A Vector Generalization of Costa's Entropy-Power Inequality with Applications

Ruoheng Liu, Tie Liu, H. Vincent Poor, and Shlomo Shamai (Shitz)

arXiv:0903.3024v1 [cs.IT] 17 Mar 2009

Abstract

This paper considers an entropy-power inequality (EPI) of Costa and presents a natural vector generalization with a real positive semidefinite matrix parameter. This new inequality is proved using a perturbation approach via a fundamental relationship between the derivative of mutual information and the minimum mean-square error (MMSE) estimate in linear vector Gaussian channels. As an application, a new extremal entropy inequality is derived from the generalized Costa EPI and then used to establish the secrecy capacity regions of the degraded vector Gaussian broadcast channel with layered confidential messages.

Index Terms

Entropy-power inequality (EPI), extremal entropy inequality, information-theoretic security, mutual information and minimum mean-square error (MMSE) estimate, vector Gaussian broadcast channel

I. INTRODUCTION

In information theory, the entropy-power inequality (EPI) of Shannon [1] and Stam [2] has played key roles in the solution of several canonical network communication problems. Celebrated examples include Bergmans's solution [3] to the Gaussian broadcast channel problem, Leung-Yan-Cheong and Hellman's solution [4] to the Gaussian wire-tap channel problem, Ozarow's solution [5] to the Gaussian two-description problem, Oohama's solution [6] to the quadratic Gaussian CEO problem, and more recently Weingarten, Steinberg and Shamai's solution [7] to the multiple-input multiple-output Gaussian broadcast channel problem.

Ruoheng Liu and H. Vincent Poor are with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544, USA. Email: {rliu,poor}@princeton.edu

Tie Liu is with the Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX 77843, USA. Email: tieliu@tamu.edu

Shlomo Shamai (Shitz) is with the Department of Electrical Engineering, Technion-Israel Institute of Technology, Technion City, Haifa 32000, Israel. Email: sshlomo@ee.technion.ac.il

Let $\mathbf{X}$ and $\mathbf{Z}$ be two independent random n-vectors with densities in $\mathbb{R}^n$, where $\mathbb{R}$ denotes the set of real numbers. The classical EPI of Shannon [1] and Stam [2] can be written as

$$\exp\left[\frac{2}{n}h(\mathbf{X}+\mathbf{Z})\right] \geq \exp\left[\frac{2}{n}h(\mathbf{X})\right] + \exp\left[\frac{2}{n}h(\mathbf{Z})\right] \tag{1}$$

where $h(\mathbf{X})$ denotes the differential entropy of $\mathbf{X}$. The equality holds if and only if $\mathbf{X}$ and $\mathbf{Z}$ are Gaussian and with proportional covariance matrices.

In network information theory, most applications focus on the special case of (1) where one of the random vectors is fixed to be Gaussian. In this setting, the classical EPI of Shannon and Stam can be further strengthened as shown by Costa [8]. Let $\mathbf{Z}$ be a Gaussian random n-vector with a positive definite covariance matrix, and let a be a real scalar such that $a \in [0, 1]$. Costa's EPI [8] can be written as

$$\exp\left[\frac{2}{n}h(\mathbf{X}+\sqrt{a}\mathbf{Z})\right] \geq (1-a)\exp\left[\frac{2}{n}h(\mathbf{X})\right] + a\exp\left[\frac{2}{n}h(\mathbf{X}+\mathbf{Z})\right] \tag{2}$$

for any random n-vector $\mathbf{X}$ independent of $\mathbf{Z}$. The equality holds if and only if $\mathbf{X}$ is also Gaussian and with a covariance matrix proportional to that of $\mathbf{Z}$'s.

Though not as widely known as the classical EPI of Shannon and Stam, Costa's EPI has found useful applications in deriving capacity bounds for the Gaussian interference channel [9] and the multiantenna flat-fading channel [10]. The original proof of Costa's EPI provided in [8] was based on rather detailed calculations. Simplified proofs based on a Fisher information inequality [11] and a fundamental relationship between the derivative of mutual information and minimum mean-square error (MMSE) in linear Gaussian channels [12] can be found in [13] and [14], respectively.

