Proceedings of Machine Learning Research vol 195:1–36, 2023 36th Annual Conference on Learning Theory

# Weak Recovery Threshold for the Hypergraph Stochastic Block Model

**Yuzhou Gu** YUZHOUGU@MIT.EDU
*Masaschusetts Institute of Technology*

**Yury Polyanskiy** YP@MIT.EDU
*Masaschusetts Institute of Technology*

**Editors:** Gergely Neu and Lorenzo Rosasco

## Abstract

We study the weak recovery problem on the $r$-uniform hypergraph stochastic block model ($r$-HSBM) with two balanced communities. In HSBM a random graph is constructed by placing hyperedges with higher density if all vertices of a hyperedge share the same binary label, and weak recovery asks to recover a non-trivial fraction of the labels. We introduce a multi-terminal version of strong data processing inequalities (SDPIs), which we call the multi-terminal SDPI, and use it to prove a variety of impossibility results for weak recovery. In particular, we prove that weak recovery is impossible below the Kesten-Stigum (KS) threshold if $r = 3, 4$, or a strength parameter $\lambda$ is at least $\frac{1}{5}$. Prior work Pal and Zhu (2021) established that weak recovery in HSBM is always possible above the KS threshold. Consequently, there is no information-computation gap for these cases, which (partially) resolves a conjecture of Angelini et al. (2015). To our knowledge this is the first impossibility result for HSBM weak recovery.

As usual, we reduce the study of non-recovery of HSBM to the study of non-reconstruction in a related broadcasting on hypertrees (BOHT) model. While we show that BOHT's reconstruction threshold coincides with KS for $r = 3, 4$, surprisingly, we demonstrate that for $r \geq 7$ reconstruction is possible also below KS. This shows an interesting phase transition in the parameter $r$, and suggests that for $r \geq 7$, there might be an information-computation gap for the HSBM. For $r = 5, 6$ and large degree we propose an approach for showing non-reconstruction below KS, suggesting that $r = 7$ is the correct threshold for onset of the new phase.

**Keywords:** hypergraph stochastic block model, weak recovery, broadcasting on hypertrees, multi-terminal strong data processing inequalities, information-computation gap

## 1. Introduction

**Hypergraph stochastic block model.** The stochastic block model (SBM) is a random graph model with community structures. It exhibits many interesting behaviors and has received a lot of attention in the last decade (see Abbe (2017) for a survey). The hypergraph stochastic block model (HSBM) is a generalization of SBM to hypergraphs, which arguably models real social networks better due to the existence of small clusters. It was first considered in Ghoshdastidar and Dukkipati (2014) and has been studied in a number of works, e.g., Angelini et al. (2015); Ghoshdastidar and Dukkipati (2015a,b, 2017); Chien et al. (2018, 2019); Lin et al. (2017); Ahn et al. (2018); Kim et al. (2018); Cole and Zhu (2020); Pal and Zhu (2021); Dumitriu et al. (2021); Zhang and Tan (2022); Zhang et al. (2022); Dumitriu and Wang (2023).

We consider the $r$-uniform HSBM, where all hyperedges have the same size $r$. The model has two parameters $a > b \in \mathbb{R}_{\geq 0}$. The HSBM hypergraph is generated as follows: Let the vertex set be $V = [n]$. Generate a random label $X_u$ for all vertices $u \in V$ i.i.d. $\sim \operatorname{Unif}(\{\pm\})$. Then, for every $S \in \binom{V}{r}$, if all vertices in $S$ have the same label, add hyperedge $S$ with probability $\frac{a}{\binom{n}{r-1}}$;

otherwise add hyperedge $S$ with probability $\frac{b}{\binom{n}{r-1}}$. We denote the model as $\mathrm{HSBM}(n, 2, r, a, b)$ (where 2 means there are two communities).

For SBM and HSBM the most important problem is to recover $X$ from observing only $G$. Due to symmetry in the labels, we can only hope for recovering the communities up to a global sign flip. Thus we define the distance between two labelings $X, Y \in \{\pm\}^V$ as

$$d_H(X, Y) = \min_{s\in\{\pm\}} \sum_{u\in V} \mathbb{1}\{X_i \neq sY_i\}. \tag{1}$$

There are three kinds of recovery guarantees commonly seen in the literature.

- Exact recovery (strong consistency): The goal is to recover the labels exactly, i.e., to design an estimator $\widehat{X} = \widehat{X}(G)$ such that

$$\lim_{n\to\infty} \mathbb{P}[d_H(\widehat{X}, X) = 0] = 1. \tag{2}$$

- Almost exact recovery (weak consistency): The goal is the recover almost all labels, i.e., to design an estimator $\widehat{X} = \widehat{X}(G)$ such that

$$\lim_{n\to\infty} \mathbb{P}[d_H(\widehat{X}, X) = o(n)] = 1. \tag{3}$$

- Weak recovery (partial recovery): The goal is to recover a non-trivial fraction of the labels, i.e., to design an estimator $\widehat{X} = \widehat{X}(G)$ such that there exists a constant $c < \frac{1}{2}$ such that

$$\lim_{n\to\infty} \mathbb{P}[d_H(\widehat{X}, X) \le (c + o(1))n] = 1. \tag{4}$$

  Note that a trivial algorithm achieves $c = \frac{1}{2}$.

Different recovery questions are relevant in different parameter regimes. For exact recovery and almost exact recovery, the phase transition occurs at expected degree of order $\log n$ (i.e., $a, b = \Theta(\log n)$ grows with $n$). In this paper, we focus on the constant degree regime ($a, b$ are absolute constants), where the weak recovery problem is relevant.

The phase transition for exact recovery is known Kim et al. (2018); Zhang and Tan (2022) for more general HSBMs. For weak recovery, Angelini et al. (2015) conjectured that a phase transition occurs at the Kesten-Stigum threshold. The positive (algorithm) part of their conjecture has been proved by Pal and Zhu (2021); Stephan and Zhu (2022) in vast generality, giving an efficient weak recovery algorithm above the Kesten-Stigum threshold. Despite the progress on the positive part, to the best of our knowledge, there are no negative (impossibility) results for any $r \ge 3$ before the current work.

For the graph (SBM) case $r = 2$, the positive part was proved by Massoulié (2014); Mossel et al. (2018) and the negative part was established by Mossel et al. (2015, 2018) via reduction to the broadcasting on trees (BOT) model. Therefore a natural idea is to study the reconstruction problem for a suitable hypergraph generalization of the BOT model, which we call the broadcasting on hypertrees (BOHT) model. Zhang et al. (2022) mentioned that the difficulty in proving negative results lies in analyzing the BOHT model. In this paper we prove impossibility of weak recovery results by proving non-reconstruction results for BOHT.

Before describing the BOHT model, we define the following useful parameters for HSBM.

- For every vertex $u$, the expected number of hyperedges containing $u$ is $d \pm o(1)$, where

$$d = \frac{(a-b) + 2^{r-1}b}{2^{r-1}}. \tag{5}$$

- Expected number of vertices adjacent to $u$ is $\alpha \pm o(1)$, where

$$\alpha = (r-1)d = (r-1)\frac{(a-b) + 2^{r-1}b}{2^{r-1}}. \tag{6}$$

- Expected number of neighbors in the same community minus the number of neighbors in the other community is $\beta \pm o(1)$ where

$$\beta = (r-1)\frac{a-b}{2^{r-1}}. \tag{7}$$

- Strength of the broadcasting channel is characterized by $\lambda \in [0,1]$, defined as

$$\lambda = \frac{\beta}{\alpha} = \frac{a-b}{a-b+2^{r-1}b}. \tag{8}$$

- Signal-to-noise ratio (SNR), which is conjectured to govern the algorithmic weak recovery threshold for HSBM:

$$\mathrm{SNR} := \alpha\lambda^2 = (r-1)d\lambda^2 = \frac{(r-1)(a-b)^2}{2^{r-1}((a-b)+2^{r-1}b)}. \tag{9}$$

The Kesten-Stigum (KS) threshold is at $\mathrm{SNR} = 1$.

**Broadcasting on hypertrees.** We define a general broadcasting on hypertrees (BOHT) model. Let $q \ge 2$ (alphabet size), $r \ge 2$ (hyperedge size) be integers. Let $\pi \in \mathcal{P}([q])$ be a distribution of full support (where $\mathcal{P}([q])$ denotes the space of distributions on $[q]$). Let $B : [q] \to [q]^{r-1}$ be a probability kernel (called the broadcasting channel), satisfying

$$\sum_{k \in [q]} \pi_k \sum_{\substack{x \in [q]^{r-1} \\ x_i = j}} B(x_1, \dots, x_{r-1}|k) = \pi_j \quad \forall i \in [r-1], j \in [q]. \tag{10}$$

Let $T$ be a (possibly random) $r$-uniform linear[1] hypertree rooted at $\rho$. The model $\mathrm{BOHT}(T, q, r, \pi, B)$ generates a label $\sigma_u$ for every vertex $u \in T$ via a downward process: (1) generate $\sigma_\rho \sim \pi$ (2) given $\sigma_u$, for every downward hyperedge $S = \{u, v_1, \dots, v_{r-1}\}$, generate $\sigma_{v_1}, \dots, \sigma_{v_{r-1}}$ according to $B(\cdot|\sigma_u)$, i.e., for every $y, x_1, \dots, x_{r-1} \in [q]$, we have

$$\mathbb{P}[\sigma_{v_i} = x_i \forall i \in [r-1] | \sigma_u = y] = B(x_1, \dots, x_{r-1}|y). \tag{11}$$

We often consider the case where $T$ is a Galton-Watson hypertree, meaning that every vertex independently has $t \sim D$ downward hyperedges, where $D$ is a distrbution on $\mathbb{Z}_{\ge 0}$. We denote the resulting model as $\mathrm{BOHT}(q, r, \pi, B, D)$. An important case is $D = \mathrm{Pois}(d)$, the Poisson distribution with mean $d$. When $D$ is a singleton at $d \in \mathbb{Z}_{\ge 0}$ we also denote the model as $\mathrm{BOHT}(q, r, \pi, B, d)$.

1. Linear means that the intersection of two distinct hyperedges has size at most one.

For $\mathrm{HSBM}(n,2,r,a,b)$, the corresponding BOHT model has $q=2$, $\pi=\mathrm{Unif}(\{\pm\})$, and $B=B_{r,\lambda}$ ($\lambda\in[0,1]$ is given by (8)) where

$$B_{r,\lambda}(x_1,\dots,x_{r-1}|y)=\begin{cases}\lambda+\frac{1}{2^{r-1}}(1-\lambda), & \text{if } x_i=y\forall i\in[r-1],\\ \frac{1}{2^{r-1}}(1-\lambda), & \text{otherwise.}\end{cases} \tag{12}$$

We denote this model as $\mathrm{BOHT}(2,r,\lambda,D)$ and call it the special BOHT model.

The reconstruction problem asks whether we can gain any non-trivial information about the root given observation of far away vertices. In other words, whether the limit

$$\lim_{k\to\infty} I(\sigma_\rho;T_k,\sigma_{L_k}) \tag{13}$$

is non-zero, where $L_k$ is the set of vertices at distance $k$ to the root $\rho$, and $T_k$ is the set of vertices at distance $\le k$ to $\rho$. When the limit is non-zero, we say reconstruction is possible for the BOHT model; when the limit is zero, we say reconstruction is impossible. It is known Pal and Zhu (2021) that the $r$-neighborhood (for any constant $r$) of a random vertex converges (in the sense of local weak convergence) to the Poisson hypertree described above. Therefore non-reconstruction on a Poisson hypertree implies impossibility of weak recovery for the corresponding HSBM.

For the case $r=2$, the reconstruction threshold for the symmetric BOT model was established by Bleher et al. (1995); Evans et al. (2000). People have also studied generalizations of the BOT model with larger alphabet or asymmetric broadcasting channel, e.g., Mossel (2001); Mossel and Peres (2003); Mézard and Montanari (2006); Borgs et al. (2006); Bhatnagar et al. (2010); Sly (2009, 2011); Külske and Formentin (2009); Liu and Ning (2019); Gu and Polyanskiy (2020); Mossel et al. (2022). Nevertheless, to our knowledge, there has been no previous work studying the reconstruction problem for BOHT.

**Belief propagation.** The BOT and BOHT models can be studied using the belief propagation operator. Consider the model $\mathrm{BOHT}(q,r,\pi,B,D)$. Let $M_k$ denote the channel $\sigma_\rho\mapsto(T_k,\sigma_{L_k})$. Then $(M_k)_{k\ge 0}$ satisfies the following recursion, called the belief propagation recursion:

$$M_{k+1}=\mathbb{E}_{t\sim D}\left(M_k^{\times(r-1)}\circ B\right)^{\star t}, \tag{14}$$

where $(\cdot)^{\times(r-1)}$ denotes tensorization power, and $(\cdot)^{\star t}$ denotes $\star$-convolution power (see Section 2). Let BP be the operator

$$\mathrm{BP}(P):=\mathbb{E}_{t\sim D}\left(P^{\times(r-1)}\circ B\right)^{\star t} \tag{15}$$

defined on the space of channels with input alphabet $[q]$. Then the reconstruction problem is equivalent to asking whether the limit $\mathrm{BP}^\infty(\mathrm{Id}):=\lim_{k\to\infty}\mathrm{BP}^k(\mathrm{Id})$ is trivial, where Id stands for the identity channel $\mathrm{Id}(y|x)=\mathbb{1}\{x=y\}$.

**Strong data processing inequalities.** A useful tool for studying BOT models is the strong data processing inequalities (SDPIs). They are quantitative versions of the data processing inequality (DPI), the most fundamental inequality in information theory. The input-restricted version of SDPI states that for any Markov chain $U-X-Y$, we have

$$I(U;Y)\le\eta_{\mathrm{KL}}(P_X,P_{Y|X})I(U;X) \tag{16}$$

where $\eta_{\mathrm{KL}}(P_X, P_{Y|X})$ is a constant (called the contraction coefficient) depending only on $P_X$ and $P_{Y|X}$. We always have $\eta_{\mathrm{KL}}(P_X, P_{Y|X}) \le 1$ by DPI, and the inequality is usually strict. For any $f$-divergence, there is a corresponding version of SDPI, by replacing $I$ with $I_f$ and $\eta_{\mathrm{KL}}$ with another constant $\eta_f$ in (16).

To apply SDPI to reconstruction problems on trees, the following equivalent form is more useful: for any Markov chain $Y - X - U$, we have

$$I(U;Y) \le \eta^{(p)}_{\mathrm{KL}}(P_Y, P_{X|Y}) I(U;X), \tag{17}$$

where $\eta^{(p)}_{\mathrm{KL}}(P_Y, P_{X|Y})$ is a constant depending only on $P_Y$ and $P_{X|Y}$. (17) is called the post-SDPI in Polyanskiy and Wu (2023+). Comparing (16) and (17) we see that $\eta_{\mathrm{KL}}(P_X, P_{Y|X}) = \eta^{(p)}_{\mathrm{KL}}(P_Y, P_{X|Y})$.

Now consider a BOT model $\mathrm{BOHT}(q, 2, \pi, B, d)$. Note that in this case $B$ is a Markov kernel from $[q]$ to $[q]$ and $\pi B = \pi$. Then the post-SDPI says that for any channel $P$ with input alphabet $[q]$, we have

$$I(\pi, P \circ B) \le \eta^{(p)}_{\mathrm{KL}}(\pi, B) I(\pi, P), \tag{18}$$

where $I(\pi, P)$ denotes the mutual information $I(X;Y)$ between two variables where $P_X = \pi$ and $P_{Y|X} = P$. By subadditivity of mutual information under $\star$-convolution, we have

$$I(\pi, \mathrm{BP}(P)) \le dI(\pi, P \circ B) \le d\eta^{(p)}_{\mathrm{KL}}(\pi, B) I(\pi, P). \tag{19}$$

When $d\eta^{(p)}_{\mathrm{KL}}(\pi, B) < 1$, we have $\lim_{k\to\infty} I(\pi, \mathrm{BP}^k(P)) = 0$ and reconstruction is impossible.

This argument first appeared in Külske and Formentin (2009); Formentin and Külske (2009) with SKL information, and Gu et al. (2020) used it with mutual information to give currently best known non-reconstruction results for the Potts model in some parameter regimes. Although we introduced the method with $\mathrm{BOHT}(q, 2, \pi, B, d)$ model, with slight modification it works for $\mathrm{BOHT}(q, 2, \pi, B, D)$ or $\mathrm{BOHT}(T, q, 2, \pi, B)$.

