Ising Model on Locally Tree-like Graphs: Uniqueness of Solutions to Cavity Equations

Qian Yu* and Yury Polyanskiy†
*Department of Electrical and Computer Engineering, Princeton University
†Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology

Abstract

In the study of Ising models on large locally tree-like graphs, in both rigorous and non-rigorous methods one is often led to understanding the so-called belief propagation distributional recursions and its fixed points. We prove that there is at most one non-trivial fixed point for Ising models with zero or certain random external fields. Previously this was only known for sufficiently "low-temperature" models. Our main innovation is in applying information-theoretic ideas of channel comparison leading to a new metric (degradation index) between binary-input-symmetric (BMS) channels under which the Belief Propagation (BP) operator is a strict contraction (albeit non-multiplicative). A key ingredient of our proof is a strengthening of the classical stringy tree lemma of [1].

Our result simultaneously closes the following 6 conjectures in the literature: 1) independence of robust reconstruction accuracy to leaf noise in broadcasting on trees [2]; 2) uselessness of global information for a labeled 2-community stochastic block model, or 2-SBM [3]; 3) optimality of local algorithms for 2-SBM under noisy side information [4]; 4) uniqueness of BP fixed point in broadcasting on trees in the Gaussian (large degree) limit [4]; 5) boundary irrelevance in broadcasting on trees [5]; 6) characterization of entropy (and mutual information) of community labels given the graph in 2-SBM [5].

CONTENTS

I. Main result and motivation

The central object of interest in this paper is a belief propagation (BP) operator $\mathcal{Q}_s$ that takes a symmetric distribution μ and produces another symmetric distribution $\mathcal{Q}_s\mu$. We call a probability distribution μ on $(-\infty,+\infty]$ *symmetric* if

$$d\mu(r) = e^r d\mu(-r), \qquad \iff \qquad \mu[E] = \int e^{-r}\mathbb{1}\{-r \in E\} d\mu(r) \tag{1}$$

for every measurable subset $E \subseteq (-\infty,+\infty)$ (see [6, Section. 15.2.2] or Section I-A for motivation). A special distribution $\mu(\{0\}) = 1$ is denoted by a Dirac-delta δ_0 and is called *trivial*. The operator $\mathcal{Q}_s$ depends on three parameters: a *crossover probability* $\delta \in [0,1]$, a symmetric *survey distribution* μ_s and a *(branching or) degree distribution* P_d on $\mathbb{Z}_+$. Given these, we define $\mathcal{Q}_s\mu$ for any symmetric μ to be the probability law of random variable R

$$R \triangleq \sum_{u=1}^{d} Z_u F_\delta(\tilde{R}_u) + S, \tag{2}$$

where $d \sim P_d$, $\tilde{R}_u \overset{iid}{\sim} \mu$, $Z_u \overset{iid}{\sim} (-1)^{\mathrm{Ber}(\delta)}$, $S \sim \mu_s$ (all jointly independent) and

$$F_\delta(x) \triangleq \ln\frac{(1-\delta)e^x + \delta}{\delta e^x + 1 - \delta} = 2\tanh^{-1}\left((1-2\delta)\tanh\frac{x}{2}\right).$$

The special case of $S = 0$ or, equivalently, $\mu_s = \delta_0$ is referred to as *BP without survey* and in this case we denote the BP operator by $\mathcal{Q}$ without the subscript s.

In this paper we consider the topic of convergence of iterations $\mathcal{Q}_s^h\mu_0$ as $h \to \infty$. Naturally, in this regard, we define distribution μ to be a *BP fixed point* if $\mathcal{Q}_s\mu = \mu$. Note that in the case of no survey ($S = 0$), and only in that case, there is a *trivial fixed point* $\mu = \delta_0$.

The main result of our work is the following.

Theorem 1. *There exists at most one non-trivial symmetric BP fixed point μ^*, unless we are in the exceptional case of $S = 0$ (no survey), $d = 1$ a.s. and $\delta \in \{0,1\}$ (in which case $\mathcal{Q}_{\mathrm{s}}$ is an identity operator). In the non-exceptional case, for all non-trivial symmetric μ, the recursion $\mathcal{Q}_{\mathrm{s}}^h\mu$ converges weakly to the same fixed point, to μ^* if it exists, or to the trivial δ_0 otherwise.*

This result is contained in Theorems 4 and 15 below. As alluded to in the abstract, Theorem 1 resolves a number of long-standing questions in the theory of Ising models on trees and locally tree-like graphs. In a nutshell, the main innovation of our work is the discovery of a (rather strange) metric between distributions (equivalently, between BMS channels) under which a finite number of applications of $\mathcal{Q}_{\mathrm{s}}$ is strictly contracting (see Definition 13 below).

We note, however, that our result says little about the actual structure of the fixed point. From prior work, though, we know that in the case of $S = 0$ (no survey) and fixed degree d the δ_0 is the unique fixed point iff $(1-2\delta)^2 d \le 1$ (a Kesten-Stigum threshold [1]). Above criticality the non-trivial fixed point μ^* emerges and it is known to be approximately Gaussian [7] in the sense that if $R_\delta \sim \mu^*$ then as $(1-2\delta)^2 d \to 1^+$:

$$\frac{R_\delta}{\sqrt{1-2\delta-d^{-\frac{1}{2}}}} \xrightarrow{d} \mathcal{N}(0,\frac{\sigma^2}{2}), \qquad \sigma^2 \triangleq \frac{16d\sqrt{d}}{d-1}.$$

We next proceed to deriving the recursion (2) and explaining its connection to Ising models, statistical physics and stochastic block model.

A. Derivation of BP recursion

Consider an inference problem for the Ising model on infinite trees (see Fig.1 for an illustration). We have a rooted tree channel that is generated recursively, where each vertex v has an i.i.d. number of children sampled from a given degree distribution P_d. Each vertex v is associated with a binary random variable X_v. The variable on the root, denoted by X_0, is $\mathrm{Ber}(\frac{1}{2})$. Then for any other vertex v, X_v is identical to the variable on their parent node with probability $1-\delta$, conditioned on all other variables that are not their descendant, where δ is a given parameter in $[0,1]$.

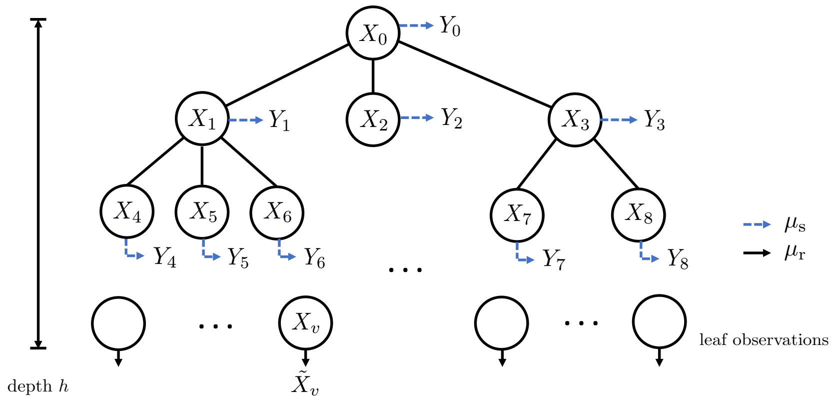


Fig. 1: Illustration of a tree channel with depth h. The solid lines represent binary symmetric channels with error probability δ. The estimation of X_0 is based on leaf observations and the possible surveys. The corresponding channels are represented by the solid arrows and the dashed arrows, with log-likelihood ratio distributions given by μ_{r} and μ_{s}, respectively. The leaf observations consist of only the vertices at depth h, not including vertex 2 in this example.

In a basic setting called broadcasting on trees (BOT), we are interested in the process of estimating X_0 given the random tree graph structure and the collection of variables on all vertices with depth h. Formally, we aim to characterize the distribution of the log-likelihood ratio (LLR) for estimating X_0 conditioned on $X_0=0$. Let L_h denote the set of all vertices at depth h from the root, $X_{L_h}=\{X_v : v\in L_h\}$ be the vector of all values at depth h, and $T_{(h)}$ denote the tree subgraph induced by all vertices in $\{L_0, L_1, ..., L_h\}$. The set of all observed information consists of $T_{(h)}$ and X_{L_h}. The LLR distribution for tree networks of depth h, denoted by $\mu_{(h)}$, is given by

$$\mu_{(h)}[E]=\mathbb{P}\left[\ln\frac{\mathbb{P}[T_{(h)},X_{L_h}|X_0=0]}{\mathbb{P}[T_{(h)},X_{L_h}|X_0=1]}\in E\middle|X_0=0\right]$$

for any measurable $E\subseteq(-\infty,+\infty]$.

Knowing the LLR distribution, we can compute fundamental quantities such as minimum error probabilities and mutual information. Particularly, the recoverability of X_0 with non-trivial error probability (bounded away from $\frac{1}{2}$ for large h) is possible if and only if the LLR distributions converge to a non-trivial distribution.

More generally, a variant of BOT is considered in [5], where in addition to the *leaf observations* (i.e., variables at depth h), each vertex above depth h is also observed through identical BMS channels, and the received information is called the *surveys*. This formulation can be further generalized to include the case where leaf observations are noisy, with corresponding observation channels being identical and BMS. The LLR distributions for such cases can be similarly defined. Formally, for any vertex v with depth less than h, there is a survey observation denoted by Y_v. For any vertex v with depth h, there is a (possibly) noisy leaf observation denoted by $\tilde{X}_v$. All Y_v's and $\tilde{X}_v$'s are jointly independent conditioned on the tree structure and the variables on the tree. The conditional distribution of each Y_v and $\tilde{X}_v$ only depends on their corresponding variable X_v. They are specified by either the transition distribution of the survey channels or the leaf observation channels, respectively. Let $Y_{(h)}\triangleq\{Y_v : v\in L_0\cup L_1\cup ...\cup L_{h-1}\}$ be the collection of all surveys and $\tilde{X}_{(h)}\triangleq\{\tilde{X}_v : v\in L_h\}$ be the collection of noisy leaf observations. The set of all observations is given by $\mathcal{O}_h\triangleq(T_h, Y_{(h)}, \tilde{X}_{(h)})$.

Now that $\mathcal{O}_h$ can belong to continuous domains, we define the LLR distribution $\mu_{(h)}$ to be the law of[1]

$$\ln\frac{f_{0,h}\left(\mathcal{O}_h\right)}{f_{1,h}\left(\mathcal{O}_h\right)} \text{ given } X_0=0, \text{ where } f_{v,h}\triangleq\frac{\mathrm{d}P_{\mathcal{O}_h|X_0=v}}{\mathrm{d}P_{\mathcal{O}_h}}.$$

Following the same definition, we define the LLR distributions for all binary-input channels by replacing X_0 with their input variables and $\mathcal{O}_h$ with their outputs.

[1]Here the LLR variable $\ln\frac{f_{0,h}(\mathcal{O}_h)}{f_{1,h}(\mathcal{O}_h)}$ is constructed using the Radon–Nikodym derivative, which is well-defined for all measurable spaces and unique up to a set of zero measure.

The BP operator arises naturally as the recursion rule for the LLR distributions of the tree channels described above. In the no survey setting, we have $\mu_{(h+1)} = \mathcal{Q}\mu_{(h)}$. When survey channels are present, we have $\mu_{(h+1)} = \mathcal{Q}_{\mathrm{s}}\mu_{(h)}$, where μ_{s} is set to be the LLR distribution of the corresponding survey channels.

The derivation of BP recursion relies on the fact that the observed information can be partitioned into subsets of independent variables given X_0 and the tree structure. Therefore, the overall LLR can be written as the summation of individual LLRs from each component. Specifically, consider the tree channel of depth $h+1$. Let d be the degree of the root vertex and let $1, ..., d$ be the labels of the d children. The individual terms consist of the LLRs from the subtrees rooted at each vertex $u \in [d]$, and the LLR from the survey when it exists.

Due to the recursive structure of the tree construction, the subtree rooted at each vertex $u \in [d]$ resembles the same network that is defined for depth h. Consequently, if we consider the LLR variable for estimating X_u with the subtree information, which is a function of all relevant surveys, leaf observations, and the subtree structure, then by definition, the law of this variable is given by $\mu_{(h)}$ conditioned on $X_u = 0$. For convenience, we denote this variable by $\hat{R}_u$. Observe that each subtree is a BMS channel. We have $-\hat{R}_u \sim \mu_{(h)}$ conditioned on $X_u = 1$. Hence, by letting $Z_u \triangleq (-1)^{X_u}$ and $\tilde{R}_u \triangleq Z_u\hat{R}_u$, we have $\tilde{R}_u \sim \mu_{(h)}$ independent of X_u. The LLR component for X_0 that corresponds to this subtree takes into account of the uncertainty of X_u, which is given by

$$\ln \frac{e^{\hat{R}_u}\mathbb{P}[X_u = 0|X_0 = 0] + \mathbb{P}[X_u = 1|X_0 = 0]}{e^{\hat{R}_u}\mathbb{P}[X_u = 0|X_0 = 1] + \mathbb{P}[X_u = 1|X_0 = 1]} = F_\delta(\hat{R}_u) = Z_u F_\delta(\tilde{R}_u).$$

Hence, the overall LLR can be written as $\sum_{u=1}^{d} Z_u F_\delta(\tilde{R}_u) + S$, where S is the LLR variable for the survey at vertex 0. Recall that $\tilde{R}_u$, Z_u, and S are jointly independent conditioned on $X_0 = 0$, which is given by the tree channel construction. Further, we also have $Z_u \sim (-1)^{\mathrm{Ber}(\delta)}$ and $S \sim \mu_s$ under the same condition. Thus, we have recovered the BP operator specified by equation (2). Finally, the initial condition of the recursion is simply given by $\mu_{(0)} = \mu_{\mathrm{r}}$, where μ_{r} is the LLR distribution of the leaf observation channels.

The symmetry condition of μ_{s}, μ_{r}, and $\mu_{(h)}$ arises from the BMS property of the corresponding channels. Generally, for any BMS channel, let μ denote its LLR distribution and let μ^- be the distribution of the same LLR function generated with input $X_0 = 1$. By the BMS property, we must have $d\mu^-(r) = d\mu(-r)$ for any $r \in [-\infty, +\infty)$. Then the definition of LLR gives $d\mu_{(h)}(r) = e^r d\mu^-_{(h)}(r)$, which implies equation (1).

B. Cavity method and previous work

The operator $\mathcal{Q}$ is also known as density evolution [8, Section 2.2], Bethe recursion [9, Definition 1.6]. It arises from a so-called cavity method [10], which (non-rigorously, but often correctly) allows one to infer important qualities and quantities of statistical physical systems based on the knowledge of the fixed points of the $\mathcal{Q}$. Correspondingly, the distributional identity $\mathcal{Q}\mu = \mu$ is known as the 1RSB cavity equation, with Parisi parameter set to $x = 1$ [6, Section 14.6, (19.72)]. The particular version of the $\mathcal{Q}$ corresponds to cavity equations for the Ising model on a locally tree-like graph for ferromagnetic $\delta < \frac{1}{2}$ or anti-ferromagnetic $\delta > \frac{1}{2}$ case. We mention that distributional recursions are not necessary for understanding the former [11] (due to absence of frustration in the boundary condition) but are necessary for the latter, see [8]. The version with survey (BP operator $\mathcal{Q}_s$) would correspond to certain random external fields, which are not independent across sites.

Both the BOT and the BP fixed point formulations have been widely studied in statistical physics [11], [12], evolutionary biology [13], [14], and information theory [1], [15]. The condition for the existence of non-trivial symmetric BP fixed points has been determined exactly, which can be described using the branching number [1], [12], [16]. Equivalently, this resolved the problem of identifying the set of (P_d, δ) for which recovery is possible in the BOT. The version of BOT with noisy leaf observations was introduced in [17], who also demonstrated that the regime of recovery (for general leaf observation channels under the Ising model case we are considering here) is unchanged.

A renewed interest in BOT was sparked by the groundbreaking works of [18], [19], which connected it to the stochastic block model with 2 communities (2-SBM). In 2-SBM the goal is to estimate a set of hidden labels by observing an associated random graph. The labels are defined on n vertices, each being i.i.d. Ber($\frac{1}{2}$). The graph is constructed by independently connecting any pair of vertices, with probability $\frac{a}{n}$ for pairs with the same labels and with probability $\frac{b}{n}$ for the rest. It turned out that in the regime of $n \to \infty$ the probability of error of estimating the label given the graph can be lower bounded by the probability of recovery error in BOT with $P_d = \mathrm{Poisson}(\frac{a+b}{2})$ and $\delta = \frac{b}{a+b}$. On the other hand, [2] shows that the 2-SBM error can be upper bounded by the recovery error in the robust BOT with the same parameters. This was established by running BP from a good initialization (for details, see [2, Section 5]). Consequently, the uniqueness of BP fixed point implies that the upper and the lower bounds coincide, showing that the performance of the optimal (but exponential time) maximum likelihood estimator can be achieved in polynomial time. This result, however, was only shown in [2] under the condition of "high SNR" or low temperature. They made a conjecture that the result (and BP uniqueness) should hold unconditionally. In [5] the range under which the uniqueness holds was further enlarged. This paper resolves the conjecture in full.

The consequence of the discussion above is that for the 2-SBM we can explicitly (modulo computing the BP fixed point) evaluate the probability of error for recovering an individual vertex label. A more global quantity is conditional entropy of *all* *vertex labels* given the graph. [5] gives a formula for the latter, but only under the assumption of *boundary irrelevance (BI)* for the problem of BOT with survey. BI refers to the effect of leaf observations becoming independent of the root variable when conditioned on the survey information. BI was conjectured to be true in [3] for binary erasure survey channels, and in [5] for general symmetric survey channels. In our language, BI is equivalent to uniqueness of BP fixed point for the $\mathcal{Q}_s$. We thus provide a positive proof of these conjectures in Section IV-B.

The setting of BOT with survey arose in the line of work [3], [4] which investigated optimality of local algorithms (of BP type) and made conjectures similar to the BI. Furthermore, using methods of [2] they were able to prove those conjectures in the regime of high-SNR. Our proof for all those conjectures closes the full spectra of the SNR and essentially follows from BI – see Section IV-E.

We also mention that uniqueness of BP fixed point has been investigated under a simplified formulation, where the recursion is approximated using central limit theorem when the degrees are large – see Conjecture 2.6 in [4]. Our proof techniques directly extend to this limiting regime as well – see Section IV-C.

In conclusion, identifying the uniqueness of the BP fixed point for either $\mathcal{Q}$ (no survey) or $\mathcal{Q}_s$ (with survey) has been a long-unsolved question, appearing in a web of interlinked problems. Our resolution of the uniqueness closes all related conjectures as well (Section IV). Although partial resolutions were already presented in [2], [5] the method here appears to be completely different and we do not believe that tightening of the previous methods would be able to close the full range of SNR – this is briefly discussed further in Section II-A.

C. Extension to non-symmetric distributions

In the context of cavity equation, the BP recursion can be defined and studied for asymmetric distributions. To understand how the general class of distributions is defined, we need to recall that the BP operator is derived in a framework where the input μ can be viewed as the LLR distribution of some binary-input leaf observation channels conditioned on $X_0 = 0$. Any such distribution must satisfy the following condition.

Definition 2. We call a probability distribution μ on $(-\infty, +\infty]$ an *LLR distribution* if

$$\int e^{-r} d\mu(r) \leq 1. \tag{3}$$

We define the *complement distribution* of μ to be any distribution μ^- on $[-\infty, +\infty)$ that satisfies

$$d\mu(r) = e^r d\mu^-(r) \qquad \iff \qquad \mu^-[E] = \int e^{-r} \mathbb{1}\{r \in E\} d\mu(r) \tag{4}$$

for every measurable subset $E \subseteq (-\infty, +\infty)$.

In other words, μ must allow the existence of a μ^- that can be served as the law of the LLR variable generated by $X_0 = 1$.[2] For general asymmetric LLR distributions, the BP operator that reflects the same process (i.e., BOT recursion specified by the same parameters but with asymmetric leaf observation channels) needs to be written in a slightly different form. It can be derived from the same steps in Section I-A. Specifically, for any LLR distribution μ and any associated complement distribution μ^-, we define $\mathcal{Q}_{\mathrm{s}}\mu$ to be the probability law of random variable R

$$R \triangleq \sum_{u=1}^{d} F_\delta(\hat{R}_u) + S, \tag{5}$$

where $d \sim P_d$, $\hat{R}_u \overset{iid}{\sim} (1-\delta)\mu + \delta\mu^-$, and $S \sim \mu_s$, all jointly independent. We call any LLR distribution μ a *BP fixed point* if $\mathcal{Q}_{\mathrm{s}}\mu = \mu$. It can be seen that for symmetric μ this definition coincides with the one we gave earlier in this section.

Our main result implies the uniqueness of BP fixed points over general distributions and the unique convergence of BP recursion with general initialization. The proof for the following result is presented in Appendix C.

Corollary 3 (Asymmetric BP fixed points)**.** *Fix any degree distribution P_d, parameter $\delta \in [0,1]$, and symmetric μ_{s} that belongs to the non-exceptional case specified in Theorem 1. There is at most one non-trivial BP fixed point μ^*, and it is symmetric. For all non-trivial LLR distribution μ, the recursion $\mathcal{Q}_{\mathrm{s}}^h\mu$ converges weakly to the same symmetric fixed point, to μ^* if it exists, or to the trivial δ_0 otherwise.*

D. Organization of the paper

The rest of the paper is organized as follows. In Section II, we define some important tools and provide the proof ideas for our main theorem. Then in Section III, we provide the proof details for the key intermediate steps. We illustrate in Section IV how our results imply the solutions to the open conjectures mentioned earlier. Finally in Section V, we extend our results to general tree structures, covering curious cases where the spin interactions can be stochastic (see the setting of i.i.d. weights in [13]), periodic (see the first example in [16]), or nonisotropic (see the illustration in Fig. 2).

[2] Conversely, any LLR distribution can be mapped to a binary-input channel, similar to that any symmetric distribution is the LLR distribution of a BMS channel.

II. Proof ideas and outline

We first present the proof of the main result for the case of no survey $\mu_s = \delta_0$. Namely, we show the following.

Theorem 4 (Uniqueness without survey)**.** *Fix any degree distribution P_d and parameter $\delta \in [0,1]$ such that either $\mathbb{P}[d=1] < 1$ or $\delta \in (0,1)$. There is at most one non-trivial symmetric BP fixed point μ^* for $\mathcal{Q}$. For all non-trivial symmetric μ, the recursion $\mathcal{Q}^h\mu$ converges weakly when $h \to \infty$ to the same fixed point, to μ^* if it exists, or to the trivial δ_0 otherwise.*

The proof of Theorem 4 builds upon ideas of channel comparison (a.k.a. comparison of experiments), which were previously used in [1] to show certain negative results, and more recently by [15] for the positive side. Here we extend this methodology in two ways: a) strengthening the *stringy tree lemma* from [1]; and b) introducing of the concept of *degradation index*. The latter allows us to define a potential function over symmetric distributions (and LLR distributions in general) that is only stabilized at a unique solution. For clarity, we illustrate the main concepts over symmetric distributions, which enables simplifications compared to their general forms. We first state the definition of degradation (see [6, Section 15.2.3] for a reference).

Definition 5. For any two symmetric distributions μ_Y, μ_Z defined on $(-\infty, +\infty]$, we say μ_Y is a *degraded version* of μ_Z, denoted by $\mu_Y \preceq \mu_Z$, if one can define a joint distribution $\mu_{Y,Z}$ with μ_Y, μ_Z as marginal distributions, such that $\mu_{Y|Z}$ is invariant under $(Y,Z) \to (-Y,-Z)$.

Intuitively, for any $\mu_Y \preceq \mu_Z$, μ_Z can be viewed as the LLR distribution of a symmetric binary hypothesis testing problem and μ_Y can be viewed as a noisy version of μ_Z where the observation is corrupted by a symmetric noise channel $\mu_{Y|Z}$. A more detailed discussion on degradation can be found in Appendix A.

