Nature* 466, 617–621. doi: 10.1038/nature09178

Pfister, J.-P., and Gerstner, W. (2006). Triplets of spikes in a model of spike timing-dependent plasticity. *J. Neurosci.* 26, 9673–9682. doi: 10.1523/JNEUROSCI.1425-06.2006

Pillow, J. W., Shlens, J., Paninski, L., Sher, A., Litke, A. M., Chichilnisky, E. J., et al. (2008). Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature* 454, 995–999. doi: 10.1038/nature07140

Reddy, G. D., Kelleher, K., Fink, R., and Saggau, P. (2008). Three-dimensional random access multiphoton microscopy for functional imaging of neuronal activity. *Nat. Neurosci.* 11, 713–720. doi: 10.1038/nn.2116

Rosenbaum, R. J., Trousdale, J., and Josić, K. (2010). Pooling and correlated neural activity. *Front. Comput. Neurosci.* 4:9. doi: 10.3389/fncom.2010.00009

Rosenbaum, R. J., Trousdale, J., and Josić K. (2011). The effects of pooling on spike train correlations. *Front. Neurosci.* 5:58. doi: 10.3389/fnins.2011.00058

Ross, S. M. (1995). *Stochastic Processes*. 2nd Edn. New York, NY: Wiley.

Roudi, Y., Nirenberg, S., and Latham, P. E. (2009). Pairwise maximum entropy models for studying large biological systems: when they can work and when they can't. *PLoS Comput. Biol.* 5:e1000380. doi: 10.1371/journal.pcbi.1000380

Salinas, E., and Sejnowski, T. J. (2001). Correlated neuronal activity and the flow of neural information. *Nat. Rev. Neurosci.* 2, 539–550. doi: 10.1038/35086012

Schneidman, E., Berry, M. J., Segev, R., and Bialek, W. (2006). Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature* 440, 1007–1012. doi: 10.1038/nature04701

Shadlen, M. N., and Newsome, W. T. (1998). The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. *J. Neurosci.* 18, 3870–3896.

Shamir, M., and Sompolinsky, H. (2004). Nonlinear population codes. *Neural Comput.* 16, 1105–1136. doi: 10.1162/089976604773717559

Shimazaki, H., Amari, S.-i., Brown, E. N., and Grün, S. (2012). State-space analysis of time-varying higher-order spike correlation for multiple neural spike train data. *PLoS Comp. Biol.* 8:e1002385. doi: 10.1371/journal.pcbi.1002385

Shlens, J., Field, G. D., Gauthier, J. L., Greschner, M., Sher, A., Litke, A. M., et al. (2009). The structure of large-scale synchronized firing in primate retina. *J. Neurosci.* 29, 5022–5031. doi: 10.1523/JNEUROSCI.5187-08.2009

Shlens, J., Field, G. D., Gauthier, J. L., Grivich, M. I., Petrusca, D., Sher, A., et al. (2006). The structure of multi-neuron firing patterns in primate retina. *J. Neurosci.* 26, 8254–8266. doi: 10.1523/JNEUROSCI.1282-06.2006

Singer, W. (1999). Neuronal synchrony: a versatile code review for the definition of relations? *Neuron* 24, 49–65. doi: 10.1016/S0896-6273(00)80821-1

Staude, B., Rotter, S., and Grün, S. (2010). CuBIC: cumulant based inference of higher-order correlations in massively parallel spike trains. *J. Comput. Neurosci.* 29, 327–350. doi: 10.1007/s10827-009-0195-x

Stratonovich, R. L., and Silverman, R. A. (1967). *Topics in the Theory of Random Noise*. Vol. 2. New York, NY: Gordon and Breach.

Tang, A., Jackson, D., Hobbs, J., Chen, W., Smith, J. L., Patel, H., et al. (2008). A maximum entropy model applied to spatial and temporal correlations from cortical networks *in vitro*. *J. Neurosci.* 28, 505–518. doi: 10.1523/JNEUROSCI.3359-07.2008

Tetzlaff, T., Buschermöhle, M., Geisel, T., and Diesmann, M. (2003). The spread of rate and correlation in stationary cortical networks. *Neurocomputing* 52, 949–954. doi: 10.1016/S0925-2312(02)00854-8

Tetzlaff, T., Rotter, S., Stark, E., Abeles, M., Aertsen, A., and Diesmann, M. (2008). Dependence of neuronal correlations on filter characteristics and marginal spike train statistics. *Neural Comput.* 20, 2133–2184. doi: 10.1162/neco.2008.05-07-525

Thorpe, S., Delorme, A., and van Rullen, R. (2001). Spike-based strategies for rapid processing. *Neural Netw.* 14, 715–725. doi: 10.1016/S0893-6080(01)00083-1

Vasquez, J. C., Marre, O., Palacios, A. G., Berry, M. J. II., and Cessac, B. (2012). Gibbs distribution analysis of temporal correlations structure in retina ganglion cells. *J. Physiol. Paris* 106, 120–127. doi: 10.1016/j.jphysparis.2011.11.001

White, B., Abbott, L. F., and Fiser, J. (2012). Suppression of

cortical neural variability is stimulus- and state-dependent. *J. Neurophysiol.* 108, 2383–2392. doi: 10.1152/jn.00723.2011

Xu, N., Harnett, M. T., Williams, S. R., Huber, D., O'Connor, D. H., Svoboda, K., et al. (2012). Nonlinear dendritic integration of sensory and motor input during an active sensing task. *Nature* 492, 247–251. doi: 10.1038/nature 11601

Yu, S., Yang, H., Nakahara, H., Santos, G. S., Nikolić, D., and Plenz, D. (2011). Higher-order interactions characterized in cortical activity. *J. Neurosci.* 31, 17514–17526. doi: 10.1523/JNEUROSCI.3127-11. 2011

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.

Received: 10 May 2013; paper pending published: 31 May 2013; accepted: 12 June 2013; published online: 17 July 2013.

Citation: Trousdale J, Hu Y, Shea-Brown E and Josić K (2013) A generative spike train model with time-structured higher order correlations. Front. Comput. Neurosci. 7:84. doi: 10.3389/fncom. 2013.00084

Copyright © 2013 Trousdale, Hu, Shea-Brown and Josić. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and subject to any copyright notices concerning any third-party graphics etc.

APPENDIX

PROOF OF THE DISTRIBUTIONAL REPRESENTATION OF THE GTaS MODEL IN EQUATION (9)

The construction of the GTaS model allows us to provide a useful distributional representation of the process. We describe this representation in a theorem that generalizes Theorem 1 in Bäuerle and Grübel (2005). This theorem also immediately implies that the GTaS process is marginally Poisson.

Some definitions are required: first, for subsets $A_1, \ldots, A_N \in \mathcal{B}(\mathbb{R})$ and $D, D' \subset \mathbb{D}$ with $D \subset D'$, let

$$M(D, D'; A_1, \ldots, A_N) := B_1 \times \cdots \times B_N \text{ with } B_i := \begin{cases} A_i, & \text{for } i \in D, \\ A_i^c, & \text{for } i \in D' \backslash D, \\ \mathbb{R}, & \text{otherwise} \end{cases}$$

In addition, setting $\mathbf{1} = (1, \ldots, 1)$ to be the N-dimensional vector with all components equal to unity, and if Q_D is a measure on $\mathbb{R}^N$, then we define the measure $\nu(Q_D)$ by

$$\begin{aligned} \nu(Q_D)(A) &:= \int Q_D(A - t\mathbf{1})dt \quad \text{for } A \in \mathcal{B}(\mathbb{R}^N) \\ &= \int P(\mathbf{Y} + t\mathbf{1} \in A | \mathbf{Y} \sim Q_D)dt. \end{aligned} \tag{A1}$$

The measure $\nu(Q_D)$ can be interpreted as giving the *expected* Lebesgue measure of the subset L of $\mathbb{R}$ for which uniform shifts by the elements of L translate a random vector $\mathbf{Y} \sim Q_D$ in to A. Heuristically, one may imagine sliding the vector $\mathbf{Y}$ over the whole real line, and counting the number of times every coordinate ends up in the "right" set—the projection of A onto that dimension. In equation form, this means

$$\nu(Q_D)(A) = \mathbf{E}_{\mathbf{Y}}[\ell(\{t \in \mathbb{R} : \mathbf{Y} + t\mathbf{1} \in A\}) | \mathbf{Y} \sim Q_D]. \tag{A2}$$

where the subscript $\mathbf{Y}$ indicates that we take the average over the distribution of $\mathbf{Y} \sim Q_D$. A short proof of this representation is presented below. We now present the theorem, with a proof indicating adjustments necessary to that of Bäuerle and Grübel (2005).

Theorem 0 *Let X be an N-dimensional counting process of GTaS type with base rate λ, thinning mechanism $p = (p_D)_{D \subset \mathbb{D}}$, and family of shift distributions $(Q_D)_{D \subset \mathbb{D}}$. Then, for any Borel subsets $A_1, \ldots, A_N$ of the real line, we have the following distributional representation:*

$$\begin{pmatrix} X_1(A_1) \\ \vdots \\ X_N(A_N) \end{pmatrix} =_{\text{distr}} \begin{pmatrix} \sum_{D \ni 1} \xi(D; A_1, \ldots, A_N) \\ \vdots \\ \sum_{D \ni d} \xi(D; A_1, \ldots, A_N) \end{pmatrix}, \tag{A3}$$

where the random variables $\xi(D; A_1, \ldots, A_N), \emptyset \neq D \subset \mathbb{D}$, are independent and Poisson distributed with

$$\mathbf{E}[\xi(D; A_1, \ldots, A_N)] = \lambda \sum_{D' \supset D} p_{D'} \nu(Q_{D'})(M(D, D'; A_1, \ldots, A_N)).$$

Proof. For each marking $D' \subset \mathbb{D}$, define $\mathbf{X}^{D'}$ to be an independent TaS (Bäuerle and Grübel, 2005) counting process with mother process rate $\lambda p_{D'}$, shift distribution $Q_{D'}$, and markings $(p_D^{D'})_{D \subset \mathbb{D}}$ where $p_D^{D'} = 1$ if $D = D'$ and is zero otherwise (i.e., the only possible marking for $\mathbf{X}^{D'}$ is D'). We first claim that

$$\mathbf{X} =_{\text{distr}} \sum_{D'} \mathbf{X}^{D'}. \tag{A4}$$

To see this, note that spikes in the mother process of the GTaS process of $\mathbf{X}$ marked for a set D' occur at a rate $\lambda p_{D'}$, which is the rate of the process $\mathbf{X}^{D'}$. In addition, these event times are then shifted by $Q_{D'}$, exactly as they are for $\mathbf{X}^{D'}$. Thus, the distribution of event times (and hence the counting process distributions) are equivalent.

Let $A_1, \ldots, A_N$ be any Borel subsets of the real line. Applying Theorem 1 of Bäuerle and Grübel (2005) to each $\mathbf{X}^{D'}$ gives the following distributional representation:

$$\begin{pmatrix} X_1^{D'}(A_1) \\ \vdots \\ X_N^{D'}(A_N) \end{pmatrix} =_{\text{distr}} \begin{pmatrix} \sum_{D \ni 1} \xi^{D'}(D; A_1, \ldots, A_N) \\ \vdots \\ \sum_{D \ni N} \xi^{D'}(D; A_1, \ldots, A_N) \end{pmatrix}, \tag{A5}$$

where the random variables $\xi^{D'}(D; , A_1, \ldots, A_N)$ are taken to be identically zero unless $D \subset D'$. In the latter case, they are independent and Poisson distributed with

$$\begin{aligned} &\mathbf{E}\left[\xi^{D'}(D; A_1, \ldots, A_N)\right] \\ &\quad = \lambda p_{D'} \sum_{D'' \supset D} p_{D''}^{D'} \nu(Q_{D'})(M(D, D''; A_1, \ldots, A_N)) \\ &\quad = \lambda p_{D'} \nu(Q_{D'})(M(D, D'; A_1, \ldots, A_N)). \end{aligned}$$

The second equality above follows from the fact that $p_{D''}^{D'} = 1$ if $D'' = D'$ and is zero otherwise.

Next, define

$$\begin{aligned} \xi(D; A_1, \ldots, A_N) &= \sum_{D'} \xi^{D'}(D; A_1, \ldots, A_N) \\ &= \sum_{D' \supset D} \xi^{D'}(D; A_1, \ldots, A_N). \end{aligned}$$

As the sum of independent Poisson variables is again Poisson with rate equal to the sum of the rates, we have that $\xi(D; A_1, \ldots, A_N)$ is Poisson with mean

$$\mathbf{E}[\xi(D; A_1, \ldots, A_N)] = \lambda \sum_{D' \supset D} p_{D'} \nu(Q_{D'})(M(D, D'; A_1, \ldots, A_N)). \tag{A6}$$

Finally, combining Equations (A4, A5), we may write

$$\begin{pmatrix} X_1(A_1) \\ \vdots \\ X_N(A_N) \end{pmatrix} =_{\text{distr}} \begin{pmatrix} \sum_{D'} \sum_{D \ni 1} \xi^{D'}(D; A_1, \ldots, A_N) \\ \vdots \\ \sum_{D'} \sum_{D \ni N} \xi^{D'}(D; A_1, \ldots, A_N) \end{pmatrix},$$
$$= \begin{pmatrix} \sum_{D \ni 1} \sum_{D'} \xi^{D'}(D; A_1, \ldots, A_N) \\ \vdots \\ \sum_{D \ni N} \sum_{D'} \xi^{D'}(D; A_1, \ldots, A_N) \end{pmatrix},$$
$$= \begin{pmatrix} \sum_{D \ni 1} \xi(D; A_1, \ldots, A_N) \\ \vdots \\ \sum_{D \ni N} \xi(D; A_1, \ldots, A_N) \end{pmatrix},$$

which, along with Equation (A6), establishes the theorem. □

A short note: The variable $\xi(D; A_1, \ldots, A_N)$ counts the number of points which are marked by a set $D' \supset D$, but after shifting, only the points attributed to the processes with indices $i \in D$ remain in the corresponding subsets A_i. Thus, to determine the number of points attributed to the ith process which lie in A_i ($X_i(A_i)$), one simply sums the variables ξ for all D containing i, as in Equation (A3). Thus, the intensity of $\xi(D; A_1, \ldots, A_N)$,

$$\lambda p_{D'} \nu(Q_{D'})(M(D, D'; A_1, \ldots, A_N)),$$

is simply the expected number of such points. Keeping in mind these natural interpretations of terms, Theorem 1 is easier to digest, and the result is not surprising.

