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 $2^n R_1$ and $2^n R_2$ 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 $X$ and $Y$ be two finite sets. Let $T_X$ be an $n$-type on $X$, i.e., an empirical distribution of sequences from $X^n$.

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 CCF–2007965; in part by the NSF Science and Technology Center (Science of Information) under Grant CCF–0939370; and in part by the National Science Foundation and the Office of Naval Research.

Lei Yu is with the School of Statistics and Data Science, LSTM, KLM-DASR, 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).

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

0018-9448 © 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.
Given \( 1 \leq M_1 \leq |T_{X_1}|, 1 \leq M_2 \leq |T_{Y_1}| \), define the maximal density of subgraphs with size \( (M_1, M_2) \) as
\[
\Gamma_n(M_1, M_2) := \max_{A \subseteq T_{X_1}, B \subseteq T_{Y_1}} \frac{|A| = M_1, |B| = M_2}{\rho(G[A, B])}.
\] (1)

Recall that \( T_{X_1} \) and \( T_{Y_1} \) denote the conditional types corresponding to the joint type \( T_{XY} \). For a sequence \( x \in T_{X_1} \), let
\[
T_{Y_1|x}(x) := \{ y \in Y^n : (x, y) \in T_{XY} \}
\]
denote the corresponding conditional type class. Since \( N_1 := |T_{Y_1|x}(x)| \) is independent of \( x \in T_{X_1} \), the degrees of the vertices \( x \in T_{X_1} \) are all equal to the constant \( N_1 \). Similarly, the degrees of the vertices \( y \in T_{Y_1} \) are all equal to the constant \( N_2 := |T_{X_1|Y}(y)| \). Hence we have
\[
|B| \rho(G[A, B]) + |B| \rho(G[A, B']) = \frac{|(A \times T_{Y_1}) \cap T_{XY}|}{|A|} = \sum_{x \in A} \frac{|T_{XY}|(x)}{|A|} = N_1,
\]
where \( B' := T_{Y_1} \setminus B \). Thus, over \( A, B \) with fixed sizes, maximizing \( \rho(G[A, B]) \) is equivalent to minimizing \( \rho(G[A, B']) \) (or \( \rho(G[A', B]) \)). 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 \( A \subseteq T_{X_1} \) and \( M_2 \), we see that
\[
\max_{B \subseteq T_{Y_1} : |B| = M_2} \rho(G[A, B]) = \frac{1}{|A|} \max_{B \subseteq T_{Y_1} : |B| = M_2} \sum_{y \in B} |A \cap T_{X_1|Y}(y)|,
\]
and the maximum is attained by \( B' \) such that \( |A \cap T_{X_1|Y}(y)| \geq |A \cap T_{X_1|Y}(y')| \) for any \( y', y' \notin B' \). Hence, \( M_2 \mapsto \max_{B \subseteq T_{Y_1} : |B| = M_2} \rho(G[A, B]) \) is nonincreasing, which implies that \( \Gamma_n(M_1, M_2) \) is nonincreasing in one parameter given the other parameter.

Let
\[
\mathcal{R}_X^{(n)} := \left\{ \frac{1}{n} \log M_1 : M_1 \in |T_{X_1}| \right\},
\]
\[
\mathcal{R}_Y^{(n)} := \left\{ \frac{1}{n} \log M_2 : M_2 \in |T_{Y_1}| \right\},
\]
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
\[
E_n(R_1, R_2) := -\frac{1}{n} \log \Gamma_n(2^{nR_1}, 2^{nR_2}).
\] (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
\[
\mathcal{R}_n(T_{XY}) := \{ (R_1, R_2) \in \mathcal{R}_X^{(n)} \times \mathcal{R}_Y^{(n)} : \Gamma_n(2^{nR_1}, 2^{nR_2}) = 1 \}.
\] (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
\[
E(R_1, R_2) := \lim_{k \to \infty} -\frac{1}{kn} \log \Gamma_{kn}(2^{knR_1}, 2^{knR_2}),
\]
and the asymptotic biclique rate region as
\[
\mathcal{R}(T_{XY}) := \text{closure} \bigcup_{k \geq 1} \mathcal{R}_{kn}(T_{XY}).
\] (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 \( (x, y) \) in the joint type class \( T_{XY} \), then the type graph can be constructed from the set of all pair sequences resulting

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

4We use the notation \([m:n] := \{m, m+1, \ldots, n\}\) and \([n] := \{1:n\}\).
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 $P_{XY}$, the noninteractive simulation problem concerns estimating the maximal and minimal joint probability $P_{XY}^{(n)}(A \times B)$ when the marginal probabilities $P_{X}^{(n)}(A)$ and $P_{Y}^{(n)}(B)$ 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 techniques and linear algebra; similar techniques were also employed 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 Ahswede 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

Our 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 supports to be exponentially small. We obtain stronger

The symmetric case $P_{X}^{(n)}(A_n) = P_{Y}^{(n)}(B_n)$ 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.
(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 Kirchner 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 =: b_n to denote a_n = b_n. We use 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 C(Q_X|UW, Q_Y|VW) to
denote the set of conditional couplings Q_{XY|UW} with con-
tional marginals Q_X|UW, Q_Y|VW. Note that, given Q_X|UW
and Q_Y|VW, for any Q_{XY|UW} ∈ C(Q_X|UW, Q_Y|VW) and
any Q_{UVW}, the joint law Q_{XYUW} = Q_{XY|UW}Q_{UVW} is
such that X ↔ (U, W) ↔ V and Y ↔ (V, W) ↔ U, where
the notation X ↔ Y ↔ 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 x, we use T_x to denote the
type of x. For an m × n matrix B = (b_{i,j}) and two subsets
H ⊆ [m], L ⊆ [n], we use B_{H,L} to denote (b_{i,j})_{i∈H,j∈L},
e., the submatrix of B consisting of the elements with
indices in H × L. For a length-n vector or sequence x and
a subset J ⊆ [n], x_J := (x_j)_{j∈J} is defined similarly. For
a distribution P_X, we use P^n_X to denote the n-fold produc-
tof P_X. We will also use notations H_Q(X) or H(Q_X) to
denote the entropy of X ∼ Q_X. If the distribution is denoted by
P_X, we sometimes write the entropy as H(X) for brevity. We
use supp(P_X) to denote the support of P_X. The logarithm
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 : X →
[0, ∞) and g : Y → [0, ∞), define their inner product
\[ \langle f, g \rangle := \text{E}[f(X)g(Y)] = \sum_{(x,y) \in X \times Y} P_{XY}(x,y)f(x)g(y). \] (8)

The L^p-norm of f for p ∈ [1, ∞) and the pseudo L^p-norm
of f for p ∈ (0, 1) are defined as
\[ \|f\|_p := \left(\text{E}[f(X)^p]\right)^{1/p} = \left(\sum_{x \in X} P_X(x)f(x)^p\right)^{1/p}. \] (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
\[ F^*(R_1, R_2) := \max_{P_{XYW} \text{: } H(X,Y|W), \quad H(X|W) \leq R_1, H(Y|W) \leq R_2} P_{XYW}. \] (10)

and
\[ E^*(R_1, R_2) := R_1 + R_2 - F^*(R_1, R_2). \] (11)

**Theorem 1:** Given a joint n-type T_{XY} with n ≥ 2(|X||Y|+ 2)|X||Y|, for (R_1, R_2) ∈ R_1 \times R_2, we have
\[ E^*(R_1, R_2) \leq E_n(R_1, R_2) \leq E^*(R_1, R_2) + \epsilon_n, \] (12)
where \( \epsilon_n := \frac{|X||Y|+2|X||Y|}{2(n+1)} \log \left(\frac{n+1}{2}\right). \) As a consequence,
for any n ≥ 1 and any joint n-type T_{XY}, we have
\[ E(R_1, R_2) = E^*(R_1, R_2). \] (13)

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

**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
\[ G^*(R_1, R_2) := \min_{P_{XYW} \text{: } H(X,Y|W) \leq R_1, H(Y|W) \leq R_2} I(X,Y;W) \] (14)
corresponding to the minimum common rate given marginal
rates \( R_1, R_2 \) in the Gray–Wyner source coding network
\[ \text{[43, Theorem 14.3].} \]

**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, \( E_n(R_1, R_2) =
E^*(R_1, R_2) + O(\log n) \), can be recovered from a more
general result given in \[ \text{[44, (6) and (7)]} \] 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 \( R_1, R_2 \geq 0, \) 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) \( F^*(R_1, R_2) \leq \min \{H_T(X,Y), R_1 + R_2, R_1 + H_T(Y|X), R_2 + H_T(Y|X)\} \).

Moreover, \( F^*(R_1, R_2) \) is concave in \( (R_1, R_2) \) on \( \{(R_1, R_2) : R_1 \geq 0, R_2 \geq 0\} \).

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

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

Theorem 1 is an edge-isoperimetric result for the bipartite
graph induced by a joint n-type T_{XY}. For the case in which
Given 1 ≤ M ≤ |T_{XY}|, define the maximal density of subgraphs with size M as

\[ \Gamma_n(M) := \max_{A \subseteq T_{XY} : |A| = M} \rho(G[A]). \]

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

\[ E_n(R) := -\frac{1}{n} \log \Gamma_n(2^{|R|}). \]

For any subsets \( \mathcal{A}, \mathcal{B} \) of \( \Lambda^n \), we have

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

On the other hand, we have

\[ \Gamma_n(M) \leq \Gamma_n(M, M). \]

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

\[ E_n(R) = E^*(R, R) + O\left(\frac{\log n}{n}\right), \]

where the asymptotic constant in the \( O\left(\frac{\log n}{n}\right) \) 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 \( x \in T_{TX} \) and there is a directed edge from \( x \) to \( y \) if and only if \( (x, y) \in T_{TXY} \). Hence, for this case, Theorem 1 can be also considered as an edge-isomietric result for a directed graph induced by \( T_{XY} \). Specifically, for a subset \( \mathcal{A} \subseteq T_{TX} \), 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

\[ \rho(G[\mathcal{A}]) := \frac{\# \text{ of directed edges in } G[\mathcal{A}]}{|\mathcal{A}|^2}. \]

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

\[ \mathcal{R}^*(T_{XY}) := \bigcup_{0 \leq \alpha \leq 1 \land P_{X^Y}, Q_{XY} : \alpha P_{X^Y} + (1 - \alpha) Q_{XY} = T_{XY}} \{(R_1, R_2) : R_1 \leq \alpha H_P(X|Y), R_2 \leq (1 - \alpha) H_Q(Y|X)\}. \]

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

\[ \mathcal{R}^*(T_{XY}) - [0, \varepsilon_{1,n}] \times [0, \varepsilon_{2,n}] \cap \big( \mathcal{R}_X^{(n)} \times \mathcal{R}_Y^{(n)} \big) \subseteq \mathcal{R}_n(T_{XY}) \subseteq \mathcal{R}^*(T_{XY}) \cap \big( \mathcal{R}_X^{(n)} \times \mathcal{R}_Y^{(n)} \big) \]

where \( \mathcal{R}_n(T_{XY}) \) is defined in (5), “-” is the Minkowski difference (i.e., for \( \mathcal{A}, \mathcal{B} \subseteq \mathbb{R}^m \), \( \mathcal{A} - \mathcal{B} := \{ \sum_{i \in \mathcal{B}} (\mathcal{A}_i - \mathcal{B}_i) \} \)), and \( \varepsilon_{1,n} := \frac{|\mathcal{X}| |\mathcal{Y}|}{n} \log \frac{n^{15(n+1)}}{16|\mathcal{X}||\mathcal{Y}|} \), and \( \varepsilon_{2,n} := \frac{|\mathcal{X}| |\mathcal{Y}|}{n} \log \frac{n^{15(n+1)}}{16|\mathcal{X}||\mathcal{Y}|} \).

