Broadcasting on trees near criticality

Yuzhou Gu, Hajir Roozbehani, and Yury Polyanskiy

***Abstract*—We revisit the problem of broadcasting on d-ary trees: starting from a Bernoulli$(1/2)$ random variable X_0 at a root vertex, each vertex forwards its value across binary symmetric channels BSC_δ to d descendants. The goal is to reconstruct X_0 given the vector X_{L_h} of values of all variables at depth h. It is well known that reconstruction (better than a random guess) is possible as $h \to \infty$ if and only if $\delta < \delta_c(d)$. In this paper, we study the behavior of the mutual information and the probability of error when δ is slightly subcritical. The innovation of our work is application of the recently introduced "less-noisy" channel comparison techniques. For example, we are able to derive the positive part of the phase transition (reconstructability when $\delta < \delta_c$) using purely information-theoretic ideas. This is in contrast with previous derivations, which explicitly analyze distribution of the Hamming weight of X_{L_h} (a so-called Kesten-Stigum bound).**

I. Introduction

We consider the following problem, also known as broadcasting on trees (BOT). Consider an infinite rooted d-ary tree, in which every vertex v has d descendants $v_1, \dots, v_d$. Let L_h denote all vertices at depth h, so that $|L_h| = d^h$. To each vertex v we associate a binary random variable X_v, whose joint distribution is described inductively as follows. The root variable $X_0 \sim \mathrm{Ber}(1/2)$ is an unbiased Bernoulli. Given all random variables X_{L_h} at depth h the variables at depth $h+1$ are generated conditionally independently as follows. If (u, v) is an edge in the tree with $u \in L_h$ and $v \in L_{h+1}$ the (conditioned on X_{L_h}) we set $X_v = X_u$ with probability $(1-\delta)$ and $X_v = 1 - X_u$ otherwise. We define the following quantities[1]:

$$P_e(\delta) = \lim_{h\to\infty} \mathbb{P}[X_0 \neq \hat{X}_0(X_{L_h})],$$

$$\hat{X}_0(y_h) = \operatorname*{argmax}_{a\in\{0,1\}} \mathbb{P}[X_0 = a | X_{L_h} = y_h]\,, \quad (1)$$

$$I(\delta) = \lim_{h\to\infty} I(X_0; X_{L_h})\,. \quad (2)$$

When $P_e < 1/2$ (equivalently, $I > 0$) we say that *reconstruction is possible*. The foundational work [1] established that the reconstruction is possible if and only if

$$\delta < \delta_c \triangleq \frac{1}{2}\left(1 - \frac{1}{\sqrt{d}}\right).$$

We note that the positive part (that $P_e < 1/2$ when $\delta < \delta_c$) follows from a so-called Kesten-Stigum bound, cf. [2], which in fact proves that reconstruction can be done by a sub-optimal detector

$$\hat{X}_{0,maj}(y_h) = 1\{\|y\|_H > d^{h-1}/2\}\,, \quad (3)$$

where $\|y\|_H = |\{j : y_j \neq 0|$ is the Hamming weight. The extension to general (non-regular) trees was done in [2] and beyond trees in [3]. There are deep connections between BOT and other problems. In statistical physics it arises in the study of the free-boundary Gibbs measure for the Ising model on a tree [1], problems on random graphs [4] and in random constraint satisfaction [5]. It was a key step for establishing the sharp thresholds for the problem of community detection (stochastic block model) [6]. It can be seen as a simple model for genetic mutations [7] and noisy computations [8].

We note that various theories (starting from Ginzburg-Landau) in statistical physics predict the type of behavior of various quantities in the vicinity of the phase transition (the so-called *critical exponents*). However, to the best of our knowledge, the behavior of $I(\delta)$ and $P_e(\delta)$ near the critical point $\delta = \delta_c - \tau$ with $\tau \ll 1$ is not understood. In particular, the best results available in the literature show that for some $0 < c_1 < c_2$ and $c_3, c_4 > 0$ we have

$$c_1\tau + o(\tau) \le I(\delta_c - \tau) \le c_2\tau + o(\tau), \quad (4)$$

$$\tfrac{1}{2} - c_3\sqrt{\tau} + o(\sqrt{\tau}) \le P_e(\delta_c - \tau) \le \tfrac{1}{2} - c_4\tau + o(\tau)\,. \quad (5)$$

The main open question is establishing the critical exponent in (5). This can be phrased equivalently as follows: The upper bound in (5) can only be tight if in the regime $\delta \uparrow \delta_c$ the induced (BMS) channel $X_0 \mapsto X_{L_h}$ resembles an erasure channel, whereas the lower bound can only be tight if the channel resembles "more diffuse" BMS channel, akin to binary-input AWGN one. Thus, settling the exponent in (5) can be rephrased as the question of understanding the kind of residual uncertainty about X_0 remaining after observing the leaves X_{L_h}.