PROOF OF EQUATION (27)

In Equation (A2), we gave a more intuitive representation of the measure $\nu(Q_D)$ than the one first defined in Bäuerle and Grübel (2005), which we prove here. Suppose that Q is a measure on $\mathcal{B}(\mathbb{R}^d)$, and $A \in \mathcal{B}(\mathbb{R}^d)$. Then we have

$$\begin{aligned} \nu(Q)(A) &= \int Q(A - t\mathbf{1})dt \\ &= \iint 1_{A - t\mathbf{1}}(\mathbf{y}) Q(d\mathbf{y})dt \\ &= \iint 1_{\{t \in \mathbb{R}: \mathbf{y} + t\mathbf{1} \in A\}}(t)dtQ(d\mathbf{y}) \\ &= \int \ell(\{t \in \mathbb{R} : \mathbf{y} + t\mathbf{1} \in A\})Q(d\mathbf{y}) \\ &= \mathbf{E}_{\mathbf{Y}}[\ell(\{t \in \mathbb{R} : \mathbf{Y} + t \in A\})|\mathbf{Y} \sim Q], \end{aligned}$$

thus proving Equation (A2)

PROOF OF THEOREM 1.1

Theorem 1.1 *Let X be a joint counting process of GTaS type with total intensity λ, marking distribution $(p_D)_{D \subset \mathbb{D}}$, and family of shift distributions $(Q_D)_{D \subset \mathbb{D}}$. Let $A_1, \ldots, A_k$ be arbitrary sets in $\mathcal{B}(\mathbb{R})$, and $\bar{D} = \{i_1, \ldots, i_k\} \subset \mathbb{D}$ with $|\bar{D}| = k$. The cross-cumulant of the counting processes may be written*

$$\kappa(X_{i_1}(A_1), \ldots, X_{i_k}(A_k)) = \lambda \sum_{D' \supset \bar{D}} p_{D'} \int P(t\mathbf{1} + \mathbf{Y}^{\bar{D}} \in A_1 \times \cdots \times A_k | \mathbf{Y} \sim Q_{D'})dt \tag{A7}$$

where $\mathbf{Y}^{\bar{D}}$ represents the projection of the random vector $\mathbf{Y}$ onto the dimensions indicated by the members of the set $\bar{D}$. Furthermore, assuming that the shift distributions possess densities $(q_D)_{D \subset 2^{\mathbb{D}}}$, the cross-cumulant density is given by

$$\kappa^X_{i_1 \cdots i_k}(\tau_1, \ldots, \tau_{k-1}) = \lambda \sum_{D' \supset \bar{D}} p_{D'} \int q^{\bar{D}}_{D'}(t, t + \tau_1, \cdots, t + \tau_{k-1})dt, \tag{A8}$$

where $q^{\bar{D}}_{D'}$ indicates the kth order joint marginal density of $q_{D'}$ in the dimensions of $\bar{D}$.

Proof. First, as noted in the text, we may rewrite the distributional representation of Theorem 0 (Equation A3) as

$$\begin{pmatrix} X_{i_1}(A_{i_1}) \\ \vdots \\ X_{i_k}(A_{i_k}) \end{pmatrix} =_{\text{distr}} \begin{pmatrix} \sum_{i_1 \in D \subset \bar{D}} \zeta_D(A_1, \ldots, A_N) \\ \vdots \\ \sum_{i_k \in D \subset \bar{D}} \zeta_D(A_1, \ldots, A_N) \end{pmatrix} \tag{A9}$$

where

$$\zeta_D(A_1, \ldots, A_N) = \sum_{\substack{D' \supset D \\ (\bar{D} \setminus D) \cap D' = \emptyset}} \xi(D'; A_1, \ldots, A_N). \tag{A10}$$

Repeating the description from the main text, the processes ζ_D are comprised of a sum of all of the processes $\xi(D')$ (defined above, in Theorem 0) such that D' contains all of the indices D, but no other indices which are part of the subset $\bar{D}$ under consideration. These sums are non-overlapping, implying that the ζ_D are also independent and Poisson.

Using the representation of Equation (A9), we first find that

$$\kappa(X_{i_1}(A_1), \ldots, X_{i_k}(A_k)) = \kappa\left[\sum_{i_1 \in D_1 \subset \bar{D}} \zeta_{D_1}, \ \ldots, \ \sum_{i_k \in D_k \subset \bar{D}} \zeta_{D_k}\right] = \sum_{i_1 \in D_1 \subset \bar{D}} \cdots \sum_{i_k \in D_k \subset \bar{D}} \kappa[\zeta_{D_1}, \ldots, \zeta_{D_k}].$$

where we suppressed the dependence of the variables ζ_D on the subsets A_i. The first equality in the previous equation is simply the representation defined in Equation (A10), and the second is from the multilinear property of cumulants (property (C1) in

the Methods). Note that the sums are over the sets $D_1, \ldots, D_k$ satisfying the given conditions. Recall that, by construction, the Poisson processes ζ_D (see Equation A10) are independent for distinct marking sets. Accordingly, the cumulant $\kappa[\zeta_{D_1}, \ldots, \zeta_{D_k}]$ is zero unless $D_1 = \ldots = D_k$, by property (C2) of cumulants—that is,

$$\kappa[\zeta_{D_1}(A_1, \ldots, A_N), \ldots, \zeta_{D_k}(A_1, \ldots, A_N)] = \begin{cases} \kappa_k(\zeta_{\bar{D}}(A_1, \ldots, A_N)) & D_j = \bar{D} \text{ for each j} \\ 0 & \text{otherwise} \end{cases}.$$

Hence,

$$\begin{aligned} \kappa(X_{i_1}(A_1), \ldots, X_{i_k}(A_k)) &= \kappa_k(\zeta_{\bar{D}}(A_1, \ldots, A_N)) \\ &= \mathbf{E}\big[\zeta_{\bar{D}}(A_1, \ldots, A_N)\big], \end{aligned} \quad \text{(A11)}$$

where we have again used that all cumulants of a Poisson-distributed random variable are equal to its mean.

For what follows, taking $D_0, D' \subset \mathbb{D}$ fixed with $D_0 \subset D'$, the sets $M(D, D'; A_1, \ldots, A_N)$ with $D_0 \subset D \subset D'$ are disjoint, and

$$\cup_{D_0 \subset D \subset D'} M(D, D'; A_1, \ldots, A_N) = B_1 \times \cdots \times B_N \quad \text{with } B_i = \begin{cases} A_i, & i \in D_0 \\ \mathbb{R}, & i \notin D_0 \end{cases}. \quad \text{(A12)}$$

In particular, note the independence of the above union from D'.

Substituting Equation (A10) in to (A11), we have

$$\begin{aligned} &\kappa(X_{i_1}(A_1), \ldots, X_{i_k}(A_k)) \\ &= \sum_{D \supset \bar{D}} \mathbf{E}[\xi(D; A_1, \ldots, A_k] \\ &= \lambda \sum_{D \supset \bar{D}} \sum_{D' \supset D} p_{D'} \nu(Q_{D'})(M(D, D'; A_1, \ldots, A_N)) \\ &= \lambda \sum_{D' \supset \bar{D}} p_{D'} \sum_{\bar{D} \subset D \subset D'} \nu(Q_{D'})(M(D, D'; A_1, \ldots, A_N)) \\ &= \lambda \sum_{D' \supset \bar{D}} p_{D'} \nu(Q_{D'})(\cup_{\bar{D} \subset D \subset D'} M(D, D'; A_1, \ldots, A_N)) \\ &= \lambda \sum_{D' \supset \bar{D}} p_{D'} \int P(t + \mathbf{Y}^{\bar{D}} \in A_1 \times \cdots \times A_k | \mathbf{Y} \sim Q_{D'}) dt, \end{aligned}$$

where the third equality above is a simple exchange of the order of summation, and the fourth equality follows from the additivity of the measure $\nu(Q_{D'})$ over the disjoint sets $M(D, D'; A_1, \ldots, A_N)$. Finally, the fifth equality makes use of the independence of the set union on the fourth line from the set D' as indicated by Equation (A12), the definition of the measure $\nu(Q_{D'})$ in Equation (A1) and the value of the set union given in Equation (A12).

This completes the proof of Equation (A7), and (A8) follows from the definition of the cross-cumulant density in Equation (24) of the Methods. □

OTHER DETAILS

Parameters for figures in the text

Figure 1. For **Figure 1**, the GTaS process of size $N = 6$ consisted of only first order and population-level events which were assigned marking probabilities

$$p_D = \begin{cases} 0.05 & D = \mathbb{D} \\ \frac{0.95}{6} & D = \{i\} \text{ for some } i \in \mathbb{D} \\ 0 & \text{otherwise} \end{cases}.$$

The rate of the mother process was $\lambda = 0.5$ kHz, and the shift times for population level events were generated as in section 2.2.2 with

$$T_i \sim \Gamma(2, 1) - 1, \quad i = 1, \ldots, 6,$$

where the Gamma distribution has density

$$f(t|k, \theta) = \frac{1}{\Gamma(k)\theta^k} x^{k-1} e^{-\frac{x}{\theta}} \Theta(t).$$

Figures 3, 4. For **Figures 3**, **4**, the GTaS process of size $N = 6$ consisted of first and second order as well as population-level events. These events had marking probabilities

$$p_D = \begin{cases} 0.05 & D = \mathbb{D} \\ \frac{0.95}{21} & D = \{i\}, \{i, j\} \text{ for some } i, j \in \mathbb{D} \\ 0 & \text{otherwise} \end{cases}.$$

The rate of the mother process was $\lambda = 0.5$ kHz, and the shift times for population level events were generated as in section 2.2.2 with

$$T_i \sim \text{Exp}(0.5), \quad i = 1, \ldots, 6.$$

The shift times of the second order events were drawn from an independent Gaussian distribution with each coordinate having standard deviation 5 ms.

Figure 5. For **Figure 5**, the network parameters were $w^{\text{in}} = 0.4$, $w^{\text{syn}} = 6$, $\tau_{\text{syn}} = 0.1$, $\tau_{\text{d}} = 1.75$. The GTaS input had the same size as the network ($N = 10$). As in the example of **Figures 3**, **4**, the GTaS input included first and second order as well as population level events. Here, we set

$$p_D = \begin{cases} 0.2 & D = \mathbb{D} \\ \frac{0.95}{5} & D = \{i\}, \{i, j\} \text{ for some } i, j \in \mathbb{D} \\ 0 & \text{otherwise} \end{cases}.$$

The rate of the mother process was $\lambda = 1.5$ kHz, and the shift times for population level events were generated as in section 2.2.2 with

$$T_i \sim \Gamma(k, \theta), \quad i = 1, \ldots, 6.$$

The shift parameters k, θ (representing shape and scale) were determined by the given shift mean μ_{shift} and standard deviation σ_{shift} as

$$\mu_{\text{shift}} = k\theta, \quad \sigma_{\text{shift}} = \sqrt{k\theta^2}.$$

The shift times of the second order events were drawn from an independent Gaussian distribution with each coordinate having standard deviation 0.3 ms.

frontiers in
COMPUTATIONAL NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 01 July 2013
doi: 10.3389/fncom.2013.00082

Interareal coupling reduces encoding variability in multi-area models of spatial working memory

Zachary P. Kilpatrick*

Department of Mathematics, University of Houston, Houston, TX, USA

Edited by:
Ruben Moreno-Bote, Foundation Sant Joan de Deu, Spain

Reviewed by:
Albert Compte, Institut d'investigacions Biomèdiques August Pi i Sunyer, Spain
Moritz Helias, Institute for Advanced Simulation, Germany

****Correspondence:***
Zachary P. Kilpatrick, Department of Mathematics, University of Houston, 651 Phillip G Hoffman Hall, Houston, 77204-3008 TX, USA
e-mail: zpkilpat@math.uh.edu

Persistent activity observed during delayed-response tasks for spatial working memory (Funahashi et al., 1989) has commonly been modeled by recurrent networks whose dynamics is described as a *bump attractor* (Compte et al., 2000). We examine the effects of interareal architecture on the dynamics of bump attractors in stochastic neural fields. Lateral inhibitory synaptic structure in each area sustains stationary bumps in the absence of noise. Introducing noise causes bumps in individual areas to wander as a Brownian walk. However, coupling multiple areas together can help reduce the variability of the bump's position in each area. To examine this quantitatively, we approximate the position of the bump in each area using a small noise expansion that also assumes weak amplitude interareal projections. Our asymptotic results show the motion of the bumps in each area can be approximated as a multivariate Ornstein–Uhlenbeck process. This shows reciprocal coupling between areas can always reduce variability, if sufficiently strong, even if one area contains much more noise than the other. However, when noise is correlated between areas, the variability-reducing effect of interareal coupling is diminished. Our results suggest that distributing spatial working memory representations across multiple, reciprocally-coupled brain areas can lead to noise cancelation that ultimately improves encoding.

Keywords: neural field, bump attractor, spatial working memory, correlations, noise cancelation

INTRODUCTION

Persistent spiking activity has been experimentally observed in prefrontal cortex (Funahashi et al., 1989; Miller et al., 1996), parietal cortex (Colby et al., 1996; Pesaran et al., 2002), superior colliculus (Basso and Wurtz, 1997), caudate nucleus (Hikosaka et al., 1989; Levy et al., 1997), and globus pallidus (Mushiake and Strick, 1995; McNab and Klingberg, 2008) during the retention interval of visuospatial working memory tasks. Often, the subject must remember a cue's location for several seconds (Funahashi et al., 1989). Delay period neurons persistently fire in response to a preferred cue orientation as described by a bell-shaped tuning curve. Networks of these neurons, with recurrent excitation between similarly tuned neurons and broadly tuned feedback inhibition, can generate spatially localized "bumps." The position of these bumps encodes the remembered location of the cue (Compte et al., 2000).

Dynamic variability can degrade the accuracy of working memory over time though. Fluctuations in membrane voltage and synaptic conductance can lead to spontaneous spike or failure events at the edge of the bump, causing the bump to wander diffusively (Compte et al., 2000; Laing and Chow, 2001). Bump attractor networks are particularly prone to such diffusive error because bump positions lie on a line attractor where each location is neutrally stable (Amari, 1977). Interestingly, psychophysical data demonstrates spatial working memory error does scale linearly with delay time, suggesting the underlying process that degrades memory is diffusive (White et al., 1994; Ploner et al., 1998). Much theoretical work has examined network properties that might limit memory degradation. Several computational studies have explored networks built from bistable neuronal units, which sustain persistent states that are less susceptible to noise (Camperi and Wang, 1998; Koulakov et al., 2002; Goldman et al., 2003). In addition, synaptic facilitation has been shown to slow the drift of bump position due to internal variability (Itskov et al., 2011). Synaptic plasticity has also be shown to reduce diffusion of bumps in (Hansel and Mato, 2013). Finally, spatially heterogeneous recurrent excitation can reduce wandering of bumps quantizing the line attractor by stabilizing a finite set of bump locations (Kilpatrick and Ermentrout, 2013; Kilpatrick et al., 2013).

