Capacity Upper Bounds for the Relay Channel via Reverse Hypercontractivity

Jingbo Liu, *Member, IEEE*, and Ayfer Özgür, *Senior Member, IEEE*

***Abstract*— We revisit the primitive relay channel, introduced by Cover in 1987. Recent work derived upper bounds on the capacity of this channel that are tighter than the classical cutset bound using the concentration of measure. In this paper, we recover, generalize, and improve upon some of these upper bounds with simpler proofs using reverse hypercontractivity. To our knowledge, this is the first application of reverse hypercontractivity in proving first-order converses in network information theory.**

***Index Terms*— Shannon theory, relay channel, reverse hypercontractivity, Markov semigroups, converses, concentration of measure.**

I. INTRODUCTION

THE primitive relay channel, introduced by Cover in 1987 [1], models the communication scenario where a source-destination pair is assisted by a single relay which is connected to the destination with an independent channel of some finite capacity.[1] See Figure 1. The primitive relay channel can be regarded as the simplest network model that intertwines channel coding with source coding. As noted by Kim [2], "on the one hand, it is the simplest channel coding problem (from the source transmitter's point of view) with a source coding constraint; on the other hand, it is the simplest source coding problem (from the relay's point of view) for a channel code". As such, even-though its capacity remains unknown, it has served as a good testbed for developing new relay coding schemes as well as new converse techniques over the last three decades [2]–[13].

The classical upper bound on the capacity of this channel is the so-called cutset bound developed by Cover and El Gamal in 1979 [14]. This result bounds the capacity of the channel by its minimal cut capacity, in a flavor similar to the famous max-flow min-cut theorem for graphical networks [15]. Recently, Wu, Ozgur and Xie in [8] and Wu and Ozgur in [9] proved new upper bounds on the capacity of the primitive relay channel. In particular, [9] considered the canonical Gaussian case and developed the first upper bounds for Gaussian relay networks that are tighter than the cutset bound. The paper [8] focused on discrete primitive relay channels and developed upper bounds that improved on the earlier bounds for discrete channels by Xue [7] and Zhang [3]. The bounds in [8] and [9], as well as the earlier bounds in [3], [7] that improve on the cutset bound for this channel model, are built upon distinct and non-trivial uses of concentration of measure in discrete or Gaussian spaces [16], [17]. For general introduction to concentration of measure, sometimes also known as the blowing-up lemma, see [18] or [19].

In this paper, we reprove and generalize these bounds using a new converse technique introduced by Liu, van Handel and Verdú in [20] (extended version [21]) relying on the reverse hypercontractivity of Markov semigroups. In a unified way (by using different Markov semigroups adapted to various channel models), we prove capacity upper bounds for Gaussian channels and channels with bounded density (including all discrete memoryless channels) which end up being slightly sharper than the corresponding bounds in [9] and [8]. We remark that reverse hypercontractivity has gained some recent interest within the information theory community. For example [22]–[25] studied the equivalent formulations of the reverse hypercontractivity and [26] computed the reverse hypercontractivity region for the erasure channel. The line of work [20], [23], [27]–[29] integrated reverse hypercontractivity with the functional representations of information measures to prove second-order (a.k.a. strong) converses for multiuser information theoretic problems. However, to the best of our knowledge, the current paper provides the first application of reverse hypercontractivity to proving *first-order* converses for multi-user problems.

The simplicity of the new proof via reverse hypercontractivity mainly comes from the saving in the extra tensorization argument. Measure concentration based proofs for the relay channel often use a so-called "lifting" (a.k.a. tensorization) argument to invoke the blowing up lemma, which entails an often complicated construction of "typical" sets. With reverse hypercontractivity this step is eliminated since reverse hypercontractivity itself tensorizes. We note that reverse hypercontractivity of a reversible semigroup is known to be equivalent to the modified log-Sobolev inequality [30],

Manuscript received November 27, 2018; revised May 24, 2020; accepted June 8, 2020. Date of publication July 13, 2020; date of current version August 18, 2020. This work was supported in part by the National Science Foundation (NSF) under Grant CCF-1350595, Grant CCF-1016625, Grant CCF-0939370, Grant CCF-1514538, and Grant DMS-1148711; in part by the Army Research Office (ARO) under Grant W911NF-15-1-0479 and Grant W911NF-14-1-0094; and in part by the Center for Science of Information (CSoI), an NSF Science and Technology Center, under Agreement CCF-0939370. This article was presented in part at the 2019 IEEE International Symposium on Information Theory (ISIT). *(Corresponding author: Jingbo Liu.)*

Jingbo Liu is with the Institute for Data, Systems, and Society, Massachusetts Institute of Technology, Cambridge, MA 02142 USA, and also with the Department of Statistics, University of Illinois at Urbana-Champaign, Champaign, IL 61820 USA (e-mail: jingbol@illinois.edu).

Ayfer Özgür is with the Department of Electrical Engineering, Stanford University, Stanford, CA 94305 USA (e-mail: aozgur@stanford.edu).

Communicated by M. Raginsky, Associate Editor for Probability and Statistics.

Digital Object Identifier 10.1109/TIT.2020.3008599

[1]The term "primitive" first appears in [2].

which in turn implies sub-Gaussian concentration in rather general settings via the Herbst argument (see e.g. the discrete [31, Lemma 3.16] and Gaussian [31, Theorem 3.25]) cases. However, in our proofs reverse hypercontractivity is directly integrated to the information theoretic converse without going through concentration of measure.

We also note that in more recent work [11], Wu, Barnes and Ozgur develop a tighter upper bound on the capacity of the Gaussian primitive relay channel which significantly improves on [9], and in particular, is strong enough to resolve an open problem posed by Cover in [1] regarding the capacity of the primitive relay channel. (An extension of their proof to the binary symmetric primitive relay channel is given in [32].) Those bounds build on a rearrangement inequality on the sphere. Note that Cover's problem concerns the regime of high relay rates. The bounds in the present paper are not sufficiently tight in that regime to resolve Cover's problem, and hence are strictly looser than the bound in [11] at least in that regime. It would be interesting to see if these stronger results can be recovered with simpler proofs based on the reverse hypercontractivity approach we develop in the current paper; see Section VI for more discussion.

The paper is organized as follows. Section II reviews the precise formulation of the primary relay channel problem. In Section III we state the reverse hypercontractivity results for the Ornstein-Uhlenbeck and the semi-simple semigroup we use in this paper. The main results are presented in Section IV and proved in Section V.