Our definition of the degradation index is based on the operator known as box-convolution (notation coming from the LDPC codes, see [20, page 181]). Consider a pair of BMS channels A_1 and A_2, with LLR distributions μ_i, $i \in \{1,2\}$. Out of them we can produce a new BMS channel A as follows: consider an input bit X and generate an independent X' as Bernoulli$(1/2)$, let $Y' = A_2(X')$ be the noisy observation of X' and let $Y = A_1(X \oplus X')$ be the noisy observation of XOR of X and X'. The A channel is a channel from X to the pair (Y, Y') and we denote its LLR by $\mu = \mu_1 \boxast \mu_2$. More formally, we have the following definition.

Definition 6 (Box Convolution for Symmetric Distributions)**.** For any $\phi \in [0,1]$, let B_ϕ denote the symmetric distribution defined on $\{-\ln\frac{1-\phi}{\phi}, \ln\frac{1-\phi}{\phi}\}$. We define box convolution $\boxast$ to be the weakly continuous bilinear operator over the space of symmetric distributions satisfying the following condition

$$B_{\phi_1} \boxast B_{\phi_2} \triangleq B_{\phi_1+\phi_2-2\phi_1\phi_2} \qquad \text{for all } \phi_1, \phi_2 \in [0,1]\,.$$

It is clear that box convolution is commutative. One can show the following alternative definition, which proves that box convolution is associative.

Proposition 7. *Let $X \sim \mu$, $Y \sim \nu$ be independent random variables with symmetric distributions, then $\mu \boxast \nu$ is identical to the law of*

$$Z \triangleq 2\tanh^{-1}\left(\tanh\frac{X}{2}\tanh\frac{Y}{2}\right). \tag{6}$$

Proof. Note that the law specified by equation (6) is weakly continuous and bilinear. It fulfills all requirements in Definition 6 (the special cases of $\mu = B_{\phi_1}$ and $\nu = B_{\phi_2}$ can be directly verified). On the other hand, the box convolution is uniquely determined through bilinear expansion where the input distributions are expressed as mixtures of $\{B_\phi\}_{\phi\in[0,1]}$. For instance, any symmetric μ is a mixture under the law of $\phi \sim \frac{1}{2}(1-\tanh\frac{|X|}{2})$. Hence, based on this uniqueness, any operator that satisfies Definition 6 needs to be identical to the instance provided in the proposition. ☐

Remark 1. We point out a convenient interpretation of the channel corresponding to $B_\phi \boxast \mu$. This channel corresponds to sequentially concatenating a binary symmetric channel (BSC) with crossover probability ϕ and a channel with LLR μ. Thus, compared to the μ channel the input bit first experiences a random ϕ-flip. The general box convolution $\mu_1 \boxast \mu$ corresponds to the same channel, except that the crossover probability ϕ is random (but known to the receiver) and $\log\frac{1-\phi}{\phi} \sim \mu_1$.

The stringy tree lemma in [1] can be stated as follows, using box convolution.

Theorem 8. *[Stringy Tree Lemma (STL) [1]] For any $\phi \in [0,1]$ and any symmetric distributions $\mu_1, \mu_2, ..., \mu_d$, we have*

$$B_\phi \boxast (\mu_1 * \mu_2 * \ldots * \mu_d) \preceq (B_\phi \boxast \mu_1) * (B_\phi \boxast \mu_2) * \ldots * (B_\phi \boxast \mu_d).$$

Remark 2. Recall the physical interpretation of box-convolution. Applying the theorem above repetitively compares any tree channel with a depth 1 tree channel, each edge formed by concatenating all channels on a path in the original tree from its root to a corresponding leaf (hence the name stringy tree).

We establish the commutation relation between the BP operator and box convolution, but instead under a stronger notion of degradation (to be specified in Definition 10). Observe that the BP operator can be expressed using elementary operations.

Proposition 9. *For any symmetric μ, we have $\mathcal{Q}\mu = \mathbb{E}_{P_d}[(B_\delta \boxast \mu)^{*(d)}]$, where $(\cdot)^{*(d)}$ denotes self convolution by d times.*

Proof. Recall the definition of BP operators. As implied by Proposition 7, for any independent $Z_u \sim (-1)^{\mathrm{Ber}(\delta)}$ and $\tilde{R}_u \sim \mu$, the law of $Z_u F_\delta(\tilde{R}_u)$ can be exactly written as $B_\delta \boxast \mu$. Then, conditioned on any fixed d, the law for the summation of those intermediate variables equals the convolution of their distributions. Finally, the distribution $\mathcal{Q}\mu$ is obtained by taking the mixture over the randomness in degree distribution. $\square$

The key result is summarized in Theorem 11 (proved in Section III), which relies on the following definition.

Definition 10 (Strict Degradation)**.** For any two distributions ν and μ, we define $\nu \prec \mu$ if $\exists\ \phi \in (0, \frac{1}{2}]$ such that $\nu \preceq B_\phi \boxast \mu$.

Theorem 11. *For any $\delta \in [0,1]$ and degree distribution P_d satisfying $\mathbb{P}[d \leq 1] < 1$, let $\mathcal{Q}$ be the associated BP operator. Then for any $\phi \in (0, \frac{1}{2})$ and symmetric μ, we have*

$$B_\phi \boxast \mathcal{Q}^2\mu \prec \mathcal{Q}^2(B_\phi \boxast \mu) \qquad \text{if } \mathbb{P}[d > 2] = 0, \tag{7}$$

$$B_\phi \boxast \mathcal{Q}\mu \prec \mathcal{Q}(B_\phi \boxast \mu) \qquad \text{otherwise.} \tag{8}$$

Remark 3. Note that the fixed point equation can be written as $\mathcal{Q}\mu = \mu$, Theorem 11 implies that μ and $B_\phi \boxast \mu$ can not be both non-trivial symmetric fixed points. As we will show later in this work (see Remark 6), Theorem 11 essentially states that either $\mathcal{Q}^2$ or $\mathcal{Q}$ will always reduce a distance function between distinct non-trivial symmetric distributions measured based on the degradation index.

Remark 4. The stringy tree construction in [1] implies that $B_\phi \boxast \mathcal{Q}\mu \preceq \mathcal{Q}(B_\phi \boxast \mu)$ for any degree distribution. Theorem 11 provides a strict version of this result for $\mathbb{P}[d > 2] > 0$. Note that this condition can not be further relaxed, as one can show that inequality (8) is not satisfied when $\mathbb{P}[d > 2] = 0$.

Remark 5. Theorem 11 shows that our proof requires special treatment when $d \leq 2$. As shown in Section V, these two possible cases can be naturally unified when viewed under a generalized model, where the infinite tree is generated from arbitrary elements. A generalized version of inequality (7) and (8) is provided, and the number of applications of the BP operator for strict inequality to hold depends on a requirement called polygon condition (see Definition 44 and Theorem 45).

We also have the following fact shown in Appendix B-A (by checking that degradation is transitive and box-convolution-preserving):

Proposition 12. *If $\nu \preceq \tau \prec \mu$ or $\nu \prec \tau \preceq \mu$, then $\nu \prec \mu$.*

Given these results, we are ready to prove the main theorem for the no survey case.

Proof of Theorem 4. Consider any two non-trivial symmetric fixed points μ, ν, we first prove that $\mu \preceq \nu$.

Definition 13 (Degradation Index)**.** For any two symmetric distributions μ and ν, we define the *degradation index* from ν to μ to be

$$\phi^*(\mu, \nu) \triangleq \inf \left\{ \phi \mid B_\phi \boxast \mu \preceq \nu \right\}. \tag{9}$$

We use the following Proposition, which is proved in Appendix B-B.

Proposition 14. *Degradation index has the following properties.*

1) *We always have $\phi^*(\mu, \nu) < \frac{1}{2}$ for ν non-trivial.*
2) *For any symmetric μ and ν, we have $B_{\phi^*(\mu,\nu)} \boxast \mu \preceq \nu$.*
3) *For any $\phi \in (0, \frac{1}{2})$ and any symmetric μ and ν satisfying $B_\phi \boxast \mu \prec \nu$, we have $\phi^*(\mu, \nu) < \phi$.*
4) *For any symmetric μ, ν, τ, we have $1 - 2\phi^*(\mu, \tau) \geq (1 - 2\phi^*(\mu, \nu))(1 - 2\phi^*(\nu, \tau))$.*

We consider the fixed point condition $\mathcal{Q}\mu = \mu$ and any $\phi \in (0, \frac{1}{2})$ satisfying $B_\phi \boxast \mu \preceq \nu$. From the first property in Proposition 14, such ϕ exists. Then from the second property, we can choose $\phi = \phi^*(\mu, \nu)$ unless $\phi^*(\mu, \nu) = 0$. We focus on the non-trivial case where $\mathbb{P}[d \leq 1] < 1$. Otherwise, any fixed point μ has to satisfy $\mu(\{0\}) = 1$, which makes them unique and trivial, unless $\mathbb{P}[d \leq 1] = 1$, which falls into the exceptional case stated in the theorems. Under this condition, Theorem 11 states that there is an integer k for any degree distribution such that

$$B_\phi \boxast \mu \prec \mathcal{Q}^k(B_\phi \boxast \mu).$$

Because $\mathcal{Q}$ describes BP, it preserves degradation. Hence,

$$\mathcal{Q}^k(B_\phi \boxast \mu) \preceq Q^k \nu = \nu.$$

Recall the transitivity property stated in Proposition 12. This implies $B_\phi \boxast \mu \prec \nu$.

However, this conclusion is mutually exclusive with $\phi = \phi^*(\mu, \nu)$ according to the third property in Proposition 14, which states that strict degradation provides strict upper bound on degradation index. Thus, we must have $\phi^*(\mu, \nu) = 0$, and $\mu \preceq \nu$ follows from the second property in Proposition 14.

By symmetry, we have $\nu \preceq \mu$ as well, which in turn implies $\nu = \mu$ since degradation satisfies antisymmetry (see the fourth property in Proposition 48). Therefore, there can be at most one non-trivial symmetric BP fixed point.

We next prove convergence of iterations $\mu_h = \mathcal{Q}^h\mu$ to μ^*, which denotes from now on either the unique non-trivial fixed point (if it exists) or δ_0 (otherwise). First, notice that if $\mu_h \to \mu_\infty$ (weakly) then μ_∞ must be a fixed point of $\mathcal{Q}$. Indeed, we have $\mu_{h+1} = \mathcal{Q}\mu_h$ and taking $h \to \infty$ we get the statement $\mathcal{Q}\mu_\infty = \mu_\infty$ after applying (weak) continuity of $\mathcal{Q}$. The weak continuity follows from the following argument. Let $\{R_{h,u}, u = 1, \dots\}$ be iid $\sim \mu_h$. From Skorokhod's representation we can assume $R_{h,u} \to R_u$ (almost surely) as $h \to \infty$. But then from equation (2) we see that conditioned on any value of degree d we have

$$\sum_{u=1}^{d} Z_u F_\delta(R_{h,u}) \overset{a.s.}{\to} \sum_{u=1}^{d} Z_u F_\delta(R_u)\,,$$

implying that $\mathcal{Q}\mu_h \to \mathcal{Q}\mu_\infty$.

Second, we notice that the sequence $\tilde{\mu}_h = \mathcal{Q}^h B_0$ (i.e. BP iteration initialized from the measure corresponding to perfectly observed leaves: $B_0[+\infty] = 1$) is monotonically decreasing (since it corresponds to a channel from X_0 to X_{L_h}, see Fig. 1). Thus, it is convergent (Prop. 48) and even $\tilde{\mu}_h \to \mu^*$. Indeed, since $B_0 \succeq \mu^*$ we must have $\lim \tilde{\mu}_h \succeq \mu^*$, which implies that the limit cannot be δ_0 unless $\mu^* = \delta_0$.

Third, we notice that for any $\phi \in [0, 1]$, the stringy tree lemma implies that the sequence $\hat{\mu}_h = \mathcal{Q}^h\hat{\mu}$, where $\hat{\mu} = B_\phi \circledast \mu^*$, is monotonically increasing and hence convergent by Prop. 48. Indeed, $\hat{\mu}_1 = \mathcal{Q}\hat{\mu} = \mathcal{Q}(B_\phi \circledast \mu^*) \succeq B_\phi \circledast \mathcal{Q}\mu^* = \hat{\mu}$. Thus applying $\mathcal{Q}^{h-1}$ to both sides we obtain $\hat{\mu}_h \succeq \hat{\mu}_{h-1}$. Further, if $\hat{\mu} \neq \delta_0$ then we can see that the limit point satisfies $\lim_{h\to\infty} \hat{\mu}_h \succeq \hat{\mu}$, which can not be δ_0. Hence, the convergence in this case is always to μ^*.

Finally, we complete the proof. If $\mu^* = \delta_0$ (i.e. no non-trivial fixed points exist) then $\delta_0 \preceq \mu \preceq B_0$ and applying $\mathcal{Q}^h$ we obtain that $\mu_h \to \delta_0$ by sandwiching (cf. Prop. 48). If $\mu^* \neq \delta_0$ and $\mu \neq \delta_0$ then let $\phi < 1/2$ be such that $B_\phi \circledast \mu^* \preceq \mu$ (Prop. 14) and we have $\hat{\mu} \triangleq B_\phi \circledast \mu^* \neq \delta_0$. In this case, $\hat{\mu}_h \preceq \mu_h \preceq \tilde{\mu}_h$ and again the sandwich property shows $\mu_h \to \mu^*$. ☐

Remark 6 (Distance contraction). We have shown that a finite number of applications of the BP operator satisfies the following contractivity condition

$$d(\mathcal{Q}^k\mu, \mathcal{Q}^k\nu) < d(\mu, \nu) \qquad \forall \mu, \nu \neq \delta_0\,,$$

for a metric d between non-trivial symmetric distributions, defined as

$$d(\mu, \nu) = |\ln(1 - 2\phi^*(\mu, \nu)) + \ln(1 - 2\phi^*(\nu, \mu))|. \tag{10}$$

That $d(\mu, \nu) = 0$ implies $\mu = \nu$ is clear, while the triangle inequality follows from Proposition 14 (fourth property). Some properties of this metric are discussed in Appendix A-A. In particular, it is strictly stronger than weak convergence (Levy-Prokhorov) metric.

A common way to prove convergence is to apply a well-known principle of Edelstein, which states that contractive self-map over a compact metric space defines recursions that converge to a unique fixed point [21, Remark 3.2]. However, as Remark 15 in Appendix A shows, the space of symmetric distributions is not compact under the degradation metric, even with a radius constraint. Thus, we proved the convergence of BP recursion directly.

We mention that classically, contraction methods have mostly been studied for linear recursions (i.e. affine combinations of independent variables), see [22]–[24]. Note that setting $S = 0$, $Z_u = c$ (constant) and $F_\delta(r) = r$ in (2) reduces the search for BP fixed points to finding stable laws. The speed of convergence to stable laws (in particular in the central limit theorems) has been studied by constructing special distances such as Zolotarev metrics [25]. (For such metrics contraction properties can be proved by analyzing the norms of the coefficient matrices [24].) Our work may be seen as identifying the appropriate counterpart (degradation distance) for the non-linear model of (2).

A. Comparison to the methods of [2] and [5]

Since partial resolutions of the BP uniqueness were already done in [2], it is natural to ask whether the method here is merely tightening of [2]. Especially since in [5] the range in which uniqueness holds was extended compared to [2] by precisely leveraging channel degradation. We want to argue, however, that the proof here is fundamentally different and explain why the methods as in [2], [5] are inherently tailored to "high-SNR" cases exclusively.

Consider the case of regular trees (i.e., with fixed d) which are key building blocks in the proofs of [2], [5]. The authors investigated contraction properties of potential functions that are either defined in the form of the L_p distance, or can be lower bounded by them. In the regime of $d(1 - 2\delta)^2 > 1$, such potential functions are non-contractive when the LLR distributions are close to the trivial distribution. For instance, taking symmetric Gaussian distributions with vanishing second moments. The application of BP operators can be approximated with $F_\delta(R) \approx (1 - 2\delta)R$, and the distributions are scaled by a factor of $d^{\frac{1}{2}}(1 - 2\delta)$. As a consequence, metrics defined in L_p norms are also increased by the same factor after each BP recursion. Therefore, to exploit any contraction property, the core of the proofs in [2], [5] is to identify cases where the BP recursion is bounded away from the trivial distribution. However, when this condition is imposed, the high SNR requirement emerges as all BP fixed points converge to δ_0 when $d(1 - 2\delta)^2 \to 1^+$.

In this work, we employed a different approach by constructing a new measure of distance between the LLR distributions. In contrast to previous works, where degradation was mainly used to assist the analysis for existing potential functions, we incorporate degradation as a part of the construction. The metric we developed is in some sense "scale-invariant" and allows us to treat the "low–SNR" cases and "high–SNR" cases simultaneously.

B. Extension to Broadcast with Survey

Now we present the proof ideas for non-trivial survey distributions. Formally, we prove the following theorem.

Theorem 15 (Uniqueness with non-trivial survey)**.** *Fix any degree distribution P_d, parameter $\delta \in [0,1]$, and non-trivial symmetric survey distribution μ_{s}. There is exactly one non-trivial symmetric BP fixed point μ^*. For all symmetric μ, the recursion $\mathcal{Q}_{\mathrm{s}}^h \mu$ converges weakly to μ^*.*

The proof of uniqueness relies on the following intermediate step, which is proved in Section III-B.

Theorem 16. *For any $\delta \in [0,1]$, non-trivial symmetric μ_{s}, and degree distribution P_d with $\mathbb{P}[d=0] < 1$, let $\mathcal{Q}_{\mathrm{s}}$ be the associated BP operator. Then for any $\phi \in (0, \frac{1}{2})$ and symmetric μ, we have*

$$B_\phi \boxast \mathcal{Q}_{\mathrm{s}}^k \mu \prec \mathcal{Q}_{\mathrm{s}}^k (B_\phi \boxast \mu) \qquad \text{if } \mathbb{P}[d>1] = 0, \tag{11}$$

$$B_\phi \boxast \mathcal{Q}_{\mathrm{s}} \mu \prec \mathcal{Q}_{\mathrm{s}} (B_\phi \boxast \mu) \qquad \text{otherwise,} \tag{12}$$

for some $k \in \mathbb{N}_+$.

Assuming the correctness of Theorem 16, the proof is obtained as follows.

Proof of Theorem 15. We focus on the non-trivial case where $\mathbb{P}[d=0] < 1$, otherwise, we have $\mathcal{Q}_{\mathrm{s}}\mu = \mu_{\mathrm{s}}$ and Theorem 15 clearly holds. The key observation here is that Theorem 16 plays the same role as Theorem 11 in the no survey case. Hence, by following the same steps in the proof of Theorem 4, one can show that there is at most one non-trivial symmetric BP fixed point. On the other hand, by the monotone convergence property of degradation, the BP recursion with noiseless leaf observation (i.e., $\mu_{\mathrm{r}}(\{+\infty\}) = 1$) converges weakly to a symmetric fixed point. Note that μ_{s} is non-trivial. Any BP fixed point with respect to $\mathcal{Q}_{\mathrm{s}}$ must also be non-trivial. This proves the existence of μ^*. To summarize, there is exactly one unique non-trivial symmetric BP fixed point.

The convergence of $\mu_h = \mathcal{Q}_{\mathrm{s}}^h \mu$ to μ^* can be proved by sandwiching between the BP recursions initialized by B_0 and δ_0. The first sequence $\mathcal{Q}_{\mathrm{s}}^h B_0$ corresponds to the noiseless leaf observation case, which converges to μ^*. The second sequence $\mathcal{Q}_{\mathrm{s}}^h \delta_0$ corresponds to the no leaf observation case (i.e., $\mu_{\mathrm{r}}(\{0\}) = 1$), which is monotonically increasing (since they each corresponds to a channel with an increasing set of surveys), and hence convergent to same the symmetric fixed point (due to uniqueness). Therefore, the sandwiching of Prop. 48 can be applied with the comparison $\mathcal{Q}_{\mathrm{s}}^h \delta_0 \preceq \mu_h \preceq \mathcal{Q}_{\mathrm{s}}^h B_0$, which is due to the natural coupling. ☐

Remark 7. The existence of survey channels significantly affects the properties of BP fixed points. As we have shown, when the survey channels have a non-zero capacity, there is always one unique BP fixed point, and it is non-trivial. However, when survey channels are absent, either the trivial and non-trivial fixed points coexist, or only the trivial solution remains. This difference is also reflected in the statements of the uniqueness theorems.

III. Technical details

In this section, we prove the key intermediate steps, i.e., Theorem 11 and Theorem 16. The gist of the proof is to characterize the commutation rules between the building blocks of the BP operators. In particular, we use the commutativity of box convolution, which is equivalently $F_\phi \circ F_\delta = F_\delta \circ F_\phi$. Then we develop an exchange rule between convolution and box convolution. For convenience, we make the following definition.

Definition 17. For any symmetric distribution μ, we define its *β-curve* as a function on domain $t \in \mathbb{R}$ given by the following equation.

$$\beta(t;\mu) \triangleq \mathbb{E}_{R\sim\mu}\left[\max\left\{\tanh\frac{|R|}{2}, |t|\right\}\right]. \tag{13}$$

We also define

$$t_{\max}(\mu) \triangleq \inf\{t \in [0,1] \mid \beta(t) = t\}.$$

The meaning of the β-curve is given by the next two results (one trivial, one classical).

Proposition 18. *For any $t \in [0,1]$ and symmetric μ, the function value $\frac{1}{2}(1-\beta(t;\mu))$ equals the minimum error probability for a binary hypothesis testing problem with LLR distribution given by μ and prior given by* $\mathrm{Ber}(\frac{1-t}{2})$.

See Appendix B-C for a proof of the first result. The next result is a celebrated Blackwell–Sherman–Stein (BSS) theorem [26]–[29] (an equivalent form is also stated in [20, Theorem 4.74]), which connects degradation and the β-curves.

Theorem 19. *[Blackwell–Sherman–Stein] For any pair of symmetric distributions ν and μ, we have $\nu \preceq \mu$ if and only if $\beta(t;\nu) \le \beta(t;\mu)$ for all $t \in [0,1]$.*

Our principal analytical tool is the following extension of the BSS result to the relation of strict degradation (see Appendix B-D for the proof).

Proposition 20. *For any non-trivial symmetric distributions μ and ν, the following statements are equivalent.*

1) $\nu \prec \mu$.
2) $\beta(t;\nu) < \beta(t;\mu)$ *for all* $t \in [0, t_{\max}(\nu)]$.
3) $\beta(t;\nu) < \beta(t;\mu)$ *for all* $t \in [0, t_{\max}(\mu))$ *and* $t_{\max}(\nu) < t_{\max}(\mu)$.

Together with Prop. 18 we can see that strict degradation between two distributions can be checked by comparing their associated Bayesian inference problems. The proof of Proposition 20 relies on a several elementary steps. One such step is the fact that box convolution with any B_ϕ can be viewed as a homothetic transformation:

Proposition 21. *For any $\phi \in [0, \frac{1}{2})$, the β-curve of B_ϕ is given by*

$$\beta(t; B_\phi) = \max\{|t|, 1-2\phi\}. \tag{14}$$

More generally, for any symmetric μ,

$$\beta(t; B_\phi \boxast \mu) = (1-2\phi)\beta\left(\frac{t}{1-2\phi}; \mu\right). \tag{15}$$

Proof. Equation (14) directly follows from the definition of β-curves. Then equation (15) can be proved from the second definition of box convolution. □

In order to prove the needed β-curve gaps for Theorem 11 and Theorem 16, we define the following refinement of strict degradation.

Definition 22. For any symmetric distributions μ, ν, and parameter s, we define $\nu \prec_s \mu$ if $\beta(t;\nu) < \beta(t;\mu)$ for all t satisfying

$$t \in \left(\tanh\left(\frac{s}{2}\right), t_{\max}(\nu)\right]$$

and $\beta(t;\nu) \le \beta(t;\mu)$ for all other t. We also define

$$r_{\max}(\mu) \triangleq \sup\{r \in [0, +\infty] \mid \mu([r, +\infty]) > 0\} = 2\tanh^{-1} t_{\max}(\mu).$$

The following propositions are proved in Appendix B-E and Appendix B-F, respectively.

