The Annals of Applied Probability
2000, Vol. 10, No. 2, 410–433

BROADCASTING ON TREES AND THE ISING MODEL

BY WILLIAM EVANS, CLAIRE KENYON, YUVAL PERES[1]
AND LEONARD J. SCHULMAN[2]

*University of Arizona, Université de Paris Sud,
Hebrew University and University of California at Berkeley
and Georgia Institute of Technology*

Consider a process in which information is transmitted from a given root node on a noisy tree network T. We start with an unbiased random bit R at the root of the tree and send it down the edges of T. On every edge the bit can be reversed with probability ε, and these errors occur independently. The goal is to reconstruct R from the values which arrive at the nth level of the tree. This model has been studied in information theory, genetics and statistical mechanics. We bound the reconstruction probability from above, using the maximum flow on T viewed as a capacitated network, and from below using the electrical conductance of T. For general infinite trees, we establish a sharp threshold: the probability of correct reconstruction tends to 1/2 as $n \to \infty$ if $(1-2\varepsilon)^2 < p_c(T)$, but the reconstruction probability stays bounded away from 1/2 if the opposite inequality holds. Here $p_c(T)$ is the critical probability for percolation on T; in particular $p_c(T) = 1/b$ for the $b+1$-regular tree. The asymptotic reconstruction problem is equivalent to purity of the "free boundary" Gibbs state for the Ising model on a tree. The special case of regular trees was solved in 1995 by Bleher, Ruiz and Zagrebnov; our extension to general trees depends on a coupling argument and on a reconstruction algorithm that weights the input bits by the electrical current flow from the root to the leaves.

1. Introduction. Consider the following broadcast process. At the root ρ of a tree T a binary random variable is chosen uniformly at random. This bit is then propagated, with error, throughout the tree as follows: for a fixed $\varepsilon \in (0, 1/2]$, each vertex receives the bit at its parent with probability $1-\varepsilon$, and the opposite bit with probability ε. These events at the vertices are statistically independent. (In the language of communication theory, each edge of the tree is functioning as a binary symmetric channel.) This model has been studied in information theory, mathematical genetics and statistical physics; some of the history is described in Section 2.

Suppose we are given the bits that arrived at some fixed set of vertices W of the tree. Using the optimal reconstruction strategy (maximum likelihood), the probability of correctly reconstructing the original bit at the root is clearly at least 1/2; denote this probability by $(1+\Delta)/2$. Our main results are a lower bound for $\Delta = \Delta(T, W, \varepsilon)$ in terms of the the effective electrical conductance

Received November 1998; revised July 1999.

[1]Supported in part by NSF Grants DMS-94-04391 and DMS-98-03597.

[2]Supported in part by a NSF Mathematical Sciences Postdoctoral Fellowship.

AMS 1991 *subject classifications*. Primary 60K35; secondary 90B15, 68R99.

Key words and phrases. Tree, percolation, Ising model, branching number, electrical network, noisy computation.

from the root ρ to W (Theorem 1.2), and an upper bound for Δ which is the maximum flow from ρ to W for certain edge capacities (Theorem 1.3.) When T is an infinite tree, these bounds allow us to determine (in Theorem 1.1) the critical parameter ε_c so that, denoting the nth level of T by T_n, we have

$$\lim_{n\to\infty} \Delta(T, T_n, \varepsilon) \begin{cases} > 0, & \text{if } \varepsilon < \varepsilon_c, \\ = 0, & \text{if } \varepsilon > \varepsilon_c. \end{cases} \tag{1}$$

As we explain in the next section, vanishing of the above limit is equivalent to extremality of the "free boundary" limiting Gibbs state for the ferromagnetic Ising model. For the special case of regular trees, the problem of determining ε_c was open for two decades, and was finally solved in 1995 by Bleher, Ruiz and Zagrebnov [2]; see Section 2.2 for background.

Because of the Ising model interpretation, we will label the vertices of T with ± 1 valued random variables $\{\sigma_v\}$, called *spins*, instead of random bits.

These spins can be constructed from independent variables $\{\eta_e\}$ labeling the edges of T, as follows. For each edge e, let $\mathbf{P}[\eta_e = -1] = \varepsilon = 1 - \mathbf{P}[\eta_e = 1]$. Let σ_ρ be a uniformly chosen spin, and for any other vertex v let

$$\sigma_v := \sigma_\rho \prod_e \eta_e, \tag{2}$$

where the product is over all edges e on the path from ρ to v. Given $\sigma_W = \{\sigma_v \colon v \in W\}$, the strategy which maximizes the probability of correctly reconstructing σ_ρ, is to decide according to the sign of $\mathbf{E}(\sigma_\rho \mid \sigma_W)$; with this strategy, the difference between the probabilities of correct and incorrect reconstruction is

$$\Delta(T, W, \varepsilon) = \mathbf{E}\Big|\mathbf{P}(\sigma_\rho = 1 \mid \sigma_W) - \mathbf{P}(\sigma_\rho = -1 \mid \sigma_W)\Big|. \tag{3}$$

Alternatively, $\Delta(T, W, \varepsilon)$ can be interpreted as the *total variation distance* between the conditional distributions of σ_W given $\sigma_\rho = 1$ and given $\sigma_\rho = -1$; see Section 4.2. The dependence between σ_ρ and σ_W is also captured by the *mutual information* (discussed in Section 4.1),

$$I(\sigma_\rho; \sigma_W) := \sum_{x,y} \mathbf{P}[\sigma_\rho = x, \sigma_W = y] \log \frac{\mathbf{P}[\sigma_\rho = x, \sigma_W = y]}{\mathbf{P}[\sigma_\rho = x]\mathbf{P}[\sigma_W = y]}.$$

For an infinite tree T, a remarkably good summary of its behavior in probabilistic contexts is provided by its *branching number* $\mathrm{br}(T)$, introduced in [24]. This is the supremum of the real numbers $\lambda \geq 1$, such that T admits a positive flow from the root to infinity, if on every edge e of T, the flow is bounded by $\lambda^{-|e|}$. Here $|e|$ denotes the number of edges, (including e) on the path from e to the root; $\mathrm{br}(T)^{-1}$ is the critical probability for Bernoulli percolation on T. The equivalent definitions of $\mathrm{br}(T)$ in terms of percolation, cutset sums and electrical conductance, are reviewed in Section 3.

THEOREM 1.1. *Let T be an infinite tree with root ρ, and suppose its vertices are assigned random spins $\{\sigma_v\}$, using the flip probability $\varepsilon < 1/2$ as in* (2).

Consider the problem of reconstructing σ_ρ *from the spins at the nth level* T_n *of* T:

(i) *If* $1-2\varepsilon > \mathrm{br}(T)^{-1/2}$ *then* $\inf_{n\ge1}\Delta(T,T_n,\varepsilon)>0$ *and* $\inf_{n\ge1} I(\sigma_\rho;\sigma_{T_n})>0$.
(ii) *If* $1-2\varepsilon < \mathrm{br}(T)^{-1/2}$ *then* $\inf_{n\ge1}\Delta(T,T_n,\varepsilon)=0$ *and* $\inf_{n\ge1} I(\sigma_\rho;\sigma_{T_n})=0$.

The tail field of the random variables $\{\sigma_v\}_{v\in T}$ *contains events with probability strictly between* 0 *and* 1 *in case* (i), *but not in case* (ii).

Thus in the notation of (1), $\varepsilon_c = (1-\mathrm{br}(T)^{-1/2})/2$. As mentioned above, this is already known when T is a $b+1$-regular tree [for which $\mathrm{br}(T) = b$]. Theorem 1.1 is considerably more general, as there are many other trees (e.g., Galton-Watson trees, subperiodic trees) for which one can calculate $\mathrm{br}(T)$ explicitly (see Section 3); and $\mathrm{br}(T)$ is well defined for any infinite tree. Simple examples show that at criticality, when $1-2\varepsilon = \mathrm{br}(T)^{-1/2}$, asymptotic solvability of the reconstruction problem is not determined by the branching number; see Section 9. To see the relevance of the quantity $1-2\varepsilon$ appearing in Theorem 1.1, note the following equivalent construction of the random variables $\{\sigma_v\}$: perform independent bond percolation on T with parameter $\theta = 1-2\varepsilon$ (the probability of open bonds), and independently assign to each of the resulting percolation clusters a uniform random spin (the same spin is assigned to all vertices in each cluster). This is a special case of the Fortuin–Kasteleyn random cluster representation of the Ising model (see, e.g., [14]); on a tree, it is elementary to verify the equivalence of this representation with the construction (2).

The following two theorems contain estimates of reconstruction probability and mutual information that imply Theorem 1.1. The notion of effective resistance, used in the next theorem, is explained in [7] and [26]. On a tree, effective resistance is easily calculated via the parallel and series laws.

THEOREM 1.2. *Let* T *be a tree with root* ρ, *and let* W *be a finite set of vertices in* T. *Given* $\varepsilon\in(0,1/2]$, *denote* $\theta := 1-2\varepsilon$, *and consider the electrical network obtained by assigning to each edge* e *of* T *the resistance* $(1-\theta^2)\theta^{-2|e|}$. *Then*

$$\left.\begin{array}{r}\Delta(T,W,\varepsilon)\\ I(\sigma_\rho;\sigma_W)\end{array}\right\} \ge \frac{1}{1+\mathscr{R}_{\mathrm{eff}}(\rho\leftrightarrow W)}, \tag{4}$$

where $\mathscr{R}_{\mathrm{eff}}$ *denotes effective resistance.*

Our proof of this theorem is based on reconstruction by weighted majority vote, that is, reconstruction according to the sign of an unbiased linear estimator of the root spin. We relate the variance of such an estimator to the energy of a corresponding unit flow from ρ to W. The unit flow of minimal energy is the electrical current flow, and its energy is the effective resistance between ρ and W. The proof is completed by invoking a general lemma, which bounds Δ and $I(\sigma_\rho;\sigma_W)$ from below by the reciprocal of the variance of any unbiased linear estimator for σ_ρ. We find it quite surprising that on any infinite tree, reconstruction using such linear estimators has the same threshold as maximum-likelihood reconstruction.

