Graphs of Joint Types, Noninteractive Simulation, and Stronger Hypercontractivity

Lei Yu, *Member, IEEE*, Venkat Anantharam, *Fellow, IEEE*, and Jun Chen, *Senior Member, IEEE*

***Abstract*— In this paper, we study the type graph, namely, a bipartite graph induced by a joint type. We investigate the maximum edge density of induced bipartite subgraphs of this graph having a number of vertices on each side on an exponential scale in the length n of the type. This can be seen as an isoperimetric problem. We provide asymptotically sharp bounds for the exponent of the maximum edge density as the length of the type goes to infinity. We also study the biclique rate region of the type graph, which is defined as the set of (R_1, R_2) such that there exists a biclique of the type graph which has respectively 2^{nR_1} and 2^{nR_2} vertices on the two sides. We provide asymptotically sharp bounds for the biclique rate region as well. We then discuss the connections of these results to noninteractive simulation and hypercontractivity inequalities. Furthermore, as an application of our results, a new outer bound for the zero-error capacity region of the binary adder channel is provided, which improves the previously best known bound, due to Austrin, Kaski, Koivisto, and Nederlof. Our proofs in this paper are based on the method of types and linear algebra.**

***Index Terms*— Graphs of joint types, noninteractive simulation, small-set expansion, isoperimetric inequalities, hypercontractivity, binary adder channel.**

I. Introduction

Let $\mathcal{X}$ and $\mathcal{Y}$ be two finite sets. Let T_X be an n-type on $\mathcal{X}$, i.e., an empirical distribution of sequences from $\mathcal{X}^n$. Let $\mathcal{T}_{T_X}^{(n)}$, or $\mathcal{T}_{T_X}$ for short, be the n-type class with respect to T_X, i.e., the set of sequences of length n having the type T_X. Similarly, let T_{XY} be a joint n-type[1] on $\mathcal{X} \times \mathcal{Y}$ and $\mathcal{T}_{T_{XY}}^{(n)}$, or $\mathcal{T}_{T_{XY}}$ for short, the joint n-type class with respect to T_{XY}. Note that $\mathcal{T}_{T_{XY}} \subseteq \mathcal{T}_{T_X} \times \mathcal{T}_{T_Y}$, where T_X, T_Y are the marginal types corresponding to the joint type T_{XY}. In this paper, we consider the undirected bipartite graph $G_{T_{XY}}$ whose vertex set is $\mathcal{T}_{T_X} \cup \mathcal{T}_{T_Y}$ and whose edge set can be identified with $\mathcal{T}_{T_{XY}}$, defined as follows. Consider $\mathbf{x} \in \mathcal{T}_{T_X}$ and $\mathbf{y} \in \mathcal{T}_{T_Y}$ as vertices of $G_{T_{XY}}$. Two vertices $\mathbf{x}, \mathbf{y}$ are joined by an edge if and only if $(\mathbf{x}, \mathbf{y}) \in \mathcal{T}_{T_{XY}}$. The graph $G_{T_{XY}}$ is termed the graph of T_{XY} or, more succinctly, a type graph [2]. For brevity, when there is no ambiguity, we use the abbreviated notation G for $G_{T_{XY}}$.

For subsets $\mathcal{A} \subseteq \mathcal{T}_{T_X}, \mathcal{B} \subseteq \mathcal{T}_{T_Y}$, we obtain an induced subgraph $G[\mathcal{A}, \mathcal{B}]$ of G, whose vertex set is the union of $\mathcal{A}$ and $\mathcal{B}$, and where $\mathbf{x}, \mathbf{y}$ are joined by an edge if and only if they are joined by an edge in G. For the induced subgraph $G[\mathcal{A}, \mathcal{B}]$, the (edge) density $\rho(G[\mathcal{A}, \mathcal{B}])$ is defined as

$$\rho(G[\mathcal{A}, \mathcal{B}]) := \frac{\#\text{ of edges in } G[\mathcal{A}, \mathcal{B}]}{|\mathcal{A}||\mathcal{B}|}.$$

Thus we have $\rho(G[\mathcal{A}, \mathcal{B}]) = \frac{|(\mathcal{A}\times\mathcal{B})\cap\mathcal{T}_{T_{XY}}|}{|\mathcal{A}||\mathcal{B}|}$. Since[2] $|\mathcal{T}_{T_X^{(n)}}| \doteq 2^{nH_{T^{(n)}}(X)}$ [3], it follows that $\rho(G) = \frac{|\mathcal{T}_{T_{XY}^{(n)}}|}{|\mathcal{T}_{T_X^{(n)}}||\mathcal{T}_{T_Y^{(n)}}|} \doteq 2^{-nI_{T^{(n)}}(X;Y)}$ for any sequence of joint types $\{T_{XY}^{(n)}\}$, where $I_{T^{(n)}}(X;Y)$ denotes the mutual information of the pair (X, Y) having the joint distribution $T_{XY}^{(n)}$, taken to the base 2. Moreover, if we only fix T_X, T_Y, $\mathcal{A}$, and $\mathcal{B}$, then $T_{XY} \in \mathcal{C}_n(T_X, T_Y) \mapsto \rho(G_{T_{XY}}[\mathcal{A}, \mathcal{B}])$ forms a probability mass function, i.e.,

$$\rho(G_{T_{XY}}[\mathcal{A}, \mathcal{B}]) \geq 0,$$
$$\sum_{T_{XY}\in\mathcal{C}_n(T_X,T_Y)} \rho(G_{T_{XY}}[\mathcal{A}, \mathcal{B}]) = 1,$$

where $\mathcal{C}_n(T_X, T_Y)$ denotes the set of joint types T_{XY} with marginals T_X, T_Y. We term this distribution a type distribution, which roughly speaking can be considered as a generalization from binary alphabets to arbitrary finite alphabets of the classic distance distribution in coding theory; please refer to [4] for the distance distribution of a single code, and [5] for the distance distribution between two codes.

[1]We attribute the parameter n to T_{XY}.

[2]Throughout this paper, we write $a_n \doteq b_n$ to denote $a_n = b_n 2^{o(n)}$.

Manuscript received 13 January 2022; revised 6 February 2023; accepted 14 January 2024. Date of publication 24 January 2024; date of current version 19 March 2024. The work of Lei Yu and Venkat Anantharam was supported in part by NSF under Grant CNS–1527846, Grant CCF–1618145, Grant CCF-1901004, and Grant CIF-2007965; in part by the NSF Science and Technology Center (Science of Information) under Grant CCF—0939370; and in part by the William and Flora Hewlett Foundation Supported Center for Long Term Cybersecurity at Berkeley. The work of Lei Yu was supported in part by NSFC under Grant 62101286 and in part by the Fundamental Research Funds for the Central Universities of China (Nankai University) under Grant 054-63233073. An earlier version of this paper was presented in part at the 2021 IEEE International Symposium on Information Theory [DOI: 10.1109/ISIT45174.2021.9517992]. *(Corresponding author: Lei Yu.)*

Lei Yu is with the School of Statistics and Data Science, LPMC, KLMDASR, and LEBPS, Nankai University, Tianjin 300071, China (e-mail: leiyu@nankai.edu.cn).

Venkat Anantharam is with the Department of Electrical Engineering and Computer Sciences, University of California at Berkeley, Berkeley, CA 94720 USA (e-mail: ananth@berkeley.edu).

Jun Chen is with the Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON L8S 4K1, Canada (e-mail: chenjun@mcmaster.ca).

Communicated by A. Somekh-Baruch, Associate Editor for Shannon Theory and Information Measures.

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TIT.2024.3357859.

Digital Object Identifier 10.1109/TIT.2024.3357859

Given $1 \leq M_1 \leq |\mathcal{T}_{T_X}|, 1 \leq M_2 \leq |\mathcal{T}_{T_Y}|$, define the maximal density of subgraphs with size (M_1, M_2) as

$$\Gamma_n(M_1, M_2) := \max_{\substack{\mathcal{A}\subseteq\mathcal{T}_{T_X}, \mathcal{B}\subseteq\mathcal{T}_{T_Y}: \\ |\mathcal{A}|=M_1, |\mathcal{B}|=M_2}} \rho(G[\mathcal{A}, \mathcal{B}]). \tag{1}$$

Recall that $T_{X|Y}$ and $T_{Y|X}$ denote the conditional types corresponding to the joint type T_{XY}. For a sequence $\mathbf{x} \in \mathcal{T}_{T_X}$, let

$$\mathcal{T}_{T_{Y|X}}(\mathbf{x}) := \{\mathbf{y} \in \mathcal{Y}^n : (\mathbf{x}, \mathbf{y}) \in \mathcal{T}_{T_{XY}}\}$$

denote the corresponding conditional type class. Since $N_1 := |\mathcal{T}_{T_{Y|X}}(\mathbf{x})|$ is independent of $\mathbf{x} \in \mathcal{T}_{T_X}$, the degrees of the vertices $\mathbf{x} \in \mathcal{T}_{T_X}$ are all equal to the constant N_1. Similarly, the degrees of the vertices $\mathbf{y} \in \mathcal{T}_{T_Y}$ are all equal to the constant $N_2 := |\mathcal{T}_{T_{X|Y}}(\mathbf{y})|$. Hence we have

$$\begin{aligned} &|\mathcal{B}|\rho(G[\mathcal{A}, \mathcal{B}]) + |\mathcal{B}^c|\rho(G[\mathcal{A}, \mathcal{B}^c]) \\ &= \frac{|(\mathcal{A} \times \mathcal{T}_{T_Y}) \cap \mathcal{T}_{T_{XY}}|}{|\mathcal{A}|} \\ &= \frac{\sum_{\mathbf{x}\in\mathcal{A}} |\mathcal{T}_{T_{Y|X}}(\mathbf{x})|}{|\mathcal{A}|} = N_1, \end{aligned}$$

where $\mathcal{B}^c := \mathcal{T}_{T_Y} \backslash \mathcal{B}$. Thus, over $\mathcal{A}, \mathcal{B}$ with fixed sizes, maximizing $\rho(G[\mathcal{A}, \mathcal{B}])$ is equivalent to minimizing $\rho(G[\mathcal{A}, \mathcal{B}^c])$ (or $\rho(G[\mathcal{A}^c, \mathcal{B}])$). In other words, determining the maximal density is in fact an edge-isoperimetric problem which concerns minimizing the number of or weighted sum of edges between a set of vertices and its complement. Furthermore, given $\mathcal{A} \subseteq \mathcal{T}_{T_X}$ and M_2, we see that

$$\begin{aligned} &\max_{\mathcal{B}\subseteq\mathcal{T}_{T_Y}:|\mathcal{B}|=M_2} \rho(G[\mathcal{A}, \mathcal{B}]) \\ &= \frac{1}{|\mathcal{A}|M_2} \max_{\mathcal{B}\subseteq\mathcal{T}_{T_Y}:|\mathcal{B}|=M_2} \sum_{\mathbf{y}\in\mathcal{B}} |\mathcal{A} \cap \mathcal{T}_{T_{X|Y}}(\mathbf{y})|, \end{aligned}$$

and the maximum is attained by $\mathcal{B}^*$ such that[3] $|\mathcal{A} \cap \mathcal{T}_{T_{X|Y}}(\mathbf{y})| \geq |\mathcal{A}\cap\mathcal{T}_{T_{X|Y}}(\mathbf{y}')|$ for any $\mathbf{y} \in \mathcal{B}^*, \mathbf{y}' \notin \mathcal{B}^*$. Hence, $M_2 \mapsto \max_{\mathcal{B}\subseteq\mathcal{T}_{T_Y}:|\mathcal{B}|=M_2} \rho(G[\mathcal{A}, \mathcal{B}])$ is nonincreasing, which implies that $\Gamma_n(M_1, M_2)$ is nonincreasing in one parameter given the other parameter.

Let[4]

$$\mathcal{R}_X^{(n)} := \{\frac{1}{n} \log M_1 : M_1 \in [|\mathcal{T}_{T_X}|]\}, \tag{2}$$

$$\mathcal{R}_Y^{(n)} := \{\frac{1}{n} \log M_2 : M_2 \in [|\mathcal{T}_{T_Y}|]\}, \tag{3}$$

where the logarithm is taken to the base 2. Given a joint n-type T_{XY}, define the exponent of maximal density for a pair $(R_1, R_2) \in \mathcal{R}_X^{(n)} \times \mathcal{R}_Y^{(n)}$ as

$$E_n(R_1, R_2) := -\frac{1}{n} \log \Gamma_n(2^{nR_1}, 2^{nR_2}). \tag{4}$$

If the edge density of a subgraph in a bipartite graph G is equal to 1, then this subgraph is called a biclique of G. Along these lines, we define the biclique rate region of T_{XY} as

$$\begin{aligned} \mathcal{R}_n(T_{XY}) := \{(R_1, R_2) \in \mathcal{R}_X^{(n)} \times \mathcal{R}_Y^{(n)} : \\ \Gamma_n(2^{nR_1}, 2^{nR_2}) = 1\}. \end{aligned} \tag{5}$$

Observe that any n-type T_{XY} can also be viewed as a kn-type for $k \geq 1$. With an abuse of notation, we continue to use T_{XY} to denote the corresponding kn-type. With this in mind, for an n-type T_{XY} define the asymptotic exponent of maximal density for a pair $(R_1, R_2) \in \mathcal{R}_X^{(n)} \times \mathcal{R}_Y^{(n)}$ as[5]

$$E(R_1, R_2) := \lim_{k\to\infty} -\frac{1}{kn} \log \Gamma_{kn}(2^{knR_1}, 2^{knR_2}), \tag{6}$$

and the asymptotic biclique rate region as[6]

$$\mathcal{R}(T_{XY}) := \text{closure} \bigcup_{k\geq 1} \mathcal{R}_{kn}(T_{XY}). \tag{7}$$

The blocklengths considered here are taken as multiples of n, since the limit and union above are taken for fixed T_{XY}, but T_{XY} is not always an m-type for an arbitrary integer m.

In this paper we are interested in characterizing the limits $E(R_1, R_2)$ and $\mathcal{R}(T_{XY})$, and in bounding the corresponding convergence rates.

A. Motivations

Our motivations for studying the type graph have the following three aspects.

1) The method of types is a classic and powerful tool in information theory. In this method, the basic unit is the (joint) type or (joint) type class. To the authors' knowledge, it is not well understood how the sequence pairs are distributed in a joint type class. The maximal density (as well as the biclique rate region) measures how concentrated are the joint-type sequence pairs by counting the number of joint-type sequence pairs in each "local" rectangular subset. Hence, our study of the type graph deepens the understanding of the distribution (or structure) of sequence pairs in a joint type class. The first study on this topic can be traced back to Han and Kobayashi's work [8], and it was also investigated in [2], [9], and [10] recently. In all these works, either a typicality graph (an approximate version of the type graph) or an approximate version of a biclique of the type graph was considered. In contrast, we consider the exact version of a biclique of the type graph, which results in a rate region different from theirs.
2) Observe that if one starts with a pair sequence $(\mathbf{x}, \mathbf{y})$ in the joint type class $\mathcal{T}_{T_{XY}}$, then the type graph can be constructed from the set of all pair sequences resulting

[3] This condition is closely related to a classic concept, the η-image of a set, which was exploited in the context of the image size characterization in [6] and [7].

[4] We use the notation $[m : n] := \{m, m + 1, \ldots, n\}$ and $[n] := [1 : n]$.

[5] The limit exists because $\log \Gamma_{kn}(2^{knR_1}, 2^{knR_2})$ is subadditive in k for a given n-type T_{XY}. Further, given T_{XY}, the limit does not depend on the value of n that we attribute to T_{XY}.

[6] Using a product construction, we see that $k\mathcal{R}_{kn}(T_{XY})$ is "superadditive" in k, i.e., $k_1\mathcal{R}_{k_1n}(T_{XY})+k_2\mathcal{R}_{k_2n}(T_{XY}) \subseteq (k_1+k_2)\mathcal{R}_{(k_1+k_2)n}(T_{XY})$. Hence $\mathcal{R}(T_{XY}) = \text{closure} \lim_{k\to\infty} \mathcal{R}_{kn}(T_{XY})$ and, moreover, $\mathcal{R}(T_{XY})$ is only dependent on T_{XY} and independent of the value of n that we attribute to T_{XY}.

from permutations of this pair sequence. Thus, unlike other well-studied large graphs, the type graph is deterministic rather than stochastic. There are relatively few works focusing on deterministic large graphs. Hence, as a purely combinatorial problem, studying the type graph is of independent interest.

3) The maximal and minimal density problems for type graphs are closely related to noninteractive simulation problems (or noise stability problems) and hypercontractivity inequalities. Hence, studying the type graph could provide more insights to these related topics.

B. Related Works

Han and Kobayashi [8] introduced a concept similar to the asymptotic biclique rate region defined in this paper. However, roughly speaking, their definition is an approximate version of our definition, in the sense that in their definition, for a distribution P_{XY} (not necessarily a type), type classes are replaced with typical sets with respect to P_{XY}, and the constraint $\Gamma_n(2^{nR_1}, 2^{nR_2}) = 1$ is replaced with $\Gamma_n(2^{nR_1^{(n)}}, 2^{nR_2^{(n)}}) \to 1$ as $n \to \infty$ for a sequence of types $T_{XY}^{(n)}$ converging to P_{XY} and a sequence of pairs $(R_1^{(n)}, R_2^{(n)})$ converging to (R_1, R_2). This approximate version was also investigated in [2], [9], and [10].

In fact, the maximal and minimal density problems on a type graph are equivalent to the noninteractive simulation problem in some sense. Given a joint distribution P_{XY}^n, the noninteractive simulation problem concerns estimating the maximal and minimal joint probability $P_{XY}^n(\mathcal{A} \times \mathcal{B})$ when the marginal probabilities $P_X^n(\mathcal{A})$ and $P_Y^n(\mathcal{B})$ are given. The study of the noninteractive simulation problem dates back to Gács and Körner's and Witsenhausen's seminal papers [11], [12]. Most of the existing works on this topic focus on doubly symmetric binary sources (DSBSes). For the DSBS, by utilizing the tensorization property of maximal correlation, Witsenhausen proved sharp bounds on $P_{XY}^n(\mathcal{A} \times \mathcal{B})$ for the case $P_X^n(\mathcal{A}) = P_Y^n(\mathcal{B}) = \frac{1}{2}$, where the upper and lower bounds are respectively attained by symmetric $(n-1)$-subcubes (e.g., $\mathcal{A} = \mathcal{B} = \{\mathbf{x} : x_1 = 1\}$) and anti-symmetric $(n-1)$-subcubes (e.g., $\mathcal{A} = -\mathcal{B} = \{\mathbf{x} : x_1 = 1\}$). Recently, by combining Fourier analysis with a coding-theoretic result, the first author and Tan [5] derived the sharp upper bound for the case $P_X^n(\mathcal{A}) = P_Y^n(\mathcal{B}) = \frac{1}{4}$, where the upper bound is attained by symmetric $(n-2)$-subcubes (e.g., $\mathcal{A} = \mathcal{B} = \{\mathbf{x} : x_1 = x_2 = 1\}$). Kahn et al. [13] first applied the single-function version of (forward) hypercontractivity inequalities to obtain bounds for the noninteractive simulation problem, by replacing nonnegative functions in the hypercontractivity inequalities with Boolean functions. Mossel and O'Donnell [14], [15] applied the two-function version of hypercontractivity inequalities to obtain bounds in a similar way. Kamath and the second author [16] improved the use of hypercontractivity inequalities in a slightly different way, specifically by replacing nonnegative functions with two-valued functions (not restricted to be $\{0, 1\}$-valued). Furthermore, as mentioned previously, Ordentlich et al. [17] studied the regime in which $P_X^n(\mathcal{A}_n), P_Y^n(\mathcal{B}_n)$ vanish exponentially fast, and they solved the limiting cases $\rho \to 0, 1$. The symmetric case $P_X^n(\mathcal{A}_n) = P_Y^n(\mathcal{B}_n)$ in this exponential regime was solved by Kirshner and Samorodnitsky [18]. Furthermore, the noninteractive simulation problem for Gaussian sources was investigated in [19] and [20], and the ones with Markov chain noise models and multi-terminal versions of noninteractive simulation problems have also been studied in the literature; e.g., [14] and [21]. We refer readers to the monograph [22] for a comprehensive introduction to this topic.

Brascamp–Lieb (BL) inequalities constitute a class of inequalities that generalize the families of Hölder inequalities. Hypercontractivity inequalities are special cases of BL inequalities. Hypercontractivity inequalities were investigated in [23], [24], [25], [26], [27], [28], [29], [30], [31], and [32] among others. Information-theoretic characterizations of the BL (and hypercontractivity) inequalities can be traced back to Ahlswede and Gács's seminal work [29], where a related quantity, known as the hypercontractivity constant, was expressed in terms of relative entropies. The information-theoretic characterization for the forward BL inequalities on Euclidean spaces was given in [33]; this was independently discovered later [34] in the case of finite alphabets. An information-theoretic characterization of the reverse BL inequalities for finite alphabets was provided in [35], [36], and [37]. By using Fenchel duality, the extension of the characterization for forward inequalities to arbitrary measurable spaces and the extension of the characterization for reverse inequalities to Polish spaces under certain compactness conditions were done in [38]. These compactness conditions were removed in [39] by using large deviations theory.

C. Main Contributions

Out main contribution in this paper is the complete characterization of the asymptotic biclique rate region for any joint type defined on finite alphabets. We observe that, in general, the asymptotic biclique rate region defined by us is a subset (in general, a strict subset) of the approximate one defined by Han and Kobayashi [8]. In fact, their definition for a distribution P_{XY} is equal to the asymptotic rate region of a sequence of n-types $\{T_{XY}^{(n)}\}$ approaching P_{XY}, which satisfy the condition $E_n(R_1^{(n)}, R_2^{(n)}) \to 0$ as $n \to \infty$. Our proof for the characterization of biclique rate region combines information-theoretic techniques and linear algebra; similar techniques were also used in [40] and [41].

We also characterize the asymptotic exponent of maximal density, and interpret it in terms of noninteractive simulation, for which the marginal probabilities are exponentially small. Note that this regime was first explicitly studied by Ordentlich et al. [17], who solved limiting cases for DSBSes. In fact, a complete characterization (involving time-sharing random variables) of this problem exists in the literature, which is a direct consequence of the existing information-theoretic characterization of Brascamp–Lieb inequalities. Applying this result to zero-error coding for the binary adder channel yields a new bound on the zero-error capacity.

Finally, we relax Boolean functions in noninteractive simulation problems to any nonnegative functions, but still restrict their suppports to be exponentially small. We obtain stronger (forward and reverse) Brascamp–Lieb and hypercontractivity inequalities, which, in asymptotic cases, reduce to the common ones when the exponents of the sizes of the supports are zero. (Note that these stronger inequalities can be also derived from the existing information-theoretic characterization of the classic Brascamp–Lieb inequalities.) Similar inequalities were previously derived by Polyanskiy and Samorodnitsky [42] and by Kirshner and Samorodnitsky [18] by different methods.

