On the Construction of Sparse Matrices From Expander Graphs

Bubacarr Bah [1,2] and Jared Tanner [3]*

[1] The Research Centre, African Institute for Mathematical Sciences, Muizenberg, South Africa, [2] Department of Mathematical Sciences, University of Stellenbosch, Stellenbosch, South Africa, [3] Mathematics Institute, University of Oxford, Oxford, United Kingdom

We revisit the asymptotic analysis of probabilistic construction of adjacency matrices of expander graphs proposed in Bah and Tanner [1]. With better bounds we derived a new reduced sample complexity for d, the number of non-zeros per column of these matrices (or equivalently the left-degree of the underlying expander graph). Precisely $d = \mathcal{O}\left(\log_s(N/s)\right)$; as opposed to the standard $d = \mathcal{O}\left(\log(N/s)\right)$, where N is the number of columns of the matrix (also the cardinality of set of left vertices of the expander graph) or the ambient dimension of the signals that can be sensed by such matrices. This gives insights into why using such sensing matrices with small d performed well in numerical compressed sensing experiments. Furthermore, we derive quantitative sampling theorems for our constructions which show our construction outperforming the existing state-of-the-art. We also used our results to compare performance of sparse recovery algorithms where these matrices are used for linear sketching.

Keywords: compressed sensing (CS), expander graphs, probabilistic construction, sample complexity, sampling theorems, combinatorial compressed sensing, linear sketching, sparse recovery algorithms

1. INTRODUCTION

Sparse binary matrices, say $\mathbf{A} \in \{0, 1\}^{n \times N}$, with $n \ll N$ are widely used in applications including graph sketching [2, 3], network tomography [4, 5], data streaming [6, 7], breaking privacy of databases via aggregate queries [8], compressed imaging of intensity patterns [9], and more generally combinatorial compressed sensing [10–15], linear sketching [7], and group testing [16, 17]. In all these areas we are interested in the case where $n \ll N$, in which case $\mathbf{A}$ is used as an efficient encoder of sparse signals $\mathbf{x} \in \mathbb{R}^N$ with sparsity $s \ll n$, where they are known to preserve ℓ^1 distance of sparse vectors [18]. Conditions that guarantee that a given encoder, $\mathbf{A}$, also referred to as a sensing matrix in compressed sensing, typically include the nullspace, coherence, and the restricted isometry conditions, see [19] and references there in. The goal is for $\mathbf{A}$ to satisfy one or more of these conditions with the minimum possible n, the number of measurements. For uniform guarantees over all $\mathbf{A}$, it has been established that n has to be $\Omega\left(s^2\right)$, but that with high probability on the draw of random $\mathbf{A}$, n can be $\mathcal{O}\left(s \log N/n\right)$ for $\mathbf{A}$ with entries drawn from a sub-gaussian distribution, see [19] for a review of such results. Matrices with entries drawn from a Bernoulli distribution fall in the family of sub-gaussian but these are *dense* as opposed the the *sparse* binary matrices considered here. For computational advantages, such as faster application and smaller storage, it is advantageous to use sparse $\mathbf{A}$ in application [1, 14, 18].

Herein we consider the n achievable when $\mathbf{A}$ is an adjacency matrix of an expander graph [18], expander graph will be defined in the next section. Since the construction of such matrices can be construed as either a linear algebra problem or equivalently a graph theory one (in this manuscript we will focus more on the linear algebra discourse). There has been significant research on expander

graphs in pure mathematics and theoretical computer science, see [20] and references therein. Both deterministic and probabilistic constructions of expander graphs have been suggested. The best known deterministic constructions achieve $n = \mathcal{O}\left(s^{1+\alpha}\right)$ for some $\alpha > 0$ [21]. On the other hand random constructions, first proven in Bassalygo and Pinsker [22], achieve the optimal $n = \mathcal{O}\left(s \log\left(N/s\right)\right)$, precisely $n = \mathcal{O}(sd)$, with $d = \mathcal{O}\left(\log\left(N/s\right)\right)$, where d is the *left degree* of the expander graph but also the number of ones in each column of $\mathbf{A}$, to be defined in the next section. However, to the best of our knowledge, it was [1] that proposed a probabilistic construction that is not only optimal but also more suitable to making quantitative statements where such matrices are applied.

This work follows the probabilistic construction proposed in Bah and Tanner[1] but with careful computation of the bounds, is able to achieve $n = \mathcal{O}\left(s \log\left(N/s\right)\right)$ with $d = \mathcal{O}\left(\frac{\log(N/s)}{\log s}\right)$. We retain the complexity of n but got a smaller complexity for d, which is novel. Related results with a similar d were derived in Indyk and Razenshteyn [23] and Bah et al.[24] but for structure sparse signals in the framework of model-based compressing sensing or sketching. In that framework, one has second order information about $\mathbf{x}$ beyond simple sparsity, which is first order information about $\mathbf{x}$. It is thus expected and established that it is possible to get a small n and hence a smaller d. Arguably, such a small complexity for d justifies in hindsight fixing d to a small number in simulations with such $\mathbf{A}$ as in Bah and Tanner [1], Bah et al. [24], and Mendoza-Smith and Tanner [14], just to mention a few.

The results derived here are asymptotic, though finite dimensional bounds follow directly. We focus on for what ratios of the problem dimensions (s, n, N) does these results hold. There is almost a standard way of interrogating such a question, i.e., phase transitions, probably introduced to the compressed sensing literature by Donoho and Tanner [25]. In other words, we derive sampling theorems numerically depicted by phase transition plots about problem size spaces for which our construction holds. This is similar to what was done in Bah and Tanner [1] but for comparison purposes we include phase transition plots from probabilistic constructions by Buhrman et al. [26] and Berinde [27]. The plots show improvement over these earlier works. Furthermore, we show implications of our results for compressed sensing by using our results with the phase transition framework to compare the performance of selected combinatorial compressed sensing algorithms as is done in Bah and Tanner [1] and Mendoza-Smith and Tanner [14].

The manuscript is organized as follows. Section 1 gives the introduction; while section 2 sets the notation and defines some useful terms. The main results are stated in section 3 and the details of the construction is given in section 4. This is followed by a discussion in section 5 about our results, comparing them to existing results and using the results to compare the performance of some combinatorial compressed sensing algorithms. In section 6 we state the remaining proofs of theorems, lemmas, corollaries, and propositions used in this manuscript. After this section is the conclusion in section 7. We include an Appendix in **Supplementary Materials**, where we summarized key relevant

materials from Bah and Tanner [1], and showed the derivation of some bounds used in the proofs.

2. PRELIMINARIES

2.1. Notation

Scalars will be denoted by lowercase letters (e.g., k), vectors by lowercase boldface letters (e.g., $\mathbf{x}$), sets by uppercase calligraphic letters (e.g., $\mathcal{S}$) and matrices by uppercase boldface letters (e.g., $\mathbf{A}$). The cardinality of a set $\mathcal{S}$ is denoted by $|\mathcal{S}|$ and $[N] := \{1, \ldots, N\}$. Given $\mathcal{S} \subseteq [N]$, its complement is denoted by $\mathcal{S}^c := [N] \setminus \mathcal{S}$ and $\mathbf{x}_{\mathcal{S}}$ is the restriction of $\mathbf{x} \in \mathbb{R}^N$ to $\mathcal{S}$, i.e., $(\mathbf{x}_{\mathcal{S}})_i = x_i$ if $i \in \mathcal{S}$ and 0 otherwise. For a matrix $\mathbf{A}$, the restriction of $\mathbf{A}$ to the columns indexed by $\mathcal{S}$ is denoted by $\mathbf{A}_{\mathcal{S}}$. For a graph, $\Gamma(\mathcal{S})$ denotes the set of *neighbors* of $\mathcal{S}$, that is the nodes that are connected to the nodes in $\mathcal{S}$, and $e_{ij} = (x_i, y_j)$ represents an edge connecting node x_i to node y_j. The ℓ_p norm of a vector $\mathbf{x} \in \mathbb{R}^N$ is defined as $\|\mathbf{x}\|_p := \left(\sum_{i=1}^N x_i^p\right)^{1/p}$.

2.2. Definitions

Below we give formal definitions that will be used in this manuscript.

Definition 2.1 (ℓ_p-norm restricted isometry property). A matrix $\mathbf{A}$ satisfies the ℓ_p-norm restricted isometry property (RIP-p) of order s and constant $\delta_s < 1$ if it satisfies the following inequality.

$$(1 - \delta_s) \|\mathbf{x}\|_p^p \leq \|\mathbf{A}\mathbf{x}\|_p^p \leq (1 + \delta_s) \|\mathbf{x}\|_p^p, \quad \forall\, s\text{–sparse } \mathbf{x}. \quad (1)$$

The most popular case is RIP-2 and was first proposed in Cand‘es et al. [28]. Typically when RIP is mentioned without qualification, it means RIP_2. In the discourse of this work though RIP-1 is the most relevant. The RIP says that $\mathbf{A}$ is a near-isometry and it is a sufficient condition to guarantee exact sparse recovery in the noiseless setting (i.e., $\mathbf{y} = \mathbf{A}\mathbf{x}$); or recovery up to some error bound, also referred to as *optimality condition*, in the noisy setting (i.e., $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}$, where $\mathbf{e}$ is the bounded noise vector). we define optimality condition more precisely below.

Definition 2.2 (Optimality condition). Given $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}$ and $\widehat{\mathbf{x}} = \Delta\left(\mathbf{A}\mathbf{x} + \mathbf{e}\right)$ for a reconstruction algorithm Δ, the optimal error guarantee is

$$\|\widehat{\mathbf{x}} - \mathbf{x}\|_p \leq C_1 \sigma_s(\mathbf{x})_q + C_2 \|\mathbf{x}\|_p, \quad (2)$$

where $C_1, C_2 > 0$ depend only on the RIP constant (RIC), i.e., δ_s, and not the problem size, $1 \leq q \leq p \leq 2$, and $\sigma_s(\mathbf{x})_q$ denote the error of the best s-term approximation in the ℓ_q-norm, that is

$$\sigma_s(\mathbf{x})_q := \min_{s-\text{sparse } \mathbf{z}} \|\mathbf{z} - \mathbf{x}\|_q. \quad (3)$$

Equation (2) is also referred to as the ℓ_p/ℓ_q optimality condition (or error guarantee). Ideally, we would like ℓ_2/ℓ_2, but the best provable is ℓ_2/ℓ_1 [28], weaker than this is the ℓ_1/ℓ_1 [18], which is what is possible with the $\mathbf{A}$ considered in this work.

To aid translation between the terminology of graph theory and linear algebra we define the *set of neighbors* in both notation.

Definition 2.3 (Definition 1.4 in Bah and Tanner [1]). Consider a bipartite graph $G = ([N], [n], \mathcal{E})$ where $\mathcal{E}$ is the set of edges and $e_{ij} = (x_i, y_j)$ is the edge that connects vertex x_i to vertex y_j. For a given set of left vertices $\mathcal{S} \subset [N]$ its set of neighbors is $\Gamma(\mathcal{S}) = \{y_j | x_i \in \mathcal{S}$ and $e_{ij} \in \mathcal{E}\}$. In terms of the adjacency matrix, $\mathbf{A}$, of $G = ([N], [n], \mathcal{E})$ the set of neighbors of $\mathbf{A}_{\mathcal{S}}$ for $|\mathcal{S}| = s$, denoted by A_s, is the set of rows with at least one non-zero.

Definition 2.4 (Expander graph). Let $G = ([N], [n], \mathcal{E})$ be a left-regular bipartite graph with N left vertexes, n right vertexes, a set of edges $\mathcal{E}$ and left degree d. If, for any $\epsilon \in (0, 1/2)$ and any $\mathcal{S} \subset [N]$ of size $|\mathcal{S}| \leq s$, we have that $|\Gamma(\mathcal{S})| \geq (1 - \epsilon)d|\mathcal{S}|$, then G is referred to as a (s, d, ϵ)-expander graph.

The ϵ is referred to as the expansion coefficient of the graph. A (s, d, ϵ)-expander graph, also called an *unbalanced* expander graph [18] or a *lossless* expander graph [29], is a highly connected bipartite graph. We denote the ensemble of $n \times N$ binary matrices with d ones per column by $\mathcal{B}(N, n; d)$, or just $\mathcal{B}$ to simplify notation. We also will denote the ensemble of $n \times N$ adjacency matrices of (s, d, ϵ)-expander graphs as $\mathbb{E}(N, n; s, d, \epsilon)$ or simply $\mathbb{E}$.

