Common Randomness and Key Generation with Limited Interaction

Jingbo Liu Paul Cuff Sergio Verdú
Dept. of Electrical Eng., Princeton University, NJ 08544
jingbo@princeton.edu

Abstract

The basic two-terminal common randomness (CR) or key generation model is considered, where the communication between the terminals may be limited, and in particular may not be enough to achieve the maximal CR/key rate. We introduce a general framework of XY-absolutely continuous distributions and XY-concave function, and characterize the first order CR/key-communication tradeoff in terms of the evaluation of the XY-concave envelope of a functional defined on a set of distributions, which is simpler than the multi-letter characterization. Two extreme cases are given special attention. First, in the regime of very small communication rates, the CR per bit of interaction (CRBI) and key per bit of interaction (KBI) are expressed with a new "symmetrical strong data processing constant", defined as the minimum of a parameter such that a certain information-theoretic functional touches its XY-concave envelope at a given source distribution. We also provide a computationally friendly strong converse bound for CRBI and a similar (but not necessarily strong) one for KBI in terms of the supremum of the maximal correlation coefficient over a set of distributions. The proof uses hypercontractivity and properties of the Rényi divergence. A criterion the tightness of the bound is given with applications to the binary symmetric sources. Second, a new characterization of the minimum interaction rate needed for achieving the maximal key rate (MIMK) is given, and we resolve a conjecture by Tyagi and Narayan [45] regarding the MIMK for binary sources. We also propose a new conjecture for the binary symmetric sources.

I. Introduction

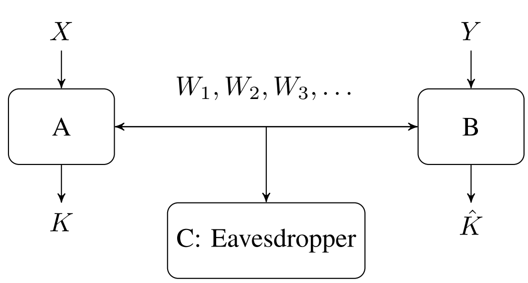

Figure 1: Terminals A and B observe sources with joint distribution Q_{XY} and interactive communication between A and B is allowed.

Generally speaking, common randomness (CR) generation [2][18] concerns the task of producing a common piece of information by several terminals accessing correlated sources, possibly allowing communications among the terminals. The related key generation problem [35][1][17] imposes an additional constrain that an eavesdropper, knowing the law of the system and the communications but not these correlated sources, can learn almost nothing about the common information generated. The importance of CR/key generation in cryptography and other areas of information theory is well appreciated [41][35][1][2]. From the purely theoretical viewpoint, they are also fascinating sources of problems because of the connections to various measures of correlation. Consider the case of two correlated i.i.d. sources with per-letter distribution Q_{XY}. The maximal rate of key that can be produced without any communication constraint equals the mutual information $I(X;Y)$ [1]. In the other extreme where the communication rate vanishes, the *key per bit of communication* under the one-way protocol in [1] is a monotonic function of the strong data processing constant [14][31]; and under the one-communicator protocol [32], a reflection of the dual convex set of the set of hypercontractive coefficients [32]. The Gács-Körner common information [20] is the maximal CR rate obtainable without any communication. The Wyner common information [48] characterizes an extreme point in the intersection between a hyperplane and the rate region in the one-communicator CR generation [2, Theorem 4.2] or key generation.

Despite the successes in those models mentioned above, which mainly consists of one directional communication among terminals, many basic problems have remained open in settings involving interactive communications or multi-terminals [19][18][17]. Most of the existing literature focused on achieving the maximal possible key rate. Csiszár and Narayan [17] showed that the maximal key rate obtainable from multi-terminals having public interactive communications equals the entropy rate of all sources minus the communication rate needed for *communication for omniscience* [17], the latter related to the subject of interactive source coding studied by Kaspi [29]. Moreover, Tyagi [45] showed that the minimum interactive communication rate needed for achieving the maximal key rate (MIMK) between two terminals with interactions equals the *interactive common information* [45] minus the mutual information rate of the sources, and provided a multi-letter characterization of MIMK. However, a complete characterization of the tradeoff between the key rate and communication rate poses more challenge, because when the communication rate is not large enough for the terminals to become omniscient, it not obvious what piece of information they have to agree on. Indeed, as mentioned at the end of Section VII in [45], a characterization of the key rate when the communication rate is less than MIMK, along with a single-letter characterization of MIMK, "remains an interesting open problem".

In this paper we consider the two-terminal interactive CR/key generation model shown in Figure 1, which is similar to the setting of [45] mentioned above, but look at the tradeoff between the key rate and communication rate rather than MIMK alone, and adopt completely different approaches. We first revisit Kaspi's original idea of multi-letter characterizations of the rate region of interactive source coding [29] where each round of communication accounts for a new auxiliary random variable and adds in a new term to the rate expressions which resembles the expressions in the one-way counterparts. In our interactive CR/key generation problem, we derive a similar multi-letter characterization as the first step. In terms of the first order region, CR generation and key generation are essentially equivalent problems.

We then simplify the multi-letter characterization of the key-communication tradeoff region using XY*-concave envelopes*, partially inspired by a similar characterization in the context of interactive source coding by Ma et al. [34], who noticed that each auxiliary random variable in the multi-letter region, which corresponds to each round of the public communications, amounts to convexifying the rate region with respect a marginal distribution. Hence in the infinite-round limit, the minimum sum rate can be described in terms of a marginally convex envelope, i.e.,the the greatest functional which is convex w.r.t. each marginal distribution and dominated by a given functional. Expressing the role of an auxiliary random variable as taking the convex envelope is very common in information theory [19][4][36][23]. At the first sight, this idea is easily overlooked as a mere restatement of the multi-letter region. However, as demonstrated by Ma et al.'s work as well as the present paper, the conceptual simplification opens the possibility of tackling specific open questions and making new connections. Moreover, we introduce a notion of X-absolute continuity, so that the marginal concave envelope approach is applicable to general non-discrete sources. In fact this framework may be applied to other problems involving convex/concave envelopes to avoid the technical difficulty of defining a conditional distribution from a given joint distribution.

In the regime of very small communication rates, the *CR per bit of interaction* (CRBI) or *key per bit of interaction* (KBI) is the fundamental limit on the maximal amount of CR or key bits that can be "unlocked" by each communication bit, which is the most befitting for the scenario of a stringent communication constraint and relatively abundant correlated resources. KBI is not completely implied by the rate region since the length of the communication bits can be a vanishing fraction of the blocklength. The concave envelope characterization implies that KBI is a monotonic function of a "symmetrical strong data processing constant" (SSDPI), defined as the minimum of a parameter such that a certain information-theoretic functional touches its XY-concave envelope at a given source distribution. It is interesting to compare SSDPI with the conventional strong data processing constant [3], which has a similar (but only one-sided) convex envelope characterization [4] and is in a similar way related to the key per bit of communication in the one-way protocol [14][31]. A more computationally friendly strong converse bound on CRBI and a similar (but not necessarily strong) one for KBI are also given, which is a monotonic function of the supremum of the maximal correlation coefficient over a set of distributions. The proof of the upper bound uses hypercontractivity and properties of the Rényi divergence. A necessary condition on the tightness is given, implying that for the binary symmetric source and the Gaussian source the KBI is not improved by increasing the number of rounds of interactions.

Returning to the MIMK problem considered by Tyagi [45], a different characterization of the minimum interaction rate needed for achieving the key capacity is given by establishing several fundamental properties of XY-concave functions. In [45] Tyagi conjectured that MIMK equals the minimum one round communication rate for achieving the maximal key rate for binary sources. We use the new characterization to prove that the necessary and sufficient condition for the conjecture to hold is that the joint distribution can be given by a binary symmetric channel with an arbitrary input distribution. We also propose a new conjecture about the complete key-communication tradeoff for binary symmetric sources.

II. Premilaries

A. Problem Setup

In this paper we consider an interactive key generation model in Figure 1. Let Q_{XY} be the joint distribution of the sources. the Terminals A and B observe X and Y, respectively. Terminal A computes an integer $W_1 = W_1(X)$ (possibly stochastically) and sends it to B. Then terminal B computes an integer $W_2 = W_2(W_1, Y)$ and send it to B, and so on, for a total of r rounds/times. Then, A and B calculate integers[1] $K = K(X, W^r)$ and $\hat{K} = \hat{K}(Y, W^r)$ possibly stochastically as keys. Both terminals want $K = \hat{K}$ with high probability, while keeping it (almost) independent of the public messages W^r which is observed by an eavesdropper.

In the case of stationary memoryless sources and block coding, we will substitute $X \leftarrow X^n$ and $Y \leftarrow Y^n$, where n is the blocklength. The performance is measured by

$$\delta_n := \frac{1}{2}|Q_{K\hat{K}} - T_{K\hat{K}}| \tag{1}$$

in CR generation, or

$$\Delta_n := \frac{1}{2}|Q_{K\hat{K}W^r} - T_{K\hat{K}}Q_{W^r}| \tag{2}$$

in the case of key generation, where $T_{K\hat{K}}$ denotes the target distribution under which $K = \hat{K}$ is equiprobable, and the total variation $|\cdot|$ is defined as the ℓ_1 distance. Such performance measure are natural when the *likelihood encoder* is used in the achievability proof, (c.f. [31]).

Definition 1. The tuple (R, R_1, R_2) is said to be *r-achievable* ($r \in \{1, 2, \dots, \infty\}$) if a sequence of generation schemes in r rounds [2] can be designed to fulfill the following conditions:

$$\liminf_{n\to\infty} \frac{1}{n} \log |\mathcal{K}| \ge R; \tag{3}$$

$$\limsup_{n\to\infty} \frac{1}{n} \log |\mathcal{W}_l| \le R_i, \quad i = 1, 2, \tag{4}$$

and $\lim_{n\to\infty} \delta_n = 0$ in CR generation or $\lim_{n\to\infty} \Delta_n = 0$ in key generation.

Remark 1. Some authors have considered other alternatives, say

$$\epsilon_n := \mathbb{P}[K \ne \hat{K}], \tag{5}$$

$$\nu_n := D(Q_{K|W^r} \| T_K | Q_{W^r}), \tag{6}$$

for the key generation problem. The relation to (2) is as follows. Clearly, $\Delta_n \to 0$ implies that $\epsilon_n \to 0$. Also, notice that for arbitrary P and Q on the same alphabet $\mathcal{X}$, [16, Lemma 2.7] gives

$$|H(P) - H(Q)| \le |P - Q| \log \frac{|\mathcal{X}|}{|P - Q|} + \log |\mathcal{X}| 1 \left\{ |P - Q| > \frac{1}{2} \right\} \tag{7}$$

[1]Notation $W_i^j := (W_i, W_{i+1}, \dots, W_j)$ denotes a vector and $W^r := W_1^r$.

[2]As a convention, we shall say "in r rounds" or "r-round" if the number of rounds of communication less than $r+1$ (or equivalently, not exceeding r for an integer r or finite for $r = \infty$). Therefore the term is not precise if $r = \infty$.

Thus by Jensen's inequality and Markov inequality, we have

$$\nu_n = H(T_K) - H(Q_{K|W^r}|Q_{W^r}) \tag{8}$$

$$\leq 2\Delta_n \log \frac{|\mathcal{X}|}{2\Delta_n} + 4\Delta_n \log |\mathcal{X}|. \tag{9}$$

Therefore exponentially vanishing Δ_n (which is usually guaranteed by the likelihood encoder based proof, c.f. [31]) ensures $\nu_n \to 0$. On the other hand, by Pinsker-Csiszár inequality, $\nu_n \to 0$ implies $|Q_{KW^r} - T_K Q_{W^r}| \to 0$, which, combined with $\epsilon_n \to 0$, implies that $\Delta_n \to 0$.

In terms of the first order rate for the stationary memoryless sources, CR generation and key generation are essentially equivalent problems and the achievable rate region for one is a linear transform of the other; see e.g. [45] for the discussion. Essentially, the maximal CR rate is the maximum of the key rate plus the communication rates (whether local randomization is allowed or not). For this reason we shall only discuss results for key generation when the first order rates are concerned.

From the standard diagonalization argument [25], the achievable region is closed. The set of r-achievable tuples for key generation is denoted by $\mathcal{R}_r(X,Y)$. Clearly $\mathcal{R}_r(X,Y)$ is "increasing" in r. We can also show that it is "continuous" at $r=\infty$, that is $\mathcal{R}_\infty(X,Y)$ equals the closure of $\bigcup_{r=1}^{\infty}\mathcal{R}_r(X,Y)$. The "$\supseteq$" part is immediate from the definitions. The "$\subseteq$" part, in essence, relies on the converse proof for $\mathcal{R}_\infty(X,Y)$.

Inspired by Kaspi's multi-letter characterization of the rate region for interactive source coding [29], we proved that $\mathcal{R}_r(X,Y)$ is the closure of the set of (R, R_1, R_2) satisfying

$$R \leq \sum_{1\leq i\leq r}^{\text{odd}} I(U_i;Y|U^{i-1}) + \sum_{1\leq i\leq r}^{\text{even}} I(U_i;X|U^{i-1}), \tag{10}$$

$$R_1 \geq \sum_{1\leq i\leq r}^{\text{odd}} I(U_i;X|U^{i-1}) - \sum_{1\leq i\leq r}^{\text{odd}} I(U_i;Y|U^{i-1}), \tag{11}$$

$$R_2 \geq \sum_{1\leq i\leq r}^{\text{even}} I(U_i;Y|U^{i-1}) - \sum_{1\leq i\leq r}^{\text{even}} I(U_i;X|U^{i-1}), \tag{12}$$

where the auxiliary r.v.'s satisfy

$$U_i - (X, U^{i-1}) - Y \quad i \in \{1,\dots,r\}\setminus 2\mathbb{Z}; \tag{13}$$

$$X - (Y, U^{i-1}) - U_i \quad i \in \{1,\dots,r\}\cap 2\mathbb{Z}. \tag{14}$$

The notation $\sum_{1\leq i\leq r}^{\text{odd}}$ is used as an abbreviation of $\sum_{i\in\{1,\dots,r\}\setminus 2\mathbb{Z}}$, and similarly for $\sum_{1\leq i\leq r}^{\text{even}}$.

This bound is quite intuitive: depending on whether i is odd or even, U_i corresponds to the messages sent by either the terminal A or B. The first round of communication contributes to the term $(I(U_1;Y), I(U_1;X)-I(U_1;Y), 0)$ in the rate tuple expressions, which is exactly the rates in one-round key generation [1]. The second round adds in similar terms but all mutual information are now conditioned on U_1, which is now shared publicly, and so on. A formal proof of this multi-letter characterization will be given in a separate note [33] using the likelihood encoder [43] and standard converse proof techniques.

B. XY-Absolutely Continuity

Recall that a nonnegative finite measure ν is absolutely continuous with respect to another one μ on the same measurable space $(\mathcal{X}, \mathscr{F})$, denoted as $\nu \ll \mu$, if there exists a measurable function f such that

$$\nu(\mathcal{A}) = \int_{\mathcal{A}} f(x) \mathrm{d}\mu \tag{15}$$

for all $\mathcal{A} \in \mathscr{F}$. We extend the idea to the following:

Definition 2. A nonnegative finite measure ν_{XY} is said to be *X-absolutely continuous with respect to μ_{XY}*, denoted by

$$\nu_{XY} \preceq_X \mu_{XY} \tag{16}$$

if there exists a measurable[3] function f such that

$$\nu_{XY}(\mathcal{A}) = \int_{\mathcal{A}} f(x) \mathrm{d}\mu_{XY}(x, y) \tag{17}$$

for any $\mathcal{A} \in \mathscr{F}$. Moreover, ν is said to be *XY-absolutely continuous with respect to μ*, denoted simply as $\nu \preceq \mu$, if there exists a measurable function f and g such that

$$\nu_{XY}(\mathcal{A}) = \int_{\mathcal{A}} f(x) g(y) \mathrm{d}\mu_{XY}(x, y). \tag{18}$$

Note that (146) implies $\nu \ll \mu$, so an equivalent definition of (16) is that the Radon-Nikodym derivative $\frac{\mathrm{d}\nu}{\mathrm{d}\mu}(x, y)$ depends only on x. Similarly, an equivalent definition of XY-marginal absolute continuity is that $\frac{\mathrm{d}\nu}{\mathrm{d}\mu}$ is a product of a function depending on x and a function depending on y.