D. Notation

We write $:=$ and occasionally $=:$ for equality by definition. Throughout this paper, for two sequences of reals, we use $a_n \doteq b_n$ to denote $a_n = b_n 2^{o(n)}$. We use $\mathcal{C}(Q_X, Q_Y)$ to denote the set of couplings Q_{XY} with marginals Q_X, Q_Y. Given $Q_{X|UW}$ and $Q_{Y|VW}$, we use $\mathcal{C}(Q_{X|UW}, Q_{Y|VW})$ to denote the set of conditional couplings $Q_{XY|UVW}$ with conditional marginals $Q_{X|UW}, Q_{Y|VW}$. Note that, given $Q_{X|UW}$ and $Q_{Y|VW}$, for any $Q_{XY|UVW} \in \mathcal{C}(Q_{X|UW}, Q_{Y|VW})$ and any Q_{UVW}, the joint law $Q_{XYUVW} = Q_{XY|UVW}Q_{UVW}$ is such that $X \leftrightarrow (U,W) \leftrightarrow V$ and $Y \leftrightarrow (V,W) \leftrightarrow U$, where the notation $X \leftrightarrow Y \leftrightarrow Z$ for a triple of random variables (X,Y,Z) denotes that X and Z are conditionally independent given Y. For a length-n sequence $\mathbf{x}$, we use $T_{\mathbf{x}}$ to denote the type of $\mathbf{x}$. For an $m \times n$ matrix $\mathbf{B} = (b_{i,j})$ and two subsets $\mathcal{H} \subseteq [m], \mathcal{L} \subseteq [n]$, we use $\mathbf{B}_{\mathcal{H},\mathcal{L}}$ to denote $(b_{i,j})_{i\in\mathcal{H}, j\in\mathcal{L}}$, i.e., the submatrix of $\mathbf{B}$ consisting of the elements with indices in $\mathcal{H} \times \mathcal{L}$. For a length-$n$ vector or sequence $\mathbf{x}$ and a subset $\mathcal{J} \subseteq [n]$, $\mathbf{x}_{\mathcal{J}} := (x_j)_{j\in\mathcal{J}}$ is defined similarly. For a distribution P_X, we use P_X^n to denote the n-fold product of P_X. We will also use notations $H_Q(X)$ or $H(Q_X)$ to denote the entropy of $X \sim Q_X$. If the distribution is denoted by P_X, we sometimes write the entropy as $H(X)$ for brevity. We use $\mathrm{supp}(P_X)$ to denote the support of P_X. The logarithm $\log$ is taken to the base 2, and $\ln$ is taken to the natural base. Note that, as is the case for many other information-theoretic results, the results in this paper can be viewed as independent of the choice of the base of the logarithm as long as exponentiation is interpreted as being with respect to the same base.

For a joint distribution P_{XY} and for functions $f : \mathcal{X} \to [0,\infty)$ and $g : \mathcal{Y} \to [0,\infty)$, define their *inner product*

$$\langle f, g\rangle := \mathbb{E}[f(X)g(Y)] = \sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} P_{XY}(x,y)f(x)g(y). \tag{8}$$

The L^p*-norm* of f for $p \in [1,\infty)$ and the *pseudo* L^p*-norm* of f for $p \in (0,1)$ are defined as

$$\|f\|_p := \left(\mathbb{E}[f(X)^p]\right)^{1/p} = \left(\sum_{x\in\mathcal{X}} P_X(x)f(x)^p\right)^{1/p}. \tag{9}$$

II. Type Graphs

In this section, we completely characterize the asymptotic exponent of maximal density and the asymptotic biclique rate region.

A. Exponents

The asymptotic behavior of the exponent of maximal density is characterized in the following theorem, whose proof is provided in Appendix B. For all nonnegative pairs (R_1, R_2), define

$$F^*(R_1,R_2) := \max_{\substack{P_{XYW}: P_{XY}=T_{XY},\\ H(X|W)\le R_1, H(Y|W)\le R_2}} H(X,Y|W), \tag{10}$$

and

$$E^*(R_1,R_2) := R_1 + R_2 - F^*(R_1,R_2). \tag{11}$$

Theorem 1: Given a joint n-type T_{XY} with $n \ge 2(|\mathcal{X}||\mathcal{Y}|+2)|\mathcal{X}||\mathcal{Y}|$, for $(R_1,R_2) \in \mathcal{R}_X^{(n)} \times \mathcal{R}_Y^{(n)}$, we have

$$E^*(R_1,R_2) \le E_n(R_1,R_2) \le E^*(R_1,R_2) + \varepsilon_n, \tag{12}$$

where $\varepsilon_n := \frac{(|\mathcal{X}||\mathcal{Y}|+2)|\mathcal{X}||\mathcal{Y}|}{n} \log \frac{(n+1)n^6}{|\mathcal{X}|^4|\mathcal{Y}|^4}$. As a consequence, for any $n \ge 1$ and any joint n-type T_{XY}, we have

$$E(R_1,R_2) = E^*(R_1,R_2). \tag{13}$$

Without loss of optimality, the alphabet size of W in the definition of $F^*(R_1,R_2)$ can be assumed to be no larger than $|\mathcal{X}||\mathcal{Y}| + 2$.

Remark 1: $E^*(R_1,R_2)$ can be also expressed as

$$E^*(R_1,R_2) = R_1 + R_2 - H_T(XY) + G^*(R_1,R_2),$$

with

$$G^*(R_1,R_2) := \min_{\substack{P_{XYW}: P_{XY}=T_{XY},\\ H(X|W)\le R_1, H(Y|W)\le R_2}} I(X,Y;W) \tag{14}$$

corresponding to the minimum common rate given marginal rates (R_1,R_2) in the Gray–Wyner source coding network [43, Theorem 14.3].

Remark 2: The explicit expression of E^* for the doubly symmetric binary source was given in Section III-C.

Remark 3: A slightly weaker statement, $E_n(R_1,R_2) = E^*(R_1,R_2) + O(\frac{\log n}{n})$, can be recovered from a more general result given in [44, (6) and (7)] via the noninteractive simulation interpretation of the maximal density problem; see Section III-A.

Before proving Theorem 1, we first list several properties of $F^*(R_1,R_2)$ in the following lemma. The proof is provided in Appendix A.

Lemma 1: For any joint n-type T_{XY} and $R_1, R_2 \ge 0$, the following properties of $F^*(R_1,R_2)$ hold.

1) Given R_1, $F^*(R_1,R_2)$ is nondecreasing in R_2 and, given R_2, $F^*(R_1,R_2)$ is nondecreasing in R_1.

2) $F^*(R_1,R_2) \le \min\{H_T(X,Y), R_1 + R_2, R_1 + H_T(Y|X), R_2 + H_T(X|Y)\}$. Moreover, $F^*(H_T(X),H_T(Y)) = H_T(X,Y)$.

3) $F^*(0,R_2) = \min\{R_2, H_T(Y|X)\}$ and, similarly, $F^*(R_1,0) = \min\{R_1, H_T(X|Y)\}$.

4) $F^*(R_1,R_2)$ is concave in (R_1,R_2) on $\{(R_1,R_2) : R_1 \ge 0, R_2 \ge 0\}$.

5) For $\delta_1,\delta_2 \ge 0$, we have $0 \le F^*(R_1+\delta_1, R_2+\delta_2) - F^*(R_1,R_2) \le \delta_1 + \delta_2$ for all $R_1 \ge 0$, $R_2 \ge 0$.

Theorem 1 is an edge-isoperimetric result for the bipartite graph induced by a joint n-type T_{XY}. For the case in which

$\mathcal{X} = \mathcal{Y}$ and $T_X = T_Y$, the bipartite graph of T_{XY} can be replaced by a non-bipartite one. Consider a directed graph[7] (allowing self-loops if $X = Y$ under T_{XY}) in which the vertices consist of $\mathbf{x} \in \mathcal{T}_{T_X}$ and there is a directed edge from[8] $\mathbf{x}$ to $\mathbf{y}$ if and only if $(\mathbf{x}, \mathbf{y}) \in \mathcal{T}_{T_{XY}}$. Hence, for this case, Theorem 1 can be also considered as an edge-isoperimetric result for a directed graph induced by T_{XY}. Specifically, for a subset $\mathcal{A} \subseteq \mathcal{T}_{T_X}$, let $G[\mathcal{A}]$ be the induced subgraph of the directed graph of T_{XY}. The (edge) density $\rho(G[\mathcal{A}])$ is defined as

$$\rho(G[\mathcal{A}]) := \frac{\#\text{ of directed edges in } G[\mathcal{A}]}{|\mathcal{A}|^2} = \frac{|(\mathcal{A} \times \mathcal{A}) \cap \mathcal{T}_{T_{XY}}|}{|\mathcal{A}|^2}.$$

Given $1 \le M \le |\mathcal{T}_{T_X}|$, define the maximal density of subgraphs with size M as[9]

$$\Gamma_n(M) := \max_{\mathcal{A} \subseteq \mathcal{T}_{T_X} : |\mathcal{A}| = M} \rho(G[\mathcal{A}]).$$

Given a joint n-type T_{XY}, for $R \in \mathcal{R}_X^{(n)}$ as defined in (2), define the exponent of maximal density as

$$E_n(R) := -\frac{1}{n} \log \Gamma_n(2^{nR}). \quad (15)$$

For any subsets $\mathcal{A}, \mathcal{B}$ of $\mathcal{X}^n$, we have

$$|\mathcal{A}||\mathcal{B}|\rho(G[\mathcal{A}, \mathcal{B}]) \le |\mathcal{A} \cup \mathcal{B}|^2 \rho(G[\mathcal{A} \cup \mathcal{B}]).$$

On the other hand,

$$\Gamma_n(M) \le \Gamma_n(M, M).$$

Hence

$$\frac{1}{4}\Gamma_n\left(\frac{M}{2}, \frac{M}{2}\right) \le \Gamma_n(M) \le \Gamma_n(M, M).$$

Combining the inequalities above with Theorem 1 yields the following result.

Corollary 1: For any joint n-type T_{XY}, and $R \in \mathcal{R}_X^{(n)}$, we have

$$E_n(R) = E^*(R, R) + O\left(\frac{\log n}{n}\right), \quad (16)$$

where the asymptotic constant in the $O(\frac{\log n}{n})$ term on the right hand side depends only on $|\mathcal{X}|$, and $E^*(R_1, R_2)$ is defined in Theorem 1.

For the case of $\mathcal{X} = \mathcal{Y}$ and $T_X = T_Y$, the bipartite graph of T_{XY} can be also considered as an undirected graph (allowing self-loops if $X = Y$ under T_{XY}) in which the vertices consist of $\mathbf{x} \in \mathcal{T}_{T_X}$ and $(\mathbf{x}, \mathbf{y})$ is an edge if and only if $(\mathbf{x}, \mathbf{y})$ or $(\mathbf{y}, \mathbf{x}) \in \mathcal{T}_{T_{XY}}$. By a similar argument to the above, Corollary 1 still holds for this case, which hence can be considered as a generalization of [18, Theorem 1.6] from binary alphabets to arbitrary finite alphabets.

B. Biclique Rate Region

The asymptotic behavior of the biclique rate region is characterized in the following theorem, whose proof is provided in Appendix C. Define

$$\mathcal{R}^*(T_{XY}) := \bigcup_{\substack{0 \le \alpha \le 1, P_{XY}, Q_{XY}: \\ \alpha P_{XY} + (1-\alpha) Q_{XY} = T_{XY}}} \{(R_1, R_2) : R_1 \le \alpha H_P(X|Y), R_2 \le (1-\alpha) H_Q(Y|X)\}. \quad (17)$$

Theorem 2: For any $n \ge 8(|\mathcal{X}||\mathcal{Y}|)^{7/5}$ and any T_{XY},

$$\begin{aligned}&(\mathcal{R}^*(T_{XY}) - [0, \varepsilon_{1,n}] \times [0, \varepsilon_{2,n}]) \cap (\mathcal{R}_X^{(n)} \times \mathcal{R}_Y^{(n)}) \\ &\subseteq \mathcal{R}_n(T_{XY}) \\ &\subseteq \mathcal{R}^*(T_{XY}) \cap (\mathcal{R}_X^{(n)} \times \mathcal{R}_Y^{(n)})\end{aligned} \quad (18)$$

where $\mathcal{R}_n(T_{XY})$ is defined in (5), "$-$" is the Minkowski difference (i.e., for $\mathcal{A}, \mathcal{B} \subseteq \mathbb{R}^m$, $\mathcal{A} - \mathcal{B} := \bigcap_{b \in \mathcal{B}} (\mathcal{A} - b)$), $\varepsilon_{1,n} := \frac{|\mathcal{X}||\mathcal{Y}|}{n} \log \frac{n^4(n+1)}{16|\mathcal{X}|}$, and $\varepsilon_{2,n} := \frac{|\mathcal{X}||\mathcal{Y}|}{n} \log \frac{n^4(n+1)}{16|\mathcal{Y}|^2}$. In particular,

$$\mathcal{R}(T_{XY}) = \mathcal{R}^*(T_{XY}), \quad (19)$$

where $\mathcal{R}(T_{XY})$ is the asymptotic biclique rate region, defined in (7).

Remark 4: Theorem 2 can be easily generalized to the k-variables case with $k \ge 3$. For this case, let $T_{X_1,\ldots,X_k}$ be a joint n-type. Then the graph G induced by $T_{X_1,\ldots,X_k}$ is in fact a k-partite hypergraph. The (edge) density of the subgraph of G with vertex sets $(\mathcal{A}_1, \ldots, \mathcal{A}_k)$ is defined as

$$\rho(G[\mathcal{A}_1, \ldots, \mathcal{A}_k]) := \frac{|(\prod_{i=1}^k \mathcal{A}_i) \cap \mathcal{T}_{T_{X_1,\ldots,X_k}}|}{\prod_{i=1}^k |\mathcal{A}_i|}.$$

It is interesting to observe that $\rho(G) \doteq 2^{-nI_{T^{(n)}}(X_1;\ldots;X_k)}$ for a sequence of joint types $\{T_{X_1,\ldots,X_k}^{(n)}\}$, where $I_{T^{(n)}}(X_1; \ldots; X_k) := \sum_{i=1}^k H_{T^{(n)}}(X_i) - H_{T^{(n)}}(X_1, \ldots, X_k)$. Given a joint n-type $T_{X_1,\ldots,X_k}$, we define the k-clique rate region as

$$\mathcal{R}_n(T_{X_1,\ldots,X_k}) := \{(\frac{1}{n}\log|\mathcal{A}_1|, \ldots, \frac{1}{n}\log|\mathcal{A}_k|) : \rho(G[\mathcal{A}_1, \ldots, \mathcal{A}_k]) = 1\}.$$

Following similar steps to our proof of Theorem 2, for this case we have

$$\begin{aligned}&(\mathcal{R}^*(T_{X_1,\ldots,X_k}) - [0, O(\frac{\log n}{n})]^k) \cap (\prod_{i=1}^k \mathcal{R}_{X_i}^{(n)}) \\ &\subseteq \mathcal{R}_n(T_{X_1,\ldots,X_k}) \\ &\subseteq \mathcal{R}^*(T_{X_1,\ldots,X_k}) \cap (\prod_{i=1}^k \mathcal{R}_{X_i}^{(n)}),\end{aligned}$$

[7]When we extend the bipartite graph to a non-bipartite one, we assume the graph to be directed, in order to ensure that the pairs of sequences $(\mathbf{x}, \mathbf{y})$ and the edges in the graph are mapped to each other in a one-to-one way.

[8]Without of loss generality, we consider the edges from $\mathbf{x}$ to $\mathbf{y}$, since we can obtain a graph with edges from $\mathbf{y}$ to $\mathbf{x}$ if we consider the type T_{YX} (instead of T_{XY}).

[9]We use the same notation as the one in (1) for the bipartite graph case, but here the edge density has only one parameter. The difference between these two maximal densities is that in (1) the maximization is taken over a pair of sets $(\mathcal{A}, \mathcal{B})$, but here only over one set (equivalently, under the restriction $\mathcal{A} = \mathcal{B}$).

where

$$\mathcal{R}^*(T_{X_1,\ldots,X_k}) := \bigcup_{\substack{\alpha_i\geq 0, P^{(i)}_{X_1,\ldots,X_k}, i\in[k]: \sum_{i=1}^k \alpha_i = 1 \\ \sum_{i=1}^k \alpha_i P^{(i)}_{X_1,\ldots,X_k} = T_{X_1,\ldots,X_k}}} \{(R_1,\ldots,R_k) : R_i \leq \alpha_i H_{P^{(i)}}(X_i|X^{\setminus i})\},$$

with $X^{\setminus i} := (X_1,\ldots,X_{i-1},X_{i+1},\ldots,X_k)$.

Proposition 1: Given T_{XY}, $\mathcal{R}^*(T_{XY})$ is a closed convex set.

Proof: Using the continuity of $H_P(X|Y)$ in P_{XY} it can be established that $\mathcal{R}^*(T_{XY})$ is closed. Convexity follows by the following argument. For any $(R_1,R_2), (\hat{R}_1,\hat{R}_2) \in \mathcal{R}^*(T_{XY})$, there exist (α, P_{XY}, Q_{XY}) and $(\hat{\alpha}, \hat{P}_{XY}, \hat{Q}_{XY})$ such that

$$\alpha P_{XY} + (1-\alpha)Q_{XY} = T_{XY},$$
$$\hat{\alpha}\hat{P}_{XY} + (1-\hat{\alpha})\hat{Q}_{XY} = T_{XY},$$
$$R_1 \leq \alpha H_P(X|Y), R_2 \leq (1-\alpha)H_Q(Y|X),$$
$$\hat{R}_1 \leq \hat{\alpha} H_{\hat{P}}(X|Y), \hat{R}_2 \leq (1-\hat{\alpha})H_{\hat{Q}}(Y|X).$$

Then for any $\lambda \in [0,1]$,

$$\lambda R_1 + (1-\lambda)\hat{R}_1$$
$$\leq \lambda\alpha H_P(X|Y) + (1-\lambda)\hat{\alpha} H_{\hat{P}}(X|Y) \quad (20)$$
$$\leq \beta H_{P^{(\theta)}}(X|Y), \quad (21)$$

where $\beta = \lambda\alpha + (1-\lambda)\hat{\alpha}$, and $P^{(\theta)}_{XY} := \theta P_{XY} + (1-\theta)\hat{P}_{XY}$ with $\theta = \frac{\lambda\alpha}{\lambda\alpha+(1-\lambda)\hat{\alpha}}$ if $\beta > 0$; $P^{(\theta)}_{XY}$ is chosen as an arbitrary distribution if $\beta = 0$. Here (21) follows since $H_P(X|Y)$ is concave in P_{XY}. By symmetry, $\lambda R_2 + (1-\lambda)\hat{R}_2 \leq (1-\beta)H_{Q^{(\hat{\theta})}}(Y|X)$, where $Q^{(\hat{\theta})}_{XY} := \hat{\theta}P_{XY} + (1-\hat{\theta})\hat{P}_{XY}$ with $\hat{\theta} = \frac{\lambda(1-\alpha)}{\lambda(1-\alpha)+(1-\lambda)(1-\hat{\alpha})}$ if $\beta < 1$; $Q^{(\hat{\theta})}_{XY}$ is chosen as an arbitrary distribution if $\beta = 1$. Since $\beta P^{(\theta)}_{XY} + (1-\beta)Q^{(\hat{\theta})}_{XY} = T_{XY}$, it follows that $\lambda(R_1,R_2) + (1-\lambda)(\hat{R}_1,\hat{R}_2) \in \mathcal{R}^*(T_{XY})$, i.e., $\mathcal{R}^*(T_{XY})$ is convex. ■

Since $\mathcal{R}^*(T_{XY})$ is convex, an extreme case of $\mathcal{R}^*(T_{XY})$ is a triangle region. We next study when the asymptotic biclique rate region is a triangle region. We obtain the following necessary and sufficient condition. The proof is provided in Appendix E.

Proposition 2: Let T_{XY} be a joint n-type such that $H_T(X|Y), H_T(Y|X) > 0$. Then the asymptotic biclique rate region $\mathcal{R}(T_{XY})$ is a triangle region, i.e.,

$$\mathcal{R}(T_{XY}) = \mathcal{R}_\triangle(T_{XY}) := \bigcup_{0\leq\alpha\leq 1} \{(R_1,R_2) : R_1 \leq \alpha H_T(X|Y), R_2 \leq (1-\alpha)H_T(Y|X)\},$$

if and only if T_{XY} satisfies that $T_{X|Y}(x|y)^{1/H_T(X|Y)} = T_{Y|X}(y|x)^{1/H_T(Y|X)}$ for all x,y.

The condition in Proposition 2 is satisfied by the joint n-types T_{XY} which have marginals $T_X = \mathrm{Unif}(\mathcal{X}), T_Y = \mathrm{Unif}(\mathcal{Y})$ and satisfy at least one of the following two conditions:

1) $|\mathcal{X}| = |\mathcal{Y}|$;
2) X, Y are independent under the distribution T_{XY}.

TABLE I
THE DISTRIBUTION OF A DSBS WITH CORRELATIONCOEFFICIENT ρ

$X\backslash Y$	0	1
0	$\frac{1+\rho}{4}$	$\frac{1-\rho}{4}$
1	$\frac{1-\rho}{4}$	$\frac{1+\rho}{4}$

Example (DSBS): A typical example that satisfies these conditions is the DSBS, whose distribution is given in Table I. Hence, the asymptotic biclique rate region is the triangle region $\{(R_1,R_2) : R_1 + R_2 \leq h(\frac{1-\rho}{2})\}$ if the joint n-type T_{XY} is a DSBS with correlation coefficient $\rho \in [0,1]$. Here, $h : t \mapsto -t\log t - (1-t)\log(1-t)$ denotes the binary entropy function.

For a joint type T_{XY} the Han and Kobayashi region (which is of course defined for any joint distribution, not necessarily a type) is given by [8]

$$\mathcal{R}^{**}(T_{XY}) := \bigcup_{\substack{P_{XYW}: P_{XY} = T_{XY}, \\ X \leftrightarrow W \leftrightarrow Y}} \{(R_1,R_2) : R_1 \leq H(X|W), R_2 \leq H(Y|W)\}. \quad (22)$$

By Theorem 1, $\mathcal{R}^{**}(T_{XY})$ also coincides with the region $\{(R_1,R_2) : E^*(R_1,R_2) = 0\}$. This implies that $\mathcal{R}^*(T_{XY}) \subseteq \mathcal{R}^{**}(T_{XY})$, i.e., that for any joint type T_{XY} the asymptotic biclique rate region defined in this paper is a subset of Han and Kobayashi's approximate version. This can also be seen by directly comparing the definition of $\mathcal{R}^*(T_{XY})$ in (17) to the definition of $\mathcal{R}^{**}(T_{XY})$ in (22). This can be seen as follows. For a joint type T_{XY}, let $Q^{(0)}_{XY}$, $Q^{(1)}_{XY}$, and $0 \leq \alpha \leq 1$ be such that $\alpha Q^{(0)}_{XY} + (1-\alpha)Q^{(1)}_{XY} = T_{XY}$, and let (R_1,R_2) be a rate pair such that $R_1 \leq \alpha H_{Q^{(0)}}(X|Y)$ and $R_2 \leq (1-\alpha)H_{Q^{(1)}}(Y|X)$. Without loss of generality, we assume that $\mathcal{X}$ and $\mathcal{Y}$ are disjoint, i.e., $\mathcal{X} \cap \mathcal{Y} = \emptyset$, since otherwise, we can respectively map them to another two sets satisfying this requirement by bijections. Let (X,Y,W) be a tuple of random variables such that W takes values in $\mathcal{X} \cup \mathcal{Y}$, and $W \in \mathcal{Y}$ with probability α and $W \in \mathcal{X}$ with probability $1-\alpha$. Moreover, under the condition $W \in \mathcal{Y}$, it holds that $W = Y$ and $(X,Y) \sim Q^{(0)}_{XY}$; under the condition $W \in \mathcal{X}$, it holds that $W = X$ and $(X,Y) \sim Q^{(1)}_{XY}$. It can be checked that $(X,Y) \sim T_{XY}$ and we have $X \leftrightarrow W \leftrightarrow Y$, $R_1 \leq H(X|W)$, and $R_2 \leq H(Y|W)$. This inclusion can be strict. For example, when the joint n-type T_{XY} is a DSBS with a positive crossover probability, the region $\mathcal{R}^{**}(T_{XY})$, which is computed in [8, Section 4], strictly contains the asymptotic biclique region $\mathcal{R}^*(T_{XY})$, which, by Proposition 2, is a triangle region. Another family of examples where the asymptotic biclique region is strictly contained in the region of Han and Kobayashi is when the joint n-type T_{XY} is $\mathrm{Unif}(\mathcal{X}\times\mathcal{Y})$. Here $\mathcal{R}^{**}(T_{XY})$ equals the rectangle region $[0,H(X)]\times[0,H(Y)]$, while Proposition 2 implies that $\mathcal{R}^{*}(T_{XY})$ is a triangle region.