Our contributions are as follows:

1) We adapt the channel comparison technique from [9] to the study of I and P_e and show in particular that the positive part can be established without analyzing either the suboptimal decoder (3) (as in Kesten-Stigum) or the belief propagation (as in [1]).
2) We improve available estimates on c_2.
3) We develop a sequence of numerically computable bounds, each provably upper and lower bounding I and P_e, whose evaluation allows us to make the following two conjectures for binary trees.

Conjecture 1. $I(\delta_c - \tau) = \frac{4\sqrt{2}}{\ln 2}\tau + o(\tau)$ *bit.*

Conjecture 2. $P_e(\delta_c - \tau) = 1/2 - \Theta(\sqrt{\tau})$.

A. Channel comparison lemmas

We quickly review the channel comparison lemmas of [9] and discuss how they relate to broadcasting. We start with reviewing some key information-theoretic notions.

The authors are affiliated with the Department of EECS and Laboratory of Decision and Information Sciences (LIDS) at MIT. Email: {yuzhougu,hajir,yp}@mit.edu. This work was supported in part by the National Science Foundation award under grant agreement CCF-17-17842 and by the Center for Science of Information (CSoI), an NSF Science and Technology Center, under grant agreement CCF-09-39370.

[1]Throughout this paper, we use log to denote binary logarithm, and ln to denote natural logarithm. Mutual information I is defined with base 2.

Definition 1 ([10, §5.6]). *Given two channels $P_{Y|X}$ and $P_{Y'|X}$ with common input alphabet, we say that $P_{Y'|X}$ is*

- less noisy *than $P_{Y|X}$, denoted by $P_{Y|X} \preceq_{\mathrm{l.n.}} P_{Y'|X}$, if for all joint distributions P_{UX} we have*
$$I(U;Y) \le I(U;Y')$$
- more capable *than $P_{Y|X}$, denoted by $P_{Y|X} \preceq_{\mathrm{m.c.}} P_{Y'|X}$, if for all marginal distributions P_X we have*
$$I(X;Y) \le I(X;Y').$$
- less degraded *than $P_{Y|X}$, denoted by $P_{Y|X} \preceq_{\mathrm{deg}} P_{Y'|X}$, if there exists a Markov chain $Y - Y' - X$.*

We refer to [11, Sections I.B, II.A] and [12, Section 6] for alternative useful characterizations of the less-noisy order.

For an arbitrary pair of random variables we define
$$I_{\chi^2}(X;Y) = \chi^2(P_{X,Y} \| P_X \otimes P_Y)\,,$$
where $P_X \otimes P_Y$ denotes the joint distribution on (X,Y) under which they are independent.

Let W be a BMS channel (cf. [9], Definition 7), $X \sim \mathrm{Ber}(1/2)$ and $Y = W(X)$ be the output induced by X. We define W's probability of error, capacity, and χ^2-capacity as follows
$$P_e(W) = \frac{1 - \mathrm{TV}(W(\cdot|0), W(\cdot|1))}{2}, \tag{6}$$
$$C(W) = I(X;Y), \tag{7}$$
$$C_{\chi^2}(W) = I_{\chi^2}(X;Y)\,. \tag{8}$$

Lemma 1. *[9, Lemma 2] The following holds:*

1) *Among all* BMS *channels with the same value of $P_e(W)$ the least degraded is* BEC *and the most degraded is* BSC*, i.e.*
$$\mathrm{BSC}_\delta \preceq_{deg} W \preceq_{deg} \mathrm{BEC}_{2\delta}\,, \tag{9}$$
where $\preceq_{deg}$ denotes the (output) degradation order.
2) *Among all* BMS *with the same capacity C the most capable is* BEC *and the least capable is* BSC*, i.e.:*
$$\mathrm{BSC}_{1-h_b^{-1}(C)} \preceq_{mc} W \preceq_{mc} \mathrm{BEC}_{1-C}\,, \tag{10}$$
where $\preceq_{mc}$ denotes the more-capable order, and $h_b^{-1} : [0,1] \to [0,1/2]$ is the functional inverse of the (base-2) binary entropy function $h_b : [0,1/2] \to [0,1]$.
3) *Among all* BMS *channels with the same value of χ^2-capacity $\eta = I_{\chi^2}(W)$ the least noisy is* BEC *and the most noisy is* BSC*, i.e.*
$$\mathrm{BSC}_{1/2-\sqrt{\eta}/2} \preceq_{ln} W \preceq_{ln} \mathrm{BEC}_{1-\eta}\,, \tag{11}$$
where $\preceq_{ln}$ denotes the less-noisy order.

The next lemma states that if the incoming messages to BP are comparable, then the output messages are comparable as well.