II. Problem Formulation

Consider a primitive memoryless relay channel, $P_{YZ|X}$, as depicted in Fig. 1. The source's input is $X \in \mathcal{X}$, the channel output at the relay is $Z \in \Omega$, and the channel output at the destination is $Y \in \Omega$. Let us assume that $P_{Y|X} = P_{Z|X}$ and $P_{Y,Z|X} = P_{Y|X}P_{Z|X}$. The symmetry condition $P_{Y|X} = P_{Z|X}$ is imposed for notional simplicity and for ease of comparison with existing results in the literature, but our method can be easily extended to asymmetric cases (see Remark 3 ahead). The channel is memoryless meaning that $P_{Y^n,Z^n|X^n} = \Pi_{i=1}^n P_{Y_i,Z_i|X_i}$. The relay Z can communicate to the destination Y via an error-free digital link of rate C_0 nats/ channel use.

For this channel, a code of rate R and blocklength n, denoted by

$$(\mathcal{C}_{(n,R)}, f_n(z^n), g_n(y^n, f_n(z^n))), \text{ or simply, } (\mathcal{C}_{(n,R)}, f_n, g_n),$$

consists of the following:

1) A codebook at the source X: $\mathcal{C}_{(n,R)} = \{x^n(m) \in \mathcal{X}^n, m \in \{1, 2, \ldots, \lceil e^{nR} \rceil\}\}$;
2) An encoding function at the relay Z:
$$f_n \colon \Omega^n \to \{1, 2, \ldots, \lceil e^{nC_0} \rceil\}; \quad (1)$$
3) A decoding function at the destination Y: $g_n : \Omega^n \times \{1, 2, \ldots, \lceil e^{nC_0} \rceil\} \to \{1, 2, \ldots, \lceil e^{nR} \rceil\}$.

The average probability of error of the code is defined as

$$P_e^{(n)} = \Pr(g_n(Y^n, f_n(Z^n)) \neq M), \quad (2)$$

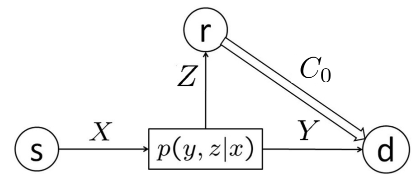

Fig. 1. Primitive relay channel.

where the message M is assumed to be uniformly drawn from the message set $\{1, 2, \ldots, \lceil e^{nR} \rceil\}$. A rate R is said to be achievable if there exists a sequence of codes

$$\{(\mathcal{C}_{(n,R)}, f_n, g_n)\}_{n=1}^{\infty}$$

such that the average probability of error $P_e^{(n)} \to 0$ as $n \to \infty$. The capacity of the primitive relay channel is the supremum of all achievable rates, denoted by $C(C_0)$.

The following proposition summarizes an intermediate step in the derivation of the cutset bound [14]. It follows immediately from (3) and (4) that $C(C_0) \le \sup_{P_X} I(X; YZ)$ and $C(C_0) \le \sup_{P_X} I(X; Y) + C_0$, which are the cut set bounds corresponding to two cut sets of the network.

Proposition 1: Consider a symmetric primitive memoryless relay channel where $P_{Y|X} = P_{Z|X}$. Suppose that there exist encoding and decoding schemes with error probability $\epsilon := P_e^{(n)}$ and blocklength n (see definition in (2)). Let X^n denote the random codeword transmitted by the source, Z^n and Y^n the relay's and the destination's observations over the n channel uses and $I = f_n(Z^n)$ the index transmitted by the relay. Let Q be equiprobable on $\{1, \ldots, n\}$ and independent of (X^n, Y^n, Z^n). Then

$$R \le I(X_Q; Y_Q, Z_Q) + \mu(\epsilon) \quad (3)$$

$$R \le I(X_Q; Y_Q) + \frac{1}{n}H(I|Y^n) - \frac{1}{n}H(I|X^n) + \mu(\epsilon) \quad (4)$$

where $\mu(\epsilon) \to 0$ as $\epsilon \to 0$.

For completeness we include the short proof here:

Proof: Using the Fano inequality and the chain rules, we have

$$R \le \frac{1}{n}I(X^n; I, Y^n) + \mu(\epsilon) \quad (5)$$
$$\le \frac{1}{n}I(X^n; Z^n, Y^n) + \mu(\epsilon) \quad (6)$$
$$\le \frac{1}{n}\sum_{i=1}^{n} I(X_i; Z_i, Y_i) + \mu(\epsilon) \quad (7)$$
$$= I(X_Q; Z_Q, Y_Q|Q) + \mu(\epsilon) \quad (8)$$

where in (7) we used the fact that the channel is memoryless; and

$$R \le \frac{1}{n}I(X^n; I, Y^n) + \mu(\epsilon) \quad (9)$$
$$= \frac{1}{n}I(X^n; Y^n) + \frac{1}{n}H(I|Y^n) - \frac{1}{n}H(I|X^n) + \mu(\epsilon) \quad (10)$$
$$\le \frac{1}{n}\sum_{i=1}^{n} I(X_i; Y_i) + \frac{1}{n}H(I|Y^n) - \frac{1}{n}H(I|X^n) + \mu(\epsilon). \quad (11)$$
$$= I(X_Q; Y_Q|Q) + \frac{1}{n}H(I|Y^n) - \frac{1}{n}H(I|X^n) + \mu(\epsilon). \quad (12)$$

The proof is completed since the Markov chain $Q - X_Q - Y_Q$ implies that $I(X_Q; Y_Q|Q) \leq I(X_Q; Y_Q)$; similarly $I(X_Q; Z_Q, Y_Q|Q) \leq I(X_Q; Z_Q, Y_Q)$. ■

Note that this is an n-letter upper bound on the capacity of the channel. However if we can establish bounds on $\frac{1}{n}H(I|Y^n)$ for any value of $\frac{1}{n}H(I|X^n) := h$ satisfying the conditions in the proposition, we can use the proposition to establish an explicit computable upper bound on the capacity of the symmetric primitive relay channel. This is the approach of [8], [9], [11], which we also adopt in this paper.

III. Preliminaries

In this section, we introduce some more notation and provide a brief overview of reverse hypercontractivity, which will be our main tool for proving upper bounds on the capacity of the relay channel defined in the previous section.

A. Notation

Given a measurable space $\mathcal{Y}$, let $\mathcal{H}_{[0,1]}(\mathcal{Y})$ (resp. $\mathcal{H}_+(\mathcal{Y})$) be the set of measurable functions on $\mathcal{Y}$ taking values in $[0, 1]$ (resp. $[0, \infty)$). Given a probability measure Q on $\mathcal{Y}$, $f \in \mathcal{H}_+(\mathcal{Y})$, and $p \in (0, \infty)$, let

$$Q(f) := \int f \, \mathrm{d}\, Q, \tag{13}$$

$$\|f\|_{L^p(Q)} = \|f\|_p := [Q(f^p)]^{\frac{1}{p}}. \tag{14}$$

Then by a limiting argument we have

$$\|f\|_{L^0(Q)} := e^{Q(\ln f)}. \tag{15}$$

Given a channel (i.e. conditional probability) $W = P_{Y|X}$, we will often write $W_{x^n} := \otimes_{i=1}^n P_{Y|X=x_i}$. Finally, the bases in all logarithms, exponentials and information-theoretic quantities in this paper are natural.