The difference between the exact and approximate versions of asymptotic biclique rate regions is caused by the "type overflow" effect, which was crystallized by the first author and Tan in [45]. Let (R_1,R_2) be a pair such that $E^*(R_1,R_2)=0$. Let $(\mathcal{A},\mathcal{B})$ be an optimal pair of subsets attaining $E^*(R_1,R_2)$. All the sequences in $\mathcal{A}$ have type T_X, and all the sequences in $\mathcal{B}$ have type T_Y. However, in general, the joint types of $(\mathbf{x},\mathbf{y})\in\mathcal{A}\times\mathcal{B}$ might "overflow" from the target joint type T_{XY}. The number of non-overflowed sequence pairs (i.e., $|(\mathcal{A}\times\mathcal{B})\cap\mathcal{T}_{T_{XY}}|$) has exponent R_1+R_2, since $E^*(R_1,R_2)=0$. This means that not too many sequence pairs have overflowed. However, if type overflow is forbidden, then we must reduce the rates of $\mathcal{A}$ and $\mathcal{B}$ to satisfy this requirement. This leads to the exact version of the asymptotic biclique rate region being strictly smaller than the approximate version. In other words, the exact asymptotic biclique rate region is more sensitive to the type overflow effect than the approximate version. A similar conclusion was previously drawn by the first author and Tan in [45] for the common information problem. Technically speaking, the type overflow effect corresponds to the fact that optimization over couplings is involved in our expressions. Intuitively, it is caused by the Markov chain constraints in the problem. We believe that the type overflow effect usually accompanies problems involving Markov chains.

III. Noninteractive Simulation

In this section, we connect the maximal density problem on type graphs to the noninteractive simulation (or noise stability) problem. We focus on two noninteractive simulation problems, one with sources uniformly distributed over a joint n-type and the other with memoryless sources.

A. Sources $\mathrm{Unif}(\mathcal{T}_{T_{XY}})$

In this subsection, we assume $(\mathbf{X},\mathbf{Y})\sim P_{\mathbf{X},\mathbf{Y}}:=\mathrm{Unif}(\mathcal{T}_{T_{XY}})$. Given two marginal probabilities $P_{\mathbf{X}}(\mathcal{A})$ and $P_{\mathbf{Y}}(\mathcal{B})$, what are the possible maximal and minimal values of the joint probability $P_{\mathbf{X},\mathbf{Y}}(\mathcal{A}\times\mathcal{B})$? This problem is termed the noninteractive binary simulation problem or the (two-set version of) noise stability problem.

Define $\mathcal{E}_X^{(n)}:=\{\frac{1}{n}\log|\mathcal{T}_{T_X}|-R_1 : R_1\in\mathcal{R}_X^{(n)}\}$ and $\mathcal{E}_Y^{(n)}:=\{\frac{1}{n}\log|\mathcal{T}_{T_Y}|-R_2 : R_2\in\mathcal{R}_Y^{(n)}\}$, where $\mathcal{R}_X^{(n)}$ and $\mathcal{R}_Y^{(n)}$ are defined in (2). Given a joint n-type T_{XY}, for $(E_1,E_2)\in\mathcal{E}_X^{(n)}\times\mathcal{E}_Y^{(n)}$, define the exponents of the maximal and minimal noise stability as

$$\underline{\Upsilon}_n(E_1,E_2):=-\frac{1}{n}\log\max_{\substack{\mathcal{A}\subseteq\mathcal{T}_{T_X},\mathcal{B}\subseteq\mathcal{T}_{T_Y}:\\ P_{\mathbf{X}}(\mathcal{A})=2^{-nE_1},\\ P_{\mathbf{Y}}(\mathcal{B})=2^{-nE_2}}}P_{\mathbf{X},\mathbf{Y}}(\mathcal{A}\times\mathcal{B}),\quad(23)$$

$$\overline{\Upsilon}_n(E_1,E_2):=-\frac{1}{n}\log\min_{\substack{\mathcal{A}\subseteq\mathcal{T}_{T_X},\mathcal{B}\subseteq\mathcal{T}_{T_Y}:\\ P_{\mathbf{X}}(\mathcal{A})=2^{-nE_1},\\ P_{\mathbf{Y}}(\mathcal{B})=2^{-nE_2}}}P_{\mathbf{X},\mathbf{Y}}(\mathcal{A}\times\mathcal{B}).\quad(24)$$

The noninteractive binary simulation problem is to determine these two quantities. This problem originates from Gács and Körner's and Witsenhausen's seminal works [11], [12] in the study of the Gács–Körner–Witsenhausen common information. This topic has also attracted independent interest from the computer science community, due to the connection with the analysis of Boolean functions [46]. We refer readers to the related works mentioned in Section I-B or the monograph [22] for more information on this problem.

We determine the asymptotic behavior of $\underline{\Upsilon}_n$ in the following theorem. However, the asymptotic behavior of $\overline{\Upsilon}_n$ is currently unclear; see the discussion in Section V. For $0\le s\le H(X), 0\le t\le H(Y)$, define

$$\underline{\Upsilon}^*(s,t):=\min_{\substack{P_{XYW}:P_{XY}=T_{XY},\\ I(X;W)\ge s,I(Y;W)\ge t}}I(XY;W).$$

Theorem 3: For any T_{XY} and $(E_1,E_2)\in\mathcal{E}_X^{(n)}\times\mathcal{E}_Y^{(n)}$, we have

$$\underline{\Upsilon}_n(E_1,E_2)=\underline{\Upsilon}^*(E_1,E_2)+O(\frac{\log n}{n}),\quad(25)$$

where the asymptotic constant in the $O(\frac{\log n}{n})$ bound depends only on $|\mathcal{X}|,|\mathcal{Y}|$.

In fact, this result can be recovered from a more general result given in [44, (6) and (7)]. The latter generalizes the $\underline{\Upsilon}_n$ for the uniform distribution over a type class to the infimum of $\underline{\Upsilon}_n$ over all distributions not too far from a product distribution.

Proof: Observe that

$$P_{\mathbf{X}}(\mathcal{A})=|\mathcal{A}|/|\mathcal{T}_{T_X}|,\quad(26)$$
$$P_{\mathbf{Y}}(\mathcal{B})=|\mathcal{B}|/|\mathcal{T}_{T_Y}|,\quad(27)$$
$$P_{\mathbf{X},\mathbf{Y}}(\mathcal{A}\times\mathcal{B})=\rho(G[\mathcal{A},\mathcal{B}])|\mathcal{A}||\mathcal{B}|/|\mathcal{T}_{T_{XY}}|.\quad(28)$$

So, Theorem 3 is implied by Theorem 1. ∎

B. Sources P_{XY}^n

In this subsection, we consider the noninteractive simulation problem with $(\mathbf{X},\mathbf{Y})\sim P_{XY}^n$, where P_{XY} is a joint distribution defined on $\mathcal{X}\times\mathcal{Y}$. We still assume that $\mathcal{X},\mathcal{Y}$ are finite sets of cardinality at least 2 and that $P_X(x)>0$, $P_Y(y)>0$ for all $(x,y)\in\mathcal{X}\times\mathcal{Y}$, where P_X,P_Y denote the marginal distributions of P_{XY}. Ordentlich et al. [17] focused on binary symmetric distributions P_{XY}, and studied the exponent of $P_{XY}^n(\mathcal{A}\times\mathcal{B})$ given that $P_X^n(\mathcal{A}),P_Y^n(\mathcal{B})$ vanish exponentially fast with exponents E_1,E_2, respectively. Let

$$E_{1,\max}:=-\log P_{X,\min},\quad E_{2,\max}:=-\log P_{Y,\min},$$

where $P_{X,\min}:=\min_x P_X(x)$, $P_{Y,\min}:=\min_y P_Y(y)$. In this subsection, we consider an arbitrary distribution $P_{X,Y}$ satisfying $P_X(x)>0$, $P_Y(y)>0$ for all $(x,y)\in\mathcal{X}\times\mathcal{Y}$ and, for $E_1\in[0,E_{1,\max}]$, $E_2\in[0,E_{2,\max}]$, we aim at characterizing[10]

$$\underline{\Theta}(E_1,E_2):=\lim_{n\to\infty}\underline{\Theta}_n(E_1,E_2),\quad(29)$$

[10] By time-sharing arguments, given (E_1,E_2), $\{n\underline{\Theta}_n(E_1,E_2)\}_{n\ge1}$ is subadditive. Hence, by Fekete's Subadditive Lemma, the first limit in (29) exists and equals $\inf_{n\ge1}\underline{\Theta}_n(E_1,E_2)$. Similar observations serve to define the second limit in (29).

$$\overline{\Theta}(E_1,E_2) := \lim_{n\to\infty} \overline{\Theta}_n(E_1,E_2), \quad (30)$$

where the exponents of the maximal and minimal noise stability are defined by

$$\underline{\Theta}_n(E_1,E_2) := -\frac{1}{n}\log \max_{\substack{\mathcal{A}\subseteq\mathcal{X}^n,\mathcal{B}\subseteq\mathcal{Y}^n:\\ P_X^n(\mathcal{A})\le 2^{-nE_1},\\ P_Y^n(\mathcal{B})\le 2^{-nE_2}}} P_{XY}^n(\mathcal{A}\times\mathcal{B}), \quad (31)$$

$$\overline{\Theta}_n(E_1,E_2) := -\frac{1}{n}\log \min_{\substack{\mathcal{A}\subseteq\mathcal{X}^n,\mathcal{B}\subseteq\mathcal{Y}^n:\\ P_X^n(\mathcal{A})\ge 2^{-nE_1},\\ P_Y^n(\mathcal{B})\ge 2^{-nE_2}}} P_{XY}^n(\mathcal{A}\times\mathcal{B}). \quad (32)$$

For $E_1\in[0,E_{1,\max}], E_2\in[0,E_{2,\max}]$, define

$$\underline{\Theta}^*(E_1,E_2) \quad (33)$$
$$:= \min_{\substack{Q_{XYW}:D(Q_{X|W}\|P_X|Q_W)\ge E_1,\\ D(Q_{Y|W}\|P_Y|Q_W)\ge E_2}} D(Q_{XY|W}\|P_{XY}|Q_W) \quad (34)$$

$$= \min_{\substack{Q_W,Q_{X|W},Q_{Y|W}:\\ D(Q_{X|W}\|P_X|Q_W)\ge E_1,\\ D(Q_{Y|W}\|P_Y|Q_W)\ge E_2}} \mathsf{D}(Q_{X|W},Q_{Y|W}\|P_{XY}|Q_W), \quad (35)$$

$$\overline{\Theta}^*(E_1,E_2) \quad (36)$$
$$:= \max_{\substack{Q_W,Q_{X|W},Q_{Y|W}:\\ D(Q_{X|W}\|P_X|Q_W)\le E_1,\\ D(Q_{Y|W}\|P_Y|Q_W)\le E_2}} \mathsf{D}(Q_{X|W},Q_{Y|W}\|P_{XY}|Q_W), \quad (37)$$

where

$$\mathsf{D}(Q_{X|W},Q_{Y|W}\|P_{XY}|Q_W) := \min_{Q_{XY|W}\in\mathcal{C}(Q_{X|W},Q_{Y|W})} D(Q_{XY|W}\|P_{XY}|Q_W)$$

and the notation $\mathcal{C}(Q_{X|W},Q_{Y|W})$ is defined in Subsection I-D. Without loss of optimality, the alphabet size of W in either (33) or (36) can be assumed to be no larger than 3. This is a consequence of the support lemma in [43].

The asymptotic exponents $\underline{\Theta}$ and $\overline{\Theta}$ are characterized in the following theorem. Note that the following theorem is not new, since it is a direct consequence of the information-theoretic characterization of Brascamp–Lieb inequalities given in [29], [33], [34], and [47] for the forward part and [35], [36], [37], [39], [47] for the reverse part. For more details see [22, Section 10.3].

Theorem 4 (Strong Small-Set Expansion Theorem): For $E_1\in[0,E_{1,\max}], E_2\in[0,E_{2,\max}]$, the following hold.

1) $\underline{\Theta}(E_1,E_2) = \underline{\Theta}^*(E_1,E_2)$. Moreover, $\underline{\Theta}_n(E_1,E_2)\ge\underline{\Theta}^*(E_1,E_2)$ for any $n\ge1$.
2)

$$\overline{\Theta}(E_1,E_2) = \overline{\Theta}^{**}(E_1,E_2) := \begin{cases}\overline{\Theta}^*(E_1,E_2), & E_1,E_2>0,\\ E_1, & E_2=0,\\ E_2, & E_1=0.\end{cases}$$

Moreover, $\overline{\Theta}_n(E_1,E_2)\le\overline{\Theta}^{**}(E_1,E_2)$ for any $n\ge1$.

Remark 5: We interpret $\overline{\Theta}^*(E_1,E_2)$ as ∞ if $P_{XY}(x,y)=0$ for some (x,y). It can be checked that this interpretation is consistent with the definition in (33) because Q_W, $Q_{X|W}$, and $Q_{Y|W}$ in the outer maximum can be chosen so that $Q_W(w)>0$, $Q_{X|W}(x|w)>0$, and $Q_{Y|W}(y|w)>0$ for some w.

Remark 6: By the convexity of $\underline{\Theta}^*$ and concavity of $\overline{\Theta}^*$, Theorem 4 implies $\underline{\Theta}_n(E_1,E_2)\ge\lim_{t\downarrow0}\frac{1}{t}\underline{\Theta}^*(tE_1,tE_2)$ and $\overline{\Theta}_n(E_1,E_2)\le\lim_{t\downarrow0}\frac{1}{t}\overline{\Theta}^{**}(tE_1,tE_2)$. In particular, for the DSBS with correlation coefficient $\rho>0$, these inequalities reduce to that

$$\underline{\Theta}_n(E_1,E_2)\ge\begin{cases}\frac{E_1+E_2-2\rho\sqrt{E_1E_2}}{1-\rho^2}, & \rho^2E_1\le E_2\le E_1/\rho^2,\\ E_1, & E_2<\rho^2E_1,\\ E_2, & E_2>E_1/\rho^2,\end{cases} \quad (38)$$

$$\overline{\Theta}_n(E_1,E_2)\le\frac{E_1+E_2+2\rho\sqrt{E_1E_2}}{1-\rho^2}. \quad (39)$$

These inequalities correspond to the small-set expansion theorem given in [29, Lemma 1] [13, Lemma 3.4] [14, Theorem 3.4] [15, Generalized Small-Set Expansion Theorem on p. 285].

Remark 7: We must have $\liminf_{n\to\infty}-\frac{1}{n}\log P_X^n(\mathcal{A}_n)\ge E_1$ for any sequence $(\mathcal{A}_n,\mathcal{B}_n)$ attaining the asymptotic exponent $\underline{\Theta}(E_1,E_2)$. If $\limsup_{n\to\infty}-\frac{1}{n}\log P_X^n(\mathcal{A}_n)>E_1$ then it must be the case that $E_1<E_{1,\max}$. So in this case, we can add sequences in $\mathcal{X}^n$ to $\mathcal{A}_n$ to get a new sequence $(\mathcal{A}_n,\mathcal{B}_n)$ such that $\lim_{n\to\infty}-\frac{1}{n}\log P_X^n(\mathcal{A}_n)=E_1$. This is possible since for each $E_1\in[0,E_{1,\max})$ there is a sequence $\tilde{\mathcal{A}}_n\subseteq\mathcal{X}^n$ such that (1) each $\tilde{\mathcal{A}}_n$ is a type class (the type can change with n), (2) $P_X^n(\tilde{\mathcal{A}}_n)\le2^{-nE_1}$, (3) $\lim_{n\to\infty}|\tilde{\mathcal{A}}_n|=\infty$, and (4) $\lim_{n\to\infty}-\frac{1}{n}\log P_X^n(\tilde{\mathcal{A}}_n)=E_1$. We can then simply replace $\mathcal{A}_n$ by $\mathcal{A}_n\cup\hat{\mathcal{A}}_n$ where $\hat{\mathcal{A}}_n$ is a maximal subset of $\tilde{\mathcal{A}}_n$ among those that continue to satisfy $P_X^n(\mathcal{A}_n\cup\hat{\mathcal{A}}_n)\le2^{-nE_1}$. Now, the resulting new sequence $(\mathcal{A}_n,\mathcal{B}_n)$ continues to attain the asymptotic exponent $\underline{\Theta}(E_1,E_2)$ (by the converse part in the theorem above). Similarly, if needed, we can also add sequences to $\mathcal{B}_n$ such that $\lim_{n\to\infty}-\frac{1}{n}\log P_Y^n(\mathcal{B}_n)=E_2$. This implies that there exists a sequence $(\mathcal{A}_n,\mathcal{B}_n)$ such that $\lim_{n\to\infty}-\frac{1}{n}\log P_X^n(\mathcal{A}_n)=E_1,\lim_{n\to\infty}-\frac{1}{n}\log P_Y^n(\mathcal{B}_n)=E_2$, and $\lim_{n\to\infty}-\frac{1}{n}\log P_{XY}^n(\mathcal{A}_n\times\mathcal{B}_n)=\underline{\Theta}(E_1,E_2)$.

Remark 8: We define the effective region of $\overline{\Theta}^*$ as the set of (E_1,E_2) for which $\overline{\Theta}^*(E_1,E_2)=\hat{\psi}(E_1,E_2)$, i.e., there exists an optimal tuple $(Q_W,Q_{X|W},Q_{Y|W})$ attaining the maximum in the definition of $\overline{\Theta}^*(E_1,E_2)$ such that $D(Q_{X|W}\|P_X|Q_W)=E_1,D(Q_{Y|W}\|P_Y|Q_W)=E_2$. Note that $D(Q_{X|W}\|P_X|Q_W)$ is the asymptotic exponent of the probability of a conditional type class with type $Q_{X|W}$ given Q_W, and $\mathsf{D}(Q_{X|W},Q_{Y|W}\|P_{XY}|Q_W)$ is the asymptotic exponent of the probability of $\mathcal{T}_{Q_{X|W}}(\mathbf{w})\times\mathcal{T}_{Q_{Y|W}}(\mathbf{w})$ with $\mathbf{w}$ having type Q_W. Hence, for (E_1,E_2) in the effective region of $\overline{\Theta}^*$, there exists a sequence of $(\mathcal{A}_n,\mathcal{B}_n)$ such that $\lim_{n\to\infty}-\frac{1}{n}\log P_X^n(\mathcal{A}_n)=E_1,\lim_{n\to\infty}-\frac{1}{n}\log P_Y^n(\mathcal{B}_n)=E_2$, and $\lim_{n\to\infty}-\frac{1}{n}\log P_{XY}^n(\mathcal{A}_n\times\mathcal{B}_n)=\overline{\Theta}^*(E_1,E_2)$. The strong small-set expansion theorem is improved in [39] by proving asymptotically sharp bounds for equality constraints in (31) and (32).

Remark 9: The discrepancy between the noninteractive simulation problem with a uniform source defined on a joint type and the one with a product source was noted in [44], and similar discrepancies were also noted and exploited in some classic works on strong converses in network information theory, e.g., [6] and [7]. Specifically, for any joint distribution P_{XY} and $0 \le E_1 \le H(X), 0 \le E_2 \le H(Y)$, we have $\underline{\Upsilon}^*(E_1, E_2) \ge \underline{\Theta}^*(E_1, E_2)$ (the inequality is strict in general), where $\underline{\Upsilon}^*$ and $\underline{\Theta}^*$ are both defined for P_{XY}. This observation follows since for any distribution Q_{XYW} with marginal $Q_{XY} = P_{XY}$, it holds that $I_Q(XY;W) = D(Q_{XY|W}\|P_{XY}|Q_W) = \mathbb{E}_{Q_W} D(Q_{XY|W}\|P_{XY})$. Similar equalities hold for $I_Q(X;W)$ and $I_Q(Y;W)$. If we drop the condition $Q_{XY} = P_{XY}$ in the definition of $\underline{\Upsilon}^*$, then we obtain $\underline{\Theta}^*$. So, we have the observation. Another observation is $\lim_{t\downarrow 0} \frac{1}{t}\underline{\Upsilon}^*(tE_1, tE_2) = \lim_{t\downarrow 0} \frac{1}{t}\underline{\Theta}^*(tE_1, tE_2)$. This follows by the two equivalent information-theoretic characterizations of hypercontractivity inequalities: one expressed in terms of relative entropies and the other expressed in terms of mutual information [34].

Since the alphabet size of W in the definition of $\underline{\Theta}^*(E_1, E_2)$ or $\overline{\Theta}^*(E_1, E_2)$ can be taken to be at most 3, both $\underline{\Theta}(E_1, E_2)$ and $\overline{\Theta}(E_1, E_2)$ are achieved by a sequence of the time-sharing of at most *three* type codes (or equivalently, a conditional type class with conditional random variable W taking at most three values). Here a type code refers to a code of the form $(\mathcal{A}, \mathcal{B}) := (\mathcal{T}_{T_X}, \mathcal{T}_{T_Y})$ for a pair of types (T_X, T_Y).

Define the optimal transport divergence[11] (or the minimum relative entropy) of (Q_X, Q_Y) with respect to P_{XY}, as

$$\mathsf{D}(Q_X, Q_Y\|P_{XY}) := \min_{Q_{XY}\in\mathcal{C}(Q_X,Q_Y)} D(Q_{XY}\|P_{XY}).$$

For $s \in [0, E_{1,\max}], t \in [0, E_{2,\max}]$, define

$$\varphi(s,t) := \min_{\substack{Q_{XY}: D(Q_X\|P_X)=s,\\ D(Q_Y\|P_Y)=t}} D(Q_{XY}\|P_{XY}) \tag{40}$$

$$= \min_{\substack{Q_X,Q_Y: D(Q_X\|P_X)=s,\\ D(Q_Y\|P_Y)=t}} \mathsf{D}(Q_X, Q_Y\|P_{XY}), \tag{41}$$

and

$$\psi(s,t) := \max_{\substack{Q_X,Q_Y: D(Q_X\|P_X)=s,\\ D(Q_Y\|P_Y)=t}} \mathsf{D}(Q_X, Q_Y\|P_{XY}). \tag{42}$$

Define $\breve{f}$ as the lower convex envelope of a function f, and $\hat{f}$ as its upper concave envelope. Then, by definition, we have

$$\underline{\Theta}^*(E_1, E_2) = \min_{\substack{(q_i,s_i,t_i)_{i\in[3]}:\\ \sum_i q_i=1,\, q_i\ge 0,\forall i\in[3]\\ \sum_i q_i s_i\ge E_1, \sum_i q_i t_i\ge E_2}} \sum_i q_i\varphi(s_i,t_i) \tag{43}$$

$$= \min_{s\ge E_1, t\ge E_2} \min_{\substack{(q_i,s_i,t_i)_{i\in[3]}:\\ \sum_i q_i=1,\, q_i\ge 0,\forall i\in[3]\\ \sum_i q_i s_i=s, \sum_i q_i t_i=t}} \sum_i q_i\varphi(s_i,t_i)$$

$$= \min_{s\ge E_1, t\ge E_2} \breve{\varphi}(s,t), \tag{44}$$

and similarly,

$$\overline{\Theta}^*(E_1, E_2) = \max_{s\le E_1, t\le E_2} \hat{\psi}(s,t). \tag{45}$$

[11]The reason for this name is due to its resemblance to the optimal transport cost [48]. In the latter, the objective function is the expected cost, instead of the relative entropy.