Lemma 2. *[9, Lemma 3] Fix some random transformation $P_{Y|X_0,X_1^m}$ and m* BMS *channels $W_1, ..., W_m$. Let $W : X_0 \mapsto (Y, Y_1^m)$ be a (possibly non-*BMS*) channel defined as follows. First, $X_1, ..., X_m$ are generated as i.i.d* $\mathrm{Ber}(1/2)$*. Second, each Y_j is generated as an observation of X_j over the W_j, i.e. $Y_j = W_j(X_j)$ (observations are all conditionally independent given X_1^m). Finally, Y is generated from X_0, X_1^m via $P_{Y|X_0,X_1^m}$ (conditionally independent of Y_1^m given X_1^m). Define the $\tilde{W}$ channel similarly, but with W_j's replaced with $\tilde{W}_j$'s. The following statements hold:*

1) *If $\tilde{W}_j \preceq_{deg} W_j$ then $\tilde{W} \preceq_{deg} W$.*
2) *If $\tilde{W}_j \preceq_{ln} W_j$ then $\tilde{W} \preceq_{ln} W$.*

Remark 1. *An analogous statement for more capable channels does not hold (see Example 2 in [9]).*

Definition 2 (Erasure function). *Consider a single layer of a d-ary tree with source X_0. Suppose that each boundary node is observed through a (memoryless)* BEC *channel, i.e., $Y^{(j)} = \mathrm{BEC}_q(X^{(j)})$ where q is the probability of erasure. The function*
$$E^{\mathrm{BEC}}(q) \triangleq \mathbb{E}[\mathbf{P}(X_0 = 1|Y^{(1)}, \cdots, Y^{(d)})|X_0 = 0].$$
is called the erasure function of the tree. Here the expectation is taken with respect to the randomization over bits as well as the noise in the observations.

Definition 3 (Error function). *In the setup of Definition 2, let $Y^{(j)} = \mathrm{BSC}_q(X^{(j)})$ where q is the crossover probability. The function*
$$E^{\mathrm{BSC}}(q) \triangleq \mathbb{E}[\mathbf{P}(X_0 = 1|Y^{(1)}, \cdots, Y^{(d)})|X_0 = 0].$$
is called the error function of the tree. Here the expectation is taken with respect to the randomization over bits as well as the noise in the observations.

Definition 4 (χ^2-entropies). *Take the setup of Definition 2. Let Y_i's be* BEC *induced observations as before. Define the erasure χ^2-entropy function to be*
$$\mathcal{H}^{\mathrm{BEC}}(q) \triangleq \mathbb{E}[1 - I_{\chi^2}(X_0; Y^{(1)}, \cdots, Y^{(d)})].$$
The corresponding error χ^2-entropy $\mathcal{H}^{\mathrm{BSC}}$ is defined in an analogous manner to Definition 3.

The next proposition shows that the broadcasting problem can be cast into the setting of comparison lemmas.

Proposition 1. *Consider a single layer of a d-ary tree with source X_0 and independent observations $X_i = \mathrm{BSC}_\delta(X_0)$ along the edges. Consider the channels $W : X_0 \mapsto Y_1^d$ with $Y_i = W_i(X_i)$ and $\tilde{W} : X_0 \mapsto \tilde{Y}_1^d$ with $\tilde{Y}_i = \tilde{W}_i(X_i)$. The following statements hold:*

1) *If $\tilde{W}_j \preceq_{deg} W_j$ then $\tilde{W} \preceq_{deg} W$.*
2) *If $\tilde{W}_j \preceq_{ln} W_j$ then $\tilde{W} \preceq_{ln} W$.*

Proof. Let $X' \sim \mathrm{Ber}(1/2)^{\otimes d}$. Define the parity codes $Y_i' = X_0 + X_i'$. Note that the channel $X_0 \to X_i$ is equivalent to $X_0 \to (Y_i', \mathrm{BSC}_\delta(X_i'))$. Likewise, the channels W_i are equivalent to $X_0 \to (Y_i', W_i(\mathrm{BSC}_\delta(X_i')))$. This latter map is of the form in Lemma 2, from which both statements follow. ■

As a consequence we have the following propositions for the broadcasting problem.

Proposition 2. *Consider the dynamical systems*
$$q_{t+1}^{\mathrm{BEC}}(x) = 2E^{\mathrm{BEC}}(q_t^{\mathrm{BEC}}(x)), \tag{12}$$
$$q_{t+1}^{\mathrm{BSC}}(x) = E^{\mathrm{BSC}}(q_t^{\mathrm{BSC}}(x)), \tag{13}$$

initialized at $q_0^{\mathrm{BEC}}(x) = q_0^{\mathrm{BSC}}(x) = x$. Let $P_e(\mathcal{T}_\ell)$ be the probability of error under BP after broadcasting on a d-ary tree of depth ℓ. Then
$$\frac{q_\ell^{\mathrm{BEC}}(0)}{2} \le P_e(\mathcal{T}_\ell) \le q_\ell^{\mathrm{BSC}}(0).$$

Proof. The proof follows from that of [9, Proposition 9] upon replacing Lemma 2 with Proposition 1. ■