B. Reverse Hypercontractivity

Let $T \colon \mathcal{H}_+(\mathcal{Y}) \to \mathcal{H}_+(\mathcal{Y})$ be a nonnegativity-preserving map (i.e., nonnegative functions are mapped to nonnegative functions) and let Q be a fixed reference measure. T is said to be reverse hypercontractive (see for example [33]) if for some $0 \leq p < q \leq 1$,

$$\|Tf\|_p \geq \|f\|_q, \quad \forall f \in \mathcal{H}_+(\mathcal{Y}). \tag{16}$$

Note that if T is a conditional expectation operator (i.e., there exists some conditional probability $P_{Y|X}$ such that

$$(Tf)(x) = P_{Y|X=x}(f) \tag{17}$$

for each $x \in \mathcal{Y}$), then (16) holds if $0 \leq p = q \leq 1$, by Jensen's inequality. Reverse hypercontractivity characterizes how T is able to increase the small values of the function (i.e. positivity improving [33]). Markov semigroups provide a rich source of operators satisfying reverse hypercontractivity (among other favorable properties).

C. Ornstein-Uhlenbeck Semigroups

For any $x^n \in \mathbb{R}^n$ and $t \geq 0$, define a linear operator $T_{x^n,t}$ by

$$T_{x^n,t} f(y^n) := \mathbb{E}[f(e^{-t}y^n + (1 - e^{-t})x^n + \sqrt{1 - e^{-2t}}V^n)] \tag{18}$$

for any $f \in \mathcal{H}_+(\mathbb{R}^n)$, where $V^n \sim \mathcal{N}(0^n, \mathbf{I}_n)$. Then $(T_{x^n,t})_{t \geq 0}$ is called the Ornstein-Uhlenbeck (OU) semigroup with stationary measure $P_{Y^n|X^n=x^n} := \mathcal{N}(x^n, \mathbf{I}_n)$ (see for example [33]). The OU semigroup is among the first examples where a reverse hypercontractivity estimate has been worked out:

Lemma 2 [33]: For any $q < p < 1$ and $t \geq \frac{1}{2} \ln \frac{1-q}{1-p}$, we have

$$\|T_{x^n,t} f\|_q \geq \|f\|_p, \quad \forall f \in \mathcal{H}_+(\mathbb{R}^n) \tag{19}$$

where the norms are with respect to the stationary measure $\mathcal{N}(x^n, \mathbf{I}_n)$.

In particular, taking $q = 0$, we see that for any $f \in \mathcal{H}_{[0,1]}(\mathbb{R}^n)$,

$$\mathbb{E}[\ln T_{x^n,t} f] \geq \ln \|f\|_{1-e^{-2t}} \tag{20}$$

$$\geq \frac{1}{1 - e^{-2t}} \ln \mathbb{E}[f] \tag{21}$$

$$\geq \left(1 + \frac{1}{2t}\right) \ln \mathbb{E}[f], \quad \forall f \in \mathcal{H}_{[0,1]}(\mathbb{R}^n) \tag{22}$$

where (21) used the fact that $f^p \geq f$, (22) follows from $\frac{1}{1-e^{-2t}} \leq \frac{1}{2t} + 1$ (note that $\ln \mathbb{E}[f]$ is negative!), and the expectations are with respect to the stationary measure.

We remark that Lemma 2 is completely dual to the (forward) hypercontractivity estimate for the OU semigroup (see e.g. [34]) in the sense that the dependence of the parameters $t \geq \frac{1}{2} \ln \frac{1-q}{1-p}$ takes on the same formula (although in the case of hypercontractivity, both p and q are greater than 1). However, this is merely a coincidence for the OU semigroup. The reverse hypercontractivity is generally weaker (and hence more common) than hypercontractivity [30], as the next example illustrates.

D. Simple and Semi-Simple Semigroups

The simplest and the most natural semigroup that can be defined for any given stationary measure P on a measurable space $\mathcal{Y}$ is the *simple semigroup* (see e.g. [30]), defined by

$$T_t f = e^{-t} f + (1 - e^{-t}) P(f), \quad \forall t \geq 0, f \in \mathcal{H}_+(\mathcal{Y}). \tag{23}$$

In the i.i.d. case, the tensor product $T_t^{\otimes n}$ of any Markov semigroup operator T_t forms a new Markov semigroup whose stationary measure is $P^{\otimes n}$. The product inherits the same reverse hypercontractive inequality as its factors, which is called *tensorization* (see e.g. [30]). We call the tensor product of a simple semigroup a *semi-simple* semigroup. We will use the semi-simple semigroup in the case of discrete memoryless channels and continious channels with bounded density.

By establishing the equivalence between the modified log-Sobolev inequality and the reverse hypercontractivity, the recent paper by Mossel et al. [30] established the following striking universal reverse hypercontractivity estimate which does not depend on the stationary measure P. We remark that the (forward) hypercontractivity is drastically different in that the bound depends on the smallest probability mass in P (see e.g. [35]).

Lemma 3 [30]: Let P_i be a probability distribution on $\mathcal{Y}$ for each $i = 1, \dots, n$, and let

$$T_t := \otimes_{i=1}^n [e^{-t} + (1 - e^{-t})P_i] \tag{24}$$

be a Markov semigroup with stationary measure $\otimes_{i=1}^n P$. For any $q < p < 1$ and $t \geq \ln \frac{1-q}{1-p}$, we have

$$\|T_t f\|_q \geq \|f\|_p, \quad \forall f \in \mathcal{H}_+(\mathcal{Y}^n) \tag{25}$$

where the norms are with respect to the stationary measure.

In particular, taking $q = 0$ and using the same arguments before, we obtain

$$\mathbb{E}[\ln T_t f] \geq \frac{1}{1 - e^{-t}} \ln \mathbb{E}[f] \tag{26}$$

$$\geq \left(1 + \frac{1}{t}\right) \ln \mathbb{E}[f], \quad \forall f \in \mathcal{H}_{[0,1]}(\mathcal{Y}^n) \tag{27}$$

where the expectations are with respect to the stationary measure.

IV. Main Results

We first state our results for the Gaussian case and then for channels with bounded density.