Hence $\underline{\Theta}^*(E_1, E_2)$ is convex in (E_1, E_2), and $\overline{\Theta}^*(E_1, E_2)$ is concave in (E_1, E_2). By definition, it holds that for $s \in [0, E_{1,\max}], t \in [0, E_{2,\max}]$,

$$\underline{\Theta}^*(s,t) \le \breve{\varphi}(s,t) \le \varphi(s,t) \le \psi(s,t) \le \hat{\psi}(s,t) \le \overline{\Theta}^*(s,t). \tag{46}$$

Proposition 3: Both $\underline{\Theta}^*(E_1, E_2)$ and $\overline{\Theta}^*(E_1, E_2)$ are continuous over $E_1 \in [0, E_{1,\max}], E_2 \in [0, E_{2,\max}]$.

Proof: By the convexity and concavity, $\underline{\Theta}^*(E_1, E_2)$ and $\overline{\Theta}^*(E_1, E_2)$ are continuous over $E_1 \in (0, E_{1,\max}), E_2 \in (0, E_{2,\max})$. On the boundary, the continuity of these two functions follows by the continuity of the constraint functions and the continuity of the objective function, i.e., the continuity of $D(Q_{X|W}\|P_X|Q_W), D(Q_{Y|W}\|P_Y|Q_W)$ and $\mathsf{D}(Q_{X|W}, Q_{Y|W}\|P_{XY}|Q_W)$ in $(Q_W, Q_{X|W}, Q_{Y|W})$.

The continuity of $D(Q_{X|W}\|P_X|Q_W), D(Q_{Y|W}\|P_Y|Q_W)$ in $(Q_W, Q_{X|W}, Q_{Y|W})$ is obvious, since as assumed, P_X and P_Y have full support. We claim that

$$f(Q_W, Q_{X|W}, Q_{Y|W}) := \mathsf{D}(Q_{X|W}, Q_{Y|W}\|P_{XY}|Q_W) \tag{47}$$

is continuous in $(Q_W, Q_{X|W}, Q_{Y|W})$, which follows by the following lemma.

Lemma 2 [49, Lemma 13]: Let P_X, Q_X be distributions on $\mathcal{X}$, and P_Y, Q_Y distributions on $\mathcal{Y}$. Then for any $Q_{XY} \in \mathcal{C}(Q_X, Q_Y)$, there exists $P_{XY} \in \mathcal{C}(P_X, P_Y)$ such that

$$\|P_{XY} - Q_{XY}\| \le \|P_X - Q_X\| + \|P_Y - Q_Y\|, \tag{48}$$

where $\|P - Q\| := \sup_{\mathcal{A}} P(\mathcal{A}) - Q(\mathcal{A})$ denotes the total variation distance between P and Q.

By Lemma 2, given $(Q_W, Q_{X|W}, Q_{Y|W}, P_W, P_{X|W}, P_{Y|W})$, for any $Q_{XY|W} \in \mathcal{C}(Q_{X|W}, Q_{Y|W})$, there exists $P_{XY|W} \in \mathcal{C}(P_{X|W}, P_{Y|W})$ such that

$$\|P_{WXY} - Q_{WXY}\| \le \|P_W - Q_W\| + \max_w \|P_{X|W=w} - Q_{X|W=w}\| + \max_w \|P_{Y|W=w} - Q_{Y|W=w}\|.$$

Hence, for any sequence $(P_W^{(k)}, P_{X|W}^{(k)}, P_{Y|W}^{(k)})$ convergent to $(Q_W, Q_{X|W}, Q_{Y|W})$, $\limsup_{k\to\infty} f(P_W^{(k)}, P_{X|W}^{(k)}, P_{Y|W}^{(k)}) \le f(Q_W, Q_{X|W}, Q_{Y|W})$, and $f(Q_W, Q_{X|W}, Q_{Y|W}) \le \liminf_{k\to\infty} f(P_W^{(k)}, P_{X|W}^{(k)}, P_{Y|W}^{(k)})$. Hence $f(Q_W, Q_{X|W}, Q_{Y|W})$ is continuous in $(Q_W, Q_{X|W}, Q_{Y|W})$. ■

C. Example: DSBS

Consider a DSBS with correlation coefficient ρ, whose distribution P_{XY} is given in Table I. We assume that $0 < \rho < 1$. Denote by $h : t \mapsto -t\log t - (1-t)\log(1-t)$ the binary entropy function, and h^{-1} as the inverse of the restriction of h to the set $[0, \frac{1}{2}]$.

The following explicit expression for $\underline{\Upsilon}^*$ (or E^* given in (11), or equivalently, the Gray–Wyner coding region or the mutual information region [50]) for the DSBS was conjectured by Gray and Wyner [50], [51], and recently confirmed

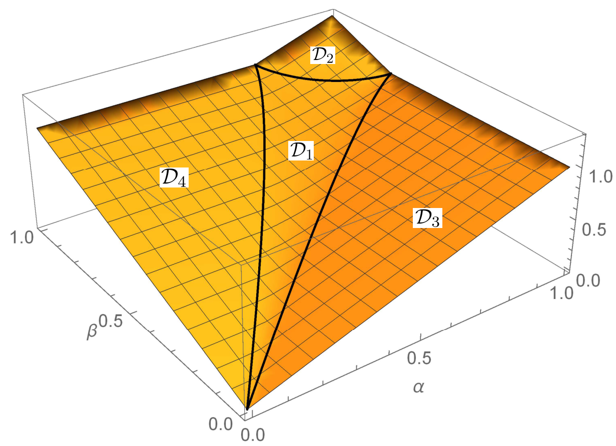

Fig. 1. Illustration of $\underline{\Upsilon}^*$ for $\rho = 0.9$.

positively by the first author [52]. For $(s,t) \in [0,1]^2$, it holds that

$$\underline{\Upsilon}^*(s,t) = \begin{cases} 1-(1-q)h(\frac{a+b-q}{2(1-q)}) \\ \quad -qh(\frac{a-b+q}{2q}), & (s,t)\in\mathcal{D}_1, \\ 1+h(q)-h(a)-h(b), & (s,t)\in\mathcal{D}_2, \\ 1-h(a), & (s,t)\in\mathcal{D}_3, \\ 1-h(b), & (s,t)\in\mathcal{D}_4, \end{cases} \quad (49)$$

where $q = \frac{1-\rho}{2}, a = h^{-1}(1-s), b = h^{-1}(1-t)$, and

$$\begin{aligned} \mathcal{D}_1 &:= \{(s,t)\in[0,1]^2 : a*q \geq b, \\ &\qquad b*q\geq a,\ a*b\geq q\}, \\ \mathcal{D}_2 &:= \{(s,t)\in[0,1]^2 : a*b<q\}, \\ \mathcal{D}_3 &:= \{(s,t)\in[0,1]^2 : a*q<b\}, \\ \mathcal{D}_4 &:= \{(s,t)\in[0,1]^2 : b*q<a\}. \end{aligned}$$

The function $\underline{\Upsilon}^*$ is plotted in Fig. 1.

We next provide explicit expressions for $\underline{\Theta}^*$ and $\overline{\Theta}^*$. Suppose $Q_X = (a, 1-a)$ and $Q_Y = (b, 1-b)$. For the DSBS, we have $D(Q_X\|P_X) = 1-h(a)$. Hence, if $D(Q_X\|P_X) = s$, then we have $a = h^{-1}(1-s)$ or $1-h^{-1}(1-s)$. Similarly, for Q_Y such that $D(Q_Y\|P_Y) = t$, we have $b = h^{-1}(1-t)$ or $1-h^{-1}(1-t)$.

Define $\kappa := (\frac{1+\rho}{1-\rho})^2$. For $\max\{0, a+b-1\} \leq p \leq \min\{a,b\}$, define

$$\begin{aligned} D_{a,b}(p) &:= D((p, a-p, b-p, 1+p-a-b)\| \\ &\qquad (\frac{1+\rho}{4}, \frac{1-\rho}{4}, \frac{1-\rho}{4}, \frac{1+\rho}{4})), & (50) \\ \mathsf{D}(a,b) &:= \min_{0,a+b-1\leq p\leq a,b} D_{a,b}(p) & (51) \\ &= D_{a,b}(p^*), & (52) \end{aligned}$$

where

$$p^* = \frac{1}{2(\kappa-1)}\Big((\kappa-1)(a+b)+1 - \sqrt{((\kappa-1)(a+b)+1)^2 - 4\kappa(\kappa-1)ab}\Big).$$

Equation (52) follows from the facts that $p \mapsto D_{a,b}(p)$ is convex (due to the convexity of the relative entropy), $\max\{0, a+b-1\} \leq p^* \leq \min\{a,b\}$, and the extreme value is taken at p^*. Furthermore, we have the following lemma, whose proof is provided in Appendix F.

Lemma 3: For $0 \leq a, b \leq \frac{1}{2}$, it holds that

$$\begin{aligned} \mathsf{D}(a,b) &= \mathsf{D}(1-a, 1-b) \\ &\leq \mathsf{D}(a, 1-b) = \mathsf{D}(1-a, b). \end{aligned} \quad (53)$$

By Lemma 3, we have

$$\begin{aligned} \varphi(s,t) &= \mathsf{D}(h^{-1}(1-s), h^{-1}(1-t)), \\ \psi(s,t) &= \mathsf{D}(h^{-1}(1-s), 1-h^{-1}(1-t)), \end{aligned}$$

Then $\underline{\Theta}^*(E_1,E_2)$ and $\overline{\Theta}^*(E_1,E_2)$ are determined by $\varphi(s,t)$ and $\psi(s,t)$ via (44) and (45). Moreover, by Theorem 4, $\underline{\Theta}(E_1,E_2)$ is attained by a sequence involving the

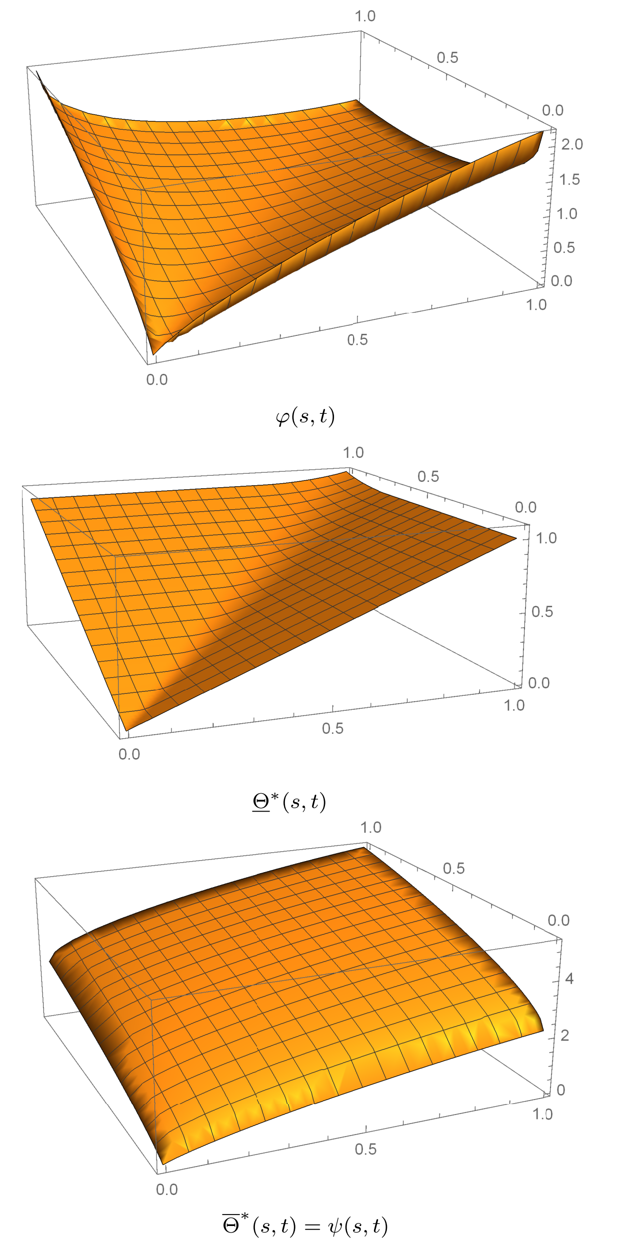

Fig. 2. Illustration of $\varphi(s,t)$, $\psi(s,t)$, $\underline{\Theta}^*(E_1,E_2)$, and $\overline{\Theta}^*(E_1,E_2)$ for the DSBS for $\rho = 0.9$. Note that $\underline{\Theta}^*(E_1,E_2)$ and $\overline{\Theta}^*(E_1,E_2)$ are expressed in terms of $\varphi(s,t)$ and $\psi(s,t)$ in (44) and (45). All the bases of logarithms are 2 for these figures.

time-sharing of at most three pairs of concentric[12] Hamming spheres, and $\overline{\Theta}(E_1, E_2)$ is attained by a sequence involving the time-sharing of at most three pairs of anti-concentric Hamming spheres.

Proposition 4 (DSBS): For the DSBS, the following hold.
1) $\underline{\Theta}(E_1, E_2)$ is achieved by a sequence of pairs of concentric Hamming spheres if $\tilde{\varphi}(E_1, E_2) := \min_{s\geq E_1, t\geq E_2} \varphi(s,t)$ is convex in (E_1, E_2).
2) $\overline{\Theta}(E_1, E_2)$ is achieved by a sequence of pairs of anti-concentric Hamming spheres if $\psi(s,t)$ is concave in (s,t).

Proof: We prove Proposition 4. For a function $f : \mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$, we use $\tilde{f}$ to denote $(x,y) \mapsto \inf_{s\geq x, t\geq y} f(s,t)$. Then, by assumption, $\tilde{\varphi}$ is convex. We now prove that $\tilde{\varphi} = \underline{\Theta}^*$, where by definition, $\underline{\Theta}^* = \breve{\tilde{\varphi}}$. On one hand, $\breve{\tilde{\varphi}} \leq \tilde{\varphi}$. On the other hand, $\breve{\tilde{\varphi}} \geq \breve{\tilde{\tilde{\varphi}}} = \tilde{\tilde{\varphi}} = \tilde{\varphi}$. Hence, $\tilde{\varphi} = \underline{\Theta}^*$. This implies that the time-sharing random variable W can be removed. By Theorem 4, time-sharing is not needed to attain $\underline{\Theta}$, completing the proof of Statement 1).

Statement 2) follows similarly, but we need to prove that $\psi(s,t) = \max_{s\leq E_1, t\leq E_2} \psi(s,t)$. This equality is equivalent to that $\max_{Q_X, Q_Y : D(Q_X\|P_X)\leq s, D(Q_Y\|P_Y)\leq t} \mathsf{D}(Q_X, Q_Y\|P_{XY})$ is always attained by some Q_X, Q_Y satisfying that both the equalities in the constraints hold. Observe that in this maximization both the objective function and the constraint functions are convex and, moreover, the set of feasible solutions is compact. By the Krein–Milman theorem, the set of feasible solutions is the closed convex hull of its extreme points. Hence, the maximization is attained by an extreme point. An extreme point here is a pair (Q_X, Q_Y) such that Q_X is either a Dirac distribution or a distribution satisfying $D(Q_X\|P_X) = s$, and so is Q_Y. For the DSBS considered here, when $s < 1$, there is no Dirac distribution in the set of feasible solutions, which means any extreme points must satisfy $D(Q_X\|P_X) = s$. Similarly, they must also satisfy $D(Q_Y\|P_Y) = t$. These are the desired, which imply Statement 2). ■

Ordentlich et al. [17] conjectured that $\underline{\Theta}(E_1, E_2)$ is achieved by a sequence of pairs of concentric Hamming spheres, and $\overline{\Theta}(E_1, E_2)$ is achieved by a sequence of pairs of anti-concentric Hamming spheres. Hence their conjecture is true under the assumptions in Proposition 4. Given Theorem 4, Ordentlich–Polyanskiy–Shayevitz's conjecture boils down to proving the convexity of $\tilde{\varphi}$ and concavity of ψ. In other words, the essence is to remove the time-sharing random variable W in both $\underline{\Theta}^*$ and $\overline{\Theta}^*$. Subsequent to the completion of this paper, the first author proved this point, and hence confirmed positively the Ordentlich–Polyanskiy–Shayevitz conjecture [39], [53]. Furthermore, noninteractive simulation in the exponential regime was also studied by Kirshner and Samorodnitsky [18] who solved the symmetric case $E_1 = E_2$.

The functions $\varphi, \psi, \underline{\Theta}^*$ and $\overline{\Theta}^*$ for $\rho = 0.9$ are plotted in Fig. 2. This figure numerically verifies the assumptions in Proposition 4.

[12]Here we call two Hamming spheres concentric if they have the same center and the radiuses are both not larger than or both not smaller than $n/2$. Similarly, two spheres are called anti-concentric if they have the same center and one of the two radiuses is not larger than $n/2$ while the other one is not smaller than $n/2$.

D. Applications to Zero-Error Coding

As mentioned in [17], the minimization part of the conjecture of Ordentlich, Polyanskiy, and Shayevitz implies a sharper outer bound for the zero-error capacity region of the binary adder channel.

Consider the two-user binary adder channel (BAC) $(\mathbf{a}, \mathbf{b}) \in \{0,1\}^{2n} \mapsto \mathbf{a} + \mathbf{b} \in \{0,1,2\}^n$ and a code $(\mathcal{A}_n, \mathcal{B}_n)$ with $\mathcal{A}_n, \mathcal{B}_n \subseteq \{0,1\}^n$ for this channel. Here $\mathbf{a}+\mathbf{b}$ denotes addition over $\mathbb{Z}^n$. When the code $(\mathcal{A}_n, \mathcal{B}_n)$ is used to transmit messages over the BAC, the receiver is able to decode the messages without any error if and only if any pair $(\mathbf{a}, \mathbf{b}) \in \mathcal{A}_n \times \mathcal{B}_n$ is mapped to a unique sequence in $\{0,1,2\}^n$, i.e., $|\mathcal{A}_n + \mathcal{B}_n| = |\mathcal{A}_n| \cdot |\mathcal{B}_n|$, where $\mathcal{A}_n + \mathcal{B}_n$ denotes the sumset $\{\mathbf{a} + \mathbf{b} : \mathbf{a} \in \mathcal{A}_n, \mathbf{b} \in \mathcal{B}_n\}$. The zero-error capacity region $\mathcal{C}$ of the BAC (or the rate region of uniquely decodable code pairs) is defined as the set of (R_1, R_2) for which there is a sequence of pairs $\mathcal{A}_n, \mathcal{B}_n \subseteq \{0,1\}^n$ with $|\mathcal{A}_n| = 2^{n(R_1+o(1))}, |\mathcal{B}_n| = 2^{n(R_2+o(1))}$ such that $|\mathcal{A}_n + \mathcal{B}_n| = |\mathcal{A}_n| \cdot |\mathcal{B}_n|$ for every n.

Finding the capacity region of the BAC is a long standing open problem; refer to [17], [54], [55], [56], [57], [58], [59], [60], [61], and [62] for details. The current progress on this topic is rather unsatisfactory. The upper bound on the sum rate $R_1 + R_2$ is still the simple bound $3/2$, which corresponds to the maximum sum rate in the Shannon capacity. However, Urbanke and Li [59] broke through the $3/2$ bound in the unbalanced case, in which it is assumed that $R_1 = 1$ (note that it does not mean $\mathcal{A}_n = \{0,1\}^n$) and they showed that $R_2 \leq 0.4921$. Later, this result was improved to $R_2 \leq 0.4798$ in [61] and $R_2 \leq 0.4228$ respectively in [62]. The latter is the best known upper bound until now. The best known lower bound for this case is $R_2 \geq 1/4$ given in [58].

In particular, the reverse small-set expansion inequality given in (39) for the DSBS was used by Austrin, Kaski, Koivisto, and Nederlof [62] to prove the best known upper bound. As mentioned by Ordentlich et al. [17], repeating the arguments in [62] with improved bounds on $\overline{\Theta}(E_1, E_2)$ will yield tighter bounds on R_2 when $R_1 = 1$. Replacing the reverse small-set expansion inequality in the proof given in [62] with the characterization of $\overline{\Theta}(E_1, E_2)$ in Theorem 4, we obtain the following result.

Theorem 5: If $(1-\epsilon, R_2) \in \mathcal{C}$, then for any $\rho \in (0,1)$ there exists some $\lambda \in \frac{1}{2} \pm \sqrt{\frac{\ln(2)\epsilon}{2}}$ such that

$$\lambda\overline{\Theta}^*\left(\frac{\epsilon}{\lambda}, \frac{\lambda+\epsilon-R_2}{\lambda}\right) \geq \lambda\left(\frac{5}{2} - \log(3-\rho)\right) - \frac{1}{2} - \epsilon - \sqrt{\frac{\ln(2)\epsilon}{2}},$$

where $\overline{\Theta}^*$ is defined for the DSBS with correlation coefficient ρ. In particular, if $\epsilon = 0$, we obtain for any $\rho \in (0,1)$,

$$\frac{1}{2}\overline{\Theta}^*(0, 1-2R_2) \geq \frac{1}{2}\left(\frac{3}{2} - \log(3-\rho)\right). \tag{54}$$

Numerical results show that if $R_1 = 1$ (i.e., $\epsilon = 0$), by choosing the almost best $\rho = 0.6933$, (54) implies $R_2 \leq 0.4177$, which improves the previously best known bound $R_2 \leq 0.4228$ established in [62]. Note that the upper bound $R_2 \leq 0.4177$ was first calculated in [17]. Subsequent to the completion of this paper, $\overline{\Theta}^* = \psi$ was proven by the first author [39], [53], which further simplifies the inequalities in Theorem 5.

IV. Brascamp–Lieb and Hypercontractivity Inequalities

In this section, we relax Boolean functions in noninteractive simulation problems to any nonnegative functions, but still restrict their supports to be exponentially small. Let $(\mathbf{X}, \mathbf{Y}) \sim P_{XY}^n$, where P_{XY} is a joint distribution defined on $\mathcal{X} \times \mathcal{Y}$. Recall the notation $\langle f, g\rangle = \mathbb{E}[f(\mathbf{X})g(\mathbf{Y})]$ and $\|f\|_p = \left(\mathbb{E}[f(\mathbf{X})^p]\right)^{1/p}$. We continue to assume that $\mathcal{X}, \mathcal{Y}$ are finite sets, each with cardinality at least 2, and with $P_X(x) > 0, P_Y(y) > 0$ for all $(x, y) \in \mathcal{X} \times \mathcal{Y}$. We next derive strengthened versions of (forward and reverse) Brascamp–Lieb and hypercontractivity inequalities by using Theorem 4. Our inequalities reduce to the usual ones when $\alpha = \beta = 0$. For $\alpha \in [0, E_{1,\max}], \beta \in [0, E_{2,\max}]$ and $p, q \in (0, \infty)$, define

$$\underline{\Lambda}_{p,q}^*(\alpha, \beta) := \min_{\substack{\alpha \le s \le E_{1,\max} \\ \beta \le t \le E_{2,\max}}} (\underline{\Theta}^*(s, t) - \frac{s}{p} - \frac{t}{q}),$$

and

$$\overline{\Lambda}_{p,q}^*(\alpha, \beta) := \min_{\substack{\alpha \le s \le E_{1,\max} \\ \beta \le t \le E_{2,\max}}} (\frac{s}{p} + \frac{t}{q} - \overline{\Theta}^*(s, t)).$$

Remark 10: Subsequent to the completion of this paper, the first author proved that $\underline{\Theta}^* = \tilde{\varphi}$ and $\overline{\Theta}^* = \psi$ for the DSBS in [53].

The strengthened (forward and reverse) Brascamp–Lieb inequalities are given in the following theorem, whose proof is provided in Appendix G.