3. RESULTS

Before stating the main results of the paper, we summarise here how we improve upon existing results in Bah and Tanner [1], which is generally true of standard random construction of such matrices. Essentially in the random constructions we want one minus the probability in (5) to go to zero as the problem size $(s, n, N) \to \infty$. In Theorem 4.1 (which was Theorem 1.6 in Bah and Tanner [1]) for this to happen we require $d = \mathcal{O}\left(\log(N/s)\right)$ based on the bound (7). In this work we derived a tighter bound than (7), which is (12), that requires d to be smaller than the standard bound by a factor of $\log(s)$.

The main result of this work is formalized in Theorem 3.1, which is an asymptotic result, where the dimensions grow while their ratios remain bounded. This is also referred to as the *proportional growth asymptotics* [30, 31]. We point out that it has been observed in practice that applying these matrices (including the simulations we do in this manuscript), which necessitate having finite dimensions, produces results that agree with the theory. However, the theoretical analysis used in this work cannot be applied to the finite setting.

Theorem 3.1. *Consider $\epsilon \in (0, \frac{1}{2})$ and let $d, s, n, N \in \mathbb{N}$, a random draw of an $n \times N$ matrix $\mathbf{A}$ from $\mathcal{B}$, i.e., for each column of $\mathbf{A}$ uniformly assign ones in d out of n positions, as $(s, n, N) \to \infty$ while $s/n \in (0, 1)$ and $n/N \in (0, 1)$, with probability approaching 1 exponentially, the matrix $\mathbf{A} \in \mathbb{E}$ with*

$$d = \mathcal{O}\left(\frac{\log(N/s)}{\epsilon \log s}\right), \quad and \quad n = \mathcal{O}\left(\frac{s \log(N/s)}{\epsilon^2}\right). \quad (4)$$

The proof of this theorem is found in section 6.1. It is worth emphasizing that the complexity of d is novel and it is the main contribution of this work.

Furthermore, in the *proportional growth asymptotics*, i.e., as $(s, n, N) \to \infty$ while $s/n \to \rho$ and $n/N \to \delta$ with $\rho, \delta \in (0, 1)$, for completeness, we derived a *phase transition* function (curve) in $\delta\rho$-space below which Theorem 3.1 holds, but above which Theorem 3.1 fails to hold. This is formalized in the following lemma.

Lemma 3.1. *Fix $\epsilon \in (0, \frac{1}{2})$ and let $d, s, n, N \in \mathbb{N}$, as $(s, n, N) \to \infty$ while $s/n \to \rho \in (0, 1)$ and $n/N \to \delta \in (0, 1)$ then for $\rho < (1 - \gamma)\rho_{BT}(\delta; d, \epsilon)$ and $\gamma > 0$, a random draw of $\mathbf{A}$ from $\mathcal{B}$ implies $\mathbf{A} \in \mathbb{E}$ with probability approaching 1 exponentially.*

The proof of this lemma is given in section 6.2. The phase transition function $\rho_{BT}(\delta; d, \epsilon)$ turned out to be significantly higher that those derived from existing probabilistic constructions, hence our results are significant improvement over earlier works. This will be graphically demonstrated with some numerical simulations in section 5.

4. CONSTRUCTION

The standard probabilistic construction is for each column of $\mathbf{A}$ to uniformly assign ones in d out of n positions; while the standard approach to derive the probability bounds is to randomly selected s columns of $\mathbf{A}$ indexed by $\mathcal{S}$ and compute the probability that $|A_s| < (1 - \epsilon)ds$, then do a *union bound* over all sets $\mathcal{S}$ of size s. Our work in Bah and Tanner [1] computed smaller bounds than previous works based on a *dyadic splitting* of $\mathcal{S}$ and derived the following bound. We changed the notation and format of Theorem 1.6 in Bah and Tanner [1] slightly to be consistent with the notation and format in this manuscript.

Theorem 4.1 (Theorem 1.6, Bah and Tanner [1]). *Consider $d, s, n, N \in \mathbb{N}$, fix $\mathcal{S}$ with $|\mathcal{S}| \leq s$, let an $n \times N$ matrix $\mathbf{A}$ be drawn from $\mathcal{B}$, then*

$$Prob\left(|A_s| \leq a_s\right) < p_n(s, d) \cdot e^{[n \cdot \Psi_n(a_s, \ldots, a_1)]} \quad (5)$$

with $a_1 := d$, and the functions defined as

$$p_n(s, d) = \frac{2}{25\sqrt{2\pi s^3 d^3}}, \quad and \quad (6)$$

$$\Psi_n(a_s, \ldots, a_1) = \frac{1}{n}\left[3s \log(5d) + \sum_{i \in \Omega} \frac{s}{2i} \psi_i\right], \quad \Omega = \{2^j\}_{j=0}^{\log_2(s)-1}, \quad (7)$$

where ψ_i is given by the following expression

$$(n - a_i) \cdot \mathcal{H}\left(\frac{a_{2i} - a_i}{n - a_i}\right) + a_i \cdot \mathcal{H}\left(\frac{a_{2i} - a_i}{a_i}\right) - n \cdot \mathcal{H}\left(\frac{a_i}{n}\right) \quad (8)$$

and $\mathcal{H}(\cdot)$ is the Shannon entropy in base e logarithms, and the index set $\Omega = \{2^j\}_{j=0}^{\log_2(s)-1}$.

a) *If no restriction is imposed on a_s then the a_i for $i > 1$ take on their expected value $\hat{a}_i$ given by*

$$\hat{a}_{2i} = \hat{a}_i\left(2 - \frac{\hat{a}_i}{n}\right), \quad for\ i \in \{2^j\}_{j=0}^{\log_2(s)-1}. \quad (9)$$

b) If a_s is restricted to be less than $\hat{a}_s$, then the a_i for $i > 1$ are the unique solutions to the following polynomial system

$$a_{2i}^3 - 2a_i a_{2i}^2 + 2a_i^2 a_{2i} - a_i^2 a_{4i} = 0, \quad \text{for } i \in \{2^j\}_{j=0}^{\log_2(s)-2}, \quad (10)$$

with $a_{2i} \geq a_i$ for each i.

In this work, based on the same approach as in Bah and Tanner [1], we derive new expressions for the $p_n(s,d)$ and $\Psi_n(a_s,\ldots,a_1)$ in Theorem 4.1, i.e., (6) and (7) respectively, and provide a simpler bound for the improved expression of $\Psi_n(a_s,\ldots,a_1)$.

Lemma 4.1. *Theorem 4.1 holds with the functions*

$$p_n(s,d) = 2^{-3} s^{9/2} e^{1/4}, \quad (11)$$

$$\Psi_n(a_s,\ldots,a_1) = \frac{1}{n}\left[\frac{3\log 2}{2}\log_2^2 s + \left(\log_2 s - \frac{3}{2}\right)\log a_s \right.$$

$$\left. + \sum_{i \in \Omega}\frac{s}{2i}\psi_i\right], \quad \text{for } \Omega = \{2^j\}_{j=0}^{\log_2(s)-1}. \quad (12)$$

The proof of the lemma is given in section 6.3. Asymptotically, the argument of the exponential term in the bound of the probability in (5) of Theorem 4.1, i.e., $\Psi_n(a_s,\ldots,a_1)$ in (12) is more important than the polynomial $p_n(s,d)$ in (11) since the exponential factor will dominate the polynomial factor. The significance of the lemma is that $\Psi_n(a_s,\ldots,a_1)$ in (12) is smaller than $\Psi_n(a_s,\ldots,a_1)$ in (7) since $\frac{3\log 2}{2}\log_2^2 s + \left(\log_2 s - \frac{3}{2}\right)\log a_s$ in (12) is asymptotically smaller than $3s\log(5d)$ in (7), because the former is $\mathcal{O}(\text{polylogs})$ while the latter is $\mathcal{O}(s)$, since we consider $a_s = \mathcal{O}(s)$.

Recall that we are interested in computing $\text{Prob}\left(|A_s| \leq a_s\right)$ when $a_s = (1-\epsilon)ds$. This means having to solve the polynomial equation (10) to compute as small a bound of $\Psi_n\left((1-\epsilon)ds,\ldots,d\right)$ as possible. We derive an asymptotic solution to (10) for $a_s = (1-\epsilon)ds$ and use that solution to get the following bounds.

Theorem 4.2. *Consider $d,s,n,N \in \mathbb{N}$, fix $\mathcal{S}$ with $|\mathcal{S}| \leq s$, for $\eta > 0$, $\beta \geq 1$, and $\epsilon \in \left(0,\frac{1}{2}\right)$, let an $n \times N$ matrix $\mathbf{A}$ be drawn from $\mathcal{B}$, then*

$$\text{Prob}\left(|A_s| \leq (1-\epsilon)ds\right) < p_n(s,d,\epsilon) \cdot \exp\left[n \cdot \Psi_n\left(s,d,\epsilon\right)\right] \quad (13)$$

where

$$p_n(s,d,\epsilon) = \frac{\sqrt[4]{e} \cdot s^{\log_2(1-\epsilon)+3}}{\sqrt{2^6(1-\epsilon)^3 d^3}}, \quad (14)$$

$$\Psi_n\left(s,d,\epsilon\right) \leq -\frac{1}{2n}\left[\eta\beta^{-1}(\beta-1)(1-\epsilon)ds\log_2(s/2)\right.$$

$$\left. -(5\log 2)\log_2^2 s - 2\log d \log_2 s\right]. \quad (15)$$

The proof of this theorem is also found in section 6.5. Since Theorem 4.2 holds for a fixed $\mathcal{S}$ of size at most s, if we want this to hold for all $\mathcal{S}$ of size at most s, we do a union bound over all $\mathcal{S}$ of size at most s. This leads to the following probability bound.

Theorem 4.3. *Consider $d,s,n,N \in \mathbb{N}$, and all $\mathcal{S}$ with $|\mathcal{S}| \leq s$, for $\tau > 0$, and $\epsilon \in \left(0,\frac{1}{2}\right)$, let an $n \times N$ matrix $\mathbf{A}$ be drawn from $\mathcal{B}$, then*

$$\text{Prob}\left(|A_s| \leq (1-\epsilon)ds\right) < p_N(s,d,\epsilon) \cdot e^{[N \cdot \Psi_N(s,d,\epsilon)]}, \quad (16)$$

where

$$p_N(s,d,\epsilon) = \frac{5 \cdot \sqrt[4]{e} \cdot s^{\log_2(1-\epsilon)+5/2}}{\sqrt{2^{10}(1-\epsilon)^3 d^3 \pi (1-s/N)}}, \quad (17)$$

$$\Psi_N\left(s,d,\epsilon\right) \leq -\frac{s}{N}\log\left(\frac{s}{N}\right) + \frac{s}{N}$$

$$-\frac{\tau(1-\epsilon)d}{2\log 2}\frac{s}{N}\log\left(\frac{s}{2}\right) + o(N). \quad (18)$$

Proof: Applying the union bound over all $\mathcal{S}$ of size at most s to (13) leads to the following.

$$\text{Prob}\left(|A_s| \leq (1-\epsilon)ds\right) < \binom{N}{s} p_n(s,d,\epsilon) e^{[n \cdot \Psi_n(s,d,\epsilon)]}. \quad (19)$$

Then we used the upper bound of (143) to bound the combinatorial term $\binom{N}{s}$ in (19). After some algebraic manipulations, we separated the the polynomial term, given in (17), from the exponential terms whose exponent is

$$\Psi_N\left(s,d,\epsilon\right) := \mathcal{H}\left(\frac{s}{N}\right) + \frac{n}{N}\Psi_n\left(s,d,\epsilon\right). \quad (20)$$

We upper bound $\Psi_N\left(s,d,\epsilon\right)$ in (18) by upper bounding $\mathcal{H}\left(\frac{s}{N}\right)$ with $-\frac{s}{N}\log\left(\frac{s}{N}\right) + \frac{s}{N}$ and the upper bound of $\Psi_n\left(s,d,\epsilon\right)$ in (15). The $o(N)$ decays to zeros with N and its a result of dividing the polylogarithmic terms of s in (15), and $\tau = \eta\beta^{-1}(\beta-1)$ in (15). This concludes the proof. $\square$

The next corollary easily follows from Theorem 4.3 and it is equivalent to Theorem 3.1. Its statement is that if the conditions therein holds, then the probability that the cardinality of the set of neighbors of any $\mathcal{S}$ with $|\mathcal{S}| \leq s$ is less than $(1-\epsilon)ds$ goes to zero as dimensions of $\mathbf{A}$ grows. On the other hand, the probability that the cardinality of the set of neighbors of any $\mathcal{S}$ with $|\mathcal{S}| \leq s$ is greater than $(1-\epsilon)ds$ goes to one as dimensions of $\mathbf{A}$ grows. Implying that $\mathbf{A}$ is the adjacency matrix of a (s,d,ϵ)-expander graph.