In particular,

\[ \mathcal{R}(T_{XY}) = \mathcal{R}^*(T_{XY}), \]

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 \( k \geq 3 \). 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 \( \langle A_1, \ldots, A_k \rangle \) is defined as

\[ \rho(G[A_1, \ldots, A_k]) := \frac{|\prod_{i=1}^k A_i \cap T_{X_1, \ldots, X_k}|}{\prod_{i=1}^k |A_i|}. \]

It is interesting to observe that \( \rho(G) \geq 2^{-n I_{T^{(n)}}(X_1; \ldots; X_k)} \) for a sequence of joint types \( \{T^{(n)}_{X_1, \ldots, X_k}\} \), 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

\[ \mathcal{R}_n(T_{X_1, \ldots, X_k}) := \left\{ \left( \frac{1}{n} \log |A_1|, \ldots, \frac{1}{n} \log |A_k| \right) : \rho(G[A_1, \ldots, A_k]) = 1 \right\}. \]

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

\[ \mathcal{R}^*(T_{X_1, \ldots, X_k}) \subseteq \mathcal{R}_n(T_{X_1, \ldots, X_k}) \subseteq \mathcal{R}^*(T_{X_1, \ldots, X_k}) \cap \left( \prod_{i=1}^k \mathcal{R}_X^{(n)} \right) \]
where

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

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

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

**Proof:** Using the continuity of \( H_{P}(X|Y) \) in \( P_{XY} \) it can be established that \( R^*(T_{XY}) \) is closed. Convexity follows by the following argument. For any \( (R_1, R_2), (\hat{R}_1, \hat{R}_2) \in R^*(T_{XY}) \), there exist \( (\alpha, P_{XY}, Q_{XY}) \) 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 \( \lambda \in [0, 1] \),
\[
\lambda R_1 + (1 - \lambda) \hat{R}_1 \leq \lambda \alpha H_P(X|Y) + (1 - \lambda) \hat{\alpha} H_{\hat{P}}(X|Y) \]
\[
\leq \beta P_{XY} + (1 - \beta) \hat{P}_{XY} \quad \text{with} \quad \beta = \frac{\lambda \alpha + (1 - \lambda) \hat{\alpha}}{\lambda \alpha + (1 - \lambda) \hat{\alpha} \alpha} \]
\[
\leq \frac{\lambda \alpha + (1 - \lambda) \hat{\alpha}}{\lambda \alpha + (1 - \lambda) \hat{\alpha} \alpha} \quad \text{if} \quad \beta > 0; \quad \beta P_{XY} + (1 - \beta) H_{Q_{\lambda}}(Y|X), \]
\[
\leq \frac{\lambda \alpha + (1 - \lambda) \hat{\alpha}}{\lambda \alpha + (1 - \lambda) \hat{\alpha} \alpha} \quad \text{if} \quad \beta < 1; \quad \beta H_{P_{\theta}}(X|Y), \]
\[
\leq \frac{\lambda \alpha + (1 - \lambda) \hat{\alpha}}{\lambda \alpha + (1 - \lambda) \hat{\alpha} \alpha} \quad \text{if} \quad \beta = 1. \]

By Theorem 1, \( R^*(T_{XY}) \) also coincides with the region
\[ \bigcup \{(R_1, R_2) : P_{XY} = T_{XY}, R_1 \leq H(X|W), R_2 \leq H(Y|W)\}. \]

**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 \( \{(R_1, R_2) : R_1 + R_2 \leq h(\frac{1}{2} - \rho)\} \) 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.

| \( X \setminus Y \) | 0 | 1 |
|---|---|---|
| 0 | 1 + \( \rho \) | 1 - \( \rho \) |
| 1 | 1 - \( \rho \) | 1 + \( \rho \) |

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]

\[ \mathcal{R}^{**}(T_{XY}) := \bigcup \{(R_1, R_2) : P_{XY} = T_{XY}, R_1 \leq H(X|W), R_2 \leq H(Y|W)\}. \]

Another family of examples where the asymptotic biclique region is strictly contained in the region of Han and Kobayashi

---

**Table I**

| \( X \setminus Y \) | 0 | 1 |
|---|---|---|
| 0 | 1 + \( \rho \) | 1 - \( \rho \) |
| 1 | 1 - \( \rho \) | 1 + \( \rho \) |
is when the joint n-type $T_{XY}$ is Unif$(X \times Y)$. Here $R^*(T_{XY})$ equals the rectangle region $[0, H(X)] \times [0, H(Y)]$, while Proposition 2 implies that $R^*(T_{XY})$ 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 $(A, B)$ be an optimal pair of subsets attaining $E^*(R_1, R_2)$. All the sequences in $A$ have type $T_X$, and all the sequences in $B$ have type $T_Y$. However, in general, the joint types of $(x, y) \in A \times B$ might "overflow" from the target joint type $T_{XY}$. The number of non-overflowed sequence pairs (i.e., $|(A \times B) \cap T_{XY}|$) 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 $A$ and $B$ 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 Unif$(T_{XY})$

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

Define $E_X^{(n)} := \{ \frac{1}{n} \log |T_{XY}| - R_1 : R_1 \in R_X^{(n)} \}$ and $E_Y^{(n)} := \{ \frac{1}{n} \log |T_{XY}| - R_2 : R_2 \in R_Y^{(n)} \}$, where $R_X^{(n)}$ and $R_Y^{(n)}$ are defined in (2). Given a joint n-type $T_{XY}$, for $(E_1, E_2) \in E_X^{(n)} \times E_Y^{(n)}$, define the exponents of the maximal and minimal noise stability as

$$\Theta_n(E_1, E_2) := -\frac{1}{n} \log \min_{A \subseteq T_X, B \subseteq T_Y : P_X(A) = 2^{-nE_1}, P_Y(B) = 2^{-nE_2}} P_{X,Y}(A \times B),$$

(23)

$$\overline{\Theta}_n(E_1, E_2) := -\frac{1}{n} \log \max_{A \subseteq T_X, B \subseteq T_Y : P_X(A) = 2^{-nE_1}, P_Y(B) = 2^{-nE_2}} P_{X,Y}(A \times B).$$

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

$$\overline{\Theta}_n(s, t) := \min_{P_{X,Y} : P_{X,Y} = T_{XY}, I(X:W) \geq s, I(Y:W) \geq t} I(X;Y;W).$$

Theorem 3: For any $T_{XY}$ and $(E_1, E_2) \in E_X^{(n)} \times E_Y^{(n)}$, we have

$$\overline{\Theta}_n(E_1, E_2) = \overline{\Theta}_n(E_1, E_2) + O\left(\frac{\log n}{n}\right),$$

(25)

where the asymptotic constant in the $O(\frac{\log n}{n})$ bound depends only on $|X|, |Y|$. In fact, this result can be recovered from a more general result given in [44, (6) and (7)]. The latter generalizes the $\overline{\Theta}_n$ for the uniform distribution over a type class to the infimum of $\overline{\Theta}_n$ over all distributions not too far from a product distribution.

Proof: Observe that

$$P_X(A) = |A|/|T_{X}|,$$

(26)

$$P_Y(B) = |B|/|T_{Y}|,$$

(27)

$$P_{X,Y}(A \times B) = \rho(G[A,B])|A||B|/|T_{XY}|.$$  

(28)

So, Theorem 3 is implied by Theorem 1. $\blacksquare$

#### B. Sources $P_{X,Y}^n$

In this subsection, we consider the noninteractive simulation problem with $(X, Y) \sim P_{X,Y}^n$, where $P_{X,Y}$ 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 $P_{X,Y}$. Ordentlich et al. [17] focused on binary symmetric distributions $P_{X,Y}$, and studied the exponent of $P_{X,Y}(A \times B)$ given that $P_X^n(A), P_Y^n(B)$ vanish exponentially fast with exponents $E_1, E_2$, respectively. Let

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

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 $E_1 \in [0, E_{1,\text{max}}]$, $E_2 \in [0, E_{2,\text{max}}]$, we aim at characterizing

$$\Theta(E_1, E_2) := \lim_{n \to \infty} \Theta_n(E_1, E_2).$$

(29)

By time-sharing arguments, given $(E_1, E_2)$, $(n \Theta_n(E_1, E_2))_{n \geq 1}$ is subadditive. Hence, by Fekete’s Subadditive Lemma, the first limit in (29) exists and equals $\Theta(E_1, E_2)$. Similar observations serve to define the second limit in (29).
where the exponents of the maximal and minimal noise stability are defined by

\[
\Theta_n(E_1, E_2) := \lim_{n \to \infty} \Theta_n(E_1, E_2), \quad (30)
\]