**Multi-terminal SDPI.** We generalize the above method to BOHT models with $r \ge 3$. To this end, we introduce a multi-terminal version of the post-SDPI. Let $\pi \in \mathcal{P}([q])$ be a distribution and $B : [q] \to [q]^{r-1}$ be a probability kernel satisfying (10). We define the multi-terminal contraction coefficient $\eta^{(m)}_{\mathrm{KL}}(\pi, B)$ (where $m$ stands for "multi") as the smallest constant such that for any channel $P$ with input alphabet $[q]$, we have

$$I(\pi, P^{\times(r-1)} \circ B) \le (r-1)\eta^{(m)}_{\mathrm{KL}}(\pi, B) I(\pi, P). \tag{20}$$

(See Figure 1 for an illustration.) Then with a similar argument as the $r = 2$ case, we can prove non-reconstruction for BOHT whenever $(r-1)d\eta^{(m,s)}_{\mathrm{KL}}(B) < 1$.

In the single-terminal setting, we usually distinguish pre-SDPI and post-SDPI. In our multi-terminal setting, $B$ has one input and multiple outputs, so a multi-terminal version of post-SDPI makes more sense than that of pre-SDPI. Therefore we call Eq. 20 multi-terminal SDPI rather than multi-terminal post-SDPI.

For a BOHT model, if $q = 2$, $\pi = \mathrm{Unif}(\{\pm\})$, and $B : \{\pm\} \to \{\pm\}^{r-1}$ together with the sign flip $\{\pm\}^{r-1} \to \{\pm\}^{r-1}$ is a BMS channel (see Section 2), then we say the model is binary symmetric. For such models, the BP operator sends BMS channels to BMS channels. So we could restrict $P$ to be a BMS channel and define $\eta_{\mathrm{KL}}^{(m,s)}(B)$ (where $s$ stands for "symmetric") to be the smallest constant such that (20) holds for all BMS channels $P$. By definition $\eta_{\mathrm{KL}}^{(m,s)}(B) \le \eta_{\mathrm{KL}}^{(m)}(\pi, B)$, so it might be able to give better non-reconstruction results than the non-BMS version. Furthermore, due to a large number of tools dealing with BMS channels, $\eta_{\mathrm{KL}}^{(m,s)}(B)$ is often easier to compute than $\eta_{\mathrm{KL}}^{(m)}(\pi, B)$.

We could replace KL divergence in the above discussion by other $f$-divergences, and define the corresponding multi-terminal contraction coefficients. See Section 3 for more discussions.

**Our results.** Our first result is non-reconstruction for BOHT based on multi-terminal SDPIs.

**Theorem 1 (Non-reconstruction for BOHT)** *Consider the model* $\mathrm{BOHT}(q, r, \pi, B, D)$ *where* $\mathbb{E}_{t\sim D} t = d$.

*(i) If*

$$(r-1)d\eta_{\mathrm{KL}}^{(m)}(\pi, B) < 1, \tag{21}$$

$$\textit{or} \quad (r-1)d\eta_{\mathrm{SKL}}^{(m)}(\pi, B) < 1 \textit{ and } I_{\mathrm{SKL}}(\pi, B) < \infty, \tag{22}$$

*then reconstruction is impossible.*

*(ii) Suppose the BOHT model is binary symmetric. If*

$$(r-1)d\eta_{\mathrm{KL}}^{(m,s)}(B) < 1, \tag{23}$$

$$\textit{or} \quad (r-1)d\eta_{\chi^2}^{(m,s)}(B) < 1, \tag{24}$$

$$\textit{or} \quad (r-1)d\eta_{\mathrm{SKL}}^{(m,s)}(B) < 1 \textit{ and } C_{\mathrm{SKL}}(B) < \infty, \tag{25}$$

*then reconstruction is impossible.*

Our method can be modified to give non-reconstruction results for the $\mathrm{BOHT}(T, q, r, \pi, B)$ model. See Section A.

We apply Theorem 1 to the special case where $\pi = \mathrm{Unif}(\{\pm\})$ and $B = B_{r,\lambda}$. We compute $\eta_f^{(m,s)}(B_{r,\lambda})$ for several cases and prove the following result.

**Theorem 2 (Non-reconstruction for special BOHT)** *Consider the special BOHT model* $\mathrm{BOHT}(2, r, \lambda, D)$ *where* $\mathbb{E}_{t\sim D} t = d$.

*(i) For* $r = 3, 4$*, if* $(r-1)d\lambda^2 \le 1$*, then reconstruction is impossible.*

*(ii) For any* $r \ge 5$*, if*

$$(r-1)d \sup_{0<\epsilon\le 1} f_{r,\lambda}(\epsilon) < 1, \tag{26}$$

$$\textit{where} \quad f_{r,\lambda}(\epsilon) := \frac{1}{r-1} \sum_{1\le i\le r-1} \binom{r-1}{i} (1-\epsilon)^{r-1-i} \epsilon^{i-1} \frac{\lambda^2}{\lambda + (1-\lambda)2^{1-i}}, \tag{27}$$

*then reconstruction is impossible.*

*(iii) For any $r \geq 5$, if $\lambda \geq \frac{1}{5}$ and $(r-1)d\lambda^2 \leq 1$, then reconstruction is impossible.*

Note that Theorem 2(iii) is non-trivial even for $(r-1)\lambda^2 > 1$, because $\mathrm{BOHT}(2, r, \lambda, D)$ allows non-integer $d$ and the hypertree is infinite with positive probability whenever $(r-1)d > 1$. The constant $\frac{1}{5}$ in Theorem 2(iii) can be improved for any fixed $r$. For example, for $r = 5$, the constant can be improved to $\frac{1}{7}$.

By a standard reduction, our non-reconstruction results for BOHT implies impossibility of weak recovery for the corresponding HSBM.

**Theorem 3 (Impossibility of weak recovery for HSBM)** *Consider the model $\mathrm{HSBM}(n, 2, r, a, b)$. Let $\mathrm{BOHT}(2, r, \lambda, \mathrm{Pois}(d))$ be the corresponding BOHT model. If any of the conditions in Theorem 2(i)(ii)(iii) holds, then weak recovery is impossible.*

In fact, the reduction holds for more general HSBMs. See Section D.

It is known that for any BOHT model above the Kesten-Stigum threshold, reconstruction is possible. By Theorem 2, for the special BOHT model, the KS threshold is tight for $r = 3, 4$ or $\lambda \geq \frac{1}{5}$. Surprisingly, we show that for $r \geq 7$ and large $d$, reconstruction is possible below the KS threshold.

**Theorem 4 (Reconstruction for BOHT with $r \geq 7$ and large $d$)** *Consider the model $\mathrm{BOHT}(2, r, \lambda, d)$ or $\mathrm{BOHT}(2, r, \lambda, \mathrm{Pois}(d))$. For $r \geq 7$, there exists a constant $d_0 = d_0(r)$ such that for all $d \geq d_0$, there exists $\lambda \in [0, 1]$ such that $(r-1)d\lambda^2 < 1$ and reconstruction is possible.*

**Our technique.** Our main technique for proving the non-reconstruction results is the multi-terminal SDPIs. Theorem 1 is a simple application of the multi-terminal SDPIs and subadditivity properties (under $\star$-convolution) of the relevant information measures.

Theorem 2 is by applying Theorem 1 and computing the relevant multi-terminal contraction coefficients, except for the critical case $(r-1)d\lambda^2 = 1$. For Theorem 2(i) we compute the SKL multi-terminal contraction coefficients for $r = 3, 4$. For Theorem 2(ii) we compute the $\chi^2$-multi-terminal contraction coefficients. Theorem 2(iii) is a corollary of (ii). To compute the contraction coefficients, we write down an explicit description of the BP operator, and use properties of BMS channels such as BSC mixture representation and extremal BMS channels.

It turns out that for the special BOHT model, the tight reconstruction threshold can be achieved for $r = 3$ using SKL or $\chi^2$-contraction, and for $r = 4$ using SKL. KL contraction does not give tight threshold for any $r \geq 3$ and $\chi^2$-contraction fails for $r \geq 4$. In particular, there exists $\lambda \in (0, 1)$ such that $\eta_{\chi^2}^{(m,s)}(B_{4,\lambda}) > \eta_{\mathrm{SKL}}^{(m,s)}(B_{4,\lambda}) = \lambda^2$. This shows an important difference between single-terminal and multi-terminal SDPIs, because for single-terminal SDPIs we always have $\eta_{\chi^2}(\pi, P) \leq \eta_{\mathrm{SKL}}(\pi, P)$ (e.g., Cohen et al. (1998); Raginsky (2016); Polyanskiy and Wu (2017)). Furthermore, before our work, to the best of our knowledge, non-reconstruction results proved via SKL information could always be also shown via other information measures ($\chi^2$-information Evans et al. (2000), KL information Gu and Polyanskiy (2020), etc.). It appears, thus, that BOHT is the first example where contraction via SKL information gives better results than any other information measures we have tried.

For the critical case in Theorem 2(i)(iii), some extra argument is needed. Roughly speaking, we show that the multi-terminal SDPI achieves equality only when the input channel $P$ is trivial. So any fixed point of the BP operator must be trivial.

Theorem 3 is a corollary of Theorem 2 via a standard reduction which says non-reconstruction for BOHT implies impossibility of weak recovery for the corresponding HSBM. This reduction was first proved by Mossel et al. (2015, 2018) for the two-community SBM, and we extend it to handle general HSBM.

Theorem 4 is proved using Gaussian approximation at large $d$ and contraction of $\chi^2$-information. This is a method introduced by Sly (2009, 2011) and has proved successful in several settings Liu and Ning (2019); Mossel et al. (2022).

**Structure of the paper.** In Section 2, we review preliminaries on information channels. In Section 3, we introduce the multi-terminal SDPI and prove Theorem 1. In Section 4 we study the special BOHT model $\mathrm{BOHT}(2, r, \lambda, D)$ and prove Theorem 2. In Section 5, we discuss a possible approach to resolve the $r = 5, 6$ case of the special BOHT model.

In Section A, we prove non-reconstruction results for BOHT models on a fixed hypertree. In Section B, we compute SKL multi-terminal contraction coefficients for $B_{r,\lambda}$ with $r = 3, 4$. In Section C, we compute $\chi^2$-multi-terminal contraction coefficients for several binary-input symmetric channels. In Section D, we give a general reduction from HSBM to BOHT, and prove Theorem 3. In Section E, we prove Theorem 4, that the KS threshold is not tight for the special BOHT model with $r \geq 7$ and large $d$.

## 2. Preliminaries

We give necessary preliminaries on information channels, especially BMS channels. Most material in this section can be found in Polyanskiy and Wu (2023+) or (Richardson and Urbanke, 2008, Chapter 4).

**Definition 5 (BMS channels)** *A channel $P : \{\pm\} \to \mathcal{Y}$ is called a binary memoryless symmetric (BMS) channel if there exists a measurable involution $\sigma : \mathcal{Y} \to \mathcal{Y}$ such that $P(E|+) = P(\sigma(E)|-)$ for all measurable subsets $E \subseteq \mathcal{Y}$.*

Binary erasure channels (BECs) and binary symmetric channels (BSCs) are the simplest examples of BMS channels. Channel $B_{r,\lambda} : \{\pm\} \to \{\pm\}^{r-1}$ defined in (12) (together with coordinate-wise sign flip) is also naturally a BMS channel.

BMS channels are equivalent to distributions on the interval $\left[0, \frac{1}{2}\right]$, via the following lemma.

**Lemma 6 (BSC mixture representation of BMS channels)** *Every BMS channel $P$ is equivalent to a channel $X \to (\Delta, Z)$ where $\Delta \in \left[0, \frac{1}{2}\right]$ is independent of $X$ and $P_{Z|\Delta,X} = \mathrm{BSC}_\Delta(\cdot|X)$.*

In the setting of Lemma 6, we say $\Delta$ is the $\Delta$-component of $P$. We define $\theta = 1 - 2\Delta \in [0, 1]$ to be the $\theta$-component of $P$, because it sometimes simplifies the notation.

Degradation is a very useful relationship between channels.

**Definition 7 (Degradation)** *Let $P : \mathcal{X} \to \mathcal{Y}$ and $Q : \mathcal{X} \to \mathcal{Z}$ be two channels with the same input alphabet. We say $P$ is a degradation of $Q$, denoted $P \leq_{\deg} Q$, if there exists channel $R : \mathcal{Z} \to \mathcal{Y}$ such that $P = R \circ Q$.*

For any $f$-divergence, distrbution $\pi \in \mathcal{P}(\mathcal{X})$ and channel $P : \mathcal{X} \to \mathcal{Y}$, we define $I_f(\pi, P)$ as the $f$-information $I_f(X; Y)$ where $X$ is a random variable with distribution $\pi$ and $Y$ is the output of $P$ when given input $X$. Every $f$-information respects degradation: if $P \leq_{\deg} Q$, then

$I_f(\pi, P) \le I_f(\pi, Q)$ for any $f$ and $\pi$. For our purpose, the most important $f$-divergences are the KL divergence $f(x) = x \log x$, $\chi^2$-divergence $f(x) = (x-1)^2$, and symmetric KL (SKL) divergence $f(x) = (x-1)\log x$. We denote the corresponding $f$-information as $I$, $I_{\chi^2}$ and $I_{\mathrm{SKL}}$ respectively.

When $P$ is a BMS channel and $\pi = \mathrm{Unif}(\{\pm\})$, we use $C_f(P)$ to denote $I_f(\pi, P)$. We can compute $C_f(P)$ using the $\Delta$-component. In particular, we have the following information measures.

**Definition 8 (Information measures for BMS channels)** *Let $P$ be a BMS channel, $\Delta$ be its $\Delta$-component, and $\theta$ be its $\theta$-component. We define the following information measures.*

$$C(P) = \mathbb{E}[\log 2 + \Delta \log \Delta + (1-\Delta)\log(1-\Delta)], \qquad \text{(capacity)}$$

$$C_{\chi^2}(P) = \mathbb{E}\theta^2, \qquad (\chi^2\text{-capacity})$$

$$C_{\mathrm{SKL}}(P) = \mathbb{E}\left[\left(\frac{1}{2} - \Delta\right)\log\frac{1-\Delta}{\Delta}\right] = \mathbb{E}\left[\theta \operatorname{arctanh}\theta\right]. \qquad \text{(SKL capacity)}$$

Let $P : \mathcal{X} \to \mathcal{Y}$ and $Q : \mathcal{X}' \to \mathcal{Y}'$ be two channels. We define the tensor product channel $P \times Q : \mathcal{X} \times \mathcal{X}' \to \mathcal{Y} \times \mathcal{Y}'$ by letting $P$ and $Q$ acting on the two inputs independently. For $n \in \mathbb{Z}_{\ge 1}$, we use $P^{\times n} : \mathcal{X}^n \to \mathcal{Y}^n$ to denote the $n$-th tensor power of $P$.

Let $P : \mathcal{X} \to \mathcal{Y}$ and $Q : \mathcal{X} \to \mathcal{Z}$ be two channels with the same input alphabet. We define the $\star$-convolution $P \star Q : \mathcal{X} \to \mathcal{Y} \times \mathcal{Z}$ by letting $P$ and $Q$ acting on the same input independently. For $n \in \mathbb{Z}_{\ge 0}$, we use $P^{\star n} : \mathcal{X} \to \mathcal{Y}^n$ to denote the $n$-th $\star$-power of $P$.

