

Noisy random Boolean formulae: A statistical physics perspective

Alexander Mozeika,^{1,*} David Saad,¹ and Jack Raymond²

¹*The Non-linearity and Complexity Research Group, Aston University, Birmingham B4 7ET, United Kingdom*

²*Department of Physics, The Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong, China*

(Received 16 March 2010; revised manuscript received 15 September 2010; published 14 October 2010)

Properties of computing Boolean circuits composed of noisy logical gates are studied using the statistical physics methodology. A formula-growth model that gives rise to random Boolean functions is mapped onto a spin system, which facilitates the study of their typical behavior in the presence of noise. Bounds on their performance, derived in the information theory literature for specific gates, are straightforwardly retrieved, generalized and identified as the corresponding macroscopic phase transitions. The framework is employed for deriving results on error-rates at various function-depths and function sensitivity, and their dependence on the gate-type and noise model used. These are difficult to obtain via the traditional methods used in this field.

DOI: [10.1103/PhysRevE.82.041112](https://doi.org/10.1103/PhysRevE.82.041112)

PACS number(s): 05.40.-a, 02.50.-r, 89.70.Eg, 05.40.Ca

I. INTRODUCTION

Computation as a physical phenomena takes many forms including classical logical circuits, quantum computing and biological neural networks. Noise is present in all practical computing systems and is a source of error with an immediate effect on the ability to represent specific functions and operations and the viability of some computing paradigms. The main sources of errors in classical computing circuits based on semiconductor technology are heat generation, cosmic rays and production defects [1]. The impact of noise becomes even more dramatic as the drive toward miniaturization of computer components causes the circuits to become more complex and of large scale [1]. The presence of decoherence-noise in quantum computers is also a significant obstacle for exploiting their full computational power [2]. The effects of noise on computing and information processing in other systems, such as biological neural networks, which are inherently noisy, remain poorly understood.

One of the first to study the effect of noise in computing systems was von Neumann [3] who attempted to explain the robustness of biological computing circuits by representing them as logical circuits comprising conventional Boolean logical gates. His model represented neural activities by a circuit (or formula) composed of ϵ -noisy Boolean gates; he suggested alternative gate-constructions to limit the resulting noise and analyzed the maximal noise tolerated before the generated functions become random. Boolean variables and gates correspond to spins and structured interactions between them within the statistical physics framework used in this paper; noise level in this context relates to temperature.

Before progressing any further, a few formal definitions are required: (i) A *circuit* may be regarded as a directed acyclic graph in which nodes of in-degree zero (representing boundary conditions, unaffected by other variables) are either Boolean constants or references to arguments, nodes of in-degree $k \geq 1$ are logical gates of k arguments and nodes of out-degree zero (that are monitored but do not affected other variables) correspond to the circuit outputs. (ii) A *formula* is

a single-output circuit where the output of each gate is used as an input to at most one gate. (iii) The ϵ -noisy gate is designed to compute a Boolean function $\alpha: \{-1, 1\}^k \rightarrow \{-1, 1\}$, representing structured k -spin interaction between variables, but for each input $S \in \{-1, 1\}^k$ there is an error probability ϵ such that $\alpha(S) \rightarrow -\alpha(S)$; this can be viewed as thermal noise. To simplify the analysis, error-probability is taken to be independent for each gate in the circuit. Clearly, a noisy circuit ($\epsilon > 0$) cannot perform any given computation in a deterministic manner: for any circuit-input there is a non-vanishing probability that the circuit will produce the wrong output. (iv) The maximum of this error probability δ over all circuit-inputs determines the *reliability* of the circuit, i.e., mismatch between expected (noiseless) and actual outputs.

In his paper, von Neumann showed that reliable computation ($\delta < 1/2$) is possible for a sufficiently small ϵ [3] and demonstrated how reliability of a Boolean noisy circuit can be improved by using gate-constructions based only on ϵ -noisy gates. There had been little development in the analysis of noisy computing systems until the seminal work of Pippenger [4] who addressed the problem from an information theory point of view. He showed that if a noisy k -ary formula is used to compute a Boolean function f with the error probability $\delta < 1/2$, then (i) there is an upper bound for the gate-error $\epsilon(k)$ which is strictly less than $1/2$ and (ii) there is a lower bound for the formula-depth $\hat{d}(k, \epsilon, \delta) \geq d$, where d is the depth of a noiseless formula computing f ; the *depth* of a formula being the number of gates on the longest path from an input node to the output node. In comparison to its noiseless counterpart, a noisy formula that computes reliably has greater depth due to the presence of restitution-gates, implying longer computation times [4].

A number of papers have followed and extended Pippenger's results. For instance, similar results were derived for circuits by Feder [5], who also improved the bounds obtained by Pippenger for formulas. The exact noise thresholds for k -ary Boolean formulas were later determined for odd k [6,7] and formulas constructed of 2-input NAND gates [8]; the latter was recently suggested as the exact noise threshold for *general* 2-input gate formulas [9].

Against this plethora of results from the information theory and theoretical computer science (TCS) literature, our

*a.s.mozeika@aston.ac.uk

aim is to provide an alternative view based on a statistical physics framework, which offers a powerful methodology that can recover and extend existing results to provide insight beyond what is accessible via the information theory methodology. The latter mainly relies on specific circuit constructions and methods that correspond to the *worst* case bounds. In contrast, our emphasis is on the *typical* case analysis of noisy circuits, which facilitates the study of formulas properties at any depth and offers flexibility in extending the results to any distribution of logical gates.

The analysis in the current paper is based on path integral methodology, specifically tailored for this task and originated in the statistical physics of disordered systems. It complements other methods that have been successfully employed in the study of similar problems from TCS and information theory [10] ranging from classical combinatorial optimization problems (graph coloring, k-SAT, reconstruction on trees and graph-isomorphism to name but a few) to source and channel coding [11], but is arguably more appropriate here due to the directed nature of the formulas studied. Nonetheless, similar results may be obtained via alternative methodology such as the replica and cavity methods [10]. As in the previous cases, we believe that our understanding of Boolean formulas will be significantly enhanced by the interaction across disciplines [10].

The study of noisy computing requires the generation of *typical* functions. Apparently, constructing typical functions by randomly connecting Boolean gates is not trivial and constitutes an area of research in its own right. Most of the familiar paradigms in the TCS literature identify gates or processes that can represent any arbitrary function, but when applied at random they tend to generate trivial functions showing weak dependence on the input variables. To generate typical formulas, which compute all Boolean functions with uniform probability, using *randomly generated* circuits, we employ a variant of the growth process suggested by Savický [12] that, under very broad conditions, produces uniformly sampled functions as the depth of the formulas becomes large.

The remainder of this paper is organized as follows. In Sec. II, we discuss generation of typical Boolean functions and the model we employ for generating them. In Sec. III, we define our model of noisy computation used for the analysis followed by the derivation of the corresponding mean-field theory in Sec. IV. Results obtained by applying our method to random formulas which use single gates or distribution over gates are presented in Sec. V followed by a summary and discussion of future work in Sec. VI. Technical aspects of the calculations which lead to our theoretic results are provided in Appendixes A and C.

II. RANDOM BOOLEAN FUNCTIONS

To investigate the effect of gate-noise on circuits representing random Boolean functions one should first identify a method for generating such circuits using basic logical gates. The importance of random Boolean functions is that they facilitate the study of average case properties, in contrast to the traditionally-studied worst-case scenario [13]. A common

approach to represent a random Boolean function is by constructing a random Boolean circuit or formula. However, finding a circuit representation of a Boolean function using a particular set of gates and of a bounded size is considered a difficult problem [14]. The most studied methods of generating random Boolean functions use random tree generation or a growth process as their core procedure. We will briefly introduce two of these methods.

In the first method, a rooted k -ary tree is sampled from the uniform distribution of all rooted k -ary trees; the leaves of this random tree are then labeled by reference to the Boolean variables and internal nodes are labeled by the Boolean gates used. This method was used to investigate typical properties of random Boolean AND/OR formulas [15–17] and obtain bounds on the probability $P(f)$ of a random formula to compute a given Boolean function f ; they also show that for a small number of inputs n such formulas result in very simple functions [16] with high probability, a behavior that becomes more pronounced for large n .

The second method uses the following growth process: First, one defines an initial distribution over a set of simple Boolean functions. Second, and in further steps, the formulas chosen from the distributions defined in previous steps are combined by Boolean gates. One such process, described by Savický [12], uses only a single Boolean gate α and is defined by the recursion on the set of formulas A_ℓ ,

$$A_0 = \{1, -1, S_1, \dots, S_n, -S_1, \dots, -S_n\},$$

$$A_{\ell+1} = \{\alpha(\phi_1, \dots, \phi_k); \phi_j \in A_\ell \text{ for } j=1, 2, \dots, k\}. \quad (1)$$

Savický showed that under very broad conditions on α the probability of computing a Boolean function by a formula $\phi \in A_\ell$ tends to the uniform distribution over all Boolean functions of n variables when $\ell \rightarrow \infty$ [12]. Furthermore, depending on the initial conditions A_0 and the gate α the process converges to a *single Boolean function* or to the *uniform distribution* over some class of Boolean functions [13].

In this framework, all Boolean functions of n variables are represented with equal statistical weight when $\ell \rightarrow \infty$, but the number of gates in formulas grows exponentially with the formula depth ℓ . Here, in order to tame this explosion in the number of gates, we propose a layered variant of the Savický growth process. The first step in our process is to sample randomly and uniformly exactly N entries of an input vector $\hat{S}^0 = (S_1^0, \dots, S_N^0)$. In the second, and all subsequent steps for $\ell = 1, \dots, L-1$, we construct a vector $\hat{S}^{\ell+1} = (S_1^{\ell+1}, \dots, S_N^{\ell+1})$ where the i th entry $S_i^{\ell+1}$ is an output of the gate $\alpha(S_{i_1}^\ell, \dots, S_{i_k}^\ell)$ with k input-indices sampled uniformly from the set of all possible (unordered) indices $\{i_1, \dots, i_k\}$. The result of the process is the layered $N \times (L+1)$ Boolean circuit shown in Fig. 1 (left construction, in blue). For large N , the variable S_i^ℓ in our model corresponds to the output of a random k -ary formula of depth ℓ , which computes a Boolean function $\{-1, 1\}^N \rightarrow \{-1, 1\}$. In the limit $N \rightarrow \infty$, with $\ell \in O(N^0)$, we expect the statistical properties of the formulas generated by this and the Savický's growth process to be equivalent; this is supported by the results reported later. The advantage of

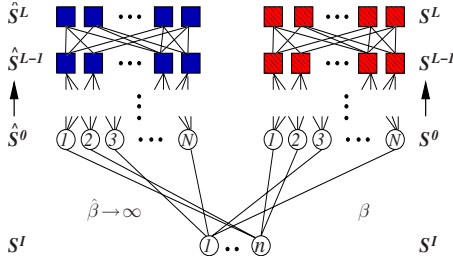


FIG. 1. (Color online) The model of two coupled systems with identical topology and different inverse temperatures β and $\hat{\beta} \rightarrow \infty$. Gates are indicated by squares, S^l and input nodes by circles. Blue (dark gray) indicates noiseless gates, red (gray) noisy gates.

using the layered representation is that it allows us to explore the typical behavior of noisy random Boolean formulas.

While the vector \hat{S}^0 represents randomly sampled single entries, one would also like to study cases where entries are statistically dependent and sampled from a smaller set. To cater for a higher level of correlation, the 0-layer boundary conditions are generated by selecting randomly \hat{S}^0 entries from the finite set $S^l = \{S^l_1, \dots, S^l_n\}$. This facilitates studying the functions generated and their dependence on properties of the set S^l .