Note that Costa's EPI (2) provides a strong relationship among the differential entropies of three random vectors: $\mathbf{X}$, $\mathbf{X}+\sqrt{a}\mathbf{Z}$ and $\mathbf{X}+\mathbf{Z}$. To apply, the increments of $\mathbf{X}+\sqrt{a}\mathbf{Z}$ and $\mathbf{X}+\mathbf{Z}$ over $\mathbf{X}$ need to be Gaussian and have *proportional* covariance matrices. For some applications in network information theory (as we will see shortly), the proportionality requirement may turn out to be overly restrictive. A main contribution of this paper is to prove a natural generalization of Costa's EPI (2) by replacing the real scalar a with a positive semidefinite *matrix* parameter. The result is summarized in the following theorem.

Theorem 1 (Generalized Costa's EPI): Let $\mathbf{Z}$ be a Gaussian random n-vector with a positive definite covariance matrix $\mathbf{N}$, and let $\mathbf{A}$ be an $n \times n$ real symmetric matrix such that $0 \preceq \mathbf{A} \preceq \mathbf{I}$. Here, $\mathbf{I}$ denotes the $n \times n$ identity matrix, and "$\preceq$" denotes "less or equal to" in the positive semidefinite partial ordering between real symmetric matrices. Then,

$$\exp\left[\frac{2}{n}h(\mathbf{X}+\mathbf{A}^{\frac{1}{2}}\mathbf{Z})\right] \geq |\mathbf{I}-\mathbf{A}|^{\frac{1}{n}}\exp\left[\frac{2}{n}h(\mathbf{X})\right] + |\mathbf{A}|^{\frac{1}{n}}\exp\left[\frac{2}{n}h(\mathbf{X}+\mathbf{Z})\right] \tag{3}$$

for any random n-vector $\mathbf{X}$ independent of $\mathbf{Z}$. The equality holds if $\mathbf{Z}$ is Gaussian and with a covariance matrix $\mathbf{B}$ such that $\mathbf{B} - \mathbf{A}\mathbf{B}$ and $\mathbf{B} + \mathbf{A}^{\frac{1}{2}}\mathbf{N}\mathbf{A}^{\frac{1}{2}}$ are proportional.

Note that when $\mathbf{A} = a\mathbf{I}$, the generalized Costa EPI (3) reduces to the original Costa EPI (2). On the other hand, when $\mathbf{A}$ is not a scaled identity, the covariance matrices of increments of $\mathbf{X} + \mathbf{A}^{\frac{1}{2}}\mathbf{Z}$ and $\mathbf{X} + \mathbf{Z}$ over $\mathbf{X}$ do not need to be proportional. As we will see, the ability to cope with a *general* matrix parameter makes the generalized Costa EPI more flexible and powerful than the original Costa EPI.

A different but related generalization of Costa's EPI was considered by Payaró and Palomar [15], where they examined the concavity of the entropy-power $\exp\left[\frac{2}{n}h(\mathbf{A}^{\frac{1}{2}}\mathbf{X} + \mathbf{Z})\right]$ with respect to the matrix parameter $\mathbf{A}$. This line of research was motivated by the observation that the original Costa EPI (2) is equivalent to the concavity of the entropy power $\exp\left[\frac{2}{n}h(\sqrt{a}\mathbf{X} + \mathbf{Z})\right]$ with respect to the scalar parameter a. Unlike the scalar case, Payaró and Palomar [15] showed that the entropy-power $\exp\left[\frac{2}{n}h(\mathbf{A}^{\frac{1}{2}}\mathbf{X} + \mathbf{Z})\right]$ is in general *not* concave with respect to the matrix parameter $\mathbf{A}$. However, the concavity does hold when $\mathbf{A}$ is restricted to be *diagonal* [15].

In information theory, a main application of the EPI is to derive extremal entropy inequalities, which can then be used to solve network communication problems. In their work [16], Liu and Viswanath derived an extremal entropy inequality based on the classical EPI of Shannon [1] and Stam [2] and used it to establish the private message capacity region of the vector Gaussian broadcast channel via the Marton outer bound [17, Theorem 5]. In this paper, we will derive a new extremal entropy inequality based on the generalized Costa EPI and use it to characterize the secrecy capacity regions of the degraded vector Gaussian broadcast channel with layered confidential messages.

The rest of the paper is organized as follows. In Section II, we summarize the main results of the paper, including a new extremal entropy inequality and its applications on the degraded vector Gaussian broadcast channel with layered confidential messages. In Section III, we prove the generalized Costa EPI, following a perturbation approach via a fundamental relationship between the derivative of mutual information and MMSE estimate in linear vector Gaussian channels [18, Theorem 2]. In Section IV, we derive the new extremal entropy inequality from the generalized Costa EPI. The coding theorems for the degraded vector Gaussian broadcast channel with layered confidential messages are proved in Section V and Section VI. Finally, in Section VII, we conclude the paper with some remarks.