Next, we present an upper bound on Δ and $I(\sigma_\rho;\sigma_W)$. Say that a set of vertices W_1 *separates* ρ *from* W if any path from ρ to W intersects W_1. For a vertex v of T, denote by $|v|$ the number of edges on the path from v to ρ.

THEOREM 1.3. *Let W be a finite set of vertices in the tree T. For any set of vertices W_1 that separates the root ρ from W, we have*

$$\Delta(T, W, \varepsilon)^2 \le 2\left(1 - \prod_{v\in W_1} \sqrt{1-\theta^{2|v|}}\right) \le 2 \sum_{v\in W_1} \theta^{2|v|} \tag{5}$$

and

$$I(\sigma_\rho;\sigma_W) \le \sum_{v\in W_1} I(\sigma_\rho;\sigma_v) \le \sum_{v\in W_1} \theta^{2|v|}. \tag{6}$$

In view of the mincut-maxflow theorem, (6) is an upper bound on mutual information in terms of the maximum flow in a capacitated network. Theorem 1.3 is proved by comparing the given tree T to a "stringy tree" $\widehat{T}$ which has an isomorphic set of paths from the root to the vertices of W_1, but these paths are pairwise edge disjoint. We show that $\Delta(T, W, \varepsilon) \le \Delta(\widehat{T}, W_1, \varepsilon)$ by constructing, in Theorem 6.1, a noisy channel that maps the spins on W_1 in $\widehat{T}$ to the spins on W in T.

Symmetric trees. A tree T is called *spherically symmetric* if for every $n \ge 1$, all vertices in T_n have the same degree. For such a tree, the effective resistance from the root to level n is easily computed, and we infer from Theorems 1.1–1.3 that

$$\left(2 + 2(1-\theta^2)\sum_{k=1}^{n} \frac{\theta^{-2k}}{|T_k|}\right)^{-1} \le I(\sigma_\rho;\sigma T_n) \le \inf_{k\le n} |T_k|\theta^{2k} \tag{7}$$

and $(1-2\varepsilon_c)^{-2} = \liminf_n |T_n|^{1/n}$.

Since reconstruction using majority vote is crucial to our proof of Theorem 1.2 (at least in the spherically symmetric case), we examined closely the distribution of the spin sum $S_n := \sum_{v\in T_n} \sigma_v$ given σ_ρ. Two other motivations for this are:

1. In some instances of the reconstruction problem, the sites where the given spins are located are unknown, and S_n is the only data available.
2. For small ε, the partial sums of the spins on the leaves of a regular tree define a stochastic process which is "less predictable than simple random walk"; see [1] and [16] for precise formulations and applications.

The Harris–FKG inequality (see, e.g., [14]) implies that the events $\sigma_\rho = 1$ and $S_n > 0$ are positively correlated. However, the following more delicate inequality is required to conclude that when we are (only) given S_n, maximum-likelihood reconstruction coincides with majority vote.

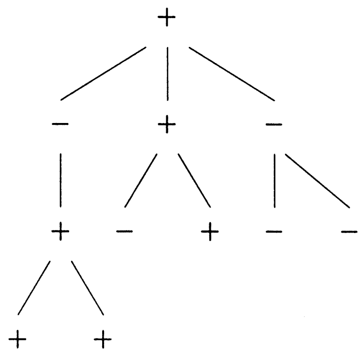

FIG. 1. *Tree with $+, -$ spins at the vertices.*

THEOREM 1.4. *Let T be a spherically symmetric tree of depth n, and denote $S_n := \sum_{v \in T_n} \sigma_v$. Then for any error probability $\varepsilon \in (0, 1/2]$,*

$$\forall k > 0 \qquad \mathbf{P}[S_n = k \mid \sigma_\rho = 1] \geq \mathbf{P}[S_n = k \mid \sigma_\rho = -1]. \tag{8}$$

This inequality also holds when the error probabilities vary, as long as they too are spherically symmetric (fixed in each level of T).

The example in Figure 3 shows that even on a regular tree, majority vote can disagree with maximum likelihood when the spin configuration σ_{T_n} is given.

Given the boundary data in Figure 3, the root spin σ_ρ is more likely to be -1 than $+1$ provided that ε is sufficiently small, since $\sigma_\rho = +1$ requires four spin flips, while $\sigma_\rho = -1$ requires only three spin flips.

Organization. In Section 2.1–2.3 we describe how the model above arose in computer science, statistical mechanics and genetics. In Section 3 we review

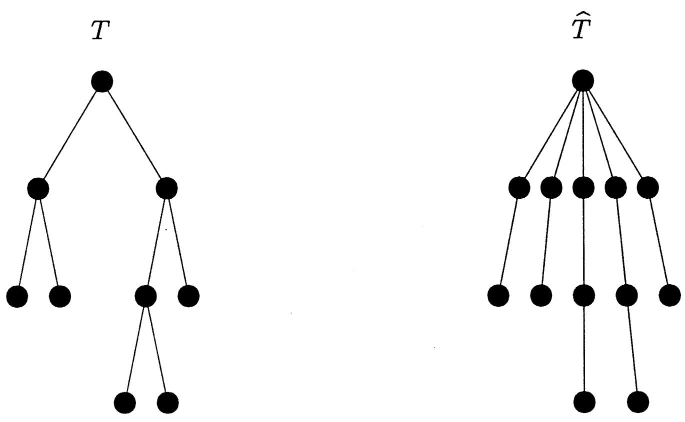

FIG. 2. *A Tree T and the corresponding stringy tree $\widehat{T}$.*

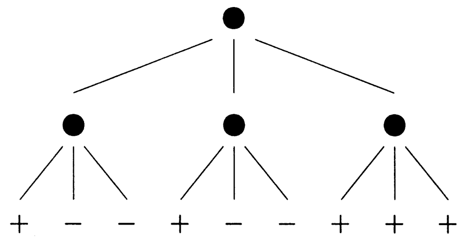

FIG. 3. *Majority vote can disagree with maximum likelihood.*

the notion of branching number and infer Theorem 1.1 from Theorems 1.2–1.3. In Section 4 we collect probabilistic and information-theoretic material that we will use in the proofs of Theorems 1.2–1.3. The conductance lower bound for reconstruction, Theorem 1.2, is generalized to allow edge-dependent error probabilities and proved in Section 5. A similar generalization of the upper bound, Theorem 1.3, is established in Section 6. The correlation inequality concerning majority vote, Theorem 1.4, is proved in Section 7. Several reconstruction algorithms are compared in Section 8. Extensions and unsolved problems are discussed in Section 9.

2. Background.

2.1. *Noisy computation.* A (computational) *circuit* is a directed acyclic graph in which each internal node is labelled by a Boolean logic gate. If Boolean values ("bits") are "input" at the sources of the graph, each edge of the graph carries the bit obtained by applying the gate at its starting node, to the values entering that gate. The output of the circuit is the sequence of bits reaching the sink nodes. The size of the circuit is the number of edges in the graph; its depth is the length of a maximal path from a source to a sink. We will focus on the case that the circuit has a single output bit. Von Neumann [38] proposed a model of computation in noisy circuits where each gate computes correctly with probability $1 - \varepsilon$, independently of all other gates. He proved that if ε is sufficiently small, then there exists $p > 1/2$ such that for any Boolean function f there is such a noisy circuit $\mathscr{C}_f$, using gates of bounded indegree, with the following property. For each input string x, the output of $\mathscr{C}_f$ equals $f(x)$ with probability at least p. Von Neumann showed how such a noisy circuit can be constructed from a noiseless circuit that computes f, using gates of the same bounded indegree, at the cost of increasing the depth of the circuit by a bounded factor. Pippenger [32] subsequently showed that increase in depth by a factor greater than 1 is necessary for some circuits, and furthermore, that such a simulation is impossible beyond a certain level of noise.

Evans and Schulman [8, 10, 11] improved these results by a modification of Pippenger's method. The proof technique was to bound the mutual information between a (random) input bit to the circuit, and the set of bits at the outputs of a set of gates W in the circuit. (The output bits depend on both the random input and the random noise in the circuit). Using a "quantified data processing lemma," Evans and Schulman proved that

$$I(\sigma_\rho; \sigma_W) \le \sum_{w\in W} \sum_{\gamma\in\Gamma(\rho, w)} \theta^{2|\gamma|}, \tag{9}$$

where $\Gamma(\rho, w)$ is the set of paths connecting the input gate ρ to the output gate w, and $|\gamma|$ is the length of the path γ (see Lemma 3.2.1 in [8]).

Note the similarity with Theorem 1.3 above. The noisy computation problem considered in [8, 10, 11] is more general than the broadcasting problem studied here; for the latter problem, the left-hand inequality in (6) gives a sharper upper bound on mutual information.