Complementary to these possibilities, we propose that interareal coupling across multiple areas of cortex may reduce error in working memory recall generated by dynamic fluctuations. Multiple representations of spatial working memory have been identified in different cortical areas (Colby et al., 1996). This distributed representation makes working memory information readily available for motor (Owen et al., 1996) and decision-making (Curtis and Lee, 2010) tasks. In addition, this redundancy may serve to reduce degrading effects of noise. It is known that several areas involved in oculomotor delayed response tasks are reciprocally coupled to one another (Constantinidis and Wang, 2004; Curtis, 2006). We presume the representation of a spatial working memory in a single area takes the form of a bump in a recurrently coupled neural field. Projections between

areas share information about bump position across the multi-area network. Recently, (Folias and Ermentrout, 2011) showed several novel activity patterns emerge when considering neural fields with multiple areas. In addition, recent analyses of spatiotemporal dynamics of perceptual rivalry have exploited dual population neural field models, where activity in each area represents a single percept (Kilpatrick and Bressloff, 2010; Bressloff and Webber, 2012b). In this study, we focus on activity patterns where bumps in each area have positions that remain close.

Our study mostly focuses on a dual area model of spatial working memory, where each area provides a replicate representation of the presented cue. We begin by demonstrating the neutral stability of the bump position in each area, in the absence of noise and interareal projections. Upon including noise and interareal projections, we use a small-noise expansion to derive an effective stochastic differential equation for the position of the bump in each area. The effective system is a multivariate Ornstein–Uhlenbeck process, which we can analyze using diagonalization. The variance of this stochastic process decreases as the strength of connections between areas increases. Variance reduction relies on cancelations arising due to averaging noise between both areas. Thus, when noise is strongly correlated between areas, the effect of interareal coupling is negligible. Lastly, we show this analysis extends to the case of N (more than two) areas and that for sufficiently strong interareal connections, variance scales as $1/N$.

MATERIALS AND METHODS

DUAL AREA MODEL OF SPATIAL WORKING MEMORY

We consider a recurrently coupled model commonly used for spatial working memory (Camperi and Wang, 1998; Ermentrout, 1998) and visual processing (Ben-Yishai et al., 1995). GABAergic inhibition (Gupta et al., 2000) typically acts faster than excitatory NMDAR kinetics (Clements et al., 1992), and we assume excitatory synapses contain a mixture of AMPA and NMDA components. Thus, we make the assumption that inhibition is slaved to excitation as in (Amari, 1977). We can then describe average activity $u_1(x,t)$ and $u_2(x,t)$ of neurons in either area by the system (Ben-Yishai et al., 1995; Folias and Ermentrout, 2011; Kilpatrick and Ermentrout, 2013)

$$\tau \mathrm{d}u_1(x,t) = \left[-u_1 + w_{11} * f(u_1) + \varepsilon^{1/2} w_{12} * f(u_2)\right]\mathrm{d}t + \varepsilon^{1/2}\mathrm{d}W_1(x,t), \tag{1a}$$

$$\tau \mathrm{d}u_2(x,t) = \left[-u_2 + w_{22} * f(u_2) + \varepsilon^{1/2} w_{21} * f(u_1)\right]\mathrm{d}t + \varepsilon^{1/2}\mathrm{d}W_2(x,t), \tag{1b}$$

where the effects of synaptic architecture are described by the convolution

$$w_{jk} * f(u_k) = \int_{-\pi}^{\pi} w_{jk}(x-y) f(u_k(y,t))\mathrm{d}y, \tag{2}$$

for $j,k = 1,2$, so the case $j = k$ describes recurrent synaptic connections within a area and $j \neq k$ describes synaptic connections between areas (interareal). Several fMRI and electrode recordings have revealed correlations between activity in multiple cortical areas during spatial working memory tasks (Constantinidis and Wang, 2004; Curtis, 2006), such as parietal and prefrontal cortex (Chafee and Goldman-Rakic, 1998). However, it seems the strength of these correlations is often not on the order of the activity itself (di Pellegrino and Wise, 1993). For this reason, we presume the strength of interareal connections is weak $0 \leq \varepsilon^{1/2} \ll 1$. Note, we could choose to make them a different magnitude than the noise, but for analytical convenience, we choose interareal connection and noise magnitude to be roughly the same. Analysis could still be performed in other cases, but it would simply be more complicated. By setting $\tau = 1$, we can assume that time evolves on units of the excitatory synaptic time constant, which we presume to be roughly 10 ms (Häusser and Roth, 1997). The function $w_{jk}(x-y)$ describes the strength (amplitude of w_{jk}) and net polarity (sign of w_{jk}) of synaptic interactions from neurons with stimulus preference y to those with preference x. Following previous studies, we presume the modulation of the recurrent synaptic strength is given by the cosine

$$w_{jj}(x-y) = w(x-y) = \cos(x-y), \quad j = 1,2, \tag{3}$$

so neurons with similar orientation preference excite one another and those with dissimilar orientation preference disynaptically inhibit one another (Ben-Yishai et al., 1995; Ferster and Miller, 2000). Lateral inhibitory type network architectures are supported by anatomical studies of the delay period neurons in prefrontal cortex (Goldman-Rakic, 1995). Our general analysis will apply to any even symmetric function of the distance $x-y$, but we typically compute things using (Equation 3) since it eases calculations. Finally, synaptic connections from area k to j are specified by the weight function $w_{jk}(x-y)$, and we typically take this to be the function

$$w_{jk}(x-y) = E_j + M_j \cos(x-y), \quad k \neq j \tag{4}$$

where E_j and M_j specify the strength of baseline excitation and modulation projecting to the jth area.

Output firing rates are given by taking the gain function $f(u)$ of the synaptic input, which we usually proscribe to be (Wilson and Cowan, 1973)

$$f(u) = \frac{1}{1+\mathrm{e}^{-\gamma(u-\theta)}},$$

and often take the high gain limit $\gamma \to \infty$ for analytical convenience, so (Amari, 1977)

$$f(u) = H(u-\theta) = \begin{cases} 0 : u < \theta, \\ 1 : u \geq \theta. \end{cases} \tag{5}$$

Effects of noise are described by the small amplitude ($0 \leq \varepsilon \ll 1$) stochastic processes $\varepsilon^{1/2} W_j(x,t)$ that are white in time and correlated in space so that $\langle \mathrm{d}W_j(x,t)\rangle = 0$ and

$$\langle \mathrm{d}W_j(x,t)\mathrm{d}W_j(y,s)\rangle = C_j(x-y)\delta(t-s)\mathrm{d}t\mathrm{d}s,$$

$$\langle \mathrm{d}W_j(x,t)\mathrm{d}W_k(y,s)\rangle = C_c(x-y)\delta(t-s)\mathrm{d}t\mathrm{d}s,$$

describing both local and shared noise in either area, $j = 1, 2$ with $j \neq k$. For simplicity, we assume the local spatial correlations have a cosine profile $C_j(x) = c_j \cos(x)$. We also typically assume the correlated noise component has cosine profile so $C_c(x) = c_c \cos(x)$. Therefore, in the limit $c_c \to 0$, there are no interareal noise correlations, and in the limit $c_c \to \min(c_1, c_2)$, noise in each area is maximally correlated. For instance, when $c_1 = c_2 = c_c = 1$, noise in each area is drawn from the same process.

MULTIPLE-AREA MODEL OF SPATIAL WORKING MEMORY

To incorporate the effects of many coupled, redundant areas encoding a spatial working memory, we consider a model with N areas and arbitrary synaptic architecture, given by

$$\tau \mathrm{d}u_j(x,t) = \left[-u_j + \varepsilon^{1/2} \sum_{k=1}^{N} w_{jk} * f(u_k) \right] \mathrm{d}t + \varepsilon^{1/2} \mathrm{d}W_j(x,t) \quad (6)$$

where u_j represents neural activity in the jth area where $j = 1, \ldots, N$. As before, we set $\tau = 1$, so each time unit corresponds to the roughly 10 ms timescale of excitatory synaptic conductance. The weight function $w_{jk}(x - y)$ represents the connection from neurons in area k with cue preference y to neurons in area j with cue preference x as described by (Equation 2). For comparison with numerical simulations, we take weight functions to be the cosines (Equation 3) and (Equation 4) and the firing rate function to be Heaviside (Equation 5). As in the dual area model, noises $W_j(x, t)$ are white in time and correlated in space so that $\langle \mathrm{d}W_j(x,t) \rangle = 0$ and

$$\langle \mathrm{d}W_j(x,t) \mathrm{d}W_k(y,s) \rangle = C_{jk}(x-y)\delta(t-s)\mathrm{d}t\mathrm{d}s,$$

with $j, k = 1, \ldots, N$, where local noise correlations are described when $j = k$ and noise correlations between areas are described when $j \neq k$. For comparison with numerical simulations, we consider $C_{jj}(x) = \cos(x)$ and $C_{jk}(x) = c_c \cos(x)$ for all $j \neq k$.

NUMERICAL SIMULATION OF STOCHASTIC DIFFERENTIAL EQUATIONS

The spatially extended model (Equation 1) was simulated using an Euler–Maruyama method with a timestep 10^{-4}, using Riemann integration on the convolution term with 2000 spatial grid points. To compute and compare the variances $\langle \Delta_1(t)^2 \rangle$ for the dual and multiple area model, we simulated the system 5000 times. The position of the bump Δ_j at each timestep, in each simulation, was determined by the position x in each area j at which the maximal value of $u_j(x, t)$ was attained. The variance was then computed at each timepoint and compared to our asymptotic calculations.

RESULTS

We will now study how interareal architecture affect the dynamics of bumps in multiple area stochastic neural fields. To start, we demonstrate that in the absence of reciprocal connectivity between areas bump attractors exist that are neutrally stable to perturbations that change their position, which has long been known (Amari, 1977; Camperi and Wang, 1998; Ermentrout, 1998). Introducing weak interareal connectivity can decrease the variability in bump position because noise that moves bumps in the opposite direction is canceled due to an attractive force introduced by connectivity. Perturbations that push bumps in the same direction are still integrated, so bumps wander due to dynamic fluctuations, but their effective variance is smaller than it would be without interareal synaptic connections. In the presence of noise correlations between areas, effects of noise cancelation are weaker since stochastic forcing in each area is increasingly similar. Our asymptotic analysis is able to explain all of this with its resulting multivariate Ornstein–Uhlenbeck process.

BUMPS IN THE NOISE-FREE SYSTEM

To begin, we seek stationary solutions to Equation (1) in the absence interareal connections and noise ($\varepsilon \to 0$). Similar analyses have been carried out for bumps in single area populations (Ermentrout, 1998; Hansel and Sompolinsky, 1998). For this study, we assume recurrent connections are identical in all areas ($w_{jj} = w$). Relaxing this assumption slightly does not dramatically alter our results. Note first stationary solutions take the form $(u_1(x,t), u_2(x,t)) = (U_1(x), U_2(x))$. In the absence of any interareal connections, we would not necessarily expect the peaks of these bumps to be at the same location. However, translation invariance of the system (Equation 1) allows us to set the center of both bumps to be $x = 0$ to ease calculations. The stationary bump solutions then satisfy the system

$$U_1 = w * f(U_1), \quad U_2 = w * f(U_2), \quad (7)$$

so the shape of each bump is only determined by the local connections w. For w given by Equation (3), since $U_1(x)$ and $U_2(x)$ are assumed to be peaked at $x = 0$, then by also assuming even symmetric solutions, we find

$$U_1(x) = \int_{-\pi}^{\pi} \cos y f(U_1(y)) \mathrm{d}y \cos x,$$
$$U_2(x) = \int_{-\pi}^{\pi} \cos y f(U_2(y)) \mathrm{d}y \cos x, \quad (8)$$

where we use $\cos(x - y) = \cos x \cos y + \sin x \sin y$. We can more easily compute the precise shape of these bumps in case of a Heaviside firing rate function (Equation 5). There is then an identical active region of each bump such that $U_1(x) > \theta$ and $U_2(x) > \theta$ when $x \in (-a, a)$, so the Equation (8) become $U_1(x) = U_2(x) = 2 \sin a \cos x$. Applying self-consistency, $U_1(\pm a) = U_2(\pm a) = \theta$, we can generate an implicit equation for the half-widths of the bumps a given by $2 \sin a \cos a = \sin(2a) = \theta$. Solving this explicitly for a, we find two solutions on $a \in [0, \pi]$: $a_u = \frac{1}{2} \sin^{-1} \theta$ and $a_s = \frac{\pi}{2} - \frac{1}{2} \sin^{-1} \theta$. Only the bump associated with a_s is stable.

The bumps (Equation 7) are neutrally stable to perturbations in both directions, which can lead to encoding error once the effects of dynamic fluctuations are considered (Kilpatrick et al., 2013). Since the two areas are uncoupled, examining bumps' stability can be reduced to studying each bump's stability individually (see Kilpatrick and Ermentrout, 2013 for details). Translating

a bump by a scaling of the spatial derivative $U'(x)$, we find $u_j(x,t) = U_j(x) + \varepsilon^{1/2} U_j'(x) e^{\lambda t}$ is associated with a zero eigenvalue ($\lambda = 0$), corresponding to neutral stability. To see this, we plug it into the corresponding bump equation of Equation (1) in the absence of noise and interareal connections and examine the linearization

$$\lambda U_j'(x) = -U_j'(x) + \int_{-\pi}^{\pi} w(x-y) f'(U_j(y)) U_j'(y) \mathrm{d}y. \quad (9)$$

Note, in the limit of infinite gain $\gamma \to \infty$, a sigmoid f becomes the Heaviside (Equation 5), and

$$f'(U(x)) = \frac{\mathrm{d}H(U(x))}{\mathrm{d}U} = \frac{\delta(x-a)}{|U'(a)|} + \frac{\delta(x+a)}{|U'(a)|},$$

in the sense of the distributions. Equation (9) still hold in this case. Differentiating (Equation 7), and integrating by parts, we find

$$\begin{aligned} -U_1' + w * [f'(U_1) U_1'] &= 0, \\ -U_2' + w * [f'(U_2) U_2'] &= 0, \end{aligned} \quad (10)$$

where the boundary terms vanish due to periodicity of the domain $[-\pi, \pi]$. Thus, the right hand side of Equation (9) vanishes, and $\lambda = 0$ is the only eigenvalue corresponding to translating perturbations. Thus, either bump (in area 1 or 2) is neutrally stable to perturbations that shifts its position in either direction (rightwards or leftwards), since the bump in each area experiences no force from the other bump.

This changes when we consider the effect of interareal connectivity. Once the two areas of Equation (1) are reciprocally coupled, bumps are stable to perturbations that translate them in opposite directions of one another (see **Figure 1**). Interareal connections act as a restoring force between the two positions of each bump. We will demonstrate this in the subsequent section by deriving a linear stochastic system for the position of either bump in the presence of small noise and weak interareal connectivity. The restorative nature of interareal connectivity is revealed by the negative eigenvalue associated with the interaction matrix (Equation 15) of our stochastic system, as shown in Equation (18).