Proposition 3. *Consider the dynamical systems*
$$q_{t+1}^{\mathrm{BEC}}(x) = \mathcal{H}^{\mathrm{BEC}}(q_t^{\mathrm{BEC}}(x)), \tag{14}$$
$$q_{t+1}^{\mathrm{BSC}}(x) = 1/2 - 1/2\sqrt{1 - \mathcal{H}^{\mathrm{BSC}}(q_t^{\mathrm{BEC}}(x))}, \tag{15}$$
initialized at $q_0^{\mathrm{BEC}}(x) = q_0^{\mathrm{BSC}}(x) = x$. Let $I(X_0; \mathcal{T}_\ell)$ be the mutual information between root and observed leaves at depth ℓ. Then
$$1 - q_\ell^{\mathrm{BEC}}(0) \ge I(X_0; \mathcal{T}_\ell) \ge 1 - h(q_\ell^{\mathrm{BSC}}(0)),$$
where h is the binary entropy function.

Proof. The proof follows easily from that of [9, Proposition 9] upon replacing Lemma 2 with Proposition 1. ■

II. The reconstruction threshold

In this section we prove the reconstruction threshold using the channel comparison method.

Proposition 4. *If $d(1-2\delta)^2 > 1$, then recovery (better than random guess) is possible on d-ary trees.*

Proof. By Proposition 3, it suffices to show that the χ^2-dynamics for BSC expand the information in a neighborhood of 0. Consider a d-ary tree with source X_0. Suppose that its children $X_1, \dots, X_d$ are observed with some probability λ through a BSC channel. Let $Y_1, \dots, Y_d$ be the observations. Because we work in a neighborhood of 0, we write $\lambda = \frac{1}{2} - \epsilon$ with $\epsilon > 0$ very small. For simplicity, write $\kappa := \delta * \lambda = \frac{1}{2} - (1-2\delta)\epsilon$. Then by definition we have
$$\begin{aligned} &I_{\chi^2}(X_0; Y) \\ &= \sum_{x_0 \in \{0,1\}} \sum_{y \in \{0,1\}^d} \frac{\mathbb{P}(X_0 = x_0, Y = y)^2}{\mathbb{P}(X_0 = x_0)\mathbb{P}(Y = y)} - 1 \\ &= 2 \sum_{0 \le i \le d} \binom{d}{i} \frac{\kappa^{2i}(1-\kappa)^{2(d-i)}}{\kappa^i(1-\kappa)^{d-i} + \kappa^{d-i}(1-\kappa)^i} - 1. \end{aligned}$$
Using the formula
$$\begin{aligned} &\kappa^a(1-\kappa)^b + \kappa^b(1-\kappa)^a \\ &= 2^{1-a-b}(1 + (\binom{a}{2} + \binom{b}{2} - ab)4(1-2\delta)^2\epsilon^2 + O(\epsilon^4)), \end{aligned}$$
we can expand in terms of ϵ and get
$$\begin{aligned} &I_{\chi^2}(X_0; Y) \\ &= \sum_{0 \le i \le d} \binom{d}{i} 2^{-d}(1 + (\binom{2i}{2} + \binom{2(d-i)}{2} - 4i(d-i) \\ &- \binom{i}{2} - \binom{d-i}{2} + i(d-i))4(1-2\delta)^2\epsilon^2) + O(\epsilon^4) - 1 \\ &= 4d(1-2\delta)^2\epsilon^2 + O(\epsilon^4). \end{aligned}$$
Note that the input χ^2-information into the local neighborhood is $4\epsilon^2$ under our parametrization. Thus denoting by $I_{\chi^2}^t$ the amount of χ^2-information between a target node and its leaves left after t iterations, we get
$$I_{\chi^2}^t = d(1-2\delta)^2 I_{\chi^2}^{t-1}(1 + o(1))$$
This means that if $d(1-2\delta)^2 > 1$, then for small enough ϵ the dynamics expand the information and hence the input information cannot contract to 0 no matter how small it is. ■

Likewise, BEC comparisons recover the following result:

Proposition 5. *If $d(1-2\delta)^2 \le 1$ and $(d, \delta) \ne (1, 0)$, then recovery (better than random guess) is impossible on d-ary trees.*

Proof. By Proposition 3, we need to show that BEC dynamics contracts information. Let $X_1, \dots, X_d$ be the children of X_0 and $Y_1, \dots, Y_d$ be their observations through a $\mathrm{BEC}_{1-\epsilon}$ channel Applying Lemma 2 to the composed channel $X \to X_i \to Y_i$, we see that we can replace Y_i with Y_i', where each $X \to Y_i'$ is an independent copy of $\mathrm{BEC}_{1-(1-2\delta)^2\epsilon}$. We have
$$I_{\chi^2}(X_0; Y') = 1 - (1 - (1-2\delta)^2\epsilon)^d.$$
The input information is ϵ under our parametrization. Consider the function $f(\epsilon) = 1 - (1 - (1-2\delta)^2\epsilon)^d$. We have $f(0) = 0$ and
$$f'(\epsilon) = d(1-2\delta)^2(1 - (1-2\delta)^2\epsilon)^{d-1}.$$
So $f'(\epsilon) \le 1$ for $\epsilon \in [0,1]$, and equality is only achieved at $\epsilon = 0$. So f has only one fixed point in $[0,1]$, which is 0. Therefore χ^2-information contracts to 0. ■