Theorem 6: Let $p, q > 0$ and $\alpha \in [0, E_{1,\max}], \beta \in [0, E_{2,\max}]$. Let f, g be nonnegative functions on $\mathcal{X}^n$ and $\mathcal{Y}^n$ respectively such that $P_X^n(\mathrm{supp}(f)) \le 2^{-n\alpha}, P_Y^n(\mathrm{supp}(g)) \le 2^{-n\beta}$. Then

$$\langle f, g\rangle \le 2^{-n\underline{\Lambda}_{p,q}^*(\alpha,\beta)} \|f\|_p \|g\|_q, \tag{55}$$

$$\langle f, g\rangle \ge 2^{n\overline{\Lambda}_{p,q}^*(\alpha,\beta)} \|f\|_p \|g\|_q. \tag{56}$$

Remark 11: Given $\alpha \in [0, E_{1,\max}], \beta \in [0, E_{2,\max}]$ and $p, q > 0$, the inequality (55) is exponentially sharp, in the sense that the exponents on the two sides of (55) are asymptotically equal as $n \to \infty$, for a sequence of Boolean functions $f_n = 1_{\mathcal{A}_n}, g_n = 1_{\mathcal{B}_n}$ with $(\mathcal{A}_n, \mathcal{B}_n)$ denoting the sets given in Remark 7 but with (E_1, E_2) there replaced by the optimal (s^*, t^*) attaining the minimum in the definition of $\underline{\Lambda}_{p,q}^*(\alpha, \beta)$. Note that if (α, β) is in the effective region of $\overline{\Theta}^*$, then the optimal (s^*, t^*) attaining the minimum in the definition of $\overline{\Lambda}_{p,q}^*(\alpha, \beta)$ is still in the effective region of $\overline{\Theta}^*$. Given (α, β) in the effective region of $\overline{\Theta}^*$ and $p, q > 0$, the inequality (56) is exponentially sharp, in the sense that the exponents on the two sides of (56) are asymptotically equal as $n \to \infty$, for a sequence of Boolean functions $f_n = 1_{\mathcal{A}_n}, g_n = 1_{\mathcal{B}_n}$ with some sequence $(\mathcal{A}_n, \mathcal{B}_n)$; see Remark 8.

Remark 12: A special case of Theorem 6 with $p \ge p_0, q \ge q_0$ for (55) and $p \le p_1, q \le q_1$ for (55) can be recovered by the information-theoretic characterization of classic Brascamp–Lieb inequalities, where $(\frac{1}{p_0}, \frac{1}{q_0})$ is a subgradient of $\underline{\Theta}^*$ and $(\frac{1}{p_1}, \frac{1}{q_1})$ is a subgradient of $\overline{\Theta}^*$. See Corollary 3 in [39] which is a consequence of Theorem 2 therein, and also the simple proof of Theorem 2 therein given in Appendix C in [39].

For $\alpha \in [0, E_{1,\max}], \beta \in [0, E_{2,\max}]$, define the forward and reverse (α, β)-hypercontractivity regions as

$$\begin{aligned}\mathcal{R}_{\alpha,\beta}^+(P_{XY}) := \{&(p, q) \in (0, \infty)^2 : \\ &\underline{\Theta}^*(E_1, E_2) \ge \frac{1}{p}E_1 + \frac{1}{q}E_2, \\ &\forall E_1 \in [\alpha, E_{1,\max}], E_2 \in [\beta, E_{2,\max}]\}, \\ \mathcal{R}_{\alpha,\beta}^-(P_{XY}) := \{&(p, q) \in (0, \infty)^2 : \\ &\overline{\Theta}^*(E_1, E_2) \le \frac{1}{p}E_1 + \frac{1}{q}E_2, \\ &\forall E_1 \in [\alpha, E_{1,\max}], E_2 \in [\beta, E_{2,\max}]\}.\end{aligned}$$

For $\alpha = \beta = 0$, $\mathcal{R}_{0,0}^+(P_{XY})$ and $\mathcal{R}_{0,0}^-(P_{XY})$ correspond to the classic hypercontractivity regions in [29], [33], [34], and [47] for the forward one and [35], [36], [37], [39], [47] for the reverse one.

As a consequence of Theorem 6, we obtain the following new version of hypercontractivity.

Theorem 7: Under the assumption in Theorem 6, it holds that

$$\langle f, g\rangle \le \|f\|_p \|g\|_q, \forall (p, q) \in \mathcal{R}_{\alpha,\beta}^+(P_{XY}), \tag{57}$$

$$\langle f, g\rangle \ge \|f\|_p \|g\|_q, \forall (p, q) \in \mathcal{R}_{\alpha,\beta}^-(P_{XY}). \tag{58}$$

Remark 13: These two inequalities are exponentially sharp in the same sense as (55) and (56); see Remark 11. That is, given $\alpha \in [0, E_{1,\max}], \beta \in [0, E_{2,\max}]$ and (p, q) in the boundary of $\mathcal{R}_{\alpha,\beta}^+(P_{XY})$, the inequality (55) is exponentially sharp, in the sense that the exponents on the two sides of (55) are asymptotically equal as $n \to \infty$, for a sequence of Boolean functions $f_n = 1_{\mathcal{A}_n}, g_n = 1_{\mathcal{B}_n}$ with some sequence $(\mathcal{A}_n, \mathcal{B}_n)$. Given (α, β) in the effective region of $\overline{\Theta}^*$ and (p, q) in the boundary of $\mathcal{R}_{\alpha,\beta}^-(P_{XY})$, the inequality (56) is exponentially sharp, in the sense that the exponents on the two sides of (56) are asymptotically equal as $n \to \infty$, for a sequence of Boolean functions $f_n = 1_{\mathcal{A}_n}, g_n = 1_{\mathcal{B}_n}$ with some sequence $(\mathcal{A}_n, \mathcal{B}_n)$.

Note that the hypercontractivity inequalities in Theorem 6 differ from the common ones in the factors $2^{-n\underline{\Lambda}_{p,q}^*(\alpha,\beta)}$ and $2^{n\overline{\Lambda}_{p,q}^*(\alpha,\beta)}$; while the ones in Theorem 7 differ from the common ones in the region of parameters p, q. Strengthening the forward hypercontractivity was previously studied in [18] and [42]. Polyanskiy and Samorodnitsky [42] strengthened the hypercontractivity inequalities in a similar sense to Theorem 6; while Kirshner and Samorodnitsky [18] strengthened the hypercontractivity inequalities in a similar sense to Theorem 7. However, both works in [18] and [42] focused on strengthening the single-function version of forward hypercontractivity. Moreover, the hypercontractivity inequalities in [42] are only sharp at extreme cases, and only DSBSes were considered in [18]. A systematic investigation of the exponentially sharp version of Brascamp–Lieb and hypercontractivity inequalities in Polish spaces and under a general measure of the "sizes" of functions (termed the two-parameter entropy) was done by the first author in [39].

V. Concluding Remarks

The maximal density of subgraphs of a type graph and the biclique rate region have been studied in this paper. One may be also interested in their counterparts—the minimal density of subgraphs of a type graph and the independent-set rate region. Here, given a joint n-type T_{XY}, $1 \le M_1 \le |\mathcal{T}_{T_X}|$, and $1 \le M_2 \le |\mathcal{T}_{T_Y}|$, we define the minimal density of subgraphs of the type graph of T_{XY} with size (M_1, M_2) as

$$\underline{\Gamma}_n(M_1, M_2) := \min_{\substack{\mathcal{A}\subseteq\mathcal{T}_{T_X},\mathcal{B}\subseteq\mathcal{T}_{T_Y}:\\ |\mathcal{A}|=M_1,|\mathcal{B}|=M_2}} \rho(G[\mathcal{A},\mathcal{B}]).$$

Similar to the biclique rate region, we define the independent-set rate region as

$$\underline{\mathcal{R}}_n(T_{XY}) := \{(R_1, R_2) \in \mathcal{R}_X^{(n)} \times \mathcal{R}_Y^{(n)} : \underline{\Gamma}_n(2^{nR_1}, 2^{nR_2}) = 0\}.$$

Then one can easily obtain the following inner bound and outer bound on $\underline{\mathcal{R}}_n(T_{XY})$.

Proposition 5: For any n and T_{XY},

$$\begin{aligned}&(\underline{\mathcal{R}}^{(i)}(T_{XY}) - [0, \varepsilon_{1,n}] \times [0, \varepsilon_{2,n}]) \cap (\mathcal{R}_X^{(n)} \times \mathcal{R}_Y^{(n)})\\ &\subseteq \underline{\mathcal{R}}_n(T_{XY})\\ &\subseteq \underline{\mathcal{R}}^{(o)}(T_{XY}) \cap (\mathcal{R}_X^{(n)} \times \mathcal{R}_Y^{(n)})\end{aligned}$$

for some positive sequences $\{\varepsilon_{1,n}\}$ and $\{\varepsilon_{2,n}\}$ which both vanish as $n \to \infty$, where

$$\begin{aligned}\underline{\mathcal{R}}^{(o)}(T_{XY}) &:= \{(R_1, R_2) : R_1 \le H(X), R_2 \le H(Y)\},\\ \underline{\mathcal{R}}^{(i)}(T_{XY}) &:= \bigcup_{\substack{P_W, P_{X|W}, P_{Y|W}:\\ P_W P_{X|W}, P_W P_{Y|W} \text{ are } n\text{-types},\\ P_X = T_X, P_Y = T_Y,\\ Q_{XY} \ne T_{XY}, \forall Q_{XY|W} \in \mathcal{C}(P_{X|W}, P_{Y|W})}} \{(R_1, R_2) : R_1 \le H(X|W), R_2 \le H(Y|W)\}.\end{aligned}$$

The inner bound above can be proven by using the codes used in proving the achievability part of Theorem 1. The outer bound above is trivial. Determining the asymptotics of $\underline{\mathcal{R}}_n(T_{XY})$ could be of interest. However, currently, we have no idea how to tackle it. In addition, if $\underline{\mathcal{R}}_n(T_{XY})$ is not asymptotically equal to $\underline{\mathcal{R}}^{(o)}(T_{XY})$, then determining the exponent of the minimal density is also interesting.

Furthermore, many other fundamental properties of type graphs remain to be investigated, including graph coloring, graph circuits, graph embedding, graph connectivity, covering and packing, etc. [63]. Thanks to good structures enjoyed by type graphs, it seems not hopeless to characterize them.

Appendix A
Proof of Lemma 1

Statements 1) and 2) follow directly from the definition of $F^*(R_1, R_2)$. Note that in Statement 2), the maximum $H_T(X, Y)$ is attained by $P_{XYW} = T_{XY} P_W$ (or set W to constant) when $R_1 = H_T(X), R_2 = H_T(Y)$.

Statement 3): By symmetry, it suffices to only consider the case $R_1 = 0$. By Statement 2), $F^*(0, R_2) \le \min\{R_2, H_T(Y|X)\}$. On the other hand, if $R_2 \ge H_T(Y|X)$, then we choose $W = X$, which leads to $H(X|W) = 0$ and $H(Y|W) = H(X, Y|W) = H_T(Y|X)$. Hence we have $F^*(0, R_2) = H_T(Y|X)$ for $R_2 \ge H_T(Y|X)$. If $R_2 \le H_T(Y|X)$, then one can find a random variable U such that $H(Y|X, U) = R_2$. For example, we choose $U = (V, J)$ with V defined on $\mathcal{X} \cup \mathcal{Y}$ and J defined on $\{0, 1\}$ such that $V = X$ if $J = 0$ and $V = Y$ if $J = 1$, where $J \sim \mathrm{Bern}(\alpha)$ for $\alpha := R_2/H_T(Y|X)$ is independent of (X, Y). Set $W = (X, U)$. We have $H(X|W) = 0$ and $H(Y|W) = H(X, Y|W) = H(X, Y|W, J) = R_2$. Hence we have $F^*(0, R_2) = R_2$ for $R_2 \le H_T(Y|X)$.

Statement 4): Let P_{XYW_0} attain $F^*(R_1, R_2)$, and P_{XYW_1} attain $F^*(\hat{R}_1, \hat{R}_2)$. For $0 < \alpha < 1$, define $J \sim \mathrm{Bern}(\alpha)$ independent of (X, Y, W_0, W_1) and let $W := W_J$, taking values in $\mathcal{W}_0 \cup \mathcal{W}_1$, where $\mathcal{W}_j$ denotes the alphabet of W_j for $j = 0, 1$. Note that J is a deterministic function of W. Then P_{XYW} induces

$$\begin{aligned}H(X, Y|W) &= \alpha H(X, Y|W_0) + (1-\alpha) H(X, Y|W_1),\\ H(X|W) &= \alpha H(X|W_0) + (1-\alpha) H(X|W_1),\\ H(Y|W) &= \alpha H(Y|W_0) + (1-\alpha) H(Y|W_1).\end{aligned}$$

Therefore,

$$\begin{aligned}&F^*(\alpha R_1 + (1-\alpha)\hat{R}_1, \alpha R_2 + (1-\alpha)\hat{R}_2)\\ &\ge \alpha F^*(R_1, R_2) + (1-\alpha) F^*(\hat{R}_1, \hat{R}_2).\end{aligned}$$

Statement 5): If $\delta_1 = \delta_2 = 0$, there is nothing to prove. If $\delta_2 > \delta_1 = 0$, then, for $t \ge 0$,

$$f(t) := F^*(R_1, t)$$

is nondecreasing and concave, by Statements 1) and 4). Hence, for fixed δ_2,

$$\frac{f(t+\delta_2) - f(t)}{\delta_2}$$

is nonincreasing in t. Combining this with Statements 2) and 3) yields

$$\begin{aligned}&\frac{f(t+\delta_2) - f(t)}{\delta_2} \le \frac{f(\delta_2) - f(0)}{\delta_2}\\ &\le \frac{\delta_2 + \min\{R_1, H_T(X|Y)\} - \min\{R_1, H_T(X|Y)\}}{\delta_2}\\ &= 1.\end{aligned}$$

Setting $t = R_2$, we obtain $F^*(R_1, R_2 + \delta_2) - F^*(R_1, R_2) \le \delta_2$, as desired.

By symmetry, the claim also holds in the case $\delta_1 > \delta_2 = 0$. Now we consider the case $\delta_1, \delta_2 > 0$. Without loss of generality, we assume $\frac{R_1}{\delta_1} \ge \frac{R_2}{\delta_2}$. For $t \ge -\frac{R_2}{\delta_2}$, define

$$g(t) := F^*(R_1 + \delta_1 t, R_2 + \delta_2 t).$$

By Statements 1) and 4), $g(t)$ is nondecreasing and concave. Hence, for fixed δ_2,

$$g(t+1) - g(t)$$

is nonincreasing in t. Combining this with Statements 2) and 3) yields that for $t \ge -\frac{R_2}{\delta_2}$ we have

$$\begin{aligned} g(t+1) - g(t) &\le g(-\frac{R_2}{\delta_2}+1) - g(-\frac{R_2}{\delta_2}) \\ &= F^*(R_1 - \frac{\delta_1 R_2}{\delta_2} + \delta_1, \delta_2) - F^*(R_1 - \frac{\delta_1 R_2}{\delta_2}, 0) \\ &\le \min\{R_1 - \frac{\delta_1 R_2}{\delta_2} + \delta_1, H_T(X|Y)\} + \delta_2 \\ &\quad - \min\{R_1 - \frac{\delta_1 R_2}{\delta_2}, H_T(X|Y)\} \\ &\le \delta_1 + \delta_2. \end{aligned}$$

Setting $t = 0$, we obtain $F^*(R_1+\delta_1, R_2+\delta_2) - F^*(R_1, R_2) \le \delta_1 + \delta_2$, as desired.

Appendix B
Proof of Theorem 1

The claim that we can restrict attention to the case $|\mathcal{W}| \le |\mathcal{X}||\mathcal{Y}| + 2$ in the definition of $F^*(R_1, R_2)$ comes from the support lemma in [43]. We next prove (12).

Lower bound: Let $\mathcal{C} := (\mathcal{A} \times \mathcal{B}) \cap \mathcal{T}_{T_{XY}}$ for some optimal $(\mathcal{A}, \mathcal{B})$ attaining $\Gamma_n(2^{nR_1}, 2^{nR_2})$. Let $(\mathbf{X}, \mathbf{Y}) \sim \mathrm{Unif}(\mathcal{C})$. Then,

$$\begin{aligned} \Gamma_n(2^{nR_1}, 2^{nR_2}) = \frac{|\mathcal{C}|}{|\mathcal{A}||\mathcal{B}|} &= \frac{2^{H(\mathbf{X},\mathbf{Y})}}{2^{nR_1}2^{nR_2}}, \\ \frac{1}{n}H(\mathbf{X}) &\le R_1, \\ \frac{1}{n}H(\mathbf{Y}) &\le R_2, \end{aligned}$$

which follow by the fact that the entropy of a random variable is no larger than the logarithm of its support size, and they are equal if the random variable is uniformly distributed over its support. Therefore,

$$\begin{aligned} E_n(R_1, R_2) &= R_1 + R_2 - \frac{1}{n}H(\mathbf{X}, \mathbf{Y}) \\ &= R_1 + R_2 - \frac{1}{n}\sum_{i=1}^{n} H(X_i, Y_i | X^{i-1}, Y^{i-1}) \\ &= R_1 + R_2 - H(X_J, Y_J | X^{J-1}, Y^{J-1}, J), \end{aligned}$$

where $J \sim \mathrm{Unif}[n]$ is a random time index independent of (X^n, Y^n) and X^{J-1} denotes a "random vector"[13] induced by (J, X^n). On the other hand,

$$H(X_J|X^{J-1}, Y^{J-1}, J) \le H(X_J|X^{J-1}J) = \frac{1}{n}H(\mathbf{X}) \le R_1,$$
$$H(Y_J|X^{J-1}, Y^{J-1}, J) \le R_2.$$

Using the notation

$$X := X_J, Y := Y_J, W := (X^{J-1}, Y^{J-1}, J),$$

we obtain $(X, Y) \sim T_{XY}$, and

$$\begin{aligned} E_n(R_1, R_2) &\ge \inf_{\substack{P_{XYW}: P_{XY}=T_{XY}, \\ H(X|W) \le R_1, \\ H(Y|W) \le R_2}} R_1 + R_2 - H(X, Y|W) \\ &= E^*(R_1, R_2). \end{aligned}$$

Upper bound: In this part, we assume that W is a random variable defined on an alphabet $\mathcal{W}$ such that $|\mathcal{W}| \le |\mathcal{X}||\mathcal{Y}| + 2$. For a joint n-type P_{XYW} such that $P_{XY} = T_{XY}, H(X|W) \le R_1, H(Y|W) \le R_2$ and for a fixed sequence $\mathbf{w}$ with type P_W, we choose $\mathcal{A}$ as the union of $\mathcal{T}_{P_{X|W}}(\mathbf{w})$ and $2^{nR_1} - |\mathcal{T}_{P_{X|W}}(\mathbf{w})|$ of arbitrary sequences outside $\mathcal{T}_{P_{X|W}}(\mathbf{w})$, which is possible because $\mathcal{T}_{P_{X|W}}(\mathbf{w}) \le 2^{nH(X|W)}$, see [3, Lemma 2.5], and choose $\mathcal{B}$ in a similar way, but with $\mathcal{T}_{P_{X|W}}(\mathbf{w})$ replaced by $\mathcal{T}_{P_{Y|W}}(\mathbf{w})$. Then $|\mathcal{A}| = 2^{nR_1}$ and $|\mathcal{B}| = 2^{nR_2}$. Observe that

$$\begin{aligned} |(\mathcal{A} \times \mathcal{B}) \cap \mathcal{T}_{T_{XY}}| &\ge |\mathcal{T}_{P_{XY|W}}(\mathbf{w})| \\ &\ge 2^{n(H(X,Y|W) - \frac{|\mathcal{W}||\mathcal{X}||\mathcal{Y}|\log(n+1)}{n})}, \end{aligned}$$

where

- the first inequality follows since for any pair $(\mathbf{x}, \mathbf{y}) \in \mathcal{T}_{P_{XY|W}}(\mathbf{w})$, the tuple $(\mathbf{w}, \mathbf{x}, \mathbf{y})$ must have joint type P_{WXY}, and hence, $(\mathbf{w}, \mathbf{x})$ has joint type P_{WX}, $(\mathbf{w}, \mathbf{y})$ has joint type P_{WY}, and $(\mathbf{x}, \mathbf{y})$ has joint type T_{XY};
- the second inequality follows from [3, Lemma 2.5].

Thus we have

$$\begin{aligned} \rho(G[\mathcal{A}, \mathcal{B}]) &= \frac{|(\mathcal{A} \times \mathcal{B}) \cap \mathcal{T}_{T_{XY}}|}{2^{nR_1}2^{nR_2}} \\ &\ge 2^{-n(R_1 + R_2 - H(X,Y|W) + \frac{|\mathcal{W}||\mathcal{X}||\mathcal{Y}|\log(n+1)}{n})}. \end{aligned} \tag{59}$$

Optimizing the exponent in (59) over all joint n-types P_{XYW} such that $P_{XY} = T_{XY}, H(X|W) \le R_1, H(Y|W) \le R_2$ yields the upper bound

$$\begin{aligned} E_n(R_1, R_2) &\le R_1 + R_2 - F_n(R_1, R_2) \\ &\quad + \frac{|\mathcal{W}||\mathcal{X}||\mathcal{Y}|\log(n+1)}{n}, \end{aligned} \tag{60}$$

where $F_n(R_1, R_2)$ is defined similarly as $F^*(R_1, R_2)$ in (10) but with the P_{XYW} in (10) restricted to be a joint n-type and $\mathcal{W}$ assumed to satisfy $|\mathcal{W}| \le |\mathcal{X}||\mathcal{Y}| + 2$.

We next show that the values of $F_n(R_1, R_2)$ and $F^*(R_1, R_2)$ do not differ too much. For a joint n-type T_{XY} and a distribution P_{XYW} with $P_{XY} = T_{XY}$, one can find a n-type Q_{XYW} with $Q_{XY} = T_{XY}$ such that $\|P_{XYW} - Q_{XYW}\| \le \frac{|\mathcal{W}||\mathcal{X}||\mathcal{Y}|}{2n}$, where $\|\cdot\|$ denotes the TV distance, see [64, Lemma 3]. Combining this with [3, Lemma 2.7] (i.e., if $\|P_X - Q_X\| \le \Theta \le \frac{1}{4}$, then $|H_P(X) - H_Q(X)| \le -2\Theta \log \frac{2\Theta}{|\mathcal{X}|}$), we have for $\frac{|\mathcal{W}||\mathcal{X}||\mathcal{Y}|}{2n} \le \frac{1}{4}$ that

$$\begin{aligned} &|H_P(X|W) - H_Q(X|W)| \\ &\le |H_P(X,W) - H_Q(X,W)| + |H_P(W) - H_Q(W)| \\ &\le -2\frac{|\mathcal{W}||\mathcal{X}||\mathcal{Y}|}{2n} \log \frac{2\frac{|\mathcal{W}||\mathcal{X}||\mathcal{Y}|}{2n}}{|\mathcal{X}||\mathcal{W}|} \\ &\quad - 2\frac{|\mathcal{W}||\mathcal{X}||\mathcal{Y}|}{2n} \log \frac{2\frac{|\mathcal{W}||\mathcal{X}||\mathcal{Y}|}{2n}}{|\mathcal{W}|} \end{aligned} \tag{61}$$

[13] Rigorously speaking, the "random vector" X^{J-1} is not well defined since for different i, the random vectors X^{i-1} are defined on different spaces. (The space of X^{i-1} is $\mathcal{X}^{i-1}$ for each i.) One way to address this issue is to map X^{i-1} to a common (measurable) space via one-to-one functions. Another simpler way is to concatenate X^{i-1} with a length-$(n-i+1)$ of constant symbols, e.g., $\hat{X}^n_{(i-1)} := (X^{i-1}, x_0, \dots, x_0)$ where x_0 is a fixed symbol and appears $n-i+1$ times here. In this case, X^{J-1} denotes $\hat{X}^n_{(J-1)}$. This convention applies throughout this paper.

$$= -\frac{|\mathcal{W}||\mathcal{X}||\mathcal{Y}|}{n}\log\frac{|\mathcal{X}||\mathcal{Y}|^2}{n^2}, \tag{62}$$

and similarly,

$$|H_P(Y|W) - H_Q(Y|W)| \le -\frac{|\mathcal{W}||\mathcal{X}||\mathcal{Y}|}{n}\log\frac{|\mathcal{X}|^2|\mathcal{Y}|}{n^2}, \tag{63}$$

$$|H_P(XY|W) - H_Q(XY|W)| \le -\frac{|\mathcal{W}||\mathcal{X}||\mathcal{Y}|}{n}\log\frac{|\mathcal{X}||\mathcal{Y}|}{n^2}. \tag{64}$$

Combining (62)-(64) yields that

$$F_n(R_1,R_2) \ge F^*(R_1 + \frac{|\mathcal{W}||\mathcal{X}||\mathcal{Y}|}{n}\log\frac{|\mathcal{X}||\mathcal{Y}|^2}{n^2}, R_2 + \frac{|\mathcal{W}||\mathcal{X}||\mathcal{Y}|}{n}\log\frac{|\mathcal{X}|^2|\mathcal{Y}|}{n^2}) + \frac{|\mathcal{W}||\mathcal{X}||\mathcal{Y}|}{n}\log\frac{|\mathcal{X}||\mathcal{Y}|}{n^2}.$$

Applying Statement 5) of Lemma 1, we obtain

$$F_n(R_1,R_2) \ge F^*(R_1,R_2) + \frac{|\mathcal{W}||\mathcal{X}||\mathcal{Y}|}{n}\log\frac{|\mathcal{X}|^4|\mathcal{Y}|^4}{n^6}.$$

Substituting this into the upper bound in (60) and combining with the assumption $|\mathcal{W}| \le |\mathcal{X}||\mathcal{Y}|+2$ yields the desired upper bound.