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 \(\Theta^*\) and concavity of \(\Theta^*\), Theorem 4 implies \(\Theta_n(E_1, E_2) \geq \lim_{n \to 0} \frac{1}{2} \Theta^*(tE_1, tE_2)\) and \(\Theta_n(E_1, E_2) \leq \lim_{n \to 0} \frac{1}{2} \Theta^{**}(tE_1, tE_2)\). In particular, for the DSBS with correlation coefficient \(\rho > 0\), these inequalities reduce to that

\[
\Theta_n(E_1, E_2) \geq \begin{cases} 
\frac{E_1 + E_2 - 2\sqrt{E_1E_2}}{1 - \rho^2}, & E_1 \leq E_2 \leq E_1/\rho^2, \\
E_2, & E_2 > E_1/\rho^2.
\end{cases} 
\]

(38)

\[
\Theta_n(E_1, E_2) \leq \frac{E_1 + E_2 + 2\sqrt{E_1E_2}}{1 - \rho^2}. 
\]

(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 \(\lim_{n \to \infty} \frac{1}{n} \log P^n_X(A_n) \geq E_1\) for any sequence \((A_n, B_n)\) attaining the asymptotic exponent \(\Theta^*(E_1, E_2)\). If \(\lim \sup_{n \to \infty} \frac{1}{n} \log P^n_X(A_n) > E_1\) then it must be the case that \(E_1 < E_{1,\text{max}}\). So in this case, we can add sequences in \(\mathcal{X}^n\) to \(A_n\) to get a new sequence \((A_n, B_n)\) such that \(\lim_{n \to \infty} \frac{1}{n} \log P^n_X(A_n) = E_1\). This is possible since for each \(E_1 \in [0, E_{1,\text{max}}]\) there is a sequence \(A_n \subseteq \mathcal{X}^n\) such that (1) each \(A_n\) is a type class, and concavity of \(\Theta^*\) (see 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 \(\Theta^*\) and \(\Theta^{**}\) 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 \(E_1 \in [0, E_{1,\text{max}}], E_2 \in [0, E_{2,\text{max}}]\), the following hold.

1) \(\Theta^*(E_1, E_2) = \Theta^{**}(E_1, E_2)\). Moreover, \(\Theta_n(E_1, E_2) \geq \Theta^*(E_1, E_2)\) for any \(n \geq 1\).

2) \(\Theta^*(E_1, E_2) = \Theta^{**}(E_1, E_2)\)

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

(32)

Moreover, \(\Theta_n(E_1, E_2) \leq \Theta^{**}(E_1, E_2)\) for any \(n \geq 1\).

**Remark 5:** We interpret \(\Theta^*(E_1, E_2)\) as \(\infty\) if \(P_{XY}(x, y) = 0\) for some \((x, y)\). It can be checked that this interpretation is

Theorem 3.4 [15, Generalized Small-Set Expansion Theorem on p. 285].

**Remark 8:** We define the effective region of \(\Theta^*\) as the set of \((E_1, E_2)\) for which \(\Theta^*(E_1, E_2) = \Theta^*(E_1, E_2)\), i.e., there exists an optimal tuple \((Q_W, Q_X|W, Q_Y|W)\) attaining the maximum in the definition of \(\Theta^*(E_1, E_2)\) such that \(D(Q_X|W||P_X|Q_W) = E_1\) and \(D(Q_Y|W||P_Y|Q_W) = E_2\). 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 \(D(Q_X|W||P_Y|Q_W)\) is the asymptotic exponent of the probability of \(T_{Q_X|W}(w)\times T_{Q_Y|W}(w)\) with \(w\) having type \(Q_W\). Hence, for \((E_1, E_2)\) the effective region of \(\Theta^*\), there exists a sequence of \((A_n, B_n)\) such that \(\lim_{n \to \infty} \frac{1}{n} \log P^n_X(A_n) = E_1\), \(\lim_{n \to \infty} \frac{1}{n} \log P^n_Y(B_n) = E_2\), and \(\lim_{n \to \infty} \frac{1}{n} \log P^n_Y(B_n) = E_2\). The 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 $0 \leq E_1 \leq H(X), 0 \leq E_2 \leq H(Y)$, we have $\Upsilon^*(E_1, E_2) \geq \Theta^*(E_1, E_2)$ (the inequality is strict in general), where $\Upsilon^*$ and $\Theta^*$ are both defined for $P_{XY}$. This observation follows since for any distribution $Q_{XWY}$ with marginal $Q_{XY} = P_{XY}$, it holds that $I_Q(XY; W) = D(Q_{XY}|W) || P_{XY} = E_{QW} D(Q_{XY}|W) || P_{XY}$. 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 $\Upsilon^*$, then we obtain $\Theta^*$. So, we have the observation. Another observation is $\lim_{t \to 0} \frac{1}{t} \Upsilon^*(tE_1, tE_2) = \lim_{t \to 0} \frac{1}{t} \Theta^*(tE_1, tE_2)$.

By the following 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 $\Theta^*(E_1, E_2)$ or $\Theta^*(E_1, E_2)$ can be taken to be at most 3, both $\Theta^*(E_1, E_2)$ and $\Theta^*(E_1, E_2)$ 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 $(A, B) := (T_{XY}, T_{XY})$ for a pair of types $(T_X, T_Y)$.

Define the optimal transport divergence¹¹ (or the minimum relative entropy) of $(Q_X, Q_Y)$ with respect to $P_{XY}$, as

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

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

$$\phi(s, t) := \min_{D(Q_X, Q_Y || P_{XY}) = t} D(Q_{XY} || P_{XY})$$

and

$$\psi(s, t) := \max_{D(Q_X, Q_Y || P_{XY}) = t} D(Q_{XY} || P_{XY}).$$

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

$$\Theta^*(E_1, E_2) = \min_{\{q_i(s_i, t_i)\}_{i=1}^{\infty}} \sum_{i} q_i \phi(s_i, t_i) = \min_{s \geq E_1, t \geq E_2} \sum_{\{q_i(s_i, t_i)\}_{i=1}^{\infty}} q_i \phi(s_i, t_i)$$

and similarly,

$$\Theta^*(E_1, E_2) = \max_{s \leq E_1, t \leq E_2} \hat{\psi}(s, t).$$

¹¹The 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 $\Theta^*(E_1, E_2)$ is convex in $(E_1, E_2)$, and $\Theta^*(E_1, E_2)$ is concave in $(E_1, E_2)$. By definition, it holds that for $s \in [0, E_{1,max}]$, $t \in [0, E_{2,max}]$,

$$\Theta^*(s, t) \leq \phi(s, t) \leq \phi(s, t) \leq \psi(s, t) \leq \hat{\psi}(s, t) \leq \Theta^*(s, t).$$

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

**Proof:** By the convexity and concavity, $\Theta^*(E_1, E_2)$ and $\Theta^*(E_1, E_2)$ 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 $D(Q_{XW} || P_X | Q_W), D(Q_{YW} || P_Y | Q_W)$ and $D(Q_{XW} || P_X | Q_Y) || P_Y | Q_W)$ in $(Q_W, Q_{XW}, Q_{YW})$. The continuity of $D(Q_{XW} || P_X | Q_W), D(Q_{YW} || P_Y | Q_W)$ in $(Q_W, Q_{XW}, Q_{YW})$ is obvious, since as assumed, $P_X$ and $P_Y$ have full support. We claim that

$$f(Q_W, Q_{XW}, Q_{YW}) := D(Q_{XW} || P_X | Q_W), D(Q_{YW} || P_Y | Q_W)$$

is continuous in $(Q_W, Q_{XW}, Q_{YW})$, 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

$$\|P_{XW} - Q_{XW}\| \leq \|P_X - Q_X\| + \|P_Y - Q_Y\|,$$

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

By Lemma 2, given $(Q_W, Q_{XW}, Q_{YW}, P_X, P_Y, P_{XW}, P_{YW})$, for any $Q_{XY} \in \mathcal{C}(Q_X, Q_Y)$, there exists $P_{XY} \in \mathcal{C}(P_X, P_Y)$ such that

$$\|P_{XW} - Q_{XW}\| \leq \|P_X - Q_X\| + \|P_Y - Q_Y\| + \|P_{XW} - Q_{XW}\| + \|P_{YW} - Q_{YW}\|.$$

Hence, for any sequence $(P_{XW}^{(k)}, P_{XW}^{(k)}, P_{YW}^{(k)})$ convergent to $(Q_{W}, Q_{XW}, Q_{YW})$, $\lim sup_{k \to \infty} f(P_{XW}^{(k)}, P_{XW}^{(k)}, P_{YW}^{(k)}) \leq f(Q_{W}, Q_{XW}, Q_{YW})$, and $f(Q_{W}, Q_{XW}, Q_{YW}) \leq \lim inf_{k \to \infty} f(P_{XW}^{(k)}, P_{XW}^{(k)}, P_{YW}^{(k)})$. Hence $f(Q_{W}, Q_{XW}, Q_{YW})$ is continuous in $(Q_{W}, Q_{XW}, Q_{YW})$.

**C. Example: DSBS**

Consider a DSBS with correlation coefficient $\rho$, 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 $[0, 1/2]$. The following explicit expression for $\Upsilon^*$ (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.

---

Author: YU et al., GrAPHS OF JOINT TYPES, NONINTERACTIVE SIMULATION, AND STRONGER HYPERCONTRACTIVITY 2295

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

\[
\mathcal{Y}^*(s, t) = \begin{cases} 
1 - (1 - q)h\left(\frac{a + b - q}{\kappa - 1}\right) \\
- q h\left(\frac{a + b - q}{2q}\right) \\
1 + h(q) - h(a) - h(b), \\
1 - h(a), \\
1 - h(b), \\
\end{cases} \quad (s, t) \in D_1, \\
(1 - q)h(q) - h(a) - h(b), \quad (s, t) \in D_2, \\
1 - h(a), \quad (s, t) \in D_3, \\
1 - h(b), \quad (s, t) \in D_4,
\]

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

\[
D_1 := \{(s, t) \in [0, 1]^2 : a * q \geq b, b * q \geq a, a * b \geq q\}, \\
D_2 := \{(s, t) \in [0, 1]^2 : a * b < q\}, \\
D_3 := \{(s, t) \in [0, 1]^2 : a * q < b\}, \\
D_4 := \{(s, t) \in [0, 1]^2 : b * q < a\}.
\]

The function \(\mathcal{Y}^*\) is plotted in Fig. 1.