2.2. *The Ising model.* Let G be a finite undirected graph with vertex set V; let $u \sim v$ indicate that vertices u and v are adjacent. In the *ferromagnetic Ising model* with no external field on G, the *interaction strength* $J > 0$ and the *temperature* $t > 0$ determine a *Gibbs distribution* $\mathscr{G} = \mathscr{G}_{J,t}$ on $\{\pm 1\}^V$ which is defined by

$$\mathscr{G}(\sigma) = Z(t)^{-1} \exp\left(\sum_{u\sim v} J\sigma_u\sigma_v/t \right), \tag{10}$$

where the normalizing factor $Z(t)$ is called the *partition function*. If the graph G is a tree, then this is equivalent to the Markovian propagation description in Section 1 for an appropriate choice of the error parameter ε. Indeed, if $u \sim v$ are adjacent vertices in a finite tree with $\sigma_u = \sigma_v$, then flipping all the spins on one side of the edge connecting u and v will multiply the probability in (10) by $\exp(-2J/t)$. Thus if we define ε by

$$\frac{\varepsilon}{1-\varepsilon} = \exp(-2J/t), \tag{11}$$

then the distributions defined by (2) and (10) coincide. For an infinite graph G, a weak limit point of the Gibbs distributions (10) on finite subgraphs $\{G_n\}$ exhausting G (possibly with boundary conditions imposed on $\sigma_{\partial G_n}$), is called a (limiting) *Gibbs state* on G. See [13] for more complete definitions, using the notion of specification.

For any infinite graph with bounded degrees, the limiting Gibbs state is unique at sufficiently high temperatures, that is, the limit from finite subgraphs exists and does not depend on boundary conditions. When $G = T$ is a tree, this means that

$$\lim_{n\to\infty} \mathbf{E}\big[\sigma_\rho \mid \sigma_{T_n} \equiv 1\big] = 0 \tag{12}$$

at high temperatures. Some graphs admit a phase transition: below a certain critical temperature, multiple Gibbs states appear and the limit in (12) is strictly positive. The critical temperature t_c^+ for this transition on a regular tree T was determined in 1974, [33]; this result was generalized in

1989 [23], which showed that $\tanh(J/t_c^+) = \mathrm{br}(T)^{-1}$; in the equivalent Markovian description, the critical parameter ε_c^+ for an all + boundary to affect σ_ρ in the limit satisfies $1 - 2\varepsilon_c^+ = \mathrm{br}(T)^{-1}$.

In general, a Gibbs state is extremal (or "pure") iff it has a trivial tail; see [13], Theorem 7.7. The tree-indexed Markov chain (2) on an infinite tree T is the limit of the Gibbs distributions (10) on finite subtrees, with no boundary conditions imposed; hence it is called the *free boundary Gibbs state* on T. In 1975 Spitzer ([34], Theorem 4) claimed that on a $b+1$-regular tree $T^{(b)}$, the free boundary Gibbs states are extremal at any temperature. A counterexample, due to T. Kamae, was published in 1977 (see [18]). Kamae showed that the sum of spins on $T_n^{(b)}$, normalized by its L^2 norm, converges to a nonconstant tail-measurable function, provided that $1 - 2\varepsilon > b^{-1/2}$. In 1978, this result was put in a broader context by Moore and Snell [27], who showed it followed from the 1966 results, [21], on multitype branching processes. Moore and Snell noted that it was open whether the free boundary Gibbs state on $T^{(b)}$ is extremal when $b^{-1} < 1 - 2\varepsilon \le b^{-1/2}$. Chayes, Chayes, Sethna and Thouless [5] successfully analyzed a closely related spin-glass model on T_b; by a gauge transformation, this is equivalent to the Ising model with i.i.d. uniform $\{\pm 1\}$ boundary conditions. Although these boundary conditions are quite different from a free boundary, they turn out to have the same critical temperature. Bleher, Ruiz and Zagrebnov [2] adapted the recursive methods of [5] to the extremality problem and showed that the free boundary Gibbs state on $T^{(b)}$ is extremal whenever $1-2\varepsilon \le b^{-1/2}$. Shortly thereafter, an alternative streamlined argument was found by Ioffe [19]. We remark that at noncritical parameters, this extremality is implied by the inequality of [8], Lemma 3.2.1, quoted in Section 2.1, but the connection was not realized at the time. Theorem 1.1 was first established by the authors in [9]. After hearing a lecture by one of us on [9], Ioffe [20] gave a completely different, and quite beautiful, proof of the upper bound in Theorem 1.1.

2.3. *Genetic reconstruction and parsimony.* Tree-indexed Markov chains as in the introduction have been studied in the mathematical biology literature, [4], [36] and others. In that literature the two "spins" are often called "colors," and correspond to traits of individuals, species, or DNA sequences. The "broadcasting errors" (color changes along edges) represent mutations, and one attempts to infer traits of ancestors from those of an observable population. The preferred method of reconstruction there is *parsimony*; given a two-coloring of the leaves, the internal nodes are colored so as to minimize the number of bicolored edges. (This particular meaning of "parsimony" is typical of its broader use as an inference criterion in the computational biology literature; see [12] and [17].) Parsimony is discussed further in Section 8.

3. The branching number and proof of Theorem 1.1. Recall from the introduction the definition of the branching number of an infinite tree. Lyons [24] describes specific trees for which the branching number is strictly smaller than the exponential growth rate, and gives several equivalent characterizations of the branching number of an infinite tree T.

1. *Percolation*. Suppose that the edges of T are independently removed with probability $1-p$ and retained with probability p. If $p < \mathrm{br}(T)^{-1}$, then all connected components of the resulting subgraph are finite a.s., while if $p > \mathrm{br}(T)^{-1}$, then this subgraph has infinite components a.s.
2. *Minimum cutset sums*. $\mathrm{br}(T)$ is the supremum of the numbers $\lambda \geq 1$ such that $\inf_\Pi \sum_{v\in\Pi} \lambda^{-|v|} > 0$, where the infimum is over all cutsets Π (a set of vertices is called a *cutset* if it intersects every infinite path emanating from ρ).
3. *Electrical resistance*. Assign each edge e of T conductance $\lambda^{-|e|}$. Then $\mathrm{br}(T)$ is the supremum of the numbers $\lambda \geq 1$ such that $\sup_{n\geq 1} \mathscr{R}_{\mathrm{eff}}(\rho \leftrightarrow T_n) < \infty$.

As noted in Section 2.2, [23] characterized the critical temperature for uniqueness of Gibbs states on T in terms of $\mathrm{br}(T)$; our Theorem 1.1 gives such a characterization for the critical temperature for extremality. Next, we derive this theorem from Theorems 1.2 and 1.3.

PROOF OF THEOREM 1.1. (i) From $\theta = 1-2\varepsilon > \mathrm{br}(T)^{-1/2}$ it follows that

$$\mathscr{R}_{\mathrm{eff}}(\rho \leftrightarrow \infty) := \sup_n \mathscr{R}_{\mathrm{eff}}(\rho \leftrightarrow T_n) < \infty$$

when each edge e is assigned conductance $\theta^{2|e|}$. By (4),

$$\inf_{n\geq 1} \Delta(T, T_n, \varepsilon) \geq \inf_{n\geq 1} \frac{1}{1+\mathscr{R}_{\mathrm{eff}}(\rho \leftrightarrow T_n)} \geq \frac{1}{1+\mathscr{R}_{\mathrm{eff}}(\rho \leftrightarrow \infty)} > 0$$

and similarly $\inf_{n\geq 1} I(\sigma_\rho; \sigma_{T_n}) > 0$, as asserted. In particular, σ_ρ is not independent of the tail field of $\{\sigma_v\}$, so this tail field is not trivial.

(ii) If $\theta = 1-2\varepsilon < \mathrm{br}(T)^{-1/2}$ then $\inf_\Pi \sum_{v\in\Pi} \theta^{2|v|} = 0$, so Theorem 1.3 implies that $\inf_{n\geq 1} \Delta(T, T_n, \varepsilon) = 0$ and $\inf_{n\geq 1} I(\sigma_\rho; \sigma_{T_n}) = 0$. Next, fix a finite set of vertices W_0. For each $w \in W_0$ and $n > |w|$, denote by $T_n(w)$ the set of vertices in T_n which connect to ρ via w. Then Lemma 4.1(iii) implies that for sufficiently large n,

$$I(\sigma_{W_0}; \sigma_{T_n}) \leq \sum_{w\in W_0} I(\sigma_{W_0}, \sigma_{T_n(w)}) = \sum_{w\in W_0} I(\sigma_w, \sigma_{T_n(w)}), \tag{13}$$

since the conditional distribution of $\sigma_{T_n(w)}$ given σ_{W_0} is the same as its conditional distribution given σ_w. For any finite W_0, the right-hand side of (13) tends to 0 as $n \to \infty$; It follows that the tail of $\{\sigma_v\}$ is trivial. $\square$

4. Tools from information theory and statistics.

4.1. *Mutual information.*

DEFINITION. Let X, Y be random variables defined on the same probability space which take finitely many values. The *entropy* of X is defined by

$$H(X) := -\sum_x \mathbf{P}[X = x] \log \mathbf{P}[X = x]$$

and the *mutual information* $I(X;Y)$ between X and Y is defined to be

$$\begin{aligned} I(X;Y) &:= H(X) + H(Y) - H(X, Y) \\ &= \sum_{x, y} \mathbf{P}[X = x, Y = y] \log \frac{\mathbf{P}[X = x, Y = y]}{\mathbf{P}[X = x]\mathbf{P}[Y = y]}. \end{aligned}$$

We collect a few basic properties of mutual information in the following lemma. See, for example, [6], Section 2.