III. MODEL

As described in Sec. II, the noisy computation model considered here is a feed-forward layered $N \times (L+1)$ Boolean circuit. The layers in the circuit are numbered from 0 (input) to L (output). Each layer $\ell \in \{1, \dots, L\}$ in the circuit is composed of exactly N ϵ -noisy, k -ary Boolean gates. Noise at gate α_i^ℓ on site (i, ℓ) operates independently and in a stochastic manner according to the microscopic law

$$P(S_i^\ell | S_{i_1}^{\ell-1}, \dots, S_{i_k}^{\ell-1}) = \frac{e^{\beta S_i^\ell \alpha_i^\ell(S_{i_1}^{\ell-1}, \dots, S_{i_k}^{\ell-1})}}{2 \cosh[\beta \alpha_i^\ell(S_{i_1}^{\ell-1}, \dots, S_{i_k}^{\ell-1})]}, \quad (2)$$

where $\beta = 1/T$ is the inverse temperature, related to the gate noise ϵ via $\tanh \beta = 1 - 2\epsilon$. The gate-output S_i^ℓ is completely random when $\beta \rightarrow 0$ ($\epsilon = 1/2$) and completely deterministic when $\beta \rightarrow \infty$ ($\epsilon = 0$). Our model is acyclic by definition, so given the state of gates at layer ℓ , gates at layer $\ell+1$ operate independently of each other. The latter suggests that the probability of the microscopic state S^0, \dots, S^L , where $S^\ell \in \{-1, 1\}^N$, is just a product of Eq. (2) over all sites and layers in the circuit. Furthermore, to investigate the properties of noisy formulas we consider two copies of the same topology, shown in Fig. 1, but with different temperatures $\beta < \infty$ (noisy) and $\hat{\beta} \rightarrow \infty$ (noiseless), comparing the two will enable one to study the effect of noise on the resulting functions. Following similar arguments to those of the single circuit case, the probability of microscopic states in the two systems is

$$P[\{S^\ell\}; \{\hat{S}^\ell\}] = P(S^0, \hat{S}^0 | S^L) \prod_{\ell=1}^L P(S^\ell | S^{\ell-1}) P(\hat{S}^\ell | \hat{S}^{\ell-1}) \quad (3)$$

where

$$P(S^\ell | S^{\ell-1}) = \prod_{i=1}^N \frac{e^{\beta S_i^\ell \sum_{j_1, \dots, j_k} A_{j_1, \dots, j_k}^{\ell, i} \alpha_i^\ell(S_{j_1}^{\ell-1}, \dots, S_{j_k}^{\ell-1})}}{2 \cosh \left[\beta \sum_{j_1, \dots, j_k} A_{j_1, \dots, j_k}^{\ell, i} \alpha_i^\ell(S_{j_1}^{\ell-1}, \dots, S_{j_k}^{\ell-1}) \right]}. \quad (4)$$

The set of connectivity tensors $\{A_{i_1, \dots, i_k}^{\ell, i}\}$, where $A_{i_1, \dots, i_k}^{\ell, i} \in \{0, 1\}$, denotes connections between gates at different layers in the circuit; the indices i_1, \dots, i_k at layer $\ell-1$ serve as the inputs of gate i at layer ℓ . The conditional probability $P(\hat{S}^\ell | \hat{S}^{\ell-1})$ is the same as in Eq. (4) but with $\beta \rightarrow \hat{\beta}$.

The sources of disorder in our model are the random connections, random boundary conditions and random gates. The former two arise in the layered growth process described in the last two paragraphs of Sec. II and represent random interactions between variables.

Random connections. The basic step in this growth process is the addition of a new gate with probability $P(A_{j_1, \dots, j_k}^{\ell, i}) = \frac{1}{N^k} \delta_{A_{j_1, \dots, j_k}^{\ell, i}; 1} + (1 - \frac{1}{N^k}) \delta_{A_{j_1, \dots, j_k}^{\ell, i}; 0}$ of being connected to exactly k gate-outputs on the previous layer $\ell-1$. This procedure is carried out independently for all gates in the circuit giving rise to the probability distribution

$$P(\{A_{i_1, \dots, i_k}^{\ell, i}\}) = \frac{1}{Z_A} \prod_{\ell, i=1}^{L, N} \left[\delta \left[1; \sum_{j_1, \dots, j_k} A_{j_1, \dots, j_k}^{\ell, i} \right] \times \prod_{i_1, \dots, i_k}^N \left[\frac{1}{N^k} \delta_{A_{i_1, \dots, i_k}^{\ell, i}; 1} + \left(1 - \frac{1}{N^k} \right) \delta_{A_{i_1, \dots, i_k}^{\ell, i}; 0} \right] \right], \quad (5)$$

where Z_A is a normalization constant. The Kronecker delta function inside the definition (5) enforces the constraint $\sum_{j_1, \dots, j_k} A_{j_1, \dots, j_k}^{\ell, i} = 1$, i.e., the gate on site (i, ℓ) is mapped to exactly one element from the set of all possible output-indices $\{i_1, \dots, i_k\}$ from the previous layer. Other sparse connectivity profiles can be easily accommodated into our framework by incorporating additional constraints into the definition (5) via the appropriate delta functions.

Random boundary conditions. In the layered growth process, these are generated by selecting members of the input set S^l , where $|S^l| \in O(N^0)$, with probability $\frac{1}{|S^l|}$, and assigning them to the input layer 0. The boundary condition is identical for the two systems which leads to the distribution

$$P(S^0, \hat{S}^0 | S^L) = \prod_{i=1}^N \delta_{S_i^0, S_{n_i}^L} \delta_{\hat{S}_i^0, \hat{S}_{n_i}^L}, \quad (6)$$

where $\{n_i\}$ are independent random indices pointing to members of input set S^l with probability $P(n_i) = \frac{1}{|S^l|}$. Further correlations can be introduced via the probability $P(S^l)$.

Random gates. In addition to the topological disorder, induced by the growth process, we assume that the gate α_i^ℓ added at each step of the process can be sampled randomly

and independently from the set G of k -ary Boolean gates. Under this assumption the distribution over gates takes the form

$$P(\{\alpha_i^\ell\}) = \prod_{\ell,i=1}^{L,N} P(\alpha_i^\ell), \quad (7)$$

where $P(\alpha_i^\ell) = \sum_{\alpha \in GP} p_\alpha \delta_{\alpha, \alpha_i^\ell}$ with $\sum_{\alpha \in GP} p_\alpha = 1$ and $p_\alpha \geq 0$.

IV. METHOD

To compute the probability distribution [Eq. (3)] directly for a circuit of finite but significant size is difficult. However, the structure of Eq. (3) is similar to the one that describes evolution of the disordered Ising spin system [18]. This similarity becomes apparent if one regards the layers in our model as discrete time steps of parallel dynamics. A common way to deal with the probabilistic objects that take this form is to use the generating functional method of statistical mechanics [19]. The generating functional is given by

$$\Gamma[\boldsymbol{\psi}; \hat{\boldsymbol{\psi}}] = \langle e^{-i \sum_{\ell,i} \psi_i^\ell S_i^\ell + \hat{\psi}_i^\ell \hat{S}_i^\ell} \rangle, \quad (8)$$

where the shorthand $\langle \dots \rangle$ denotes the average over the joint probability [Eq. (3)]. The generating functional [Eq. (8)] can be regarded as a characteristic function of [Eq. (3)] from which moments of the distribution can be obtained by taking

partial derivatives with respect to the generating fields $\{\psi_i^\ell, \hat{\psi}_j^{\ell'}\}$, for example $\langle S_i^\ell \hat{S}_j^{\ell'} \rangle = -\lim_{\psi, \hat{\psi} \rightarrow 0} \frac{\partial^2}{\partial \psi_i^\ell \partial \hat{\psi}_j^{\ell'}} \Gamma[\boldsymbol{\psi}; \hat{\boldsymbol{\psi}}]$. Following prescription of [19], we assume that for $N \rightarrow \infty$ the system is self-averaging, i.e., the macroscopic behavior of the system depends only on the statistical properties of the disorder, and compute $\Gamma[\boldsymbol{\psi}; \hat{\boldsymbol{\psi}}]$ where $[\dots]$ denotes an average over the disorder. The disorder-averaged generating function [Eq. (8)] gives rise to the following macroscopic observables

$$m(\ell) = \frac{1}{N} \sum_{i=1}^N \overline{\langle S_i^\ell \rangle} = \lim_{\psi, \hat{\psi} \rightarrow 0} \frac{i}{N} \sum_{i=1}^N \frac{\partial}{\partial \psi_i^\ell} \overline{\Gamma[\boldsymbol{\psi}; \hat{\boldsymbol{\psi}}]},$$

$$C(\ell) = \frac{1}{N} \sum_{i=1}^N \overline{\langle S_i^\ell \hat{S}_i^\ell \rangle} = - \lim_{\psi, \hat{\psi} \rightarrow 0} \frac{1}{N} \sum_{i=1}^N \frac{\partial^2}{\partial \psi_i^\ell \partial \hat{\psi}_i^\ell} \overline{\Gamma[\boldsymbol{\psi}; \hat{\boldsymbol{\psi}}]}, \quad (9)$$

where $m(\ell)$ is the average activity (magnetization) on layer ℓ and $C(\ell)$ is the overlap between two systems. Averaging over the disorder in Eq. (8) leads to the saddle-point integral (see Appendix A for details)

$$\bar{\Gamma} = \int \{dP d\hat{P} d\Omega d\hat{\Omega}\} e^{N\Psi[P, \hat{P}, \Omega, \hat{\Omega}]}, \quad (10)$$

where Ψ is the macroscopic saddle-point surface

$$\Psi[\dots] = i \sum_{\ell=0}^{L-1} \sum_{S, \hat{S}} \hat{P}^\ell(S, \hat{S}) P^\ell(S, \hat{S}) + i \sum_{\ell=0}^{L-1} \int dx d\hat{x} d\omega \hat{\Omega}^\ell(x, \hat{x}, \omega) \Omega^\ell(x, \hat{x}, \omega) + \sum_{\ell=0}^{L-1} \sum_{\{S_j, \hat{S}_j\}} \prod_{j=1}^k [P^\ell(S_j, \hat{S}_j)] \int dx d\hat{x} d\omega \Omega^\ell(x, \hat{x}, \omega)$$

$$\times \langle e^{-i\{x\alpha(\{S_j\}) + \hat{x}\alpha(\{\hat{S}_j\}) + \omega\}} \rangle_\alpha + \sum_n P(n) \ln \int \{dH d\hat{H} d\hat{x}\} \int D\omega \sum_{S, \hat{S}} M_n[H, \mathbf{x}; \hat{H}, \hat{\mathbf{x}}; \omega; S, \hat{S}], \quad (11)$$

where $\langle \cdot \rangle_\alpha$ represents an average over gate distribution and M is an effective single-site measure

$$M_n[\dots] = \delta_{S^0, S^L} \delta_{\hat{S}^0, \hat{S}^L} \prod_{\ell=0}^{L-1} [e^{ix^\ell H^\ell + i\hat{x}^\ell \hat{H}^\ell + \beta S^{\ell+1} H^\ell + \beta \hat{S}^{\ell+1} \hat{H}^\ell} e^{-\ln 2 \cosh(\beta H^\ell) - \ln 2 \cosh(\beta \hat{H}^\ell) - iP^\ell(S^\ell, \hat{S}^\ell) - i\Omega^\ell(x^\ell, \hat{x}^\ell, \omega^{\ell+1}) + i\omega^{\ell+1}}]. \quad (12)$$

The generating fields $\boldsymbol{\psi}, \hat{\boldsymbol{\psi}}$ have been removed from the above as they are no longer needed. For $N \rightarrow \infty$ the path-integral [Eq. (10)] is dominated by the extremum of the functional $\Psi[\dots]$ of Eq. (11). Functional variation of Eq. (11) with respect to the order parameters $\{P, \hat{P}, \Omega, \hat{\Omega}\}$ gives rise to four saddle-point equations

$$P^\ell(S, \hat{S}) = \sum_n P(n) \langle \delta_{S^\ell, S} \delta_{\hat{S}^\ell, \hat{S}} \rangle_{M_n}, \quad (13)$$

$$\hat{P}^\ell(S, \hat{S}) = i \sum_{i=1}^k \sum_{\{S_j, \hat{S}_j\}} \delta_{S_i, S} \delta_{\hat{S}_i, \hat{S}} \prod_{j \neq i} [P^\ell(S_j, \hat{S}_j)] \int dx d\hat{x} d\omega \Omega^\ell(x, \hat{x}, \omega) \langle e^{-i\{x\alpha(\{S_j\}) + \hat{x}\alpha(\{\hat{S}_j\}) + \omega\}} \rangle_\alpha, \quad (14)$$

$$\Omega^\ell(x, \hat{x}, \omega) = \sum_n P(n) \langle \delta(x - x^\ell) \delta(\hat{x} - \hat{x}^\ell) \delta(\omega - \omega^{\ell+1}) \rangle_{M_n}, \quad (15)$$

$$\hat{\Omega}^\ell(x, \hat{x}, \omega) = i \sum_{\{S_j, \hat{S}_j\}} \prod_{j=1}^k [P^\ell(S_j, \hat{S}_j)] \langle e^{-i[x\alpha(\{S_j\}) + \hat{x}\alpha(\{\hat{S}_j\}) + \omega]} \rangle_\alpha \quad (16)$$

where $\langle \dots \rangle_{M_n}$ is the average over the probability distribution resulting from Eq. (12). The saddle-point Eqs. (13)–(16) can be simplified significantly (see Appendix B for details) and it turns out that in order to solve this problem we only need to compute the order parameter [Eq. (13)]. The physical meaning of this order parameter is given by $P^\ell(S, \hat{S}) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \langle \delta_{S_i^\ell, S} \delta_{\hat{S}_i^\ell, \hat{S}} \rangle_{S^i}$, i.e., the disorder-averaged joint probability of sites in the two systems. The single-site effective measure [Eq. (12)] also benefits from the simplification; in particular, if we integrate out the continuous variables in Eq. (12) we are led to the expression.