Appendix C
Proof of Theorem 2

We now prove Theorem 2. Since (19) follows from (18), it suffices to prove (18).

Inner Bound: The inner bound proof here uses a standard time-sharing argument. Let d be an integer such that $1 \le d \le n-1$. Let (P_{XY}, Q_{XY}) be a pair comprised of a d-joint type and an $(n-d)$-joint type on $\mathcal{X}\times\mathcal{Y}$ such that $\frac{d}{n}P_{XY} + (1-\frac{d}{n})Q_{XY} = T_{XY}$. For a fixed length-$d$ sequence $\mathbf{y}$ with type P_Y an a fixed length-$(n-d)$ sequence $\mathbf{x}$ with type Q_X, we choose $\mathcal{A} = \mathcal{T}_{P_{X|Y}}(\mathbf{y}) \times \{\mathbf{x}\}$ and $\mathcal{B} = \{\mathbf{y}\} \times \mathcal{T}_{Q_{Y|X}}(\mathbf{x})$. Then, from [3, Lemma 2.5], we have $|\mathcal{A}| \ge 2^{d(H_P(X|Y) - \frac{|\mathcal{X}||\mathcal{Y}|\log(d+1)}{d})}$ and similarly $|\mathcal{B}| \ge 2^{(n-d)(H_Q(Y|X) - \frac{|\mathcal{X}||\mathcal{Y}|\log(n-d+1)}{n-d})}$. On the other hand, for this code we have $\mathcal{A}\times\mathcal{B} \subseteq \mathcal{T}_{T_{XY}}$. Hence any rate pair $(R_1,R_2) \in (\mathcal{R}_X^{(n)} \times \mathcal{R}_Y^{(n)})$ with

$$R_1 \le \frac{d}{n}(H_P(X|Y) - \frac{|\mathcal{X}||\mathcal{Y}|\log(d+1)}{d}),$$
$$R_2 \le (1-\frac{d}{n})(H_Q(Y|X) - \frac{|\mathcal{X}||\mathcal{Y}|\log(n-d+1)}{n-d}),$$

is achievable (i.e., it is in $\mathcal{R}_n(T_{XY})$), which in turn implies that a pair of smaller rates $(R_1,R_2) \in (\mathcal{R}_X^{(n)} \times \mathcal{R}_Y^{(n)})$ with

$$R_1 \le \frac{d}{n}H_P(X|Y) - \frac{|\mathcal{X}||\mathcal{Y}|\log(n+1)}{n}, \tag{65}$$

$$R_2 \le (1-\frac{d}{n})H_Q(Y|X) - \frac{|\mathcal{X}||\mathcal{Y}|\log(n+1)}{n}, \tag{66}$$

is achievable.

We next remove the constraint that (P_{XY}, Q_{XY}) are joint types. For $0 \le \alpha \le 1$, let $(\hat{P}_{XY}, \hat{Q}_{XY})$ be a pair of distributions such that $\alpha\hat{P}_{XY} + (1-\alpha)\hat{Q}_{XY} = T_{XY}$. Define $d := \|\lfloor n\alpha\hat{P}_{XY}\rfloor\|_1$. Note that we have

$$n\alpha - |\mathcal{X}||\mathcal{Y}| \le d \le n\alpha. \tag{67}$$

We first consider the case

$$4|\mathcal{X}||\mathcal{Y}| \le d \le n - 4|\mathcal{X}||\mathcal{Y}|. \tag{68}$$

Define $P_{XY} := \frac{\lfloor n\alpha\hat{P}_{XY}\rfloor}{d}$. Then P_{XY} is a joint d-type and $\|P_{XY} - \hat{P}_{XY}\| \le \frac{|\mathcal{X}||\mathcal{Y}|}{d} \le \frac{1}{4}$. Define $Q_{XY} := \frac{nT_{XY} - dP_{XY}}{n-d}$, which is a joint $(n-d)$-type and satisfies $\|Q_{XY} - \hat{Q}_{XY}\| \le \frac{|\mathcal{X}||\mathcal{Y}|}{n-d} \le \frac{1}{4}$. Combining [3, Lemma 2.7] with the equality $H(X|Y) = H(X,Y) - H(Y)$, we have

$$H_P(X|Y) \ge H_{\hat{P}}(X|Y) + \frac{2|\mathcal{X}||\mathcal{Y}|}{d}\log\frac{4|\mathcal{X}|}{d^2},$$
$$H_Q(Y|X) \ge H_{\hat{Q}}(Y|X) + \frac{2|\mathcal{X}||\mathcal{Y}|}{n-d}\log\frac{4|\mathcal{Y}|}{(n-d)^2}.$$

These inequalities, together with (65) and (66), imply that

$$\text{RHS of (65)} \ge \frac{d}{n}H_{\hat{P}}(X|Y) - \frac{2|\mathcal{X}||\mathcal{Y}|}{n}\log\frac{n^2}{4|\mathcal{X}|} - \frac{|\mathcal{X}||\mathcal{Y}|\log(n+1)}{n} \tag{69}$$

$$\ge \alpha H_{\hat{P}}(X|Y) - \frac{|\mathcal{X}||\mathcal{Y}|}{n}\log|\mathcal{X}| - \frac{2|\mathcal{X}||\mathcal{Y}|}{n}\log\frac{n^2}{4|\mathcal{X}|} - \frac{|\mathcal{X}||\mathcal{Y}|\log(n+1)}{n} \tag{70}$$

$$= \alpha H_{\hat{P}}(X|Y) - \epsilon_{1,n}; \tag{71}$$

$$\text{RHS of (66)} \ge (1-\frac{d}{n})H_{\hat{Q}}(Y|X) - \frac{2|\mathcal{X}||\mathcal{Y}|}{n}\log\frac{n^2}{4|\mathcal{Y}|} - \frac{|\mathcal{X}||\mathcal{Y}|\log(n+1)}{n} \tag{72}$$

$$\ge (1-\alpha)H_{\hat{Q}}(Y|X) - \frac{2|\mathcal{X}||\mathcal{Y}|}{n}\log\frac{n^2}{4|\mathcal{Y}|} - \frac{|\mathcal{X}||\mathcal{Y}|\log(n+1)}{n} \tag{73}$$

$$= (1-\alpha)H_{\hat{Q}}(Y|X) - \epsilon_{2,n}. \tag{74}$$

Recall the definitions of $\epsilon_{1,n}$ and $\epsilon_{2,n}$ in Theorem 2.

Combining (65), (66), (71), and (74) yields that any rate pair $(R_1,R_2) \in (\mathcal{R}_X^{(n)} \times \mathcal{R}_Y^{(n)})$ with $R_1 \le \alpha H_{\hat{P}}(X|Y) - \epsilon_{1,n}$ and $R_2 \le (1-\alpha)H_{\hat{Q}}(Y|X) - \epsilon_{2,n}$, for any $0 \le \alpha \le 1$ and $(\hat{P}_{XY}, \hat{Q}_{XY})$ a pair of distributions such that $\alpha\hat{P}_{XY} + (1-\alpha)\hat{Q}_{XY} = T_{XY}$, is achievable as long as the condition in (68) holds.

We next consider the case $0 \le d < 4|\mathcal{X}||\mathcal{Y}|$. For this case, we have

$$\alpha H_{\hat{P}}(X|Y) \le \frac{d + |\mathcal{X}||\mathcal{Y}|}{n}\log|\mathcal{X}| \le \frac{5|\mathcal{X}||\mathcal{Y}|}{n}\log|\mathcal{X}| \le \varepsilon_{1,n},$$

where the first inequality follows by (67) and the fact that $H_{\hat{P}}(X|Y) \le \log|\mathcal{X}|$. Hence

$$\{(R_1, R_2) : R_1 \le \alpha H_{\hat{P}}(X|Y), R_2 \le (1-\alpha)H_{\hat{Q}}(Y|X)\} - [0, \varepsilon_{1,n}] \times [0, \varepsilon_{2,n}]$$

is empty, and so its intersection with $(\mathcal{R}_X^{(n)} \times \mathcal{R}_Y^{(n)})$ is also empty. Therefore, there is nothing to prove in this case. The case when $n - 4|\mathcal{X}||\mathcal{Y}| < d \le n$ can be handled similarly. This completes the proof for the inner bound.

Outer Bound: We next prove the outer bound by combining information-theoretic methods and linear algebra. Observe that the biclique rate region only depends on the probability values of T_{XY}, rather than the alphabets $\mathcal{X}, \mathcal{Y}$. With this in mind, we observe that we can identify $\mathcal{X}$ and $\mathcal{Y}$ with subsets of $\mathbb{R}$ by one-to-one mappings such that, for any probability distribution P_{XY}, if $(X,Y) \in \mathcal{X} \times \mathcal{Y}$ satisfies $(X,Y) \sim P_{XY}$ we can talk about the expectations $\mathbb{E}_P[X]$, $\mathbb{E}_P[Y]$, the covariance $\mathrm{Cov}_P(X,Y)$, and the correlation $\mathbb{E}_P[XY]$. Translating the choices of $\mathcal{X}$ and/or $\mathcal{Y}$ (as subsets of $\mathbb{R}$) does not change $\mathrm{Cov}_P(X,Y)$, so we can ensure that we make these choices in such a way that $\mathbb{E}_P[XY] = \mathrm{Cov}_P(X,Y) + \mathbb{E}_P[X]\mathbb{E}_P[Y] = 0$.

Let us now choose $\mathcal{X}, \mathcal{Y} \subseteq \mathbb{R}$ in this way, such that for the given joint n-type T_{XY} we have $E_T[XY] = 0$. Then, for $\mathcal{A} \times \mathcal{B} \subseteq \mathcal{T}_{T_{XY}}$, we will have $\langle \mathbf{x}, \mathbf{y} \rangle = 0$ for any $(\mathbf{x}, \mathbf{y}) \in \mathcal{A} \times \mathcal{B}$, where $\mathbf{x}, \mathbf{y}$ are now viewed as row vectors in $\mathbb{R}^n$. Let $\overline{\mathcal{A}}$ denote the linear space spanned by all the vectors in $\mathcal{A}$, and let $\overline{\mathcal{B}}$ denote the linear space spanned by all the vectors in $\mathcal{B}$. Hence $\overline{\mathcal{B}} \subseteq \overline{\mathcal{A}}^{\perp}$, where $\overline{\mathcal{A}}^{\perp}$ denotes the orthogonal complement of a subspace $\overline{\mathcal{A}}$. As an important property of the orthogonal complement, $\dim(\overline{\mathcal{A}}) + \dim(\overline{\mathcal{A}}^{\perp}) = n$. Hence $\dim(\overline{\mathcal{A}}) + \dim(\overline{\mathcal{B}}) \le n$.

We next establish the following exchange lemma. The proof is provided in Appendix D, and is based on the well-known exchange lemma in linear algebra.

Lemma 4: Let $\mathcal{V}_1, \mathcal{V}_2$ be mutually orthogonal linear subspaces of $\mathbb{R}^n$ with dimensions, denoted as n_1, n_2, satisfying $n_1 + n_2 = n$. Then there always exists a partition $\{\mathcal{J}_1, \mathcal{J}_2\}$ of $[n]$ such that $|\mathcal{J}_i| = n_i$ and $\mathbf{x} = f_i(\mathbf{x}_{\mathcal{J}_i}), \forall \mathbf{x} \in \mathcal{V}_i, i = 1,2$ for some deterministic linear functions $f_i : \mathbb{R}^{n_i} \to \mathbb{R}^n$, where $\mathbf{x}_{\mathcal{J}_i} := (x_j)_{j \in \mathcal{J}_i}$.

Remark 14: The natural generalization of this lemma also holds for k mutually orthogonal linear subspaces of $\mathbb{R}^n$ with total dimensions equal to n, and can be proved using Lemma 5 in Appendix D. Furthermore, the condition "mutually orthogonal linear subspaces of $\mathbb{R}^n$" can be replaced by "mutually (linearly) independent linear subspaces of $\mathbb{R}^n$" (i.e., such that the dimension of the span of the subspaces equals the sum of the dimensions of the subspaces), or, more generally, affine subspaces each of which is a translate of one of a mutually independent family of linear subspaces of $\mathbb{R}^n$.

Remark 15: In other words, under the assumption in this lemma there always exists a permutation σ of $[n]$ such that $\mathbf{x}^{(\sigma)} = f_1(\mathbf{x}^{(\sigma)}_{[1:n_1]}), \forall \mathbf{x} \in V_1$ and $\mathbf{x}^{(\sigma)} = f_2(\mathbf{x}^{(\sigma)}_{[n_1+1:n]}), \forall \mathbf{x} \in V_2$ for some deterministic functions $f_i : \mathbb{R}^{n_i} \to \mathbb{R}^n$, where $\mathbf{x}^{(\sigma)}$ is obtained by permuting the components of $\mathbf{x}$ using σ.

Let d denote $\dim(\overline{\mathcal{A}})$, so we have $\dim(\overline{\mathcal{A}}^{\perp}) = n - d$. Let $\mathbf{X} \sim \mathrm{Unif}(\mathcal{A}), \mathbf{Y} \sim \mathrm{Unif}(\mathcal{B})$ be two independent random vectors, i.e., $(\mathbf{X}, \mathbf{Y}) \sim P_{\mathbf{X},\mathbf{Y}} := \mathrm{Unif}(\mathcal{A})\mathrm{Unif}(\mathcal{B})$. Now we choose $V_1 = \overline{\mathcal{A}}, V_2 = \overline{\mathcal{A}}^{\perp}, \mathbf{X}_1 = \mathbf{X}, \mathbf{X}_2 = \mathbf{Y}$ in Lemma 4. Then there exists a partition $\{\mathcal{J}, \mathcal{J}^c\}$ of $[n]$ such that $|\mathcal{J}| = d$ and $\mathbf{X} = f_1(\mathbf{X}_{\mathcal{J}}), \mathbf{Y} = f_2(\mathbf{Y}_{\mathcal{J}^c})$ for some deterministic functions $f_1 : \mathbb{R}^d \to \mathbb{R}^n, f_2 : \mathbb{R}^{n-d} \to \mathbb{R}^n$. By this property, on the one hand we have

$$\begin{aligned} R_1 &= \frac{1}{n}H(\mathbf{X}) = \frac{1}{n}H(\mathbf{X}|\mathbf{Y}) = \frac{1}{n}H(\mathbf{X}_{\mathcal{J}}|\mathbf{Y}) \\ &\le \frac{1}{n}H(\mathbf{X}_{\mathcal{J}}|\mathbf{Y}_{\mathcal{J}}) \le \frac{1}{n}\sum_{j\in\mathcal{J}} H(X_j|Y_j) \\ &= \frac{d}{n}H(X_J|Y_J, J) \le \frac{d}{n}H(X_J|Y_J) = \frac{d}{n}H(\tilde{X}|\tilde{Y}), \end{aligned}$$

where $J \sim \mathrm{Unif}(\mathcal{J})$, $\tilde{X} := X_J, \tilde{Y} := Y_J$, with J being independent of $(\mathbf{X}, \mathbf{Y})$. Similarly, we have

$$\begin{aligned} R_2 &= \frac{1}{n}H(\mathbf{Y}) = \frac{1}{n}H(\mathbf{Y}|\mathbf{X}) = \frac{1}{n}H(\mathbf{Y}_{\mathcal{J}^c}|\mathbf{X}) \\ &\le \frac{1}{n}H(\mathbf{Y}_{\mathcal{J}^c}|\mathbf{X}_{\mathcal{J}^c}) \le \frac{1}{n}\sum_{j\in\mathcal{J}^c} H(Y_j|X_j) \\ &= (1-\frac{d}{n})H(Y_{\hat{J}}|X_{\hat{J}}, \hat{J}) \\ &\le (1-\frac{d}{n})H(Y_{\hat{J}}|X_{\hat{J}}) = (1-\frac{d}{n})H(\hat{Y}|\hat{X}), \end{aligned}$$

where $\hat{J} \sim \mathrm{Unif}(\mathcal{J}^c)$, $\hat{X} := X_{\hat{J}}, \hat{Y} := Y_{\hat{J}}$, with $\hat{J}$ being independent of $(\mathbf{X}, \mathbf{Y}, J)$. On the other hand,

$$\begin{aligned} &\frac{d}{n}P_{\tilde{X}\tilde{Y}} + (1-\frac{d}{n})P_{\hat{X}\hat{Y}} \\ &= \frac{1}{n}\sum_{j\in\mathcal{J}} P_{X_jY_j} + \frac{1}{n}\sum_{j\in\mathcal{J}^c} P_{X_jY_j} \\ &= \frac{1}{n}\sum_{j=1}^{n} P_{X_jY_j} = \mathbb{E}_{(\mathbf{X},\mathbf{Y})}[T_{\mathbf{XY}}] = T_{XY}, \end{aligned}$$

where $T_{\mathbf{XY}}$ denotes the joint type of a random pair $(\mathbf{X}, \mathbf{Y})$ which is hence also random (but equals T_{XY} pointwise). This completes the proof of the outer bound.

Appendix D
Proof of Lemma 4

For a pair of orthogonal subspaces $(\mathcal{V}, \mathcal{V}^{\perp})$ with dimensions respectively $n_1, n - n_1$, let $\{\mathbf{u}_j : 1 \le j \le n_1\}$ be an orthogonal basis of $\mathcal{V}$, and $\{\mathbf{u}_j : n_1 + 1 \le j \le n\}$ be an orthogonal basis of $\mathcal{V}^{\perp}$. Then $\{\mathbf{u}_j : 1 \le j \le n\}$ forms an orthogonal basis of $\mathbb{R}^n$. Denote by $\mathbf{U}$ the $n \times n$ matrix with j-th row being $\mathbf{u}_j$. Then $\mathbf{U}$ is orthogonal. We now express $\mathbf{x} \in \mathcal{V}$ and $\mathbf{y} \in \mathcal{V}^{\perp}$, thought of as row vectors, in terms of this orthogonal basis, i.e.,

$$\mathbf{x} = \hat{\mathbf{x}}\mathbf{U}, \qquad \mathbf{y} = \hat{\mathbf{y}}\mathbf{U}, \tag{75}$$

where $\hat{\mathbf{x}} := \mathbf{x}\mathbf{U}^{\top}, \hat{\mathbf{y}} := \mathbf{y}\mathbf{U}^{\top}$, and $\mathbf{U}^{\top}$ is the transpose of $\mathbf{U}$. Since for any $\mathbf{x} \in \mathcal{V}$ we have $\langle \mathbf{x}, \mathbf{u}_j \rangle = 0$ for all $n_1 + 1 \le j \le n$, we obtain that $\hat{x}_j = 0$ for all $n_1 + 1 \le j \le n$. Similarly,

$\hat{y}_j = 0$ for all $1 \le j \le n_1$. Hence we can rewrite $\hat{\mathbf{x}} = (\hat{\mathbf{x}}_1, \mathbf{0}), \hat{\mathbf{y}} = (\mathbf{0}, \hat{\mathbf{y}}_2)$. We write $\mathbf{U}$ in a block form: $\mathbf{U} = \begin{bmatrix}\mathbf{U}_1\\ \mathbf{U}_2\end{bmatrix}$ where $\mathbf{U}_1, \mathbf{U}_2$ are respectively of size $n_1 \times n, (n-n_1) \times n$. Then

$$\mathbf{x} = \hat{\mathbf{x}}_1 \mathbf{U}_1, \qquad \mathbf{y} = \hat{\mathbf{y}}_2 \mathbf{U}_2. \tag{76}$$

We now need the following well-known exchange lemma.

Lemma 5 (Exchange Lemma): [65, Theorem 3.2] Let $k \ge 2$ be an integer. Let $\mathbf{B}$ be an $n \times n$ nonsingular matrix, and $\{\mathcal{H}_l, 1 \le l \le k\}$ be a partition of $[n]$. Then there always exists another partition $\{\mathcal{L}_l, 1 \le l \le k\}$ of $[n]$ with $|\mathcal{L}_l| = |\mathcal{H}_l|$ such that all the sub-matrices $\mathbf{B}_{\mathcal{H}_l, \mathcal{L}_l}, 1 \le l \le k$ are nonsingular.

The proof of this lemma follows easily from repeated use of the Laplace expansion for determinants. A short proof in the case $k = 2$, which is the only case we use, goes as follows. Let $\mathbf{B} = (b_{i,j})$ be an $n \times n$ matrix and $\mathcal{H}$ a subset of $[n]$. Then the determinant of $\mathbf{B}$ can be expanded as follows:

$$\det(\mathbf{B}) = \sum_{\mathcal{L} \subseteq [n] : |\mathcal{L}| = |\mathcal{H}|} \varepsilon^{\mathcal{H},\mathcal{L}} \det(\mathbf{B}_{\mathcal{H},\mathcal{L}}) \det(\mathbf{B}_{\mathcal{H}^c,\mathcal{L}^c})$$

where $\varepsilon^{\mathcal{H},\mathcal{L}}$ is the sign of the permutation determined by $\mathcal{H}$ and $\mathcal{L}$, equal to $(-1)^{(\sum_{h\in\mathcal{H}} h) + (\sum_{\ell \in \mathcal{L}} \ell)}$. Since $\mathbf{B}$ is nonsingular, there must be at least one choice of $|\mathcal{L}|$, with $|\mathcal{L}| = |\mathcal{H}|$, such that both $\mathbf{B}_{\mathcal{H},\mathcal{L}}$ and $\mathbf{B}_{\mathcal{H}^c,\mathcal{L}^c}$ are nonsingular, which is what is being claimed.