Proposition 23. *Let μ, ν, τ be symmetric distributions.*

1) *We have $\nu \prec_s \mu$ if $\beta(t;\nu) < \beta(t;\mu)$ for all $t \in \left(\tanh\left(\frac{s}{2}\right), t_{\max}(\mu)\right)$, $\beta(t;\nu) \le \beta(t;\mu)$ for all other t, and $t_{\max}(\nu) < t_{\max}(\mu)$; when $s < r_{\max}(\mu)$, the converse also holds true.*
2) *If $\nu \preceq \tau \prec_s \mu$ or $\nu \prec_s \tau \preceq \mu$, then $\nu \prec_s \mu$.*
3) *$\nu \prec \mu$ implies $\nu \prec_s \mu$ for any $s \ge -r_{\max}(\mu)$, the latter implies $\nu \preceq \mu$.*
4) *$\nu \prec_s \mu$ is always true when $s \ge r_{\max}(\nu)$ and $\nu \preceq \mu$; if $\nu \prec_s \mu$ for any $s < 0$, then $\nu \prec \mu$.*

Proposition 24. *For any $\phi \in (0,1)$ and $\delta_1, \delta_2 \in [0,1]$, let $s_{\min} = |F_\phi(r_{\max}(B_{\delta_1})) - F_\phi(r_{\max}(B_{\delta_2}))|$. Then we have*

$$B_\phi \boxast (B_{\delta_1} * B_{\delta_2}) \prec_{s_{\min}} (B_\phi \boxast B_{\delta_1}) * (B_\phi \boxast B_{\delta_2}). \tag{16}$$

The β-curves for the distributions on both sides of inequality (16) are piecewise linear. Moreover, the proof in Appendix B-F provides the exact condition for the inequality between β-curves to be strict. By averaging over δ_1, δ_2 the β-curves in the above special case, one can prove the following corollary (see Appendix B-G for details).

Corollary 25. *For any $\phi \in (0,1)$ and any two symmetric distributions μ_1, μ_2, we have*

$$B_\phi \boxast (\mu_1 * \mu_2) \prec_s (B_\phi \boxast \mu_1) * (B_\phi \boxast \mu_2) \tag{17}$$

for

$$s = |F_\phi(r_{\max}(\mu_1)) - F_\phi(r_{\max}(\mu_2))|.$$

Now we establish connections between β-curve and convolution.

Proposition 26. *For any $\phi \in [0, \frac{1}{2}]$, $t \in [0, 1)$, and symmetric μ, we have*

$$\beta(t; B_\phi * \mu) = \left(\frac{1+t-2t\phi}{2}\right)\beta(t_0; \mu) + \left(\frac{1-t+2t\phi}{2}\right)\beta(t_1; \mu),$$

where

$$t_0 \triangleq \frac{1+t-2\phi}{1+t-2t\phi} = \tanh\left(\frac{r_{\max}(B_\phi)}{2} + \tanh^{-1}(t)\right),$$
$$t_1 \triangleq \left|\frac{1-t-2\phi}{1-t+2t\phi}\right| = \tanh\left(\left|\frac{r_{\max}(B_\phi)}{2} - \tanh^{-1}(t)\right|\right).$$

Proposition 26 is proved in Appendix B-H. By averaging the β-curves in Proposition 26 over ϕ, we have the following "rule of convolution", which is proved in Appendix B-I.

Definition 27. For any symmetric μ, define $\text{supp}(\mu) \triangleq \{v \in \mathbb{R} \mid \mu([v-\epsilon, v+\epsilon]) > 0 \ \forall \epsilon > 0\}$.

Proposition 28 (Rule of Convolution)**.** *For any symmetric μ, ν, τ, and $\ell \in \text{supp}(\tau)$, if $\nu \prec_s \mu$, then*

$$\beta\left(\tanh\frac{r}{2}; \nu * \tau\right) < \beta\left(\tanh\frac{r}{2}; \mu * \tau\right) \tag{18}$$

for any $r \in (s-\ell, r_{\max}(\mu)-\ell)$; if $\nu \prec \mu$, then inequality (18) holds for $r \in [-\ell, r_{\max}(\mu)-\ell)$.

Notice that the above β-curve analysis can be used to show the expected convolution-preserving properties of degradation. We extend them to $\prec_s$ and $\prec$, stated as follows.

Proposition 29. *For any symmetric μ, ν, and τ, we have the following facts.*

1) *If $\nu \prec \mu$, then $\tau \boxast \nu \prec \tau \boxast \mu$.*
2) *If $\nu \prec_s \mu$, then $\tau \boxast \nu \prec_{s_\tau} \tau \boxast \mu$, where $s_\tau \triangleq \ln \frac{e^s e^{r_{\max}(\tau)}+1}{e^s + e^{r_{\max}(\tau)}}$.*
3) *If $\nu \prec \mu$, then unless $r_{\max}(\tau) = r_{\max}(\mu) = +\infty$, we have $\tau * \nu \prec_s \tau * \mu$ with $s = r_{\max}(\tau) - r_{\max}(\mu)$; if in addition, $r_{\max}(\tau) < r_{\max}(\mu)$, then $\tau * \nu \prec \tau * \mu$.*

Proof. The first two properties are based on the fact that box convolution with any symmetric τ can be viewed as a mixture of homothetic transformations in β-curve (see Proposition 21). Hence, any strict inequality is preserved, except when τ is trivial, where the proof is straightforward.

The third property can be proved with Proposition 28. Specifically, for the non-trivial case where $r_{\max}(\tau) < +\infty$, we can choose $\ell = \pm r_{\max}(\tau)$ and obtain that the β-curve inequalities are strict for $t \in [-r_{\max}(\tau), r_{\max}(\mu) - r_{\max}(\tau)) \cup [r_{\max}(\tau), r_{\max}(\mu) + r_{\max}(\tau))$. Note that β-curves are even functions. We have the strict condition for any $t \in (r_{\max}(\tau) - r_{\max}(\mu),\ r_{\max}(\mu) + r_{\max}(\tau))$. Then we can apply Proposition 23 to complete the proof. □

The above results can be used to prove the following statements.

Proposition 30. *For any $\phi \in (0, 1)$ and any symmetric ν, we have*

$$B_\phi \boxast (\nu^{*(2)}) \prec_0 (B_\phi \boxast \nu)^{*(2)}, \tag{19}$$
$$B_\phi \boxast (\nu^{*(d)}) \prec (B_\phi \boxast \nu)^{*(d)} \qquad \forall\ d > 2. \tag{20}$$

Proof. Inequality (19) directly follows from Corollary 25. Inequality (20) clearly holds when $\phi = \frac{1}{2}$ or ν is trivial. We prove inequality (20) by induction for both B_ϕ and ν are non-trivial.

(a). Consider the base case where $d = 3$. Let $r_\nu = r_{\max}(\nu)$. Because convolution preserves degradation, by stringy tree lemma, or inequality (19) and Proposition 23, we have

$$(B_\phi \boxast \nu) * (B_\phi \boxast (\nu^{*(2)})) \preceq (B_\phi \boxast \nu)^{*(3)}. \tag{21}$$

Next, from Corollary 25, we have

$$B_\phi \boxast (\nu^{*(3)}) \prec_s (B_\phi \boxast \nu) * (B_\phi \boxast (\nu^{*(2)})), \tag{22}$$

where

$$s = F_\phi(r_{\max}(\nu^{*(2)})) - F_\phi(r_{\max}(\nu)) = F_\phi(2r_\nu) - F_\phi(r_\nu).$$

The above two steps form a chain of degradation. By transitivity stated in Proposition 23, this implies

$$B_\phi \boxast (\nu^{*(3)}) \prec_s (B_\phi \boxast \nu)^{*(3)}.$$

To prove the needed statement, it suffices to show that any of these two steps has strict inequality in β-curves for $t \in [0, s]$. We apply Rule of Convolution to inequality (21) and let $\ell = F_\phi(r_\nu) = r_{\max}(B_\phi \boxast \nu) \in \operatorname{supp}(B_\phi \boxast \nu)$, the strict condition holds for

$$t \in \left[0, \tanh \frac{r_{\max}((B_\phi \boxast \nu)^{*(2)}) - \ell}{2}\right) = \left[0, \tanh \frac{F_\phi(r_\nu)}{2}\right).$$

Because $F_\phi(r_\nu) > s$ for both B_ϕ and ν non-trivial, the strict degradation statement is implied by Proposition 20.

(b) Assume inequality (20) is proved for some $d = d_0 \geq 3$. By induction assumption, we have the following chain similar to the base case.

$$B_\phi \boxast (\nu^{*(d_0+1)}) \preceq (B_\phi \boxast \nu) * B_\phi \boxast (\nu^{*(d_0)}) \prec (B_\phi \boxast \nu)^{*(d_0+1)}.$$

In particular, we apply the third property in Proposition 29 to show strict degradation in the second step. Therefore, the induction step follows from Proposition 12.

(c) To conclude, inequality (20) is proved for any $d \geq 3$. ☐

Proposition 30 implies the following degradation relationships between $B_\phi \boxast (\mathcal{Q}\mu)$ and $\mathcal{Q}(B_\phi \boxast \mu)$ when d is deterministic.

Corollary 31. *If $\mathbb{P}[d = d_0] = 1$ for some fixed d_0, then for any $\phi \in (0, \frac{1}{2})$ and symmetric μ, we have*

$$B_\phi \boxast (\mathcal{Q}\mu) \prec_0 \mathcal{Q}(B_\phi \boxast \mu) \qquad \text{if } d_0 = 2, \tag{23}$$

$$B_\phi \boxast (\mathcal{Q}\mu) \prec \mathcal{Q}(B_\phi \boxast \mu) \qquad \text{if } d_0 > 2. \tag{24}$$

Proof. Recall that $\mathcal{Q}$ can be expressed as in Proposition 9. The results in the corollary can be exactly obtained by letting $\nu = B_\delta \boxast \mu$ in Proposition 30 and applying the commutativity of box convolution. ☐

A. Proof of Theorem 11

Proof. For brevity, we focus on non-trivial cases where μ is non-trivial and $\delta \neq \frac{1}{2}$. We first consider the deterministic d case and fill in the gap for $d = 2$. Note that $\mathcal{Q}$ preserves degradation. By Corollary 31 and Proposition 23, we have the following chain of degradation.

$$B_\phi \boxast \mathcal{Q}^2\mu \prec_0 \mathcal{Q}(B_\phi \boxast \mathcal{Q}\mu) \preceq \mathcal{Q}^2(B_\phi \boxast \mu).$$

In particular, the first step is obtained by replacing μ with $\mathcal{Q}\mu$ in Corollary 31. The above chain implies strict inequality in β-curves for $t \in (0, t_{\max}(B_\phi \boxast \mathcal{Q}^2\mu)]$. By non-trivial condition, from Proposition 20, it remains to prove strict inequality of β-curves at $t = 0$.

To that end, we zoom in on the second step and apply Rule of Convolution to the first inequality of the following chain.

$$\mathcal{Q}(B_\phi \boxast \mathcal{Q}\mu) = (B_\phi \boxast B_\delta \boxast \mathcal{Q}\mu)^{*(2)} \tag{25}$$

$$\preceq (B_\phi \boxast B_\delta \boxast \mathcal{Q}\mu) * (B_\delta \boxast \mathcal{Q}(B_\phi \boxast \mu)) \tag{26}$$

$$\preceq (B_\delta \boxast \mathcal{Q}(B_\phi \boxast \mu))^{*(2)} = \mathcal{Q}^2(B_\phi \boxast \mu). \tag{27}$$

Note that Corollary 31 and Proposition 29 implies

$$B_\phi \boxast B_\delta \boxast \mathcal{Q}\mu \prec_0 B_\delta \boxast \mathcal{Q}(B_\phi \boxast \mu),$$

and the non-trivial condition implies that the $r_{\max}$ functions for both sides are different. Therefore, we can choose $\ell = r_{\max}(B_\phi \boxast B_\delta \boxast \mathcal{Q}\mu)$ for the Rule of Convolution and apply it to inequality (26), which leads to the needed strict condition at $t = 0$.

Now we consider general degree distributions. First for $\mathbb{P}[d > 2] = 0$, recall that our formulation assumes non-trivial cases where $\mathbb{P}[d \leq 1] < 1$. We have $d = 2$ with non-zero probability. Then the $r_{\max}$ function of $B_\phi \boxast (\mathcal{Q}^2\mu)$ is identical to that of its $d = 2$ component. Thus, by linearity, our earlier proof for the deterministic $d = 2$ case implies strict inequality of β-curves for the full range $t \in [0, t_{\max}(B_\phi \boxast (\mathcal{Q}^2\mu))]$, and the needed statement is implied.

On the other case, we have $\mathbb{P}[d > 2] > 0$. If d is upper bounded by some fixed integer almost surely, we can let d_0 be the largest possible d for such degree distribution and apply the same linearity argument to inequality (24) to prove the statement. Otherwise, d is unbounded, and we have strict inequality on β-curves for any $t < t_{\max}(\mathcal{Q}(B_\phi \boxast \mu)) = 1$. Note that $t_{\max}(B_\phi \boxast (\mathcal{Q}^2\mu) = 1 - 2\phi < 1$. The statement follows from Proposition 23. ☐

B. Proof of Theorem 16

We start by formulating two useful results.

Proposition 32. *For any $\phi \in (0,1)$ and any symmetric distributions μ, ν, let $s_{\min} \triangleq F_\phi(r_{\max}(\mu)) - r_{\max}(\nu)$. We have*

$$B_\phi \circledast (\mu * \nu) \prec (B_\phi \circledast \mu) * \nu \tag{28}$$

if $s_{\min} < 0$, and

$$B_\phi \circledast (\mu * \nu) \prec_{s_{\min}} (B_\phi \circledast \mu) * \nu \tag{29}$$

otherwise.

Proof. Similar to the basic setting, our technique is to prove strict inequalities for beta-curves by forming a chain of degradation. First observe that $B_\phi \circledast \nu \prec \nu$ for any $\phi \in (0,1)$. By convolving $B_\phi \circledast \mu$ on both sides and note that degradation is convolution-preserving, we have

$$(B_\phi \circledast \mu) * (B_\phi \circledast \nu) \preceq (B_\phi \circledast \mu) * \nu. \tag{30}$$

On the other hand, Corollary 25 implies following step, which completes the chain.

$$B_\phi \circledast (\mu * \nu) \preceq (B_\phi \circledast \mu) * (B_\phi \circledast \nu) \tag{31}$$

When $s_{\min} < 0$, we have $r_{\max}(B_\phi \circledast \mu) = F_\phi(r_{\max}(\mu)) < r_{\max}(\nu)$. Hence, we can apply the third statement in Proposition 29 to obtain a strict version of inequality (30), i.e., $(B_\phi \circledast \mu) * (B_\phi \circledast \nu) \prec (B_\phi \circledast \mu) * \nu$. Then, inequality (28) follows from the transitivity statement in Proposition 12.

More generally, the strict version of inequality (30) can be written as $(B_\phi \circledast \mu) * (B_\phi \circledast \nu) \prec_{s_{\min}} (B_\phi \circledast \mu) * \nu$ according to Proposition 29. Then, inequality (29) is proved by the second statement in Proposition 23. ☐

Next, consider the case of $d = 1$, so that $\mathcal{Q}_{\mathrm{s}}\mu = (B_\delta \circledast \mu) * \mu_{\mathrm{s}}$. By induction, one can derive the following result (proof in Appendix B-J).

Proposition 33. *For $\phi \in (0,1)$ and $d = 1$, let $r_{\mathrm{s}} \triangleq r_{\max}(\mu_{\mathrm{s}})$, $s_0 \triangleq F_\phi(r_{\max}(\mu))$, then for any $k \in \mathbb{N}$, we have*

$$B_\phi \circledast (\mathcal{Q}_{\mathrm{s}}^k \mu) \prec \mathcal{Q}_{\mathrm{s}}^k (B_\phi \circledast \mu)$$

if $s_k \triangleq F_\delta(s_{k-1}) - r_{\mathrm{s}} < 0$, and

$$B_\phi \circledast (\mathcal{Q}_{\mathrm{s}}^k \mu) \prec_{s_k} \mathcal{Q}_{\mathrm{s}}^k (B_\phi \circledast \mu)$$

otherwise.

One can show that there is a finite k for $s_k < 0$. For example, as a rough estimate, we have $s_0 \le F_\phi(+\infty) < +\infty$ and $s_k \le s_{k-1} - r_{\mathrm{s}}$ for any positive s_k. Because μ_{s} is non-trivial, we also have $r_{\mathrm{s}} > 0$. Hence, we can find $k \le 1 + s_0 / r_{\mathrm{s}}$ for strict degradation to hold, which gives the needed statement.

Proof of Theorem 16. First, consider the case $\mathbb{P}[d > 1] = 0$. Recall that the theorem statement assumes $\mathbb{P}[d = 0] < 1$, so that we have $d = 1$ with non-zero probability. Consequently, the β-curve analysis is dominated by the $d = 1$ component, and strict inequalities in the full range of t can be obtained using Proposition 33.

Formally, $\mathcal{Q}_{\mathrm{s}}$ can be written as a linear combination of two operators, each corresponds to the BP operator for a deterministic $d \in \{0,1\}$. For brevity, we denote them by $\mathcal{Q}_{\mathrm{s},0}$ and $\mathcal{Q}_{\mathrm{s},1}$. Then, $\mathcal{Q}_{\mathrm{s}}^k$ can be expanded into a linear combination of 2^k chains, and each side of inequality (11) can be decomposed into the following 2^k terms.

$$\sum_{d_1,\ldots,d_k \in \{0,1\}} B_\phi \circledast \mathcal{Q}_{\mathrm{s},d_1} \mathcal{Q}_{\mathrm{s},d_2} \ldots \mathcal{Q}_{\mathrm{s},d_k} \mu \cdot \prod_{j=1}^k \mathbb{P}[d = d_j] \prec \sum_{d_1,\ldots,d_k \in \{0,1\}} \mathcal{Q}_{\mathrm{s},d_1} \mathcal{Q}_{\mathrm{s},d_2} \ldots \mathcal{Q}_{\mathrm{s},d_k} (B_\phi \circledast \mu) \cdot \prod_{j=1}^k \mathbb{P}[d = d_j]$$

Each corresponding terms in the above inequality can be compared individually. Recall that $\mathcal{Q}_{\mathrm{s}}\mu = (\mathcal{Q}\mu) * \mu_{\mathrm{s}}$. By Corollary 31 and Proposition 32, the inequality $B_\phi \circledast (\mathcal{Q}_{\mathrm{s}}\mu) \preceq \mathcal{Q}_{\mathrm{s}}(B_\phi \circledast \mu)$ holds for any BP operator $\mathcal{Q}_{\mathrm{s}}$, which includes $\mathcal{Q}_{\mathrm{s},0}$ and $\mathcal{Q}_{\mathrm{s},1}$. By applying this inequality recursively, we have the following individual comparisons.

$$B_\phi \circledast \mathcal{Q}_{\mathrm{s},d_1} \mathcal{Q}_{\mathrm{s},d_2} \ldots \mathcal{Q}_{\mathrm{s},d_k} \mu \preceq \mathcal{Q}_{\mathrm{s},d_1} \mathcal{Q}_{\mathrm{s},d_2} \ldots \mathcal{Q}_{\mathrm{s},d_k} (B_\phi \circledast \mu)$$

Among all terms, the ones with all BP operators corresponds to $d = 1$ achieves strict degradation due to Proposition 33. Note that these terms have the largest $t_{\max}$ values on both sides, because $t_{\max}(\mathcal{Q}_{\mathrm{s},d}\nu)$ is non-decreasing with respect to d and $t_{\max}(\nu)$ for any symmetric ν. The interval $t \in [0, t_{\max}(B_\phi \circledast (\mathcal{Q}_{\mathrm{s}}^k \mu))]$ must be contained within the range where the β-curve inequality between these two terms is strict. The $\mathbb{P}[d = 0] < 1$ condition ensures that this gap has non-zero weights in the overall β functions. Then inequality (11) is proved by Proposition 20.

It remains to consider general degree distributions with $\mathbb{P}[d > 1] > 0$. Note that in the basic setting, we have essentially proved that if $\mathbb{P}[d > 1] > 0$, then

$$B_\phi \circledast \mathcal{Q}\mu \prec_0 \mathcal{Q}(B_\phi \circledast \mu). \tag{32}$$

Specifically, the above inequality is directly implied by Theorem 11 if $\mathbb{P}[d > 2] > 0$. In the other case, we have that the $r_{\max}$ function of $B_\phi \circledast \mathcal{Q}\mu$ is dominated by its $d = 2$ component. Thus, inequality (23) implies non-zero gaps in β-curves for all $t \in (0, t_{\max}(B_\phi \circledast \mathcal{Q}\mu)]$, which proves inequality (32).

Let $r_{\rm s} = r_{\max}(\mu_{\rm s})$, $r_Q = r_{\max}(\mathcal{Q}\mu)$, $\tilde{r}_Q = r_{\max}(\mathcal{Q}(B_\phi \circledast \mu))$. We first apply Proposition 32 and then the Rule of Convolution to obtain

$$B_\phi \circledast \mathcal{Q}_{\rm s}\mu \preceq (B_\phi \circledast \mathcal{Q}\mu) * \mu_{\rm s} \preceq \mathcal{Q}_{\rm s}(B_\phi \circledast \mu). \tag{33}$$

Consider the first step of inequality (33), the statement of Proposition 32 implies that the gap between the β-curves on both sides is strict for any $t = \tanh \frac{|s|}{2}$ with

$$F_\phi(r_Q) - r_{\rm s} < s < F_\phi(r_Q) + r_{\rm s}. \tag{34}$$

Then by the rule of convolution, the β-curve inequality for the second step is strict for $t = \tanh \frac{|s|}{2}$ with

$$-r_{\rm s} < s < \tilde{r}_Q - r_{\rm s}. \tag{35}$$

Note that inequality (32) implies that $\tilde{r}_Q > r_{\max}(B_\phi \circledast \mathcal{Q}\mu) = F_\phi(r_Q)$. Thus, (34) and (35) cover all $0 \leq s < F_\phi(r_Q) + r_{\rm s}$, concluding the proof of

$$B_\phi \circledast \mathcal{Q}_{\rm s}\mu \prec \mathcal{Q}_{\rm s}(B_\phi \circledast \mu). \tag{36}$$

□

IV. Implications

A. Implications on Robust Reconstruction

A variant of BOT was formulated in [17], called robust reconstruction, where all leaf observations are obtained through some identical noisy channels. The estimation problem is to infer the root variable given the tree structure and the noisy leaf observations. Robust reconstruction for the Ising model case was studied in [2], which can be described as in the setting of Fig. 1, except that we select $\mu_{\rm s}$ to be the trivial distribution δ_0. As derived earlier, the LLR distributions for this formulation at depth h is exactly given by $\mathcal{Q}^h\mu_{\rm r}$, where $\mathcal{Q}$ is the BP operator defined in Section I.

Theorem 4 (or Corollary 3 for the asymmetric case) implies the following statement.

Theorem 34. *For any fixed P_d and $\delta \in [0, 1]$, the distributions in the following classes all exist and are identical, unless $d = 1$ a.s. and $\delta \in \{0, 1\}$.*

(a) The limiting LLR distribution for the basic setting.

(b) The limiting LLR distribution for robust construction with any non-trivial initialization.

(c) The dominant BP fixed point, i.e., a fixed point μ^* of $\mathcal{Q}$ where any other fixed point μ satisfies $\mu \preceq \mu^*$.