Corollary 4.1. *Given $d,s,n,N \in \mathbb{N}$, and $\epsilon \in \left(0,\frac{1}{2}\right)$, for $d \geq \frac{c_d \log(N/s)}{\epsilon \log s}$ and $n \geq \frac{c_n s \log(N/s)}{\epsilon^2}$, with $c_d, c_n > 0$. Let an $n \times N$ matrix $\mathbf{A}$ be drawn from $\mathcal{B}$, in the proportional growth asymptotics*

$$\text{Prob}\left(|A_s| \leq (1-\epsilon)ds\right) \to 0. \quad (21)$$

Proof: It suffice to focus on the exponent of (16), more precisely on the bound of $\Psi_N\left(s,d,\epsilon\right)$ in (18), i.e.,

$$-\frac{s}{N}\log\left(\frac{s}{N}\right) + \frac{s}{N} - \frac{\tau(1-\epsilon)d}{2\log 2}\frac{s}{N}\log\left(\frac{s}{2}\right) + o(N). \quad (22)$$

We can ignore the $o(N)$ term as this goes to zero as N grows, and show that the remaining sum is negative. The remaining sum is

$$-\frac{s}{N}\log\left(\frac{s}{N}\right) + \frac{s}{N} - \frac{\tau(1-\epsilon)d}{2\log 2}\frac{s}{N}\log\left(\frac{s}{2}\right)$$
$$= \frac{s}{N}\left[-\log\left(\frac{s}{N}\right) + 1 - \frac{\tau(1-\epsilon)d}{2\log 2}\log\left(\frac{s}{2}\right)\right]. \quad (23)$$

Hence, we can further focus on the sum in the square brackets, and find conditions on d that will make it negative. We require

$$-\log\left(\frac{s}{N}\right) + 1 - \frac{\tau(1-\epsilon)d}{2\log 2}\log\left(\frac{s}{2}\right) < 0,$$

$$\Rightarrow \quad d > \frac{2\log 2\left(\log\left(N/s\right)+1\right)}{\tau(1-\epsilon)\log(s/2)} \quad (24)$$

$$\Leftrightarrow \quad d \geq \frac{c_d\log(N/s)}{\epsilon\log s}, \quad \exists\, c_d > 0. \quad (25)$$

Recall $\tau = \eta\beta^{-1}(\beta-1)$ and β is a function of ϵ, with $\beta(\epsilon) \approx 1+\epsilon$. Therefore, τ is a function of ϵ, and $\tau(\epsilon) = \mathcal{O}(\epsilon)$, hence there exists a $c_d > 0$ for (25) to hold.

With regards to the complexity of n, we go back to the right hand side (RHS) of (24) and we substitute $\frac{C_d\log(N/s)}{\epsilon\log(s/2)}$ with $C_d > 0$ for d in the RHS of (24) to get the following.

$$-\log\left(\frac{s}{N}\right) + 1 - \frac{C_d\tau(1-\epsilon)\log\left(N/s\right)}{2\epsilon\log 2} < 0. \quad (26)$$

Now we assume $n = \frac{C_n s\log(N/s)}{\epsilon^2}$ with $C_n > 0$ for n and substitute this in (26) to get the following.

$$-\log\left(\frac{s}{N}\right) + 1 - \frac{C_d\tau(1-\epsilon)\epsilon n}{2C_n s\log 2} < 0,$$

$$\Rightarrow \quad n > \frac{2C_n s\log 2\left(\log\left(N/s\right)+1\right)}{C_d\tau\epsilon(1-\epsilon)} \quad (27)$$

$$\Leftrightarrow \quad n \geq \frac{c_n s\log\left(N/s\right)}{\epsilon^2}, \quad \exists\, c_n > 0. \quad (28)$$

Again since $\tau(\epsilon) = \mathcal{O}(\epsilon)$, hence there exists a $c_n > 0$ for (28) to hold. The bound of n in (28) agrees with our earlier assumption, thus concluding the proof. □

5. DISCUSSION

5.1. Comparison to Other Constructions

In addition to being the first probabilistic construction of adjacency matrices of expander graphs with such a small degree, quantitatively our results compares favorably to existing probabilistic constructions. We use the standard tool of phase transitions to compare our construction to the construction proposed in Berinde [27] and those proposed in Buhrman et al. [26]. The phase transition curve $\rho_{BT}(\delta; d, \epsilon)$ we derived in Lemma 3.1 is the ρ that solves the following equation.

$$-\rho\log\left(\delta\rho\right) + \rho - \frac{\tau(1-\epsilon)d\rho\log(\delta\rho)}{2\log 2} - \frac{\tau\epsilon(1-\epsilon)d}{2c_n\log 2}$$

$$+\frac{\tau(1-\epsilon)d\rho}{2} = 0, \quad (29)$$

where $c_n > 0$ is as in (28). Equation (29) comes from taking the limit, in the proportional growth asymptotics, of the bound in (18), setting that to zero and simplifying. Similarly, for any $\mathcal{S}$ with $|\mathcal{S}| \leq s$, Berinde [27] derived the following bound on the set of neigbors of $\mathcal{S}$, i.e., A_s.

$$\text{Prob}\left(|A_s| \leq (1-\epsilon)ds\right) < \binom{N}{s}\binom{ds}{\epsilon ds}\left(\frac{ds}{n}\right)^{\epsilon ds}. \quad (30)$$

We then express the bound in (30) as the product of a polynomial term and an exponential term as we did for the right hand side of (16). A bound of the exponent is carefully derived as in the derivations above. We set the limit, in the proportional growth asymptotics, of this bound to zero and simplify to get the following.

$$(\epsilon d - 1)\log\rho - \log\delta + (1+\epsilon d) - \epsilon d\log\left(\epsilon/d\right) = 0. \quad (31)$$

We refer to the ρ that solves (31) as the phase transition for the construction proposed by Berinde [27] and denote this ρ (the phase transition function) as $\rho_{BI}(\delta; d, \epsilon)$. Another probabilistic construction was proposed by Burhman et al. [26]. In conformity with the notation used in this manuscript their bound is equivalent to the following, also stated in a similar form by Indyk and Razenshteyn [23].

$$\text{Prob}\left(|A_s| \leq (1-\epsilon)ds\right) < \binom{N}{s}\left(\frac{\nu\epsilon n}{ds}\right)^{-\epsilon ds}, \quad (32)$$

where $\nu > 0$. We again express the bound in (32) as the product of a polynomial term and an exponential term. A bound of the exponent is carefully derived as in the derivations above. We set the limit, in the proportional growth asymptotics, of this bound to zero and simplify to get the following.

$$(\epsilon d - 1)\log\rho - \log\delta + 1 - \epsilon d\log\left(\nu\epsilon/d\right) = 0. \quad (33)$$

Similarly, we refer to the ρ that solves (33) as the phase transition for the construction proposed by Burhman et al. [26] and denote this ρ as $\rho_{BM}(\delta; d, \epsilon)$. We compute numerical solutions to (29), (31), and (33) to derive the phase transitions, for the existence (probabilistic construction) of expanders, $\rho_{BT}(\delta; d, \epsilon)$, $\rho_{BI}(\delta; d, \epsilon)$, and $\rho_{BM}(\delta; d, \epsilon)$ respectively. These are plotted in the *left panel* of **Figure 1**. It is clear that our construction has a much higher phase transition than the others, as expected from our previous results, see Figure 9 in Bah and Tanner [1]. Recall that the phase transition curves in these plots depict construction of adjacency matrices of (s, d, ϵ)-expanders with high probability for ratios of s, n, and N (since $\rho := s/n$, and $\delta := n/N$) below the curve; and the failure to construct adjacency matrices of (s, d, ϵ)-expanders with high probability for ratios of s, n, and N above the curve. Essentially, the larger the area under the curve the better.

Remark 5.1. It is easy to see that $\rho_{BI}(\delta; d, \epsilon)$ is a special case of $\rho_{BM}(\delta; d, \epsilon)$ since the two phase transitions will coincide, or
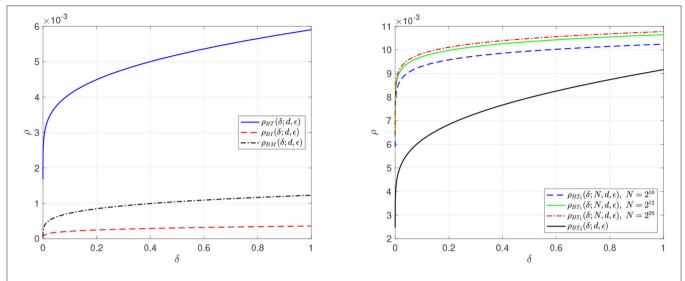
FIGURE 1 | Plots of phase transitions for probabilistic construction of expanders, with fixed $d = 2^5$, $\epsilon = 1/6$, and $\delta \in \left[10^{-6}, 1\right]$ on a logarithmically spaced grid of 100 points. **(Left)** A comparison of $\rho_{BT}(\delta; d, \epsilon)$, $\rho_{BI}(\delta; d, \epsilon)$, and $\rho_{BM}(\delta; d, \epsilon)$, where $c_n = 2$. **(Right)** A comparison of $\rho_{BT}(\delta; d, \epsilon)$ denoted as $\rho_{BT_2}(\delta; d, \epsilon)$ to our previous $\rho_{BT}(\delta; d, \epsilon)$ denoted as $\rho_{BT_1}(\delta; d, \epsilon)$ in Bah and Tanner [1] with different values of N, (i.e., 2^{10}, 2^{12}, and 2^{20}) and $c_n = 2/3$.

equivalently (31) and (33) will be the same, when $\nu = e^{-1}$. One could argue that Berinde's derivation in [27] suffers from over counting.

Given that this work is an improvement of our work in Bah and Tanner [1] in terms of simplicity in computing $\rho_{BT}(\delta; d, \epsilon)$, for completeness we compare our new phase transition $\rho_{BT}(\delta; d, \epsilon)$ denoted as $\rho_{BT_2}(\delta; d, \epsilon)$ to our previous $\rho_{BT}(\delta; d, \epsilon)$ denoted as $\rho_{BT_1}(\delta; d, \epsilon)$ in the *right panel* of **Figure 1**. Each computation of $\rho_{BT_1}(\delta; d, \epsilon)$ requires the specification of N, which is not needed in the computation of $\rho_{BT_2}(\delta; d, \epsilon)$, hence the simplification. However, the simplification led to a lower phase transition as expected, which is confirmed by the plots in the *right panel* of **Figure 1**.

Remark 5.2. *These simulations also inform us about the size of c_n. See from the plots of $\rho_{BT}(\delta; d, \epsilon)$ and $\rho_{BT_2}(\delta; d, \epsilon)$ that the smaller the value of c_n the higher the phase transition but since $\rho_{BT_2}(\delta; d, \epsilon)$ has to be a lower bound of $\rho_{BT_1}(\delta; d, \epsilon)$, for values of c_n much smaller than 2/3, the lower bound will fail to hold. This informed the choice of $c_n = 2$ in the plot of $\rho_{BT}(\delta; d, \epsilon)$ in the left panel of* **Figure 1**.

5.2. Implications for Combinatorial Compressed Sensing

When the sensing matrices are restricted to the sparse binary matrices considered in this manuscript, compressed sensing is usually referred to as *combinatorial compressed sensing* a term introduced in Berinde et al. [18] and used extensively in Mendoza-Smith and Tanner [14] and Mendoza-Smith et al. [15]. In this setting, compressed sensing is more-or-less equivalent to linear sketching. The implications of our results on combinatorial compressed sensing are two-fold. One is on the ℓ_1-norm RIP, we donate as RIP-1; while the second is in the comparison

of performance of recovery algorithms for combinatorial compressed sensing.

5.2.1. RIP-1
As can be seen from (2), the recovery errors in compressed sensing depend on the RIC, i.e., δ_s. The following lemma deduced from Theorem 1 of Berinde et al.[18] shows that a scaled **A** drawn from $\mathbb{E}$ have RIP with $\delta_s = 2\epsilon$.

Lemma 5.1. *Consider $\epsilon \in (0, \frac{1}{2})$ and let **A** be drawn from $\mathbb{E}$, then $\mathbf{\Phi} = \mathbf{A}/d$ satisfies the following RIP-1 condition*

$$(1 - 2\epsilon) \|\mathbf{x}\|_1 \leq \|\mathbf{\Phi}\mathbf{x}\|_1 \leq \|\mathbf{x}\|_1, \quad \forall \, s\text{–sparse } \mathbf{x}. \quad (34)$$

The interested reader is referred to the proof of Theorem 1 in Berinde et al. [18] for the proof of this lemma. Key to the holding of Lemma 5.1 is the existence of (s, d, ϵ)-expander graphs, hence one can draw corollaries from our results on this.