A. Gaussian Channels

Lemma 4: Let $I - Z^n - X^n - Y^n$ and $P_{Y^n|X^n=x^n} = P_{Z^n|X^n=x^n} = \mathcal{N}(x^n, \mathbf{I}_n)$. Define $h := \frac{1}{n} H(I|X^n)$. Then

$$H(I|Y^n) \leq n \min_{t>0} \left\{ t + \frac{1}{1 - e^{-2t}} h \right\} \tag{28}$$

$$= \frac{n}{2} \ln\left(1 + h + \sqrt{h^2 + 2h}\right) + \frac{n}{2}\left(h + \sqrt{h^2 + 2h}\right) \tag{29}$$

$$\leq n(h + \sqrt{2h}). \tag{30}$$

The relaxed bound (30) is the same as [9, Lemma 4.1].[2] The inequality in (30) is strict when $h > 0$.[3] Thus (29) is a strict improvement of [9, Lemma 4.1]. When combined with Proposition 1, this results immediately yield the following upper bound on the capacity of the Gaussian symmetric primitive relay channel. We say $P > 0$ is an average power constraint if the codewords satisfy

$$\frac{1}{2^{nR}} \sum_{m=1}^{2^{nR}} \|x^n(m)\|_2^2 \leq nP. \tag{31}$$

[2]Note that the entropy and rates in [9] are defined in bits, while we use nats in the current paper.

[3]To see (30), first use the change of variable $t := h + \sqrt{h^2 + 2h}$. Then (30) is equivalent to $\ln(1 + t) \leq \frac{-t + 2t\sqrt{1+t}}{1+t}$ when $t \geq 0$. We can then verify that both sides vanish as $t = 0$ and their derivatives satisfy strict inequality for $t > 0$.

Corollary 5: The capacity of the symmetric Gaussian primitive relay channel with $P_{Y|X=x} = P_{Z|X=x} = \mathcal{N}(0, N)$, average power constraint P and relay rate $C_0 \geq 0$ satisfies

$$C(C_0) \leq \left\{ \frac{1}{2} \ln\left(1 + \frac{2P}{N}\right), \frac{1}{2} \ln\left(1 + \frac{P}{N}\right) + C_0 - c^{-1}(C_0) \right\} \tag{32}$$

where $c^{-1}(\cdot)$ denotes the inverse of the surjective function $c\colon [0, \infty) \to [0, \infty)$, $h \mapsto \frac{1}{2} \ln\left(1 + h + \sqrt{h^2 + 2h}\right) + \frac{1}{2}\left(h + \sqrt{h^2 + 2h}\right)$.

Proof: Let X^n be an equiprobably selected codeword, and Q be the time sharing random variable as in Proposition 1, and note that

$$E[X_Q^2] = \frac{1}{n} \sum_{i=1}^n E[X_i^2] = \frac{1}{n} E\left[\sum_{i=1}^n X_i^2\right] \leq P. \tag{33}$$

Then Proposition 1 gives

$$R \leq I(X_Q; Y_Q, Z_Q) + \mu(\epsilon) \tag{34}$$

$$R \leq I(X_Q; Y_Q) + \frac{1}{n} H(I|Y^n) - \frac{1}{n} H(I|X^n) + \mu(\epsilon) \tag{35}$$

$$\leq I(X_Q; Y_Q) + \min\{C_0, c(h)\} - h + \mu(\epsilon) \tag{36}$$

where we defined $h = \frac{1}{n} H(I|X^n)$, and (36) used the fact that $\frac{1}{n} H(I|Y^n) \leq C_0$ and applied Lemma 4. To finish, note that the mutual information terms are maximized by choosing $X_Q \sim \mathcal{N}(0, P)$, and $\min\{C_0, c(h)\} - h$ is maximized at $h = c^{-1}(C_0)$ which can be seen from the fact that $c(h) - h$ is an increasing function. ■

Using similar lines of argument as in the proof of Lemma 4, we will be able to obtain bounds for channels with bounded densities in Section IV-B (by using a different semigroup adapted to that class of channels). For the same Gaussian setting as Lemma 4, however, we can capitalize on certain scaling properties of the Gaussian channel and use slightly different lines of proof, to show the following sharper bound:

Lemma 6: Let $I - Z^n - X^n - Y^n$ where $P_{Y^n|X^n=x^n} = P_{Z^n|X^n=x^n} = \mathcal{N}(x^n, \mathbf{I}_n)$, and define $h_2 := \frac{1}{n} H(I|Y^n)$ and $h_1 := \frac{1}{n} H(I|X^n)$. Then

$$h_2 - h_1 \leq \frac{1}{2} \ln(1 + 2h_2). \tag{37}$$

Asymptotically, [9, Lemma 4.1] (which is (30)) gives[4]

$$h_2 \lesssim \sqrt{2h_1}, \quad h_1 \ll 1; \tag{38}$$

$$h_2 \lesssim 2h_1, \quad h_1 \gg 1. \tag{39}$$

The bound in (29) has the same asymptotics (38); but with (39) replaced by $h_2 \lesssim h_1$, $h_1 \gg 1$. In contrast, the bound in Lemma 6 gives

$$h_2 \lesssim \sqrt{h_1}, \quad h_1 \ll 1; \tag{40}$$

$$h_2 \lesssim h_1, \quad h_1 \gg 1, \tag{41}$$

so Lemma 6 essentially yields improvements by constant factors. A numerical comparison is shown in Figure 2.

Combined with Proposition 1, this results yields the following upper bound on the capacity of the Gaussian symmetric

[4]We write $c(h) \lesssim h$, $h \ll 1$, if $\limsup_{h \downarrow 0} \frac{c(h)}{h} \leq 1$.

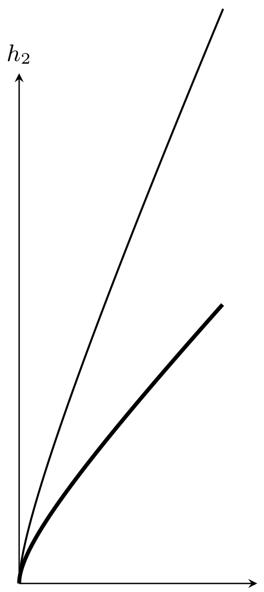

Fig. 2. The bound in [9, Lemma 4.1] (see (30)) is plotted in the thin line and the bound in Lemma 6 is plotted in the thick line, for $h_1 \in (0, 3)$.

primitive relay channel. The proof is immediate and follows similar arguments as in the proof of Corollary 7.