It is straightforward to see that $\nu \preceq \mu$ if there exists $(\theta^i_{XY})_{i=1}^t$ for some odd integer t such that

$$\nu \preceq_Y \theta^t; \tag{19}$$

$$\theta^i \preceq_X \theta^{i-1}, \quad i \in \{1, \dots, t\} \setminus 2\mathbb{Z}; \tag{20}$$

$$\theta^i \preceq_Y \theta^{i-1}, \quad i \in \{1, \dots, t\} \cap 2\mathbb{Z}; \tag{21}$$

$$\theta^1 \preceq_X \mu. \tag{22}$$

In fact, the converse is also true, and one can choose $t \leq 3$. In the case of finite alphabets, one can even improve the bound to $t = 1$. The latter cannot always be achieved for general alphabet because $\int f(x) \mathrm{d}\mu_X(x)$ can be infinite even if $\int f(x) g(y) \mathrm{d}\mu(x, y)$ is finite.

The relation $\preceq_X$ is a *preorder* relation [4] on the set of nonnegative finite measures. We shall denote by

$$\mathcal{M}_X(\mu) := \{\nu \colon \nu \preceq_X \mu\} \tag{23}$$

[3]More precisely, we have assumed σ-algebras $\mathscr{F}$ and $\mathscr{G}$ on $\mathcal{X}$ and $\mathcal{Y}$, respectively, and the σ-algebra on $\mathcal{X} \times \mathcal{Y}$ is generated by $\mathcal{A} \times \mathcal{B}$ where $\mathcal{A} \in \mathscr{F}$ and $\mathcal{B} \in \mathscr{G}$, so here the measurability of f is w.r.t. $\mathscr{F}$.

[4]A preorder relation satisfies reflexivity and transitivity, but not antisymmetry. The more familiar notion of partially ordered set is a preordered set satisfying antisymmetry.

the *lower set* of μ in the set of nonnegative finite measures. Similarly, $\mathcal{M}(\mu)$ is defined as the lower set of μ with respect to $\preceq$. Both relations also make the set of probability distributions a preordered set. Denote by $\mathcal{P}_X(Q_{XY})$ or $\mathcal{P}(Q_{XY})$ the corresponding lower sets.

Remark 2. The lower set $\mathcal{P}(Q_{XY})$ appears frequently information theory (with different notations and names). Csiszár [15] showed that the I-projection of Q_{XY} onto the linear set of distributions having given marginal distributions, if exists, must belong to $\mathcal{P}(Q_{XY})$. Due to this fact, $\mathcal{P}(Q_{XY})$ has emerged, e.g. in the context of hypercontractivity [27] and multiterminal hypothesis testing [37]. In interactive source coding [34] the set $\mathcal{P}(Q_{XY})$ has been defined for discrete distributions, without introducing the preorder relation. In both [34] and the present paper, the appearance of $\mathcal{P}(Q_{XY})$ is due to the conditioning on auxiliary random variables satisfying Markov structures, c.f. (10)-(12).

Next, we introduce notions of concave functions and concave envelopes with respect to the marginal distributions, the discrete case being defined in [34]. We refine those definitions using the XY-absolute continuity framework to resolve the technicality of defining a conditional distribution from a joint distribution.

Definition 3. A functional σ on a set $\mathcal{P}$ of distributions is said to be X*-concave* if for any $P_{XY} \in \mathcal{P}$, $(P^i_{XY})_{i=0,1}$ and $\alpha \in [0,1]$ satisfying [5]

$$P^i_{XY} \preceq_X P_{XY}, \quad i = 0, 1; \tag{24}$$

$$P_{XY} = \bar{\alpha} P^0_{XY} + \alpha P^1_{XY}, \tag{25}$$

it holds that

$$\sigma(P_{XY}) \geq \bar{\alpha}\sigma(P^0_{XY}) + \alpha\sigma(P^1_{XY}). \tag{26}$$

Moreover, σ is said to be XY*-concave* if it is both X-concave and Y-concave.

Definition 4. Given a functional σ on a set $\mathcal{P}$ of distributions, the functional σ' is said to be the X-concave envelope of σ, denoted as $\mathrm{env_X}(\sigma)$, if σ' is X-concave and is dominated by any other X-concave functional which dominates σ. The XY*-concave envelope*, denoted by $\mathrm{env_{XY}}(\sigma)$, is defined similarly.

The existence of X-concave envelope is immediate from the existence of the conventional concave envelope for a function. For the existence of XY-concave envelope, we can take the X-concave envelope and Y-concave envelope of the given functional alternatingly. The pointwise limit exists by the monotone convergence theorem and satisfies the condition in Definition 4.

III. Convex Geometric Characterizations of the Rate Regions

In this section we study the tradeoff between the key rate and the interactive communication rate, that is, the set of achievable pair $(R, R_1 + R_2)$ in key generation. The set of achievable tuple (R, R_1, R_2) clearly can be

[5] In this paper $\bar{\alpha} := 1 - \alpha$ for $\alpha \in [0, 1]$.

handled with the same approach. Moreover, corresponding results for CR generation will not be discussed since, as mentioned before, the achievable rate region for CR generation is just a linear transformation of the one for key generation.

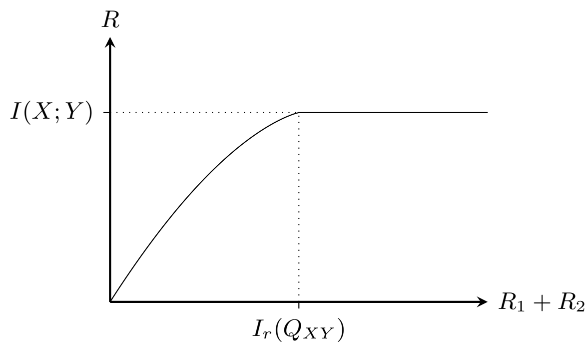

Figure 2: Achievable region (below the curve) of the key rate R and the sum interactive communication rate R_1+R_2. The minimum interaction needed for maximal key rate is denoted as $I_r(Q_{XY})$ (see Section V).

Define the *total sum rate*

$$S := R + R_1 + R_2, \tag{27}$$

and we consider the problem of characterizing $\mathcal{S}_r(X,Y)$, which is defined as the set of achievable (S,R). For any Q_{XYU^r} where U^r satisfies the given Markov chains, denote by $R(Q_{XYU^r})$ the right side of (10) and $S(Q_{XYU^r})$ is defined similarly for the total sum rate. Observe that

$$R(Q_{XYU^r}) = \sum_{1\le i\le r}^{\text{odd}} I(U_i;Y|U^{i-1}) + \sum_{1\le i\le r}^{\text{even}} I(U_i;X|U^{i-1}) \tag{28}$$

$$= I(U_1;Y) + \sum_{2\le i\le r}^{\text{even}} I(U_i;X|U^{i-1}) + \sum_{1\le i\le r}^{\text{odd}} I(U_i;Y|U^{i-1}) \tag{29}$$

$$= I(X;Y) - I(X;Y|U_1) + \sum_{1\le i\le r}^{\text{odd}} I(U_i;X|U^{i-1}) + \sum_{2\le i\le r}^{\text{even}} I(U_i;Y|U^{i-1}), \tag{30}$$

where the last step used the Markov condition $U_1 - X - Y$. Hence by rearranging,

$$I(X;Y) - R(Q_{XYU^r}) = I(Y;X|U_1) - \int R(Q_{YXU_2^r|U_1=u})\mathrm{d}Q_{U_1}(u) \tag{31}$$

Now the key observation is that the right side above is similar to the left except that each term is conditioned on U_1, the role of X and Y are switched, and (conditioned on U_1) there are $r-1$ auxiliary random variables left. Similarly, we also have

$$H(X,Y) - S(Q_{XYU^r}) = H(Y,X|U_1) - \int S(Q_{YXU_2^r|U_1=u})\mathrm{d}Q_{U_1}(u), \tag{32}$$

and similar observations can be made. In the case of non-discrete (X,Y), we can choose a reference measure and replace the entropy/conditional entropy terms above with relative entropy/conditional relative entropy, at the cost of slightly more cumbersome notation, so there is no loss of generality with this approach.

Given Q_{XY}, and $s > 0$, define a functional on $\mathcal{P}(Q_{XY})$ by

$$\omega_0^s(P_{XY}) := sH(\hat{X}, \hat{Y}) - I(\hat{X}; \hat{Y}). \tag{33}$$

where $P_{XY} \preceq Q_{XY}$ and $(\hat{X}, \hat{Y}) \sim P_{XY}$. For $r \in \{1, 2, \dots\}$, define

$$\omega_r^s := \begin{cases} \mathrm{env}_X(\omega_{r-1}^s) & r \text{ is odd;} \\ \mathrm{env}_Y(\omega_{r-1}^s) & r \text{ is even,} \end{cases} \tag{34}$$

and define ω_∞^s as the XY-concave envelope of ω_0^s.

From (31), (32) and the fact that $S(Q_{XY}) = R(Q_{XY}) = 0$ when $r = 0$, we immediately obtain

Theorem 1.

$$\omega_r^s(Q_{XY}) := \sup_{U^r} \{s[H(X,Y) - S(Q_{XYU^r})] - [I(X;Y) - R(Q_{XYU^r})]\} \tag{35}$$

$$= sH(X,Y) - I(X;Y) + \sup_{U^r} \{R(Q_{XYU^r}) - sS(Q_{XYU^r})\} \tag{36}$$

$$= sH(X,Y) - I(X;Y) + \sup_{(S,R) \in \mathcal{S}_r(X,Y)} \{R - sS\} \tag{37}$$

where U^r satisfies (13)-(14).

Because $H(X,Y)$ and $I(X;Y)$ are independent of U^r, characterizing the closed convex set $\mathcal{S}_r(X,Y)$ is equivalent to computing $\omega_r^s(Q_{XY})$ for each $s > 0$.

The significance of Theorem 1 is that we can sometimes come up with a XY-concave function that upper-bounds ω_0^s. If, out of luck, the upper-bounding function evaluated at Q_{XY} can also be achieved by a known scheme, we will be able to determine $\omega_\infty^s(Q_{XY})$.

IV. CR/Key per Bit of Interaction

Similar to the capacity per unit energy/cost [42][46] in the context of channel coding, in this section we consider the following fundamental limit in interaction CR/key generation.

Definition 5. For $r \in \{1, 2, \dots, \infty\}$, $\delta \in [0, 1]$, define the *δ-CR per bit of r-round interaction* $\gamma_r^\delta(X;Y)$ as the maximum real number $\gamma \geq 0$ such that there exists a sequence (indexed by k) of r-round (possibly stochastic) CR generation schemes which fulfill the following conditions:

$$\liminf_{k \to \infty} \frac{\log |\mathcal{K}|}{\log |\mathcal{W}^r|} \geq \gamma; \tag{38}$$

$$\lim_{k \to \infty} \log |\mathcal{K}| = \infty; \tag{39}$$

$$\limsup_{k \to \infty} \delta_k \leq \delta. \tag{40}$$

where δ_k is as in (2). The *CR per bit of r-round interaction* is defined as

$$\gamma_r(X;Y) := \inf_{\delta > 0} \gamma_r^\delta(X;Y). \tag{41}$$

We shall use *CR per bit of interaction* (CRBI) as an abbreviation of $\gamma_\infty(X;Y)$. The *key per bit of interaction* (KBI), denoted as $\Gamma_\infty(X;Y)$ is defined similarly with δ_k above replaced by Δ_k.

Note that there is no constraint on the blocklength in Definition 5. In particular, the blocklength can grow super-linearly in $\log|\mathcal{W}^r|$, in which case the rates are zero and the fraction in (38) cannot be written as $\frac{R}{R_1+R_2}$. Nevertheless, for stationary memoryless sources, one can still show the relation

$$\Gamma_r(X;Y) := \sup\left\{\frac{R}{R_1+R_2}\colon (R,R_1,R_2)\in\mathcal{R}_r(X,Y)\right\} \tag{42}$$

by a careful reexamination of the proof of the achievable rate region (c.f. a similar result in the context of channel coding with costs [46]).

We shall provide compact formulas for γ_∞ and Γ_∞ by introducing a symmetrical data processing constant, and also derive computational friendly upper-bounds on γ_∞^δ and Γ_∞ ($\delta\in(0,1)$).

A. Symmetrical SDPI and the Exact Formulas

To begin with, recall that the *key per bit of communication* (c.f. [14][31][6]) is the $r=1$ special case of KBI, and according to (10)-(12), has the formula

$$\Gamma_1(X;Y) = \sup_{U\colon U-X-Y}\frac{I(U;Y)}{I(U;X)-I(U;Y)} \tag{43}$$

$$= \frac{s^*(X;Y)}{1-s^*(X;Y)}. \tag{44}$$

where the strong data processing constant (c.f. [3][4][38]) is commonly defined as

$$s_1^*(X;Y) := \sup_{U\colon U-X-Y}\frac{I(U;Y)}{I(U;X)} \tag{45}$$

$$= \sup\left\{\frac{R}{S}\colon (S,R)\in\mathcal{S}_1(X,Y)\right\} \tag{46}$$

and we always assume that the supremums are over auxiliary random variables such that the fraction is well-defined. Conventionally, s_1^* is denoted as s^* [3][4]. From (45), it is not hard to see that $s^*(X;Y)$ has the following equivalent characterization. Recall (33) defined a functional on $\mathcal{P}(Q_{XY})$ by

$$\omega_0^s(P_{XY}) = sH(\hat{X}) - H(\hat{Y}) + (s+1)H(\hat{Y}|\hat{X}) \tag{47}$$

where $P_{XY}\preceq Q_{XY}$ and $(\hat{X},\hat{Y})\sim P_{XY}$.

Proposition 1. *$s^*(X;Y)$ is the infimum of $s>0$ such that*

$$\omega^s(Q_{XY}) := \operatorname*{env}_X \omega_0^s(Q_{XY}) = \omega_0^s(Q_{XY}). \tag{48}$$

If X or Y is non-discrete, we may choose an arbitrary reference measure and replace the entropies with (the negative of) the relative entropies, so there is no loss of generality with the concave envelope characterization

[6] Incidentally, Ahlswede made pioneering contributions to both the strong data processing constant [3] and key generation [1], although it appears that he never explicitly reported a connection between the two.

approach. In the discrete case, the concave envelope characterization in Proposition 1 is essentially shown by Anantharam et al. [4], noting that the third term in (47) is linear in P_X for fixed $Q_{Y|X}$. However, by using the framework in Section II-B, our Proposition 1 avoids the challenge of defining a conditional distribution $Q_{Y|X}$ from the possibly non-discrete joint distribution Q_{XY}.