Corollary 5.1. *Consider $\epsilon \in (0, \frac{1}{2})$ and let $d, s, n, N \in \mathbb{N}$. In the proportional growth asymptotics with a random draw of an $n \times N$ matrix **A** from $\mathcal{B}$, the matrix $\mathbf{\Phi} := \mathbf{A}/d$ has RIP-1 with probability approaching 1 exponentially, if*

$$d = \mathcal{O}\left(\frac{\log(N/s)}{\epsilon \log s}\right), \quad and \quad n = \mathcal{O}\left(\frac{s \log(N/s)}{\epsilon^2}\right). \quad (35)$$

Proof: Note that the upper bound of (34) holds trivially for any $\mathbf{\Phi} = \mathbf{A}/d$ where **A** has d ones per column, i.e., $\mathbf{A} \in \mathcal{B}$. But for the lower bound of (34) to hold for any $\mathbf{\Phi} = \mathbf{A}/d$, we need **A** to be an (s, d, ϵ)-expander matrix, i.e., $\mathbf{A} \in \mathbb{E}$. Note that the event $|A_s| \geq (1 - \epsilon)ds$ is equal to the event $\|A_{\mathcal{S}}\mathbf{x}\|_1 \geq (1 - 2\epsilon)d\|\mathbf{x}\|_1$, which is equivalent to $\|\mathbf{\Phi}_{\mathcal{S}}\mathbf{x}\|_1 \geq (1 - 2\epsilon)\|\mathbf{x}\|_1$, for a fixed $\mathcal{S}$, with

$|\mathcal{S}| \leq s$. For $\mathbf{A}$ to be in $\mathbb{E}$, we need expansion for all sets $\mathcal{S}$, with $|\mathcal{S}| \leq s$, i.e., $\mathbf{A} \in \mathbb{E}$. The key thing to remember is that

$$
\begin{aligned}
&\text{Prob}\left(\|\mathbf{\Phi x}\|_1 \geq (1 - 2\epsilon)\|\mathbf{x}\|_1\right) \\
&\quad = \text{Prob}\left(|A_s| \geq (1 - \epsilon)ds\right), \quad \text{for all } \{\mathcal{S} : |\mathcal{S}| \leq s\}. \quad (36)
\end{aligned}
$$

The probability in (36) going to 1 exponentially in the proportional growth asymptotics, i.e., the existence of $\mathbf{A} \in \mathbb{E}$ with parameters as given in (35), is what is stated in Theorem 3.1. Therefore, the rest of the proof follows from the proof of Theorem 3.1, hence concluding the proof of the corollary. □

Notably, Lemma 5.1 holds with $\mathbf{\Phi}$ having much smaller number of non-zeros per column due to our construction. More over, we can derive sampling theorems for which Lemma 5.1 holds as thus.

Corollary 5.2. *Fix* $\epsilon \in (0, \frac{1}{2})$ *and let* $d, s, n, N \in \mathbb{N}$. *In the proportional growth asymptotics, for any* $\rho < (1 - \gamma)\rho_{BT}(\delta; d, \epsilon)$ *and* $\gamma > 0$, *a random draw of* $\mathbf{A}$ *from* $\mathcal{B}$ *implies* $\mathbf{\Phi} := \mathbf{A}/d$ *has RIP-1 with probability approaching* 1 *exponentially.*

Proof: The proof of this corollary follows from the proof of Corollary 5.1 above, and it is related to the proof of Lemma 3.1 as the proof of Corollary 5.1 is to the proof of Theorem 3.1. The details of the proof is thus skipped. □

5.2.2. Performance of Algorithms

We wish to compare the performance of selected combinatorial compressed sensing algorithms in terms of the possible problem sizes (s, n, N) that these algorithms can reconstruct sparse/compressible signals/vectors up to their respective error guarantees. The comparison is typically done in the framework of phase transitions, which depict a boundary curve where ratios of problems sizes above this curve are recovered with probability approaching 0 exponentially; while problems sizes below the curve are recovered with probability approaching 1 exponentially. The list of combinatorial compressed sensing algorithms includes Expander Matching Pursuit (EMP) [32], Sparse Matching Pursuit [33], Sequential Sparse Matching Pursuit (SSMP) [13], Left Degree Dependent Signal Recovery (LDDSR) [11], Expander Recovery (ER) [12], Expander Iterative Hard-Thresholding (EIHT) [19, Section 13.4], and Expander ℓ_0-decoding (ELD) with both serial and parallel versions [14]. For reason similar to those used in Bah and Tanner [1] and Mendoza-Smith and Tanner [14], we selected out of this list four of the algorithms: (i) SSMP, (ii) ER, (ii) EIHT, (iv) ELD. Descriptions of these algorithms is skipped here but the interested reader is referred to the original papers or their summarized details in Bah and Tanner [1] and Mendoza-Smith and Tanner [14]. We were also curious as to how ℓ_1-minimization's performance compares to these selected combinatorial compressed sensing algorithms, since ℓ_1-minimization (ℓ_1-min) can be used to solve the combinatorial problem solved by these algorithms, see [18, Theorem 3].

The phase transitions are based on conditions on the RIC of the sensing matrices used. Consequent to Lemma 5.1, this becomes conditions on the expansion coefficient (i.e., ϵ) of the underlying (s, d, ϵ)-expander graphs of the sparse sensing

matrices used. Where this condition on ϵ is not explicitly given it is easily deducible from the recovery guarantees given for each algorithms. The conditions are summarized in the table below.

The *theoretical values* are what will be found in the reference given in the table; while the *computational values* are what we used in our numerical experiments to compute the phase transition curves of the algorithms. The value for e was set to be 10^{-15}, to make the ϵ_k as large as possible under the given condition. With these values we computed the phase transitions in **Figure 2**.

The two figures are the same except for the different sparsity value used. The performance of the algorithms in this framework are thus ranked as follows: ELD, ER, ℓ_1-min, EIHT, and SSMP.

Remark 5.3. *We point out that there are many way to compare performance of algorithms, this is just one way. For instance, we can compare runtime complexities or actual computational runtimes as in Mendoza-Smith and Tanner [14]; phase transitions of different probabilities, here the probability of recovery is 1 but this could be set to something else, like 1/2 in the simulations in Mendoza-Smith and Tanner [14]; one could also compare number of iterations and iteration cost as was also done in Mendoza-Smith and Tanner [14]. We further point out that such comaprisons could be conducted with real-world data simulations, as was done in Mendoza-Smith and Tanner [14].*

6. PROOFS OF THEOREMS, LEMMAS, AND PROPOSITIONS

6.1. Theorem 3.1

The proof of the theorem follows trivially from Corollary 4.1. Based on (21) of Corollary 4.1 we deduce that $\mathbf{A} \in \mathbb{E}$ with probability approaching 1 exponentially with $d \geq \frac{c_d \log(N/s)}{\epsilon \log s}$ and $n \geq \frac{c_n s \log(N/s)}{\epsilon^2}$, with $c_d, c_n > 0$, hence concluding the proof. □

6.2. Lemma 3.1

The phase transition curve $\rho_{BT}(\delta; d, \epsilon)$ is based the bound of the exponent of (16), which is

$$
-\frac{s}{N}\log\left(\frac{s}{N}\right) + \frac{s}{N} - \frac{\tau(1 - \epsilon)d}{2\log 2}\frac{s}{N}\log\left(\frac{s}{2}\right) + o(N). \quad (37)
$$

In the propotional growth asymptotics $(s, n, N) \to \infty$ while $s/n \to \rho \in (0, 1)$ and $n/N \to \delta \in (0, 1)$. This implies that $o(N) \to 0$ and (37) becomes

$$
\begin{aligned}
&-\rho\log(\delta\rho) + \rho - \frac{\tau(1 - \epsilon)d\rho\log(\delta\rho)}{2\log 2} \\
&-\frac{\tau\epsilon(1 - \epsilon)d}{2c_n\log 2} + \frac{\tau(1 - \epsilon)d\rho}{2}, \quad (38)
\end{aligned}
$$

where $c_n > 0$ is as in (28). If (38) is negative then as the problem size grows we have

$$
\text{Prob}\left(|A_s| \leq (1 - \epsilon)ds\right) \to 0. \quad (39)
$$

Therefore, setting (38) to zero and solving for ρ gives us a critical ρ below which (38) is negative and positive above it.
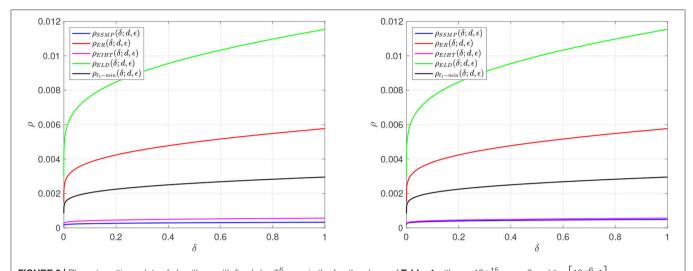
FIGURE 2 | Phase transitions plots of algorithms with fixed $d = 2^5$, ϵ as in the fourth column of **Table 1** with $e = 10^{-15}$, $c_n = 2$ and $\delta \in \left[10^{-6}, 1\right]$ on a logarithmically spaced grid of 100 points. **(Left)** $k = 3s$ for $\rho_{SSMP}(\delta; d, \epsilon)$. **(Right)** $k = \lceil(2 + e)s\rceil$ for $\rho_{SSMP}(\delta; d, \epsilon)$.

TABLE 1 | Recovery conditions on ϵ.

Algorithm	Theoretical values		Computational values	
	Condition	**Sparsity**	**Condition**	**Sparsity**
SSMP [13]	$\epsilon_k < 1/16$	$k = (c + 1)s,\ c > 1$	$\epsilon_k = 1/16 - e$	$k = \lceil(2 + e)s\rceil,\ k = 3s$
ER [12]	$\epsilon_k < 1/4$	$k = 2s$	$\epsilon_k = 1/4 - e$	$k = 2s$
EIHT [19]	$\epsilon_k < 1/12$	$k = 3s$	$\epsilon_k = 1/12 - e$	$k = 3s$
ELD [14]	$\epsilon_k \leq 1/4$	$k = s$	$\epsilon_k = 1/4$	$k = s$
ℓ_1-min [18]	$\epsilon_k < 1/6$	$k = 2s$	$\epsilon_k = 1/6 - e$	$k = 2s$

The critical ρ is the phase transition ρ, i.e., $\rho_{BT}(\delta; d, \epsilon)$, where below $\rho_{BT}(\delta; d, \epsilon)$ is parameterized by the γ in the lemma. This concludes the proof. □

6.3. Lemma 4.1

By the dyadic splitting proposed in Bah and Tanner [1], we let $A_s = A^1_{\lceil\frac{s}{2}\rceil} \cup A^2_{\lfloor\frac{s}{2}\rfloor}$ such that $|A_s| = \left|A^1_{\lceil\frac{s}{2}\rceil} \cup A^2_{\lfloor\frac{s}{2}\rfloor}\right|$ and therefore

$$\text{Prob}\left(|A_s| \leq a_s\right) = \text{Prob}\left(\left|A^1_{\lceil\frac{s}{2}\rceil} \cup A^2_{\lfloor\frac{s}{2}\rfloor}\right| \leq a_s\right) \quad (40)$$

$$= \sum_{l_s=d}^{a_s} \text{Prob}\left(\left|A^1_{\lceil\frac{s}{2}\rceil} \cup A^2_{\lfloor\frac{s}{2}\rfloor}\right| = l_s\right) \quad (41)$$

In (41) we sum over all possible events, i.e., all possible sizes of l_s. In line with the splitting technique, we simplify the probability to the product of the probabilities of the cardinalities of $\left|A^1_{\lceil\frac{s}{2}\rceil}\right|$ and $\left|A^2_{\lfloor\frac{s}{2}\rfloor}\right|$ and their intersection. Using the definition of $P_n(\cdot)$ in Lemma 8.2 (Appendix), thus leads to the following.

$$\text{Prob}\left(|A_s| \leq a_s\right)$$
$$= \sum_{l_s=2d}^{a_s} \sum_{l^1_{\lceil\frac{s}{2}\rceil}=d}^{a_{\lceil\frac{s}{2}\rceil}} \sum_{l^2_{\lfloor\frac{s}{2}\rfloor}=d}^{a_{\lfloor\frac{s}{2}\rfloor}} P_n\left(l_s, l^1_{\lceil\frac{s}{2}\rceil}, l^2_{\lfloor\frac{s}{2}\rfloor}\right)$$

$$\times \text{Prob}\left(\left|A^1_{\lceil\frac{s}{2}\rceil}\right| = l^1_{\lceil\frac{s}{2}\rceil}\right) \text{Prob}\left(\left|A^2_{\lfloor\frac{s}{2}\rfloor}\right| = l^2_{\lfloor\frac{s}{2}\rfloor}\right). \quad (42)$$