LEMMA 4.1. (i) $I(X;Y) \geq 0$, *with equality iff X and Y are independent.*
(ii) *Data processing inequality*: *if $X \mapsto Y \mapsto Z$ form a Markov chain* (*i.e.*, *X and Z are conditionally independent given Y*), *then $I(X;Y) \geq I(X;Z)$.*
(iii) *Subadditivity*: *if $Y_1, \dots, Y_n$ are conditionally independent given X, then $I(X;(Y_1, \dots, Y_n)) \leq \sum_{j=1}^n I(X;Y_j)$.*

The assumption of conditional independence in part (iii) cannot be omitted, as is shown by standard examples of three dependent random variables which are pairwise independent (e.g., Boolean variables satisfying $X = Y_1 + Y_2 \bmod 2$). Nevertheless, inequality (6) in Theorem 1.3 extends (iii) to a setting where this conditional independence need not hold.

4.2. *Distances between probability measures.* Let ν_+ and ν_- be two probability measures on the same space Ω. (In our application Ω is finite, but it is convenient to use notation that applies more generally.) Set $\nu := (\nu_+ + \nu_-)/2$ and denote $f = d\nu_+/d\nu$, $g = d\nu_-/d\nu$, so that $f + g \equiv 2$. Suppose that ξ is uniform in $\{\pm 1\}$, and X has distribution ν_ξ. Inferring ξ from X is a basic problem of Bayesian hypothesis testing. (In our application, ξ will be the root spin σ_ρ, and X will be some function of the spin configuration σ_W on a finite vertex set W.)

There are several important notions of distance between ν_+ and ν_-, that can be related to this inference problem.

1. *Total variation distance.* $D_V(\nu_+, \nu_-) := \frac{1}{2} \int |f - g| \, d\nu$ can be interpreted as the difference between the probabilities of correct and erroneous inference. Indeed, among all functions $\widehat{\xi}$ of the observations, the probability of error

$\mathbf{P}[\widehat{\xi} \neq \xi]$ is minimized by taking $\widehat{\xi} = 1$ if $f(X) \geq g(X)$, and $\widehat{\xi} = -1$ otherwise. We then have

$$\Delta := \mathbf{P}[\widehat{\xi} = \xi] - \mathbf{P}[\widehat{\xi} \neq \xi] = \frac{1}{2}\left(\int \widehat{\xi} f \, d\nu - \int \widehat{\xi} g \, d\nu\right) = \frac{1}{2}\int |f - g| \, d\nu. \tag{14}$$

2. χ^2 *distance* $D_\chi(\nu_+, \nu_-) := \frac{1}{2}\{\int (f-g)^2 \, d\nu\}^{1/2}$ represents the L^2 norm of the conditional expectation $\mathbf{E}(\xi | X) = \frac{1}{2}(f(X) - g(X))$.
3. *Mutual information* between ξ and X,

$$D_I(\nu_+, \nu_-) := I(\xi; X) = \frac{1}{2}\int (f \log f + g \log g) \, d\nu \tag{15}$$

is a symmetrized version of the Kullback–Leibler divergence (see [37]).

4. *The Hellinger distance*,

$$D_H(\nu_+, \nu_-) := \int \left(\sqrt{f} - \sqrt{g}\right)^2 d\nu = 2\left(1 - \int \sqrt{fg} \, d\nu\right), \tag{16}$$

derives its importance from the simple behavior of the Hellinger integrals

$$\mathrm{Int}_H(\nu_+, \nu_-) := \int \sqrt{fg} \, d\nu$$

for product measures

$$\mathrm{Int}_H(\nu_+ \times \mu_+, \nu_- \times \mu_-) = \mathrm{Int}_H(\nu_+, \nu_-)\mathrm{Int}_H(\mu_+, \mu_-). \tag{17}$$

These distances appear in different sources under different names and with different normalizations. We collect here some well-known inequalities between them that will be useful below. For more on this topic, see, for example, [22] or [39].

LEMMA 4.2. *With the notation above*:

(i) $D_\chi^2 \leq D_V \leq D_\chi \leq \sqrt{D_H}$.
(ii) $D_\chi^2 \leq D_I \leq 2D_\chi^2$.
(iii) *If* ν_+ *and* ν_- *are measures on* $\mathbb{R}$, *then*

$$\left\{\int x d(\nu_+ - \nu_-)\right\}^2 = \left\{\int x[f(x) - g(x)] \, d\nu\right\}^2 \leq 4\int x^2 \, d\nu D_\chi^2.$$

PROOF. (i) The left-hand inequality follows from $|f(x) - g(x)| \leq 2$, and the middle inequality from Cauchy–Schwarz. The right-hand inequality follows from the identity $f - g = (\sqrt{f} - \sqrt{g})(\sqrt{f} + \sqrt{g})$ and the concavity relation $(\sqrt{f} + \sqrt{g})/2 \leq \sqrt{(f+g)/2} = 1$.

(ii) Setting $\psi = (f-g)/2$, the assertion follows from the pointwise inequalities

$$\frac{\psi^2}{2} \leq \frac{1+\psi}{2}\log(1+\psi) + \frac{1-\psi}{2}\log(1-\psi) \leq \psi^2. \tag{18}$$

Here the left-hand inequality is verified for $\psi \in [0, 1)$ by comparing second derivatives, and the right-hand inequality follows from $\log(1 + y) \le y$.

(iii) This is just the Cauchy–Schwarz inequality. $\square$

Finally, we interpret the data processing inequality in terms of distances. Suppose that we are given transition probabilities on the state space, that is a stochastic matrix M (the entries of M are nonnegative and the row sums are all 1). Write $M^*\mu(y) := \sum_x M(x, y)\mu(x)$. Then Lemma 4.1(ii) implies that

$$D_I(M^*\nu_+, M^*\nu_-) \le D_I(\nu_+, \nu_-).$$

An analogous inequality holds for total variation,

$$\begin{aligned} D_V(M^*\nu_+, M^*\nu_-) &= \tfrac{1}{2}\sum_y \bigl|M^*\nu_+(y) - M^*\nu_-(y)\bigr| \\ &\le \tfrac{1}{2}\sum_y\sum_x M(x, y)\bigl|\nu_+(x) - \nu_-(x)\bigr| \qquad (19) \\ &= \tfrac{1}{2}\sum_x \bigl|\nu_+(x) - \nu_-(x)\bigr| = D_V(\nu_+, \nu_-). \end{aligned}$$

5. Conductance lower bounds: Proof of Theorem 1.2. We consider a more general model, where the switching probabilities ε_e vary from edge to edge. Let σ_ρ be a uniform spin, and take independent random variables $\{\eta_e\}$ with $\mathbf{P}[\eta_e = -1] = \varepsilon_e = 1 - \mathbf{P}[\eta_e = 1]$. For any vertex $v \ne \rho$, define

$$\sigma_v := \sigma_\rho \prod_{e \in \mathrm{path}\,(\rho, v)} \eta_e, \tag{20}$$

where path(v) is the path from ρ to v. [More generally let path(u, v) be the path from u to v.] In the language of the Ising model, this corresponds to using varying interaction strengths $\{J_e\}$ as in [23], so that (10) is replaced by $\mathscr{G}(\sigma) = Z(t)^{-1}\exp(\sum_{u\sim v} J_{uv}\sigma_u\sigma_v/t)$, and the conversion equation (11) holds separately on each edge.

For each edge e, define $\theta_e := \mathbf{E}[\eta_e] = 1 - 2\varepsilon_e$. As before, there is also an equivalent construction of the random field $\{\sigma_v\}$ based on percolation: perform independent bond percolation on T where the edge e is open with probability θ_e, and assign independent uniform spins to the open percolation clusters.

Next, assign to each edge e the resistance

$$R(e) := (1 - \theta_e^2) \prod_{f \in \mathrm{path}\,(\rho, e)} \theta_f^{-2}, \tag{21}$$

where path(ρ, e) is the path from ρ to e (inclusive). Also, for each vertex v in T, let

$$\Theta_v := \prod_{e \in \mathrm{path}\,(v)} \theta_e.$$

Say that a set of vertices W is an *antichain* if no vertex in W is a descendant of another.

LEMMA 5.1. *Let W be a finite antichain in T. For any unit flow μ from ρ to W, the weighted sum*

$$S_\mu := \sum_{v \in W} \frac{\mu(v)\sigma_v}{\Theta_v} \tag{22}$$

satisfies $\mathbf{E}[S_\mu|\sigma_\rho] = \sigma_\rho$ *and*

$$\mathbf{E}\big[S_\mu^2\big] = \mathbf{E}\big[S_\mu^2|\sigma_\rho\big] = 1 + \sum_e R(e)\mu(e)^2. \tag{23}$$

Consequently,

$$\min_\mu \mathbf{E}\big[S_\mu^2\big] = 1 + \mathscr{R}_{\mathrm{eff}}(\rho \leftrightarrow W), \tag{24}$$

and the minimum is attained precisely when μ is the unit current flow from ρ to W.

PROOF. From the product representation (20), we infer that

$$\mathbf{E}[\sigma_v|\sigma_\rho] = \sigma_\rho \Theta_v$$

for any vertex v. The formula for $\mathbf{E}[S_\mu|\sigma_\rho]$ follows by linearity. Similarly, for any two vertices v, w in T,

$$\mathbf{E}[\sigma_v \sigma_w] = \prod_{e \in \mathrm{path}\,(v,w)} \theta_e = \frac{\Theta_v \Theta_w}{\Theta_{v\wedge w}^2}, \tag{25}$$

where $v \wedge w$, the *meeting point* of v and w, is the vertex farthest from the root ρ on $\mathrm{path}(v) \cap \mathrm{path}(w)$. The percolation representation can also be invoked to justify (25).