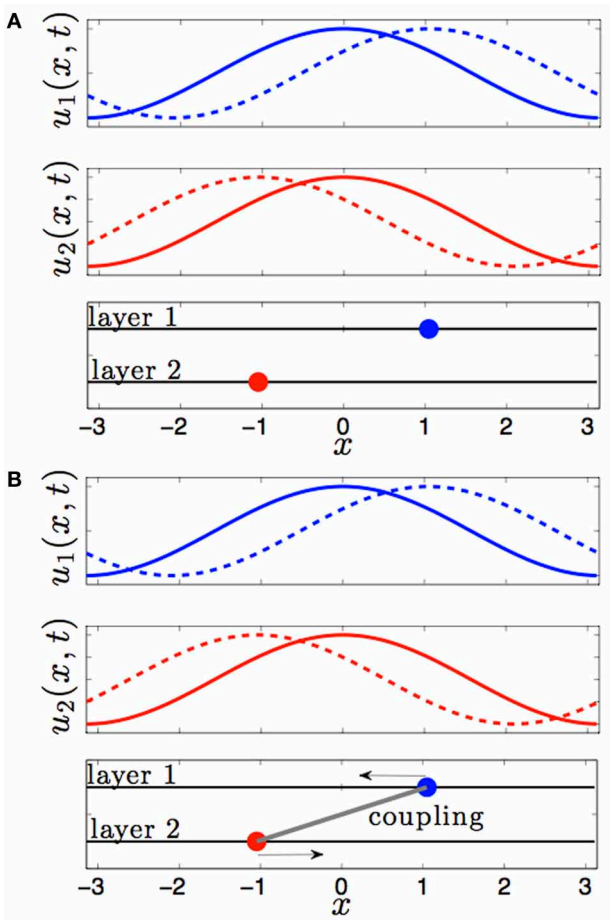


FIGURE 1 | Effect of interareal coupling on the stability of bumps to translating perturbations. (A) In the absence of interareal coupling, bumps (solid) are neutrally stable to perturbations (dashed) that translate them in opposite directions. **(B)** In the presence of interareal coupling, bumps are linearly stable, as revealed by the negative eigenvalue in Equation (18), to perturbations that translate them in opposite directions.

NOISE-INDUCED WANDERING OF BUMPS

Now we consider the effects of small noise on the position of bumps in the presence of weak interareal connections. We start by presuming noise generates two distinct effects in the bumps (see **Figure 2**). First, noise causes both bumps to wander away from their initial positions, while still being pulled back into place by the bump in the other area. Bump position in areas 1 and 2 will be described by the time-varying stochastic variables $\Delta_1(t)$ and $\Delta_2(t)$. Second, noise causes fluctuations in the shape of both bumps, described by a correction Φ_j. To account for this, we consider the ansatz

$$\begin{aligned} u_1 &= U_1(x - \Delta_1(t)) + \varepsilon^{1/2} \Phi_1(x - \Delta_1(t), t) + \cdots \\ u_2 &= U_2(x - \Delta_2(t)) + \varepsilon^{1/2} \Phi_2(x - \Delta_2(t), t) + \cdots \end{aligned} \quad (11)$$

Armero et al. (1998) originally developed this approach to analyze of front propagation in stochastic PDE models. In stochastic neural fields, it has been modified to analyze wave propagation (Bressloff and Webber, 2012a) and bump wandering (Kilpatrick and Ermentrout, 2013). Plugging the ansatz (Equation 11) into the system (Equation 1) and expanding in powers of $\varepsilon^{1/2}$, we find that at $\mathcal{O}(1)$, we have the bump solution (Equation 7). Proceeding to $\mathcal{O}(\varepsilon^{1/2})$, we find

$$\mathrm{d}\mathbf{\Phi} - \mathcal{L}\mathbf{\Phi} = \begin{pmatrix} \varepsilon^{-1/2} \dot{\Delta}_1 U_1' + \mathrm{d}W_1 \\ \varepsilon^{-1/2} \dot{\Delta}_2 U_2' + \mathrm{d}W_2 \end{pmatrix} + \mathcal{K}(x,t), \quad (12)$$

where $\mathcal{K}(x,t)$ is the 2×1 vector function

$$\mathcal{K}(x,t) = \begin{pmatrix} w_{12} * \left[f(U_2) + f'(U_2) U_2' \cdot (\Delta_2 - \Delta_1) \right] \mathrm{d}t \\ w_{21} * \left[f(U_1) + f'(U_1) U_1' \cdot (\Delta_1 - \Delta_2) \right] \mathrm{d}t \end{pmatrix}$$

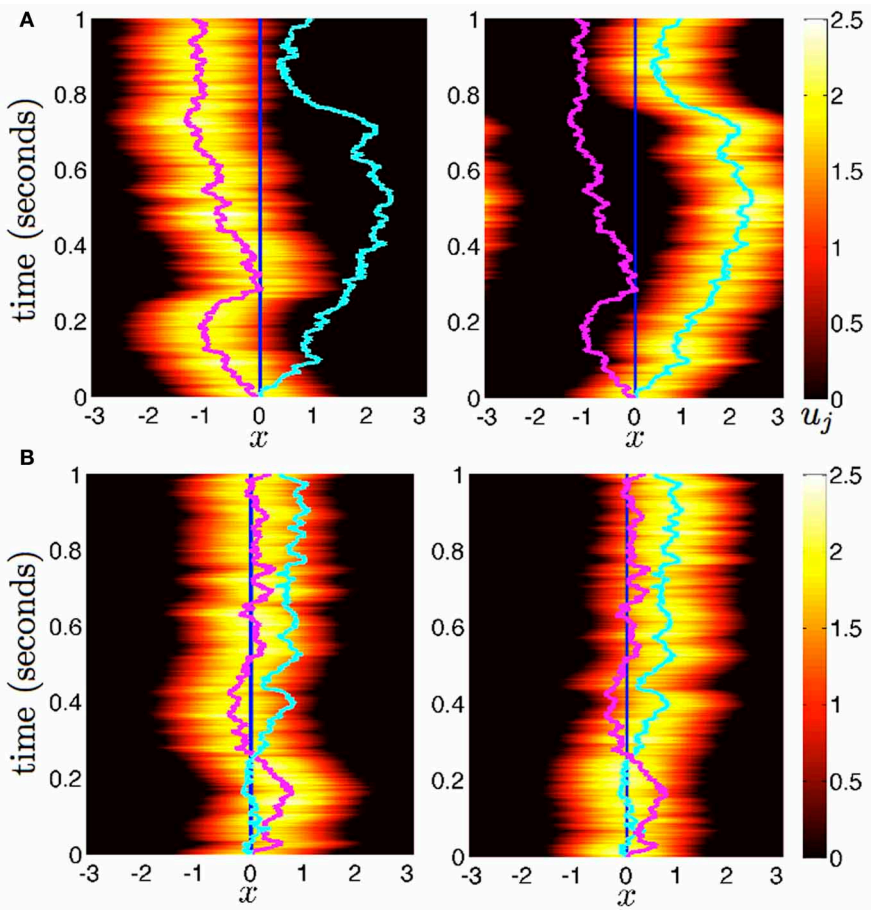


FIGURE 2 | Diffusion of bumps in the dual area stochastic neural field (Equation 1). (A) Without interareal connections ($w_{12} = w_{21} \equiv 0$), each bump executes Brownian motion about the domain, due to stochastic forces. **(B)** In the presence of interareal connections $\sqrt{\varepsilon}w_{12}(x) = \sqrt{\varepsilon}w_{21}(x) = 0.01(\cos(x) + 1)$, the position of bump 1 (magenta) is attracted to the position of bump 2 (cyan) and vice versa. Due to the reversion of each bump to the position of the other, both bumps effectively wander the domain less. Local connectivity is described by the cosine (Equation 3); the firing rate function is Equation (5). Other parameters are threshold $\theta = 0.5$ and noise amplitude $\varepsilon = 0.025$.

$\mathbf{\Phi} = (\Phi_1(x,t), \Phi_2(x,t))^T$; and $\mathcal{L}$ is the linear operator

$$\mathcal{L}\mathbf{u} = \begin{pmatrix} -u(x) + w(x) * [f'(U_1(x))u(x)] \\ -v(x) + w(x) * [f'(U_2(x))v(x)] \end{pmatrix}$$

for any vector $\mathbf{u} = (u(x)\ v(x))^T$ of integrable functions. Note that the nullspace of $\mathcal{L}$ includes the vectors $(U_1', 0)^T$ and $(0, U_2')^T$, due to Equation (10). The last terms in the right hand side vector of Equation (12) arise due to interareal connections. We have linearized them under the assumption $|\Delta_1 - \Delta_2|$ remains small, so

$$f(U_j(x + \Delta_k - \Delta_j)) \approx f(U_j(x)) + f'(U_j(x))U_j'(x) \cdot (\Delta_k - \Delta_j),$$

where $j = 1, 2$ and $k \neq j$. To make sure that a solution to Equation (12) exists, we require the right hand side is orthogonal to all elements of the null space of the adjoint $\mathcal{L}^*$, which is defined

$$\int_{-\pi}^{\pi} \mathbf{p}^T \mathcal{L}\mathbf{u}\mathrm{d}x = \int_{-\pi}^{\pi} \mathbf{u}^T \mathcal{L}^*\mathbf{p}\mathrm{d}x,$$

for any integrable vector $\mathbf{p} = \big(p(x)\ q(x)\big)^T$. It then follows

$$\mathcal{L}^*\mathbf{p} = \begin{pmatrix} -p(x) + f'(U_1(x))[w(x) * p(x)] \\ -q(x) + f'(U_2(x))[w(x) * q(x)] \end{pmatrix}. \tag{13}$$

We can show that the nullspace of $\mathcal{L}^*$ contains the vector $\mathbf{f}_1 = (f'(U_1)U_1', 0)^T$ by plugging it into Equation (13) to yield

$$\mathcal{L}^*\mathbf{f}_1 = \begin{pmatrix} -f'(U_1)U_1' + f'(U_1)[w * [f'(U_1)U_1'] \\ 0 \end{pmatrix} = \mathbf{0}$$

where $\mathbf{0} = (0, 0)^T$ and we use Equation (10). We can also show the nullspace of $\mathcal{L}^*$ contains $\mathbf{f}_2 = (0, f'(U_2)U_2')^T$ in the same way. Thus, we can ensure Equation (12) has a solution by taking the inner product of both sides of Equation (12) with the two null vectors to yield

$$\begin{aligned}&\langle f'(U_1)U_1', \varepsilon^{-1/2}\dot{\Delta}_1 U_1' + \mathrm{d}W_1\\ &+ w_{12} * \left[f(U_2) + f'(U_2)U_2' \cdot (\Delta_2 - \Delta_1)\right]\mathrm{d}t\rangle = 0\\ &\langle f'(U_2)U_2', \varepsilon^{-1/2}\dot{\Delta}_2 U_2' + \mathrm{d}W_2\\ &+ w_{21} * \left[f(U_1) + f'(U_1)U_1' \cdot (\Delta_1 - \Delta_2)\right]\mathrm{d}t\rangle = 0,\end{aligned}$$

where we define the inner product $\langle u, v\rangle = \int_{-\pi}^{\pi} u(x)v(x)\mathrm{d}x$. Therefore, the stochastic vector $\mathbf{\Delta}(t) = (\Delta_1(t), \Delta_2(t))^T$ obeys the multivariate Ornstein–Uhlenbeck process

$$\mathrm{d}\mathbf{\Delta}(t) = \mathbf{K}\mathbf{\Delta}(t)\mathrm{d}t + \mathrm{d}\mathbf{W}(t) \tag{14}$$

where effects of interareal connections are described by the matrix

$$\mathbf{K} = \begin{pmatrix} -\kappa_1 & \kappa_1 \\ \kappa_2 & -\kappa_2 \end{pmatrix}, \tag{15}$$

with

$$\begin{aligned}\kappa_1 &= \frac{\langle f'(U_1)U_1', \varepsilon^{1/2} w_{12} * \left[f'(U_2)U_2'\right]\rangle}{\langle f'(U_1)U_1', U_1'\rangle},\\ \kappa_2 &= \frac{\langle f'(U_2)U_2', \varepsilon^{1/2} w_{21} * \left[f'(U_1)U_1'\right]\rangle}{\langle f'(U_2)U_2', U_2'\rangle},\end{aligned} \tag{16}$$

and $(w_{12} * f(U_2)) \cdot U_1'$ and $(w_{21} * f(U_1)) \cdot U_2'$ vanish upon integration since they are odd. Noise is described by the vector $\mathrm{d}\mathbf{W}(t) = (\mathrm{d}\mathcal{W}_1, \mathrm{d}\mathcal{W}_2)^T$ with

$$\begin{aligned}\mathrm{d}\mathcal{W}_1(t) &= -\varepsilon^{1/2}\frac{\langle f'(U_1)U_1, \mathrm{d}W_1\rangle}{\langle f'(U_1)U_1', U_1'\rangle},\\ \mathrm{d}\mathcal{W}_2(t) &= -\varepsilon^{1/2}\frac{\langle f'(U_2)U_2, \mathrm{d}W_2\rangle}{\langle f'(U_2)U_2', U_2'\rangle}.\end{aligned}$$

The white noise term $\mathbf{W}$ has zero mean $\langle \mathbf{W}(t)\rangle = \mathbf{0}$ and variance described by pure diffusion so $\langle \mathbf{W}(t)\mathbf{W}^T(t)\rangle = \mathbf{D}t$ with

$$\mathbf{D} = \begin{pmatrix} D_1 & D_c \\ D_c & D_2 \end{pmatrix} \tag{17}$$

where the associated diffusion coefficients of the variance are

$$\begin{aligned}D_1 &= \varepsilon\frac{\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} F_1(x)F_1(y)C_1(x-y)\mathrm{d}x\mathrm{d}y}{\left[\int_{-\pi}^{\pi} F_1(x)U_1'(x)\mathrm{d}x\right]},\\ D_2 &= \varepsilon\frac{\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} F_2(x)F_2(y)C_2(x-y)\mathrm{d}x\mathrm{d}y}{\left[\int_{-\pi}^{\pi} F_2(x)U_2'(x)\mathrm{d}x\right]}.\end{aligned}$$

where $F_j(x) = f'(U_j(x))U_j'(x)$ and covariance is described by the coefficient

$$D_c = \varepsilon\frac{\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} F_1(x)f'(U_2(y))F_2(y)C_c(x-y)\mathrm{d}x\mathrm{d}y}{\left[\int_{-\pi}^{\pi} F_1(x)U_1'(x)\mathrm{d}x\right]\left[\int_{-\pi}^{\pi} F_2(x)U_2'(x)\mathrm{d}x\right]}.$$

In the next section, we analyze this stochastic system (Equation 14), showing how coupling between areas can reduce the variability of the bump positions $\Delta_1(t)$ and $\Delta_2(t)$.