In a slight abuse of notation we write $\displaystyle\sum_{\substack{\tilde{l}j \\ j\in[m]}}$ to denote applying the sum m times. We also drop the limits of the summation indices henceforth. Now we use Lemma 8.1 in Appendix to simplify (42) as follows.

$$\sum_{l^{j_1}_{Q_0}} \sum_{l^{j_2}_{Q_1}} \sum_{l^{j_3}_{R_1}} P_n\left(l^{j_1}_{Q_0}, l^{2j_1-1}_{\lceil\frac{Q_0}{2}\rceil}, l^{2j_1}_{\lfloor\frac{Q_0}{2}\rfloor}\right) \times$$
$$\scriptstyle j_1\in[q_0]\, j_2\in[q_1]\, j_3\in[r_1]$$

$$\prod_{j_2=1}^{q_1} \text{Prob}\left(\left|A^{j_2}_{Q_1}\right| = l^{j_2}_{Q_1}\right) \prod_{j_3=q_1+1}^{q_1+r_1} \text{Prob}\left(\left|A^{j_3}_{R_1}\right| = l^{j_3}_{R_1}\right) \quad (43)$$

Now we proceed with the splitting—note (43) stopped only at the first level. At the next level, the second, we will have q_2 sets with Q_2 columns and r_2 sets with R_2 columns which leads to the following expression.

$$\sum_{l^{j_1}_{Q_0}} \sum_{l^{j_2}_{Q_1}} \sum_{l^{j_3}_{R_1}} P_n\left(l^{j_1}_{Q_0}, l^{2j_1-1}_{\lceil\frac{Q_0}{2}\rceil}, l^{2j_1}_{\lfloor\frac{Q_0}{2}\rfloor}\right)$$
$$\scriptstyle j_1\in[q_0]\, j_2\in[q_1]\, j_3\in[r_1]$$

$$\left[\sum_{\substack{l_{Q_2}^{j_4}\\ j_4\in[q_2]}}\sum_{\substack{l_{R_2}^{j_5}\\ j_5\in[r_2]}} P_n\left(l_{Q_1}^{j_2}, l_{\lceil\frac{Q_1}{2}\rceil}^{2j_2-1}, l_{\lfloor\frac{Q_1}{2}\rfloor}^{2j_2}\right) \times P_n\left(l_{R_1}^{j_3}, l_{\lceil\frac{R_1}{2}\rceil}^{2j_3-1}, l_{\lfloor\frac{R_1}{2}\rfloor}^{2j_3}\right)\right.$$

$$\left.\prod_{j_4=1}^{q_2}\text{Prob}\left(\left|A_{Q_2}^{j_4}\right|=l_{Q_2}^{j_4}\right)\prod_{j_5=q_2+1}^{q_2+r_2}\text{Prob}\left(\left|A_{R_2}^{j_5}\right|=l_{R_2}^{j_5}\right)\right]. \quad (44)$$

We continue this splitting of each instance of Prob(·) for $\lceil\log_2 s\rceil-1$ levels until reaching sets with single columns where, by construction, the probability that the single column has d non-zeros is one. Note that at this point we drop the subscripts j_i, as they are no longer needed. This process gives a complicated product of nested sums of $P_n(\cdot)$ which we express as

$$\sum_{l_{Q_0}}\sum_{l_{Q_1}}\sum_{l_{R_1}} P_n\left(l_{Q_0}, l_{\lceil\frac{Q_0}{2}\rceil}, l_{\lfloor\frac{Q_0}{2}\rfloor}\right)$$

$$\left[\sum_{l_{Q_2}}\sum_{l_{R_2}} P_n\left(l_{Q_1}, l_{\lceil\frac{Q_1}{2}\rceil}, l_{\lfloor\frac{Q_1}{2}\rfloor}\right)P_n\left(l_{R_1}, l_{\lceil\frac{R_1}{2}\rceil}, l_{\lfloor\frac{R_1}{2}\rfloor}\right)\right.$$

$$\left.\left[\times\cdots\left[\sum_{l_{Q_{\lceil\log_2 s\rceil-1}}} P_n\left(l_4, l_2, l_2\right)P_n\left(l_3, l_2, d\right)P_n\left(l_2, d, d\right)\right]\cdots\right]. \quad (45)$$

Using the expression for $P_n(\cdot)$ in (133) of Lemma 8.2 (Appendix) we bound (45) by bounding each $P_n(\cdot)$ as in (134) with a product of a polynomial, $\pi(\cdot)$, and an exponential with exponent $\psi_n(\cdot)$.

$$\sum_{l_{Q_0}}\sum_{l_{Q_1}}\sum_{l_{R_1}}\pi\left(l_{Q_0}, l_{\lceil\frac{Q_0}{2}\rceil}, l_{\lfloor\frac{Q_0}{2}\rfloor}\right) e^{\psi_n\left(l_{Q_0}, l_{\lceil\frac{Q_0}{2}\rceil}, l_{\lfloor\frac{Q_0}{2}\rfloor}\right)}\times$$

$$\left[\sum_{l_{Q_2}}\sum_{l_{R_2}}\pi\left(l_{Q_1}, l_{\lceil\frac{Q_1}{2}\rceil}, l_{\lfloor\frac{Q_1}{2}\rfloor}\right)\times\right.$$

$$e^{\psi_n\left(l_{Q_1}, l_{\lceil\frac{Q_1}{2}\rceil}, l_{\lfloor\frac{Q_1}{2}\rfloor}\right)}\pi\left(l_{R_1}, l_{\lceil\frac{R_1}{2}\rceil}, l_{\lfloor\frac{R_1}{2}\rfloor}\right)\cdot e^{\psi_n\left(l_{R_1}, l_{\lceil\frac{R_1}{2}\rceil}, l_{\lfloor\frac{R_1}{2}\rfloor}\right)}\times$$

$$\left[\cdots\times\left[\sum_{l_{Q_{\lceil\log_2 s\rceil-1}}}\pi\left(l_4, l_2, l_2\right)\times e^{\psi_n(l_4,l_2,l_2)}\pi\left(l_3, l_2, d\right)\times\right.\right.$$

$$\left.\left. e^{\psi_n(l_3,l_2,d)}\pi\left(l_2, d, d\right)\cdot e^{\psi_n(l_2,d,d)}\right]\cdots\right]. \quad (46)$$

Using Lemma 8.4 in Appendix we maximize the $\psi_n(\cdot)$ and hence the exponentials in (46). We maximize each $\psi_n(\cdot)$ by choosing $l_{(\cdot)}$ to be $a_{(\cdot)}$. Then (46) will be upper bounded by the following.

$$\sum_{l_{Q_0}}\sum_{l_{Q_1}}\sum_{l_{R_1}}\pi\left(l_{Q_0}, l_{\lceil\frac{Q_0}{2}\rceil}, l_{\lfloor\frac{Q_0}{2}\rfloor}\right) e^{\psi_n\left(a_{Q_0}, a_{\lceil\frac{Q_0}{2}\rceil}, a_{\lfloor\frac{Q_0}{2}\rfloor}\right)}\times$$

$$\left[\sum_{l_{Q_2}}\sum_{l_{R_2}}\pi\left(l_{Q_1}, l_{\lceil\frac{Q_1}{2}\rceil}, l_{\lfloor\frac{Q_1}{2}\rfloor}\right)\times e^{\psi_n\left(a_{Q_1}, a_{\lceil\frac{Q_1}{2}\rceil}, a_{\lfloor\frac{Q_1}{2}\rfloor}\right)}\right.$$

$$\left.\pi\left(l_{R_1}, l_{\lceil\frac{R_1}{2}\rceil}, l_{\lfloor\frac{R_1}{2}\rfloor}\right)\cdot e^{\psi_n\left(a_{R_1}, a_{\lceil\frac{R_1}{2}\rceil}, a_{\lfloor\frac{R_1}{2}\rfloor}\right)}\left[\cdots\times\right.\right.$$

$$\left[\sum_{l_{Q_{\lceil\log_2 s\rceil-1}}}\pi\left(l_4, l_2, l_2\right)\times e^{\psi_n(a_4,a_2,a_2)}\pi\left(l_3, l_2, d\right) e^{\psi_n(a_3,a_2,d)}\right.$$

$$\left.\pi\left(l_2, d, d\right)\cdot e^{\psi_n(a_2,d,d)}\right]\cdots\right]. \quad (47)$$

We then factor the product of exponentials. This product becomes an exponential where exponent is the summation of the $\psi_n(\cdot)$, we will denote this exponent as $\widetilde{\Psi}_n\left(a_s,\ldots,a_2,d\right)$. Then (47) simplifies to the following.

$$e^{\widetilde{\Psi}_n(a_s,\ldots,a_2,d)}\cdot\sum_{l_{Q_0}}\sum_{l_{Q_1}}\sum_{l_{R_1}}\pi\left(l_{Q_0}, l_{\lceil\frac{Q_0}{2}\rceil}, l_{\lfloor\frac{Q_0}{2}\rfloor}\right)\times$$

$$\left[\sum_{l_{Q_2}}\sum_{l_{R_2}}\pi\left(l_{Q_1}, l_{\lceil\frac{Q_1}{2}\rceil}, l_{\lfloor\frac{Q_1}{2}\rfloor}\right)\times\pi\left(l_{R_1}, l_{\lceil\frac{R_1}{2}\rceil}, l_{\lfloor\frac{R_1}{2}\rfloor}\right)\times\right[$$

$$\left.\cdots\times\left[\sum_{l_{Q_{\lceil\log_2 s\rceil-1}}}\pi\left(l_4, l_2, l_2\right)\pi\left(l_3, l_2, d\right)\pi\left(l_2, d, d\right)\right]\cdots\right] \quad (48)$$

We denote the factor multiplying the exponential term by $\Pi\left(l_s,\ldots,l_2,d\right)$, therefore we have the following bound.

$$\text{Prob}\left(|A_s|\leq a_s\right)\leq\Pi\left(l_s,\ldots,l_2,d\right)\cdot e^{\left(\widetilde{\Psi}_n(a_s,\ldots,a_2,d)\right)}, \quad (49)$$

where $\widetilde{\Psi}_n\left(a_s,\ldots,a_2,d\right)$ is exactly $\Psi_n\left(a_s,\ldots,a_2,d\right)$ given by (78) in [1, Proof of Theorem 1.6]. Consequently, we state the bound of $\widetilde{\Psi}_n\left(a_s,\ldots,a_2,d\right)$ and skip the proof, which is as thus.

$$\widetilde{\Psi}_n\left(a_s,\ldots,a_2,d\right)\leq\sum_{i\in\Omega}\frac{s}{2i}\psi_i,\quad\text{for }\Omega=\{2^j\}_{j=0}^{\log_2(s)-1}, \quad (50)$$

where ψ_i is given by (8). The upper bound of $\Pi\left(l_s,\ldots,l_2,d\right)$ is given by the following proposition.