Remark 8. In [2], it was conjectured that the error probability for the maximum likelihood estimator is independent of the observation channels when $h \to \infty$, as long as their channel capacity is non-zero. Note that this error probability can be written as an expectation over the LLR distribution (see Proposition 18). The unique convergence stated in Theorem 34 provides a positive proof to this conjecture. More generally, the same guarantee holds for any quantity that can be written as the expectation of a bounded continuous function on $(-\infty, +\infty]$, such as mutual information and Bayesian estimation errors under different prior distributions. Theorem 34 also provides a proof of Proposition 1 in [7].

B. Boundary Irrelevance for Broadcast with Survey

We first present a definition of boundary irrelevance in terms of LLR distributions.

Definition 35. For any degree distribution P_d, $\delta \in [0, 1]$, and symmetric non-trivial survey distribution $\mu_{\rm s}$, let $\mathcal{Q}_{\rm s}$ be the associated BP operator. We say boundary irrelevance (BI) is satisfied if both $\mu_{(h)} \triangleq \mathcal{Q}_{\rm s}^h B_0$ and $\underline{\mu}_{(h)} \triangleq \mathcal{Q}_{\rm s}^h B_{\frac{1}{2}}$ weakly converges to the same distribution on domain $(-\infty, +\infty]$ as $h \to \infty$.

Note that $\mu_{(h)}$ represents the LLR distribution for estimation with full leaf information, and $\underline{\mu}_{(h)}$ represents the corresponding LLR distribution with no leaf information (see Fig. 1). The above definition essentially states that ignoring leaf information will not affect estimation as $h \to \infty$, which is consistent with the notation of BI defined in the literature. In particular, one can show that our definition is equivalent to the version in [5], and is stronger than the error-probability guarantee in [3]. Therefore, we present the following Theorem, which is a direct consequence of Theorem 15, which simultaneously resolves Conjecture 1 in [5] and Conjecture 1 in [3].

Theorem 36. *BI holds for any combination of P_d, $\delta \in [0, 1]$, and symmetric non-trivial $\mu_{\rm s}$.*

C. Uniqueness and Convergence in the Large d Limit

We consider a recursion process characterized by the following operator $\mathcal{Q}_{\mathrm{L}}$: For any symmetric μ_{s} and any distribution $P_{\overline{d}}$ on domain $[0,+\infty]$, we set

$$\mathcal{Q}_{\mathrm{L}}\mu = \mathbb{E}_{\overline{d}\sim P_{\overline{d}}}\left[\mathcal{N}\left(\overline{d}\cdot V_\mu\right)\right] * \mu_{\mathrm{s}},$$

where $\mathcal{N}(s) \triangleq \mathcal{N}(\frac{s}{2}, s)$ for any $s\in[0,+\infty]$, $V_\mu \triangleq \mathbb{E}_{R\sim\mu}\left[4\tanh\left(\frac{R}{2}\right)\right]$ for any symmetric μ, and $\mathbb{E}_{\overline{d}\sim P_{\overline{d}}}[\cdot]$ represents a mixture of distributions over the law of $\bar{d}$. This operator was considered in [4] as a limit of $\mathcal{Q}_{\mathrm{s}}$ (or $\mathcal{Q}$ when μ_{s} is trivial) for $\delta\to\frac{1}{2}$, where the degree distribution P_d is parameterized by δ, and $\overline{d}\triangleq d(1-\delta)^2$ converges in distribution to $P_{\overline{d}}$ on domain $[0,+\infty]$. Similar to earlier sections, one can define the fixed point equation to be $\mu=\mathcal{Q}_{\mathrm{L}}\mu$ and define BP recursion as $\mu_{(h+1)}=\mathcal{Q}_{\mathrm{L}}\mu_{(h)}$. The operator $\mathcal{Q}_{\mathrm{L}}$ can also be defined for asymmetric distributions by setting $V_\mu=\mathbb{E}_{R\sim\frac{1}{2}(\mu+\mu^-)}[4\tanh^2\left(\frac{R}{2}\right)]$, where μ^- is the complement of μ (see Definition 2).

To extend our earlier results to $\mathcal{Q}_{\mathrm{L}}$, we prove its contractivity in terms of the degradation index. In particular, note that the contraction implied by the BP operator is non-multiplicative, a careful investigation is needed to show that strict inequalities in β-curves are maintained in the limit of large d. We present the results in the following theorem, and provide a proof in Appendix D.

Theorem 37. *Consider the large d regime defined by any $P_{\overline{d}}$ and any symmetric μ_{s}.*

1) *There is at most one unique non-trivial BP fixed point, and it is symmetric. (Uniqueness of non-trivial BP fixed point)*
2) *Non-trivial symmetric BP fixed point exists if and only if either $\mathbb{E}\left[\overline{d}\right]\in(1,+\infty]$ or μ_{s} is non-trivial. (Existence of non-trivial BP fixed point)*
3) *BP recursion with any non-trivial initialization converges to the unique non-trivial fixed point if it exists, or to the trivial δ_0 otherwise. (Independence of Convergence and Initialization)*
4) *When μ_{s} is non-trivial, the above convergence statement also applies for the trivial initialization. (Boundary Irrelevance)*

Remark 9. The uniqueness statement in the above theorem resolves Conjecture 2.6 in [4], by applying the special case where $P_{\overline{d}}$ is a delta distribution and $\mu_{\mathrm{s}}=B_\alpha$. Generally, our formulation does not assume d scales with $(1-2\delta)^{-2}$ in high probability. The sublinear and superlinear components of P_d are naturally captured by non-zero mass points in $P_{\overline{d}}$ at 0 and $+\infty$.

D. Full Characterizations of Accuracy and Entropy in Stochastic Block Model

Consider a 2-SBM problem with a set of n vertices $\mathcal{V}=\{v_1,v_2,...,v_n\}$. Let X_v denote the label on vertex v, and $X=(X_{v_1},X_{v_2},...,X_{v_n})$ denote the collection of all labels. The entries of X are i.i.d. Ber($\frac{1}{2}$). A random graph G is generated based on the labels according to the rules of SBM. Formally, we represent G using its adjacency matrix, i.e., $G_{ij}=1$ if and only if v_i and v_j are adjacent. Then all $\{G_{ij}\}_{i\le j}$ are independent Bernoulli random variables, with

$$\mathbb{P}[G_{ij}=1]=\begin{cases}\frac{a}{n} & \text{if } X_{v_i}=X_{v_j} \text{ and } i\neq j,\\ \frac{b}{n} & \text{if } X_{v_i}\neq X_{v_j},\\ 0 & \text{otherwise.}\end{cases}$$

The goal in this setting is to design algorithms that use the random graph to produce an estimate of X.

There are two main quantities of interests. For any estimator $\hat{X}$, its *estimation accuracy*, denoted by $\mathrm{acc}_n(\hat{X})$, is defined as follows [2].

$$\mathrm{acc}_n(\hat{X})\triangleq\frac{1}{2}+\left|\frac{1}{n}\sum_{v\in\mathcal{V}}\left|X_v-\hat{X}_v\right|-\frac{1}{2}\right|. \tag{37}$$

In particular, note that the conditioned graph distribution is invariant under a global bit flip of hidden labels. No algorithm can achieve a non-trivial estimation in the expected number of correctly estimated labels. The accuracy defined in equation (37) captures the correlation between the partitions induced by the labels, which removes the global bit-flip effect.

Note that $\mathrm{acc}_n(\hat{X})$ is random. The quantity $p_{\mathrm{G}}(a,b)$ was introduced in [2] to measure the performance of estimators, defined as the maximum accuracy that can be achieved by any estimator for large n with non-zero probability. Formally, let f denote the function that takes the observations and returns $\hat{X}$, $p_{\mathrm{G}}(a,b)$ is defined as follows.

$$p_{\mathrm{G}}(a,b)=\lim_{\epsilon\to 0}\limsup_{n\to\infty}\sup_{f}\sup\left\{p\;\middle|\;\mathbb{P}\left[\mathrm{acc}_n(\hat{X})\geq p\right]\geq\epsilon\right\}.$$

The problems of interests are to characterize $p_{\mathrm{G}}(a,b)$ and to prove whether it can be achieved by any algorithm with high-probability. Both were only resolved when a and b satisfy certain conditions. However, with the leaf-independence result proved in Section IV-A, the proofs in [2], [19], [30] can be extended to all regimes.

The other quantity of interest is the so called *SBM entropy*, denoted by $\mathcal{H}(a,b)$, which is defined to be the limit of the normalized conditional entropy of all labels X given the graph G, as $n \to \infty$:

$$\mathcal{H}(a,b) \triangleq \lim_{n\to\infty} \frac{1}{n} H(X|G).$$

The SBM entropy also characterizes the normalized mutual information between the labels and the graph defined as

$$\mathcal{I}(a,b) \triangleq \lim_{n\to\infty} \frac{1}{n} I(X;G)\,.$$

Similar to the accuracy metric, it was an open problem to characterize $\mathcal{H}(a,b)$ and $\mathcal{I}(a,b)$ for all parameter values using BP fixed points. It was pointed out in [5] that a complete characterization can be obtained once the BI result stated in Section IV-B is proved.

To summarize, we have the following theorem, which strengthens Theorem 2.9 in [2] and Theorem 1 in [5].

Theorem 38. *For any a and b,*

1) *we have $p_{\mathrm{G}}(a,b) = p(\mu^*_{a,b})$, where $\mu^*_{a,b}$ is the dominant BP fixed point (as specified in Theorem 34) for the broadcast on tree problem with $P_d = \mathrm{Pois}(\frac{a+b}{2})$ and $\delta = \frac{b}{a+b}$, and $p(\mu^*_{a,b}) \triangleq \mu^*_{a,b}((0,+\infty]) + \frac{1}{2}\mu^*_{a,b}(\{0\})$;*
2) *there is a polynomial time algorithm that achieves $p_{\mathrm{G}}(a,b)$ with high probability, i.e., with $\mathrm{acc}_n(\hat{X})$ converges in probability to $p_{\mathrm{G}}(a,b)$ as $n \to \infty$;*
3) *we have $\mathcal{H}(a,b) = \log 2 - \mathcal{I}(a,b) = \int_0^1 \mathbb{E}_{R\sim\mu_{\epsilon,a,b}}[\log(2\cosh\frac{R}{2}) - \frac{R}{2}\tanh(\frac{R}{2})]d\epsilon$, where $\mu_{\epsilon,a,b}$ is the unique BP fixed point for the broadcast with survey setting with $P_d = \mathrm{Pois}(\frac{a+b}{2})$, $\delta = \frac{b}{a+b}$, and BEC survey $\mu_{\mathrm{s}} = \epsilon B_0 + (1-\epsilon)B_{\frac{1}{2}}$.*

E. Stochastic Block Model with Side Information

Consider a variant of the 2-SBM formulation, where the estimator has additional access to a noisy version of all hidden labels. Similar to the broadcast with survey setting, each label is observed through an independent symmetric channel, and we denote their LLR distribution by μ_{s}.

In the presence of this side information, a different notion of accuracy was considered in the literature. In [3], [4], the authors considered estimators that asymptotically maximizes the expected fraction of correctly estimated labels. Formally, we denote this function by $p_n(\hat{X})$, which can be defined by the following equation.

$$p_n(\hat{X}) \triangleq \mathbb{E}\left[\frac{1}{n}\sum_{v\in\mathcal{V}}\left|X_v - \hat{X}_v\right|\right] = \frac{1}{n}\sum_{v\in\mathcal{V}} \mathbb{P}\left[X_v = \hat{X}_v\right]. \tag{38}$$

The estimation accuracy defined in equation (38) can be maximized by applying the ML estimator individually for each X_v. However, the ML estimator becomes computationally intractable when the graph is large as it relies on global information. Therefore, local algorithms have been studied, and they have been conjectured to be optimal [3], [4]. In particular, for any fixed parameter $t \in \mathbb{N}_+$, an algorithm is called t-local if it estimates each X_v only using the information within the subgraph induced by vertices with a distance from v less than t. Such local information resembles the distribution of local observation in the broadcast with survey setting as $n \to \infty$ for fixed a and b, up to a graph isomorphism. Hence, one can estimate each X_v using the same belief propagation whenever the graph is locally tree-like.

Local BP is asymptotically optimal among local algorithms. We present the following theorem, which states that there are no gaps between the estimation accuracies of local and global algorithms.[3]

Theorem 39. *For 2-SBM with any fixed a, b and side information generated based on any non-trivial μ_{s}, we have*

$$\lim_{t\to\infty}\lim_{n\to\infty} p_n(\hat{X}^{(t)}_{\mathrm{BP}}) = \lim_{n\to\infty} p^*_n, \tag{39}$$

where $\hat{X}^{(t)}_{\mathrm{BP}}$ is any estimator that runs local BP with parameter $\delta = \frac{b}{a+b}$, and p^*_n is the optimal estimation accuracy over all estimators.

Remark 10. Conjecture 1 in [3] corresponds to taking $\mu_{\mathrm{s}} = pB_0 + (1-p)B_{\frac{1}{2}}$ for $p > 0$ (binary erasure channels). Conjecture 2.5 in [4] corresponds to taking $\mu_{\mathrm{s}} = B_\alpha$ for $\alpha \in [0, \frac{1}{2})$ (binary symmetric channels). Thus, Theorem 39 closes both of them.

Proof of Theorem 39. The proof can be established by first connecting SBM to the broadcast with survey setting, similar to the approach presented in [4, Section 2.4]. Formally, one can show that Lemma 3.7 and Lemma 3.9 in [4] holds for general

[3] In certain parts of [4], a generalized setting was considered, where a, b are n-dependent. It is clear that the same generalization is not considered in their Conjecture 1, otherwise the stated limits may not converge. However, one can still prove a similar optimality result using the BI presented in Section IV-B and IV-C, stated in terms of the absolute difference between estimation accuracies. More generally, this asymptotical optimality can hold whenever the local tree-like condition is satisfied for large n.

symmetric $\mu_{\rm s}$, which bound the accuracies on both sides of equation (39) using limiting distributions of BP recursion. I.e., for any fixed t, a, b, and symmetric $\mu_{\rm s}$, we have

$$\lim_{n\to\infty} p_n\left(\hat{X}_{\rm BP}^{(t)}\right) = p\left(\mu_{(t)}\right),$$
$$\limsup_{n\to\infty} p_n^* \le p\left(\underline{\mu}_{(t)}\right),$$

where $\mu_{(t)}$ and $\underline{\mu}_{(t)}$ are the LLR distributions in the BP recursions initialized by $\mu_{(0)} = B_0$ and $\underline{\mu}_{(0)} = B_{\frac{1}{2}}$. As a consequence, the needed equality condition is implied if all limiting distributions are identical, which is essentially the BI condition stated in Section IV-B. In this work, we proved that the BI condition holds for all regimes (Theorem 36), which completes the proof of optimality of local BP algorithms. $\square$

V. Further Extensions

Consider an infinite tree channel generated by the following class of structures.

Definition 40 (Element Tree)**.** An element tree T is defined to be a tuple that consist of

(a) a finite rooted tree;

(b) a parameter δ_e for each edge e;

(c) a subset of leaves $V_{\rm G}$ called the *growing points*;

(d) a symmetric distribution μ_v for each vertex $v \notin V_{\rm G}$.

The tree network is generated recursively based on a distribution of element trees, denoted by P_T. Initially, the tree has a single growing point at the root vertex. Then for each step $h \in \mathbb{N}_+$, all growing points are replaced by some random subtrees, that are i.i.d. with distribution P_T.[4] Given the tree structure, a single bit X_0 is broadcast downlink from the root, and each edge e serves as a BSC channel with crossover probability δ_e. We consider the process of estimating X_0 for the tree channel generated at each step h, where the estimator has access to the tree structure, the variables on the boundary of the network (i.e., the growing points) and survey observations at all other vertices each through a channel characterized by the corresponding μ_v. Generally, we also allow that the observations on the growing points are noisy, i.e., independently though identical binary-input channels.

Similar to earlier formulations, the LLR distribution of this problem can be determined through BP recursion. For any fixed element tree T, we denote its BP operator by $\mathcal{Q}_T$, which is defined as follows. If T is a single growing point, then $\mathcal{Q}_T$ is the identity map. Otherwise, let o denote the root vertex, $e_1, ..., e_d$ denote its incident edges, and $T_1, ..., T_d$ denote the corresponding subtrees rooted at the d children vertices, then $\mathcal{Q}_T\mu$ is given by the law of

$$R \triangleq R_{\rm s} + \sum_{u=1}^{d} F_{\delta_{e_u}}(\hat{R}_u), \tag{40}$$

where $R_{\rm s} \sim \mu_o$, $\hat{R}_u \sim (1-\delta_{e_u})\hat{\mu}_u + \delta_{e_u}\hat{\mu}_u^-$ (all jointly independent), $\hat{\mu}_u \triangleq \mathcal{Q}_{T_u}\mu$, and $\hat{\mu}_u^-$ is the complement (see Definition 2) of $\hat{\mu}_u$. For a general distribution P_T, we define $\mathcal{Q}_{P_T}\mu \triangleq \mathbb{E}_{P_T}[\mathcal{Q}_T\mu]$, and we say μ is a BP fixed point if $\mu = \mathcal{Q}_{P_T}\mu$.

Proposition 41. *Let $\mu_{(h)}$ be the LLR distribution of the tree model at step h, we have $\mu_{(h+1)} = \mathcal{Q}_{P_T}\mu_{(h)}$.*

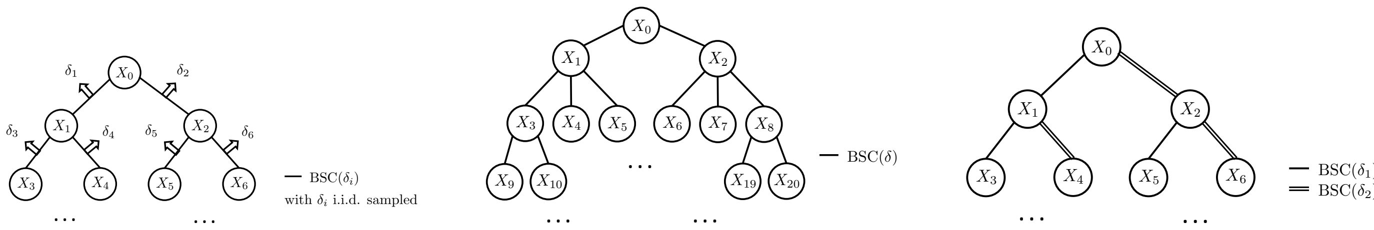
Proof. The derivation follows the same steps in Section I-A, as BP applies to any tree graphical models. $\square$

The above framework generalizes the settings in earlier sections, which allows us to cover several models of interest as special cases (see Fig. 2). We provide a full characterization for the uniqueness of BP fixed points in all cases. For ease of discussion, we assume non-zero capacity for all edges, i.e., $\delta_e \neq \frac{1}{2}$, otherwise the channel can be simplified by removing the corresponding subtrees. Among all tree distributions, there is one subclass where the recursion is easy to analyze, but the results need to be stated separately.

Definition 42. We say an element tree *simple*, if the tree network reduces to a noiseless simple path, i.e., there is exactly one growing point, all edges on the path from the growing point to the root has $\delta_e \in \{0,1\}$, and all survey distributions μ_v are trivial. We say an element tree distribution P_T is *trivial* if T is simple w.p.1.

This subclass of trivial tree distributions corresponds to the exceptional cases specified in Theorem 1 in the basic formulation. Generally, $\mathcal{Q}_{P_T}$ is the identity operator for any trivial P_T. Given this definition, we state our main theorem for non-trivial cases as follows.

[4]For readers interested in models where the tree can grow from non-leaf vertices, any such growing point can be treated equivalently as a leaf vertex connected to the tree through an edge with $\delta_e = 0$.



(a) Crossover probabilities being i.i.d. random and known to the estimator.

(b) Periodic, i.e., the degree of each vertex depends periodically on their depths.

(c) Nonisotropic, i.e., each vertex is connected through a fixed list of channels, each with distinct crossover probabilities.

Fig. 2: An Illustration of three different classes of Ising models. All presented examples can be generated from certain element-tree distributions.

Theorem 43. *Fix any non-trivial distribution P_T. The operator $\mathcal{Q}_{P_T}$ has at most one unique non-trivial fixed point μ^*, and it is symmetric. For all non-trivial μ, the recursion $\mathcal{Q}^h_{P_T}\mu$ converges weakly to the same fixed point, to μ^* if it exists, and to the trivial δ_0 otherwise.*

Remark 11. Note that the trivial fixed point $\mu = \delta_0$ exists if and only if all survey distributions are trivial w.p.1. This corresponds to the no survey case in the basic formulation. In all other cases, the operator $\mathcal{Q}_{P_T}$ has exactly one unique non-trivial fixed point, and all BP recursions converge to μ^*, including ones initialized with the trivial distribution.

To state the key intermediate results, we introduce the following concepts.

Definition 44. For any fixed tree channel T and any $r > 0$, we say a growing point v is *dominant*, if the sign of the LLR message returned by the BP algorithm remains consistent whenever the input message at v is r and the inputs at all other growing points are within $[-r, r]$. For a given parameter r, we say T satisfies the *polygon condition* if there are no dominant growing points, and we say T is *representative* if it also satisfies

$$F_T(r) \triangleq r_{\max}\left(\mathcal{Q}_T B_{\frac{1}{1+e^r}}\right) > r.$$

Theorem 45. *For any distribution P_T, any non-trivial symmetric fixed point μ, and any $\phi \in (0, \frac{1}{2})$, let $r = r_{\max}(B_\phi \boxast \mu)$, we have*

$$B_\phi \boxast \mathcal{Q}_{P_T}\mu \prec \mathcal{Q}_{P_T}(B_\phi \boxast \mu) \tag{41}$$

if T is representative with non-zero probability. Moreover, for any P_T that is non-trivial, there is an integer $h > 0$ that the tree network created by distribution P_T in h steps is representative with non-zero probability, implying

$$B_\phi \boxast \mathcal{Q}^h_{P_T}\mu \prec \mathcal{Q}^h_{P_T}(B_\phi \boxast \mu). \tag{42}$$

Remark 12. Intuitively, for any growing point v to be dominant, it is equivalent to state that the estimation error probability in a robust reconstruction setting can be minimized by solely measuring X_v. In particular, the estimation error is Bayes with $\mathrm{Ber}(\frac{1}{2})$ prior, and the observing channels are BSC with crossover probability $\frac{1}{1+e^r}$. Theorem 45 states that whenever an element tree is not equivalent to a trivial network, its self-concatenation will forbid such optimal estimators as the tree grows large.

Remark 13. The settings in earlier sections can be interpreted as special cases where the heights of the element trees are no greater than 1 and all δ_e parameters are constant. The contraction results stated in our general formulation provide a natural interpretation on the need of special treatments for $d \leq 2$. When no survey channels are present, the polygon condition requires $F_{\delta_e}(r)$ for all edges to be the side lengths of a polygon. It requires at least $d \geq 3$ equilateral edges to form a polygon. Otherwise, self-concatenation or non-trivial survey distributions are needed to meet the condition.

We also state some useful properties. which can be proved by induction over the structure of element trees.

Proposition 46. *The following statements hold for any distribution P_T.*

1) *If $\mu \preceq \nu$, then $\mathcal{Q}_{P_T}\mu \preceq \mathcal{Q}_{P_T}\nu$.*
2) *If $\{\mu_n\}_{n\in\mathbb{N}}$ converges weakly to μ, then $\mathcal{Q}_{P_T}\mu_n$ converges weakly to $\mathcal{Q}_{P_T}\mu$.*
3) *For any symmetric μ and $\phi \in [0, 1]$, we have $B_\phi \boxast \mathcal{Q}_{P_T}\mu \preceq \mathcal{Q}_{P_T}(B_\phi \boxast \mu)$.*

Proof. The first property follows from the fact that BP preserves degradation. The second property is due to the continuity of convolution and box convolution for weak convergence. Formally, one can repeat the same relevant steps in the proof of Theorem 4 as follows. For any fixed rooted tree, we write $\mathcal{Q}_T\mu_n$ as the law of a function that depends on a list of independent random variables with distributions given by μ_n and μ_v's. The value of this function converges in probability as $n \to \infty$, due to the weak convergence of μ_n, and this convergence is uniform with respect to all parameter values of δ_e's. Then the weak convergence of $\mathcal{Q}_{P_T}\mu_n$ follows from the fact that finite rooted trees are countable. The third property follows from the stringy tree lemma and the commutativity of box convolution. □

In the rest of this section, we prove the uniqueness theorem assuming the correctness of above intermediate results. Then we provide a proof of Theorem 45 in Appendix E.