$$\begin{aligned} M_n[S^L, \hat{S}^L, \dots, S^0, \hat{S}^0] &= \delta_{S^0, S^L} \delta_{\hat{S}^0, \hat{S}^L} \prod_{\ell=0}^{L-1} \left\{ \sum_{\{S_j, \hat{S}_j\}} \prod_{j=1}^k [P^\ell(S_j, \hat{S}_j)] \right. \\ &\quad \times \left. \left\langle \frac{e^{\beta S^{\ell+1} \alpha(\{S_j\})}}{2 \cosh \beta [\alpha(\{S_j\})]} \frac{e^{\hat{\beta} \hat{S}^{\ell+1} \alpha(\{\hat{S}_j\})}}{2 \cosh \hat{\beta} [\alpha(\{\hat{S}_j\})]} \right\rangle_\alpha \right\}. \end{aligned} \quad (17)$$

Using Eq. (17) the macroscopic observables Eq. (9) can be easily computed from the joint probability distribution Eq. (13), resulting in the set of equations

$$m(\ell+1) = \sum_{\{S_j\}} \prod_{j=1}^k \left[\frac{1 + S_j m(\ell)}{2} \right] \langle \tanh[\beta \alpha(S_1, \dots, S_k)] \rangle_\alpha, \quad (18)$$

$$\begin{aligned} C(\ell+1) &= \sum_{\{S_j, \hat{S}_j\}} \prod_{j=1}^k \left[\frac{1 + S_j m(\ell) + \hat{S}_j \hat{m}(\ell) + S_j \hat{S}_j C(\ell)}{4} \right] \\ &\quad \times \langle \tanh[\beta \alpha(S_1, \dots, S_k)] \tanh[\hat{\beta} \alpha(\hat{S}_1, \dots, \hat{S}_k)] \rangle_\alpha, \end{aligned} \quad (19)$$

where the magnetization $\hat{m}(\ell)$ is computed by a similar equation to Eq. (18) but with $\beta \rightarrow \hat{\beta}$. Initial conditions for these equations are given by $m(0) = \hat{m}(0) = \frac{1}{|S^0|} \sum_{S \in S^0} S$, $C(0) = 1$.

The connectivity profile considered in our model leads to a simple mean-field theory, where the macroscopic behaviors of the two copies of the same system is completely determined by the set of observables $\{m(\ell), \hat{m}(\ell), C(\ell)\}$; which relate to the order parameter Eq. (13) via $P^\ell(S, \hat{S}) = \frac{1}{4} [1 + S m(\ell) + \hat{S} \hat{m}(\ell) + S \hat{S} C(\ell)]$, while the single system behavior is described by $\{m(\ell)\}$. Furthermore, since $\langle \prod_j S_j^\ell \rangle \rightarrow \prod_j \langle S_j^\ell \rangle$ for finite j , when $N \rightarrow \infty$ [directly computed from Eq. (8)] the

spins on layer ℓ are uncorrelated as any site (i, ℓ) is a root of a full k -ary tree growing from the input layer $\ell=0$, which in turn points to the input set S^i . Loops in the circuit are rare and trees can be regarded as independent random Boolean formulas. The output of a typical formula at layer ℓ is determined by the probability $P^\ell(S)$.

Overlap and function error. The overlap order parameter of Eq. (9) is related to the normalized Hamming distance $D(\ell)$ between the states S^ℓ and \hat{S}^ℓ via the identity $D(\ell) = \frac{1}{2} [1 - C(\ell)]$. This allows one to define the order parameter $\Delta(\ell) = \lim_{\beta, \hat{\beta} \rightarrow \infty} \frac{1}{2} [1 - C(\ell)]$, used to probe sensitivity of the circuit with respect to its input, an indication to the complexity of the functions represented by the given circuit. The Hamming distance $D(\ell)$ is also related to the probability $P(S_i^\ell \neq \hat{S}_i^\ell)$ and facilitates the estimate of the error probability $\delta(\ell)$ on the ℓ th layer of a noisy circuit. More specifically, we define this error probability $\delta(\ell) = \max_{S^i} \lim_{\hat{\beta} \rightarrow \infty} \frac{1}{2} (1 - C(\ell))$, comparing the maximal error between the noisy and noiseless version of the same circuit with respect to all possible inputs. Obviously, in the absence of noise ($\beta \rightarrow \infty$) one trivially obtains $\delta(\ell) = 0$ for all ℓ .

V. RESULTS

In this section, we apply the main equations of our theory Eqs. (18) and (19) to study a Boolean formulas constructed from a particular single gate (MAJ- k and NAND) or from a distribution of gates (AND/OR). Our choice of single gates is motivated by the rigorous results of [7–9] for noisy computation by Boolean formulas based on this particular subset of logical gates. Here we extend these results to the typical formulas generated by the layered growth process of Sec. II. In particular, we identify the Boolean functions computed in the absence of noise and study their complexity and sensitivity as a function of their depths and types of gates used. For the noisy formulas, we compute phase diagrams and study dependence of the formula-output error on the formula-depth for a different levels of gate-noise. In addition, we study the impact of a quenched noise on a noisy computation and also introduce the noisy computation model where thermal noise enters the system via random gate-thresholds. The noiseless random AND/OR formulas have been studied extensively in the area of Boolean function complexity [20], here we consider random formulas both with and without noise.

A. Majority vote (MAJ- k) gate

1. Critical behavior

In this section our choice of α is a universal majority gate with k inputs (MAJ- k). The reasons for choosing this gate are twofold. First, it was proved in [6,7] that the majority gate is optimal for the noisy computation in formulas. Second, formulas constructed from the majority gates can in principle compute any Boolean function [12]. A convenient representation of the MAJ- k gate is given by the identity $\text{MAJ}(S_1, \dots, S_k) = \text{sgn}[\sum_{j=1}^k S_j]$, where k is odd. Using this representation, Eqs. (18) and (19) become

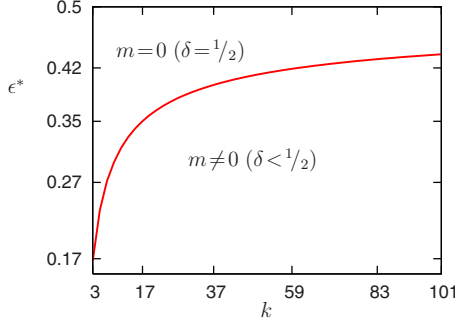


FIG. 2. (Color online) Phase diagram of a circuit based on MAJ- k gates.

$$m(\ell + 1) = (1 - 2\epsilon) \sum_{n=0}^k \binom{k}{n} \times \left[\frac{1 + m(\ell)}{2} \right]^n \left[\frac{1 - m(\ell)}{2} \right]^{k-n} \text{sgn}[2n - k], \quad (20)$$

$$C(\ell + 1) = (1 - 2\epsilon) \sum_{k_1 + \dots + k_4 = k} \frac{k!}{k_1! \times \dots \times k_4!} \times P^{k_1}(-1, -1) P^{k_2}(1, -1) \times P^{k_3}(-1, 1) P^{k_4}(1, 1) \text{sgn}[k_1 - k_2 + k_3 - k_4] \times \text{sgn}[k_1 + k_2 - k_3 - k_4], \quad (21)$$

where $P(S, \hat{S}) = \frac{1}{4}[1 + Sm(\ell) + \hat{S}\hat{m}(\ell) + S\hat{S}C(\ell)]$. To obtain Eqs. (20) and (21) we used the identity $\tanh \beta = 1 - 2\epsilon$ relating the gate-error ϵ to the inverse temperature β , and the limit $\hat{\beta} \rightarrow \infty$ ($\hat{\epsilon} = 0$) for comparing noisy and noiseless circuit outputs later on.

We now concentrate only on Eq. (20) that describes the evolution of magnetization from layer to layer. The point $m(\infty) = 0$ is always a stationary solution of this equation, i.e., $m(\ell + 1) = m(\ell) = m(\infty)$. Expanding Eq. (20) around this stationary solution gives the critical noise value $\epsilon^*(k) = 1/2 - 2^{k-2}/k \binom{k-1}{(k-1)/2}$, identical to the results of [6,7]; at the critical noise the asymptotic solution $m(\infty) = 0$ becomes unstable and two stable solutions $\pm m(\infty)$ emerge (e.g., for $k=3$ we find $m(\infty) = \pm \sqrt{\frac{1-6\epsilon}{1-2\epsilon}}$ [21]). In the case when $\epsilon > \epsilon^*(k)$ the magnetization $m(\ell)$ decays to 0 asymptotically; while for $\epsilon < \epsilon^*(k)$ we have $\lim_{\ell \rightarrow \infty} m(\ell) = \pm m(\infty)$ where the positive and negative stationary solutions correspond to the positive and negative initial magnetizations $m(0) = \frac{1}{|S^l|} \sum_{S \in S^l} S$, respectively. Thus the critical noise level $\epsilon^*(k)$ separates the unordered and ordered phases (see Fig. 2).

The relation between the new stable solutions and the reliability of the computation follows from the ability to preserve one bit of information presented at the input, by setting $S^l = \{S\}$; the phase transition observed in Eq. (20) implies that the circuit can preserve one bit of information for arbitrarily many layers only when $\epsilon < \epsilon^*(k)$. The error probability $P^\ell(-S) = \frac{1}{2}(1 - Sm(\ell))$ is a measure of how well this one bit is preserved after passing through ℓ layers. A complicated

computational task may require significant number of layers, hence only relatively simple operations can be performed by the circuit reliably when $\epsilon > \epsilon^*(k)$.

Now we turn to Eq. (21) that describes evolution of the overlap between two systems. The initial conditions are the same for both systems, so we have $m(0) = \hat{m}(0)$ and $C(0) = 1$. The magnetization in the noisy system [$\epsilon < \epsilon^*(k)$] converges to $\pm m(\infty)$ and for the noiseless system we have $\hat{m}(\infty) = \pm 1$, depending on the sign of $m(0)$. Inserting these stationary points into the Eq. (21) results in $C(\infty) = \pm m(\infty)$. The overlap $C(\infty)$ relates to the probability of error $\delta(\infty) = \frac{1}{2}[1 - C(\infty)]$. Thus the error $\delta(\infty) < 1/2$ only when $\epsilon < \epsilon^*(k)$.

2. Boolean functions generated

The analysis of Eq. (20) can also reveal the type of Boolean functions generated in the layered growth process. In particular, in the noiseless case ($\epsilon = 0$) the stationary solutions of this equation are given by $m(\infty) = 1$ and $m(\infty) = -1$ which correspond to the initial conditions $m(0) > 0$ and $m(0) < 0$, respectively. For $m(0) = 0$ we obtain $m(\infty) = 0$. Each site in our model can be associated with an output of the formula that computes some Boolean function. The average formula on layer ℓ provides outputs S with probability $P^\ell(S) = \frac{1}{2}[1 - Sm(\ell)]$. This suggests that for the noiseless case $\epsilon = 0$ the average formula on layer ℓ converges to a constant output (± 1) if the initial magnetization $m(0)$ is positive/negative, respectively, and to a random output (± 1) if $m(0) = 0$, where $m(0) = \frac{1}{|S^l|} \sum_{S \in S^l} S$. An alternative representation of the functions generated takes the form $F = \text{sgn}[\sum_{i=1}^n \theta_i S_i^l + \theta_0]$, where $\theta_i \in \mathbb{Z}$, using the definition: $\text{sgn}[0] = \pm 1$ with probability $1/2$. This means that depending on the initial conditions the formulas converge to a *single* Boolean function or to the *uniform distribution* over some set of functions. For example, if $\sum_{i=1}^n \theta_i S_i^l + \theta_0 \neq 0$ for $\forall (S_1^l, \dots, S_n^l) \in \{-1, 1\}^n$ then all formulas in the circuit converge to a single *linear threshold function* and can compute any linearly separable Boolean function [22].

In general, when $\sum_{i=1}^n \theta_i S_i^l + \theta_0 = 0$, it is difficult to say if the generated formulas compute *all* Boolean functions or only the subset of these functions. However, this result is consistent with the study of Savický [12], where it has been shown that when $S^l = \{-1, 1, S_1^l, \dots, S_n^l, -S_1^l, \dots, -S_n^l\}$ the formulas in the stationary state of the process compute all Boolean functions of n variables. It is also consistent with results [13] showing that for $S^l = \{-1, 1, S_1^l, \dots, S_n^l\}$ the formulas converge to the MAJ- n function when n is odd and to the uniform distribution over so-called slice functions when n is even [13] (also in the absence of the constants $\{-1, 1\}$ in the set S^l).