Corollary 7: The capacity of the symmetric Gaussian primitive relay channel with $P_{Y|X=x} = P_{Z|X=x} = \mathcal{N}(x, N)$ with average power constraint P satisfies

$$C(C_0) \le \min\left\{\frac{1}{2}\ln\left(1+\frac{2P}{N}\right), \frac{1}{2}\ln\left(1+\frac{P}{N}\right)+\frac{1}{2}\ln(1+2C_0)\right\}. \tag{42}$$

B. Channel Distributions With Bounded Densities

In this section, we state our results for discrete memoryless channels. More generally, our bounds apply to channels with bounded conditional density, or more precisely when one can find a reference measure such that the density of the output distribution of a stationary memoryless channel can be bounded by a constant independent of the input distribution.

Lemma 8: Fix $W = P_{Y|X}$. Suppose that $I - Z^n - X^n - Y^n$ where $P_{Y^n|X^n} = P_{Z^n|X^n} = W^{\otimes n}$, and

$$\alpha := \sup_x \left\| \frac{\mathrm{d}\, P_{Y|X=x}}{\mathrm{d}\, Q_Y} \right\|_\infty < \infty \tag{43}$$

for some probability measure Q_Y. (In the discrete case, we can always take $\alpha = \sum_y \max_x W_x(y)$.) Define $h := \frac{1}{n}H(I|X^n)$. Then

$$\begin{aligned} &\frac{1}{n}H(I|Y^n) \\ &\le c_\alpha(h) := \min_{t>0}\left\{(\alpha-1)t + \frac{1}{1-e^{-t}}h\right\} && (44) \\ &= (\alpha-1)\left[\ln\left(1+\frac{h}{2(\alpha-1)}+\sqrt{\frac{h}{\alpha-1}+\frac{h^2}{4(\alpha-1)^2}}\right)\right. \\ &\quad \left. + \frac{h}{2(\alpha-1)} + \sqrt{\frac{h}{\alpha-1}+\frac{h^2}{4(\alpha-1)^2}}\right]. && (45) \end{aligned}$$

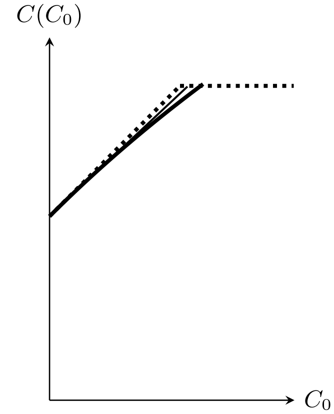

Fig. 3. The bound on capacity using in [9, Lemma 4.1] (see (30)) is plotted in the thin line. The bound in Corollary 7 is plotted in the thick line. The cutset bound is plotted in the dotted line. The signal to noise ratio is chosen as $\frac{P}{N} = 0.5$. The range of the relay rate is $C_0 \in (0, 0.27)$.

Remark 1: In [8, Lemma 7.1], a weaker bound of

$$\frac{1}{n}H(I|Y^n) \le O\left(\sqrt{h}\ln\frac{1}{h}\right), \quad h \to 0 \tag{46}$$

was derived using the blowing-up lemma for discrete memoryless channels. Here we got rid of the a logarithmic factor.

Remark 2: Recall that the ∞-mutual information for a given $P_{Y|X}$ and P_X is defined as

$$I_\infty(X;Y) := \inf_{Q_Y} \ln \left\| \frac{\mathrm{d}P_{XY}}{\mathrm{d}(P_X \times Q_Y)} \right\|_\infty; \tag{47}$$

see e.g. [36] and the references therein. Under reasonable regularity conditions (e.g. P_X is a fully supported distribution on a countably infinite set), we have $I_\infty(X;Y) = \inf_{Q_Y} \ln \sup_x \left\| \frac{\mathrm{d}P_{Y|X=x}}{\mathrm{d}Q_Y} \right\|_\infty$ and hence the optimal α in (43) is $e^{I_\infty(X;Y)}$. Note that the particular choice of P_X is immaterial for the calculation of $I_\infty(X;Y)$ as long as P_X is fully supported. Also note that in the case of Gaussian channel without a power constraint, it is not possible to find Q_Y for which α defined in (43) is finite.

The above lemma immediately yields the following upper bound on the capacity of primitive relay channels. The proof is similar to the proof of Corollary 7.

Corollary 9: Consider a stationary memoryless primitive relay channel with $P_{Y|X=x} = P_{Z|X=x}$ and suppose that the condition (43) is satisfied. Let α and $c_\alpha(h)$ be as defined in Lemma 8. Then the capacity satisfies

$$C(C_0) \le \min\{I(X;Y,Z),\, I(X;Y) + C_0 - c_\alpha^{-1}(C_0)\}$$

for some random variable $X \in \mathcal{X}$ and $c_\alpha^{-1}(\cdot)$ denotes the inverse function.

V. Proofs

A. Proof of Lemma 4

Proof: For each integer (relay message) $i \in \mathcal{I}$, let $f_i \in \mathcal{H}_{[0,1]}(\mathbb{R}^n)$ be the probability of sending i upon observing the channel output at the relay Z^n (in the case of deterministic relay decoder, f_i will be an indicator function). Then

$$\mathbb{P}[I = i | X^n = x^n] = W_{x^n}(f_i). \tag{48}$$

Hence

$$H(I|X^n) = \mathbb{E}_{X^n I}\left[\ln \frac{1}{W_{X^n}(f_I)}\right]. \tag{49}$$

But from (22), we see that for any x^n and i,

$$W_{x^n}\left(\ln(T_{x^n,t}f_i)\right) \geq \frac{1}{1-e^{-2t}} \ln W_{x^n}(f_i). \tag{50}$$

To finish the proof, let us define a convolution operator T_t that does not depend on x^n:

$$T_t f(y^n) := \mathbb{E}[f(y^n + \sqrt{1-e^{-2t}}V^n)]. \tag{51}$$

In words, the action of $T_{x^n,t}$ can be viewed as consisting of two steps: first T_t, and then dilate (with center x^n) by a factor e^t. It is easy to see a basic fact: when a function is integrated against a measure, the integral is invariant if both the function and the measure contract by the same factor; this observation proves that

$$W_{x^n}\left(\ln(T_{x^n,t}f_i)\right) = \bar{W}_{x^n}\left(\ln(T_t f_i)\right) \tag{52}$$

where we defined a new channel $\bar{W}_{x^n} = \mathcal{N}(x^n, e^{-2t}\mathbf{I}_n)$. Now define a new Markov chain $I - Z^n - X^n - \bar{Y}^n$ where $P_{\bar{Y}^n|X^n} = \bar{W}$. We have