EFFECT OF COUPLING ON BUMP POSITION VARIANCE

To analyze the Ornstein–Uhlenbeck process (Equation 14), we start by diagonalizing the matrix $\mathbf{K} = \mathbf{V}\Lambda\mathbf{V}^{-1}$ using the eigenvalue decomposition

$$\begin{aligned}\Lambda &= \begin{pmatrix} 0 & 0 \\ 0 & -\kappa_1 - \kappa_2 \end{pmatrix},\\ \mathbf{V} &= \frac{1}{\kappa_1 + \kappa_2}\begin{pmatrix} 1 & \kappa_1 \\ 1 & -\kappa_2 \end{pmatrix},\\ \mathbf{V}^{-1} &= \begin{pmatrix} \kappa_2 & \kappa_1 \\ 1 & -1 \end{pmatrix},\end{aligned} \tag{18}$$

such that Λ is the diagonal matrix of eigenvalues; columns of $\mathbf{V}$ are right eigenvectors; and rows of $\mathbf{V}^{-1}$ are left eigenvectors. Eigenvalues λ_1, λ_2 and eigenvectors $\mathbf{v}_1$, $\mathbf{v}_2$ inform us of the effect of interareal coupling on linear stability. The eigenvalue $\lambda_1 = 0$ corresponds to the neutral stability of the positions $(\Delta_1, \Delta_2)^T$ to translations in the same direction $\mathbf{v}_1 = (1, 1)^T$. The negative eigenvalue $\lambda_2 = -(\kappa_1 + \kappa_2)$ corresponds to the linear stability introduced by interareal connections. The positions $(\Delta_1, \Delta_2)^T$ revert to one another when perturbations translate them in opposite directions $\mathbf{v}_2 = (\kappa_1, -\kappa_2)^T$.

Diagonalizing $\mathbf{K} = \mathbf{V}\Lambda\mathbf{V}^{-1}$ using Equation (18), we can compute the mean and variance of the vector $\mathbf{\Delta}(t)$ given by Equation (14). First, note that the mean $\langle \mathbf{\Delta}(t)\rangle = \mathrm{e}^{\mathbf{K}t}\mathbf{\Delta}(0)$ (Gardiner, 2003), which we can compute

$$\langle \mathbf{\Delta}\rangle = \begin{pmatrix} (\kappa_2 + \kappa_1\mathrm{e}^{\lambda_2 t})\Delta_1(0) + (\kappa_1 - \kappa_1\mathrm{e}^{\lambda_2 t})\Delta_2(0) \\ (\kappa_2 - \kappa_2\mathrm{e}^{\lambda_2 t})\Delta_1(0) + (\kappa_1 + \kappa_2\mathrm{e}^{\lambda_2 t})\Delta_2(0) \end{pmatrix}$$

using the diagonalization $\mathrm{e}^{\mathbf{K}t} = \mathbf{V}\mathrm{e}^{\Lambda t}\mathbf{V}^{-1}$. Since $\lambda_2 = -(\kappa_1 + \kappa_2) < 0$,

$$\lim_{t\to\infty}\langle \mathbf{\Delta}(t)\rangle = [\kappa_2\Delta_1(0) + \kappa_1\Delta_2(0)]\begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Thus, the means of $\Delta_1(t)$ and $\Delta_2(t)$ always relax to the same position in long time, due to the linear stability introduced by connections between areas. Under the assumption they both begin at $\Delta_1(0) = \Delta_2(0) = 0$, the covariance matrix is given (Gardiner, 2003)

$$\langle \mathbf{\Delta}(t)\mathbf{\Delta}^T(t)\rangle = \int_0^t \mathrm{e}^{\mathbf{K}(t-s)}\mathbf{D}\mathrm{e}^{\mathbf{K}^T(t-s)}\mathrm{d}s, \tag{19}$$

where $\mathbf{D}$ is the covariance coefficient matrix of the white noise vector $\mathbf{W}(t)$ given by Equation (17). To compute Equation (19), we additionally need the diagonalization $\mathbf{K}^T = (\mathbf{V}^{-1})^T \Lambda \mathbf{V}^T$, so $e^{\mathbf{K}^T t} = (\mathbf{V}^{-1})^T e^{\Lambda t} \mathbf{V}^T$. After multiplying and integrating (Equation 19), we find the elements of the covariance matrix

$$\langle \mathbf{\Delta}(t)\mathbf{\Delta}^T(t)\rangle = \begin{pmatrix} \langle \Delta_1(t)^2\rangle & \langle \Delta_1(t)\Delta_2(t)\rangle \\ \langle \Delta_1(t)\Delta_2(t)\rangle & \langle \Delta_2(t)^2\rangle \end{pmatrix}$$

are

$$\langle \Delta_1(t)^2\rangle = D_+ t + 2\kappa_1 r_1(t) + \frac{\kappa_1}{\kappa_2} r_2(t) \tag{20}$$

$$\langle \Delta_2(t)^2\rangle = D_+ t - 2\kappa_2 r_1(t) + \frac{\kappa_2}{\kappa_1} r_2(t) \tag{21}$$

$$\langle \Delta_1(t)\Delta_2(t)\rangle = D_+ t + (\kappa_1 - \kappa_2) r_1(t) - r_2(t)$$

where the effective diffusion coefficients are

$$D_+ = \frac{\kappa_2^2 D_1 + 2\kappa_1\kappa_2 D_c + \kappa_1^2 D_2}{(\kappa_1+\kappa_2)^2}, \tag{22}$$

$$D_r = \frac{\kappa_2 D_1 - \kappa_1 D_2 + (\kappa_1 - \kappa_2) D_c}{(\kappa_1+\kappa_2)^2}, \tag{23}$$

$$D_- = \frac{D_1 - 2D_c + D_2}{(\kappa_1+\kappa_2)^2}, \tag{24}$$

so that D_+ and D_- are variances of noises occurring along the eigendirections $\mathbf{v}_1$ and $\mathbf{v}_2$. The functions $r_1(t), r_2(t)$ are exponentially saturating

$$r_1(t) = \frac{D_r}{\kappa_1+\kappa_2}\left[1 - e^{-(\kappa_1+\kappa_2)t}\right],$$

$$r_2(t) = \frac{\kappa_1\kappa_2 D_-}{2(\kappa_1+\kappa_2)}\left[1 - e^{-2(\kappa_1+\kappa_2)t}\right].$$

The main quantities of interest to us are the variances (Equation 20) and (Equation 21) with which we can make a few observations concerning the effect of interareal connections on the variance of bump positions.

First, note the long term variance of either bump's position $\Delta_1(t)$ and $\Delta_2(t)$ will be the same, described by the averaged diffusion coefficient D_+, since

$$\lim_{t\to\infty} \langle \Delta_1(t)^2\rangle = \lim_{t\to\infty} \langle \Delta_2(t)^2\rangle = D_+ t. \tag{25}$$

As the effective coupling strengths κ_j are increased, we can expect the variances $\langle \Delta_j(t)^2\rangle$ approach these limits at faster rates since other portions of the variance decay at a rate proportional to $|\lambda_2| = \kappa_1 + \kappa_2$.

Next, we study the case, across all times t, where connections between areas are the same ($w_{12} \equiv w_{21} = w_r$) and noise within areas is identical ($D_1 \equiv D_2 = D_l$), the mean reversion rates will be the same ($\kappa_1 = \kappa_2 = \kappa$) and terms in Equation (23) cancel so $D_r = 0$. Thus, the variances will be identical ($\langle \Delta_1(t)^2\rangle = \langle \Delta_2(t)^2\rangle = \langle \Delta(t)^2\rangle$) and

$$\langle \Delta(t)^2\rangle = \frac{D_l + D_c}{2} t + \frac{D_l - D_c}{8\kappa}\left[1 - e^{-4\kappa t}\right].$$

This demonstrates the way in which correlated noise (D_c) contributes to the variance. When noise within each area is shared ($D_c \to D_l$), there is no benefit to interareal coupling and $\langle \Delta(t)^2\rangle = D_l t$ (see Kilpatrick and Ermentrout, 2013). However, when any noise is not shared between areas ($D_c < D_l$), variance can be reduced by increasing coupling strength κ between areas. The variance $\langle \Delta(t)^2\rangle$ is monotone decreasing in κ since

$$\frac{\partial}{\partial\kappa}\langle \Delta(t)^2\rangle = \frac{D_l - D_c}{8}\frac{(1+4\kappa t)e^{-4\kappa t} - 1}{\kappa^2} \le 0.$$

Inequality holds because $(1 + 4\kappa t) \le e^{4\kappa t}$ is ensured by the Taylor series expansion of $e^{4\kappa t}$ when $\kappa t > 0$.

Thus, variance is minimized in the limit

$$\lim_{\kappa\to\infty} \langle \Delta(t)^2\rangle = \frac{D_l + D_c}{2} t. \tag{26}$$

Therefore, strengthening interareal connections in *both* directions reduces the variance in bump position. On the other hand, in the limit of no interareal connections, we find $\lim_{\kappa\to 0}\langle \Delta(t)^2\rangle = D_l t$, and the variance in a bump's position is determined entirely by local sources of noise.

Returning to asymmetric connectivity ($\kappa_1 \neq \kappa_2$), we consider the case of feedforward connectivity from area 1 to 2 ($w_{12} \equiv 0$), $\kappa_1 = 0$, so $D_+ = D_1$ and the formulas for the variances reduce to

$$\langle \Delta_1(t)^2\rangle = D_1 t,$$

$$\langle \Delta_2(t)^2\rangle = D_1 t + \frac{2(D_1 - D_c)}{\kappa_2}\left[1 - e^{-\kappa_2 t}\right] + \frac{D_1 - 2D_c + D_2}{2\kappa_2}\left[1 - e^{-2\kappa_2 t}\right],$$

so the pure diffusive term of both variances is wholly determined by the local noise of area 1. Then, only the position of the bump in area 2 possesses additional mean-reverting fluctuations in its position, which arise from local sources of noise that force it away from the position of the bump in area 1. In this situation, the variance of the bump in area 2's position is minimized when

$$\lim_{\kappa_2\to\infty} \langle \Delta_1(t)^2\rangle = \lim_{\kappa_2\to\infty} \langle \Delta_2(t)^2\rangle = D_1 t.$$

Comparing this with Equation (26) we see that, since $D_c \le D_1$, the variances $\langle \Delta_j(t)^2\rangle$ will always be higher in this case than in

the case of very strong reciprocal coupling between both areas. Averaging information and noise between both areas decreases positional variance as opposed to one area simply receiving noise and information from another. Similar results have been recently identified in the context of studying synchrony of reciprocally coupled noisy oscillators (Ly and Ermentrout, 2010).

One important caveat is that if area 1 has more noise than area 2, the weighting of reciprocal connectivity, κ_1 and κ_2, should be balanced to minimize the variance. If the average diffusion coefficient D_+ is weighted too heavily with the area having the larger variance, the area with less intrinsic noise can end up noisier than it would be without reciprocal connectivity. To see this in the extreme case feedforward coupling, note that if $D_2 < D_1$, then $D_2 t < D_1 t < \langle \Delta_2(t)^2 \rangle$. Thus, the variance of $\Delta_2(t)$ increases as opposed to the uncoupled case where $\langle \Delta_2(t)^2 \rangle = D_2 t$.

We now derive the optimal weighting of κ_1 and κ_2 to minimize the long term variance (Equation 25) for general asymmetric connectivity, in the absence of correlated noise $D_c = 0$. To do so, we fix κ_2 and find the κ_1 that minimizes D_+, which happens to be

$$\kappa_1 = \kappa_2 \frac{D_1}{D_2}.$$

Thus, for identical noise $D_1 = D_2$, setting $\kappa_1 = \kappa_2$ minimizes D_+. For much stronger noise in area 2 ($D_2 \gg D_1$), κ_1 should be made relatively small. In the case of noise correlations between areas ($D_c > 0$), the optimal value of κ_1 that minimizes (Equation 25) is

$$\kappa_1 = \kappa_2 \frac{D_1 - D_c}{D_2 - D_c}.$$

CALCULATING THE STOCHASTIC MOTION OF BUMPS

We now compute the effective variances (Equation 20) and (Equation 21), considering the specific case of Heaviside firing rate functions (Equation 5), cosine synaptic weights (Equation 3) and (Equation 4). Doing so, we can compare our asymptotic results to those computed from numerical simulations. We compute the mean reversion terms κ_1 and κ_2 by noting the spatial derivative of each bump will be $U_1'(x) = U_2'(x) = -2 \sin a \sin x$ and the null vector components are

$$f'(U_j(x)) U_j'(x) = \delta(x + a) - \delta(x - a).$$

for $j = 1, 2$. Plugging these formulae into Equation (16), we find $\kappa_1 = \varepsilon^{1/2} M_1$ and $\kappa_2 = \varepsilon^{1/2} M_2$.

We first consider the case of uncorrelated noise between areas, so $c_c \equiv 0$, meaning $D_c = 0$. We can compute the diffusion coefficients associated with the local noise in each area assuming cosine spatial correlations

$$D_1 = \frac{c_1 \varepsilon}{2 + 2\sqrt{1 - \theta^2}}, \quad D_2 = \frac{c_2 \varepsilon}{2 + 2\sqrt{1 - \theta^2}}. \tag{27}$$

We can then compute Equations (20) and (21) directly, for the case of no noise correlations between areas, by plugging in Equation (27).