Substituting $\mathbf{B} \leftarrow \mathbf{U}, \mathcal{H}_1 \leftarrow [n_1], \mathcal{H}_2 \leftarrow [n_1+1:n]$ in this lemma, we obtain that there exists a partition $\{\mathcal{J}, \mathcal{J}^c\}$ of $[n]$ with $|\mathcal{J}| = n_1$ such that both the sub-matrices $\mathbf{U}_{[n_1],\mathcal{J}}, \mathbf{U}_{[n_1+1:n],\mathcal{J}^c}$ are nonsingular. Denote $\mathbf{U}_{1,\mathcal{J}}$ as the submatrix of $\mathbf{U}_1$ consisting of $\mathcal{J}$- indexed columns of $\mathbf{U}_1$, and define $\mathbf{U}_{1,\mathcal{J}^c}, \mathbf{U}_{2,\mathcal{J}}, \mathbf{U}_{2,\mathcal{J}^c}$ similarly. Then, by definition, $\mathbf{U}_{1,\mathcal{J}} = \mathbf{U}_{[n_1],\mathcal{J}}, \mathbf{U}_{2,\mathcal{J}^c} = \mathbf{U}_{[n_1+1:n],\mathcal{J}^c}$. Therefore, from (76), we have

$$\hat{\mathbf{x}}_1 = \mathbf{x}_{\mathcal{J}} \mathbf{U}_{1,\mathcal{J}}^{-1}, \qquad \hat{\mathbf{y}}_2 = \mathbf{y}_{\mathcal{J}^c} \mathbf{U}_{2,\mathcal{J}^c}^{-1}.$$

Substituting these back into (76), we obtain that

$$\begin{aligned}(\mathbf{x}_{\mathcal{J}}, \mathbf{x}_{\mathcal{J}^c}) &= \mathbf{x}_{\mathcal{J}} \mathbf{U}_{1,\mathcal{J}}^{-1} (\mathbf{U}_{1,\mathcal{J}}, \mathbf{U}_{1,\mathcal{J}^c}) \\ &= (\mathbf{x}_{\mathcal{J}}, \mathbf{x}_{\mathcal{J}} \mathbf{U}_{1,\mathcal{J}}^{-1} \mathbf{U}_{1,\mathcal{J}^c})\end{aligned}$$

and

$$(\mathbf{y}_{\mathcal{J}}, \mathbf{y}_{\mathcal{J}^c}) = (\mathbf{y}_{\mathcal{J}^c} \mathbf{U}_{2,\mathcal{J}^c}^{-1} \mathbf{U}_{2,\mathcal{J}}, \mathbf{y}_{\mathcal{J}^c}).$$

Hence the proof is completed.

Appendix E
Proof of Proposition 2

From Theorem 2, we know that $\mathcal{R}(T_{XY}) = \mathcal{R}^*(T_{XY})$, where $\mathcal{R}(T_{XY})$ is the asymptotic biclique rate region, defined in (7) and $\mathcal{R}^*(T_{XY})$ is defined in Theorem 2. Furthermore, $\mathcal{R}^*(T_{XY})$ is a closed convex set (see Proposition 1). Hence

$$\mathcal{R}(T_{XY}) = \mathcal{R}_{\triangle}(T_{XY})$$

if and only if

$$\max_{\substack{0 \le \alpha \le 1, P_{XY}, Q_{XY}: \\ \alpha P_{XY} + (1-\alpha) Q_{XY} = T_{XY}}} \varphi_\alpha(P_{XY}, Q_{XY}) \le 1, \tag{77}$$

where $\varphi_\alpha(P_{XY}, Q_{XY}) := \frac{\alpha}{\beta_1} H_P(X|Y) + \frac{1-\alpha}{\beta_2} H_Q(Y|X)$ with $\beta_1 := H_T(X|Y), \beta_2 := H_T(Y|X)$. Here the domain of φ_α can be taken to be the set of pairs of probability distributions (P_{XY}, Q_{XY}) such that $\mathrm{supp}(P_{XY}) = \mathrm{supp}(Q_{XY}) \subseteq \mathrm{supp}(T_{XY})$. Moreover, (77) can be rewritten as that for any $0 \le \alpha \le 1$,

$$\max_{P_{XY}, Q_{XY} : \alpha P_{XY} + (1-\alpha) Q_{XY} = T_{XY}} \varphi_\alpha(P_{XY}, Q_{XY}) \le 1. \tag{78}$$

Observe that $\varphi_\alpha(T_{XY}, T_{XY}) = 1$. Hence (78) can be rewritten as that for any $0 < \alpha < 1$, $P_{XY} = Q_{XY} = T_{XY}$ is an optimal solution to the LHS of (78). Next we study for what kind of T_{XY} it holds for all $0 < \alpha < 1$ that $P_{XY} = Q_{XY} = T_{XY}$ is an optimal solution to the LHS of (78).

Given $0 < \alpha < 1$, observe that $\alpha P_{XY} + (1-\alpha) Q_{XY}$ is linear in (P_{XY}, Q_{XY}), and $\varphi_\alpha(P_{XY}, Q_{XY})$ is concave in (P_{XY}, Q_{XY}) (which can be shown by the log sum inequality [66, Theorem 2.7.1]). Hence the LHS of (78) is a linearly-constrained convex optimization problem. This means that showing that the pair (T_{XY}, T_{XY}) is an extremum for this convex optimization problem iff T_{XY} satisfies the conditions given in Corollary 2, is equivalent to establishing that (T_{XY}, T_{XY}) is an optimum for the convex optimization problem (thus establishing (78) for $0 < \alpha < 1$) iff T_{XY} satisfies the conditions given in Corollary 2. Since the notion of extremality is local, to show this it suffices to consider the modified version of this convex optimization problem where the domain of φ_α is taken to be the set of pairs of probability distributions (P_{XY}, Q_{XY}) such that $\mathrm{supp}(P_{XY}) = \mathrm{supp}(Q_{XY}) = \mathrm{supp}(T_{XY})$.

We are thus led to consider the Lagrangian

$$\begin{aligned} L = \varphi_\alpha(P_{XY}, Q_{XY}) & \\ + \sum_{(x,y)\in\mathrm{supp}(T_{XY})} & \eta(x,y)\big(\alpha P(x,y) \\ & + (1-\alpha) Q(x,y) - T(x,y)\big) \\ + \mu_1 \Big(& \sum_{(x,y)\in\mathrm{supp}(T_{XY})} P(x,y) - 1\Big) \\ + \mu_2 \Big(& \sum_{(x,y)\in\mathrm{supp}(T_{XY})} Q(x,y) - 1\Big). \end{aligned}$$

By checking the feasible solution (P_{XY}, Q_{XY}) with $P_{XY} = Q_{XY} = T_{XY}$, one can find that Slater's condition for the modified version of the convex optimization problem in (78) (described above) is satisfied, which implies that extrema of the modified version of the optimization problem in (78) are given by the Karush–Kuhn–Tucker (KKT) conditions:

$$\begin{aligned}\frac{\partial L}{\partial P(x,y)} &= -\frac{\alpha}{\beta_1} \log P(x|y) + \alpha \eta(x,y) + \mu_1 \\ &= 0, \forall (x,y) \in \mathrm{supp}(T_{XY}),\end{aligned} \tag{79}$$

$$\begin{aligned}\frac{\partial L}{\partial Q(x,y)} &= -\frac{1-\alpha}{\beta_2} \log Q(y|x) + (1-\alpha) \eta(x,y) + \mu_2 \\ &= 0, \forall (x,y) \in \mathrm{supp}(T_{XY}),\end{aligned} \tag{80}$$

$$\alpha P(x,y) + (1-\alpha)Q(x,y) = T(x,y), \quad \forall (x,y) \in \mathrm{supp}(T_{XY}), \tag{81}$$

$$\sum_{(x,y)\in\mathrm{supp}(T_{XY})} P(x,y) = 1, \tag{82}$$

$$\sum_{(x,y)\in\mathrm{supp}(T_{XY})} Q(x,y) = 1, \tag{83}$$

$$P(x,y), Q(x,y) > 0, \forall (x,y) \in \mathrm{supp}(T_{XY}), \tag{84}$$

for some reals $\eta(x,y), \mu_1, \mu_2$ with $(x,y) \in \mathrm{supp}(T_{XY})$. Here the conditions in (84) come from the restriction we have imposed on the domain of φ_α.

We first prove "if" part. That is, for T_{XY} satisfying the conditions given in Corollary 2, given any $0 < \alpha < 1$, $P_{XY} = Q_{XY} = T_{XY}$ together with some reals $\eta(x,y), \mu_1, \mu_2$ must satisfy (79)-(84). To this end, we choose $\eta(x,y) = \frac{1}{\beta_1}\log T(x|y) = \frac{1}{\beta_2}\log T(y|x)$, $\mu_1 = \mu_2 = 0$, which satisfy (79) and (80).

We next consider the "only if" part. Substituting $P = Q = T$ and taking expectations with respect to the type T_{XY} for the both sides of (79) and (80), we obtain that

$$\frac{\mu_1}{\alpha} = \frac{\mu_2}{1-\alpha}. \tag{85}$$

Substituting this back to (79) and (80) yields that $T_{X|Y}(x|y)^{1/H_T(X|Y)} = T_{Y|X}(y|x)^{1/H_T(Y|X)}$ for all x, y.

Appendix F
Proof of Lemma 3

The two equalities above can be verified easily. Here we only prove the inequality above. Without loss of generality, we assume $0 \le \beta \le \alpha \le \frac{1}{2}$. Then, in the definition of $\mathsf{D}(\alpha,\beta)$, we minimize $D_{\alpha,\beta}(p)$ over $0 \le p \le \beta$. Furthermore,

$$\begin{aligned} D_{\alpha,\beta}(p) = &-H(p, \alpha-p, \beta-p, 1+p-\alpha-\beta) \\ &- (1+2p-\alpha-\beta)\log(1+\rho) \\ &- (\alpha+\beta-2p)\log(1-\rho) + \log 4. \end{aligned}$$

Let $s := \alpha+\beta-2p$. Then we have

$$\begin{aligned} D_{\alpha,\beta}(p) = &-H(p, \alpha-p, \beta-p, 1+p-\alpha-\beta)|_{p=\frac{\alpha+\beta-s}{2}} \\ &- (1-s)\log(1+\rho) - s\log(1-\rho) + \log 4. \end{aligned}$$

By definition, $\mathsf{D}(\alpha,\beta)$ can be rewritten as the minimum of $D_{\alpha,\beta}(p)$ over $\alpha-\beta \le s \le \alpha+\beta$. Given (α,β), $D_{\alpha,\beta}(p)$ is convex in s which follows by the convexity of the relative entropy. Moreover, $H(p, \alpha-p, \beta-p, 1+p-\alpha-\beta)$ is maximized at $p = \alpha\beta$, i.e., at $s = \alpha+\beta-2\alpha\beta$. Hence, the derivative of $D_{\alpha,\beta}(p)$ w.r.t. s at $s = \alpha+\beta-2\alpha\beta$ is $\log\frac{1+\rho}{1-\rho}$, which is nonnegative. This implies that the minimum of $D_{\alpha,\beta}(p)$ is attained at some point s such that $\alpha-\beta \le s \le \alpha+\beta-2\alpha\beta$ (or equivalently, at some $p \in (\alpha\beta, \beta]$). In other words, without changing the value of $\mathsf{D}(\alpha,\beta)$, one can replace $H(p, \alpha-p, \beta-p, 1+p-\alpha-\beta)$ above with

$$\tilde{H}(p, \alpha-p, \beta-p, 1+p-\alpha-\beta) := \begin{cases} H(p, \alpha-p, \beta-p, \\ \quad 1+p-\alpha-\beta), & p \in (\alpha\beta, \beta], \\ h(\alpha)+h(\beta), & p \in (-\infty, \alpha\beta]. \end{cases}$$

That is, $\mathsf{D}(\alpha,\beta)$ is equal to the minimum of

$$\begin{aligned} &-\tilde{H}(p, \alpha-p, \beta-p, 1+p-\alpha-\beta)|_{p=\frac{\alpha+\beta-s}{2}} \\ &- (1-s)\log(1+\rho) - s\log(1-\rho) + \log 4 \end{aligned}$$

over $s \ge \alpha-\beta$.

We next deal with $\mathsf{D}(1-\alpha,\beta)$. In the definition of $\mathsf{D}(1-\alpha,\beta)$, we minimize $D_{1-\alpha,\beta}(p)$ over the same range $0 \le p \le \beta$. Furthermore,

$$\begin{aligned} D_{1-\alpha,\beta}(p) = &-H(p, 1-\alpha-p, \beta-p, \alpha+p-\beta) \\ &- (\alpha+2p-\beta)\log(1+\rho) \\ &- (1-\alpha+\beta-2p)\log(1-\rho) + \log 4. \end{aligned}$$

Let $t := 1-\alpha+\beta-2p$. Then, similarly to the above, we have

$$\begin{aligned} D_{1-\alpha,\beta}(p) = &-H(p, \alpha-p, \beta-p, \\ &\quad 1+p-\alpha-\beta)|_{p=\frac{\alpha+\beta-(1-t)}{2}} \\ &- (1-t)\log(1+\rho) - t\log(1-\rho) + \log 4, \end{aligned}$$

which can be seen by verifying that

$$\begin{aligned} &-H(p, \alpha-p, \beta-p, 1+p-\alpha-\beta)|_{p=\frac{\alpha+\beta-(1-t)}{2}} \\ &= -H(p, 1-\alpha-p, \beta-p, \alpha+p-\beta)|_{p=\frac{1-\alpha+\beta-t}{2}}. \end{aligned}$$

Hence, $\mathsf{D}(1-\alpha,\beta)$ is equal to the minimum of $D_{1-\alpha,\beta}(p)$ over $1-\alpha-\beta \le t \le 1-\alpha+\beta$. Given (α,β), $D_{1-\alpha,\beta}(p)$ is convex in t. Moreover, $H(p, \alpha-p, \beta-p, 1+p-\alpha-\beta)$ is maximized at $p = \alpha\beta$, i.e., at $t = 1-\alpha-\beta+2\alpha\beta$. Hence the derivative of $D_{1-\alpha,\beta}(p)$ w.r.t. t at $t = 1-\alpha-\beta+2\alpha\beta$ is still $\log\frac{1+\rho}{1-\rho}$ which is nonnegative. Hence $\mathsf{D}(1-\alpha,\beta)$ is equal to the minimum of $D_{1-\alpha,\beta}(p)$ over $1-\alpha-\beta \le t \le 1-\alpha-\beta+2\alpha\beta$.

To prove $\mathsf{D}(1-\alpha,\beta) \ge \mathsf{D}(\alpha,\beta)$, it suffices to show that

$$\begin{aligned} &-H(p, \alpha-p, \beta-p, 1+p-\alpha-\beta)|_{p=\frac{\alpha+\beta-(1-s)}{2}} \\ &\ge -\tilde{H}(p, \alpha-p, \beta-p, 1+p-\alpha-\beta)|_{p=\frac{\alpha+\beta-s}{2}} \end{aligned} \tag{86}$$

for all $1-\alpha-\beta \le s \le 1-\alpha-\beta+2\alpha\beta$. By the definition of $\tilde{H}$, we only need to check

$$\begin{aligned} g(s) := &H(p, \alpha-p, \beta-p, 1+p-\alpha-\beta)|_{p=\frac{\alpha+\beta-s}{2}} \\ &- H(p, \alpha-p, \beta-p, \\ &\quad 1+p-\alpha-\beta)|_{p=\frac{\alpha+\beta-(1-s)}{2}} \\ \ge\ &0 \end{aligned} \tag{87}$$

for $1-\alpha-\beta \le s \le \alpha+\beta-2\alpha\beta$. If $\alpha+\beta-2\alpha\beta < 1-\alpha-\beta$ there is nothing to show. We may therefore assume that $1-\alpha-\beta \le \alpha+\beta-2\alpha\beta$.

Computing the derivative of g, we see that g is nonincreasing on $[1-\alpha-\beta, \alpha+\beta-2\alpha\beta]$. On the other hand, observe that $g(\alpha+\beta-2\alpha\beta) \ge 0$ since the maximum of the first entropy in (87) is attained at $s = \alpha+\beta-2\alpha\beta$. Hence, we have $g \ge 0$ on $[1-\alpha-\beta, \alpha+\beta-2\alpha\beta]$. This completes the proof.

Remark 16: Although the proof above seems complicated, the intuition behind it is simple. Observe that $\mathsf{D}(\alpha,\beta)$ is equal to the asymptotic exponent of $P_{XY}^n(\mathcal{A}_n \times \mathcal{B}_n)$ where

$\mathcal{A}_n, \mathcal{B}_n$ are type classes with types asymptotically converging to $(\alpha, 1-\alpha)$ and $(\beta, 1-\beta)$ respectively. In other words, $\mathcal{A}_n, \mathcal{B}_n$ are the concentric Hamming spheres with common center $(0, 0, \ldots, 0)$ with radii r_n, s_n satisfying $r_n/n \to \alpha, s_n/n \to \beta$ as $n \to \infty$. Similarly, $\mathsf{D}(1-\alpha, \beta)$ is equal to the asymptotic exponent of $P_{XY}^n(\hat{\mathcal{A}}_n \times \mathcal{B}_n)$ where $\hat{\mathcal{A}}_n$ is the anti-concentric Hamming sphere of $\mathcal{A}_n$. Hence, the type of $\hat{\mathcal{A}}_n$ converges to $(1-\alpha, \alpha)$ asymptotically. On the other hand, we can write for $\mathbf{y} \in \mathcal{B}_n$,

$$P_{X|Y}^n(\mathcal{A}_n|\mathbf{y}) = (\frac{1+\rho}{2})^n \sum_{\mathbf{x} \in \mathcal{A}_n} (\frac{1-\rho}{1+\rho})^{d(\mathbf{x},\mathbf{y})}.$$

By permutation, one can observe that the expression above remains the same for all $\mathbf{y} \in \mathcal{B}_n$. Hence,

$$P_{XY}^n(\mathcal{A}_n \times \mathcal{B}_n) = P_Y^n(\mathcal{B}_n)(\frac{1+\rho}{2})^n \sum_{\mathbf{x} \in \mathcal{A}_n} (\frac{1-\rho}{1+\rho})^{d(\mathbf{x},\mathbf{y})}.$$

Denote $\eta := \frac{1-\rho}{1+\rho}$ and denote F as the CDF of the distance $d(\mathbf{X}, \mathbf{y})$ with $\mathbf{X} \sim \mathrm{Unif}(\mathcal{A}_n)$. Then, we have

$$\begin{aligned}\frac{1}{|\mathcal{A}_n|} \sum_{\mathbf{x} \in \mathcal{A}_n} \eta^{d(\mathbf{x},\mathbf{y})} &= \mathbb{E}_{\mathbf{X} \sim \mathrm{Unif}(\mathcal{A}_n)} \eta^{d(\mathbf{X},\mathbf{y})} \\ &= \sum_{d=0}^{\infty} (F(d) - F(d-1))\eta^d \\ &= \sum_{d=0}^{\infty} F(d)(\eta^d - \eta^{d+1}). \end{aligned} \quad (88)$$

Similarly,

$$\frac{1}{|\hat{\mathcal{A}}_n|} \sum_{\mathbf{x} \in \hat{\mathcal{A}}_n} \eta^{d(\mathbf{x},\mathbf{y})} = \sum_{d=0}^{\infty} G(d)(\eta^d - \eta^{d+1}), \quad (89)$$

where G is the CDF of the distance $d(\mathbf{X}, \mathbf{y})$ with $\mathbf{X} \sim \mathrm{Unif}(\hat{\mathcal{A}}_n)$. Since the sphere $\mathcal{A}_n$ is "closer" to $\mathbf{y} \in \mathcal{B}_n$ than the sphere $\hat{\mathcal{A}}_n$, intuitively, $F(d) \geq G(d)$ for all $d \geq 0$ which implies $P_{XY}^n(\mathcal{A}_n \times \mathcal{B}_n) \geq P_{XY}^n(\hat{\mathcal{A}}_n \times \mathcal{B}_n)$. This further implies $\mathsf{D}(1-\alpha, \beta) \geq \mathsf{D}(\alpha, \beta)$. In the proof above, we showed that the asymptotic exponent of (88) is not larger than that of (89), with ns, nt denoting the distances. This is a weaker version of $P_{XY}^n(\mathcal{A}_n \times \mathcal{B}_n) \geq P_{XY}^n(\hat{\mathcal{A}}_n \times \mathcal{B}_n)$, but it still implies $\mathsf{D}(1-\alpha, \beta) \geq \mathsf{D}(\alpha, \beta)$.

APPENDIX G
PROOF OF THEOREM 6

Our proof combines Theorem 4 with ideas from [18, Proof of Theorem 1.8]. Observe that by the product construction, the optimal exponents

$$\underline{\Lambda}_{p,q}^{(n)}(\alpha, \beta) := -\frac{1}{n} \log \sup_{\substack{f,g: P_X^n(\mathrm{supp}(f)) \leq 2^{-n\alpha}, \\ P_Y^n(\mathrm{supp}(g)) \leq 2^{-n\beta}}} \frac{\langle f, g \rangle}{\|f\|_p \|g\|_q}, \quad (90)$$

$$\overline{\Lambda}_{p,q}^{(n)}(\alpha, \beta) := -\frac{1}{n} \log \inf_{\substack{f,g: P_X^n(\mathrm{supp}(f)) \leq 2^{-n\alpha}, \\ P_Y^n(\mathrm{supp}(g)) \leq 2^{-n\beta}}} \frac{\langle f, g \rangle}{\|f\|_p \|g\|_q}, \quad (91)$$

satisfy that $n\underline{\Lambda}_{p,q}^{(n)}(\alpha, \beta)$ is subadditive and $n\overline{\Lambda}_{p,q}^{(n)}(\alpha, \beta)$ is superadditive in n. So, by Fekete's lemma, $\inf_{n \geq 1} \underline{\Lambda}_{p,q}^{(n)}(\alpha, \beta) = \lim_{n \to \infty} \underline{\Lambda}_{p,q}^{(n)}(\alpha, \beta)$ and $\sup_{n \geq 1} \overline{\Lambda}_{p,q}^{(n)}(\alpha, \beta) = \lim_{n \to \infty} \overline{\Lambda}_{p,q}^{(n)}(\alpha, \beta)$, which means that we only need to focus on the asymptotic case.