$$\begin{aligned} -H(I|\bar{Y}^n) &= \mathbb{E}_{I\bar{Y}^n}[\ln P_{I|\bar{Y}^n}(I|\bar{Y}^n)] && (53)\\ &\geq \mathbb{E}_{I\bar{Y}^n}\left[\ln\left((T_t f_I)(\bar{Y}^n)\right)\right] && (54)\\ &= \mathbb{E}_{IX^n}\left[\bar{W}_{X^n}\left(\ln(T_t f_I)\right)\right] && (55)\\ &= \mathbb{E}_{IX^n}\left[W_{X^n}\left(\ln(T_{X^n,t}f_I)\right)\right] && (56)\\ &\geq \frac{1}{1-e^{-2t}}\mathbb{E}_{IX^n}\left[\ln W_{X^n}(f_I)\right] && (57)\\ &= -\frac{1}{1-e^{-2t}}H(I|X^n) && (58) \end{aligned}$$

where

- (54) used the fact that for any $\bar{y}^n$,

$$\sum_{i\in\mathcal{I}}(T_t f_i)(\bar{y}^n) = \left(T_t \sum_{i\in\mathcal{I}} f_i\right)(\bar{y}^n) = 1. \tag{59}$$

  Indeed, upon rearrangements, (54) is reduced to the nonnegativity of the conditional relative entropy

$$D(P_{I|\bar{Y}^n}\|Q_{I|\bar{Y}^n}|P_{\bar{Y}^n}) \geq 0 \tag{60}$$

  where we defined the conditional distribution $Q_{I|\bar{Y}^n}$ as $Q_{I|\bar{Y}^n}(i|\bar{y}^n) = T_t f_i(\bar{y}^n)$, for each $i, \bar{y}^n$.
- (56) is from (52).
- (57) is from (50) and (49).

However,

$$\begin{aligned} &H(I|Y^n) - H(I|\bar{Y}^n)\\ &\leq h(Y^n|I) - h(\bar{Y}^n|I) && (61)\\ &= I(\sqrt{1-e^{-2t}}G^n; Y^n|I) && (62)\\ &\leq I(\sqrt{1-e^{-2t}}G^n; \sqrt{1-e^{-2t}}G^n + e^{-t}\bar{G}^n|I) && (63)\\ &= nt && (64) \end{aligned}$$

where (61) can be seen from EPI; In (62) we defined G^n and $\bar{G}^n$ to be independent according to $\mathcal{N}(\mathbf{0}, \mathbf{I}_n)$, and independent of (I, X^n), and put

$$\bar{Y}^n = X^n + e^{-t}\bar{G}^n; \tag{65}$$

$$Y^n = \bar{Y}^n + \sqrt{1-e^{-2t}}G^n. \tag{66}$$

This constructs a coupling of

$$I - X^n - \bar{Y}^n - Y^n \tag{67}$$

with the desired marginal distributions. The conclusion then follows by optimizing t. ■

Remark 3: As alluded, we focused on the primitive relay channel for simplicity, while the argument can easily be extended to the general (asymmetric) case. Indeed, suppose that $P_{Y|X}$ is the additive Gaussian channel with variance σ^2. Then (64) will be modified to $H(I|Y^n) - H(I|\bar{Y}^n) \leq n(t - \ln\sigma)$, and the final bound can be obtained by optimizing t for which $\sigma^2 \geq e^{-2t}$. Similar comments applies to other lemmas: In (74) ahead, we define $\hat{W}_{x^n} := \mathcal{N}(x^n, \sigma^2 e^{2t}\mathbf{I}_n)$ instead, and obtain

$$H(I|X^n) \geq \mathbb{E}\left[\ln\frac{1}{\hat{W}_{X^n}(f_I)}\right] + n(\ln\sigma + t + \frac{1}{2\sigma^2}e^{-2t} - \frac{1}{2}) \tag{68}$$

so that the rest of the proof continue and t is optimized over $(0,\infty)$. In (87) ahead, suppose that $V_{x^n} := P_{Y^n|X^n=x^n}$ for each x^n, then we have

$$H(I|X^n) \geq \mathbb{E}_{IX^n}\left[\ln\frac{1}{V_{X^n}(f_I)}\right] + n\ln\sup_x\left\|\frac{V_x}{W_x}\right\|_\infty \tag{69}$$

and the rest of the proof continues.

B. Proof of Lemma 6

Proof: The high-level idea of the proof is roughly as follows: in the proof of Lemma 4 we established the following result: $H(I|\bar{Y}^n) \leq \frac{1}{1-e^{-2t}}H(I|X^n)$, where $\bar{Y}^n$ is obtained by scaling the noise by a factor e^{-t}. Then we bound $H(I|Y^n)$ in terms of $H(I|\bar{Y}^n)$. If, instead, we somehow scale the noise by a factor of e^t beforehand to cancel this effect, then we can directly obtain a bound on $H(I|Y^n)$.

Analogous to (52), define the conditional distribution

$$\hat{W}_{x^n} := \mathcal{N}(x^n, e^{2t}\mathbf{I}_n), \tag{70}$$

then we have the following scaling invariance:

$$\hat{W}_{x^n}\left(\ln(\hat{T}_{x^n,t}f)\right) = W_{x^n}\left(\ln(\hat{T}_t f)\right) \tag{71}$$

where $\hat{T}_{x^n,t}$ is the semigroup with stationary measure $\hat{W}_{x^n}$, and $\hat{T}_t$ is defined by

$$\hat{T}_t f(y^n) := \mathbb{E}[f(y^n + e^t\sqrt{1-e^{-2t}}V^n)]. \tag{72}$$

Then we can use the similar steps as before:

$$H(I|X^n) = \mathbb{E}\left[\ln \frac{1}{W_{X^n}(f_I)}\right] \tag{73}$$
$$\geq \mathbb{E}\left[\ln \frac{1}{\hat{W}_{X^n}(f_I)}\right] - nt + \frac{n}{2}(1-e^{-2t}) \tag{74}$$
$$\geq (1-e^{-2t})\mathbb{E}\left[\hat{W}_{X^n}\left(\ln \frac{1}{\hat{T}_{X^n,t} f_I}\right)\right] - nt + \frac{n}{2}(1-e^{-2t}) \tag{75}$$
$$= (1-e^{-2t})\mathbb{E}\left[W_{X^n}\left(\ln \frac{1}{\hat{T}_t f_I}\right)\right] - nt + \frac{n}{2}(1-e^{-2t}) \tag{76}$$
$$= (1-e^{-2t})\mathbb{E}\left[\ln \frac{1}{\hat{T}_t f_I(Y^n)}\right] - nt + \frac{n}{2}(1-e^{-2t}) \tag{77}$$
$$\geq (1-e^{-2t})H(I|Y^n) - nt + \frac{n}{2}(1-e^{-2t}). \tag{78}$$

where

- (74) follows from

$$\mathbb{E}\left[\ln \frac{W_{X^n}(f_I)}{\hat{W}_{X^n}(f_I)}\right] = D(P_{I|Z^n} \circ W_{X^n} \| P_{I|Z^n} \circ \hat{W}_{X^n} | P_{X^n}) \tag{79}$$
$$\leq D(W_{X^n} \| \hat{W}_{X^n} | P_{X^n}) \tag{80}$$
$$= nt - \frac{n}{2}(1-e^{-2t}). \tag{81}$$