3. Sensitivity of the generated functions

We will now turn to Eq. (21) when $m(0) = \hat{m}(0) = 0$ and $C(0) = 1$. For the noiseless case $\epsilon = 0$ the stationary solution of this equation is given by $C(\infty) = 1$. This solution, however, is unstable and a small perturbation to the initial state $C(0) = 1$ leads to the stationary state $C(\infty) = 0$, which is stable. This implies that the circuit is very sensitive to its input

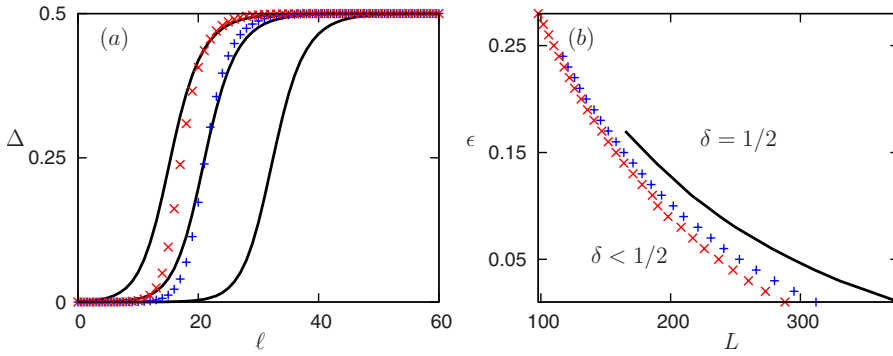


FIG. 3. (Color online) (a) Evolution of the Hamming distance as a function of the layer ℓ for $k=3$ (solid line) with input mismatch $\Delta(0)=10^{-3}, 10^{-4}, 10^{-6}$ (left to right) and, for $k=5$ (+) and $k=7$ (\times) with $\Delta(0)=10^{-6}$. (b) Phase boundaries for gate-noise ϵ at layer L when $k=3, 5, 7$, using the same symbols.

when $m(0)=0$. In particular, the Hamming distance $\Delta(\ell) = \frac{1}{2}[1 - C(\ell)]$ increases for small perturbations to $\Delta(0)$ [see Figs. 3(a)], i.e., a small perturbation to the input is amplified by the circuit. This in turn means that when $\epsilon > 0$ the circuit also amplifies the noise-perturbation and the error $\delta(\ell)$ is growing. The error, however, can be kept under control for many layers by making ϵ sufficiently small [see Fig. 3(b)].

4. Dynamics

We will now examine how the computation in the circuit proceeds from layer to layer. As an example we take $S^I = \{S_1, \dots, S_{11}\}$ and study the evolution of magnetization $m(\ell)$ and error $\delta(\ell)$; in the noiseless case the corresponding circuit computes the MAJ-11 function. The initial magnetization $|m(0)|$ takes its values from the set $\{1, 9/11, 7/11, 5/11, 3/11, 1/11\}$ for this choice of S^I . The input with the smallest magnetization in this set is very important. On the one hand, when $\epsilon=0$ the initial state with the smallest $|m(0)|$ is also the furthest from the stationary state $|m(\infty)|=1$. So it will take for the magnetization $m(\ell)$ the largest number of layers to converge for this input. On the other hand, when $\epsilon > 0$ the input with the smallest $|m(0)|$ is more likely to be suppressed by the noise. For these reasons, in what follows we study the evolution of the magnetization and errors only for $m(0) = 1/11$.

Magnetization. In Fig. 4(a) we examine how the magnetization $m(\ell)$ evolves from layer to layer in circuits with MAJ-7 gates. We observe that when $\epsilon=0$ the magnetization converges to its stationary value $m(\infty)=1$ relatively quickly. Since we use $m(0)=1/11$ the convergence to $m(\ell)=1$ indicates that all formulas in the circuit compute the MAJ-11 function. For noise values $\epsilon > 0$ the speed of convergence is decreasing as ϵ increases and becomes very slow as we approach $\epsilon^*(k)$. In general, increasing k [$0 \leq \epsilon < \epsilon^*(k)$] leads

to a reduction in relaxation times because of the inequality $F_\epsilon^{k+2}[m] \geq F_\epsilon^k[m] \geq m$, where $F_\epsilon^k[m]$ is the right hand side of Eq. (20). Finally, when we increase the noise level to $\epsilon \gg \epsilon^*(k)$ the magnetization relaxes to its stationary 0 value exponentially fast.

Function error. Figure 4(b) shows the evolution of the error $\delta(\ell)$. In the region $0 < \epsilon < \epsilon^*(k)$ we observe two distinct stages in the dynamics. Initially, the error is increasing until it reaches its maximum value. Note that this happens before the MAJ-11 function is computed exactly when $\epsilon=0$ [see Fig. 4(a)]. Also, the location of this maximum is only weakly affected by noise. These two observations suggest that initially the inputs to the gates are very inhomogeneous which leads to noise amplification. After the error reaches its maximum value the inputs become more and more homogeneous leading to the suppression of noise and as a result the error decreases until it eventually becomes stationary. As we approach the critical boundary $\epsilon^*(k)$ the number of layers needed for the error to equilibrate increases. The dynamic behavior of the error changes from nonmonotonically increasing to monotonically increasing when we approach the critical boundary $\epsilon^*(k)$ from below. The evolution of error becomes strictly monotonic when $\epsilon \gg \epsilon^*(k)$ and in this region the error relaxes to its stationary value $1/2$ exponentially fast. This example is highly representative of the situation when all formulas in the circuit converge to a single Boolean function. Here we can tell exactly how many layers it takes for the circuit to compute this function when $\epsilon=0$. This number is given by L such that $m(L) \approx m(\infty)$ starting with the smallest initial magnetization $m(0)$ induced by the inputs S^I . Obviously, adding more layers to the circuit with L layers will not affect the computation when $\epsilon=0$. However, in the region $0 < \epsilon < \epsilon^*(k)$ we can reduce the error $\delta(\ell)$ by adding more layers. This is not true for all ϵ and $m(0)$. The error can be reduced only when $\epsilon < \epsilon^0(k)$, where $\epsilon^0(k)$ is the solution of

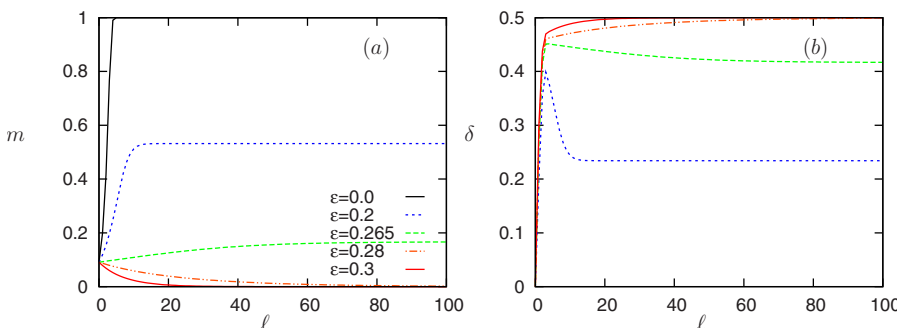


FIG. 4. (Color online) Evolution of magnetization (m) and error (δ) in MAJ- k formulas for $k=7$ [$\epsilon^*(7) \approx 0.271$] with different noise (ϵ) values.

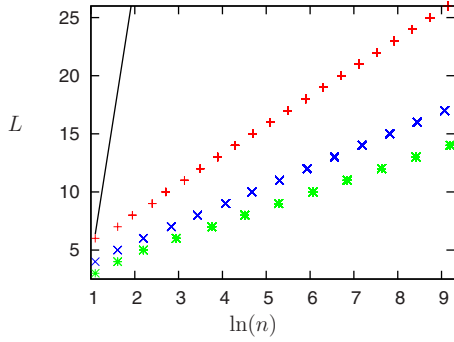


FIG. 5. (Color online) Number of layers in the noiseless MAJ- k based circuit computing MAJ- n function. Theoretically obtained data-points are represented by the symbols + ($k=3$), \times ($k=5$) and * ($k=7$). Slopes of the respective straight lines (not shown) fitted to the data are 2.473, 1.595, and 1.282, respectively. The straight line corresponds to the bound $k^2 \ln(n)$, derived in [13], plotted here for $k=3$ only.

$m(0) = F_{\epsilon^0(k)}^k[m(0)]$ for a given initial magnetization $m(0)$. At $\epsilon = \epsilon^0(k)$ the dynamic behavior of $m(\ell)$ changes from the monotonically increasing [when $\epsilon < \epsilon^0(k)$] to monotonically decreasing [when $\epsilon > \epsilon^0(k)$]. Respectively, the dynamic behavior of the error $\delta(\ell)$ changes from the nonmonotonically increasing to the monotonically increasing. Only in the former regime one can reduce the error $\delta(\ell)$ by adding more layers to the circuit. However, this strategy fails for inputs with $m(0)=0$ and the circuit computes more than one Boolean function. For $m(0)=0$ the probability of error $\delta(\ell)$ increases toward its stationary value $\delta(\infty)=1/2$ ($m(\ell)=0$) even when $\epsilon < \epsilon^*(k)$. The error $\delta(\ell)$ can be bounded away from 1/2 only by reducing the gate-error ϵ , which depends on the formula depth L [Fig. 3(b)].

5. Convergence rates

In this section we study convergence rates at $\epsilon=0$ and $\epsilon = \epsilon^*(k) \pm \Delta\epsilon$ regions of the phase diagram plotted in Fig. 2. The former allows us to estimate the number of layers required in a formula to represent the asymptotically corresponding function, which is directly related to its *size*, when all formulas in the circuit converge to a single Boolean function. The latter probes the regime where the computation is expected to be very slow, but the error can be still reduced by adding more layers to the circuit.

Noiseless formulas. First, we study the rate of convergence when $\epsilon=0$ and $m(0)=1/n$, where $n \in \mathbb{N}$ is odd, and the MAJ- k based circuit computes MAJ- n function. In general, we find that the number of layers needed for the magnetization to converge scales as $O[f(k)\ln(n)]$. This rate of convergence is consistent with rigorous results [13] for the growth process defined by Savický [12]. However, the *worst case bound* $f(k)$ derived in [13] grows as $k2^k$ with the gate in-degree k , while in our study we find that in the typical case $f(k)$ is *decreasing* with increasing k (see Fig. 5). Furthermore, this result holds not only for MAJ- n , but for *any* linear threshold function (with integer weights) computed by the MAJ- k circuit. It is natural to expect that when $k \rightarrow \infty$ the function $f(k)$ is vanishing and the numbers of layers L in the

circuit approaches 1. The discrepancy in the asymptotic behavior of the worst case bounds [13] and the *typical* asymptotic behavior observed in our work is due to the *average topology* considered here, which turns out to be more realistic.

Near the critical noise. Second, we study convergence rate for $\epsilon = \epsilon^*(k) \pm \Delta\epsilon$, where $0 < \Delta\epsilon \ll 1$. Very close to the phase boundary $\epsilon^*(k)$ the differences $m(\ell+1) - m(\ell)$ are very small and the difference Eq. (20) can be well approximated by a differential equation. For $k=3$ the differential equation reduces to $\frac{d}{d\ell}m(\ell) = -m(\ell) + \frac{1}{2}(1-2\epsilon)[3m(\ell) - m^3(\ell)]$, where ℓ is continuous, which can be solved exactly. The solution is given by $m^2(\ell) = \left\{ \left[\frac{1}{m^2(0)} - \frac{1-2\epsilon}{1-6\epsilon} \right] e^{-(1-6\epsilon)\ell} + \frac{1-2\epsilon}{1-6\epsilon} \right\}^{-1}$. This approach is only accurate in the region $\epsilon = 1/6 \pm \Delta\epsilon$, where it gives us the asymptotic form $|m(\ell) - m(\infty)| \approx e^{-\gamma(3)\Delta\epsilon\ell}$. The $\gamma(3)$ coefficient equals 3 in the paramagnetic region and 6 in the ferromagnetic region. Thus the convergence to the asymptotic paramagnetic solution $m(\infty)=0$ is slower than to the stationary ferromagnetic solutions $m(\infty) = \pm \sqrt{\frac{1-6\epsilon}{1-2\epsilon}}$ for $k=3$. The differential version of the difference Eq. (20) is difficult to solve analytically for $k > 3$ and a different method is employed (see Appendix C), resulting for large k in the form

$$|m(\ell) - m(\infty)| \approx e^{-O(k^0)\sqrt{k}\Delta\epsilon\ell}, \quad (22)$$

convergence in both paramagnetic and ferromagnetic regimes speeds up with k .

6. Hard noise

The model studied so far can be regarded as a model of computation where errors result from single-event upsets (soft noise). Imperfections in real integrated circuits [1] that are introduced during production are an additional source of permanent errors (hard noise).