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

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

\[
D_{a,b}(p) := D((p, a - p, b - p, 1 - p + a - b)), \\
D_{a,b} := \min_{0 < a + b - 1 \leq p \leq a, b} D_{a,b}(p).
\]

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

\[
D(a, b) = D(1 - a, 1 - b) \leq D(a, 1 - b) = D(1 - a, b).
\]

By Lemma 3, we have

\[
\varphi(s, t) = D(h^{-1}(1 - s), h^{-1}(1 - t)), \\
\psi(s, t) = D(h^{-1}(1 - s), 1 - h^{-1}(1 - t)).
\]

Then \(\varphi_1^*\) and \(\varphi_2^*\) are determined by \(\varphi(s, t)\) and \(\psi(s, t)\) via (44) and (45). Moreover, by Theorem 4, \(\Theta^*(E_1, E_2)\) is attained by a sequence involving the
time-sharing of at most three pairs of concentric\footnote{Here we call two Hamming spheres concentric if they have the same center and the radiiuses 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 radiiuses is not larger than $n/2$ while the other one is not smaller than $n/2$.} Hamming spheres, and $\Theta(E_1, E_2)$ 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) $\Theta(E_1, E_2)$ 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) $\Theta(E_1, E_2)$ 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 $\varphi = \Theta^*$, where by definition, $\Theta^* = \tilde{\varphi}$. On one hand, $\tilde{\varphi} \leq \varphi$. On the other hand, $\varphi \geq \tilde{\varphi} = \tilde{\varphi}$. Hence, $\varphi = \Theta^*$. This implies that the time-sharing random variable $W$ can be removed. By Theorem 4, time-sharing is not needed to attain $\Theta$, completing the proof of Statement 1.

Statement 2) follows similarly, but we need to prove that $\varphi(s, t) = \max s \leq E_1, t \leq E_2 \psi(s, t)$. This equality is equivalent to $\max Q_X, Q_Y : \mathbb{D}(Q_X || P_X) \leq s, \mathbb{D}(Q_Y || P_Y) \leq t \mathbb{D}(Q_X, Q_Y || P_{XY})$ is always attained by some $Q_X, Q_Y$ satisfying 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 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 $\mathbb{D}(Q_X || P_X) = s$, 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 $\mathbb{D}(Q_X || P_X) = s$. Similarly, they must also satisfy $\mathbb{D}(Q_Y || P_Y) = t$. These are the desired, which imply Statement 2.

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

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

**D. Applications to Zero-Error Coding**

As mentioned in [17], the minimization part of the conjecture of Ordentlich, Polyniak, 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) $(a, b) \in \{0, 1\}^n \mapsto a + b \in \{0, 1, 2\}$ and a code $(A_n, B_n)$ with $A_n, B_n \subseteq \{0, 1\}^n$ for this channel. Here $a + b$ denotes addition over $\mathbb{Z}$. When the code $(A_n, B_n)$ 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 $(a, b) \in A_n \times B_n$ is mapped to a unique sequence in $\{0, 1, 2\}$, i.e., $|A_n + B_n| = |A_n| \cdot |B_n|$, where $A_n + B_n$ denotes the sumset $\{a + b : a \in A_n, b \in B_n\}$. The zero-error capacity region $C$ 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 $A_n, B_n \subseteq \{0, 1\}^n$ with $|A_n| = 2^{n(R_1 + o(1))}, |B_n| = 2^{n(R_2 + o(1))}$ such that $|A_n + B_n| = |A_n| \cdot |B_n|$ 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 $A_n = \{0, 1\}^n$ 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\cite{62} to prove the best known upper bound. As mentioned by Ordentlich et al.\cite{17}, repeating the arguments in [62] with improved bounds on $\Theta(E_1, E_2)$ 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 $\Theta(E_1, E_2)$ in Theorem 4, we obtain the following result.

**Theorem 5:** If $(1 - \rho, R_2) \in C$, then for any $\rho \in (0, 1)$ there exists some $\lambda \in [0, \sqrt{\ln(2)/\rho} + 1]$ such that

$$\Theta^*(\epsilon, \rho) \leq \frac{(\lambda + \epsilon - R_2)}{\lambda},$$

$$\geq \lambda \left(1 - \frac{1}{2} - \log(3 - \rho)\right) - \frac{1}{2} - \epsilon - \sqrt{\frac{\ln(2)\epsilon}{2}},$$

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

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

Numerical results show that if $R_1 = 1$ (i.e., $\epsilon = 0$), by choosing the almost best $\rho = 0.6933$,\footnote{Equation (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}
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 \((X, Y) \sim P_{XY}^n\), where \(P_{XY}\) is a joint distribution defined on \(X \times Y\). Recall the notation \((f, g) = \mathbb{E}[f(X)g(Y)]\) and \(\|f\|_p = (\mathbb{E}|f(X)|^p)^{1/p}\). We continue to assume that \(X, Y\) are finite sets, each with cardinality at least 2, and with \(P_X(x) > 0, P_Y(y) > 0\) for all \((x, y) \in X \times Y\). 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 \(\alpha \in [0, E_{1,\text{max}}], \beta \in [0, E_{2,\text{max}}]\) and \(p, q \in (0, \infty)\), define

\[
\Lambda^{*}_{p,q}(\alpha, \beta) := \min_{\alpha \leq s \leq E_{1,\text{max}}, \beta \leq t \leq E_{2,\text{max}}} \left(\Theta^*(s, t) - \frac{s}{p} - \frac{t}{q}\right),
\]

and

\[
\overline{\Lambda}^{*}_{p,q}(\alpha, \beta) := \min_{\alpha \leq s \leq E_{1,\text{max}}, \beta \leq t \leq E_{2,\text{max}}} \left(\frac{s}{p} + \frac{t}{q} - \Theta^*(s, t)\right).
\]

Remark 10: Subsequent to the completion of this paper, the first author proved that \(\Theta^* = \varphi\) 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 \(\alpha \in [0, E_{1,\text{max}}], \beta \in [0, E_{2,\text{max}}]\). Let \(f, g\) be nonnegative functions on \(X^n\) and \(Y^n\) respectively such that \(P_X^n(\text{supp}(f)) \leq 2^{-\alpha n}, P_Y^n(\text{supp}(g)) \leq 2^{-\beta n}\). Then

\[
\langle f, g \rangle \leq 2^{-n\Lambda^*_{p,q}(\alpha, \beta)} \|f\|_p \|g\|_q, \quad \langle f, g \rangle \geq 2^{-n\overline{\Lambda}^*_{p,q}(\alpha, \beta)} \|f\|_p \|g\|_q.
\]

**Remark 11:** Given \(\alpha \in [0, E_{1,\text{max}}], \beta \in [0, E_{2,\text{max}}]\) 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 \(f_n = 1_{A_n}, g_n = 1_{B_n}\) with \((A_n, B_n)\) denoting the sets given in Remark 7 but with \((E_{1,\text{max}}, E_{2,\text{max}})\) there replaced by the optimal \((s^*, t^*)\) attaining the minimum in the definition of \(\Lambda^*_{p,q}(\alpha, \beta)\). Note that if \((\alpha, \beta)\) 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 \((\alpha, \beta)\) 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 \(f_n = 1_{A_n}, g_n = 1_{B_n}\) with some sequence \((A_n, B_n)\); see Remark 8.

**Remark 12:** A special case of Theorem 6 with \(p \geq p_0, q \geq q_0\) for (55) and \(p \leq p_1, q \leq q_1\) 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 \(\overline{\Theta}^*\) 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 \(\alpha \in [0, E_{1,\text{max}}], \beta \in [0, E_{2,\text{max}}]\), define the forward and reverse \((\alpha, \beta)\)-hypercontractivity regions as

\[
\mathcal{R}^+_{\alpha,\beta}(P_{XY}) := \{(p, q) \in (0, \infty)^2 : \Theta^*(E_1, E_2) \geq \frac{1}{p}E_1 + \frac{1}{q}E_2, \forall E_1 \in [\alpha, E_{1,\text{max}}], E_2 \in [\beta, E_{2,\text{max}}]\},
\]

\[
\mathcal{R}^-_{\alpha,\beta}(P_{XY}) := \{(p, q) \in (0, \infty)^2 : \overline{\Theta}^*(E_1, E_2) \leq \frac{1}{p}E_1 + \frac{1}{q}E_2, \forall E_1 \in [\alpha, E_{1,\text{max}}], E_2 \in [\beta, E_{2,\text{max}}]\}.
\]

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

\[
\langle f, g \rangle \leq \|f\|_p \|g\|_q, \forall (p, q) \in \mathcal{R}^+_{\alpha,\beta}(P_{XY}),
\]

\[
\langle f, g \rangle \geq \|f\|_p \|g\|_q, \forall (p, q) \in \mathcal{R}^-_{\alpha,\beta}(P_{XY}).
\]

**Remark 13:** These two inequalities are exponentially sharp in the same sense as (55) and (56); see Remark 11. That is, given \(\alpha \in [0, E_{1,\text{max}}], \beta \in [0, E_{2,\text{max}}]\) 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 \(f_n = 1_{A_n}, g_n = 1_{B_n}\) with some sequence \((A_n, B_n)\).

Given \((\alpha, \beta)\) 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 \(f_n = 1_{A_n}, g_n = 1_{B_n}\) with some sequence \((A_n, B_n)\).

Note that the hypercontractivity inequalities in Theorem 6 differ from the common ones in the factors \(2^{-n\Lambda^*_{p,q}(\alpha, \beta)}\) 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 have been studied in this paper. One may type graphs, it seems not hopeless to characterize them.

The maximal density of subgraphs of a type graph of $T_{XY}$ with size $(M_1, M_2)$ as

$$\Gamma_n(M_1, M_2) := \min_{A \subseteq T_{XY}, B \subseteq T_{XY}} \rho(G[A, B]).$$

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

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

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

For some positive sequences $\{\varepsilon_1, n\}$ and $\{\varepsilon_2, n\}$ which both vanish as $n \to \infty$, where

$$\Gamma_n^{(l)}(T_{XY}) := \{ (R_1, R_2) : R_1 \leq H(X), R_2 \leq H(Y) \},$$

$$\Gamma_n^{(l)}(T_{XY}) := \bigcup_{P_X, P_{X|W}, P_{Y|W} : P_W, P_{X|W}, P_{Y|W} \text{ are } n\text{-types}, P_X = P_Y = T_Y, Q_X \neq T_{XY} \neq Q_Y \neq T_{XY} \in \mathcal{P}_{X|W}, P_{Y|W}} \{ (R_1, R_2) : R_1 \leq H(X|W), R_2 \leq H(Y|W) \}.$$

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 $\Gamma_n(T_{XY})$ could be of interest. However, currently, we have no idea how to tackle it. In addition, if $\Gamma_n(T_{XY})$ is not asymptotically equal to $\Gamma_n^{(l)}(T_{XY})$, then 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 $P_{XY|W} = T_{XY} P_W$ (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), $F^*(0, R_2) \leq \min\{R_2, H_T(Y|X)\}$. On the other hand, if $R_2 \geq H_T(Y|X)$, then we choose $W = X$, which leads to $H(X|W) = 0$ and $H(Y|W) = H(X, Y|W) = H(Y|X)$. Hence we have $F^*(0, R_2) = H_T(Y|X)$ for $R_2 \geq H_T(Y|X)$. If $R_2 < H_T(Y|X)$, 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 $X \cup \emptyset$ and $J$ defined on $0, 1$ such that $V = J = 0$ if $J = 1$, where $J \sim \text{Bern}(\alpha)$ for $\alpha := R_2/H_T(Y|X)$ 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 $R_2 < H_T(Y|X)$.