Thus

$$\frac{1}{n}I(I;X^n|Y^n) \leq \inf_{t>0}\left\{t - \frac{1}{2} + (\frac{1}{2} + \frac{1}{n}H(I|Y^n))e^{-2t}\right\}$$
$$= \frac{1}{2}\ln\left(1 + \frac{2}{n}H(I|Y^n)\right). \tag{82}$$

■

C. Proof of Lemma 8

Proof: As before, define $(f_i)_{i\in\mathcal{I}}$ as the relay decoding probability functions as before. For any $t > 0$, define the linear operators

$$T_{x^n,t} := \otimes_{i=1}^n (e^{-t} + (1-e^{-t})W_x); \tag{83}$$
$$\Lambda_t := \otimes_{i=1}^n (e^{-t} + \alpha(1-e^{-t})Q_Y). \tag{84}$$

Note that since $\alpha \geq 1$, Λ_t is not a conditional expectation operator in the sense that it can send the constant 1 function to a nonnegative function exceeding 1 somewhere. However, we can show that the factor of increase is not too big:

$$(\Lambda_t \cdot 1)(y^n) \leq (e^{-t} + \alpha(1-e^{-t}))^n \tag{85}$$
$$\leq e^{(\alpha-1)nt}, \quad \forall y^n. \tag{86}$$

The rest of the proof then follows analogously to the Gaussian case:

$$H(I|X^n) = \mathbb{E}_{IX^n}\left[\ln \frac{1}{W_{X^n}(f_I)}\right] \tag{87}$$
$$\geq (1-e^{-t})\mathbb{E}_{IX^n}\left[W_{X^n}\left(\ln \frac{1}{T_{X^n,t} f_I}\right)\right] \tag{88}$$
$$\geq (1-e^{-t})\mathbb{E}_{IX^n}\left[W_{X^n}\left(\ln \frac{1}{\Lambda_t f_I}\right)\right] \tag{89}$$
$$\geq (1-e^{-t})\mathbb{E}_{IY^n}\left[\ln \frac{1}{e^{-(\alpha-1)nt} p_{I,Y^n}}\right] \tag{90}$$
$$\geq (1-e^{-t})\mathbb{E}_{IY^n}\left[\ln P_{I|Y^n}(I|Y^n) - (\alpha-1)nt\right] \tag{91}$$
$$\geq (1-e^{-t})\left[H(I|Y^n) - (\alpha-1)nt\right] \tag{92}$$

where

- (89) follows since Λ_t dominates $T_{x^n,t}$.
- In (90), we defined, for each i, y^n,

$$p_{i,y^n} = e^{(\alpha-1)nt}\Lambda_t f_i(y^n). \tag{93}$$

  For each y^n, since $\sum_i f_i(y^n) = 1$, we see from (86) and the linearity of Λ_t that $\sum_i p_{i,y^n} \leq 1$.
- (91) follows since it is equivalent to the non-negativity of relative entropy, by rearrangements.

■

VI. Discussion

As mentioned in the introduction, recently Wu, Barnes and Ozgur [11] used the rearrangement inequalities to prove a tighter upper bound on the capacity of the Gaussian primitive relay channel which remains bounded away from the cutset bound also in the high relay rate regime (corresponding to the case of large C_0 or equivalently $\frac{1}{n}H(I|Y^n)$). (See [32] for the treatment of the binary symmetric channel.) Currently, the reverse hypercontractivity argument does not seem to be powerful enough in that regime. In particular, the bound

$$H(I|\bar{Y}^n) \leq \frac{1}{1-e^{-2t}}H(I|X^n),$$

obtained in (58) is looser than the trivial inequality

$$H(I|\bar{Y}^n) \leq H(I|X^n) + \sup_{P_{X^n}} I(I;X^n|\bar{Y}^n)$$

in the high entropy regime. One possibility is that for highly symmetric measures (e.g. Gaussian), for which a rearrangement inequality that characterizes the extremal sets/functions exists, the rearrangement approach is inherently stronger than reverse hypercontractivity. Another possibility is that there still exists certain semigroup argument that simplifies the proof [11]. Indeed, we note that there exist semigroups versions of rearrangement inequalities which are known to imply (reverse) hypercontractive inequalities (see [37, P117]). This remains an interesting open direction for future research.

Acknowledgment

The authors would like to thank Prof. Chandra Nair for helpful comments on the manuscript. Jingbo Liu would also like to thank Prof. Ramon van Handel and Sergio Verdú for their generous support, encouragement, and enlightening discussions during the time of this research. In particular, the reference [37, P117] mentioned in Section VI was suggested by Ramon van Handel.

REFERENCES

[1] T. M. Cover, "The capacity of the relay channel," in *Open Problems in Communication and Computation*, T. M. Cover and B. Gopinath, Eds. New York, NY, USA: Springer-Verlag, 1987, pp. 72–73.

[2] Y.-H. Kim, "Coding techniques for primitive relay channels," in *Proc. Annu. Allerton Conf. Commun., Control, Comput.*, 2007, pp. 129–135.

[3] Z. Zhang, "Partial converse for a relay channel," *IEEE Trans. Inf. Theory*, vol. 34, no. 5, pp. 1106–1110, Sep. 1988.

[4] M. Aleksic, P. Razaghi, and W. Yu, "Capacity of a class of modulo-sum relay channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 3, pp. 921–930, Mar. 2009.

[5] R. Tandon and S. Ulukus, "Capacity of a class of semi-deterministic primitive relay channels," in *Proc. 48th Annu. Allerton Conf. Commun., Control, Comput. (Allerton)*, Sep. 2010, pp. 931–935.

[6] N. Fawaz and M. Medard, "A converse for the wideband relay channel with physically degraded broadcast," in *Proc. IEEE Inf. Theory Workshop*, Oct. 2011, pp. 425–429.

[7] F. Xue, "A new upper bound on the capacity of a primitive relay channel based on channel simulation," *IEEE Trans. Inf. Theory*, vol. 60, no. 8, pp. 4786–4798, Aug. 2014.