Mutual information and SKL information are useful for the study of BOHT models because they are subadditive under $\star$-convolution (SKL information is even additive). For any two channels $P, Q$ with input alphabet $\mathcal{X}$ and any $\pi \in \mathcal{P}(\mathcal{X})$, we have

$$I(\pi, P \star Q) \le I(\pi, P) + I(\pi, Q), \qquad I_{\mathrm{SKL}}(\pi, P \star Q) = I_{\mathrm{SKL}}(\pi, P) + I_{\mathrm{SKL}}(\pi, Q). \tag{28}$$

The mutual information part is standard, and the SKL information part first appeared in Külske and Formentin (2009). For $\chi^2$-information, subadditivity does not hold in general, but Abbe and Boix-Adserà (2019) proved that subadditivity holds when $\mathcal{X} = \{\pm\}$ and $\pi = \mathrm{Unif}(\{\pm\})$. That is, for any two binary-input channels $P, Q$, we have

$$I_{\chi^2}(\mathrm{Unif}(\{\pm\}), P \star Q) \le I_{\chi^2}(\mathrm{Unif}(\{\pm\}), P) + I_{\chi^2}(\mathrm{Unif}(\{\pm\}), Q). \tag{29}$$

## 3. Multi-terminal SDPI

Let $q \ge 2$, $r \ge 2$ be integers, $\pi \in \mathcal{P}([q])$, and $B : [q] \to [q]^{r-1}$ be a channel satisfying (10). For any $f$-divergence, we define the multi-terminal contraction coefficient as

$$\eta_f^{(m)}(\pi, B) := \sup_P \frac{I_f(\pi, P^{\times(r-1)} \circ B)}{(r-1) I_f(\pi, P)} \tag{30}$$

where $P$ goes over all channels with input alphabet $[q]$ for which $0 < I_f(\pi, P) < \infty$.

In other words, $\eta_f^{(m)}(\pi, B)$ is the smallest constant such that for any diagram as in Figure 1 where $P_X = \pi$, $P_{Y^{r-1}|X} = B$, $P_{U_i|Y_i} = P$, we have

$$I_f(X; U^{r-1}) \le (r-1)\eta_f^{(m)}(\pi, B) I_f(Y_1; U_1). \tag{31}$$

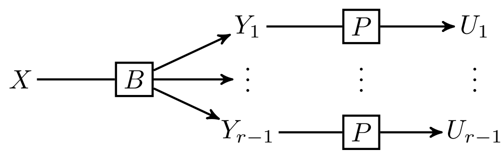


Figure 1: Setting for homogeneous multi-terminal SDPI

In the above definition, we assume $P_{U_i|Y_i}$ are the same for all $i \in [r-1]$. Therefore (30) defines a homogeneous multi-terminal contraction coefficient. It is also possible to define a heterogeneous version, where the input channels can be different. We define

$$\eta_f^{(m,ht)}(\pi, B) := \sup_{P_1,\dots,P_{r-1}} \frac{I_f(\pi, (P_1 \times \cdots \times P_{r-1}) \circ B)}{\sum_{i\in[r-1]} I_f(\pi, P_i)} \tag{32}$$

where $P_1, \dots, P_{r-1}$ goes over all channels with input alphabet $[q]$ for which $0 < \sum_{i\in[r-1]} I_f(\pi, P_i) < \infty$ (here $ht$ stands for "heterogeneous"). In other words, $\eta_f^{(m,ht)}(\pi, B)$ is the smallest constant such that for any diagram as in Figure 2 where $P_X = \pi$, we have

$$I_f(X; U^{r-1}) \le \eta_f^{(m,ht)}(\pi, B) \sum_{i\in[r-1]} I_f(Y_i; U_i). \tag{33}$$

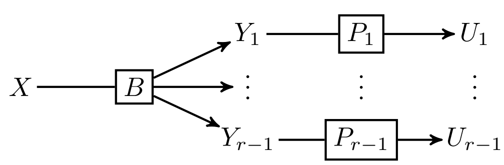


Figure 2: Setting for heterogeneous multi-terminal SDPI

It is clear from definition that

$$\eta_f^{(m)}(\pi, B) \le \eta_f^{(m,ht)}(\pi, B). \tag{34}$$

Unlike the usual contraction coefficients, it is not true in general that $\eta_f^{(m)}(\pi, B) \le 1$. Nevertheless, this holds when the $f$-information is subadditive under $\star$-convolution. For the mutual information, we have

$$I(X; U^{r-1}) \le I(Y^{r-1}; U^{r-1}) \le \sum_{i\in[r-1]} I(Y^{r-1}; U_i) = \sum_{i\in[r-1]} I(Y_i, U_i), \tag{35}$$

where the first step is by DPI and the second step is by subadditivity. Therefore

$$\eta_{\mathrm{KL}}^{(m)}(\pi, B) \le \eta_{\mathrm{KL}}^{(m,ht)}(\pi, B) \le 1. \tag{36}$$

The same holds for the SKL mutual information, so

$$\eta_{\mathrm{SKL}}^{(m)}(\pi, B) \le \eta_{\mathrm{SKL}}^{(m,ht)}(\pi, B) \le 1. \tag{37}$$

When $q = 2$, $\pi = \mathrm{Unif}(\{\pm\})$, and $B$ together with the sign flip is a BMS channel, for any BMS channels $P_1, \ldots, P_{r-1}$, the combined channel $(P_1 \times \cdots \times P_{r-1}) \circ B$ is also a BMS channel. In this case we can define versions of multi-terminal contraction coefficients restricted to BMS channels. We define

$$\eta_f^{(m,s)}(B) := \sup_P \frac{C_f(P^{\times(r-1)} \circ B)}{(r-1)C_f(P)} \tag{38}$$

where $P$ goes over BMS channels with $0 < C_f(P) < \infty$, and

$$\eta_f^{(m,ht,s)}(B) := \sup_{P_1,\ldots,P_{r-1}} \frac{C_f((P_1 \times \cdots \times P_{r-1}) \circ B)}{\sum_{i \in [r-1]} C_f(P_i)} \tag{39}$$

where $P_1, \ldots, P_{r-1}$ goes over BMS channels with $0 < \sum_{i\in[r-1]} C_f(P_i) < \infty$. Because $\chi^2$-capacity is subadditive over BMS channels, by a similar computation as (35) we have

$$\eta_{\chi^2}^{(m,s)}(B) \le \eta_{\chi^2}^{(m,ht,s)}(B) \le 1. \tag{40}$$

With these definitions, it is very easy to prove Theorem 1.

**Proof** [Proof of Theorem 1] Let $M_k$ denote the channel $\sigma_\rho \mapsto (T_k, \sigma_{L_k})$. Then $(M_k)_{k\ge 0}$ satisfies the BP recursion $M_{k+1} = \mathrm{BP}(M_k)$, where BP is defined in (15).

**Part (i), mutual information:** For any channel $P$ with input alphabet $[q]$ we have

$$I(\pi, \mathrm{BP}(P)) \le dI(\pi, P^{\times(r-1)} \circ B) \le (r-1)d\eta_{\mathrm{KL}}^{(m)}(\pi, B)I(\pi, P), \tag{41}$$

where the first step is by subadditivity of mutual information, and the second step is by definition of multi-terminal contraction coefficients. If (21) holds, then $I(\pi, M_{k+1}) \le cI(\pi, M_k)$ for $c = (r-1)d\eta_{\mathrm{KL}}^{(m)}(\pi, B) < 1$. Because $I(\pi, M_0) < \infty$, we have $\lim_{k\to\infty} I(\pi, M_k) = 0$ and non-reconstruction holds.

**Part (i), SKL information:** If (22) holds, then

$$I_{\mathrm{SKL}}(\pi, M_1) = dI_{\mathrm{SKL}}(\pi, P^{\times(r-1)} \circ B) \le dI_{\mathrm{SKL}}(\pi, B) < \infty, \tag{42}$$

where the first step is by additivity of SKL information, and the second step is by DPI. The rest of the proof is similar to the previous case.

**Part (ii), KL and SKL capacity:** The channels $M_k$ are all BMS channels, so we can use the BMS version of multi-terminal contraction coefficients. The rest of the proof is the same as Part (i).

**Part (ii), $\chi^2$-capacity:** Use subadditivity of $\chi^2$-capacity for BMS channels and the proof is similar to previous cases. ■

See Section A for a variation of Theorem 1 to the fixed-hypertree model $\mathrm{BOHT}(T, q, r, \pi, B)$.

## 4. Non-reconstruction for the special BOHT model

In this section we focus on the special BOHT model $\mathrm{BOHT}(2, r, \lambda, D)$, which is the BOHT model with $q = 2$, $\pi = \mathrm{Unif}(\{\pm\})$, and $B = B_{r,\lambda}$ as defined in (12). We prove non-reconstruction results for this model using Theorem 1(ii) by computing the relevant multi-terminal contraction coefficients.

To compute the contraction coefficients, we need to descibe $P^{\times(r-1)} \circ B_{r,\lambda}$ (where $P$ is a BMS channel) in a more explicit way. By BSC mixture representation, we only need to describe $(\mathrm{BSC}_{\Delta_1} \times \cdots \times \mathrm{BSC}_{\Delta_{r-1}}) \circ B_{r,\lambda}$. Let $\theta_i := 1 - 2\Delta_i$ for $i \in [r-1]$. For $x \in \{\pm\}^{r-1}$, we have

$$\begin{aligned}
&((\mathrm{BSC}_{\Delta_1} \times \cdots \times \mathrm{BSC}_{\Delta_{r-1}}) \circ B_{r,\lambda})(x_1, \ldots, x_{r-1}|+) \\
&= \sum_{y\in\{\pm\}^{r-1}} B_{r,\lambda}(y_1, \ldots, y_{r-1}|+) \prod_{i\in[r-1]} \mathrm{BSC}_{\Delta_i}(x_i|y_i) \\
&= \lambda \prod_{i\in[r-1]} \mathrm{BSC}_{\Delta_i}(x_i|+) + \frac{1}{2^{r-1}}(1-\lambda) \prod_{i\in[r-1]} \sum_{y_i\in\{\pm\}} \mathrm{BSC}_{\Delta_i}(x_i|y_i) \\
&= \lambda \prod_{i\in[r-1]} \left(\frac{1}{2} + \left(\frac{1}{2} - \Delta_i\right) x_i\right) + \frac{1}{2^{r-1}}(1-\lambda) \\
&= \lambda \prod_{i\in[r-1]} \left(\frac{1}{2} + \frac{1}{2}\theta_i x_i\right) + \frac{1}{2^{r-1}}(1-\lambda).
\end{aligned} \tag{43}$$

So $(\mathrm{BSC}_{\Delta_1} \times \cdots \times \mathrm{BSC}_{\Delta_{r-1}}) \circ B_{r,\lambda}$ is a mixture of $2^{r-2}$ BSCs, indexed by the set

$$\left\{x : x \in \{\pm\}^{r-1}, x_1 = +\right\}, \tag{44}$$

where the BSC corresponding to $x$ has weight (probability)

$$\begin{aligned}
&\left(\lambda \prod_{i\in[r-1]} \left(\frac{1}{2} + \frac{1}{2}\theta_i x_i\right) + \frac{1}{2^{r-1}}(1-\lambda)\right) + \left(\lambda \prod_{i\in[r-1]} \left(\frac{1}{2} - \frac{1}{2}\theta_i x_i\right) + \frac{1}{2^{r-1}}(1-\lambda)\right) \\
&= \lambda \left(\prod_{i\in[r-1]} \left(\frac{1}{2} + \frac{1}{2}\theta_i x_i\right) + \prod_{i\in[r-1]} \left(\frac{1}{2} - \frac{1}{2}\theta_i x_i\right)\right) + \frac{1}{2^{r-2}}(1-\lambda)
\end{aligned} \tag{45}$$

and $\theta$ parameter equal to the absolute value of

$$\begin{aligned}
&\frac{\left(\lambda \prod_{i\in[r-1]} \left(\frac{1}{2} + \frac{1}{2}\theta_i x_i\right) + \frac{1}{2^{r-1}}(1-\lambda)\right) - \left(\lambda \prod_{i\in[r-1]} \left(\frac{1}{2} - \frac{1}{2}\theta_i x_i\right) + \frac{1}{2^{r-1}}(1-\lambda)\right)}{\left(\lambda \prod_{i\in[r-1]} \left(\frac{1}{2} + \frac{1}{2}\theta_i x_i\right) + \frac{1}{2^{r-1}}(1-\lambda)\right) + \left(\lambda \prod_{i\in[r-1]} \left(\frac{1}{2} - \frac{1}{2}\theta_i x_i\right) + \frac{1}{2^{r-1}}(1-\lambda)\right)} \\
&= \frac{\lambda \left(\prod_{i\in[r-1]} \left(\frac{1}{2} + \frac{1}{2}\theta_i x_i\right) - \prod_{i\in[r-1]} \left(\frac{1}{2} - \frac{1}{2}\theta_i x_i\right)\right)}{\lambda \left(\prod_{i\in[r-1]} \left(\frac{1}{2} + \frac{1}{2}\theta_i x_i\right) + \prod_{i\in[r-1]} \left(\frac{1}{2} - \frac{1}{2}\theta_i x_i\right)\right) + \frac{1}{2^{r-2}}(1-\lambda)} \\
&= \frac{\lambda \left(\prod_{i\in[r-1]} (1 + \theta_i x_i) - \prod_{i\in[r-1]} (1 - \theta_i x_i)\right)}{\lambda \left(\prod_{i\in[r-1]} (1 + \theta_i x_i) + \prod_{i\in[r-1]} (1 - \theta_i x_i)\right) + 2(1-\lambda)}.
\end{aligned} \tag{46}$$

Although we need to take absolute value, information measures (except for $P_e$) in Definition 8 are all even functions in $\theta$. So we do not need to worry about the sign.

Using this explicit description of $P^{\times(r-1)} \circ B_{r,\lambda}$, we are able to compute several multi-terminal contraction coefficients of $B_{r,\lambda}$.

For Theorem 2(i), we compute the SKL contraction coefficients.

**Proposition 9 (SKL contraction coefficient)** *Fix $r = 3$ or $4$ and $\lambda \in [0, 1]$. Then*

$$\eta^{(m,ht,s)}_{\mathrm{SKL}}(B_{r,\lambda}) = \lambda^2. \tag{47}$$

*Furthermore, if $0 < \lambda < 1$ and $P_1, \ldots, P_{r-1}$ are BMS channels with at least one $P_i$ non-trivial, then*

$$C_{\mathrm{SKL}}((P_1 \times \cdots \times P_{r-1}) \circ B_{r,\lambda}) < \lambda^2 \sum_{i\in[r-1]} C_{\mathrm{SKL}}(P_i). \tag{48}$$

Proof of Prop. 9 is deferred to Section B.

**Proof** [Proof of Theorem 2(i)] If $(r-1)d \le 1$, then the BOHT model extincts almost surely (e.g., (Lyons and Peres, 2017, Prop. 5.4)). This resolves the case $\lambda = 1$. In the following, assume that $0 \le \lambda < 1$.

For $0 \le \lambda < 1$, we have $C_{\mathrm{SKL}}(B_{r,\lambda}) < \infty$. Therefore Theorem 1(ii) together with Prop. 9 implies that non-reconstruction holds for $(r-1)d\lambda^2 < 1$.

For the critical case $(r-1)d\lambda^2 = 1$ we need some extra argument. See Section B.3. ■

For Theorem 2(ii), we compute the $\chi^2$-contraction coefficients.

**Proposition 10 ($\chi^2$-contraction coefficient)** *Fix $r \in \mathbb{Z}_{\ge 3}$. Then*

$$\eta^{(m,s)}_{\chi^2}(B_{r,\lambda}) = \sup_{0<\epsilon\le 1} f_{r,\lambda}(\epsilon), \tag{49}$$

*where $f_{r,\lambda}$ is defined in (27).*

Proof of Prop. 10 is deferred to Section C.

**Proof** [Proof of Theorem 2(ii)] Follows from Theorem 1(ii) and Prop. 10. ■

Interestingly, RHS of (49) can be computed exactly when $\lambda$ is not too small.

**Lemma 11** *Fix $r \in \mathbb{Z}_{\ge 3}$ and $\lambda \in \left[\frac{1}{5}, 1\right]$. For all $0 < \epsilon \le 1$, we have $f_{r,\lambda}(\epsilon) < \lambda^2$.*

Proof of Lemma 11 is deferred to Section C.

**Proof** [Proof of Theorem 2(iii)] Prop. 10 together with Lemma 11 implies that for $\lambda \in \left[\frac{1}{5}, 1\right]$ we have

$$\eta^{(m,s)}_{\chi^2}(B_{r,\lambda}) = \lambda^2. \tag{50}$$

(For the lower bound, take $\epsilon \to 0^+$ in (49).) Then Theorem 1(ii) implies that non-reconstruction holds for $(r-1)d\lambda^2 < 1$.