Statement 4): Let $P_{XY|W}$ attain $F^*(R_1, R_2)$, and $P_{XY|W}$ attain $F^*(R_1, R_2)$. For $0 < \alpha < 1$, define $J \sim \text{Bern}(\alpha)$ independent of $(X, Y, W_0, W_1)$ and let $W := W_j$, taking values in $W_0 \cup W_1$, where $W_j$ denotes the alphabet of $W_j$ for $j = 0, 1$. Note that $J$ is a deterministic function of $W$. Then $P_{XY|W}$ induces

$$H(X, Y|W) = \alpha H(X, Y|W_0) + (1 - \alpha) H(X, Y|W_1),$$

$$H(X|W) = \alpha H(X|W_0) + (1 - \alpha) H(X|W_1),$$

$$H(Y|W) = \alpha H(Y|W_0) + (1 - \alpha) H(Y|W_1).$$

Therefore,

$$F^* (\alpha R_1 + (1 - \alpha) \hat{R}_1, \alpha R_2 + (1 - \alpha) \hat{R}_2) \geq \alpha F^*(R_1, R_2) + (1 - \alpha) F^*(\hat{R}_1, \hat{R}_2).$$

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

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

is nonincreasing and concave, by Statements 1) and 4). Hence, for fixed $\delta_2$,

$$f(t + \delta_2) - f(t) \leq \frac{f(\delta_2) - f(0)}{\delta_2} \leq \frac{\delta_2 + \min\{R_1, H_T(X|Y)\} - \min\{R_1, H_T(X|Y)\}}{\delta_2} = 1.$$

Setting $t = R_2$, we obtain $F^*(R_1, R_2 + \delta_2) - F^*(R_1, R_2) \leq \delta_2$, 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 $\frac{R_1}{\delta_1} \geq \frac{R_2}{\delta_2}$. For $t \geq -\frac{R_2}{\delta_2}$, define

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

By Statements 1) and 4), $g(t)$ is nonincreasing and concave. Hence, for fixed $\delta_2$,

$$g(t + 1) - g(t)$$
is nonincreasing in $t$. Combining this with Statements 2) and 3) yields that for $t \geq -\frac{R_2}{\delta_2}$ we have

$$g(t+1) - g(t) \leq g(-\frac{R_2}{\delta_2} + 1) - g(-\frac{R_2}{\delta_2})$$

$$= F^*(R_1 - \frac{\delta_1 R_2}{\delta_2}) + \delta_1 - F^*(R_1 - \frac{\delta_1 R_2}{\delta_2}, 0)$$

$$\leq \min\{R_1 - \frac{\delta_1 R_2}{\delta_2} + \delta_1, H_T(X|Y)\} + \delta_2$$

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

**APPENDIX B
PROOF OF THEOREM 1**

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

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

$$\Gamma_n(2^nR_1, 2^nR_2) = \frac{2^nH(X, Y)}{2^nR_1, 2^nR_2},$$

$$\frac{1}{n} H(X) \leq R_1,$n H(Y) \leq R_2,$

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,

$$E_n(R_1, R_2) = R_1 + R_2 - \frac{1}{n} H(X,Y)$$

$$= R_1 + R_2 - \frac{1}{n} \sum_{i=1}^{n} H(X_i, Y_i | X_i^{i-1}, Y_i^{i-1})$$

$$= R_1 + R_2 - H(X, Y | X^{J-1}, Y^{J-1}, J),$$

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

$$H(X, Y | X^{J-1}, Y^{J-1}, J) \leq H(X, Y | X^{J-1}) = \frac{1}{n} H(X) \leq R_1,$$

$$H(Y, X^{J-1}, Y^{J-1}, J) \leq R_2.$$ 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

$$E_n(R_1, R_2) \geq \inf_{H(X, W) \leq R_1, H(Y, W) \leq R_2} -H(X, Y | W)$$

$$= E^*(R_1, R_2).$$

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

$$|(A \times B) \cap T_{T_{XY}}| \geq |T_{P_{X,Y,W}}(w)| \geq 2^nH(X,Y|W) - \frac{|\mathcal{W}||\mathcal{X}||\mathcal{Y}|}{n} \log(n+1).$$

where

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

Thus we have

$$\rho(G[A,B]) = \frac{|(A \times B) \cap T_{T_{XY}}|}{2^nR_1, 2^nR_2} \geq 2^nR_1 - H(X,Y|W) - \frac{|\mathcal{W}||\mathcal{X}||\mathcal{Y}|}{n} \log(n+1),$$

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

$$E_n(R_1, R_2) \leq R_1 + R_2 - F_n(R_1, R_2)$$

$$+ |\mathcal{W}||\mathcal{X}||\mathcal{Y}| \frac{\log(n+1)}{n},$$

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

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 $P_{X,Y,W}$ with $P_{XY} = T_{XY}$, one can find a $n$-type $Q_{X,Y,W}$ with $Q_{XY} = T_{XY}$ such that $||P_{X,Y,W} - Q_{X,Y,W}|| \leq \frac{|\mathcal{W}||\mathcal{X}||\mathcal{Y}|}{2n}$, where $|| \cdot ||$ denotes the TV distance, see [64, Lemma 3]. Combining this with [3, Lemma 2.7] (i.e., if $\|P_X - Q_X\| \leq \Theta \leq \frac{1}{4}$, then $\|H_P(X) - H_Q(X)\| \leq -2\Theta \log \frac{2\Theta}{|\mathcal{X}|}$), we have for $\frac{|\mathcal{W}||\mathcal{X}||\mathcal{Y}|}{2n}$, that

$$|H_{P_{X,W}} - H_{Q_{X,W}}|$$

$$\leq |H_{P_{X,W}} - H_{Q_{X,W}}| + |H_{P(W)} - H_{Q(W)}|$$

$$\leq -2 \frac{|\mathcal{W}||\mathcal{X}||\mathcal{Y}|}{2n} \log \frac{2|\mathcal{W}||\mathcal{X}||\mathcal{Y}|}{|\mathcal{W}|} - 2 \frac{|\mathcal{W}||\mathcal{X}||\mathcal{Y}|}{2n} \log \frac{2|\mathcal{W}||\mathcal{X}||\mathcal{Y}|}{|\mathcal{W}|}$$

$$\leq -2 \frac{|\mathcal{W}||\mathcal{X}||\mathcal{Y}|}{2n} \log \frac{2|\mathcal{W}||\mathcal{X}||\mathcal{Y}|}{|\mathcal{W}|} - 2 \frac{|\mathcal{W}||\mathcal{X}||\mathcal{Y}|}{2n} \log \frac{2|\mathcal{W}||\mathcal{X}||\mathcal{Y}|}{|\mathcal{W}|}$$

(61)
...for $\alpha \leq \frac{n}{2}$, 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 $1 \leq d \leq n - 1$. Let $(P_{XY}, Q_{XY})$ be a pair comprised of a $d$-joint type and an $(n-d)$-joint type on $X \times Y$ such that $\frac{d}{n} P_{XY} + (1 - \frac{d}{n}) Q_{XY} = T_{XY}$. For a fixed length-$d$ sequence $y$ with type $P_{y}$ and a fixed length-$(n-d)$ sequence $x$ with type $Q_{x}$, we choose $A = T_{0_{x}y} \times \{x\}$ and $B = \{y\} \times T_{Q_{y}X(x)}$. Then, from [3, Lemma 2.5], we have $|A| \geq 2^{d(H_{P}(X|Y) - \frac{1}{d}(n-d) \log (n-d+1))}$ and similarly $|B| \geq 2^{(n-d)(H_{Q}(Y|X) - \frac{1}{d}(n-d) \log (n-d+1))}$. On the other hand, for this code we have $A \times B \subseteq T_{XY}$. Hence any rate pair $(R_{1}, R_{2}) \in (\mathcal{R}_{X}^{(n)} \times \mathcal{R}_{Y}^{(n)})$ with

$$R_{1} \leq \frac{d}{n} (H_{P}(X|Y) - \frac{|X||Y| \log (d+1)}{n})$$

and

$$R_{2} \leq \frac{1 - \frac{d}{n}}{n} (H_{Q}(Y|X) - \frac{|X||Y| \log (n-d+1)}{n - d})$$

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

$$R_{1} \leq \frac{d}{n} H_{P}(X|Y) - \frac{|X||Y| \log (n+1)}{n}$$

and

$$R_{2} \leq \frac{1 - \frac{d}{n}}{n} H_{Q}(Y|X) - \frac{|X||Y| \log (n+1)}{n}$$

is achievable.

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

$$\alpha - |X||Y| \leq d \leq \alpha.$$  

We first consider the case

$$4 |X||Y| \leq d \leq n - 4 |X||Y|.$$  

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

$$H_{P}(X|Y) \geq H_{P}(X|Y) + \frac{2|X||Y|}{d} \log \frac{4|X|}{d^{2}},$$

$$H_{Q}(Y|X) \geq H_{Q}(Y|X) + \frac{2|X||Y|}{n - d} \log \frac{4|X|}{(n - d)^{2}}.$$  

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

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

$$\geq \alpha H_{P}(X|Y) - \frac{|X||Y|}{n} \log |X| - \frac{2|X||Y|}{n} \log \frac{n^{2}}{4|X|} - \frac{|X||Y| \log (n+1)}{n}$$

$$= \alpha H_{P}(X|Y) - \epsilon_{1,n}.$$  

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 $R_{1} \leq \frac{d}{n} H_{P}(X|Y) - \epsilon_{1,n}$ and $R_{2} \leq (1 - \alpha) H_{Q}(Y|X) - \epsilon_{2,n}$, for any $0 \leq \alpha \leq 1$ and $(\bar{P}_{XY}, \bar{Q}_{XY})$ a pair of distributions such that $\alpha \bar{P}_{XY} + (1 - \alpha) \bar{Q}_{XY} = T_{XY}$, is achievable as long as the condition in (68) holds.