For symmetric connections between areas, $\kappa = \varepsilon^{1/2} M_1 = \varepsilon^{1/2} M_2$, as well as identical noise, $c_1 = c_2 = 1$, we have $\langle \Delta_1(t)^2 \rangle = \langle \Delta_2(t)^2 \rangle = \langle \Delta(t)^2 \rangle$ and

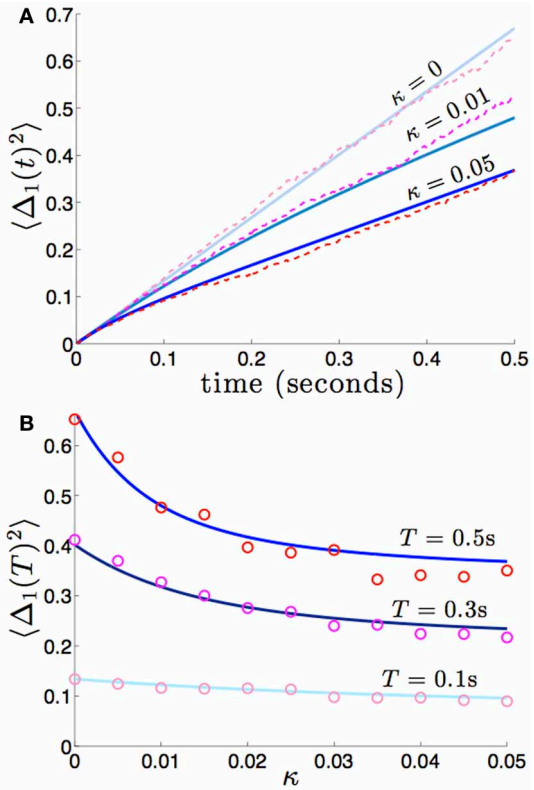


FIGURE 3 | Variance in the position of bumps as computed numerically (red shades) and from theory (blue shades) using Equation (28). Coupling between areas is symmetric $\sqrt{\varepsilon} w_{12}(x) = \sqrt{\varepsilon} w_{21}(x) = \kappa(\cos(x) + 1)$, so $\langle \Delta_1(t)^2 \rangle = \langle \Delta_2(t)^2 \rangle$, and there is no shared noise ($c_c = 0$). **(A)** The increase in variance is slower for stronger amplitudes of interareal coupling κ. Notice variance climbs sublinearly for $\kappa > 0$, due to the mean-reversion caused by coupling. **(B)** Variance drops considerably more over low values of κ that over high values. Other constituent functions and parameters are the same as in **Figure 2**.

$$\langle \Delta(t)^2 \rangle = \frac{\varepsilon t}{4(1 + \sqrt{1 - \theta^2})} + \frac{\varepsilon}{16(1 + \sqrt{1 - \theta^2})\kappa} \left[1 - e^{-4\kappa t}\right]. \tag{28}$$

We compare the formula (28) to results we obtain from numerical simulations in **Figure 3**, finding our asymptotic formula (28) matches quite well. In addition, we compare our results for general (possibly asymmetric) reciprocal connectivity to results from numerical simulations in **Figure 4**. We also show in **Figure 5**, as predicted, when κ_2 is held fixed, there is a finite optimal value of κ_1 that minimizes variance $\langle \Delta_1(t)^2 \rangle$. Therefore, reciprocal connectivity in multi-area networks should be balanced, in order to minimize positional variance of the stored bump.

Next, we consider the case of correlated noise between areas, so $c_c > 0$, meaning $D_c > 0$. In this case, the covariance terms in D_+ and D_- are non-zero. We can thus compute the diffusion coefficient associated with correlated noise

$$D_c = \frac{c_c \varepsilon}{2 + 2\sqrt{1 - \theta^2}}.$$

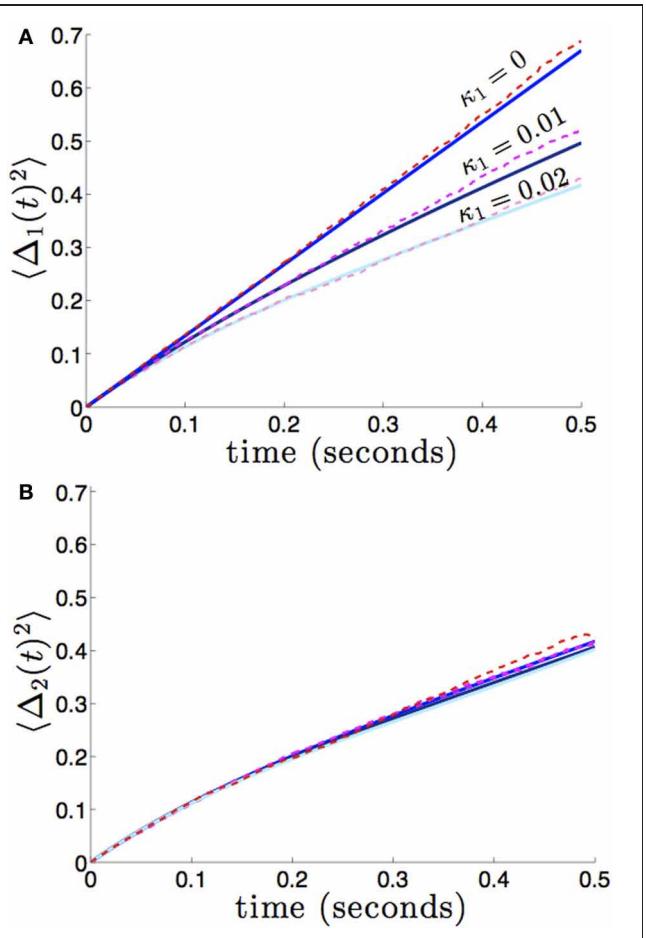


FIGURE 4 | Variance in the position of bumps as it depends on asymmetric reciprocal connectivity ($\kappa_1 \neq \kappa_2$) when noise in each area is independent and identical ($c_1 = c_2 = 1$). Fixing $\kappa_2 = 0.02$ and varying κ_1, we find **(A)** the variance $\langle \Delta_1(t)^2$ of bump 1 decreases as coupling from area 2 to 1 (κ_1) increases; **(B)** variance $\langle \Delta_2(t)^2 \rangle$ of bump 2 remain relatively unchanged. Other constituent functions and parameters are the same as in **Figure 2**.

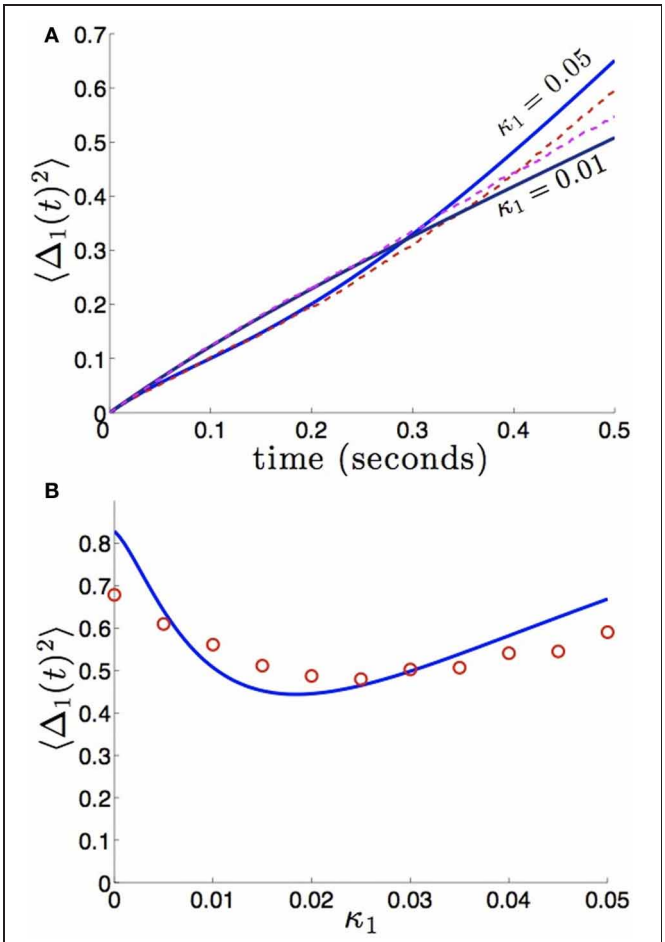


FIGURE 5 | Bump position variance depends non-monotonically on asymmetric connectivity strength. (A) For $\kappa_2 = 0.01$ and high enough values of coupling ($\kappa_1 = 0.05$), variance $\langle \Delta_1(t)^2 \rangle$ scales more quickly than for symmetric coupling ($\kappa_1 = 0.01$). Layer 1 is being sourced by the noisier area 2. **(B)** Non-monotonic dependence of variance $\langle \Delta_1(t)^2 \rangle$ on projection strength from area 2 to area 1 κ_1 is shown for fixed time $T = 50$ and $\kappa_2 = 0.01$ fixed. Amplitude of noise in area 2 is twice that of area 1 ($c_1 = 1$ and $c_2 = 2$). Other constituent functions and parameters are the same as in **Figure 2**.

In the case of symmetric connections between areas, $\kappa = \varepsilon^{1/2} M_1 = \varepsilon^{1/2} M_2$, and identical internal noise, $c_1 = c_2 = 1$, we have $\langle \Delta_1(t)^2 \rangle = \langle \Delta_2(t)^2 \rangle = \langle \Delta(t)^2 \rangle$ and

$$\langle \Delta(t)^2 \rangle = \frac{(1+c_c)\varepsilon}{4\left(1+\sqrt{1-\theta^2}\right)} t + \frac{(1-c_c)\varepsilon}{16\left(1+\sqrt{1-\theta^2}\right)\kappa}\left[1-\mathrm{e}^{-4\kappa t}\right], \tag{29}$$

which reflects the fact that interareal connections do not reduce variability as much when there are strong noise correlations c_c between areas. We demonstrate the accuracy of the theoretical calculation (Equation 29) as compared to numerical simulations in **Figure 6**. Numerical simulations also reveal the fact that stronger noise correlations between areas diminish the effectiveness of interareal connections at reducing bump position variance.

REDUCTION OF BUMP WANDERING IN MULTIPLE AREAS

We now examine the effect of interareal connections in networks with more than two areas using the system (Equation 6). As with the dual area network without noise or interareal connectivity, stationary bump solutions take the form $(u_1, \ldots, u_N) = (U_1(x), \ldots, U_N(x))$, and translation invariance let us to set all bump peaks to be located at $x = 0$ so

$$U_j = w * f(U_j), \quad j = 1, \ldots, N. \tag{30}$$

As before, we presume $w_{jj} = w$, and relaxing this assumption does not dramatically alter our results. Linear stability analysis of bumps proceeds along similar lines to the dual area network, so we omit those calculations and summarize the results. In the absence of interareal connections, each bump is neutrally stable to

perturbation in either direction. In the presence of interareal connections, all bumps are only neutrally stable to translations that move them all in the same direction. Therefore, networks with more areas provide more perturbation cancelations.

To study how noise and interareal connections affect the trajectory of bump positions, we again note noise causes all bumps to wander away from their initial position, while being pulled back into place by projections from other areas (see **Figure 7**). The position of the bump in area j is described by the stochastic variable Δ_j. Noise also causes fluctuations in the shape of both bumps, which is described by the correction term Φ_j. Therefore, we presume the resulting state of the system satisfies the ansatz

$$u_j = U_j(x - \Delta_j(t)) + \varepsilon^{1/2}\Phi_j(x - \Delta_j(t), t) + \cdots,$$

where $j = 1, \ldots, N$. Plugging this ansatz into Equation (6) and expanding in powers of $\varepsilon^{1/2}$, we find that at $\mathcal{O}(1)$, we simply have the system of Equation (30) for the bump solutions. Proceeding to $\mathcal{O}(\varepsilon^{1/2})$, we find

$$d\mathbf{\Phi} - \mathcal{L}\mathbf{\Phi} = \mathcal{K}(x,t) + \begin{pmatrix} \varepsilon^{-1/2}\dot{\Delta}_1 U_1' + dW_1 \\ \vdots \\ \varepsilon^{-1/2}\dot{\Delta}_j U_j' + dW_j \\ \vdots \\ \varepsilon^{-1/2}\dot{\Delta}_N U_N' + dW_N \end{pmatrix} \tag{31}$$

where $\mathcal{K}(x, t)$ is an $N \times 1$ vector whose jth entry is

$$\mathcal{K}_j = \sum_{k \neq j} w_{jk} * \left[f(U_k) + f'(U_k)U_k' \cdot (\Delta_k - \Delta_j) \right] dt;$$

$\mathbf{\Phi} = (\Phi_1(x, t), \cdots, \Phi_N(x, t))^T$; and $\mathcal{L}$ is the linear operator

$$\mathcal{L}\mathbf{\Psi} = \begin{pmatrix} -\Psi_1(x) + w * [f'(U_1(x))\Psi_1(x) \\ \vdots \\ -\Psi_N(x) + w * [f'(U_N)\Psi_N(x) \end{pmatrix}$$

for any integrable vector $\mathbf{\Psi} = (\Psi_1(x), \ldots, \Psi_N(x))^T$. The nullspace of $\mathcal{L}$ is spanned by the vectors $(U_1', 0, \ldots, 0)^T$; $(0, U_2', 0, \ldots, 0)^T$; …; and $(0, \ldots, 0, U_N')^T$, which can be seen

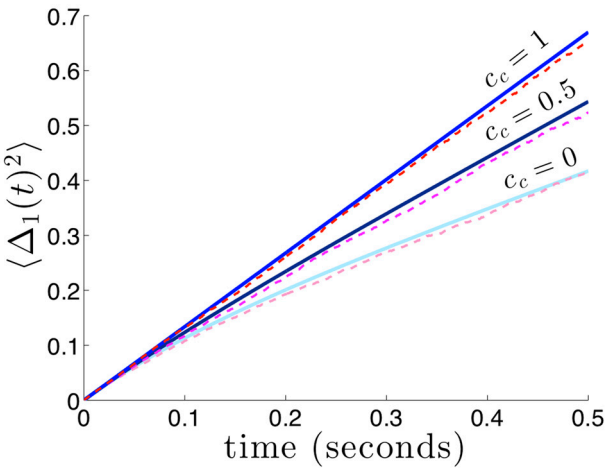


FIGURE 6 | Variance in the position of bumps as noise correlation between areas is increased. Numerically computed variance (red shades) match theoretical curves from Equation (29), blue shades, very well. Reciprocal connectivity reduces variability the most when there is no correlated noise ($c_c = 0$) between areas. As the shared noise between areas increased is amplitude ($c_c = 0.5, 1$), the advantage of reciprocal connectivity is diminished. When $c_c = 1$ changing κ does not affect the variance $\langle \Delta(t)^2 \rangle$ (see formula (29) in the limit $c_c \to 1$). Other constituent functions and parameters are the same as in **Figure 2**.