For the critical case $(r-1)d\lambda^2 = 1$ we need some extra argument. See Section C.4. ■

## 5. Discussions

For the special BOHT model, we have left the $r = 5, 6$ case open. Our preliminary computations suggest that for $r = 5, 6$, there exists an absolute constant $d_0 \in \mathbb{R}_{\geq 0}$ such that the BOHT model has non-reconstruction when $d \geq d_0$ and $(r-1)d\lambda^2 \leq 1$. We believe that a generalization of Sly's method Sly (2011); Mossel et al. (2022) can be used to prove this. In Sly's method, we compute the first few orders of the BP recursion formula. Combined with Gaussian approximation this would imply contraction of $\chi^2$-capacity. One technical challenge is that in the BOHT case we need a two-step application of Sly's method, in contrast with previous works.

## Acknowledgments

We thank Kunal Marwaha for resolving a conjecture in a preliminary version of the paper. We thank the anonymous reviewers for helpful comments.

Research was sponsored by the United States Air Force Research Laboratory and the United States Air Force Artificial Intelligence Accelerator and was accomplished under Cooperative Agreement Number FA8750-19-2-1000. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the United States Air Force or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

## References

Emmanuel Abbe. Community detection and stochastic block models: recent developments. *The Journal of Machine Learning Research*, 18(1):6446–6531, 2017.

Emmanuel Abbe and Enric Boix-Adserà. Subadditivity beyond trees and the chi-squared mutual information. In *2019 IEEE International Symposium on Information Theory (ISIT)*, pages 697–701. IEEE, 2019.

Kwangjun Ahn, Kangwook Lee, and Changho Suh. Hypergraph spectral clustering in the weighted stochastic block model. *IEEE Journal of Selected Topics in Signal Processing*, 12(5):959–974, 2018.

Maria Chiara Angelini, Francesco Caltagirone, Florent Krzakala, and Lenka Zdeborová. Spectral detection on sparse hypergraphs. In *2015 53rd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 66–73. IEEE, 2015.

Nayantara Bhatnagar, Allan Sly, and Prasad Tetali. Reconstruction threshold for the hardcore model. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques: 13th International Workshop, APPROX 2010, and 14th International Workshop, RANDOM 2010, Barcelona, Spain, September 1-3, 2010. Proceedings*, pages 434–447. Springer, 2010.

Pavel M. Bleher, Jean Ruiz, and Valentin A. Zagrebnov. On the purity of the limiting gibbs state for the ising model on the bethe lattice. *Journal of Statistical Physics*, 79:473–482, 1995.

Charles Bordenave, Marc Lelarge, and Laurent Massoulié. Non-backtracking spectrum of random graphs: community detection and non-regular ramanujan graphs. In *2015 IEEE 56th Annual Symposium on Foundations of Computer Science*, pages 1347–1357. IEEE, 2015.

Christian Borgs, Jennifer Chayes, Elchanan Mossel, and Sébastien Roch. The kesten-stigum reconstruction bound is tight for roughly symmetric binary channels. In *2006 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS'06)*, pages 518–530. IEEE, 2006.

Francesco Caltagirone, Marc Lelarge, and Léo Miolane. Recovering asymmetric communities in the stochastic block model. *IEEE Transactions on Network Science and Engineering*, 5(3):237–246, 2017.

I Chien, Chung-Yi Lin, and I-Hsiang Wang. Community detection in hypergraphs: Optimal statistical limit and efficient algorithms. In *International Conference on Artificial Intelligence and Statistics*, pages 871–879. PMLR, 2018.

I Eli Chien, Chung-Yi Lin, and I-Hsiang Wang. On the minimax misclassification ratio of hypergraph community detection. *IEEE Transactions on Information Theory*, 65(12):8095–8118, 2019.

Byron Chin and Allan Sly. Optimal recovery of block models with $q$ communities. *arXiv preprint arXiv:2010.10672*, 2020.

Byron Chin and Allan Sly. Optimal reconstruction of general sparse stochastic block models. *arXiv preprint arXiv:2111.00697*, 2021.

Joel Cohen, Johannes HB Kempermann, and Gheorghe Zbaganu. *Comparisons of stochastic matrices with applications in information theory, statistics, economics and population*. Springer Science & Business Media, 1998.

Sam Cole and Yizhe Zhu. Exact recovery in the hypergraph stochastic block model: A spectral algorithm. *Linear Algebra and its Applications*, 593:45–73, 2020.

Ioana Dumitriu and Haixiao Wang. Exact recovery for the non-uniform hypergraph stochastic block model. *arXiv preprint arXiv:2304.13139*, 2023.

Ioana Dumitriu, Haixiao Wang, and Yizhe Zhu. Partial recovery and weak consistency in the non-uniform hypergraph stochastic block model. *arXiv preprint arXiv:2112.11671*, 2021.

William Evans, Claire Kenyon, Yuval Peres, and Leonard J. Schulman. Broadcasting on trees and the ising model. *Annals of Applied Probability*, pages 410–433, 2000.

Marco Formentin and Christof Külske. On the purity of the free boundary condition potts measure on random trees. *Stochastic Processes and their Applications*, 119(9):2992–3005, 2009.

Debarghya Ghoshdastidar and Ambedkar Dukkipati. Consistency of spectral partitioning of uniform hypergraphs under planted partition model. *Advances in Neural Information Processing Systems*, 27, 2014.

Debarghya Ghoshdastidar and Ambedkar Dukkipati. A provable generalized tensor spectral method for uniform hypergraph partitioning. In *International Conference on Machine Learning*, pages 400–409. PMLR, 2015a.

Debarghya Ghoshdastidar and Ambedkar Dukkipati. Spectral clustering using multilinear SVD: Analysis, approximations and applications. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 29, 2015b.

Debarghya Ghoshdastidar and Ambedkar Dukkipati. Consistency of spectral hypergraph partitioning under planted partition model. *The Annals of Statistics*, 45(1):289–315, 2017.

Yuzhou Gu. *Channel Comparison Methods and Statistical Problems on Graphs*. PhD thesis, Massachusetts Institute of Technology, 2023.

Yuzhou Gu and Yury Polyanskiy. Non-linear log-Sobolev inequalities for the Potts semigroup and applications to reconstruction problems. *arXiv preprint arXiv:2005.05444*, 2020.

Yuzhou Gu, Hajir Roozbehani, and Yury Polyanskiy. Broadcasting on trees near criticality. In *2020 IEEE International Symposium on Information Theory (ISIT)*, pages 1504–1509. IEEE, 2020.

Chiheon Kim, Afonso S. Bandeira, and Michel X. Goemans. Stochastic block model for hypergraphs: Statistical limits and a semidefinite programming approach. *arXiv preprint arXiv:1807.02884*, 2018.

J. Körner and K. Marton. Comparison of two noisy channels. *Topics in Information Theory, I. Csiszár and P. Elias, Eds., Amsterdam, The Netherlands*, pages 411–423, 1977.

Christof Külske and Marco Formentin. A symmetric entropy bound on the non-reconstruction regime of Markov chains on Galton-Watson trees. *Electronic Communications in Probability*, 14:587–596, 2009.

Chung-Yi Lin, I Eli Chien, and I-Hsiang Wang. On the fundamental statistical limit of community detection in random hypergraphs. In *2017 IEEE International Symposium on Information Theory (ISIT)*, pages 2178–2182. IEEE, 2017.

Wenjian Liu and Ning Ning. Large degree asymptotics and the reconstruction threshold of the asymmetric binary channels. *Journal of Statistical Physics*, 174:1161–1188, 2019.

Russell Lyons. Random walks and percolation on trees. *The Annals of Probability*, 18(3):931–958, 1990.

Russell Lyons and Yuval Peres. *Probability on trees and networks*, volume 42. Cambridge University Press, 2017.

Anuran Makur and Yury Polyanskiy. Comparison of channels: Criteria for domination by a symmetric channel. *IEEE Transactions on Information Theory*, 64(8):5704–5725, 2018.

Kunal Marwaha. A useful inequality of inverse hyperbolic tangent. *arXiv preprint arXiv:2305.18348*, 2023.

Laurent Massoulié. Community detection thresholds and the weak ramanujan property. In *Proceedings of the forty-sixth annual ACM Symposium on Theory of Computing*, pages 694–703, 2014.

Marc Mézard and Andrea Montanari. Reconstruction on trees and spin glass transition. *Journal of statistical physics*, 124:1317–1350, 2006.

Elchanan Mossel. Reconstruction on trees: beating the second eigenvalue. *The Annals of Applied Probability*, 11(1):285–300, 2001.

Elchanan Mossel and Yuval Peres. Information flow on trees. *The Annals of Applied Probability*, 13(3):817–844, 2003.

Elchanan Mossel, Joe Neeman, and Allan Sly. Reconstruction and estimation in the planted partition model. *Probability Theory and Related Fields*, 162:431–461, 2015.

Elchanan Mossel, Joe Neeman, and Allan Sly. A proof of the block model threshold conjecture. *Combinatorica*, 38(3):665–708, 2018.

Elchanan Mossel, Allan Sly, and Youngtak Sohn. Exact phase transitions for stochastic block models and reconstruction on trees. *arXiv preprint arXiv:2212.03362*, 2022.

Soumik Pal and Yizhe Zhu. Community detection in the sparse hypergraph stochastic block model. *Random Structures & Algorithms*, 59(3):407–463, 2021.

Yury Polyanskiy and Yihong Wu. Strong data-processing inequalities for channels and bayesian networks. In *Convexity and Concentration*, pages 211–249. Springer, 2017.

Yury Polyanskiy and Yihong Wu. *Information Theory: From Coding to Learning*. Cambridge University Press, 2023+.

Maxim Raginsky. Strong data processing inequalities and $\phi$-sobolev inequalities for discrete channels. *IEEE Transactions on Information Theory*, 62(6):3355–3389, 2016.

Tom Richardson and Rüdiger Urbanke. *Modern coding theory*. Cambridge university press, 2008.

Hajir Roozbehani and Yury Polyanskiy. Low density majority codes and the problem of graceful degradation. *arXiv preprint arXiv:1911.12263*, 2019.

Allan Sly. Reconstruction of random colourings. *Communications in Mathematical Physics*, 288 (3):943–961, 2009.

Allan Sly. Reconstruction for the Potts model. *The Annals of Probability*, 39(4):1365 – 1406, 2011. doi: 10.1214/10-AOP584. URL https://doi.org/10.1214/10-AOP584.

Ludovic Stephan and Laurent Massoulié. Robustness of spectral methods for community detection. In *Conference on Learning Theory*, pages 2831–2860. PMLR, 2019.

Ludovic Stephan and Laurent Massoulié. Non-backtracking spectra of weighted inhomogeneous random graphs. *Mathematical Statistics and Learning*, 5(3):201–271, 2022.

Ludovic Stephan and Yizhe Zhu. Sparse random hypergraphs: Non-backtracking spectra and community detection. In *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 567–575. IEEE, 2022.

David Sutter and Joseph M. Renes. Universal polar codes for more capable and less-noisy channels and sources. In *2014 IEEE International Symposium on Information Theory*, pages 1461–1465. IEEE, 2014.

Erchuan Zhang, David Suter, Giang Truong, and Syed Zulqarnain Gilani. Sparse hypergraph community detection thresholds in stochastic block model. In *Thirty-Sixth Conference on Neural Information Processing Systems (NeurIPS)*, 2022.

Qiaosheng Zhang and Vincent Y. F. Tan. Exact recovery in the general hypergraph stochastic block model. *IEEE Transactions on Information Theory*, 69(1):453–471, 2022.

## Appendix A. Non-reconstruction for the fixed-hypertree BOHT model

In this section we give a fixed-hypertree version of Theorem 1. Consider the model $\mathrm{BOHT}(T, q, r, \pi, B)$ where $T$ is a fixed rooted $r$-uniform linear hypertree. Recall the definition of the branching number.

**Definition 12 (Branching number Lyons (1990))** *Let $T$ be a possibly infinite tree rooted at $\rho$. Define a flow to be a function $f : V(T) \to \mathbb{R}_{\geq 0}$ such that for every vertex $u$, we have*

$$f_u = \sum_{v \in c(u)} f_v. \tag{51}$$

*Define $\mathrm{br}(T)$ to be the $\sup$ of all numbers $\lambda$ such that there exists a flow $f$ with $f_\rho > 0$, and $f_u \leq \lambda^{-d(u)}$ for all vertices $u$, where $d(u)$ denotes the depth of $u$ (i.e., distance to $\rho$).*

For an $r$-uniform linear hypertree, we can split every downward hyperedge into $(r-1)$ downward edges, and apply the above definition. In this way we extend the definition of branching number to linear hypertrees.

**Theorem 13 (Non-reconstruction for BOHT)** *Consider the model $\mathrm{BOHT}(T, q, r, \pi, B)$.*

*(i) If*

$$\mathrm{br}(T)\eta^{(m,ht)}_{\mathrm{KL}}(\pi, B) < 1, \tag{52}$$

$$\text{or} \quad \mathrm{br}(T)\eta^{(m,ht)}_{\mathrm{SKL}}(\pi, B) < 1, I_{\mathrm{SKL}}(\pi, B) < \infty, \text{and } T \text{ has bounded maximum degree}, \tag{53}$$

*then reconstruction is impossible.*

*(ii) Suppose the BOHT model is binary symmetric. If*

$$\mathrm{br}(T)\eta^{(m,ht,s)}_{\mathrm{KL}}(B) < 1, \tag{54}$$

$$\text{or} \quad \mathrm{br}(T)\eta^{(m,ht,s)}_{\chi^2}(B) < 1, \tag{55}$$

$$\text{or} \quad \mathrm{br}(T)\eta^{(m,ht,s)}_{\mathrm{SKL}}(B) < 1, C_{\mathrm{SKL}}(B) < \infty, \text{and } T \text{ has bounded maximum degree}, \tag{56}$$

*then reconstruction is impossible.*

**Proof** The proof is a generalization of the argument from Gu and Polyanskiy (2020).

**Part (i), mutual information:** For any vertex $u$, let $L_{u,k}$ denote the set of descendants of $u$ at distance $k$ to $\rho$. Define

$$a_u := H(\pi)^{-1} \left(\eta_{\mathrm{KL}}^{(m,ht)}(\pi, B)\right)^{d(u)} \lim_{k\to\infty} I(\sigma_u; \sigma_{L_{u,k}}). \tag{57}$$

By DPI, $I(\sigma_u; \sigma_{L_{u,k}})$ is non-increasing for $k \ge d(u)$, so the limit exists.

Let $\gamma(u)$ denote the set of downward hyperedges of $u$ and $c(u)$ denotes the set of children of $u$. For any $e = \{u, v_1, \dots, v_{r-1}\} \in \gamma(u)$, we have a diagram as in Figure 3, where $P_{\sigma_u} = \pi$, $P_i = P_{\sigma_{L_{v_i,k}}|\sigma_{v_i}}$. Define $L_{e\backslash u,k} := \bigcup_{i\in[r-1]} L_{v_i,k}$. By definition of multi-terminal contraction

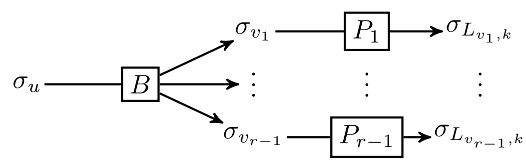


Figure 3: Apply multi-terminal SDPI to BOHT with a fixed hypertree

coefficients, we have

$$I(\sigma_u; \sigma_{L_{e\backslash u,k}}) \le \eta_{\mathrm{KL}}^{(m,ht)}(\pi, B) \sum_{i\in[r-1]} I(\sigma_{v_i}; \sigma_{L_{v_i,k}}). \tag{58}$$

Summing over all $e \in \gamma(u)$ and using subadditivity, we have

$$I(\sigma_u; \sigma_{L_{u,k}}) \le \sum_{e\in\gamma(u)} I(\sigma_u; \sigma_{L_{e\backslash u,k}}) \le \eta_{\mathrm{KL}}^{(m,ht)}(\pi, B) \sum_{v\in c(u)} I(\sigma_v; \sigma_{L_{v,k}}). \tag{59}$$

Comparing (57) and (59) we see that

$$a_u \le \sum_{v\in c(u)} a_v. \tag{60}$$

By definition, we have

$$a_u \le \left(\eta_{\mathrm{KL}}^{(m,ht)}(\pi, B)\right)^{d(u)}. \tag{61}$$

However, $a$ is not a flow yet. We define a flow $b$ from $a$. For a vertex $u$, let $u_0 = \rho, \dots, u_\ell = u$ be the shortest path from $\rho$ to $u$. Define

$$b_u := a_u \prod_{0\le j\le \ell-1} \frac{a_{u_j}}{\sum_{v\in c(u_j)} a_v}. \tag{62}$$

(If the denominator is zero, let $b_u = 0$.) Then we have

$$b_u = \sum_{v\in c(u)} b_v, \quad \text{and} \quad b_u \le a_u \le \left(\eta_{\mathrm{KL}}^{(m,ht)}(\pi, B)\right)^{d(u)}. \tag{63}$$

By definition of branching number, we have $b_\rho = 0$. Therefore

$$\lim_{k\to\infty} I(\sigma_\rho; \sigma_{L_k}) = 0 \tag{64}$$

and non-reconstruction holds.