Remark 2. *In the proof of Proposition 4, we showed that when the input information is close to 0, in the limit the information would contract to a non-zero value. Therefore our proof in fact shows that robust reconstruction (a stronger condition than reconstruction) on such trees is possible. By [13], for broadcasting on trees, the robust reconstruction threshold coincides with the Kesten-Stigum bound. It is shown in [14] that when the alphabet size is at least five, the Kesten-Stigum bound is never tight for the (non-robust) reconstruction problem. So for large alphabet size, our method does not yield tight reconstruction threshold.*

III. Bounds on mutual information

Proposition 6. *Let $d \ge 2$ and $\delta = \delta_c - \tau$ where $d(1-2\delta_c)^2 = 1$. Let T_ℓ be the d-ary tree channel as in above. Then*
$$\begin{aligned} \frac{2d\sqrt{d}}{(d-1)\ln 2}\tau + o(\tau) &\le \lim_\ell I(X_0; T_\ell) \\ &\le \frac{4(d+1)\sqrt{d}}{d-1}\tau + o(\tau). \end{aligned}$$

Proof. The proof is by analyzing the recursion in the proof of Proposition 4 and 5 more carefully.

In the setting of proof of Proposition 4, expanding everything to the order of ϵ^4 and computing a binomial sum, we get
$$\begin{aligned} &I_{\chi^2}(X_0; Y) \\ &= 4d(1-2\delta)^2\epsilon^2 + 16d(d-1)(1-2\delta)^4\epsilon^4 + O(\epsilon^6) \\ &= 4(1 + 4\sqrt{d}\tau + o_\tau(\tau))\epsilon^2 + 16(\frac{d-1}{d} + o_\tau(1))\epsilon^4 + O(\epsilon^6). \end{aligned}$$

The input information is $4\epsilon^2$ under this parametrization. Solving the dynamics, we get

$$\epsilon^* = (\sqrt{\frac{d\sqrt{d}}{d-1}} + o(1))\sqrt{\tau}.$$

This gives

$$\lim_{\ell} I(X_0; T_\ell) \ge 1 - h(\frac{1}{2} - (\sqrt{\frac{d\sqrt{d}}{d-1}} + o(1))\sqrt{\tau}) = \frac{2d\sqrt{d}}{(d-1)\ln 2}\tau + o(\tau).$$

Following the proof of proof of Proposition 5, let us consider the function $f(\epsilon) = 1 - (1 - (1-2\delta)^2\epsilon)^d$. Now the function $f(\epsilon)$ is concave on $[0,1]$, and there is a unique fixed point in $(0,1)$. By expanding in terms of ϵ, we have

$$f(\epsilon) = d(1-2\delta)^2\epsilon - \binom{d}{2}(1-2\delta)^4\epsilon^2 + O(\epsilon^3) = (1 + 4\sqrt{d}\tau + o_\tau(\tau))\epsilon - (\frac{d-1}{2d} + o_\tau(1))\epsilon^2 + O(\epsilon^3).$$

So the unique fixed point is at

$$\epsilon^* = \frac{8d\sqrt{d}}{d-1}\tau + o(\tau).$$

This gives

$$\lim_{\ell} I(X_0; T_\ell) \le \frac{8d\sqrt{d}}{d-1}\tau + o(\tau).$$

In fact, knowing that the limit is linear in τ, we can improve this upper bound. Instead of considering $I(X; Y')$ in the proof of Proposition 5, let us consider $I(X; Y)$ directly. We can compute that

$$I_{\chi^2}(X_0; Y) = \sum_{x_0 \in \{0,1\}} \sum_{y \in \{0,1,*\}^d} \frac{\mathbb{P}(X_0 = x_0, Y = y)^2}{\mathbb{P}(X_0 = x_0)\mathbb{P}(Y = y)} - 1 = 2 \sum_{0 \le j \le i \le d} \binom{d}{i} \epsilon^i (1-\epsilon)^{d-i} \binom{i}{j} \cdot \frac{(1-\delta)^{2j}\delta^{2(i-j)}}{(1-\delta)^j\delta^{i-j} + (1-\delta)^{i-j}\delta^j} - 1.$$

Let us call this function $g(\epsilon)$. Note that by Lemma 2, we always have $g(\epsilon) \le f(\epsilon)$ on $[0,1]$. So the largest fixed point of g is upper bounded by the non-trivial fixed point of f, which is of order $\Theta(\tau)$. This justifies performing series expansion in ϵ.