Proposition 6.1. *Given* $l_s\leq a_s$, $l_{\lceil s/2\rceil}\leq a_{\lceil s/2\rceil},\cdots, l_2\leq a_2$, *we have*

$$\Pi\left(l_s,\ldots,l_2,d\right)\leq 2^{-3}\cdot e^{\frac{1}{4}}\cdot s^{\frac{3}{2}\log_2 s+\frac{9}{2}}\cdot(a_s)^{\log_2 s-\frac{3}{2}}. \quad (51)$$

The proof of the proposition is found in section 6.4. Taking log of right hand side of (51) and then exponentiating the results yields

$$\Pi\left(l_s,\ldots,l_2,d\right)$$

$$\leq\exp\left[\frac{1}{4}-3\log 2+\left(\frac{9}{2}+\frac{3}{2}\log_2 s\right)\log s\right.$$

$$\left.+\left(\log_2 s-\frac{3}{2}\right)\log a_s\right] \quad (52)$$

$$=2^{-3}s^{9/2}e^{1/4}\cdot e^{\left[\frac{3}{2}\log_2 s\log s+\left(\log_2 s-\frac{3}{2}\right)\log a_s\right]} \quad (53)$$

$$=2^{-3}s^{9/2}e^{1/4}\cdot e^{\left[\frac{3\log 2}{2}\log_2^2 s+\left(\log_2 s-\frac{3}{2}\right)\log a_s\right]}. \quad (54)$$

Combining (54) and (50) gives the following bound for (49)

$$
\begin{aligned}
\text{Prob}\left(|A_s| \le a_s\right) \le\ & 2^{-3} s^{9/2} e^{1/4} \cdot \exp\left[\frac{\log 8}{2} \log_2^2 s\right. \\
& \left. + \left(\log_2 s - \frac{3}{2}\right) \log a_s + \widetilde{\Psi}_n\left(a_s, \ldots, d\right)\right]
\end{aligned}
\tag{55}
$$

It follows therefore that $p_n(s,d) = 2^{-3} s^{9/2} e^{1/4}$ as in (11) and the exponent in (55) is $n \cdot \Psi_n\left(a_s, \ldots, a_1\right)$, which implies (12). This concludes the proof of the lemma. □

6.4. Proposition 6.1

By definition, from (48), we have

$$
\begin{aligned}
\Pi\left(l_s, \ldots, l_2, d\right) := & \sum_{l_{Q_0}} \sum_{l_{Q_1}} \sum_{l_{R_1}} \pi\left(l_{Q_0}, l_{\lceil \frac{Q_0}{2} \rceil}, l_{\lfloor \frac{Q_0}{2} \rfloor}\right) \times \\
& \left[\sum_{l_{Q_2}} \sum_{l_{R_2}} \pi\left(l_{Q_1}, l_{\lceil \frac{Q_1}{2} \rceil}, l_{\lfloor \frac{Q_1}{2} \rfloor}\right) \times \pi\left(l_{R_1}, l_{\lceil \frac{R_1}{2} \rceil}, l_{\lfloor \frac{R_1}{2} \rfloor}\right) \cdot \left[\cdots\right.\right. \\
& \left.\left. \times \left[\sum_{l_{Q_{\lceil \log_2 s \rceil - 1}}} \pi\left(l_4, l_2, l_2\right) \pi\left(l_3, l_2, d\right) \pi\left(l_2, d, d\right)\right] \cdots \right]\right].
\end{aligned}
\tag{56}
$$

From (138) we see that $\pi(\cdot)$ is maximized when all the three arguments are the same and using Corollary 8.1 we take largest possible arguments that are equal in the range of the summation. Before we write out the resulting bound for $\Pi\left(l_s, \ldots, l_2, d\right)$, we simplify notation by denoting $\pi(x,x,x)$ as $\pi(x)$, and noting that $Q_{\lceil \log_2 s \rceil - 1} = 2$. Therefore, the bound becomes the following.

$$
\begin{aligned}
\Pi\left(l_s, \ldots, l_2, d\right) \le & \sum_{l_{Q_0}} \sum_{l_{Q_1}} \sum_{l_{R_1}} \pi\left(l_{Q_0}\right) \left[\sum_{l_{Q_2}} \sum_{l_{R_2}} \pi\left(l_{Q_1}\right)\right. \\
& \left. \times \pi\left(l_{R_1}\right) \left[\cdots \left[\sum_{l_2} \pi\left(l_4\right) \pi\left(l_3\right) \pi\left(l_2\right)\right] \cdots \right]\right].
\end{aligned}
\tag{57}
$$

Properly aligning the $\pi(\cdot)$ with their relevant summations simplifies the right hand side (RHS) of (57) to the following.

$$
\begin{aligned}
& \left[\sum_{l_{Q_0}} \pi\left(l_{Q_0}\right)\right] \cdot \left[\sum_{l_{Q_1}} \pi\left(l_{Q_1}\right)\right] \cdot \left[\sum_{l_{R_1}} \pi\left(l_{R_1}\right)\right] \times \cdots \\
& \times \left[\sum_{l_4} \pi\left(l_4\right)\right] \cdot \left[\sum_{l_3} \pi\left(l_3\right)\right] \cdot \left[\sum_{l_2} \pi\left(l_2\right)\right].
\end{aligned}
\tag{58}
$$

From (138) we have $\pi\left(y\right)$ defined as

$$
\pi(y,y,y) = \left(\frac{5}{4}\right)^2 \sqrt{\frac{2\pi y(n-y)}{n}} < \left(\frac{5}{4}\right)^2 \sqrt{2\pi y}.
\tag{59}
$$

We use the RHS of (59) to upper bound each term in (58), leading to the following bound.

$$
\left[\sqrt{2\pi}\left(\frac{5}{4}\right)^2 \sum_{l_{Q_0}} \sqrt{l_{Q_0}}\right] \cdot \left[\sqrt{2\pi}\left(\frac{5}{4}\right)^2 \sum_{l_{Q_1}} \sqrt{l_{Q_1}}\right].
$$

$$
\left[\sqrt{2\pi}\left(\frac{5}{4}\right)^2 \sum_{l_{R_1}} \sqrt{l_{R_1}}\right] \times \cdots \times \left[\sqrt{2\pi}\left(\frac{5}{4}\right)^2 \sum_{l_4} \sqrt{l_4}\right] \times
$$

$$
\left[\sqrt{2\pi}\left(\frac{5}{4}\right)^2 \sum_{l_3} \sqrt{l_3}\right] \cdot \left[\sqrt{2\pi}\left(\frac{5}{4}\right)^2 \sum_{l_2} \sqrt{l_2}\right].
\tag{60}
$$

For each Q_i and R_i, $i = 1, \ldots, \lceil \log_2 s \rceil - 2$, which means we have $\lceil \log_2 s \rceil - 2$ pairs plus one Q_0, hence (60) simplifies to the following.

$$
\begin{aligned}
& \left[\sqrt{2\pi}\left(\frac{5}{4}\right)^2\right]^{2\lceil \log_2 s \rceil - 3} \left[\sum_{l_{Q_0}} \sqrt{l_{Q_0}}\right] \cdot \left[\sum_{l_{Q_1}} \sqrt{l_{Q_1}}\right] \times \cdots \\
& \times \left[\sum_{l_3} \sqrt{l_3}\right] \cdot \left[\sum_{l_2} \sqrt{l_2}\right]
\end{aligned}
\tag{61}
$$

$$
\le \left[\sqrt{2\pi}\left(\frac{5}{4}\right)^2\right]^{2\lceil \log_2 s \rceil - 3} \cdot \left(q_0 \sqrt{a_{Q_0}}\right) \cdot \left(q_1 \sqrt{a_{Q_1}}\right) \cdots
$$

$$
\times \left(q_{\lceil \log_2 s \rceil - 2} \sqrt{a_3}\right) \cdot \left(r_{\lceil \log_2 s \rceil - 2} \sqrt{a_2}\right)
\tag{62}
$$

$$
= \left[\sqrt{2\pi}\left(\frac{5}{4}\right)^2\right]^{2\lceil \log_2 s \rceil - 3} \cdot \left(q_0 q_1 r_1 \cdots q_{\lceil \log_2 s \rceil - 2} r_{\lceil \log_2 s \rceil - 2}\right) \cdot
$$

$$
\left(a_{Q_0} a_{Q_1} a_{R_1} \cdots a_3 a_2\right)^{1/2}.
\tag{63}
$$

From (61) to (62) we upper each sum by taking the largest possible value of $l_{(\cdot)}$, which is $a_{(\cdot)}$, and multiplied it with the total number terms in the summation given by Lemma 8.1 in Appendix . We did upper bound the following two terms of (63).

$$
q_0 q_1 r_1 q_2 r_2 q_3 r_3 \cdots q_{\lceil \log_2 s \rceil - 2} r_{\lceil \log_2 s \rceil - 2} \le s^{\log_2 s - 1},
\tag{64}
$$

$$
\left(a_{Q_0} a_{Q_1} a_{R_1} a_{Q_2} a_{R_2} a_{Q_3} a_{R_3} \cdots a_3 a_2\right)^{1/2} \le 2^{-1} \cdot e^{\frac{1}{4}} \cdot \left(a_s\right)^{\log_2 s - \frac{3}{2}} \times s^{\frac{1}{2} \log_2 s + \frac{3}{2}}
\tag{65}
$$

Details of the derivation of the bounds (64) and (65) is in the Appendix . Using these bounds from (63) we have the following upper bound for $\Pi\left(l_s, \ldots, l_2, d\right)$.

$$
\Pi\left(l_s, \ldots, l_2, d\right) \le \left[\sqrt{2\pi}\left(\frac{5}{4}\right)^2\right]^{2\lceil \log_2 s \rceil - 3} \cdot \left(s^{\log_2 s - 1}\right) \times
$$

$$
\left(2^{-1} \cdot e^{\frac{1}{4}} \cdot \left(a_s\right)^{\log_2 s - \frac{3}{2}} \cdot s^{\frac{1}{2} \log_2 s + \frac{3}{2}}\right)
\tag{66}
$$

$$
\le 2^{-3} \cdot e^{\frac{1}{4}} \cdot \left(a_s\right)^{\log_2 s - \frac{3}{2}} \cdot s^{\frac{3}{2} \log_2 s + \frac{9}{2}}.
\tag{67}
$$

From (66) to (67) we used the following upper bound.

$$
\left[\sqrt{2\pi}\left(\frac{5}{4}\right)^2\right]^{2\lceil \log_2 s \rceil - 3} \le \left[\sqrt{2\pi}\left(\frac{5}{4}\right)^2\right]^{2(\log_2 s + 1) - 3}
$$

$$
\le 4^{2\log_2 s - 1} = 2^{4\log_2 s - 2} = 2^{-2} s^4.
\tag{68}
$$

The bound (67) coincides with (51), hence concluding the proof.

□

6.5. Theorem 4.2

The following lemma is a key input in this proof.

Lemma 6.1. *Let* $0 < \alpha \le 1$, *and* $\varepsilon_n > 0$ *such that* $\varepsilon_n \to 0$ *as* $n \to \infty$. *Then for* $a_s < \hat{a}_s$,

$$a_{2i} = 2a_i + ca_i^2, \quad for \ c = -\beta n^{-1}, \tag{69}$$

where

$$\beta = \frac{1 + \sqrt{1 - 4(1 - \varepsilon_n)^2 \left(1 - e^{-\alpha d}\right) e^{-\alpha d}}}{2(1 - \varepsilon_n)\left(1 - e^{-\alpha d}\right)}. \tag{70}$$

The proof of the lemma is found in section 6.6. Recall from Theorem 4.1 that

$$\Psi_n(a_s, \ldots, a_1) = \frac{1}{n}\left[\frac{3\log 2}{2}\log_2^2 s + \left(\log_2 s - \frac{3}{2}\right)\log a_s + \sum_{i \in \Omega} \frac{s}{2i}\psi_i\right], \quad for \ \Omega = \{2^j\}_{j=0}^{\log_2(s)-1}, \tag{71}$$

where ψ_i is given by the following expression

$$(n - a_i)\cdot\mathcal{H}\left(\frac{a_{2i} - a_i}{n - a_i}\right) + a_i\cdot\mathcal{H}\left(\frac{a_{2i} - a_i}{a_i}\right) - n\cdot\mathcal{H}\left(\frac{a_i}{n}\right) \tag{72}$$

We use Lemma 6.1 to upper bound ψ_i in (72) away from zero from above as $n \to 0$. We formalize this bound in the following proposition.