Proof of Theorem 43. First, as explained in the proof of Theorem 4, a contraction statement in the form of inequality (42) implies the uniqueness of non-trivial symmetric BP fixed point. Then, as noted in the proof for asymmetric distributions (see proof of Corollary 3), the non-existence of asymmetric fixed point and the needed BP convergence can be proved by focusing on a certain subclass of the initial distributions. In particular, let $\mu_{(h)} \triangleq \mathcal{Q}^h_{P_T} B_0$ denote the recursion with a noiseless channel initialization. We only need to show that when $\mu_{(h)}$ converges to the non-trivial μ^*, the following recursion converges to the same limit for any non-trivial μ.

$$\underline{\mu}_{(h)} \triangleq \mathcal{Q}^h_{P_T}\underline{\mu}_{(0)},$$
$$\underline{\mu}_{(0)} \triangleq B_{\phi^*(\mu^*,\mu)} \boxast \mu^*.$$

This generalization is guaranteed by the fact that $\mathcal{Q}_{P_T}$ preserves degradation.

Recall that μ^* is a symmetric fixed point. We can apply the third property of Proposition 46 to show that

$$\underline{\mu}_{(0)} = B_{\phi^*(\mu^*,\mu)} \boxast (\mathcal{Q}_{P_T}\mu^*) \preceq \mathcal{Q}_{P_T}(B_{\phi^*(\mu^*,\mu)} \boxast \mu^*) = \underline{\mu}_{(1)}.$$

More generally, we have the following chain by applying the first property of Proposition 46.

$$\underline{\mu}_{(h)} = \mathcal{Q}^h_{P_T}\underline{\mu}_{(0)} \preceq \mathcal{Q}^h_{P_T}\underline{\mu}_{(1)} = \underline{\mu}_{(h+1)}. \tag{43}$$

Therefore, the monotone convergence property of degradation can be applied and the recursion $\underline{\mu}_{(h)}$ converges to a symmetric distribution. By the second property of Proposition 46, this limit is a BP fixed point (see relevant steps in the proof of Theorem 4), and we denote it by $\underline{\mu}^*$.

Because the degradation index $\phi^*(\mu^*,\mu)$ always belongs to $[0,\frac{1}{2})$ for non-trivial μ. We have that $\underline{\mu}_{(0)}$ is non-trivial. Given the degradation chain stated by inequality (43), all $\underline{\mu}_{(h)}$ are bounded by $\underline{\mu}_{(0)}$, so is their limit. Hence, $\underline{\mu}^*$ is non-trivial as well. Then by the uniqueness of non-trivial symmetric fixed point we have $\underline{\mu}^* = \mu^*$. □

Acknowledgement

This work was supported in part by the MIT-IBM Watson AI Lab and by the National Science Foundation under Grant No CCF-2131115.

Appendix A
Properties of degradation, β-curves and weak convergence

When discussing symmetric distributions, here we consider the most general formulation where distributions can have non-zero mass at $+\infty$, similar to [31], [32], we adopt the natural definition of weak convergence over domain $(-\infty,+\infty]$.

Proposition 47. *The set of symmetric probability distributions on $(-\infty,\infty]$ and the set of convex, monotone functions $\beta : [0,1] \to [0,1]$ such that $\beta(t) \geq t$ are in 1-1 correspondence given by Def. 17. Furthermore, the following are equivalent:*

1) *The sequence of symmetric distributions μ_n weakly converge to μ*
2) *$\beta(t;\mu_n) \to \beta(t;\mu)$ for all $t \in [0,1]$.*
3) *$\beta(t;\mu_n) \to \beta(t;\mu)$ uniformly over $t \in [0,1]$.*

Proof. First, notice that every symmetric $R \sim \mu$ is completely determined by the distribution of $T = \tanh\frac{|R|}{2}$. Indeed, for any bounded function f that is continuous on $(-\infty,+\infty]$ we have

$$\mathbb{E}_{R\sim\mu}[f(R)] = \mathbb{E}_{R\sim\mu}\left[\frac{e^R f(R) + f(-R)}{e^R+1}\right].$$

The function under the latter expectation is even and thus can be written as $h(T)$ for some continuous function $h : [0,1] \to \mathbb{R}$. Hence, we can equivalently study correspondence between measures ν of T on $[0,1]$ and $\beta(t;\mu) = \mathbb{E}_{T\sim\nu}[T \vee t]$. A simple integration by parts shows

$$\beta(t;\mu) = 1 - \int_t^1 dy \mathbb{P}[T \leq y] = t + \int_t^1 dy \mathbb{P}[T > y]. \tag{44}$$

From this identity it is clear how to recover the CDF of T from the $\beta(t;\mu)$ by differentiation.

For $1 \Rightarrow 2$ we only need to use the fact that $\tanh\frac{|R|}{2} \vee t$ is a bounded continuous function of R. For $2 \Rightarrow 3$ we simply notice that β-curves are always 1-Lipschitz, and thus the family $\beta(t;\mu_n), n \geq 1$ is equicontinuous (hence pointwise convergence and uniform convergence coincide). For $3 \Rightarrow 1$ we first notice that $\mu_n \to \mu$ (weakly) is equivalent to convergence of corresponding distributions of $\tanh \frac{|R|}{2}$ (as discussed above). Denote this (to be shown) convergence by $\nu_n \to \nu$ (weakly). To that end, notice that from (44) we have $\beta(0) = \mathbb{E}[T]$ and

$$\int_0^1 t^{s-1}(\beta(t) - t)dt = \frac{1}{s(s+1)}\mathbb{E}[T^{s+1}] \qquad \forall\, s > 0.$$

Therefore, uniform convergence of β-curves implies convergence of moments of T and thus (since T is supported on $[0, 1]$) the weak convergence as well. □

Proposition 48. *Degradation has the following properties.*

1) *(Continuity) For any two sequences of distributions $\{\mu_n\}_{n\in\mathbb{N}}$ and $\{\nu_n\}_{n\in\mathbb{N}}$ that weakly converge to μ and ν respectively and satisfy $\mu_n \preceq \nu_n$ for any n, we have $\mu \preceq \nu$.*
2) *(Sandwich Theorem) For any sequences of distributions $\{\underline{\mu}_n\}_{n\in\mathbb{N}}$, $\{\mu_n\}_{n\in\mathbb{N}}$, $\{\overline{\mu}_n\}_{n\in\mathbb{N}}$ satisfying $\underline{\mu}_n \preceq \mu_n \preceq \overline{\mu}_n$ for any n and $\underline{\mu}_n, \overline{\mu}_n$ weakly converges to μ^* for some μ^*, we have μ_n weakly converges to μ^*.*
3) *(Monotone Convergence) For any sequence of symmetric distributions $\{\mu_n\}_{n\in\mathbb{N}}$, if either $\mu_n \preceq \mu_{n+1}$ or $\mu_{n+1} \preceq \mu_n$ holds for all n, then the sequence converges weakly to a symmetric distribution.*
4) *(Poset Structure and Antisymmetry) Degradation defines a partial order on the set of all symmetric distributions. Especially, $\mu \preceq \nu$ and $\nu \preceq \mu$ implies $\mu = \nu$.*
5) *(Convolution-preserving) For any symmetric distributions μ, ν, and τ satisfying $\mu \preceq \nu$, we have $\tau * \mu \preceq \tau * \nu$ and $\tau \boxast \mu \preceq \tau \boxast \nu$.*
6) *(Invertibility over BSC Operator) For any $\phi \in [0, \frac{1}{2})$ and any symmetric distributions μ and ν, the statements $\mu \preceq \nu$ and $B_\phi \boxast \mu \preceq B_\phi \boxast \nu$ are equivalent.*
7) *(BP-preserving) For any $\delta \in [0, 1]$ and degree distribution P_d, let $\mathcal{Q}_s$ be the operator defined in Section I. Then for any two symmetric distributions $\mu \preceq \nu$, we have*
$$\mathcal{Q}_s\mu \preceq \mathcal{Q}_s\nu.$$

Remark 14. The concept of degradation has been studied as early as in [26]–[29], under a topic called comparison of experiments. It has also appeared later in [20], [33]–[38] for communication channels. Under those contexts, degradation serves as a preorder that divides experiments or communication channels into equivalent classes. The relationship given by Definition 5 permits antisymmetry due to the fact that LLR is a minimal sufficient statistic.

Proof of Proposition 48. The first property follows from the sequential compactness of joint distributions. Consider any choice of joint distributions between μ_n and ν_n that satisfy Definition 5. We can find a subsequence of these distributions that converges weakly as $n \to \infty$. Their limit provides a valid construction for degradation. Specifically, let $P_{Y,Z}$ be any such limit and by marginal convergence we assume that $Y \sim \mu$ and $Z \sim \nu$. The invariance of $P_{Y|Z}$ under $(Y, Z) \to (-Y, -Z)$ is preserved under weak convergence as it is equivalent to $dP_{Y,Z} = e^Z dP_{-Y,-Z}$.

The second property is clear from Theorem 19 and Prop. 47. Specifically, we only use the part of Theorem 19 that states β-curve inequalities are implied by degradation. This directly follows from the fact that degradation implies inequalities in Bayes estimation errors, which lead to the needed inequalities in β-curves (Proposition 18).

The third follows from the identification with β-curves (see [20, Lemma 4.75]). The fourth property is again via the β-curve inequalities stated in Theorem 19.

The rest of the properties state that degradation is preserved under several elementary operations and their compositions. They can be proved by considering the binary-input-channel equivalence of the related distributions, and the needed Markov chain constructions naturally follow from the probabilistic interpretations of these operations. In particular, box convolution can be viewed as channel concatenation (see Remark 1), and convolution can be viewed as parallel concatenation. □

A. Convergence in Degradation Metric

The next result explains that topology of the degradation metric is strictly finer than that of weak convergence.

Proposition 49. *Let $\{\mu_n\}_{n\in\mathbb{N}}$ and μ be non-trivial symmetric distributions, then the following are equivalent*

1) $d(\mu_n, \mu) \to 0$
2) $\mu_n \to \mu$ *(weakly) and* $t_{\max}(\mu_n) \to t_{\max}(\mu)$

Proof. We first prove that convergence in degradation metric implies weak convergence. Recall Proposition 47. It suffices to show the pointwise convergence of β-curves. The proof follows from the fact that both $\phi^*(\mu_n, \mu)$ and $\phi^*(\mu, \mu_n)$ converges to 0 as $d(\mu_n, \mu) \to 0$, and each degradation index implies a uniform bound on $\beta(t; \mu) - \beta(t; \mu_n)$.

Consider any pair of symmetric distributions μ and ν, we derive a bound on $\beta(t; \mu) - \beta(t; \nu)$ using $\phi^*(\mu, \nu)$. Recall the second property of Proposition 14 states that

$$B_{\phi^*(\mu,\nu)} \boxast \mu \preceq \nu.$$

By applying Theorem 19 and equation (15), we have

$$\beta(t;\nu) \geq (1-2\phi^*(\mu,\nu))\beta\left(\frac{t}{1-2\phi^*(\mu,\nu)};\mu\right).$$

Note that β-curves are non-negative, non-decreasing function of $|t|$ that are upper bounded by 1 for any $|t|\leq 1$. If $\phi^*(\mu,\nu)<\frac{1}{2}$, we have the following inequality.

$$\beta(t;\nu) \geq (1-2\phi^*(\mu,\nu))\beta\left(t;\mu\right) \geq \beta\left(t;\mu\right) - 2\phi^*(\mu,\nu) \qquad \text{for } |t|\leq 1.$$

Because $\beta(t;\nu)=\beta(t;\nu)=|t|$ for $|t|>1$, we have obtained an bound of $\sup_t(\beta(t;\mu)-\beta(t;\nu))$, which converges to 0 as $\phi^*(\mu,\nu)\rightarrow 0$.

By symmetry, we also have

$$\beta(t;\mu) \geq \beta\left(t;\nu\right) - 2\phi^*(\nu,\mu)$$

for $\phi^*(\nu,\mu)<\frac{1}{2}$. Therefore, the convergence of degradation index function implies that

$$\lim_{n\rightarrow\infty}\sup_t|\beta(t;\mu_n)-\beta(t;\mu)| \leq \lim_{n\rightarrow\infty} 2(\phi^*(\mu,\mu_n)+\phi^*(\mu_n,\mu))=0,$$

which proves the weak convergence.

Similarly, we provide an analysis of $t_{\max}$ functions based on the BSS Theorem 19 and equation (15). For any symmetric μ and ν, Proposition 14 implies that

$$t_{\max}(\nu) \geq (1-2\phi^*(\mu,\nu))t_{\max}(\mu) \geq t_{\max}(\mu) - 2\phi^*(\mu,\nu).$$

Therefore, by symmetry, we have

$$\lim_{n\rightarrow\infty}|t_{\max}(\mu_n)-t_{\max}(\mu)| \leq \lim_{n\rightarrow\infty} 2(\phi^*(\mu,\mu_n)+\phi^*(\mu_n,\mu))=0.$$

Now we prove the opposite direction. First, we show that unconditionally, the weak convergence of μ_n to μ implies

$$\lim_{n\rightarrow\infty}\phi^*(\mu,\mu_n)=0. \tag{45}$$

Recall that weak convergence implies L_∞-convergence in β-curves (Proposition 47). We can find a sequence of non-negative numbers $\epsilon_1,\epsilon_2,...$ converging to 0 such that

$$\beta(t;\mu_n) \geq \beta\left(t;\mu\right) - \epsilon_n \tag{46}$$

for all $t\in[0,1]$ and $n\in\mathbb{N}_+$. The above lower bound can be realized by symmetric distributions. Formally, let $\beta_n(t)=\max\{\beta\left(t;\mu\right)-\epsilon_n,t\}$. It is clear from Proposition 47 that each β_n is a valid β-curve and we can find a symmetric ν_n such that $\beta\left(t;\nu_n\right)=\beta_n(t)$. Specifically, we have

$$\nu_n[(-r,r)] = \begin{cases}\mu[(-r,r)] & \text{if } r\in(0,r_{\max}(\nu_n)],\\ 1 & \text{if } r\in(r_{\max}(\nu_n),+\infty).\end{cases} \tag{47}$$

For μ being non-trivial, we have $t_{\max}(\mu)>0$. Therefore, we can choose

$$\phi_n=\frac{1}{2}\left(1-\frac{t_{\max}(\nu_n)}{t_{\max}(\mu)}\right)$$

so that $r_{\max}(\nu_n)=F_{\phi_n}(r_{\max}(\mu))$ and

$$\nu_n[(-F_{\phi_n}(r),F_{\phi_n}(r))] \leq \mu[(-r,r)] = B_{\phi_n}\boxast\mu\left[(-F_{\phi_n}(r),F_{\phi_n}(r))\right]$$

for all $r\in(0,+\infty)$. By the integration law in equation (44) and the BSS Theorem, the above inequality implies

$$B_{\phi_n}\boxast\mu\preceq\nu_n\preceq\mu_n,$$

which proves that $\phi^*(\mu,\mu_n)\leq\phi_n$. From the continuity of β-curves, the convergence of ϵ_n to 0 implies that $\lim_{n\rightarrow\infty}t_{\max}(\nu_n)=t_{\max}(\mu)$. Hence,

$$\lim_{n\rightarrow\infty}\phi^*(\mu,\mu_n)\leq\lim_{n\rightarrow\infty}\phi_n=0.$$

On the other hand, we prove the following convergence with the additional assumption of $t_{\max}(\mu_n)\rightarrow t_{\max}(\mu)$.

$$\lim_{n\rightarrow\infty}\phi^*(\mu_n,\mu)=0. \tag{48}$$

By the same arguments, we pick any sequence of non-negative numbers $\epsilon_1, \epsilon_2, \ldots$ converging to 0 such that

$$\beta(t;\mu) \geq \beta(t;\mu_n) - \epsilon_n \tag{49}$$

holds for all $t \in [0,1]$ and $n \in \mathbb{N}_+$. Let $\tilde{\mu}_n$, $\tilde{\nu}_n$ be the symmetric distributions with $\beta(t;\tilde{\mu}_n) = \max\{\beta(t;\mu_n), \beta(t;\mu)\}$ and $\beta(t;\tilde{\nu}_n) = \max\{\beta(t;\tilde{\mu}_n) - \epsilon_n, t\}$. From BSS theorem we have $\mu_n \preceq \tilde{\mu}_n$ and $\tilde{\nu}_n \preceq \mu$.

Recall that μ is non-trivial, which implies that $t_{\max}(\tilde{\mu}_n) \geq t_{\max}(\mu) > 0$. We can choose

$$\tilde{\phi}_n = \frac{1}{2}\left(1 - \frac{t_{\max}(\tilde{\nu}_n)}{t_{\max}(\tilde{\mu}_n)}\right),$$

so the earlier proof steps imply that $B_{\tilde{\phi}_n} \circledast \tilde{\mu}_n \preceq \tilde{\nu}_n$. Hence, we have the following chain

$$B_{\tilde{\phi}_n} \circledast \mu_n \preceq B_{\tilde{\phi}_n} \circledast \tilde{\mu}_n \preceq \tilde{\nu}_n \preceq \mu,$$

which leads to the upper bound $\phi^*(\mu, \mu_n) \leq \tilde{\phi}_n$. Note that inequality (49) and the definition of $\tilde{\nu}_n$ imply that

$$\beta(t;\tilde{\nu}_n) \in [\beta(t;\mu) - \epsilon_n, \beta(t;\mu)].$$

We have the convergence of $t_{\max}(\tilde{\nu}_n)$ to $t_{\max}(\mu)$ by the continuity of β-curves and the convergence of ϵ_n. Note that $t_{\max}(\tilde{\mu}_n) = \max\{t_{\max}(\mu_n), t_{\max}(\mu)\}$. We also have the convergence of $t_{\max}(\tilde{\mu}_n)$ to $t_{\max}(\mu)$. Therefore,

$$\lim_{n\to\infty} \phi^*(\mu_n, \mu) \leq \lim_{n\to\infty} \tilde{\phi}_n = 0.$$

To conclude, the convergence of both $\phi^*(\mu, \mu_n)$ and $\phi^*(\mu_n, \mu)$ to 0 proves the convergence in degradation metric. □

Remark 15. The convergence in degradation metric is not implied solely by the weak convergence (or L_∞-convergence in β-curves). For example, the sequence $\mu_n = 2^{-n} \cdot B_{\frac{1}{6}} + (1 - 2^{-n}) \cdot B_{\frac{1}{3}}$ converges to $\mu = B_{\frac{1}{3}}$ both weakly and in β-curves, but $d(\mu_n, \mu) = \ln 2$ for any n. However, as we have shown in the earlier proof, a converse can be proved for a single-sided function $\phi(\mu, \mu_n)$, which can be viewed as a potential function that is stabilized by the BP recursion. We present a slightly more general version of this result in Proposition 50.

Besides, note that even the sequence μ_n has a bounded radius in d, it has no convergent subsequence under the same metric, which shows that bounded closed sets are not compact in the d-metric. Further, all pairwise distances in this sequence are bounded away from zero as $d(\mu_m, \mu_n) = \ln \frac{2}{1+2^{-|m-n|}}$. Thus, while d-convergence is almost equivalent to weak convergence as per Prop. 49, the induced topologies are quite different.

Proposition 50. *If a sequence of LLR distributions $\{\mu_n\}_{n\in\mathbb{N}}$ converges weakly to μ, then for any symmetric ν*

$$\lim_{n\to\infty} \phi^*(\nu, \mu_n) = \phi^*(\nu, \mu).$$

Proof. The following lower bound follows from the first property in Proposition 48 and the definition of degradation index,

$$\liminf_{n\to\infty} \phi^*(\nu, \mu_n) \geq \phi^*(\nu, \mu).$$

Hence, it remains to show that

$$\limsup_{n\to\infty} \phi^*(\nu, \mu_n) \leq \phi^*(\nu, \mu). \tag{50}$$

Recall that in the proof of Proposition 49 we have proved equation (45), which covers the special case of $\nu = \mu$. We bound the degradation index for general ν by applying the triangle inequality, i.e., the fourth property of Proposition 14.

$$\phi^*(\nu, \mu_n) \leq \phi^*(\mu, \mu_n) + \phi^*(\nu, \mu) - 2\phi^*(\mu, \mu_n)\phi^*(\nu, \mu). \tag{51}$$

Then the needed result is proved using equation (45), particularly,

$$\limsup_{n\to\infty} \phi^*(\nu, \mu_n) \leq \limsup_{n\to\infty} \phi^*(\mu, \mu_n) + \phi^*(\nu, \mu) = \phi^*(\nu, \mu).$$

□

Appendix B
Supplementary Proofs for Theorem 1

A. Proof of Proposition 12

Proof. We first prove the proposition for condition $\nu \preceq \tau \prec \mu$. Recall the definition of strict degradation. We can find $\phi \in (0, \frac{1}{2}]$ such that $\nu \preceq \tau \preceq B_\phi \circledast \mu$. By applying the transitivity of degradation, we have $\nu \preceq B_\phi \circledast \mu$, which proves $\nu \prec \mu$.

On the other hand, if $\nu \prec \tau \preceq \mu$, we have $\phi \in (0, \frac{1}{2}]$ such that $\nu \preceq B_\phi \circledast \tau$. Now we need the fifth property in Proposition 48 to obtain $B_\phi \circledast \tau \preceq B_\phi \circledast \mu$. Then the rest follows from the transitivity of degradation through the same arguments. □

B. Proof of Proposition 14

Proof. The first property is proved by a probability-of-error argument. Recall the definition of degradation index. We only need to show the existence of a $\phi \in [0, \frac{1}{2})$ that satisfies $B_\phi \boxast \mu \preceq \nu$. We prove this fact by choosing $\phi = \mathbb{E}_{R\sim\nu}[\frac{1-\text{sgn}(R)}{2}]$. Because ν is non-trivial and symmetric, we have $\phi \in [0, \frac{1}{2})$.

Following the commutativity of box convolution and the physical interpretation, we have $B_\phi \boxast \mu \preceq B_\phi$. On the other hand, B_ϕ can be viewed as the LLR distribution of the 1-bit maximum likelihood estimator for the estimation problem characterized by ν. Hence, there is a natural joint distribution that implies $B_\phi \preceq \nu$. Then the statement $B_\phi \boxast \mu \preceq \nu$ follows from the transitivity of degradation.

The second property states that the set defined in equation (9) has a minimum. It follows from the continuity of degradation under weak convergence, see the first property in Proposition 48. The third property can be proved using the sixth property in Proposition 48. The fourth property is due to the transitivity of degradation and associativity of box convolution. □

C. Proof of Proposition 18

Proof. Recall that the optimal Bayes estimation error is achieved by the maximum a posteriori (MAP) estimator, which compares the LLR with a fixed threshold $\ln \frac{1-t}{1+t} = -2\,\tanh^{-1} t$ for the given prior distribution. Therefore, the achieved error can be written as

$$\frac{1+t}{2} \cdot \mu\left[\left(-\infty, -2\,\tanh^{-1} t\right]\right] + \frac{1-t}{2} \cdot \int e^{-r} \mathbb{1}\{r \in (-2\,\tanh^{-1} t, +\infty)\} d\mu(r),$$

where the second term is derived from the definition of LLR. For brevity, we denote this quantity by $P_{\mathrm{e}}(t)$.

