

# Strengthening the Entropy Power Inequality

Thomas A. Courtade

Department of Electrical Engineering and Computer Sciences  
University of California, Berkeley

## Abstract

We tighten the Entropy Power Inequality (EPI) when one of the random summands is Gaussian. Our strengthening is closely connected to the concept of strong data processing for Gaussian channels and generalizes the (vector extension of) Costa's EPI. This leads to a new reverse entropy power inequality and, as a corollary, sharpens Stam's inequality relating entropy power and Fisher information. Applications to network information theory are given, including a short self-contained proof of the rate region for the two-encoder quadratic Gaussian source coding problem.

Our argument is based on weak convergence and a technique employed by Geng and Nair for establishing Gaussian optimality via rotational-invariance, which traces its roots to a 'doubling trick' that has been successfully used in the study of functional inequalities.

## Index Terms

Entropy power inequality, Costa's EPI, Stam's Inequality, Strong Data Processing, Gaussian Source Coding

## I. INTRODUCTION AND MAIN RESULT

For a random variable  $X$  with density  $f$ , the differential entropy of  $X$  is defined by

$$h(X) = - \int f(x) \log f(x) dx. \quad (1)$$

Similarly,  $h(\mathbf{X})$  is defined to be the differential entropy of a random vector  $\mathbf{X} \in \mathbb{R}^n$  having density on  $\mathbb{R}^n$ . The celebrated Entropy Power Inequality (EPI) put forth by Shannon [1] and rigorously established by Stam [2] and Blachman [3] asserts that for  $X, W$  independent

$$2^{2h(X+W)} \geq 2^{2h(X)} + 2^{2h(W)}. \quad (2)$$

Under the assumption that  $W$  is Gaussian, we prove the following strengthening of (2):

**Theorem 1.** *Let  $X \sim P_X$ , and let  $W \sim N(0, \sigma^2)$  be independent of  $X$ . For any  $V$  satisfying  $X \rightarrow (X+W) \rightarrow V$ ,*

$$2^{2(h(X+W) - I(X;V))} \geq 2^{2(h(X) - I(X+W;V))} + 2^{2h(W)}. \quad (3)$$

The notation  $X \rightarrow (X+W) \rightarrow V$  in Theorem 1 follows the usual convention, indicating that the random variables  $X, X+W$  and  $V$  form a Markov chain, in that order. Throughout, we write  $X \rightarrow Y \rightarrow V|Q$  to denote random variables  $X, Y, V, Q$  with joint distribution factoring as  $P_{XYVQ} = P_{XQ}P_{Y|XQ}P_{V|YQ}$ . That is,  $X \rightarrow Y \rightarrow V$  form a Markov chain conditioned on  $Q$ .

When the integral (1) does not exist, or if  $X$  does not have density, then we adopt the convention that  $h(X) = -\infty$ . In this case, the inequality (3) is a trivial consequence of the data processing inequality. So, as with the classical EPI, Theorem 1 is only informative when  $X$  has density and  $h(X)$  exists.

A conditional version of the EPI is often useful in applications. Theorem 1 easily generalizes along these lines. Indeed, due to joint convexity of  $\log(2^x + 2^y)$  in  $x, y$ , we obtain the following corollary of Theorem 1:

This work was supported by NSF Grants CCF-1528132 and CCF-0939370 (Center for Science of Information).  
Email: courtade@berkeley.edu

**Corollary 1.** Suppose  $X, W$  are conditionally independent given  $Q$ , and moreover that  $W$  is conditionally Gaussian given  $Q$ . Then, for any  $V$  satisfying  $X \rightarrow (X + W) \rightarrow V|Q$ ,

$$2^{2(h(X+W|Q)-I(X;V|Q))} \geq 2^{2(h(X|Q)-I(X+W;V|Q))} + 2^{2h(W|Q)}. \quad (4)$$

It is interesting to note that the conditional version of the classical EPI assumes a form symmetric to (3). In particular, for  $Q \rightarrow X \rightarrow (X + W)$ , it holds that

$$2^{2(h(X+W)-I(X+W;Q))} \geq 2^{2(h(X)-I(X;Q))} + 2^{2h(W)}. \quad (5)$$

Note that the mutual informations in the exponents on the LHS and RHS of (3) and (5) respectively correspond to the smaller and larger mutual informations in the corresponding data processing inequalities  $I(X; V) \leq I(X + W; V)$  and  $I(X + W; Q) \leq I(X; Q)$ .

As one would expect, Theorem 1 also admits a vector generalization, which may be regarded as our main result:

**Theorem 2.** Suppose  $\mathbf{X}, \mathbf{W}$  are  $n$ -dimensional random vectors that are conditionally independent given  $Q$ , and moreover that  $\mathbf{W}$  is conditionally Gaussian given  $Q$ . Then, for any  $V$  satisfying  $\mathbf{X} \rightarrow (\mathbf{X} + \mathbf{W}) \rightarrow V|Q$ ,

$$2^{\frac{2}{n}(h(\mathbf{X}+\mathbf{W}|Q)-I(\mathbf{X};V|Q))} \geq 2^{\frac{2}{n}(h(\mathbf{X}|Q)-I(\mathbf{X}+\mathbf{W};V|Q))} + 2^{\frac{2}{n}h(\mathbf{W}|Q)}. \quad (6)$$

In the following section, we will see that the strengthening of the classical EPI afforded by Theorem 2 generalizes Costa's EPI [4] (and the vector generalization [5]), which has found applications ranging from interference channels to secrecy capacity (e.g., [5]–[8]). It also leads to a new reverse EPI, which can be applied to improve Stam's inequality or, equivalently, the Gaussian logarithmic Sobolev inequality. Moreover, we will see that Theorem 2 leads to a very brief proof of the converse for the rate region of the quadratic Gaussian two-encoder source-coding problem [9], [10]. Applications to one-sided interference channels and strong data processing inequalities are also given.

We remark that the restriction of  $\mathbf{W}$  to be conditionally Gaussian in Theorem 2 should not be a severe limitation in practice. Indeed, in applications of the EPI, it is typically the case that one of the variables is Gaussian. As noted by Rioul [11], examples include the scalar Gaussian broadcast channel problem [12] and its generalization to the multiple-input multiple-output case [13], [14]; the secrecy capacity of the Gaussian wiretap channel [15] and its multiple access extension [16]; determination of the corner points for the scalar Gaussian interference channel problem [6], [7]; the scalar Gaussian source multiple-description problem [17]; and characterization of the rate-distortion regions for several multiterminal Gaussian source coding schemes [10], [18], [19]. It is tempting to conjecture that (6) holds when the distribution of  $\mathbf{W}$  is unconstrained, however we suspect this is not true (but no counterexample was immediately apparent).

## II. APPLICATIONS

### A. Generalized Costa's Entropy Power Inequality

Costa's EPI [4] states that, for independent  $n$ -dimensional random vectors  $\mathbf{X} \sim P_{\mathbf{X}}$  and  $\mathbf{W} \sim N(0, \Sigma)$ ,

$$2^{\frac{2}{n}h(\mathbf{X}+\alpha\mathbf{W})} \geq (1 - \alpha^2)2^{\frac{2}{n}h(\mathbf{X})} + \alpha^2 2^{\frac{2}{n}h(\mathbf{X}+\mathbf{W})} \quad \text{for } |\alpha| \leq 1. \quad (7)$$

This result was generalized to a vector setting by Liu *et al.* using perturbation and I-MMSE arguments [5]. We demonstrate below that this generalization follows as an easy corollary to Theorem 2 by taking  $V$  equal to  $\mathbf{X}$  contaminated by additive Gaussian noise. In this sense, Theorem 2 may be interpreted as a further generalization of Costa's EPI, where the additive noise is no longer restricted to be Gaussian.

**Theorem 3.** [5] Let  $\mathbf{X} \sim P_{\mathbf{X}}$  and  $\mathbf{W} \sim N(0, \Sigma)$  be independent,  $n$ -dimensional random vectors. For a positive semidefinite matrix  $A \preceq I$ ,

$$2^{\frac{2}{n}h(\mathbf{X}+A^{1/2}\mathbf{W})} \geq |I - A|^{1/n} 2^{\frac{2}{n}h(\mathbf{X})} + |A|^{1/n} 2^{\frac{2}{n}h(\mathbf{X}+\mathbf{W})}. \quad (8)$$

*Proof:* Let  $\mathbf{W}_1, \mathbf{W}_2$  denote two independent copies of  $\mathbf{W}$ , and put  $\mathbf{Y} = \mathbf{X} + A^{1/2}\mathbf{W}_1$  and  $V = \mathbf{Y} + (I - A)^{1/2}\mathbf{W}_2$ . Note that  $V = \mathbf{X} + \mathbf{W}$  in distribution so that  $I(\mathbf{X}; V) = h(\mathbf{X} + \mathbf{W}) - h(\mathbf{W})$ . Similarly,  $I(\mathbf{Y}; V) = h(\mathbf{X} + \mathbf{W}) - h((I - A)^{1/2}\mathbf{W})$ . Now, (8) follows from Theorem 2 since

$$2^{\frac{2}{n}(h(\mathbf{X} + A^{1/2}\mathbf{W}) - h(\mathbf{X} + \mathbf{W}) + h(\mathbf{W}))} = 2^{\frac{2}{n}(h(\mathbf{Y}) - I(\mathbf{X}; V))} \quad (9)$$

$$\geq 2^{\frac{2}{n}(h(\mathbf{X}) - I(\mathbf{Y}; V))} + 2^{\frac{2}{n}h(A^{1/2}\mathbf{W}_1)} \quad (10)$$

$$= 2^{\frac{2}{n}(h(\mathbf{X}) - h(\mathbf{X} + \mathbf{W}) + h((I - A)^{1/2}\mathbf{W}))} + |A|^{1/n} 2^{\frac{2}{n}h(\mathbf{W})} \quad (11)$$

$$= |I - A|^{1/n} 2^{\frac{2}{n}(h(\mathbf{X}) - h(\mathbf{X} + \mathbf{W}) + h(\mathbf{W}))} + |A|^{1/n} 2^{\frac{2}{n}h(\mathbf{W})}. \quad (12)$$

Multiplying both sides by  $2^{\frac{2}{n}(h(\mathbf{X} + \mathbf{W}) - h(\mathbf{W}))}$  completes the proof.  $\blacksquare$

Costa's EPI may be interpreted as a concavity property enjoyed by entropy powers. The proof of Theorem 3 suggests a generalization of this property to non-Gaussian noise. Indeed, we have the following, which may be viewed as a reverse EPI:

**Theorem 4.** *Let  $\mathbf{X} \sim P_{\mathbf{X}}, \mathbf{Z} \sim P_{\mathbf{Z}}$  and  $\mathbf{W} \sim N(0, \Sigma)$  be independent,  $n$ -dimensional random vectors. Then*

$$2^{\frac{2}{n}(h(\mathbf{X} + \mathbf{W}) + h(\mathbf{Z} + \mathbf{W}))} \geq 2^{\frac{2}{n}(h(\mathbf{X}) + h(\mathbf{Z}))} + 2^{\frac{2}{n}(h(\mathbf{X} + \mathbf{Z} + \mathbf{W}) + h(\mathbf{W}))}. \quad (13)$$

*Proof:* This is an immediate consequence of Theorem 2 by putting  $V = \mathbf{X} + \mathbf{Z} + \mathbf{W}$  and rearranging exponents.  $\blacksquare$

We briefly remark that Madiman observed the following inequality on submodularity of differential entropy [20], which can be proved via data processing: if  $X, Z, W$  are independent random variables, then

$$2^{2(h(X+W) + h(Z+W))} \geq 2^{2(h(X+Z+W) + h(W))}. \quad (14)$$

When  $W$  is Gaussian, Theorem 4 sharpens inequality (14) by reducing the LHS by a factor of  $2^{2(h(X) + h(Z))}$ .

#### B. A Reverse EPI and a Refinement of Stam's Inequality

Theorem 4 admits several interesting corollaries which are deeply connected to the celebrated Gaussian Logarithmic Sobolev Inequality (LSI). To start, define the entropy power  $N(\mathbf{X})$  and the Fisher Information  $J(\mathbf{X})$  of a random vector  $\mathbf{X}$  with density  $f$  with respect to Lebesgue measure as follows:

$$N(\mathbf{X}) \triangleq \frac{1}{2\pi e} 2^{\frac{2}{n}h(\mathbf{X})} \quad J(\mathbf{X}) \triangleq \mathbb{E} \left[ \frac{\|\nabla f(\mathbf{X})\|^2}{f(\mathbf{X})} \right]. \quad (15)$$

To avoid degeneracy, we assume throughout this section that entropies and Fisher informations exist and are finite.

In exploring the similarity between the Brunn-Minkowski inequality and the EPI, Costa and Cover [21] proved the following ‘‘information isoperimetric inequality’’ for  $n$ -dimensional  $\mathbf{X}$

$$N(\mathbf{X})J(\mathbf{X}) \geq n. \quad (16)$$

This inequality is commonly referred to as Stam's inequality, due to the fact that he first observed it in his classic 1959 paper [2] in the one-dimensional case. In 1975, Gross rediscovered (16) by establishing the (mathematically equivalent) LSI for the standard Gaussian measure  $\gamma_n$  on  $\mathbb{R}^n$  [22]: For every  $h$  on  $\mathbb{R}^n$  with gradient in  $L^2(\gamma_n)$

$$\int_{\mathbb{R}^n} h^2 \log h^2 d\gamma_n \leq 2 \int_{\mathbb{R}^n} |\nabla h|^2 d\gamma_n + \left( \int_{\mathbb{R}^n} h^2 d\gamma_n \right) \log \left( \int_{\mathbb{R}^n} h^2 d\gamma_n \right). \quad (17)$$

In the same paper, Gross also proved that (17) is equivalent to the hypercontractivity of the Ornstein-Uhlenbeck semigroup [23]. It wasn't until the 1990's that Carlen [24] showed the equivalence between Stam's inequality and Gross' LSI. We refer the reader to [25] for a concise proof and further historical details.

Since (16) is proved using de Bruijn's identity and the special case of Shannon's EPI when one summand is Gaussian, Theorem 4 naturally leads to a sharpening of (16). Surprisingly, this strengthening takes the form of a reverse EPI, which upper bounds  $N(\mathbf{X} + \mathbf{Z})$  in terms of the marginal entropies and Fisher informations.

**Theorem 5.** *If  $\mathbf{X}$  and  $\mathbf{Z}$  are independent  $n$ -dimensional random vectors, then*

$$N(\mathbf{X})N(\mathbf{Z})(J(\mathbf{X}) + J(\mathbf{Z})) \geq nN(\mathbf{X} + \mathbf{Z}). \quad (18)$$

*Proof:* We may assume  $J(\mathbf{X}) < \infty$  and  $J(\mathbf{Z}) < \infty$ , else there is nothing to prove. To begin, let  $\mathbf{G} \sim N(0, I)$  be independent of  $\mathbf{X}, \mathbf{Z}$  and recall de Bruijn's identity [2]:  $\frac{d}{dt}h(\mathbf{X} + \sqrt{t}\mathbf{G}) = \frac{1}{2\ln 2}J(\mathbf{X} + \sqrt{t}\mathbf{G})$ . In particular, we have

$$\left. \frac{d}{dt}N(\mathbf{X} + \sqrt{t}\mathbf{G}) \right|_{t=0} = \frac{1}{n}N(\mathbf{X})J(\mathbf{X}). \quad (19)$$

Identifying  $\mathbf{W} = \sqrt{t}\mathbf{G}$  in Theorem 4 and rearranging, we find

$$\frac{N(\mathbf{X} + \sqrt{t}\mathbf{G})N(\mathbf{Z} + \sqrt{t}\mathbf{G}) - N(\mathbf{X})N(\mathbf{Z})}{t} \geq N(\mathbf{X} + \mathbf{Z} + \sqrt{t}\mathbf{G}) \geq N(\mathbf{X} + \mathbf{Z}). \quad (20)$$

Letting  $t \rightarrow 0$  and applying (19) proves the claim.  $\blacksquare$

It is straightforward to recover Stam's inequality from Theorem 5. Indeed, let  $\mathbf{Z} \sim N(0, \sigma^2 I)$  with variance chosen such that  $N(\mathbf{Z}) = N(\mathbf{X})$ , then (18) reduces to

$$N(\mathbf{X})J(\mathbf{X}) + N(\mathbf{Z})J(\mathbf{Z}) \geq n \frac{N(\mathbf{X} + \mathbf{Z})}{N(\mathbf{X})} \geq n \frac{2N(\mathbf{X})}{N(\mathbf{X})} = 2n, \quad (21)$$

where the second inequality follows from the EPI. Since  $N(\mathbf{Z})J(\mathbf{Z}) = n$ , (16) follows.