Proposition 6.2. *Let* $\eta > 0$ *and* $\beta > 1$ *as defined in Lemma 6.1. Then*

$$\psi_i \le -a_i\eta(\beta - 1)\beta^{-1}\left(1 - \beta\frac{a_i}{n}\right)^{-1}. \tag{73}$$

The proof of Proposition 6.2 is found in section 6.7. Using the bound of ψ_i in Proposition 6.2, we upper $\Psi(a_s, \ldots, a_1)$ as follows.

$$\Psi(a_s, \ldots, a_1) \le \frac{1}{n}\left[\frac{3\log 2}{2}\log_2^2 s + \left(\log_2 s - \frac{3}{2}\right)\log a_s - \sum_{i \in \Omega}\frac{s}{2i}\cdot\frac{a_i\eta(\beta - 1)}{\beta\left(1 - \beta\frac{a_i}{n}\right)}\right], \tag{74}$$

$$\le -\frac{\eta(\beta - 1)}{\beta}\sum_{i \in \Omega}\left(\frac{s}{2i}\cdot\frac{a_i}{n}\right) + \frac{1}{n}\left[\frac{3\log 2}{2}\log_2^2 s + \left(\log_2 s - \frac{3}{2}\right)\log a_s\right]. \tag{75}$$

Then setting $a_s = (1 - \epsilon)ds$ and substituting in (75), the factor multiplying $\frac{1}{n}$ becomes

$$\frac{3\log 2}{2}\log_2^2 s + \left(\log_2 s - \frac{3}{2}\right)\log\left[(1 - \epsilon)ds\right] \tag{76}$$

$$= \frac{3\log 2}{2}\log_2^2 s + \log(1 - \epsilon)\log_2 s + \log d\log_2 s +$$

$$\log_2 s\log s - \frac{3}{2}\log\left[(1 - \epsilon)d\right] - \frac{3}{2}\log s \tag{77}$$

$$= \frac{5\log 2}{2}\log_2^2 s + \left(\log_2(1 - \epsilon) - 3/2\right)\log s + \log d\log_2 s + \log\left[(1 - \epsilon)^{-3/2}d^{-3/2}\right] \tag{78}$$

$$= \frac{5\log 2}{2}\log_2^2 s + \log d\log_2 s + \log\left[(1 - \epsilon)^{-3/2}d^{-3/2}\right] + \log s^{\log_2(1 - \epsilon) - 3/2}. \tag{79}$$

The last two terms of (79) become polynomial in s, d and ϵ, when exponentiated hence they are incorporated into $p_n(s, d, \epsilon)$ in (14), which means

$$p_n(s, d, \epsilon) = p_n(s, d)\cdot\exp\left[\log\left[(1 - \epsilon)^{-3/2}d^{-3/2}\right] + \log s^{\log_2(1 - \epsilon) - 3/2}\right] \tag{80}$$

$$= 2^{-3}s^{9/2}e^{1/4}\cdot(1 - \epsilon)^{-3/2}d^{-3/2}s^{\log_2(1 - \epsilon) - 3/2} \tag{81}$$

$$= \frac{\sqrt[4]{e}\cdot s^{\log_2(1 - \epsilon) + 3}}{\sqrt{2^6(1 - \epsilon)^3 d^3}}, \tag{82}$$

which is (14). The first two terms of (79) will grow faster than a polynomial in s, d and ϵ when exponentiated, hence they replace in (75), the factor multiplying $\frac{1}{n}$. Therefore, (79) is modified as thus

$$-\frac{\eta(\beta - 1)}{\beta}\sum_{i \in \Omega}\left(\frac{s}{2i}\cdot\frac{a_i}{n}\right) + \frac{1}{n}\left[\frac{5\log 2}{2}\log_2^2 s + \log d\log_2 s\right]$$

$$=: \Psi_n(s, d, \epsilon). \tag{83}$$

The factor $\sum_{i \in \Omega}\left(\frac{s}{2i}\cdot\frac{a_i}{n}\right)$ in (83) is lower bounded as follows, see proof in section 6.8.

$$\sum_{i \in \Omega}\left(\frac{s}{2i}\cdot\frac{a_i}{n}\right) \ge \frac{\log_2(s/2)}{2n}(1 - \epsilon)ds. \tag{84}$$

Using this bound in (83) gives (15), thus concluding the proof. $\square$

6.6. Lemma 6.1

Recall that we have a formula for the expected values of the a_i as

$$\hat{a}_{2i} = \hat{a}_i\left(2 - \frac{\hat{a}_i}{n}\right) \quad for \ i \in \{2^j\}_{j=0}^{\log_2(s)-1}, \tag{85}$$

which follow a relatively simple formulas, and then the coupled system of cubics as

$$a_{2i}^3 - 2a_ia_{2i}^2 + 2a_i^2a_{2i} - a_i^2a_{4i} = 0 \quad for \ i \in \{2^j\}_{j=0}^{\log_2(s)-2}, \tag{86}$$

for when the final a_s is constrained to be less than $\hat{a}_s$. To simplify the notation of the indexing in (86), observe that if $i = 2^j$ for a fixed j, then $2i = 2^{j+1}$ and $4i = 2^{j+2}$. Therefore, it suffice to use the index a_j, a_{j+1}, and a_{j+2} rather than a_i, a_{2i}, and a_{4i}. Moving the second two terms in (86) to the right and dividing the quadratic multiples we get the relation

$$\frac{a_{j+2} - 2a_{j+1}}{a_{j+1}^2} = \frac{a_{j+1} - 2a_j}{a_j^2}, \tag{87}$$

which is the same expression on the right and left, but with j increased by one on the left. This implies that the fraction is independent of j, so

$$\frac{a_{j+1} - 2a_j}{a_j^2} = c, \quad \Rightarrow \quad a_{j+1} = a_j(2 + ca_j), \quad (88)$$

for some constant c independent of j (though not necessarily of n). This is in fact the relation (85), if we set c to be equal to $-1/n$. One can then wonder what is the behavior of c if we fix the final a_s. Moreover, (88) is equivalent to

$$ca_{j+1} + 1 = (ca_j + 1)^2, \quad (89)$$

which inductively leads to

$$ca_l + 1 = (ca_0 + 1)^{2^l}, \quad l > 0, \quad (90)$$

so that one has a relation of the lth stage in terms of the first stage. Note this does not require the a_s to be fixed, (90) is how one simply computes all a_l for $l > 0$ once one has a_0 and c. The point is that c to match the a_s one has to select c appropriately. So the way we calculate c is by knowing a_0 and a_s, then solving (90) for $l = s$. Unfortunately there is not an easy way to solve for c in (90) so we need to do some asymptotic approximation. Let's assume that a_l is close to $\hat{a}_l$. So we do an asymptotic expansion in terms of the difference from $\hat{a}_l$.

To simplify things a bit lets insert $a_0 = d$ (since a_0 is a_1 in our standard notation) and then we insert what we know for $\hat{a}_l$. For $\hat{a}_l$ we have $c = -n^{-1}$, see (85). We then have from (90) that

$$a_l = c^{-1}(cd + 1)^{2^l} - c^{-1} \quad \text{and} \quad \hat{a}_l = -n(-d/n + 1)^{2^l} + n. \quad (91)$$

So if we write $a_l = (1 - \varepsilon_n)\hat{a}_l$ and consider the case of $\varepsilon_n \to 0$ as $n \to 0$. The point of this is that instead of working with $\hat{a}_l$ we can now work in terms of ε_n. Setting $a_l = (1 - \varepsilon_n)\hat{a}_l$ gives

$$c^{-1}(cd + 1)^{2^l} - c^{-1} = -n(1 - \varepsilon_n)\left[(-d/n + 1)^{2^l} - 1\right]. \quad (92)$$

We now solve for c as a function of ε_n and d. As ε_n goes to zero we should have c converging to $-n^{-1}$.

Let $\alpha n = 2^l$, for $0 < \alpha \leq 1$ and $c = -\beta(\varepsilon_n, d)/n$, then, dropping the argument of $\beta(\cdot, \cdot)$, (92) becomes

$$\frac{n}{\beta}\left(1 - \frac{\beta d}{n}\right)^{\alpha n} - \frac{n}{\beta} = n(1 - \varepsilon_n)\left(1 - \frac{d}{n}\right)^{\alpha n} - n(1 - \epsilon). \quad (93)$$

Multiplying through by β/n and performing a change of variables of $k = \alpha n$, (93) becomes

$$\left(1 - \frac{\alpha\beta d}{k}\right)^k - 1 = \beta(1 - \varepsilon_n)\left(1 - \frac{\alpha d}{k}\right)^k - \beta(1 - \varepsilon_n). \quad (94)$$

The left hand side of (94) simplifies to

$$e^{-\alpha\beta d} - 1 - \frac{e^{-\alpha\beta d}\alpha^2\beta^2 d^2}{2k} + \mathcal{O}(k^{-2}). \quad (95)$$

The right hand side of (94) simplifies to

$$\beta(1 - \varepsilon_n)e^{-\alpha d} - \beta(1 - \varepsilon_n) - \frac{\beta(1 - \varepsilon_n)e^{-\alpha d}\alpha^2 d^2}{2k} + \mathcal{O}(k^{-2}). \quad (96)$$

Matching powers of k in (95) and (96) for k^0 and k^{-1} yields the following.

$$e^{-\alpha\beta d} - 1 = \beta(1 - \varepsilon_n)e^{-\alpha d} - \beta(1 - \varepsilon_n), \quad \text{and} \quad (97)$$

$$\alpha^2\beta^2 d^2 e^{-\alpha\beta d} = \beta(1 - \varepsilon_n)\alpha^2 d^2 e^{-\alpha d}. \quad (98)$$

Both of which respectively simplify to the following.

$$e^{-\alpha\beta d} - 1 = \beta(1 - \varepsilon_n)\left(e^{-\alpha d} - 1\right), \quad \text{and} \quad (99)$$

$$\beta e^{-\alpha\beta d} = (1 - \varepsilon_n)e^{-\alpha d}. \quad (100)$$

Multiply (99) by β and subtract the two equations, (99) and (100), to get

$$(1 - \varepsilon_n)\left(1 - e^{-\alpha d}\right)\beta^2 - \beta + (1 - \varepsilon_n)e^{-\alpha d} = 0. \quad (101)$$

This yields

$$\beta = \frac{1 \pm \sqrt{1 - 4(1 - \varepsilon_n)^2\left(1 - e^{-\alpha d}\right)e^{-\alpha d}}}{2(1 - \varepsilon_n)\left(1 - e^{-\alpha d}\right)}. \quad (102)$$

To be consistent with what c ought to be as $\varepsilon_n \to 0$, we choose

$$\beta(\varepsilon_n, d) = \frac{1 + \sqrt{1 - 4(1 - \varepsilon_n)^2\left(1 - e^{-\alpha d}\right)e^{-\alpha d}}}{2(1 - \varepsilon_n)\left(1 - e^{-\alpha d}\right)}, \quad (103)$$

as required – concluding the proof. $\square$

6.7. Proposition 6.2

We use Lemma 6.1 to express ψ_i in (72) as follows

$$\psi_i = -n\left[\mathcal{H}\left(\frac{a_i}{n}\right) - \mathcal{H}\left(\frac{a_i + ca_i^2}{n - a_i}\right)\right]$$
$$+ a_i\left[\mathcal{H}\left(\frac{a_i + ca_i^2}{a_i}\right) - \mathcal{H}\left(\frac{a_i + ca_i^2}{n - a_i}\right)\right] \quad (104)$$

$$= -n\left[\mathcal{H}\left(\frac{a_i}{n}\right) - \mathcal{H}\left(\frac{a_i}{n} \cdot \frac{1 + ca_i}{1 - \frac{a_i}{n}}\right)\right]$$
$$+ a_i\left[\mathcal{H}(1 + ca_i) - \mathcal{H}\left(\frac{a_i}{n} \cdot \frac{1 + ca_i}{1 - \frac{a_i}{n}}\right)\right]. \quad (105)$$

Note that for regimes of small s/n considered

$$-\frac{\beta}{n} = c \leq -\frac{1}{n}, \quad \Rightarrow \quad ca_i \leq -\frac{a_i}{n}, \quad \text{and} \quad 1 + ca_i \leq 1 - \frac{a_i}{n}. \quad (106)$$

We need the following expressions for the Shannon entropy and it's first and second derivatives

$$\mathcal{H}(z) = -z\log z - (1 - z)\log(1 - z), \quad (107)$$

$$\mathcal{H}'(z) = \log\left(\frac{1-z}{z}\right), \quad \text{and} \tag{108}$$

$$\mathcal{H}''(z) = -\frac{1}{z(1-z)}. \tag{109}$$

But also $\mathcal{H}(z) = \mathcal{H}(1-z)$ due to the symmetry about $z = 1/2$. Similarly, $\mathcal{H}''(z)$ is symmetric about $z = 1/2$; while $\mathcal{H}'(z)$ is anti-symmetric, i.e., $\mathcal{H}'(z) = -\mathcal{H}'(1-z)$. Using the symmetry of $\mathcal{H}(z)$ we rewrite ψ_i in (105) as follows.

$$\psi_i = -n\left[\mathcal{H}\left(\frac{a_i}{n}\right) - \mathcal{H}\left(\frac{a_i}{n} \cdot \frac{1+ca_i}{1-\frac{a_i}{n}}\right)\right]$$
$$+ a_i\left[\mathcal{H}(-ca_i) - \mathcal{H}\left(\frac{a_i}{n} \cdot \frac{1+ca_i}{1-\frac{a_i}{n}}\right)\right]. \tag{110}$$

From (106), we deduce the following ordering

$$\frac{a_i}{n} \cdot \frac{1+ca_i}{1-\frac{a_i}{n}} \leq \frac{a_i}{n} \leq -ca_i \leq 1/2. \tag{111}$$