Example 1. *For the binary symmetric sources (BSS) with error probability ϵ, $s^*(X;Y) = (1-2\epsilon)^2$. The scalar Gaussian sources with correlation coefficient ρ has $s^*(X;Y) = \rho^2$. For an erasure channel with erasure probability ϵ and equiprobable input distribution, we have $s^*(X;Y) = 1-\epsilon$ and numerical simulation suggests that $s^*(Y;X) = \frac{1}{\log \frac{2}{1-\epsilon}}$ for small enough $1-\epsilon$. Additional examples including the Z-channel or the binary symmetric channel (BSC) with non-equiprobable inputs can be found in [5].*

Returning to the key per bit of interaction, we can define a similar notion of data processing constant from a multi-letter expression, or equivalently according to the analysis in Section III, with the following concave envelope characterization:

Definition 6. Define the *symmetrical data processing constant* (SSDPI) $s^*_\infty(X;Y)$ as the infimum of $s > 0$ such that

$$\omega^s_\infty(Q_{XY}) := \mathrm{env}_{\mathrm{XY}}\, \omega^s_0(Q_{XY}) = \omega^s_0(Q_{XY}). \tag{49}$$

By the conventional data processing inequality, it is immediate to show that $s^*_\infty(X;Y) \in [0,1]$. Moreover from (49), we clearly have the symmetric property $s^*_\infty(X;Y) = s^*_\infty(Y;X)$, in contrast to $s^*(\cdot)$ [4]. It will certainly be illuminating to list and compare the properties of $s^*_\infty(X;Y)$ and $s(X;Y)$ in the future work.

The symmetrical SDPI is related to the operational quantities by the following, whose proof follows from Theorem 1 and (42).

Theorem 2.

$$s^*_\infty(X;Y) = \sup\left\{\frac{R}{S} \colon (S,R) \in \mathcal{S}_\infty(X,Y)\right\} \tag{50}$$

$$\gamma_\infty(X;Y) = \frac{1}{1 - s^*_\infty(X;Y)}. \tag{51}$$

$$\Gamma_\infty(X;Y) = \frac{s^*_\infty(X;Y)}{1 - s^*_\infty(X;Y)}. \tag{52}$$

B. A Useful Upper-bound

For U^r satisfying (13)-(14), the following upper-bound follows from the definition (45) of SDPI:

$$\frac{I(U_i;Y|U^{i-1})}{I(U_i;X|U^{i-1})} \le \sup_{u^{i-1}} s^*(X;Y|U^{i-1} = u^{i-1}) \tag{53}$$

$$\le \sup_{P_{XY} \preceq Q_{XY}} s^*(\hat{X};\hat{Y}) \tag{54}$$

where the last inequality follows since it is trivial to check by induction that $Q_{XY|U^{i-1}=u^{i-1}}$ for any u^{i-1}.

Definition 7. For $(X,Y) \sim Q_{XY}$, the *maximal correlation coefficient* [26][22][39] is defined as

$$\rho_{\rm m}^2(X;Y) := \sup_{f,g} \mathbb{E}[f(X)g(Y)] \tag{55}$$

where the supremum is over measurable real valued functions f and g satisfying $\mathbb{E}[f(X)] = \mathbb{E}[g(Y)] = 0$ and $\mathbb{E}[f^2(X)] = \mathbb{E}[g^2(Y)] = 1$.

Ahlswede and Gács [3, Theorem 8] (see also [12]) proved a useful relation between SDPI and the maximal correlation coefficient, which, in the language of Section II-B, is that

$$\sup_{P_{XY} \preceq_X Q_{XY}} s^*(\hat{X};\hat{Y}) = \sup_{P_{XY} \preceq_X Q_{XY}} \rho_{\rm m}^2(\hat{X};\hat{Y}). \tag{56}$$

From Definition 2 and (56), we have

$$\sup_{P\colon P \preceq Q} s^*(\hat{X};\hat{Y}) = \sup_{T\colon T \preceq_Y Q} \sup_{P\colon P \preceq_X T} s^*(\hat{X};\hat{Y}) \tag{57}$$

$$= \sup_{T\colon T \preceq_Y Q} \sup_{P\colon P \preceq_X T} \rho_{\rm m}^2(\hat{X};\hat{Y}) \tag{58}$$

$$= \sup_{P\colon P \preceq Q} \rho_{\rm m}^2(\hat{X};\hat{Y}) \tag{59}$$

Theorem 3. *Given Q_{XY},*

$$\gamma_\infty(X;Y) \le \sup_{P_{XY} \preceq Q_{XY}} \frac{1}{1-\rho_{\rm m}^2(\hat{X};\hat{Y})} \tag{60}$$

$$\Gamma_\infty(X;Y) \le \sup_{P_{XY} \preceq Q_{XY}} \frac{\rho_{\rm m}^2(\hat{X};\hat{Y})}{1-\rho_{\rm m}^2(\hat{X};\hat{Y})} \tag{61}$$

where $(\hat{X},\hat{Y}) \sim P_{XY}$. Moreover, if $P_{XY} = Q_{XY}$ supremizes $\rho_{\rm m}(\hat{X};\hat{Y})$, than the equalities hold and in fact are achieved by the one-way communication protocol (in either way).

Proof: By symmetry both $\frac{I(U_i;Y|U^{i-1})}{I(U_i;X|U^{i-1})}$ and $\frac{I(U_i;X|U^{i-1})}{I(U_i;Y|U^{i-1})}$ are bounded by (54). The inequality follows from the rate region (10)-(12), the property (83) and the basic inequality $\frac{\sum_i a_i}{\sum_i b_i} \le \sup_i \frac{a_i}{b_i}$ for nonnegative (a_i) and (b_i) such that the fractions are defined.

The sufficient condition for the equality can be seem from (44) and the fact that $\rho_{\rm m}^2(\hat{X};\hat{Y}) \le s^*(X;Y)$ for any P_{XY}. ■

In general, the maximal correlation coefficient is much easier to compute than the strong data processing constant. Let us use boldface to denote a matrix corresponding to a discrete joint distribution. e.g.

$$\mathbf{P}_{XY} := [P_{XY}(x,y)]_{xy}, \tag{62}$$

with the marginal distributions always thought of as a column vector. Define

$$\mathbf{A} := \mathrm{diag}(\mathbf{P}_X)^{-\frac{1}{2}} \mathbf{P}_{XY}\, \mathrm{diag}(\mathbf{P}_Y)^{-\frac{1}{2}} \tag{63}$$

and let $\mathbf{M} := \mathbf{A}^\top \mathbf{A}$. Then $\rho_{\rm m}^2(\hat{X};\hat{Y})$ is the second largest eigenvalue value of $\mathbf{M}$ (c.f. [4]). See also [30] for an extension to non-discrete distributions.

Using the calculus of variation, we show next a necessary condition that the discrete distribution $P_{XY} = Q_{XY}$ achieves the supremum in (61), whose proof is deferred to Appendix C.

Definition 8. The *graph* of a discrete distribution distribution Q_{XY} is defined as the bipartite graph whose adjacency matrix is the sign of $\mathbf{Q}_{XY}$. We say Q_{XY} is *indecomposable* [47] if its graph is connected.

Theorem 4. *If $P_{XY} = Q_{XY}$ achieves the supremum in (61), then*

$$\mathbf{u}^{\circ 2} = \mathbf{Q}_{X|Y}\mathbf{v}^{\circ 2}; \tag{64}$$

$$\mathbf{v}^{\circ 2} = \mathbf{Q}_{Y|X}\mathbf{u}^{\circ 2}. \tag{65}$$

Moreover, if Q_{XY} is indecomposable and both Q_X and Q_Y are fully supported, then

$$\mathbf{u}^{\circ 2} = \mathbf{Q}_X; \tag{66}$$

$$\mathbf{v}^{\circ 2} = \mathbf{Q}_Y. \tag{67}$$

Remark 3. The conditions (66) and (67) need not be satisfied when Q_{XY} is not indecomposable (e.g. consider $X = Y$ binary but not equiprobable under P_{XY}).

Applying Theorem 4 to BSS, we have the following result, which may also be proved directly without invoking Theorem 4 (omitted here).

Theorem 5. *If Q_{XY} is a BSS with error probability $\epsilon \in [0,1]$, then*

$$\sup_{P_{XY} \preceq Q_{XY}} \rho_{\rm m}^2(\hat{X};\hat{Y}) = (1-2\epsilon)^2. \tag{68}$$

As a consequence, interaction does not increase CRBI or KBI for BSS:

$$\gamma_r(X;Y) = \frac{1}{1-(1-2\epsilon)^2}; \tag{69}$$

$$\Gamma_r(X;Y) = \frac{(1-2\epsilon)^2}{1-(1-2\epsilon)^2}. \tag{70}$$

where $r \in \{1, 2, \dots, \infty\}$.

Proof: We may assume without loss of generality that $\epsilon \in (0,1)$. Then by [28], the maximal correlation coefficient is continuous at any P_{XY} with fully supported marginal distribution. It is also elementary to show that $\rho_{\rm m}^2(\hat{X};\hat{Y})$ vanishes as either P_X or P_Y tends to a deterministic distribution. Therefore, the supremum in the definition of $\bar{\rho}_{\rm m}$ is achieved. By from Theorem 4, only $P_{XY} = Q_{XY}$ can possibly achieve the supremum. ■

Direct proof of Theorem 5: We only need to show that $\rho_{\rm m}^2(\hat{X};\hat{Y}) \le (1-2\epsilon)^2$ for any $P_{XY} \in \mathcal{P}(Q_{XY})$. Suppose

$$\mathbf{M} = \begin{pmatrix} x & \gamma \\ \beta & y \end{pmatrix} \tag{71}$$

is the matrix such that $\mathbf{P}_{XY}$ equals the Hadamard product

$$\begin{pmatrix} \bar{\epsilon} & \epsilon \\ \epsilon & \bar{\epsilon} \end{pmatrix} \circ \mathbf{M} \tag{72}$$

Although $\mathbf{M}$ is parameterized by 4 scalars, it only has two degrees of freedom because $P_{XY} \in Q_{XY}$ implies $\mathbf{M}$ is rank-one, and the sum of the coordinates of $\mathbf{P}_{XY}$ equals one. In fact, given the sum $s = \beta + \gamma$ and product $p = \beta\gamma$ of the two parameters, we can express the sum and product of x and y:

$$xy = p, \tag{73}$$

$$x + y = \frac{1 - \epsilon s}{\bar{\epsilon}}. \tag{74}$$

We know $\rho_{\mathrm{m}}(\hat{X};\hat{Y})$ is the second largest singular value of $\left[\frac{1}{\sqrt{P_{\hat{X}}(x)P_{\hat{Y}}(y)}} P_{\hat{X}\hat{Y}}(x,y)\right]_{x,y}$. After some systematic calculations, we can express it in terms of s and p:

$$\rho_{\mathrm{m}}^2(\hat{X};\hat{Y}) = \frac{(1-2\epsilon)^2 p}{(1-2\epsilon)^2 p + \epsilon(1-2\epsilon)s + \epsilon^2}. \tag{75}$$

For $P_{\hat{X}\hat{Y}} \in \mathcal{P}(Q_{XY})$, the admissible s and p satisfy

$$0 \le s \le \frac{1}{\epsilon}, \tag{76}$$

$$0 \le p \le \frac{1}{4} \min\left\{ \left(\frac{1-\epsilon s}{\bar{\epsilon}}\right)^2, s^2 \right\}. \tag{77}$$

Under the above conditions, it's elementary to show that (75) is maximized when $p = \frac{1}{4}$ and $s = 1$. ■

Remark 4. A celebrate central limit theorem argument by Gross [24] showed that Gaussian hypercontractivity can be obtained by BSS hypercontractivity. A similar argument for the key generation problem, c.f. [31] applied Theorem 5 implies that one-round communication is optimal for KBI for Gaussian sources as well. Moreover, we may define the *key per unit cost* if the communication costs in the two directions differ, and it is easy to from Theorem 5 that one-round communication is also optimal for achieving this quantity in the case of BSS or Gaussian sources.

C. *Strong Converse Property of the Upper-bound for CRBI*

The simple analysis in the previous section, essentially based on the Fano's inequality, gives a maximal correlation based upper-bound only for γ_r but not γ_r^δ. Proving strong converses for problems whose rate region involves auxiliary random variables is generally hard [44]. In this section, however, we prove that the maximal correlation bound also applies to γ_r^δ . The proof mainly draws on four ideas.

- The one-way communication model ($r = 1$) is viewed as a limiting degenerate case of the omniscient helper introduced in our previous paper [32]. Moreover, SDPI can be obtained as a limiting degenerate case of the hypercontractivity region [3, Theorem 5a]. In fact, we shall prove a clean one-shot converse for interactive CR generation where blocklength plays no role.

- For the last round of communication, we use the bounding technique for the omniscient helper problem in [32].
- The contribution from the previous rounds of communication is upper-bounded using certain chain rule properties of the Rényi divergence.

Consider the CR generation model in Figure 3, which is a special case of the omniscient helper CR generation problem in [32] where the helper (Terminal H) does not send any message to A. The only difference between the model in Figure 3 and the one-way model is that the helper knows (X, Y), whereas A only knows X. Hence the performance of the model in Figure 3 clearly dominates that of the one-way model. It turns out, however, that asymptotically they are equivalent [32].

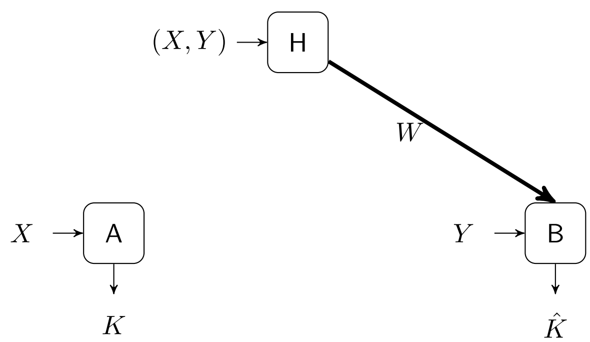

Figure 3: CR generation with an omniscient helper

The strong converse in [32] for CR per bit of communication uses the equivalence of the following two inequalities, which were proved independently by [11] and [36] using different methods.

Proposition 2 (Equivalent characterizations of hypercontractivity)**.** *Fix* Q_{XY} *and* $b, c \in (0, \infty)$. *Then*

$$\mathbb{E}[f(X)g(Y)] \leq \|f\|_{1/b}\|g\|_{1/c} \tag{78}$$

for all nonnegative f *and* g *if and only if*

$$D(S_{XY}\|P_{XY}) \geq bD(S_X\|P_X) + cD(S_Y\|P_Y) \tag{79}$$

for all $S_{XY} \ll P_{XY}$.

While (79) appears more connected to the single-letter solutions of the first order rate region, the functional characterization (78) provides a powerful tool in proving strong converses.

Definition 9. Given P_{XY}, define $\mathcal{G}(P_{XY})$ as the set of (b, c) for which (79) holds.

Now fix $(b, c) \in \mathcal{G}(P_{XY})$. If P_{XY} is not indecomposable, then $\Gamma_1(X;Y) = \infty$ and the problem is trivial. Below we assume that P_{XY} is indecomposable in which case for each $c \in (0,1)$ it is always possible to choose b such that $b + c > 1$ [4]. From the proof of [32, Theorem 10], we have the following bound for the model in Figure 3 (for possibly stochastic decoders):

$$\frac{1}{|\mathcal{K}|}\sum_k \mathbb{P}[K = \hat{K} = k]^{\frac{1}{b+c}} \le \frac{1}{|\mathcal{K}|}|\mathcal{W}|^{\frac{c}{b+c}} \tag{80}$$

Letting T be the correct distribution under which $K = \hat{K}$ is equiprobability distributed on $\mathcal{K}$. Then (80) can represented equivalently in terms of Rényi divergence:

$$D_\alpha(P_{K\hat{K}}\|T) = \log|\mathcal{K}| - \frac{1}{1-\alpha}\log\left(\sum_k \mathbb{P}[K = \hat{K} = k]^{\frac{1}{b+c}}\right) \tag{81}$$

$$\ge \log|\mathcal{K}| - \frac{1}{1-\alpha}\cdot\frac{c}{b+c}\log|\mathcal{W}| \tag{82}$$

$$= -\frac{c}{b+c-1}\log|\mathcal{W}| + \log|\mathcal{K}| \tag{83}$$

where $\alpha := \frac{1}{b+c} < 1$.