We next consider the case $0 \leq d \leq 4 |X||Y|$. For this case, we have

$$\alpha H_{P}(X|Y) \leq \frac{d}{n} + \frac{|X||Y|}{n} \log |X|$$

$$\leq \frac{5|X||Y|}{n} \log |X| \leq \epsilon_{1,n},$$

$$\leq \frac{5|X||Y|}{n} \log |X| \leq \epsilon_{1,n}.$$
where the first inequality follows by (67) and the fact that $H_p(X|Y) \leq \log |\mathcal{X}|$. Hence

$$\{(R_1, R_2) : R_1 \leq \alpha H_p(X|Y), R_2 \leq (1-\alpha)H_Q(Y|X)\}$$

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 $n - 4|\mathcal{X}||\mathcal{Y}| < d \leq n$ 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 $P_{XY}$, 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 $E_P[X], E_P[Y]$, the covariance $\text{Cov}_P(X, Y)$, and the correlation $E_P[XY]$. Translating the choices of $\mathcal{X}$ and/or $\mathcal{Y}$ (as subsets of $\mathbb{R}$) does not change $\text{Cov}_P(X, Y)$, so we can ensure that we make these choices in such a way that $E_P[XY] = \text{Cov}_P(X, Y) + E_P[X]E_P[Y] = 0$.

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 $A \times B \subseteq T_{XY}$, we will have $(x, y) = 0$ for any $(x, y) \in A \times B$, where $x, y$ are now viewed as row vectors in $\mathbb{R}^n$. Let $\mathcal{A}$ denote the linear space spanned by all the vectors in $A$, and let $\mathcal{B}$ denote the linear space spanned by all the vectors in $B$. Hence $\mathcal{B} \subseteq \mathcal{A}^\perp$, where $\mathcal{A}^\perp$ denotes the orthogonal complement of a subspace $\mathcal{A}$. As an important property of the orthogonal complement, $\dim(\mathcal{A}) + \dim(\mathcal{A}^\perp) = n$. Hence $\dim(\mathcal{A}) = \dim(\mathcal{B}) \leq n$.

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 $V_1, V_2$ 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 $\{J_1, J_2\}$ of $[n]$ such that $|J_1| = n_1$ and $x = f_1(x_J), \forall x \in V_1, i = 1, 2$ for some deterministic linear functions $f_i : \mathbb{R}^{n_i} \to \mathbb{R}^n$, where $x_J := (x_j)_{j \in J}$.

**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 $\sigma$ of $[n]$ such that $\tilde{x}(\sigma) = f_1(x_{\sigma_1}^{[1:n_1]}), \forall x \in V_1$ and $\tilde{x}(\sigma) = f_2(x_{\sigma_{n_1+1:n_2]}^{[n_1+1:n_2]}), \forall x \in V_2$ for some deterministic functions $f_1 : \mathbb{R}^{n_1} \to \mathbb{R}^n$, where $x(\sigma)$ is obtained by permuting the components of $x$ using $\sigma$.

Let $d$ denote $\dim(\mathcal{A})$, so we have $\dim(\mathcal{A}^\perp) = n - d$. Let $X \sim \text{Unif}(A), Y \sim \text{Unif}(B)$ be two independent random vectors, i.e., $(X, Y) \sim P_{XY} := \text{Unif}(A)\text{Unif}(B)$. Now we choose $V_1 = \mathcal{A}, V_2 = \mathcal{A}^\perp, X_1 = X, X_2 = Y$ in Lemma 4. Then there exists a partition $\{J, J^c\}$ of $[n]$ such that $|J| = d$ and $X = f_1(x_J, Y = f_2(Y_{J^c})$ 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

$$R_1 = \frac{1}{n}H(X) = \frac{1}{n}H(X|Y) = \frac{1}{n}H(X_J|Y)$$

$$\leq \frac{1}{n}H(X_J|Y_J) \leq \frac{1}{n} \sum_{j \in J} H(X_j|Y_j)$$

$$= \frac{d}{n}H(X_j|Y_j, J) = \frac{d}{n}H(X_j|Y_j) = \frac{d}{n}H(X_j|Y)$$

where $J \sim \text{Unif}(J), X_j := X_j, Y_j := Y_j$, with $J$ being independent of $(X, Y)$. Similarly, we have

$$R_2 = \frac{1}{n}H(Y) = \frac{1}{n}H(Y|X) = \frac{1}{n}H(Y_{J^c}|X)$$

$$\leq \frac{1}{n}H(Y_{J^c}|X_J) \leq \frac{1}{n} \sum_{j \in J^c} H(Y_j|X_j)$$

$$= (1 - \frac{d}{n})H(Y_j|X_J, J)$$

$$\leq (1 - \frac{d}{n})H(Y_j|X_J) = (1 - \frac{d}{n})H(Y|X)$$

where $\tilde{J} \sim \text{Unif}(J^c), \tilde{X}_j := X_j, \tilde{Y}_j := Y_j$, with $\tilde{J}$ being independent of $(X, Y, J)$. On the other hand,

$$\frac{d}{n}P_{XY} + (1 - \frac{d}{n})P_{\tilde{X}\tilde{Y}}$$

$$= \frac{1}{n} \sum_{j \in J} P_{X_JY_j} + \frac{1}{n} \sum_{j \in J^c} P_{X_JY_j}$$

$$= \frac{1}{n} \sum_{j=1} P_{X_JY_j} = E_{(X, Y)}[T_{XY}] = T_{XY},$$

where $T_{XY}$ denotes the joint type of a random pair $(X, Y)$ 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 $(V, V^\perp)$ with dimensions respectively $n_1, n- n_1$, let $\{u_j : 1 \leq j \leq n_1\}$ be an orthogonal basis of $V$, and $\{u_j : n_1 + 1 \leq j \leq n\}$ be an orthogonal basis of $V^\perp$. Then $\{u_j : 1 \leq j \leq n\}$ forms an orthogonal basis of $\mathbb{R}^n$. Denote by $U$ the $n \times n$ matrix with $j$-th row being $u_j$. Then $U$ is orthogonal. We now express $x \in V$ and $y \in V^\perp$, thought of as row vectors, in terms of this orthogonal basis, i.e.,

$$x = \tilde{x}U, \quad y = \tilde{y}U^\top,$$

where $\tilde{x} := xU^\top, \tilde{y} := yU^\top$, and $U^\top$ is the transpose of $U$. Since for any $x \in V$ we have $\langle x, u_j \rangle = 0$ for all $n_1 + 1 \leq j \leq n$, we obtain that $\tilde{x}_j = 0$ for all $n_1 + 1 \leq j \leq n$. Similarly,
\[\hat{y}_j = 0 \text{ for all } 1 \leq j \leq n_1. \text{ Hence we can rewrite } \hat{x} = (\hat{x}_1, 0), \hat{y} = (0, \hat{y}_2). \text{ We write } U \text{ in a block form: } U = \begin{bmatrix} U_1 \\ U_2 \end{bmatrix} \text{ where } U_1, U_2 \text{ are respectively of size } n_1 \times n, (n - n_1) \times n. \text{ Then}
\]
\[x = \hat{x}_1 U_1, \quad y = \hat{y}_2 U_2. \quad (76)\]

We now need the following well-known exchange lemma.

**Lemma 5 (Exchange Lemma):** [65, Theorem 3.2] Let \( k \geq 2 \) be an integer. Let \( B \) be an \( n \times n \) nonsingular matrix, and \( \{H_i, 1 \leq i \leq k\} \) be a partition of \([n] \). Then there always exists another partition \( \{L_i, 1 \leq l \leq k\} \) of \([n]\) such that all the sub-matrices \( B_{H_i,L_i}, 1 \leq l \leq k \) are nonsingular.

The proof of this lemma follows easily from repeated use of the Laplace expansion for determinants. A short proof in the Appendix, there must be at least one choice of \( \phi \) such that both \( \phi = \) 0 and \( \phi = \) 1.

We now need the following well-known exchange lemma.

**Lemma 5 (Exchange Lemma):** [65, Theorem 3.2] Let \( k \geq 2 \) be an integer. Let \( B \) be an \( n \times n \) nonsingular matrix, and \( \{H_i, 1 \leq i \leq k\} \) be a partition of \([n] \). Then there always exists another partition \( \{L_i, 1 \leq l \leq k\} \) of \([n]\) such that all the sub-matrices \( B_{H_i,L_i}, 1 \leq l \leq k \) are nonsingular.

The proof of this lemma follows easily from repeated use of the Laplace expansion for determinants. A short proof in the Appendix, there must be at least one choice of \( \phi \) such that both \( \phi = \) 0 and \( \phi = \) 1.

Substituting these back into (76), we obtain that
\[x = \hat{x}_1 U_{1,j}, \quad y = \hat{y}_{j'} U_{2,j'.} \quad (76)\]

Substituting these back into (76), we obtain that
\[(x_j, y_{j'}) = x_j U_{1,j}^{-1} U_{1,j}, \quad y_{j'} = y_{j'} U_{2,j'}^{-1} U_{2,j'} \quad (76)\]

Hence the proof is completed.

**APPENDIX E**

**PROOF OF PROPOSITION 2**

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

if and only if
\[\max_{0 \leq \alpha \leq 1, P_{XY}, Q_{XY}} \varphi_\alpha(P_{XY}, Q_{XY}) = \varphi_\alpha(P_{XY}, Q_{XY}) = \leq 1. \quad (77)\]

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 \( \beta_1 := H_P(X|Y), \beta_2 := H_P(Y|X) \). Here the domain of \( \varphi_\alpha \) can be taken to be the set of pairs of probability distributions \((P_{XY}, Q_{XY})\) such that \( \text{supp}(P_{XY}) = \text{supp}(Q_{XY}) \subseteq \text{supp}(T_{XY}) \). Moreover, (77) can be rewritten as that for any \( 0 \leq \alpha \leq 1, \quad (78) \]

\[\max_{P_{XY}, Q_{XY}: \alpha P_{XY} + (1-\alpha)Q_{XY} = T_{XY}} \varphi_\alpha(P_{XY}, Q_{XY}) = \leq 1. \quad (78)\]

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 \( \varphi_\alpha \) is taken to be the set of pairs of probability distributions \((P_{XY}, Q_{XY})\) such that \( \text{supp}(P_{XY}) = \text{supp}(Q_{XY}) = \text{supp}(T_{XY}) \).