A natural way to introduce hard noise into our model is to define *quenched* random variables $\{\xi_i^\ell\}$, where $P(\xi_i^\ell) = p\delta_{\xi_i^\ell, -1} + (1-p)\delta_{\xi_i^\ell, 1}$, which can invert the gate output α_i^ℓ permanently. Using transformation $\alpha_i^\ell \rightarrow \xi_i^\ell \alpha_i^\ell$ in Eq. (2) and following the steps of calculation in Sec. IV, we find that the inclusion of hard noise in our model leads to $(1-2\epsilon) \rightarrow (1-2p)(1-2\epsilon)$ in Eqs. (20) and (21). As a result, the effect of quenched noise is to reduce the critical noise $\epsilon^*(k)$. In particular, the new critical noise value is given by $\epsilon^*(k, p) = \frac{1}{2} - \frac{2^{k-2}}{(1-2p)k \binom{k-1}{(k-1)/2}}$ when $0 \leq p < \epsilon^*(k, 0)$ and $\epsilon^*(k, p) = 0$ when $p \geq \epsilon^*(k, 0)$.

The hard noise can also be introduced by making a fraction of the gates insensitive to the inputs, i.e., gates that produce constants. In particular, by taking $P(\alpha) = p_0\delta_{\alpha, \text{MAJ-}k} + p_-\delta_{\alpha, -1} + p_+\delta_{\alpha, 1}$, where p_\pm are the probabilities of constant ± 1 outputs and $p_0, p_-, p_+ \geq 0$ with $p_0 + p_- + p_+ = 1$, Eq. (18) for $p_+ - p_- = 0$ results in $\epsilon^*(k, p_0) = \frac{1}{2} - \frac{2^{k-2}}{p_0 k \binom{k-1}{(k-1)/2}}$ when $0 \leq (1-p_0)/2 < \epsilon^*(k, p_0=0)$ and $\epsilon^*(k, p_0) = 0$ when $(1-p_0)/2 \geq \epsilon^*(k, p_0=0)$. So the introduction of constant gates reduces the critical noise value $\epsilon^*(k)$ when $p_+ = p_-$ by effectively reducing the number of active gates. For $p_+ - p_- \neq 0$ the effect of hard noise is more drastic. For

$\epsilon=0$ the circuit, irrespective of its input, is biased toward one of its outputs ± 1 , depending on the value of $p_+ - p_-$.

7. Threshold noise

The MAJ- k function can be seen as a special case of a linear threshold function. The linear threshold functions (or binary perceptrons) are widely used in modeling the neuronal activities of the brain such as memory and learning [23]. The noise in these models is usually introduced via random thresholds [23] giving rise to the microscopic law

$$P(S_i^\ell | S_{i_1}^{\ell-1}, \dots, S_{i_k}^{\ell-1}) = \frac{e^{\beta S_i^\ell \sum_{j=1}^k S_{i_j}^{\ell-1}}}{2 \cosh \left[\beta \sum_{j=1}^k S_{i_j}^{\ell-1} \right]}, \quad (23)$$

which in the limit of $\beta \rightarrow \infty$ corresponds to the noiseless MAJ- k gate. Using the microscopic law [Eq. (23)] in the model of noisy computation [Eq. (3)] requires only minor change to the calculations in Sec. IV and leads to the corresponding equations for magnetization and overlap

$$m(\ell+1) = \sum_{n=0}^k \binom{k}{n} \left[\frac{1+m(\ell)}{2} \right]^n \left[\frac{1-m(\ell)}{2} \right]^{k-n} \times \tanh[\beta(2n-k)], \quad (24)$$

$$C(\ell+1) = \sum_{k_1+\dots+k_4=k} \frac{k!}{k_1! \times \dots \times k_4!} P^{k_1}(-1, -1) P^{k_2}(1, -1) \times P^{k_3}(-1, 1) P^{k_4}(1, 1) \tanh[\beta(k_1 - k_2 + k_3 - k_4)] \times \text{sgn}[k_1 + k_2 - k_3 - k_4]. \quad (25)$$

In the limit $\beta \rightarrow \infty$ ($\epsilon=0$), Eqs. (24) and (25) and Eqs. (20) and (21) are identical. So that all the results derived for the noiseless MAJ- k circuit are also valid here.

Critical noise level. However, the macroscopic behavior of the model with noisy thresholds is different from the noisy MAJ- k model for any $\epsilon > 0$. Analysis of Eq. (24) reveals that the point $m(\infty)=0$ is always a solution of this equation. Expanding Eq. (24) around this point leads to the condition

$$1 = 2^{-k} \sum_{n=0}^k \binom{k}{n} |2n-k| \tanh(\beta|2n-k|), \quad (26)$$

where $m(\infty)=0$ becomes unstable and two stable $\pm m(\infty)$ solutions emerge. In Fig. 6, we compare the resulting phase boundary with that of the MAJ- k based circuits. The MAJ- k gate is more resilient to the threshold noise [Eq. (23)] than to the flip noise [Eq. (2)]. This is not surprising since the effect of flip noise on the MAJ- k gate (inverting the gate-output regardless of the input) is more drastic than the effect of threshold noise where gate-outputs S_i^ℓ corresponding to inputs with high input-magnetization $|\sum_{j=1}^k S_{i_j}^{\ell-1}|$ are less likely to be flipped. Furthermore, as $k \rightarrow \infty$ the critical noise level $\epsilon^*(k)$ in both models approaches $1/2$ as $1/2 - \epsilon^*(k) = O(k^{-\gamma})$, but for the threshold noise model it can be shown that $\gamma=1$ while for the model with flip noise $\gamma=1/2$.

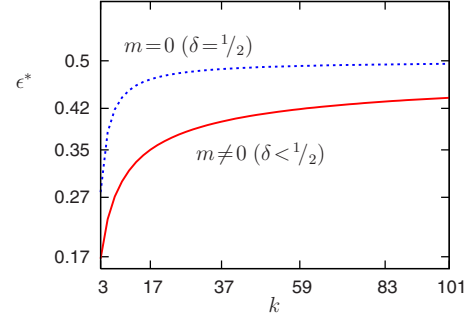


FIG. 6. (Color online) The critical noise level ϵ^* as a function of k for the perceptron (dotted line) and MAJ- k (solid line) circuits.

Convergence rate. For the noisy threshold model considered here the evolution of magnetization and δ -error is qualitatively similar to the evolution of these order parameters in the flip-noise model. However, we find that the convergence to all stationary states is much quicker in the noisy threshold model; for large k convergence is dominated by $|m(\ell) - m(\infty)| \approx e^{-O(k^0)k\Delta\epsilon\ell}$, which is significantly faster than Eq. (22).

Finally, we note that the magnetization Eq. (24) is identical to the one derived for the parallel dynamics of an Ising ferromagnet in a fully asymmetric Bethe lattice [18,24] if layers in the circuit-model are regarded as the time-steps of parallel dynamics. This suggests that the site-time topology (recurrent network) generated by the parallel dynamics is similar to that of layered networks considered here when $N \rightarrow \infty$. Based on this observation, which was exploited for instance in [25], we expect the computation performed by the recurrent and layered networks to be the same at least for $m(0) \neq 0$. Potential differences may arise from the non-vanishing connected correlations between the different times (layers).

B. NAND gate

Formulas constructed by 2-input universal NAND gates are optimal for noisy computation using 2-ary Boolean gates [8,9]. Using the NAND gate representation as the linear threshold function $-\text{sgn}[S_1 + S_2 + 1]$ the magnetization [Eq. (18)] and overlap [Eq. (19)] equations become (using $\hat{\beta} \rightarrow \infty$ and $\tanh \beta = 1 - 2\epsilon$)

$$m(\ell+1) = \frac{1}{2}(1-2\epsilon)\{[1-m(\ell)]^2 - 2\} \quad (27)$$

$$C(\ell+1) = \frac{1}{4}(1-2\epsilon)[1 + 2m(\ell) + 2\hat{m}(\ell) + 2C(\ell) - m^2(\ell) + 2m(\ell)\hat{m}(\ell) - 2m(\ell)C(\ell) - \hat{m}^2(\ell) - 2\hat{m}(\ell)C(\ell) + C^2(\ell)], \quad (28)$$

where $m(0)=\hat{m}(0)=\frac{1}{|S|}\sum_{S \in S} S$, $C(0)=1$. Here, the equation for $\hat{m}(\ell)$ is identical to Eq. (27), but with $\epsilon=0$.

Critical noise level. The magnetization [Eq. (27)] admits only one steady state $[m(\infty)=m(\ell+1)=m(\ell)]$ solution in the region $m(\infty) \in [-1, 1]$. However, the solution

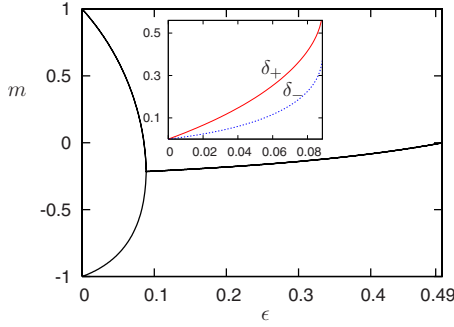


FIG. 7. (Color online) Magnetization m and error δ as a function of gate-noise ϵ in NAND-gate formulas. Inset—The dependence of δ_{\pm} on the noise level in the range $0 \leq \epsilon \leq (3 - \sqrt{7})/4$.

$m(\infty) = 1 - \frac{1 - \sqrt{8\epsilon^2 - 12\epsilon + 5}}{2\epsilon - 1}$ to Eq. (27) becomes unstable for the values of noise $\epsilon \leq \epsilon^* = (3 - \sqrt{7})/4$, which identifies ϵ^* as the critical threshold. Above this threshold the (output) magnetization on layer $\ell \rightarrow \infty$ converges to the value $m(\infty)$, which is independent of the initial (input) magnetization $m(0)$. Below the threshold ϵ^* the magnetization $m(\ell)$ oscillates from layer to layer and the properties of this oscillation depends on $m(0)$. The latter suggests that the circuit performs computation only in the region $\epsilon < \epsilon^*$, in agreement with [8]. To explore the properties of this computation we follow the difference equation of the probability $P_-(\ell) = [1 - m(\ell)]/2$ at even layers via Eq. (27) to obtain

$$P_-(\ell + 2) = 1 - \epsilon - (1 - 2\epsilon)(1 - \epsilon)^2 + 2(1 - \epsilon)(1 - 2\epsilon)^2 P_-^2(\ell) - (1 - 2\epsilon)^3 P_-^4(\ell), \quad (29)$$

where $P_-(0) = [1 - m(0)]/2$. Equation (29) also describes the evolution of $P_-(\ell)$ on odd layers with the initial condition being $P_-(1) = \frac{1}{2}(1 - 2\epsilon)[P_-^2(0) - \frac{1}{2}]$. In the region $P_- \in [0, 1]$ Eq. (29) admits three steady state solutions $P_-(\infty) = \left\{ \frac{1 - \sqrt{8\epsilon^2 - 12\epsilon + 5}}{4\epsilon - 2}, \frac{-1 \pm \sqrt{8\epsilon^2 - 12\epsilon + 1}}{4\epsilon - 2} \right\}$. The first solution becomes unstable at the noise threshold $\epsilon^* = (3 - \sqrt{7})/4$ and the other two solutions are stable for the noise values $\epsilon < \epsilon^*$. Plotting these solutions with respect to the noise ϵ in terms of the corresponding magnetization variables $m(\infty)$ gives the phase diagram depicted in Fig. 7. The stationary solutions of Eq. (29)

also allow one to compute the δ -error which, due to oscillatory behavior of the magnetization $m(\ell)$ in Eq. (27), depends on the sign of the output. In particular the error $\delta(\infty)$ takes its values from the set $\left\{ \delta_- = \frac{-1 + \sqrt{8\epsilon^2 - 12\epsilon + 1}}{4\epsilon - 2}, \delta_+ = 1 - \frac{-1 - \sqrt{8\epsilon^2 - 12\epsilon + 1}}{4\epsilon - 2} \right\}$ when $m(0) \in (2 - \sqrt{5}, 1 - \frac{1 - \sqrt{8\epsilon^2 - 12\epsilon + 5}}{2\epsilon - 1})$ and $\delta(\infty) \in \{1 - \delta_+, 1 - \delta_-\}$ when $m(0) \in (2 - \sqrt{5}, 1 - \frac{1 - \sqrt{8\epsilon^2 - 12\epsilon + 5}}{2\epsilon - 1})$, respectively. The dependence of the error-functions δ_{\pm} on the gate-noise ϵ is shown in the inset of Fig. 7.

Functions generated: For $\epsilon = 0$ the basins of attraction of the fixed points $P_-(\infty) \in \{1, 0\}$ are given by $P_-(0) \in [1, \frac{\sqrt{5}-1}{2})$ and $P_-(0) \in (\frac{\sqrt{5}-1}{2}, 0]$, respectively. Thus for $\epsilon = 0$ and $\ell \rightarrow \infty$ the NAND formulas results in -1 if $m(0) < 2 - \sqrt{5}$ and $+1$ otherwise, where $m(0) = \frac{1}{|S|} \sum_{S \in S} S$, when ℓ is even and its inverse $-F$ when ℓ is odd. In contrast to the result for MAJ- k circuit, the variety of functions generated by the NAND formulas is rather limited. All formulas in the NAND circuit converge to linear threshold functions, which due to the threshold value of $2 - \sqrt{5}$ cannot compute all linearly separable Boolean functions.