**Part (i), SKL information:** Suppose every vertex $u$ has at most $\gamma_{\max}$ downward hyperedges. Then

$$\lim_{k\to\infty} I_{\mathrm{SKL}}(\sigma_u; \sigma_{L_{u,k}}) \le I_{\mathrm{SKL}}(\sigma_u; \sigma_{L_{u,d(u)+1}}) = |\gamma(u)| I_{\mathrm{SKL}}(\pi, B) \le \gamma_{\max} I_{\mathrm{SKL}}(\pi, B) < \infty. \tag{65}$$

We define

$$a_u := (\gamma_{\max} I_{\mathrm{SKL}}(\pi, B))^{-1} \left(\eta_{\mathrm{SKL}}^{(m,ht)}(\pi, B)\right)^{d(u)} \lim_{k\to\infty} I_{\mathrm{SKL}}(\sigma_u; \sigma_{L_{u,k}}). \tag{66}$$

By (65), we have

$$a_u \le \left(\eta_{\mathrm{SKL}}^{(m,ht)}(\pi, B)\right)^{d(u)}. \tag{67}$$

The rest of the proof is similar to the mutual information case.

**Part (ii), KL and SKL capacity:** In this case, channels appeared in the above proof (e.g., $\sigma_u \mapsto \sigma_{L_{u,k}}$) are all BMS channels. Using the same proof with BMS version of multi-terminal contraction coefficients leads to the desired result.

**Part (ii), $\chi^2$-capacity:** Note that $C_{\chi^2}(P) \le 1$ for all BMS channels $P$. So we can define

$$a_u := \left(\eta_{\chi^2}^{(m,ht,s)}(B)\right)^{d(u)} \lim_{k\to\infty} C_{\chi^2}(\sigma_{L_{u,k}}) \tag{68}$$

and it satisfies

$$a_u \le \left(\eta_{\chi^2}^{(m,ht,s)}(\pi, B)\right)^{d(u)}. \tag{69}$$

The rest of the proof is similar to the previous cases. ■

## Appendix B. Computation of SKL contraction coefficients

In this section we prove Prop. 9, which says that for any BMS channels $P_1, \ldots, P_{r-1}$ we have

$$C_{\mathrm{SKL}}((P_1 \times \cdots \times P_{r-1}) \circ B_{r,\lambda}) \le \lambda^2 \sum_{i\in[r-1]} C_{\mathrm{SKL}}(P_i). \tag{70}$$

By BSC mixture representation of BMS channels, (70) is equivalent to

$$C_{\mathrm{SKL}}\left((\mathrm{BSC}_{\Delta_1} \times \cdots \times \mathrm{BSC}_{\Delta_{r-1}}) \circ B_{r,\lambda}\right) \le \lambda^2 \sum_{i\in[r-1]} C_{\mathrm{SKL}}(\mathrm{BSC}_{\Delta_i}) \tag{71}$$

for all $\Delta_1, \ldots, \Delta_{r-1} \in \left[0, \frac{1}{2}\right]$.

### B.1. Case $r = 3$

We first prove the $r = 3$ case.

**Lemma 14** *For any* $\Delta_1, \Delta_2 \in \left[0, \frac{1}{2}\right]$, *we have*

$$C_{\mathrm{SKL}}((\mathrm{BSC}_{\Delta_1} \times \mathrm{BSC}_{\Delta_2}) \circ B_{r,\lambda}) \le \lambda^2 (C_{\mathrm{SKL}}(\mathrm{BSC}_{\Delta_1}) + C_{\mathrm{SKL}}(\mathrm{BSC}_{\Delta_2})). \tag{72}$$

*Furthermore, the inequality is strict when* $0 < \lambda < 1$ *and* $\min\{\Delta_1, \Delta_2\} < \frac{1}{2}$.

**Proof** We expand LHS of (72) using the BP recursion formula established in Section 4. Let $\theta_i = 1 - 2\Delta_i$ for $i = 1, 2$. Then

$$\begin{aligned}
&C_{\mathrm{SKL}}((\mathrm{BSC}_{\Delta_1} \times \mathrm{BSC}_{\Delta_2}) \circ B_{r,\lambda}) \\
&= \sum_{x_1 = +, x_2 \in \{\pm\}} \frac{1}{2} \lambda(\theta_1 x_1 + \theta_2 x_2) \operatorname{arctanh} \frac{\lambda(\theta_1 x_1 + \theta_2 x_2)}{\lambda(1 + \theta_1 x_1 \theta_2 x_2) + (1 - \lambda)} \\
&= \lambda \left( \frac{1}{2}(\theta_1 + \theta_2) \operatorname{arctanh} \frac{\lambda(\theta_1 + \theta_2)}{1 + \lambda\theta_1\theta_2} + \frac{1}{2}(\theta_1 - \theta_2) \operatorname{arctanh} \frac{\lambda(\theta_1 - \theta_2)}{1 - \lambda\theta_1\theta_2} \right) \\
&= \lambda \left( \frac{1}{2}(\theta_1 + \theta_2) F_\lambda(\theta_1, \theta_2) + \frac{1}{2}(\theta_1 - \theta_2) F_\lambda(\theta_1, -\theta_2) \right)
\end{aligned} \tag{73}$$

where

$$F_\lambda(\theta_1, \theta_2) := \operatorname{arctanh} \frac{\lambda(\theta_1 + \theta_2)}{1 + \lambda\theta_1\theta_2}. \tag{74}$$

Note that by definition, $F_\lambda(\theta_1, \theta_2) = -F_\lambda(-\theta_1, -\theta_2)$ and $F_\lambda(\theta_1, \theta_2) = F_\lambda(\theta_2, \theta_1)$.

We have

$$\begin{aligned}
&\frac{1}{2}(\theta_1 + \theta_2) F_\lambda(\theta_1, \theta_2) + \frac{1}{2}(\theta_1 - \theta_2) F_\lambda(\theta_1, -\theta_2) \\
&= \frac{1}{2}\theta_1 (F_\lambda(\theta_1, \theta_2) + F_\lambda(\theta_1, -\theta_2)) + \frac{1}{2}\theta_2 (F_\lambda(\theta_1, \theta_2) + F_\lambda(-\theta_1, \theta_2)) \\
&\le \theta_1 F_\lambda(\theta_1, 0) + \theta_2 F_\lambda(0, \theta_2) \\
&= \theta_1 \operatorname{arctanh}(\lambda\theta_1) + \theta_2 \operatorname{arctanh}(\lambda\theta_2) \\
&\le \lambda(\theta_1 \operatorname{arctanh} \theta_1 + \theta_2 \operatorname{arctanh} \theta_2) \\
&= \lambda(C_{\mathrm{SKL}}(\mathrm{BSC}_{\Delta_1}) + C_{\mathrm{SKL}}(\mathrm{BSC}_{\Delta_2})),
\end{aligned} \tag{75}$$

where the second step follows from Lemma 15, and the fourth step follows convexity of arctanh in $[0, 1]$. Combining (73)(75) we finish the proof. ■

**Lemma 15** *For* $\lambda, \theta_1, \theta_2 \in [0, 1]$, *we have*

$$\frac{1}{2}\left(F_\lambda(\theta_1, \theta_2) + F_\lambda(\theta_1, -\theta_2)\right) \le F_\lambda(\theta_1, 0). \tag{76}$$

*Furthermore, the inequality is strict when* $0 < \lambda < 1$ *and* $\theta_1, \theta_2 > 0$.

**Proof** We use the formula

$$\operatorname{arctanh} x + \operatorname{arctanh} y = \operatorname{arctanh} \frac{x+y}{1+xy} \tag{77}$$

to expand both sides of (76). LHS is

$$\begin{aligned} & F_\lambda(\theta_1,\theta_2) + F_\lambda(\theta_1,-\theta_2) \\ = & \operatorname{arctanh} \frac{\lambda(\theta_1+\theta_2)}{1+\lambda\theta_1\theta_2} + \operatorname{arctanh} \frac{\lambda(\theta_1-\theta_2)}{1-\lambda\theta_1\theta_2} \\ = & \operatorname{arctanh} \frac{2\lambda\theta_1(1-\lambda\theta_2^2)}{\lambda^2(\theta_1^2-\theta_2^2)+1-\lambda^2\theta_1^2\theta_2^2}. \end{aligned} \tag{78}$$

RHS is

$$2F_\lambda(\theta_1,0) = \operatorname{arctanh} \frac{2\lambda\theta_1}{1+\lambda^2\theta_1^2}. \tag{79}$$

By comparing (78)(79) and using monotonicity of $\operatorname{arctanh}$, it suffices to prove that

$$\frac{1-\lambda\theta_2^2}{\lambda^2(\theta_1^2-\theta_2^2)+1-\lambda^2\theta_1^2\theta_2^2} \le \frac{1}{1+\lambda^2\theta_1^2}. \tag{80}$$

We have

$$(\lambda^2(\theta_1^2-\theta_2^2)+1-\lambda^2\theta_1^2\theta_2^2) - (1-\lambda\theta_2^2)(1+\lambda^2\theta_1^2) = \lambda(1-\lambda)(1-\lambda\theta_1^2)\theta_2^2 \ge 0. \tag{81}$$

This finishes the proof. ■

**Proof** [Proof of Prop. 9, case $r=3$] By BSC mixture representation of BMS channels (Lemma 6) and Lemma 14. ■

### B.2. Case $r = 4$

Now we prove the $r = 4$ case.

**Lemma 16** *For all $\Delta_1, \Delta_2, \Delta_3 \in [0, \frac{1}{2}]$, we have*

$$C_{\mathrm{SKL}}((\mathrm{BSC}_{\Delta_1} \times \mathrm{BSC}_{\Delta_2} \times \mathrm{BSC}_{\Delta_3}) \circ B_{r,\lambda}) \le \lambda^2 \sum_{i\in[3]} C_{\mathrm{SKL}}(\mathrm{BSC}_{\Delta_i}). \tag{82}$$

*Furthermore, the inequality is strict when $0 < \lambda < 1$ and $\min\{\Delta_1, \Delta_2, \Delta_3\} < \frac{1}{2}$.*

The proof is based on the following inequality. In a preliminary version of the current paper, we proposed this inequality as a conjecture based on numerical computation. Shortly after that, Marwaha (2023) gave a beautiful analytical proof.

**Lemma 17 (Marwaha (2023))** *For* $\lambda, \theta_1, \theta_2, \theta_3 \in [0, 1]$, *we have*

$$\begin{aligned}&\frac{1}{4}(G_\lambda(\theta_1, \theta_2, \theta_3) + G_\lambda(\theta_1, -\theta_2, \theta_3) + G_\lambda(\theta_1, \theta_2, -\theta_3) + G_\lambda(\theta_1, -\theta_2, -\theta_3)) \\ &\le \lambda \sum_{i\in[3]} \theta_i \operatorname{arctanh} \theta_i,\end{aligned} \tag{83}$$

*where*

$$G_\lambda(\theta_1, \theta_2, \theta_3) := (\theta_1 + \theta_2 + \theta_3 + \theta_1\theta_2\theta_3)F_\lambda(\theta_1, \theta_2, \theta_3), \tag{84}$$

$$F_\lambda(\theta_1, \theta_2, \theta_3) := \operatorname{arctanh} \frac{\lambda(\theta_1 + \theta_2 + \theta_3 + \theta_1\theta_2\theta_3)}{1 + \lambda(\theta_1\theta_2 + \theta_2\theta_3 + \theta_3\theta_1)}. \tag{85}$$

*Furthermore, the inequality is strict when* $0 < \lambda < 1$ *and* $\max\{\theta_1, \theta_2, \theta_3\} > 0$.

**Proof** [Proof of Lemma 16] We expand LHS of (82) using BP recursion formula established in Section 4. Let $\theta_i = 1 - 2\Delta_i$ for $i \in [3]$. Then

$$\begin{aligned}&C_{\mathrm{SKL}}((\mathrm{BSC}_{\Delta_1} \times \mathrm{BSC}_{\Delta_2} \times \mathrm{BSC}_{\Delta_3}) \circ B_{r,\lambda}) \\ &= \sum_{x_1=+, x_2, x_3 \in \{\pm\}} \frac{1}{4}\lambda(\theta_1x_1 + \theta_2x_2 + \theta_3x_3 + \theta_1\theta_2\theta_3x_1x_2x_3) \\ &\quad \cdot \operatorname{arctanh} \frac{\lambda(\theta_1x_1 + \theta_2x_2 + \theta_3x_3 + \theta_1\theta_2\theta_3x_1x_2x_3)}{1 + \lambda(\theta_1x_1\theta_2x_2 + \theta_2x_3\theta_3x_3 + \theta_3x_3\theta_1x_1)} \\ &= \frac{\lambda}{4}(G_\lambda(\theta_1, \theta_2, \theta_3) + G_\lambda(\theta_1, -\theta_2, \theta_3) + G_\lambda(\theta_1, \theta_2, -\theta_3) + G_\lambda(\theta_1, -\theta_2, -\theta_3))\end{aligned} \tag{86}$$

where

$$G_\lambda(\theta_1, \theta_2, \theta_3) := (\theta_1 + \theta_2 + \theta_3 + \theta_1\theta_2\theta_3)F_\lambda(\theta_1, \theta_2, \theta_3), \tag{87}$$

$$F_\lambda(\theta_1, \theta_2, \theta_3) := \operatorname{arctanh} \frac{\lambda(\theta_1 + \theta_2 + \theta_3 + \theta_1\theta_2\theta_3)}{1 + \lambda(\theta_1\theta_2 + \theta_2\theta_3 + \theta_3\theta_1)}. \tag{88}$$

By Lemma 17, we have

$$\begin{aligned}&\frac{1}{4}(G_\lambda(\theta_1, \theta_2, \theta_3) + G_\lambda(\theta_1, -\theta_2, \theta_3) + G_\lambda(\theta_1, \theta_2, -\theta_3) + G_\lambda(\theta_1, -\theta_2, -\theta_3)) \\ &\le \lambda \sum_{i\in[3]} \theta_i \operatorname{arctanh} \theta_i \\ &= \lambda \sum_{i\in[3]} C_{\mathrm{SKL}}(\mathrm{BSC}_{\Delta_i}).\end{aligned} \tag{89}$$

Combining (86)(89) we finish the proof. ■

**Proof** [Proof of Prop. 9, case $r = 4$] By BSC mixture representation of BMS channels (Lemma 6) and Lemma 16. ■

### B.3. Handle the critical case

In section we prove the critical case of Theorem 2(i), that for $r = 3, 4$ and $(r-1)d\lambda^2 = 1$, reconstruction is impossible for the BOHT model $\mathrm{BOHT}(2, r, \lambda, D)$.

Before giving the proof, we introduce some more preliminaries on BMS channels.

**Definition 18 (Limit of BMS channels)** *Let $(P_k)_{k\ge 0}$ be a sequence of BMS channels and $P_\infty$ be a BMS channel. For $k \in \mathbb{Z}_{\ge 0} \cup \{\infty\}$, let $\Delta_k \in \left[0, \frac{1}{2}\right]$ denote the $\Delta$-component of $P_k$ and $P_{\Delta_k} \in \mathcal{P}\left(\left[0, \frac{1}{2}\right]\right)$ denote its distribution. We say $(P_k)_{k\ge 0}$ converges weakly to $P_\infty$ if $(P_{\Delta_k})_{k\to\infty}$ converges weakly to $P_{\Delta_\infty}$ as distributions on $\left[0, \frac{1}{2}\right]$.*

The following lemma is a direct consequence of (Gu, 2023, Lemma 11.2) (see also (Richardson and Urbanke, 2008, Theorem 7.24)).