It is now easy to determine the second moment of S_μ:

$$\mathbf{E}\big[S_\mu^2\big] = \sum_{v,\, w \in W} \frac{\mu(v)\mu(w)}{\Theta_v \Theta_w} \mathbf{E}[\sigma_v \sigma_w] = \sum_{v,\, w \in W} \frac{\mu(v)\mu(w)}{\Theta_{v\wedge w}^2}. \tag{26}$$

Next, insert the identity,

$$\frac{1}{\Theta_u^2} = 1 + \sum_{e \in \mathrm{path}(u)} R(e)$$

with $u = v \wedge w$, into (26). Changing the order of summation, and using the fact that W is an antichain, we obtain

$$\mathbf{E}\big[S_\mu^2\big] = 1 + \sum_e R(e) \sum_{v,\, w \in W} \mathbf{1}_{\{e \in \mathrm{path}(v\wedge w)\}} \mu(v)\mu(w). \tag{27}$$

Since $\mathrm{path}(v \wedge w) = \mathrm{path}(v) \cap \mathrm{path}(w)$ and

$$\begin{aligned}\sum_{v,\, w \in W} \mathbf{1}_{\{e \in \mathrm{path}(v\wedge w)\}} \mu(v)\mu(w) &= \Big(\sum_{v \in W} \mathbf{1}_{\{e \in \mathrm{path}(v)\}} \mu(v)\Big)\Big(\sum_{w \in W} \mathbf{1}_{\{e \in \mathrm{path}(w)\}} \mu(w)\Big) \\ &= \mu(e)^2,\end{aligned}$$

(27) is equivalent to (23). Finally, (24) follows from Thompson's principle; see [7] or [26].

For convenience, we restate Theorem 1.2 in its extended form.

THEOREM 1.2′. *With the notation above,*

$$\left.\begin{array}{l}\Delta(T, W, \{\varepsilon_e\}) \\ I(\sigma_\rho; \sigma_W)\end{array}\right\} \geq \frac{1}{1 + \mathscr{R}_{\text{eff}}(\rho \leftrightarrow W)}. \tag{28}$$

PROOF. We may assume that W is an antichain. (Otherwise, remove from W all vertices which have an ancestor in W.) Let μ be the unit current flow from ρ to W for the resistances $R(e)$ as in the preceding lemma, and let S_μ be the weighted sum (22). In order to apply Lemma 4.2, denote by ν_+ the conditional distribution of S_μ given that $\sigma_\rho = 1$; define ν_- analogously by conditioning that $\sigma_\rho = -1$, so that $\nu = (\nu_+ + \nu_-)/2$ is the unconditioned distribution of S_μ. We then have by Lemma 4.2(iii) that

$$D_\chi^2(\nu_+, \nu_-) \geq \frac{\{\int x\, d(\nu_+ - \nu_-)\}^2}{4\int x^2\, d\nu} = \frac{(\mathbf{E}[S_\mu | \sigma_\rho = 1] - \mathbf{E}[S_\mu | \sigma_\rho = -1])^2}{4\mathbf{E}[S_\mu^2]}.$$

Applying Lemma 5.1, we deduce that

$$D_\chi^2(\nu_+, \nu_-) \geq \frac{1}{1 + \mathscr{R}_{\text{eff}}(\rho \leftrightarrow W)}. \tag{29}$$

By Lemma 4.2, the difference $\Delta = \Delta(T, W, \{\varepsilon_e\})$ between the probabilities of correct and incorrect reconstruction, satisfies $\Delta = D_V(\nu_+, \nu_-) \geq D_\chi^2(\nu_+, \nu_-)$, and the mutual information between σ_ρ and σ_W also satisfies $I(\sigma_\rho; \sigma_W) = D_I(\nu_+, \nu_-) \geq D_\chi^2(\nu_+, \nu_-)$. In conjunction with (29), this completes the proof. □

6. Mincut upper bound: Proof of Theorem 1.3. A (communication) *channel* is a stochastic matrix describing the conditional distribution $\mathbf{P}(Y|X)$ of the output variable Y given the input X; see [6]. Often a channel is realized by a relation of the form $Y = f(X, Z)$, where f is a function and Z is a random variable (representing the "noise") which is independent of X. A *noisy tree* is a tree with flip probabilities labeling the edges. The *stringy tree* $\widehat{T}$ associated with a finite noisy tree T is the tree which has the same set of root-leaf paths as T but in which these paths act as independent channels. More precisely, for every root-leaf path in T, there corresponds an identical (in terms of length and flip probabilities on the edges) root-leaf path in $\widehat{T}$, and furthermore, all the root-leaf paths in $\widehat{T}$ are edge disjoint (see Figure 2 in the Introduction).

THEOREM 6.1. *Given a finite noisy tree T with leaves W, let $\widehat{T}$, with leaves $\widehat{W}$ and root $\hat{\rho}$, be the stringy tree associated with T. There is a channel which, for $\xi \in \{\pm 1\}$, transforms the conditional distribution $\sigma_{\widehat{W}}|(\sigma_{\hat{\rho}} = \xi)$ into the conditional distribution $\sigma_W|(\sigma_\rho = \xi)$. Equivalently, we say that $\widehat{T}$ dominates T.*

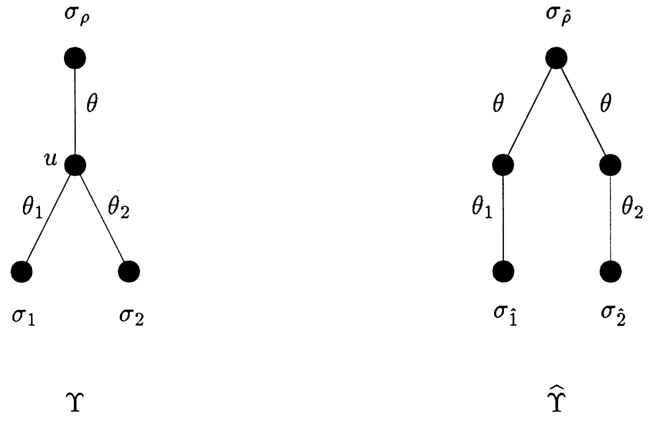

FIG. 4. *Υ is dominated by $\widehat{\Upsilon}$.*

PROOF. We start by establishing a key special case of the theorem: namely, that the tree Υ shown in Figure 4, is dominated by the corresponding stringy tree $\widehat{\Upsilon}$. The root of Υ has just one child u, and the edge leading to it has flip probability $(1-\theta)/2$. The vertex u has two children: on the edge leading to the left child the flip probability is $(1-\theta_1)/2 \le 1/2$, and on the edge leading to the right child the flip probability is $(1-\theta_2)/2 \le 1/2$.

The degenerate case $\theta_1 = \theta_2 = 0$ is handled by the identity channel, and excluded in the sequel. Assume w.l.o.g. that $\theta_2 \le \theta_1$ and let z be a ± 1 valued random variable (independent of the spins on $\widehat{\Upsilon}$) with mean θ_2/θ_1. Given $0 \le \alpha \le 1$, to be specified below, we define the channel as follows:

$$\sigma_1^* = \sigma_{\hat 1},$$

$$\sigma_2^* = \begin{cases} \sigma_{\hat 2}, & \text{with probability } \alpha, \\ \sigma_{\hat 1} z, & \text{with probability } 1-\alpha. \end{cases}$$

To prove that $(\sigma_{\hat\rho}, \sigma_1^*, \sigma_2^*)$ has the same distribution as $(\sigma_\rho, \sigma_1, \sigma_2)$, it suffices to show that the means of corresponding products are equal. (The sufficiency is easy to establish directly and is a special case of the fact that the characters on

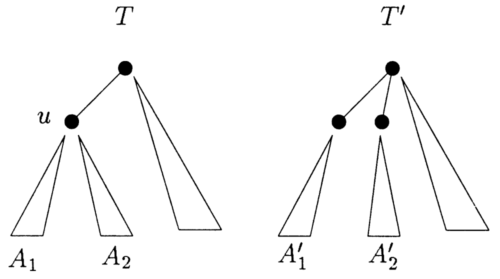

FIG. 5. *Basic step in proof of Theorem* 6.1.

a finite Abelian group G form a basis for the vector space of complex functions on G.) By symmetry,

$$\begin{aligned}\mathbf{E}(\sigma_\rho) = \mathbf{E}(\sigma_1) = \mathbf{E}(\sigma_2) = \mathbf{E}(\sigma_\rho \sigma_1 \sigma_2) = \mathbf{E}(\sigma_{\hat\rho}) \\ = \mathbf{E}(\sigma_1^*) = \mathbf{E}(\sigma_2^*) = \mathbf{E}(\sigma_{\hat\rho}\sigma_1^*\sigma_2^*) = 0\end{aligned}$$

and thus we only need to check pair correlations. First note that $E(\sigma_\rho\sigma_1) = \theta\theta_1$, $E(\sigma_\rho\sigma_2) = \theta\theta_2$ and $E(\sigma_1\sigma_2) = \theta_1\theta_2$. Now clearly, $\mathbf{E}(\sigma_{\hat\rho}\sigma_1^*) = \mathbf{E}(\sigma_\rho\sigma_1)$. Our choice of z guarantees that $\mathbf{E}(\sigma_{\hat\rho}\sigma_{\hat 1} z) = \mathbf{E}(\sigma_{\hat\rho}\sigma_{\hat 1})\theta_2/\theta_1 = \theta\theta_2$, whence $\mathbf{E}(\sigma_{\hat\rho}\sigma_2^*) = \theta\theta_2 = \mathbf{E}(\sigma_\rho\sigma_2)$ for any choice of α. Finally, since $\mathbf{E}(\sigma_1^*\sigma_{\hat 2}) = \theta_1\theta^2\theta_2 \le \theta_1\theta_2 = \mathbf{E}(\sigma_1\sigma_2)$ and

$$\mathbf{E}(\sigma_1^*\sigma_{\hat 1} z) = \mathbf{E}(z) = \theta_2/\theta_1 \ge \theta_1\theta_2,$$

we can choose $\alpha \in [0,1]$ so that $\mathbf{E}(\sigma_1^*\sigma_2^*) = \theta_1\theta_2 = \mathbf{E}(\sigma_1\sigma_2)$; explicitly,

$$\alpha = (1-\theta_1^2)/(1-\theta^2\theta_1^2). \tag{30}$$

This proves that $\widehat{\Upsilon}$ dominates Υ.