Example. As an example of a noisy computation by the NAND circuit we consider the input set $S^I = \{S_1, S_2, S_3\}$; the initial magnetization is given by $m(0) = (S_1 + S_2 + S_3)/3 \in \{-1, -1/3, 1/3, 1\}$ and the noiseless circuit converges to the MAJ-3 Boolean function on even layers and to its dual on odd layers. In Fig. 8 we plot the evolution of magnetization m and δ -error only for $m(0) = -1/3$ where the number of layers ℓ needed for the magnetization $m(\ell)$ to converge to its stationary value -1 (for ℓ even) is maximal.

We observe that for $\epsilon = 0$ the magnetization approaches its stationary value -1 in approximately 16 layers, beyond which all formulas compute MAJ-3 function. For gate-noise values $\epsilon > 0$, the number of layers L needed for the magnetization to become stationary increases as we increase ϵ toward its critical value ϵ^* , while the stationary magnetization value $m(\infty)$ decreases ($|m(\infty)| < 1$). As a result, the stationary values of the δ -error, which is directly related to $m(\infty)$, grow monotonically toward $\delta(\infty)$. The error grows first then reduces with the addition of layers (see Fig. 8); this error reduction is only possible when $\epsilon < \epsilon(0)$, where $\epsilon(0) = \frac{m^2(0) - m(0) - 3 + \sqrt{-3m^2(0) + 4m(0) + 8}}{2(m^2(0) - 2m(0) - 1)}$. Above the critical noise level ϵ^* the stationary magnetization $m(\infty)$ is independent of $m(0)$ and the computation becomes unreliable.

Convergence rate. Following the method outlined in Sec. V A, we obtain convergence rates to the stationary solutions of Eq. (29) in the range $0 < \epsilon < \epsilon^*$; for $\Delta = \epsilon^* - \epsilon$ we find

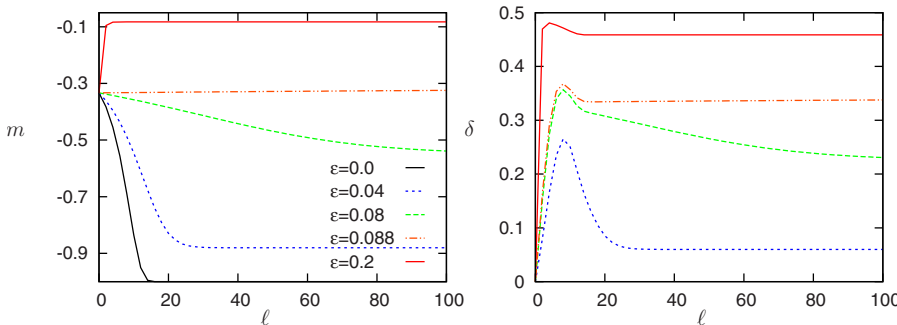


FIG. 8. (Color online) Evolution of magnetization (m) and δ -error on even layers in NAND formulas computing MAJ-3 function.

$$|P_-(\ell) - P_-(\infty)| \geq \text{const} \times \exp \left\{ \frac{\ell}{2} \ln \left[8 \left(\frac{3 - \sqrt{7}}{4} - \Delta \right) \times \left(\frac{3 + \sqrt{7}}{4} + \Delta \right) \right] \right\}. \quad (30)$$

Close to the critical noise level ($\Delta \rightarrow 0$) the argument of the log function approaches unity and the number of layers needed to converge to the stationary solution (the point of intersection of all magnetization curves in Fig. 7) diverges. In the opposite limit of $\Delta \rightarrow \epsilon^* = \frac{3 - \sqrt{7}}{4}$ the argument of the log function approaches zero and the convergence to the stationary states $m(\infty) = \pm 1$ is very fast.

Finally, we note that the results obtained provide definite answers to some of the conjectures proposed in [8], in particular: (a) The threshold $\epsilon^* = \frac{3 - \sqrt{7}}{4}$ is valid for random NAND-based formulas with reliable (hard) inputs. (b) Reliable computation at $\epsilon = \epsilon^*$ is not possible as Eq. (29) has only one fixed point, which is both unstable and attractive.

C. AND/OR gates

In the model of computation which we consider here the gate α is sampled from the distribution $P(\alpha) = p \delta_{\alpha; \text{sgn}[S_1 + S_2 + 1]} + (1 - p) \delta_{\alpha; \text{sgn}[S_1 + S_2 - 1]}$. Using this definition in the gate-averages of Eqs. (18) and (19) gives

$$m(\ell + 1) = \frac{1}{2}(1 - 2\epsilon)[2p - 1 + 2m(\ell) - (2p - 1)m^2(\ell)], \quad (31)$$

$$C(\ell + 1) = \frac{1}{4}(1 - 2\epsilon)[1 - 2m(\ell) - 2\hat{m}(\ell) + 2C(\ell) + 4pm(\ell) + 4p\hat{m}(\ell) - 4pm(\ell)C(\ell) - 4p\hat{m}(\ell)C(\ell) - m^2(\ell) + 2m(\ell)\hat{m}(\ell) + 2m(\ell)C(\ell) - \hat{m}^2(\ell) + 2\hat{m}(\ell)C(\ell) + C^2(\ell)]. \quad (32)$$

Equation (31) can be written in a more convenient form

$$P_-(\ell + 1) = \epsilon + 2(1 - 2\epsilon)(1 - p)P_-(\ell) + (1 - 2\epsilon)(2p - 1)P_-^2(\ell), \quad (33)$$

where $P_-(\ell) = [1 - m(\ell)]/2$ is the probability of output taking value of -1 . For $\epsilon > 0$ Eq. (33) has only one (stable) steady state solution $P_-(\infty)$. Thus there is no phase transition in this model for any noise value $\epsilon > 0$ and the information about the input cannot be preserved for infinitely many layers.

The noiseless balanced AND/OR trees were studied in [20]. Here we only show how to recover their results [30] from the Eqs. (31)–(33).

First, we note that by setting $\epsilon = 0$ in Eq. (33) one obtains the equation of Lemma 3.1 in [20]. Equation (33) has two fixed points $P_-(\infty) \in \{1, 0\}$ ($m(\infty) \in \{-1, 1\}$) when $p \neq 1/2$. The first point $P_-(\infty) = 1$ is stable while the second point $P_-(\infty) = 0$ is unstable when $p < 1/2$, so the circuit computes the OR function of the variables belonging to the input set S^l . For $p > 1/2$ the point $P_-(\infty) = 1$ is unstable and the point

$P_-(\infty) = 0$ is stable, so the circuit computes AND function.

Second, we set $p = 1/2$ and allow for $m(0) \neq \hat{m}(0)$, i.e., we have two copies of the same circuit but with different inputs in Eqs. (31) and (32). For $p = 1/2$ the magnetization is conserved from layer to layer [$m(\ell) = m(0)$] and the overlap Eq. (32) reduces to

$$C(\ell + 1) = \frac{1}{4} + \frac{1}{2}C(\ell) - \frac{1}{4}m^2(0) + \frac{1}{2}m(0)\hat{m}(0) - \frac{1}{4}\hat{m}^2(0) + \frac{1}{4}C^2(\ell). \quad (34)$$

Computing the joint probability $P^\ell(-1, 1) = [1 - m(0) + \hat{m}(0) - C(\ell)]/4$ from Eq. (34) one obtains

$$P^\ell(-1, 1) = P^\ell(-1, 1) \left[1 - \frac{1}{2}m(0) + \frac{1}{2}\hat{m}(0) - P^\ell(-1, 1) \right], \quad (35)$$

which corresponds to Lemma 3.2 in [20]. In general, for $\ell \rightarrow \infty$, we have $P^\infty(S, \hat{S}) = [1 + Sm(0) + \hat{S}\hat{m}(0) + S\hat{S}C(\infty)]/4$, where $C(\infty) = 1 - |\hat{m}(0) - m(0)|$. The analysis [20] of Eq. (35) reveals that AND/OR-based formulas compute constant ± 1 functions when $S^l = \{S_1, \dots, S_n, -S_1, \dots, -S_n\}$ and linear threshold functions $\text{sgn}[\sum_{j=1}^n S_j - n + 2i]$, where $i \in \{1, \dots, n\}$ and $S^l = \{S_1, \dots, S_n\}$. Finally, convergence to the functions computed is exponential in ℓ except for the special case of $p = 1/2$ when it is logarithmic ($\sim \ell^{-1}$) [20].

VI. DISCUSSION

Random formulas generated by a layered variant of the Savický formula-growth process are used to study the typical properties of noisy formulas. The layered framework facilitates a direct mapping to the physical Ising spin system, which can be seen as a dynamical system where time steps correspond to circuit layers. This analogy facilitates the use of generating functional analysis (GFA) methodology to analyze properties of noisy formulas.

Here, we use GFA to study the typical properties of noisy random Boolean formulas constructed from single gates or distributions of gates. All exact noise thresholds, which were derived in TCS using rigorous methods, are recovered within our framework and identified as macroscopic phase transitions. On a quantitative side, we attribute this to the exact correspondence of the mean-field Eq. (18) to the single-gate probability of error (assuming independence of inputs) used to derive these noise thresholds in TCS.

In the noiseless case ($\epsilon = 0$), we have identified the Boolean functions generated by the layered growth process, but only when the input-set magnetization $m(0) = 0$ is not a fixed point of the dynamics [Eq. (18)]. For inputs with $m(0) = 0$ our results are consistent with those of Savický's growth process that generates Boolean functions of arbitrary complexity. Furthermore, we have established that functions generated in the layered growth process are sensitive to the input variables, an indication of their complexity. In order to find exactly which functions are generated for the input set with $m(0) = 0$, with or without noise, one needs a more sophisti-

cated version of the mean-field theory presented here [26].

In the noisy case $\epsilon > 0$ we studied the evolution of the output magnetization and computation-error from layer to layer and their dependence on the input-set magnetization and noise. We identified a range of the critical gate-noise parameter values, and their dependence on the input-set magnetization $m(0)$, where adding more layers to the circuit may reduce the computation error. The speed of convergence to the equilibrium was studied both numerically (for $\epsilon=0$) and analytically (for $\epsilon > 0$). Results for $\epsilon=0$ are consistent with the existing bounds, but show that tighter bounds can be derived in a typical case.

The standard noisy computation model was extended to include “production errors” (uncorrelated hard noise). We have found that the effect of hard noise on the critical behavior of noisy circuit is to effectively reduce its critical noise threshold. We expect that the critical noise threshold (if it exists) in any noisy circuit will be affected in this way. Also, in our work a standard flip-noise is compared with threshold noise. In particular, we have found that the MAJ- k gate is more resilient against threshold noise than to flip-noise. We expect that any gate which can be represented as the linear

threshold function is more robust against threshold noise, at least for the threshold noise distribution considered here.

We believe that much can be learned about the typical properties of noisy computation via this approach, which complements the rigorous bounds derived in the TCS literature and provides insight that may help in the development of new rigorous techniques. Also, the problem studied and the method presented here may offer a different perspective on the dynamics of physical systems with binary degrees of freedom, in the same way as the exchange of ideas has benefited both the statistical physics and TCS in the past [27,28].

ACKNOWLEDGMENTS

Support by the Leverhulme trust (Grant No. F/00 250/H) is gratefully acknowledged.