We are thus led to consider the Lagrangian
\[L = \varphi_\alpha(P_{XY}, Q_{XY}) + \sum_{(x,y) \in \text{supp}(T_{XY})} \eta(x,y)(\alpha P(x,y) + (1-\alpha)Q(x,y) - T(x,y)) + \mu_1 \left( \sum_{(x,y) \in \text{supp}(T_{XY})} P(x,y) - 1 \right) + \mu_2 \left( \sum_{(x,y) \in \text{supp}(T_{XY})} Q(x,y) - 1 \right).\]

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:
\[\frac{\partial L}{\partial P(x,y)} = -\frac{\alpha}{\beta_1} \log P(x,y) + \alpha \eta(x,y) + \mu_1 = 0, \forall (x,y) \in \text{supp}(T_{XY}), \quad (79)\]
\[\frac{\partial L}{\partial Q(x,y)} = -\frac{1-\alpha}{\beta_2} \log Q(y|x) + (1-\alpha) \eta(x,y) + \mu_2 = 0, \forall (x,y) \in \text{supp}(T_{XY}), \quad (80)\]
\[
\alpha P(x, y) + (1 - \alpha)Q(x, y) = T(x, y), \\
\forall (x, y) \in \text{supp}(T_{XY}),
\]
\[
\sum_{(x, y) \in \text{supp}(T_{XY})} P(x, y) = 1,
\]
\[
\sum_{(x, y) \in \text{supp}(T_{XY})} Q(x, y) = 1,
\]
\[
P(x, y), Q(x, y) > 0, \forall (x, y) \in \text{supp}(T_{XY}),
\]
for some reals \(\eta(x, y), \mu_1, \mu_2\) with \((x, y) \in \text{supp}(T_{XY})\). Here the conditions in (84) come from the restriction we have imposed on the domain of \(\varphi_\alpha\).

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 \(\eta(x, y) = \frac{1}{\mu_1} \log T(x|y) = \frac{1}{\mu_2} \log T(y|x)\), \(\mu_1 = \mu_2 = 0\), 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
\[
\frac{\mu_1}{1 - \alpha} = \frac{\mu_2}{1 - \beta}.
\]
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 \(0 \leq \beta \leq \alpha \leq \frac{1}{2}\). Then, in the definition of \(D(\alpha, \beta)\), we minimize \(D_{\alpha, \beta}(p)\) over \(0 \leq p \leq \beta\). Furthermore,
\[
D_{\alpha, \beta}(p) = -H(p, \alpha - p, \beta - p, 1 + p - \alpha - \beta) - (1 + 2p - \alpha - \beta) \log(1 + p) - (\alpha + 2p) \log(1 - p) + \log 4.
\]
Let \(s := \alpha + \beta - 2p\). Then we have
\[
D_{\alpha, \beta}(p) = -H(p, \alpha - p, \beta - p, 1 + p - \alpha - \beta)\big|_{p = \frac{s + \alpha - \beta}{2}}
- (1 - s) \log(1 + p) - s \log(1 - p) + \log 4.
\]
By definition, \(D(\alpha, \beta)\) can be rewritten as the minimum of \(D_{\alpha, \beta}(p)\) over \(\alpha - \beta \leq s \leq \alpha + \beta\). Given \((\alpha, \beta)\), \(D_{\alpha, \beta}(p)\) 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 \(\log \frac{1 + p}{1 - p}\), which is nonnegative. This implies that the minimum of \(D_{\alpha, \beta}(p)\) is attained at some point \(s\) such that \(\alpha - \beta \leq s \leq \alpha + \beta - 2\alpha\beta\) (or equivalently, at some \(p \in (\alpha, \beta)\)). In other words, without changing the value of \(D(\alpha, \beta)\), one can replace \(H(p, \alpha - p, \beta - p, 1 + p - \alpha - \beta)\) above with
\[
H(p, \alpha - p, \beta - p, 1 + p - \alpha - \beta) := \begin{cases} 
H(p, \alpha - p, \beta - p, 1 + p - \alpha - \beta), & p \in (\alpha, \beta, \beta), \\
h(\alpha) + h(\beta), & p \in (-\infty, \alpha, \beta].
\end{cases}
\]
That is, \(D(\alpha, \beta)\) is equal to the minimum of
\[
-H(p, \alpha - p, \beta - p, 1 + p - \alpha - \beta)\big|_{p = \frac{s + \alpha - \beta}{2}}
- (1 - s) \log(1 + p) - s \log(1 - p) + \log 4
\]
over \(s \geq \alpha - \beta\).

We next deal with \((1 - \alpha, \beta)\). In the definition of \(D(1 - \alpha, \beta)\), we minimize \(D_{1 - \alpha, \beta}(p)\) over the same range \(0 \leq p \leq \beta\). Furthermore,
\[
D_{1 - \alpha, \beta}(p) = -H(p, 1 - \alpha - p, \beta - p, \alpha + p - \beta) - (1 + 2p - \beta) \log(1 + p) - (1 - \alpha + \beta - 2p) \log(1 - p) + \log 4.
\]
Let \(t := 1 - \alpha + \beta - 2p\). Then, similarly to the above, we have
\[
D_{1 - \alpha, \beta}(p) = -H(p, \alpha - p, \beta - p, 1 + p - \alpha - \beta)\big|_{p = \frac{s + \alpha - \beta}{2}}
- (1 - t) \log(1 + p) - t \log(1 - p) + \log 4,
\]
which can be seen by verifying that
\[
-H(p, \alpha - p, \beta - p, 1 + p - \alpha - \beta)\big|_{p = \frac{s + \alpha - \beta}{2}}
= -H(p, 1 - \alpha - p, \beta - p, \alpha + p - \beta)\big|_{p = \frac{s + \alpha - \beta}{2}}.
\]
Hence, \(D(1 - \alpha, \beta)\) is equal to the minimum of \(D_{1 - \alpha, \beta}(p)\) over \(1 - \alpha - \beta \leq t \leq 1 - \alpha + \beta\). Given \((\alpha, \beta)\), \(D_{1 - \alpha, \beta}(p)\) 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 \(\log \frac{1 + p}{1 - p}\) which is nonnegative. Hence \(D(1 - \alpha, \beta)\) is equal to the minimum of \(D_{1 - \alpha, \beta}(p)\) over \(1 - \alpha - \beta \leq t \leq 1 - \alpha - \beta + 2\alpha\beta\).

To prove \(D(1 - \alpha, \beta) \geq D(\alpha, \beta)\), it suffices to show that
\[
-H(p, \alpha - p, \beta - p, 1 + p - \alpha - \beta)\big|_{p = \frac{s + \alpha - \beta}{2}}
\geq -H(p, \alpha - p, \beta - p, 1 + p - \alpha - \beta)\big|_{p = \frac{s + \alpha - \beta}{2}}
\]
for all \(1 - \alpha - \beta \leq s \leq 1 - \alpha - \beta + 2\alpha\beta\). By the definition of \(H\), we only need to check
\[
g(s) := H(p, \alpha - p, \beta - p, 1 + p - \alpha - \beta)\big|_{p = \frac{s + \alpha - \beta}{2}}
-H(p, \alpha - p, \beta - p, 1 + p - \alpha - \beta)\big|_{p = \frac{s + \alpha - \beta}{2}}
\geq 0
\]
for \(1 - \alpha - \beta \leq s \leq 1 - \alpha - \beta + 2\alpha\beta\). If \(\alpha + \beta - 2\alpha\beta < 1 - \alpha - \beta\) there is nothing to show. We may therefore assume that \(1 - \alpha - \beta \leq \alpha + \beta - 2\alpha\beta\).

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 \(g(\alpha + \beta - 2\alpha\beta) \geq 0\) since the maximum of the first entropy in (87) is attained at \(s = \alpha + \beta - 2\alpha\beta\). Hence, we have \(g \geq 0\) 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 \(D(\alpha, \beta)\) is equal to the asymptotic exponent of \(P_{XY}^n(\mathcal{A}_n \times \mathcal{B}_n)\) where
\(A_n, B_n\) are type classes with types asymptotically converging to \((\alpha, 1-\alpha)\) and \((\beta, 1-\beta)\) respectively. In other words, \(A_n, B_n\) are the concentric Hamming spheres with common center \((0, \ldots, 0)\) with radii \(r_n, s_n\) satisfying \(r_n/n \rightarrow \alpha, s_n/n \rightarrow \beta\) as \(n \rightarrow \infty\). Similarly, \(D(1-\alpha, \beta)\) is equal to the asymptotic exponent of \(P_{XY}^n(A_n \times B_n)\) where \(A_n\) is the anti-concentric Hamming sphere of \(A_n\). Hence, the type of \(A_n\) converges to \((1-\alpha, \alpha)\) asymptotically. On the other hand, we can write \(y \in B_n\),
\[
P^n_{XY}(A_n \times B_n) = P^n_Y(B_n) \left(\frac{1}{2^n} \right) \sum_{x \in A_n} (1 - \frac{1}{2^n})^d(x,y).
\]

By permutation, one can observe that the expression above remains the same for all \(y \in B_n\). Hence,
\[
P^n_{XY}(A_n \times B_n) = P^n_Y(B_n) \left(\frac{1}{2^n} \right) \sum_{x \in \mathbb{A}_n} (1 - \frac{1}{2^n})^d(x,y).
\]

Denote \(\eta := \frac{1-\rho}{2^n}\) and denote \(F\) as the CDF of the distance \(d(X,Y)\) with \(X \sim \text{Unif}(A_n)\). Then, we have
\[
\mathbb{E}_x \left[ \sum_{x \in \mathbb{A}_n} \eta^d(x,y) \right] = \mathbb{E}_x \left[ \sum_{x \in \mathbb{A}_n} \eta^d(x,y) \right] = \frac{1}{|\mathbb{A}_n|} \sum_{x \in \mathbb{A}_n} \eta^d(x,y).
\]

Similarly,
\[
\mathbb{E}_x \left[ \sum_{x \in \mathbb{A}_n} \eta^d(x,y) \right] = \mathbb{E}_x \left[ \sum_{x \in \mathbb{A}_n} \eta^d(x,y) \right] = \frac{1}{|\mathbb{A}_n|} \sum_{x \in \mathbb{A}_n} \eta^d(x,y).
\]

where \(G\) is the CDF of the distance \(d(X,Y)\) with \(X \sim \text{Unif}(A_n)\). Since the sphere \(A_n\) is “closer” to \(y \in B_n\) than the sphere \(X_n\), intuitively, \(F(d) \geq G(d)\) for all \(d \geq 0\) which implies \(P_{XY}^n(A_n \times B_n) \geq P_{XY}^n(X_n \times B_n)\). This further implies \(D(1-\alpha, \beta) \geq \Omega(\alpha, \beta)\). In the proof above, we showed that the asymptotic exponent of (88) is not larger than that of (89), with \(n,s\), \(nt\) denoting the distances. This is a weaker version of \(P_{XY}^n(A_n \times B_n) \geq P_{XY}^n(X_n \times B_n)\), but it still implies \(D(1-\alpha, \beta) \geq \Omega(\alpha, \beta)\).