Stated another way, (16) reads  $\frac{1}{n}J(\mathbf{X}) \geq \frac{1}{N(\mathbf{X})}$ . Using the EPI, we may sandwich the (appropriately normalized) entropy power of the sum  $\mathbf{X} + \mathbf{Z}$  according to

$$\frac{1}{n}J(\mathbf{X}) + \frac{1}{n}J(\mathbf{Z}) \geq \frac{N(\mathbf{X} + \mathbf{Z})}{N(\mathbf{X})N(\mathbf{Z})} \geq \frac{1}{N(\mathbf{X})} + \frac{1}{N(\mathbf{Z})}, \quad (22)$$

which is met with equality throughout if  $\mathbf{X}$  and  $\mathbf{Z}$  are Gaussian with proportional covariance matrices.

Next, let  $\mathbf{X}', \mathbf{X}$  be independent and identically distributed with finite entropy, and define the *doubling constant* of  $\mathbf{X}$  (cf. [26]), denoted by  $d(\mathbf{X})$ , as

$$d(\mathbf{X}) \triangleq \frac{N\left(\frac{\mathbf{X} + \mathbf{X}'}{\sqrt{2}}\right)}{N(\mathbf{X})}. \quad (23)$$

We remark that the doubling constant and its relationship to other functionals is discussed in [26] for the one-dimensional setting, and in [27] for general dimension.

By letting  $\mathbf{Z}$  and  $\mathbf{X}$  be independent and identically distributed, Theorem 5 yields the following inequality, which expresses the deficit in (16) in terms of the doubling constant  $d(\mathbf{X})$ :

**Corollary 2.** *For any  $n$ -dimensional random vector  $\mathbf{X}$ ,*

$$N(\mathbf{X})J(\mathbf{X}) \geq n d(\mathbf{X}). \quad (24)$$

Recalling the conditions for equality in the EPI,  $d(\mathbf{X}) \geq 1$  with equality if and only if  $\mathbf{X}$  is Gaussian. Therefore, Corollary 2 represents a strict strengthening of (16). Since (16) is equivalent to the Gaussian LSI, Corollary 2 provides a bound on the deficit in the LSI. Such bounds have been of recent interest [28]–[30], and are interpreted as a stability estimate for the LSI.

Let  $\mathbf{X}$  be a random vector having density  $f$  with respect to  $\gamma_n$ . Then the LSI (17) may be written as<sup>1</sup>

$$\int_{\mathbb{R}^n} f \log f \, d\gamma_n \leq \frac{1}{2} \int_{\mathbb{R}^n} \frac{|\nabla f|^2}{f} d\gamma_n. \quad (25)$$

<sup>1</sup>In fact, this is completely equivalent to Gross' formulation (17).

Under the assumptions that  $\mathbf{X}$  is zero-mean and satisfies the Poincaré inequality

$$\zeta \mathbb{E} [s^2(\mathbf{X})] \leq \mathbb{E} [|\nabla s(\mathbf{X})|^2] \quad (26)$$

for every smooth  $s : \mathbb{R}^n \rightarrow \mathbb{R}$  such that  $\mathbb{E}[s(\mathbf{X})] = 0$ , the LSI (25) may be improved to

$$\int_{\mathbb{R}^n} f \log f \, d\gamma_n \leq \frac{c(\zeta)}{2} \int_{\mathbb{R}^n} \frac{|\nabla f|^2}{f} d\gamma_n, \quad (27)$$

where  $c(\zeta) < 1$  for ‘spectral gap’  $\zeta > 0$  [29]. In fact, by using Corollary 2 and the self-strengthening argument of [28], the constant  $c(\zeta)$  established in [29] may be improved by incorporating  $d(\mathbf{X})$ .

When  $n = 1$ ,  $\text{Var}(X) = 1$  and  $h(X) > -\infty$ , Ball, Barthe and Naor [31] showed that the Poincaré inequality (26) implies

$$d(X) \geq (N(X))^{-\frac{\zeta}{2+2\zeta}}. \quad (28)$$

Since  $N(X) \leq 1$  due to  $\text{Var}(X) = 1$ , we obtain a sharpening of Stam’s inequality:

**Corollary 3.** *Let  $X$  be a zero-mean random variable with  $\text{Var}(X) = 1$  and finite entropy. If  $X$  satisfies the Poincaré inequality (26), then*

$$(N(X))^{1+\frac{3}{2}\zeta} (J(X))^{1+\zeta} \geq 1. \quad (29)$$

On account of [31], a doubling constant  $d(X) > 1$  is a weaker assumption than presence of a spectral gap. Therefore, inequality (24) may be viewed as an improvement on the stability estimate (27) in the sense that a less restrictive hypothesis is required.

In closing, we remark that the inequality (28) also holds for  $\mathbb{R}^n$  (with  $2 + 2\zeta$  replaced by  $4 + 4\zeta$ ), provided the density of  $\mathbf{X}$  is log-concave [32]. Thus, Corollary 3 can be modified accordingly.

### C. Converse for the Two-Encoder Quadratic Gaussian Source Coding Problem

Characterizing the rate region for the two-encoder quadratic Gaussian source coding problem was a longstanding open problem in the field of network information theory until its ultimate resolution by Wagner *et al.* in their landmark paper [9], which established that a separation-based scheme [33], [34] was optimal. Wagner *et al.*’s work built upon Oohama’s earlier solution to the one-helper problem [10] and the independent solutions to the Gaussian CEO problem due to Prabhakaran, Tse and Ramachandran [19] and Oohama [18] (see [35] for a self-contained treatment). Since Wagner *et al.*’s original proof of the sum-rate constraint, other proofs have been proposed based on estimation-theoretic arguments and semidefinite programming (e.g., [36]), however all known proofs are quite complex. Below, we show that the converse result for the entire rate region is a direct consequence of Theorem 2, thus unifying the results of [9] and [10] under a common and succinct inequality.

**Theorem 6.** [9] *Let  $\mathbf{X}, \mathbf{Y} = \{X_i, Y_i\}_{i=1}^n$  be independent identically distributed pairs of jointly Gaussian random variables with correlation  $\rho$ . Let  $\phi_X : \mathbb{R}^n \rightarrow \{1, \dots, 2^{nR_X}\}$  and  $\phi_Y : \mathbb{R}^n \rightarrow \{1, \dots, 2^{nR_Y}\}$ , and define*

$$d_X \triangleq \frac{1}{n} \mathbb{E} [\|\mathbf{X} - \mathbb{E}[\mathbf{X}|\phi_X(\mathbf{X}), \phi_Y(\mathbf{Y})]\|^2] \quad (30)$$

$$d_Y \triangleq \frac{1}{n} \mathbb{E} [\|\mathbf{Y} - \mathbb{E}[\mathbf{Y}|\phi_X(\mathbf{X}), \phi_Y(\mathbf{Y})]\|^2]. \quad (31)$$

Then

$$R_X \geq \frac{1}{2} \log \left( \frac{1}{d_X} (1 - \rho^2 + \rho^2 2^{-2R_Y}) \right) \quad (32)$$

$$R_Y \geq \frac{1}{2} \log \left( \frac{1}{d_Y} (1 - \rho^2 + \rho^2 2^{-2R_X}) \right) \quad (33)$$

$$R_X + R_Y \geq \frac{1}{2} \log \frac{(1 - \rho^2)\beta(d_X d_Y)}{2d_X d_Y}, \quad (34)$$

where and  $\beta(D) \triangleq 1 + \sqrt{1 + \frac{4\rho^2 D}{(1-\rho^2)^2}}$ .

The key ingredient is the following consequence of Theorem 2:

**Proposition 1.** For  $\mathbf{X}, \mathbf{Y}$  as above,

$$2^{-\frac{2}{n}(I(\mathbf{Y};U)+I(\mathbf{X};V|U))} \geq \rho^2 2^{-\frac{2}{n}(I(\mathbf{X};U)+I(\mathbf{Y};V|U))} + 1 - \rho^2 \quad (35)$$

for any  $U, V$  satisfying  $U \rightarrow \mathbf{X} \rightarrow \mathbf{Y} \rightarrow V$ .

*Proof:* Since mutual information is invariant to scaling, we may assume without loss of generality that  $Y_i = \rho X_i + Z_i$ , where  $X_i \sim N(0, 1)$  and  $Z_i \sim N(0, 1 - \rho^2)$ , independent of  $X_i$ . Now, Theorem 2 implies

$$2^{\frac{2}{n}(h(\mathbf{Y}|U)-I(\mathbf{X};V|U))} \geq 2^{\frac{2}{n}(h(\rho\mathbf{X}|U)-I(\mathbf{Y};V|U))} + 2^{\frac{2}{n}h(\mathbf{Z})} \quad (36)$$

$$= \rho^2 2^{\frac{2}{n}(h(\mathbf{X}|U)-I(\mathbf{Y};V|U))} + (2\pi e)(1 - \rho^2). \quad (37)$$

Since  $2^{-\frac{2}{n}h(\mathbf{Y})} = 2^{-\frac{2}{n}h(\mathbf{X})} = \frac{1}{2\pi e}$ , multiplying through by  $\frac{1}{2\pi e}$  establishes the claim. ■

*Proof of Theorem 6:* For convenience, put  $U = \phi_X(\mathbf{X})$  and  $V = \phi_Y(\mathbf{Y})$ . Using the Markov relationship  $U \rightarrow \mathbf{X} \rightarrow \mathbf{Y} \rightarrow V$ , we may rearrange the exponents in Proposition 1 to obtain the equivalent inequality

$$2^{-\frac{2}{n}(I(\mathbf{X};U,V)+I(\mathbf{Y};U,V))} \geq 2^{-\frac{2}{n}I(\mathbf{X},\mathbf{Y};U,V)} \left(1 - \rho^2 + \rho^2 2^{-\frac{2}{n}I(\mathbf{X},\mathbf{Y};U,V)}\right). \quad (38)$$

The left- and right-hand sides of (38) are monotone decreasing in  $\frac{1}{n}(I(\mathbf{X};U,V)+I(\mathbf{Y};U,V))$  and  $\frac{1}{n}I(\mathbf{X},\mathbf{Y};U,V)$ , respectively. Therefore, if

$$\frac{1}{n}(I(\mathbf{X};U,V) + I(\mathbf{Y};U,V)) \geq \frac{1}{2} \log \frac{1}{D} \quad \text{and} \quad \frac{1}{n}I(\mathbf{X},\mathbf{Y};U,V) \leq R \quad (39)$$

for some pair  $(R, D)$ , then we have  $D \geq 2^{-2R} (1 - \rho^2 + \rho^2 2^{-2R})$ , which is a quadratic inequality with respect to the term  $2^{-2R}$ . This is easily solved using the quadratic formula to obtain:

$$2^{-2R} \leq \frac{2D}{(1 - \rho^2)\beta(D)} \Rightarrow R \geq \frac{1}{2} \log \frac{(1 - \rho^2)\beta(D)}{2D}, \quad (40)$$

where  $\beta(D) \triangleq 1 + \sqrt{1 + \frac{4\rho^2 D}{(1-\rho^2)^2}}$ . Note that Jensen's inequality and the maximum-entropy property of Gaussians imply  $\frac{1}{n}I(\mathbf{X};U,V) \geq \frac{1}{2} \log \frac{1}{d_X}$  and  $\frac{1}{n}I(\mathbf{Y};U,V) \geq \frac{1}{2} \log \frac{1}{d_Y}$ , so that

$$\frac{1}{n}(I(\mathbf{X};U,V) + I(\mathbf{Y};U,V)) \geq \frac{1}{2} \log \frac{1}{d_X d_Y}, \quad (41)$$

establishing (34) since  $\frac{1}{n}I(\mathbf{X},\mathbf{Y};U,V) \leq \frac{1}{n}(H(U) + H(V)) \leq R_X + R_Y$ . Similarly, Proposition 1 implies

$$2^{2R_X + \log d_X} \geq 2^{\frac{2}{n}(I(\mathbf{X};U|V)-I(\mathbf{X};U,V))} = 2^{-\frac{2}{n}I(\mathbf{X};V)} \geq (1 - \rho^2) + \rho^2 2^{-\frac{2}{n}I(\mathbf{Y};V)} \geq (1 - \rho^2) + \rho^2 2^{-2R_Y}. \quad (42)$$

Rearranging (and symmetry) yields (32)-(33). ■

**Remark 1.** Proposition 1 (a special case of Theorem 2) was first established by the author and Jiao in [37]. In fact, Proposition 1 establishes a stronger result than the converse for the two-terminal Gaussian source coding problem; it shows that the rate regions coincide for the problems when distortion is measured under quadratic loss and logarithmic loss [38], [39].



#### D. One-sided Gaussian Interference Channel

The one-sided Gaussian interference channel (IC) (or Z-Gaussian IC) is a discrete memoryless channel, with input-output relationship given by

$$Y_1 = X_1 + W \quad (43)$$

$$Y_2 = \alpha Y_1 + X_2 + W_2, \quad (44)$$

where  $X_i$  and  $Y_i$  are the channel inputs and observations corresponding to Encoder  $i$  and Decoder  $i$ , respectively, for  $i = 1, 2$ . Here,  $W \sim N(0, 1)$  and  $W_2 \sim N(1 - \alpha^2)$  are independent of each other and of the channel inputs  $X_1, X_2$ . We have assumed  $|\alpha| < 1$  since the setting where  $|\alpha| \geq 1$  is referred to as the *strong interference* regime, and the capacity is known to coincide with the Han-Kobayashi inner bound [6], [35], [40], [41]. Observe that we have expressed the one-sided Gaussian IC in *degraded form*, which has capacity region identical to the corresponding non-degraded version as proved by Costa [6]. Despite receiving significant attention from researchers over several decades, the capacity region of the one-sided Gaussian IC remains unknown in the regime of  $|\alpha| < 1$  described above.

Having already discussed connections between Costa's EPI (7) and Theorem 2 above, we remark that Costa's EPI was apparently motivated by the Gaussian IC [6]. Since Theorem 2 generalizes Costa's result, the one-sided Gaussian IC presents itself as a natural application. Toward this end, we establish a new multi-letter outer bound to give a simple demonstration of how Theorem 2 might be applied to the one-sided Gaussian IC.

**Theorem 7.**  $(R_1, R_2) \in \mathcal{C}(\alpha, P_1, P_2)$  only if

$$R_1 \leq \frac{1}{2} \log(1 + P_1) \quad (45)$$

$$R_2 \leq \frac{1}{2} \log(1 + P_2) \quad (46)$$

$$2^{-2R_2 + o(1)} \geq 2^{-\frac{2}{n}I(X_1^n, X_2^n; Y_2^n)} \sup_{V: Y_1^n \rightarrow Y_0^n \rightarrow V} \left\{ \alpha^2 2^{2R_1 - \frac{2}{n}I(Y_0^n; V|Y_1^n)} + (1 - \alpha^2) 2^{\frac{2}{n}I(Y_1^n; V)} \right\}, \quad (47)$$

for some independent  $X_1^n, X_2^n$  satisfying the power constraints  $\mathbb{E}[\|X_i^n\|^2] \leq nP_i$ ,  $i = 1, 2$ .