$$\begin{aligned} g(\epsilon) &= (1-\epsilon)^d + 2d((1-\delta)^2 + \delta^2)\epsilon(1-\epsilon)^{d-1} \\ &+ d(d-1)(\frac{(1-\delta)^4 + \delta^4}{(1-\delta)^2 + \delta^2} + (1-\delta)\delta)\epsilon^2(1-\epsilon)^{d-2} \\ &+ O(\epsilon^3) - 1 \\ &= d(1-2\delta)^2\epsilon - d(d-1)\frac{(1-2\delta)^4}{(1-2\delta)^2+1}\epsilon^2 + O(\epsilon^3) \\ &= (1 + 4\sqrt{d}\tau + o_\tau(\tau))\epsilon - (\frac{d-1}{d+1} + o_\tau(1))\epsilon^2 + O(\epsilon^3). \end{aligned}$$

We see that the largest fixed point of g must satisfy

$$\epsilon^* = \frac{4(d+1)\sqrt{d}}{d-1}\tau + o(\tau).$$

In this way we get

$$\lim_{\ell} I(X_0; T_\ell) \le \frac{4(d+1)\sqrt{d}}{d-1}\tau + o(\tau).$$

■

Remark 3. *We compare the above lower bound with (7) in [2].*[2] *We note that the lower bound of [2] can in the limit be simplified into*

$$\lim_{\ell\to\infty} I_{\chi^2}(X_0; T_\ell) \ge \frac{1}{1 + \frac{1-(1-2\delta)^2}{d(1-2\delta)^2 - 1}}.$$

Near the critical threshold, RHS behaves as $\frac{4d\sqrt{d}}{d-1}\tau$. *So they obtained the the same* χ^2*-information lower bound, thus the same mutual information lower bound, as in Proposition 6.*

[2] did not state explicitly an upper bound on mutual information. Nonetheless, their upper bound is by comparison with percolation, and that leads to an upper bound of

$$\lim_{\ell\to\infty} I(X_0; T_\ell) \le \frac{8d\sqrt{d}}{d-1}\tau + o(\tau).$$

In this case we see that channel comparison leads to a better upper bound.

In the case of binary trees, we perform a more refined analysis to improve the upper bound.

Proposition 7. *Let* $\delta = \delta_c - \tau$ *with* $2(1-2\delta_c)^2 = 1$. *Let* T_ℓ *be the binary tree channel as in above. Then*

$$\lim_{\ell} I(X_0; T_\ell) \le 8(\sqrt{2}+1)(1 - h(\frac{1}{2} - \sqrt{\frac{1}{\sqrt{2}} - \frac{1}{2}}))\tau + o(\tau).$$

Proof. Suppose the input distribution is a mixture of BSC_Δ for Δ supported at $\{1/2 - \alpha_t, 1/2\}$. We iterate the dynamics of Proposition 3 while finding the best (w.r.t the less noisy order) channel within this family. This family contains BEC (corresponding to $\alpha = 1/2$), so this approach may lead to a better bound. We define

$$\bar{\delta} := (1/2 - \alpha) * \delta = 1/2 - \alpha(1 - 2\delta).$$

The output distribution has support $\{\frac{\bar{\delta}^2}{\bar{\delta}^2 + (1-\bar{\delta})^2)}, \bar{\delta}, 1/2\}$. Using Lemma 1, we replace $\mathrm{BSC}_{\bar{\delta}}$ with a mixture of $\mathrm{BSC}_{1/2}$ and $\mathrm{BSC}_{\frac{\bar{\delta}^2}{\bar{\delta}^2 + (1-\bar{\delta})^2)}}$, while preserving χ^2-information. Therefore

$$1/2 - \alpha_{t+1} = \frac{\bar{\delta}^2}{\bar{\delta}^2 + (1-\bar{\delta})^2}.$$

Solving this, we get that in the ℓ limit

$$\alpha^* = \frac{\sqrt{1 - 4\delta}}{2(1-2\delta)}.$$

For $\alpha = \alpha^*$, we have

$$C_{\chi^2}(\mathrm{BSC}_{\bar{\delta}}) = (1-2\delta)^2 C_{\chi^2}(\mathrm{BSC}_{\frac{\bar{\delta}^2}{\bar{\delta}^2 + (1-\bar{\delta})^2}}).$$

So when applying Lemma 1, every unit weight for the former becomes $(1-2\delta)^2$ weight for the latter.

Let ϵ_t be the weight of $\mathrm{BSC}_{1/2}$ in iteration t. Then in the ℓ limit ϵ should satisfy

$$1 - \epsilon = (1-\epsilon)^2(\bar{\delta}^2 + (1-\bar{\delta})^2) + 2\epsilon(1-\epsilon)(1-2\delta)^2.$$

Solving this we get $\epsilon^* = 1 - 8(\sqrt{2}+1)\tau + o(\tau)$.