Now we consider the general case. Fix a finite noisy tree T with leaves W. We construct a sequence of channels whose composition is the desired channel. The intermediate stages in this composition, when applied to the law of $\sigma_{\widehat W}$, yield the distributions of the spins on the leaves of the intermediate trees (which are more "stringy" than T and less so than $\widehat T$). We describe the last channel in the sequence; this channel transforms the conditional distribution of the spins on the leaves of a slightly more stringy tree T' (where $\widehat{T'} = \widehat T$) to the conditional distribution of the spins on the leaves of T. The theorem then follows by induction.

Let u be a vertex with more than one child which is closest to the root in T. Let A_1 be the leaves of one child's subtree and A_2 the leaves of the other children's subtrees. Let T' be the tree which is identical to T except that T' replaces the path from the root to u with two independent paths from the root to two copies of u: one the parent of the child whose subtree has leaves A_1 and the other the parent of the children whose subtrees have (collectively) leaves A_2. Let A_1' and A_2' represent these sets of leaves in T'.

We describe a channel M which takes the distribution (given $\sigma_{\rho'} = \xi$) of $(\sigma_{A_1'}, \sigma_{A_2'})$ and transforms it into the distribution (given $\sigma_\rho = \xi$) of $(\sigma_{A_1}, \sigma_{A_2})$ (the other leaf values are unchanged).

Let $\varepsilon = (1-\theta)/2$ denote the probability that $\sigma_u \ne \xi$. Let p_s (respectively q_s) denote the distribution of σ_{A_1} (respectively of σ_{A_2}) given that $\sigma_u = s$. Since σ_{A_1} and σ_{A_2} are independent given σ_u, we have

$$\begin{aligned}\mathbf{P}(\sigma_{A_1} = a_1, \sigma_{A_2} = a_2 \mid \sigma_\rho = -1) &= (1-\varepsilon)p_-(a_1)q_-(a_2) + \varepsilon p+(a_1)q+(a_2),\\ \mathbf{P}(\sigma_{A_1'} = a_1, \sigma_{A_2'} = a_2 \mid \sigma_{\rho'} = -1) &= \Big((1-\varepsilon)p_-(a_1) + \varepsilon p_+(a_1)\Big)\\ &\quad\times\Big((1-\varepsilon)q_-(a_2) + \varepsilon q_+(a_2)\Big).\end{aligned}$$

Since these probabilities are typically not equal, the channel must change the spins (a_1, a_2) obtained at the leaves of T' into different spins with some probability.

An important observation, easily verified, is that, for a given spin ξ and a given spin configuration (a_1, a_2) at the leaves, the probability given $\sigma_\rho = \xi$ of the event "$\sigma_{A_1} \in \{a_1, -a_1\}$ and $\sigma_{A_2} \in \{a_2, -a_2\}$" is equal to the probability given $\sigma_{\rho'} = \xi$ of the event "$\sigma_{A_1'} \in \{a_1, -a_1\}$ and $\sigma_{A_2'} \in \{a_2, -a_2\}$." The channel M will be composed of (typically many) channels $M_{(\pm a_1, \pm a_2)}$, one for each set of four configurations $\{(\pm a_1, \pm a_2)\} = \{(a_1, a_2), (a_1, -a_2), (-a_1, a_2), (-a_1, -a_2)\}$. The channel $M_{(\pm a_1, \pm a_2)}$ transforms the distribution given $\sigma_{\rho'} = \xi$ of $(\sigma_{A_1'}, \sigma_{A_2'})$ on $\{(\pm a_1, \pm a_2)\}$, into the distribution given $\sigma_\rho = \xi$ of $(\sigma_{A_1}, \sigma_{A_2})$ on $\{(\pm a_1, \pm a_2)\}$. It mimics the channel showing that $\widehat{\Upsilon}$ dominates the tree Υ (Figure 4).

For a pair a_1, a_2 define $\theta_1 := 1 - 2(p_+(a_1)/p_+(a_1) + p_+(-a_1))$ and $\theta_2 := 1 - 2(q_+(a_2)/q_+(a_2) + q_+(-a_2))$. We may assume that $(1 - \theta_1)/2 \le 1/2$, $(1 - \theta_2)/2 \le 1/2$ and $\theta_2 \le \theta_1$ (renaming variables if necessary). With these values of θ_1 and θ_2, let $\sigma_{\hat{1}}, \sigma_{\hat{2}}, \sigma_1$, and σ_2 be the spins at the leaves of the trees $\widehat{\Upsilon}$ and Υ as shown in Figure 4. Now, given that $\sigma_{A_1'}, \sigma_{A_1} \in \{\pm a_1\}$ and $\sigma_{A_2'}, \sigma_{A_2} \in \{\pm a_2\}$, the key is that $(\sigma_{\rho'}, \sigma_{A_1'}, \sigma_{A_2'})$ and $(\sigma_\rho, \sigma_{A_1}, \sigma_{A_2})$ have the same distribution as $(\sigma_{\hat{\rho}}, \sigma_{\hat{1}} a_1, \sigma_{\hat{2}} a_2)$ and $(\sigma_\rho, \sigma_1 a_1, \sigma_2 a_2)$. The channel constructed above, which demonstrated domination of Υ by $\widehat{\Upsilon}$, therefore serves as the channel $M_{(\pm a_1, \pm a_2)}$ which converts the distribution on $(\sigma_{A_1'}, \sigma_{A_2'})$ into the distribution on $(\sigma_{A_1}, \sigma_{A_2})$, conditional on $\sigma_{A_1'}, \sigma_{A_1} \in \{\pm a_1\}$ and $\sigma_{A_2'}, \sigma_{A_2} \in \{\pm a_2\}$. Combining all such channels $M_{(\pm a_1, \pm a_2)}$ results in a channel which converts the distribution on $(\sigma_{\rho'}, \sigma_{A_1'}, \sigma_{A_2'})$ into the distribution on $(\sigma_\rho, \sigma_{A_1}, \sigma_{A_2})$.

(Note that the argument did not rely on u being closest to the root; the same argument would work for any u, with the spin of u's parent in the role of σ_ρ.) $\square$

We will establish Theorem 1.3 in the more general setting described in the previous section, and use the notation introduced there. In particular, recall that $\Theta_v := \prod_{e \in \mathrm{path}(v)} \theta_e$. Thus, we will prove the following extension.

THEOREM 1.3$'$. *Let W be a finite set of vertices in the tree T. For any set of vertices W_1 in a tree T that separates the root ρ from a finite set of vertices W, we have*

$$\Delta(T, W, \varepsilon)^2 \le 2\left(1 - \prod_{v \in W_1} \sqrt{1 - \Theta_v^2}\right) \le 2 \sum_{v \in W_1} \Theta_v^2 \tag{31}$$

and

$$I(\sigma_\rho; \sigma_W) \le \sum_{v \in W_1} I(\sigma_\rho; \sigma_v) \le \sum_{v \in W_1} \Theta_v^2. \tag{32}$$

PROOF. We first prove (32). Since W_1 separates ρ from W, the data processing inequality [Lemma 4.1(ii)] yields $I(\sigma_\rho; \sigma_W) \le I(\sigma_\rho; \sigma_{W_1})$. Let T_1 be the tree obtained from T by retaining only W_1 and ancestors of nodes in W_1. Let

$\widehat{T_1}$ be the stringy tree associated with T_1. From Theorem 6.1 applied to T_1 and the data processing inequality, we obtain $I(\sigma_\rho;\sigma_{W_1}) \le I(\sigma_{\hat\rho};\sigma_{\widehat{W}_1})$. Since the spins on leaves of $\widehat{T_1}$ are conditionally independent given $\sigma_{\hat\rho}$, subadditivity [Lemma 4.1(iii)] gives

$$I(\sigma_{\hat\rho};\sigma_{\widehat{W}_1}) \le \sum_{\hat v \in \widehat{W}_1} I(\sigma_{\hat\rho};\sigma_{\hat v}).$$

However because of the definition of the stringy tree, the mutual information between $\sigma_{\hat\rho}$ and $\sigma_{\hat v}$ is identical to the mutual information between σ_ρ and σ_v in T_1, hence the left inequality in (32).

Since $\mathbf{E}(\sigma_\rho\sigma_v) = \Theta_v$ for each v, the right-hand inequality in (32) follows from the right-hand inequality in (18).

We now turn to the total variation inequality (31). Recall that $\Delta(T, W, \varepsilon)$, the difference between the probabilities of correct and incorrect reconstruction, equals $D_V(\nu_+^W, \nu_-^W)$, the total variation distance between the two distributions of the spins on W given $\sigma_\rho = \pm 1$.

By (19), Theorem 6.1, and Lemma 4.2,

$$D_V(\nu_+^W,\nu_-^W) \le D_V(\nu_+^{W_1},\nu_-^{W_1}) \le D_V(\nu_+^{\widehat{W}_1},\nu_-^{\widehat{W}_1}) \le \sqrt{D_H(\nu_+^{\widehat{W}_1},\nu_-^{\widehat{W}_1})}.$$

Now, $D_H(\nu_+^{\widehat{W}_1},\nu_-^{\widehat{W}_1})$ on the stringy tree $\widehat{T_1}$ is easily calculated using the multiplicative property of Hellinger integrals: $\nu_+^{\widehat{W}_1}$ is just the product over $w \in \widehat{W}_1$ of ν_+^w, the distribution of σ_w given $\sigma_\rho = 1$, and similarly $\nu_-^{\widehat{W}_1} = \Pi_w \nu_-^w$. Since $\mathrm{Int}_H(\nu_+^w,\nu_-^w) = \sqrt{1-\Theta_w^2}$, the left-hand inequality in (31) follows; the right-hand inequality there is a consequence of the standard inequality $\Pi(1-x_j) \ge 1 - \sum x_j$. $\square$

7. The correlation inequality: Proof of Theorem 1.4. Define $\mathbf{P}_+(\cdot) := \mathbf{P}(\cdot \mid \sigma_\rho = 1)$. The statement we need to prove is

$$\forall y > 0 \qquad \mathbf{P}_+(S_n = y) \ge \mathbf{P}_+(S_n = -y). \tag{33}$$

We show this by induction on n. The base case $n = 0$ is clear.

Rewriting the probabilities in (33) by conditioning on S_{n-1}, it suffices to prove that for all $x \ge 0$,

$$\begin{aligned}
&\mathbf{P}(S_n = y \mid S_{n-1} = x)\mathbf{P}_+(S_{n-1} = x)\\
&\quad + \mathbf{P}(S_n = y \mid S_{n-1} = -x)\mathbf{P}_+(S_{n-1} = -x)\\
&\qquad \ge \mathbf{P}(S_n = -y \mid S_{n-1} = x)\mathbf{P}_+(S_{n-1} = x)\\
&\quad + \mathbf{P}(S_n = -y \mid S_{n-1} = -x)\mathbf{P}_+(S_{n-1} = -x).
\end{aligned}$$

[We used the spherical symmetry of T to replace $\mathbf{P}_+(\cdot \mid S_{n-1} = x)$ by $\mathbf{P}(\cdot \mid S_{n-1} = x)$.] By symmetry, $\mathbf{P}(S_n = y \mid S_{n-1} = z) = \mathbf{P}(S_n = -y \mid S_{n-1} = -z)$

for any y, z. Thus it suffices to prove that

$$[\mathbf{P}(S_n = y \mid S_{n-1} = x) - \mathbf{P}(S_n = -y \mid S_{n-1} = -x)] \\ \times [\mathbf{P}_+(S_{n-1} = x) - \mathbf{P}_+(S_{n-1} = -x)] \geq 0.$$

For $x \geq 0$, the first term in this product is nonnegative by Lemma 7.1 below, and the second term is nonnegative by induction, hence the proof. □

LEMMA 7.1. *Define a coin to be a* ± 1 *valued random variable. Let* $m, n \geq 0$ *and let* S *denote the sum of* $2n + m$ *random independent coins, of which* $n + m$ *have mean* $\theta \geq 0$ *and* n *have mean* $-\theta$. *Then for any* $y \geq 0$, *we have*

$$\mathbf{P}(S = y) \geq \mathbf{P}(S = -y).$$

PROOF. Start by considering a random variable $S^*_{l,k} = l + S^*_{0,k}$ where $l \geq 0$ and $S^*_{0,k}$ is a sum of k independent fair coins. In this case the inequality $\mathbf{P}(S^*_{l,k} = y) \geq \mathbf{P}(S^*_{l,k} = -y)$ for $y \geq 0$ is immediate from unimodality of the binomial coefficients. We will prove the lemma by showing the law of S is a mixture (i.e., a convex combination) of laws of such variables $S^*_{l,k}$.

Write $S = X + Y$, where $X = X_1 + \cdots + X_m$ is the sum of m coins of mean θ, and $Y = Y_1 + \cdots + Y_n$, with each Y_j being the sum of a coin of mean θ and a coin of mean $-\theta$ (all these coins are independent).

Each random variable X_i or Y_i can be represented as follows:

$$X_i = \begin{cases} \text{a fair coin,} & \text{with prob. } 1 - \theta, \\ +1, & \text{with prob. } \theta, \end{cases}$$

$$Y_i = \begin{cases} \text{sum of 2 fair coins,} & \text{with prob. } 1 - \theta^2, \\ 0, & \text{with prob. } \theta^2. \end{cases}$$

The second identity in law follows from

$$\left(\frac{1-\theta^2}{4}, \frac{1+\theta^2}{2}, \frac{1-\theta^2}{4}\right) = (1-\theta^2)\left(\frac{1}{4}, \frac{1}{2}, \frac{1}{4}\right) + \theta^2\Big(0, 1, 0\Big).$$

Consequently, the law of S (denoted $\mathscr{L}(S)$) is a convex combination of laws of variables $S^*_{l,k}$; explicitly,

$$\mathscr{L}(S) = \sum_{a=0}^{m}\sum_{b=0}^{n}\binom{m}{a}\binom{n}{b}(1-\theta)^a\theta^{m-a}(1-\theta^2)^b\theta^{2(n-b)}\mathscr{L}(S^*_{m-a,\,a+2b}).$$

This completes the proof. □

8. A comparison of reconstruction algorithms. Theorem 1.1 and our proof of Theorem 1.2 imply that on infinite spherically symmetric trees, reconstruction by global majority vote has the same threshold for success as maximum likelihood reconstruction. (On general trees, the same applies to weighted majority.) In this section we compare these methods to other methods currently in use: parsimony and certain recursive algorithms. Although maximum-likelihood reconstruction has the smallest probability of error among all reconstruction algorithms, alternative methods have certain advantages, such as robustness (precise knowledge of the flip probability ε is not required), lower complexity, and ease of analysis.

We start by reviewing parsimony. Given a bicoloring of the boundary of a tree T, a *Parsimonious* coloring of the internal nodes is any assignment of the two colors to these nodes that minimizes the total number of bicolored edges. There may be several parsimonious colorings; the following recursive procedure (equivalent to "Fitch's algorithm," cf. [12], [17]) determines whether they all assign the same color to the root. Suppose that the boundary nodes are colored $\{+1, -1\}$. Starting from the parents of the boundary nodes, assign recursively to each internal node the color of the majority of its ± 1-colored children. In case of a tie, assign the noncolor "?."

This procedure assigns the root a value in $\{+1, -1\}$ if, and only if, all parsimonious colorings assign the root that value.

On a fixed finite tree, parsimony coincides with maximum likelihood reconstruction if the error probability ε is small enough. (Hence in this setting, it is superior to majority vote.) However, for larger ε, parsimony can perform significantly worse than maximum likelihood and majority vote.

To illustrate this, suppose that a binary tree of depth n is bicolored using a mutation ($\equiv$ flip) probability ε. Denote by $\Psi(n, \varepsilon)$ the difference between the probability that parsimony will reconstruct the correct root color and the probability that the opposite color will be reconstructed, given the colors at level n. Steel [35] showed that $\inf_n \Psi(n, \varepsilon) > 0$ if $\varepsilon < 1/8$, but $\Psi(n, \varepsilon) \to 0$ as $n \to \infty$ if $\varepsilon \geq 1/8$. Thus, for flip rates ε such that $1 - 2\varepsilon \in (2^{-1/2}, 3/4]$ on the binary tree, majority vote will detect the dependence between the boundary and root colors, while parsimony will miss it.

On a tree where each vertex has k children with k odd, parsimony reduces to recursive majority; [28] has shown that reconstruction via this method succeeds asymptotically if and only if

$$\varepsilon < \beta_k := \frac{1}{2} - \frac{2^k}{4k} \binom{k-1}{\frac{k-1}{2}}^{-1}.$$

In fact, this result can be deduced from earlier arguments concerning noisy computation, [16], (for $k = 3$) and [8], [Theorem 5.0.3] (for larger odd k). It is shown in [16], [8] that β_k is the noise threshold for reliable computation of all Boolean functions by noisy formulas with k-input gates. This is a different setting (in the formula case noise affects computation of the recursive majority of the inputs, as contrasted with noiseless computation of the recursive majority of bits generated by a noisy broadcast process), but the recursions that occur in the proofs are precisely the same.

Mossel [28] also analyzed reconstruction algorithms on regular trees (and more generally, on l-periodic trees) that in order to determine the color assigned to a node v, are allowed to examine the colors of its descendants l generations down. (However, only a single bit can be stored at each node). He showed that among these algorithms, recursively applying majority vote of the descendants l generations down is optimal, yet it succeeds asymptotically only for flip probabilities ε below a threshold which is strictly lower than ε_c.

We stress that the comparisons above apply only to randomly *bicolored* trees, where the colors propagate via binary symmetric channels. Mossel [29] recently showed that the behavior of the algorithms discussed above may be quite different when the number of colors exceeds two, as well as for asymmetric binary channels.

9. Concluding remarks and unsolved problems. *Thresholds for recursive reconstruction algorithms.* The results of Steel, Hajek and Weller, Evans and Mossel quoted in Section 8, lead us to conjecture that on a randomly bicolored regular b-tree, *any* recursive reconstruction algorithm A that stores only a bounded number of bits at each node, must fail asymptotically above a certain noise threshold $\varepsilon_c(A)$ which is *strictly* lower than $\varepsilon_c = (1 - b^{-1/2})/2$.

Note that on a regular tree of depth n, global majority can be computed recursively provided $O(n)$ bits can be stored at each node. Bayesian reconstruction can also be implemented recursively, but requires arithmetic over the real numbers; it would be worth examining how roundoff error in this recursive calculation would affect the reconstruction threshold.