*Proof:* The only nontrivial inequality to prove is (47). Thus, we begin by noting that Theorem 2 implies

$$2^{\frac{2}{n}(h(Y_2^n|X_2^n) - I(Y_1^n; V|X_2^n))} \geq 2^{\frac{2}{n}(h(\alpha Y_1^n|X_2^n) - I(Y_2^n; V|X_2^n))} + 2^{\frac{2}{n}h(W_2^n|X_2^n)} \quad (48)$$

$$= \alpha^2 2^{\frac{2}{n}(h(Y_1^n) - I(Y_2^n; V|X_2^n))} + (1 - \alpha^2) 2^{\frac{2}{n}h(W^n)} \quad (49)$$

for all  $V$  such that  $Y_1^n \rightarrow Y_2^n \rightarrow V|X_2^n$ . Since  $h(W^n) = h(Y_2^n|X_1^n, X_2^n) = h(Y_1^n|X_1^n)$ ,  $I(Y_2^n; V|X_2^n) = I(Y_0^n; V, X_2^n)$  and  $I(Y_1^n; V|X_2^n) = I(Y_1^n; V, X_2^n)$ , this can be rewritten as

$$2^{-\frac{2}{n}I(X_2^n; Y_2^n) + \frac{2}{n}I(X_1^n, X_2^n; Y_2^n)} \geq \sup_{V: Y_1^n \rightarrow Y_0^n \rightarrow V} \left\{ \alpha^2 2^{\frac{2}{n}I(X_1^n; Y_1^n) - \frac{2}{n}I(Y_0^n; V|Y_1^n)} + (1 - \alpha^2) 2^{\frac{2}{n}I(Y_1^n; V)} \right\}. \quad (50)$$

Therefore,

$$2^{-2(R_2 - \epsilon_n)} \geq 2^{-\frac{2}{n}I(X_2^n; Y_2^n)} \quad (51)$$

$$\geq 2^{-\frac{2}{n}I(X_1^n, X_2^n; Y_2^n)} \sup_{V: Y_1^n \rightarrow Y_0^n \rightarrow V} \left\{ \alpha^2 2^{\frac{2}{n}I(X_1^n; Y_1^n) - \frac{2}{n}I(Y_0^n; V|Y_1^n)} + (1 - \alpha^2) 2^{\frac{2}{n}I(Y_1^n; V)} \right\} \quad (52)$$

$$\geq 2^{-\frac{2}{n}I(X_1^n, X_2^n; Y_2^n)} \sup_{V: Y_1^n \rightarrow Y_0^n \rightarrow V} \left\{ \alpha^2 2^{2(R_1 - \epsilon_n) - \frac{2}{n}I(Y_0^n; V|Y_1^n)} + (1 - \alpha^2) 2^{\frac{2}{n}I(Y_1^n; V)} \right\}, \quad (53)$$

where (51) and (53) hold for  $\epsilon_n \rightarrow 0$  due to Fano's inequality. Multiplying both sides by  $2^{2\epsilon_n}$  proves the claim. ■

The Han-Kobayashi achievable region [35], [41] evaluated for Gaussian inputs (without power control) can be expressed as the set of rate pairs  $(R_1, R_2)$  satisfying (45), (46) and

$$2^{-2R_2} \geq \frac{\alpha^2 P_2 2^{2R_1}}{(P_2 + 1 - \alpha^2)(1 + \alpha^2 P_1 + P_2)} + \frac{1 - \alpha^2}{P_2 + 1 - \alpha^2}. \quad (54)$$

Interestingly, (54) takes a similar form to (47); however, it is known that transmission without power control is suboptimal for the Gaussian Z-interference channel in general [42], [43]. Nevertheless, it may be possible to identify a random variable  $V$  in (47), possibly depending on  $X_2$ , which ultimately improves known bounds. We leave this for future work.

#### E. Relationship to Strong Data Processing

Strong data processing inequalities and their connection to hypercontractivity have garnered much recent attention [44]–[51]. For random variables  $A, B$ , the standard data processing inequality asserts that  $I(V; A) \leq I(V; B)$  for any random variable  $V$  satisfying  $A \rightarrow B \rightarrow V$ . For  $A, B \sim P_{AB}$ , it is natural to define the best-possible data processing function

$$g_I(t, P_{AB}) = \sup_{V: A \rightarrow B \rightarrow V} \{I(V; A) : I(V; B) \leq t\}, \quad (55)$$

so that  $I(V; A) \leq g_I(I(V; B), P_{AB}) \leq 1$  is the sharpest possible data processing inequality for the joint distribution  $P_{AB}$ . Thus, Theorem 1 may be rephrased as:

$$2^{2(h(Y) - g_I(t, P_{XY}))} \geq 2^{2(h(X) - t)} + 2^{2h(W)} \quad \forall t \geq 0, \quad (56)$$

where  $Y = X + W$ . Given the close relationship between the sharpened EPI (56) and strong data processing, it might be appropriate to call Theorem 1 a *strong entropy power inequality*. In any case, on rearranging, we find the following simple bound on  $g_I$  for Gaussian channels:

**Corollary 4.** *Let  $X \sim P_X$  and  $Z \sim N(0, 1)$  be independent. For  $Y = X + Z$ ,*

$$g_I(t, P_{XY}) \leq I(X; Y) - \frac{1}{2} \log \left( 1 + \frac{1}{2\pi e} 2^{2(h(X) - t)} \right). \quad (57)$$

*Moreover, for Gaussian  $X$ , the inequality (57) is an equality.*

We remark that Calmon, Polyanskiy and Wu [48], [49] have recently considered a complementary setting where they bound the best-possible data processing function defined according to

$$F_I(t, \gamma) = \sup \{I(Y; U) : I(X; U) \leq t, U \rightarrow X \rightarrow Y\}, \quad (58)$$

where  $Y = X + Z$ , and the supremum is over all  $P_{UX}$  such that  $\mathbb{E}[X^2] \leq \gamma$ .

### III. PROOF OF MAIN RESULTS

Here we give the main ideas behind proving Theorem 1. Technical details are provided in Section V and referred to as needed. For random variables  $X, Y \sim P_{XY}$ , we write  $X|Y = y$  to denote the random variable  $X$  conditional on  $\{Y = y\}$ . Note that  $X|Y = y$  is uniquely defined in the sense that different versions of the same are equal  $P_Y$ -a.e. A sequence of random variables  $X_1, X_2, \dots$  indexed by  $n \in \mathbb{N}$  will be denoted by the shorthand  $\{X_n\}$ , and convergence of  $\{X_n\}$  in distribution to a random variable  $X_*$  is written  $X_n \xrightarrow{\mathcal{D}} X_*$ .

In order to minimize the difference in inequality (3), we would like to simultaneously minimize the exponent  $h(X + W) - I(X; V)$ , while maximizing the exponent  $h(X) - I(X + W; V)$  over all valid choices of  $X, V$ . Toward this end, for a random variable  $X \sim P_X$ , let  $Y$  be defined via the additive Gaussian noise channel  $P_{Y|X}$  given by  $Y = \sqrt{\text{snr}}X + Z$ , where  $Z \sim N(0, 1)$  and define the family of functionals

$$s_\lambda(X, \text{snr}) = -h(X) + \lambda h(Y) + \inf_{V: X \rightarrow Y \rightarrow V} \{I(Y; V) - \lambda I(X; V)\} \quad (59)$$

parameterized by  $\lambda \geq 1$ . Similarly, for  $(X, Y, Q) \sim P_{XQ}P_{Y|X}$ , define the functional of  $P_{XQ}$

$$s_\lambda(X, \text{snr}|Q) = -h(X|Q) + \lambda h(Y|Q) + \inf_{V: X \rightarrow Y \rightarrow V|Q} \{I(Y; V|Q) - \lambda I(X; V|Q)\}, \quad (60)$$



and let  $\mathfrak{C}(\mathfrak{s}_\lambda(X, \text{snr}))$  denote the lower convex envelope of  $\mathfrak{s}_\lambda(\cdot, \text{snr})$  at  $X$ . That is,

$$\mathfrak{C}(\mathfrak{s}_\lambda(X, \text{snr})) = \inf_{P_{Q|X}} \mathfrak{s}_\lambda(X, \text{snr}|Q). \quad (61)$$

We consider the optimization problem

$$V_\lambda(\text{snr}) = \inf_{P_X : \mathbb{E}[X^2] \leq 1} \mathfrak{C}(\mathfrak{s}_\lambda(X, \text{snr})) = \inf_{P_{XQ} : \mathbb{E}[X^2] \leq 1} \mathfrak{s}_\lambda(X, \text{snr}|Q). \quad (62)$$

**Remark 2.** Note that, in the optimization problem (62), it suffices to consider  $Q \in \mathcal{Q}$  with  $|\mathcal{Q}| \leq 2$ . Indeed, by Fenchel-Caratheodory-Bunt [52, Theorem 18(ii)], taking  $Q$  supported on two points is sufficient to preserve the values of  $\mathbb{E}[X^2] = \sum_q p(q) \mathbb{E}[X^2|Q = q]$  and  $\mathfrak{s}_\lambda(X, \text{snr}|Q) = \sum_q p(q) \mathfrak{s}_\lambda(X, \text{snr}|Q = q)$ .

We have the following explicit characterization of  $V_\lambda(\text{snr})$ :

**Theorem 8.**

$$V_\lambda(\text{snr}) = \begin{cases} \frac{1}{2} \left[ \lambda \log \left( \frac{\lambda 2\pi e}{\lambda - 1} \right) - \log \left( \frac{2\pi e}{\lambda - 1} \right) + \log(\text{snr}) \right] & \text{if } \text{snr} \geq \frac{1}{\lambda - 1} \\ \frac{1}{2} \left[ \lambda \log(2\pi e(1 + \text{snr})) - \log(2\pi e) \right] & \text{if } \text{snr} \leq \frac{1}{\lambda - 1}. \end{cases} \quad (63)$$

The essential idea needed to establish Theorem 8 is that we only need to consider Gaussian random variables in optimization problem (62). We establish this using a weak convergence argument; the critical ingredients are proved in Sections V-C and V-D, and respectively assert:

Claim I: There exists a sequence  $\{X_n, Q_n\}$  satisfying

$$\lim_{n \rightarrow \infty} \mathfrak{s}_\lambda(X_n, \text{snr}|Q_n) = V_\lambda(\text{snr}) \quad (64)$$

$$\mathbb{E}[X_n^2] \leq 1 \quad n = 1, 2, \dots \quad (65)$$

and  $(X_n, Q_n) \xrightarrow{\mathcal{D}} (X_*, Q_*)$ , with  $X_*| \{Q_* = q\} \sim N(\mu_q, \sigma_X^2)$  for  $P_{Q_*}$ -a.e.  $q$ , with  $\sigma_X^2 \leq 1$  not depending on  $q$ .

Claim II: If  $X_n \xrightarrow{\mathcal{D}} X_* \sim N(\mu, \sigma_X^2)$  and  $\sup_n \mathbb{E}[X_n^2] < \infty$ , then

$$\liminf_{n \rightarrow \infty} \mathfrak{s}_\lambda(X_n, \text{snr}) \geq \mathfrak{s}_\lambda(X_*, \text{snr}). \quad (66)$$

In words, Claim I states that there exists a sequence  $\{X_n, Q_n\}$  which approaches the infimum of the optimization problem (62), with  $X_n$  converging weakly to Gaussian. Claim II notes that the functional  $\mathfrak{s}_\lambda(X, \text{snr})$  is weakly lower semicontinuous at Gaussian  $X$ . Combining the two claims allows us to restrict attention to Gaussian  $X$  in optimization problem (62).

With these facts in hand, the proof of Theorem 8 follows from elementary calculus and the classical EPI. We require the following proposition, which is a consequence of the conditional EPI, and a dual formulation of an inequality observed by Oohama [10].

**Proposition 2.** Let  $X \sim N(0, \gamma)$  and  $Z \sim N(0, 1)$  be independent, and define  $Y = \sqrt{\text{snr}}X + Z$ . Then for  $\lambda \geq 1$ ,

$$\inf_{V: X \rightarrow Y \rightarrow V} \left( I(Y; V) - \lambda I(X; V) \right) = \begin{cases} \frac{1}{2} \left[ \log((\lambda - 1)\gamma \text{snr}) - \lambda \log\left(\frac{\lambda - 1}{\lambda} (1 + \gamma \text{snr})\right) \right] & \text{if } \gamma \text{snr} \geq \frac{1}{\lambda - 1} \\ 0 & \text{if } \gamma \text{snr} \leq \frac{1}{\lambda - 1}. \end{cases}$$

*Proof:* Let  $V$  be such that  $X \rightarrow Y \rightarrow V$ , and let  $X| \{V = v\}, Y| \{V = v\}$  denote the random variables conditioned on  $\{V = v\}$ . Since  $X, Y$  are jointly Gaussian and  $V \rightarrow Y \rightarrow X$ , we have  $X| \{V = v\} = \rho Y| \{V = v\} + W$ , where  $\rho := \frac{\gamma \sqrt{\text{snr}}}{1 + \gamma \text{snr}}$  and  $W \sim N\left(0, \gamma - \frac{\gamma^2 \text{snr}}{1 + \gamma \text{snr}}\right)$  is independent of  $Y| \{V = v\}$ . By the entropy power inequality, it holds that

$$2^{2h(X|V=v)} \geq 2^{2h(\rho Y|V=v)} + 2^{2h(W)} = \frac{\gamma^2 \text{snr}}{(1 + \gamma \text{snr})^2} 2^{2h(Y|V=v)} + 2\pi e \left( \gamma - \frac{\gamma^2 \text{snr}}{1 + \gamma \text{snr}} \right) \quad (67)$$

which, upon applying Jensen's inequality and rearranging, yields

$$2^{-2I(X;V)} \geq \frac{1 + \gamma \text{snr} 2^{-2I(Y;V)}}{1 + \gamma \text{snr}}. \quad (68)$$

It follows that

$$I(Y;V) - \lambda I(X;V) \geq I(Y;V) + \frac{\lambda}{2} \log \left( 1 + \gamma \text{snr} 2^{-2I(Y;V)} \right) - \frac{\lambda}{2} \log(1 + \gamma \text{snr}) \quad (69)$$

$$\geq \begin{cases} \frac{1}{2} [\log((\lambda - 1)\gamma \text{snr}) - \lambda \log(\frac{\lambda-1}{\lambda}(1 + \gamma \text{snr}))] & \text{if } \gamma \text{snr} > \frac{1}{\lambda-1} \\ 0 & \text{if } \gamma \text{snr} \leq \frac{1}{\lambda-1}, \end{cases} \quad (70)$$

where the second inequality follows by minimizing over  $I(Y;V) \geq 0$ . When  $\gamma \text{snr} \leq \frac{1}{\lambda-1}$ , this is trivially achieved by setting  $V = \text{constant}$ . On the other hand, if  $\gamma \text{snr} > \frac{1}{\lambda-1}$ , then it is easy to see that the lower bound is achieved by taking  $V = Y + U$ , where  $U \sim N(0, \frac{1+\gamma \text{snr}}{\gamma \text{snr}(\lambda-1)-1})$ . ■

*Proof of Theorem 8:* Noting that  $s_\lambda(X, \text{snr})$  is invariant to translations of  $\mathbb{E}[X]$ , it follows from Claims I and II that

$$V_\lambda(\text{snr}) = \inf_{0 \leq \gamma \leq 1} s_\lambda(X_\gamma, \text{snr}), \quad \text{where } X_\gamma \sim N(0, \gamma). \quad (71)$$

Recalling the definition of  $s_\lambda(\cdot, \text{snr})$ , Proposition 2 implies

$$s_\lambda(X_\gamma, \text{snr}) = \begin{cases} \frac{1}{2} \left[ \lambda \log \left( \frac{\lambda 2\pi e}{\lambda-1} \right) - \log \left( \frac{2\pi e}{\lambda-1} \right) + \log(\text{snr}) \right] & \text{if } \gamma \text{snr} \geq \frac{1}{\lambda-1} \\ \frac{1}{2} [\lambda \log(2\pi e(1 + \gamma \text{snr})) - \log(2\pi e\gamma)] & \text{if } \gamma \text{snr} \leq \frac{1}{\lambda-1}. \end{cases} \quad (72)$$

Differentiating with respect to the quantity  $\gamma$ , we find that  $\frac{1}{2} [\lambda \log(2\pi e(1 + \gamma \text{snr})) - \log(2\pi e\gamma)]$  is decreasing in  $\gamma$  provided  $\gamma \text{snr} \leq \frac{1}{\lambda-1}$ . Therefore, taking  $\gamma = 1$  minimizes  $s_\lambda(X_\gamma, \text{snr})$  over the interval  $\gamma \in [0, 1]$ , proving the claim. ■

Given the explicit characterization of  $V_\lambda(\text{snr})$ , which is a dual form of inequality (3), we are now in a position to prove Theorem 1.

*Proof of Theorem 1:* We first establish (3) under the additional assumption that  $\mathbb{E}[X^2] < \infty$ , and generalize at the end via a truncation argument. Toward this goal, since mutual information is invariant to scaling, it is sufficient to prove that, for  $Y = \sqrt{\text{snr}}X + Z$  with  $\mathbb{E}[X^2] \leq 1$  and  $Z \sim N(0, 1)$  independent of  $X$ , we have

$$2^{2(h(Y)-I(X;V))} \geq \text{snr} 2^{2(h(X)-I(Y;V))} + 2^{2h(Z)} \quad (73)$$

for  $V$  satisfying  $X \rightarrow Y \rightarrow V$ . Multiplying both sides by  $\sigma^2$  and choosing  $\text{snr} := \frac{\text{Var}(X)}{\sigma^2}$  gives the desired inequality (3) when  $\mathbb{E}[X^2] < \infty$ . Thus, to prove (73), observe by definition of  $V_\lambda(\text{snr})$  that

$$-h(X) + I(Y;V) \geq \lambda(I(X;V) - h(Y)) + V_\lambda(\text{snr}). \quad (74)$$

Minimizing the RHS over  $\lambda$  proves the inequality. In particular, the RHS of (74) is concave in  $\lambda$ , with derivative given by

$$\frac{\partial}{\partial \lambda} \left\{ \lambda(I(X;V) - h(Y)) + V_\lambda(\text{snr}) \right\} = \begin{cases} I(X;V) - h(Y) + \frac{1}{2} \log \left( \frac{\lambda 2\pi e}{\lambda-1} \right) & \text{if } \text{snr} \geq \frac{1}{\lambda-1} \\ I(X;V) - h(Y) + \frac{1}{2} \log(2\pi e(1 + \text{snr})) & \text{if } \text{snr} < \frac{1}{\lambda-1}. \end{cases}$$

Since  $h(Y) \leq \frac{1}{2} \log(2\pi e(1 + \text{snr}))$  by the maximum entropy property of Gaussians, it follows that  $I(X;V) - h(Y) + \frac{1}{2} \log(2\pi e(1 + \text{snr})) \geq 0$ , implying that  $\frac{\partial}{\partial \lambda} \left\{ \lambda(I(X;V) - h(Y)) + V_\lambda(\text{snr}) \right\} = 0$  for  $\lambda$  satisfying  $\text{snr} \geq \frac{1}{\lambda-1}$ . In particular, the RHS of (74) is minimized when  $\lambda$  satisfies

$$\frac{\lambda}{\lambda-1} = \frac{1}{2\pi e} 2^{-2(I(X;V)-h(Y))}. \quad (75)$$

Substituting into (74) and recalling that  $2^{2h(Z)} = 2\pi e$  proves (73).

Now, we eliminate the assumption that  $\mathbb{E}[X^2] < \infty$ . Toward this end, let  $X$  have density, let  $W$  be Gaussian independent of  $X$ , and consider  $V$  satisfying  $X \rightarrow Y \rightarrow V$ , where  $Y = X + W$ . Define  $X_n$  to be the random variable  $X$  conditioned on the event  $\{|X| \leq n\}$ , let  $Y_n = X_n + W$  and define  $V_n$  via  $P_{V|Y} : Y_n \mapsto V_n$ . Since  $X_n$  is bounded,  $\mathbb{E}[X_n^2] < \infty$  so that

$$2^{2(h(Y_n) - I(X_n; V_n))} \geq 2^{2(h(X_n) - I(Y_n; V_n))} + 2^{2h(W)}. \quad (76)$$

It follows by [53, Lemma 3] that  $\lim_{n \rightarrow \infty} h(X_n) = h(X)$ , provided  $h(X)$  exists. Moreover, since  $X_n \xrightarrow{\mathcal{D}} X$ , Lemma 2 (see Section V-A) asserts that  $\lim_{n \rightarrow \infty} h(X_n + W) = h(X + W)$ , so that  $h(Y_n) \rightarrow h(Y)$ . It is easy to see that  $(X_n, V_n) \xrightarrow{\mathcal{D}} (X, V)$ , so  $\liminf_{n \rightarrow \infty} I(X_n; V_n) \geq I(X; V)$  by lower semicontinuity of relative entropy. Finally, the chain rule for mutual information implies

$$I(Y; V) + H(\mathbb{1}_{\{|X| \leq n\}}) \geq I(Y; V | \mathbb{1}_{\{|X| \leq n\}}) \geq I(Y_n; V_n) \mathbb{P}\{|X| \leq n\}, \quad (77)$$

giving  $\limsup_{n \rightarrow \infty} I(Y_n; V_n) \leq I(Y; V)$ . Putting these observations together, we have established

$$2^{2(h(Y) - I(X; V))} \geq 2^{2(h(X) - I(Y; V))} + 2^{2h(W)} \quad (78)$$

as desired.  $\blacksquare$

#### IV. EXTENSION TO RANDOM VECTORS

The vector generalization of the classical EPI is usually proved by a combination of conditioning, Jensen's inequality and induction (e.g., [35, Problem 2.9]). The same argument does not appear to readily apply in generalizing Theorem 1 to its vector version due to complications arising from the Markov constraint  $\mathbf{X} \rightarrow (\mathbf{X} + \mathbf{W}) \rightarrow V$ . However, the desired generalization may be established by noting an additivity property enjoyed by the dual form.

For a random vector  $\mathbf{X} \sim P_{\mathbf{X}}$ , let  $\mathbf{Y}$  be defined via the additive Gaussian noise channel  $\mathbf{Y} = \Gamma^{1/2} \mathbf{X} + \mathbf{Z}$ , where  $\mathbf{Z} \sim N(0, I)$  is independent of  $\mathbf{X}$  and  $\Gamma$  is a diagonal matrix with nonnegative diagonal entries. Analogous to the scalar case, define the family of functionals

$$s_{\lambda}(\mathbf{X}, \Gamma) = -h(\mathbf{X}) + \lambda h(\mathbf{Y}) + \inf_{V: \mathbf{X} \rightarrow \mathbf{Y} \rightarrow V} \left\{ I(\mathbf{Y}; V) - \lambda I(\mathbf{X}; V) \right\} \quad (79)$$

parameterized by  $\lambda \geq 1$ . Similarly, for  $(\mathbf{X}, \mathbf{Y}, Q) \sim P_{\mathbf{X}Q} P_{\mathbf{Y}|\mathbf{X}}$ , define

$$s_{\lambda}(\mathbf{X}, \Gamma | Q) = -h(\mathbf{X} | Q) + \lambda h(\mathbf{Y} | Q) + \inf_{V: \mathbf{X} \rightarrow \mathbf{Y} \rightarrow V | Q} \left\{ I(\mathbf{Y}; V | Q) - \lambda I(\mathbf{X}; V | Q) \right\}, \quad (80)$$

and consider the optimization problem

$$V_{\lambda}(\Gamma) = \inf_{P_{\mathbf{X}Q} : \mathbb{E}[X_i^2] \leq 1, i \in [n]} s_{\lambda}(\mathbf{X}, \Gamma | Q). \quad (81)$$

**Theorem 9.** *If  $\Gamma = \text{diag}(\text{snr}_1, \text{snr}_2, \dots, \text{snr}_n)$ , then*

$$V_{\lambda}(\Gamma) = \sum_{i=1}^n V_{\lambda}(\text{snr}_i). \quad (82)$$

*Proof:* Let  $\Gamma$  be a block diagonal matrix with blocks given by  $\Gamma = \text{diag}(\Gamma_1, \Gamma_2)$ . Partition  $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2)$  and  $\mathbf{Z} = (\mathbf{Z}_1, \mathbf{Z}_2)$  such that  $\mathbf{Y}_i = \Gamma_i^{1/2} \mathbf{X}_i + \mathbf{Z}_i$  for  $i = 1, 2$ . Then, for any  $V$  such that  $\mathbf{X} \rightarrow \mathbf{Y} \rightarrow V | Q$ , it follows from Lemma 10 (see Section V-C) that

$$s_{\lambda}(\mathbf{X}, \Gamma | Q) \geq s_{\lambda}(\mathbf{X}_1, \Gamma_1 | \mathbf{X}_2, Q) + s_{\lambda}(\mathbf{X}_2, \Gamma_2 | \mathbf{Y}_1, Q). \quad (83)$$

Hence,  $V_{\lambda}(\Gamma) \geq V_{\lambda}(\Gamma_1) + V_{\lambda}(\Gamma_2)$  by definition, so induction proves the claim.  $\blacksquare$

*Proof of Theorem 2:* Define  $\mathbf{Y} = \mathbf{X} + \mathbf{W}$  for convenience. As in the scalar setting, we establish the unconditional claim (where  $Q$  is constant) under the constraint  $\mathbb{E}[\|\mathbf{X}\|^2] < \infty$ . The general result follows by a

truncation argument exactly as in the scalar setting. Moreover, we may assume  $\Sigma_{\mathbf{W}} \succ 0$ , else the inequality reduces to  $h(\mathbf{Y}) + I(\mathbf{Y}; V) \geq h(\mathbf{X}) + I(\mathbf{X}; V)$ , which is trivially true by the data processing inequality and the fact that conditioning reduces entropy.

Thus, due to positive definiteness of  $\Sigma_{\mathbf{W}}$  and invariance of mutual information under one-one transformations, we may multiply both sides of (6) by  $|\Sigma_{\mathbf{W}}|^{-1/n}$  to obtain the equivalent inequality

$$2^{\frac{2}{n}(h(\Sigma_{\mathbf{W}}^{-1/2}\mathbf{Y}) - I(\Sigma_{\mathbf{W}}^{-1/2}\mathbf{X}; V))} \geq 2^{\frac{2}{n}(h(\Sigma_{\mathbf{W}}^{-1/2}\mathbf{X}) - I(\Sigma_{\mathbf{W}}^{-1/2}\mathbf{Y}; V))} + 2^{\frac{2}{n}h(\Sigma_{\mathbf{W}}^{-1/2}\mathbf{W})}. \quad (84)$$

However,  $\Sigma_{\mathbf{W}}^{-1/2}\mathbf{W} \sim N(0, I)$  and  $\mathbb{E}[\|\Sigma_{\mathbf{W}}^{-1/2}\mathbf{X}\|^2] < \infty$  provided  $\mathbb{E}[\|\mathbf{X}\|^2] < \infty$ , so we may assume without loss of generality that  $\mathbf{W} \sim N(0, I)$  in establishing the unconditional version of (6).

To simplify further, put  $\text{snr} := \max_{1 \leq i \leq n} \mathbb{E}[X_i^2]$ . Note that we may assume  $\text{snr} > 0$ , else the claimed inequality is trivial since  $h(\mathbf{X}) = -\infty$  and  $h(\mathbf{Y}) - I(\mathbf{X}; V) \geq h(\mathbf{Y}) - I(\mathbf{X}; \mathbf{Y}) = h(\mathbf{W})$ . Therefore, (6) is equivalent to

$$2^{\frac{2}{n}(h(\mathbf{Y}) - I(\mathbf{X}; V))} \geq \text{snr} 2^{\frac{2}{n}(h(\mathbf{X}) - I(\mathbf{Y}; V))} + 2^{\frac{2}{n}h(\mathbf{Z})} \quad (85)$$

holding for  $\mathbf{X} \rightarrow \mathbf{Y} \rightarrow V$ , where  $\mathbf{Y} = \sqrt{\text{snr}}\mathbf{X} + \mathbf{Z}$ ,  $\mathbf{Z} \sim N(0, I)$  is independent of  $\mathbf{X}$ , and  $\max_{1 \leq i \leq n} \mathbb{E}[X_i^2] \leq 1$ . This is established exactly as in the proof of Theorem 1, since  $V_{\lambda}(\text{snr} \cdot I) = nV_{\lambda}(\text{snr})$ . ■

## V. PROOF OF CLAIMS I AND II

This section is dedicated to the proof of Claims I and II of Section III. Several of the steps in the proof require properties and characterizations of Gaussian random variables, which are recalled and proved as needed in the first two subsections. The third subsection is dedicated to the proof of Claim I, and the fourth subsection is dedicated to the proof of Claim II.

### A. Properties of Gaussian Perturbation

We collect below a few facts about random variables that are contaminated by Gaussian noise. Of particular interest to us will be weakly convergent sequences of random variables, and corresponding continuity properties under perturbation by Gaussian noise.

**Lemma 1.** [54, Lemma 5.1.3] *If  $X, Z$  are independent random variables and  $Z$  is normal, then  $X + Z$  has a non-vanishing probability density function which has derivatives of all orders.*

**Lemma 2.** [55, Propositions 16 and 18] *Let  $\mathbf{X}_n \xrightarrow{\mathcal{D}} \mathbf{X}_*$  with  $\sup_n \mathbb{E}[\|\mathbf{X}_n\|^2] < \infty$ , and let  $\mathbf{Z} \sim N(0, \sigma^2 I)$  be a non-degenerate Gaussian, independent of  $\{\mathbf{X}_n\}, \mathbf{X}_*$ . Let  $\mathbf{Y}_n = \mathbf{X}_n + \mathbf{Z}$  and  $\mathbf{Y}_* = \mathbf{X}_* + \mathbf{Z}$ . Finally, let  $f_n(\mathbf{y})$  and  $f_*(\mathbf{y})$  denote the density of  $\mathbf{Y}_n$  and  $\mathbf{Y}_*$ , respectively. Then*

1.  $\mathbf{Y}_n \xrightarrow{\mathcal{D}} \mathbf{Y}_*$
2.  $\|f_n(\mathbf{y}) - f_*(\mathbf{y})\|_{\infty} \rightarrow 0$
3.  $h(\mathbf{Y}_n) \rightarrow h(\mathbf{Y}_*)$ .

**Lemma 3.** *Suppose  $(X_{1,n}, X_{2,n}) \xrightarrow{\mathcal{D}} (X_{1,*}, X_{2,*})$  with  $\sup_n \mathbb{E}[X_{i,n}^2] < \infty$  for  $i = 1, 2$ . Let  $(Z_1, Z_2) \sim N(0, \sigma^2 I)$  be pairwise independent of  $(X_{1,n}, X_{2,n})$  and  $(X_{1,*}, X_{2,*})$ , and, for  $i = 1, 2$ , define  $Y_{i,n} = X_{i,n} + Z_i$  and  $Y_{i,*} = X_{i,*} + Z_i$ . Then  $(Y_{1,n}, Y_{2,n}) \xrightarrow{\mathcal{D}} (Y_{1,*}, Y_{2,*})$  and*

$$\liminf_{n \rightarrow \infty} I(X_{1,n}; X_{2,n} | Y_{1,n}, Y_{2,n}) \geq I(X_{1,*}; X_{2,*} | Y_{1,*}, Y_{2,*}). \quad (86)$$

*Proof:* The fact that  $(Y_{1,n}, Y_{2,n}) \xrightarrow{\mathcal{D}} (Y_{1,*}, Y_{2,*})$  follows from Lemma 2. Lemma 2 also establishes that

$$h(Y_{1,n}, Y_{2,n}) \rightarrow h(Y_{1,*}, Y_{2,*}). \quad (87)$$

On account of the Markov chains  $(X_{2,n}, Y_{2,n}) \rightarrow X_{1,n} \rightarrow Y_{1,n}$  and  $(X_{1,n}, Y_{1,n}) \rightarrow X_{2,n} \rightarrow Y_{2,n}$ , we have the identity

$$I(X_{1,n}; X_{2,n} | Y_{1,n}, Y_{2,n}) = I(X_{1,n}, Y_{2,n}; Y_{1,n}, X_{2,n}) - I(X_{1,n}, X_{2,n}; Y_{1,n}, Y_{2,n}). \quad (88)$$

Observe that  $\liminf_{n \rightarrow \infty} I(X_{1,n}, Y_{2,n}; Y_{1,n}, X_{2,n}) \geq I(X_{1,*}, Y_{2,*}; Y_{1,*}, X_{2,*})$  due to lower semicontinuity of relative entropy, and  $\lim_{n \rightarrow \infty} I(X_{1,n}, X_{2,n}; Y_{1,n}, Y_{2,n}) = I(X_{1,*}, X_{2,*}; Y_{1,*}, Y_{2,*})$  due to (87) and the fact that  $h(Y_{1,*}, Y_{2,*} | X_{1,*}, X_{2,*}) = h(Y_{1,n}, Y_{2,n} | X_{1,n}, X_{2,n}) = h(Z_1, Z_2)$  is constant. Thus, (86) is proved by applying the identity (88) again for  $(X_{1,*}, X_{2,*}, Y_{1,*}, Y_{2,*})$ . ■

**Lemma 4.** Let  $\{Y_n\}, Y_*$  be as in Lemma 2. Fix  $b > 0$  and a channel  $P_{V|Y}$ , and define  $\{V_n\}, V_*$  according to  $P_{V|Y} : Y_n \mapsto V_n$  and  $P_{V|Y} : Y_* \mapsto V_*$ . There exists a sequence  $\{\epsilon_n\}$  depending on  $b$  and  $\{Y_n\}$ , but not on  $P_{V|Y}$ , satisfying  $\lim_{n \rightarrow \infty} \epsilon_n = 0$  and

$$I(V_n; Y_n | |Y_n| \leq b) \leq (1 + \epsilon_n)^2 I(V_*; Y_* | |Y_*| \leq b) - (1 + \epsilon_n)^2 \log(1 - \epsilon_n)^2, \quad (89)$$

$$I(V_n; Y_n | |Y_n| \leq b) \geq (1 - \epsilon_n)^2 I(V_*; Y_* | |Y_*| \leq b) - (1 - \epsilon_n)^2 \log(1 + \epsilon_n)^2, \quad (90)$$

$$\left| \frac{\mathbb{P}(|Y_n| \leq b)}{\mathbb{P}(|Y_*| \leq b)} - 1 \right| \leq \epsilon_n. \quad (91)$$

*Proof:* Let  $f_n(y)$  and  $f_*(y)$  denote the density of  $Y_n$  and  $Y_*$ , respectively. By Lemma 1, the density  $f_*$  is continuous and does not vanish, and is therefore bounded away from zero on the interval  $B = [-b, b]$ . By Lemma 2,  $\|f_n(y) - f_*(y)\|_\infty \rightarrow 0$ , so it follows that

$$\sup_{y \in B} \left| 1 - \frac{f_n(y)}{f_*(y)} \right| \leq \epsilon_n \quad \text{and} \quad \sup_{y \in B} \left| 1 - \frac{f_*(y)}{f_n(y)} \right| \leq \epsilon_n \quad (92)$$

for some  $\epsilon_n \rightarrow 0$  as  $n \rightarrow \infty$  (note that  $\epsilon_n$  does not depend on  $P_{V|Y}$ ). As a consequence,

$$\mathbb{P}(Y_* \in B) = \int_B f_*(y) dy \leq (1 + \epsilon_n) \int_B f_n(y) dy = (1 + \epsilon_n) \mathbb{P}(Y_n \in B). \quad (93)$$

Hence, for  $y \in B$ , the conditional densities of  $Y_n | \{Y_n \in B\}$  and  $Y_* | \{Y_* \in B\}$  satisfy

$$\frac{f_{Y_n | \{Y_n \in B\}}(y)}{f_{Y_* | \{Y_* \in B\}}(y)} = \frac{f_n(y)}{f_*(y)} \cdot \frac{\mathbb{P}(Y_* \in B)}{\mathbb{P}(Y_n \in B)} \leq (1 + \epsilon_n)^2. \quad (94)$$

By a symmetric argument,  $f_{Y_n | \{Y_n \in B\}}(y) \geq (1 - \epsilon_n)^2 f_{Y_* | \{Y_* \in B\}}(y)$  for all  $y \in B$ . Therefore, for any Borel set  $A$ ,

$$\mathbb{P}(V_n \in A | Y_n \in B) = \int_B \int_A P_{V|Y=y}(dv) f_{Y_n | \{Y_n \in B\}}(y) dy \geq (1 - \epsilon_n)^2 \int_B \int_A P_{V|Y=y}(dv) f_{Y_* | \{Y_* \in B\}}(y) dy \quad (95)$$

$$= (1 - \epsilon_n)^2 \mathbb{P}(V_* \in A | Y_* \in B). \quad (96)$$

As a consequence,  $\frac{dP_{V_n | Y_n \in B}}{dP_{V_* | Y_* \in B}}(v) \geq (1 - \epsilon_n)^2$ . Combining the above observations we have

$$I(V_n; Y_n | Y_n \in B) = \int_B \int f_{Y_n | \{Y_n \in B\}}(y) \log \left( \frac{dP_{V|Y=y}}{dP_{V_n | Y_n \in B}}(v) \right) P_{V|Y=y}(dv) dy \quad (97)$$

$$\leq (1 + \epsilon_n)^2 \int_B \int f_{Y_* | \{Y_* \in B\}}(y) \log \left( \frac{1}{(1 - \epsilon_n)^2} \frac{dP_{V|Y=y}}{dP_{V_* | Y_* \in B}}(v) \right) P_{V|Y=y}(dv) dy \quad (98)$$

$$= (1 + \epsilon_n)^2 I(V_*; Y_* | Y_* \in B) - (1 + \epsilon_n)^2 \log(1 - \epsilon_n)^2. \quad (99)$$

By a symmetric argument, we also have

$$I(V_n; Y_n | Y_n \in B) \geq (1 - \epsilon_n)^2 I(V_*; Y_* | Y_* \in B) - (1 - \epsilon_n)^2 \log(1 + \epsilon_n)^2, \quad (100)$$

which proves (89)-(90). Inequality (91) is established by the same logic as (93). ■

**Lemma 5.** Let  $X \sim P_X$  and let  $Z \sim N(0, \sigma^2)$  be a non-degenerate Gaussian, independent of  $X$ . It holds that

$$\lim_{b \rightarrow \infty} \mathbb{P}(|X| > b) I(X; X + Z | |X| > b) = 0. \quad (101)$$

*Proof:* The proof follows that of [56, Theorem 6]. By lower semicontinuity of relative entropy, we have

$$\liminf_{b \rightarrow \infty} I(X; X + Z \mid |X| \leq b) \geq I(X; X + Z). \quad (102)$$

Also,

$$I(X; X + Z) \geq \mathbb{P}(|X| \leq b)I(X; X + Z \mid |X| \leq b), \quad (103)$$

so that  $\lim_{b \rightarrow \infty} I(X; X + Z \mid |X| \leq b) = \lim_{b \rightarrow \infty} \mathbb{P}(|X| \leq b)I(X; X + Z \mid |X| \leq b) = I(X; X + Z)$ . By the chain rule

$$\mathbb{P}(|X| > b)I(X; X + Z \mid |X| > b) = I(X; X + Z) - I(\mathbb{1}_{\{|X| \leq b\}}; X + Z) - \mathbb{P}(|X| \leq b)I(X; X + Z \mid |X| \leq b),$$

so the claim is proved since  $I(\mathbb{1}_{\{|X| \leq b\}}; X + Z)$  vanishes as  $b \rightarrow \infty$ .  $\blacksquare$

### B. Characterizations of Gaussian Random Variables

The goal of this subsection is to establish the following characterization of Gaussian random variables:

**Lemma 6.** Suppose  $(X_{1,n}, X_{2,n}) \xrightarrow{\mathcal{D}} (X_{1,*}, X_{2,*})$  with  $\sup_n \mathbb{E}[X_{i,n}^2] < \infty$  for  $i = 1, 2$ . Let  $(Z_1, Z_2) \sim N(0, \sigma^2 I)$  be pairwise independent of  $(X_{1,n}, X_{2,n})$  and, for  $i = 1, 2$ , define  $Y_{i,n} = X_{i,n} + Z_i$ . If  $X_{1,n}, X_{2,n}$  are independent and

$$\liminf_{n \rightarrow \infty} I(X_{1,n} + X_{2,n}; X_{1,n} - X_{2,n} \mid Y_{1,n}, Y_{2,n}) = 0, \quad (104)$$

then  $X_{1,*}, X_{2,*}$  are independent Gaussian random variables with identical variances.

We require two facts. First, a fundamental result of Bernstein [57] asserts the following:

**Lemma 7.** [54, Theorem 5.1.1] If  $X_1, X_2$  are independent random variables such that  $X_1 + X_2$  and  $X_1 - X_2$  are independent, then  $X_1$  and  $X_2$  are normal, with identical variances.

**Remark 3.** Formally, Bernstein's theorem does not comment on the identical variances of  $X_1, X_2$ . However, assuming without loss of generality that  $X_1, X_2$  are zero-mean, the observation that  $X_1$  and  $X_2$  have identical variances is immediate since  $\mathbb{E}[X_1^2] - \mathbb{E}[X_2^2] = \mathbb{E}[(X_1 - X_2)(X_1 + X_2)] = 0$ . This fact was explicitly noted by Geng and Nair [55].

Second, we will need the following observation:

**Lemma 8.** Let  $Y = X + Z$ , where  $Z \sim N(0, \sigma^2)$  is a non-degenerate Gaussian, independent of  $X$ . If  $X \mid \{Y = y\}$  is normal for  $P_Y$ -a.e.  $y$ , with variance  $\sigma_X^2$  not depending on  $y$ , then  $X$  is normal with variance  $\frac{\sigma^2 \sigma_X^2}{\sigma^2 - \sigma_X^2}$ .

*Proof:* If  $X \mid \{Y = y\}$  is normal for  $P_Y$ -a.e.  $y$  with variance  $\sigma_X^2$  not depending on  $Y$ , then  $X = \mathbb{E}[X \mid Y] + W$  a.s., where  $W \sim N(0, \sigma_X^2)$  is independent of  $Y$ . In particular,  $X$  has density  $f_X$  by Lemma 1. Also, by Lemma 1,  $Y$  has density  $f_Y$ . The conditional density  $f_{Y \mid X}$  exists and is Gaussian by definition, and  $f_{X \mid Y}$  is a valid Gaussian density for  $P_Y$ -a.e.  $y$ , with corresponding variance  $\sigma_X^2$  not depending on  $y$ . Thus, we have

$$\log f_X(x) = \log f_Y(y) + \log f_{Y \mid X}(y \mid x) - \log f_{X \mid Y}(x \mid y). \quad (105)$$

The key observation is that the RHS of (105) is a quadratic function in  $x$ . Since  $f_X$  is a density and must integrate to unity, it must therefore be Gaussian. Direct computation reveals that  $X$  has variance  $\frac{\sigma^2 \sigma_X^2}{\sigma^2 - \sigma_X^2}$ .  $\blacksquare$

*Proof of Lemma 6:* Let  $Y_{i,*}$  be as in the statement of Lemma 3, and recall that the same lemma asserts  $(Y_{1,n}, Y_{2,n}) \xrightarrow{\mathcal{D}} (Y_{1,*}, Y_{2,*})$ . By definition of  $Z_1, Z_2$ , the random variables  $(Z_1 + Z_2)$  and  $(Z_1 - Z_2)$  are independent and Gaussian with respective variances  $2\sigma^2$ . Thus, noting that assumption (104) is equivalent to

$$\liminf_{n \rightarrow \infty} I(X_{1,n} + X_{2,n}; X_{1,n} - X_{2,n} \mid Y_{1,n} + Y_{2,n}, Y_{1,n} - Y_{2,n}) = 0, \quad (106)$$



we may apply Lemma 3 to the sequences  $\{X_{1,n} + X_{2,n}, X_{1,n} - X_{2,n}\}$  and  $\{Y_{1,n} + Y_{2,n}, Y_{1,n} - Y_{2,n}\}$  to obtain

$$I(X_{1,*} + X_{2,*}; X_{1,*} - X_{2,*} | Y_{1,*}, Y_{2,*}) = I(X_{1,*} + X_{2,*}; X_{1,*} - X_{2,*} | Y_{1,*} + Y_{2,*}, Y_{1,*} - Y_{2,*}) = 0. \quad (107)$$

Using independence of  $X_{1,n}, X_{2,n}$ , Lemma 3 applied directly yields

$$I(X_{1,*}; X_{2,*} | Y_{1,*}, Y_{2,*}) = 0. \quad (108)$$

In particular, for  $P_{Y_{1,*}Y_{2,*}}$ -a.e.  $y_1, y_2$ , the random variables  $X_{1,*} | \{Y_{1,*}, Y_{2,*} = y_1, y_2\}$  and  $X_{2,*} | \{Y_{1,*}, Y_{2,*} = y_1, y_2\}$  are independent, and  $(X_{1,*} + X_{2,*}) | \{Y_{1,*}, Y_{2,*} = y_1, y_2\}$  and  $(X_{1,*} - X_{2,*}) | \{Y_{1,*}, Y_{2,*} = y_1, y_2\}$  are independent. Therefore, Lemma 7 implies that  $X_{1,*} | \{Y_{1,*}, Y_{2,*} = y_1, y_2\}$  and  $X_{2,*} | \{Y_{1,*}, Y_{2,*} = y_1, y_2\}$  are normal with identical variances. Starting with the third claim of Lemma 2 and applying lower semicontinuity of relative entropy, we observe

$$I(X_{1,*}; Y_{1,*}) = \lim_{n \rightarrow \infty} I(X_{1,n}; Y_{1,n}) = \lim_{n \rightarrow \infty} I(X_{1,n}; Y_{1,n}, Y_{2,n}) \quad (109)$$

$$\geq I(X_{1,*}; Y_{1,*}, Y_{2,*}) \quad (110)$$

$$= I(X_{1,*}; Y_{1,*}) + I(X_{1,*}; Y_{2,*} | Y_{1,*}), \quad (111)$$

so it follows that  $X_{1,*} \rightarrow Y_{1,*} \rightarrow Y_{2,*}$ , and therefore  $X_{1,*} | \{Y_{1,*}, Y_{2,*} = y_1, y_2\} \sim X_{1,*} | \{Y_{1,*} = y_1\}$ . Similarly,  $X_{2,*} | \{Y_{1,*}, Y_{2,*} = y_1, y_2\} \sim X_{2,*} | \{Y_{2,*} = y_2\}$ . So, we may conclude that the random variables  $X_{1,*} | \{Y_{1,*} = y_1\}$  and  $X_{2,*} | \{Y_{2,*} = y_2\}$  are normal, with identical variances not depending on  $y_1, y_2$ . Invoking Lemma 8, we find that both  $X_{1,*}$  and  $X_{2,*}$  are normal with identical variances, completing the proof. ■

### C. Existence of sequences satisfying $\lim_{n \rightarrow \infty} s_\lambda(X_n, \text{snr} | Q_n) = V_\lambda(\text{snr})$ that converge weakly to Gaussian

The goal of this section is to prove the following result, which was the first essential ingredient needed for the proof of Theorem 8 (i.e., Claim I).

**Lemma 9.** *There exists a sequence  $\{X_n, Q_n\}$  satisfying*

$$\lim_{n \rightarrow \infty} s_\lambda(X_n, \text{snr} | Q_n) = V_\lambda(\text{snr}) \quad (112)$$

$$\mathbb{E}[X_n^2] \leq 1 \quad n \geq 1 \quad (113)$$

and  $(X_n, Q_n) \xrightarrow{\mathcal{D}} (X_*, Q_*)$ , with  $X_* | \{Q_* = q\} \sim N(\mu_q, \sigma_X^2)$  for  $P_{Q_*}$ -a.e.  $q$ , with  $\sigma_X^2 \leq 1$  not depending on  $q$ .

A rough outline of the proof is as follows: We first establish a superadditivity property of  $s_\lambda(X, \text{snr} | Q)$ , and then exploit this property in conjunction with the characterization of Gaussians proved in Lemma 6 to verify the existence of sequence  $\{X_n, Q_n\}$  satisfying  $\lim_{n \rightarrow \infty} s_\lambda(X_n, \text{snr} | Q_n) = V_\lambda(\text{snr})$  which converges weakly to Gaussian. We begin with a straightforward observation:

**Lemma 10.** *Let  $\mathbf{X} = (X_1, X_2)$ ,  $\mathbf{Y} = (Y_1, Y_2)$ , and  $Q$  have joint distribution  $P_{\mathbf{X}\mathbf{Y}Q} = P_{X_1X_2Q}P_{Y_1|X_1}P_{Y_2|X_2}$ . If  $V$  satisfies  $\mathbf{X} \rightarrow \mathbf{Y} \rightarrow V | Q$ , then for  $\lambda \geq 1$ , we have*

$$\begin{aligned} & I(Y_1, Y_2; V | Q) - h(X_1, X_2 | Q) - \lambda (I(X_1, X_2; V | Q) - h(Y_1, Y_2 | Q)) \\ & \geq I(Y_1; V | X_2, Q) - h(X_1 | X_2, Q) - \lambda (I(X_1; V | X_2, Q) - h(Y_1 | X_2, Q)) \\ & \quad + I(Y_2; V | Y_1, Q) - h(X_2 | Y_1, Q) - \lambda (I(X_2; V | Y_1, Q) - h(Y_2 | Y_1, Q)). \end{aligned} \quad (114)$$

Moreover,  $X_1 \rightarrow Y_1 \rightarrow V | (X_2, Q)$  and  $X_2 \rightarrow Y_2 \rightarrow V | (Y_1, Q)$ .

*Proof:* The second claim is straightforward. Indeed, using  $P_{\mathbf{X}\mathbf{Y}Q} = P_{X_1X_2Q}P_{Y_1|X_1}P_{Y_2|X_2}$ , we can factor the joint distribution of  $(\mathbf{X}, \mathbf{Y}, V, Q)$  as  $P_{\mathbf{X}\mathbf{Y}VQ} = P_{X_1X_2Q}P_{Y_1|X_1}P_{Y_2|X_2}P_{V|Y_1Y_2Q} = P_{X_1X_2Q}P_{Y_1|X_1}P_{Y_2V|Y_1X_2Q}$ . Marginalizing over  $Y_2$ , we find that  $X_1 \rightarrow Y_1 \rightarrow V | (X_2, Q)$ . The symmetric Markov chain follows similarly by writing  $P_{\mathbf{X}\mathbf{Y}VQ} = P_{X_1X_2Y_1Q}P_{Y_2|X_2}P_{V|Y_1Y_2Q}$  and marginalizing over  $X_1$ .

To prove the claimed inequality, note the following identities:

$$\begin{aligned} I(Y_1, Y_2; V|Q) - h(X_1, X_2|Q) \\ = I(Y_1; V|Q) + I(Y_2; V|Q, Y_1) - h(X_2|Q) - h(X_1|Q, X_2) \end{aligned} \quad (115)$$

$$= I(Y_1; V|Q) + I(Y_2; V|Q, Y_1) - h(X_2|Q, Y_1) - h(X_1|Q, X_2) - I(X_2; Y_1|Q) \quad (116)$$

$$= I(Y_1; V|Q, X_2) + I(Y_2; V|Q, Y_1) - h(X_2|Q, Y_1) - h(X_1|Q, X_2) - I(X_2; Y_1|Q, V), \quad (117)$$

and

$$\begin{aligned} I(X_1, X_2; V|Q) - h(Y_1, Y_2|Q) \\ = I(X_2; V|Q) + I(X_1; V|Q, X_2) - h(Y_1|Q) - h(Y_2|Q, Y_1) \end{aligned} \quad (118)$$

$$= I(X_2; V|Q) + I(X_1; V|Q, X_2) - h(Y_1|Q, X_2) - h(Y_2|Q, Y_1) - I(X_2; Y_1|Q) \quad (119)$$

$$= I(X_2; V|Q, Y_1) + I(X_1; V|Q, X_2) - h(Y_1|Q, X_2) - h(Y_2|Q, Y_1) - I(X_2; Y_1|Q, V). \quad (120)$$

Therefore,

$$I(Y_1, Y_2; V|Q) - h(X_1, X_2|Q) - \lambda (I(X_1, X_2; V|Q) - h(Y_1, Y_2|Q)) \quad (121)$$

$$\begin{aligned} = & I(Y_1; V|X_2, Q) - h(X_1|X_2, Q) - \lambda (I(X_1; V|X_2, Q) - h(Y_1|X_2, Q)) \\ & + I(Y_2; V|Y_1, Q) - h(X_2|Y_1, Q) - \lambda (I(X_2; V|Y_1, Q) - h(Y_2|Y_1, Q)) \\ & + (\lambda - 1)I(X_2; Y_1|V, Q), \end{aligned} \quad (122)$$

which proves the inequality (114) since  $\lambda \geq 1$ . ■

Lemma 10 leads to the desired superadditivity property of  $s_\lambda(X, \text{snr}|Q)$ :

**Lemma 11.** Let  $P_{Y|X}$  be the Gaussian channel  $Y = \sqrt{\text{snr}}X + Z$ , where  $Z \sim N(0, 1)$  is independent of  $X$ . Now, suppose  $(X, Y, Q) \sim P_{XQ}P_{Y|X}$ , and let  $(X_1, Y_1, Q_1)$  and  $(X_2, Y_2, Q_2)$  denote two independent copies of  $(X, Y, Q)$ . Define

$$X_+ = \frac{X_1 + X_2}{\sqrt{2}}, \quad X_- = \frac{X_1 - X_2}{\sqrt{2}}, \quad (123)$$

and in a similar manner, define  $Y_+, Y_-$ . Letting  $\mathbf{Q} = (Q_1, Q_2)$ , we have for  $\lambda \geq 1$

$$2s_\lambda(X, \text{snr}|Q) \geq s_\lambda(X_+, \text{snr}|X_-, \mathbf{Q}) + s_\lambda(X_-, \text{snr}|Y_+, \mathbf{Q}) \quad (124)$$

and

$$2s_\lambda(X, \text{snr}|Q) \geq s_\lambda(X_+, \text{snr}|Y_-, \mathbf{Q}) + s_\lambda(X_-, \text{snr}|X_+, \mathbf{Q}). \quad (125)$$

*Proof:* The crucial observation is that the unitary transformation  $(Y_1, Y_2) \mapsto (Y_+, Y_-)$  preserves the Gaussian nature of the channel. That is, if  $Y_i = \sqrt{\text{snr}}X_i + Z_i$ , then  $Y_+ = \sqrt{\text{snr}}X_+ + \frac{1}{\sqrt{2}}(Z_1 + Z_2)$  and  $Y_- = \sqrt{\text{snr}}X_- + \frac{1}{\sqrt{2}}(Z_1 - Z_2)$ , where the pair  $(\frac{1}{\sqrt{2}}(Z_1 + Z_2), \frac{1}{\sqrt{2}}(Z_1 - Z_2))$  is equal in distribution to  $(Z_1, Z_2)$ .

Thus, consider an arbitrary  $V$  satisfying  $(X_+, X_-) \rightarrow (Y_+, Y_-) \rightarrow V|\mathbf{Q}$ . By Lemma 10 and the above observation, we have

$$\begin{aligned} I(Y_1, Y_2; V|\mathbf{Q}) - h(X_1, X_2|\mathbf{Q}) - \lambda (I(X_1, X_2; V|\mathbf{Q}) - h(Y_1, Y_2|\mathbf{Q})) \\ = I(Y_+, Y_-; V|\mathbf{Q}) - h(X_+, X_-|\mathbf{Q}) - \lambda (I(X_+, X_-; V|\mathbf{Q}) - h(Y_+, Y_-|\mathbf{Q})) \end{aligned} \quad (126)$$

$$\geq I(Y_+; V|X_-, \mathbf{Q}) - h(X_+|X_-, \mathbf{Q}) - \lambda (I(X_+; V|X_-, \mathbf{Q}) - h(Y_+|X_-, \mathbf{Q})) \quad (127)$$

$$\begin{aligned} & + I(Y_-; V|Y_+, \mathbf{Q}) - h(X_-|Y_+, \mathbf{Q}) - \lambda (I(X_-; V|Y_+, \mathbf{Q}) - h(Y_-|Y_+, \mathbf{Q})) \\ & \geq s_\lambda(X_+, \text{snr}|X_-, \mathbf{Q}) + s_\lambda(X_-, \text{snr}|Y_+, \mathbf{Q}). \end{aligned} \quad (128)$$

This proves (124) since

$$\inf_{V: \mathbf{X} \rightarrow (\mathbf{Y}, \mathbf{Q}) \rightarrow V} I(Y_1, Y_2; V | \mathbf{Q}) - h(X_1, X_2 | \mathbf{Q}) - \lambda (I(X_1, X_2; V | \mathbf{Q}) - h(Y_1, Y_2 | \mathbf{Q})) \quad (129)$$

$$\leq \sum_{i=1}^2 \inf_{V: X_i \rightarrow Y_i \rightarrow V | Q_i} I(Y_i; V | Q_i) - h(X_i | Q_i) - \lambda (I(X_i; V | Q_i) - h(Y_i | Q_i)) \quad (130)$$

$$= 2s_\lambda(X, \text{snr} | Q), \quad (131)$$

where the inequality follows since the infimum is taken over a smaller set.  $\blacksquare$

**Remark 4.** In some sense, Lemma 11 is the key to the whole proof. The subadditivity property ultimately implies that the optimizing distribution in optimization problem (62) is rotationally invariant, and therefore Gaussian. This idea was introduced to the information theory literature by Geng and Nair [55], but has origins in a ‘doubling trick’ which has been used to great success in the literature on functional inequalities [24], [58] and has been attributed to K. Ball [59]. The reader is referred to [60], [61] for a detailed discussion of the duality between extremisation of information measures and functional inequalities.

We are now ready to prove Lemma 9.

*Proof of Lemma 9:* For convenience, we will refer to any sequence  $\{X_n, Q_n\}$  satisfying (112)-(113) as *admissible*. Since  $s_\lambda(X_n, \text{snr} | Q_n)$  is invariant to translations of the mean of  $X_n$ , we may restrict our attention to admissible sequences satisfying  $\mathbb{E}[X_n] = 0$  without any loss of generality.

Begin by letting  $\{X_n, Q_n\}$  be an admissible sequence with the property that

$$\lim_{n \rightarrow \infty} (h(Y_n | Q_n) - h(X_n | Q_n)) \leq \liminf_{n \rightarrow \infty} (h(Y'_n | Q'_n) - h(X'_n | Q'_n)) \quad (132)$$

for any other admissible sequence  $\{X'_n, Q'_n\}$ . Clearly, such a sequence can always be constructed by a diagonalization argument. Moreover, the LHS of (132) must be finite. To see this, note first that  $h(Y_n | Q_n) - h(X_n | Q_n) \geq 0$  since conditioning reduces entropy. On the other hand,  $s_\lambda(X_n, \text{snr} | Q_n) < V_\lambda(\text{snr}) + 1$  for  $n$  sufficiently large. Hence, there is some  $V_n$  satisfying  $X_n \rightarrow Y_n \rightarrow V_n | Q_n$  for which

$$h(Y_n | Q_n) - h(X_n | Q_n) \leq V_\lambda(\text{snr}) + 1 + \lambda I(X_n; V_n | Q_n) - I(Y_n; V_n | Q_n) - (\lambda - 1)h(Y_n | Q_n) \quad (133)$$

$$\leq V_\lambda(\text{snr}) + 1 + (\lambda - 1)I(X_n; V_n | Q_n) - (\lambda - 1)h(Y_n | Q_n) \quad (134)$$

$$\leq V_\lambda(\text{snr}) + 1 + (\lambda - 1)I(X_n; Y_n | Q_n) - (\lambda - 1)h(Y_n | Q_n) \quad (135)$$

$$= V_\lambda(\text{snr}) + 1 - (\lambda - 1)h(Y_n | X_n), \quad (136)$$

where (134) and (135) are both due to the data processing inequality. Since  $V_\lambda(\text{snr}) < \infty$  trivially and  $h(Y_n | X_n) = h(Z) = \frac{1}{2} \log 2\pi e$ , we conclude that the LHS of (132) is finite as claimed.

By the same logic as in the remark following (62), we may assume that  $Q_n \in \mathcal{Q}$ , where  $|\mathcal{Q}| = 3$ , since this is sufficient to preserve the values of  $\mathbb{E}[X_n^2]$ ,  $s_\lambda(X_n, \text{snr} | Q_n)$  and  $(h(Y_n | Q_n) - h(X_n | Q_n))$ . Thus, since  $\mathcal{Q}$  is finite and  $\mathbb{E}[X_n^2] \leq 1$ , the sequence  $\{X_n, Q_n\}$  is tight. By Prokhorov’s theorem [62], we may assume that there is some  $(X_*, Q_*)$  for which  $(X_n, Q_n) \xrightarrow{\mathcal{D}} (X_*, Q_*)$  by restricting our attention to a subsequence of  $\{X_n, Q_n\}$  if necessary. Moreover,  $\mathbb{E}[X_*^2] \leq \liminf_{n \rightarrow \infty} \mathbb{E}[X_n^2] \leq 1$  by Fatou’s lemma.

Next, for a given  $n$ , let  $(X_{1,n}, Q_{1,n})$  and  $(X_{2,n}, Q_{2,n})$  denote two independent copies of  $(X_n, Q_n)$ . Define

$$X_{+,n} = \frac{X_{1,n} + X_{2,n}}{\sqrt{2}}, \quad X_{-,n} = \frac{X_{1,n} - X_{2,n}}{\sqrt{2}}, \quad (137)$$

In a similar manner, define  $Y_{+,n}, Y_{-,n}$ , and put  $\mathbf{Q}_n = (Q_{1,n}, Q_{2,n})$ . Applying Lemma 11 to the variables  $\mathbf{Q}_n \rightarrow (X_{+,n}, X_{-,n}) \rightarrow (Y_{+,n}, Y_{-,n})$ , we obtain

$$2s_\lambda(X_n, \text{snr} | Q_n) \geq s_\lambda(X_{+,n}, \text{snr} | X_{-,n} \mathbf{Q}_n) + s_\lambda(X_{-,n}, \text{snr} | Y_{+,n} \mathbf{Q}_n), \quad (138)$$

and the symmetric inequality

$$2s_\lambda(X_n, \text{snr}|Q_n) \geq s_\lambda(X_{+,n}, \text{snr}|Y_{-,n}\mathbf{Q}_n) + s_\lambda(X_{-,n}, \text{snr}|X_{+,n}\mathbf{Q}_n). \quad (139)$$

By independence of  $X_{1,n}$  and  $X_{2,n}$  and the assumption that  $\mathbb{E}[X_n] = 0$ , we have

$$\mathbb{E}[X_{+,n}^2] = \mathbb{E}[X_{-,n}^2] = \frac{1}{2}\mathbb{E}[X_{1,n}^2] + \frac{1}{2}\mathbb{E}[X_{2,n}^2] = \mathbb{E}[X_n^2] \leq 1. \quad (140)$$

Hence, it follows that the terms in the RHS of (138) and the RHS of (139) are each lower bounded by  $V_\lambda(\text{snr})$ . Since  $\lim_{n \rightarrow \infty} s_\lambda(X_n, \text{snr}|Q_n) = V_\lambda(\text{snr})$  by definition, we must also have

$$\lim_{n \rightarrow \infty} \frac{1}{2} \left( s_\lambda(X_{+,n}, \text{snr}|Y_{-,n}\mathbf{Q}_n) + s_\lambda(X_{-,n}, \text{snr}|Y_{+,n}\mathbf{Q}_n) \right) = V_\lambda(\text{snr}). \quad (141)$$

In particular, by letting the random pair  $(X'_n, Q'_n)$  correspond to equal time-sharing between the pairs  $(X_{+,n}, (Y_{-,n}\mathbf{Q}_n))$  and  $(X_{-,n}, (Y_{+,n}\mathbf{Q}_n))$ , we have constructed an admissible sequence  $\{X'_n, Q'_n\}$  which satisfies

$$\lim_{n \rightarrow \infty} s_\lambda(X'_n, \text{snr}|Q'_n) = V_\lambda(\text{snr}). \quad (142)$$

Using Markovity, the following identity is readily established

$$h(Y_n|Q_n) - h(X_n|Q_n) = \frac{1}{2} (h(Y_{+,n}, Y_{-,n}|\mathbf{Q}_n) - h(X_{+,n}, X_{-,n}|\mathbf{Q}_n)) \quad (143)$$

$$= \frac{1}{2} (h(Y_{-,n}|Y_{+,n}, \mathbf{Q}_n) - h(X_{-,n}|Y_{+,n}, \mathbf{Q}_n)) \quad (144)$$

$$+ \frac{1}{2} (h(Y_{+,n}|Y_{-,n}, \mathbf{Q}_n) - h(X_{+,n}|Y_{-,n}, \mathbf{Q}_n))$$

$$+ \frac{1}{2} I(X_{+,n}; X_{-,n}|Y_{+,n}, Y_{-,n}, \mathbf{Q}_n)$$

$$= h(Y'_n|Q'_n) - h(X'_n|Q'_n) + \frac{1}{2} I(X_{+,n}; X_{-,n}|Y_{+,n}, Y_{-,n}, \mathbf{Q}_n). \quad (145)$$

Since the sequence  $\{X'_n, Q'_n\}$  is admissible, it must also satisfy (132). Therefore, in view of (145) and the fact that the LHS of (132) is finite, this implies that

$$\liminf_{n \rightarrow \infty} I(X_{1,n} + X_{2,n}; X_{1,n} - X_{2,n}|Y_{1,n}, Y_{2,n}, \mathbf{Q}_n) = 0. \quad (146)$$

In particular, for  $P_{Q_*} \times P_{Q_*}$ -a.e.  $(q_1, q_2)$ ,

$$\liminf_{n \rightarrow \infty} I(X_{1,n} + X_{2,n}; X_{1,n} - X_{2,n}|Y_{1,n}, Y_{2,n}, (\mathbf{Q}_n = q_1, q_2)) = 0. \quad (147)$$

This completes the proof since Lemma 6 guarantees that, for  $P_{Q_*}$ -a.e.  $q$ , the random variable  $X_*|\{Q_* = q\}$  is normal with variance not depending on  $q$ , and moreover we have already observed that  $\mathbb{E}[X_*^2] \leq 1$ , so the variance of  $X_*|\{Q_* = q\}$  is at most unity as claimed.  $\blacksquare$

#### D. Weak Semicontinuity of $s_\lambda(\cdot, \text{snr})$

This subsection is devoted to establishing the following semicontinuity property of  $s_\lambda(\cdot, \text{snr})$ , which was the second essential ingredient needed for the proof of Theorem 8 (i.e., Claim II).

**Lemma 12.** *If  $X_n \xrightarrow{\mathcal{D}} X_* \sim N(\mu, \sigma_X^2)$  and  $\sup_n \mathbb{E}[X_n^2] < \infty$ , then*

$$\liminf_{n \rightarrow \infty} s_\lambda(X_n, \text{snr}) \geq s_\lambda(X_*, \text{snr}). \quad (148)$$

Recall that  $s_\lambda(X, \text{snr})$  is defined in terms of the Gaussian channel  $Y = \sqrt{\text{snr}}X + Z$ . However, for the purposes of the proof, it will be convenient to omit the snr scaling factor, and instead parametrize the channel in terms of

the noise variance. Toward this end, let  $Z \sim N(0, \sigma^2)$ . For  $\lambda > 0$  and a random variable  $X \sim P_X$ , independent of  $Z$ , define  $Y = X + Z$  and the functionals

$$F_{\lambda, \sigma^2}(X) = \inf_{V: X \rightarrow Y \rightarrow V} (I(Y; V) - \lambda I(X; V)) \quad (149)$$

$$G_{\lambda, \sigma^2}(X) = -h(X) + \lambda h(Y). \quad (150)$$

Lemma 12 is an immediate corollary of weak lower semicontinuity of  $G_{\lambda, \sigma^2}(X)$  and  $F_{\lambda, \sigma^2}(X)$  at Gaussian  $X$ . These facts are established separately below in Lemmas 13 and 15, respectively. The former is straightforward, while the latter requires some effort.

**Lemma 13.** *If  $X_n \xrightarrow{\mathcal{D}} X_* \sim N(\mu, \sigma_X^2)$  and  $\sup_n \mathbb{E}[X_n^2] < \infty$ , then*

$$\liminf_{n \rightarrow \infty} G_{\lambda, \sigma^2}(X_n) \geq G_{\lambda, \sigma^2}(X_*). \quad (151)$$

*Proof:* Fix  $\delta > 0$  and define  $N_\delta \sim N(0, \delta)$ , pairwise independent of  $\{X_n\}, X_*$ . Observe that

$$G_{\lambda, \sigma^2}(X_n) = -h(X_n) + \lambda h(Y_n) \geq -h(X_n + N_\delta) + \lambda h(Y_n). \quad (152)$$

By the third claim of Lemma 2, we have  $-h(X_n + N_\delta) + \lambda h(Y_n) \rightarrow -h(X_* + N_\delta) + \lambda h(Y_*)$  as  $n \rightarrow \infty$ . Thus,

$$\liminf_{n \rightarrow \infty} G_{\lambda, \sigma^2}(X_n) \geq -h(X_* + N_\delta) + \lambda h(Y_*). \quad (153)$$

Since  $h(X_* + N_\delta) = \frac{1}{2} \log(2\pi e(\sigma_X^2 + \delta))$  is continuous in  $\delta$ , we may take  $\delta \downarrow 0$  to prove the claim.  $\blacksquare$

**Lemma 14.**  *$F_{\lambda, \sigma^2}(X)$  is continuous in  $\lambda$ . Furthermore, if  $X \sim N(\mu, \sigma_X^2)$ , then*

$$F_{\lambda, \sigma^2}(X) = \begin{cases} \frac{1}{2} \left[ \log \left( (\lambda - 1) \frac{\sigma_X^2}{\sigma^2} \right) - \lambda \log \left( \frac{\lambda - 1}{\lambda} \left( 1 + \frac{\sigma_X^2}{\sigma^2} \right) \right) \right] & \text{If } \lambda \geq 1 + \frac{\sigma^2}{\sigma_X^2} \\ 0 & \text{If } 0 \leq \lambda \leq 1 + \frac{\sigma^2}{\sigma_X^2}. \end{cases} \quad (154)$$

*In particular,  $F_{\lambda, \sigma^2}(X)$  is continuous in the parameters  $\sigma^2, \sigma_X^2$  and  $\lambda$  for Gaussian  $X$ .*

*Proof:* The function  $F_{\lambda, \sigma^2}(X)$  is the pointwise infimum of linear functions in  $\lambda$ , and is therefore concave and continuous on the open interval  $\lambda \in (0, \infty)$  for any distribution  $P_X$ . The explicit expression (154) follows by identifying  $\gamma \text{snr} \leftarrow \frac{\sigma_X^2}{\sigma^2}$  in Proposition 2.  $\blacksquare$

**Lemma 15.** *If  $X_n \xrightarrow{\mathcal{D}} X_* \sim N(\mu, \sigma_X^2)$  and  $\sup_n \mathbb{E}[X_n^2] < \infty$ , then*

$$\liminf_{n \rightarrow \infty} F_{\lambda, \sigma^2}(X_n) \geq F_{\lambda, \sigma^2}(X_*). \quad (155)$$

*Proof:* Fix an interval  $B = [-b, b]$ , a channel  $P_{V|Y}$ , and  $\delta$  satisfying  $0 < \delta < \sigma^2/2$ . Recalling the definition of  $Z \sim N(0, \sigma^2)$ , decompose  $Z = N_1 + N_2 + N_3$ , where  $N_1 \sim N(0, \delta)$ ,  $N_2 \sim N(0, \sigma^2 - 2\delta)$  and  $N_3 \sim N(0, \delta)$  are mutually independent. Define  $X_n^\delta = X_n + N_1$  and  $Y_n^\delta = Y_n - N_3 = X_n + N_1 + N_2$ . Note that we have  $X_n \rightarrow X_n^\delta \rightarrow Y_n^\delta \rightarrow Y_n \rightarrow V_n$ , where  $V_n$  is defined by the stochastic transformation  $P_{V|Y} : Y_n \mapsto V_n$ . Using the notation of Lemma 2, we also have  $X_* \rightarrow X_*^\delta \rightarrow Y_*^\delta \rightarrow Y_* \rightarrow V_*$ , where  $Y_* = X_* + Z$ ,  $X_*^\delta = X_* + N_1$ ,  $Y_*^\delta = Y_* - N_3$  and  $V_*$  is defined via  $P_{V|Y} : Y_* \mapsto V_*$ . With these definitions in hand, we may apply Lemma 4 to the processes  $\{X_n^\delta\}, \{Y_n^\delta\}$  to conclude the existence of a sequence  $\epsilon_n \rightarrow 0$ , not depending on  $P_{V|Y}$ , that satisfies

$$I(V_n; X_n^\delta | X_n^\delta \in B) \leq (1 + \epsilon_n)^2 I(V_*; X_*^\delta | X_*^\delta \in B) - (1 + \epsilon_n)^2 \log(1 - \epsilon_n)^2 \quad (156)$$

$$I(V_n; Y_n^\delta | Y_n^\delta \in B) \geq (1 - \epsilon_n)^2 I(V_*; Y_*^\delta | Y_*^\delta \in B) - (1 - \epsilon_n)^2 \log(1 + \epsilon_n)^2 \quad (157)$$

$$\mathbb{P}(X_n^\delta \in B) \leq (1 + \epsilon_n) \mathbb{P}(X_*^\delta \in B) \quad (158)$$

$$\mathbb{P}(Y_n^\delta \in B) \geq (1 - \epsilon_n) \mathbb{P}(Y_*^\delta \in B). \quad (159)$$

Now, we have the following sequence of inequalities

$$I(Y_n; V_n) - \lambda I(X_n; V_n) \geq I(Y_n^\delta; V_n) - \lambda I(X_n^\delta; V_n) \quad (160)$$

$$= I(Y_n^\delta, \mathbb{1}_{\{Y_n^\delta \in B\}}; V_n) - \lambda I(X_n^\delta, \mathbb{1}_{\{X_n^\delta \in B\}}; V_n) \quad (161)$$

$$= \mathbb{P}(Y_n^\delta \in B) I(Y_n^\delta; V_n | Y_n^\delta \in B) + \mathbb{P}(Y_n^\delta \notin B) I(Y_n^\delta; V_n | Y_n^\delta \notin B) + I(\mathbb{1}_{\{Y_n^\delta \in B\}}; V_n) - \lambda \left( \mathbb{P}(X_n^\delta \in B) I(X_n^\delta; V_n | X_n^\delta \in B) + \mathbb{P}(X_n^\delta \notin B) I(X_n^\delta; V_n | X_n^\delta \notin B) + I(\mathbb{1}_{\{X_n^\delta \in B\}}; V_n) \right) \quad (162)$$

$$\geq \mathbb{P}(Y_n^\delta \in B) I(Y_n^\delta; V_n | Y_n^\delta \in B) - \lambda \left( \mathbb{P}(X_n^\delta \in B) I(X_n^\delta; V_n | X_n^\delta \in B) + \mathbb{P}(X_n^\delta \notin B) I(X_n^\delta; V_n | X_n^\delta \notin B) + H(\mathbb{1}_{\{X_n^\delta \in B\}}) \right) \quad (163)$$

$$\geq \mathbb{P}(Y_n^\delta \in B) (1 - \epsilon_n)^2 I(Y_*^\delta; V_* | Y_*^\delta \in B) - \mathbb{P}(Y_n^\delta \in B) (1 - \epsilon_n)^2 \log(1 + \epsilon_n)^2 - \lambda \left( \mathbb{P}(X_n^\delta \in B) (1 + \epsilon_n)^2 I(X_*^\delta; V_* | X_*^\delta \in B) - \mathbb{P}(X_n^\delta \in B) (1 + \epsilon_n)^2 \log(1 - \epsilon_n)^2 \right) \quad (164)$$

$$- \lambda \left( \mathbb{P}(X_n^\delta \notin B) I(X_n^\delta; V_n | X_n^\delta \notin B) + H(\mathbb{1}_{\{X_n^\delta \in B\}}) \right) \geq \frac{\mathbb{P}(Y_n^\delta \in B)}{\mathbb{P}(Y_*^\delta \in B)} (1 - \epsilon_n)^2 \left( I(Y_*^\delta; V_*) - \mathbb{P}(Y_*^\delta \notin B) I(Y_*^\delta; V_* | Y_*^\delta \notin B) - I(\mathbb{1}_{\{Y_*^\delta \in B\}}; V_*) \right) \quad (165)$$

$$- \lambda (1 + \epsilon_n)^2 \frac{\mathbb{P}(X_n^\delta \in B)}{\mathbb{P}(X_*^\delta \in B)} I(X_*^\delta; V_*) - \mathbb{P}(Y_n^\delta \in B) (1 - \epsilon_n)^2 \log(1 + \epsilon_n)^2 + \lambda \mathbb{P}(X_n^\delta \in B) (1 + \epsilon_n)^2 \log(1 - \epsilon_n)^2 - \lambda \left( \mathbb{P}(X_n^\delta \notin B) I(X_n^\delta; V_n | X_n^\delta \notin B) + H(\mathbb{1}_{\{X_n^\delta \in B\}}) \right) \geq (1 - \epsilon_n)^3 I(Y_*^\delta; V_*) - \lambda (1 + \epsilon_n)^3 I(X_*^\delta; V_*) \quad (166)$$

$$- (1 - \epsilon_n)^3 \left( \mathbb{P}(Y_*^\delta \notin B) I(Y_*^\delta; V_* | Y_*^\delta \notin B) + H(\mathbb{1}_{\{Y_*^\delta \in B\}}) \right) - \mathbb{P}(Y_n^\delta \in B) (1 - \epsilon_n)^2 \log(1 + \epsilon_n)^2 + \lambda \mathbb{P}(X_n^\delta \in B) (1 + \epsilon_n)^2 \log(1 - \epsilon_n)^2 - \lambda \left( \mathbb{P}(X_n^\delta \notin B) I(X_n^\delta; V_n | X_n^\delta \notin B) + H(\mathbb{1}_{\{X_n^\delta \in B\}}) \right) \geq (1 - \epsilon_n)^3 F_{\lambda_n, (\sigma^2 - 2\delta)}(X_*^\delta) \quad (167)$$

where  $\lambda_n := \lambda \left( \frac{1 + \epsilon_n}{1 - \epsilon_n} \right)^3$ . The above steps are justified as follows:

- (160) follows by the data processing inequality.
- (161) follows since  $\mathbb{1}_{\{Y_n^\delta \in B\}}$  and  $\mathbb{1}_{\{X_n^\delta \in B\}}$  are functions of  $Y_n^\delta$  and  $X_n^\delta$ , respectively.
- (162) follows from the chain rule for mutual information.
- (163) follows from non-negativity of mutual information, the fact that  $I(\mathbb{1}_{\{X_n^\delta \in B\}}; V_n) \leq H(\mathbb{1}_{\{X_n^\delta \in B\}})$ , and the data processing inequality which implies  $I(X_n^\delta; V_n | X_n^\delta \notin B) \leq I(X_n^\delta; V_n | X_n^\delta \in B)$ .
- (164) follows from (156) and (157).
- (165) follows from the chain rule for mutual information, which implies

$$I(Y_*^\delta; V_* | Y_*^\delta \in B) = \frac{1}{\mathbb{P}(Y_*^\delta \in B)} \left( I(Y_*^\delta; V_*) - \mathbb{P}(Y_*^\delta \notin B) I(Y_*^\delta; V_* | Y_*^\delta \notin B) - I(\mathbb{1}_{\{Y_*^\delta \in B\}}; V_*) \right) \quad (168)$$



and, combined with non-negativity of mutual information,

$$I(X_*^\delta; V_* | X_*^\delta \in B) \leq \frac{1}{\mathbb{P}(X_*^\delta \in B)} I(X_*^\delta; V_*). \quad (169)$$

- (166) follows from (158), (159), the fact that  $I(\mathbb{1}_{\{Y_*^\delta \in B\}}; V_n) \leq H(\mathbb{1}_{\{Y_*^\delta \in B\}})$ , and the data processing inequality which implies  $I(Y_*^\delta; V_* | Y_*^\delta \notin B) \leq I(Y_*^\delta; Y_* | Y_*^\delta \notin B)$ .
- (167) follows from the definition of  $F_{\lambda_n, (\sigma^2 - 2\delta)}(X_*^\delta)$  by taking the infimum over  $V_*$  satisfying  $X_*^\delta \rightarrow Y_*^\delta \rightarrow V_*$ .

Summarizing above, we have shown

$$\begin{aligned} I(Y_n; V_n) - \lambda I(X_n; V_n) &\geq (1 - \epsilon_n)^3 F_{\lambda_n, (\sigma^2 - 2\delta)}(X_*^\delta) \\ &\quad - (1 - \epsilon_n)^3 \left( \mathbb{P}(Y_*^\delta \notin B) I(Y_*^\delta; Y_* | Y_*^\delta \notin B) + H(\mathbb{1}_{\{Y_*^\delta \in B\}}) \right) \\ &\quad - \mathbb{P}(Y_n^\delta \in B) (1 - \epsilon_n)^2 \log(1 + \epsilon_n)^2 + \lambda \mathbb{P}(X_n^\delta \in B) (1 + \epsilon_n)^2 \log(1 - \epsilon_n)^2 \\ &\quad - \lambda \left( \mathbb{P}(X_n^\delta \notin B) I(X_n^\delta; Y_n | X_n^\delta \notin B) + H(\mathbb{1}_{\{X_n^\delta \in B\}}) \right), \end{aligned} \quad (170)$$

Note that the RHS of (170) does not depend on  $V_n$  (i.e.,  $P_{V|Y}$ ). Thus, taking the infimum over  $V_n$  satisfying  $X_n \rightarrow Y_n \rightarrow V_n$  and then letting  $n \rightarrow \infty$ , we arrive at

$$\begin{aligned} \liminf_{n \rightarrow \infty} F_{\lambda, \sigma^2}(X_n) &\geq F_{\lambda, (\sigma^2 - 2\delta)}(X_*^\delta) - \left( \mathbb{P}(Y_*^\delta \notin B) I(Y_*^\delta; Y_* | Y_*^\delta \notin B) + H(\mathbb{1}_{\{Y_*^\delta \in B\}}) \right) \\ &\quad - \lambda \left( \mathbb{P}(X_*^\delta \notin B) I(X_*^\delta; Y_* | X_*^\delta \notin B) + H(\mathbb{1}_{\{X_*^\delta \in B\}}) \right), \end{aligned} \quad (171)$$

which follows due to  $\epsilon_n \rightarrow 0$  and the following:

- $F_{\lambda_n, (\sigma^2 - 2\delta)}(X_*^\delta) \rightarrow F_{\lambda, (\sigma^2 - 2\delta)}(X_*^\delta)$  by continuity of  $F_{\lambda, \sigma^2}(X)$  in  $\lambda$  (Lemma 14).
- $\mathbb{P}(X_n^\delta \notin B) \rightarrow \mathbb{P}(X_*^\delta \notin B)$  since  $X_n^\delta \xrightarrow{\mathcal{D}} X_*^\delta$  by the first claim of Lemma 2. By the same token,  $H(\mathbb{1}_{\{X_n^\delta \in B\}}) \rightarrow H(\mathbb{1}_{\{X_*^\delta \in B\}})$  by continuity of the binary entropy function.
- $I(X_n^\delta; Y_n | X_n^\delta \notin B) \rightarrow I(X_*^\delta; Y_* | X_*^\delta \notin B)$  by the third claim of Lemma 2 since  $\limsup_n \mathbb{E}[(X_n^\delta)^2 | X_n^\delta \notin B] < \infty$  due to the fact that  $\sup_n \mathbb{E}[X_n^2] < \infty$  and  $\mathbb{P}(X_n^\delta \notin B) \rightarrow \mathbb{P}(X_*^\delta \notin B)$ , a positive constant.

As we take  $b \rightarrow \infty$ , continuity of the binary entropy function and Lemma 5 together imply the latter two terms in the RHS of (171) vanish, yielding the inequality

$$\liminf_{n \rightarrow \infty} F_{\lambda}(X_n) \geq F_{\lambda, (\sigma^2 - 2\delta)}(X_*^\delta). \quad (172)$$

Since  $\delta$  was arbitrary and  $F_{\lambda, (\sigma^2 - 2\delta)}(X_*^\delta)$  is continuous in  $\delta$  by Lemma 14, the proof is complete by letting  $\delta \downarrow 0$ .  $\blacksquare$

**Remark 5.** Given the tedious chain of inequalities in the proof of Lemma 15, it is easy to lose sight of the overall picture. The crucial idea is that perturbing  $X_n \rightarrow X_n^\delta$  and  $Y_n \rightarrow Y_n^\delta$  allows us to eventually eliminate dependence on the channel  $P_{V|Y}$  in the RHS of (167). Resisting the temptation to take limits  $n \rightarrow \infty$  or  $b \rightarrow \infty$  until after dependence on any particular channel  $P_{V|Y}$  is eliminated (i.e., inequality (170)) is also essential.

We note that the hypothesis that  $X_* \sim N(0, \sigma^2)$  was not needed in the proof of Lemma 15 until the very last step. Indeed, we may actually conclude that the following general result holds, which may be of independent interest:

**Proposition 3.** Suppose  $X_n \xrightarrow{\mathcal{D}} X_*$  and  $\sup_n \mathbb{E}[X_n^2] < \infty$ , then for all  $0 < \delta < \delta' < \sigma^2$ , the following holds:

$$\liminf_{n \rightarrow \infty} F_{\lambda, \sigma^2}(X_n) \geq F_{\lambda, (\sigma^2 - \delta')}(X_* + N_\delta) \quad (173)$$

where  $N_\delta \sim N(0, \delta)$  is independent of  $X_*$ .

*Proof:* The claim follows from the proof of Lemma 15, but stopping at (172) and not particularizing to Gaussian  $X_*$ . The replacement of  $2\delta$  by  $\delta'$  is straightforward by decomposing  $Z$  differently in the first step of the proof.  $\blacksquare$

**Remark 6.** It is possible to establish weak upper semicontinuity of  $F_{\lambda, \sigma^2}(\cdot)$ , but that is not needed for our purposes.

## REFERENCES

- [1] C.E. Shannon. A mathematical theory of communication. *Bell System Technical Journal, The*, 27(4):623–656, Oct 1948.
- [2] A. J. Stam. Some inequalities satisfied by the quantities of information of Fisher and Shannon. *Information and Control*, 2(2):101–112, 1959.
- [3] Nelson M. Blachman. The convolution inequality for entropy powers. *Information Theory, IEEE Transactions on*, 11(2):267–271, Apr 1965.
- [4] M.H.M. Costa. A new entropy power inequality. *Information Theory, IEEE Transactions on*, 31(6):751–760, Nov 1985.
- [5] Ruoheng Liu, Tie Liu, H.V. Poor, and S. Shamai. A vector generalization of costa’s entropy-power inequality with applications. *Information Theory, IEEE Transactions on*, 56(4):1865–1879, April 2010.
- [6] M.H.M. Costa. On the Gaussian interference channel. *Information Theory, IEEE Transactions on*, 31(5):607–615, Sep 1985.
- [7] Yury Polyanskiy and Yihong Wu. Wasserstein continuity of entropy and outer bounds for interference channels. *arXiv preprint arXiv:1504.04419*, 2015.
- [8] G. Bagherikaram, A.S. Motahari, and A.K. Khandani. The secrecy capacity region of the Gaussian mimo broadcast channel. *Information Theory, IEEE Transactions on*, 59(5):2673–2682, May 2013.
- [9] A.B. Wagner, S. Tavildar, and P. Viswanath. Rate region of the quadratic Gaussian two-encoder source-coding problem. *Information Theory, IEEE Transactions on*, 54(5):1938–1961, May 2008.
- [10] Y. Oohama. Gaussian multiterminal source coding. *Information Theory, IEEE Transactions on*, 43(6):1912–1923, Nov 1997.
- [11] O. Rioul. Information theoretic proofs of entropy power inequalities. *Information Theory, IEEE Transactions on*, 57(1):33–55, Jan 2011.
- [12] P. Bergmans. A simple converse for broadcast channels with additive white Gaussian noise (corresp.). *Information Theory, IEEE Transactions on*, 20(2):279–280, Mar 1974.
- [13] H. Weingarten, Y. Steinberg, and S. Shamai. The capacity region of the Gaussian multiple-input multiple-output broadcast channel. *Information Theory, IEEE Transactions on*, 52(9):3936–3964, Sept 2006.
- [14] M. Mohseni and J.M. Cioffi. A proof of the converse for the capacity of Gaussian mimo broadcast channels. In *Information Theory, 2006 IEEE International Symposium on*, pages 881–885, July 2006.
- [15] S. Leung-Yan-Cheong and M.E. Hellman. The Gaussian wire-tap channel. *Information Theory, IEEE Transactions on*, 24(4):451–456, Jul 1978.
- [16] E. Tekin and A. Yener. The Gaussian multiple access wire-tap channel. *Information Theory, IEEE Transactions on*, 54(12):5747–5755, Dec 2008.
- [17] L. Ozarow. On a source-coding problem with two channels and three receivers. *Bell System Technical Journal, The*, 59(10):1909–1921, Dec 1980.
- [18] Y. Oohama. Rate-distortion theory for Gaussian multiterminal source coding systems with several side informations at the decoder. *Information Theory, IEEE Transactions on*, 51(7):2577–2593, July 2005.
- [19] V. Prabhakaran, D. Tse, and K. Ramachandran. Rate region of the quadratic Gaussian CEO problem. In *Information Theory, 2004. ISIT 2004. Proceedings. International Symposium on*, pages 119–, June 2004.
- [20] Mokshay Madiman. On the entropy of sums. In *Proc. IEEE Inform. Theory Workshop*, pages 303–307, 2008.
- [21] Maice Costa and Thomas Cover. On the similarity of the entropy power inequality and the Brunn-Minkowski inequality (corresp.). *Information Theory, IEEE Transactions on*, 30(6):837–839, 1984.
- [22] Leonard Gross. Logarithmic Sobolev inequalities. *American Journal of Mathematics*, 97(4):1061–1083, 1975.
- [23] Edward Nelson. The free Markoff field. *Journal of Functional Analysis*, 12(2):211–227, 1973.
- [24] Eric A Carlen. Superadditivity of Fisher’s information and logarithmic Sobolev inequalities. *Journal of Functional Analysis*, 101(1):194–211, 1991.
- [25] Maxim Raginsky and Igal Sason. Concentration of measure inequalities in information theory, communications, and coding. *Foundations and Trends in Communications and Information Theory*, 10(1-2):1–247, 2013.
- [26] Ioannis Kontoyiannis and Mokshay Madiman. Sumset and inverse sumset inequalities for differential entropy and mutual information. *Information Theory, IEEE Transactions on*, 60(8):4503–4514, 2014.
- [27] Mokshay Madiman and Ioannis Kontoyiannis. The Ruzsa divergence for random elements in locally compact abelian groups. *arXiv preprint arXiv:1508.04089*, 2015.
- [28] Sergey G Bobkov, Nathael Gozlan, Cyril Roberto, and P-M Samson. Bounds on the deficit in the logarithmic Sobolev inequality. *Journal of Functional Analysis*, 267(11):4110–4138, 2014.
- [29] Max Fathi, Emanuel Indrei, and Michel Ledoux. Quantitative logarithmic Sobolev inequalities and stability estimates. *arXiv preprint arXiv:1410.6922*, 2014.
- [30] Jean Dolbeault and Giuseppe Toscani. Stability results for logarithmic Sobolev and Gagliardo-Nirenberg inequalities. *International Mathematics Research Notices*, page rnv131, 2015.
- [31] Keith Ball, Franck Barthe, and Assaf Naor. Entropy jumps in the presence of a spectral gap. *Duke Mathematical Journal*, 119(1):41–63, 2003.
- [32] Keith Ball and Van Hoang Nguyen. Entropy jumps for isotropic log-concave random vectors and spectral gap. *Studia Mathematica*, 213(1):81–96, 2012.

- [33] T. Berger. *Multiterminal Source Coding*. In G. Longo (Ed.), *The Information Theory Approach to Communications*. Springer-Verlag, New York, NY, USA, 1977.
- [34] S.-Y. Tung. *Multiterminal Source Coding*. PhD thesis, Cornell University, Ithaca, NY, 1978.
- [35] Abbas El Gamal and Young-Han Kim. *Network Information Theory*. Cambridge University Press, 2012.
- [36] Jia Wang, Jun Chen, and Xiaolin Wu. On the sum rate of Gaussian multiterminal source coding: New proofs and results. *Information Theory, IEEE Transactions on*, 56(8):3946–3960, 2010.
- [37] Thomas Courtade and Jiantao Jiao. An extremal inequality for long Markov chains. In *Communication, Control, and Computing (Allerton), 2014 52nd Annual Allerton Conference on*, pages 763–770. IEEE, 2014.
- [38] Thomas A Courtade and Tsachy Weissman. Multiterminal source coding under logarithmic loss. *IEEE Transactions on Information Theory*, 60(1), 2014.
- [39] Jiantao Jiao, Thomas A Courtade, Kartik Venkat, and Tsachy Weissman. Justification of logarithmic loss via the benefit of side information. *Information Theory, IEEE Transactions on*, 61(10):5357–5365, 2015.
- [40] Hiroshi Sato. The capacity of the Gaussian interference channel under strong interference (corresp.). *Information Theory, IEEE Transactions on*, 27(6):786–788, 1981.
- [41] Han Te Sun and Kingo Kobayashi. A new achievable rate region for the interference channel. *IEEE transactions on information theory*, 27(1):49–60, 1981.
- [42] Max HM Costa. Noisebergs in Z-Gaussian interference channels. In *Information Theory and Applications Workshop (ITA), 2011*, pages 1–6. IEEE, 2011.
- [43] Chandra Nair and Max HM Costa. Gaussian Z-interference channel: Around the corner. In *Information Theory and Applications Workshop (ITA), 2016*. IEEE, 2016.
- [44] Rudolf Ahlswede and Peter Gács. Spreading of sets in product spaces and hypercontraction of the Markov operator. *The Annals of Probability*, pages 925–939, 1976.
- [45] Chandra Nair. Equivalent formulations of hypercontractivity using information measures. In *International Zurich Seminar on Communications*, page 42, 2014.
- [46] Venkat Anantharam, Amin Aminzadeh Gohari, Sudeep Kamath, and Chandra Nair. On hypercontractivity and the mutual information between boolean functions. In *Allerton*, pages 13–19, 2013.
- [47] Maxim Raginsky. Strong data processing inequalities and  $\phi$ -Sobolev inequalities for discrete channels. *arXiv preprint arXiv:1411.3575*, 2014.
- [48] Yury Polyanskiy and Yihong Wu. Dissipation of information in channels with input constraints. *arXiv preprint arXiv:1405.3629*, 2014.
- [49] Flávio du Pin Calmon, Yury Polyanskiy, and Yihong Wu. Strong data processing inequalities in power-constrained Gaussian channels. In *ISIT*, pages 2558–2562. IEEE, 2015.
- [50] Venkat Anantharam, Amin Gohari, Sudeep Kamath, and Chandra Nair. On maximal correlation, hypercontractivity, and the data processing inequality studied by Erkip and Cover. *arXiv preprint arXiv:1304.6133*, 2013.
- [51] Thomas Courtade. Outer bounds for multiterminal source coding via a strong data processing inequality. In *Information Theory Proceedings (ISIT), 2013 IEEE International Symposium on*, pages 559–563. IEEE, 2013.
- [52] Harold Gordon Eggleston. *Convexity*. Number 47. CUP Archive, 1958.
- [53] S.G. Bobkov and G.P. Chistyakov. Entropy power inequality for the Rényi entropy. *Information Theory, IEEE Transactions on*, 61(2):708–714, Feb 2015.
- [54] Włodzimierz Bryc. *The normal distribution: characterizations with applications*, volume 100. Springer Science & Business Media, 2012.
- [55] Yanlin Geng and C. Nair. The capacity region of the two-receiver Gaussian vector broadcast channel with private and common messages. *Information Theory, IEEE Transactions on*, 60(4):2087–2104, April 2014.
- [56] Yihong Wu and S. Verdu. Functional properties of minimum mean-square error and mutual information. *Information Theory, IEEE Transactions on*, 58(3):1289–1301, March 2012.
- [57] S. N. Bernstein. On a property characteristic of the normal law. *Trudy Leningrad. Polytech. Inst*, 3:21–22, 1941.
- [58] Elliott H Lieb. Gaussian kernels have only Gaussian maximizers. *Inventiones mathematicae*, 102(1):179–208, 1990.
- [59] Franck Barthe. Optimal Young’s inequality and its converse: a simple proof. *Geometric & Functional Analysis GAFA*, 8(2):234–242, 1998.
- [60] Eric A Carlen and Dario Cordero-Erausquin. Subadditivity of the entropy and its relation to BrascampLieb type inequalities. *Geometric and Functional Analysis*, 19(2):373–405, 2009.
- [61] Jingbo Liu, Thomas A. Courtade, Paul Cuff, and Sergio Verdú. Information theoretic perspectives on Brascamp-Lieb inequalities. *2016 International Symposium on Information Theory (submitted)*, 2016.
- [62] Rick Durrett. *Probability: theory and examples*. Cambridge university press, 2010.