[8] X. Wu, A. Ozgiur, and L.-L. Xie, "Improving on the cut-set bound via geometric analysis of typical sets," *IEEE Trans. Inf. Theory*, vol. 63, no. 4, pp. 2254–2277, Apr. 2017.

[9] X. Wu and A. Ozgur, "Cut-set bound is loose for Gaussian relay networks," *IEEE Trans. Inf. Theory*, vol. 64, no. 2, pp. 1023–1037, Feb. 2018.

[10] X. Wu and A. Ozgur, "Improving on the cut-set bound for general primitive relay channels," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2016, pp. 1675–1679.

[11] X. Wu, L. P. Barnes, and A. Ozgur, "'The capacity of the relay channel': Solution to cover's problem in the Gaussian case," *IEEE Trans. Inf. Theory*, vol. 65, no. 1, pp. 255–275, Jan. 2019.

[12] X. Wu, L. P. Barnes, and A. Ozgur, "The geometry of the relay channel," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2017, pp. 2233–2237.

[13] Y. Chen and N. Devroye, "Zero-error relaying for primitive relay channels," *IEEE Trans. Inf. Theory*, vol. 63, no. 12, pp. 7708–7715, Dec. 2017.

[14] T. Cover and A. E. Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Inf. Theory*, vol. 25, no. 5, pp. 572–584, Sep. 1979.

[15] L. R. Ford and D. R. Fulkerson, "Maximal flow through a network," *Can. J. Math.*, vol. 8, pp. 399–404, 1956.

[16] K. Marton, "A simple proof of the blowing-up lemma (Corresp.)," *IEEE Trans. Inf. Theory*, vol. 32, no. 3, pp. 445–446, May 1986.

[17] M. Talagrand, "Transportation cost for Gaussian and other product measures," *Geometric Funct. Anal.*, vol. 6, no. 3, pp. 587–600, May 1996.

[18] M. Ledoux, "Concentration of measure and logarithmic Sobolev inequalities," in *Séminaire de Probabilités XXXIII*. Berlin, Germany: Springer, 1999, pp. 120–216.

[19] M. Raginsky and I. Sason, "Concentration of measure inequalities in information theory, communications, and coding," *Found. Trends Commun. Inf. Theory*, vol. 10, nos. 1–2, pp. 1–250, 2013.

[20] J. Liu, R. van Handel, and S. Verdu, "Beyond the blowing-up lemma: Sharp converses via reverse hypercontractivity," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2017, pp. 943–947.

[21] J. Liu, R. van Handel, and S. Verdú, "Second-order converses via reverse hypercontractivity," 2018, *arXiv:1812.10129*. [Online]. Available: http://arxiv.org/abs/1812.10129

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

[23] J. Liu, T. A. Courtade, P. Cuff, and S. Verdú, "Smoothing Brascamp-Lieb inequalities and strong converses for common randomness generation," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Barcelona, Spain, Jul. 2016, pp. 1043–1047.

[24] J. Liu, T. Courtade, P. Cuff, and S. Verdú, "A forward-reverse brascamp-lieb inequality: Entropic duality and Gaussian optimality," *Entropy*, vol. 20, no. 6, p. 418, May 2018.

[25] S. Beigi and C. Nair, "Equivalent characterization of reverse Brascamp-Lieb-type inequalities using information measures," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Barcelona, Spain, Jul. 2016, pp. 1–5.

[26] C. Nair and Y. N. Wang, "Reverse hypercontractivity region for the binary erasure channel," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Aachen, Germany, Jun. 2017, pp. 938–942.

[27] J. Liu, P. Cuff, and S. Verdu, "Secret key generation with one communicator and a one-shot converse via hypercontractivity," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Hong Kong, Jun. 2015, pp. 710–714.

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

[29] J. Liu, "Dispersion bound for the Wyner-Ahlswede-Körner network via reverse hypercontractivity on types," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2018, pp. 1854–1858.

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

[31] R. van Handel, "Probability in high dimension," Princeton Univ., Princeton, NJ, USA, Lecture Notes, Dec. 2016. [Online]. Available: https://web.math.princeton.edu/~rvan/APC550.pdf

[32] L. P. Barnes, X. Wu, and A. Ozgur, "A solution to cover's problem for the binary symmetric relay channel: Geometry of sets on the Hamming sphere," in *Proc. 55th Annu. Allerton Conf. Commun., Control, Comput. (Allerton)*, Oct. 2017, pp. 844–851.

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

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

[35] S. Boucheron, G. Lugosi, and O. Bousquet, *Concentration Inequalities*. Berlin, Germany: Springer, 2004.

[36] S. Verdú, "α-mutual information," in *Proc. IEEE Inf. Theory Appl. Workshop (ITA)*, Feb. 2015, pp. 1–6.

[37] M. Ledoux, "Isoperimetry and Gaussian analysis," in *Lectures on Probability Theory and Statistics*. Berlin, Germany: Springer, 1996, pp. 165–294.

Jingbo Liu (Member, IEEE) received the B.E. degree from Tsinghua University, Beijing, China, in 2012, and the Ph.D. degree from Princeton University, Princeton, NJ, USA, in 2018, both in electrical engineering. He is currently a Postdoctoral Fellow at the MIT Institute for Data, Systems, and Society. In Fall 2020, he will join the Department of Statistics, University of Illinois, Urbana-Champaign, as an Assistant Professor (affiliate: Electrical & Computer Engineering). His research interests include information theory, statistical inference, high-dimensional probability, and the related fields. He gave a semi-plenary presentation at the 2015 IEEE International Symposium on Information Theory, Hong Kong. His bachelor thesis received the Best Undergraduate Thesis Award at Tsinghua University in 2012. He was a recipient of the Princeton University Wallace Memorial Fellowship in 2016. His Ph.D. thesis received the Bede Liu Best Dissertation Award of Princeton and the Thomas M. Cover Dissertation Award of the IEEE Information Theory Society in 2018.

Ayfer Özgür (Senior Member, IEEE) received the B.Sc. degrees in electrical engineering and physics and the M.Sc. degree in electrical engineering from Middle East Technical University, Turkey, in 2001 and 2004, respectively, and the Ph.D. degree from the Information Processing Group, EPFL, in 2009.

She is currently an Associate Professor of electrical engineering and the Chambers Faculty Scholar at the School of Engineering, Stanford University. Her current research interests include distributed communication and learning, wireless systems, and information theory. She received the EPFL Best Ph.D. Thesis Award in 2010, the NSF CAREER Award in 2013, the Okawa Foundation Research Grant and the IEEE Communication Theory Technical Committee (CTTC) Early Achievement Award in 2018, and the Google Faculty Research Award in 2019, and was selected as the inaugural Goldsmith Lecturer of the IEEE IT Society in 2020.