**Lemma 19** *Let $(P_k)_{k\ge 0}$ be a sequence of BMS channels. If $P_{k+1} \le_{\deg} P_k$ for all $k \ge 0$, then $(P_k)_{k\ge 0}$ converges weakly to some BMS channels $P_\infty$.*

Recall that $M_k$ is the BMS channel $\sigma_\rho \mapsto (T_k, \sigma_{L_k})$. Then $M_{k+1} \le_{\deg} M_k$. By Lemma 19, the limit $M_\infty := \lim_{k\to\infty} M_k$ exists. For $k \in \mathbb{Z}_{\ge 0} \cup \{\infty\}$, let $\Delta_k$ denote the $\Delta$-component of $M_k$. Then $P_{\Delta_k}$ converges weakly to $P_{\Delta_\infty}$ as $k \to \infty$.

The proof idea is as follows. By definition of the limit, we have $\mathrm{BP}(M_\infty) = M_\infty$. If $M_\infty$ is non-trivial, then by Eq. (48), we have $C_{\mathrm{SKL}}(\mathrm{BP}(M_\infty)) < C_{\mathrm{SKL}}(M_\infty)$, which leads to contradiction. The actual argument is more involved because SKL capacity can be infinite for some BMS channels.

**Proof** [Proof of Theorem 2(i) critical case] The case $\lambda = 1$ is already handled in Section 4. So we can wlog assume $0 < \lambda < 1$.

Suppose for the sake of contradiction that reconstruction holds. Then the limit channel $M_\infty$ is non-trivial.

We first prove that $\mathbb{P}[\Delta_\infty = 0] = 0$. If not, then by weak convergence, there exists $\delta > 0$ such that for all $\epsilon > 0$, $\lim_{k\to\infty} \mathbb{P}[\Delta_k < \epsilon] > \delta$. Because $\lim_{\epsilon\to 0^+} C_{\mathrm{SKL}}(\mathrm{BSC}_\epsilon) = \infty$, this implies $\lim_{k\to\infty} C_{\mathrm{SKL}}(M_k) = \infty$. However, for all $k \ge 1$, we have $C_{\mathrm{SKL}}(M_k) \le C_{\mathrm{SKL}}(M_1) = dC_{\mathrm{SKL}}(B_{r,\lambda}) < \infty$. Contradiction. So $\mathbb{P}[\Delta_\infty = 0] = 0$.

Because $M_\infty$ is non-trivial, $\mathbb{P}\left[\Delta_\infty = \frac{1}{2}\right] < 1$. So $\mathbb{P}\left[0 < \Delta_\infty < \frac{1}{2}\right] > 0$. By weak convergence, there exists $c > 0$ and a closed interval $I \subseteq \left(0, \frac{1}{2}\right)$ such that $\mathbb{P}[\Delta_k \in I] \ge c$ for $k$ large enough. By Prop. 9, $\forall \delta_1, \ldots, \delta_{r-1} \in I$, we have

$$C_{\mathrm{SKL}}((\mathrm{BSC}_{\delta_1} \times \cdots \times \mathrm{BSC}_{\delta_{r-1}}) \circ B_{r,\lambda}) < \lambda^2 \sum_{i\in[r-1]} C_{\mathrm{SKL}}(\mathrm{BSC}_{\delta_i}). \tag{90}$$

Because $I$ is compact, there exists $\epsilon > 0$ such that $\forall \delta_1, \ldots, \delta_{r-1} \in I$,

$$C_{\mathrm{SKL}}((\mathrm{BSC}_{\delta_1} \times \cdots \times \mathrm{BSC}_{\delta_{r-1}}) \circ B_{r,\lambda}) \le \lambda^2 \sum_{i\in[r-1]} C_{\mathrm{SKL}}(\mathrm{BSC}_{\delta_i}) - \epsilon. \tag{91}$$

For any $k$ satisfying $\mathbb{P}[\Delta_k \in I] \ge c$, we have

$$\begin{aligned}
& C_{\mathrm{SKL}}(\mathrm{BP}(M_k)) \qquad (92)\\
&= d C_{\mathrm{SKL}}\left(M_k^{\times(r-1)} \circ B_{r,\lambda}\right)\\
&= d \mathop{\mathbb{E}}_{\delta_1,\ldots,\delta_{r-1}\sim P_{\Delta_k}} C_{\mathrm{SKL}}((\mathrm{BSC}_{\delta_1}\times\cdots\times \mathrm{BSC}_{\delta_{r-1}})\circ B_{r,\lambda})\\
&\le d \mathop{\mathbb{E}}_{\delta_1,\ldots,\delta_{r-1}\sim P_{\Delta_k}} \left[\lambda^2 \sum_{i\in[r-1]} C_{\mathrm{SKL}}(\mathrm{BSC}_{\delta_i}) - \epsilon \mathbb{1}\{\delta_1,\ldots,\delta_{r-1}\in I\}\right]\\
&\le d((r-1)\lambda^2 C_{\mathrm{SKL}}(M_k) - \epsilon c^{r-1})\\
&= C_{\mathrm{SKL}}(M_k) - d\epsilon c^{r-1}.
\end{aligned}$$

So for $k$ large enough,

$$C_{\mathrm{SKL}}(M_{k+1}) \le C_{\mathrm{SKL}}(M_k) - d\epsilon c^{r-1}. \qquad (93)$$

Because $d\epsilon c^{r-1} > 0$ and $C_{\mathrm{SKL}}(M_k) < \infty$ for all $k \ge 1$, this implies $C_{\mathrm{SKL}}(M_k) < 0$ for $k$ large enough, which cannot be true. This finishes the proof. ■

## Appendix C. Computation of $\chi^2$-contraction coefficients

In this section we prove Prop. 10 and Lemma 11, which computes the $\chi^2$-multi-terminal contraction coefficient $\eta_{\chi^2}^{(m,s)}(B_{r,\lambda})$. In fact, our method works for the more general setting where $B : \{\pm\} \to \{\pm\}^{r-1}$ together with the sign flip $\{\pm\}^{r-1} \to \{\pm\}^{r-1}$ is a BMS channel.

### C.1. Less-noisy preorder

Our method uses the less-noisy preorder, a very useful channel preorder, especially for BMS channels.

**Definition 20 (Less-noisy preorder Körner and Marton (1977))** *Let $P : \mathcal{X} \to \mathcal{Y}$ and $Q : \mathcal{X} \to \mathcal{Z}$ be two channels with the same input alphabet. We say $P$ is less noisy than $Q$, denoted $P \ge_{\ln} Q$, if for every measurable space $\mathcal{W}$, distribution $\pi \in \mathcal{P}(\mathcal{W})$, and channel $R : \mathcal{W} \to \mathcal{X}$, we have $I(\pi, P \circ R) \ge I(\pi, Q \circ R)$.*

Less-noisy preorder behaves nicely under channel transformations, summarized as follows.

- (Composition) Let $P, Q$ be two channels with the same input alphabet $\mathcal{X}$. Let $R : \mathcal{W} \to \mathcal{X}$ be a channel. If $P \le_{\ln} Q$, then $P \circ R \le_{\ln} Q \circ R$.
- (Tensorization) Sutter and Renes (2014); Polyanskiy and Wu (2017) Let $P_1$ and $Q_1$ be two channels with the same input alphabet $\mathcal{X}$, and $P_2$ and $Q_2$ be two channels with the same input alphabet $\mathcal{Y}$. If $P_1 \le_{\ln} Q_1$ and $P_2 \le_{\ln} Q_2$, then $P_1 \times Q_1 \le_{\ln} P_2 \times Q_2$.
- ($\star$-convolution) Let $P_1, P_2, Q_1, Q_2$ be four channels with the same input alphabet. If $P_1 \le_{\ln} Q_1$ and $P_2 \le_{\ln} Q_2$, then $P_1 \star Q_1 \le_{\ln} P_2 \star Q_2$.

Although defined using mutual information, less-noisy preorder is closely related to $\chi^2$-divergence. (Makur and Polyanskiy, 2018, Theorem 1) implies that for BMS channels $P, Q$, if $P \le_{\ln} Q$, then $C_{\chi^2}(P) \le C_{\chi^2}(Q)$. Furthermore, under $\chi^2$-capacity constraint, BEC and BSC are the extremal channels in less-noisy preorder.

**Lemma 21 ((Roozbehani and Polyanskiy, 2019, Lemma 2))** *Among all BMS channels with the same $\chi^2$-capacity $C_{\chi^2}(W) = \eta$ the least noisy one is* BEC *and the most noisy one is* BSC*, i.e.*

$$\mathrm{BSC}_{1/2-\sqrt{\eta}/2} \le_{\ln} W \le_{\ln} \mathrm{BEC}_{1-\eta}\,. \tag{94}$$

### C.2. Binary symmetric model

We consider the more general setting where $B : \{\pm\} \to \{\pm\}^{r-1}$ is a BMS channel. Recall that this is the setting for the binary symmetric BOHT model.

**Proposition 22** *Suppose $B : \{\pm\} \to \{\pm\}^{r-1}$ together with the sign flip $\{\pm\}^{r-1} \to \{\pm\}^{r-1}$ is a BMS channel. Then*

$$\eta_{\chi^2}^{(m,s)}(B) = \sup_{0<\epsilon\le 1} f_B(\epsilon) \tag{95}$$

*where*

$$f_B(\epsilon) := \frac{1}{(r-1)\epsilon} C_{\chi^2}\left(\mathrm{BEC}_{1-\epsilon}^{\times(r-1)} \circ B\right) \tag{96}$$

*is a polynomial of degree $r-2$.*

**Proof** Let $P$ be a non-trivial BMS channel and $\epsilon = C_{\chi^2}(P)$. By Lemma 21, $P \le_{\ln} \mathrm{BEC}_{1-\epsilon}$. Because less-noisy preorder is preserved under tensorization and composition, we have

$$P^{\times(r-1)} \circ B \le_{\ln} \mathrm{BEC}_{1-\epsilon}^{\times(r-1)} \circ B. \tag{97}$$

Then by (Makur and Polyanskiy, 2018, Theorem 1) we have

$$C_{\chi^2}(P^{\times(r-1)} \circ B) \le C_{\chi^2}\left(\mathrm{BEC}_{1-\epsilon}^{\times(r-1)} \circ B\right). \tag{98}$$

So

$$\eta_{\chi^2}^{(m,s)}(B) = \sup_P \frac{C_{\chi^2}(P^{\times(r-1)} \circ B)}{(r-1)C_{\chi^2}(P)} = \sup_{0<\epsilon\le 1} \frac{C_{\chi^2}\left(\mathrm{BEC}_{1-\epsilon}^{\times(r-1)} \circ B\right)}{(r-1)C_{\chi^2}(\mathrm{BEC}_{1-\epsilon})} = \sup_{0<\epsilon\le 1} f_B(\epsilon). \tag{99}$$

It remains to prove that $f_B(\epsilon)$ is a polynomial of degree of $r-2$. By BSC mixture representation, we have

$$\begin{aligned}\epsilon f_B(\epsilon) &= \frac{1}{r-1} C_{\chi^2}\left(\mathrm{BEC}_{1-\epsilon}^{\times(r-1)} \circ B\right) \qquad (100)\\ &= \frac{1}{r-1} \sum_{x\in\{0,1\}^{r-1}} \left(\prod_{i\in[r-1]} \left(\epsilon^{x_i}(1-\epsilon)^{1-x_i}\right)\right) C_{\chi^2}\left(\left(\prod_{i\in[r-1]} \mathrm{BSC}_{(1-x_i)/2}\right) \circ B\right),\end{aligned}$$

which is a degree-$(r-1)$ polynomial. Furthermore, the constant coefficient of $\epsilon f_B(\epsilon)$ is

$$\frac{1}{r-1} C_{\chi^2}\left(\mathrm{BSC}_{1/2}^{\times(r-1)} \circ B\right) = 0. \tag{101}$$

So $f_B(\epsilon)$ is a polynomial of degree $r-2$. ■

For $r = 2$, $f_B(\epsilon)$ is a constant, and we get $\eta_{\chi^2}^{(m,s)}(B) = f_B(1) = C_{\chi^2}(B)$. For $r = 3$, $f_B(\epsilon)$ is a linear function, so $\eta_{\chi^2}^{(m,s)}(B) = \max\{f_B(0), f_B(1)\}$.

Prop. 22 immediately leads to the following corollary.

**Corollary 23 (Non-reconstruction for binary symmetric BOHT)** *Consider a binary symmetric BOHT model* $\mathrm{BOHT}(2, r, \mathrm{Unif}(\{\pm\}), B, D)$ *where* $\mathbb{E}_{t\sim D} t = d$. *If*

$$(r-1)d \sup_{0<\epsilon\le 1} f_B(\epsilon) < 1, \tag{102}$$

*where* $f_B$ *is defined in* (96)*, then reconstruction is impossible.*

**Proof** By Theorem 1(ii) and Prop. 22. ■

### C.3. Special BOHT model

We apply Prop. 22 to the special case where $B = B_{r,\lambda}$.

**Proof** [Proof of Prop. 10] It suffices to prove that $f_{B_{r,\lambda}} = f_{r,\lambda}$, where LHS is defined in (96) and RHS is defined in (27). We have

$$\begin{aligned}
&f_{B_{r,\lambda}}(\epsilon) \\
&= \frac{1}{(r-1)\epsilon} \sum_{x\in\{0,1\}^{r-1}} \left(\prod_{i\in[r-1]} \left(\epsilon^{x_i}(1-\epsilon)^{1-x_i}\right)\right) C_{\chi^2}\left(\left(\prod_{i\in[r-1]} \mathrm{BSC}_{(1-x_i)/2}\right) \circ B_{r,\lambda}\right) \\
&= \frac{1}{(r-1)\epsilon} \sum_{1\le i\le r-1} \binom{r-1}{i} \epsilon^i (1-\epsilon)^{r-1-i} C_{\chi^2}\left(\left(\mathrm{Id}^{\times i} \times 0^{\times(r-1-i)}\right) \circ B_{r,\lambda}\right) \\
&= \frac{1}{r-1} \sum_{1\le i\le r-1} \binom{r-1}{i} \epsilon^{i-1} (1-\epsilon)^{r-1-i} C_{\chi^2}\left(\left(\mathrm{Id}^{\times i} \times 0^{\times(r-1-i)}\right) \circ B_{r,\lambda}\right),
\end{aligned} \tag{103}$$

where $\mathrm{Id} = \mathrm{BSC}_0$ denotes the identity channel and $0 = \mathrm{BSC}_{1/2}$ denotes the trivial channel. Using the BP recursion formula established in Section 4, we have

$$C_{\chi^2}\left(\left(\mathrm{Id}^{\times i} \times 0^{\times(r-1-i)}\right) \circ B_{r,\lambda}\right) = \frac{\lambda^2}{\lambda + (1-\lambda)2^{1-i}} \tag{104}$$

for $1 \le i \le r-1$. Therefore

$$f_{B_{r,\lambda}}(\epsilon) = \frac{1}{r-1} \sum_{1\le i\le r-1} \binom{r-1}{i} \epsilon^{i-1}(1-\epsilon)^{r-1-i} \frac{\lambda^2}{\lambda + (1-\lambda)2^{1-i}} = f_{r,\lambda}(\epsilon). \tag{105}$$

This finishes the proof. ■

**Proof** [Proof of Lemma 11] Note that $f_{r,\lambda}(0) = \lambda^2$. Let us compute $f'_{r,\lambda}(\epsilon)$.

$$\begin{aligned} f'_{r,\lambda}(\epsilon) &= \frac{1}{r-1} \sum_{1 \le i \le r-1} \binom{r-1}{i} \left( \frac{d}{d\epsilon} \left( (1-\epsilon)^{r-1-i} \epsilon^{i-1} \right) \right) \frac{\lambda^2}{\lambda + (1-\lambda)2^{1-i}} \qquad (106) \\ &= \frac{1}{r-1} \sum_{1 \le i \le r-1} \binom{r-1}{i} \left( (i-1)(1-\epsilon)^{r-1-i} \epsilon^{i-2} \right. \\ &\quad \left. -(r-1-i)(1-\epsilon)^{r-2-i} \epsilon^{i-1} \right) \frac{\lambda^2}{\lambda + (1-\lambda)2^{1-i}} \\ &= \frac{1}{r-1} \sum_{1 \le i \le r-1} \binom{r-1}{i} (i-1)(1-\epsilon)^{r-1-i} \epsilon^{i-2} \frac{\lambda^2}{\lambda + (1-\lambda)2^{1-i}} \\ &\quad - \frac{1}{r-1} \sum_{2 \le i \le r} \binom{r-1}{i-1} (r-i)(1-\epsilon)^{r-1-i} \epsilon^{i-2} \frac{\lambda^2}{\lambda + (1-\lambda)2^{2-i}} \\ &= \frac{1}{r-1} \sum_{2 \le i \le r-1} \binom{r-1}{i} (1-\epsilon)^{r-1-i} \epsilon^{i-2} \left( \frac{(i-1)\lambda^2}{\lambda + (1-\lambda)2^{1-i}} - \frac{i\lambda^2}{\lambda + (1-\lambda)2^{2-i}} \right). \end{aligned}$$

When $\lambda \in \left[\frac{1}{5}, 1\right]$, we have

$$\frac{i-1}{\lambda + (1-\lambda)2^{1-i}} \le \frac{i}{\lambda + (1-\lambda)2^{2-i}} \qquad (107)$$

for all integer $i \ge 2$, and the inequality is strict for $i \in \{2\} \cup \mathbb{Z}_{\ge 5}$. Therefore $f'_{r,\lambda}(\epsilon) < 0$ for all $\epsilon \in [0, 1]$. So for $0 < \epsilon \le 1$ we have $f_{r,\lambda}(\epsilon) < f_{r,\lambda}(0) = \lambda^2$. ■

We remark that for fixed $r$, the range $\lambda \in \left[\frac{1}{5}, 1\right]$ could be improved. For example, for $r = 5$ and $\lambda \in \left[\frac{1}{7}, 1\right]$ we have $f_{r,\lambda}(\epsilon) \le \lambda^2$ for all $\epsilon \in [0, 1]$.