APPENDIX A: DISORDER AVERAGE

Here, we outline the calculation steps which lead to the saddle-point integral Eq. (10). The starting point of this calculation is the generating functional

$$\Gamma[\boldsymbol{\psi}; \hat{\boldsymbol{\psi}}] = \sum_{\{\mathbf{S}^\ell, \hat{\mathbf{S}}^\ell\}} P(\mathbf{S}^0, \hat{\mathbf{S}}^0 | \mathbf{S}^L) \exp \left[\sum_{\ell, i=1}^{L, N} \{ \beta S_i^\ell H_i^{\ell-1}(\mathbf{S}^{\ell-1}) + \hat{\beta} \hat{S}_i^\ell \hat{H}_i^{\ell-1}(\hat{\mathbf{S}}^{\ell-1}) \} \right] \\ \times \exp \left[- \sum_{\ell, i=1}^{L, N} \{ \ln 2 \cosh[\beta H_i^{\ell-1}(\mathbf{S}^{\ell-1})] + \ln 2 \cosh[\hat{\beta} \hat{H}_i^{\ell-1}(\hat{\mathbf{S}}^{\ell-1})] \} \right] e^{-i \sum_{\ell, i} \{ \psi_i^\ell S_i^\ell + \hat{\psi}_i^\ell \hat{S}_i^\ell \}} \quad (\text{A1})$$

defining $H_i^{\ell-1}(\mathbf{S}^{\ell-1}) = \sum_{j_1, \dots, j_k} A_{j_1, \dots, j_k}^{\ell, i} \alpha_i^\ell(S_{j_1}^{\ell-1}, \dots, S_{j_k}^{\ell-1})$ and $\hat{H}_i^{\ell-1}(\hat{\mathbf{S}}^{\ell-1}) = \sum_{j_1, \dots, j_k} \hat{A}_{j_1, \dots, j_k}^{\ell, i} \alpha_i^\ell(\hat{S}_{j_1}^{\ell-1}, \dots, \hat{S}_{j_k}^{\ell-1})$. Introducing the fields via the integral identities

$$\prod_{i=1}^N \prod_{\ell=0}^{L-1} \left\{ \int \frac{dH_i^\ell dx_i^\ell}{2\pi} e^{ix_i^\ell [H_i^\ell - H_i^\ell(\mathbf{S}^\ell)]} \right\} = \prod_{i=1}^N \prod_{\ell=0}^{L-1} \left\{ \int \frac{d\hat{H}_i^\ell d\hat{x}_i^\ell}{2\pi} e^{i\hat{x}_i^\ell [\hat{H}_i^\ell - \hat{H}_i^\ell(\hat{\mathbf{S}}^\ell)]} \right\} = 1 \quad (\text{A2})$$

leads to

$$\Gamma[\boldsymbol{\psi}; \hat{\boldsymbol{\psi}}] = \prod_{i=1}^N \prod_{\ell=0}^{L-1} \left\{ \int \frac{dH_i^\ell dx_i^\ell d\hat{H}_i^\ell d\hat{x}_i^\ell}{(2\pi)^2} \right\} \sum_{\{\mathbf{S}^\ell, \hat{\mathbf{S}}^\ell\}} P(\mathbf{S}^0, \hat{\mathbf{S}}^0 | \mathbf{S}^L) e^{-i \sum_{\ell, i} \{ \psi_i^\ell S_i^\ell + \hat{\psi}_i^\ell \hat{S}_i^\ell \}} \quad (\text{A3})$$

$$\times e^{\sum_{i=1}^N \sum_{\ell=0}^{L-1} \{ \beta S_i^{\ell+1} H_i^\ell + \hat{\beta} \hat{S}_i^{\ell+1} \hat{H}_i^\ell - \ln 2 \cosh[\beta H_i^\ell] - \ln 2 \cosh[\hat{\beta} \hat{H}_i^\ell] + ix_i^\ell H_i^\ell + i\hat{x}_i^\ell \hat{H}_i^\ell \}} \\ \times \prod_{\ell, i=1}^{L, N} \prod_{j_1, \dots, j_k}^N e^{-i A_{j_1, \dots, j_k}^{\ell, i} \{ x_i^{\ell-1} \alpha_i^\ell(S_{j_1}^{\ell-1}, \dots, S_{j_k}^{\ell-1}) + \hat{x}_i^{\ell-1} \alpha_i^\ell(\hat{S}_{j_1}^{\ell-1}, \dots, \hat{S}_{j_k}^{\ell-1}) \}}. \quad (\text{A4})$$

Averaging Eq. (A4) over the connectivity [Eq. (5)] and gate disorders [Eq. (7)] results to leading order in

$$\begin{aligned}
Z_A & \sum_{\{A_{i_1, \dots, i_k}^{\ell, i}\}} P(\{A_{i_1, \dots, i_k}^{\ell, i}\}) \sum_{\{\alpha_i^\ell\}} P(\{\alpha_i^\ell\}) \prod_{\ell=1}^{L,N} \prod_{j_1, \dots, j_k}^N e^{-iA_{j_1, \dots, j_k}^{\ell, i} \{x_i^\ell \alpha_i^\ell (S_{j_1}^{\ell-1}, \dots, S_{j_k}^{\ell-1}) + \hat{x}_i^\ell \alpha_i^\ell (\hat{S}_{j_1}^{\ell-1}, \dots, \hat{S}_{j_k}^{\ell-1})\}} \\
& = \sum_{\{A_{i_1, \dots, i_k}^{\ell, i}\}} \prod_{\ell=1}^{L,N} \left\{ \int_{-\pi}^{\pi} \frac{d\omega_i^\ell}{2\pi} e^{i\omega_i^\ell} \prod_{i_1, \dots, i_k}^N \left[\frac{1}{N^k} \delta_{A_{i_1, \dots, i_k}^{\ell, i}; 1} + \left(1 - \frac{1}{N^k}\right) \delta_{A_{i_1, \dots, i_k}^{\ell, i}; 0} \right] \right. \\
& \quad \left. \times \sum_{\alpha_i^\ell} P(\alpha_i^\ell) e^{-iA_{i_1, \dots, i_k}^{\ell, i} \{x_i^{\ell-1} \alpha_i^\ell (S_{i_1}^{\ell-1}, \dots, S_{i_k}^{\ell-1}) + \hat{x}_i^{\ell-1} \alpha_i^\ell (\hat{S}_{i_1}^{\ell-1}, \dots, \hat{S}_{i_k}^{\ell-1}) + \omega_i^\ell\}} \right\} \\
& = \prod_{\ell=1}^{L,N} \int_{-\pi}^{\pi} \frac{d\omega_i^\ell}{2\pi} e^{i\omega_i^\ell} \exp \left[\frac{1}{N^k} \sum_{i_1, \dots, i_k}^N \langle e^{-i\{x_i^{\ell-1} \alpha (S_{i_1}^{\ell-1}, \dots, S_{i_k}^{\ell-1}) + \hat{x}_i^{\ell-1} \alpha (\hat{S}_{i_1}^{\ell-1}, \dots, \hat{S}_{i_k}^{\ell-1}) + \omega_i^\ell\}} - 1 \rangle_\alpha \right]. \tag{A5}
\end{aligned}$$

In the first line of Eq. (A5) we have used the integral representation of Kronecker delta function $\delta_{n,m} = \int_{-\pi}^{\pi} \frac{d\omega}{2\pi} e^{i\omega(n-m)}$; the final exponential form is valid for large N . By setting all x_i^ℓ and \hat{x}_i^ℓ variables to 0 in Eq. (A5), we find that the normalization constant Z_A contributes the factor e^{NL} to the generating functional [Eq. (A1)] for large N . Using the result of disorder-average [Eq. (A5)] in Eq. (A1) we obtain the disorder-averaged generating functional

$$\begin{aligned}
\overline{\Gamma[\boldsymbol{\psi}; \hat{\boldsymbol{\psi}}]} & = \prod_{i=1}^N \prod_{\ell=0}^{L-1} \left\{ \int \frac{iH_i^\ell dx_i^\ell d\hat{H}_i^\ell d\hat{x}_i^\ell}{(2\pi)^2} \int_{-\pi}^{\pi} \frac{d\omega_i^{\ell+1}}{2\pi} e^{i\omega_i^{\ell+1}} \right\} \\
& \quad \times \sum_{\{S^\ell, \hat{S}^\ell\}} P(S^0, \hat{S}^0 | S^L) e^{-i\sum_{\ell,i} \{\psi_i^\ell S_i^\ell + \hat{\psi}_i^\ell \hat{S}_i^\ell\} + NL} e^{\sum_{i=1}^N \sum_{\ell=0}^{L-1} \{\beta S_i^{\ell-1} H_i^\ell + \hat{\beta} \hat{S}_i^{\ell-1} \hat{H}_i^\ell - \ln 2 \cosh[\beta H_i^\ell] - \ln 2 \cosh[\hat{\beta} \hat{H}_i^\ell] + i x_i^\ell H_i^\ell + i \hat{x}_i^\ell \hat{H}_i^\ell\}} \\
& \quad \times \prod_{\ell=1}^L \exp \left[\frac{1}{N^k} \sum_{i_1, \dots, i_k}^N \langle e^{-i\{x_i^{\ell-1} \alpha (S_{i_1}^{\ell-1}, \dots, S_{i_k}^{\ell-1}) + \hat{x}_i^{\ell-1} \alpha (\hat{S}_{i_1}^{\ell-1}, \dots, \hat{S}_{i_k}^{\ell-1}) + \omega_i^\ell\}} - 1 \rangle_\alpha + O(N^{-k+1}) \right]. \tag{A6}
\end{aligned}$$

In order to achieve factorization over sites in the equation we isolate the densities

$$P^\ell(S, \hat{S}) = \frac{1}{N} \sum_{i=1}^N \delta_{S, S_i^\ell} \delta_{\hat{S}, \hat{S}_i^\ell}, \tag{A7}$$

$$\Omega^\ell(x, \hat{x}, \omega) = \frac{1}{N} \sum_{i=1}^N \delta(x - x_i^\ell) \delta(\hat{x} - \hat{x}_i^\ell) \delta(\omega - \omega_i^{\ell+1}) \tag{A8}$$

via the respective integrofunctional representations of unity

$$\begin{aligned}
& \int \{dP^\ell d\hat{P}^\ell\} e^{iN \sum_{S, \hat{S}} \hat{P}^\ell(S, \hat{S}) [P^\ell(S, \hat{S}) - (1/N) \sum_{i=1}^N \delta_{S, S_i^\ell} \delta_{\hat{S}, \hat{S}_i^\ell}]} = 1, \tag{A9} \\
& \int \{d\Omega^\ell d\hat{\Omega}^\ell\} e^{iN \int dx d\hat{x} d\omega \Omega^\ell(x, \hat{x}, \omega) [\Omega^\ell(x, \hat{x}, \omega) - (1/N) \sum_{i=1}^N \delta(x - x_i^\ell) \delta(\hat{x} - \hat{x}_i^\ell) \delta(\omega - \omega_i^{\ell+1})]} = 1,
\end{aligned}$$

which leads to

$$\begin{aligned}
\overline{\Gamma[\boldsymbol{\psi}; \hat{\boldsymbol{\psi}}]} & = \int \{dP d\hat{P} d\Omega d\hat{\Omega}\} \exp \left[N \sum_{\ell=0}^{L-1} \left\{ i \sum_{S, \hat{S}} \hat{P}^\ell(S, \hat{S}) P^\ell(S, \hat{S}) + i \int dx d\hat{x} d\omega \Omega^\ell(x, \hat{x}, \omega) \Omega^\ell(x, \hat{x}, \omega) \right\} \right] \\
& \quad \times \exp \left[N \sum_{\ell=0}^{L-1} \sum_{\{S_j, \hat{S}_j\}}^k \{P^\ell(S_j, \hat{S}_j)\} \int dx d\hat{x} d\omega \Omega^\ell(x, \hat{x}, \omega) \langle e^{-i\{x\alpha(S_1, \dots, S_k) + \hat{x}\alpha(\hat{S}_1, \dots, \hat{S}_k) + \omega\}} \rangle_\alpha \right] \\
& \quad \times \prod_{i=1}^N \left[\sum_{S_i^0, \hat{S}_i^0} \delta_{S_i^0, S_i^L} \delta_{\hat{S}_i^0, \hat{S}_i^L} \prod_{\ell=0}^{L-1} \left\{ \sum_{S_i^\ell, \hat{S}_i^\ell} \int \frac{dH_i^\ell dx_i^\ell d\hat{H}_i^\ell d\hat{x}_i^\ell}{(2\pi)^2} \int_{-\pi}^{\pi} \frac{d\omega_i^{\ell+1}}{2\pi} e^{i\omega_i^{\ell+1}} \right. \right. \\
& \quad \left. \left. \times e^{\beta S_i^{\ell+1} H_i^\ell + \hat{\beta} \hat{S}_i^{\ell+1} \hat{H}_i^\ell - \ln 2 \cosh[\beta H_i^\ell] - \ln 2 \cosh[\hat{\beta} \hat{H}_i^\ell] + i x_i^\ell H_i^\ell + i \hat{x}_i^\ell \hat{H}_i^\ell - i P^\ell(S_i^\ell, \hat{S}_i^\ell) - i \Omega^\ell(x_i^\ell, \hat{x}_i^\ell, \omega_i^{\ell+1})} \right\} e^{-i \sum_{\ell=0}^L \{\psi_i^\ell S_i^\ell + \hat{\psi}_i^\ell \hat{S}_i^\ell\}} \right] e^{O(LN^{-k+1})} \tag{A10}
\end{aligned}$$