Remark that the Rényi divergence may be regarded as a natural performance measure itself, and it will of interest to investigate its properties, e.g. universal composability in the future work. Here we only point out that it is related to the total variation distance via the following basic result, which was essentially presented in [32, Theorem 10]:

Proposition 3. *Suppose Q is the equiprobable on $\{1, \dots, M\}$ and P is an arbitrary distribution on the same alphabet. For any $\alpha \in (0,1)$,*

$$D_\alpha(P\|Q) \le \frac{1}{1-\alpha}\log\frac{1}{1 - \frac{1}{M} - \frac{1}{2}|P - Q|}. \tag{84}$$

Now return to the interactive CR generation model, and let Q_{XYW^r} be the joint distribution of the source and the messages. Assume without loss of generality that r is an odd number, that is, the last round of communication is from A to B. It is trivial to show by induction that for any $(w_1, \dots, w_{r-1})$, we have

$$Q_{XY|W^{r-1}=w^{r-1}} \preceq Q_{XY}. \tag{85}$$

In the last round of the communication, we can apply the bound in (83) with $P_{XY} \leftarrow Q_{XY|W^{r-1}=w^{r-1}}$. Suppose $(b, c) \in \mathcal{G}(P_{XY|W^{r-1}=w^{r-1}})$ then (83) we obtain

$$D_{\frac{1}{b+c}}(Q_{K\hat{K}|W^{r-1}=w^{r-1}}\|T) \ge -\frac{c}{b+c-1}\log|\mathcal{W}_r| + \log|\mathcal{K}|. \tag{86}$$

This above analysis takes care of the contribution to CR from the last round of communication. Next, we handle the contribution from the previous rounds.

Proposition 4 (Chain rule for Rényi divergence)**.**

$$D_\alpha(P_{XY}\|Q_{XY}) = D_\alpha(P_{Y|X}\|Q_{Y|X}|P_X^{(\alpha)}) + D_\alpha(P_X\|Q_X) \tag{87}$$

where

$$P_X^{(\alpha)} := \frac{P_X^\alpha Q_X^{1-\alpha}}{\int \mathrm{d}P_X^\alpha \mathrm{d}Q_X^{1-\alpha}}. \tag{88}$$

Using the monotonicity of the norm, we can show that

Proposition 5. *For discrete X,*

$$D_\alpha(P_Y\|Q_Y) + H_\alpha(P_X) \geq D_\alpha(P_{Y|X}\|Q_Y|P_X^{(\alpha)}) \tag{89}$$

where $P_X^{(\alpha)} := \frac{P_X^\alpha}{\int \mathrm{d}P_X^\alpha}$.

From Proposition 4, using (86) and letting $Q_{W^{r-1}}$ be the equiprobable distribution,

$$\begin{aligned} D_{\frac{1}{b+c}}(Q_{K\hat{K}}\|T_{K\hat{K}}) &\geq \min_{w^{r-1}} D_{\frac{1}{b+c}}(Q_{K\hat{K}|W^{r-1}=w^{r-1}}\|T_{K\hat{K}}) - H_{\frac{1}{b+c}}(Q_{W^{r-1}}) && (90)\\ &\geq -\frac{c}{b+c-1}\log|\mathcal{W}_r| + \log|\mathcal{K}| - \log|\mathcal{W}^{r-1}| && (91)\\ &\geq \log|\mathcal{K}| - \frac{c}{b+c-1}\log|\mathcal{W}^r|. && (92)\end{aligned}$$

Denote by $\mathcal{W} := \mathcal{W}^r$ the alphabet of total communications. We have the following one-shot converse bound

Theorem 6. *If $(b,c) \in \bigcap_{P_{XY}\preceq Q_{XY}} \mathcal{G}(P_{XY})$ and $b+c>1$, then for CR generation*

$$D_{\frac{1}{b+c}}(Q_{K\hat{K}}\|T_{K\hat{K}}) \geq \log|\mathcal{K}| - \frac{c}{b+c-1}\log|\mathcal{W}|. \tag{93}$$

The connection to the maximal correlation bound is seen through the following result, whose proof is follows from (56) and the argument in the proof of [3, Theorem 5a] regarding a relation between hypercontractivity and SDI for finite-alphabet distributions.

Proposition 6. *For any indecomposable finite-alphabet source Q_{XY} and $\epsilon > 0$, there exists (b,c) such that $(b,c) \in \bigcap_{P_{XY}\preceq Q_{XY}} \mathcal{G}(P_{XY})$, $b+c>1$, $b\leq 1$ and that*

$$\frac{1-b}{c} \leq \sup_{P_{XY}\preceq Q_{XY}} \rho_{\mathrm{m}}^2(\hat{X};\hat{Y}) + \epsilon. \tag{94}$$

Combining Theorem 6 Proposition 3 and Theorem 6, we obtain the following asymptotic result:

Corollary 1. *For any finite-alphabet stationary memoryless source with per-letter distribution Q_{XY}, if the ratio*

$$\frac{\log|\mathcal{K}|}{\log|\mathcal{W}|} > \frac{1}{1-\sup_{P_{\mathsf{XY}}\preceq Q_{\mathsf{XY}}}\rho_{\mathsf{m}}^2(\hat{X};\hat{Y})} \tag{95}$$

is fixed where $(\hat{X},\hat{Y}) \sim P_{XY}$, then, for CR generation (allowing local randomization), regardless of the blocklength, we have the following as $|\mathcal{K}| \to \infty$:

1) *There is some $0 \leq \alpha < 1$ such that $D_\alpha(Q_{K\hat{K}}\|T_{K\hat{K}})$ grows at least linearly with a strictly positive slope in $\log|\mathcal{K}|$.*
2) *$1 - \frac{1}{2}|Q_{K\hat{K}} - T_{K\hat{K}}|$ vanishes at least polynomially in $|\mathcal{K}|$.*

In particular,

$$\gamma_\infty^\delta(X;Y) \le \sup_{P_{XY} \preceq Q_{XY}} \frac{1}{1-\rho_{\mathrm{m}}^2(\hat{X};\hat{Y})} \tag{96}$$

holds for any $\delta \in (0,1)$.

V. Minimum Interaction for Maximal Key Rate

Define $I_r(Q_{XY})$ as the minimal interactive communication rate needed for maximal key rate in r rounds, starting from A to B. More precisely, it can be defined in the following ways from the rate region or the multi-letter characterization of the rate region:

$$I_r(Q_{XY}) = \inf\{r_1, r_2 \colon (r, r_1, r_2) \in \mathcal{R}_r(X,Y),\ \exists r\} \tag{97}$$

$$= \inf\{d - I(X;Y) \colon (d, r) \in \mathcal{S}_r(X,Y),\ \exists r\} \tag{98}$$

$$= \inf\left\{d \colon \sup_{U^r \colon S(Q_{XYU^r}) - I(X;Y) \le d} R(Q_{XYU^r}) = I(X;Y)\right\} \tag{99}$$

where $S(Q_{XYU^r})$ and $R(Q_{XYU^r})$ were defined in Section III. We then have the following general concave envelope characterization. Its proof is essentially based on a very simple geometric fact about the supporting hyperplane of a convex set (see Figure 4), which should be applicable to other similar problems as well.

Theorem 7. *For a stationary memoryless source with per-letter distribution* Q_{XY}*,*

$$I_r(Q_{XY}) = H(X|Y) + H(Y|X) - \lim_{s \downarrow 0} \frac{1}{s} \omega_r^s(Q_{\mathsf{XY}}). \tag{100}$$

where ω_r^s *is as in* (34).

Again, for non-discrete distributions we may choose a reference measure and replace the entropy with the relative entropy in the analysis, so a similar result holds, *mutatis mutandis*. However, it should be pointed out that $I_r(Q_{XY})$ is *usually* infinite for non-discrete sources, such as the Gaussian source.

Proof: From (36), the right side of (100) equals

$$\lim_{s \downarrow 0} \inf_{U^r} \left\{S(Q_{XYU^r}) - I(X;Y) + \frac{1}{s}[I(X;Y) - R(Q_{XYU^r})]\right\} \tag{101}$$

$$= \sup_{s > 0} \inf_{U^r} \left\{S(Q_{XYU^r}) - I(X;Y) + \frac{1}{s}[I(X;Y) - R(Q_{XYU^r})]\right\}. \tag{102}$$

From (99), the infimum in (102) is upper-bounded by $I_r(Q_{XY})$ for any s, establishing the $\ge$ part of (100). For the other direction, choose an arbitrary $\epsilon > 0$. Here, note that $\mathcal{S}_r(X,Y)$ is a closed convex subset of $[0,\infty) \times [0, I(X;Y)]$, and it contains the line $[I(X;Y) + I_r(Q_{XY})] \times \{I(X;Y)\}$. It is easy to check that $(I(X;Y) + I_r(Q_{XY}) - \epsilon, I(X;Y)) \notin \mathcal{S}_r(X,Y)$, so by the Hahn-Banach theorem (hyperplane separation theorem), there exists an $s > 0$ such that for all U^r,

$$R(Q_{XYU^r}) \le s(S(Q_{XYU^r}) - I(X;Y) - I_r(Q_{XYU^r}) + \epsilon) + I(X;Y). \tag{103}$$

For such an s, the infimum in (102) is lower-bounded by $I_r(Q_{XY}) - \epsilon$, as desired. ∎

Next we shall provide a even simpler characterization of the MIMK. Define

$$\sigma_0(P_{XY}) := \begin{cases} H(\hat{X},\hat{Y}) & I(\hat{X};\hat{Y})=0; \\ -\infty & \text{otherwise.} \end{cases} \tag{104}$$

For $r \in \{1,2,\dots\}$, define

$$\sigma_r := \begin{cases} \text{env}_X(\sigma_{r-1}) & r \text{ is odd;} \\ \text{env}_Y(\sigma_{r-1}) & r \text{ is even,} \end{cases} \tag{105}$$

and define σ_∞ as the XY-concave envelope of σ_0. In view of (31) and (32), we can express $\sigma_r(Q_{XY})$ in a form similar to (102).

$$\sigma_r(Q_{XY}) = \inf_{U^r}\sup_{s>0}\left\{S(Q_{XYU^r}) - I(X;Y) + \frac{1}{s}[I(X;Y) - R(Q_{XYU^r})]\right\}. \tag{106}$$

The result below shows that in the finite alphabet case, we can indeed switched the order of the supremum and the infimum. As is often the case, compactness guarantees such saddle point properties. The proof is rather simple for fully supported Q_{XY}, but much more ideas are needed for general discrete alphabets; see Appendix A.

Lemma 1. *Fix a Q_{XY} on a finite alphabet. For any $P_{XY} \in \mathcal{P}(Q_{XY})$ and $r \in \{0,1,2,3,\dots,\infty\}$,*

$$\sigma_r(P_{XY}) = \lim_{s\downarrow 0}\frac{1}{s}\omega_r^s(P_{XY}). \tag{107}$$

Proof: The pointwise convergence (107) trivially holds when $r = 0$, in view of the definition (104). For other values of r, the proof follows by induction, using the fact that $\omega_r^s(P_{XY})$ monotonically decreases in s and Proposition 7. Note that the nonnegativity assumption in Proposition 7 because $\frac{1}{s}\omega_r^s(P_{XY}) = H(P_{XY}) > 0$ when either $\hat{X}$ or $\hat{Y}$ is constant. ■

Theorem 8. *If Q_{XY} is a distribution on a finite alphabet, then for $r \in \{1,\dots,\infty\}$*

$$I_r(Q_{XY}) = H(Y|X) + H(X|Y) - \sigma_r(Q_{XY}). \tag{108}$$

Proof: Immediate from Lemma 1 and Theorem 7. ■

Corollary 2. *If X and Y are both binary under Q_{XY}, then the necessary and sufficient condition for*

$$\min\{I^1(Q_{XY}), I^1(Q_{YX})\} = I^\infty(Q_{YX}) \tag{109}$$

is either $Q_{Y|X}$ or $Q_{X|Y}$ is a binary symmetric channel (Q_X-almost surely or Q_Y almost surely).

Remark 5. Tyagi [45] introduced a concept called "interactive common randomness" and showed its relation to the minimum rate of interactive communication needed to generate the maximal amount of key. Then by drawing an elegant connection to sufficient statistics, Tyagi [45, Theorem 9] proved Corollary 2 above in the case of binary symmetric Q_{XY}, and *conjectured* that (109) holds for all binary sources. Here we have provided the necessary and sufficient condition for the conjecture to hold with an entirely different approach.

Proof: The sufficiency is convenient to check. Suppose without loss of generality that $Q_{Y|X}$ is a BSC with crossover probability ϵ, and consider only the nontrivial case that Q_X is fully supported. There exists $\epsilon \in [0,1]$ such that $\mathcal{P}(Q_{XY})$ can be parameterized by f and g as

$$P_{XY} = \frac{1}{Z}\begin{pmatrix} \bar{\epsilon}\bar{f}\bar{g} & \epsilon\bar{f}g \\ \epsilon f\bar{g} & \bar{\epsilon}fg \end{pmatrix} \tag{110}$$

where the normalization constant

$$Z := f * \bar{g} * \epsilon. \tag{111}$$

That is, there exists a one to one correspondence from (f,g) to $P_{XY} \in \mathcal{P}(Q_{XY})$. Let π be such a bijection, and $\pi^X(f,g)$ (resp. $\pi^Y(f,g)$) be the X-marginal (resp. Y-marginal) of $\pi(f,g)$. To to avoid cumbersome notations, we will sometimes write functionals like $\sigma_r(f,g)$ instead of $\sigma_r(\pi(f,g))$, but keep in mind that concavity are always w.r.t. to the probability distributions rather than (f,g).

There exists a real number c such that

$$h\left(\frac{\epsilon g}{\bar{g} * \epsilon}\right) \le h(\epsilon) - \frac{c}{2}\cdot\frac{g-\frac{1}{2}}{\bar{g} * \epsilon} \tag{112}$$

$$= h(\epsilon) - \frac{c}{4\epsilon} + \frac{c}{4\epsilon\bar{\epsilon}}\cdot\frac{\bar{g}\bar{\epsilon}}{\bar{g} * \epsilon} \tag{113}$$

for all g, since the function on the right hand side is linear and the function on the left hand side is concave (both viewed as functions of the binary distribution $\left(\frac{\epsilon g}{\bar{g}*\epsilon}, \frac{\bar{\epsilon}\bar{g}}{\bar{g}*\epsilon}\right)$), and for any c both functions have the same evaluation at the equiprobable distribution $(\frac{1}{2},\frac{1}{2})$. On the other hand, $\sigma_0(f,g) = -\infty$ if $fg\bar{f}\bar{g} > 0$, and $\sigma_0(f,g) = h\left(\frac{\epsilon g}{\bar{g}*\epsilon}\right)$ when $f = 0$ (and similar expressions for $f = 1$, $g = 0$, or $g = 1$ cases). These imply that the XY-concave function

$$\ell(f,g)\colon \mathcal{P}(Q_{XY}) \to \mathbb{R}, \tag{114}$$

$$(f,g) \mapsto h(\epsilon) + \frac{c(f-\frac{1}{2})(g-\frac{1}{2})}{f * \bar{g} * \epsilon} \tag{115}$$

upper bounds σ_0, hence upper bounds σ_∞. Then we have, noting that $g = \frac{1}{2}$ for Q_{XY},

$$h(\epsilon) = \sigma_1(Q_{XY}) \le \sigma_\infty(Q_{XY}) \le \ell\left(f, \frac{1}{2}\right) = h(\epsilon) \tag{116}$$

Thus $\sigma_1(Q_{XY}) = \sigma_\infty(Q_{XY})$, which, by Theorem 8, implies that (109) holds.