**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
\[
\Lambda_{p,q}^{(n)}(\alpha, \beta) := -\frac{1}{n} \log \sup_{f,g: P_X^k(supp(f)) \leq \varepsilon} \frac{\langle f, g \rangle}{\|f\|_p \|g\|_q},
\]
\[
\Xi_{p,q}^{(n)}(\alpha, \beta) := -\frac{1}{n} \log \inf_{f,g: P_X^k(supp(f)) \leq \varepsilon} \frac{\langle f, g \rangle}{\|f\|_p \|g\|_q},
\]
satisfy that \(n\Lambda_{p,q}^{(n)}(\alpha, \beta)\) is subadditive and \(n\Xi_{p,q}^{(n)}(\alpha, \beta)\) is superadditive in \(n\). So, by Fekete’s lemma, \(\inf_{n \geq 1} \Lambda_{p,q}^{(n)}(\alpha, \beta) = \lim_{n \rightarrow \infty} \Lambda_{p,q}^{(n)}(\alpha, \beta)\) and

\[
\sup_{n \geq 1} \Xi_{p,q}^{(n)}(\alpha, \beta) = \lim_{n \rightarrow \infty} \Xi_{p,q}^{(n)}(\alpha, \beta),
\]

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 \(f \leq \frac{1}{n} P_X^k g \leq \frac{1}{n} P_{XY}^k\), and moreover, \(\frac{1}{n} \log \|f\|_p + \frac{1}{n} \log \|g\|_q\) are uniformly bounded for all \(n \geq 1\). This is because given \(|f|_1 = 1\), for \(p \geq 1\), we have
\[
1 = \|f\|_1 \leq \|f\|_p \leq \|f\|_\infty \leq \frac{1}{P_{XY}^k},
\]
and for \(0 < p \leq 1\), we have
\[
\|f\|_p^p \leq \|f\|_p \leq \|f\|_1 = 1.
\]

For sufficiently large \(a > 0\), the points at which \(f \leq 2^{-na}\) 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 \(o(n_i)\), where \(o(n_i)\) denotes a term vanishing as \(n \rightarrow \infty\) uniformly over all \(f\) and \(g\) with \(|f|_1 = |g|_1 = 1\). This is because,
\[
|f_a|_p \leq |f_a|_p^p \leq |f_a|_p^p + 2^{-npa},
\]
and
\[
\mathbb{E}[f_a g_a] \leq \mathbb{E}[f g] \leq \mathbb{E}[f_a g_a] + 3 \cdot 2^{-an}.
\]

All the remaining points of \(\mathbb{X}^n\) can be partitioned into \(r = r(a, b)\) level sets \(A_1, \ldots, A_r\) 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 \(\mathbb{Y}^n\) can be partitioned into \(s = s(a, b)\) level sets \(B_1, \ldots, B_s\) such that \(g\) varies by a factor of at most \(2^{nb}\) in each level set. Let \(\alpha_i := -\frac{1}{n} \log P_X^k (A_i), \beta_i := -\frac{1}{n} \log P_{XY}^k (A_i), \mu_i := \frac{1}{n} \log (u_i), \nu_i := \frac{1}{n} \log (v_i)\), where \(u_i, v_i\) are respectively the median value of \(f\) on \(A_i\) and the median value of \(g\) on \(B_i\). (If \(A_i\) is empty then \(u_i\) can be chosen to be any value within the level set defining \(A_i\), and similarly for \(B_i\) and \(v_i\).) Note that \(f(x) \in [u_i 2^{-nb}, u_i 2^{nb}]\) on the set \(A_i\), and \(g(y) \in [v_i 2^{-nb}, v_i 2^{nb}]\) on the set \(B_i\). Moreover, \(\alpha_i \geq \alpha, \beta_i \geq \beta, \forall i, j\). Then,
\[
\frac{1}{n} \log \|f\|_p \geq \frac{1}{n} \log \|f_a\|_p \geq \frac{1}{n} \log \|f_a\|_p - b \geq N_X(p) - b,
\]
where \(N_X(p) := \max_{1 \leq i \leq r} \{-\alpha_a + \mu_i\}\). Similarly,
\[
\frac{1}{n} \log \|g\|_q \geq N_Y(q) - b,
\]
where \(N_Y(q) := \max_{1 \leq i \leq s} \{-\beta_a + \nu_i\}\). Utilizing these equations, we obtain
\[
\frac{1}{n} \log \langle f_a, g_a \rangle \leq 3 \cdot 2^{-an}.
\]

\[
\frac{1}{n} \log \left( \sum_{i=1}^{r} \sum_{j=1}^{s} P_{XY}^k(A_i \times B_j) u_i v_j 2^{2nb + 3 \cdot 2^{-an}} \right)
\]
\[
\frac{1}{n} \log \left( \sum_{i=1}^{r} \sum_{j=1}^{s} P_{XY}^k(A_i \times B_j) u_i v_j 2^{2nb + 3 \cdot 2^{-an}} \right)
\]
Combining the inequalities above, we have

\[
\frac{1}{n} \log \|f\|_p \leq \frac{1}{n} \log \|f_a\|_p + \frac{1}{np} \log [1 + 2^{-npa}]
\]

\[
\leq \frac{1}{np} \log \left( \sum_{i=1}^{r} \left( 1 + 2^{-npa} \right) \right)
\]

\[
\leq N_X(p) + b + \epsilon_n,
\]

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

Similarly, we have

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

where \(\bar{\epsilon}_n := \frac{1}{nq} \log s + \frac{1}{nq} \log [1 + 2^{-npa}]\), which tends to zero as \(n \to \infty\) for large enough \(a\) and any fixed \(b\).

On the other hand,

\[
\frac{1}{n} \log \langle f, g \rangle \geq \frac{1}{n} \log \langle f_a, g_a \rangle + \frac{1}{np} \log \|f\|_p \|g\|_q
\]

\[
\geq \sum_{i=1}^{r} \sum_{j=1}^{s} \left[ \left( -\Theta^{\ast}(\alpha_i, \beta_j) + \mu_i + \nu_j \right) \cdot 2^{-nb} \right]
\]

\[
\geq \max_{1 \leq i \leq r, 1 \leq j \leq s} \left\{ -\Theta^{\ast}(\alpha_i, \beta_j) + \mu_i + \nu_j \right\} - 2b
\]

(95)

\[
\geq -\Theta^{\ast}(\alpha_i^{\ast}, \beta_j^{\ast}) + \mu_i^{\ast} + \nu_j^{\ast} - 2b
\]

(96)

\[
= -\Theta^{\ast}(\alpha_i^{\ast}, \beta_j^{\ast}) + \frac{\alpha_i^{\ast}}{p} + \frac{\beta_j^{\ast}}{q} + N_X(p) + N_Y(q) - 2b,
\]

where (95) follows from Theorem 4, with the maximum being taken only over those pairs \((i, j)\) for which \(P^{\ast}_X(A_i) > 0\) and \(P^{\ast}_Y(B_j) > 0\), since \(\Theta^{\ast}(\alpha_i, \beta_j)\) is defined only for \(\alpha_i \in [0, E_{1, max}], \beta_j \in [0, E_{2, max}]\); also, in (96), \(i^{\ast}\) is defined as the optimal \(i\) attaining \(N_X(p)\) and \(j^{\ast}\) as the optimal \(j\) attaining \(N_Y(q)\). Combining the inequalities above, we have

\[
\frac{1}{n} \log \|f\|_p \geq \max_{p,q} \langle \alpha, \beta \rangle - 4\epsilon_n - \hat{\epsilon}_n.
\]

(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. Csiszár 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. Csiszári 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. Pradhan, “On large deviation analysis of sampling from typical sets,” in Proc. Workshop Inf. Theory Appl. (ITA), 2007. [Online]. Available: https://public.webiste.unc.edu/~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 Annua. 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. 290–336, Aug. 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 their extensions,” Found. Trends Commun. Inf. Theory, vol. 19, no. 2, pp. 107–389, 2022.

[19] A. Bonami, “Inequalités en probabilité de certains sous-espaces d’un espace L2 application aux chaos de Wiener,” Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete, vol. 14, no. 1, pp. 36–48, 1969.

[20] A. Bonami, “Etude des coefficients de Fourier des fonctions de Lp(G),” Annales de L’Institut Fourier, vol. 20, no. 2, pp. 335–402, 1970.

[21] W. Beckner, “Inequalities in Fourier analysis,” Ann. Math., vol. 102, no. 1, pp. 159–182, 1975.

[22] G. Gross, “Logarithmic sobolev inequalities,” Amer. J. Math., vol. 97, no. 4, pp. 1061–1073, 1975.

[23] 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.

[24] C. Borell, “positivity improving operators and hypercontractivity,” Mathematische Zeitschrift, vol. 180, no. 3, pp. 225–234, Sep. 1982.

[25] D. Bakry, “L’hypercontractivité et son utilisation en thèorie des semi-groupes,” in Lectures on Probability Theory, Berlin, Germany: Springer, 1994, pp. 1–114.

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

[27] 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.

[28] 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.

[29] 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.

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

[31] L. Yu, “Strong Brascamp–Lieb inequalities,” 2021, arXiv:2102.06935.

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

[33] 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.
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.