### C.4. Handle the critical case

In this section we prove the critical case of Theorem 2(iii), that for $\lambda \in \left[\frac{1}{5}, 1\right]$ and $(r-1)d\lambda^2 = 1$, reconstruction is impossible.

The proof idea is similar to the SKL case (Section B.3). Because $\chi^2$-capacity is a bounded function for BMS channels, the proof is easier than the SKL case.

**Proof** [Proof of Theorem 2(iii) critical case] Suppose for the sake of contraction that reconstruction holds. Then the limit channel $M_\infty := \lim_{k \to \infty} M_k$ is non-trivial. Let $\epsilon := C_{\chi^2}(M_\infty) > 0$. Then $M_\infty \le_{\ln} \mathrm{BEC}_{1-\epsilon}$, and thus $\mathrm{BP}(M_\infty) \le_{\ln} \mathrm{BP}(\mathrm{BEC}_{1-\epsilon})$. So

$$\begin{aligned} C_{\chi^2}(\mathrm{BP}(M_\infty)) &\le C_{\chi^2}(\mathrm{BP}(\mathrm{BEC}_{1-\epsilon})) \le d C_{\chi^2}\left(\mathrm{BEC}_{1-\epsilon}^{\times(r-1)} \circ B_{r,\lambda}\right) \qquad (108) \\ &\le (r-1)d\epsilon f_{r,\lambda}(\epsilon) < (r-1)d\lambda^2\epsilon = \epsilon = C_{\chi^2}(M_\infty), \qquad (109) \end{aligned}$$

where the first step is by (Makur and Polyanskiy, 2018, Theorem 1), the second step is by subadditivity, the third step is by $f_{r,\lambda} = f_{B_{r,\lambda}}$, the fourth step is by Lemma 11, the fifth step is by $(r-1)d\lambda^2 = 1$, and the sixth step is by definition of $\epsilon$. On the other hand, $\mathrm{BP}(M_\infty) = M_\infty$, so $C_{\chi^2}(\mathrm{BP}(M_\infty)) = C_{\chi^2}(M_\infty)$. Contradiction. ■

## Appendix D. Weak recovery threshold for HSBM

In this section we prove Theorem 3. Our proof uses a reduction from HSBM to BOHT, which works in a very general setting. Let us define the general HSBM.

**Definition 24 (General HSBM Angelini et al. (2015); Stephan and Zhu (2022))** *Let $n \ge 1$ (number of vertices), $q \ge 2$ (number of communities), $r \ge 2$ (hyperedge size) be integers. Let $\pi \in \mathcal{P}([q])$ be a distribution with full support. Let $\mathbf{A} \in \left(\mathbb{R}^q_{\ge 0}\right)^{\otimes r}$ be a tensor satisfying*

$$a_{i_1,\ldots,i_r} = a_{i_{\sigma(1)},\ldots,i_{\sigma(r)}} \tag{110}$$

*for any $i_1, \ldots, i_r \in [q]$, $\sigma \in \mathrm{Aut}([r])$. The hypergraph stochastic block model $\mathrm{HSBM}(n, q, r, \pi, \mathbf{A})$ is defined as follows: Let $V = [n]$ be the set of vertices. Generate a random label $X_u$ for all vertices $u \in V$ i.i.d. $\sim \pi$. Then for every $S = \{u_1, \ldots, u_r\} \in \binom{V}{r}$, add hyperedge $S$ to the hypergraph with probability $\frac{a_{X_{u_1},\ldots,X_{u_r}}}{\binom{n}{r-1}}$. The resulting pair $(X, G = (V, E))$ is the output of the model.*

Clearly the above definition generalizes the model $\mathrm{HSBM}(n, 2, r, a, b)$ defined in the introduction.

Let $(X, G) \sim \mathrm{HSBM}(n, q, r, \pi, \mathbf{A})$. We say the model admits weak recovery if there exists an estimator outputting a subset $S \subseteq V$ such that for some $\epsilon > 0$, with probability $1 - o(1)$, there exists $i, j \in [q]$ such that

$$\frac{\#\{v \in S : X_v = i\}}{\#\{v \in V : X_v = i\}} - \frac{\#\{v \in S : X_v = j\}}{\#\{v \in V : X_v = j\}} \ge \epsilon. \tag{111}$$

For $\mathrm{HSBM}(n, 2, r, a, b)$, this definition agrees with the one we gave in the introduction.

In the HSBM, the expected degree (number of hyperedges containing a vertex) of a vertex with label $i \in [q]$ is $d_i \pm o(1)$, where

$$d_i = \sum_{i_1,\ldots,i_{r-1} \in [q]} a_{i,i_1,\ldots,i_{r-1}} \prod_{j \in [r-1]} \pi_{i_j}. \tag{112}$$

If $d_i \ne d_j$ for some $i, j \in [q]$, we can distinguish community $i$ and $j$ using a classifier based on degree, which trivially solves the weak recovery problem. Therefore, we make the following standard assumption.

**Condition 25** *We say the model $\mathrm{HSBM}(n, q, r, \pi, \mathbf{A})$ is degree indistinguishable if $d_i = d_j$ for all $i, j \in [q]$, where $d_i$ is defined in Eq. (112). For such models, we define $d = d_i$ for any $i$.*

For a degree indistinguishable HSBM, the local neighborhood of any vertex corresponds to a BOHT model. This relationship was first shown in Massoulié (2014); Mossel et al. (2015) in the case of two-community symmetric SBMs, and later generalized to various settings Bordenave et al. (2015); Caltagirone et al. (2017); Stephan and Massoulié (2019, 2022); Gu and Polyanskiy (2020); Chin and Sly (2020, 2021); Mossel et al. (2022); Pal and Zhu (2021); Stephan and Zhu (2022).

**Proposition 26 (HSBM-BOHT coupling (Stephan and Zhu, 2022, Prop. 3))** *Let $(X, G) \sim \mathrm{HSBM}(n, q, r, \pi, \mathbf{A})$ be a model satisfying Condition 25. Let $v \in V$ and $k = c \log n$ for some small enough constant $c > 0$ not depending on $n$. Let $B(v, k)$ be the set of vertices with distance $\le k$ to $v$.*

*Let* $(T,\sigma) \sim \mathrm{BOHT}(q,r,\pi,M,\mathrm{Pois}(d))$*, and*

$$M_{i,(i_1,\ldots,i_{r-1})} = \frac{1}{d} a_{i,i_1,\ldots,i_{r-1}} \prod_{j\in[r-1]} \pi_j. \tag{113}$$

*Let $\rho$ be the root of $T$, and $T_k$ be the set of vertices at distance $\le k$ to $\rho$.*

*Then $(G|_{B(v,k)}, X_{B(v,k)})$ can be coupled to $(T_k, \sigma_{T_k})$ with $o(1)$ TV distance.*

In the setting of Prop. 26, we say the model $\mathrm{BOHT}(q,r,\pi,M,\mathrm{Pois}(d))$ is the BOHT model corresponding to $\mathrm{HSBM}(n,q,r,\pi,\mathbf{A})$.

Now we can state the general reduction. This reduction was first established by Mossel et al. (2015) in the case of two-community symmetric SBMs, and later generalized to various settings Gu and Polyanskiy (2020); Mossel et al. (2022).

**Theorem 27 ((Gu, 2023, Theorem 5.15))** *Let $\mathrm{HSBM}(n,q,r,\pi,\mathbf{A})$ be a model satisfying Condition 25. Let $\mathrm{BOHT}(q,r,\pi,M,\mathrm{Pois}(d))$ be the corresponding BOHT model. If reconstruction for the BOHT model is impossible, then weak recovery for the HSBM is impossible.*

The proof of Theorem 27 uses Prop. 26 and that HSBMs have no long range correlations, a fact first established by Mossel et al. (2015) in the case of two-community symmetric SBMs.

**Proposition 28 ((Gu, 2023, Prop. 5.6))** *Let $(X, G=(V,E)) \sim \mathrm{HSBM}(n,q,r,\pi,\mathbf{A})$. Let $A = A(G), B = B(G), C = C(G) \subseteq V$ be a (random) partition of $V$ such that $B$ separates $A$ and $C$ in $G$ (i.e., there exists no hyperedges $S \in E$ intersecting both $A$ and $C$). If $|A \cup B| = o(\sqrt{n})$ a.a.s., then*

$$\mathbb{P}(X_A | X_{B\cup C}, G) = (1 \pm o(1))\mathbb{P}(X_A|X_B, G) \text{ a.a.s.} \tag{114}$$

We omit the proofs of Theorem 27 and Prop. 28 and refer the reader to Gu (2023),

Using Theorem 27 and multi-terminal contraction coefficients we can prove impossibility of weak recovery results for HSBMs.

**Corollary 29 (Impossibility of weak recovery for general HSBM)** *Let $\mathrm{HSBM}(n,q,r,\pi,\mathbf{A})$ be a model satisfying Condition 25. Let $\mathrm{BOHT}(q,r,\pi,M,\mathrm{Pois}(d))$ be the corresponding BOHT model. If any of the conditions in Theorem 1(i)(ii) holds, then weak recovery is impossible.*

**Proof** By Theorem 27 and Theorem 1. ■

In particular, for $\mathrm{HSBM}(n,2,r,a,b)$, we get Theorem 3.

**Proof** [Proof of Theorem 3] By Theorem 27 and Theorem 2. ■

## Appendix E. Reconstruction below the KS threshold

In this section we prove Theorem 4, that for the special BOHT model on a regular or Poisson hypertree, reconstruction is possible below the KS threshold for $r \ge 7$ and $d$ large enough. Our proof is an analysis of evolution of $\chi^2$-capacity (also called magnetization in literature) and Gaussian approximation for large degree.

### E.1. Behavior of $\chi^2$-capacity

**Proposition 30 (Large degree asymptotics)** *Fix $r \in \mathbb{Z}_{\ge 2}$. For any $\epsilon > 0$, there exists $d_0 = d_0(r,\epsilon) > 0$ such that for any $d \ge d_0$ and $\lambda \in [0,1]$ with $(r-1)d\lambda^2 \le 1$, for any BMS channel $P$ we have*

$$|C_{\chi^2}(\mathrm{BP}(P)) - g_{r,d,\lambda}(C_{\chi^2}(P))| \le \epsilon, \tag{115}$$

*where*

$$g_{r,d,\lambda}(x) := \mathbb{E}_{Z\sim\mathcal{N}(0,1)} \tanh\left(s_{r,d,\lambda}(x) + \sqrt{s_{r,d,\lambda}(x)}Z\right), \tag{116}$$

$$s_{r,d,\lambda}(x) := d\lambda^2 \cdot \frac{1}{2}\left((1+x)^{r-1} - (1-x)^{r-1}\right). \tag{117}$$

The rest of this section is devoted to the proof of Prop. 30.

We first describe $\mathrm{BP}(P)$ in terms of the $\theta$-component. Let $P$ be a BMS channel and $P_\theta$ be the $\theta$-component of $P$. Let $t$ be the offspring ($t = d$ for regular hypertrees, $t \sim \mathrm{Pois}(d)$ for Poisson hypertrees). Let $(\theta_{ij})_{i\in[t],j\in[r-1]}$ generated $\stackrel{\text{i.i.d.}}{\sim} P_\theta$, where $\theta_{ij}$ is the $\theta$-component of the $j$-th vertex in the $i$-th downward hyperedge. Let $\theta_i$ be the $\theta$-component of $i$-th hyperedge $P^{\times(r-1)} \circ B$. As discussed in Section 4, given $(\theta_{ij})_{j\in[r-1]}$, $\theta_i$ is equal to (the absolute value of)

$$\frac{\lambda\left(\prod_{j\in[r-1]}(1+\theta_{ij}x_{ij}) - \prod_{j\in[r-1]}(1-\theta_{ij}x_{ij})\right)}{\lambda\left(\prod_{j\in[r-1]}(1+\theta_{ij}x_{ij}) + \prod_{j\in[r-1]}(1-\theta_{ij}x_{ij})\right) + 2(1-\lambda)}. \tag{118}$$

with probability

$$\lambda\left(\prod_{j\in[r-1]}\left(\frac{1}{2}+\frac{1}{2}\theta_{ij}x_{ij}\right) + \prod_{j\in[r-1]}\left(\frac{1}{2}-\frac{1}{2}\theta_{ij}x_{ij}\right)\right) + 2^{2-r}(1-\lambda) \tag{119}$$

for $(x_{ij})_{j\in[r-1]} \in \{\pm\}^{r-1}$, $x_{i1} = +$.

Let $\overline{\theta}$ be the $\theta$-component of the full channel $\mathrm{BP}(P)$. Let $P_{\overline{\theta}}$ denote the distribution of $\overline{\theta}$. Then given $(\theta_i)_{i\in[t]}$, $\overline{\theta}$ is equal to (the absolute value of)

$$\frac{\prod_{i\in[t]}(1+\theta_i x_i) - \prod_{i\in[t]}(1-\theta_i x_i)}{\prod_{i\in[t]}(1+\theta_i x_i) + \prod_{i\in[t]}(1-\theta_i x_i)} \tag{120}$$

with probability

$$\prod_{i\in[t]}\left(\frac{1}{2}+\frac{1}{2}\theta_i x_i\right) + \prod_{i\in[t]}\left(\frac{1}{2}-\frac{1}{2}\theta_i x_i\right) \tag{121}$$

for $(x_1,\dots,x_t)\in\{\pm\}^t$, $x_1=+$. In other words,

$$\begin{aligned}
&P_{\widetilde{\theta}|\theta_1,\dots,\theta_i} && (122)\\
&= \sum_{(x_1,\dots,x_t)\in\{\pm\}^t}\left(\prod_{i\in[t]}\left(\frac{1}{2}+\frac{1}{2}\theta_i x_i\right)\right)\mathbb{1}\left\{\left|\frac{\prod_{i\in[t]}(1+\theta_i x_i)-\prod_{i\in[t]}(1-\theta_i x_i)}{\prod_{i\in[t]}(1+\theta_i x_i)+\prod_{i\in[t]}(1-\theta_i x_i)}\right|\right\}\\
&= \sum_{(x_1,\dots,x_t)\in\{\pm\}^t}\left(\prod_{i\in[t]}\left(\frac{1}{2}+\frac{1}{2}\theta_i x_i\right)\right)\mathbb{1}\left\{\left|\tanh\left(\sum_{i\in[t]}\operatorname{arctanh}(\theta_i x_i)\right)\right|\right\}.
\end{aligned}$$

Write $\widetilde{\theta}_i=\theta_i x_i$. Then $\mathbb{P}[\widetilde{\theta}_i=s\theta_i|\theta_i]=\frac{1}{2}+\frac{1}{2}\theta_i s$ for $s\in\{\pm\}$. So $\widetilde{\theta}_i$ for $i\in[t]$ are iid generated from the same distribution. Let us call this distribution $D$. Then

$$P_{\widetilde{\theta}}=\mathbb{E}_t\mathbb{E}_{\widetilde{\theta}_1,\dots,\widetilde{\theta}_t\overset{\text{i.i.d.}}{\sim}D}\mathbb{1}\left\{\left|\tanh\left(\sum_{i\in[t]}\operatorname{arctanh}\widetilde{\theta}_i\right)\right|\right\} \tag{123}$$

This allows us to use central limit theorems to control the behavior of $\sum_{i\in[t]}\operatorname{arctanh}\widetilde{\theta}_i$.