To show the necessity, notice first that if either X or Y is deterministic, the key capacity is zero hence both sides of (109) are zero. There are two cases remaining regarding the support of Q_{XY}:

1) Q_{XY} is fully supported. In this case there exists $\epsilon \in (0,1)$ such that $\mathcal{P}(Q_{XY})$ can again be parameterized as (110). Observe that

$$\pi^X(f,g) = \left(1 - \frac{g(\epsilon * f)}{f * \bar{\epsilon} * g}, \frac{g(\epsilon * f)}{f * \bar{\epsilon} * g}\right) \tag{117}$$

so it is straightforward to check that the solution of $\lambda \in [0,1]$ to

$$\lambda\pi^X(1,g) + \bar{\lambda}\pi^X(0,g) = \pi^X\left(\frac{1}{2}, g\right) \tag{118}$$

is given by

$$\lambda = \epsilon * g. \tag{119}$$

Then by definition,

$$\sigma_1\left(\frac{1}{2}, g\right) = \bar{\lambda}\sigma_0(0, g) + \lambda\sigma_0(1, g) \tag{120}$$

$$= \bar{\lambda}h\left(\frac{\epsilon g}{\bar{\epsilon} * g}\right) + \lambda h\left(\frac{\epsilon \bar{g}}{\epsilon * g}\right) \tag{121}$$

$$= -h(\epsilon * g) + h(\epsilon) + h(g) \tag{122}$$

$$\leq h(\epsilon) \tag{123}$$

$$= \sigma_2\left(\frac{1}{2}, g\right) \tag{124}$$

where the inequality in (123) is strict unless $g = \frac{1}{2}$, and (124) follows along the same line as (116). Note that for $g \in (0,1)$, $\sigma_1(\cdot, g)$ is X-linear and $\sigma_3(\cdot, g)$ is X-concave. If, additionally, $g \neq \frac{1}{2}$, then (124) shows that $\sigma_3\left(\frac{1}{2}, g\right) > \sigma_1\left(\frac{1}{2}, g\right)$, which implies that

$$\sigma_3(\cdot, g) > \sigma_1(\cdot, g) \tag{125}$$

except possibly at the endpoints (i.e. when $f \in \{0, 1\}$). In sum, we have shown

$$\sigma_3(f, g) > \sigma_1(f, g) \tag{126}$$

except when $f \in \left\{0, \frac{1}{2}, 1\right\}$ or $g \in \left\{0, \frac{1}{2}, 1\right\}$. In other words, if neither $Q_{Y|X}$ or $Q_{X|Y}$ is a BSC, then

$$\sigma_\infty(Q_{XY}) \geq \sigma_3(Q_{XY}) > \sigma_1(Q_{XY}), \tag{127}$$

and by symmetry, we also have

$$\sigma_\infty(Q_{XY}) \geq \sigma_3(Q_{YX}) > \sigma_1(Q_{YX}) \tag{128}$$

which implies that the left of (109) is strictly larger than the right.

2) $|\operatorname{supp}(Q_{XY})| = 3$. Assume without loss of generality that $Q_{XY}(0,0) = 0$. We can parameterize $\mathcal{P}(Q_{XY})$ with f, g via the map

$$\pi\colon [0,1]^2 \setminus \{(0,0)\} \to \Delta(\mathcal{X} \times \mathcal{Y}) \tag{129}$$

$$(f, g) \mapsto \frac{1}{f + \bar{f}g}\begin{bmatrix} 0 & \bar{f}g \\ f\bar{g} & fg \end{bmatrix}. \tag{130}$$

Observe that

$$\pi^X(f, g) = \left(1 - \frac{f}{f + \bar{f}g}, \frac{f}{f + \bar{f}g}\right), \tag{131}$$

so it is straightforward to check that the solution of $\lambda \in [0,1]$ to

$$\lambda\pi^X(1, g) + \bar{\lambda}\pi^X(0, g) = \pi^X(f, g) \tag{132}$$

is given by

$$\lambda = \frac{f}{f + \bar{f}g}. \tag{133}$$

Thus,

$$\sigma_1(f, g) = \lambda \sigma_0(1, g) + \bar{\lambda} \cdot 0 \tag{134}$$

$$= \lambda \sigma_0(1, g) \tag{135}$$

$$= \frac{f}{f + \bar{f}g} h(g). \tag{136}$$

Next, for fixed $f \neq 0$ put $g = g_f(x) := \frac{xf}{1 - \bar{f}x}$. Then

$$\sigma_1(f, g_f(x)) = (1 - \bar{f}x) h \left(\frac{fx}{1 - \bar{f}x} \right). \tag{137}$$

A short computation shows

$$\frac{\mathrm{d}^2}{\mathrm{d}x^2} \sigma_1(f, g_f(x)) = -\frac{f}{x(1 - x)(1 - \bar{f}x)} < 0. \tag{138}$$

Noticing that $\pi^Y(f, g) = (1 - x, x)$, this implies that $\sigma_0(f, \cdot)$ is *strictly* Y-concave for $f \neq 0$. Now suppose there exist some (f_0, g_0), $(f_0 g_0 \bar{f}_0 \bar{g}_0 \neq 0)$ such that

$$\sigma_1(f_0, g_0) = \sigma_\infty(f_0, g_0). \tag{139}$$

Since $\sigma_0(\cdot, g_0)$ is linear (caution: in the distribution rather than in f), $\sigma_\infty(\cdot, g_0)$ is concave, and both functions agree on the endpoints, (139) implies that, actually,

$$\sigma_1(\cdot, g_0) = \sigma_\infty(\cdot, g_0), \tag{140}$$

and in particular $\sigma_\infty(\cdot, g_0)$ is linear. By symmetry, $\sigma_\infty(g_0, \cdot)$ is also linear. This is a contradiction since $\sigma_\infty(g_0, \cdot)$ and $\sigma_1(g_0, \cdot)$ agree at two points $g = 0, g_0$ but the former linear functions dominates the latter strictly concave function. Thus (139) is impossible, and in particular, we conclude by symmetry that

$$\sigma_1(Q_{XY}) < \sigma_\infty(Q_{XY}); \tag{141}$$

$$\sigma_1(Q_{YX}) < \sigma_\infty(Q_{YX}) = \sigma_\infty(Q_{XY}), \tag{142}$$

as desired.

■

VI. Discussion

Theorem 5 and Corollary 2 each says that allowing interaction does not improve one-way communication scheme for BSS either when the communication rate is very low or high enough to achieve the maximal key rate. These two pieces of facts naturally lead to:

Conjecture 1. For an i.i.d. symmetric Bernoulli source Q_{XY}, $S_1(X, Y) = S_\infty(X, Y)$.

For a BSS (X, Y) and under the protocol of one way communication from A to B, the optimal key-rate–communication rate is $(I(U;Y), I(U;X) - I(U;Y))$, parameterized by the symmetric Bernoulli auxiliary random variable U satisfying $U - X - Y$. The optimality of such auxiliary random variables can be shown using the concavity of the function $x \mapsto h(\epsilon * h^{-1}(x))$; see also the proof of Proposition 5.3 in [13] for the case involving an eavesdropper. What is less obvious is that such a scheme is also optimal among protocols allowing interactions, as Conjecture 1 postulates. If Conjecture 1 holds, we will also be able to conclude that $S_1(X', Y') = S_\infty(X', Y')$ for any $Q_{X'Y'}$ such that $Q_{Y'|X'}$ is a BSC, and in fact $S_\infty(X', Y')$ will be the intersection between a translation of $S_r(X, Y)$ and the first quadrant. In Appendix D, we argue that Conjecture 1 will be implied by a conjectured inequality involving four parameters, whose validity has been supported by reasonably extensive numerical computations. From the numerical results, the inequality is close to failure only in the regime of very small communication rates and very noisy BSS, but in former case, Theorem 5 has guaranteed the validity of the conjecture, while in the latter case, we proved the inequality using Taylor expansion in Appendix E.

Beyond the scope of CR/key generation problem in this paper, we hope some of our methods to become useful in other areas. For example, we have already seen that the XY-absolute continuity framework allows us to define the strong data processing constant directly from a joint distribution without worrying about the technical difficulty of determining the conditional distribution from the joint. The newly introduced symmetrical strong data processing constant (Definition 6) has a concave envelope definition very similar to the conventional strong data processing constant, and it worth exploring its significance in other contexts as well as its mathematical properties. The Rényi divergence based performance measure for CR generation in Section IV-C seems new, and this as well as the hypercontractivity based converse proof techniques opens a direction of future research. Moreover, techniques used for analyzing the concave envelope characterization, such as expressing the MIMK as a limit as the slope of the supporting line vanishes in Theorem 7 and the minimax result for finite-alphabet distributions in Lemma 1 (based on fundamental properties of XY-concave envelopes in Appendix A-B) are likely to be useful in the related interactive source coding problem or the broader area of interactive function computation originally studied in computer science [49], which has gained increasing popularity through some recent works in the CS community including [7][8][10][21] (see also the ISIT tutorial [9]).

APPENDIX A
SEMICONTINUITY OF XY-CONCAVE FUNCTIONS

Recall that a concave function on a simplex which is lower bounded (or more or less equivalently, nonnegative) on the vertices is necessarily lower semicontinuous (c.f. [40, Theorem 10.2]). For XY-concave functions, we prove a similar basic result, which will be used in the proof of Proposition 7.

Lemma 2. *Given a distribution Q_{XY} where Q_X and Q_Y are fully supported and $\mathcal{X} = \{1, \dots, m\}$, $\mathcal{Y} = \{1, \dots, n\}$.*

1) *If Q_{XY} is indecomposable, then there exists unique $(\mathbf{f}, \mathbf{g}) \in \mathbb{R}^m \times \mathbb{R}^n$ such that*

$$\mathbf{f}\mathbf{g}^\top \circ \mathbf{Q}_{XY} = \mathbf{Q}_{XY} \tag{143}$$

and the first coordinate $f_1 = 1$.

2) *Fix an* $S_{XY} \in \mathcal{P}(Q_{XY})$. *For any* $\delta \in (0,1)$, *there exists an* $\epsilon > 0$ *such that for any (possibly unnormalized)* $\mu_{XY} \preceq S_{XY}$ *satisfying* $|\mu_{XY} - S_{XY}| \le \epsilon$, *we can find* T_{XY} *satisfying*

$$S_{XY} \preceq_X T_{XY}; \tag{144}$$

$$T_{XY} \preceq_Y \mu_{XY}; \tag{145}$$

$$|S - T| \le \delta; \tag{146}$$

$$|T - \mu| \le \delta. \tag{147}$$

Note that (144) *and* (145) *imply that, actually,* $S \sim T \sim \mu$.

3) *A* XY*-concave function on* $\mathcal{P}(Q_{XY})$ *which is nonnegative when either* X *or* Y *is deterministic is necessarily lower semicontinuous.*

Proof:

1) Since the graph of Q_{XY} is connected, we can start from f_1 and visit all vertices of the bipartite graph to see that all the coordinates of $\mathbf{f}$ and $\mathbf{g}$ are uniquely determined.
2) It is without loss of generality to only prove the case of $S_{XY} = Q_{XY}$. Suppose the graph of Q_{XY} has k connected components, and assume without loss of generality that $1, \dots, k$ are X-vertices belonging to different connected components. Consider

$$\pi \colon \mathbb{R}^{m-k} \times \mathbb{R}^n \to \mathbb{R}^{mn} \tag{148}$$

$$(\bar{\mathbf{f}}, \mathbf{g}) \mapsto \mathbf{f}\,\mathbf{g}^\top \circ \mathbf{Q}_{XY} \tag{149}$$

where $\mathbf{f}$ is an m-vector whose first k coordinates are 1 and last $(m-k)$-coordinates are $\bar{\mathbf{f}}$. Denote by $\mathbf{e}_l$ the l-vector ($l \ge 1$) whose coordinates are all 1. Then π is an embedding from a neighborhood of $\mathbf{e}_{m+n-k}$ to $\mathbb{R}^{mn}$ (c.f. [6]), because it is standard to check that the rank of the differential of π at $\mathbf{e}_{m+n-k}$ is $m-k+n$ (full rank), where the calculation is essentially reduced to the case of an indecomposable distribution and property 1) can be used. Thus there is an open neighborhood $\mathcal{O}$ of $\mathbf{e}_{m+n-k}$ homeomorphic to its image under π, and in particular π has a continuous inverse on $\pi(\mathcal{O})$. Consequently, there exists $\epsilon \in (0,1)$ such that if $\mu_{XY} \preceq Q_{XY}$ and $|\mu - Q| \le \epsilon$, then $\mu \in \pi(\mathcal{O})$ and, with $\mathbf{M}$ defined as the matrix of μ_{XY}, $(\bar{\mathbf{f}}, \mathbf{g}) = (\pi|_{\mathcal{O}})^{-\mathbf{1}}(\mathbf{P})$ satisfies

$$\|\mathbf{f} - \mathbf{e}_m\|_\infty < \delta/2; \tag{150}$$

$$\|\mathbf{g} - \mathbf{e}_n\|_\infty < \delta/4, \tag{151}$$

where $\mathbf{f} = (\mathbf{e}_k, \bar{\mathbf{f}})$ as before. Put $\nu_{XY}(x,y) = f(x)Q_{XY}(x,y)$, and observe that (144)-(145) are satisfied because $\mathbf{f}$ and $\mathbf{g}$ have strictly positive coordinates. Also,

$$|Q-\nu| \le \|\mathbf{e}_m - \mathbf{f}\|_\infty \sum_{x,y} Q_{XY}(x,y) \tag{152}$$

$$\le \delta/2; \tag{153}$$

$$|\nu-\mu| \le \|\mathbf{e}_n - \mathbf{g}\|_\infty \sum_{x,y} f(x) Q_{XY}(x,y) \tag{154}$$

$$\le \|\mathbf{e}_n - \mathbf{g}\|_\infty \sum_{x,y} 2Q_{XY}(x,y) \tag{155}$$

$$\le \delta/2. \tag{156}$$

But from (150), $1-\delta/2 < |\nu| < 1+\delta/2$, so the probability distribution $T := \frac{1}{|\nu|}\nu$ satisfies

$$|\nu - T| < \delta/2. \tag{157}$$

Then (146)-(147) holds by the triangle inequality.