Reconstruction at criticality. It is shown in [2] and [19] that on infinite regular trees, $\lim_n \Delta(T, T_n, \varepsilon_c) = 0$. On general trees, Theorem 1.2 implies that finite effective resistance from the root to infinity (when each edge at level l is assigned the resistance $(1-2\varepsilon)^{-2l}$) is sufficient for $\lim_n \Delta(T, T_n, \varepsilon) > 0$. In [31], a recursive method is used to show this condition is also necessary.

Multicolored trees and the Potts model. The most natural generalization of the two-state tree-indexed Markov chain model studied in this paper involves multicolored trees, where the coloring propagates according to any finite state tree-indexed Markov chain. For instance, if this Markov chain is defined by a $q \times q$ stochastic matrix where all entries off the main diagonal equal ε, then the q-state Potts model arises. Our proof of Theorem 1.2 extends to general Markov chains, and shows that the tail of the tree-indexed chain is nontrivial if $\mathrm{br}(T) > \lambda_2^{-2}$, where λ_2 is the second eigenvalue of the transition matrix (e.g., for the q-state Potts model, $\lambda_2 = 1 - q\varepsilon$). However, calculations in [29] indicate that this lower bound is not sharp in general. Furthermore, we do not know a reasonable upper bound on mutual information between root and boundary variables. In particular, it seems that the critical parameter for tail triviality in the Potts model on a regular tree is not known. A detailed analysis of tree-indexed Markov chains arising from Hardcore restrictions was recently made in [3]; determining the critical parameters for tail triviality appears to be open in that setting as well.

An information inequality. Theorem 1.3 implies that the spins in the ferromagnetic Ising model on a tree satisfy

$$I(\sigma_v;\sigma_W) \le \sum_{w\in W} I(\sigma_v;\sigma_w),$$

for any vertex v and any finite set of vertices W. Does this inequality hold in other graphs as well?

More generally, are there natural assumptions on random variables X, $Y_1,\ldots,Y_n$ that imply the inequality $I(X;(Y_1,\ldots,Y_n)) \le \sum_{j=1}^{n} I(X,Y_j)$?

Conditions for an intermediate phase. On which infinite graphs Γ is there an interval of temperatures where the Ising model admits multiple Gibbs states, yet the free-boundary limiting Gibbs state is extremal? (Perhaps when Γ has a transitive automorphism group, the relevant property is *nonamenability*.)

Domination. In the statement of Theorem 6.1 we defined a domination relation between trees. A different notion of domination between trees was analyzed in [30]. It would be quite interesting to determine the precise relation of the two notions.

Acknowledgments. We are grateful to J. Chang, O. Häggström, S. Grossman, D. Ioffe, D. Levin, R. Lyons, E. Mossel, R. Pemantle, M. Steel and J. Steif for useful comments on earlier versions of this work and for references to the literature. Part of this work was supported by the Institute for Elementary Studies.

REFERENCES

[1] BENJAMINI, I., PEMANTLE, R. and PERES, Y. (1998). Unpredictable paths and percolation. *Ann. Probab.* **26** 1198–1211.

[2] BLEHER, P. M., RUIZ, J. and ZAGREBNOV, V. A. (1995). On the purity of the limiting Gibbs state for the Ising model on the Bethe lattice. *J. Statist. Phys.* **79** 473–482.

[3] BRIGHTWELL, G. R. and WINKLER, P. (1999). Graph homomorphisms and phase transitions. *J. Combin. Theory Ser. A.* To appear.

[4] CAVENDER, J. (1978). Taxonomy with confidence. *Math. Biosci.* **40** 271–280.

[5] CHAYES, J. T., CHAYES, L., SETHNA, J. P. and THOULESS, D. J. (1986). A mean field spin glass with short range interactions, *Comm. Math. Phys.* **106** 41–89.

[6] COVER, T. M. and THOMAS, J. A. (1991). *Elements of Information Theory*. Wiley, New York.

[7] DOYLE, P. G. and SNELL, E. J. (1984). *Random Walks and Electrical Networks.* Math. Assoc. Amer., Washington, D.C.

[8] EVANS, W. (1994). Information theory and noisy computation. Ph.D. dissertation, Dept. Computer Science, Univ. California, Berkeley.

[9] EVANS, W., KENYON, C., PERES, Y. and SCHULMAN, L. J. (1995). A critical phenomenon in a broadcast process. Unpublished manuscript.

[10] EVANS, W. and SCHULMAN, L. J. (1993). Signal propagation, with application to a lower bound on the depth of noisy formulas. In *Proceedings of the 34th Annual Symposium on Foundations of Computer Science* 594–603.

[11] EVANS, W. and SCHULMAN, L. J. (1994). Lower bound on the depth of noisy circuits. Unpublished manuscript.

[12] FITCH, W. M. (1971). Toward defining the course of evolution: minimum change for a specific tree topology. *Syst. Zool.* **20** 406–416.

[13] GEORGII, H. O. (1988). *Gibbs Measures and Phase Transitions*. de Gruyter, Berlin.

[14] GRIMMETT, G. R. (1996). Percolation and disordered systems. *Lectures in Probability Theory and Statistics, Ecole d'Eté de Probabilités de Saint Flour XXVI*. *Lecture Notes in Math*. **1665** 153–300. Springer, Berlin.

[15] HÄGGSTRÖM, O. and MOSSEL, E. (1998). A nearest neighbor process with low predictability and percolation in $2+\varepsilon$ dimensions. *Ann. Probab*. **26** 1212–1231.

[16] HAJEK, B. and WELLER, T. (1991). On the maximum tolerable noise for reliable computation by formulas. *IEEE Trans. Inform. Theory* **37** 388–391.

[17] HARTIGAN, J. A. (1971). Minimum mutation fits to a given tree. *Biometrics* **29** 53–65.

[18] HIGUCHI, Y. (1977). Remarks on the limiting Gibbs state on a $(d+1)$-tree. *Publ. RIMS Kyoto Univ*. **13** 335–348.

[19] IOFFE, D. (1996). A note on the extremality of the disordered state for the Ising model on the Bethe lattice. *Lett. Math. Phys*. **37** 137–143.

[20] IOFFE, D. (1996). A note on the extremality of the disordered state for the Ising model on the Bethe lattice. In *Trees* (B. Chauvin, S. Cohen, A. Roualt, eds.). Birkhäuser, Boston.

[21] KESTEN, H. and STIGUM, B. P. (1966). Additional limit theorems for indecomposable multidimensional Galton–Watson processes. *Ann. Math. Statist.* **37** 1463–1481.

[22] LE CAM, L. (1974). *Notes on Asymptotic Methods in Statistical Decision Theory*. Centre de Rech. Math., Univ. Montréal.

[23] LYONS, R. (1989). The Ising model and percolation on trees and tree-like graphs. *Comm. Math. Phys*. **125** 337–352.

[24] LYONS, R. (1990). Random walks and percolation on trees. *Ann. Probab*. **18** 931–958.

[25] LYONS, R. (1992). Random walks, capacity, and percolation on trees. *Ann. Probab*. **20** 2043–2088.

[26] LYONS, R. and PERES, Y. (1997). *Probability on Trees and Networks*. To appear. Available at http://php.indiana.edu/~rdlyons/.

[27] MOORE, T. and SNELL, J. L. (1979). A branching process showing a phase transition. *J. Appl. Probab*. **16** 252–260.

[28] MOSSEL, E. (1998). Recursive reconstruction on periodic trees. *Random Structures Algorithms* **13** 81–97.

[29] MOSSEL, E. (1999). Reconstruction on trees: beating the second eigenvalue. Unpublished manuscript.

[30] PEMANTLE, R. and PERES, Y. (1994). Domination between trees and application to an explosion problem. *Ann. Probab*. **22** 180–194.

[31] PEMANTLE, R. and PERES, Y. (1995). Recursions on trees and the Ising model at critical temperatures. Unpublished manuscript.

[32] PIPPENGER, N. (1988). Reliable computation by formulas in the presence of noise. *IEEE Trans. Inform. Theory* **34** 194–197.

[33] PRESTON, C. J. (1974). *Gibbs States on Countable Sets*. Cambridge Univ. Press.

[34] SPITZER, F. (1975). Markov random fields on an infinite tree. *Ann Probab*. **3** 387–394.

[35] STEEL, M. (1989). Distribution in bicolored evolutionary trees. Ph.D. thesis, Massey Univ., Palmerston North, New Zealand.

[36] STEEL, M. and CHARLESTON, M. (1995). Five surprising properties of parsimoniously colored trees. *Bull. Math. Biology* **57** 367–375.

[37] VAJDA, I. (1989). *Theory of Statistical Inference and Information*. Kluwer, Dordrecht.

[38] VON NEUMANN, J. (1956). Probabilistic logics and the synthesis of reliable organisms from unreliable components. In *Automata Studies* (C. E. Shannon and J. McCarthy, eds.) 43–98. Princeton Univ. Press.

W. EVANS
DEPARTMENT OF COMPUTER SCIENCE
UNIVERSITY OF ARIZONA
TUCSON, ARIZONA 85721-0077
E-MAIL: will@cs.arizona.edu

C. KENYON
LRI, UNIVERSITÉ DE PARIS-SUD
ORSAY
FRANCE
E-MAIL: Claire.Kenyon@lri.fr

Y. Peres
Department of Mathematics
Hebrew University, Jerusalem
Israel
and
Department of Statistics
University of California
Berkeley, California 94720-0001
E-mail: peres@math.huji.ac.il

L. J. Schulman
College of Computing
Georgia Institute of Technology
Atlanta, Georgia 30332-0280
E-mail: schulman@cc.gatech.edu