**Lemma 31** *There exists a constant $d_0=d_0(r)>0$ such that for any $d>d_0$, $\lambda\in[0,1]$ with $(r-1)d\lambda^2\le 1$, and any BMS channel $P$, we have*

$$\left|C_{\chi^2}(P^{\times(r-1)}\circ B)-s_{r,\lambda}(C_{\chi^2}(P))\right|\le O_r(\lambda^3), \tag{124}$$

$$s_{r,\lambda}(x):=\lambda^2\cdot\frac{1}{2}\left((1+x)^{r-1}-(1-x)^{r-1}\right). \tag{125}$$

*where $O_r$ hides a multiplicative factor depending only on $r$.*

**Proof** We have

$$\begin{aligned}
&C_{\chi^2}(P^{\times(r-1)}\circ B)\\
&=\mathbb{E}\theta_i^2\\
&=\mathop{\mathbb{E}}_{\theta_{i1},\dots,\theta_{i,r-1}\overset{\text{i.i.d.}}{\sim}P_\theta}\sum_{\substack{(x_{ij})_{j\in[r-1]}\in\{\pm\}^{r-1}\\ x_{i1}=+}}\\
&\qquad\left(2^{1-r}\cdot\frac{\lambda^2\left(\prod_{j\in[r-1]}(1+\theta_{ij}x_{ij})-\prod_{j\in[r-1]}(1-\theta_{ij}x_{ij})\right)^2}{\lambda\left(\prod_{j\in[r-1]}(1+\theta_{ij}x_{ij})+\prod_{j\in[r-1]}(1-\theta_{ij}x_{ij})\right)+2(1-\lambda)}\right)\\
&=\mathop{\mathbb{E}}_{\theta_{i1},\dots,\theta_{i,r-1}\overset{\text{i.i.d.}}{\sim}P_\theta}\sum_{\substack{(x_{ij})_{j\in[r-1]}\in\{\pm\}^{r-1}\\ x_{i1}=+}}\\
&\qquad\left(2^{-r}\lambda^2\left(\prod_{j\in[r-1]}(1+\theta_{ij}x_{ij})-\prod_{j\in[r-1]}(1-\theta_{ij}x_{ij})\right)^2\right)+O_r(\lambda^3).
\end{aligned}$$

The inner summation satisfies

$$\begin{aligned}
&\sum_{\substack{(x_{ij})_{j\in[r-1]}\in\{\pm\}^{r-1}\\ x_{i1}=+}} \left(\prod_{j\in[r-1]}(1+\theta_{ij}x_{ij}) - \prod_{j\in[r-1]}(1-\theta_{ij}x_{ij})\right)^2 \\
&= \frac{1}{2}\sum_{(x_{ij})_{j\in[r-1]}\in\{\pm\}^{r-1}} \left(\prod_{j\in[r-1]}(1+\theta_{ij}x_{ij}) - \prod_{j\in[r-1]}(1-\theta_{ij}x_{ij})\right)^2 \\
&= \frac{1}{2}\sum_{(x_{ij})_{j\in[r-1]}\in\{\pm\}^{r-1}} \left(\prod_{j\in[r-1]}\left(1+2\theta_{ij}x_{ij}+\theta_{ij}^2\right) - 2\prod_{j\in[r-1]}\left(1-\theta_{ij}^2\right) \right.\\
&\qquad \left. + \prod_{j\in[r-1]}\left(1-2\theta_{ij}x_{ij}+\theta_{ij}^2\right)\right) \\
&= 2^{r-1}\left(\prod_{j\in[r-1]}(1+\theta_{ij}^2) - \prod_{j\in[r-1]}(1-\theta_{ij}^2)\right).
\end{aligned}$$

Therefore

$$\begin{aligned}
&\mathop{\mathbb{E}}_{\theta_{i1},\ldots,\theta_{i,r-1}\overset{\text{i.i.d.}}{\sim}P_\theta}\sum_{\substack{(x_{ij})_{j\in[r-1]}\in\{\pm\}^{r-1}\\ x_{i1}=+}} \left(\prod_{j\in[r-1]}(1+\theta_{ij}x_{ij}) - \prod_{j\in[r-1]}(1-\theta_{ij}x_{ij})\right)^2 \\
&= 2^{r-1}\mathop{\mathbb{E}}_{\theta_{i1},\ldots,\theta_{i,r-1}\overset{\text{i.i.d.}}{\sim}P_\theta}\left(\prod_{j\in[r-1]}(1+\theta_{ij}^2) - \prod_{j\in[r-1]}(1-\theta_{ij}^2)\right) \\
&= 2^{r-1}\left(\left(1+C_{\chi^2}(P)\right)^{r-1} - \left(1-C_{\chi^2}(P)\right)^{r-1}\right).
\end{aligned}$$

Combining everything we finish the proof. ■

**Lemma 32** *There exists a constant $d_0 = d_0(r) > 0$ such that for any $d > d_0$, $\lambda \in [0,1]$ with $(r-1)d\lambda^2 \le 1$, and any BMS channel $P$, we have*

$$\left|\mathbb{E}\operatorname{arctanh}\widetilde{\theta}_i - s_{r,\lambda}(C_{\chi^2}(P))\right| = O_r(\lambda^3), \tag{126}$$

$$\left|\operatorname{Var}(\operatorname{arctanh}\widetilde{\theta}_i) - s_{r,\lambda}(C_{\chi^2}(P))\right| = O_r(\lambda^3). \tag{127}$$

**Proof** Note that $\theta_i = O_r(\lambda)$ almost surely. When $d$ is large enough, $\lambda$ is small enough, and $\operatorname{arctanh}\theta_i = \theta_i + O_r(\lambda^3)$ almost surely by Taylor expansion. Then

$$\mathbb{E}\operatorname{arctanh}\widetilde{\theta}_i = \mathbb{E}[\theta_i\operatorname{arctanh}\theta_i] = \mathbb{E}\theta_i^2 + O_r(\lambda^4), \tag{128}$$

$$\mathbb{E}(\operatorname{arctanh}\widetilde{\theta}_i)^2 = \mathbb{E}(\operatorname{arctanh}\theta_i)^2 = \mathbb{E}\theta_i^2 + O_r(\lambda^4). \tag{129}$$

By Lemma 31, we have

$$\mathbb{E}\theta_i^2 = s_{r,\lambda}(C_{\chi^2}(P)) + O_r(\lambda^3). \tag{130}$$

This already implies the statement on $\mathbb{E} \operatorname{arctanh} \widetilde{\theta}_i$. For the statement on $\operatorname{Var}(\operatorname{arctanh} \widetilde{\theta}_i)$, we note that

$$\mathbb{E} \operatorname{arctanh} \widetilde{\theta}_i = s_{r,\lambda}(C_{\chi^2}(P)) + O_r(\lambda^3) = O_r(\lambda^2). \tag{131}$$

So

$$\begin{aligned} \operatorname{Var}(\operatorname{arctanh} \widetilde{\theta}_i) &= \mathbb{E}(\operatorname{arctanh} \widetilde{\theta}_i)^2 - \left(\mathbb{E} \operatorname{arctanh} \widetilde{\theta}_i\right)^2 \\ &= s_{r,\lambda}(C_{\chi^2}(P)) + O_r(\lambda^3). \end{aligned} \tag{132}$$

This finishes the proof. ■

Now we recall a normal approximation result from (Mossel et al., 2022, Prop. 5.3). We only need the scalar version of it.

**Lemma 33 (Mossel et al. (2022))** *Let $\phi : \mathbb{R} \to \mathbb{R}$ be a thrice differentiable and bounded function with bounded derivatives up to third order. Let $V_1, \ldots, V_t \in \mathbb{R}$ be independent random real numbers. Suppose there exists deterministic numbers $\mu, \sigma \in \mathbb{R}$ such that the following holds: for some constant $C > 0$, almost surely*

$$\max \left\{ \left| \sum_{j\in[t]} \mathbb{E}V_j - \mu \right|, \left| \sum_{j\in[t]} \operatorname{Var}(V_j) - \sigma^2 \right| \right\} \le Ct^{-1/2}, \tag{133}$$

$$\max \left\{ |\mu|, |\sigma^2| \right\} \le C, \qquad \max_{j\in[t]} |V_j| \le Ct^{-1/2}. \tag{134}$$

*Then for any $\epsilon > 0$, there exists $t_0 = t_0(\epsilon, \phi, C)$ such that if $t > t_0$, then*

$$\left| \mathbb{E}\phi\left( \sum_{j\in[t]} V_j \right) - \mathbb{E}_{W\sim\mathcal{N}(\mu,\sigma^2)}\phi(W) \right| \le \epsilon. \tag{135}$$

We now have everything we need for the proof of Prop. 30.

**Proof** [Proof of Prop. 30] **Regular hypertree:** Define $\widetilde{\theta}$ as $\mathbb{P}[\widetilde{\theta} = s\theta | \theta] = \frac{1}{2} + \theta s$ for $s \in \{\pm\}$. Then

$$C_{\chi^2}(\mathrm{BP}(P)) = \mathbb{E}\widetilde{\theta} = \mathbb{E}_{\widetilde{\theta}_1,\ldots,\widetilde{\theta}_t \overset{\text{i.i.d.}}{\sim} D} \tanh\left( \sum_{i\in[t]} \operatorname{arctanh} \widetilde{\theta}_i \right). \tag{136}$$

In fact, the equality holds with $\tanh$ replaced by $\tanh^2$. We use the $\tanh$ form here because it is slightly simpler.

Now we apply Lemma 33 with

$$\phi(x) = \tanh x, \quad V_i = \operatorname{arctanh} \widetilde{\theta}_i, \quad \mu = \sigma^2 = d s_{r,\lambda}(C_{\chi^2}(P)) = s_{r,d,\lambda}(C_{\chi^2}(P)). \tag{137}$$

The conditions in Lemma 33 are satisfied by Lemma 32 and because $\lambda = O(d^{-1/2})$. This finishes the proof.

**Poisson hypertree:** Fix $\epsilon > 0$. Let $t \sim \mathrm{Pois}(d)$. By Poisson tail bounds, we have $\mathbb{P}[|t-d| > d^{0.6}] < \epsilon/3$ for large enough $d$ (depending only on $\epsilon$). We apply Lemma 33 for every $t \in [d - d^{0.6}, d + d^{0.6}]$, with $\mu = \sigma^2 = s_{r,t,\lambda}(C_{\chi^2}(P))$ and error tolerance $\epsilon/3$. Note that

$$\left|s_{r,d,\lambda}(C_{\chi^2}(P)) - s_{r,t,\lambda}(C_{\chi^2}(P))\right| = O_r(d^{-0.4}). \tag{138}$$

So for $d$ large enough (depending only on $\epsilon, r$), we have

$$\left|g_{r,d,\lambda}(C_{\chi^2}(P)) - g_{r,t,\lambda}(C_{\chi^2}(P))\right| \le \epsilon/3 \tag{139}$$

by continuity of $g_r$ (Lemma 34).

Therefore we have

$$\begin{aligned}
&\left|C_{\chi^2}(\mathrm{BP}(P)) - g_{r,d,\lambda}(C_{\chi^2}(P))\right| \\
&= \left|\mathbb{E}_{t\sim\mathrm{Pois}(d)} C_{\chi^2}((P^{\times(r-1)} \circ B)^{\star t}) - g_{r,d,\lambda}(C_{\chi^2}(P))\right| \\
&\le \mathbb{E}_{t\sim\mathrm{Pois}(d)} \mathbb{1}\{|t-d| \le d^{0.6}\} \left|C_{\chi^2}((P^{\times(r-1)} \circ B)^{\star t}) - g_{r,t,\lambda}(C_{\chi^2}(P))\right| \\
&\quad + \mathbb{E}_{t\sim\mathrm{Pois}(d)} \mathbb{1}\{|t-d| \le d^{0.6}\} \left|g_{r,t,\lambda}(C_{\chi^2}(P)) - g_{r,d,\lambda}(C_{\chi^2}(P))\right| \\
&\quad + \mathbb{E}_{t\sim\mathrm{Pois}(d)} \mathbb{1}\{|t-d| > d^{0.6}\} |C_{\chi^2}((P^{\times(r-1)} \circ B)^{\star t}) - g_{r,d,\lambda}(C_{\chi^2}(P))| \\
&\le \epsilon/3 + \epsilon/3 + \epsilon/3 = \epsilon.
\end{aligned}$$

Note that $C_{\chi^2}(P) \in [0,1]$ for any BMS channel $P$, and $g_{r,d,\lambda}(x) \in [0,1]$ for all $x \in [0,1]$. ■

### E.2. Properties of functions

Theorem 4 then follows from analyzing properties of the function $g_{r,d,\lambda}$. For $r \ge 2$, we define

$$g_r(x) := \mathbb{E}_{Z\sim\mathcal{N}(0,1)} \tanh\left(s_r(x) + \sqrt{s_r(x)} Z\right), \tag{140}$$

$$s_r(x) := \frac{1}{2(r-1)} \left((1+x)^{r-1} - (1-x)^{r-1}\right). \tag{141}$$

**Lemma 34** *For any $r \ge 2$, the function $g_r$ is strictly increasing and continuous differentiable on $[0,1]$.*

**Proof** Note that $s_r(x)$ is continuous and increasing on $[0,1]$. Therefore it suffices to prove that

$$g(s) := \mathbb{E}_{Z\sim\mathcal{N}(0,1)} \tanh\left(s + \sqrt{s} Z\right) \tag{142}$$

is continuous and increasing on $\mathbb{R}_{\ge 0}$. This statement is in fact equivalent to the $q = 2$ case in (Sly, 2011, Lemma 4.4), after a suitable change of variables. ■

**Lemma 35** *For $r \ge 7$, there exists $x \in (0,1)$ such that $g_r(x) > x$.*

**Proof** We can numerically verify that $g_7(0.8) > 0.8$. Note that $s_r(0.8)$ is increasing for $r \ge 7$. Therefore for $r \ge 7$, we have $g_r(0.8) \ge g_7(0.8) > 0.8$. ■

### E.3. Proof of Theorem 4

We are now ready to prove Theorem 4.

**Proof** [Proof of Theorem 4] Choose $x \in (0,1)$ so that $g_r(x) > x$ via Lemma 35. By continuity of $g_r$ (Lemma 34), there exists $\epsilon > 0$ such that $g_{r,d,\lambda}(x) > x + \epsilon$ for $(r-1)d\lambda^2 = 1-\epsilon$. Note that $g_{r,d,\lambda}(x)$'s dependence on $d$ and $\lambda$ is only through $d\lambda^2$.

Take $d_0 = d_0(r,\epsilon)$ in Prop. 30. For any $d > d_0$, choose $\lambda \in [0,1]$ such that $(r-1)d\lambda^2 = 1-\epsilon$. By Prop. 30, choice of $\epsilon$, and Lemma 34, for all BMS $P$ with $C_{\chi^2}(P) \ge x$ we have

$$C_{\chi^2}(\mathrm{BP}(P)) \ge g_{r,d,\lambda}(C_{\chi^2}(P)) - \epsilon \ge x. \tag{143}$$

Therefore

$$\lim_{k\to\infty} I_{\chi^2}(\sigma_\rho; T_k, \sigma_{L_k}) = \lim_{k\to\infty} C_{\chi^2}(M_k) \ge x. \tag{144}$$

Finally

$$\lim_{k\to\infty} I(\sigma_\rho; T_k, \sigma_{L_k}) \ge \lim_{k\to\infty} \frac{\log e}{2} I_{\chi^2}(\sigma_\rho; T_k, \sigma_{L_k}) \ge \frac{x \log e}{2}, \tag{145}$$

where the first step is because $C(P) \ge \frac{\log e}{2} C_{\chi^2}(P)$ for any BMS $P$. ■