We may assume, by homogeneity, that $\|f\|_1 = \|g\|_1 = 1$. This means that $f \leq 1/P_{X,\min}^n, g \leq 1/P_{Y,\min}^n$, and moreover, $\frac{1}{n} \log \|f\|_p$ and $\frac{1}{n} \log \|g\|_q$ are uniformly bounded for all $n \geq 1$. This is because given $\|f\|_1 = 1$, for $p \geq 1$, we have

$$1 = \|f\|_1 \leq \|f\|_p \leq \|f\|_\infty \leq 1/P_{X,\min}^n, \quad (92)$$

and for $0 < p \leq 1$, we have

$$P_{X,\min}^{n(1-p)/p} \leq \|f\|_p \leq \|f\|_1 = 1. \quad (93)$$

For sufficiently large $a > 0$, the points at which f or $g < 2^{-na}$ contribute little to $\|f\|_p$, $\|g\|_q$, and $\langle f, g \rangle$, in the sense that if we set f, g to be zero at these points (the resulting functions denoted as f_a, g_a), then $\frac{1}{n} \log \|f\|_p$, $\frac{1}{n} \log \|g\|_q$, and $\frac{1}{n} \log \langle f, g \rangle$ only change by amounts of the order of $o_n(1)$, where $o_n(1)$ denotes a term vanishing as $n \to \infty$ uniformly over all f and g with $\|f\|_1 = \|g\|_1 = 1$. This is because,

$$\|f_a\|_p^p \leq \|f\|_p^p \leq \|f_a\|_p^p + 2^{-npa},$$

and

$$\mathbb{E}[f_a g_a] \leq \mathbb{E}[fg] \leq \mathbb{E}[f_a g_a] + 3 \cdot 2^{-an}.$$

All the remaining points of $\mathcal{X}^n$ can be partitioned into $r = r(a, b)$ level sets $\mathcal{A}_1, \ldots, \mathcal{A}_r$ such that f varies by a factor of at most 2^{nb} in each level set, where $b > 0$. Similarly, all the remaining points of $\mathcal{Y}^n$ can be partitioned into $s = s(a, b)$ level sets $\mathcal{B}_1, \ldots, \mathcal{B}_s$ such that g varies by a factor of at most 2^{nb} in each level set. Let $\alpha_i := -\frac{1}{n} \log P_X^n(\mathcal{A}_i), \beta_i := -\frac{1}{n} \log P_Y^n(\mathcal{B}_i)$, and let $\mu_i = \frac{1}{n} \log(u_i), \nu_i = \frac{1}{n} \log(v_i)$, where u_i, v_i are respectively the median value of f on $\mathcal{A}_i$ and the median value of g on $\mathcal{B}_i$. (If $\mathcal{A}_i$ is empty then u_i can be chosen to be any value within the level set defining $\mathcal{A}_i$, and similarly for $\mathcal{B}_i$ and v_i.) Note that $f(\mathbf{x}) \in [u_i 2^{-nb}, u_i 2^{nb}]$ on the set $\mathcal{A}_i$ and $g(\mathbf{y}) \in [v_i 2^{-nb}, v_i 2^{nb}]$ on the set $\mathcal{B}_i$. Moreover, $\alpha_i \geq \alpha, \beta_j \geq \beta, \forall i, j$. Then,

$$\begin{aligned}\frac{1}{n} \log \|f\|_p &\geq \frac{1}{n} \log \|f_a\|_p \\ &\geq \frac{1}{np} \log[\sum_{i=1}^r P_X^n(\mathcal{A}_i) u_i^p] - b \\ &\geq N_X(p) - b, \end{aligned}$$

where $N_X(p) := \max_{1 \leq i \leq r}\{-\frac{\alpha_i}{p} + \mu_i\}$. Similarly,

$$\frac{1}{n} \log \|g\|_q \geq N_Y(q) - b,$$

where $N_Y(q) := \max_{1 \leq i \leq s}\{-\frac{\beta_i}{q} + \nu_i\}$.

Utilizing these equations, we obtain

$$\begin{aligned}&\frac{1}{n} \log \langle f, g \rangle \\ &\leq \frac{1}{n} \log[\langle f_a, g_a \rangle + 3 \cdot 2^{-an}] \\ &\leq \frac{1}{n} \log[\sum_{i=1}^r \sum_{j=1}^s P_{XY}^n(\mathcal{A}_i \times \mathcal{B}_j) u_i v_j \cdot 2^{2nb} + 3 \cdot 2^{-an}] \\ &\leq \frac{1}{n} \log[rs \cdot 2^{n(\max_{1 \leq i \leq r, 1 \leq j \leq s}\{-\underline{\Theta}^*(\alpha_i, \beta_j) + \mu_i + \nu_j\} + 2b)} \end{aligned}$$

$$+ 3 \cdot 2^{-an}]$$
$$\leq \frac{1}{n}\log[rs \cdot 2^{n\max_{1\leq i\leq r,1\leq j\leq s}\{-\underline{\Theta}^*(\alpha_i,\beta_j)+\frac{\alpha_i}{p}+\frac{\beta_j}{q}\}} \cdot 2^{n(N_X(p)+N_Y(q)+2b)} + 3\cdot 2^{-an}].$$

Combining the inequalities above, we have

$$\frac{1}{n}\log\frac{\langle f,g\rangle}{\|f\|_p\|g\|_q}$$
$$\leq \frac{1}{n}\log[rs\cdot 2^{n(-\underline{\Lambda}^*_{p,q}(\alpha,\beta)+N_X(p)+N_Y(q)+2b)} \cdot 2^{-\log(\|f\|_p\|g\|_q)} + \frac{3\cdot 2^{-an}}{\|f\|_p\|g\|_q}]$$
$$\leq \frac{1}{n}\log[rs\cdot 2^{n(-\underline{\Lambda}^*_{p,q}(\alpha,\beta)+4b)} + \frac{3\cdot 2^{-an}}{\|f\|_p\|g\|_q}]. \quad (94)$$

From (92), we know that if we choose a sufficiently large, the (negative) exponent of the second term in the last line above can be arbitrarily large. On the other hand, if a, b are fixed, then r, s are also fixed. Hence, (94) is upper bounded by

$$-\underline{\Lambda}^*_{p,q}(\alpha,\beta) + 4b + o_n(1).$$

Letting $n\to\infty$ and then $b\to 0$, we obtain (55).

We next prove (56). First, observe that

$$\frac{1}{n}\log\|f\|_p \leq \frac{1}{n}\log\|f_a\|_p + \frac{1}{np}\log[1+\frac{2^{-npa}}{\|f_a\|_p^p}]$$
$$\leq \frac{1}{np}\log[\sum_{i=1}^{r} P_X^n(\mathcal{A}_i)u_i^p] + b + \frac{1}{np}\log[1+\frac{2^{-npa}}{\|f_a\|_p^p}]$$
$$\leq N_X(p) + b + \epsilon_n,$$

where $\epsilon_n := \frac{1}{np}\log r + \frac{1}{np}\log[1+\frac{2^{-npa}}{\|f_a\|_p^p}]$, which tends to zero as $n\to\infty$ for large enough a and any fixed b

Similarly, we have

$$\frac{1}{n}\log\|g\|_q \leq N_Y(q) + b + \hat{\epsilon}_n,$$

where $\hat{\epsilon}_n := \frac{1}{nq}\log s + \frac{1}{nq}\log[1+\frac{2^{-nqa}}{\|g_a\|_q^q}]$, which tends to zero as $n\to\infty$ for large enough a and any fixed b.

On the other hand,

$$\frac{1}{n}\log\langle f,g\rangle$$
$$\geq \frac{1}{n}\log\langle f_a,g_a\rangle$$
$$\geq \frac{1}{n}\log\mathbb{E}[\sum_{i=1}^{r}\sum_{j=1}^{s} P_{XY}^n(\mathcal{A}_i\times\mathcal{B}_j)u_iv_j\cdot 2^{-2nb}]$$
$$\geq \max_{1\leq i\leq r,1\leq j\leq s}\{-\overline{\Theta}^*(\alpha_i,\beta_j)+\mu_i+\nu_j\} - 2b \quad (95)$$
$$\geq -\overline{\Theta}^*(\alpha_{i^*},\beta_{j^*}) + \mu_{i^*} + \nu_{j^*} - 2b \quad (96)$$
$$= -\overline{\Theta}^*(\alpha_{i^*},\beta_{j^*}) + \frac{\alpha_{i^*}}{p} + \frac{\beta_{j^*}}{q} + N_X(p) + N_Y(q) - 2b$$
$$\geq \overline{\Lambda}^*_{p,q}(\alpha,\beta) + N_X(p) + N_Y(q) - 2b,$$

where (95) follows from Theorem 4, with the maximum being taken only over those pairs (i,j) for which $P_X^n(\mathcal{A}_i) > 0$ and $P_Y^n(\mathcal{B}_j) > 0$, since $\overline{\Theta}^*(\alpha_i,\beta_j)$ is defined only for $\alpha_i \in [0, E_{1,\max}]$, $\beta_j \in [0, E_{2,\max}]$; also, in (96), i^* is defined as the optimal i attaining $N_X(p)$ and j^* as the optimal j attaining $N_Y(q)$. Combining the inequalities above, we have

$$\frac{1}{n}\log\frac{\langle f,g\rangle}{\|f\|_p\|g\|_q} \geq \overline{\Lambda}^*_{p,q}(\alpha,\beta) - 4b - \epsilon_n - \hat{\epsilon}_n. \quad (97)$$

We first choose a sufficiently large, fix a, b, and let $n\to\infty$. We have both $\epsilon_n, \hat{\epsilon}_n \to 0$. We then let $b\to 0$, and hence we obtain (56).

Acknowledgment

The authors are grateful to Or Ordentlich, Yury Polyanskiy, and Ofer Shayevitz for sharing their code and the details about their calculation in [17] to help them find out and fix an error in the previous version of Theorem 5. They also thank Amin Gohari and Sandeep Pradhan for pointing out the related references [2], [9], [10] to them. They would also like to thank anonymous reviewers for pointing out related references, and especially thank one of reviewers for pointing out that the strong small-set theorem is not new, and in fact it is a direct consequence of the information-theoretic characterization of Brascamp–Lieb inequalities.

References

[1] L. Yu, V. Anantharam, and J. Chen, "Type graphs and small-set expansion," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2021, pp. 993–998.
[2] A. Nazari, S. S. Pradhan, and A. Anastasopoulos, "New bounds on the maximal error exponent for multiple-access channels," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2009, pp. 1704–1708.
[3] I. Csiszar and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Cambridge, U.K.: Cambridge Univ. Press, 2011.
[4] F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error-Correcting Codes*, vol. 16. Amsterdam, The Netherlands: Elsevier, 1977.
[5] L. Yu and V. Y. F. Tan, "On non-interactive simulation of binary random variables," *IEEE Trans. Inf. Theory*, vol. 67, no. 4, pp. 2528–2538, Apr. 2021.
[6] R. Ahlswede, P. Gács, and J. Körner, "Bounds on conditional probabilities with applications in multi-user communication," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, vol. 34, no. 3, pp. 157–177, 1976.
[7] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Cambridge, U.K.: Cambridge Univ. Press, 2011.
[8] T. S. Han and K. Kobayashi, "Maximal rectangular subsets contained in the set of partially jointly typical sequences for dependent random variables," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, vol. 70, no. 1, pp. 15–32, 1985.
[9] D. Krithivasan and S. S. Pradhan, "On large deviation analysis of sampling from typical sets," in *Proc. Workshop Inf. Theory Appl. (ITA)*, 2007. [Online]. Available: https://public.websites.umich.edu/~dineshk/dinesh_ITA_report.pdf
[10] A. Nazari, D. Krithivasan, S. S. Pradhan, A. Anastasopoulos, and R. Venkataramanan, "Typicality graphs and their properties," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2010, pp. 520–524.
[11] P. Gács and J. Körner, "Common information is far less than mutual information," *Problems Control Inf. Theory*, vol. 2, no. 2, pp. 149–162, 1973.
[12] H. S. Witsenhausen, "On sequences of pairs of dependent random variables," *SIAM J. Appl. Math.*, vol. 28, no. 1, pp. 100–113, 1975.

[13] J. Kahn, G. Kalai, and N. Linial, "The influence of variables on Boolean functions," in *Proc. 29th Annu. Symp. Found. Comput. Sci.*, Oct. 1988, pp. 68–80.

[14] E. Mossel, R. O'Donnell, O. Regev, J. E. Steif, and B. Sudakov, "Non-interactive correlation distillation, inhomogeneous Markov chains, and the reverse Bonami-Beckner inequality," *Isr. J. Math.*, vol. 154, no. 1, pp. 299–336, Dec. 2006.

[15] R. O'Donnell, *Analysis of Boolean Functions*. Cambridge, U.K.: Cambridge Univ. Press, 2014.

[16] S. Kamath and V. Anantharam, "On non-interactive simulation of joint distributions," *IEEE Trans. Inf. Theory*, vol. 62, no. 6, pp. 3419–3435, Jun. 2016.

[17] O. Ordentlich, Y. Polyanskiy, and O. Shayevitz, "A note on the probability of rectangles for correlated binary strings," *IEEE Trans. Inf. Theory*, vol. 66, no. 12, pp. 7878–7886, Dec. 2020.

[18] N. Kirshner and A. Samorodnitsky, "A moment ratio bound for polynomials and some extremal properties of krawchouk polynomials and Hamming spheres," *IEEE Trans. Inf. Theory*, vol. 67, no. 6, pp. 3509–3541, Jun. 2021.

[19] C. Borell, "Geometric bounds on the Ornstein–Uhlenbeck velocity process," *Probab. Theory Rel. Fields*, vol. 70, no. 1, pp. 1–13, 1985.

[20] E. Mossel and J. Neeman, "Robust optimality of Gaussian noise stability," *J. Eur. Math. Soc.*, vol. 17, no. 2, pp. 433–482, Feb. 2015.

[21] E. Mossel and R. O'Donnell, "Coin flipping from a cosmic source: On error correction of truly random bits," *Random Struct. Algorithms*, vol. 26, no. 4, pp. 418–436, Jul. 2005.

[22] L. Yu and V. Y. F. Tan, "Common information, noise stability, and their extensions," *Found. Trends Commun. Inf. Theory*, vol. 19, no. 2, pp. 107–389, 2022.

[23] A. Bonami, "Ensembles $\Lambda(p)$ dans le dual de D^{∞}," in *Annales de L'Institut Fourier*, vol. 18. Paris, France: Assocation des Annales de l'Institut Fourier, 1968, pp. 193–204.

[24] K. Kiener, "Uber Produkte von quadratisch integrierbaren funktionen endlicher vielfalt," Ph.D. dissertation, Dept. Math., Universität Innsbruck, Innsbruck, Austria, 1969.

[25] M. Schreiber, "Fermeture en probabilité de certains sous-espaces d'un espace L^2 application aux chaos de Wiener," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, vol. 14, no. 1, pp. 36–48, 1969.

[26] A. Bonami, "Étude des coefficients de Fourier des fonctions de $L^p(G)$," *Annales de L'Institut Fourier*, vol. 20, no. 2, pp. 335–402, 1970.

[27] W. Beckner, "Inequalities in Fourier analysis," *Ann. Math.*, vol. 102, no. 1, pp. 159–182, 1975.

[28] L. Gross, "Logarithmic sobolev inequalities," *Amer. J. Math.*, vol. 97, no. 4, pp. 1061–1083, 1975.

[29] R. Ahlswede and P. Gács, "Spreading of sets in product spaces and hypercontraction of the Markov operator," *Ann. Probab.*, vol. 4, no. 6, pp. 925–939, 1976.

[30] C. Borell, "Positivity improving operators and hypercontractivity," *Mathematische Zeitschrift*, vol. 180, no. 3, pp. 225–234, Sep. 1982.

[31] D. Bakry, "L'hypercontractivité et son utilisation en théorie des semigroupes," in *Lectures on Probability Theory*. Berlin, Germany: Springer, 1994, pp. 1–114.

[32] E. Mossel, K. Oleszkiewicz, and A. Sen, "On reverse hypercontractivity," *Geometric Funct. Anal.*, vol. 23, no. 3, pp. 1062–1097, Jun. 2013.

[33] E. A. Carlen and D. Cordero-Erausquin, "Subadditivity of the entropy and its relation to Brascamp–Lieb type inequalities," *Geometric Funct. Anal.*, vol. 19, no. 2, pp. 373–405, 2009.

[34] C. Nair, "Equivalent formulations of hypercontractivity using information measures," in *International Zurich Seminar*. Zurich, Switzerland: ETH-Zürich, 2014.

[35] S. Kamath, "Reverse hypercontractivity using information measures," in *Proc. 53rd Annu. Allerton Conf. Commun., Control, Comput. (Allerton)*, Sep. 2015, pp. 627–633.

[36] S. Beigi and C. Nair, "Equivalent characterization of reverse Brascamp–Lieb-type inequalities using information measures," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2016, pp. 1038–1042.

[37] J. Liu, T. A. Courtade, P. Cuff, and S. Verdú, "Brascamp–Lieb inequality and its reverse: An information theoretic view," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2016, pp. 1048–1052.

[38] J. Liu, T. A. Courtade, P. W. Cuff, and S. Verdú, "A forward-reverse Brascamp–Lieb inequality: Entropic duality and Gaussian optimality," *Entropy*, vol. 20, no. 6, p. 418, 2018.

[39] L. Yu, "Strong Brascamp–Lieb inequalities," 2021, *arXiv:2102.06935*.

[40] L. Lovász, "On the Shannon capacity of a graph," *IEEE Trans. Inf. Theory*, vol. IT-25, no. 1, pp. 1–7, Jan. 1979.

[41] J. Liu, "Minoration via mixed volumes and Cover's problem for general channels," *Probab. Theory Rel. Fields*, vol. 183, no. 1, pp. 315–357, 2022.

[42] Y. Polyanskiy and A. Samorodnitsky, "Improved log-sobolev inequalities, hypercontractivity and uncertainty principle on the hypercube," *J. Funct. Anal.*, vol. 277, no. 11, Dec. 2019, Art. no. 108280.

[43] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2011.

[44] J. Liu, T. A. Courtade, P. Cuff, and S. Verdú, "Smoothing Brascamp–Lieb Inequal. And strong converses of coding theorems," *IEEE Trans. Inf. Theory*, vol. 66, no. 2, pp. 704–721, Feb. 2020.

[45] L. Yu and V. Y. F. Tan, "On exact and ∞-Rényi common informations," *IEEE Trans. Inf. Theory*, vol. 66, no. 6, pp. 3366–3406, Feb. 2020.

[46] R. O'Donnell, *Analysis of Boolean Functions*. Cambridge, U.K.: Cambridge Univ. Press, 2014.

[47] J. Liu, "Information theory from a functional viewpoint," Ph.D. dissertation, Dept. Electr. Eng., Princeton Univ., Princeton, NJ, USA, 2018.

[48] C. Villani, *Optimal Transport: Old and New*, vol. 338. Berlin, Germany: Springer, 2008.

[49] L. Yu, "Asymptotics for Strassen's optimal transport problem," *Annales de l'Institut Henri Poincare (B) Probabilites et Statistiques*, vol. 59, no. 4, p. 1745– 1777, 2023.

[50] R. M. Gray and A. D. Wyner, "Source coding for a simple network," *Bell Syst. Tech. J.*, vol. 53, no. 9, pp. 1681–1721, Nov. 1974.

[51] A. Wyner, "Recent results in the Shannon theory," *IEEE Trans. Inf. Theory*, vol. IT-20, no. 1, pp. 2–10, Jan. 1974.

[52] L. Yu, "Gray–Wyner and mutual information regions for doubly symmetric binary sources and Gaussian sources," *IEEE Trans. Inf. Theory*, vol. 69, no. 10, pp. 6251–6268, Oct. 2023.

[53] L. Yu, "The convexity and concavity of envelopes of the minimum-relative-entropy region for the DSBS," 2021, *arXiv:2106.03654*.

[54] B. Lindström, "Determination of two vectors from the sum," *J. Combinat. Theory*, vol. 6, no. 4, pp. 402–407, May 1969.

[55] H. van Tilborg, "An upper bound for codes in a two-access binary erasure channel (Corresp.)," *IEEE Trans. Inf. Theory*, vol. IT-24, no. 1, pp. 112–116, Jan. 1978.

[56] T. Kasami and S. Lin, "Bounds on the achievable rates of block coding for a memoryless multiple-access channel," *IEEE Trans. Inf. Theory*, vol. IT-24, no. 2, pp. 187–197, Mar. 1978.

[57] E. J. Weldon Jr, "Coding for a multiple-access channel," *Informationa Control*, vol. 36, no. 3, pp. 256–274, 1978.

[58] T. Kasami, S. Lin, V. Wei, and S. Yamamura, "Graph theoretic approaches to the code construction for the two-user multiple- access binary adder channel," *IEEE Trans. Inf. Theory*, vol. IT-29, no. 1, pp. 114–130, Jan. 1983.

[59] R. Urbanke and Q. Li, "The zero-error capacity region of the 2-user synchronous BAC is strictly smaller than its Shannon capacity region," in *Proc. Inf. Theory Workshop*, Jun. 1998, p. 61.

[60] G. Ajjanagadde and Y. Polyanskiy, "Adder MAC and estimates for Rényi entropy," in *Proc. 53rd Annu. Allerton Conf. Commun., Control, Comput. (Allerton)*, Oct. 2015, pp. 434–441.

[61] O. Ordentlich and O. Shayevitz, "An upper bound on the sizes of multiset-union-free families," *SIAM J. Discrete Math.*, vol. 30, no. 2, pp. 1032–1045, 2016.

[62] P. Austrin, P. Kaski, M. Koivisto, and J. Nederlof, "Sharper upper bounds for unbalanced uniquely decodable code pairs," *IEEE Trans. Inf. Theory*, vol. 64, no. 2, pp. 1368–1373, Feb. 2018.

[63] D. B. West, *Introduction to Graph Theory*, vol. 2. Upper Saddle River, NJ, USA: Prentice-Hall, 2001.

[64] L. Yu and V. Y. F. Tan, "Rényi resolvability and its applications to the wiretap channel," *IEEE Trans. Inf. Theory*, vol. 65, no. 3, pp. 1862–1897, Mar. 2019.

[65] C. Greene and T. L. Magnanti, "Some abstract pivot algorithms," *SIAM J. Appl. Math.*, vol. 29, no. 3, pp. 530–539, 1975.

[66] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Hoboken, NJ, USA: Wiley, 2006.

Lei Yu (Member, IEEE) received the B.E. and Ph.D. degrees in electronic engineering from the University of Science and Technology of China (USTC) in 2010 and 2015, respectively. From 2015 to 2020, he was a Post-Doctoral Researcher with USTC, the National University of Singapore, and the University of California at Berkeley. He is currently an Associate Professor with the School of Statistics and Data Science, LPMC, KLMDASR, and LEBPS, Nankai University, China. His research interests lie in the intersection of probability theory, information theory, and combinatorics.

Venkat Anantharam (Fellow, IEEE) received the B.Tech. degree in electronics from IIT Madras in 1980, and the M.S. degree in electrical engineering, the M.A. and C.Phil. degrees in mathematics, and the Ph.D. degree in electrical engineering from UC Berkeley, in 1982, 1983, 1984, and 1986, respectively. From 1986 to 1994, he was on the Faculty of the School of EE, Cornell University, before moving to the Department of Electrical Engineering and Computer Sciences, UC Berkeley. He is currently on the Faculty with UC Berkeley. His research interests include communication networking, game theory, information theory, probability theory, and stochastic control.

Jun Chen (Senior Member, IEEE) received the B.E. degree in communication engineering from Shanghai Jiao Tong University, Shanghai, China, in 2001, and the M.S. and Ph.D. degrees in electrical and computer engineering from Cornell University, Ithaca, NY, USA, in 2004 and 2006, respectively.

From September 2005 to July 2006, he was a Post-Doctoral Research Associate with the Coordinated Science Laboratory, University of Illinois at Urbana–Champaign, Urbana, IL, USA; and a Post-Doctoral Fellow with the IBM Thomas J. Watson Research Center, Yorktown Heights, NY, USA, from July 2006 to August 2007. Since September 2007, he has been with the Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON, Canada, where he is currently a Professor. His research interests include information theory, machine learning, wireless communications, and signal processing.

Dr. Chen was a recipient of the Josef Raviv Memorial Postdoctoral Fellowship in 2006, the Early Researcher Award from the Province of Ontario in 2010, the IBM Faculty Award in 2010, the ICC Best Paper Award in 2020, and the JSPS Invitational Fellowship in 2021. He held the title of the Barber-Gennum Chair of Information Technology, from 2008 to 2013, and the title of the Joseph Ip Distinguished Engineering Fellow, from 2016 to 2018. He was an Associate Editor of IEEE TRANSACTIONS ON INFORMATION THEORY (2014–2016 and 2021–2024), an Editor of IEEE TRANSACTIONS ON GREEN COMMUNICATIONS AND NETWORKING (2020–2021), and a Guest Editor of the Special Issue on Modern Compression for the IEEE JOURNAL ON SELECTED AREAS IN INFORMATION THEORY (2022). He is serving as an Associate Editor for IEEE TRANSACTIONS ON COMMUNICATIONS; a Lead Editor for the Special Issue Dedicated to the Memory of Toby Berger for the IEEE JOURNAL ON SELECTED AREAS IN INFORMATION THEORY; and a Guest Editor of the Special Issue on Rethinking the Information Identification, Representation, and Transmission Pipeline for the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS.