Broadcasting on Random Directed Acyclic Graphs

Anuran Makur, Elchanan Mossel, and Yury Polyanskiy

Abstract

We study the following generalization of the well-known model of broadcasting on trees to the case of directed acyclic graphs (DAGs). At time 0, a source vertex X sends out a uniform bit along binary symmetric channels (BSCs) to a set of vertices called layer 1. Each vertex except X is assumed to have indegree d. At time $k \geq 1$, vertices at layer k apply d-input Boolean processing functions to their received bits and send out the results to vertices at layer $k+1$. We say that broadcasting is possible if it is possible to reconstruct the original bit X with probability of error bounded away from $\frac{1}{2}$ using knowledge of the values of all vertices at an arbitrarily deep layer k. This question is also related to models of reliable computation, reliable storage, and information flow in biological networks.

In this paper, we treat the case of randomly constructed DAGs, for which we show that broadcasting is only possible if the BSC noise level is below a certain (degree and function dependent) critical threshold. For $d \geq 3$, and random DAGs with layers of size $\Omega(\log(k))$ and majority processing functions, we identify the critical threshold. For $d = 2$, we establish a similar result for the NAND processing function. We also prove a partial converse result for odd $d \geq 3$ illustrating that the identified thresholds are impossible to improve by selecting different processing functions if the decoder is restricted to using a single vertex's value.

Finally, for any BSC noise level, we construct explicit DAGs (using regular bipartite lossless expander graphs) with bounded degree and layers of size $\Theta(\log(k))$ admitting reconstruction. In particular, we show that the first r layers of such DAGs can be generated in either deterministic quasi-polynomial time or randomized polylogarithmic time in r. These results portray a doubly-exponential advantage for storing a bit in bounded degree DAGs compared to trees, where $d = 1$ but layer sizes need to grow exponentially with depth in order for broadcasting to be possible.

Index Terms

Broadcasting, reliable computation, noisy circuits, expander graphs, phase transition.

A. Makur and Y. Polyanskiy are with the Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139, USA (e-mail: a_makur@mit.edu; yp@mit.edu).

E. Mossel is with the Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA (e-mail: elmos@mit.edu).

I. INTRODUCTION

In this paper, we study a generalization of the well-known problem of broadcasting on trees [1]. In the broadcasting on trees problem, we are given a noisy tree T whose vertices are Bernoulli random variables and edges are independent binary symmetric channels (BSCs) with common crossover probability $\delta \in \left(0, \frac{1}{2}\right)$. Given that the root is an unbiased random bit, the objective is to decode the bit at the root from the bits at the kth layer of the tree. The authors of [1] characterize the sharp threshold for when such reconstruction is possible:

- If $(1 - 2\delta)^2 \operatorname{br}(T) < 1$, then the minimum probability of error in decoding tends to $\frac{1}{2}$ as $k \to \infty$,
- If $(1 - 2\delta)^2 \operatorname{br}(T) > 1$, then the minimum probability of error in decoding is bounded away from $\frac{1}{2}$ for all k,

where $\operatorname{br}(T)$ denotes the branching number of the tree, and the condition $(1 - 2\delta)^2 \operatorname{br}(T) \lesseqgtr 1$ determines the *Kesten-Stigum threshold* in the regular tree setting. A consequence of this result is that reconstruction is impossible for trees with sub-exponentially many vertices at each layer. Indeed, if L_k denotes the number of vertices at layer k and $\lim_{k \to \infty} L_k^{1/k} = 1$, then it is straightforward to show that $\operatorname{br}(T) = 1$, which in turn implies that $(1 - 2\delta)^2 \operatorname{br}(T) < 1$.

This result on reconstruction on trees generalizes results from random processes and statistical physics that hold for regular trees, cf. [2] (which proves achievability) and [3] (which proves the converse), and has had numerous extensions and further generalizations including [4]–[12]. (We refer readers to [1, Section 2.2] for further references on the Ising model literature.) Furthermore, reconstruction on trees plays a crucial role in understanding phylogenetic reconstruction, see e.g. [13]–[16], and in understanding phase transitions for random constraint satisfaction problems, see e.g. [17]–[20] and follow-up work.

Instead of analyzing trees, we consider the problem of broadcasting on *bounded degree directed acyclic graphs* (DAGs). As in the setting of trees, all vertices in our graphs are Bernoulli random variables and all edges are independent BSCs. Furthermore, variables located at vertices with indegree 2 or more are the values of a function on their noisy inputs.

Notice that compared to the setting of trees, broadcasting on DAGs has two principal differences: (a) in trees, layer sizes scale exponentially in depth, while in DAGs they are polynomial; (b) in trees, the indegree of each vertex is 1, while in DAGs each vertex has several incoming signals. The latter enables the possibility of information fusion at the vertices and our main goal is to understand whether the benefits of (b) overpower the harm of (a).

This paper has two main contributions. Firstly, by a probabilistic argument, we demonstrate the existence of bounded degree DAGs with $L_k = \Omega(\log(k))$ which permit recovery of the root bit for sufficiently low δ's. Secondly, we provide explicit deterministic constructions of such DAGs using regular bipartite lossless expander graphs. In particular, the constituent expander graphs for the first r layers of such DAGs can be constructed in either deterministic quasi-polynomial time or randomized polylogarithmic time in r. Together, these results imply that in terms of economy of storing information, *DAGs are doubly-exponentially more efficient than trees*.

A. Motivation

The problem of broadcasting on trees is closely related to the problem of reliable computation using noisy circuits, e.g. [21], [22]. Indeed, it can be thought of in the following way: suppose we want to remember a bit in a noisy

circuit of depth k. How big should the circuit be? Von Neumann [21] asked this question assuming that we take multiple clones of the original bit and recursively apply gates in order to reduce the noise. The broadcasting on trees perspective is to start from a single bit and repeatedly clone it so that one can recover it well from the vertices at depth k. The model we consider here again starts from a single bit but we are allowed to use bounded degree gates to reduce noise as well as to duplicate. This leads to much smaller circuits than the tree circuits.

As mentioned earlier, the broadcasting process on trees plays a fundamental role in phylogenetic reconstruction. The positive results obtained here suggest it might be possible to reconstruct other biological networks, such as phylogenetic networks (see e.g. [23]) or pedigrees (see e.g. [24], [25]), even if the growth of the network is very mild. It is interesting to explore if there are also connections between broadcasting on DAGs and random constraint satisfaction problems. Currently, we are not aware that such connections have been established.

Another motivation for this problem is to understand whether it is possible to propagate information in regular grids. Inspired by the work on one-dimensional probabilistic cellular automata, cf. [26], our conjecture is that such propagation is impossible for a two-dimensional grid regardless of the noise level and the choice of processing function (which is the same for every vertex). Our results towards establishing this conjecture will be the focus of a forthcoming paper.

B. Outline

We briefly outline the rest of this paper. Since we will use probabilistic arguments to establish the existence of bounded degree DAGs where reconstruction of root bit is possible, we will prove many of our results for random DAGs. So, the next subsection I-C formally defines the random DAG model. In section II, we present our three main results (as well as and some auxiliary results) pertaining to the random DAG model, and discuss several related results in the literature. Then, we prove these main results in sections III, IV, and V, respectively. In particular, section III analyzes broadcasting with majority processing functions when the indegree of each vertex is 3 or more, section IV analyzes broadcasting with AND and OR processing functions when the indegree of each vertex is 2, and section V illustrates our explicit constructions of DAGs where reconstruction of the root bit is possible using expander graphs. Finally, we conclude our discussion and list some open problems in section VI.

C. Random DAG Model

A *random DAG model* consists of an infinite DAG with fixed vertices that are Bernoulli ($\{0, 1\}$-valued) random variables and randomly generated edges which are independent BSCs. We first define the vertex structure of this model, where each vertex is identified with the corresponding random variable. Let the root (or "source") random variable be $X_{0,0} \sim \text{Bernoulli}\left(\frac{1}{2}\right)$. Furthermore, we define $X_k = (X_{k,0}, \ldots, X_{k,L_k-1})$ as the vector of vertex random variables at distance (i.e. length of shortest path) $k \in \mathbb{N} \triangleq \{0, 1, 2, \ldots\}$ from the root, where $L_k \in \mathbb{N}\backslash\{0\}$ denotes the number of vertices at distance k. In particular, we have $X_0 = (X_{0,0})$ so that $L_0 = 1$, and we are typically interested in the regime where $L_k \to \infty$ as $k \to \infty$.

We next define the edge structure of the random DAG model. For any $k \in \mathbb{N}\backslash\{0\}$ and any $j \in [L_k] \triangleq \{0, \ldots, L_k - 1\}$, we independently and uniformly select $d \in \mathbb{N}\backslash\{0\}$ vertices $X_{k-1,i_1}, \ldots, X_{k-1,i_d}$ from X_{k-1} (i.e. $i_1, \ldots, i_d$

are i.i.d. uniform on $[L_{k-1}]$), and then construct d directed edges: $(X_{k-1,i_1}, X_{k,j}), \ldots, (X_{k-1,i_d}, X_{k,j})$. (Here, $i_1, \ldots, i_d$ are independently chosen for each $X_{k,j}$.) This random process generates the underlying DAG structure. In the sequel, we will let G be a random variable representing this underlying (infinite) random DAG, i.e. G encodes the random configuration of the edges between the vertices.

To define a *Bayesian network* (or directed graphical model) on this random DAG, we fix some sequence of Boolean functions $f_k : \{0,1\}^d \to \{0,1\}$ for $k \in \mathbb{N}\backslash\{0\}$ (that depend on the level index k, but typically not on the realization of G), and some crossover probability $\delta \in \left(0, \frac{1}{2}\right)$ (since this is the interesting regime of δ). Then, for any $k \in \mathbb{N}\backslash\{0\}$ and $j \in [L_k]$, given $i_1, \ldots, i_d$ and $X_{k-1,i_1}, \ldots, X_{k-1,i_d}$, we define:

$$X_{k,j} = f_k(X_{k-1,i_1} \oplus Z_{k,j,1}, \ldots, X_{k-1,i_d} \oplus Z_{k,j,d}) \tag{1}$$

where $\oplus$ denotes addition modulo 2, and $\{Z_{k,j,i} : k \in \mathbb{N}\backslash\{0\}, j \in [L_k], i \in \{1, \ldots, d\}\}$ are i.i.d Bernoulli(δ) random variables that are independent of everything else. This means that each edge is a BSC with parameter δ, denoted BSC(δ). Moreover, (1) characterizes the conditional distribution of $X_{k,j}$ given its parents. In this model, the Boolean processing function used at a vertex $X_{k,j}$ depends only on the level index k. A more general model can be defined where each vertex $X_{k,j}$ has its own Boolean processing function $f_{k,j} : \{0,1\}^d \to \{0,1\}$ for $k \in \mathbb{N}\backslash\{0\}$ and $j \in [L_k]$. However, with the exception of a few converse results, we will mainly analyze instances of the simpler model in this paper.

Note that although we will analyze this model for convenience, as stated, our underlying graph is really a directed multigraph rather than a DAG, because we select the parents of a vertex with replacement. It is straightforward to construct an equivalent model where the underlying graph is truly a DAG. For each vertex $X_{k,j}$ with $k \in \mathbb{N}\backslash\{0\}$ and $j \in [L_k]$, we first construct d intermediate parent vertices $\{X_{k,j}^i : i = 1, \ldots, d\}$ that live between layers k and $k-1$, where each $X_{k,j}^i$ has a single edge pointing to $X_{k,j}$. Then, for each $X_{k,j}^i$, we independently and uniformly select a vertex from layer $k-1$, and construct a directed edge from that vertex to $X_{k,j}^i$. This defines a valid (random) DAG. As a result, every realization of G can be perceived as either a directed multigraph or its equivalent DAG. Furthermore, the Bayesian network on this true DAG is defined as follows: each $X_{k,j}$ is the output of a Boolean processing function f_k with inputs $\{X_{k,j}^i : i = 1, \ldots, d\}$, and each $X_{k,j}^i$ is the output of a BSC whose input is the unique parent of $X_{k,j}^i$ in layer $k-1$.

Finally, we define the "empirical probability of unity" at level $k \in \mathbb{N}$ as:

$$\sigma_k \triangleq \frac{1}{L_k} \sum_{m=0}^{L_k-1} X_{k,m} \tag{2}$$

where $\sigma_0 = X_{0,0}$ is just the root vertex. Observe that given $\sigma_{k-1} = \sigma$, $X_{k-1,i_1}, \ldots, X_{k-1,i_d}$ are i.i.d. Bernoulli(σ), and as a result, $X_{k-1,i_1} \oplus Z_{k,j,1}, \ldots, X_{k-1,i_d} \oplus Z_{k,j,d}$ are i.i.d. Bernoulli($\sigma * \delta$), where $\sigma * \delta \triangleq \sigma(1-\delta) + \delta(1-\sigma)$ is the convolution of σ and δ. Therefore, $X_{k,j}$ is the output of f_k upon inputting a d-length i.i.d. Bernoulli($\sigma * \delta$) string.

Under this setup, our objective is to determine whether or not the value at the root $\sigma_0 = X_{0,0}$ can be decoded from the observations X_k as $k \to \infty$. Since X_k is an exchangeable sequence of random variables given σ_0, for

any $x_{0,0}, x_{k,0}, \ldots, x_{k,L_k-1} \in \{0,1\}$ and any permutation π of $[L_k]$, we have $P_{X_k|\sigma_0}(x_{k,0}, \ldots, x_{k,L_k-1}|x_{0,0}) = P_{X_k|\sigma_0}(x_{k,\pi(0)}, \ldots, x_{k,\pi(L_k-1)}|x_{0,0})$. Letting $\sigma = \frac{1}{L_k} \sum_{j=0}^{L_k-1} x_{k,j}$, we can factorize $P_{X_k|\sigma_0}$ as:

$$P_{X_k|\sigma_0}(x_{k,0}, \ldots, x_{k,L_k-1}|x_{0,0}) = \binom{L_k}{L_k\sigma}^{-1} P_{\sigma_k|\sigma_0}(\sigma|x_{0,0}). \tag{3}$$

Using the Fisher-Neyman factorization theorem [27, Theorem 3.6], this implies that σ_k is a *sufficient statistic* of X_k for performing inference about σ_0. Therefore, we restrict our attention to the Markov chain $\{\sigma_k : k \in \mathbb{N}\}$ in our achievability proofs, since if decoding is possible from σ_k, then it is also possible from X_k. Given σ_k, inferring the value of σ_0 is a binary hypothesis testing problem with minimum achievable probability of error:

$$\mathbb{P}\big(f_{\mathsf{ML}}^k(\sigma_k) \neq \sigma_0\big) = \frac{1}{2}\left(1 - \left\|P_{\sigma_k}^+ - P_{\sigma_k}^-\right\|_{\mathsf{TV}}\right) \tag{4}$$

where $f_{\mathsf{ML}}^k : \{m/L_k : m = 0, \ldots, L_k\} \to \{0,1\}$ is the maximum likelihood (ML) decision rule based on the empirical probability of unity at level k in the absence of knowledge of the random DAG realization G, $P_{\sigma_k}^+$ and $P_{\sigma_k}^-$ are the conditional distributions of σ_k given $\sigma_0 = 1$ and $\sigma_0 = 0$ respectively, and for any two probability measures P and Q on the same measurable space $(\Omega, \mathcal{F})$, their *total variation (TV) distance* is defined as:

$$\|P - Q\|_{\mathsf{TV}} \triangleq \sup_{A \in \mathcal{F}} |P(A) - Q(A)| = \frac{1}{2}\|P - Q\|_1 \tag{5}$$

where $\|\cdot\|_1$ denotes the $\mathcal{L}^1$-norm. We say that reconstruction of the root bit σ_0 is possible when:[1]

$$\lim_{k\to\infty} \mathbb{P}\big(f_{\mathsf{ML}}^k(\sigma_k) \neq \sigma_0\big) < \frac{1}{2} \quad \Leftrightarrow \quad \lim_{k\to\infty} \left\|P_{\sigma_k}^+ - P_{\sigma_k}^-\right\|_{\mathsf{TV}} > 0. \tag{6}$$

In the sequel, to simplify our analysis when proving that reconstruction is possible, we will sometimes use other (sub-optimal) decision rules rather than the ML decision rule.

On the other hand, we will consider the Markov chain $\{X_k : k \in \mathbb{N}\}$ conditioned on G in our converse proofs. We say that reconstruction of the root bit X_0 is impossible when:

$$\lim_{k\to\infty} \mathbb{P}\big(h_{\mathsf{ML}}^k(X_k, G) \neq X_0\big|G\big) = \frac{1}{2} \quad G\text{-a.s.} \quad \Leftrightarrow \quad \lim_{k\to\infty} \left\|P_{X_k|G}^+ - P_{X_k|G}^-\right\|_{\mathsf{TV}} = 0 \quad G\text{-a.s.} \tag{7}$$

where $h_{\mathsf{ML}}^k(\cdot, G) : \{0,1\}^{L_k} \to \{0,1\}$ is the ML decision rule based on the full state at level k given knowledge of the random DAG realization G, $P_{X_k|G}^+$ and $P_{X_k|G}^-$ denote the conditional distributions of X_k given $\{X_0 = 1, G\}$ and $\{X_0 = 0, G\}$ respectively, and the notation G-a.s (almost surely) implies that the conditions in (7) hold with probability 1 with respect to the distribution of the random DAG G. Note that applying the bounded convergence theorem to the TV distance condition in (7) yields $\lim_{k\to\infty} \mathbb{E}[\|P_{X_k|G}^+ - P_{X_k|G}^-\|_{\mathsf{TV}}] = 0$, and employing Jensen's inequality here establishes the weaker impossibility result:

$$\lim_{k\to\infty} \mathbb{P}\big(h_{\mathsf{ML}}^k(X_k) \neq X_0\big) = \frac{1}{2} \quad \Leftrightarrow \quad \lim_{k\to\infty} \left\|P_{X_k}^+ - P_{X_k}^-\right\|_{\mathsf{TV}} = 0 \tag{8}$$

where $h_{\mathsf{ML}}^k : \{0,1\}^{L_k} \to \{0,1\}$ is the ML decision rule based on the full state at level k in the absence of knowledge of the random DAG realization G, and $P_{X_k}^+$ and $P_{X_k}^-$ are the conditional distributions of X_k given $X_0 = 1$ and

---

[1]The limits in (6), (7), and (8) exist because $\mathbb{P}\big(f_{\mathsf{ML}}^k(\sigma_k) \neq \sigma_0\big)$, $\mathbb{P}\big(h_{\mathsf{ML}}^k(X_k) \neq X_0\big)$, and $\mathbb{P}\big(h_{\mathsf{ML}}^k(X_k, G) \neq X_0\big|G\big)$ (for any fixed realization G) are monotone non-decreasing sequences in k that are bounded above by $\frac{1}{2}$. This can be deduced either from the data processing inequality for TV distance, or from the fact that a randomized decoder at level k can simulate the stochastic transition to level $k+1$.

$X_0 = 0$, respectively. Since σ_k is a sufficient statistic of X_k for performing inference about σ_0 when we average over G, we have $\mathbb{P}\big(h_{\mathsf{ML}}^k(X_k) \neq X_0\big) = \mathbb{P}\big(f_{\mathsf{ML}}^k(\sigma_k) \neq \sigma_0\big)$ (or equivalently, $\big\|P_{X_k}^+ - P_{X_k}^-\big\|_{\mathsf{TV}} = \big\|P_{\sigma_k}^+ - P_{\sigma_k}^-\big\|_{\mathsf{TV}}$), and the condition in (8) is a counterpart of (6).

II. MAIN RESULTS AND DISCUSSION

In this section, we state our main results, briefly delineate the main techniques or intuition used in the proofs, and discuss related literature.

A. Results on Random DAG Models

We prove two main results on the random DAG model. The first considers the setting where the indegree of each vertex (except the root) is $d \geq 3$. In this scenario, taking a majority vote of the inputs at each vertex intuitively appears to have good "local error correction" properties. So, we fix all Boolean functions in the random DAG model to be the (d-input) *majority* rule, and prove that this model exhibits a phase transition phenomenon around a critical threshold of:

$$\delta_{\mathsf{maj}} \triangleq \frac{1}{2} - \frac{2^{d-2}}{\left\lceil \dfrac{d}{2} \right\rceil \dbinom{d}{\lceil \frac{d}{2} \rceil}} \tag{9}$$

where $\lceil \cdot \rceil$ denotes the ceiling function. Indeed, the theorem below illustrates that for $\delta < \delta_{\mathsf{maj}}$, the majority decision rule $\hat{S}_k \triangleq \mathbb{1}\{\sigma_k \geq \frac{1}{2}\}$ can asymptotically decode σ_0, but for $\delta > \delta_{\mathsf{maj}}$, the ML decision rule with knowledge of G cannot asymptotically decode σ_0.

Theorem 1 (Phase Transition in Random DAG Model with Majority Rule Processing). *Let $C(\delta, d)$ and $D(\delta, d)$ be the constants defined in* (37) *and* (33) *in section* III. *For a random DAG model with $d \geq 3$ and majority processing functions (where ties are broken by outputting random bits), the following phase transition phenomenon occurs around δ_{maj}:*

1) *If $\delta \in (0, \delta_{\mathsf{maj}})$, and the number of vertices per level satisfies $L_k \geq C(\delta, d) \log(k)$ for all sufficiently large k (depending on δ and d), then reconstruction is possible in the sense that:*[2]

$$\limsup_{k \to \infty} \mathbb{P}(\hat{S}_k \neq \sigma_0) < \frac{1}{2}$$

where we use the majority decoder $\hat{S}_k = \mathbb{1}\{\sigma_k \geq \frac{1}{2}\}$ at level k.

2) *If $\delta \in \big(\delta_{\mathsf{maj}}, \frac{1}{2}\big)$, and the number of vertices per level satisfies $L_k = o\big(D(\delta, d)^{-k}\big)$, then reconstruction is impossible in the sense of* (7):

$$\lim_{k \to \infty} \big\|P_{X_k|G}^+ - P_{X_k|G}^-\big\|_{\mathsf{TV}} = 0 \quad G\text{-}a.s.$$

Theorem 1 is proved in section III. Intuitively, the proof considers the conditional expectation function, $g : [0, 1] \to [0, 1]$, $g(\sigma) = \mathbb{E}[\sigma_k | \sigma_{k-1} = \sigma]$ (see (28) and (29) in section III), which provides the approximate value of σ_k given the value of σ_{k-1} for large k. This function turns out to have three fixed points when $\delta \in (0, \delta_{\mathsf{maj}})$,

---

[2]Throughout this paper, $\log(\cdot)$ and $\exp(\cdot)$ denote the natural logarithm and natural exponential (with base e), respectively.

and only one fixed point when $\delta \in \left(\delta_{\mathsf{maj}}, \frac{1}{2}\right)$. In the former case, σ_k "moves" to the largest fixed point when $\sigma_0 = 1$, and to the smallest fixed point when $\sigma_0 = 0$. In the latter case, σ_k "moves" to the unique fixed point of $\frac{1}{2}$ regardless of the value of σ_0 (see Proposition 5 in section III).[3] This provides the guiding intuition for why we can asymptotically decode σ_0 when $\delta \in (0, \delta_{\mathsf{maj}})$, but not when $\delta \in \left(\delta_{\mathsf{maj}}, \frac{1}{2}\right)$.

The recursive (or fixed point) structure of g in the special case where $d = 3$ and $\delta_{\mathsf{maj}} = \frac{1}{6}$ can be traced back to the work of von Neumann in [21]. So, it is worth comparing Theorem 1 with von Neumann's results in [21, Section 8], where the threshold of $\frac{1}{6}$ is also significant. In [21, Section 8], von Neumann demonstrates the possibility of reliable computation by constructing a circuit with successive layers of computation and local error correction using 3-input δ-noisy majority gates (i.e. the gates independently make errors with probability δ). In his analysis, he first derives a simple recursion that captures the effect on the probability of error after applying a single noisy majority gate. Then, he uses a "heuristic" fixed point argument to show that as the depth of the circuit grows, the probability of error asymptotically stabilizes at a fixed point value less than $\frac{1}{2}$ if $\delta < \frac{1}{6}$, and the probability of error tends to $\frac{1}{2}$ if $\delta \geq \frac{1}{6}$. Moreover, he rigorously proves that reliable computation is possible for $\delta < 0.0073$.

As we mentioned in subsection I-A, von Neumann's approach to remembering a random initial bit entails using multiple clones of the initial bit as inputs to a noisy circuit with one output, where the output equals the initial bit with probability greater than $\frac{1}{2}$ for "good" choices of noisy gates. It is observed in [28, Section 2] that a balanced ternary tree circuit, with k layers of 3-input noisy majority gates and 3^k inputs that are all equal to the initial bit, can be used to remember the initial bit. In fact, von Neumann's heuristic fixed point argument that yields a critical threshold of $\frac{1}{6}$ for reconstruction is rigorous in this scenario. From this starting point, Hajek and Weller prove the stronger impossibility result that reliable computation is impossible for formulae (i.e. circuits where the output of each intermediate gate is the input of only one other gate) with general 3-input δ-noisy gates when $\delta \geq \frac{1}{6}$ [28, Proposition 2]. This development can be generalized for any odd $d \geq 3$, and [29, Theorem 1] conveys that reliable computation is impossible for formulae with general d-input δ-noisy gates when $\delta \geq \delta_{\mathsf{maj}}$.

The discussion heretofore reveals that the critical thresholds in von Neumann's circuit for remembering a bit and in our model in Theorem 1 are both δ_{maj}. It turns out that this is a consequence of the common fixed point iteration structure of the two problems (as we will explain below). Indeed, the general recursive structure of g for any odd value of d was analyzed in [29, Section 2]. On a related front, the general recursive structure of g was also analyzed in [6] in the context of performing recursive reconstruction on periodic trees, where the critical threshold of δ_{maj} again plays a crucial role. In fact, we will follow the analysis in [6] to develop these recursions in section III.

We now elucidate the common fixed point iteration structure between von Neumann's model and our model in Theorem 1. Suppose $d \geq 3$ is odd, and define the function $h : [0, 1] \to [0, 1]$, $h(p) \triangleq \mathbb{P}(\mathsf{majority}(Y_1, \ldots, Y_d) = 1)$ for $Y_1, \ldots, Y_d$ i.i.d. Bernoulli(p). Consider von Neumann's balanced d-ary tree circuit with k layers of d-input δ-noisy majority gates and d^k inputs that are all equal to the initial bit. In this model, it is straightforward to verify

---

[3]Note, however, that $\sigma_k \to \frac{1}{2}$ almost surely as $k \to \infty$ does not imply the impossibility of reconstruction in the sense of (8), let alone (7). So, a different argument is required to establish such impossibility results.

that the probability of error (i.e. output vertex $\neq$ initial bit) is $f^{(k)}(0)$, where $f : [0,1] \to [0,1]$ is given by [29, Equation (3)]:

$$f(\sigma) \triangleq \delta * h(\sigma), \tag{10}$$

and $f^{(k)}$ denotes the k-fold composition of f with itself. On the other hand, as explained in the brief intuition for our proof of Theorem 1 earlier, assuming that $\sigma_0 = 0$, the relevant recursion for our model is given by the repeated composition $g^{(k)}(0)$ (which captures the average position of σ_k after k layers). According to (28) in section III, $g(\sigma) = h(\delta * \sigma)$, which yields the relation:

$$\forall k \in \mathbb{N}\backslash\{0\}, \;\; f^{(k+1)}(0) = \delta * g^{(k)}(0) \tag{11}$$

by induction. Therefore, the fixed point iteration structures of f and g are identical, and δ_{maj} is the common critical threshold that determines when there is a unique fixed point. In particular, the fact that gates (or vertices) are noisy in von Neumann's model, while edges (or wires) are noisy in our model, has no bearing on this fixed point structure.[4]

Although both the aforementioned models use majority gates and share a common fixed point structure, it is important to recognize that our overall analysis differs from von Neumann's analysis in a crucial way. Since our recursion pertains to conditional expectations of the proportion of 1's in different layers (rather than the probabilities of error in von Neumann's setting), our proof requires exponential concentration inequalities to formalize the intuition provided by the fixed point analysis.

We now make several other pertinent remarks about Theorem 1. Firstly, reconstruction is possible in the sense of (6) when $\delta \in (0, \delta_{\mathsf{maj}})$ since the ML decision rule achieves lower probability of error than the majority decision rule,[5] and reconstruction is impossible in the sense of (8) when $\delta \in \left(\delta_{\mathsf{maj}}, \frac{1}{2}\right)$ (as explained at the end of subsection I-C). Furthermore, while part 1 of Theorem 1 only shows that the ML decoder $f^k_{\mathsf{ML}}(\sigma_k)$ based on σ_k is optimal in the absence of knowledge of the particular graph realization G, part 2 establishes that even if the ML decoder knows the graph G and has access to the full k-layer state X_k, it cannot beat the δ_{maj} threshold in all but a zero measure set of DAGs.

Secondly, the following conjecture is still open: In the random DAG model with $L_k = O(\log(k))$ and fixed $d \geq 3$, reconstruction is impossible for all choices of Boolean processing functions when $\delta \geq \delta_{\mathsf{maj}}$. A consequence of this conjecture is that majority processing functions are optimal, i.e. they achieve the δ_{maj} reconstruction threshold. The results in [6] provide strong evidence that this conjecture is true when all vertices in the random DAG use the same odd Boolean processing function. Indeed, for fixed $\delta \in \left(0, \frac{1}{2}\right)$ and any odd Boolean function gate : $\{0,1\}^d \to \{0,1\}$, let $\tilde{g} : [0,1] \to [0,1]$ be defined as $\tilde{g}(\sigma) \triangleq \mathbb{P}(\mathsf{gate}(Y_1, \ldots, Y_d) = 1)$ for $Y_1, \ldots, Y_d$ i.i.d. Bernoulli($\delta * \sigma$).[6] Then,

---

[4] We refer readers to [30] for general results on the relation between vertex noise and edge noise.

[5] It can be seen from monotonicity and symmetry considerations that without knowledge of the random DAG realization G, the ML decision rule $f^k_{\mathsf{ML}}(\sigma_k)$ is equal to the majority decision rule $\hat{S}_k$. (So, the superior limit in part 1 of Theorem 1 can be replaced by a true limit.) On the other hand, with knowledge of the random DAG realization G, the ML decision rule $f^k_{\mathsf{ML}}(\sigma_k, G)$ based on σ_k is not the majority decision rule.

[6] A Boolean function is said to be *odd* if flipping all its input bits also flips the output bit. The assumption that gate is odd ensures that the function $R^\delta_{\mathsf{gate}}(\sigma)$ in [6, Definition 2.1] is precisely equal to the function $\tilde{g}(\sigma)$.

[6, Lemma 2.4] establishes that $\tilde{g}(\sigma) \leq g(\sigma)$ for all $\sigma \geq \frac{1}{2}$ and $\tilde{g}(\sigma) \geq g(\sigma)$ for all $\sigma \leq \frac{1}{2}$, where the function g is given in (28) (and corresponds to the majority rule). Hence, if g has a single fixed point at $\sigma = \frac{1}{2}$, $\tilde{g}$ also has a single fixed point at $\sigma = \frac{1}{2}$. This intuitively suggests that if reconstruction of the root bit is impossible using majority processing functions, it is also impossible using any odd processing function. Furthermore, our proof of part 2 of Theorem 1 in section III yields that reconstruction is impossible for all choices of odd and monotone non-decreasing Boolean processing functions when $\delta > \delta_{\mathsf{maj}}$, modulo the following conjecture (which we did not verify): among all odd and monotone non-decreasing Boolean functions, the maximum Lipschitz constant of $\tilde{g}$ is attained by the majority rule at $\sigma = \frac{1}{2}$.

Thirdly, the sub-exponential layer size condition $L_k = o\big(D(\delta,d)^{-k}\big)$ in part 2 of Theorem 1 is intuitively necessary. Suppose every Boolean processing function in our random DAG model simply outputs the value of its first input bit. This effectively sets $d = 1$, and reduces our problem to one of broadcasting on a random tree model. If $L_k = \Omega(E(\delta)^k)$ for some large enough constant $E(\delta)$, then most realizations of the random tree will have branching numbers greater than $(1 - 2\delta)^{-2}$. As a result, reconstruction will be possible for most realizations of the random tree (cf. the Kesten-Stigum threshold delineated at the outset of section I). Thus, when we are proving impossibility results, L_k (at least intuitively) cannot be exponential in k with a very large base.

Fourthly, it is worth mentioning that for any fixed DAG with indegree $d \geq 3$ and sub-exponential L_k, for any choices of Boolean processing functions, and any choice of decoder, it is impossible to reconstruct the root bit when $\delta > \frac{1}{2} - \frac{1}{2\sqrt{d}}$. This follows from Evans and Schulman's result in [22], which we will discuss further in subsection II-C.

Lastly, in the context of the random DAG model studied in Theorem 1, the ensuing proposition illustrates that the problem of reconstruction using the information contained in just a single vertex, e.g. $X_{k,0}$, exhibits a similar phase transition phenomenon to that in Theorem 1.

Proposition 1 (Single Vertex Reconstruction). *Let $C(\delta, d)$ be the constant defined in (37) in section III. For a random DAG model with $d \geq 3$, the following phase transition phenomenon occurs around δ_{maj}:*

1) *If $\delta \in (0, \delta_{\mathsf{maj}})$, the number of vertices per level satisfies $L_k \geq C(\delta, d) \log(k)$ for all sufficiently large k (depending on δ and d), and all Boolean processing functions are the majority rule (where ties are broken by outputting random bits), then reconstruction is possible in the sense that:*

$$\limsup_{k \to \infty} \mathbb{P}(X_{k,0} \neq X_{0,0}) < \frac{1}{2}$$

   where we use a single vertex $X_{k,0}$ as the decoder at level k.

2) *If $\delta \in \big[\delta_{\mathsf{maj}}, \frac{1}{2}\big)$, d is odd, and the number of vertices per level satisfies $\lim_{k \to \infty} L_k = \infty$ and $R_k \triangleq \inf_{n \geq k} L_n = O\big(d^{2k}\big)$, then for all choices of Boolean processing functions (which may vary between vertices and be graph dependent), reconstruction is impossible in the sense that:*

$$\lim_{k \to \infty} \mathbb{E}\Big[\big\| P^+_{X_{k,0}|G} - P^-_{X_{k,0}|G} \big\|_{\mathsf{TV}}\Big] = 0$$

   where $P^+_{X_{k,0}|G}$ and $P^-_{X_{k,0}|G}$ are the conditional distributions of $X_{k,0}$ given $\{X_{0,0} = 1, G\}$ and $\{X_{0,0} = 0, G\}$, respectively.

Proposition 1 is proved in Appendix A. In particular, part 2 of Proposition 1 demonstrates that when $\delta \geq \delta_{\mathsf{maj}}$, the ML decoder based on a single vertex $X_{k,0}$ (with knowledge of the random DAG realization G) cannot reconstruct $X_{0,0}$ in all but a vanishing fraction of DAGs. Thus, if reconstruction is possible in the range $\delta \geq \delta_{\mathsf{maj}}$, the decoder should definitely use more than one vertex. This converse result relies on the aforementioned impossibility results on reliable computation. Specifically, the exact threshold δ_{maj} that determines whether or not reliable computation is possible using formulae is known for odd $d \geq 3$, cf. [28], [29]. Therefore, we can exploit such results to obtain a converse for odd $d \geq 3$ which holds for all choices of Boolean processing functions and at the critical value $\delta = \delta_{\mathsf{maj}}$ (although only for single vertex decoding). In contrast, when $d \geq 4$ is even, it is not even known whether such a critical threshold exists (as noted in [29, Section 7]), and hence, we cannot easily prove such converse results for even $d \geq 4$.[7]

We next present an immediate corollary of Theorem 1 which states that there exist constant indegree deterministic DAGs with $L_k = \Omega(\log(k))$ (i.e. $L_k \geq C(\delta, d) \log(k)$ for some large constant $C(\delta, d)$ and all sufficiently large k) such that reconstruction of the root bit is possible. Note that *deterministic DAGs* refer to Bayesian networks on specific realizations of G in the sequel. We will use the same notation as subsection I-C to analyze deterministic DAGs with the understanding that the randomness is engendered by $X_{0,0}$ and the edge BSCs, but not G. Formally, we have the following result which is proved in Appendix B.

Corollary 1 (Existence of DAGs where Reconstruction is Possible). *For every indegree $d \geq 3$, every noise level $\delta \in (0, \delta_{\mathsf{maj}})$, and every sequence of level sizes satisfying $L_k \geq C(\delta, d) \log(k)$ for all sufficiently large k, there exists a deterministic DAG $\mathcal{G}$ with these parameters such that if we use majority rules as our Boolean processing functions, then there exists $\epsilon = \epsilon(\delta, d) > 0$ (that depends on δ and d) such that the probability of error in ML decoding is bounded away from $\frac{1}{2} - \epsilon$:*

$$\forall k \in \mathbb{N}, \ \ \mathbb{P}\big(h_{\mathsf{ML}}^k(X_k, \mathcal{G}) \neq X_0\big) \leq \frac{1}{2} - \epsilon$$

where $h_{\mathsf{ML}}^k(\cdot, \mathcal{G}) : \{0,1\}^{L_k} \to \{0,1\}$ denotes the ML decision rule at level k based on the full k-layer state X_k (given knowledge of the DAG $\mathcal{G}$).

Since the critical threshold $\delta_{\mathsf{maj}} \to \frac{1}{2}$ as $d \to \infty$, a consequence of Corollary 1 is that for any $\delta \in \big(0, \frac{1}{2}\big)$, any sufficiently large indegree d (that depends on δ), and any sequence of level sizes satisfying $L_k \geq C(\delta, d) \log(k)$ for all sufficiently large k, there exists a deterministic DAG $\mathcal{G}$ with these parameters and all majority processing functions such that reconstruction of the root bit is possible in the sense shown above.

Until now, we have restricted ourselves to the $d \geq 3$ case of the random DAG model. Our second main result considers the setting where the indegree of each vertex (except the root) is $d = 2$, because it is not immediately obvious that deterministic DAGs (for which reconstruction is possible) exist for $d = 2$. Indeed, it is not entirely clear which Boolean processing functions are good for "local error correction" in this scenario. We choose to fix

---

[7]Note, however, that if all Boolean processing functions are the majority rule and the conditions of part 2 of Theorem 1 are satisfied, then part 2 of Theorem 1 implies (using the data processing inequality for TV distance and the bounded convergence theorem) that single vertex reconstruction is also impossible in the sense presented in part 2 of Proposition 1.

all Boolean functions at even levels of the random DAG model to be the AND rule, and all Boolean functions at odd levels of the model to be the OR rule. We then prove that this random DAG model also exhibits a phase transition phenomenon around a critical threshold of $\delta_{\mathsf{andor}} \triangleq \frac{3-\sqrt{7}}{4}$. As before, the next theorem illustrates that for $\delta < \delta_{\mathsf{andor}}$, the "biased" majority decision rule $\hat{T}_k \triangleq \mathbb{1}\{\sigma_k \geq t\}$, where $t \in (0,1)$ is defined in (53) in section IV, can asymptotically decode σ_0, but for $\delta > \delta_{\mathsf{andor}}$, the ML decision rule with knowledge of G cannot asymptotically decode σ_0. For simplicity, we only analyze this model at even levels.

Theorem 2 (Phase Transition in Random DAG Model with AND-OR Rule Processing). *Let $C(\delta)$ and $D(\delta)$ be the constants defined in (59) and (49) in section IV. For a random DAG model with $d = 2$, AND processing functions at even levels, and OR processing functions at odd levels, the following phase transition phenomenon occurs around δ_{andor}:*

1) *If $\delta \in (0, \delta_{\mathsf{andor}})$, and the number of vertices per level satisfies $L_k \geq C(\delta)\log(k)$ for all sufficiently large k (depending on δ), then reconstruction is possible in the sense that:*

$$\limsup_{k \to \infty} \mathbb{P}(\hat{T}_{2k} \neq \sigma_0) < \frac{1}{2}$$

where we use the decoder $\hat{T}_{2k} = \mathbb{1}\{\sigma_{2k} \geq t\}$ at level $2k$, which recovers the root bit by thresholding at the value $t \in (0,1)$ in (53).

2) *If $\delta \in \left(\delta_{\mathsf{andor}}, \frac{1}{2}\right)$, and the number of vertices per level satisfies $L_k = o\left(E(\delta)^{-\frac{k}{2}}\right)$ and $\liminf_{k \to \infty} L_k > \frac{2}{E(\delta)-D(\delta)}$ for any $E(\delta) \in (D(\delta), 1)$ (that depends on δ), then reconstruction is impossible in the sense of (7):*

$$\lim_{k \to \infty} \left\| P^{+}_{X_{2k}|G} - P^{-}_{X_{2k}|G} \right\|_{\mathsf{TV}} = 0 \quad G\text{-}a.s.$$

Theorem 2 is proved in section IV, and many of the remarks pertaining to Theorem 1 as well as the general intuition for Theorem 1 also hold for Theorem 2. Furthermore, a proposition analogous to part 1 of Proposition 1 and a corollary analogous to Corollary 1 also hold here (but we omit explicit statements of these results for brevity).

It is straightforward to verify that the random DAG in Theorem 2 with alternating layers of AND and OR processing functions is equivalent a random DAG with all NAND processing functions for the purposes of broadcasting.[8] Recall that in the discussion following Theorem 1, we noted how the critical threshold δ_{maj} was already known in the reliable computation literature (because it characterized when reliable computation is possible), cf. [29]. It turns out that δ_{andor} has also appeared in the reliable computation literature in a similar vein. In particular, although the existence of critical thresholds on δ for reliable computation using formulae of δ-noisy gates is not known for any even $d \geq 4$, the special case of $d = 2$ has been resolved. Indeed, Evans and Pippenger showed in [31] that reliable computation using formulae consisting of δ-noisy NAND gates is possible when $\delta < \delta_{\mathsf{andor}}$ and impossible

---

[8]Indeed, we can introduce pairs of NOT gates into every edge of our DAG that goes from an AND gate to an OR gate without affecting the statistics of the model. Since an AND gate followed by a NOT gate and an OR gate whose inputs pass through NOT gates are both NAND gates, we obtain an equivalent model where all processing functions are NAND gates. We remark that analyzing this random DAG model with NAND processing functions yields a version of Theorem 2 with the same essential characteristics (albeit with possibly weaker conditions on L_k).

when $\delta > \delta_{\mathsf{andor}}$. Moreover, Unger established in [32] that reliable computation using formulae with general 2-input δ-noisy gates is impossible when $\delta \geq \delta_{\mathsf{andor}}$.

B. Explicit Construction of Deterministic DAGs where Broadcasting is Possible

Although Corollary 1 illustrates the existence of DAGs where broadcasting (i.e. reconstruction of the root bit) is possible, it does not elucidate the structure of such DAGs. Moreover, Theorem 1 suggests that reconstruction on such deterministic DAGs should be possible using the algorithmically simple majority decision rule, but Corollary 1 is proved for the typically more complex ML decision rule. In this subsection, we address these deficiencies of Corollary 1 by presenting an explicit construction of deterministic bounded degree DAGs such that $L_k = \Theta(\log(k))$ and reconstruction of the root bit is possible using the majority decision rule.

Our construction is based on *regular bipartite lossless expander graphs*. Historically, the notion of an expander graph goes back to the work of Kolmogorov and Barzdin in [33]. Soon afterwards, Pinsker independently discovered such graphs and coined the term "expander graph" in [34].[9] Both [33] and [34, Lemma 1] prove the existence of expander graphs using probabilistic techniques. On the other hand, the first explicit construction of expander graphs appeared in [35], and more recently, lossless expander graphs were constructed using simpler ideas in [36]. We next define a pertinent variant of lossless expander graphs.

Consider a *d-regular bipartite graph* $B = (U, V, E)$, where U and V are two disjoint sets of vertices such that $|U| = |V| = n \in \mathbb{N}\backslash\{0\}$, every vertex in $U \cup V$ has degree $d \in \mathbb{N}\backslash\{0\}$, and E is the set of undirected edges between U and V. Note that we allow multiple edges to exist between two vertices in B. For any subset of vertices $S \subseteq U$, we define the *neighborhood* of S as:

$$\Gamma(S) \triangleq \{v \in V : \exists u \in S, \{u, v\} \in E\} \tag{12}$$

which is the set of all vertices in V that are adjacent to some vertex in S. For any fraction $\alpha \in (0, 1)$ and any expansion factor $\beta > 0$, B is called an (α, β)-*expander graph* if for every subset of vertices $S \subseteq U$, we have:

$$|S| \leq \alpha n \ \Rightarrow \ |\Gamma(S)| \geq \beta|S|. \tag{13}$$

Note that we only require subsets of vertices in U to expand (not V). Intuitively, such expander graphs are sparse due to the d-regularity constraint, but have high connectivity due to the expansion property (13). Furthermore, when $\alpha \leq \frac{1}{d}$, the best expansion factor one can hope for is β as close as possible to d. Hence, $(\alpha, (1-\epsilon)d)$-expander graphs with $\alpha \leq \frac{1}{d}$ and very small $\epsilon > 0$ are known as *lossless* expander graphs [36, Section 1.1].

We utilize a slightly relaxed version of lossless expander graphs in our construction. In particular, using existing results from the literature, we establish in Corollary 2 of section V that for large values of the degree d and any sufficiently large n (depending on d), there exists a d-regular bipartite graph $B = (U, V, E)$ with $|U| = |V| = n$ such that for every subset of vertices $S \subseteq U$, we have:

$$|S| = \frac{n}{d^{6/5}} \ \Rightarrow \ |\Gamma(S)| \geq (1 - \epsilon)\frac{n}{d^{1/5}} \ \text{with } \epsilon = \frac{2}{d^{1/5}} \,. \tag{14}$$

---

[9]In fact, expander graphs are called "expanding" graphs in [34].

Unlike (13), the expansion in (14) only holds for subsets $S \subseteq U$ with cardinality exactly $|S| = nd^{-6/5}$. However, we can still (loosely) perceive the graph B as a d-regular bipartite lossless (α, β)-expander graph with $\alpha = d^{-6/5}$ and $\beta = (1 - \epsilon)d$. (Strictly speaking, $nd^{-6/5}$ must be an integer, but we neglect this detail throughout our exposition for simplicity.) In the remainder of our discussion, we refer to graphs like B that satisfy (14) as d-*regular bipartite lossless* $(d^{-6/5}, d - 2d^{4/5})$-*expander graphs* with abuse of standard nomenclature.

A d-regular bipartite lossless $(d^{-6/5}, d - 2d^{4/5})$-expander graph B can be construed as representing two consecutive levels of a deterministic DAG upon which we are broadcasting. Indeed, we can make every edge in E directed by making them point from U to V, where U represents a particular level in the DAG and V the next level. In fact, we can construct deterministic DAGs where broadcasting is possible by concatenating several such d-regular bipartite lossless expander graphs together. The ensuing theorem details our DAG construction, and illustrates that reconstruction of the root bit is possible when we use majority Boolean processing functions and the majority decision rule $\hat{S}_k = \mathbb{1}\{\sigma_k \geq \frac{1}{2}\}$, where σ_k is defined in (2).

Theorem 3 (DAG Construction using Expander Graphs). *Fix any noise level $\delta \in \left(0, \frac{1}{2}\right)$, any sufficiently large odd degree $d = d(\delta) \geq 5$ (that depends on δ) satisfying:*

$$\frac{8}{d^{1/5}} + d^{6/5} \exp\left(-\frac{(1 - 2\delta)^2(d - 4)^2}{8d}\right) \leq \frac{1}{2}, \tag{15}$$

and any sufficiently large constant $N = N(\delta) \in \mathbb{N}$ (that depends on δ) such that the constant $M \triangleq \exp(N/(4d^{12/5}))$ ≥ 2 and for every $n \geq N$, there exists a d-regular bipartite lossless $(d^{-6/5}, d - 2d^{4/5})$-expander graph $B_n = (U_n, V_n, E_n)$ with $|U_n| = |V_n| = n$ that satisfies (14) for every subset $S \subseteq U_n$. Let the sequence of level sizes $\{L_k : k \in \mathbb{N}\}$ be given by $L_0 = 1$, $L_1 = N$, and:

$$\forall m \in \mathbb{N}, \forall k \in \mathbb{N} \text{ such that } M^{\lfloor 2^{m-1} \rfloor} < k \leq M^{2^m}, \quad L_k = 2^m N \tag{16}$$

where $\lfloor \cdot \rfloor$ denotes the floor function, and $L_k = \Theta(\log(k))$. Then, either in deterministic quasi-polynomial time $O(\exp(\Theta(\log(r) \log\log(r))))$, or if N additionally satisfies (81), in randomized polylogarithmic time $O(\log(r) \log\log(r))$ with strictly positive success probability (83), we can construct the constituent expander graphs for levels $0, \ldots, r$ of an infinite deterministic DAG with level sizes $\{L_k : k \in \mathbb{N}\}$ defined above, indegrees bounded by d, outdegrees bounded by $2d$, and the following edge configuration:

1) *Every vertex in X_1 has one directed edge coming from $X_{0,0}$.*
2) *For every pair of consecutive levels k and $k + 1$ such that $L_{k+1} = L_k$, the directed edges from X_k to X_{k+1} are given by the edges of B_{L_k}, where we identify the vertices in U_{L_k} with X_k and the vertices in V_{L_k} with X_{k+1}, respectively.*
3) *For every pair of consecutive levels k and $k+1$ such that $L_{k+1} = 2L_k$, we partition the vertices in X_{k+1} into two sets, $X_{k+1}^1 = (X_{k+1,0}, \ldots, X_{k+1,L_k})$ and $X_{k+1}^2 = (X_{k+1,L_k+1}, \ldots, X_{k+1,L_{k+1}})$, so that the directed edges from X_k to X_{k+1}^i are given by the edges of B_{L_k} for $i = 1, 2$, where we identify the vertices in U_{L_k} with X_k and the vertices in V_{L_k} with X_{k+1}^i, respectively, as before.*

Furthermore, for the Bayesian network defined on this infinite deterministic DAG with $X_{0,0} \sim \mathsf{Bernoulli}\left(\frac{1}{2}\right)$, independent $\mathsf{BSC}(\delta)$ edges, all identity Boolean processing functions in level $k = 1$, and all majority rule Boolean processing functions in levels $k \geq 2$ (as defined in subsection I-C), reconstruction is possible in the sense that:

$$\limsup_{k \to \infty} \mathbb{P}\left(\hat{S}_k \neq X_0\right) < \frac{1}{2}$$

where we use the majority decoder $\hat{S}_k = \mathbb{1}\left\{\sigma_k \geq \frac{1}{2}\right\}$ at level k.

Theorem 3 is proved in section V. The proof of feasibility of reconstruction follows the same overarching strategy as the proof of Theorem 1, but obviously makes essential use of the expansion property (14). We emphasize that Theorem 3 portrays that the constituent expander graphs of a deterministic DAG where broadcasting is possible can be constructed either in quasi-polynomial time or in randomized polylogarithmic time in the number of levels. Once the DAG is constructed however, reconstruction of the root bit is guaranteed to succeed using the majority decoder in the sense presented above. Finally, we note that the question of finding a deterministic polynomial time algorithm to construct DAGs where reconstruction is possible remains open.

C. Further Discussion and Impossibility Results

In this subsection, we present and discuss some impossibility results pertaining to both deterministic and random DAGs. The first result illustrates that if $L_k \leq \log(k)/(d \log(1/(2\delta)))$ for every sufficiently large k (i.e. L_k grows very "slowly"), then reconstruction is impossible regardless of the choices of Boolean processing functions and the choice of decision rule.

Proposition 2 (Slow Growth of Layers). *For any noise level $\delta \in \left(0, \frac{1}{2}\right)$ and indegree $d \in \mathbb{N}\backslash\{0\}$, if the number of vertices per level satisfies $L_k \leq \log(k)/(d \log(1/(2\delta)))$ for all sufficiently large k, then for all choices of Boolean processing functions (which may vary between vertices and be graph dependent), reconstruction is impossible in the sense that:*

1) *for a deterministic DAG:*

$$\lim_{k \to \infty} \left\| P_{X_k}^+ - P_{X_k}^- \right\|_{\mathsf{TV}} = 0$$

   where $P_{X_k}^+$ and $P_{X_k}^-$ denote the conditional distributions of X_k given $X_0 = 1$ and $X_0 = 0$, respectively.

2) *for a random DAG:*

$$\lim_{k \to \infty} \left\| P_{X_k|G}^+ - P_{X_k|G}^- \right\|_{\mathsf{TV}} = 0 \quad pointwise$$

   which means that the condition holds for every realization of the random DAG G.

This proposition is proved in Appendix C. Part 1 of Proposition 2 illustrates that when L_k is sub-logarithmic, the ML decoder based on the entire k-layer state X_k with knowledge of the deterministic DAG fails to reconstruct the root bit. Similarly, part 2 of Proposition 2 shows that reconstruction is impossible for random DAGs even if the particular DAG realization G is known and the ML decoder can access X_k. Therefore, Proposition 2 illustrates that our assumption that $L_k \geq C \log(k)$, for some constant C (that depends on δ and d) and all sufficiently large k, for reconstruction to be possible in Theorems 1 and 2 is in fact necessary.

In contrast, consider a deterministic DAG with no restrictions (i.e. no bounded indegree assumption) except for the size of L_k. Then, each vertex at level k of this DAG is connected to all L_{k-1} vertices at level $k-1$. The next proposition illustrates that $L_k = \Theta\big(\sqrt{\log(k)}\big)$ is the critical scaling of L_k in this scenario. In particular, reconstruction is possible when $L_k = \Omega\big(\sqrt{\log(k)}\big)$ (i.e. $L_k \geq A(\delta)\sqrt{\log(k)}$ for some large constant $A(\delta)$ and all sufficiently large k), and reconstruction is impossible when $L_k = O\big(\sqrt{\log(k)}\big)$ (i.e. $L_k \leq B(\delta)\sqrt{\log(k)}$ for some small constant $B(\delta)$ and all sufficiently large k). The proof of this result is deferred to Appendix D.

Proposition 3 (Broadcasting in Unbounded Degree DAG Model)**.** *Let $A(\delta)$ and $B(\delta)$ be the constants defined in* (111) *and* (112) *in Appendix* D. *Consider a deterministic DAG $\mathcal{G}$ such that for every $k \in \mathbb{N}\backslash\{0\}$, each vertex at level k has one incoming edge from all L_{k-1} vertices at level $k-1$. Then, for any noise level $\delta \in \big(0, \frac{1}{2}\big)$, we have:*

1) *If the number of vertices per level satisfies $L_k \geq A(\delta)\sqrt{\log(k)}$ for all sufficiently large k, and all Boolean processing function in $\mathcal{G}$ are the majority rule (where ties are broken by outputting 1), then reconstruction is possible in the sense that:*

$$\limsup_{k \to \infty} \mathbb{P}(\hat{S}_k \neq X_0) < \frac{1}{2}$$

   where we use the majority decoder $\hat{S}_k = \mathbb{1}\big\{\sigma_k \geq \frac{1}{2}\big\}$ at level k.

2) *If the number of vertices per level satisfies $L_k \leq B(\delta)\sqrt{\log(k)}$ for all sufficiently large k, then for all choices of Boolean processing functions (which may vary between vertices), reconstruction is impossible in the sense that:*

$$\lim_{k \to \infty} \big\|P^+_{X_k} - P^-_{X_k}\big\|_{\mathsf{TV}} = 0\,.$$

The last impossibility result we present here is an important result from the reliable computation literature due to Evans and Schulman [22]. Evans and Schulman studied von Neumann's noisy computation model (which we briefly discussed in subsection II-A), and established general conditions under which reconstruction is impossible in deterministic DAGs due to the decay of mutual information between X_0 and X_k. Recall that for two discrete random variables $X \in \mathcal{X}$ and $Y \in \mathcal{Y}$ (where $|\mathcal{X}|, |\mathcal{Y}| < \infty$), with joint probability mass function $P_{X,Y}$ and marginals P_X and P_Y respectively, the *mutual information* (in bits) between them is defined as:

$$I(X;Y) \triangleq \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} P_{X,Y}(x,y) \log_2\bigg(\frac{P_{X,Y}(x,y)}{P_X(x)P_Y(y)}\bigg) \tag{17}$$

where $\log_2(\cdot)$ is the binary logarithm, and we assume that $0 \log_2\big(\frac{0}{q}\big) = 0$ for any $q \geq 0$, and $p \log_2\big(\frac{p}{0}\big) = \infty$ for any $p > 0$ (due to continuity considerations). We present a specialization of [22, Lemma 2] for our setting as Proposition 4 below. This proposition portrays that if L_k is sub-exponential and the parameters δ and d satisfy $(1 - 2\delta)^2 d < 1$, then reconstruction is impossible in deterministic DAGs regardless of the choices of Boolean processing functions and the choice of decision rule.

Proposition 4 (Decay of Mutual Information [22, Lemma 2])**.** *For any deterministic DAG model, we have:*

$$I(X_0; X_k) \leq L_k \big((1 - 2\delta)^2 d\big)^k$$

where $L_k d^k$ is the total number of paths from X_0 to layer X_k, and $(1-2\delta)^{2k}$ can be construed as the overall contraction of mutual information along each path. Therefore, if $(1-2\delta)^2 d < 1$ and $L_k = o\big(1/((1-2\delta)^2 d)^k\big)$, then for all choices of Boolean processing functions (which may vary between vertices), we have $\lim_{k\to\infty} I(X_0; X_k) = 0$, which implies that $\lim_{k\to\infty} \big\|P_{X_k}^+ - P_{X_k}^-\big\|_{\mathsf{TV}} = 0$.

We make some pertinent remarks about this result. Firstly, Evans and Schulman's original analysis assumes that gates are noisy as opposed to edges (in accordance with von Neumann's setup), but the re-derivation of [22, Lemma 2] in [37, Corollary 7] illustrates that the result also holds for our model. In fact, the *site percolation* analysis in [37, Section 3] (which we will briefly delineate later) improves upon Evans and Schulman's estimate. Furthermore, this analysis illustrates that the bound in Proposition 4 also holds for all choices of random Boolean processing functions.

Secondly, while Proposition 4 holds for deterministic DAGs, we can easily extend it for random DAG models. Indeed, the random DAG model inherits the inequality in Proposition 4 pointwise:

$$I(X_0; X_k | G = \mathcal{G}) \le L_k \left((1-2\delta)^2 d\right)^k \tag{18}$$

for every realization of the random DAG $G = \mathcal{G}$, where $I(X_0; X_k | G = \mathcal{G})$ is the mutual information between X_0 and X_k computed using the joint distribution of X_0 and X_k given $G = \mathcal{G}$. This implies that if L_k is sub-exponential and $(1-2\delta)^2 d < 1$, then reconstruction based on X_k is impossible regardless of the choices of Boolean processing functions (which may vary between vertices and be graph dependent) and the choice of decision rule even if the decoder knows the particular random DAG realization, i.e. $\lim_{k\to\infty} \|P_{X_k|G}^+ - P_{X_k|G}^-\|_{\mathsf{TV}} = 0$ pointwise (which trivially implies (7)). Taking expectations with respect to G in (18), we get:

$$I(X_0; X_k) \le I(X_0; X_k | G) \le L_k \left((1-2\delta)^2 d\right)^k \tag{19}$$

where $I(X_0; X_k | G)$ is the conditional mutual information (i.e. the expected value of $I(X_0; X_k | G = \mathcal{G})$ with respect to G), and the first inequality follows from the chain rule for mutual information and the fact that X_0 is independent of G. Since the second inequality in (19) implies (41), invoking the argument at the end of the proof of part 2 of Theorem 1 in section III also yields that reconstruction is impossible in the sense of (7) when L_k is sub-exponential and $(1-2\delta)^2 d < 1$. Thus, $\lim_{k\to\infty} I(X_0; X_k | G) = 0$ is a sufficient condition for (7). In contrast, the first inequality in (19) only yields the impossibility of reconstruction in the sense of (8) when L_k is sub-exponential and $(1-2\delta)^2 d < 1$.

Thirdly, Evans and Schulman's result in Proposition 4 provides an upper bound on the critical threshold of δ above which reconstruction of the root bit is impossible. Indeed, the condition, $(1-2\delta)^2 d < 1$, under which mutual information decays can be rewritten as (cf. the discussion in [22, p.2373]):

$$\delta_{\mathsf{ES}}(d) \triangleq \frac{1}{2} - \frac{1}{2\sqrt{d}} < \delta < \frac{1}{2} \tag{20}$$

and reconstruction is impossible for deterministic or random DAGs in this regime of δ provided L_k is sub-exponential. As a sanity check, we can verify that $\delta_{\mathsf{ES}}(2) = 0.14644... > 0.08856... = \delta_{\mathsf{andor}}$ in the context of Theorem 2, and $\delta_{\mathsf{ES}}(3) = 0.21132... > 0.16666... = \delta_{\mathsf{maj}}$ in the context of Theorem 1 with $d = 3$. Although

$\delta_{\mathsf{ES}}(d)$ is a general upper bound on the critical threshold for reconstruction, in this paper, it is not particularly useful because we analyze explicit processing functions and decision rules, and derive specific bounds that characterize the corresponding thresholds.

Fourthly, it is worth comparing $\delta_{\mathsf{ES}}(d)$ (which comes from a site percolation argument, cf. [37, Section 3]) to an upper bound on the critical threshold for reconstruction derived from *bond percolation*. To this end, consider the random DAG model, and recall that the $\mathsf{BSC}(\delta)$'s along each edge generate independent bits with probability 2δ (as shown in the proof of Proposition 2 in Appendix C). So, we can perform bond percolation so that each edge is independently "removed" with probability 2δ. It can be shown by analyzing this bond percolation process that reconstruction is impossible (in a certain sense) when $\frac{1}{2} - \frac{1}{2d} < \delta < \frac{1}{2}$. Therefore, the Evans-Schulman upper bound of $\delta_{\mathsf{ES}}(d)$ is tighter than the bond percolation upper bound: $\delta_{\mathsf{ES}}(d) < \frac{1}{2} - \frac{1}{2d}$.

Finally, we briefly delineate how the site percolation approach in [37, Section 3] allows us to prove that reconstruction is impossible in the random DAG model for the $(1-2\delta)^2 d = 1$ case as well. Consider a site percolation process where each vertex $X_{k,j}$ (for $k \in \mathbb{N}\backslash\{0\}$ and $j \in [L_k]$) is independently "open" with probability $(1-2\delta)^2$, and "closed" with probability $1-(1-2\delta)^2$. (Note that $X_{0,0}$ is open almost surely.) For every $k \in \mathbb{N}\backslash\{0\}$, let p_k denote the probability that there is an "open connected path" from X_0 to X_k (i.e. there exist $j_1 \in [L_1], \ldots, j_k \in [L_k]$ such that $(X_{0,0}, X_{1,j_1}), (X_{1,j_1}, X_{2,j_2}), \ldots, (X_{k-1,j_{k-1}}, X_{k,j_k})$ are directed edges in the random DAG G and $X_{1,j_1}, \ldots, X_{k,j_k}$ are all open). It can be deduced from [37, Theorem 5] that for any $k \in \mathbb{N}\backslash\{0\}$:

$$I(X_0; X_k | G) \leq p_k. \tag{21}$$

Next, for each $k \in \mathbb{N}$, define the random variable:

$$\lambda_k \triangleq \frac{1}{L_k} \sum_{j \in [L_k]} \mathbb{1}\{X_{k,j} \text{ is open and connected}\} \tag{22}$$

which is the proportion of open vertices at level k that are connected to the root by an open path. (Note that $\lambda_0 = 1$.) It is straightforward to verify (using Bernoulli's inequality) that for any $k \in \mathbb{N}\backslash\{0\}$:

$$\mathbb{E}[\lambda_k | \lambda_{k-1}] = (1 - 2\delta)^2 \big(1 - (1 - \lambda_{k-1})^d \big) \leq (1 - 2\delta)^2 d \lambda_{k-1}. \tag{23}$$

Observe that by Markov's inequality and the recursion from (23), $\mathbb{E}[\lambda_k] \leq (1 - 2\delta)^2 d \, \mathbb{E}[\lambda_{k-1}]$, we have:

$$p_k = \mathbb{P}\left(\lambda_k \geq \frac{1}{L_k} \right) \leq L_k \mathbb{E}[\lambda_k] \leq L_k \big((1 - 2\delta)^2 d \big)^k \tag{24}$$

which recovers Evans and Schulman's result (Proposition 4) in the context of the random DAG model. Indeed, if $(1-2\delta)^2 d < 1$ and $L_k = o\big(1/((1 - 2\delta)^2 d)^k \big)$, then $\lim_{k \to \infty} p_k = 0$, and as a result, $\lim_{k \to \infty} I(X_0; X_k | G) = 0$ by (21). On the other hand, when $(1-2\delta)^2 d = 1$, taking expectations and applying Jensen's inequality to the equality in (23) produces:

$$\mathbb{E}[\lambda_k] \leq (1 - 2\delta)^2 \big(1 - (1 - \mathbb{E}[\lambda_{k-1}])^d \big). \tag{25}$$

This implies that $\mathbb{E}[\lambda_k] \leq F^{-1}(k)$ for every $k \in \mathbb{N}$ using the estimate in [38, Appendix A], where $F : [0,1] \to$
$\mathbb{R}_+$, $F(t) = \int_t^1 \frac{1}{f(\tau)} \, d\tau$ with $f : [0,1] \to [0,1]$, $f(t) = t - (1-2\delta)^2 \big(1 - (1-t)^d\big)$, and $F^{-1} : \mathbb{R}_+ \to [0,1]$ is
well-defined. Since $f(t) \geq \frac{d-1}{2} t^2$ for all $t \in [0,1]$, it is straightforward to show that:

$$\mathbb{E}[\lambda_k] \leq F^{-1}(k) \leq \frac{2}{(d-1)k} \, . \tag{26}$$

Therefore, the Markov's inequality argument in (24) illustrates that if $(1-2\delta)^2 d = 1$ and $L_k = o(k)$, then
$\lim_{k \to \infty} p_k = 0$ and reconstruction is impossible in the random DAG model due to (21). Furthermore, the condition
on L_k can be improved to $L_k = O(k \log(k))$ using a more sophisticated Borel-Cantelli type of argument.

III. ANALYSIS OF MAJORITY RULE PROCESSING IN RANDOM DAG MODEL

In this section, we prove Theorem 1. To this end, we first make some pertinent observations. Recall that we
have a random DAG model with $d \geq 3$, and all Boolean functions are the majority rule, i.e. $f_k(x_1, \ldots, x_d) =$
$\mathsf{majority}(x_1, \ldots, x_d)$ for every $k \in \mathbb{N}\backslash\{0\}$. Note that when the number of 1's is equal to the number of 0's, the
majority rule outputs an independent Bernoulli$\big(\frac{1}{2}\big)$ bit.[10] Suppose we are given that $\sigma_{k-1} = \sigma$ for any $k \in \mathbb{N}\backslash\{0\}$.
Then, for every $j \in [L_k]$, $X_{k,j} = \mathsf{majority}(Y_1, \ldots, Y_d)$ where $Y_1, \ldots, Y_d$ are i.i.d. Bernoulli(p) random variables
with $p = \sigma * \delta$. Define the function $g : [0,1] \to [0,1]$ as follows:

$$g(\sigma) \triangleq \mathbb{E}[\mathsf{majority}(Y_1, \ldots, Y_d)] = \mathbb{P}\left(\sum_{i=1}^d Y_i > \frac{d}{2}\right) + \frac{1}{2}\mathbb{P}\left(\sum_{i=1}^d Y_i = \frac{d}{2}\right) \tag{27}$$

$$= \begin{cases} \displaystyle\sum_{i=\frac{d}{2}+1}^d \binom{d}{i}(\sigma*\delta)^i(1-\sigma*\delta)^{d-i} + \frac{1}{2}\binom{d}{\frac{d}{2}}(\sigma*\delta)^{\frac{d}{2}}(1-\sigma*\delta)^{\frac{d}{2}} & , \ d \text{ even} \\[4mm] \displaystyle\sum_{i=\frac{d+1}{2}}^d \binom{d}{i}\underbrace{(\sigma*\delta)}_{p}{}^i\underbrace{(1-\sigma*\delta)}_{1-p}{}^{d-i} & , \ d \text{ odd} \end{cases} \tag{28}$$

which implies that $X_{k,j}$ are i.i.d. Bernoulli$(g(\sigma))$ for $j \in [L_k]$, and $L_k \sigma_k \sim \mathsf{binomial}(L_k, g(\sigma))$, since we have:

$$\mathbb{P}(X_{k,j} = 1 | \sigma_{k-1} = \sigma) = \mathbb{E}[\sigma_k | \sigma_{k-1} = \sigma] = g(\sigma) \, . \tag{29}$$

To compute the first derivative of g, we follow the analysis in [6, Section 2]. Recall that a Boolean function $h :$
$\{0,1\}^d \to \{0,1\}$ is *monotone non-decreasing* (respectively, *non-increasing*) if its value either increases (respectively,
decreases) or remains the same whenever any of its input bits is flipped from 0 to 1. For any such monotone function
$h : \{0,1\}^d \to \{0,1\}$, the *Margulis-Russo formula* states that [39], [40] (alternatively, see [41, Section 4.1]):

$$\frac{d}{dp}\mathbb{E}[h(Y_1, \ldots, Y_d)] = \sum_{i=1}^d \mathbb{E}[h(Y_1, \ldots, Y_{i-1}, 1, Y_{i+1}, \ldots, Y_d) - h(Y_1, \ldots, Y_{i-1}, 0, Y_{i+1}, \ldots, Y_d)]. \tag{30}$$

---

[10]Although generating a random bit is a natural approach to breaking ties in the majority rule, this means that the rule is no longer purely
deterministic when d is even.

Hence, since $h = \text{majority}$ is a non-decreasing function, $g' : [0, 1] \to \mathbb{R}^+$ is given by:

$$
\begin{aligned}
g'(\sigma) &= \frac{dp}{d\sigma} \frac{d}{dp} \mathbb{E}[h(Y_1, \ldots, Y_d)] \\
&= (1 - 2\delta) \sum_{i=1}^{d} \mathbb{E}[h(Y_1, \ldots, Y_{i-1}, 1, Y_{i+1}, \ldots, Y_d) - h(Y_1, \ldots, Y_{i-1}, 0, Y_{i+1}, \ldots, Y_d)] \\
&= (1 - 2\delta) \, d \, \mathbb{E}[h(1, Y_2, \ldots, Y_d) - h(0, Y_2, \ldots, Y_d)] \\
&= (1 - 2\delta) \, d \, \mathbb{P}(h(1, Y_2, \ldots, Y_d) = 1, \, h(0, Y_2, \ldots, Y_d) = 0)
\end{aligned}
\tag{31}
$$

$$
= \begin{cases}
(1 - 2\delta) \dfrac{d}{2} \left(\mathbb{P}\left(\displaystyle\sum_{i=2}^{d} Y_i = \dfrac{d}{2} - 1 \right) + \mathbb{P}\left(\displaystyle\sum_{i=2}^{d} Y_i = \dfrac{d}{2} \right) \right) & , \; d \text{ even} \\[4ex]
(1 - 2\delta) \, d \, \mathbb{P}\left(\displaystyle\sum_{i=2}^{d} Y_i = \dfrac{d-1}{2} \right) & , \; d \text{ odd}
\end{cases}
$$

$$
= \begin{cases}
(1 - 2\delta) \dfrac{d}{2} \left(\dbinom{d-1}{\frac{d}{2}-1} p^{\frac{d}{2}-1}(1-p)^{\frac{d}{2}} + \dbinom{d-1}{\frac{d}{2}} p^{\frac{d}{2}}(1-p)^{\frac{d}{2}-1} \right) & , \; d \text{ even} \\[4ex]
(1 - 2\delta) \, d \dbinom{d-1}{\frac{d-1}{2}} p^{\frac{d-1}{2}}(1-p)^{\frac{d-1}{2}} & , \; d \text{ odd}
\end{cases}
$$

$$
= \begin{cases}
(1 - 2\delta) \dfrac{d}{4} \dbinom{d}{\frac{d}{2}} (p(1-p))^{\frac{d}{2}-1} & , \; d \text{ even} \\[4ex]
(1 - 2\delta) \dfrac{d+1}{2} \dbinom{d}{\frac{d+1}{2}} (p(1-p))^{\frac{d-1}{2}} & , \; d \text{ odd}
\end{cases}
$$

$$
= \begin{cases}
(1 - 2\delta) \dfrac{d}{4} \dbinom{d}{\frac{d}{2}} ((\sigma * \delta)(1 - \sigma * \delta))^{\frac{d}{2}-1} & , \; d \text{ even} \\[4ex]
(1 - 2\delta) \dfrac{d+1}{2} \dbinom{d}{\frac{d+1}{2}} ((\sigma * \delta)(1 - \sigma * \delta))^{\frac{d-1}{2}} & , \; d \text{ odd}
\end{cases}
\tag{32}
$$

where the second equality follows from $dp/d\sigma = 1 - 2\delta$ and (30), the third equality holds because $h = \text{majority}$ is symmetric in its input bits, the fourth equality holds because $h = \text{majority}$ is non-decreasing, and the fifth equality follows from the definition of the majority rule. Since $p \mapsto p(1 - p)$ is increasing on $\left[0, \frac{1}{2} \right]$ and decreasing on $\left[\frac{1}{2}, 1 \right]$, and $p = \sigma * \delta$ is linear in σ with derivative $1 - 2\delta > 0$ such that $p = \frac{1}{2}$ when $\sigma = \frac{1}{2}$, it is straightforward to verify from (32) that g' is positive on $[0, 1]$, increasing on $\left[0, \frac{1}{2} \right]$, and decreasing on $\left[\frac{1}{2}, 1 \right]$. As a result, g is increasing on $[0, 1]$, convex on $\left[0, \frac{1}{2} \right]$, and concave on $\left[\frac{1}{2}, 1 \right]$. Furthermore, the Lipschitz constant of g over $[0, 1]$, or equivalently, the maximum value of g' over $[0, 1]$ is:

$$
D(\delta, d) \triangleq \max_{\sigma \in [0,1]} g'(\sigma) = g'\left(\frac{1}{2} \right) = (1 - 2\delta) \left(\frac{1}{2} \right)^{d-1} \left\lceil \frac{d}{2} \right\rceil \binom{d}{\lceil \frac{d}{2} \rceil}
\tag{33}
$$

regardless of whether d is even or odd.

There are two regimes of interest when we consider the contraction properties and fixed point structure of g. As defined in (9), let δ_{maj} be the critical noise level such that the Lipschitz constant $g'\left(\frac{1}{2} \right)$ is equal to 1.[11] Then, in the $\delta \in (0, \delta_{\text{maj}})$ regime, the Lipschitz constant $g'\left(\frac{1}{2} \right)$ is greater than 1. Furthermore, since $g\left(\frac{1}{2} \right) = \frac{1}{2}$ and $g(1 - \sigma) = 1 - g(\sigma)$ (which are straightforward to verify from (28)), the aforementioned properties of g imply

---

[11] We can also view δ_{maj} as the critical value such that the d-input majority gate with independent $\text{BSC}(\delta)$'s at each input is an *amplifier* if and only if $\delta < \delta_{\text{maj}}$. We refer readers to [42] for more information about amplifiers, and in particular, the relationship between amplifiers and reliable computation.

that g has three fixed points at $\sigma = 1 - \hat{\sigma}, \frac{1}{2}, \hat{\sigma}$, where the largest fixed point of g is some $\hat{\sigma} \in \left(\frac{1}{2}, 1\right)$ that depends on δ (e.g. $\hat{\sigma} = \left(1 + \sqrt{(1 - 6\delta)/(1 - 2\delta)^3}\right)/2$ when $d = 3$). In contrast, in the $\delta \in \left(\delta_{\mathsf{maj}}, \frac{1}{2}\right)$ regime, the Lipschitz constant $g'\left(\frac{1}{2}\right)$ is less than 1, and the only fixed point of g is $\sigma = \frac{1}{2}$. (We also mention that when $\delta = \delta_{\mathsf{maj}}$, g has only one fixed point at $\sigma = \frac{1}{2}$.)

Using these observations, we now prove Theorem 1.

Proof of Theorem 1. We begin by constructing a useful "monotone Markovian coupling" that will help establish both achievability and converse directions (see [43, Chapter 5] for basic definitions of Markovian couplings). Let $\{X_k^+ : k \in \mathbb{N}\}$ and $\{X_k^- : k \in \mathbb{N}\}$ denote versions of the Markov chain $\{X_k : k \in \mathbb{N}\}$ (i.e. with the same transition kernels) initialized at $X_0^+ = 1$ and $X_0^- = 0$, respectively. In particular, the marginal distributions of X_k^+ and X_k^- are $P_{X_k}^+$ and $P_{X_k}^-$, respectively. The monotone Markovian coupling $\{(X_k^-, X_k^+) : k \in \mathbb{N}\}$ between the Markov chains $\{X_k^+ : k \in \mathbb{N}\}$ and $\{X_k^- : k \in \mathbb{N}\}$ is generated as follows. First, condition on any random DAG realization $G = \mathcal{G}$. Recall that each edge $\mathsf{BSC}(\delta)$ of $\mathcal{G}$ either copies its input bit with probability $1 - 2\delta$, or produces an independent $\mathsf{Bernoulli}\left(\frac{1}{2}\right)$ bit with probability 2δ (as demonstrated in the proof of Proposition 2 in Appendix C). Next, couple $\{X_k^+ : k \in \mathbb{N}\}$ and $\{X_k^- : k \in \mathbb{N}\}$ so that along any edge BSC of $\mathcal{G}$, say $(X_{k,j}, X_{k+1,i})$, $X_{k,j}^+$ and $X_{k,j}^-$ are either both copied with probability $1 - 2\delta$, or a shared independent $\mathsf{Bernoulli}\left(\frac{1}{2}\right)$ bit is produced with probability 2δ that becomes the value of both $X_{k+1,i}^+$ and $X_{k+1,i}^-$. In other words, $\{X_k^+ : k \in \mathbb{N}\}$ and $\{X_k^- : k \in \mathbb{N}\}$ "run" on the same underlying DAG $\mathcal{G}$ and have common BSCs. Hence, after averaging over all realizations of G, it is straightforward to verify that the Markovian coupling $\{(X_k^-, X_k^+) : k \in \mathbb{N}\}$ has the following properties:

1) The "marginal" Markov chains are $\{X_k^+ : k \in \mathbb{N}\}$ and $\{X_k^- : k \in \mathbb{N}\}$.
2) For every $k \in \mathbb{N}$, X_{k+1}^+ is conditionally independent of X_k^- given X_k^+, and X_{k+1}^- is conditionally independent of X_k^+ given X_k^-.
3) For every $k \in \mathbb{N}$ and every $j \in [L_k]$, $X_{k,j}^+ \geq X_{k,j}^-$ almost surely—this is the monotonicity property of the coupling.

In particular, the third property holds because $1 = X_{0,0}^+ \geq X_{0,0}^- = 0$ is true by assumption, each edge BSC preserves monotonicity (whether it copies its input or generates a new shared bit), and the majority processing functions are symmetric and monotone non-decreasing. In the sequel, probabilities of events that depend on the coupled vertex random variables $\{(X_{k,j}^-, X_{k,j}^+) : k \in \mathbb{N}, j \in [L_k]\}$ are defined with respect to this Markovian coupling. Note that this coupling also induces a monotone Markovian coupling $\{(\sigma_k^+, \sigma_k^-) : k \in \mathbb{N}\}$ between the Markov chains $\{\sigma_k^+ : k \in \mathbb{N}\}$ and $\{\sigma_k^- : k \in \mathbb{N}\}$ (where $\{\sigma_k^+ : k \in \mathbb{N}\}$ and $\{\sigma_k^- : k \in \mathbb{N}\}$ denote versions of the Markov chain $\{\sigma_k : k \in \mathbb{N}\}$ initialized at $\sigma_0^+ = 1$ and $\sigma_0^- = 0$, respectively) such that:

1) The "marginal" Markov chains are $\{\sigma_k^+ : k \in \mathbb{N}\}$ and $\{\sigma_k^- : k \in \mathbb{N}\}$.
2) For every $j > k \geq 1$, σ_j^+ is conditionally independent of $\sigma_0^-, \ldots, \sigma_k^-, \sigma_0^+, \ldots, \sigma_{k-1}^+$ given σ_k^+, and σ_j^- is conditionally independent of $\sigma_0^+, \ldots, \sigma_k^+, \sigma_0^-, \ldots, \sigma_{k-1}^-$ given σ_k^-.
3) For every $k \in \mathbb{N}$, $\sigma_k^+ \geq \sigma_k^-$ almost surely.

Part 1: We first prove that $\delta \in (0, \delta_{\mathsf{maj}})$ implies $\limsup_{k \to \infty} \mathbb{P}(\hat{S}_k \neq \sigma_0) < \frac{1}{2}$. To this end, we start by showing

that there exists $\epsilon = \epsilon(\delta, d) > 0$ (that depends on δ and d such that:

$$\forall k \in \mathbb{N}\backslash\{0\}, \ \mathbb{P}\big(\sigma_k^+ \geq \hat{\sigma} - \epsilon \,\big|\, \sigma_{k-1}^+ \geq \hat{\sigma} - \epsilon, A_{k,j}\big) \geq 1 - \exp\big(-2L_k \gamma(\epsilon)^2\big) \tag{34}$$

where $\gamma(\epsilon) \triangleq g(\hat{\sigma} - \epsilon) - (\hat{\sigma} - \epsilon) > 0$, and $A_{k,j}$ with $0 \leq j < k$ is the non-zero probability event defined as:

$$A_{k,j} \triangleq \begin{cases} \{\sigma_j^- \leq 1 - \hat{\sigma} + \epsilon\} & , \quad 0 \leq j = k - 1 \\ \{\sigma_{k-2}^+ \geq \hat{\sigma} - \epsilon, \ldots, \sigma_j^+ \geq \hat{\sigma} - \epsilon\} \cap \{\sigma_j^- \leq 1 - \hat{\sigma} + \epsilon\} & , \quad 0 \leq j \leq k - 2 \end{cases}.$$

Since $g'(\hat{\sigma}) < 1$ and $g(\hat{\sigma}) = \hat{\sigma}$, $g(\hat{\sigma} - \epsilon) > \hat{\sigma} - \epsilon$ for sufficiently small $\epsilon > 0$. Fix any such $\epsilon > 0$ (which depends on δ and d because g depends on δ and d such that $\gamma(\epsilon) > 0$. Recall that $L_k \sigma_k \sim \text{binomial}(L_k, g(\sigma))$ given $\sigma_{k-1} = \sigma$. This implies that for every $k \in \mathbb{N}\backslash\{0\}$ and every $0 \leq j < k$:

$$\mathbb{P}\big(\sigma_k^+ < g\big(\sigma_{k-1}^+\big) - \gamma(\epsilon) \,\big|\, \sigma_{k-1}^+ = \sigma, A_{k,j}\big) = \mathbb{P}(\sigma_k < g(\sigma_{k-1}) - \gamma(\epsilon)|\sigma_{k-1} = \sigma) \leq \exp\big(-2L_k \gamma(\epsilon)^2\big)$$

where the equality follows from property 2 of our Markovian coupling $\{(\sigma_k^+, \sigma_k^-) : k \in \mathbb{N}\}$, and the inequality follows from (29) and Hoeffding's inequality [44, Theorem 1]. As a result, we have:

$$\sum_{\sigma \geq \hat{\sigma} - \epsilon} \mathbb{P}\big(\sigma_k^+ < g\big(\sigma_{k-1}^+\big) - \gamma(\epsilon) \,\big|\, \sigma_{k-1}^+ = \sigma, A_{k,j}\big) \mathbb{P}\big(\sigma_{k-1}^+ = \sigma \,\big|\, A_{k,j}\big)$$

$$\leq \exp\big(-2L_k \gamma(\epsilon)^2\big) \sum_{\sigma \geq \hat{\sigma} - \epsilon} \mathbb{P}\big(\sigma_{k-1}^+ = \sigma \,\big|\, A_{k,j}\big)$$

$$\mathbb{P}\big(\sigma_k^+ < g\big(\sigma_{k-1}^+\big) - \gamma(\epsilon), \sigma_{k-1}^+ \geq \hat{\sigma} - \epsilon \,\big|\, A_{k,j}\big) \leq \exp\big(-2L_k \gamma(\epsilon)^2\big) \mathbb{P}\big(\sigma_{k-1}^+ \geq \hat{\sigma} - \epsilon \,\big|\, A_{k,j}\big)$$

$$\mathbb{P}\big(\sigma_k^+ < g\big(\sigma_{k-1}^+\big) - \gamma(\epsilon) \,\big|\, \sigma_{k-1}^+ \geq \hat{\sigma} - \epsilon, A_{k,j}\big) \leq \exp\big(-2L_k \gamma(\epsilon)^2\big) \,.$$

Finally, notice that $\sigma_k^+ < \hat{\sigma} - \epsilon = g(\hat{\sigma} - \epsilon) - \gamma(\epsilon)$ implies that $\sigma_k^+ < g(\sigma_{k-1}^+) - \gamma(\epsilon)$ when $\sigma_{k-1}^+ \geq \hat{\sigma} - \epsilon$ (since g is non-decreasing and $g(\sigma_{k-1}^+) \geq g(\hat{\sigma} - \epsilon)$). This produces:

$$\mathbb{P}\big(\sigma_k^+ < \hat{\sigma} - \epsilon \,\big|\, \sigma_{k-1}^+ \geq \hat{\sigma} - \epsilon, A_{k,j}\big) \leq \exp\big(-2L_k \gamma(\epsilon)^2\big)$$

which in turn establishes (34).

Now fix any $\tau > 0$, and choose a sufficiently large value $K = K(\epsilon, \tau) \in \mathbb{N}$ (that depends on ϵ and τ) such that:

$$\sum_{m=K+1}^{\infty} \exp\big(-2L_m \gamma(\epsilon)^2\big) \leq \tau \,. \tag{35}$$

Note that such K exists because $\sum_{m=1}^{\infty} 1/m^2 = \pi^2/6 < +\infty$, and for all sufficiently large m (depending on δ and d), we have:

$$\exp\big(-2L_m \gamma(\epsilon)^2\big) \leq \frac{1}{m^2} \quad \Leftrightarrow \quad L_m \geq \frac{\log(m)}{\gamma(\epsilon)^2} \,. \tag{36}$$

In (36), we use the assumption that $L_m \geq C(\delta, d) \log(m)$ for all sufficiently large m (depending on δ and d), where we define the constant $C(\delta, d)$ as:

$$C(\delta, d) \triangleq \frac{1}{\gamma(\epsilon(\delta, d))^2} > 0 \,. \tag{37}$$

Using the continuity of probability measures, observe that:

$$
\mathbb{P}\left(\bigcap_{k>K} \{\sigma_k^+ \geq \hat{\sigma} - \epsilon\} \,\Big|\, \sigma_K^+ \geq \hat{\sigma} - \epsilon, \sigma_K^- \leq 1 - \hat{\sigma} + \epsilon \right) = \prod_{k>K} \mathbb{P}\big(\sigma_k^+ \geq \hat{\sigma} - \epsilon \,\big|\, \sigma_{k-1}^+ \geq \hat{\sigma} - \epsilon, A_{k,K}\big)
$$

$$
\geq \prod_{k>K} 1 - \exp\big(-2L_k \gamma(\epsilon)^2\big)
$$

$$
\geq 1 - \sum_{k>K} \exp\big(-2L_k \gamma(\epsilon)^2\big)
$$

$$
\geq 1 - \tau
$$

where the first inequality follows from (34), the second inequality is straightforward to establish using induction, and the final inequality follows from (35). Therefore, we have for any $k > K$:

$$
\mathbb{P}\big(\sigma_k^+ \geq \hat{\sigma} - \epsilon \,\big|\, \sigma_K^+ \geq \hat{\sigma} - \epsilon, \sigma_K^- \leq 1 - \hat{\sigma} + \epsilon\big) \geq 1 - \tau. \tag{38}
$$

Likewise, we can also prove mutatis mutandis that for any $k > K$:

$$
\mathbb{P}\big(\sigma_k^- \leq 1 - \hat{\sigma} + \epsilon \,\big|\, \sigma_K^+ \geq \hat{\sigma} - \epsilon, \sigma_K^- \leq 1 - \hat{\sigma} + \epsilon\big) \geq 1 - \tau \tag{39}
$$

where the choices of ϵ, τ, and K in (39) are the same as those in (38) without loss of generality.

We need to show that $\limsup_{k\to\infty} \mathbb{P}(\hat{S}_k \neq \sigma_0) < \frac{1}{2}$, or equivalently, that there exists $\lambda > 0$ such that for all sufficiently large $k \in \mathbb{N}$:

$$
\mathbb{P}\big(\hat{S}_k \neq \sigma_0\big) = \frac{1}{2}\mathbb{P}\big(\hat{S}_k \neq \sigma_0 \,\big|\, \sigma_0 = 1\big) + \frac{1}{2}\mathbb{P}\big(\hat{S}_k \neq \sigma_0 \,\big|\, \sigma_0 = 0\big) \leq \frac{1-\lambda}{2}
$$

$$
\Leftrightarrow \quad \mathbb{P}\left(\sigma_k < \frac{1}{2} \,\bigg|\, \sigma_0 = 1\right) + \mathbb{P}\left(\sigma_k \geq \frac{1}{2} \,\bigg|\, \sigma_0 = 0\right) \leq 1 - \lambda
$$

$$
\Leftrightarrow \quad \mathbb{P}\left(\sigma_k^+ \geq \frac{1}{2}\right) - \mathbb{P}\left(\sigma_k^- \geq \frac{1}{2}\right) \geq \lambda.
$$

To this end, let $E = \big\{\sigma_K^+ \geq \hat{\sigma} - \epsilon, \sigma_K^- \leq 1 - \hat{\sigma} + \epsilon\big\}$, and observe that for all $k > K$:

$$
\mathbb{P}\left(\sigma_k^+ \geq \frac{1}{2}\right) - \mathbb{P}\left(\sigma_k^- \geq \frac{1}{2}\right) = \mathbb{E}\left[\mathbb{1}\left\{\sigma_k^+ \geq \frac{1}{2}\right\} - \mathbb{1}\left\{\sigma_k^- \geq \frac{1}{2}\right\}\right]
$$

$$
\geq \mathbb{E}\left[\left(\mathbb{1}\left\{\sigma_k^+ \geq \frac{1}{2}\right\} - \mathbb{1}\left\{\sigma_k^- \geq \frac{1}{2}\right\}\right)\mathbb{1}\{E\}\right]
$$

$$
= \mathbb{E}\left[\mathbb{1}\left\{\sigma_k^+ \geq \frac{1}{2}\right\} - \mathbb{1}\left\{\sigma_k^- \geq \frac{1}{2}\right\} \,\bigg|\, E\right]\mathbb{P}(E)
$$

$$
= \left(\mathbb{P}\left(\sigma_k^+ \geq \frac{1}{2} \,\bigg|\, E\right) - \mathbb{P}\left(\sigma_k^- \geq \frac{1}{2} \,\bigg|\, E\right)\right)\mathbb{P}(E)
$$

$$
\geq \big(\mathbb{P}\big(\sigma_k^+ \geq \hat{\sigma} - \epsilon \,\big|\, E\big) - \mathbb{P}\big(\sigma_k^- > 1 - \hat{\sigma} + \epsilon \,\big|\, E\big)\big)\mathbb{P}(E)
$$

$$
\geq (1 - 2\tau)\mathbb{P}(E) \triangleq \lambda > 0
$$

where the first inequality holds because $\mathbb{1}\{\sigma_k^+ \geq \frac{1}{2}\} - \mathbb{1}\{\sigma_k^- \geq \frac{1}{2}\} \geq 0$ almost surely due to the monotonicity (property 3) of the Markovian coupling $\{(\sigma_k^+, \sigma_k^-) : k \in \mathbb{N}\}$, the second inequality holds because $1 - \hat{\sigma} + \epsilon < \frac{1}{2} < \hat{\sigma} - \epsilon$ (since $\epsilon > 0$ is small), and the final inequality follows from (38) and (39). This completes the proof for the $\delta \in (0, \delta_{\mathsf{maj}})$ regime.

Part 2: We next prove that $\delta \in \left(\delta_{\mathsf{maj}}, \frac{1}{2}\right)$ implies (7). First, notice that conditioned on any realization of the random DAG G, we have $X_{k,j}^+ \geq X_{k,j}^-$ almost surely for every $k \in \mathbb{N}$ and $j \in [L_k]$ (by construction of our coupling). Hence, conditioned on G, we obtain:

$$\left\|P_{X_k|G}^+ - P_{X_k|G}^-\right\|_{\mathsf{TV}} \leq \mathbb{P}\big(X_k^+ \neq X_k^- \big| G\big)$$

$$= \mathbb{P}\Big(\exists j \in [L_k], \, X_{k,j}^+ \neq X_{k,j}^- \Big| G\Big)$$

$$\leq \sum_{j=0}^{L_k-1} \mathbb{P}\Big(X_{k,j}^+ \neq X_{k,j}^- \Big| G\Big)$$

$$= \mathbb{E}\left[\left|\sum_{j=0}^{L_k-1} X_{k,j}^+ - X_{k,j}^-\right| \Bigg| G\right]$$

$$= L_k \, \mathbb{E}\big[\sigma_k^+ - \sigma_k^- \big| G\big]$$

where the first inequality follows from Dobrushin's maximal coupling representation of TV distance [43, Chapter 4.2], the third inequality follows from the union bound, and the fourth equality holds because $\mathbb{P}(X_{k,j}^+ \neq X_{k,j}^-|G) = \mathbb{P}(X_{k,j}^+ - X_{k,j}^- = 1|G) = \mathbb{E}[X_{k,j}^+ - X_{k,j}^-|G]$ due to the monotonicity of our coupling. Then, taking expectations with respect to G yields:

$$\mathbb{E}\left[\left\|P_{X_k|G}^+ - P_{X_k|G}^-\right\|_{\mathsf{TV}}\right] \leq L_k \, \mathbb{E}\big[\sigma_k^+ - \sigma_k^-\big]. \tag{40}$$

We can bound $\mathbb{E}\big[\sigma_k^+ - \sigma_k^-\big]$ as follows. Firstly, we use the Lipschitz continuity of g (with Lipschitz constant $D(\delta, d)$) and the monotonicity of our coupling to get:

$$0 \leq \mathbb{E}\big[\sigma_k^+ - \sigma_k^- \big| \sigma_{k-1}^+, \sigma_{k-1}^-\big] = g\big(\sigma_{k-1}^+\big) - g\big(\sigma_{k-1}^-\big) \leq D(\delta, d) \left(\sigma_{k-1}^+ - \sigma_{k-1}^-\right).$$

Then, we can take expectations with respect to $\big(\sigma_{k-1}^+, \sigma_{k-1}^-\big)$ on both sides of this inequality (and use the tower property on the left hand side) to obtain:

$$0 \leq \mathbb{E}\big[\sigma_k^+ - \sigma_k^-\big] \leq D(\delta, d) \, \mathbb{E}\big[\sigma_{k-1}^+ - \sigma_{k-1}^-\big].$$

Therefore, we recursively have:

$$0 \leq \mathbb{E}\big[\sigma_k^+ - \sigma_k^-\big] \leq D(\delta, d)^k$$

where we use the fact that $\mathbb{E}\big[\sigma_0^+ - \sigma_0^-\big] = 1$. Using (40) with this bound, we get:

$$\mathbb{E}\left[\left\|P_{X_k|G}^+ - P_{X_k|G}^-\right\|_{\mathsf{TV}}\right] \leq L_k D(\delta, d)^k$$

where letting $k \to \infty$ yields:

$$\lim_{k\to\infty} \mathbb{E}\left[\left\|P_{X_k|G}^+ - P_{X_k|G}^-\right\|_{\mathsf{TV}}\right] = 0 \tag{41}$$

because $L_k = o(D(\delta, d)^{-k})$ by assumption. (It is worth mentioning that although $L_k = o(D(\delta, d)^{-k})$ in this regime, it can diverge to infinity because the Lipschitz constant $D(\delta, d) < 1$.)

Finally, observe that $\|P_{X_k|G}^+ - P_{X_k|G}^-\|_{\mathsf{TV}} \in [0, 1]$ forms a non-increasing sequence in k for every realization of the random DAG G (since $\{X_k : k \in \mathbb{N}\}$ forms a Markov chain given G, and the data processing inequality for TV distance yields the desired monotonicity). Hence, the pointwise limit (over realizations of G) random variable,

$\lim_{k\to\infty}\|P^+_{X_k|G} - P^-_{X_k|G}\|_{\mathsf{TV}} \in [0,1]$, has mean $\mathbb{E}[\lim_{k\to\infty}\|P^+_{X_k|G} - P^-_{X_k|G}\|_{\mathsf{TV}}] = 0$ due to (41) and the bounded convergence theorem. Since a non-negative random variable that has zero mean must be equal to zero almost surely, we have (7):

$$\lim_{k\to\infty}\left\|P^+_{X_k|G} - P^-_{X_k|G}\right\|_{\mathsf{TV}} = 0 \quad G\text{-}a.s.$$

This completes the proof. ∎

Finally, the next proposition portrays that the Markov chain $\{\sigma_k : k \in \mathbb{N}\}$ converges almost surely when $\delta \in \left(\delta_{\mathsf{maj}}, \frac{1}{2}\right)$, $L_k = \omega(\log(k))$, and all processing functions are majority.

Proposition 5 (Majority Random DAG Model Almost Sure Convergence). *If* $\delta \in \left(\delta_{\mathsf{maj}}, \frac{1}{2}\right)$ *and* $L_k = \omega(\log(k))$, *then* $\lim_{k\to\infty}\sigma_k = \frac{1}{2}$ *almost surely.*

Proposition 5 is proved in Appendix E. It can be construed as a "weak" impossibility result since it demonstrates that the average number of 1's tends to $\frac{1}{2}$ in the $\delta \in \left(\delta_{\mathsf{maj}}, \frac{1}{2}\right)$ regime regardless of the initial state of the Markov chain $\{\sigma_k : k \in \mathbb{N}\}$.

IV. ANALYSIS OF AND-OR RULE PROCESSING IN RANDOM DAG MODEL

In this section, we prove Theorem 2. As before, we begin by making some pertinent observations. Recall that we have a random DAG model with $d = 2$, and all Boolean functions at even levels are the AND rule, and all Boolean functions at odd levels are the OR rule, i.e. $f_k(x_1, x_2) = x_1 \wedge x_2$ for every $k \in 2\mathbb{N}\backslash\{0\}$, and $f_k(x_1, x_2) = x_1 \vee x_2$ for every $k \in \mathbb{N}\backslash 2\mathbb{N}$. Suppose we are given that $\sigma_{k-1} = \sigma$ for any $k \in \mathbb{N}\backslash\{0\}$. Then, for every $j \in [L_k]$:

$$X_{k,j} = \begin{cases} \mathsf{Bernoulli}(\sigma * \delta) \wedge \mathsf{Bernoulli}(\sigma * \delta) &, \text{ if } k \text{ even} \\ \mathsf{Bernoulli}(\sigma * \delta) \vee \mathsf{Bernoulli}(\sigma * \delta) &, \text{ if } k \text{ odd} \end{cases} \tag{42}$$

for two i.i.d. Bernoulli random variables. Since we have:

$$\mathbb{P}(X_{k,j} = 1|\sigma_{k-1} = \sigma) = \begin{cases} (\sigma * \delta)^2 &, \text{ if } k \text{ even} \\ 1 - (1 - \sigma * \delta)^2 &, \text{ if } k \text{ odd} \end{cases} \tag{43}$$

$$= \mathbb{E}[\sigma_k|\sigma_{k-1} = \sigma], \tag{44}$$

$X_{k,j}$ are i.i.d. $\mathsf{Bernoulli}(g_{k\,(\mathsf{mod}\,2)}(\sigma))$ for $j \in [L_k]$, and $L_k\sigma_k \sim \mathsf{binomial}(L_k, g_{k\,(\mathsf{mod}\,2)}(\sigma))$, where we define $g_0 : [0,1] \to [0,1]$ as $g_0(\sigma) \triangleq (\sigma * \delta)^2$, and $g_1 : [0,1] \to [0,1]$ as $g_1(\sigma) \triangleq 1 - (1 - \sigma * \delta)^2 = 2(\sigma * \delta) - (\sigma * \delta)^2$. The derivatives of g_0 and g_1 are:

$$g_0'(\sigma) = 2(1 - 2\delta)(\sigma * \delta) \geq 0, \tag{45}$$

$$g_1'(\sigma) = 2(1 - 2\delta)(1 - \sigma * \delta) \geq 0. \tag{46}$$

Consider the composition of g_0 and g_1, $g \triangleq g_0 \circ g_1 : [0,1] \to [0,1]$, $g(\sigma) = \left(\left(2(\sigma * \delta) - (\sigma * \delta)^2\right) * \delta\right)^2$, which has derivative $g' : [0,1] \to \mathbb{R}^+$ given by:

$$g'(\sigma) = g_0'(g_1(\sigma))g_1'(\sigma)$$

$$= 4(1 - 2\delta)^2(g_1(\sigma) * \delta)(1 - \sigma * \delta) \geq 0. \tag{47}$$

This is a cubic function of σ with maximum value:

$$D(\delta) \triangleq \max_{\sigma \in [0,1]} g'(\sigma) = \begin{cases} g'\left(\frac{1-\delta}{1-2\delta} - \sqrt{\frac{1-\delta}{3(1-2\delta)^3}}\right) & , \quad \delta \in \left(0, \frac{9-\sqrt{33}}{12}\right] \\ g'(0) & , \quad \delta \in \left(\frac{9-\sqrt{33}}{12}, \frac{1}{2}\right) \end{cases} \tag{48}$$

$$= \begin{cases} \left(\frac{4(1-\delta)(1-2\delta)}{3}\right)^{\frac{3}{2}} & , \quad \delta \in \left(0, \frac{9-\sqrt{33}}{12}\right] \\ 4\delta(1-\delta)^2(1-2\delta)^2(3-2\delta) < 1 & , \quad \delta \in \left(\frac{9-\sqrt{33}}{12}, \frac{1}{2}\right) \end{cases} \tag{49}$$

which follows from standard calculus and algebraic manipulations, and Wolfram Mathematica computations. Hence, $D(\delta)$ in (49) is the Lipschitz constant of g over $[0,1]$. Since $4(1-\delta)(1-2\delta)/3 \in (0,1) \Leftrightarrow \delta \in ((3-\sqrt{7})/4, (9-\sqrt{33})/12]$, $D(\delta) < 1$ if and only if $\delta \in ((3-\sqrt{7})/4, 1/2)$. Moreover, $D(\delta) > 1$ if and only if $\delta \in (0, (3-\sqrt{7})/4)$ (and $D(\delta) = 1$ when $\delta = (3-\sqrt{7})/4$).

We next summarize the fixed point structure of g. Solving the equation $g(\sigma) = \sigma$ in Wolfram Mathematica produces:

$$\sigma = \frac{1 - 6\delta + 4\delta^2 \pm \sqrt{1 - 12\delta + 8\delta^2}}{2(1-2\delta)^2}, \frac{3 - 6\delta + 4\delta^2 \pm \sqrt{5 - 12\delta + 8\delta^2}}{2(1-2\delta)^2} \tag{50}$$

where the first pair is real when $\delta \in [0, (3-\sqrt{7})/4]$, and the second pair is always real. From these solutions, it is straightforward to verify that the only fixed points of g in the interval $[0,1]$ are:

$$t_0 \triangleq \frac{2(1-\delta)(1-2\delta) - 1 - \sqrt{4(1-\delta)(1-2\delta) - 3}}{2(1-2\delta)^2} \quad \text{(valid when } \delta \in \left(0, \frac{3-\sqrt{7}}{4}\right]) \tag{51}$$

$$t_1 \triangleq \frac{2(1-\delta)(1-2\delta) - 1 + \sqrt{4(1-\delta)(1-2\delta) - 3}}{2(1-2\delta)^2} \quad \text{(valid when } \delta \in \left(0, \frac{3-\sqrt{7}}{4}\right]) \tag{52}$$

$$t \triangleq \frac{2(1-\delta)(1-2\delta) + 1 - \sqrt{4(1-\delta)(1-2\delta) + 1}}{2(1-2\delta)^2} \tag{53}$$

which satisfy $t_0 = t_1 = t$ when $\delta = (3-\sqrt{7})/4$, and $t_0 = 0$, $t_1 = 1$ when $\delta = 0$. Furthermore, observe that:

$$t_1 - t = \frac{\sqrt{a} + \sqrt{a+4} - 2}{2(1-2\delta)^2} > 0 \quad \text{and} \quad t - t_0 = \frac{\sqrt{a} - \sqrt{a+4} + 2}{2(1-2\delta)^2} > 0 \tag{54}$$

where $a = 4(1-\delta)(1-2\delta) - 3 > 0$ for $\delta \in (0, (3-\sqrt{7})/4)$, $t_1 - t > 0$ because $x \mapsto \sqrt{x}$ is strictly increasing ($\Rightarrow \sqrt{a} + \sqrt{a+4} > 2$), and $t - t_0 > 0$ because $x \mapsto \sqrt{x}$ is strictly subadditive ($\Rightarrow \sqrt{a} + 2 > \sqrt{a+4}$). Hence, $0 < t_0 < t < t_1 < 1$ when $\delta \in (0, (3-\sqrt{7})/4)$.

Therefore, there are again two regimes of interest. Define the critical threshold $\delta_{\text{andor}} \triangleq \frac{3-\sqrt{7}}{4}$. In the regime $\delta \in (0, \delta_{\text{andor}})$, g has three fixed points $0 < t_0 < t < t_1 < 1$, and $D(\delta) > 1$. In contrast, in the regime $\delta \in \left(\delta_{\text{andor}}, \frac{1}{2}\right)$, g has only one fixed point at $t \in (0,1)$, and $D(\delta) < 1$.

We now prove Theorem 2. (The proof closely resembles the proof of Theorem 1 in section III.)

Proof of Theorem 2. As in the proof of Theorem 1, we begin by constructing a monotone Markovian coupling $\{(X_k^-, X_k^+) : k \in \mathbb{N}\}$ between the Markov chains $\{X_k^+ : k \in \mathbb{N}\}$ and $\{X_k^- : k \in \mathbb{N}\}$ (which are versions of the Markov chain $\{X_k : k \in \mathbb{N}\}$ initialized at $X_0^+ = 1$ and $X_0^- = 0$, respectively), and this coupling induces a monotone Markovian coupling $\{(\sigma_k^+, \sigma_k^-) : k \in \mathbb{N}\}$ between the Markov chains $\{\sigma_k^+ : k \in \mathbb{N}\}$ and $\{\sigma_k^- : k \in \mathbb{N}\}$ (which are versions of the Markov chain $\{\sigma_k : k \in \mathbb{N}\}$ initialized at $\sigma_0^+ = 1$ and $\sigma_0^- = 0$, respectively). This monotone Markovian coupling satisfies the following properties:

1) The "marginal" Markov chains are $\{X_k^+ : k \in \mathbb{N}\}$ and $\{X_k^- : k \in \mathbb{N}\}$.

2) For every $k \in \mathbb{N}$, X_{k+1}^+ is conditionally independent of X_k^- given X_k^+, and X_{k+1}^- is conditionally independent of X_k^+ given X_k^-.

3) For every $j > k \geq 1$, σ_j^+ is conditionally independent of $\sigma_0^-, \ldots, \sigma_k^-, \sigma_0^+, \ldots, \sigma_{k-1}^+$ given σ_k^+, and σ_j^- is conditionally independent of $\sigma_0^+, \ldots, \sigma_k^+, \sigma_0^-, \ldots, \sigma_{k-1}^-$ given σ_k^-.

4) For every $k \in \mathbb{N}$ and every $j \in [L_k]$, $X_{k,j}^+ \geq X_{k,j}^-$ almost surely.

5) Due to the previous property, $\sigma_k^+ \geq \sigma_k^-$ almost surely for every $k \in \mathbb{N}$.

As before, the fourth property above holds because $1 = X_{0,0}^+ \geq X_{0,0}^- = 0$ is true by assumption, each edge BSC preserves monotonicity, and the AND and OR processing functions are symmetric and monotone non-decreasing.

Part 1: We first prove that $\delta \in (0, \delta_{\mathsf{andor}})$ implies $\limsup_{k \to \infty} \mathbb{P}(\hat{T}_{2k} \neq \sigma_0) < \frac{1}{2}$. To this end, we start by establishing that there exists $\epsilon = \epsilon(\delta) > 0$ (that depends on δ) such that:

$$\forall k \in \mathbb{N}\backslash\{0\}, \quad \mathbb{P}\big(\sigma_{2k}^+ \geq t_1 - \epsilon \,\big|\, \sigma_{2k-2}^+ \geq t_1 - \epsilon, A_{k,j}\big) \geq 1 - 4\exp\left(-\frac{(L_{2k} \wedge L_{2k-1})\gamma(\epsilon)^2}{8}\right) \tag{55}$$

where $\wedge$ denotes the minimum operation (not to be confused with the AND operation), $\gamma(\epsilon) \triangleq g(t_1 - \epsilon) - (t_1 - \epsilon) > 0$, and $A_{k,j}$ with $0 \leq j < k$ is the non-zero probability event defined as:

$$A_{k,j} \triangleq \begin{cases} \{\sigma_{2j}^- \leq t_0 + \epsilon\} & , \quad 0 \leq j = k - 1 \\ \{\sigma_{2k-4}^+ \geq t_1 - \epsilon, \sigma_{2k-6}^+ \geq t_1 - \epsilon, \ldots, \sigma_{2j}^+ \geq t_1 - \epsilon\} \cap \{\sigma_{2j}^- \leq t_0 + \epsilon\} & , \quad 0 \leq j \leq k - 2 \end{cases}.$$

Since $g'(t_1) = 4\delta(3 - 2\delta) < 1$ and $g(t_1) = t_1$, $g(t_1 - \epsilon) > t_1 - \epsilon$ for sufficiently small $\epsilon > 0$. Fix any such $\epsilon > 0$ (which depends on δ because g depends on δ) such that $\gamma(\epsilon) > 0$. Observe that for every $k \in \mathbb{N}\backslash\{0\}$ and $\xi > 0$, we have:

$$\mathbb{P}(|\sigma_{2k} - g(\sigma_{2k-2})| > \xi|\sigma_{2k-2} = \sigma)$$

$$\leq \mathbb{P}(|\sigma_{2k} - g_0(\sigma_{2k-1})| + |g_0(\sigma_{2k-1}) - g_0(g_1(\sigma_{2k-2}))| > \xi|\sigma_{2k-2} = \sigma)$$

$$\leq \mathbb{P}(|\sigma_{2k} - g_0(\sigma_{2k-1})| + 2(1-\delta)(1-2\delta)|\sigma_{2k-1} - g_1(\sigma_{2k-2})| > \xi|\sigma_{2k-2} = \sigma)$$

$$\leq \mathbb{P}\left(\left\{|\sigma_{2k} - g_0(\sigma_{2k-1})| > \frac{\xi}{2}\right\} \cup \left\{2(1-\delta)(1-2\delta)|\sigma_{2k-1} - g_1(\sigma_{2k-2})| > \frac{\xi}{2}\right\}\bigg|\sigma_{2k-2} = \sigma\right)$$

$$\leq \mathbb{P}\left(|\sigma_{2k} - g_0(\sigma_{2k-1})| > \frac{\xi}{2}\bigg|\sigma_{2k-2} = \sigma\right) + \mathbb{P}\left(|\sigma_{2k-1} - g_1(\sigma_{2k-2})| > \frac{\xi}{4(1-\delta)(1-2\delta)}\bigg|\sigma_{2k-2} = \sigma\right)$$

$$\leq \mathbb{E}\left[\mathbb{P}\left(|\sigma_{2k} - g_0(\sigma_{2k-1})| > \frac{\xi}{2}\bigg|\sigma_{2k-1}\right)\bigg|\sigma_{2k-2} = \sigma\right] + 2\exp\left(-\frac{L_{2k-1}\xi^2}{8(1-\delta)^2(1-2\delta)^2}\right)$$

$$\leq 2\exp\left(-\frac{L_{2k}\xi^2}{2}\right) + 2\exp\left(-\frac{L_{2k-1}\xi^2}{8(1-\delta)^2(1-2\delta)^2}\right)$$

$$\leq 4\exp\left(-\frac{(L_{2k} \wedge L_{2k-1})\xi^2}{8}\right) \tag{56}$$

where the first inequality follows from the triangle inequality and the fact that $g = g_0 \circ g_1$, the second inequality holds because the Lipschitz constant of g_0 on $[0, 1]$ is $\max_{\sigma \in [0,1]} g_0'(\sigma) = g_0'(1) = 2(1-\delta)(1-2\delta)$ using (45), the fourth inequality follows from the union bound, the fifth and sixth inequalities follow from the Markov property

and Hoeffding's inequality (as well as the fact that $L_k \sigma_k \sim \text{binomial}(L_k, g_{k \,(\text{mod } 2)}(\sigma))$ given $\sigma_{k-1} = \sigma$), and the final inequality holds because $(1-\delta)^2(1-2\delta)^2 \leq 1$. Hence, for any $k \in \mathbb{N}\backslash\{0\}$ and any $0 \leq j < k$, we have:

$$
\mathbb{P}\big(\sigma_{2k}^+ < g\big(\sigma_{2k-2}^+\big) - \gamma(\epsilon)\,\big|\,\sigma_{2k-2}^+ = \sigma, A_{k,j}\big) = \mathbb{P}\big(\sigma_{2k} < g(\sigma_{2k-2}) - \gamma(\epsilon)\,\big|\,\sigma_{2k-2} = \sigma\big)
$$
$$
\leq \mathbb{P}\big(|\sigma_{2k} - g(\sigma_{2k-2})| > \gamma(\epsilon)|\sigma_{2k-2} = \sigma\big)
$$
$$
\leq 4\exp\bigg(-\frac{(L_{2k} \wedge L_{2k-1})\gamma(\epsilon)^2}{8}\bigg)
$$

where the first equality follows from property 3 of the Markovian coupling, and the final inequality follows from (56). As shown in the proof of Theorem 1, this produces:

$$
\mathbb{P}\big(\sigma_{2k}^+ < g\big(\sigma_{2k-2}^+\big) - \gamma(\epsilon)\,\big|\,\sigma_{2k-2}^+ \geq t_1 - \epsilon, A_{k,j}\big) \leq 4\exp\bigg(-\frac{(L_{2k} \wedge L_{2k-1})\gamma(\epsilon)^2}{8}\bigg)
$$
$$
\mathbb{P}\big(\sigma_{2k}^+ < t_1 - \epsilon\,\big|\,\sigma_{2k-2}^+ \geq t_1 - \epsilon, A_{k,j}\big) \leq 4\exp\bigg(-\frac{(L_{2k} \wedge L_{2k-1})\gamma(\epsilon)^2}{8}\bigg)
$$

where the second inequality follows from the first because $\sigma_{2k}^+ < t_1 - \epsilon = g(t_1 - \epsilon) - \gamma(\epsilon)$ implies that $\sigma_{2k}^+ < g(\sigma_{2k-2}^+) - \gamma(\epsilon)$ when $\sigma_{2k-2}^+ \geq t_1 - \epsilon$ (since g is non-decreasing and $g(\sigma_{2k-2}^+) \geq g(t_1 - \epsilon)$). This proves (55).

Now fix any $\tau > 0$, and choose a sufficiently large even integer $K = K(\epsilon, \tau) \in 2\mathbb{N}$ (that depends on ϵ and τ) such that:

$$
4\sum_{m=\frac{K}{2}+1}^{\infty} \exp\bigg(-\frac{(L_{2m} \wedge L_{2m-1})\gamma(\epsilon)^2}{8}\bigg) \leq \tau. \tag{57}
$$

Note that such K exists because $\sum_{m=1}^{\infty} 1/(2m-1)^2 \leq 1 + \sum_{m=2}^{\infty} 1/(2m-2)^2 = 1 + (\pi^2/24) < +\infty$, and for sufficiently large m (depending on δ), we have:

$$
\exp\bigg(-\frac{(L_{2m} \wedge L_{2m-1})\gamma(\epsilon)^2}{8}\bigg) \leq \frac{1}{(2m-1)^2} \quad \Leftrightarrow \quad L_{2m} \wedge L_{2m-1} \geq \frac{16\log(2m-1)}{\gamma(\epsilon)^2}. \tag{58}
$$

As before, in (58), we use the assumption that $L_m \geq C(\delta)\log(m)$ for all sufficiently large m (depending on δ), where we define the constant $C(\delta)$ as:

$$
C(\delta) \triangleq \frac{16}{\gamma(\epsilon(\delta))^2} > 0. \tag{59}
$$

Using the continuity of probability measures, observe that:

$$
\mathbb{P}\Bigg(\bigcap_{k>\frac{K}{2}} \{\sigma_{2k}^+ \geq t_1 - \epsilon\}\,\Bigg|\,\sigma_K^+ \geq t_1 - \epsilon, \sigma_K^- \leq t_0 + \epsilon\Bigg) = \prod_{k>\frac{K}{2}} \mathbb{P}\big(\sigma_{2k}^+ \geq t_1 - \epsilon\,\big|\,\sigma_{2k-2}^+ \geq t_1 - \epsilon, A_{k,\frac{K}{2}}\big)
$$
$$
\geq \prod_{k>\frac{K}{2}} 1 - 4\exp\bigg(-\frac{(L_{2k} \wedge L_{2k-1})\gamma(\epsilon)^2}{8}\bigg)
$$
$$
\geq 1 - 4\sum_{k>\frac{K}{2}} \exp\bigg(-\frac{(L_{2k} \wedge L_{2k-1})\gamma(\epsilon)^2}{8}\bigg)
$$
$$
\geq 1 - \tau
$$

where the first inequality follows from (55), and the final inequality follows from (57). Therefore, we have for any $k > \frac{K}{2}$:

$$
\mathbb{P}\big(\sigma_{2k}^+ \geq t_1 - \epsilon\,\big|\,\sigma_K^+ \geq t_1 - \epsilon, \sigma_K^- \leq t_0 + \epsilon\big) \geq 1 - \tau. \tag{60}
$$

Likewise, we can also prove mutatis mutandis that for any $k > \frac{K}{2}$:

$$\mathbb{P}\big(\sigma_{2k}^- \leq t_0 + \epsilon \,\big|\, \sigma_K^+ \geq t_1 - \epsilon, \sigma_K^- \leq t_0 + \epsilon\big) \geq 1 - \tau \tag{61}$$

where ϵ, τ, and K in (61) can be chosen to be the same as those in (60) without loss of generality.

Finally, we let $E = \big\{\sigma_K^+ \geq t_1 - \epsilon, \sigma_K^- \leq t_0 + \epsilon\big\}$, and observe that for all $k > \frac{K}{2}$:

$$\mathbb{P}\big(\sigma_{2k}^+ \geq t\big) - \mathbb{P}\big(\sigma_{2k}^- \geq t\big) \geq \mathbb{E}\big[\big(\mathbb{1}\big\{\sigma_{2k}^+ \geq t\big\} - \mathbb{1}\big\{\sigma_{2k}^- \geq t\big\}\big)\,\mathbb{1}\{E\}\big]$$

$$= \big(\mathbb{P}\big(\sigma_{2k}^+ \geq t \,\big|\, E\big) - \mathbb{P}\big(\sigma_{2k}^- \geq t \,\big|\, E\big)\big)\,\mathbb{P}(E)$$

$$\geq \big(\mathbb{P}\big(\sigma_{2k}^+ \geq t_1 - \epsilon \,\big|\, E\big) - \mathbb{P}\big(\sigma_{2k}^- > t_0 + \epsilon \,\big|\, E\big)\big)\,\mathbb{P}(E)$$

$$\geq (1 - 2\tau)\mathbb{P}(E) > 0$$

where the first inequality holds because $\mathbb{1}\big\{\sigma_{2k}^+ \geq t\big\} - \mathbb{1}\big\{\sigma_{2k}^- \geq t\big\} \geq 0$ almost surely due to the monotonicity (property 5) of our Markovian coupling, the second inequality holds because $t_0 + \epsilon < t < t_1 - \epsilon$ (since $\epsilon > 0$ is small), and the final inequality follows from (60) and (61). As argued in the proof of Theorem 1, this illustrates that $\limsup_{k \to \infty} \mathbb{P}(\hat{T}_{2k} \neq \sigma_0) < \frac{1}{2}$.

Part 2: We next prove that $\delta \in \big(\delta_{\mathrm{andor}}, \frac{1}{2}\big)$ implies:

$$\lim_{k \to \infty} \big\|P_{X_{2k}|G}^+ - P_{X_{2k}|G}^-\big\|_{\mathsf{TV}} = 0 \quad G\text{-a.s.} \tag{62}$$

Following the proof of Theorem 1, we can show that:

$$\mathbb{E}\Big[\big\|P_{X_{2k}|G}^+ - P_{X_{2k}|G}^-\big\|_{\mathsf{TV}}\Big] \leq L_{2k}\,\mathbb{E}\big[\sigma_{2k}^+ - \sigma_{2k}^-\big]. \tag{63}$$

In order to bound $\mathbb{E}\big[\sigma_{2k}^+ - \sigma_{2k}^-\big]$, we proceed as follows. Firstly, for any $k \in \mathbb{N}\backslash\{0\}$, we have:

$$\mathbb{E}\big[\sigma_{2k}^+ - \sigma_{2k}^- \,\big|\, \sigma_{2k-2}^+, \sigma_{2k-2}^-\big] = \mathbb{E}\big[\mathbb{E}\big[\sigma_{2k}^+ - \sigma_{2k}^- \,\big|\, \sigma_{2k-1}^+, \sigma_{2k-1}^-\big] \,\big|\, \sigma_{2k-2}^+, \sigma_{2k-2}^-\big]$$

$$= \mathbb{E}\big[g_0\big(\sigma_{2k-1}^+\big) - g_0\big(\sigma_{2k-1}^-\big) \,\big|\, \sigma_{2k-2}^+, \sigma_{2k-2}^-\big] \tag{64}$$

where the first equality follows from the tower and Markov properties, and the second equality holds because $L_{2k}\sigma_{2k} \sim \mathrm{binomial}(L_{2k}, g_0(\sigma))$ given $\sigma_{2k-1} = \sigma$. Then, recalling that $g_0(\sigma) = (\sigma * \delta)^2 = (1 - 2\delta)^2\sigma^2 + 2\delta(1 - 2\delta)\sigma + \delta^2$, we can compute:

$$\mathbb{E}\big[g_0\big(\sigma_{2k-1}^+\big) \,\big|\, \sigma_{2k-2}^+, \sigma_{2k-2}^-\big] = \mathbb{E}\big[g_0\big(\sigma_{2k-1}^+\big) \,\big|\, \sigma_{2k-2}^+\big]$$

$$= \mathbb{E}\big[(1 - 2\delta)^2\sigma_{2k-1}^{+^2} + 2\delta(1 - 2\delta)\sigma_{2k-1}^+ + \delta^2 \,\big|\, \sigma_{2k-2}^+\big]$$

$$= (1 - 2\delta)^2\Big(\mathbb{VAR}\big(\sigma_{2k-1}^+ \,\big|\, \sigma_{2k-2}^+\big) + \mathbb{E}\big[\sigma_{2k-1}^+ \,\big|\, \sigma_{2k-2}^+\big]^2\Big)$$

$$\quad + 2\delta(1 - 2\delta)\mathbb{E}\big[\sigma_{2k-1}^+ \,\big|\, \sigma_{2k-2}^+\big] + \delta^2$$

$$= (1 - 2\delta)^2 g_1\big(\sigma_{2k-2}^+\big)^2 + 2\delta(1 - 2\delta)g_1\big(\sigma_{2k-2}^+\big) + \delta^2$$

$$\quad + (1 - 2\delta)^2 \frac{g_1\big(\sigma_{2k-2}^+\big)\big(1 - g_1\big(\sigma_{2k-2}^+\big)\big)}{L_{2k-1}}$$

$$= g\big(\sigma_{2k-2}^+\big) + (1 - 2\delta)^2 \frac{g_1\big(\sigma_{2k-2}^+\big)\big(1 - g_1\big(\sigma_{2k-2}^+\big)\big)}{L_{2k-1}} \tag{65}$$

where the first equality uses property 3 of the monotone Markovian coupling, and the fourth equality uses the fact that $L_{2k-1}\sigma_{2k-1} \sim \text{binomial}(L_{2k-1}, g_1(\sigma))$ given $\sigma_{2k-2} = \sigma$. Using (64) and (65), we get:

$$
\begin{aligned}
\mathbb{E}\big[\sigma_{2k}^+ - \sigma_{2k}^- \,\big|\, \sigma_{2k-2}^+, \sigma_{2k-2}^-\big] &= g\big(\sigma_{2k-2}^+\big) - g\big(\sigma_{2k-2}^-\big) \\
&\quad + (1-2\delta)^2\left(\frac{g_1\big(\sigma_{2k-2}^+\big)\big(1 - g_1\big(\sigma_{2k-2}^+\big)\big) - g_1\big(\sigma_{2k-2}^-\big)\big(1 - g_1\big(\sigma_{2k-2}^-\big)\big)}{L_{2k-1}}\right) \\
&= g\big(\sigma_{2k-2}^+\big) - g\big(\sigma_{2k-2}^-\big) \\
&\quad + (1-2\delta)^2\left(\frac{g_1\big(\sigma_{2k-2}^+\big) - g_1\big(\sigma_{2k-2}^-\big) - \big(g_1\big(\sigma_{2k-2}^+\big)^2 - g_1\big(\sigma_{2k-2}^-\big)^2\big)}{L_{2k-1}}\right) \\
&\leq g\big(\sigma_{2k-2}^+\big) - g\big(\sigma_{2k-2}^-\big) + (1-2\delta)^2\left(\frac{g_1\big(\sigma_{2k-2}^+\big) - g_1\big(\sigma_{2k-2}^-\big)}{L_{2k-1}}\right) \\
&\leq \left(D(\delta) + \frac{2(1-\delta)(1-2\delta)^3}{L_{2k-1}}\right)\big(\sigma_{2k-2}^+ - \sigma_{2k-2}^-\big) \\
&\leq \left(D(\delta) + \frac{2}{L_{2k-1}}\right)\big(\sigma_{2k-2}^+ - \sigma_{2k-2}^-\big)
\end{aligned}
\tag{66}
$$

where the first inequality holds because $g_1\big(\sigma_{2k-2}^+\big)^2 - g_1\big(\sigma_{2k-2}^-\big)^2 \geq 0$ almost surely (since g_1 is non-negative and non-decreasing by (46), and $\sigma_{2k-2}^+ \geq \sigma_{2k-2}^-$ almost surely by property 5 of the monotone Markovian coupling), the second inequality holds because $\sigma_{2k-2}^+ \geq \sigma_{2k-2}^-$ almost surely and g and g_1 have Lipschitz constants $D(\delta)$ and $\max_{\sigma \in [0,1]} g_1'(\sigma) = 2(1-\delta)(1-2\delta)$ respectively, and the final inequality holds because $(1-\delta)(1-2\delta)^3 \leq 1$. Then, as in the proof of Theorem 1, we can take expectations in (66) to obtain:

$$
0 \leq \mathbb{E}\big[\sigma_{2k}^+ - \sigma_{2k}^-\big] \leq \left(D(\delta) + \frac{2}{L_{2k-1}}\right)\mathbb{E}\big[\sigma_{2k-2}^+ - \sigma_{2k-2}^-\big]
$$

which recursively produces:

$$
0 \leq \mathbb{E}\big[\sigma_{2k}^+ - \sigma_{2k}^-\big] \leq \prod_{i=1}^{k}\left(D(\delta) + \frac{2}{L_{2i-1}}\right)
$$

where we use the fact that $\mathbb{E}\big[\sigma_0^+ - \sigma_0^-\big] = 1$.

Next, using (63) with this bound, we get:

$$
\mathbb{E}\Big[\big\|P_{X_{2k}|G}^+ - P_{X_{2k}|G}^-\big\|_{\mathsf{TV}}\Big] \leq L_{2k}\prod_{i=1}^{k}\left(D(\delta) + \frac{2}{L_{2i-1}}\right).
\tag{67}
$$

Recall that $L_k = o\big(E(\delta)^{-\frac{k}{2}}\big)$ and $\liminf_{k\to\infty} L_k > \frac{2}{E(\delta)-D(\delta)}$ for some $E(\delta) \in (D(\delta), 1)$ (that depends on δ). Hence, there exists $K = K(\delta) \in \mathbb{N}$ (that depends on δ) such that for all $i > K$, $L_{2i-1} \geq \frac{2}{E(\delta)-D(\delta)}$. This means that we can further upper bound (67) as follows:

$$
\forall k > K, \;\; \mathbb{E}\Big[\big\|P_{X_{2k}|G}^+ - P_{X_{2k}|G}^-\big\|_{\mathsf{TV}}\Big] \leq L_{2k}\,E(\delta)^{k-K}\prod_{i=1}^{K}\left(D(\delta) + \frac{2}{L_{2i-1}}\right)
$$

and letting $k \to \infty$ produces:

$$
\lim_{k\to\infty} \mathbb{E}\Big[\big\|P_{X_{2k}|G}^+ - P_{X_{2k}|G}^-\big\|_{\mathsf{TV}}\Big] = 0.
\tag{68}
$$

Finally, as in the proof of Theorem 1, $\|P_{X_{2k}|G}^+ - P_{X_{2k}|G}^-\|_{\mathsf{TV}} \in [0,1]$ forms a non-increasing sequence in k for every realization of the random DAG G, and the pointwise limit random variable, $\lim_{k\to\infty}\|P_{X_{2k}|G}^+ - P_{X_{2k}|G}^-\|_{\mathsf{TV}} \in$

$[0, 1]$, has mean $\mathbb{E}[\lim_{k \to \infty} \|P^+_{X_{2k}|G} - P^-_{X_{2k}|G}\|_{\mathsf{TV}}] = 0$ due to (68) and the bounded convergence theorem. Therefore, we must have (62), which completes the proof.   ∎

We remark that when $\delta \in \left(\delta_{\mathsf{andor}}, \frac{1}{2}\right)$ and the condition $\liminf_{k \to \infty} L_k > \frac{2}{E(\delta) - D(\delta)}$ cannot be satisfied by any $E(\delta)$, if L_k satisfies the condition of Proposition 2 (in subsection II-C), then part 2 of Proposition 2 still yields the desired converse result. Finally, the next proposition demonstrates that the Markov chain $\{\sigma_{2k} : k \in \mathbb{N}\}$ converges almost surely when $\delta \in \left(\delta_{\mathsf{andor}}, \frac{1}{2}\right)$, $L_k = \omega(\log(k))$, all processing functions at even levels are the AND rule, and all processing functions at odd levels are the OR rule.

Proposition 6 (AND-OR Random DAG Model Almost Sure Convergence). *If $\delta \in \left(\delta_{\mathsf{andor}}, \frac{1}{2}\right)$ and $L_k = \omega(\log(k))$, then $\lim_{k \to \infty} \sigma_{2k} = t$ almost surely.*

Proposition 6 is proved in Appendix F, and much like Proposition 5, it can also be construed as a "weak" impossibility result.

V. Deterministic Quasi-Polynomial Time and Randomized Polylogarithmic Time Constructions of DAGs where Broadcasting is Possible

In this section, we prove Theorem 3 by constructing deterministic bounded degree DAGs with $L_k = \Theta(\log(k))$ where broadcasting is possible. As mentioned in subsection II-B, our construction is based on d-regular bipartite lossless $(d^{-6/5}, d - 2d^{4/5})$-expander graphs. So, we first verify that such graphs actually exist. Recall that we represent a d-regular bipartite graph as $B = (U, V, E)$, where U and V are disjoint sets of vertices and E is the set of undirected edges. The next proposition is a specialization of [45, Proposition 1, Appendix II] which illustrates that randomly generated regular bipartite graphs are good expanders with high probability.

Proposition 7 (Random Expander Graph [34, Lemma 1], [45, Proposition 1, Appendix II]). *Fix any fraction $\alpha \in (0, 1)$ and any degree $d \in \mathbb{N} \backslash \{0\}$. Then, for every sufficiently large n (depending on α and d), the randomly generated d-regular bipartite graph $\mathsf{B} = (U, V, \mathsf{E})$ with $|U| = |V| = n$ satisfies:*

$$\mathbb{P}\left(\forall S \subseteq U \text{ with } |S| = \alpha n, \ |\Gamma(S)| \geq n\left(1 - (1 - \alpha)^d - \sqrt{2d\alpha H(\alpha)}\right)\right) > 1 - \binom{n}{n\alpha} \exp(-nH(\alpha))$$

$$\geq 1 - \frac{e}{2\pi\sqrt{\alpha(1 - \alpha)n}}$$

where $H(\alpha) \triangleq -\alpha \log(\alpha) - (1 - \alpha) \log(1 - \alpha)$ is the binary entropy function.

We note that the probability measure $\mathbb{P}$ in Proposition 7 is defined by the random d-regular bipartite graph B, whose vertices U and V are fixed and edges E are random. In particular, B is generated as follows (cf. *configuration model* in [46, Section 2.4]):

1) Fix a complete bipartite graph $\hat{B} = (\hat{U}, \hat{V}, \hat{E})$ such that $|\hat{U}| = |\hat{V}| = dn$.
2) Randomly and uniformly select a perfect matching $\mathsf{M} \subseteq \hat{E}$ in $\hat{B}$.
3) Group sets of d consecutive vertices in $\hat{U}$, respectively $\hat{V}$, to generate a set of n super-vertices U, respectively V.

4) This yields a random d-regular bipartite graph $\mathsf{B} = (\mathsf{U}, \mathsf{V}, \mathsf{E})$, where every edge in E is an edge between super-vertices in M.

Note that we allow for the possibility that two vertices in B have multiple edges between them. The first inequality in Proposition 7 is proved in [45, Appendix II]. On the other hand, the second inequality in Proposition 7 is a straightforward consequence of estimating the binomial coefficient using precise Stirling's formula bounds, cf. [47, Chapter II, Section 9, Equation (9.15)]:

$$\forall n \in \mathbb{N}\backslash\{0\}, \quad \sqrt{2\pi n}\left(\frac{n}{e}\right)^n \leq n! \leq e\sqrt{n}\left(\frac{n}{e}\right)^n. \tag{69}$$

The second inequality portrays that the probability in Proposition 7 tends to 1 as $n \to \infty$. Moreover, strictly speaking, αn must be an integer, but we will neglect this detail throughout our exposition for simplicity (as in subsection II-B). We next use this proposition to establish the existence of pertinent regular bipartite lossless expander graphs.

Corollary 2 (Lossless Expander Graph). *Fix any $\epsilon \in (0, 1)$ and any degree $d \geq \left(\frac{2}{\epsilon}\right)^5$. Then, for every sufficient large n (depending on d), the randomly generated d-regular bipartite graph $\mathsf{B} = (\mathsf{U}, \mathsf{V}, \mathsf{E})$ with $|\mathsf{U}| = |\mathsf{V}| = n$ satisfies:*

$$\mathbb{P}\left(\forall S \subseteq U \text{ with } |S| = \frac{n}{d^{6/5}}, \ |\Gamma(S)| \geq (1 - \epsilon)d|S|\right) > 1 - \frac{e}{2\pi\sqrt{d^{-6/5}\left(1 - d^{-6/5}\right)n}}.$$

Hence, for every sufficient large n (depending on d), there exists a d-regular bipartite lossless $(d^{-6/5}, (1-\epsilon)d)$-expander graph $B = (U, V, E)$ with $|U| = |V| = n$ such that for every subset of vertices $S \subseteq U$, we have:

$$|S| = \frac{n}{d^{6/5}} \quad \Rightarrow \quad |\Gamma(S)| \geq (1 - \epsilon)d|S| = (1 - \epsilon)\frac{n}{d^{1/5}}.$$

Corollary 2 is proved in Appendix G. We remark that explicit constructions of bipartite lossless expander graphs B where only the vertices in U are d-regular can be found in the literature, cf. [36], but we require the vertices in V to be d-regular in our construction.

As we discussed in subsection II-B, d-regular bipartite lossless expander graphs can be concatenated to produce a DAG where broadcasting is possible. To formally establish this, we first argue that a single d-regular bipartite lossless expander graph, when perceived as two successive layers of a deterministic DAG, exhibits a "one-step broadcasting" property. Fix any crossover probability $\delta \in \left(0, \frac{1}{2}\right)$, and choose any sufficiently large *odd degree* $d = d(\delta)$ (that depends on δ) such that (15) (reproduced below) holds:

$$\frac{8}{d^{1/5}} + d^{6/5}\exp\left(-\frac{(1 - 2\delta)^2(d - 4)^2}{8d}\right) \leq \frac{1}{2} \tag{70}$$

where the left hand side tends to 0 as $d \to \infty$ for fixed δ, and the minimum value of d satisfying this inequality increases as $\delta \to \frac{1}{2}^-$. Then, Corollary 2 demonstrates that for any sufficiently large n (depending on d), there exists a d-regular bipartite lossless $(d^{-6/5}, d - 2d^{4/5})$-expander graph $B = (U, V, E)$ with $|U| = |V| = n$ such that the expansion property in (14) (reproduced below) holds:

$$\forall S \subseteq U, \ |S| = \frac{n}{d^{6/5}} \quad \Rightarrow \quad |\Gamma(S)| \geq (1 - \epsilon)\frac{n}{d^{1/5}} \ \text{ with } \epsilon = \frac{2}{d^{1/5}}. \tag{71}$$

Note that in the statements of Lemma 1 (see below) and Theorem 3, we assume the existence of such d-regular bipartite lossless $(d^{-6/5}, d - 2d^{4/5})$-expander graphs without proof due to Corollary 2. Let us assume that the

undirected edges in E are actually all directed from U to V, and construe B as two consecutive levels of a deterministic DAG upon which we are broadcasting (as in subsection II-B). In particular, let the Bernoulli random variable corresponding to any vertex $v \in U \cup V$ be denoted by X_v, and suppose each (directed) edge of B is an independent $\mathsf{BSC}(\delta)$ as before. Furthermore, let the Boolean processing function at each vertex in V be the majority rule, which is always well-defined as d is odd. This defines a Bayesian network on B, and the ensuing lemma establishes the feasibility of "one-step broadcasting" down this Bayesian network.

Lemma 1 (One-Step Broadcasting in Expander DAG). *For any noise level $\delta \in \left(0, \frac{1}{2}\right)$, any sufficiently large odd degree $d = d(\delta) \geq 5$ (that depends on δ) satisfying (15), and any sufficiently large n (depending on d), consider the Bayesian network, with independent $\mathsf{BSC}(\delta)$ noise on the edges and majority Boolean processing functions at the vertices, defined above on a d-regular bipartite lossless $(d^{-6/5}, d - 2d^{4/5})$-expander graph $B = (U, V, E)$ such that $|U| = |V| = n$. Then, for every input distribution on $\{X_u : u \in U\}$, we have:*

$$\mathbb{P}\left(\sum_{v \in V} X_v > \frac{n}{d^{6/5}} \;\middle|\; \sum_{u \in U} X_u \leq \frac{n}{d^{6/5}}\right) \leq \exp\left(-\frac{n}{2d^{12/5}}\right).$$

Proof. We begin with some useful definitions. For any vertex $v \in V$, let $\mathsf{pa}(v)$ denote the multiset of vertices in U that are parents of v. (Note that $\mathsf{pa}(v)$ is a multiset because there may be multiple edges between two vertices, and $|\mathsf{pa}(v)| = d$.) Let $S \triangleq \{u \in U : X_u = 1\} \subseteq U$ denote the subset of vertices in U that take value 1, which implies that $|S| = \sum_{u \in U} X_u$. Furthermore, for any vertex $v \in V$, let $N_v \triangleq \sum_{u \in \mathsf{pa}(v)} X_u$ denote the number of parents of v in S that have value 1 (counting with repetition). Finally, let $T \triangleq \{v \in V : N_v \geq t\} \subseteq V$ denote the subset of vertices in V with at least $t \in \mathbb{N}\backslash\{0, 1\}$ parents in S. We will assign an appropriate value to t below.

Suppose $|S| = \sum_{u \in U} X_u \leq n/d^{6/5}$ (which is the event we condition upon in the lemma statement). Consider the case where $|S| = n/d^{6/5}$. Then, applying the expansion property in (14) yields (the "vertex counting" bound):

$$|\Gamma(S)| = |T| + |\Gamma(S)\backslash T| \geq (1 - \epsilon)\frac{n}{d^{1/5}} \tag{72}$$

where $T \subseteq \Gamma(S)$ by definition of T, and $\epsilon = 2d^{-1/5}$. Moreover, we also have the "edge counting" bound:

$$t|T| + |\Gamma(S)\backslash T| \leq d|S| = \frac{n}{d^{1/5}} \tag{73}$$

since each vertex in T has at least t edges from S, each vertex in $\Gamma(S)\backslash T$ has at least 1 edge from S, and the total number of outgoing edges from S is $d|S|$. Combining (72) and (73) produces:

$$(1 - \epsilon)\frac{n}{d^{1/5}} - |T| \leq |\Gamma(S)\backslash T| \leq \frac{n}{d^{1/5}} - t|T|$$

which implies that:

$$|T| \leq \frac{n\epsilon}{d^{1/5}(t - 1)} = \frac{2n}{d^{2/5}(t - 1)}. \tag{74}$$

On the other hand, in the the case where $|S| < n/d^{6/5}$, if we flip the values of vertices in $U\backslash S$ to 1 and subsequently increase the cardinality of S, then the cardinality of T also increases or remains the same. Hence, if $|S| = \sum_{u \in U} X_u \leq n/d^{6/5}$, then (74) also holds.

Now, for any input distribution on $\{X_u : u \in U\}$, observe that:

$$\mathbb{P}\left(\sum_{v \in V} X_v > \frac{n}{d^{6/5}} \,\middle|\, \sum_{u \in U} X_u \leq \frac{n}{d^{6/5}}\right)$$

$$= \mathbb{P}\left(\sum_{v \in T} X_v + \sum_{v \in V\backslash T} X_v > \frac{n}{d^{6/5}} \,\middle|\, |S| \leq \frac{n}{d^{6/5}}\right)$$

$$\leq \mathbb{P}\left(\sum_{v \in V\backslash T} X_v > \frac{n}{d^{6/5}} - |T| \,\middle|\, |S| \leq \frac{n}{d^{6/5}}\right)$$

$$\leq \mathbb{P}\left(\sum_{v \in V\backslash T} X_v > \frac{n}{d^{6/5}} - \frac{2n}{d^{2/5}(t-1)} \,\middle|\, |S| \leq \frac{n}{d^{6/5}}\right)$$

$$= \mathbb{E}\left[\mathbb{P}\left(\sum_{v \in V\backslash T} X_v > \frac{n}{d^{6/5}} - \frac{2n}{d^{2/5}(t-1)} \,\middle|\, V\backslash T, \{N_v : v \in V\backslash T\}\right)\middle|\, |S| \leq \frac{n}{d^{6/5}}\right] \tag{75}$$

$$\leq \mathbb{E}\left[\mathbb{P}\left(\sum_{v \in V\backslash T} X_v > \frac{n}{d^{6/5}} - \frac{2n}{d^{2/5}(t-1)} \,\middle|\, V\backslash T, \{\forall v \in V\backslash T, N_v = t-1\}\right)\middle|\, |S| \leq \frac{n}{d^{6/5}}\right]$$

$$= \mathbb{E}\left[\mathbb{P}\left(\text{binomial}(|V\backslash T|, \mathbb{P}(X_v = 1|N_v = t-1)) > \frac{n}{d^{6/5}} - \frac{2n}{d^{2/5}(t-1)} \,\middle|\, V\backslash T\right)\middle|\, |S| \leq \frac{n}{d^{6/5}}\right]$$

$$\leq \mathbb{P}\left(\text{binomial}(n, \mathbb{P}(X_v = 1|N_v = t-1)) > \frac{n}{d^{6/5}} - \frac{2n}{d^{2/5}(t-1)}\right)$$

$$\leq \mathbb{P}\left(\text{binomial}\left(n, \exp\left(-2d(1-2\delta)^2\left(\frac{1}{2} - \frac{t-1}{d}\right)^2\right)\right) > \frac{n}{d^{6/5}} - \frac{2n}{d^{2/5}(t-1)}\right) \tag{76}$$

where the steps hold due to the following reasons:

1) In the first equality, T and $V\backslash T$ are random sets.

2) The second inequality holds because $X_v \in \{0, 1\}$ for all $v \in T$.

3) The third inequality follows from (74).

4) The fourth equality uses the fact that $\{X_v : v \in V\backslash T\}$ are conditionally independent of the event $\{|S| \leq n/d^{6/5}\}$ given $V\backslash T$ and $\{N_v : v \in V\backslash T\}$, and the conditional expectation in the fourth equality is over the random set $V\backslash T$ and the random variables $\{N_v : v \in V\backslash T\}$.

5) The fifth inequality holds because $N_v \leq t-1$ for every $v \in V\backslash T$, and a straightforward monotone coupling argument shows that the distribution $P_{X_v|N_v=t-1}$ stochastically dominates the distribution $P_{X_v|N_v=k}$ for any $k < t-1$. Furthermore, the conditional expectation in the fifth inequality is only over the random set $V\backslash T$.

6) The sixth equality holds because $\{X_v : v \in V\backslash T\}$ are conditionally i.i.d. given $V\backslash T$ and the event $\{\forall v \in V\backslash T, N_v = t-1\}$.

7) The seventh inequality holds because $|V\backslash T| \leq n$, and a simple monotone coupling argument establishes that a $\text{binomial}(n, \mathbb{P}(X_v = 1|N_v = t-1))$ random variable stochastically dominates a $\text{binomial}(|V\backslash T|, \mathbb{P}(X_v = 1|N_v = t-1))$ random variable.

8) The eighth inequality holds because a $\text{binomial}(n, p)$ random variable stochastically dominates a $\text{binomial}(n, q)$

random variable when $p \geq q$ (again via a monotone coupling argument), and Hoeffding's inequality yields:

$$\mathbb{P}(X_v = 1 | N_v = t - 1) = \mathbb{P}\left(\sum_{i=1}^{t-1} Z_i + \sum_{j=1}^{d-t+1} Y_j > \frac{d}{2}\right)$$

$$\leq \exp\left(-2d(1 - 2\delta)^2 \left(\frac{1}{2} - \frac{t-1}{d}\right)^2\right) \tag{77}$$

where Z_i are i.i.d. Bernoulli$(1 - \delta)$, Y_j are i.i.d. Bernoulli(δ), $\{Z_i : i \in \{1, \ldots, t-1\}\}$ and $\{Y_j : j \in \{1, \ldots, d - t + 1\}\}$ are independent, we assume that $\frac{t-1}{d} < \frac{1}{2}$, and we use the fact that X_v is the majority of its parents' values after passing them through independent BSC(δ)'s.

Finally, applying Hoeffding's inequality once more to (76) yields:

$$\mathbb{P}\left(\sum_{v \in V} X_v > \frac{n}{d^{6/5}} \,\middle|\, \sum_{u \in U} X_u \leq \frac{n}{d^{6/5}}\right)$$

$$\leq \exp\left(-2n\left(\frac{1}{d^{6/5}} - \frac{2}{d^{2/5}(t-1)} - \exp\left(-2d(1-2\delta)^2\left(\frac{1}{2} - \frac{t-1}{d}\right)^2\right)\right)^2\right) \tag{78}$$

where we assume that:

$$\frac{1}{d^{6/5}} - \frac{2}{d^{2/5}(t-1)} > \exp\left(-2d(1-2\delta)^2\left(\frac{1}{2} - \frac{t-1}{d}\right)^2\right). \tag{79}$$

Next, let $t = 1 + \left\lceil \frac{d}{4} \right\rceil$ so that:[12]

$$\frac{1}{4} \leq \frac{t-1}{d} \leq \frac{1}{4} + \frac{1}{d}.$$

Since we have assumed in the lemma statement that $d \geq 5$, the upper bound on $\frac{t-1}{d}$ illustrates that $\frac{t-1}{d} < \frac{1}{2}$, which ensures that (77) is valid. Furthermore, using both the upper and lower bounds on $\frac{t-1}{d}$, notice that (79) is also valid if we have:

$$\frac{1}{d^{6/5}} - \frac{8}{d^{7/5}} > \exp\left(-\frac{(1-2\delta)^2(d-4)^2}{8d}\right) \quad \Leftrightarrow \quad 1 > \frac{8}{d^{1/5}} + d^{6/5}\exp\left(-\frac{(1-2\delta)^2(d-4)^2}{8d}\right)$$

which is true by our assumption in (15). In fact, a simple computation shows that:

$$\frac{1}{d^{6/5}} - \frac{2}{d^{2/5}(t-1)} - \exp\left(-2d(1-2\delta)^2\left(\frac{1}{2} - \frac{t-1}{d}\right)^2\right)$$

$$\geq \frac{1}{d^{6/5}} - \frac{8}{d^{7/5}} - \exp\left(-\frac{(1-2\delta)^2(d-4)^2}{8d}\right) \geq \frac{1}{2d^{6/5}}$$

where the second inequality is equivalent to (15). Therefore, we have from (78):

$$\mathbb{P}\left(\sum_{v \in V} X_v > \frac{n}{d^{6/5}} \,\middle|\, \sum_{u \in U} X_u \leq \frac{n}{d^{6/5}}\right) \leq \exp\left(-\frac{n}{2d^{12/5}}\right)$$

which completes the proof. ∎

Intuitively, Lemma 1 parallels (34) in the proof of Theorem 1 in section III. The lemma portrays that if the proportion of 1's is small in a given layer, then it remains small in the next layer with high probability when the edges between the layers are defined by a regular bipartite lossless expander graph. We next prove Theorem 3

---

[12]The choice t is arbitrary and we could have chosen any t such that $0 < \frac{t-1}{d} < \frac{1}{2}$.

by constructing deterministic bounded degree DAGs with $L_k = \Theta(\log(k))$ and showing using Lemma 1 that the root bit can be reconstructed using the majority decision rule $\hat{S}_k = \mathbb{1}\{\sigma_k \geq \frac{1}{2}\}$. In particular, we delineate two simple algorithms to construct the constituent expander graphs of such DAGs: a *deterministic quasi-polynomial time* algorithm and a *randomized polylogarithmic time* algorithm.

Proof of Theorem 3. Fix any $\delta \in \left(0, \frac{1}{2}\right)$, any sufficiently large $d = d(\delta) \geq 5$ satisfying (15), and any sufficiently large constant $N = N(\delta) \in \mathbb{N}$ such that $M = \exp(N/(4d^{12/5})) \geq 2$ and for every $n \geq N$, there exists a d-regular bipartite lossless $(d^{-6/5}, d - 2d^{4/5})$-expander graph $B_n = (U_n, V_n, E_n)$ with $|U_n| = |V_n| = n$ that satisfies (14) for every subset $S \subseteq U_n$. Furthermore, fix the level sizes so that $L_0 = 1$, $L_1 = N$, and $\{L_k : k \in \mathbb{N}\backslash\{0,1\}\}$ are defined by (16). It is straightforward to verify that $L_k = \Theta(\log(k))$ (for fixed δ). The remainder of the proof is split into three parts. We first present two simple algorithms to generate the constituent expander graphs of the deterministic DAG described in the theorem statement, and then argue that broadcasting is possible on the resulting DAG.

Deterministic Quasi-Polynomial Time Algorithm: We will require two useful facts:

1) For fixed sets of labeled vertices U_n and V_n with $|U_n| = |V_n| = n$, the total number of d-regular bipartite graphs $B_n = (U_n, V_n, E_n)$ is given by the multinomial coefficient:

$$\binom{nd}{d, d, \ldots, d} = \frac{(nd)!}{(d!)^n} \leq n^{nd} = \exp(dn \log(n))$$

where we allow multiple edges between two vertices, and the inequality follows from e.g. [48, Lemma 2.2]. To see this, first attach d edges to each vertex in U_n, and then successively count the number of ways to choose d edges for each vertex in V_n.[13]

2) Checking whether a given d-regular bipartite graph $B_n = (U_n, V_n, E_n)$ with $|U_n| = |V_n| = n$ satisfies (14) for all subsets $S \subseteq U_n$ using brute force takes $O(n^2 \exp(nH(d^{-6/5})))$ time. To see this, note that there are $\binom{n}{nd^{-6/5}} \leq \exp(nH(d^{-6/5}))$ (cf. [48, Lemma 2.2]) subsets $S \subseteq U_n$ with $|S| = nd^{-6/5}$, and verifying (14) takes $O(n^2)$ time for each such subset S.

Consider any level $M^{2^{m-1}} < r \leq M^{2^m}$ with some associated $m \in \mathbb{N}\backslash\{0\}$. We show that the distinct expander graphs making up levels $0, \ldots, r$ of the deterministic DAG in the theorem statement can be constructed in quasi-polynomial time in r. In particular, we need to generate $m+1$ d-regular bipartite lossless $(d^{-6/5}, d-2d^{4/5})$-expander graphs $B_N, B_{2N}, \ldots, B_{2^m N}$. So, for each $i \in \{0, \ldots, m\}$, we generate $B_{2^i N}$ by exhaustively enumerating over the all possible d-regular bipartite graphs with $|U_{2^i N}| = |V_{2^i N}| = 2^i N$ until we find one that satisfies the desired expansion condition. (Note that such expander graphs are guaranteed to exist due to Corollary 2.) Using the aforementioned facts 1 and 2, generating all $m + 1$ desired graphs takes running time:

$$O((m + 1)L_r^2 \exp(L_r H(d^{-6/5})) \exp(dL_r \log(L_r))) = O(\exp(\Theta(\log(r) \log \log(r)))) \tag{80}$$

---

[13] Since the vertices in U_n and V_n are labeled, the total number of non-isomorphic d-regular bipartite graphs is smaller than $(nd)!/(d!)^n$. However, it is larger than $(nd)!/((2n)!(d!)^n)$, and the quasi-polynomial nature of our running time does not change with a more careful calculation of the number of non-isomorphic d-regular bipartite graphs.

where we use the facts that $L_r = 2^m N = \Theta(\log(r))$ and $m = \Theta(\log\log(r))$ since $M^{2^{m-1}} < r \leq M^{2^m}$.[14] Therefore, we can construct the constituent expander graphs in levels $0, \ldots, r$ of our DAG in quasi-polynomial time with brute force. Note that we neglect details of how intermediate graphs are represented in our analysis. Moreover, we are not concerned with optimizing the quasi-polynomial running time.

Randomized Polylogarithmic Time Algorithm: We will require another useful fact:

3) A random d-regular bipartite graph $\mathsf{B} = (U_n, V_n, \mathsf{E})$ with $|U_n| = |V_n| = n$ can be generated according to the distribution described after Proposition 7 in $O(n)$ time. To see this, as outlined after Proposition 7, we must first generate a uniform random perfect matching in a complete bipartite graph $\hat{B} = (\hat{U}_{dn}, \hat{V}_{dn}, \hat{E})$ such that $|\hat{U}_{dn}| = |\hat{V}_{dn}| = dn$. Observe that the edges in a perfect matching can be written as a permutation of the sequence $(1, 2, \ldots, dn)$, because each index and its corresponding value in the (permuted) sequence encodes an edge. So, perfect matchings in $\hat{B}$ are in bijective correspondence with permutations of the sequence $(1, 2, \ldots, dn)$. Therefore, we can generate a uniform random perfect matching by generating a uniform random permutation of $(1, 2, \ldots, dn)$ in $O(dn)$, or equivalently $O(n)$, time using the *Fisher-Yates-Durstenfeld-Knuth shuffle*, cf. [49, Section 3.4.2, p.145] and the references therein. (Note that we do not take the running time of the random number generation process into account.) All that remains is to create super-vertices, which can also be done in $O(n)$ time.

Suppose that the constant $N = N(\delta)$ also satisfies the additional condition:

$$N > \frac{e^2}{\left(6 - 4\sqrt{2}\right)\pi^2 d^{-6/5}\left(1 - d^{-6/5}\right)} \tag{81}$$

where N still depends only on δ (through the dependence of d on δ). Consider any level $M^{2^{m-1}} < r \leq M^{2^m}$ with some associated $m \in \mathbb{N}\backslash\{0\}$. We present a *Monte Carlo algorithm* that constructs the distinct expander graphs making up levels $0, \ldots, r$ of the deterministic DAG in the theorem statement with strictly positive success probability (that depends on δ but not on r) in polylogarithmic time in r. As in the previous algorithm, we ideally want to output $m + 1$ d-regular bipartite lossless $(d^{-6/5}, d - 2d^{4/5})$-expander graphs $B_N, B_{2N}, \ldots, B_{2^m N}$. So, using the aforementioned fact 3, for each $i \in \{0, \ldots, m\}$, we can generate a random d-regular bipartite graph $\mathsf{B} = (U_{2^i N}, V_{2^i N}, \mathsf{E})$ with $|U_{2^i N}| = |V_{2^i N}| = 2^i N$ according to the distribution in Corollary 2 in at most $O(2^m N)$ time. The total running time of the algorithm is thus:

$$O((m+1)2^m N) = O(\log(r)\log\log(r)) \tag{82}$$

since $2^m N = \Theta(\log(r))$ and $m = \Theta(\log\log(r))$ as before. Furthermore, by Corollary 2, the outputted random graphs satisfy (14) for all relevant subsets of vertices with probability at least:

$$\prod_{i=0}^{m}\left(1 - \frac{e}{2\pi\sqrt{d^{-6/5}\left(1 - d^{-6/5}\right)2^i N}}\right) \geq 1 - \frac{e}{2\pi\sqrt{d^{-6/5}\left(1 - d^{-6/5}\right)N}}\sum_{i=0}^{m}\left(\frac{1}{\sqrt{2}}\right)^i$$

---

[14]In our descriptions and analyses of the two algorithms, the big-O and big-Θ asymptotic notation conceal constants that depend on the fixed crossover probability parameter δ.

$$\geq 1 - \frac{e}{2\pi\sqrt{d^{-6/5}\left(1 - d^{-6/5}\right)N}} \sum_{i=0}^{\infty} \left(\frac{1}{\sqrt{2}}\right)^i$$

$$= 1 - \frac{e}{\left(2 - \sqrt{2}\right)\pi\sqrt{d^{-6/5}\left(1 - d^{-6/5}\right)N}} > 0 \qquad (83)$$

where the first inequality is easily proved by induction, and the quantity in the final equality is strictly positive by assumption (81). Hence, our Monte Carlo algorithm constructs the constituent expander graphs in levels $0, \ldots, r$ of our DAG with strictly positive success probability in polylogarithmic time. Once again, note that we neglect details of how intermediate graphs are represented in our analysis. Moreover, we do not account for the running time of actually printing out levels $0, \ldots, r$ of the DAG.

Finally, the aforementioned fact 2 conveys that testing whether the $m + 1$ d-regular random bipartite graphs our Monte Carlo algorithm generates are lossless $(d^{-6/5}, d - 2d^{4/5})$-expander graphs takes polynomial running time:

$$O\left((m+1)2^{2m}N^2 \exp(2^m N H(d^{-6/5}))\right) = O\left(\log\log(r)\log(r)^2 r^{8d^{12/5}H(d^{-6/5})}\right) \qquad (84)$$

where we use the fact that $2^m N < 2N\log(r)/\log(M) = 8d^{12/5}\log(r)$ since $r > M^{2^{m-1}}$ and $\log(M) = N/(4d^{12/5})$. Therefore, by repeatedly running our Monte Carlo algorithm until a valid set of $m + 1$ d-regular bipartite lossless $(d^{-6/5}, d - 2d^{4/5})$-expander graphs is produced, we obtain a *Las Vegas algorithm* that runs in expected polynomial time $O\left(\log\log(r)\log(r)^2 r^{8d^{12/5}H(d^{-6/5})}\right)$.

Feasibility of Broadcasting: We now prove that broadcasting is possible on the Bayesian network defined on the DAG constructed in the theorem statement. As before, we follow the proof of Theorem 1 in section III. So, we first construct a monotone Markovian coupling $\{(X_k^-, X_k^+) : k \in \mathbb{N}\}$ between the Markov chains $\{X_k^+ : k \in \mathbb{N}\}$ and $\{X_k^- : k \in \mathbb{N}\}$ (which denote versions of the Markov chain $\{X_k : k \in \mathbb{N}\}$ initialized at $X_0^+ = 1$ and $X_0^- = 0$, respectively) such that along any edge BSC of the deterministic DAG, say $(X_{k,j}, X_{k+1,i})$, $X_{k,j}^+$ and $X_{k,j}^-$ are either both copied with probability $1 - 2\delta$, or a shared independent Bernoulli$\left(\frac{1}{2}\right)$ bit is produced with probability 2δ that becomes the value of both $X_{k+1,i}^+$ and $X_{k+1,i}^-$. This coupling satisfies the three properties delineated at the outset of the proof of Theorem 1 in section III. Furthermore, let σ_k^+ and σ_k^- for $k \in \mathbb{N}$ be random variables with distributions $P_{\sigma_k|\sigma_0=1}$ and $P_{\sigma_k|\sigma_0=0}$, respectively (which means that $\sigma_0^+ = 1$ and $\sigma_0^- = 0$).

Notice that Lemma 1 implies the following result:

$$\mathbb{P}\left(\sigma_k^- \leq \frac{1}{d^{6/5}} \,\bigg|\, \sigma_{k-1}^- \leq \frac{1}{d^{6/5}}\right) \geq 1 - \exp\left(-\frac{L_{k-1}}{2d^{12/5}}\right) \qquad (85)$$

for every pair of consecutive levels $k - 1$ and k such that $L_k = L_{k-1}$. Moreover, for every pair of consecutive levels $k - 1$ and k such that $L_k = 2L_{k-1}$, we have:

$$\mathbb{P}\left(\sigma_k^- > \frac{1}{d^{6/5}} \,\bigg|\, \sigma_{k-1}^- \leq \frac{1}{d^{6/5}}\right)$$

$$= \mathbb{P}\left(\frac{1}{L_{k-1}}\sum_{i=0}^{L_{k-1}} X_{k,i}^- + \frac{1}{L_{k-1}}\sum_{j=L_{k-1}+1}^{L_k} X_{k,j}^- > \frac{2}{d^{6/5}} \,\bigg|\, \sigma_{k-1}^- \leq \frac{1}{d^{6/5}}\right)$$

$$= \mathbb{P}\left(\left\{\frac{1}{L_{k-1}}\sum_{i=0}^{L_{k-1}} X_{k,i}^- > \frac{1}{d^{6/5}}\right\} \cup \left\{\frac{1}{L_{k-1}}\sum_{j=L_{k-1}+1}^{L_k} X_{k,j}^- > \frac{1}{d^{6/5}}\right\} \,\bigg|\, \sigma_{k-1}^- \leq \frac{1}{d^{6/5}}\right)$$

$$\leq \mathbb{P}\left(\frac{1}{L_{k-1}} \sum_{i=0}^{L_{k-1}} X_{k,i}^- > \frac{1}{d^{6/5}} \,\Big|\, \sigma_{k-1}^- \leq \frac{1}{d^{6/5}} \right) + \mathbb{P}\left(\frac{1}{L_{k-1}} \sum_{j=L_{k-1}+1}^{L_k} X_{k,j}^- > \frac{1}{d^{6/5}} \,\Big|\, \sigma_{k-1}^- \leq \frac{1}{d^{6/5}} \right)$$

$$\leq 2 \exp\left(-\frac{L_{k-1}}{2d^{12/5}} \right)$$

where the first inequality follows from the union bound, and the final inequality follows from Lemma 1 and the construction of our DAG (recall that two separate d-regular bipartite lossless $(d^{-6/5}, d - 2d^{4/5})$-expander graphs make up the edges between X_{k-1} and X_k^1, and between X_{k-1} and X_k^2, respectively). This implies that:

$$\mathbb{P}\left(\sigma_k^- \leq \frac{1}{d^{6/5}} \,\Big|\, \sigma_{k-1}^- \leq \frac{1}{d^{6/5}} \right) \geq 1 - 2 \exp\left(-\frac{L_{k-1}}{2d^{12/5}} \right) \tag{86}$$

for every pair of consecutive levels $k - 1$ and k such that $L_k = 2L_{k-1}$, as well as for every pair of consecutive levels $k - 1$ and k such that $L_k = L_{k-1}$ (by slackening the bound in (85)). Hence, the bound in (86) holds for all levels $k \geq 2$.

Now fix any $\tau > 0$, and choose a sufficiently large value $K = K(\delta, \tau) \in \mathbb{N}$ (that depends on δ and τ) such that:

$$2 \sum_{k=K+1}^{\infty} \exp\left(-\frac{L_{k-1}}{2d^{12/5}} \right) \leq \tau. \tag{87}$$

Note that such K exists because $2 \sum_{k=1}^{\infty} 1/k^2 = \pi^2/3 < +\infty$, and for every $m \in \mathbb{N}$ and every $M^{\lfloor 2^{m-1} \rfloor} < k \leq M^{2^m}$, we have:

$$\exp\left(-\frac{L_k}{2d^{12/5}} \right) \leq \frac{1}{k^2} \quad \Leftrightarrow \quad k \leq \exp\left(\frac{2^m N}{4d^{12/5}} \right) = M^{2^m}$$

where the right hand side holds due to the construction of our deterministic DAG. Using the continuity of probability measures, observe that:

$$\mathbb{P}\left(\bigcap_{k>K} \left\{ \sigma_k^- \leq \frac{1}{d^{6/5}} \right\} \,\Big|\, \sigma_K^+ \geq 1 - \frac{1}{d^{6/5}}, \sigma_K^- \leq \frac{1}{d^{6/5}} \right) = \prod_{k>K} \mathbb{P}\left(\sigma_k^- \leq \frac{1}{d^{6/5}} \,\Big|\, \sigma_{k-1}^- \leq \frac{1}{d^{6/5}}, A_k \right)$$

$$\geq \prod_{k>K} 1 - 2 \exp\left(-\frac{L_{k-1}}{2d^{12/5}} \right)$$

$$\geq 1 - 2 \sum_{k>K} \exp\left(-\frac{L_{k-1}}{2d^{12/5}} \right)$$

$$\geq 1 - \tau$$

where A_k for $k > K$ is the non-zero probability event defined as:

$$A_k \triangleq \begin{cases} \left\{ \sigma_K^+ \geq 1 - \frac{1}{d^{6/5}} \right\} & , \quad k = K + 1 \\ \left\{ \sigma_{k-2}^- \leq \frac{1}{d^{6/5}}, \dots, \sigma_K^- \leq \frac{1}{d^{6/5}} \right\} \cap \left\{ \sigma_K^+ \geq 1 - \frac{1}{d^{6/5}} \right\} & , \quad k \geq K + 2 \end{cases},$$

the first inequality follows from (86), and the final inequality follows from (87). When using (86) in the calculation above, we can neglect the effect of the conditioning event A_k, because a careful perusal of the proof of Lemma 1 (which yields (86) as a consequence) shows that (86) continues to hold when we condition on events like A_k. Indeed, in step (75) of the proof, the random variables $\{X_v : v \in V \setminus T\}$ are conditionally independent of the σ-*algebra generated by random variables in previous layers of the DAG* given $V \setminus T$ and $\{N_v : v \in V \setminus T\}$. Moreover,

this observation extends appropriately to the current Markovian coupling setting. We have omitted these details from Lemma 1 for the sake of clarify. Therefore, we have for any $k > K$:

$$\mathbb{P}\left(\sigma_k^- \leq \frac{1}{d^{6/5}} \,\middle|\, \sigma_K^+ \geq 1 - \frac{1}{d^{6/5}}, \, \sigma_K^- \leq \frac{1}{d^{6/5}}\right) \geq 1 - \tau. \tag{88}$$

Likewise, due to the symmetry of the role of 0's and 1's in our deterministic DAG model, we can also prove mutatis mutandis that for any $k > K$:

$$\mathbb{P}\left(\sigma_k^+ \geq 1 - \frac{1}{d^{6/5}} \,\middle|\, \sigma_K^+ \geq 1 - \frac{1}{d^{6/5}}, \, \sigma_K^- \leq \frac{1}{d^{6/5}}\right) \geq 1 - \tau \tag{89}$$

where τ and K in (89) can be chosen to be the same as those in (88) without loss of generality.

Finally, define the event $E = \{\sigma_K^+ \geq 1 - \frac{1}{d^{6/5}}, \sigma_K^- \leq \frac{1}{d^{6/5}}\}$, and observe that for all $k > K$:

$$\mathbb{P}\left(\sigma_k^+ \geq \frac{1}{2}\right) - \mathbb{P}\left(\sigma_k^- \geq \frac{1}{2}\right) \geq \mathbb{E}\left[\left(\mathbb{1}\left\{\sigma_k^+ \geq \frac{1}{2}\right\} - \mathbb{1}\left\{\sigma_k^- \geq \frac{1}{2}\right\}\right)\mathbb{1}\{E\}\right]$$

$$= \left(\mathbb{P}\left(\sigma_k^+ \geq \frac{1}{2} \,\middle|\, E\right) - \mathbb{P}\left(\sigma_k^- \geq \frac{1}{2} \,\middle|\, E\right)\right)\mathbb{P}(E)$$

$$\geq \left(\mathbb{P}\left(\sigma_k^+ \geq 1 - \frac{1}{d^{6/5}} \,\middle|\, E\right) - \mathbb{P}\left(\sigma_k^- > \frac{1}{d^{6/5}} \,\middle|\, E\right)\right)\mathbb{P}(E)$$

$$\geq (1 - 2\tau)\mathbb{P}(E) > 0$$

where the first inequality holds because $\mathbb{1}\{\sigma_k^+ \geq \frac{1}{2}\} - \mathbb{1}\{\sigma_k^- \geq \frac{1}{2}\} \geq 0$ almost surely due to the monotonicity of our Markovian coupling, the second inequality holds because $\frac{1}{d^{6/5}} < \frac{1}{2} < 1 - \frac{1}{d^{6/5}}$ (since $d \geq 5$), and the final inequality follows from (88) and (89). As argued in the proof of Theorem 1 in section III, this illustrates that $\limsup_{k\to\infty} \mathbb{P}(\hat{S}_k \neq X_0) < \frac{1}{2}$, which completes the proof. ∎

VI. CONCLUSION

To conclude, we recapitulate the main contributions of this work. For random DAG models with indegree $d \geq 3$, we considered the intuitively reasonable setting where all Boolean processing functions are the majority rule. We proved in Theorem 1 that reconstruction of the root bit for this model is possible using the majority decision rule when $\delta < \delta_{\mathsf{maj}}$ and $L_k = \Omega(\log(k))$, and impossible using the ML decision rule in all but a zero measure set of DAGs when $\delta > \delta_{\mathsf{maj}}$ and L_k is sub-exponential. On the other hand, when the indegree $d = 2$ so that the choices of Boolean processing functions are unclear, we derived a similar phase transition in Theorem 2 for random DAG models with AND processing functions at all even levels and OR processing functions at all odd levels. These main results on random DAG models established the existence of deterministic DAGs where broadcasting is possible via the probabilistic method. For example, we conveyed in Corollary 1 that for any indegree $d \geq 3$, any noise level $\delta < \delta_{\mathsf{maj}}$, and $L_k = \Theta(\log(k))$, there exists a deterministic DAG with all majority processing functions such that reconstruction of the root bit is possible. In fact, Proposition 2 showed that the scaling $L_k = \Theta(\log(k))$ is optimal for such DAGs where broadcasting is possible. Finally, for any $\delta \in (0, \frac{1}{2})$ and any sufficiently large bounded indegrees and outdegrees, we constructed explicit deterministic DAGs with $L_k = \Theta(\log(k))$ and all majority processing functions such that broadcasting is possible in Theorem 3. Our construction utilized regular bipartite lossless expander graphs between successive layers of the DAGs, and we showed that the constituent expander

graphs can be generated in either deterministic quasi-polynomial time or randomized polylogarithmic time in the number of levels.

We close our discussion with a brief list of open problems that could serve as compelling directions for future research:

1) We conjectured in subsection II-A that in the random DAG model with $L_k = O(\log(k))$ and fixed $d \geq 3$, reconstruction is impossible for all choices of Boolean processing functions when $\delta \geq \delta_{\mathsf{maj}}$. Naturally, the analogous question for $d = 2$ is also open. Based on the reliable computation literature (see the discussion in II-A), we can conjecture that majority processing functions are optimal for odd $d \geq 3$, and alternating levels of AND and OR processing is optimal for $d = 2$, but it is not obvious which processing functions are optimal for general even $d \geq 4$.

2) We provided some evidence for the previous conjecture in the odd $d \geq 3$ case in part 2 of Proposition 1. A potentially simpler open question is to extend the proof of part 2 of Proposition 1 in Appendix A to show the impossibility of reconstruction using two (or more) vertices in the odd $d \geq 3$ case regardless of the choices of Boolean processing functions.

3) It is unknown whether a result similar to part 2 of Proposition 1 holds for even $d \geq 2$. For the $d = 2$ setting, a promising direction is to try and exploit the *potential function contraction* approach in [32] instead of the TV distance contraction approach in [28], [29].

4) As mentioned in subsection II-B, it is an open problem to find a deterministic polynomial time algorithm to construct deterministic DAGs with sufficiently large d and $L_k = \Theta(\log(k))$ given some δ for which broadcasting is possible. Indeed, the deterministic algorithm in Theorem 3 takes quasi-polynomial time.

5) As indicated above, for fixed δ, Theorem 3 can only construct deterministic DAGs with sufficiently large d such that broadcasting is possible. However, Corollary 1 elucidates that such deterministic DAGs exist for every $d \geq 3$ as long as $\delta < \delta_{\mathsf{maj}}$. It is an open problem to efficiently construct deterministic DAGs with $L_k = \Theta(\log(k))$ for arbitrary $d \geq 3$ and $\delta < \delta_{\mathsf{maj}}$, or $d = 2$ and $\delta < \delta_{\mathsf{andor}}$, such that broadcasting is possible.

APPENDIX A
PROOF OF PROPOSITION 1

Proof. In this proof, we assume familiarity with the development and notation in section III and the proof of Theorem 1.

Part 1: We first prove part 1. Observe that for any $k \in \mathbb{N} \backslash \{0\}$:

$$
\begin{aligned}
\mathbb{P}(X_{k,0} \neq X_{0,0}) &= \frac{1}{2}\mathbb{P}\Big(X_{k,0}^+ = 0\Big) + \frac{1}{2}\mathbb{P}\Big(X_{k,0}^- = 1\Big) \\
&= \frac{1}{2}\mathbb{E}\Big[\mathbb{P}\Big(X_{k,0}^+ = 0\Big|\sigma_k^+\Big)\Big] + \frac{1}{2}\mathbb{E}\Big[\mathbb{P}\Big(X_{k,0}^- = 1\Big|\sigma_k^-\Big)\Big] \\
&= \frac{1}{2}\mathbb{E}\big[1 - \sigma_k^+\big] + \frac{1}{2}\mathbb{E}\big[\sigma_k^-\big] \\
&= \frac{1}{2}\big(1 - \mathbb{E}\big[\sigma_k^+ - \sigma_k^-\big]\big)
\end{aligned}
\tag{90}
$$

where the third equality holds because $X_{k,0} \sim \text{Bernoulli}(\sigma)$ given $\sigma_k = \sigma$. To see this, recall the relation (3) from subsection I-C. Using this relation, it is straightforward to verify that X_k is conditionally independent of $X_{0,0}$ given σ_k. Moreover, the conditional distribution $P_{X_k|\sigma_k}$ can be computed using (3), and this yields the desired conditional distribution $P_{X_{k,0}|\sigma_k}$ mentioned above. (We omit these calculations because it is intuitively obvious that random bits at level k can be generated by first generating σ_k, then setting a uniformly and randomly chosen subset of vertices in X_k of size $L_k \sigma_k$ to be 1, and finally setting the remaining vertices in X_k to be 0.)

Due to (90), it suffices to prove that $\liminf_{k \to \infty} \mathbb{E}[\sigma_k^+ - \sigma_k^-] > 0$. To this end, recall from the proof of Theorem 1 that for any sufficiently small $\epsilon = \epsilon(\delta, d) > 0$ (that depends on δ and d) and any $\tau > 0$, there exists $K = K(\epsilon, \tau) \in \mathbb{N}$ (that depends on ϵ and τ) such that for all $k > K$, (38) and (39) (which are reproduced below) hold:

$$\mathbb{P}(A|E) \geq 1 - \tau \tag{91}$$

$$\mathbb{P}(B|E) \geq 1 - \tau \tag{92}$$

where the events are $A = \{\sigma_k^+ \geq \hat{\sigma} - \epsilon\}$, $B = \{\sigma_k^- \leq 1 - \hat{\sigma} + \epsilon\}$, and $E = \{\sigma_K^+ \geq \hat{\sigma} - \epsilon, \sigma_K^- \leq 1 - \hat{\sigma} + \epsilon\}$, respectively. Now notice that for all $k > K$:

$$
\begin{aligned}
\mathbb{E}[\sigma_k^+ - \sigma_k^-] &= \mathbb{E}[\sigma_k^+ - \sigma_k^- | E] \mathbb{P}(E) + \mathbb{E}[\sigma_k^+ - \sigma_k^- | E^c] \mathbb{P}(E^c) \\
&\geq \mathbb{E}[\sigma_k^+ - \sigma_k^- | E] \mathbb{P}(E) \\
&= \mathbb{P}(E) \left(\mathbb{E}[\sigma_k^+ | E, A] \mathbb{P}(A|E) + \mathbb{E}[\sigma_k^+ | E, A^c] \mathbb{P}(A^c|E) - \mathbb{E}[\sigma_k^- | E] \right) \\
&\geq \mathbb{P}(E) \left(\mathbb{E}[\sigma_k^+ | E, A] \mathbb{P}(A|E) - \mathbb{E}[\sigma_k^- | E] \right) \\
&= \mathbb{P}(E) \left(\mathbb{E}[\sigma_k^+ | E, A] \mathbb{P}(A|E) - \mathbb{E}[\sigma_k^- | E, B] \mathbb{P}(B|E) - \mathbb{E}[\sigma_k^- | E, B^c] \mathbb{P}(B^c|E) \right) \\
&\geq \mathbb{P}(E) \left(\mathbb{E}[\sigma_k^+ | E, A] \mathbb{P}(A|E) - \mathbb{E}[\sigma_k^- | E, B] - \mathbb{P}(B^c|E) \right) \\
&\geq \mathbb{P}(E) \left(\mathbb{E}[\sigma_k^+ | E, A] (1 - \tau) - \mathbb{E}[\sigma_k^- | E, B] - \tau \right) \\
&\geq \mathbb{P}(E) \left((\hat{\sigma} - \epsilon)(1 - \tau) - (1 - \hat{\sigma} + \epsilon) - \tau \right) \\
&= \mathbb{P}(E) \left(\hat{\sigma} - (1 - \hat{\sigma}) - 2\epsilon - \tau(1 + \hat{\sigma} - \epsilon) \right) > 0
\end{aligned}
$$

where the second line holds because $\sigma_k^+ \geq \sigma_k^-$ almost surely (monotonicity), the fourth line holds because $\sigma_k^+ \geq 0$ almost surely, the sixth line holds because $\sigma_k^- \leq 1$ almost surely, the seventh line follows from (91) and (92), the eighth line follows from the definitions of A and B, and the quantity in the ninth line does not depend on k and is strictly positive for sufficiently small ϵ and τ (which now depends on δ and d) because $\hat{\sigma} > 1 - \hat{\sigma}$. Therefore, $\liminf_{k \to \infty} \mathbb{E}[\sigma_k^+ - \sigma_k^-] > 0$, which completes the proof of part 1.