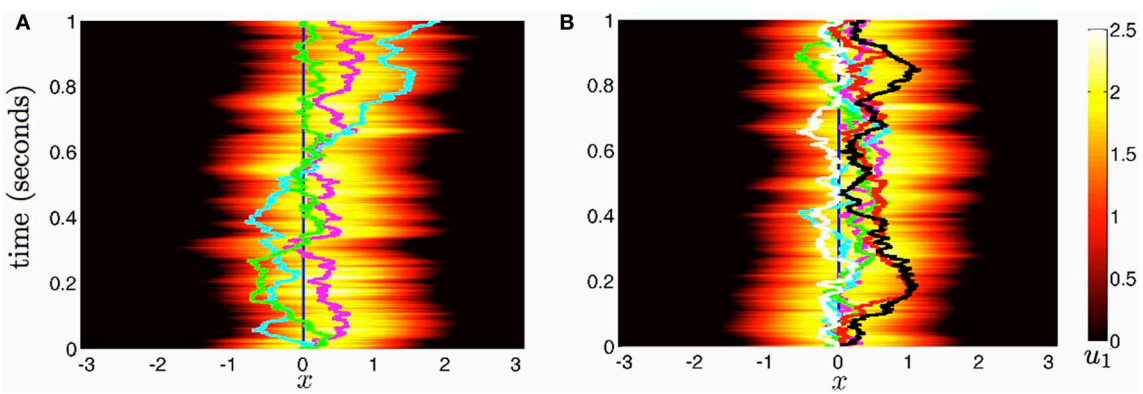


FIGURE 7 | Stochastic evolution of bump position in multi-area networks. (A) With weak coupling ($\sqrt{\varepsilon}w_{jk}(x) = 0.01(\cos(x) + 1)$ for $j \neq k$) between $N = 3$ areas, the position of bumps 1 (magenta), 2 (cyan), and 3 (green) reverts to one another. We show only the evolution of activity $u(x, t)$ in area 1. **(B)** For $N = 6$ areas and the same interareal coupling, the reduction in bump wandering is even more considerable. The trajectories of bumps in all areas (colored lines) stay close together. All other parameters are as in **Figure 2**.

by differentiating (Equation 30). The last terms on the right hand side of Equation (31) arise due to interareal connections. We have linearized them under the assumption that $|\Delta_k - \Delta_j|$ remains small for all j, k. To ensure a solution to Equation (31), we require the right hand side is orthogonal to all elements of the null space of the adjoint operator $\mathcal{L}^*$. The adjoint is defined with respect to the inner product

$$\int_{-\pi}^{\pi} \boldsymbol{\Upsilon}^T \mathcal{L}\boldsymbol{\Psi} \mathrm{d}x = \int_{-\pi}^{\pi} \boldsymbol{\Psi}^T \mathcal{L}^* \boldsymbol{\Upsilon} \mathrm{d}x$$

where $\boldsymbol{\Upsilon} = (\Upsilon_1(x), \ldots, \Upsilon_N(x))^T$ is integrable. It then follows

$$\mathcal{L}^*\boldsymbol{\Upsilon} = \begin{pmatrix} -\Upsilon_1(x) + f'(U_1(x))[w * \Upsilon_1] \\ \vdots \\ -\Upsilon_N(x) + f'(U_N)[w * \Upsilon_N] \end{pmatrix}.$$

The nullspace of $\mathcal{L}^*$ contains the vectors $(f'(U_1)U_1', 0, \ldots, 0)^T$; $(0, f'(U_2)U_2', 0, \ldots 0)^T$; …; and $(0, \ldots, 0, f'(U_N)U_N')$, which can be shown by applying $\mathcal{L}^*$ to them and using the formula generated by differentiating (Equation 30). Thus, to be sure (Equation 31) has a solution, we take the inner product of both sides of the equation with all N null vectors and isolate $\mathrm{d}\Delta_j$ terms to yield the multivariate Ornstein–Uhlenbeck process

$$\mathrm{d}\boldsymbol{\Delta}(t) = \mathbf{K}\boldsymbol{\Delta}(t)\mathrm{d}t + \mathrm{d}\mathbf{W}(t), \tag{32}$$

where effects of interareal connections are described by the matrix $\mathbf{K} \in \mathbb{R}^{N\times N}$ where the diagonal and off-diagonal entries are given

$$\mathbf{K}_{jj} = -\sum_{k\neq j} \kappa_{jk}, \qquad \mathbf{K}_{jk} = \kappa_{jk}$$

for $j = 1, \ldots, N$ and $k \neq j$, where

$$\kappa_{jk} = \frac{\langle f'(U_j)U_j', \varepsilon^{1/2} w_{jk} * [f'(U_k)U_k']\rangle}{\langle f'(U_j)U_j', U_j'\rangle},$$

and we have used the fact that $w_{jk} * f(U_k)\cdot U_j'$ is an odd function for all j, k, so they vanish on integration. Stochastic forces are described by the vector

$$\mathrm{d}\mathbf{W}(t) = \begin{pmatrix} \mathrm{d}\mathcal{W}_1(t) \\ \vdots \\ \mathrm{d}\mathcal{W}_N(t) \end{pmatrix},$$

$$\mathrm{d}\mathcal{W}_j(t) = -\varepsilon^{1/2}\frac{\langle f'(U_j)U_j', \mathrm{d}W_j\rangle}{\langle f'(U_j)U_j', U_j'\rangle}.$$

The white noise vector $\mathbf{W}(t)$ has zero mean $\langle \mathbf{W}(t)\rangle = \mathbf{0}$, and covariance matrix $\langle \mathbf{W}(t)\mathbf{W}^T(t)\rangle = \mathbf{D}t$ where associated coefficients of the matrix $\mathbf{D}$ are

$$D_{jj} = \varepsilon\frac{\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} F_j(x)F_j(y)C_j(x-y)\mathrm{d}x\mathrm{d}y}{\left[\int_{-\pi}^{\pi} F_j(x)U_j'(x)\mathrm{d}x\right]^2}.$$

where $F_j(x) = f'(U_j(x))U_j'(x)$, which describe the variance within an area and

$$D_{jk} = \varepsilon\frac{\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} F_j(x)F_k(y)C_{jk}(x-y)\mathrm{d}x\mathrm{d}y}{\left[\int_{-\pi}^{\pi} F_j(x)U_j'(x)\mathrm{d}x\right]\left[\int_{-\pi}^{\pi} F_k(x)U_k'(x)\mathrm{d}x\right]},$$

which describes covariance between areas. Since correlations are symmetric $C_{jk}(x) = C_{kj}(x)$ for all j, k, then $D_{jk} = D_{kj}$ for all j, k.

A detailed analysis of the linear stochastic system (Equation 32) is difficult without some knowledge of the entries κ_{jk}. However, we can make a few general statements. We note that all eigenvalues of $\mathbf{K}$ must have negative real part or be zero, due to the Gerschgorin circle theorem (Feingold and Varga, 1962), which states that all eigenvalues a matrix $\mathbf{K}$ must lie in one of the disks with center K_{jj} and radius $\sum_{k\neq j}|K_{jk}|$. Since $K_{jj} = -\sum_{k\neq j}\kappa_{jk}$ and $K_{jk} = \kappa_{jk}$, then

$$K_{jj} + \sum_{k\neq j} K_{jk} = -\sum_{k\neq j}\kappa_{jk} + \sum_{k\neq j}|\kappa_{jk}| = 0 \tag{33}$$

is the maximal possible eigenvalue, since $\kappa_{jk} \geq 0$ for all j, k. Therefore, we expect N eigenpairs λ_j, $\mathbf{v}_k$ associated with $\mathbf{K}$, where $\lambda_N \leq \lambda_{N-1} \leq \cdots \leq \lambda_2 \leq \lambda_1 = 0$. This means we can perform the diagonalization $\mathbf{K} = \mathbf{V}\Lambda\mathbf{V}^{-1}$, where Λ is the diagonal matrix of eigenvalues; columns of $\mathbf{V}$ are right eigenvectors; and rows of $\mathbf{V}^{-1}$ are left eigenvectors. Therefore, we can decompose the stochastic solution to Equation (32), when $\boldsymbol{\Delta}(0) = \mathbf{0}$ as

$$\boldsymbol{\Delta}(t) = \int_0^t \mathrm{e}^{\mathbf{K}(t-s)}\mathrm{d}\mathbf{W}(s) = \int_0^t \mathbf{V}\mathrm{e}^{\Lambda(t-s)}\mathbf{V}^{-1}\mathrm{d}\mathbf{W}(s),$$

Thus, as we expect, any stochastic fluctuations in Equation (32) will be integrated or decay over time due to the exponential filters $\mathrm{e}^{\lambda_j(t-s)}$. In addition, when $\boldsymbol{\Delta}(0) = \mathbf{0}$ the covariance matrix can be computed as

$$\langle \boldsymbol{\Delta}(t)\boldsymbol{\Delta}^T(t)\rangle = \int_0^t \mathrm{e}^{\mathbf{K}(t-s)}\mathbf{D}\mathrm{e}^{\mathbf{K}^T(t-s)}\mathrm{d}s, \tag{34}$$

where $\mathbf{D}$ is the matrix of diffusion coefficients for the covariance $\langle \mathbf{W}(t)\mathbf{W}^T(t)\rangle$. We now compute the covariance in the specific case of symmetric connectivity.

In the case of symmetric connectivity between areas, $w_{jk} = w_r$ for all $j \neq k$, so $\kappa_{jk} = \kappa$ for all $j \neq k$. Effects of connectivity

between areas are described by the symmetric matrix

$$\mathbf{K} = \kappa J_N - N\kappa I$$

where J_N is the $N \times N$ matrix of ones and I is the identity. The eigenvalues of J_N are N, with multiplicity one, and zero, with multiplicity $N-1$. Thus, the largest eigenvalue of $\mathbf{K} = \kappa J_N - N\kappa I$ is $\lambda_1 = 0$ with associated eigenvector $\mathbf{v}_1 = (1, \ldots, 1)^T$. All other eigenvalues are $\lambda_j = -N\kappa$ for $j \geq 2$, with associated eigenvectors $\mathbf{v}_j = \mathbf{e}_1 - \mathbf{e}_j$, where $j = 2, \ldots, N$ and $\mathbf{e}_j$ is the unit vector with a one in the jth row and zeros elsewhere. Our diagonalization of the symmetric matrix $\mathbf{K} = \mathbf{K}^T = \mathbf{V}\Lambda\mathbf{V}^{-1}$ then involves the diagonal matrix Λ of eigenvalues λ_j; the symmetric matrix $\mathbf{V}$ whose columns $\mathbf{v}_j$ are right eigenvectors; and the symmetric matrix $\mathbf{V}^{-1}$ whose rows are left eigenvectors. The matrix $\mathbf{V}^{-1}$ takes the form

$$\mathbf{V}^{-1} = \frac{1}{N}\begin{pmatrix} 1 & 1 & \cdots & \\ 1 & -(N-1) & 1 & \cdots \\ & & \ddots & 1 \\ 1 & \cdots & 1 & -(N-1) \end{pmatrix}.$$

We can thus compute the covariance using the diagonalization $e^{\mathbf{K}t} = e^{\mathbf{K}^T t} = \mathbf{V}e^{\Lambda t}\mathbf{V}^{-1}$. In addition, we will assume each area receives noise with identical statistics ($D_{jj} = D_l$) and there are identical noise correlations between areas ($D_{jk} = D_c$ for $j \neq k$), so $\mathbf{D} = (D_l - D_c)I + D_c J_N$. Multiplying and integrating (Equation 34), we find the diagonal entries (variances) of $\langle \mathbf{\Delta}(t)\mathbf{\Delta}^T(t)\rangle$ are

$$\langle \Delta_j(t)^2 \rangle = \frac{D_l + (N-1)D_c}{N} t + \frac{(N-1)(D_l - D_c)}{2N^2\kappa}\left[1 - e^{-2N\kappa t}\right], \tag{35}$$

and the off-diagonal entries (true covariances) are

$$\langle \Delta_j(t)\Delta_k(t)\rangle = \frac{D_l + (N-1)D_c}{N} t - \frac{(D_l - D_c)}{2N^2\kappa}\left[1 - e^{-2N\kappa t}\right].$$

As revealed by the diffusive term in Equation (35), the system still possesses a rotational symmetry, given by the action of rotating all the bumps in the same direction. Thus, the component of noise in this direction is not damped out by coupling. Thus, note that the long term variance of any bump's position $\Delta_j(t)$ will be approximately described by the averaged diffusion

$$\lim_{t\to\infty} \langle \Delta_j(t)^2 \rangle = \frac{D_l + (N-1)D_c}{N} t.$$

As the strength of coupling κ or number of areas N is increased, the variances $\langle \Delta_j(t)^2 \rangle$ approach this limit at a faster rate, since the other portions of variance decay at a rate proportional to $|\lambda_2| = N\kappa$. Note also that in the limit $D_c \to D_l$, effects of coupling are negligible and the long term variance of each bump is determined by the diffusion introduced by its area's internal noise.

Returning to study the full variance Equation (35) for symmetric coupling and noise, we make a few observations. First, in the limit of purely correlated noise across areas ($D_c \to D_l$), interareal connections have no effect, and $\langle \Delta_j(t)^2 \rangle = D_l t$ for all areas and arbitrary coupling strength. However, if there is any independent noise in each area ($D_c < D_l$), variance $\langle \Delta_j(t)^2 \rangle$ can always be reduced further by increasing coupling strength or the number of areas since

$$\frac{d}{d\kappa}\langle \Delta_j(t)^2 \rangle = \frac{(N-1)(D_l - D_c)}{2N^2} \times \frac{(1 + 2N\kappa)e^{-2N\kappa t}) - 1}{\kappa^2} \leq 0,$$

where inequality $(1 + 2N\kappa t) \geq e^{2N\kappa t}$ holds due to the Taylor expansion of $e^{2N\kappa t}$ when $N\kappa t \geq 0$, and

$$\frac{d}{dN}\langle \Delta_j(t)^2 \rangle = -\frac{D_l - D_c}{N^2} + \frac{D_l - D_c}{2N^3\kappa}\left[2(1 + N\kappa t)e^{-2N\kappa t} - N\right] \leq 0$$

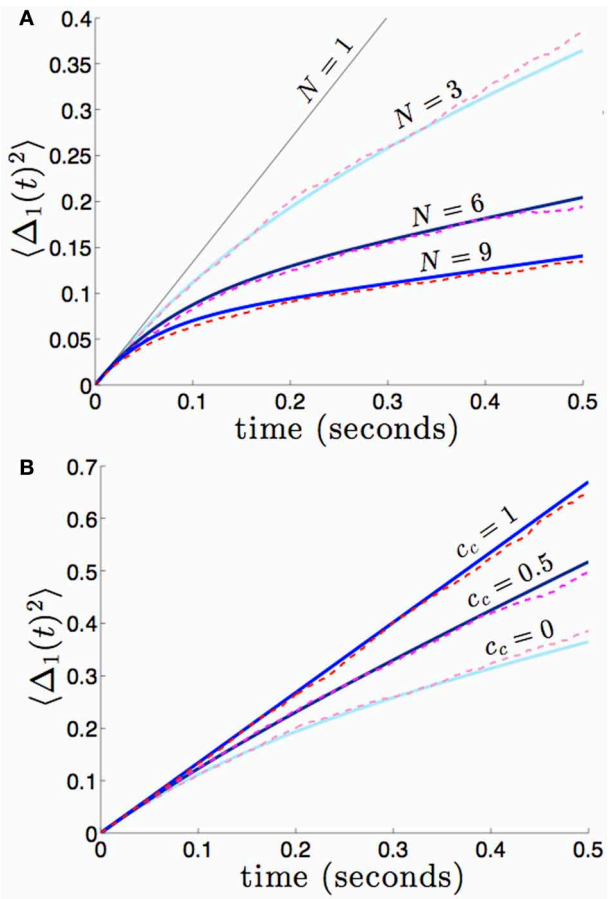


FIGURE 8 | (A) Variance in the position of the bump in the first area $\langle \Delta_1(t)^2 \rangle$ builds up more slowly in networks with more areas N, and we expect similar behavior in all other areas. Fixing the strength of interareal connections, $\sqrt{\varepsilon}w_{jk}(x) = 0.01(\cos(x) + 1)$ for $j \neq k$, we see that varying N decreases the variance $\langle \Delta_j(t)^2 \rangle$. **(B)** As in dual area networks, increasing the level of noise correlations between areas diminishes the effectiveness of interareal connectivity as a noise cancelation mechanism. Other parameters are as in **Figure 2**.

when $N \geq 2$, since $D_l \geq D_c$ and due to the Taylor expansion of $e^{2N\kappa t}$. Note, we have temporarily treated N as a continuous variable. Thus, we know the variance $\langle \Delta_j(t)^2 \rangle$ to decrease with increasing κ and expect it to decrease with increasing N.