Note that for any bounded Borel function $f : (-\infty, +\infty] \to \mathbb{R}$, its expectation over μ can be written as follows using the symmetry condition.

$$\begin{aligned}\mathbb{E}_{R\sim\mu}\left[f(R)\right] &= \mathbb{E}_{R\sim\mu}\left[\frac{e^{\frac{R}{2}}}{e^{\frac{R}{2}} + e^{-\frac{R}{2}}} f(R) + \frac{e^{-\frac{R}{2}}}{e^{\frac{R}{2}} + e^{-\frac{R}{2}}} f(R)\right] \\ &= \mathbb{E}_{R\sim\mu}\left[\frac{e^{\frac{R}{2}}}{e^{\frac{R}{2}} + e^{-\frac{R}{2}}} f(R) + \frac{e^{-\frac{R}{2}}}{e^{\frac{R}{2}} + e^{-\frac{R}{2}}} f(-R)\right].\end{aligned}$$

We specialize the above formula for $P_{\mathrm{e}}(t)$, which can be viewed as the expectation of the following function

$$p(R) \triangleq \frac{1+t}{2} \mathbb{1}\left\{t \leq -\tanh \frac{R}{2}\right\} + \frac{1-t}{2} e^{-R} \mathbb{1}\left\{t > -\tanh \frac{R}{2}\right\}.$$

The resulting expectation is given by

$$\mathbb{E}_{R\sim\mu}\left[\min\left\{\frac{e^{-|R|}}{e^{\frac{R}{2}} + e^{-\frac{R}{2}}}, \frac{1-t}{2}\right\}\right]$$

which equals $\frac{1}{2}(1 - \beta(t;\mu))$. □

D. Proof of Proposition 20

Proof. Proof of 1) ⇒ 3). From $\nu \prec \mu$ and the non-trivial condition, we can choose $\phi \in (0, \frac{1}{2})$ to satisfy $\nu \preceq B_\phi \boxast \mu$. By either applying Theorem 19 or Jensen's inequality over any joint distribution consistent with Definition 5, we have $\beta(t;\nu) \leq \beta(t; B_\phi \boxast \mu)$ for any $t \in \mathbb{R}$. Using equation (15), we have

$$(1 - 2\phi)\beta(t;\mu) \geq \beta\left((1 - 2\phi)t; \nu\right).$$

Therefore, following the definition of $t_{\max}$ and non-trivial condition, we have $t_{\max}(\nu) \leq (1 - 2\phi)t_{\max}(\mu) < t_{\max}(\mu)$. On the other hand, note that any β-curve is 1-Lipschitz. For $t \in [0, t_{\max}(\mu))$, we have

$$\begin{aligned}\beta\left((1 - 2\phi)t; \nu\right) &\geq \beta\left(t; \nu\right) - 2\phi t. \\ &> \beta\left(t; \nu\right) - 2\phi\beta(t;\mu).\end{aligned}$$

Combining the above inequalities, we obtain the stated gap condition.

Proof of 3) ⇒ 2). If $t_{\max}(\nu) < t_{\max}(\mu)$, then $[0, t_{\max}(\nu)]$ is a subset of $[0, t_{\max}(\mu))$. Thus, the same gap condition applies.

Proof of 2) ⇒ 1). Recall that both μ and ν are non-trivial. The β-curves for both distributions are positive and continuous over $\mathbb{R}$. Therefore, the following quantity is well-defined

$$\phi \triangleq \min_{t\in[0,t_{\max}(\nu)]} \frac{1}{2}\left(1 - \frac{\beta(t;\nu)}{\beta(t;\mu)}\right).$$

From the gap condition, we have $\phi \in \left(0, \frac{1}{2}\right)$. It remains to show $\nu \preceq B_\phi \circledast \mu$.

One can verify that the β-curve is non-decreasing for any measure. Hence, for any $t \in [0, t_{\max}(\nu)]$, the above definition implies the following inequality.

$$\beta(t;\nu) \leq (1-2\phi)\beta\left(t;\mu\right) \leq (1-2\phi)\beta\left(\frac{t}{1-2\phi};\mu\right).$$

Using equation (15), the above inequality can be written as

$$\beta(t;\nu) \leq \beta\left(t; B_\phi \circledast \mu\right). \tag{52}$$

Because the β-curve is lower bounded by the identity function, recall the definition of $t_{\max}(\nu)$, inequality (52) holds for $t > t_{\max}(\nu)$ as well. Apply this conclusion to Theorem 19, we have proved that $\nu \preceq B_\phi \circledast \mu$, which implies strict degradation. $\square$

E. Proof of Proposition 23

Proof. Proof of 1). The forward statement is proved by noting that when $t_{\max}(\nu) < t_{\max}(\mu)$, the set $\left(\tanh\left(\frac{s}{2}\right), t_{\max}(\nu)\right]$ is contained in the set $\left(\tanh\left(\frac{s}{2}\right), t_{\max}(\mu)\right)$. For the converse statement, we have $\beta(t;\nu) = t < \beta(t;\mu)$ for all $t \in (t_{\max}(\nu), t_{\max}(\mu))$ from the definition of $t_{\max}$. So it remains to prove that $t_{\max}(\nu) < t_{\max}(\mu)$. If $s \geq r_{\max}(\nu)$, this inequality is implied by $r_{\max}(\nu) \leq s < r_{\max}(\mu)$. Otherwise, we have $\beta(t_{\max}(\nu), \mu) > \beta(t_{\max}(\nu), \nu) = t$, and the inequality is implied due to the continuity of β-curves.

Proof of 2). The stated condition provides a chain of degradation, so we have $\beta(t;\nu) \leq \beta(t;\tau) \leq \beta(t;\mu)$ for all t. In either cases, the individual steps in the condition implies that the inequality is strict for $t \in \left(\tanh\left(\frac{s}{2}\right), t_{\max}(\nu)\right]$.

Proof of 3). The first statement is a direct consequence of Proposition 20 and the fact that β-curves are even functions. The second statement is implied by the BSS Theorem.

Proof of 4). The first statement directly follows from the definition. The second statement is implied by Proposition 20. $\square$

F. Proof of Proposition 24

Proof. For convenience, we define

$$\begin{aligned}
\tau_{(\mathrm{T})} &= B_\phi \circledast (B_{\delta_1} * B_{\delta_2}),\\
\tau_{(\mathrm{S})} &= (B_\phi \circledast B_{\delta_1}) * (B_\phi \circledast B_{\delta_2}).
\end{aligned}$$

To prove Proposition 24, we first derive closed-form expressions for the β-curves. Note that $\tau_{(\mathrm{T})}$ is a symmetric discrete distribution that can be written as a linear combination of $B_{\delta^-_{(\mathrm{T})}}$ and $B_{\delta^+_{(\mathrm{T})}}$ for some $0 \leq \delta^+_{(\mathrm{T})} \leq \delta^-_{(\mathrm{T})} \leq \frac{1}{2}$. From Proposition 21, the β-curve of $\tau_{(\mathrm{T})}$ must be a piecewise linear function with at most two corner points. Concretely, let[5]

$$\begin{aligned}
t^-_{(\mathrm{T})} &\triangleq t_{\max}\left(B_{\delta^-_{(\mathrm{T})}}\right) = (1-2\phi)\tanh\frac{|r_{\max}(B_{\delta_1}) - r_{\max}(B_{\delta_2})|}{2},\\
t^+_{(\mathrm{T})} &\triangleq t_{\max}\left(B_{\delta^+_{(\mathrm{T})}}\right) = (1-2\phi)\tanh\frac{r_{\max}(B_{\delta_1}) + r_{\max}(B_{\delta_2})}{2},\\
\beta_{(\mathrm{T}),0} &\triangleq \beta\left(0;\tau_{(\mathrm{T})}\right) = (1-2\phi)\max\{(1-2\delta_1),(1-2\delta_2)\}.
\end{aligned}$$

The value-function pair $(t, \beta(t;\tau_{(\mathrm{S})}))$ for $t \in [0,1]$ is on the lower convex envelope of the following finite set of points.

$$\left\{(0,\beta_{(\mathrm{T}),0}), \left(t^-_{(\mathrm{T})}, \beta_{(\mathrm{T}),0}\right), \left(t^+_{(\mathrm{T})}, t^+_{(\mathrm{T})}\right), (1,1)\right\}.$$

Similarly, let

$$\begin{aligned}
t^-_{(\mathrm{S})} &\triangleq t_{\max}\left(B_{\delta^-_{(\mathrm{S})}}\right) = \tanh\frac{|F_\phi\left(r_{\max}(B_{\delta_1})\right) - F_\phi\left(r_{\max}(B_{\delta_1})\right)|}{2},\\
t^+_{(\mathrm{S})} &\triangleq t_{\max}\left(B_{\delta^+_{(\mathrm{S})}}\right) = \tanh\frac{F_\phi\left(r_{\max}(B_{\delta_1})\right) + F_\phi\left(r_{\max}(B_{\delta_1})\right)}{2},\\
\beta_{(\mathrm{S}),0} &\triangleq \beta\left(0;\tau_{(\mathrm{S})}\right) = (1-2\phi)\max\{(1-2\delta_1),(1-2\delta_2)\}.
\end{aligned}$$

Then function $\beta(t;\tau_{(\mathrm{S})})$ for $t \in [0,1]$ is given by the lower convex envelope of the following finite set.

$$\left\{(0,\beta_{(\mathrm{S}),0}), \left(t^-_{(\mathrm{S})}, \beta_{(\mathrm{S}),0}\right), \left(t^+_{(\mathrm{S})}, t^+_{(\mathrm{S})}\right), (1,1)\right\}.$$

[5]When $r_{\max}(B_{\delta_1}) = r_{\max}(B_{\delta_2}) = +\infty$, $\tau_{(\mathrm{T})}$ is simply $B_{\delta^+_{(\mathrm{T})}}$, and we can let $t^-_{(\mathrm{T})}$ take any value in $[0, t^+_{(\mathrm{T})}]$.

By some elementary calculus[6], one can prove the following equation and inequalities, which shows that $\beta(t;\tau_{(\mathrm{T})}) \leq \beta(t;\tau_{(\mathrm{S})})$ for all $t\in\mathbb{R}$.

$$\begin{aligned} t^-_{(\mathrm{S})} &\leq t^-_{(\mathrm{T})},\\ t^+_{(\mathrm{S})} &\geq t^+_{(\mathrm{T})}.\\ \beta_{(\mathrm{S}),0} &= \beta_{(\mathrm{T}),0}. \end{aligned}$$

Moreover, we always have $t^+_{(\mathrm{S})} > t^+_{(\mathrm{T})}$ except when any of ϕ,δ_1,δ_2 is $\frac{1}{2}$. In all cases, this leads to $\beta(t;\tau_{(\mathrm{T})}) < \beta(t;\tau_{(\mathrm{S})})$ for any $t\in(t^-_{(\mathrm{S})},t^+_{(\mathrm{T})}]$. Note that $t^-_{(\mathrm{S})}=\tanh\frac{s_{\min}}{2}$ and $t^+_{(\mathrm{T})}=t_{\max}(\tau_{(\mathrm{T})})$. We have $\tau_{(\mathrm{T})}\prec_{s_{\min}}\tau_{(\mathrm{S})}$. $\square$

G. Proof of Corollary 25

Proof. Let $R_1\sim\mu_1$, $R_2\sim\mu_2$ be independent variables. Then each μ_j for $j\in\{1,2\}$ can be expressed as a mixture of distributions B_{δ_j} with $\delta_j=\frac{1}{1+e^{R_j}}$. From the bilinearity of convolution, the distributions on both sides of inequality (17) can be expressed as mixtures of the corresponding terms in inequality (16), i.e.,

$$\begin{aligned} B_\phi \boxast (\mu_1 * \mu_2) &= \mathbb{E}_{R_1,R_2}[B_\phi \boxast (B_{\delta_1} * B_{\delta_2})],\\ (B_\phi \boxast \mu_1) * (B_\phi \boxast \mu_2) &= \mathbb{E}_{R_1,R_2}[(B_\phi \boxast B_{\delta_1}) * (B_\phi \boxast B_{\delta_2})]. \end{aligned}$$

The β-curves can be compared using the (obvious) property: β-curve of a mixture is given by a mixture of the β-curves. Thus, the degradation relation stated in Proposition 24 implies that $B_\phi \boxast (\mu_1 * \mu_2) \preceq (B_\phi \boxast \mu_1) * (B_\phi \boxast \mu_2)$.

It remains to verify the strict condition for the inequality on β-curves. For brevity, we focus on the non-trivial cases where $\phi\neq\frac{1}{2}$ and both μ_1 and μ_2 are non-trivial. By the strict concavity of F_ϕ on $\mathbb{R}_{\geq 0}$, and the definition of $r_{\max}$, we have

$$\begin{aligned} r_{\max}(B_\phi \boxast (\mu_1 * \mu_2)) &= F_\phi(r_{\max}(\mu_1)+r_{\max}(\mu_2))\\ &< F_\phi(r_{\max}(\mu_1)) + F_\phi(r_{\max}(\mu_2))\\ &= r_{\max}((B_\phi \boxast \mu_1) * (B_\phi \boxast \mu_2)), \end{aligned}$$

which implies that $t_{\max}(B_\phi \boxast (\mu_1 * \mu_2)) < t_{\max}((B_\phi \boxast \mu_1) * (B_\phi \boxast \mu_2))$. Therefore, it suffices to compare the β-curves at any fixed $t\in\left(\tanh\left(\frac{s}{2}\right), t_{\max}((B_\phi \boxast \mu_1) * (B_\phi \boxast \mu_2))\right)$. By the definition of $r_{\max}$ function, there is a non-zero probability that R_1, R_2 are sufficiently close to $r_{\max}(\mu_1)$ and $r_{\max}(\mu_2)$, respectively, such that

$$t\in\left(\tanh\left(\frac{|F_\phi(r_{\max}(B_{\delta_1}))-F_\phi(r_{\max}(B_{\delta_2}))|}{2}\right), t_{\max}((B_\phi \boxast B_{\delta_1}) * (B_\phi \boxast B_{\delta_2}))\right).$$

According to Proposition 24 and the first statement in Proposition 23, this condition implies strict inequality in β-curves at point t between $B_\phi \boxast (B_{\delta_1} * B_{\delta_2})$ and $(B_\phi \boxast B_{\delta_1}) * (B_\phi \boxast B_{\delta_2})$. Hence, their integration also contributes to a strict inequality. $\square$

H. Proof of Proposition 26

Proof. Recall that β-curves are characterized by the error probabilities of MAP estimators (see Proposition 18). We construct an estimation problem with prior $X\sim\mathrm{Ber}(\frac{1-t}{2})$, and independent observations Y,Z such that Y is measured through a symmetric channel characterized by μ and Z is measured through a BSC with crossover probability ϕ. Let $\hat{X}$ be the MAP estimator. Because the LLR distribution of this experiment is characterized by $B_\phi * \mu$, we have

$$\mathbb{P}[\hat{X}=X] = \frac{1+\beta(t;B_\phi * \mu)}{2}.$$

Conditioned on $Z=0$ or $Z=1$, the inference problem reduces to estimating X given Y with different priors. Note that the MAP estimator remains the same. We have

$$\begin{aligned} \mathbb{P}[\hat{X}=X|Z=0] &= \frac{1+\beta(t_0;B_\phi * \mu)}{2},\\ \mathbb{P}[\hat{X}=X|Z=1] &= \frac{1+\beta(t_1;B_\phi * \mu)}{2}. \end{aligned}$$

Combining above results, we obtain the following equation, which is identical to the needed statement.

$$\beta(t;B_\phi * \mu) = \mathbb{P}[Z=0]\beta(t_0;B_\phi * \mu) + \mathbb{P}[Z=1]\beta(t_1;B_\phi * \mu).$$

$\square$

[6]In particular, the concavity of F_ϕ on $\mathbb{R}_{\geq 0}$.

I. Proof of Proposition 28

Proof. Consider any fixed $r \in (s-\ell, r_{\max}(\mu)-\ell)$. Let $t = \tanh \frac{|r|}{2}$, we have $t \in [0, 1)$. Then we can apply Proposition 26 to evaluate $\beta(t; \tau * \mu) - \beta(t; \tau * \nu)$ by writing τ as an integration of B_ϕ distributions. In particular, let $Z \sim \tau$ and $\phi = \frac{1}{e^{|Z|}+1}$. In general, we have

$$\beta(t; \tau * \mu) = \mathbb{E}\left[\left(\frac{1+t-2t\phi}{2}\right)\beta(t_0; \mu) + \left(\frac{1-t+2t\phi}{2}\right)\beta(t_1; \mu)\right] \tag{53}$$

for any symmetric μ, where

$$t_0 = \tanh\left(\frac{|Z|+|r|}{2}\right),$$

$$t_1 = \tanh\left(\left|\frac{|Z|-|r|}{2}\right|\right).$$

When $\nu \prec_s \mu$, we always have $\beta(t_0; \nu) \leq \beta(t_0; \mu)$ and $\beta(t_1; \nu) \leq \beta(t_1; \mu)$. To prove $\beta(t; \tau * \nu) < \beta(t; \tau * \mu)$ using equation (53), it remains to show that $\beta(t_0; \nu) < \beta(t_0; \mu)$ or $\beta(t_1; \nu) < \beta(t_1; \mu)$ with non-zero probability. Let $t_Z = \tanh\left(\left|\frac{r+Z}{2}\right|\right)$. Because $r + \ell \in (s, r_{\max}(\mu))$ and $\ell \in \mathrm{supp}(\tau)$, we have $\beta(t_Z; \mu) > \beta(t_Z; \nu)$ for Z in a neighbourhood of ℓ, which holds with non-zero probability. Note that $t_Z \in \{t_0, t_1\}$. The needed inequality is proved.

For the second statement, we can assume μ is non-trivial, otherwise the stated interval is an empty set and nothing needs to be proved. Note that in this case, $\nu \prec \mu$ is equivalent to $\nu \prec_s \mu$ for any $s \in (-r_{\max}(\mu), 0)$. Therefore, following exactly the same steps, we have the needed strict inequality holds for $r \in (s-\ell, r_{\max}(\mu)-\ell)$ for any such s. Hence, inequality (18) holds if r belongs to $r \in (-r_{\max}(\mu)-\ell, r_{\max}(\mu)-\ell)$, which contains $[-\ell, r_{\max}(\mu)-\ell)$ as a subset. □

J. Proof of Proposition 33

Proof. We apply induction over $k \in \mathbb{N}$. The base case $k = 1$ follows directly from Proposition 32. Indeed, when $s_1 < 0$, it implies

$$B_\phi \boxast \mathcal{Q}_{\mathrm{s}}\mu = B_\phi \boxast ((B_\delta \boxast \mu) * \mu_{\mathrm{s}}) \prec (B_\phi \boxast B_\delta \boxast \mu) * \mu_{\mathrm{s}} = \mathcal{Q}_{\mathrm{s}}(B_\phi \boxast \mu)$$

as $s_{\min} = F_\phi(r_{\max}(B_\delta \boxast \mu)) - r_{\mathrm{s}} = F_\phi(F_\delta(r_{\max}(\mu))) - r_{\mathrm{s}} = s_1 < 0$. Here we used the fact that $r_{\max}(B_\tau \boxast \nu) = F_\tau(r_{\max}(\nu))$ and

$$F_\phi \circ F_\delta = F_\delta \circ F_\phi\,,$$

due to commutativity of $B_\phi \boxast B_\delta$. Similarly, when $s_1 \geq 0$, the statement of Proposition 32 implies $B_\phi \boxast \mathcal{Q}_{\mathrm{s}}\mu \prec_{s_1} \mathcal{Q}_{\mathrm{s}}(B_\phi \boxast \mu)$.

Assume Proposition 33 holds for some $k \in \mathbb{N}$. We prove that it holds for $k+1$. As everywhere before, our method is to start from a non-strict degradation chain given by

$$B_\phi \boxast (\mathcal{Q}_{\mathrm{s}}^{k+1}\mu) \preceq \mathcal{Q}_{\mathrm{s}}(B_\phi \boxast \mathcal{Q}_{\mathrm{s}}^k\mu) \preceq \mathcal{Q}_{\mathrm{s}}(\mathcal{Q}_{\mathrm{s}}^k(B_\phi \boxast \mu)) \tag{54}$$

and keep track of areas where the comparison of β-curves is strict. Denote for convenience $\mu_k = \mathcal{Q}_{\mathrm{s}}^k\mu$, $\nu_k = \mathcal{Q}_{\mathrm{s}}^k(B_\phi \boxast \mu)$ and $r_k = r_{\max}(\mu_k)$. We will repeatedly use the fact that $r_{\max}(B_\tau \boxast \nu) = F_\tau(r_{\max}(\nu))$, so that for example $r_{k+1} = F_\delta(r_k) + r_{\mathrm{s}}$.

Observe the first step of inequality (54), which essentially states that

$$B_\phi \boxast ((B_\delta \boxast \mu_k) * \mu_{\mathrm{s}}) \preceq (B_\delta \boxast B_\phi \boxast \mu_k) * \mu_{\mathrm{s}}\,.$$

This is simply an instance of applying Proposition 32. Thus, we have that the comparison of β-curves is strict for

$$F_\phi(F_\delta(r_k)) - r_{\mathrm{s}} < s < F_\phi(F_\delta(r_k)) + r_{\mathrm{s}}\,. \tag{55}$$

Therefore, we can assume r_{s} is finite for the rest of the proof because otherwise we have already established strict comparison for the entire $s \in \mathbb{R}$, which includes $[0, r_{\max}(B_\phi \boxast \mu_{k+1})]$ as a subset.

Now we analyze the second step in (54). From the induction hypothesis we know that

$$B_\phi \boxast \mu_k \prec_{s_k} \nu_k \tag{56}$$

for $s_k \geq 0$, and $B_\phi \boxast \mu_k \prec \nu_k$ otherwise. Applying box convolution with B_δ to both sides of these inequalities and by equation (15), we obtain

$$B_\delta \boxast B_\phi \boxast \mu_k \prec_{s'_k} B_\delta \boxast \nu_k\,,$$

for $s'_k \triangleq F_\delta(s_k)$. Next, by convolving with μ_{s} on both sides, and then Proposition 28 with $\ell = r_{\mathrm{s}}$, we get that

$$(B_\delta \boxast B_\phi \boxast \mu_k) * \mu_{\mathrm{s}} \preceq (B_\delta \boxast \nu_k) * \mu_{\mathrm{s}} = \nu_{k+1}$$

with inequality for β-curves strict for all $t = \tanh \frac{|s|}{2}$ with

$$s_{k+1} = s'_k - r_{\mathrm{s}} < s < F_\delta(r_{\max}(\nu_k)) - r_{\mathrm{s}}\,. \tag{57}$$

From induction hypothesis (56) we have $r_{\max}(\nu_k) > r_{\max}(B_\phi \circledast \mu_k) = F_\phi(r_k)$. Therefore, because of the strictness of the inequality we have that (57) and (55) together imply comparison of first and last β-curves in (54) for all $t = \tanh(|s|/2)$ with

$$s_{k+1} < s < F_\phi(F_\delta(r_k)) + r_{\mathrm{s}}\,. \tag{58}$$

Finally, notice that $F_\phi(x+y) < F_\phi(x)+y$ for $x \geq 0$, $y > 0$, and $\phi \in (0,1)$. Thus the right-hand side of (58) is strictly bigger than $r_{\max}(B_\phi \circledast \mu_{k+1}) = F_\phi(r_{k+1}) = F_\phi(F_\delta(r_k) + r_{\mathrm{s}})$. In all, we have established strict comparison of β-curves for

$$s_{k+1} < s \leq r_{\max}(B_\phi \circledast \mu_{k+1})\,.$$

□

Appendix C
Proof of Corollary 3

Proposition 51. *The following statements are true.*

1) *All symmetric distributions are LLR distributions.*
2) *Any LLR distribution has a unique complement.*
3) *Any distribution μ on $(-\infty, +\infty]$ is an LLR distribution if and only if it has a complement.*

Proof. For the first statement, consider any symmetric μ. The LHS of inequality (3) equals $\mu[(-\infty, +\infty)]$, which is no greater than 1. For the second statement, the CDF of any complement distribution is completely determined by equation (4), which proves the uniqueness. Then, inequality (3) ensures that the CDF specified by equation (4) is always bounded within $[0, 1]$, which proves the existence. For third statement, the LHS of inequality (3) equals $\mu^-[(-\infty, +\infty)]$, which is no greater than 1. □

The notion of degradation can be generalized to LLR distributions. One can verify that the following definition is consistent with Definition 5 on symmetric distributions.