3) Consider an $S_{XY} \in \mathcal{P}(Q_{XY})$. Denote by $a > 0$ the minimum nonzero entry of S_{XY}, and assume without loss of generality that S_X and S_Y are supported on $\{1,\dots,m_1\}$ and $\{1,\dots,n_1\}$, respectively. For any $\delta \in (0, a/4)$, find $\epsilon > 0$ as in 2). For any $R \in \mathcal{P}(Q_{XY})$ satisfying

$$|R - S| \le \epsilon, \tag{158}$$

define, for $x \in \{1,\dots,m\}$ and $y \in \{1,\dots,n\}$,

$$\mu(x,y) := R(x,y) 1\{x \le m_1, y \le n_1\}. \tag{159}$$

Invoke 2) and find T satisfying (144)-(147). We have

$$T \ge \frac{a-\delta}{a} S, \tag{160}$$

so that

$$T = \frac{a-\delta}{a} S + \sum_{x\in\mathcal{X}} \lambda_x D_x \tag{161}$$

where each $D_x \preceq_X T$ is a distribution under which X is deterministic, and $\sum_x \lambda_x = \frac{\delta}{a}$. Denote by σ the XY-concave function in question. By its marginal concavity,

$$\sigma(T) \ge \left(1 - \frac{\delta}{a}\right)\sigma(S) + \sum_{x\in\mathcal{X}} \lambda_x \sigma(D_x) \tag{162}$$

$$\ge \left(1 - \frac{\delta}{a}\right)\sigma(S). \tag{163}$$

Since the minimum nonzero entry in T is at least $a - \delta > a/2$, we have

$$\tilde{R} \ge \mu \ge \frac{a-2\delta}{a} T, \tag{164}$$

where $\tilde{R} := \frac{1}{|\mu|}\mu = R_{XY|X\le m_1, Y\le n_1}$, so a similar argument also shows that

$$\sigma\left(\tilde{R}\right) \ge \left(1 - \frac{2\delta}{a}\right)\sigma(T). \tag{165}$$

Moreover, consider $\tilde{R}^1 := R_{XY|Y\le n_1}$. Since $1-\epsilon \le |\mu| \le 1$ by (158), we have

$$\tilde{R} \preceq_X \tilde{R}^1; \tag{166}$$

$$\tilde{R}^1 \preceq_Y R; \tag{167}$$

$$(1-\epsilon)\tilde{R} \le \tilde{R}^1; \tag{168}$$

$$(1-\epsilon)\tilde{R}^1 \le R, \tag{169}$$

so applying the similar argument again,

$$\sigma(\tilde{R}^1) \ge (1-\epsilon)\sigma(\tilde{R}); \tag{170}$$

$$\sigma(R) \ge (1-\epsilon)\sigma(\tilde{R}^1). \tag{171}$$

Assembling (163), (165), (170), (171) and noting that δ and ϵ can be chosen to be arbitrarily small, we must have

$$\liminf_{R\to S} \sigma(R) \ge \sigma(S). \tag{172}$$

■

Appendix B
Pointwise Convergence of Marginally Convex Envelopes

The following result forms the basis of the proof of Lemma 1.

Proposition 7. 1) *Suppose $(f_k)_{k\in(0,\infty)}$ is a family of continuous functions on a simplex Δ, where $f_k(x)$ is nondecreasing in k for any $x \in \Delta$. Define $f(x) := \lim_{k\downarrow 0} f_k(x)$. If $\operatorname{env} f$ is nowhere $-\infty$, then*

$$\operatorname{env} f(x) = \lim_{k\downarrow 0} \operatorname{env} f_k(x) \tag{173}$$

for any $x \in \Delta$.

2) *Consider a Q_{XY} on a finite alphabet with fully supported Q_X and Q_Y. Suppose $(f_k)_{k\in(0,\infty)}$ is a family of continuous functions on a $\mathcal{P}(Q_{XY})$, where $f_k(P_{XY})$ is nondecreasing in k for any $P_{XY} \in \mathcal{P}(Q_{XY})$, and f_k is nonnegative when either X or Y is deterministic. Define $f(P_{XY}) := \lim_{k\downarrow 0} f_k(P_{XY})$ for each $P_{XY} \in \mathcal{P}(Q_{XY})$. Then*

$$\operatorname{env}_{XY} f(Q_{XY}) = \lim_{k\downarrow 0} \operatorname{env}_{XY} f_k(Q_{XY}). \tag{174}$$

Remark 6. There are simple counterexamples to show that, in general, taking the limit and the concave envelope can not be switched if a sequence of continuous functions is only assumed to converge pointwise to a certain continuous function. Moreover if the sequence of functions are decreasing but not necessarily continuous, the switching can also fail. Therefore both the monotonicity of $\omega_r^s(P_{XY})$ in s and the continuity in P_{XY} play an essential role in the proof of Lemma 1.

Proof: For 1), the $\operatorname{env} f(x) \le \lim_{k\downarrow 0} \operatorname{env} f_k(x)$ part is trivial. For the opposite direction, notice that the following statements are equivalent:

$$\operatorname{env} f + \epsilon > f_k \quad \text{for some } k > 0 \tag{175}$$

$$\Longleftrightarrow \{x\colon \operatorname{env} f(x) + \epsilon - f_k(x) \le 0\} = \emptyset \quad \text{for some } k > 0 \tag{176}$$

$$\Longleftrightarrow \bigcap_{k>0} \{x\colon \operatorname{env} f(x) + \epsilon - f_k(x) \le 0\} = \emptyset \tag{177}$$

$$\Longleftrightarrow \sup_{k>0} (\operatorname{env} f + \epsilon - f_k) > 0 \tag{178}$$

$$\Longleftrightarrow \operatorname{env} f + \epsilon > \inf f_k = f \tag{179}$$

where (177) is the main step which follows from Cantor's intersection theorem. More precisely, notice that a concave function on a simplex is lower semicontinuous [40, Theorem 10.2], so $\operatorname{env} f + \epsilon - f_k$ is lower semicontinuous, and the set in (176) is closed in Δ, hence compact. Then (177) follows because a decreasing nested sequence of non-empty compact subsets of the Euclidean space has nonempty intersection. Since (179) holds for all ϵ, we have from (175) and the concavity of $\operatorname{env} f + \epsilon$ that

$$\operatorname{env} f + \epsilon \ge \operatorname{env} f_k \tag{180}$$

for some k. Therefore $\operatorname{env} f(x) \ge \lim_{k\downarrow 0} \operatorname{env} f_k(x)$ must hold because ϵ is arbitrary.

The proof of 2) is similar. We need the semicontinuity of the XY-concave function proved in Lemma 2.3. ■

Appendix C
Proof of Theorem 4

Consider a small perturbation, so that

$$\mathrm{d}\mathbf{P}_{XY} = \operatorname{diag}(\mathrm{d}\mathbf{f}) \circ \mathbf{P}_{XY} \tag{181}$$

where $\mathbf{P}_X^\top \mathrm{d}\mathbf{f} = 0$, and we used $\circ$ to denote the Hadamard product (pointwise product) of matrices, which has a lower priority than the conventional matrix product.

Let $\mathbf{u}$ and $\mathbf{v}$ be the left and right singular vectors of $\mathbf{A}$ corresponding to the second largest singular value of $\mathbf{A}$, both normalized, so that $\rho_{\mathsf{m}}(X;Y) = \mathbf{u}^\top \mathbf{A}\mathbf{v}$. Then

$$\mathrm{d}\rho_{\mathsf{m}}(X;Y) = \frac{1}{2\rho_{\mathsf{m}}(X;Y)} \mathrm{d}\rho_{\mathsf{m}}^2(X;Y) \tag{182}$$

$$= \frac{1}{2\rho_{\mathsf{m}}(X;Y)} \mathbf{v}^\top \mathrm{d}\mathbf{M}\mathbf{v} \tag{183}$$

$$= \frac{1}{2\rho_{\mathsf{m}}(X;Y)} \mathbf{v}^\top (\mathrm{d}\mathbf{A}^\top \mathbf{A} + \mathbf{A}^\top \mathrm{d}\mathbf{A})\mathbf{v} \tag{184}$$

$$= \mathbf{u}^\top \mathrm{d}\mathbf{A}\mathbf{v}. \tag{185}$$

But

$$\mathrm{d}\mathbf{P}_X := \mathrm{d}\mathbf{f} \circ \mathbf{P}_X; \tag{186}$$

$$\mathrm{d}\mathbf{P}_Y := \mathrm{d}\mathbf{f}^\top \mathbf{P}_{XY}, \tag{187}$$

so

$$\mathbf{u}^\top \mathrm{d}\mathbf{A}\mathbf{v} = \frac{\mathrm{d}\mathbf{f}^\top}{2} \operatorname{diag}(\mathbf{u})\mathbf{A}\mathbf{v} + \mathrm{d}\mathbf{f}^\top \operatorname{diag}(\mathbf{u})\mathbf{A}\mathbf{v} - \frac{1}{2}\mathbf{u}^\top \mathbf{A} \operatorname{diag}(\mathrm{d}\mathbf{f}^\top \mathbf{P}_{X|Y})\mathbf{v} \tag{188}$$

$$= \frac{\rho_{\mathsf{m}}(X;Y)}{2}(\mathbf{u}^{\circ 2})^\top \mathrm{d}\mathbf{f} - \frac{\rho_{\mathsf{m}}(X;Y)}{2}(\mathbf{v}^{\circ 2})^\top (\mathbf{P}_{X|Y})^\top \mathrm{d}\mathbf{f} \tag{189}$$

where, e.g. $\mathbf{u}^{\circ 2}$ represents entry-wise square. This implies that we must have

$$\mathbf{u}^{\circ 2} - \mathbf{P}_{X|Y}\mathbf{v}^{\circ 2} = a\mathbf{P}_X \tag{190}$$

for some real number a. Summing up the entries on each side on both sides gives $a = 0$. Thus

$$\mathbf{u}^{\circ 2} = \mathbf{P}_{X|Y}\mathbf{v}^{\circ 2}. \tag{191}$$

The necessity of (64) and (65) have been shown. To show the further simplification (66)-(67) under additional assumptions, notice that

$$Q_X = Q_{X|Y} Q_Y, \tag{192}$$

which, combined with (64), shows that

$$D(\mathbf{u}^{\circ 2} \| Q_X) \le D(\mathbf{v}^{\circ 2} \| Q_Y) \tag{193}$$

where we abuse the notation by considering, e.g. $\mathbf{u}^{\circ 2}$ as a probability distribution. However, the by symmetry we also have $D(\mathbf{u}^{\circ 2} \| Q_X) \ge D(\mathbf{v}^{\circ 2} \| Q_Y)$, so (193) is actually achieved with equality. Denote by P_{XY} the joint distribution associated with $\mathbf{Q}_{X|Y}\mathbf{v}^{\circ 2}$. The necessary and sufficient condition for the data processing inequality (193) to hold with equality is that $Q_{Y|X} = P_{Y|X}$ holds P_X-almost surely. In the case of indecomposable Q_{XY} and fully supported Q_X and Q_Y, it is elementary to show that $\mathbf{v}^{\circ 2} = \mathbf{P}_Y$. The other condition follows from the same reasoning.

Appendix D
An Inequality Related to Conjecture 1 and its Numerical Validation

Let

$$\bar{\mathcal{S}}_r(X,Y) := (H(X,Y), I(X;Y)) - \mathcal{S}_r(X,Y) \tag{194}$$

be the reflection of $\mathcal{S}_r(X,Y)$ with respect to a point. The functional ω_r^s defined in (34)-(33) can then be represented as

$$\omega_r^s(Q_{XY}) := \max_{(\bar{S},\bar{R}) \in \bar{\mathcal{S}}_r(X,Y)} \{s\bar{S} - \bar{R}\}. \tag{195}$$

Then geometrically, $\omega_r^s(Q_{XY})$ is as illustrated in Figure 4. Notice that the slope of the supporting line intersecting the upper-right point of $\mathcal{S}_1^s(X,Y)$ (resp. $\mathcal{S}_\infty^s(X,Y)$) is exactly the SDPI $s_1^*(X;Y)$ (resp. the SSDPI $s_\infty^*(X;Y)$), both equal to $(1-2\epsilon)^2$ for a BSS with error probability ϵ.

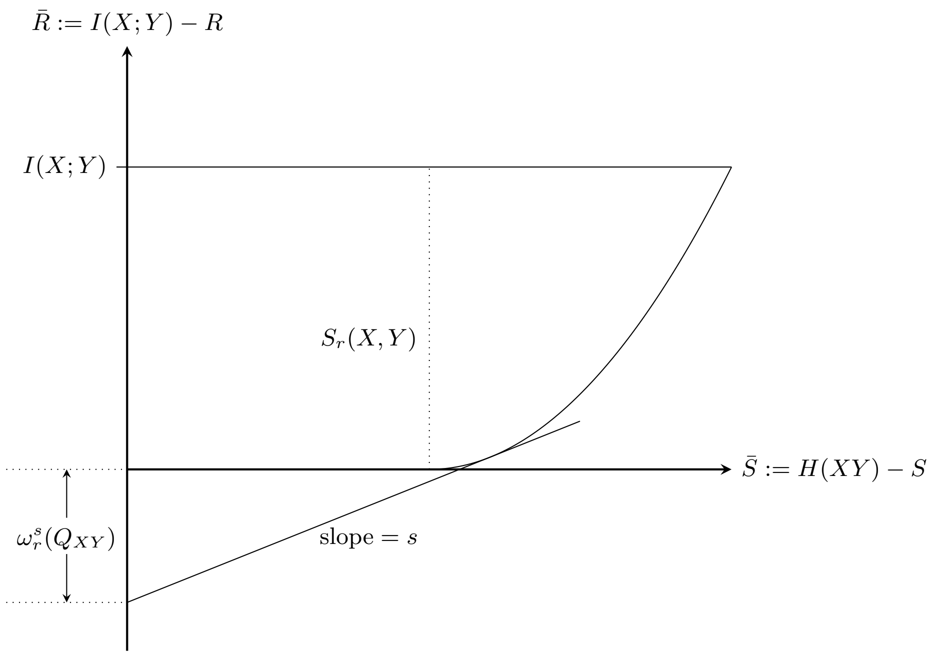

Figure 4: Geometric illustration of $\omega_r^s(Q_{XY})$.

We shall need to parameterize the lower set $\mathcal{P}(Q_{XY})$ with two parameters as in (110), via the bijection $(f,g)\mapsto P_{XY}$.

It can be easily verified that for fixed g, the transitional probability $P_{Y|X}$ is also fixed, hence f only controls the marginal P_X. Further, the function

$$\chi(f,g) := A + \frac{c}{Z}\left(f-\frac{1}{2}\right)\left(g-\frac{1}{2}\right) \tag{196}$$

is XY-*linear* (defined similarly as XY-concave with obvious changes) for any real numbers A and c. If $\alpha\in[0,\frac{1}{2}]$ is the number that maximizes $sH(X,Y|U)-I(X;Y|U)$ where U is symmetric Bernoulli satisfying $U-X-Y$, then straight forward calculations show that

$$s = \frac{(\bar{\epsilon}-\epsilon)(\log(\alpha*\epsilon)-\log(\bar{\alpha}*\epsilon))}{\log\alpha-\log\bar{\alpha}}. \tag{197}$$

Moreover, $sH(X,Y|U)-I(X;Y|U)$ is equal to ω_0^s at four points:

$$(f,g) =(\alpha,\frac{1}{2}), \tag{198}$$

$$(f,g) =(\bar{\alpha},\frac{1}{2}), \tag{199}$$

$$(f,g) =(\frac{1}{2},\alpha), \tag{200}$$

$$(f,g) = (\frac{1}{2}, \bar{\alpha}). \tag{201}$$

We can choose a unique A such that χ and ω_0^s have the same values at those four points, and a unique c such that the two functions have the same first order derivatives at those four points. It is an elementary exercise to figure out the values of such A and c. If with these values of A and c the XY-linear functional χ dominates ω_0^s, then Conjecture 1 will follow. In other words, Conjecture 1 will be implied by the following conjectured inequality:

Conjecture 2. Suppose $\alpha, \epsilon, f, g \in (0,1)$, and

$$s := \frac{(\bar{\epsilon}-\epsilon)[\log(\alpha * \epsilon) - \log(\bar{\alpha} * \epsilon)]}{\log \alpha - \log \bar{\alpha}}. \tag{202}$$

(We remind the reader the notations $\bar{\epsilon} := 1-\epsilon$ and $\alpha * \epsilon := \alpha\bar{\epsilon} + \bar{\alpha}\epsilon$.) Also define

$$c := \frac{4k\alpha\bar{\alpha}(\bar{\epsilon}-\epsilon)}{\bar{\alpha}-\alpha}\log\frac{\alpha}{\bar{\alpha}} + 4(k+1)\epsilon\bar{\epsilon}\log\frac{\epsilon}{\bar{\epsilon}} - \frac{4(\epsilon*\alpha)(\epsilon*\bar{\alpha})}{\bar{\alpha}-\alpha}\log\frac{\epsilon*\alpha}{\epsilon*\bar{\alpha}} \tag{203}$$

$$= \frac{4\epsilon\bar{\epsilon}}{\bar{\alpha}-\alpha}\log\frac{\bar{\alpha}*\epsilon}{\alpha*\epsilon} - \frac{4\epsilon\bar{\epsilon}(\bar{\epsilon}-\epsilon)\log\frac{\bar{\epsilon}}{\epsilon}\log\frac{\bar{\alpha}*\epsilon}{\alpha*\epsilon}}{\log\frac{\bar{\alpha}}{\alpha}} - 4\bar{\epsilon}\epsilon\log\frac{\bar{\epsilon}}{\epsilon}. \tag{204}$$

When $\alpha = \frac{1}{2}$ the above are defined via continuity. Then we have

$$sH(\hat{X},\hat{Y}) - I(\hat{X};\hat{Y}) \le s[h(\epsilon)+h(\alpha)] - [h(\alpha*\epsilon) - h(\epsilon)] + \frac{c(f-\frac{1}{2})(g-\frac{1}{2})}{f*\bar{g}*\epsilon}. \tag{205}$$

(remember that h is the binary entropy function, P_{XY} was defined in (110), and $(\hat{X},\hat{Y}) \sim P_{XY}$) and the equality holds at the four points (198)-(201).