So an upper bound for mutual information is

$$(1-\epsilon^*)(1 - h(1/2 - \alpha^*)) = 8(\sqrt{2}+1)(1 - h(\frac{1}{2} - \sqrt{\frac{1}{\sqrt{2}} - \frac{1}{2}}))\tau + o(\tau).$$

[2] [2] contains an error stating that $I \ge I_{\chi^2}$, which should be $I \ge \frac{1}{2}I_{\chi^2}$. (Note that they define mutual information with natural logarithm.) This leads to lower bounds on I (e.g., (4)(28) in [2]) to be off by a factor of 2. (7) in [2] is correct as stated.

■

Remark 4. *The same method can be applied to the lower bound, leading to* $\alpha_* = (\sqrt{3\sqrt{2}}+o(1))\sqrt{\tau}$ *and* $\epsilon_* = \frac{1}{3}+o(1)$*, giving*

$$\lim_\ell I(X_0; T_\ell) \geq \frac{4\sqrt{2}}{\ln 2}\tau + o(\tau).$$

Surprisingly, although we lower bound using a larger family, and the limiting distribution is different, we get the same lower bound as Proposition 6.

We have shown that $I(X_0; T_\ell) = c\tau + o(\tau)$ *for some* $c \in [8.16, 14.21]$*. The improvement over Proposition 6 can be attributed to a finer "quantization" since we try to work with less noisy channels while staying closer to the true output of BP. We shall explore this idea further in Section IV and show (numerically) that the correct slope is* $c \approx 8.16$*.*

IV. Improved bounds via local comparisons

One advantage of the comparison method is that it allows us to analyze BP, rather than some suboptimal algorithm. On the other hand, we incur some loss in each step of the analysis due to the crude approximations that are made to the input distribution in order to simplify the analysis. In some cases these losses can be significant. For instance, a naive application of the comparison method while matching probabilities of error (i.e., using least degraded channels and Proposition 2) does not even recover the right threshold. One way to avoid this issue is to do local comparisons. We first define a few quantizing operators.

Definition 5 (Q-Operators)**.** *Consider a binary random variable* X *with probability law* μ *along with quantization intervals* (a_i, b_i)*. Define the quantized BSC operator* $Q^{\mathrm{BSC}}(X)$ *as follows: replace the support of* μ *along each* (a_i, b_i) *with a single point at* $\delta_i := \frac{\int_{a_i}^{b_i} \delta d\mu}{\int_{a_i}^{b_i} d\mu}$ *with probability mass* $\int_{a_i}^{b_i} d\mu$*. Likewise, define the quantized BEC operator* $Q^{\mathrm{BEC}}(X)$ *as follows: replace the support along* (a_i, b_i) *with two quantization points* a_i, b_i *with probabilities* $p_{a_i} := \alpha_i p_i$*,* $p_{b_i} = (1-\alpha_i)p_i$*, where* $p_i = \int_{a_i}^{b_i} d\mu$ *and* $\alpha_i = \frac{b_i - \int_{a_i}^{b_i}\delta d\mu / \int_{a_i}^{b_i} d\mu}{b_i - a_i}$*. Furthermore, define* $Q^{\mathrm{BSC}}_{\chi^2}$ *(resp.* $Q^{\mathrm{BEC}}_{\chi^2}$*) similarly by matching the* χ^2*-information along each interval while contracting (resp. spreading) probability masses.*

The main idea is presented in the next proposition:

Proposition 8. *Consider broadcasting on a tree with parameter* δ*. Suppose that* μ_0^{BSC} *(the law at the boundary of the tree) is induced by* BSC_{δ_0}*, where* δ_0 *is chosen so that* $\delta_0 \geq \lim_\ell P_e(\mathcal{T}_\ell)$*. Let* $\mu_t^{\mathrm{BSC}} = Q^{\mathrm{BSC}}(\mathrm{BP}(\mu_{t-1}^{\mathrm{BSC}}))$ *be obtained by quantizing the output of BP operating on* μ_{t-1}^{BSC}*. Let* q_t^{BSC} *be the corresponding probability of error* $q_t^{\mathrm{BSC}} := \sum_i \delta_{a_i b_i}\mu_{t,i}^{\mathrm{BSC}}$*. Similarly, define* $\nu_t^{\mathrm{BSC}} = Q^{\mathrm{BSC}}_{\chi^2}(\nu_{t-1}^{\mathrm{BSC}})$*. Let* ι_t^{BSC} *be the corresponding mutual information. Likewise, define* $\mu_t^{\mathrm{BEC}} = Q^{\mathrm{BEC}}(\mathrm{BP}(\mu_{t-1}^{\mathrm{BEC}}))$ *with probability of error* q_t^{BEC} *with* $P(\mu_0 = 0) = 1$*. Define* $\nu_t^{\mathrm{BEC}}, \iota_t^{\mathrm{BEC}}$ *similarly. The following statements hold:*

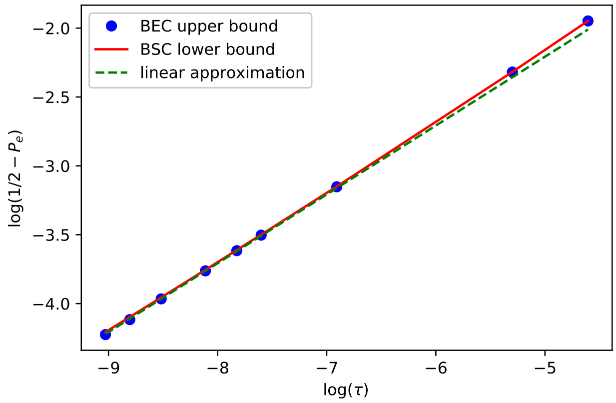

Fig. 1: Bounds on probability of error using local comparisons for $\delta = \delta_c - \tau$. The linear approximation has a slope of 1/2.