The last equality in (111) follows from that we assume $2\beta a_i \leq n$, where $1 < \beta < 2$ but very close to 1. This is a valid constraint on the cardinality of the supports a_i. To simplify notation, let $x_1 = \frac{a_i}{n} \cdot \frac{1+ca_i}{1-\frac{a_i}{n}}$, $x_2 = \frac{a_i}{n}$, and $x_3 = -ca_i$, which implies that $x_1 \leq x_2 \leq x_3 \leq 1/2$. Therefore, from (110), we have

$$\psi_i = -n\left[\mathcal{H}(x_2) - \mathcal{H}(x_1)\right] + a_i\left[\mathcal{H}(x_3) - \mathcal{H}(x_1)\right] \tag{112}$$
$$= -n\left[\mathcal{H}(x_2) - \mathcal{H}(x_1)\right] + a_i\left[\mathcal{H}(x_3) - \mathcal{H}(x_2)\right.$$
$$\left. + \mathcal{H}(x_2) - \mathcal{H}(x_1)\right] \tag{113}$$
$$= -(n - a_i)\left[\mathcal{H}(x_2) - \mathcal{H}(x_1)\right] + a_i\left[\mathcal{H}(x_3) - \mathcal{H}(x_2)\right] \tag{114}$$
$$\leq -(n - a_i)(x_2 - x_1)\mathcal{H}'(x_2) + a_i(x_3 - x_2)\mathcal{H}'(x_2) \tag{115}$$
$$= \left[a_i(x_3 - x_2) - (n - a_i)(x_2 - x_1)\right]\mathcal{H}'(x_2). \tag{116}$$

Observe that the expression in the square brackets on the right hand side of (116) is zero, which implies that

$$a_i(x_3 - x_2) = (n - a_i)(x_2 - x_1). \tag{117}$$

This is very easy to check by substituting the values of x_1, x_2, and x_3. So instead of bound (115), we alternatively upper bound (114) as follows

$$\psi_i \leq -(n - a_i)(x_2 - x_1)\mathcal{H}'(\xi_{21}) + a_i(x_3 - x_2)\mathcal{H}'(\xi_{32}), \tag{118}$$

where $\xi_{21} \in (x_1, x_2)$, and $\xi_{32} \in (x_2, x_3)$, which implies

$$\xi_{21} < \xi_{32}, \quad \text{and} \quad \mathcal{H}'(\xi_{21}) > \mathcal{H}'(\xi_{32}). \tag{119}$$

Using relation (117), bound (118) simplifies to the following.

$$\psi_i \leq -a_i(x_3 - x_2)\mathcal{H}'(\xi_{21}) + a_i(x_3 - x_2)\mathcal{H}'(\xi_{32}) \tag{120}$$
$$= -a_i(x_3 - x_2)\left[\mathcal{H}'(\xi_{21}) - \mathcal{H}'(\xi_{32})\right] \tag{121}$$
$$\leq -a_i(x_3 - x_2)(\xi_{21} - \xi_{32})\mathcal{H}''(\xi_{31}), \tag{122}$$

for $\xi_{31} \in (x_1, x_3)$. Since $\xi_{21} < \xi_{32}$, we rewrite bound (122) as follows.

$$\psi_i \leq a_i(x_3 - x_2)(\xi_{32} - \xi_{21})\mathcal{H}''(\xi_{31}) \tag{123}$$

$$\leq a_i\eta(x_3 - x_2)\mathcal{H}''(x_3), \tag{124}$$

where $\eta = \xi_{32} - \xi_{21} > 0$, and the last bound is due to the fact that $x_3 > \xi_{31}$.

Going back to our normal notation, we rewrite bound (124) as follows.

$$\psi_i \leq a_i\eta\left(-ca_i - \frac{a_i}{n}\right)\mathcal{H}''(-ca_i) \tag{125}$$

$$= a_i\eta\frac{a_i}{n}(\beta - 1)\frac{-1}{\beta\frac{a_i}{n}\left(1 - \beta\frac{a_i}{n}\right)} \tag{126}$$

$$= -\frac{a_i\eta(\beta - 1)}{\beta\left(1 - \beta\frac{a_i}{n}\right)}, \tag{127}$$

This concludes the proof. □

6.8. Inequality 84
The series bound (84) is derived as follows.

$$\sum_{i\in\Omega}\left(\frac{s}{2i} \cdot \frac{a_i}{n}\right) = \left(\frac{s}{2} \cdot \frac{a_1}{n}\right) + \left(\frac{s}{4} \cdot \frac{a_2}{n}\right) + \cdots + \left(\frac{s}{s} \cdot \frac{a_{s/2}}{n}\right) \tag{128}$$

$$\geq \left(\frac{s}{2n} \cdot \frac{(1-\epsilon)ds}{2^{\log_2 s}}\right) + \left(\frac{s}{4n} \cdot \frac{(1-\epsilon)ds}{2^{\log_2(s/2)}}\right) + \cdots$$
$$+ \left(\frac{s}{sn} \cdot \frac{(1-\epsilon)ds}{2}\right) \tag{129}$$

$$= \left[\left(\frac{s}{2n} \cdot \frac{1}{s}\right) + \left(\frac{s}{4n} \cdot \frac{2}{s}\right) + \cdots\right.$$
$$\left. + \left(\frac{s}{sn} \cdot \frac{1}{2}\right)\right](1-\epsilon)ds \tag{130}$$

$$= \left(\frac{1}{2n} + \frac{1}{2n} + \cdots + \frac{1}{2n}\right)(1-\epsilon)ds$$

$$= \frac{\log_2(s/2)}{2n}(1-\epsilon)ds. \tag{131}$$

That is the required bounds, hence concluding the proof. □

7. CONCLUSION

We considered the construction of sparse matrices that are invaluable for dimensionality reduction with application in diverse fields. These sparse matrices are more efficient computationally compared to their dense counterparts also used for the purpose of dimensionality reduction. Our construction is probabilistic based on the dyadic splitting method we introduced in Bah and Tanner [1]. By better approximation of the bounds we achieve a novel result, which is a reduced complexity of the sparsity per column of these matrices. Precisely, a complexity that is a state-of-the-art divided by $\log s$, where s is the intrinsic dimension of the problem.

Our approach is one of a few that gives quantitative sampling theorems for existence of such sparse matrices. Moreover, using the phase transition framework comparison, our construction is better than existing probabilistic constructions. We are also able to compare performance of combinatorial compressed sensing algorithms by comparing their phase transition curves. This is

one perspective in algorithm comparison amongst a couple of others like runtime and iteration complexities.

Evidently, our results holds true for the construction of expander graphs, which is a graph theory problem and is of interest to communities in theoretical computer science and pure mathematics.

AUTHOR CONTRIBUTIONS

JT contributed to conception of the idea. Both JT and BB worked on the derivation of the initial results, which was substantially improved by BB. BB wrote the first draft of the manuscript. JT did some revisions and gave feedback.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fams.2018.00039/full#supplementary-material

REFERENCES

1. Bah B, Tanner J. Vanishingly sparse matrices and expander graphs, with application to compressed sensing. *IEEE Trans Inform Theor.* (2013) **59**:7491–508. doi: 10.1109/TIT.2013.2274267
2. Ahn KJ, Guha S, McGregor A. Graph sketches: sparsification, spanners, and subgraphs. In: *Proceedings of the 31st Symposium on Principles of Database Systems.* Scottsdale, AZ: ACM (2012). p. 5–14.
3. Gilbert AC, Levchenko K. Compressing network graphs. In: *Proceedings of the LinkKDD Workshop at the 10th ACM Conference on KDD.* Seatle, WA (2004).
4. Vardi Y. Network tomography: estimating source-destination traffic intensities from link data. *J Am Stat Assoc.* (1996) **91**:365–77.
5. Castro R, Coates M, Liang G, Nowak R, Yu B. Network tomography: recent developments. *Stat Sci.* (2004) **19**:499–517. doi: 10.1214/088342304000000422
6. Muthukrishnan S. *Data Streams: Algorithms and Applications.* Now Publishers Inc (2005).
7. Indyk P. *Sketching, Streaming and Sublinear-Space Algorithms.* Graduate course notest (2007).
8. Dwork C, McSherry F, Talwar K. The price of privacy and the limits of LP decoding. In: *Proceedings of the Thirty-ninth Annual ACM Symposium on Theory of Computing.* San Diego, CA: ACM (2007). p. 85–94.
9. Donoho DL, Johnstone IM, Hoch JC, Stern AS. Maximum entropy and the nearly black object. *J R Stat Soc Ser B* (1992) **54**:41–81.
10. Donoho DL. Compressed sensing. *IEEE Trans Inform Theor.* (2006) **52**:1289–306. doi: 10.1109/TIT.2006.871582
11. Xu W, Hassibi B. Efficient compressive sensing with deterministic guarantees using expander graphs. In: *Information Theory Workshop, 2007. ITW'07.* Tahoe City, CA: IEEE (2007). p. 414–9.
12. Jafarpour S, Xu W, Hassibi B, Calderbank R. Efficient and robust compressed sensing using optimized expander graphs. *IEEE Trans Inform Theor.* (2009) **55**:4299–308. doi: 10.1109/TIT.2009.2025528
13. Berinde R, Indyk P. Sequential sparse matching pursuit. In: *47th Annual Allerton Conference on Communication, Control, and Computing.* IEEE (2009). p. 36–43.
14. Mendoza-Smith R, Tanner J. Expander ℓ_0-decoding. *Appl. Comput. Harmon. Anal.* (2017) 45:642–67. doi: 10.1016/j.acha.2017.03.001
15. Mendoza-Smith R, Tanner J, Wechsung F. A robust parallel algorithm for combinatorial compressed sensing. *arXiv [preprint]* arXiv:170409012 (2017).
16. Du DZ, Hwang FK. *Combinatorial Group Testing and Its Applications.* World Scientific (2000).
17. Gilbert AC, Iwen MA, Strauss MJ. Group testing and sparse signal recovery. In: *Signals, Systems and Computers, 2008 42nd Asilomar Conference on.* Pacific Grove, CA: IEEE (2008). p. 1059–63.
18. Berinde R, Gilbert AC, Indyk P, Karloff H, Strauss MJ. Combining geometry and combinatorics: A unified approach to sparse signal recovery. In: *Communication, Control, and Computing, 2008 46th Annual Allerton Conference on.* Urbana-Champaign, IL: IEEE (2008). p. 798–805.
19. Foucart S, Rauhut H. *A Mathematical Introduction to Compressive Sensing.* New York, NY: Springer (2013).

20. Hoory S, Linial N, Wigderson A. Expander graphs and their applications. *Bull Am Math Soc.* (2006) **43**:439–562. doi: 10.1090/S0273-0979-06-01126-8
21. Guruswami V, Umans C, Vadhan S. Unbalanced expanders and randomness extractors from Parvaresh–Vardy codes. *J ACM* (2009) **56**:20. doi: 10.1145/1538902.1538904
22. Bassalygo LA, Pinsker MS. Complexity of an optimum nonblocking switching network without reconnections. *Problemy Peredachi Informatsii* (1973) **9**:84–87.
23. Indyk P, Razenshteyn I. On model-based RIP-1 matrices. In: *International Colloquium on Automata, Languages, and Programming.* Berlin; Heidelberg: Springer (2013). p. 564–75.
24. Bah B, Baldassarre L, Cevher V. Model-based sketching and recovery with expanders. In: *SODA.* SIAM (2014). p. 1529–43.
25. Donoho DL, Tanner J. Thresholds for the recovery of sparse solutions via l_1 minimization. In: *Information Sciences and Systems, 2006 40th Annual Conference on.* IEEE (2006). p. 202–6.
26. Buhrman H, Miltersen PB, Radhakrishnan J, Venkatesh S. Are bitvectors optimal? *SIAM J Comput.* (2002) **31**:1723–44. doi: 10.1137/S0097539702405292
27. Berinde R. *Advances in Sparse Signal Recovery Methods.* Massachusetts Institute of Technology (2009).
28. Candès EJ, Romberg J, Tao T. Stable signal recovery from incomplete and inaccurate measurements. *Commun Pure Appl Math.* (2006) **59**:1207–23. doi: 10.1002/cpa.20124
29. Capalbo M, Reingold O, Vadhan S, Wigderson A. Randomness conductors and constant-degree lossless expanders. In: *Proceedings of the Thirty-fourth Annual ACM Symposium on Theory of Computing.* New York, NY: ACM (2002). p. 659–68.
30. Blanchard JD, Cartis C, Tanner J. Compressed sensing: how sharp is the restricted isometry property? *SIAM Rev.* (2011) **53**:105–25. doi: 10.1137/090748160
31. Bah B, Tanner J. Improved bounds on restricted isometry constants for Gaussian matrices. *SIAM J Matrix Anal Appl.* (2010) **31**:2882–98. doi: 10.1137/100788884
32. Indyk P, Ruzic M. Near-optimal sparse recovery in the ℓ_1-norm. In: *Foundations of Computer Science, 2008. FOCS'08. IEEE 49th Annual IEEE Symposium on.* Philadelphia, PA: IEEE (2008). p. 199–207.
33. Berinde R, Indyk P, Ruzic M. Practical near-optimal sparse recovery in the ℓ_1-norm. In: *46th Annual Allerton Conference on Communication, Control, and Computing.* Piscataway, NJ: IEEE (2008). p. 198–205.