Remark 7. By symmetry of the functions involved, we only have to verify for $\alpha, \epsilon, f \in (0, \frac{1}{2})$ and $g \in (0,1)$.

Remark 8. Conjecture 2 is stronger than Conjecture 1. On the other hand, it can be shown that Conjecture 1 implies the inequality in Conjecture 2 for $(f,g) \in [0,1]\times[\alpha,\bar{\alpha}]\bigcup[\alpha,\bar{\alpha}]\times[0,1]$.

Remark 9. From

$$\mathbb{E}[\hat{X}\hat{Y}] = \frac{\bar{\epsilon}(\bar{f}\bar{g}+fg) - \epsilon(\bar{f}g + f\bar{g})}{Z} \tag{206}$$

we obtain

$$\frac{c(f-\frac{1}{2})(g-\frac{1}{2})}{f*\bar{g}*\epsilon} = \frac{c}{4}\left[\frac{1}{2\bar{\epsilon}} - \frac{1}{2\epsilon} + \left(\frac{1}{2\bar{\epsilon}} + \frac{1}{2\epsilon}\right)\mathbb{E}[\hat{X}\hat{Y}]\right] \tag{207}$$

Therefore the conjecture inequality is equivalent to

$$(s+1)H(\hat{X},\hat{Y}) \le H(\hat{X}) + H(\hat{Y}) + s[h(\epsilon)+h(\alpha)] - [h(\alpha*\epsilon)-h(\epsilon)] + \frac{c}{8\epsilon\bar{\epsilon}}[\epsilon - \bar{\epsilon} + \mathbb{E}[\hat{X}\hat{Y}]] \tag{208}$$

Although Conjecture 2 seems elementary, we have not been able to find a full proof. Nevertheless, since it only involves four parameters we can parameterize the space $(0,1)^4$ and verify numerically. We computed the difference between the right hand side of (205) and the left hand side. From the choice of A we know that the difference is exactly zero at the four points (198)-(201). Using Matlab we computed difference between the right hand side of

(205) and the left hand side for f, g, ϵ, α ranging from vectors

$$F = [ss/3 : ss : 0.5 - ss/3]'; \tag{209}$$

$$G = [ss/3 : ss : 1 - ss/3]'; \tag{210}$$

$$E = [ss/3 : ss : 0.5 - ss/3]'; \tag{211}$$

$$A = E; \tag{212}$$

where the step size $ss := 0.001$. As the result the minimum value of the difference is `-5.841478017444557e-17` with double precision, which is quite small. Moreover negativity of the difference occurs only when $0.4963333333333333 \le \epsilon < 0.5$ and $0.4993333333333333 \le \alpha < 0.5$. If we make ϵ and α closer to 0.5, then the magnitude of the difference can further increase, up to about 10^{-9} at most; however in this case the image of the left hand side becomes noise-like of the magnitude about 10^{-9} as well, so the error is most likely due to the limit of the double precision. In fact, when we use variable precision arithmetic (vpa), the images become smooth and good looking again, and the minimum difference becomes zero.

To visualize what is happening in Conjecture 2, we plotted ω_0^s, χ and their difference in Fig. 5-7 for a particular instance of ϵ and α (the value of k is then uniquely determined).

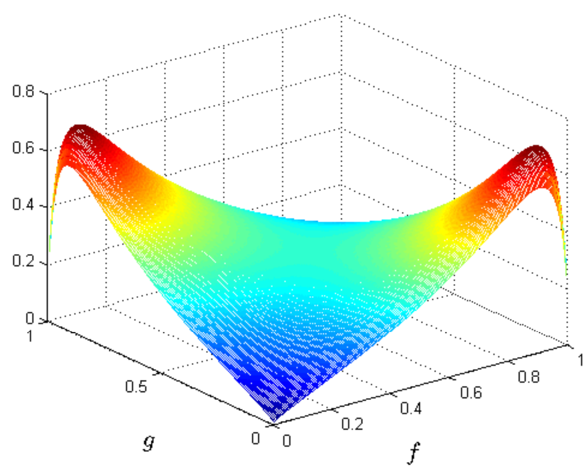

Figure 5: Plot of ω_0^s against f and g when $\alpha = \epsilon = 0.11$

Appendix E
Proof of Conjecture 2 for $\epsilon \to \frac{1}{2}$

We fix f, g and α and let $\epsilon \to \frac{1}{2}$. Let u be such that

$$\epsilon = \frac{1-u}{2}. \tag{213}$$

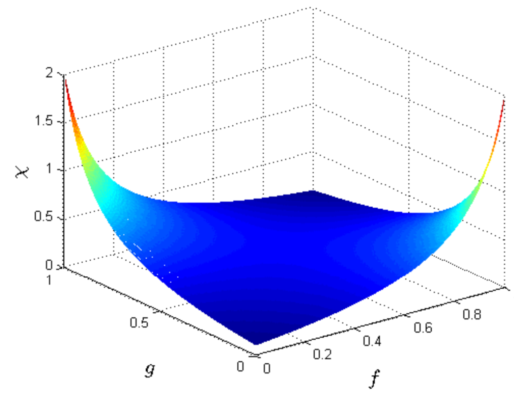

Figure 6: Plot of χ against f and g when $\alpha = \epsilon = 0.11$

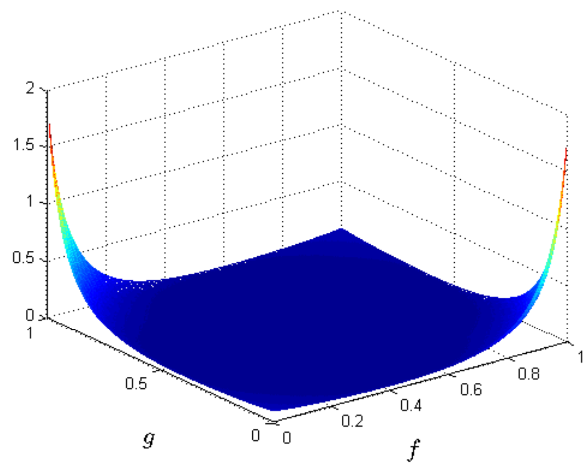

Figure 7: Plot of $\chi - \omega_0^s$ against f and g when $\alpha = \epsilon = 0.11$

Introduce the notation

$$\epsilon(x, y) := \begin{cases} 1 - \epsilon & x = y; \\ \epsilon & \text{otherwise.} \end{cases} \tag{214}$$

Then,

$$\frac{\mathrm{d}}{\mathrm{d}\epsilon} \log \epsilon(x, y) = \frac{\log e}{\epsilon(x, y)} (-1)^{x-y+1}, \tag{215}$$

$$\frac{\mathrm{d}^2}{\mathrm{d}\epsilon^2} \log \epsilon(x, y) = -\frac{\log e}{\epsilon^2(x, y)}, \tag{216}$$

$$\frac{\mathrm{d}}{\mathrm{d}\epsilon}\log Z = \frac{\log e}{f*\bar{g}*\epsilon}(f*g - f*\bar{g}) \tag{217}$$

$$= -\frac{\log e}{f*\bar{g}*\epsilon}(\bar{f}-f)(\bar{g}-g), \tag{218}$$

$$\frac{\mathrm{d}^2}{\mathrm{d}\epsilon^2}\log Z = -\frac{\log e}{(f*\bar{g}*\epsilon)}(\bar{f}-f)^2(\bar{g}-g)^2 \tag{219}$$

$$\frac{\mathrm{d}}{\mathrm{d}\epsilon}\log\sum_{x'}\epsilon(x',y)f(x') = \frac{\log e}{\sum_{x'}\epsilon(x',y)f(x')}\sum_{x'}(-1)^{x'-y+1}f(x') \tag{220}$$

$$\frac{\mathrm{d}^2}{\mathrm{d}\epsilon^2}\log\sum_{x'}\epsilon(x',y)f(x') = -\frac{\log e(\sum_{x'}(-1)^{x'-y+1}f(x'))^2}{(\sum_{x'}\epsilon(x',y)f(x'))^2} \tag{221}$$

In particular,

$$\frac{\mathrm{d}}{\mathrm{d}\epsilon}\log\epsilon(x,y)\bigg|_{\epsilon=\frac{1}{2}} = 2\log e(-1)^{x-y+1}, \tag{222}$$

$$\frac{\mathrm{d}^2}{\mathrm{d}\epsilon^2}\log\epsilon(x,y)\bigg|_{\epsilon=\frac{1}{2}} = -4\log e, \tag{223}$$

$$\frac{\mathrm{d}}{\mathrm{d}\epsilon}\log Z\bigg|_{\epsilon=\frac{1}{2}} = -2\log e(\bar{f}-f)(\bar{g}-g), \tag{224}$$

$$\frac{\mathrm{d}^2}{\mathrm{d}\epsilon^2}\log Z\bigg|_{\epsilon=\frac{1}{2}} = -4\log e(\bar{f}-f)^2(\bar{g}-g)^2, \tag{225}$$

$$\frac{\mathrm{d}}{\mathrm{d}\epsilon}\log\sum_{x'}\epsilon(x',y)f(x')\bigg|_{\epsilon=\frac{1}{2}} = 2\log e\sum_{x'}(-1)^{x'-y+1}f(x'), \tag{226}$$

$$\frac{\mathrm{d}^2}{\mathrm{d}\epsilon^2}\log\sum_{x'}\epsilon(x',y)f(x')\bigg|_{\epsilon=\frac{1}{2}} = -4\log e(\sum_{x'}(-1)^{x'-y+1}f(x'))^2. \tag{227}$$

When $\epsilon \to \frac{1}{2}$, we show that both sides of the inequality is of the order of u^2. It's easy to compute

$$s = \frac{2\log e(\bar{\alpha}-\alpha)u^2}{\log\bar{\alpha}-\log\alpha} + o(u^2), \tag{228}$$

$$c = o(u^2), \tag{229}$$

$$I(\hat{X},\hat{Y}) = \sum_{x,y}\frac{\epsilon(x,y)f(x)g(x)}{Z}\log\frac{\epsilon(x,y)Z}{\sum_{x'}\epsilon(x',y)f(x')\sum_{y'}\epsilon(x,y')g(y')} \tag{230}$$

where the summations are over $(x,y)\in\{0,1\}^2$, $x'\in\{0,1\}$ and $y'\in\{0,1\}$, respectively. Define

$$T(\epsilon) := I(\hat{X};\hat{Y})Z. \tag{231}$$

Since

$$I(\hat{X};\hat{Y})\bigg|_{\epsilon=\frac{1}{2}} = 0, \tag{232}$$

and from the minimality of $I(\hat{X};\hat{Y})$ at $\epsilon = \frac{1}{2}$,

$$\frac{\mathrm{d}}{\mathrm{d}\epsilon} I(\hat{X};\hat{Y})\Big|_{\epsilon=\frac{1}{2}} = 0, \tag{233}$$

we have from Leibniz's rule

$$\frac{\mathrm{d}^2}{\mathrm{d}\epsilon^2} T\Big|_{\epsilon=\frac{1}{2}} = \frac{\mathrm{d}^2}{\mathrm{d}\epsilon^2} I(\hat{X};\hat{Y})\Big|_{\epsilon=\frac{1}{2}} Z\Big|_{\epsilon=\frac{1}{2}} + 2\,\frac{\mathrm{d}}{\mathrm{d}\epsilon} I(\hat{X};\hat{Y})\Big|_{\epsilon=\frac{1}{2}} \frac{\mathrm{d}}{\mathrm{d}\epsilon} Z\Big|_{\epsilon=\frac{1}{2}} + I(\hat{X};\hat{Y})\Big|_{\epsilon=\frac{1}{2}} \frac{\mathrm{d}^2}{\mathrm{d}\epsilon^2} Z\Big|_{\epsilon=\frac{1}{2}} \tag{234}$$

$$= \frac{1}{2}\,\frac{\mathrm{d}^2}{\mathrm{d}\epsilon^2} I(\hat{X};\hat{Y})\Big|_{\epsilon=\frac{1}{2}}. \tag{235}$$

Thus we obtain

$$\frac{\mathrm{d}^2}{\mathrm{d}\epsilon^2} I(\hat{X};\hat{Y})\Big|_{\epsilon=\frac{1}{2}} = 2\,\frac{\mathrm{d}^2}{\mathrm{d}\epsilon^2} T\Big|_{\epsilon=\frac{1}{2}} \tag{236}$$

which is useful because the differential on the right hand side above is easier to compute than the left hand side. From (230) and (231),

$$\frac{\mathrm{d}^2}{\mathrm{d}\epsilon^2} T\Big|_{\epsilon=\frac{1}{2}} = \sum_{x,y} \frac{\mathrm{d}^2}{\mathrm{d}\epsilon^2}[\epsilon(x,y)f(x)g(x)]\Big|_{\epsilon=\frac{1}{2}} \log \frac{\epsilon(x,y)Z}{\sum_{x'}\epsilon(x',y)f(x')\sum_{y'}\epsilon(x,y')g(y')}\Bigg|_{\epsilon=\frac{1}{2}} \tag{237}$$

$$+ 2\sum_{x,y} \frac{\mathrm{d}}{\mathrm{d}\epsilon}[\epsilon(x,y)f(x)g(x)]\Big|_{\epsilon=\frac{1}{2}} \frac{\mathrm{d}}{\mathrm{d}\epsilon} \log \frac{\epsilon(x,y)Z}{\sum_{x'}\epsilon(x',y)f(x')\sum_{y'}\epsilon(x,y')g(y')}\Bigg|_{\epsilon=\frac{1}{2}} \tag{238}$$

$$+ \sum_{x,y} \epsilon(x,y)f(x)g(x)\Big|_{\epsilon=\frac{1}{2}} \frac{\mathrm{d}^2}{\mathrm{d}\epsilon^2} \log \frac{\epsilon(x,y)Z}{\sum_{x'}\epsilon(x',y)f(x')\sum_{y'}\epsilon(x,y')g(y')}\Bigg|_{\epsilon=\frac{1}{2}} \tag{239}$$