Part 2: We next prove part 2. We begin with a few seemingly unrelated observations that will actually be quite useful later. Recall that $R_k = \inf_{n \geq k} L_n$ for every $k \in \mathbb{N}$ and $R_k = O(d^{2k})$. Hence, there exists a constant $\alpha = \alpha(\delta, d) > 0$ (that depends on δ and d) such that for all sufficiently large k (depending on δ and d), we have:

$$R_k \leq \alpha d^{2k}. \tag{93}$$

Let $\beta = \frac{\log(\alpha)}{6\log(d)}$, and define the sequence $\{m(k) \in \mathbb{N}\}$ (indexed by k) as:

$$m = m(k) \triangleq \left\lfloor \frac{\log\big(R_{\lfloor(2k/3)-\beta\rfloor}\big)}{4\log(d)} \right\rfloor \tag{94}$$

where $\lfloor \cdot \rfloor$ denotes the floor function, and $\frac{2k}{3} \geq \beta$ for all sufficiently large k (depending on δ and d) so that the sequence is eventually well-defined. This sequence satisfies the following conditions:

$$\lim_{k\to\infty} m(k) = \infty \,, \tag{95}$$

$$\lim_{k\to\infty} \frac{d^{2m}}{R_{k-m}} = 0 \,. \tag{96}$$

The first limit (95) holds because $\lim_{k\to\infty} R_k = \liminf_{k\to\infty} L_k = \infty$ (by assumption), and the second limit (96) is true because for all sufficiently large k (depending on δ and d):

$$\frac{d^{2m}}{R_{k-m}} \leq \frac{\sqrt{R_{\lfloor(2k/3)-\beta\rfloor}}}{R_{k-m}} \leq \frac{\sqrt{R_{\lfloor(2k/3)-\beta\rfloor}}}{R_{\lfloor(2k/3)-\beta\rfloor}} = \frac{1}{\sqrt{R_{\lfloor(2k/3)-\beta\rfloor}}}$$

where the first inequality follows from (94), and the second inequality holds because $\{R_k : k \in \mathbb{N}\}$ is non-decreasing, and $m \leq \big(\log\big(\alpha d^{(4k/3)-2\beta}\big)\big)/(4\log(d)) = \frac{k}{3} + \beta$ for all sufficiently large k using (93) and (94).

We next establish that a small portion of the random DAG G above the vertex $X_{k,0}$ is a directed tree with high probability. To this end, for any sufficiently large $k \in \mathbb{N}$ (depending on δ and d) such that $k - m \geq 0$, let G_k denote the (random) induced subgraph of the random DAG G consisting of all vertices in levels $k - m, \ldots, k$ that have a path to $X_{k,0}$, where $m = m(k)$ is defined in (94). (Note that $X_{k,0}$ always has a path to itself.) Moreover, define the event $T_k \triangleq \{G_k \text{ is a directed tree}\}$. Now, for any sufficiently large k (depending on δ and d) such that $d^{2r} \leq R_{k-r} \leq L_{k-r}$ for every $r \in \{1, \ldots, m\}$ (which is feasible due to (96), and ensures that the ensuing steps are valid), notice that:

$$\begin{aligned}
\mathbb{P}(T_k) &= \prod_{r=1}^{m} \prod_{s=0}^{d^r-1} \left(1 - \frac{s}{L_{k-r}}\right) \\
&\geq \prod_{r=1}^{m} \left(1 - \frac{1}{L_{k-r}} \sum_{s=0}^{d^r-1} s\right) \\
&= \prod_{r=1}^{m} \left(1 - \frac{d^r(d^r-1)}{2L_{k-r}}\right) \\
&\geq 1 - \frac{1}{2} \sum_{r=1}^{m} \frac{d^r(d^r-1)}{L_{k-r}} \\
&\geq 1 - \frac{1}{2R_{k-m}} \sum_{r=1}^{m} d^{2r} \\
&= 1 - \frac{1}{2R_{k-m}} \left(\frac{d^2(d^{2m}-1)}{d^2-1}\right) \\
&\geq 1 - \left(\frac{d^2}{2(d^2-1)}\right) \frac{d^{2m}}{R_{k-m}} \tag{97}
\end{aligned}$$

where the first equality holds because the edges of G are chosen randomly and independently and we must ensure that the parents of every vertex in G_k are distinct, the second and fourth inequalities are straightforward to prove by induction, and the third and sixth equalities follow from arithmetic and geometric series computations, respectively.

The bound in (97) conveys that $\lim_{k\to\infty} \mathbb{P}(T_k) = 1$ due to (96), i.e. G_k is a directed tree with high probability for large k.

We introduce some useful notation for the remainder of this proof. First, condition on any realization of the random DAG G such that the event T_k occurs (for sufficiently large k such that (97) holds). This also fixes the choices of Boolean processing functions at the vertices (which may vary between vertices and be graph dependent). For any vertex $X_{n,j}$ in the tree G_k with $n < k$, let $\tilde{X}_{n,j}$ denote the output of the edge $\mathsf{BSC}(\delta)$ with input $X_{n,j}$ in G_k. (Hence, $\tilde{X}_{n,j}$ is the input of a Boolean processing function at a single vertex in level $n+1$ of G_k.) On the other hand, let $\tilde{X}_{k,0}$ be the output of an independent $\mathsf{BSC}(\delta)$ channel (which is not necessarily in G) with input $X_{k,0}$. Since G_k is a tree, the random variables $\{\tilde{X}_{n,j} : X_{n,j}$ is a vertex of $G_k\}$ describe the values at the gates of a *noisy formula* $\tilde{G}_k$, where the Boolean functions in G_k correspond to d-input δ-noisy gates in $\tilde{G}_k$ (and we think of the independent BSC errors as occurring at the gates rather than the edges). Next, in addition to conditioning on G and T_k, we also condition on one of two realizations $X_{k-m} = x_0$ or $X_{k-m} = x_1$ for any $x_0, x_1 \in \{0,1\}^{L_{k-m}}$. In particular, corresponding to any binary random variable Y in $\tilde{G}_k$, define the following 2-tuple in $[0,1]^2$, cf. [28], [29]:

$$\lambda^Y \triangleq \left(\mathbb{P}(Y \neq 0 | X_{k-m} = x_0, G, T_k), \mathbb{P}(Y \neq 1 | X_{k-m} = x_1, G, T_k) \right). \tag{98}$$

Lastly, for any constant $a \in [0,1]$, let $S(a) \subseteq [0,1]^2$ be the convex hull of the points $\{(a,a), (1-a, 1-a), (0,1), (1,0)\}$, cf. [28], [29]. With these definitions, we can state a version of the pivotal lemma in [29, Lemma 2], which was proved in the $d = 3$ case in [28].

Lemma 2 (TV Distance Contraction in Noisy Formulae [29, Lemma 2]). *If $d \geq 3$ is odd and $\delta \geq \delta_{\mathsf{maj}}$, then for every possible d-input δ-noisy gate in $\tilde{G}_k$ with inputs $Y_1, \ldots, Y_d$ and output Y, we have:*

$$\lambda^{Y_1}, \ldots, \lambda^{Y_d} \in S(a) \text{ with } a \in \left[0, \frac{1}{2}\right] \implies \lambda^Y \in S(f(a))$$

where the function $f : [0,1] \to [0,1]$ is defined in (10).

We remark that Lemma 2 differs from [29, Lemma 2] in the definition of the 2-tuple λ^Y for any binary random variable Y in the noisy formula. Since [29, Lemma 2] is used to yield the impossibility results on reliable computation discussed in subsection II-A, [29, Section III] defines λ^Y for this purpose as $\lambda^Y = (\mathbb{P}(Y \neq X | X = 0), \mathbb{P}(Y \neq X | X = 1))$, where X is a single relevant binary input random variable of the noisy formula (and all other inputs are fixed). In contrast, we define λ^Y in (98) by conditioning on any two realizations of the random variables X_{k-m}. This ensures that the inputs, say $\tilde{X}_{n,j_1}, \ldots, \tilde{X}_{n,j_d}$ for some $k-m \leq n < k$ and $j_1, \ldots, j_d \in [L_n]$, of every d-input δ-noisy gate in the noisy formula $\tilde{G}_k$ are conditionally independent given X_{k-m}, which is a crucial property required by the proof of [29, Lemma 2]. We omit the proof of Lemma 2 because it is virtually identical to the proof of [29, Lemma 2] in [29, Sections IV and V]. (The reader can verify that every step in the proofs in [29, Sections IV and V] continues to hold with our definition of λ^Y.)

Lemma 2 indeed demonstrates a strong data processing inequality style of contraction for TV distance, cf. [32, Equation (1)]. To see this, observe that $(x,y) \in S(a)$ with $a \in \left[0, \frac{1}{2}\right]$ if and only if $a \leq ax + (1-a)y \leq 1-a$ and

$a \leq ay + (1 - a)x \leq 1 - a$. This implies that $a \leq \frac{x+y}{2} \leq 1 - a$, and hence, $|1 - x - y| \leq 1 - 2a$. Furthermore, for any binary random variable Y in $\tilde{G}_k$, we have using (5):

$$
\begin{aligned}
\left\| P_{Y|G,T_k,X_{k-m}=x_1} - P_{Y|G,T_k,X_{k-m}=x_0} \right\|_{\mathsf{TV}} = |1 - \mathbb{P}(Y \neq 0 | X_{k-m} = x_0, G, T_k) \\
- \mathbb{P}(Y \neq 1 | X_{k-m} = x_1, G, T_k)|
\end{aligned}
\tag{99}
$$

where $P_{Y|G,T_k,X_{k-m}=x}$ denotes the conditional distribution of Y given $\{X_{k-m} = x, G, T_k\}$ for any $x \in \{0,1\}^{L_{k-m}}$. Thus, if $\lambda^Y \in S(a)$ with $a \in \left[0, \frac{1}{2}\right]$, then we get:

$$
\left\| P_{Y|G,T_k,X_{k-m}=x_1} - P_{Y|G,T_k,X_{k-m}=x_0} \right\|_{\mathsf{TV}} \leq 1 - 2a \,.
$$

Now notice that $\lambda^{\tilde{X}_{k-m,j}} \in S(0)$ for every random variable $\tilde{X}_{k-m,j}$ in $\tilde{G}_k$, where $j \in [L_{k-m}]$. As a result, a straightforward induction argument using Lemma 2 (much like that in the proof in [29, Section III]) yields $\lambda^{\tilde{X}_{k,0}} \in S(f^{(m)}(0))$. This implies that:[15]

$$
\left\| P_{\tilde{X}_{k,0}|G,T_k,X_{k-m}=x_1} - P_{\tilde{X}_{k,0}|G,T_k,X_{k-m}=x_0} \right\|_{\mathsf{TV}} \leq 1 - 2f^{(m)}(0) = 1 - 2\left(\delta * g^{(m-1)}(0)\right)
\tag{100}
$$

where the function $g : [0,1] \to [0,1]$ is given in (28) in section III, and the equality follows from (11). Moreover, since $\mathbb{P}(\tilde{X}_{k,0} \neq y | G, T_k, X_{k-m} = x) = \delta * \mathbb{P}(X_{k,0} \neq y | G, T_k, X_{k-m} = x)$ for any $y \in \{0,1\}$ and any $x \in \{0,1\}^{L_{k-m}}$, a simple calculation using (99) shows that:

$$
\left\| P_{\tilde{X}_{k,0}|G,T_k,X_{k-m}=x_1} - P_{\tilde{X}_{k,0}|G,T_k,X_{k-m}=x_0} \right\|_{\mathsf{TV}} = (1 - 2\delta) \left\| P_{X_{k,0}|G,T_k,X_{k-m}=x_1} - P_{X_{k,0}|G,T_k,X_{k-m}=x_0} \right\|_{\mathsf{TV}}
$$

which, using (100), produces:

$$
\left\| P_{X_{k,0}|G,T_k,X_{k-m}=x_1} - P_{X_{k,0}|G,T_k,X_{k-m}=x_0} \right\|_{\mathsf{TV}} \leq \frac{1 - 2\left(\delta * g^{(m-1)}(0)\right)}{1 - 2\delta}
\tag{101}
$$

for any $x_0, x_1 \in \{0,1\}^{L_{k-m}}$. The inequality in (101) conveys a contraction of the TV distance on the left hand side. Since g has only one fixed point at $\frac{1}{2}$ when $\delta \geq \delta_{\mathsf{maj}}$ (see section III), and (95) holds, the fixed point theorem gives us $\lim_{k\to\infty} g^{(m-1)}(0) = \frac{1}{2}$ (where $g^{(m-1)}(0)$ increases to $\frac{1}{2}$). Hence, the upper bound in (101) decreases to 0 as $k \to \infty$. Furthermore, note that (101) holds for all choices of Boolean processing functions (which may vary between vertices and be graph dependent), because Lemma 2 is agnostic to the particular gates used in $\tilde{G}_k$.

Finally, recall that the *Dobrushin contraction coefficient* of any Markov transition kernel $P_{Z|W}$ with input alphabet $\mathcal{W}$ and output alphabet $\mathcal{Z}$, such that $2 \leq |\mathcal{W}|, |\mathcal{Z}| < +\infty$, is defined as [50]:

$$
\eta_{\mathsf{TV}}\left(P_{Z|W}\right) \triangleq \sup_{\substack{P_W, Q_W: \\ P_W \neq Q_W}} \frac{\|P_Z - Q_Z\|_{\mathsf{TV}}}{\|P_W - Q_W\|_{\mathsf{TV}}}
\tag{102}
$$

$$
= \max_{w,w' \in \mathcal{W}} \left\| P_{Z|W=w} - P_{Z|W=w'} \right\|_{\mathsf{TV}} \in [0,1]
\tag{103}
$$

where the supremum in the first equality is over all pairs of distinct probability distributions P_W and Q_W on $\mathcal{W}$, P_Z and Q_Z denote the output distributions on $\mathcal{Z}$ induced by passing P_W and Q_W through $P_{Z|W}$ respectively, the second equality is Dobrushin's two-point characterization of η_{TV} [50], and for any $w \in \mathcal{W}$, $P_{Z|W=w}$ denotes the

---

[15]The inequality in (100) can be perceived as a repeated application of a *tensorized* universal upper bound on the *Dobrushin curve* of any d-input δ-noisy gate, cf. [38, Section II-A].

wth conditional distribution on $\mathcal{Z}$ in $P_{Z|W}$. For any fixed realization of the random DAG G such that T_k occurs (for sufficiently large k such that (97) holds), observe that:

$$
\begin{aligned}
\left\|P_{X_{k,0}|G,T_k,X_{0,0}=1} - P_{X_{k,0}|G,T_k,X_{0,0}=0}\right\|_{\mathsf{TV}} &= \eta_{\mathsf{TV}}\big(P_{X_{k,0}|G,T_k,X_0}\big) \\
&\leq \eta_{\mathsf{TV}}\big(P_{X_{k,0}|G,T_k,X_{k-m}}\big)\,\eta_{\mathsf{TV}}\big(P_{X_{k-m}|G,T_k,X_0}\big) \\
&\leq \max_{x_0,x_1}\left\|P_{X_{k,0}|G,T_k,X_{k-m}=x_1} - P_{X_{k,0}|G,T_k,X_{k-m}=x_0}\right\|_{\mathsf{TV}} \\
&\leq \frac{1 - 2\big(\delta * g^{(m-1)}(0)\big)}{1 - 2\delta}
\end{aligned}
\tag{104}
$$

where $P_{X_{k,0}|G,T_k,X_0}$, $P_{X_{k,0}|G,T_k,X_{k-m}}$, and $P_{X_{k-m}|G,T_k,X_0}$ are transition kernels from X_0 to $X_{k,0}$, from X_{k-m} to $X_{k,0}$, and from X_0 to X_{k-m} respectively, the first line follows from (103) where $P_{X_{k,0}|G,T_k,X_{0,0}=y}$ denotes the conditional distribution of $X_{k,0}$ given $\{X_{0,0} = y, G, T_k\}$ for any $y \in \{0,1\}$, the second line holds because $X_0 \to X_{k-m} \to X_{k,0}$ forms a Markov chain (given G and T_k) and η_{TV} is sub-multiplicative in its input stochastic matrix (this follows easily from (102)—see e.g. [51, Lemma 4.3]), the third line follows from (103) and the maximum here is over all $x_0, x_1 \in \{0,1\}^{L_{k-m}}$, and the last line follows from (101). Taking conditional expectations with respect to G given T_k in (104) yields:

$$
\mathbb{E}\left[\left\|P^+_{X_{k,0}|G} - P^-_{X_{k,0}|G}\right\|_{\mathsf{TV}}\Big|T_k\right] \leq \frac{1 - 2\big(\delta * g^{(m-1)}(0)\big)}{1 - 2\delta}
$$

where $P^+_{X_{k,0}|G}$ and $P^-_{X_{k,0}|G}$ inside the conditional expectation correspond to the conditional probability distributions $P_{X_{k,0}|G,T_k,X_{0,0}=1}$ and $P_{X_{k,0}|G,T_k,X_{0,0}=0}$, respectively (as we condition on T_k). Therefore, we have:

$$
\begin{aligned}
\mathbb{E}\left[\left\|P^+_{X_{k,0}|G} - P^-_{X_{k,0}|G}\right\|_{\mathsf{TV}}\right] &= \mathbb{E}\left[\left\|P^+_{X_{k,0}|G} - P^-_{X_{k,0}|G}\right\|_{\mathsf{TV}}\Big|T_k\right]\mathbb{P}(T_k) \\
&\quad + \mathbb{E}\left[\left\|P^+_{X_{k,0}|G} - P^-_{X_{k,0}|G}\right\|_{\mathsf{TV}}\Big|T_k^c\right](1 - \mathbb{P}(T_k)) \\
&\leq \frac{1 - 2\big(\delta * g^{(m-1)}(0)\big)}{1 - 2\delta} + \left(\frac{d^2}{2(d^2-1)}\right)\frac{d^{2m}}{R_{k-m}}
\end{aligned}
$$

using the tower property, the fact that TV distance is bounded by 1, and (97). Letting $k \to \infty$ establishes the desired result:

$$
\lim_{k\to\infty}\mathbb{E}\left[\left\|P^+_{X_{k,0}|G} - P^-_{X_{k,0}|G}\right\|_{\mathsf{TV}}\right] \leq \frac{1 - 2\big(\delta * \lim_{k\to\infty}g^{(m-1)}(0)\big)}{1 - 2\delta} + \left(\frac{d^2}{2(d^2-1)}\right)\lim_{k\to\infty}\frac{d^{2m}}{R_{k-m}} = 0
$$

because $\lim_{k\to\infty}g^{(m-1)}(0) = \frac{1}{2}$ (as noted earlier) and (96) holds. This completes the proof. ∎

APPENDIX B

PROOF OF COROLLARY 1

Proof. This follows from applying the probabilistic method. Fix any $d \geq 3$, any $\delta \in (0, \delta_{\mathsf{maj}})$, and any sequence of level sizes satisfying $L_k \geq C(\delta, d)\log(k)$ for all sufficiently large k. We know from Theorem 1 that for the random DAG model with these parameters and majority processing functions, there exist $\epsilon = \epsilon(\delta, d) > 0$ and $K = K(\delta, d) \in \mathbb{N}$ (which depend on δ and d) such that:

$$
\forall k \geq K, \ \ \mathbb{P}\Big(\hat{S}_k \neq X_0\Big) \leq \frac{1}{2} - 2\epsilon.
$$

Now define $P_k(G) \triangleq \mathbb{P}(h_{\mathsf{ML}}^k(X_k, G) \neq X_0 | G)$ for $k \in \mathbb{N}$ as the conditional probability that the ML decision rule based on the full k-layer state X_k makes an error given the random DAG G, and let E_k for $k \in \mathbb{N}$ be the set of all deterministic DAGs $\mathcal{G}$ with indegree d and level sizes $\{L_m : m \in \mathbb{N}\}$ such that $P_k(\mathcal{G}) \leq \frac{1}{2} - \epsilon$. Observe that for every $k \geq K$:

$$
\begin{aligned}
\frac{1}{2} - 2\epsilon \geq \mathbb{P}\left(\hat{S}_k \neq X_0\right) &= \mathbb{E}\left[\mathbb{P}\left(\hat{S}_k \neq X_0 \middle| G\right)\right] \\
&\geq \mathbb{E}[P_k(G)] \\
&= \mathbb{E}[P_k(G)|G \in E_k]\,\mathbb{P}(G \in E_k) + \mathbb{E}[P_k(G)|G \notin E_k]\,\mathbb{P}(G \notin E_k) \\
&\geq \mathbb{E}[P_k(G)|G \notin E_k]\,\mathbb{P}(G \notin E_k) \\
&\geq \left(\frac{1}{2} - \epsilon\right)\mathbb{P}(G \notin E_k)
\end{aligned}
$$

where the first and third lines follow from the law of total expectation, the second line holds because the ML decision rule minimizes the probability of error, the fourth line holds because the first term in the previous line is non-negative, and the final line holds because $G \notin E_k$ implies that $P_k(G) > \frac{1}{2} - \epsilon$. Then, we have for every $k \geq K$:

$$
\mathbb{P}(G \in E_k) \geq \frac{2\epsilon}{1 - 2\epsilon} > 0.
$$

Since $\{E_k : k \in \mathbb{N}\}$ form a non-increasing sequence of sets (because $P_k(G)$ is non-decreasing in k), we get via continuity:

$$
\mathbb{P}\left(G \in \bigcap_{k \in \mathbb{N}} E_k\right) = \lim_{k \to \infty} \mathbb{P}(G \in E_k) \geq \frac{2\epsilon}{1 - 2\epsilon} > 0
$$

which means that there exists a deterministic DAG $\mathcal{G}$ with indegree d, noise level δ, level sizes $\{L_k : k \in \mathbb{N}\}$, and majority processing functions such that $P_k(\mathcal{G}) \leq \frac{1}{2} - \epsilon$ for all $k \in \mathbb{N}$. This completes the proof. ∎

APPENDIX C

PROOF OF PROPOSITION 2

Proof.

Part 1: We first prove part 1, where we are given a fixed deterministic DAG $\mathcal{G}$. Observe that the BSC along each edge of this DAG produces its output bit by either copying its input bit exactly with probability $1 - 2\delta$, or generating an independent Bernoulli$\left(\frac{1}{2}\right)$ output bit with probability 2δ. This is because the BSC transition matrix can be decomposed as:[16]

$$
\begin{bmatrix} 1 - \delta & \delta \\ \delta & 1 - \delta \end{bmatrix} = (1 - 2\delta) \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + (2\delta) \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}. \tag{105}
$$

Now consider the events:

$$
A_k \triangleq \{\text{all } dL_k \text{ edges from level } k - 1 \text{ to level } k \text{ generate independent output bits}\}
$$

---

[16]This simple, but useful, idea is a specialization of more sophisticated Fortuin-Kasteleyn random cluster representations of Ising models [41].

for $k \in \mathbb{N}\backslash\{0\}$, which have probabilities $\mathbb{P}(A_k) = (2\delta)^{dL_k}$ since the BSCs on the edges are independent. These events are mutually independent (once again because the BSCs on the edges are independent). Since the condition on L_k in the proposition statement is equivalent to:

$$(2\delta)^{dL_k} \geq \frac{1}{k} \quad \text{for all sufficiently large } k \,,$$

we must have:

$$\sum_{k=1}^{\infty} \mathbb{P}(A_k) = \sum_{k=1}^{\infty} (2\delta)^{dL_k} = +\infty \,.$$

The second Borel-Cantelli lemma then tells us that infinitely many of the events $\{A_k : k \in \mathbb{N}\backslash\{0\}\}$ occur almost surely, i.e. $\mathbb{P}(\bigcap_{m=1}^{\infty} \bigcup_{k=m}^{\infty} A_k) = 1$. In particular, if we define $B_m \triangleq \bigcup_{k=1}^{m} A_k$ for $m \in \mathbb{N}\backslash\{0\}$, then by continuity:

$$\lim_{m \to \infty} \mathbb{P}(B_m) = \mathbb{P}\left(\bigcup_{k=1}^{\infty} A_k\right) = 1 \,. \tag{106}$$

Finally, observe that:

$$\begin{aligned}
\lim_{m \to \infty} \mathbb{P}(h_{\mathsf{ML}}^m(X_m, \mathcal{G}) \neq X_0) &= \lim_{m \to \infty} \mathbb{P}(h_{\mathsf{ML}}^m(X_m, \mathcal{G}) \neq X_0 | B_m) \, \mathbb{P}(B_m) \\
&\quad + \mathbb{P}(h_{\mathsf{ML}}^m(X_m, \mathcal{G}) \neq X_0 | B_m^c) \, \mathbb{P}(B_m^c) \\
&= \lim_{m \to \infty} \mathbb{P}(h_{\mathsf{ML}}^m(X_m, \mathcal{G}) \neq X_0 | B_m) \\
&= \lim_{m \to \infty} \frac{1}{2} \mathbb{P}(h_{\mathsf{ML}}^m(X_m, \mathcal{G}) = 1 | B_m) + \frac{1}{2} \mathbb{P}(h_{\mathsf{ML}}^m(X_m, \mathcal{G}) = 0 | B_m) \\
&= \frac{1}{2}
\end{aligned} \tag{107}$$

where $h_{\mathsf{ML}}^m(\cdot, \mathcal{G}) : \{0, 1\}^{L_m} \to \{0, 1\}$ denotes the ML decision rule at level m based on X_m (given knowledge of the DAG $\mathcal{G}$), the second equality uses (106), and the third equality holds because $X_{0,0} \sim \mathsf{Bernoulli}\left(\frac{1}{2}\right)$ is independent of B_m, and X_m is conditionally independent of X_0 given B_m. The condition in (107) is equivalent to the TV distance condition in part 1 of the proposition statement; this proves part 1.

Part 2: To prove part 2, notice that part 1 immediately yields:

$$\lim_{k \to \infty} \left\| P_{X_k|G}^+ - P_{X_k|G}^- \right\|_{\mathsf{TV}} = 0 \quad \textit{pointwise}$$

which completes the proof. $\blacksquare$

APPENDIX D

PROOF OF PROPOSITION 3

Proof.

Part 1: Fix any noise level $\delta \in \left(0, \frac{1}{2}\right)$ and any constant $\epsilon \in \left(0, \frac{1}{4}\right)$. Furthermore, tentatively suppose that $L_k \geq A(\epsilon, \delta)\sqrt{\log(k)}$ for all sufficiently large k, where the constant $A(\epsilon, \delta)$ is defined as:

$$A(\epsilon, \delta) \triangleq \frac{2}{(1 - 2\delta)\epsilon\sqrt{1 - 2\epsilon}} \,. \tag{108}$$

Now consider the deterministic DAG $\mathcal{G}$ such that each vertex at level $k \in \mathbb{N}\backslash\{0\}$ is connected to all L_{k-1} vertices at level $k - 1$ and all Boolean processing functions are the majority rule. (Note that when there is only one input, the

majority rule behaves like the identity map.) For all $k \in \mathbb{N} \backslash \{0\}$, since X_k is an exchangeable sequence of random variables given σ_0, σ_k is a sufficient statistic of X_k for performing inference about σ_0, where σ_k is defined in (2) (cf. subsection I-C). We next prove a useful "one-step broadcasting" lemma involving σ_k's for this model.

Lemma 3 (One-Step Broadcasting in Unbounded Degree DAG). *Under the aforementioned assumptions, there exists $K = K(\epsilon, \delta) \in \mathbb{N}$ (that depends on ϵ and δ) such that for all $k \geq K$, we have:*

$$\mathbb{P}\left(\sigma_k \geq \frac{1}{2} + \epsilon \,\middle|\, \sigma_{k-1} \geq \frac{1}{2} + \epsilon\right) \geq 1 - \left(\frac{1}{k-1}\right)^2.$$

Proof. Suppose we are given that $\sigma_{k-1} = \sigma \geq \frac{1}{2} + \epsilon$ for any $k \in \mathbb{N} \backslash \{0\}$. Then, $\{X_{k,j} : j \in [L_k]\}$ are conditionally i.i.d. Bernoulli($\mathbb{P}(X_{k,0} = 1|\sigma_{k-1} = \sigma)$) and $L_k \sigma_k \sim$ binomial($L_k, \mathbb{P}(X_{k,0} = 1|\sigma_{k-1} = \sigma)$), where $\mathbb{P}(X_{k,0} = 1|\sigma_{k-1} = \sigma) = \mathbb{E}[\sigma_k|\sigma_{k-1} = \sigma]$. Furthermore, since $X_{k,0}$ is the majority of the values of $X_{k-1,0}, \ldots, X_{k-1,L_{k-1}}$ after passing them through independent BSC(δ)'s, we have:

$$\mathbb{E}[\sigma_k|\sigma_{k-1} = \sigma] = \mathbb{P}(X_{k,0} = 1|\sigma_{k-1} = \sigma)$$

$$= 1 - \mathbb{P}\left(\sum_{i=1}^{L_{k-1}\sigma} Z_i + \sum_{j=1}^{L_{k-1}(1-\sigma)} Y_j < \frac{L_{k-1}}{2}\right)$$

$$\geq 1 - \exp\left(-2L_{k-1}\left(\frac{1}{2} - \sigma * \delta\right)^2\right)$$

$$\geq 1 - \exp\left(-2L_{k-1}\left(\frac{1}{2} - \left(\frac{1}{2} + \epsilon\right) * \delta\right)^2\right)$$

$$= 1 - \exp\left(-2L_{k-1}\epsilon^2(1 - 2\delta)^2\right) \tag{109}$$

where Z_i are i.i.d. Bernoulli($1 - \delta$), Y_j are i.i.d. Bernoulli(δ), $\{Z_i : i \in \{1, \ldots, L_{k-1}\sigma\}\}$ and $\{Y_j : j \in \{1, \ldots, L_{k-1}(1-\sigma)\}\}$ are independent, the first inequality follows from Hoeffding's inequality using the fact that $\sigma * \delta > \frac{1}{2}$ (because $\sigma > \frac{1}{2}$), and the second inequality holds because $\sigma \geq \frac{1}{2} + \epsilon$, which implies that $\sigma * \delta \geq \left(\frac{1}{2} + \epsilon\right) * \delta > \frac{1}{2}$.