1) $q_\ell^{\mathrm{BEC}} \leq \lim_\ell P_e(\mathcal{T}_\ell) \leq q_\ell^{\mathrm{BSC}}$.
2) $\iota_\ell^{\mathrm{BEC}} \geq \lim_\ell I(\mathcal{T}_\ell) \geq \iota_\ell^{\mathrm{BSC}}$.

Remark 5. *To choose* $\delta_0 < \frac{1}{2}$ *we may, for example, use a Kesten-Stigum upper bound on* P_e*, corresponding to a suboptimal algorithm as in* (3)*.*

Proof. Note that $Q^{\mathrm{BSC}}(\mu)$ is obtained from μ by the transformation

$$(y, \delta) \mapsto \begin{cases} (y, \delta_i) & \delta \in [a_i, b_i], \\ (y, \delta) & \text{o.w.} \end{cases}$$

The probabilities of error match by construction. This shows that $Q^{\mathrm{BSC}}(\mu_t)$ is a degradation of μ_t. This proves the upper bound since if the initial input is degraded w.r.t $\mathcal{T}_\ell$ then all the subsequent iterations remain degraded. For the lower bound, we note that a probability distribution with its masses at the center of an interval is a degradation of one with two spikes at the boundaries. Note that indeed when the original distribution has a single atom in some interval this follows directly from the above transformation. The general case follows since if there are more than one atoms, we can degrade sequentially. This proves the first statement. The second statement can be proved similarly. ■

Using uniform quantization in the $[0, 1/2]$ interval with 1024 points, we were able to show that

$$I(X_0; \mathcal{T}_\ell) = c\tau + o(\tau)$$

with $c \approx 8.16$. This is the basis of our Conjecture 1.

Using a degradation argument (or Fano's inequality), one can also show

$$1/2 - c'\sqrt{\tau} + o(\sqrt{\tau}) \leq P_e(\mathcal{T}_\ell) \leq 1/2 - c\tau + o(\tau).$$

It is natural to ask what is the correct exponent for P_e. Using the same approach we were able to show (see Fig. 1)

$$\log(1 - 2P_e) \geq 0.504 \log \tau + c.$$

We thus conjecture that $\sqrt{\tau}$ is the correct exponent.

REFERENCES

[1] P. M. Bleher, J. Ruiz, and V. A. Zagrebnov, "On the purity of the limiting gibbs state for the ising model on the bethe lattice," *Journal of Statistical Physics*, vol. 79, no. 1, pp. 473–482, Apr 1995. [Online]. Available: https://doi.org/10.1007/BF02179399

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

[3] A. Makur, E. Mossel, and Y. Polyanskiy, "Broadcasting on random directed acyclic graphs," *IEEE Transactions on Information Theory*, 2019.

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

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

[6] E. Mossel, J. Neeman, and A. Sly, "Belief propagation, robust reconstruction and optimal recovery of block models," in *Conference on Learning Theory*, 2014, pp. 356–370.

[7] E. Mossel, "On the impossibility of reconstructing ancestral data and phylogenies," *Journal of computational biology*, vol. 10, no. 5, pp. 669–676, 2003.

[8] W. Evans and L. J. Schulman, "Signal propagation, with application to a lower bound on the depth of noisy formulas," in *Proceedings of 1993 IEEE 34th Annual Foundations of Computer Science*. IEEE, 1993, pp. 594–603.

[9] H. Roozbehani and Y. Polyanskiy, "Low density majority codes and the problem of graceful degradation," *arXiv preprint arXiv:1911.12263*, 2019.

[10] A. El Gamal and Y.-H. Kim, *Network information theory*. Cambridge university press, 2011.

[11] A. Makur and Y. Polyanskiy, "Comparison of channels: criteria for domination by a symmetric channel," *IEEE Transactions on Information Theory*, vol. 64, no. 8, pp. 5704–5725, 2018.

[12] Y. Polyanskiy and Y. Wu, "Strong data-processing inequalities for channels and bayesian networks," in *Convexity and Concentration*. Springer, 2017, pp. 211–249.

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

[14] A. Sly, "Reconstruction for the potts model," in *Proceedings of the forty-first annual ACM symposium on Theory of computing*. ACM, 2009, pp. 581–590.