We can compute the variance $\langle \Delta_j(t)^2 \rangle$ explicitly in the case of Heaviside firing rate functions (Equation 5), cosine synaptic weights (Equation 3) and (Equation 4). With these assumptions, as well as there being identical noise to all areas ($c_{jj} = 1$ for all j, $c_{jk} = c_c$ for $j \neq k$), we find

$$D_l = \frac{\varepsilon}{2 + 2\sqrt{1-\theta^2}}, \quad D_c = \frac{c_c \varepsilon}{2 + 2\sqrt{1-\theta^2}},$$

so that

$$\langle \Delta_j(t)^2 \rangle = \frac{(1 + (N-1)c_c)\varepsilon}{2N(1+\sqrt{1-\theta^2})} t + \frac{(1-c_c)\varepsilon}{4N^2\kappa} \left[1 - e^{-2N\kappa t}\right], \tag{36}$$

which reflects the fact that increasing the number of areas will decrease variability, when noise between areas is not too strongly correlated. We demonstrate the accuracy of this formula (36) in **Figure 8**. In numerical simulations, as predicted by our asymptotic calculations, the variance scales more slowly in time in networks with more areas.

DISCUSSION

We have shown that interareal coupling in multi-area stochastic networks can reduce the diffusive wandering of bumps. Since bump attractors offer a well studied model of persistent activity underlying spatial working memory (Compte et al., 2000), our results provide a novel suggestion for how the memory networks may reduce error. Our calculations have exploited a small noise approximation for the position of the bump in each area (Armero et al., 1998; Bressloff and Webber, 2012a). Assuming connectivity between areas is weak, we have shown the equations describing bump positions reduce to a multivariate Ornstein–Uhlenbeck process. In this formulation, we find interareal connectivity stabilizes all but one eigendirection in the space of bump position movements. Neutral stability does still exist, so stochastic forces that move bumps in all areas in the same direction do not decay away. However, sources of noise that force bumps in opposite directions create bump movements that will decay with time. Thus, interareal connectivity provides a noise cancelation mechanism that operates by stabilizing the bumps in each area to stochastic forces that push them in opposite directions. (Polk et al., 2012) recently explored noise correlation statistics in persistent state networks that reduce wandering. Our work complements these results by studying synaptic architectures that limit persistent state diffusion.

Storing spatial working memories with neural activity that spans multiple brain areas does serve other purposes than potential noise cancelation. Delayed response tasks that lead to limb motion can generate persistent activity in the parietal cortex (Colby et al., 1996; Pesaran et al., 2002) so that motor responses can be readily executed. In addition, superior colliculus demonstrates sustained activity (Basso and Wurtz, 1997), which is an area also thought to underlie directed behavioral responses. Therefore, activity is distributed between areas providing short term information storage, like prefrontal cortex (Goldman-Rakic, 1995), and those responsible for motor responses and/or behavior. An additional effect of this delegation of activity is that reciprocal connections between areas may provide noise cancelation during the storage period of working memory. However, our work suggests distributing working memory-serving neural activity between areas that receive strongly correlated noise will not provide as effective cancelation.

Our work should be contrasted with several other results concerning the stabilization of networks that encode a continuous variable (Koulakov et al., 2002; Goldman et al., 2003; Cain and Shea-Brown, 2012; Kilpatrick et al., 2013). Pure integrators, which are usually line attractors, are notoriously fragile to parametric perturbations, so (Koulakov et al., 2002) suggested they may be made more robust by considering networks that integrate in discrete bursts, rather than continuously. This can be implemented by considering a population of bistable neural units so that firing rate integration of a stimulus occurs in a stairstep fashion, rather than a ramplike fashion (see Goldman et al., 2003 for example). Related ideas were recently implemented in a bump attractor model of spatial working memory (Kilpatrick et al., 2013), but quantization was implemented with synaptic architecture rather than single neural unit properties. As opposed to the approach of quantizing the space of possible stimulus representations, we have kept the representation space a continuum. Deleterious effects of noise are reduced by considering reciprocal connectivity between encoding areas that redundantly represent the stimulus. Due to noise cancelations, the encoding error of the network decreases as the number of areas is increased.

REFERENCES

Amari, S. (1977). Dynamics of pattern formation in lateral-inhibition type neural fields. *Biol. Cybern.* 27, 77–87. doi: 10.1007/BF00337259

Armero, J., Casademunt, J., Ramirez-Piscina, L., and Sancho, J. M. (1998). Ballistic and diffusive corrections to front propagation in the presence of multiplicative noise. *Phys. Rev. E* 58, 5494–5500. doi: 10.1103/PhysRevE.58.5494

Basso, M. A., and Wurtz, R. H. (1997). Modulation of neuronal activity by target uncertainty. *Nature* 389, 66–69. doi: 10.1038/37975

Ben-Yishai, R., Bar-Or, R. L., and Sompolinsky, H. (1995). Theory of orientation tuning in visual cortex. *Proc. Natl. Acad. Sci. U.S.A.* 92, 3844–3848. doi: 10.1073/pnas.92.9.3844

Bressloff, P. C., and Webber, M. A. (2012a). Front propagation in stochastic neural fields. *SIAM J. Appl. Dyn. Syst.* 11, 708–740. doi: 10.1137/110851031

Bressloff, P. C., and Webber, M. A. (2012b). Neural field model of binocular rivalry waves. *J. Comput. Neurosci.* 32, 233–252. doi: 10.1007/s10827-011-0351-y

Cain, N., and Shea-Brown, E. (2012). Computational models of decision making: integration, stability, and noise. *Curr. Opin. Neurobiol.* 22, 1047–1053. doi: 10.1016/j.conb.2012.04.013

Camperi, M., and Wang, X. J. (1998). A model of visuospatial working memory in prefrontal cortex: recurrent network and cellular bistability. *J. Comput. Neurosci.* 5, 383–405. doi: 10.1023/A:1008837311948

Chafee, M. V., and Goldman-Rakic, P. S. (1998). Matching patterns of activity in primate prefrontal area 8a and parietal area 7ip neurons during

a spatial working memory task. *J. Neurophysiol.* 79, 2919–2940.

Clements, J. D., Lester, R. A., Tong, G., Jahr, C. E., and Westbrook, G. L. (1992). The time course of glutamate in the synaptic cleft. *Science* 258, 1498–1501. doi: 10.1126/science.1359647

Colby, C. L., Duhamel, J. R., and Goldberg, M. E. (1996). Visual, presaccadic, and cognitive activation of single neurons in monkey lateral intraparietal area. *J. Neurophysiol.* 76, 2841–2852.

Compte, A., Brunel, N., Goldman-Rakic, P. S., and Wang, X. J. (2000). Synaptic mechanisms and network dynamics underlying spatial working memory in a cortical network model. *Cereb. Cortex* 10, 910–923. doi: 10.1093/cercor/10.9.910

Constantinidis, C., and Wang, X.-J. (2004). A neural circuit basis for spatial working memory. *Neuroscientist* 10, 553–565. doi: 10.1177/1073858404268742

Curtis, C. E. (2006). Prefrontal and parietal contributions to spatial working memory. *Neuroscience* 139, 173–180. doi: 10.1016/j.neuroscience.2005.04.070

Curtis, C. E., and Lee, D. (2010). Beyond working memory: the role of persistent activity in decision making. *Trends Cogn. Sci.* 14, 216–222. doi: 10.1016/j.tics.2010.03.006

di Pellegrino, G., and Wise, S. P. (1993). Visuospatial versus visuomotor activity in the premotor and prefrontal cortex of a primate. *J. Neurosci.* 13, 1227–1243.

Ermentrout, B. (1998). Neural networks as spatio-temporal pattern-forming systems. *Rep. Progress Phys.* 61, 353. doi: 10.1088/0034-4885/61/4/002

Feingold, D. G., and Varga, R. S. (1962). Block diagonally dominant matrices and generalizations of the gerschgorin circle theorem. *Pacific J. Math.* 12, 1241–1250.

Ferster, D., and Miller, K. D. (2000). Neural mechanisms of orientation selectivity in the visual cortex. *Annu. Rev. Neurosci.* 23, 441–471. doi: 10.1146/annurev.neuro.23.1.441

Folias, S. E., and Ermentrout, G. B. (2011). New patterns of activity in a pair of interacting excitatory-inhibitory neural fields. *Phys. Rev. Lett.* 107:228103. doi: 10.1103/PhysRevLett.107.228103

Funahashi, S., Bruce, C. J., and Goldman-Rakic, P. S. (1989). Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *J. Neurophysiol.* 61, 331–349.

Gardiner, C. W. (2003). *Handbook of Stochastic Methods*. New York, NY: Springer.

Goldman, M. S., Levine, J. H., Major, G., Tank, D. W., and Seung, H. S. (2003). Robust persistent neural activity in a model integrator with multiple hysteretic dendrites per neuron. *Cereb. Cortex* 13, 1185–1195. doi: 10.1093/cercor/bhg095

Goldman-Rakic, P. S. (1995). Cellular basis of working memory. *Neuron* 14, 477–485. doi: 10.1016/0896-6273(95)90304-6

Gupta, A., Wang, Y., and Markram, H. (2000). Organizing principles for a diversity of gabaergic interneurons and synapses in the neocortex. *Science* 287, 273–278. doi: 10.1126/science.287.5451.273

Hansel, D., and Mato, G. (2013). Short-term plasticity explains irregular persistent activity in working memory tasks. *J. Neurosci.* 33, 133–149. doi: 10.1523/JNEUROSCI.3455-12.2013

Hansel, D., and Sompolinsky, H. (1998). "Modeling feature selectivity in local cortical circuits," in *Methods in Neuronal Modeling: From Ions to Networks*, chapter 13, eds C. Koch, and I. Segev (Cambridge: MIT), 499–567.

Häusser, M., and Roth, A. (1997). Estimating the time course of the excitatory synaptic conductance in neocortical pyramidal cells using a novel voltage jump method. *J. Neurosci.* 17, 7606–7625.

Hikosaka, O., Sakamoto, M., and Usui, S. (1989). Functional properties of monkey caudate neurons. iii. activities related to expectation of target and reward. *J. Neurophysiol.* 61, 814–832.

Itskov, V., Hansel, D., and Tsodyks, M. (2011). Short-term facilitation may stabilize parametric working memory trace. *Front. Comput. Neurosci.* 5:40. doi: 10.3389/fncom.2011.00040

Kilpatrick, Z. P., and Bressloff, P. C. (2010). Binocular rivalry in a competitive neural network with synaptic depression. *SIAM J. Appl. Dyn. Syst.* 9, 1303–1347. doi: 10.1137/100788872

Kilpatrick, Z. P., and Ermentrout, B. (2013). Wandering bumps in stochastic neural fields. *SIAM J. Appl. Dyn. Syst.* 12, 61–94. doi: 10.1137/120877106

Kilpatrick, Z. P., Ermentrout, B., and Doiron, B. (2013). Optimizing working memory with spatial heterogeneity of recurrent cortical excitation. (Submitted).

Koulakov, A. A., Raghavachari, S., Kepecs, A., and Lisman, J. E. (2002). Model for a robust neural integrator. *Nat. Neurosci.* 5, 775–782. doi: 10.1038/nn893

Laing, C. R., and Chow, C. C. (2001). Stationary bumps in networks of spiking neurons. *Neural. Comput.* 13, 1473–1494. doi: 10.1162/089976601750264974

Levy, R., Friedman, H. R., Davachi, L., and Goldman-Rakic, P. S. (1997). Differential activation of the caudate nucleus in primates performing spatial and nonspatial working memory tasks. *J. Neurosci.* 17, 3870–3882.

Ly, C., and Ermentrout, G. B. (2010). Coupling regularizes individual units in noisy populations. *Phys. Rev. E Stat. Nonlin. Soft. Matter. Phys.* 81(1 Pt 1):011911. doi: 10.1103/PhysRevE.81.011911

McNab, F., and Klingberg, T. (2008). Prefrontal cortex and basal ganglia control access to working memory. *Nat. Neurosci.* 11, 103–107. doi: 10.1038/nn2024

Miller, E. K., Erickson, C. A., and Desimone, R. (1996). Neural mechanisms of visual working memory in prefrontal cortex of the macaque. *J. Neurosci.* 16, 5154–5167.

Mushiake, H., and Strick, P. L. (1995). Pallidal neuron activity during sequential arm movements. *J. Neurophysiol.* 74, 2754–2758.

Owen, A. M., Evans, A. C., and Petrides, M. (1996). Evidence for a two-stage model of spatial working memory processing within the lateral frontal cortex: a positron emission tomography study. *Cereb. Cortex* 6, 31–38. doi: 10.1093/cercor/6.1.31

Pesaran, B., Pezaris, J. S., Sahani, M., Mitra, P. P., and Andersen, R. A. (2002). Temporal structure in neuronal activity during working memory in macaque parietal cortex. *Nat. Neurosci.* 5, 805–811. doi: 10.1038/nn890

Ploner, C. J., Gaymard, B., Rivaud, S., Agid, Y., and Pierrot-Deseilligny, C. (1998). Temporal limits of spatial working memory in humans. *Eur. J. Neurosci.* 10, 794–797. doi: 10.1046/j.1460-9568.1998.00101.x

Polk, A., Litwin-Kumar, A., and Doiron, B. (2012). Correlated neural variability in persistent state networks. *Proc. Natl. Acad. Sci. U.S.A.* 109, 6295–6300. doi: 10.1073/pnas.1121274109

White, J. M., Sparks, D. L., and Stanford, T. R. (1994). Saccades to remembered target locations: an analysis of systematic and variable errors. *Vis. Res.* 34, 79–92. doi: 10.1016/0042-6989(94)90259-3

Wilson, H. R., and Cowan, J. D. (1973). A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue. *Biol. Cybern.* 13, 55–80. doi: 10.1007/bf00288786

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.

Received: 29 April 2013; paper pending published: 24 May 2013; accepted: 11 June 2013; published online: 01 July 2013.

Citation: Kilpatrick ZP (2013) Interareal coupling reduces encoding variability in multi-area models of spatial working memory. Front. Comput. Neurosci. 7:82. doi: 10.3389/fncom.2013.00082

Copyright © 2013 Kilpatrick. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and subject to any copyright notices concerning any third-party graphics etc.