Next, observe that there exists $K = K(\epsilon, \delta) \in \mathbb{N}$ (that depends on ϵ and δ) such that for all $k \geq K$, we have:

$$\exp\left(-2L_{k-1}\epsilon^2(1 - 2\delta)^2\right) \leq \epsilon$$

because $\lim_{k \to \infty} L_k = \infty$ by assumption. So, for any $k \geq K$, this yields the bound:

$$\mathbb{E}[\sigma_k|\sigma_{k-1} = \sigma] \geq 1 - \epsilon > \frac{1}{2} + \epsilon$$

using (109) and the fact that $\epsilon < \frac{1}{4}$. As a result, we can apply the Chernoff-Hoeffding bound, cf. [44, Theorem 1], to σ_k for any $k \geq K$ and get:

$$\mathbb{P}\left(\sigma_k < \frac{1}{2} + \epsilon \,\middle|\, \sigma_{k-1} = \sigma\right) \leq \exp\left(-L_k D\left(\frac{1}{2} + \epsilon \,\middle\|\, \mathbb{E}[\sigma_k|\sigma_{k-1} = \sigma]\right)\right)$$

where $D(\alpha||\beta) \triangleq \alpha \log(\alpha/\beta) + (1-\alpha) \log((1-\alpha)/(1-\beta))$ for $\alpha, \beta \in (0,1)$ denotes the binary *Kullback-Leibler divergence* (or relative entropy) function. Notice that:

$$D\left(\frac{1}{2} + \epsilon \,\bigg|\bigg|\, \mathbb{E}[\sigma_k | \sigma_{k-1} = \sigma]\right) \geq -H\left(\frac{1}{2} + \epsilon\right) - \left(\frac{1}{2} - \epsilon\right) \log(1 - \mathbb{E}[\sigma_k | \sigma_{k-1} = \sigma])$$

$$\geq L_{k-1} \epsilon^2 (1 - 2\epsilon)(1 - 2\delta)^2 - H\left(\frac{1}{2} + \epsilon\right)$$

where the first inequality holds because $\log(\mathbb{E}[\sigma_k | \sigma_{k-1} = \sigma]) < 0$, and the second inequality follows from (109). Hence, we have for any $k \geq K$:

$$\mathbb{P}\left(\sigma_k < \frac{1}{2} + \epsilon \,\bigg|\, \sigma_{k-1} = \sigma\right) \leq \exp\left(-L_{k-1} L_k \epsilon^2 (1 - 2\epsilon)(1 - 2\delta)^2 + L_k H\left(\frac{1}{2} + \epsilon\right)\right)$$

where we can multiply both sides by $\mathbb{P}(\sigma_{k-1} = \sigma)$ and then sum over all $\sigma \geq \frac{1}{2} + \epsilon$ (as in the proof of (34) within the proof of Theorem 1 in section III) to get:

$$\mathbb{P}\left(\sigma_k < \frac{1}{2} + \epsilon \,\bigg|\, \sigma_{k-1} \geq \frac{1}{2} + \epsilon\right) \leq \exp\left(-L_{k-1} L_k \left(\epsilon^2 (1 - 2\epsilon)(1 - 2\delta)^2 - \frac{1}{L_{k-1}} H\left(\frac{1}{2} + \epsilon\right)\right)\right).$$

Since $\lim_{k \to \infty} L_k = \infty$ by assumption, we can choose $K = K(\epsilon, \delta)$ to be sufficiently large so that for all $k \geq K$, we also have $H\left(\frac{1}{2} + \epsilon\right)/L_{k-1} \leq \epsilon^2 (1 - 2\epsilon)(1 - 2\delta)^2/2$. Thus, for every $k \geq K$:

$$\mathbb{P}\left(\sigma_k < \frac{1}{2} + \epsilon \,\bigg|\, \sigma_{k-1} \geq \frac{1}{2} + \epsilon\right) \leq \exp\left(-L_{k-1} L_k \frac{\epsilon^2 (1 - 2\epsilon)(1 - 2\delta)^2}{2}\right). \tag{110}$$

Finally, we once again increase $K = K(\epsilon, \delta)$ if necessary to ensure that $L_{k-1} \geq A(\epsilon, \delta)\sqrt{\log(k-1)}$ for every $k \geq K$ (as presumed earlier). This implies that for all $k \geq K$:

$$L_{k-1} L_k \geq A(\epsilon, \delta)^2 \sqrt{\log(k-1)\log(k)} \geq A(\epsilon, \delta)^2 \log(k-1)$$

which, using (108) and (110), produces:

$$\mathbb{P}\left(\sigma_k < \frac{1}{2} + \epsilon \,\bigg|\, \sigma_{k-1} \geq \frac{1}{2} + \epsilon\right) \leq \exp(-2 \log(k-1)) = \left(\frac{1}{k-1}\right)^2$$

for all $k \geq K$. This proves Lemma 3. $\blacksquare$

Lemma 3 is an analogue of (34) in the proof of Theorem 1 in section III. It illustrates that if the proportion of 1's is large in a given layer of $\mathcal{G}$, then it remains large in the next layer of $\mathcal{G}$ with high probability.

To proceed, we specialize Lemma 3 by arbitrarily selecting a particular value of ϵ, say $\epsilon = \frac{7}{32} \in \left(0, \frac{1}{4}\right)$. This implies that the constant $A(\epsilon, \delta)$ becomes:

$$A(\delta) = A\left(\frac{7}{32}, \delta\right) = \frac{256}{21(1 - 2\delta)} \tag{111}$$

using (108). In the proposition statement, it is assumed that $L_k \geq A(\delta)\sqrt{\log(k)}$ for all sufficiently large k. Thus, Lemma 3 holds with $\epsilon = \frac{7}{32} \in \left(0, \frac{1}{4}\right)$ under the assumptions of part 1 of Proposition 3. At this point, we can execute the proof of part 1 of Theorem 1 in section III mutatis mutandis (with Lemma 3 playing the pivotal role of (34)) to establish part 1 of Proposition 3. We omit the details of this proof for brevity.

Part 2: To prove part 2, we use the proof technique of part 1 of Proposition 2 in Appendix C. Recall that the BSC(δ) along each edge of the DAG $\mathcal{G}$ produces its output bit by either copying its input bit with probability

$1 - 2\delta$, or generating an independent Bernoulli$\left(\frac{1}{2}\right)$ output bit with probability 2δ. As before, consider the mutually independent events:

$$A_k \triangleq \{\text{all } L_{k-1}L_k \text{ edges from level } k-1 \text{ to level } k \text{ generate independent output bits}\}$$

for $k \in \mathbb{N}\backslash\{0\}$, which have probabilities $\mathbb{P}(A_k) = (2\delta)^{L_{k-1}L_k}$. Define the constant $B(\delta)$ as:

$$B(\delta) \triangleq \frac{1}{\sqrt{\log\left(\frac{1}{2\delta}\right)}} . \tag{112}$$

Since we assume in the proposition statement that $L_k \leq B(\delta)\sqrt{\log(k)}$ for all sufficiently large k, we have:

$$L_{k-1}L_k \leq \frac{\sqrt{\log(k-1)\log(k)}}{\log\left(\frac{1}{2\delta}\right)} \leq \frac{\log(k)}{\log\left(\frac{1}{2\delta}\right)}$$

for all sufficiently large k, which implies that:

$$(2\delta)^{L_{k-1}L_k} \geq \frac{1}{k}$$

for all sufficiently large k. Hence, we get $\sum_{k=1}^{\infty} \mathbb{P}(A_k) = \sum_{k=1}^{\infty}(2\delta)^{L_{k-1}L_k} = +\infty$, and the second Borel-Cantelli lemma establishes that infinitely many of the events $\{A_k : k \in \mathbb{N}\backslash\{0\}\}$ occur almost surely, or equivalently, $\mathbb{P}(\bigcap_{m=1}^{\infty}\bigcup_{k=m}^{\infty} A_k) = 1$. As a result, we can define $B_m \triangleq \bigcup_{k=1}^{m} A_k$ for $m \in \mathbb{N}\backslash\{0\}$ such that $\lim_{m\to\infty} \mathbb{P}(B_m) = 1$. Therefore, we have (as before):

$$\lim_{m\to\infty} \mathbb{P}(h_{\mathsf{ML}}^m(X_m, \mathcal{G}) \neq X_0) = \frac{1}{2}$$

where $h_{\mathsf{ML}}^m(\cdot, \mathcal{G}) : \{0,1\}^{L_m} \to \{0,1\}$ denotes the ML decision rule at level m based on X_m (given knowledge of the DAG $\mathcal{G}$). This completes the proof. ∎

APPENDIX E

PROOF OF PROPOSITION 5

Proof. Recall that $L_k\sigma_k \sim \text{binomial}(L_k, g(\sigma))$ given $\sigma_{k-1} = \sigma$. This implies via Hoeffding's inequality and (29) that for every $k \in \mathbb{N}\backslash\{0\}$ and $\epsilon_k > 0$:

$$\mathbb{P}(|\sigma_k - g(\sigma_{k-1})| > \epsilon_k | \sigma_{k-1} = \sigma) \leq 2\exp\left(-2L_k\epsilon_k^2\right)$$

where we can take expectations with respect to σ_{k-1} to get:

$$\mathbb{P}(|\sigma_k - g(\sigma_{k-1})| > \epsilon_k) \leq 2\exp\left(-2L_k\epsilon_k^2\right) . \tag{113}$$

Now fix any $\tau > 0$, and choose a sufficiently large integer $K = K(\tau) \in \mathbb{N}$ (that depends on τ) such that:

$$\mathbb{P}(\exists k > K, |\sigma_k - g(\sigma_{k-1})| > \epsilon_k) \leq \sum_{k=K+1}^{\infty} \mathbb{P}(|\sigma_k - g(\sigma_{k-1})| > \epsilon_k) \leq 2\sum_{k=K+1}^{\infty} \exp\left(-2L_k\epsilon_k^2\right) \leq \tau$$

where we use the union bound, and let $\epsilon_k = \sqrt{\log(k)/L_k}$ (or equivalently, $\exp\left(-2L_k\epsilon_k^2\right) = 1/k^2$). This implies that for any $\tau > 0$:

$$\mathbb{P}(\forall k > K, |\sigma_k - g(\sigma_{k-1})| \leq \epsilon_k) \geq 1 - \tau . \tag{114}$$

Since for every $k > K$, $|\sigma_k - g(\sigma_{k-1})| \leq \epsilon_k$, we can recursively obtain the following relation:

$$\forall k \in \mathbb{N} \backslash [K+1], \quad \left| \sigma_k - g^{(k-K)}(\sigma_K) \right| \leq \sum_{m=K+1}^{k} D(\delta, d)^{k-m} \epsilon_m \tag{115}$$

where $D(\delta, d)$ denotes the Lipschitz constant of g on $[0, 1]$ as defined in (33), and $D(\delta, d) \in (0, 1)$ since $\delta \in \left(\delta_{\mathsf{maj}}, \frac{1}{2} \right)$. Since $L_m = \omega(\log(m))$, for any $\epsilon > 0$, we can take $K = K(\epsilon, \tau)$ (which depends on both ϵ and τ) to be sufficiently large so that $\sup_{m > K} \epsilon_m \leq \epsilon(1 - D(\delta, d))$. Now observe that we have:

$$\forall k \in \mathbb{N} \backslash [K+1], \quad \sum_{m=K+1}^{k} D(\delta, d)^{k-m} \epsilon_m \leq \left(\sup_{m > K} \epsilon_m \right) \sum_{j=0}^{\infty} D(\delta, d)^j = \left(\sup_{m > K} \epsilon_m \right) \frac{1}{1 - D(\delta, d)} \leq \epsilon.$$

Moreover, since $g : [0, 1] \to [0, 1]$ is a contraction when $\delta \in \left(\delta_{\mathsf{maj}}, \frac{1}{2} \right)$, it has a unique fixed point $\sigma = \frac{1}{2}$, and $\lim_{m \to \infty} g^{(m)}(\sigma_K) = \frac{1}{2}$ almost surely by the fixed point theorem [52, Chapter 5, Exercise 22(c)]. As a result, for any $\tau > 0$ and any $\epsilon > 0$, there exists $K = K(\epsilon, \tau) \in \mathbb{N}$ such that:

$$\mathbb{P}\left(\forall k > K, \left| \sigma_k - g^{(k-K)}(\sigma_K) \right| \leq \epsilon \right) \geq 1 - \tau$$

which implies, after letting $k \to \infty$, that:

$$\mathbb{P}\left(\frac{1}{2} - \epsilon \leq \liminf_{k \to \infty} \sigma_k \leq \limsup_{k \to \infty} \sigma_k \leq \frac{1}{2} + \epsilon \right) \geq 1 - \tau.$$

Lastly, we can first let $\epsilon \to 0$ and employ the continuity of $\mathbb{P}$, and then let $\tau \to 0$ to obtain:

$$\mathbb{P}\left(\lim_{k \to \infty} \sigma_k = \frac{1}{2} \right) = 1.$$

This completes the proof.   ∎

APPENDIX F

PROOF OF PROPOSITION 6

Proof. This proof is analogous to the proof of Proposition 5. For every $k \in \mathbb{N} \backslash \{0\}$ and $\epsilon_k > 0$, we have after taking expectations in (56) that:

$$\mathbb{P}(|\sigma_{2k} - g(\sigma_{2k-2})| > \epsilon_k) \leq 4 \exp\left(-\frac{(L_{2k} \wedge L_{2k-1})\epsilon_k^2}{8} \right). \tag{116}$$

Now fix any $\tau > 0$, and choose a sufficiently large integer $K = K(\tau) \in \mathbb{N}$ (that depends on τ) such that:

$$\mathbb{P}(\exists k > K, |\sigma_{2k} - g(\sigma_{2k-2})| > \epsilon_k) \leq \sum_{m=K+1}^{\infty} \mathbb{P}(|\sigma_{2m} - g(\sigma_{2m-2})| > \epsilon_m)$$

$$\leq 4 \sum_{m=K+1}^{\infty} \exp\left(-\frac{(L_{2m} \wedge L_{2m-1})\epsilon_m^2}{8} \right) \leq \tau$$

where we use the union bound and (116), and we set $\epsilon_m = 4\sqrt{\log(m)/(L_{2m} \wedge L_{2m-1})}$ (which ensures that $\exp(-(L_{2m} \wedge L_{2m-1})\epsilon_m^2/8) = 1/m^2$). This implies that for any $\tau > 0$:

$$\mathbb{P}(\forall k > K, |\sigma_{2k} - g(\sigma_{2k-2})| \leq \epsilon_k) \geq 1 - \tau. \tag{117}$$

Since for every $k > K$, $|\sigma_{2k} - g(\sigma_{2k-2})| \leq \epsilon_k$, we can recursively obtain the following relation:

$$\forall k \in \mathbb{N} \backslash [K+1], \quad \left| \sigma_{2k} - g^{(k-K)}(\sigma_{2K}) \right| \leq \sum_{m=K+1}^{k} D(\delta)^{k-m} \epsilon_m \tag{118}$$

where $g^{(k-K)}$ denotes g composed with itself $k - K$ times, and $D(\delta)$ denotes the Lipschitz constant of g on $[0, 1]$ as shown in (49), which is in $(0, 1)$ since $\delta \in \left(\delta_{\mathsf{andor}}, \frac{1}{2}\right)$. Since $L_m = \omega(\log(m))$, for any $\epsilon > 0$, we can take $K = K(\epsilon, \tau) \in \mathbb{N}$ (which depends on both ϵ and τ) to be sufficiently large so that $\sup_{m > K} \epsilon_m \le \epsilon(1 - D(\delta))$. This implies that:

$$\forall k \in \mathbb{N}\backslash[K + 1], \quad \sum_{m=K+1}^{k} D(\delta)^{k-m} \epsilon_m \le \epsilon$$

as shown in the proof of Proposition 5. Moreover, since $g : [0, 1] \to [0, 1]$ is a contraction when $\delta \in \left(\delta_{\mathsf{andor}}, \frac{1}{2}\right)$, it has a unique fixed point $\sigma = t \in [0, 1]$, and $\lim_{m \to \infty} g^{(m)}(\sigma_{2K}) = t$ almost surely by the fixed point theorem. As a result, for any $\tau > 0$ and any $\epsilon > 0$, there exists $K = K(\epsilon, \tau) \in \mathbb{N}$ such that:

$$\mathbb{P}\left(\forall k > K, \left|\sigma_{2k} - g^{(k-K)}(\sigma_{2K})\right| \le \epsilon\right) \ge 1 - \tau$$

which implies that:

$$\mathbb{P}\left(\lim_{k \to \infty} \sigma_{2k} = t\right) = 1$$

once again as shown in the proof of Proposition 5. This completes the proof. ∎

APPENDIX G

PROOF OF COROLLARY 2

Proof. Fix any $\epsilon > 0$ and any $d \ge \left(\frac{2}{\epsilon}\right)^5$, and let $\alpha = d^{-6/5}$. To establish the first part of the corollary, it suffices to prove that for every $n \in \mathbb{N}\backslash\{0\}$:

$$n\left(1 - (1 - \alpha)^d - \sqrt{2d\alpha H(\alpha)}\right) \ge (1 - \epsilon)d\alpha n$$

$$\Leftrightarrow \quad 1 - \frac{1 - (1 - \alpha)^d - \sqrt{2d\alpha H(\alpha)}}{d\alpha} \le \epsilon.$$

Indeed, if this is true, then Proposition 7 immediately implies the desired lower bound on the probability that B is a d-regular bipartite lossless $(d^{-6/5}, (1 - \epsilon)d)$-expander graph. To this end, observe that:

$$\begin{aligned}
1 - \frac{1 - (1 - \alpha)^d - \sqrt{2d\alpha H(\alpha)}}{d\alpha} &\le 1 - \frac{1 - e^{-d\alpha}}{d\alpha} + \frac{\sqrt{2d\alpha H(\alpha)}}{d\alpha} \\
&\le \frac{d\alpha}{2} + \sqrt{\frac{2H(\alpha)}{d\alpha}} \\
&\le \frac{d\alpha}{2} + \sqrt{\frac{4\log(2)}{d\sqrt{\alpha}}} \\
&= \frac{1}{2d^{1/5}} + \frac{2\sqrt{\log(2)}}{d^{1/5}} \\
&\le \frac{2}{d^{1/5}} \\
&\le \epsilon
\end{aligned}$$

where the first inequality follows from the standard bound $(1 - \alpha)^d \le e^{-d\alpha}$ for $\alpha \in (0, 1)$ and $d \in \mathbb{N}\backslash\{0\}$, the second inequality follows from the easily verifiable bounds $0 \le 1 - \frac{1-e^{-x}}{x} \le \frac{x}{2}$ for $x > 0$, the third inequality follows from the well-known bound $H(\alpha) \le 2\log(2)\sqrt{\alpha(1 - \alpha)} \le 2\log(2)\sqrt{\alpha}$ for $\alpha \in (0, 1)$, the fourth equality

follows from substituting $\alpha = d^{-6/5}$, the fifth inequality follows from direct computation, and the final inequality holds because $d \geq \left(\frac{2}{\epsilon}\right)^5$. This proves the first part of the corollary.

The existence of d-regular bipartite lossless $(d^{-6/5}, (1-\epsilon)d)$-expander graphs for every sufficiently large n (depending on d) in the second part of the corollary follows from the first part by invoking the probabilistic method. This completes the proof. ∎

REFERENCES

[1] W. Evans, C. Kenyon, Y. Peres, and L. J. Schulman, "Broadcasting on trees and the Ising model," *The Annals of Applied Probability*, vol. 10, no. 2, pp. 410–433, May 2000.

[2] H. Kesten and B. P. Stigum, "A limit theorem for multidimensional Galton-Watson processes," *The Annals of Mathematical Statistics*, vol. 37, no. 5, pp. 1211–1223, October 1966.

[3] P. M. Bleher, J. Ruiz, and V. A. Zagrebnov, "On the purity of the limiting Gibbs state for the Ising model on the Bethe lattice," *Journal of Statistical Physics*, vol. 79, no. 1-2, pp. 473–482, April 1995.

[4] D. Ioffe, "Extremality of the disordered state for the Ising model on general trees," in *Trees: Workshop in Versailles, June 14-16 1995*, ser. Progress in Probability, B. Chauvin, S. Cohen, and A. Rouault, Eds., vol. 40. Basel, Switzerland: Birkhäuser, 1996, pp. 3–14.

[5] ——, "On the extremality of the disordered state for the Ising model on the Bethe lattice," *Letters in Mathematical Physics*, vol. 37, no. 2, pp. 137–143, June 1996.

[6] E. Mossel, "Recursive reconstruction on periodic trees," *Random Structures and Algorithms*, vol. 13, no. 1, pp. 81–97, August 1998.

[7] ——, "Reconstruction on trees: Beating the second eigenvalue," *The Annals of Applied Probability*, vol. 11, no. 1, pp. 285–300, February 2001.

[8] R. Pemantle and Y. Peres, "The critical Ising model on trees, concave recursions and nonlinear capacity," *The Annals of Probability*, vol. 38, no. 1, pp. 184–206, January 2010.

[9] A. Sly, "Reconstruction of random colourings," *Communications in Mathematical Physics*, vol. 288, no. 3, pp. 943–961, June 2009.

[10] ——, "Reconstruction for the Potts model," *The Annals of Probability*, vol. 39, no. 4, pp. 1365–1406, July 2011.

[11] S. Janson and E. Mossel, "Robust reconstruction on trees is determined by the second eigenvalue," *The Annals of Probability*, vol. 32, no. 3B, pp. 2630–2649, July 2004.

[12] N. Bhatnagar, J. Vera, E. Vigoda, and D. Weitz, "Reconstruction for colorings on trees," *SIAM Journal on Discrete Mathematics*, vol. 25, no. 2, pp. 809–826, July 2011.

[13] E. Mossel, "On the impossibility of reconstructing ancestral data and phylogenies," *Journal of Computational Biology*, vol. 10, no. 5, pp. 669–676, July 2003.

[14] ——, "Phase transitions in phylogeny," *Transactions of the American Mathematical Society*, vol. 356, no. 6, pp. 2379–2404, June 2004.

[15] C. Daskalakis, E. Mossel, and S. Roch, "Optimal phylogenetic reconstruction," in *Proceedings of the 38th Annual ACM Symposium on Theory of Computing (STOC)*, Seattle, WA, USA, May 21-23 2006, pp. 159–168.

[16] S. Roch, "Toward extracting all phylogenetic information from matrices of evolutionary distances," *Science*, vol. 327, no. 5971, pp. 1376–1379, March 2010.

[17] M. Mézard and A. Montanari, "Reconstruction on trees and spin glass transition," *Journal of Statistical Physics*, vol. 124, no. 6, pp. 1317–1350, September 2006.

[18] F. Krzakała, A. Montanari, F. Ricci-Tersenghi, G. Semerjian, and L. Zdeborová, "Gibbs states and the set of solutions of random constraint satisfaction problems," *Proceedings of the National Academy of Sciences (PNAS)*, vol. 104, no. 25, pp. 10 318–10 323, June 2007.

[19] A. Gerschenfeld and A. Montanari, "Reconstruction for models on random graphs," in *Proceedings of the 48th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, Providence, RI, USA, October 20-23 2007, pp. 194–204.

[20] A. Montanari, R. Restrepo, and P. Tetali, "Reconstruction and clustering in random constraint satisfaction problems," *SIAM Journal on Discrete Mathematics*, vol. 25, no. 2, pp. 771–808, July 2011.

[21] J. von Neumann, "Probabilistic logics and the synthesis of reliable organisms from unreliable components," in *Automata Studies*, ser. Annals of Mathematics Studies, C. E. Shannon and J. McCarthy, Eds., vol. 34.   Princeton, NJ, USA: Princeton University Press, 1956, pp. 43–98.

[22] W. S. Evans and L. J. Schulman, "Signal propagation and noisy circuits," *IEEE Transactions on Information Theory*, vol. 45, no. 7, pp. 2367–2373, November 1999.

[23] D. H. Huson, R. Rupp, and C. Scornavacca, *Phylogenetic Networks: Concepts, Algorithms and Applications*.   New York, NY, USA: Cambridge University Press, 2010.

[24] E. A. Thompson, *Pedigree Analysis in Human Genetics*, ser. Johns Hopkins Series in Contemporary Medicine and Public Health.   Baltimore, MD, USA: Johns Hopkins University Press, 1986.

[25] M. Steel and J. Hein, "Reconstructing pedigrees: A combinatorial perspective," *Journal of Theoretical Biology, Elsevier*, vol. 240, no. 3, pp. 360–367, June 2006.

[26] L. F. Gray, "A reader's guide to Gacs's "positive rates" paper," *Journal of Statistical Physics*, vol. 103, no. 1-2, pp. 1–44, April 2001.

[27] R. W. Keener, *Theoretical Statistics: Topics for a Core Course*, ser. Springer Texts in Statistics.   New York, NY, USA: Springer, 2010.

[28] B. Hajek and T. Weller, "On the maximum tolerable noise for reliable computation by formulas," *IEEE Transactions on Infomation Theory*, vol. 37, no. 2, pp. 388–391, March 1991.

[29] W. S. Evans and L. J. Schulman, "On the maximum tolerable noise of k-input gates for reliable computation by formulas," *IEEE Transactions on Information Theory*, vol. 49, no. 11, pp. 3094–3098, November 2003.

[30] R. L. Dobrushin and S. I. Ortyukov, "Lower bound for the redundancy of self-correcting arrangements of unreliable functional elements," *Problemy Peredachi Informatsii*, vol. 13, no. 1, pp. 82–89, 1977, in Russian.

[31] W. Evans and N. Pippenger, "On the maximum tolerable noise for reliable computation by formulas," *IEEE Transactions on Information Theory*, vol. 44, no. 3, pp. 1299–1305, May 1998.

[32] F. Unger, "Noise threshold for universality of 2-input gates," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT)*, Nice, France, June 24-29 2007, pp. 1901–1905.

[33] A. N. Kolmogorov and Y. M. Barzdin, "On the realization of networks in three-dimensional space," in *Selected Works of A. N. Kolmogorov – Volume III: Information Theory and the Theory of Algorithms*, ser. Mathematics and Its Applications (Soviet Series), A. N. Shiryayev, Ed., vol. 27.   Dordrecht, Netherlands: Springer, 1993, pp. 194–202, originally published in: Problemy Kibernetiki, no. 19, pp. 261-268, 1967.

[34] M. S. Pinsker, "On the complexity of a concentrator," in *Proceedings of the 7th International Teletraffic Congress (ITC)*, Stockholm, Sweden, June 13-20 1973, pp. 318/1–318/4.

[35] G. A. Margulis, "Explicit constructions of concentrators," *Problemy Peredachi Informatsii*, vol. 9, no. 4, pp. 71–80, 1973, in Russian.

[36] M. R. Capalbo, O. Reingold, S. P. Vadhan, and A. Wigderson, "Randomness conductors and constant-degree lossless expanders," in *Proceedings of the 34th Annual ACM Symposium on Theory of Computing (STOC)*, Montréal, QC, Canada, May 19-21 2002, pp. 659–668.

[37] Y. Polyanskiy and Y. Wu, "Strong data-processing inequalities for channels and Bayesian networks," in *Convexity and Concentration*, ser. The IMA Volumes in Mathematics and its Applications, E. Carlen, M. Madiman, and E. M. Werner, Eds., vol. 161.   New York, NY, USA: Springer, 2017, pp. 211–249.

[38] ——, "Dissipation of information in channels with input constraints," *IEEE Transactions on Information Theory*, vol. 62, no. 1, pp. 35–55, January 2016.

[39] G. A. Margulis, "Probabilistic characteristics of graphs with large connectivity," *Problemy Peredachi Informatsii*, vol. 10, no. 2, pp. 101–108, 1974, in Russian.

[40] L. Russo, "On the critical percolation probabilities," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, vol. 56, no. 2, pp. 229–237, June 1981.

[41] G. Grimmett, "Percolation and disordered systems," in *Lectures on Probability Theory and Statistics: Ecole d'Eté de Probabilités de Saint-Flour XXVI-1996*, ser. Lecture Notes in Mathematics, P. Bernard, Ed., vol. 1665.   Berlin, Heidelberg, Germany: Springer, 1997, pp. 153–300.

[42] N. Shutty, M. Wootters, and P. Hayden, "Noise thresholds for amplification: From quantum foundations to classical fault-tolerant computation," September 2018, arXiv:1809.09748 [cs.IT].

[43] D. A. Levin, Y. Peres, and E. L. Wilmer, *Markov Chains and Mixing Times*, 1st ed.   Providence, RI, USA: American Mathematical Society, 2009.

[44] W. Hoeffding, "Probability inequalities for sums of bounded random variables," *Journal of the American Statistical Association*, vol. 58, no. 301, pp. 13–30, March 1963.

[45] M. Sipser and D. A. Spielman, "Expander codes," *IEEE Transactions on Information Theory*, vol. 42, no. 6, pp. 1710–1722, November 1996.

[46] B. Bollobás, *Random Graphs*, 2nd ed., ser. Cambridge Studies in Advanced Mathematics. Cambridge, United Kingdom: Cambridge University Press, 2001, vol. 73.

[47] W. Feller, *An Introduction to Probability Theory and Its Applications*, 3rd ed. New York, NY, USA: John Wiley & Sons, Inc., 1968, vol. 1.

[48] I. Csiszár and P. C. Shields, *Information Theory and Statistics: A Tutorial*, ser. Foundations and Trends in Communications and Information Theory, S. Verdú, Ed. Hanover, MA, USA: now Publishers Inc., 2004, vol. 1, no. 4.

[49] D. E. Knuth, *The Art of Computer Programming: Seminumerical Algorithms*, 3rd ed. Reading, MA, USA: Addison-Wesley, 1997, vol. 2.

[50] R. L. Dobrushin, "Central limit theorem for nonstationary Markov chains. I," *Theory of Probability and Its Applications*, vol. 1, no. 1, pp. 65–80, 1956.

[51] E. Seneta, *Non-negative Matrices and Markov Chains*, 2nd ed., ser. Springer Series in Statistics. New York, NY, USA: Springer, 1981.

[52] W. Rudin, *Principles of Mathematical Analysis*, 3rd ed., ser. International Series in Pure and Applied Mathematics. New York, NY, USA: McGraw-Hill, 1976.