From (222)-(227), we see that the first term is zero. The second term is equal to

$$2\sum_{x,y} f(x)g(y)(-1)^{x-y+1} \cdot 2\log e \left((-1)^{x-y+1} - (\bar{f}-f)(\bar{g}-g) + \sum_{x'}(-1)^{x'-y} f(x') + \sum_{y'}(-1)^{x-y'} g(y') \right) \tag{240}$$

$$= 4\log e\left(1 + (\bar{f}-f)^2(\bar{g}-g)^2 - (\bar{f}-f)^2 - (\bar{g}-g)^2\right) \tag{241}$$

The third term in (239) can be simplified as

$$\sum_{x,y} \frac{f(x)g(y)}{2} \cdot 4\log e \left[-1 - (\bar{f}-f)^2(\bar{g}-g)^2 + (\sum_{x'}(-1)^{x'-y+1}f(x'))^2 + (\sum_{y'}(-1)^{x-y'+1}g(y'))^2 \right] \tag{242}$$

$$= 2\log e[-1 - (\bar{f}-f)^2(\bar{g}-g)^2 + (\bar{f}-f)^2 + (\bar{g}-g)^2] \tag{243}$$

Hence

$$\frac{\mathrm{d}^2}{\mathrm{d}\epsilon^2} T\Big|_{\epsilon=\frac{1}{2}} = 4\log e[1 - (\bar{f}-f)^2][1 - (\bar{g}-g)^2] \tag{244}$$

Thus we find the left hand side of (205) is

$$\frac{2\log e(\bar{\alpha}-\alpha)u^2}{\log\bar{\alpha} - \log\alpha}[h(f) + h(g)] - \frac{u^2}{2}\log e[1 - (\bar{f}-f)^2][1 - (\bar{g}-g)^2] + o(u^2). \tag{245}$$

The right hand side of (205) is

$$\frac{2\log e(\bar{\alpha}-\alpha)u^2}{\log\bar{\alpha}-\log\alpha}(1+h(\alpha))-\frac{\log e}{2}[1-(\bar{\alpha}-\alpha)^2]u^2. \tag{246}$$

Thus the following inequality implies the validity of (205) for fixed $\alpha, f, g \in (0, \frac{1}{2})$ and vanishing $\epsilon - \frac{1}{2}$:

$$\frac{2(\bar{\alpha}-\alpha)}{\log\bar{\alpha}-\log\alpha}[h(f)+h(g)]-8f\bar{f}g\bar{g} \le \frac{2(\bar{\alpha}-\alpha)}{\log\bar{\alpha}-\log\alpha}(1+h(\alpha))-2\alpha\bar{\alpha} \tag{247}$$

Note that now we only have to verify the inequality for $g \in (0, \frac{1}{2})$, in contrast to Remark 7. Consider fixed α. The values of f and g that maximizes the left hand side of (247) must be the solution of the following optimization problem:

$$\text{minimize } \eta(f,g) := 16f\bar{f}g\bar{g} \quad \text{subject to } \phi(f,g) := h(f)+h(g) = C \tag{248}$$

for some constant C. We solve this minimization problem using Lagrange multiplier method. Define

$$L(f,g) := \eta(f,g) - \lambda\phi(f,g). \tag{249}$$

Suppose (f^*, g^*) is a local minimum, then for some value of α, we have

$$\begin{cases} \partial_1 L(f^*,g^*) = 0, \\ \partial_2 L(f^*,g^*) = 0. \end{cases} \tag{250}$$

for some $\lambda = \lambda^* \neq 0$, which implies that

$$\begin{cases} \partial_1\eta(f^*,g^*) = \lambda^*\partial_1\phi(f^*,g^*), \\ \partial_2\eta(f^*,g^*) = \lambda^*\partial_2\phi(f^*,g^*). \end{cases} \tag{251}$$

1) If $f^* \neq \frac{1}{2}$ and $g^* \neq \frac{1}{2}$, we can cancel λ^* from (251) and obtain after rearrangement

$$\frac{\log\bar{f}^*-\log f^*}{\bar{f}^*-f^*}\bar{f}^*f^* = \frac{\log\bar{g}^*-\log g^*}{\bar{g}^*-g^*}\bar{g}^*g^*. \tag{252}$$

It is elementary to check (e.g. by writing it as Taylor series in terms of $1-2x$ that the function

$$T(x) := \frac{\log\bar{x}-\log x}{\bar{x}-x}\bar{x}x \tag{253}$$

is monotonically increasing on $(0, \frac{1}{2})$. Thus (252) implies that

$$g^* = f^*. \tag{254}$$

Recall that (f^*, f^*) being a local minimum point implies that the Hessian matrix $[\partial^2_{i,j}L(f^*, f^*)]$ is positive-semidefinite on the orthogonal complement of the span of $\nabla\phi(f^*, f^*)$. In our case, this means that the matrix

$$\begin{pmatrix} \frac{\lambda^*\log e}{f^*\bar{f}^*} - 32f^*\bar{f}^* & 16(\bar{f}^*-f^*)^2 \\ 16(\bar{f}^*-f^*)^2 & \frac{\lambda^*\log e}{f^*\bar{f}^*} - 32f^*\bar{f}^* \end{pmatrix} \tag{255}$$

is positive-semidefinite on the span of $(1, -1)^\top$, or equivalently,

$$\frac{\lambda^*\log e}{f^*\bar{f}^*} - 32f^*\bar{f}^* \ge 16(\bar{f}^*-f^*)^2 \tag{256}$$

Substituting (254) into (251), we obtain

$$\lambda^* = \frac{16(\bar{f}^* - f^*)f^*\bar{f}^*}{\log \bar{f}^* - \log f^*}, \tag{257}$$

hence (256) is equivalent to

$$\frac{\log e(\bar{f}^* - f^*)}{\log \bar{f}^* - \log f^*} - 2f^*\bar{f}^* \geq (\bar{f}^* - f^*)^2 \tag{258}$$

However for any $u^* := 1 - 2f^* \neq 0$, we show that (258) fails:

$$\begin{aligned}
\text{LHS of (258)} =& \frac{u^*}{\ln(1+u^*) - \ln(1-u^*)} - 2f^*\bar{f}^* && (259)\\
=& \frac{u^*}{\sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k} u^{*k} - \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k} u^{*k}} - 2f^*\bar{f}^* && (260)\\
=& \frac{u^*}{\sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k} u^{*k} + \sum_{k=1}^{\infty} \frac{1}{k} u^{*k}} - 2f^*\bar{f}^* && (261)\\
=& \frac{1}{\sum_{l \in 2\mathcal{N}} \frac{2}{l+1} u^{*l}} - 2f^*\bar{f}^* && (262)\\
<& \frac{1}{2} - 2f^*\bar{f}^* && (263)\\
=& \frac{(\bar{f}^* - f^*)^2}{2} && (264)\\
<& \text{RHS of (258)}. && (265)
\end{aligned}$$

Therefore, the solution to (248) must belong to the following case:

2) If either $f^* = \frac{1}{2}$ or $g^* = \frac{1}{2}$, by the symmetry of (247) we may assume without loss of generality that $g^* = \frac{1}{2}$. The left hand side of (247) becomes

$$\frac{2(\bar{\alpha} - \alpha)}{\log \bar{\alpha} - \log \alpha}(1 + h(f)) - 2f\bar{f}. \tag{266}$$

When viewed as a function of f, it is maximized by $f = \alpha$ using Calculus, in which case it agrees with the right hand side of (247). Thus (247) is proved.

References

[1] Ahlswede, R. and Csiszár, I. (Apr. 1993). Common randomness in information theory and cryptography. I. Secret sharing. *IEEE Transactions on Information Theory*, 39(4):1121–1132.

[2] Ahlswede, R. and Csiszár, I. (Jan. 1998). Common randomness in information theory and cryptography. Part II. CR capacity. *IEEE Transactions on Information Theory*, 44(1):225–240.

[3] Ahlswede, R. and Gacs, P. (1976). Spreading of sets in product spaces and hypercontraction of the Markov operator. *The Annals of Probability*, pages 925–939.

[4] Anantharam, V., Gohari, A., Kamath, S., and Nair, C. (2013). On Maximal Correlation, Hypercontractivity, and the Data Processing Inequality studied by Erkip and Cover. *arXiv preprint arXiv:1304.6133*.

[5] Anantharam, V., Gohari, A. A., Kamath, S., and Nair, C. On hypercontractivity and the mutual information between Boolean functions. *The 51st Annual Allerton Conference on Communication, Control, and Computing*, pages 13–19.

[6] Boothby, W. M. (2003). *An introduction to differentiable manifolds and Riemannian geometry*, volume 120. Gulf Professional Publishing.

[7] Braverman, M. (2012a). Coding for interactive computation: Progress and challenges. In *Allerton Conference*, pages 1914–1921.

[8] Braverman, M. (2012b). Interactive information complexity. In *Proceedings of the forty-fourth annual ACM symposium on Theory of computing*, pages 505–524. ACM.

[9] Braverman, M., Oshman, R., and Weinstein, O. (September 2015). Information and communication complexity. *IEEE Information Theory Society Newsletter*, 65(3).

[10] Braverman, M. and Rao, A. (2014). Information equals amortized communication. *IEEE Transactions on Information Theory*, 60(10):6058–6069.

[11] Carlen, E. A. and Cordero-Erausquin, D. (2009). Subadditivity of the entropy and its relation to Brascamp–Lieb type inequalities. *Geometric and Functional Analysis*, 19(2):373–405.

[12] Choi, M.-D., Ruskai, M. B., and Seneta, E. (1994). Equivalence of certain entropy contraction coefficients. *Linear algebra and its applications*, 208:29–36.

[13] Chou, R. A. and Bloch, M. R. (Aug. 2014). Separation of Reliability and Secrecy in Rate-Limited Secret Key Generation. *IEEE Transactions on Information Theory*, 60(8):4941–4957.

[14] Courtade, T. A. (Istanbul, Turkey, July 2013). Outer bounds for multiterminal source coding via a strong data processing inequality. In *Proceedins of 2013 IEEE International Symposium on Information Theory (ISIT)*, pages 559–563.

[15] Csiszár, I. (1975). I-divergence geometry of probability distributions and minimization problems. *The Annals of Probability*, pages 146–158.

[16] Csiszar, I. and Körner, J. (2011). *Information theory: coding theorems for discrete memoryless systems*. Cambridge University Press.

[17] Csiszár, I. and Narayan, P. (2004). Secrecy capacities for multiple terminals. *IEEE Transactions on Information Theory*, 50(12):3047–3061.

[18] Csiszár, I. and Narayan, P. (Feb. 2000). Common randomness and secret key generation with a helper. *IEEE Transactions on Information Theory*, 46(2):344–366.

[19] El Gamal, A. and Kim, Y.-H. (2011). *Network Information Theory*. Cambridge University Press.

[20] Gács, P. and Körner, J. (1973). Common information is far less than mutual information. *Problems of Control and Information Theory*, 2(2):149–162.

[21] Ganor, A., Kol, G., and Raz, R. (2014). Exponential separation of information and communication. In *2014 IEEE 55th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 176–185. IEEE.

[22] Gebelein, H. (1941). Das statistische problem der korrelation als variations- und eigenwert-problem und sein

zusammenhang mit der ausgleichungsrechnung. *Zeitschrift fur angew. Math. und Mech.*, 21:364–379.

[23] Geng, Y. and Nair, C. (2014). The capacity region of the two-receiver Gaussian vector broadcast channel with private and common messages. *IEEE Transactions on Information Theory*, 60(4):2087–2104.

[24] Gross, L. (1975). Logarithmic Sobolev Inequalities. *American Journal of Mathematics*, 97(4):1061–1083.

[25] Han, T. S. (2003). *Te Sun Han, Information-Spectrum Method in Information Theory*. Springer.

[26] Hirschfeld, H. O. A connection between correlation and contingency, journal=Proc. Cambridge Philosophical Soc., volume=31, pages=520–524, year=1935.

[27] Kamath, S. Reverse hypercontractivity using information measures. In *Proceedings of the 53rd Annual Allerton Conference on Communications, Control and Computing 2015, Monticello, Illinois*.

[28] Kamath, S. and Anantharam, V. (2012). Non-interactive simulation of joint distributions: The Hirschfeld-Gebelein-Rényi maximal correlation and the hypercontractivity ribbon. In *50th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 1057–1064. IEEE.

[29] Kaspi, A. H. (1985). Two-way source coding with a fidelity criterion. *IEEE Transactions on Information Theory*, 31(6):735–740.

[30] Lancaster, H. O. (1957). Some properties of the bivariate normal distribution considered in the form of a contingency table. *Biometrika*, 44(1–2):289–292.

[31] Liu, J., Cuff, P., and Verdú, S. Key Capacity for Product Sources with Application to Stationary Gaussian Processes. *arXiv:1409.5844v1 [cs.IT]*.

[32] Liu, J., Cuff, P., and Verdú, S. (2015b). Secret Key Generation with One Communicator and a One-Shot Converse via Hypercontractivity. In *Proceedings of 2015 IEEE International Symposium on Information Theory (ISIT)*, pages 710–714.

[33] Liu, J., Cuff, P., and Verdú, S. (Novemeber 2015a). Rate region for interactive key generation and common randomness generation. *private note*.

[34] Ma, N., Ishwar, P., and Gupta, P. (2012). Interactive source coding for function computation in collocated networks. *IEEE Transactions on Information Theory*, 58(7):4289–4305.

[35] Maurer, U. M. (Mar. 1993). Secret key agreement by public discussion from common information. *IEEE Transactions on Information Theory*, 39(3):733–742.

[36] Nair, C. (2014). Equivalent Formulations of Hypercontractivity Using Information Measures. *IZS Workshop, 2014*.

[37] Polyanskiy, Y. (Cambridge, MA, July 2012). Hypothesis testing via a comparator. In *Proceedings of 2012 IEEE International Symposium on Information Theory Proceedings (ISIT)*, pages 2206–2210. IEEE.

[38] Polyanskiy, Y. and Wu, Y. A note on the strong data-processing inequalities in Bayesian networks. *http://arxiv.org/pdf/1508.06025v1.pdf*.

[39] Renyi, A. (1959). On measures of dependence. *Acta Math. Hung.*, 10:441–451.

[40] Rockafellar, R. T. (1970). *Convex analysis*. Princeton university press.

[41] Shannon, C. E. (1949b). Communication Theory of Secrecy Systems. *Bell system technical journal*, 28(4):656–

715.

[42] Shannon, C. E. (Jan. 1949a). Communication in the presence of noise. *Proc. IRE*, 37:10–21.

[43] Song, E. C., Cuff, P., and Poor, H. V. The likelihood encoder for lossy compression. *arXiv:1408.4522v2*.

[44] Tan, V. Y. F. (2014). Asymptotic Estimates in Information Theory with Non-Vanishing Error Probabilities. *Foundations and Trends in Communications and Information Theory*, 11(1-2):1–184.

[45] Tyagi, H. (2013). Common information and secret key capacity. *IEEE Transactions on Information Theory*, 59(9):5627–5640.

[46] Verdú, S. (May 1990). On channel capacity per unit cost. *IEEE Transactions on Information Theory*, 36(5):1019–1030.

[47] Witsenhausen, H. S. On Sequences of Pairs of Dependent Random Variables. *SIAM Journal on Applied Mathematics*, 28(1):100–113.

[48] Wyner, A. D. (Feb. 1975). The common information of two dependent random variables. *IEEE Transactions on Information Theory*, 21(2):163–179.

[49] Yao, A. C.-C. (1979). Some complexity questions related to distributive computing. In *STOC*, pages 209–213.