Definition 52. For any two LLR distributions μ_Y, μ_Z, we say μ_Y is a *degraded version* of μ_Z, denoted by $\mu_Y \preceq \mu_Z$, if one can define joint distributions $\mu_{Y,Z}$ and μ_{Y^-,Z^-}, with μ_Y, μ_Z, and their complements being the marginal distributions, such that $\mu_{Y|Z}$ and $\mu_{Y^-|Z^-}$ are identical.

Moreover, we have the following generalization.

Proposition 53. *Degradation defines a partial order on the set of all LLR distributions, which satisfies continuity, sandwich theorem, and has monotone convergence.*

Proof. Our proof relies on the following generalization of β-curves.

Definition 54. For any LLR distribution μ, we define its *β-curve* as a function on domain $t \in \mathbb{R}$ given by the following equation.

$$\beta(t;\mu) \triangleq \mathbb{E}_{R\sim\frac{1}{2}(\mu+\mu^-)}\left[\left|\tanh\frac{R}{2} - t\right|\right], \tag{59}$$

where μ^- is the complement distribution.

Given this definition, one can prove that LLR distributions and their β-curves are in one-to-one correspondence. This is because every symmetric μ is completely determined by the distribution of $T = \tanh\frac{R}{2}$ for $R \sim \frac{1}{2}(\mu + \mu^-)$, and their CDF can be obtained by differentiating the β-curve based on the following equation.

$$\beta(t;\mu) = 1 - t + \int_{-1}^{t} dy \mathbb{P}[T \leq y] - \int_{t}^{1} dy \mathbb{P}[T \leq y]\,.$$

The BSS theorem can be generalized for asymmetric distributions for the above β-curve definition. In particular, we only require $\beta(t;\nu) \leq \beta(t;\mu)$ for any $t \in \mathbb{R}$ and $\nu \preceq \mu$. This can be proved by the fact that Proposition 18 holds for asymmetric distributions, then the comparison of β-curves is implied by the comparison of corresponding hypothesis testing problems.

Combining the above results, for any $\mu \preceq \nu \preceq \mu$ the β-curves of μ and ν are identical. Therefore, we have $\mu = \nu$, which proves antisymmetry. The rest of the properties are obtained by the corresponding steps in Proof of Proposition 48, except that for continuity, the $(Y, Z) \to (-Y, -Z)$ transformation is replaced by the comparison to the joint distribution generated by complements. □

Proof of Corollary 3. The corollary is proved by first showing the unique convergence, i.e., the BP recursion converges to the same symmetric fixed point for all non-trivial initializations. Once it is proved, the non-existence of asymmetric fixed point follows naturally from the fact that any fixed point is the limiting distribution of the BP recursion with itself as the initialization.

Consider any fixed non-trivial initialization μ. We denote its recursion by $\tilde{\mu}_{(h)} \triangleq \mathcal{Q}_{\mathrm{s}}^h\mu$. As a reference, we also consider the BP recursion in the setting with perfect leaf observation, which is given by $\mu_{(h)} = \mathcal{Q}_{\mathrm{s}}^h B_0$. Note that $\tilde{\mu}_{(h)}$ can be interpreted as the LLR distributions with noisy leaf observations. By the natural coupling, we have $B_{\frac{1}{2}} \preceq \tilde{\mu}_{(h)} \preceq \mu_{(h)}$. Recall that the sandwich theorem of degradation holds for general LLR distributions. This implies the needed convergence if $\mu_{(h)}$ converges to the trivial distribution.

For the other case, we have that $\mu_{(h)}$ converges to a non-trivial symmetric fixed point μ^* by the statement of Theorem 1. When the limit of $\mu_{(h)}$ is non-trivial, we need to replace $B_{\frac{1}{2}}$ with degradation bounds that also converge to μ^*. To that end, we extend the notion of degradation index. Because μ^* is symmetric, we can define $\phi^*(\mu^*, \tilde{\mu}_{(0)})$ exactly the same way as in equation (9), and we let $\underline{\mu}_{(h)}$ be the recursion initialized by

$$\underline{\mu}_{(0)} \triangleq B_{\phi^*(\mu^*,\tilde{\mu}_{(0)})} \circledast \mu^*.$$

The first and the second property in Proposition 14 hold under this extension, implying that $\underline{\mu}_{(0)}$ is non-trivial and $\underline{\mu}_{(0)} \preceq \mu$. The proof for the first property can be generalized by choosing ϕ in the proof of in Proposition 14 as follows.

$$\phi = \inf_{P_{Z|Y}} \max\left\{\mathbb{P}_{Y\sim\nu}[Z=0], \mathbb{P}_{Y\sim\nu^-}[Z=1]\right\},$$

where ν^- is the complement distribution of ν. From elementary statistics, it is known that B_ϕ is the LLR distribution of a thresholding quantized version of the estimation problem characterized by ν, where the quantizer is chosen to achieve the minimax error probability. Therefore, we have $B_\phi \preceq \nu$, and the rest of the proof follows the same steps.

Because degradation is still preserved under BP recursion for general distributions. We have

$$\underline{\mu}_{(h)} = \mathcal{Q}^h \underline{\mu}_{(0)} \preceq \mathcal{Q}^h \mu = \tilde{\mu}_{(h)}.$$

Note that $\underline{\mu}_{(0)}$ is symmetric. The uniqueness theorem states that the the constructed lower bounds converge to μ^*. Then the unique convergence of $\mathcal{Q}^h\mu$ follows from the sandwich theorem. □

APPENDIX D
PROOF OF THEOREM 37

Note that the definition of degradation index remain unchanged. Theorem 37 can be proved using the same steps upon the following proposition.[7]

Proposition 55. *For any symmetric μ and ν,*

1) *if $\nu \preceq \mu$, then $\mathcal{Q}_{\mathrm{L}}\nu \preceq \mathcal{Q}_{\mathrm{L}}\mu$;*
2) *we have $B_\phi \circledast \mathcal{Q}_{\mathrm{L}}\mu \prec \mathcal{Q}_{\mathrm{L}}(B_\phi \circledast \mu)$ for any $\phi \in (0, \frac{1}{2})$;*
3) *if ν is nontrivial, then $\phi^*(\mathcal{Q}_{\mathrm{L}}\mu, \mathcal{Q}_{\mathrm{L}}\nu) < \phi^*(\mu, \nu)$ or $\mu \preceq \nu$.*

Proof. The first statement in Proposition 55 is proved by viewing $\mathcal{Q}_{\mathrm{L}}$ as a limit of $\mathcal{Q}$ or $\mathcal{Q}_{\mathrm{s}}$. The third statement can be proved using the second statement and Proposition 14. Therefore, we focus on the proof for the second statement, which is obtained by deriving and comparing related β-curve functions. For brevity, we shall ignore cases where μ is trivial or $P_{\overline{d}}$ is a delta distribution at 0, where the statement is obviously true. Hence, the LHS of the needed inequality is non-trivial, and it suffices to examine the β-curves on both sides due to Proposition 20.

We first focus on the case where μ_{s} is trivial, which implies that any $\mathcal{Q}_{\mathrm{L}}\mu$ is a mixture of Gaussians, and their β-curves can be written using integrals of elementary functions. We have the following statement, which is derived from the fact that β-curves are linear functions of the errors in Bayes estimation. The error function for each fixed prior distribution is the minimum of estimation error over all threshold decoders, and the optimal threshold is the LLR that is consistent with the prior.

Proposition 56. *For any $t \in [0,1]$ and $s \in [0, +\infty]$, we have*

$$\beta(t; \mathcal{N}(s)) = \max_{r\in[0,+\infty]} \frac{1}{2}\left((1-t)\mathrm{erf}\left(\frac{\sqrt{s}}{2\sqrt{2}} - r\right) + (1+t)\mathrm{erf}\left(\frac{\sqrt{s}}{2\sqrt{2}} + r\right)\right),$$

where $\mathrm{erf}(z) \triangleq \frac{2}{\sqrt{\pi}}\int_0^z e^{-x^2}dx$, and a maximizer is given by $r = r^*(s,t) \triangleq \sqrt{\frac{2}{s}}\mathrm{arctanh}(t)$. Moreover, for any fixed $t \in [0,1]$,

$$\frac{d}{d\sqrt{s}}\beta(t; \mathcal{N}(s)) = \sqrt{\frac{1-t^2}{2\pi}} e^{-r^*(s,t)^2 - \frac{s}{8}}.$$

Due to convexity of arctanh, we have $r^*(\theta^2 s, \theta t) \leq r^*(s,t)$ for any $\theta \in (0,1)$ and $t \in [0,1]$. Therefore, the following proposition is implied by the Lagrange mean value theorem.

[7] Except for the existence of non-trivial fixed points, which can be proved by analyzing the evolution of V_{μ_h} with noiseless initialization.

Proposition 57. *The following inequality holds for any $\theta \in (0,1)$, $t \in [0,\theta]$, and $s \in (0,+\infty]$.*

$$\beta(t;\mathcal{N}(\theta^2 s)) > \theta\beta(t/\theta;\mathcal{N}(s)).$$

From linearity of β-curves, for μ_{s} being the trivial distribution, we have

$$\beta(t;\mathcal{Q}_{\mathrm{L}}\mu) = \mathbb{E}_{P_{\overline{d}}}\left[\beta(t;\mathcal{N}\left(\overline{d}\cdot V_\mu\right)\right].$$

For brevity, let $\theta \triangleq (1-2\phi)^2$. Note that $V_{B_\phi \boxast \mu} = \theta^2 V_\mu$ and recall Proposition 21. For any $t \in [0,\theta]$ we have

$$\beta(t;B_\phi \boxast \mathcal{Q}_{\mathrm{L}}\mu) = \theta\mathbb{E}_{P_{\overline{d}}}\left[\beta(t/\theta;\mathcal{N}\left(\overline{d}\cdot V_\mu\right)\right],$$
$$\beta(t;\mathcal{Q}_{\mathrm{L}}(B_\phi \boxast \mu)) = \mathbb{E}_{P_{\overline{d}}}\left[\beta(t;\mathcal{N}\left(\overline{d}\cdot \theta^2 V_\mu\right)\right].$$

Recall that we only need to provide a prove for $\mathbb{P}[\overline{d}>0]>0$ and non-trivial μ. We have $V_\mu > 0$ and $t_{\max}(B_\phi \boxast \mathcal{Q}_{\mathrm{L}}\mu) = \theta$. By an integration argument, the above results implies that

$$\beta(t;\mathcal{Q}_{\mathrm{L}}(B_\phi \boxast \mu)) > \beta(t;B_\phi \boxast \mathcal{Q}_{\mathrm{L}}\mu)$$

for any $t \in [0, t_{\max}(B_\phi \boxast \mathcal{Q}_{\mathrm{L}}\mu)]$, which proves the needed statement.

Now we take non-trivial μ_{s} into account. Let $\mathcal{Q}_{\mathrm{L},\delta_0}\mu \triangleq \mathbb{E}_{\overline{d}\sim P_{\overline{d}}}\left[\mathcal{N}\left(\overline{d}\cdot V_\mu\right)\right]$, we have that $\mathcal{Q}_{\mathrm{L}}\mu = \mathcal{Q}_{\mathrm{L},\delta_0}\mu * \mu_{\mathrm{s}}$. The earlier proof steps have essentially shown that

$$B_\phi \boxast \mathcal{Q}_{\mathrm{L},\delta_0}\mu \prec \mathcal{Q}_{\mathrm{L},\delta_0}(B_\phi \boxast \mu)$$

for any $\phi \in (0,\frac{1}{2})$. Compared to the case with fixed P_d, the above inequality exactly leads to a form that corresponds to inequality (32). Therefore, the proof can be completed using the same steps from inequality (33) to inequality (36). ☐

APPENDIX E
PROOF OF THEOREM 45

In this appendix, we prove the intermediate results for the uniqueness theorem stated in Section V. For brevity, we shall assume $\delta_e < \frac{1}{2}$ throughout this section, as the more general cases produce the same set of BP operators. We start by investigating the case of deterministic T, where a generalized version of inequality (41) can be proved based on the following concepts.

Definition 58. For any element tree T and parameter $r>0$, we define the *polygon number* of any growing point v to be essential infimum of the LLR message returned by the BP algorithm, constrained on the input message at v being r and at all other growing points being $-r$. We define the polygon number of the pair (T,r) to be maximum polygon number over all growing points, and denote it by $p(T,r)$.

Proposition 59. *For any $r>0$, an element tree T satisfies the polygon condition if and only if $p(T,r)<0$.*

Proof. Recall Definition 44. By the monotonicity of BP, any growing point is domiant if and only if their polygon number non-negative. Thus, the polygon coditon holds if and only if all of them are negative, which is equivalent to having $p(T,r)<0$. ☐

Proposition 60. *Let $\mathcal{Q}_T$ denote the BP operator for any deterministic element tree T. Then for any non-trivial symmetric μ and any $\phi \in (0,\frac{1}{2})$, we have*

$$B_\phi \boxast \mathcal{Q}_T\mu \prec_p \mathcal{Q}_T(B_\phi \boxast \mu), \tag{60}$$

where $p = p(T, r_{\max}(B_\phi \boxast \mu))$.

Proof. We prove the proposition by induction over the size of the tree. For the base cases where the element tree only has one single vertex, there are two possibilities. If the tree has a growing point, we have $p = r_{\max}(B_\phi \boxast \mu)$ and $\mathcal{Q}_T$ being the identity function. In this case, inequality (60) reduces to $\nu \prec_{r_{\max}(\nu)} \nu$ with $\nu = B_\phi \boxast \mu$, which directly follows from the definition of $\prec_p$. Otherwise, let μ_o denote the survey distribution on the only vertex. We have $p = -r_{\max}(\mu_o)$ and $\mathcal{Q}_T$ being the constant operator that returns μ_o. In this case, inequality (60) reduces to $B_\phi \boxast \mu_o \prec_{-r_{\max}(\mu_o)} \mu_o$, which can be proved using the rule of box convolution on β-curves (Proposition 21).

Now we prove the induction step by considering three possible cases. Again, let o denote the root vertex. We first prove the inequality when μ_o is trivial and the root has at most one child. Note that the base cases have already covered one-vertex trees. The root has exactly one incident edge in this regime. We denote this edge by e, and the sub element tree rooted at the unique child by $\tilde{T}$. The induction assumption implies

$$B_\phi \boxast \mathcal{Q}_{\tilde{T}}\mu \prec_{\tilde{p}} \mathcal{Q}_{\tilde{T}}(B_\phi \boxast \mu),$$

where $\tilde{p} = p(\tilde{T}, r_{\max}(B_\phi \boxast \mu))$. It is clear that $p = F_{\delta_e}(\tilde{p})$ and $\mathcal{Q}_T\nu = B_{\delta_e} \boxast \mathcal{Q}_{\tilde{T}}\nu$ for any symmetric ν. Therefore, the needed inequality in this case can be obtained by box convolving both sides of the above inequality with B_{δ_e}, and the analysis follows from the commutativity of box convolution and the second property of Proposition 29.

For the second regime, we assume that μ_o is trivial, but the root has at least two children. We create a new element tree for each incident edge of the root, by removing all other incident edges and their subtrees, and we denote them by $\tilde{T}_1, ..., \tilde{T}_d$. Further, let j^* be any index that minimizes $r_{\max}(\mathcal{Q}_{T_j}\mu)$. We create another element tree, denoted by $\hat{T}_{j^*}$, to be the sub element tree of T that excludes everything from $\tilde{T}_{j^*}$, but not the root vertex. It is clear that $\mathcal{Q}_T\mu = (\mathcal{Q}_{\tilde{T}_{j^*}}\mu) * (\mathcal{Q}_{\hat{T}_{j^*}}\mu)$. Hence, the following is implied by Corollary 25.

$$B_\phi \circledast \mathcal{Q}_T\mu \prec_{p^*} (B_\phi \circledast \mathcal{Q}_{\tilde{T}_{j^*}}\mu) * (B_\phi \circledast \mathcal{Q}_{\hat{T}_{j^*}}\mu), \tag{61}$$

where $p^* = |F_\phi(r_{\max}(\mathcal{Q}_{\hat{T}_{j^*}}\mu)) - F_\phi(r_{\max}(\mathcal{Q}_{\tilde{T}_{j^*}}\mu))|$. Recall that both $\tilde{T}_{j^*}$ and $\hat{T}_{j^*}$ are not identical to the full tree in this regime. The induction assumptions can be applied to each, and we have the following inequalities due to that convolution preserves degradation (see Proposition 48).

$$\begin{aligned}(B_\phi \circledast \mathcal{Q}_{\tilde{T}_{j^*}}\mu) * (B_\phi \circledast \mathcal{Q}_{\hat{T}_{j^*}}\mu) &\preceq (\mathcal{Q}_{\tilde{T}_{j^*}}(B_\phi \circledast \mu)) * (B_\phi \circledast \mathcal{Q}_{\hat{T}_{j^*}}\mu) && (62)\\ &\preceq (\mathcal{Q}_{\tilde{T}_{j^*}}(B_\phi \circledast \mu)) * (\mathcal{Q}_{\hat{T}_{j^*}}(B_\phi \circledast \mu)) && (63)\\ &= \mathcal{Q}_T(B_\phi \circledast \mu).\end{aligned}$$

Note that the above steps and inequality (61) exactly complete a full chain of degradation. Similar to the basic settings, we can prove the overall strict inequality in β-curves by keeping track of the strict conditions for each individual step. We first look at inequality (63). By induction assumptions, we have

$$B_\phi \circledast \mathcal{Q}_{\hat{T}_{j^*}}\mu \prec_{\hat{p}} \mathcal{Q}_{\hat{T}_{j^*}}(B_\phi \circledast \mu),$$

where $\hat{p} = p(\hat{T}_{j^*}, r_{\max}(B_\phi \circledast \mu))$. Therefore, by choosing $\ell = r_{\max}(\mathcal{Q}_{\tilde{T}_{j^*}}(B_\phi \circledast \mu))$ for the rule of convolution, inequality (63) provides strict conditions for

$$s \in (\hat{p} - r_{\max}(\mathcal{Q}_{\tilde{T}_{j^*}}(B_\phi \circledast \mu)), r_{\max}(B_\phi \circledast \mathcal{Q}_{\hat{T}_{j^*}}\mu) - r_{\max}(\mathcal{Q}_{\tilde{T}_{j^*}}(B_\phi \circledast \mu))].$$

The lower boundary of the above set is no greater than p, because it is exactly the maximum polygon number over growing points that are within $\hat{T}_{j^*}$. As mentioned earlier, inequality (61) already provides non-zero gaps for $s \in (p^*, r_{\max}(B_\phi \circledast \mathcal{Q}_T\mu)]$. Given our choice of j^*, we also have $p^* = r_{\max}(B_\phi \circledast \mathcal{Q}_{\hat{T}_{j^*}}\mu) - r_{\max}(B_\phi \circledast \mathcal{Q}_{\tilde{T}_{j^*}}\mu)$. Therefore, when $\tilde{T}_{j^*}$ is simple, we have

$$\mathcal{Q}_{\tilde{T}_{j^*}}(B_\phi \circledast \mu) = B_\phi \circledast \mu = B_\phi \circledast \mathcal{Q}_{\tilde{T}_{j^*}}\mu,$$

and we have already established strict inequality for $s \in (p, r_{\max}(B_\phi \circledast \mathcal{Q}_T\mu)]$.

When $\tilde{T}_{j^*}$ is not simple, we also need to investigate inequality (62). Similarly, by applying rule of convolution to the induction assumption, with $\ell = r_{\max}(B_\phi \circledast \mathcal{Q}_{\hat{T}_{j^*}}\mu)$, and using the fact that β-curves are even functions for symmetric distributions, we obtain strict inequalities for

$$s \in (r_{\max}(B_\phi \circledast \mathcal{Q}_{\hat{T}_{j^*}}\mu) - r_{\max}(\mathcal{Q}_{\tilde{T}_{j^*}}(B_\phi \circledast \mu)), r_{\max}(B_\phi \circledast \mathcal{Q}_{\hat{T}_{j^*}}\mu) - p(\tilde{T}_{j^*}, r_{\max}(B_\phi \circledast \mu))).$$

To proceed, we need to prove a simple fact: for any $\tilde{T}_{j^*}$ being non-simple, we have

$$p(\tilde{T}_{j^*}, r_{\max}(B_\phi \circledast \mu)) < r_{\max}(B_\phi \circledast \mathcal{Q}_{\tilde{T}_{j^*}}\mu). \tag{64}$$

Let v be a growing point in $\tilde{T}_{j^*}$ that maximizes the polygon number for $r = r_{\max}(B_\phi \circledast \mu)$, and let $\delta_1, ..., \delta_k$ denote the parameters of the edges on the path from the root to v. We define $p_0 = r_{\max}(B_\phi \circledast B_{\delta_1} \circledast ... \circledast B_{\delta_k} \circledast \mu)$. We have $p_0 > p(\tilde{T}_{j^*}, r_{\max}(B_\phi \circledast \mu))$, because p_0 is exactly the output LLR message of the BP algorithm with the input at v being r and all other inputs setting to 0 (including the surveys). Here we used the commutativity of box convolution. On the other hand, note that $B_{\delta_1} \circledast ... \circledast B_{\delta_k} \circledast \mu \preceq \mathcal{Q}_{\tilde{T}_{j^*}}\mu$, because the first quantity corresponds to a network with strictly less observation points. Hence, $p_0 < r_{\max}(B_\phi \circledast \mathcal{Q}_{\tilde{T}_{j^*}}\mu)$, and inequality (64) is proved. This implies that the upper boundary of the set obtained from inequality (62) is greater than p^*. Thus, the union of all three intervals covers the entire $(p, r_{\max}(B_\phi \circledast \mathcal{Q}_T\mu)]$.

Having proved the induction step for any element tree with trivial μ_o, we consider the third regime where μ_o can be any symmetric distribution. Let $\tilde{T}$ be the element tree obtained by removing the survey channel at the root from T. We have $\mathcal{Q}_T\nu = \mu_o * \mathcal{Q}_{\tilde{T}}\nu$ for any symmetric ν, and we have already proved that $\mathcal{Q}_{\tilde{T}}$.

$$B_\phi \circledast \mathcal{Q}_{\tilde{T}}\mu \prec_{\tilde{p}} \mathcal{Q}_{\tilde{T}}(B_\phi \circledast \mu), \tag{65}$$

where $\tilde{p} = p(\tilde{T}, r_{\max}(B_\phi \circledast \mu))$. The above inequality leads to the following chain

$$B_\phi \circledast \mathcal{Q}_{\tilde{T}}\mu = B_\phi \circledast (\mu_o * \mathcal{Q}_{\tilde{T}}\mu) \preceq \mu_o * (B_\phi \circledast \mathcal{Q}_{\tilde{T}}\mu) \preceq \mu_o * \mathcal{Q}_{\tilde{T}}(B_\phi \circledast \mu) = \mathcal{Q}_T(B_\phi \circledast \mu).$$

The first inequality is a direct application of Proposition 32, and it states strict inequality in β-curves for

$$s \in (F_\phi(r_{\max}(\mathcal{Q}_{\tilde{T}}\mu)) - r_{\max}(\mu_o), r_{\max}(B_\phi \circledast \mathcal{Q}_{\tilde{T}}\mu)].$$

Apply rule of convolution to the second inequality with $\ell = r_{\max}(\mu_o)$, we obtain strict conditions for

$$s \in (\tilde{p} - r_{\max}(\mu_o), F_\phi(r_{\max}(\mathcal{Q}_{\tilde{T}}\mu)) - \ell].$$

By the definition of polygon number, the lower boundary of the above interval is exactly p. Hence, we have proved that the overall degradation implied by the above chain has strict inequalities in β-curves for all $s \in (p, r_{\max}(B_\phi \boxast \mathcal{Q}_{\tilde{T}}\mu)]$. ☐

We present another important fact.