The site-dependent part of the above equation can be written in the form

$$\exp \left[\sum_n \frac{1}{N} \sum_{i=1}^N \delta_{n;n_i} \ln \int \{d\mathbf{H}_i d\mathbf{x}_i d\hat{\mathbf{H}}_i d\hat{\mathbf{x}}_i\} \int D\boldsymbol{\omega}_i \sum_{\mathbf{S}_i, \hat{\mathbf{S}}_i} M_n[\mathbf{H}_i, \mathbf{x}_i; \hat{\mathbf{H}}_i, \hat{\mathbf{x}}_i; \boldsymbol{\omega}_i; \mathbf{S}_i, \hat{\mathbf{S}}_i] \right] \quad (\text{A11})$$

where we have defined the effective single-site measure

$$M_{n_i}[\mathbf{H}_i, \mathbf{x}_i; \hat{\mathbf{H}}_i, \hat{\mathbf{x}}_i; \boldsymbol{\omega}_i; \mathbf{S}_i, \hat{\mathbf{S}}_i] = \delta_{S_i^0, S_i^t} \delta_{\hat{S}_i^0, \hat{S}_i^t} e^{-i \sum_{\ell=0}^L \{\psi_i^{\ell} S_i^{\ell} + \hat{\psi}_i^{\ell} \hat{S}_i^{\ell}\}} \\ \times \prod_{\ell=0}^{L-1} e^{\beta S_i^{\ell+1} H_i^{\ell} + \hat{\beta} \hat{S}_i^{\ell+1} \hat{H}_i^{\ell} - \ln 2 \cosh[\beta H_i^{\ell}] - \ln 2 \cosh[\hat{\beta} \hat{H}_i^{\ell}] + i x_i^{\ell} H_i^{\ell} + i \hat{x}_i^{\ell} \hat{H}_i^{\ell} - i \hat{P}^{\ell}(S_i^{\ell}, \hat{S}_i^{\ell}) - i \hat{\Omega}^{\ell}(x_i^{\ell}, \hat{x}_i^{\ell}, \omega_i^{\ell+1})} \quad (\text{A12})$$

and use the notation $\int D\boldsymbol{\omega}_i = \prod_{\ell=1}^L \int_{-\pi}^{\pi} \frac{d\omega_i^{\ell}}{2\pi} e^{i\omega_i^{\ell}}$. Using Eq. (A11) in the disorder-averaged generating functional [Eq. (A10)] with all the generating fields $\{\psi_i^{\ell}, \hat{\psi}_i^{\ell}\}$ set to 0 and assuming that the law of large numbers for the random index-variables $\{n_i\}$ holds, i.e., $\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \delta_{n;n_i} \rightarrow P(n)$, we arrive at Eq. (10).

APPENDIX B: SIMPLIFICATION OF THE SADDLE-POINT PROBLEM

To solve the saddle-point Eqs. (13)–(16), we first use the saddle-point Eq. (16) to eliminate the conjugate order parameter $\hat{\Omega}^{\ell}$ from the effective measure [Eq. (12)], giving

$$M_n[\dots] = \delta_{S^0, S_n^t} \delta_{\hat{S}^0, \hat{S}_n^t} \prod_{\ell=0}^{L-1} \frac{e^{\beta S^{\ell+1} H^{\ell} + \hat{\beta} \hat{S}^{\ell+1} \hat{H}^{\ell}}}{2 \cosh(\beta H^{\ell}) 2 \cosh(\hat{\beta} \hat{H}^{\ell})} e^{i x^{\ell} H^{\ell} + i \hat{x}^{\ell} \hat{H}^{\ell} + i \omega^{\ell+1}} \\ \times \exp \left[\sum_{\{S_j, \hat{S}_j\}} \prod_{j=1}^k [P^{\ell}(S_j, \hat{S}_j)] \langle e^{-i \{x^{\ell} \alpha(\{S_j\}) + \hat{x}^{\ell} \alpha(\{\hat{S}_j\}) + \omega^{\ell+1}\}} \rangle_{\alpha} \right] e^{-i \hat{P}^{\ell}(S^{\ell}, \hat{S}^{\ell})}. \quad (\text{B1})$$

Second, using the above result we compute the Fourier transform

$$F_{\gamma}^{\ell'}[y, z] = \int \{d\mathbf{H} d\mathbf{x} d\hat{\mathbf{H}} d\hat{\mathbf{x}} D\boldsymbol{\omega}\} \sum_{\mathbf{S}, \hat{\mathbf{S}}} M_n[\mathbf{H}, \mathbf{x}; \hat{\mathbf{H}}, \hat{\mathbf{x}}; \boldsymbol{\omega}; \mathbf{S}, \hat{\mathbf{S}}] e^{-i \gamma \{x^{\ell'} y + \hat{x}^{\ell'} z + \omega^{\ell'+1}\}} \quad (\text{B2})$$

where $\gamma \in \{0, 1\}$. For $\gamma=0, 1$ we obtain

$$F_0^{\ell'}[y, z] = \sum_{\{S^{\ell}, \hat{S}^{\ell}\}} \delta_{S^0, S_n^t} \delta_{\hat{S}^0, \hat{S}_n^t} \prod_{\ell=0}^{L-1} W[S^{\ell+1}; \hat{S}^{\ell+1}] N[S^{\ell}; \hat{S}^{\ell}], \quad (\text{B3})$$

$$F_1^{\ell'}[y, z] = \sum_{\{S^{\ell}, \hat{S}^{\ell}\}} \delta_{S^0, S_n^t} \delta_{\hat{S}^0, \hat{S}_n^t} \tilde{W}[S^{\ell'+1}; \hat{S}^{\ell'+1}] N[S^{\ell'}; \hat{S}^{\ell'}] \prod_{\ell \neq \ell'}^{L-1} W[S^{\ell+1}; \hat{S}^{\ell+1}] N[S^{\ell}; \hat{S}^{\ell}], \quad (\text{B4})$$

respectively, where

$$W[S^{\ell+1}; \hat{S}^{\ell+1}] = \sum_{\{S_j, \hat{S}_j\}} \prod_{j=1}^k [P^{\ell}(S_j, \hat{S}_j)] \left\langle \frac{e^{\beta S^{\ell+1} \alpha(\{S_j\}) + \hat{\beta} \hat{S}^{\ell+1} \alpha(\{\hat{S}_j\})}}{2 \cosh\{\beta \alpha(\{S_j\})\} 2 \cosh\{\hat{\beta} \alpha(\{\hat{S}_j\})\}} \right\rangle_{\alpha} \quad (\text{B5})$$

$$\tilde{W}[S^{\ell'+1}; \hat{S}^{\ell'+1}] = \frac{e^{\beta S^{\ell'+1} y + \hat{\beta} \hat{S}^{\ell'+1} z}}{2 \cosh(\beta y) 2 \cosh(\hat{\beta} z)} \quad \text{and} \quad N[S^{\ell}; \hat{S}^{\ell}] = e^{-i \hat{P}^{\ell}(S^{\ell}, \hat{S}^{\ell})}. \quad (\text{B6})$$

Next we notice that $\int dx d\hat{x} d\omega \Omega^{\ell'}(x, \hat{x}, \omega) e^{-i\{xy+\hat{x}z+\omega\}} = \frac{F_1^{\ell'}[y,z]}{F_0^{\ell'}[y,z]}$. Using the fact that Eqs. (B5) and (B6) are both probability distributions the computation for $\ell'=L-1$ gives $\frac{F_1^{\ell'}[y,z]}{F_0^{\ell'}[y,z]}=1$. Plugging in this result into the saddle-point Eq. (14) gives us $\hat{P}^{L-1}(S, \hat{S})=ik$ implying that $N[S^{L-1}; \hat{S}^{L-1}]=e^k$. The latter is used to show that for $\ell'=L-2$ gives $\frac{F_1^{\ell'}[y,z]}{F_0^{\ell'}[y,z]}=1$ and so on until we conclude that $\hat{P}^{\ell}(S, \hat{S})=ik$ for all ℓ .

APPENDIX C: CONVERGENCE RATE

To show how the asymptotic form [Eq. (22)] can be obtained we first note that

$$\text{const} \times \left| \left\{ \frac{d}{dm} F_{\epsilon}^k[m] \right\}_{m=m(\infty)} \right|^{\ell} \leq |m(\ell) - m(\infty)|, \quad (\text{C1})$$

where $F_{\epsilon}^k[m]$ is the right hand side of Eq. (20), i.e., the distance $|m(\ell) - m(\infty)|$ for an arbitrary point $m(\ell)$ is always

sgreater than the distance for $m(\ell)=m(\infty)+\Delta m(\ell)$. The lower bound [Eq. (C1)] can be made into an upper bound by choosing an appropriate constant [29]. Computing the left hand side in [Eq. (C1)] for the MAJ- k circuit with $\epsilon = \epsilon^*(k) \pm \Delta\epsilon$ results in

$$\left\{ \left[1 - \frac{2\Delta\epsilon}{1-2\epsilon^*(k)} \right] [1 - m^2(\infty)]^{(k-1)/2} \right\}^{\ell} \leq |m(\ell) - m(\infty)|. \quad (\text{C2})$$

Since $[1 - \frac{2\Delta\epsilon}{1-2\epsilon^*(k)}][1 - m^2(\infty)]^{(k-1)/2} \leq [1 - \frac{2\Delta\epsilon}{1-2\epsilon^*(k)}]$, the convergence rate in the paramagnetic region [$\epsilon > \epsilon^*(k)$] is slower than in the ferromagnetic one [$\epsilon < \epsilon^*(k)$]. The latter is due to the amplification of thermal fluctuations which are only suppressed for $\ell \rightarrow \infty$. The critical noise $\epsilon^*(k)$ for large k can be approximated by $\epsilon^*(k) \approx \frac{1}{2}(1 - \frac{\sqrt{\pi}}{2k})$. Inserting this into the Eq. (C2) for $m(\infty)=0$ gives us the asymptotic form [Eq. (22)].

-
- [1] S. Borkar, *IEEE MICRO* **25**, 10 (2005).
[2] A. Ekert and R. Jozsa, *Rev. Mod. Phys.* **68**, 733 (1996).
[3] J. Von Neumann, *Probabilistic Logics and the Synthesis of Reliable Organisms from Unreliable Components* (Princeton University Press, Princeton, NJ, 1956), p. 43–98, Automata Studies.
[4] N. Pippenger, *IEEE Trans. Inf. Theory* **34**, 194 (1988).
[5] T. Feder, *IEEE Trans. Inf. Theory* **35**, 569 (1989).
[6] B. Hajek and T. Weller, *IEEE Trans. Inf. Theory* **37**, 388 (1991).
[7] W. Evans and L. Schulman, *IEEE Trans. Inf. Theory* **49**, 3094 (2003).
[8] W. Evans and N. Pippenger, *IEEE Trans. Inf. Theory* **44**, 1299 (1998).
[9] F. Unger, *IEEE Trans. Inf. Theory* **54**, 3693 (2008).
[10] M. Mézard and A. Montanari, *Information, Physics, and Computation* (Oxford University Press, Oxford, 2009).
[11] Y. Kabashima and D. Saad, *J. Phys. A* **37**, R1 (2004).
[12] P. Savický, *Discrete Math.* **83**, 95 (1990).
[13] A. Brodsky and N. Pippenger, *Random Struct. Algorithms* **27**, 490 (2005).
[14] I. Wegener, *The Complexity of Boolean Functions* (B. G. Teubner, Stuttgart, 1987).
[15] H. Lefmann and P. Savický, *Random Struct. Algorithms* **10**, 337 (1997).
[16] B. Chauvin, P. Flajolet, D. Gardy, and B. Gittenberger, *Combinatorics, Probab. Comput.* **13**, 475 (2004).
[17] D. Gardy and A. Woods, *DMTCS Proceedings* (2005), Vol. AD, pp. 139 [<http://www.emis.de/journals/DMTCS/proceedings/abstracts/dmAD0113.abs.html>].
[18] K. Mimura and A. C. C. Coolen, *J. Phys. A: Math. Theor.* **42**, 415001 (2009).
[19] C. De Dominicis, *Phys. Rev. B* **18**, 4913 (1978).
[20] H. Fournier, D. Gardy, and A. Genitrini, in *6th SIAM Workshop on Analytic Algorithms and Combinatorics (ANALCO)* (SIAM, New York, 2009), pp. 51–57.
[21] A. Mozeika, D. Saad, and J. Raymond, *Phys. Rev. Lett.* **103**, 248701 (2009).
[22] M. Minsky and S. Papert, *Perceptrons: An Introduction to Computational Geometry*, 2nd ed. (MIT Press, Cambridge, MA, 1972).
[23] A. C. C. Coolen, R. Kühn, and P. Sollich, *Theory of Neural Information Processing Systems* (Oxford University Press, Oxford, 2005).
[24] I. Neri and D. Bollé, *J. Stat. Mech.: Theory Exp.* (2009) P08009.
[25] B. Derrida, E. Gardner, and A. Zippelius, *EPL* **4**, 167 (1987).
[26] A. Mozeika and D. Saad (to be published).
[27] F. Barahona, *J. Phys. A* **15**, 3241 (1982).
[28] J. Machta and R. Greenlaw, *J. Stat. Phys.* **82**, 1299 (1996).
[29] J. M. Ortega, *SIAM (Soc. Ind. Appl. Math.) J. Numer. Anal.* **10**, 268 (1973).
[30] To establish the connection between [20] and our work we use the mapping $S_i = 1 - 2x_i$ from $x_i \in \{1, 0\}$ to $S_i \in \{-1, 1\}$.