Proposition 61. *For any distribution P_T, non-trivial symmetric point μ, and any $\phi \in (0, \frac{1}{2})$, let $r = r_{\max}(B_\phi \boxast \mu)$, we have*

$$\operatorname{esssup} F_T(r) \geq r. \tag{66}$$

Moreover, the above inequality is strict if and only if P_T is non-trivial.

Proof. The stated inequality is a direct consequence from the third statement in Proposition 46, by taking $r_{\max}$ on both sides and using the fact that μ is a BP fixed point. So, our proof will be focused on the strict condition. Specifically, it is clear that the equality condition holds for trivial P_T, because F_T is the identity function w.p.1 under such condition. We only need to prove the converse, showing that the equality condition implies the trivialness of P_T.

For convenience, we define

$$\alpha(s; \nu) \triangleq \beta\left(\tanh\frac{s}{2}; \nu\right) - \tanh\frac{|s|}{2}$$

for symmetric ν. Our proof relies on the following property, which is proved in Appendix E-A.

Proposition 62. *For any element tree T, symmetric ν, and $s \in (0, r_{\max}(\nu))$, we have*

$$\alpha\left(F_T(s); \mathcal{Q}_T\nu\right) \leq \tanh\frac{F_T(s)}{2} \cdot \alpha\left(s; \nu\right) / \tanh\frac{s}{2}\,, \tag{67}$$

where $\mathcal{Q}_T$ denotes the BP operator for the fixed T. The equality condition holds if and only if T has no more than one growing point and all survey distributions are trivial.

Assume the equality condition of inequality (66). We have $F_T(r) \leq r$ w.p.1. Note that $\alpha(s, \nu)$ is non-increasing for $s \geq 0$. This condition implies

$$\alpha\left(r; \mathcal{Q}_T\mu\right) \leq \alpha\left(F_T(r); \mathcal{Q}_T\mu\right).$$

Recall that μ is non-trivial and $\phi \in (0, \frac{1}{2})$. We have $r = F_\phi(r_{\max}(\mu)) \in (0, r_{\max}(\mu))$. Hence, one can apply Proposition 62 to the above inequality and obtain

$$\alpha\left(r; \mathcal{Q}_T\mu\right) \leq \tanh\frac{F_T(r)}{2} \cdot \alpha\left(r; \mu\right) / \tanh\frac{r}{2} \leq \alpha\left(r; \mu\right). \tag{68}$$

Because μ is a BP fixed point, two sides of the above inequalities are identical in expectation. Therefore, we need all equality conditions to hold in probability. First, $r \in (0, r_{\max}(\nu))$ implies $\alpha\left(r; \mu\right) > 0$. We must have $F_T(s) = s$ w.p.1 for the second step in inequality (68). Then, combined with the equality condition stated in Proposition 62, the tree T must be simple w.p.1, implying that P_T is trivial. ☐

Now we use these results to prove Theorem 45.

Proof. The first statement relies on the linearity of β-curves and Proposition 60. As proved earlier in this appendix (recall Proposition 59 and Proposition 60), any element tree T that satisfies the polygon condition with parameter $r = r_{\max}(B_\phi \boxast \mu)$ implies

$$B_\phi \boxast \mathcal{Q}_T\mu \prec_p \mathcal{Q}_T(B_\phi \boxast \mu)$$

for some $p = p(T, r) < 0$. From the fourth property of Proposition 23, we essentially have

$$B_\phi \boxast \mathcal{Q}_T\mu \prec \mathcal{Q}_T(B_\phi \boxast \mu).$$

Given this relation, any T that is representative implies strict inequalities in β-curves for all $s \in [0, r]$. By linearity, the same non-zero gap condition holds for two sides of inequality (41). Hence, the needed strict degradation is implied by the fact that $r_{\max}(B_\phi \boxast \mathcal{Q}_{P_T}\mu)$ is exactly r, as μ is a BP fixed point.

To prove that random self-concatenation leads to representative trees with non-zero probability, it suffices to find a subset of element trees with non-zero measure on P_T, such that all self-concatenations within this subset is representative for some fixed h. Recall that P_T is not simple. According to Proposition 61, we can find $\tilde{r} > r$ such that $\mathbb{P}[F_T(r) > \tilde{r}] > 0$. Hence, let $\mathcal{T}$ be the subset of element trees satisfying $F_T(r) > \tilde{r}$, the set $\mathcal{T}$ has non-zero measure.

Now we show that there is a fixed h such that all self-concatenations within $\mathcal{T}$ with h steps lead to element trees satisfying the polygon condition. Assume the contrary, by Definition 44, for each sufficiently large h there is a concatenation $T_{(h)}$ of trees from $\mathcal{T}$ and a growing point v at the leaf such that the polygon number of v is non-negative. We utilize the following proposition, proved in Appendix E-B, which essentially provides an upper bound of polygon numbers using $F_T(r)$.

Proposition 63. *Consider any element tree with k growing points and a list of non-negative numbers $r_0, r_1, ..., r_k$. Let s denote the essential supremum of the output LLR message for the BP algorithm with k inputs at the growing points given by $r_1, r_2, ..., r_k$, and let p denote the essential infimum with k inputs given by $r_0, -r_2, -r_3, ..., -r_k$. If $p \geq 0$, then $p+s \leq r_0+r_1$.*

Let L denote the path from the root to v, and let $T_1, ..., T_h$ denote the element trees that overlaps with L in the sequential order that the root belongs to T_1 and v belongs to T_h. Let o_{j-1} denote the root of T_j, and let o_h denote v. We consider the BP algorithm on $T_{(h)}$ where the input is r for v and $-r$ for all other growing points. Let p_j denote essential infimum of the message sent from vertex o_j. Because all inputs except for the one at v are negative and p_0 is non-negative in the end, the essential infimum for all messages sent from vertices on L must be non-negative, i.e., $p_0, ..., p_h \geq 0$.

Now we use Proposition 63 to derive a recursive relation between these quantities. Consider each T_j, let p_j be the input at its growing point on L (i.e., o_j). For each other growing point, let their input be the essential infimum of their message sent in the above process. Then the essential infimum of the message sent from o_{j-1} is exactly p_{j-1}. We view the input at o_j to be r_0, and the inputs at all other growing points to be $-r_2, -r_3, ..., -r_k$. Then the variable p in Proposition 63 is identical to p_{j-1}. Consider the other set of input configurations with $r_1 = r$. Because of the monotonicity of BP algorithms and $F_T(r) > r$ for all element trees in $\mathcal{T}$, all $r_2, ..., r_k$ are lower bounded by r. Again, due to monotonicity, we have $s \geq \tilde{r}$. Therefore, Proposition 63 states that $p_{j-1} \leq p_j - (\tilde{r} - r)$.

Because $p_h = r$ is finite, and $\tilde{r} - r$ is a fixed positive number given $\mathcal{T}$. There can be at most finitely many h for the recursion to holds with $p_0 \geq 0$, which leads to a contradiction. $\square$

A. Proof of Proposition 62

Note that any BP operator is a composition of convolution and box convolution. We prove the inequality by keeping track of all the composition steps. We first state the specialized statements for each building block in the following proposition.

Proposition 64. *Consider any symmetric ν and $s > 0$.*

1) *For any $\delta \in [0, 1]$, we have*

$$\alpha(F_\delta(s); B_\delta \boxast \nu) = \left(\tanh \frac{F_\delta(s)}{2}\right) \cdot \frac{\alpha(s; \nu)}{\tanh \frac{s}{2}}. \tag{69}$$

2) *For any symmetric τ and $r \geq 0$, we have*

$$\alpha(s + r; \nu * \tau) \leq \left(\tanh \frac{s + r}{2}\right) \cdot \frac{\alpha(s; \nu) + \alpha(r; \tau)}{\tanh \frac{s}{2} + \tanh \frac{r}{2}}, \tag{70}$$

where the equality condition holds if and only if either $r = 0$ and τ is trivial, or $\alpha(s; \nu) = \alpha(r; \tau) = 0$.

Proof. The first property follows from the rule of box convolution on β-curves (Proposition 21). To prove the second property, consider independent variables $R_\nu \sim \nu$ and $R_\tau \sim \tau$. We have

$$\alpha(s + r; \nu * \tau) = \mathbb{E}\left[\max\left\{\tanh \frac{|R_\nu + R_\tau|}{2} - \tanh \frac{s + r}{2}, 0\right\}\right].$$

By symmetry condition, the above quantity can be written as

$$\alpha(s + r; \nu * \tau) = \mathbb{E}[f(|R_\nu|, |R_\tau|)],$$

where

$$f(|R_\nu|, |R_\tau|) \triangleq \begin{cases} \tanh \max\left\{\frac{|R_\nu|}{2}, \frac{|R_\tau|}{2}\right\} - \tanh \frac{s+r}{2} & \text{if } s + r < ||R_\nu| - |R_\tau|| \\ \frac{1}{2}\left(\tanh \frac{|R_\nu| + |R_\tau|}{2} - \tanh \frac{s+r}{2}\right)\left(1 + \tanh \frac{|R_\nu|}{2} \tanh \frac{|R_\tau|}{2}\right) & \text{if } ||R_\nu| - |R_\tau|| \leq s + r < |R_\nu| + |R_\tau| \\ 0 & \text{otherwise} \end{cases}.$$

Note that the RHS of inequality (70) can be written similarly as the expectation of the following function

$$g(|R_\nu|, |R_\tau|) \triangleq \left(\tanh \frac{s + r}{2}\right) \cdot \left(\frac{\tanh \max\left\{\frac{|R_\nu|}{2}, \frac{s}{2}\right\} + \tanh \max\left\{\frac{|R_\tau|}{2}, \frac{r}{2}\right\}}{\tanh \frac{s}{2} + \tanh \frac{r}{2}} - 1\right).$$

By elementary calculus, we have

$$f(|R_\nu|, |R_\tau|) \leq g(|R_\nu|, |R_\tau|).$$

Then by taking the expectation, we have proved inequality (70).

When equality holds, we need $f(|R_\nu|, |R_\tau|) = g(|R_\nu|, |R_\tau|)$ w.p.1. We prove that either one of the equality conditions stated in Proposition 64 is needed for this statement. When $s \in (0, r_{\max}(\nu))$, we have $\mathbb{P}[|R_\nu| > s] > 0$. Then for any fixed $|R_\nu| > s$, the equality $f(|R_\nu|, |R_\tau|) = g(|R_\nu|, |R_\tau|)$ is equivalent to $r = 0$ and trivial τ. Consequently, the first condition is needed in this regime.

For the other case, we have $s \geq r_{\max}(\nu)$, which implies $\alpha(s; \nu) = 0$. By symmetry, one can prove that inequality (70) is strict for $r < r_{\max}(\tau)$, given that we assumed $s > 0$. Hence, in this case, equality also requires $r \geq r_{\max}(\tau)$, which implies $\alpha(r; \tau) = 0$. To summarize, the second equality condition is needed in this regime. □

Proof of Proposition 62. If T has no growing points, then $\mathcal{Q}_T$ returns a fixed distribution and $F_T(s) = r_{\max}(\mathcal{Q}_T \nu)$. Hence, the LHS of inequality (67) is zero. Then, because $s < r_{\max}(\nu)$, we have $\alpha(s; \nu)$ being positive, and the inequality is proved. The equality condition in this case is given by $F_T(s) = 0$. Recall that we assumed all edges to have non-zero capacity. This condition is equivalent to having all survey distributions be trivial, which is the same condition stated in the proposition.

In the rest of this proof, we assume T has at least one growing point. Again, we prove the stated inequality by induction over the size of the tree and the number of non-trivial survey distributions. Consider the base case where T only has one vertex. It has to be a single growing point. Therefore, we have $F_T(s) = s$ and $\mathcal{Q}_T$ being the identity function. Hence, the equality condition holds.

For the induction step, T has at least two vertices. We can choose one incident edge e of the root vertex, such that the corresponding subtree contains at least one growing point. We consider two possible cases. First, if the root has only one incident edge and μ_o is trivial, we let $\tilde{T}$ denote the subtree obtained by removing the the root and edge e. The BP operator can be written as $\mathcal{Q}_T \nu = B_{\delta_e} \boxast \mathcal{Q}_{\tilde{T}} \nu$ and $F_T(s) = F_{\delta_e}(F_{\tilde{T}}(s))$. Recall that $\tilde{T}$ has at least one growing point and all edges have non-zero capacity. We have $F_{\tilde{T}}(s) > 0$ for any $s > 0$. Hence, the rule of box convolution on β-curves implies

$$\alpha(F_T(s); \mathcal{Q}_T \nu) = \alpha(F_\delta(F_{\tilde{T}}(s)); B_\delta \boxast \mathcal{Q}_{\tilde{T}} \nu) = \left(\tanh \frac{F_T(s)}{2}\right) \cdot \frac{\alpha(F_{\tilde{T}}(s); \mathcal{Q}_{\tilde{T}} \nu)}{\tanh \frac{F_{\tilde{T}}(s)}{2}}.$$

By applying the induction assumption on $\tilde{T}$, the RHS of the above inequality is exactly upper bounded by the RHS of inequality (67). Furthermore, the equality holds if and only if $\tilde{T}$ has exactly one growing point and all survey distributions are trivial. This is consistent with the statement in Proposition 62.

For the other case, let T_1 denote the subtree consists of $\tilde{T}$, e, and a root with a trivial survey distribution; let T_2 denote the subtree that contains all other contents of T. We have $\mathcal{Q}_T \nu = \mathcal{Q}_{T_1} \nu * \mathcal{Q}_{T_2} \nu$ and $F_T(s) = F_{T_1}(s) + F_{T_2}(s)$. Because T_1 contains at least one growing point, we have $F_{T_1}(s) > 0$. Thus, we have the following expression from the second statement of Proposition 64.

$$\alpha(F_T(s); \mathcal{Q}_T \nu) = \alpha(F_{T_1}(s) + F_{T_2}(s); \mathcal{Q}_{T_1} \nu * \mathcal{Q}_{T_2} \nu) \leq \left(\tanh \frac{F_T(s)}{2}\right) \cdot \frac{\alpha(F_{T_1}(s); \mathcal{Q}_{T_1} \nu) + \alpha(F_{T_2}(s); \mathcal{Q}_{T_2} \nu)}{\tanh \frac{F_{T_1}(s)}{2} + \tanh \frac{F_{T_2}(s)}{2}}.$$

Note that both T_1 and T_2 contain either at least one less vertex or one less non-trivial survey distribution compared to T. Hence, induction assumptions can be applied, which exactly gives inequality (67). On the other hand, recall that $\alpha(s; \nu) > 0$ and $F_T(s) > 0$. The overall equality requires $\alpha(F_{T_1}(s); \mathcal{Q}_{T_1} \nu) + \alpha(F_{T_2}(s); \mathcal{Q}_{T_2} \nu) > 0$. Hence, to satisfy the equality condition of Proposition 64, we need $F_{T_2}(s) = 0$ and $\mathcal{Q}_{T_2} \nu$ is trivial. Because s is positive, these conditions imply that T_2 has no growing points, and all its survey distributions are trivial. Again, because $F_T(s) > 0$, overall equality requires the equality condition for the T_1 induction. Therefore, T_1 can have at most one growing point, and all its survey distributions are trivial as well. Combining the requirements on T_1 and T_2, the element tree T must have have at most one growing point and all trivial survey distributions.

□

B. Proof of Proposition 63

Proof. Let v_0 denote the root, and $L = (v_0, v_1, v_2, ..., v_\ell)$ denote the path from the root to the first growing point. When the inputs are given by $r_1, r_2, r_3, ..., r_k$, let q_j be the essential supremum of the summation of all messages received at v_j excluding the one received from the path. Because messages are degraded when passing through each edge, the final output is upper bounded by the summation of all external messages received on the path, i.e.,

$$s \leq r_1 + q_0 + ... + q_{k-1}.$$

When the inputs are given by $r_0, -r_2, -r_3, ..., -r_k$, let p_j denote the essential infimum of the message sent from r_j. Because $p \geq 0$ and all input messages except for the first growing point are non-positive. We must have $p_0, ..., p_k \geq 0$. Thus, the essential infimum of the message received at r_j from the path is no greater than p_{j+1}. By symmetry, the essential infimum of the summation of all other received messages at r_j is $-q_j$. Hence, we have $p_j \leq p_{j+1} - q_j$, which implies that

$$p \leq r_0 - q_0 - ... - q_{k-1}.$$

Combine these two bounds, we have $p + s \leq r_0 + r_1$. □

References

[1] W. Evans, C. Kenyon, Y. Peres, and L. J. Schulman, "Broadcasting on trees and the ising model," *The Annals of Applied Probability*, vol. 10, no. 2, pp. 410–433, 2000. [Online]. Available: http://www.jstor.org/stable/2667156

[2] E. Mossel, J. Neeman, and A. Sly, "Belief propagation, robust reconstruction and optimal recovery of block models," *The Annals of Applied Probability*, vol. 26, no. 4, pp. 2211 – 2256, 2016. [Online]. Available: https://doi.org/10.1214/15-AAP1145

[3] V. Kanade, E. Mossel, and T. Schramm, "Global and local information in clustering labeled block models," *IEEE Transactions on Information Theory*, vol. 62, no. 10, pp. 5906–5917, 2016.

[4] E. Mossel and J. Xu, "Local algorithms for block models with side information," in *Proceedings of the 2016 ACM Conference on Innovations in Theoretical Computer Science*, ser. ITCS '16. New York, NY, USA: Association for Computing Machinery, 2016, p. 71–80. [Online]. Available: https://doi.org/10.1145/2840728.2840749

[5] E. Abbe, C. Elisabetta, Y. Gu, and Y. Polyanskiy, "Stochastic block model entropy and broadcasting on trees with survey," *arXiv preprint arXiv:2101.12601*, Jan. 2021.

[6] M. Mezard and A. Montanari, *Information, physics, and computation*. Oxford University Press, 2009.

[7] Q. Yu and Y. Polyanskiy, "Broadcasting on trees near criticality: Perturbation theory," in *2021 IEEE International Symposium on Information Theory (ISIT)*, 2021, pp. 2101–2106.

[8] A. Coja-Oghlan, F. Krzakala, W. Perkins, and L. Zdeborová, "Information-theoretic thresholds from the cavity method," *Advances in Mathematics*, vol. 333, pp. 694–795, 2018.

[9] A. Dembo, A. Montanari, and N. Sun, "Factor models on locally tree-like graphs," *The Annals of Probability*, vol. 41, no. 6, pp. 4162–4213, 2013.

[10] M. Mézard, G. Parisi, and M. A. Virasoro, *Spin glass theory and beyond: An Introduction to the Replica Method and Its Applications*. World Scientific Publishing Company, 1987, vol. 9.

[11] A. Montanari, E. Mossel, and A. Sly, "The weak limit of ising models on locally tree-like graphs," *Probability Theory and Related Fields*, vol. 152, no. 1, pp. 31–51, 2012. [Online]. Available: https://doi.org/10.1007/s00440-010-0315-6

[12] 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, 1995. [Online]. Available: https://doi.org/10.1007/BF02179399

[13] S. Roch and K.-C. Wang, "Sufficient condition for root reconstruction by parsimony on binary trees with general weights," *Electronic Communications in Probability*, vol. 26, no. none, pp. 1 – 13, 2021. [Online]. Available: https://doi.org/10.1214/21-ECP423

[14] C. Daskalakis, E. Mossel, and S. Roch, "Evolutionary trees and the ising model on the bethe lattice: a proof of steel's conjecture," *Probability Theory and Related Fields*, vol. 149, no. 1-2, pp. 149–189, 2011.

[15] Y. Gu, H. Roozbehani, and Y. Polyanskiy, "Broadcasting on trees near criticality," in *2020 IEEE International Symposium on Information Theory (ISIT)*, 2020, pp. 1504–1509.

[16] R. Lyons, "Random walks and percolation on trees," *The Annals of Probability*, vol. 18, no. 3, pp. 931–958, 1990. [Online]. Available: http://www.jstor.org/stable/2244410

[17] S. Janson and E. Mossel, "Robust reconstruction on trees is determined by the second eigenvalue," *Ann. Probab.*, vol. 32, no. 3B, pp. 2630–2649, 07 2004. [Online]. Available: https://doi.org/10.1214/009117904000000153

[18] A. Decelle, F. Krzakala, C. Moore, and L. Zdeborová, "Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications," *Physical Review E*, vol. 84, no. 6, p. 066106, 2011.

[19] E. Mossel, J. Neeman, and A. Sly, "Reconstruction and estimation in the planted partition model," *Probability Theory and Related Fields*, vol. 162, no. 3, pp. 431–461, 2015. [Online]. Available: https://doi.org/10.1007/s00440-014-0576-6

[20] T. Richardson and R. Urbanke, *Modern coding theory*. Cambridge university press, 2008.

[21] M. Edelstein, "On fixed and periodic points under contractive mappings," *Journal of the London Mathematical Society*, vol. s1-37, no. 1, pp. 74–79, 1962. [Online]. Available: https://londmathsoc.onlinelibrary.wiley.com/doi/abs/10.1112/jlms/s1-37.1.74

[22] U. Rösler, "A fixed point theorem for distributions," *Stochastic Processes and their Applications*, vol. 42, no. 2, pp. 195–214, 1992. [Online]. Available: https://www.sciencedirect.com/science/article/pii/030441499290035O

[23] U. Rösler, "A limit theorem for "quicksort"," *Applications/Theoretical Informatics and Applications*, vol. 25, pp. 85–100, 1999.

[24] R. Neininger and L. Rüschendorf, "A general limit theorem for recursive algorithms and combinatorial structures," *The Annals of Applied Probability*, vol. 14, no. 1, pp. 378–418, 2004.

[25] V. M. Zolotarev, "Approximation of distributions of sums of independent random variables with values in infinite-dimensional spaces," *Theory of Probability & Its Applications*, vol. 21, no. 4, pp. 721–737, 1977. [Online]. Available: https://doi.org/10.1137/1121086

[26] D. Blackwell, "Comparison of experiments," in *2nd Berkeley Symposium on Mathematical Statistics and Probability*, 1951, pp. 93–102.

[27] S. Sherman, "On a theorem of hardy, littlewood, polya, and blackwell," in *Proceedings of the National Academy of Sciences of the United States of America*, vol. 37, no. 12, 1951, pp. 826–831.

[28] C. Stein, "Notes on a seminar on theoretical statistics. i. comparison of experiments," *Report, University of Chicago*, 1951.

[29] D. Blackwell, "Equivalent comparisons of experiments," *The annals of mathematical statistics*, pp. 265–272, 1953.

[30] E. Mossel, J. Neeman, and A. Sly, "A proof of the block model threshold conjecture," *Combinatorica*, vol. 38, no. 3, pp. 665–708, 2018. [Online]. Available: https://doi.org/10.1007/s00493-016-3238-8

[31] T. Richardson and R. Urbanke, "Fixed points and stability of density evolution," *Communications in Information & Systems*, vol. 4, no. 1, pp. 103–116, 2004.

[32] T. Richardson, M. Shokrollahi, and R. Urbanke, "Design of capacity-approaching irregular low-density parity-check codes," *IEEE Transactions on Information Theory*, vol. 47, no. 2, pp. 619–637, 2001.

[33] C. E. Shannon, "A note on a partial ordering for communication channels," *Information and control*, vol. 1, no. 4, pp. 390–397, 1958.

[34] T. Cover, "Broadcast channels," *IEEE Transactions on Information Theory*, vol. 18, no. 1, pp. 2–14, 1972.

[35] E. Arikan, "Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *IEEE Transactions on Information Theory*, vol. 55, no. 7, pp. 3051–3073, 2009.

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

[37] K. Ou, "A numerical study on a noisy order of channels and information quantities," Master's thesis, The University of Electro-Communications, 2012.

[38] 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.