II. SUMMARY OF MAIN RESULTS

The following notation will be used throughout the paper. A random vector is denoted with an upper-case letter (e.g., $\mathbf{X}$), its realization is denoted with the corresponding lower-case letter (e.g., $\mathbf{x}$), and its probability density function is denoted with $p(\mathbf{x}) = p_{\mathbf{X}}(\mathbf{x})$. We use $\mathsf{E}[\mathbf{X}]$ to denote the expectation of $\mathbf{X}$. Thus, the covariance matrix of $\mathbf{X}$ is given by

$$\mathsf{Cov}(\mathbf{X}) = \mathsf{E}\left[(\mathbf{X} - \mathsf{E}[\mathbf{X}])(\mathbf{X} - \mathsf{E}[\mathbf{X}])^{\mathsf{T}}\right].$$

Given any jointly distributed random vectors $(\mathbf{X}, \mathbf{Y})$, the MMSE estimate of $\mathbf{X}$ from the observation $\mathbf{Y}$ is the conditional mean $\mathsf{E}[\mathbf{X}|\mathbf{Y}]$. The MMSE (matrix) is given by:

$$\mathrm{Cov}(\mathbf{X}|\mathbf{Y}) = \mathsf{E}\left[(\mathbf{X} - \mathsf{E}[\mathbf{X}|\mathbf{Y}])(\mathbf{X} - \mathsf{E}[\mathbf{X}|\mathbf{Y}])^{\mathsf{T}}\right].$$

A. A New Extremal Entropy Inequality

The following extremal entropy inequality is a consequence of the generalized Costa EPI.

Theorem 2: Let $\mathbf{Z}_k$, $k = 0, \ldots, K$, be a total of $K + 1$ Gaussian random n-vectors with positive definite covariance matrices $\mathbf{N}_k$, respectively. Assume that $\mathbf{N}_1 \preceq \ldots \preceq \mathbf{N}_K$. If there exists an $n \times n$ positive semidefinite matrix $\mathbf{B}^*$ such that

$$\sum_{k=1}^{K} \mu_k (\mathbf{B}^* + \mathbf{N}_k)^{-1} + \mathbf{M}_1 = (\mathbf{B}^* + \mathbf{N}_0)^{-1} + \mathbf{M}_2 \tag{4}$$

for some $n \times n$ positive semidefinite matrices $\mathbf{M}_1$, $\mathbf{M}_2$ and $\mathbf{S}$ with

$$\mathbf{B}^* \mathbf{M}_1 = 0 \tag{5}$$

$$\text{and} \quad (\mathbf{S} - \mathbf{B}^*)\mathbf{M}_2 = 0 \tag{6}$$

and real scalars $\mu_k \geq 0$ with $\sum_{k=1}^{K} \mu_k = 1$, then

$$\sum_{k=1}^{K} \mu_k h(\mathbf{X} + \mathbf{Z}_k | U) - h(\mathbf{X} + \mathbf{Z}_0 | U) \leq \sum_{k=1}^{K} \frac{\mu_k}{2} \log |\mathbf{B}^* + \mathbf{N}_k| - \frac{1}{2} \log |\mathbf{B}^* + \mathbf{N}_0| \tag{7}$$

for any $(\mathbf{X}, U)$ independent of $(\mathbf{Z}_0, \ldots, \mathbf{Z}_K)$ such that $\mathsf{E}[\mathbf{X}\mathbf{X}^{\mathsf{T}}] \preceq \mathbf{S}$.

Note that (4)–(6) are precisely the Karush-Kuhn-Tucker (KKT) conditions (see [7, Appendix IV] and [19, Section 5.2]) for the optimization program:

$$\max_{0 \preceq \mathbf{B} \preceq \mathbf{S}} \left[\sum_{k=1}^{K} \frac{\mu_k}{2} \log |\mathbf{B} + \mathbf{N}_k| - \frac{1}{2} \log |\mathbf{B} + \mathbf{N}_0| \right].$$

Therefore, (7) implies that a jointly *Gaussian* $(U, \mathbf{X})$ such that for each $U = u$, $\mathbf{X}$ has the *same* covariance matrix is an optimal solution to the optimization program:

$$\max_{(U, \mathbf{X})} \left[\sum_{k=1}^{K} \mu_k h(\mathbf{X} + \mathbf{Z}_k | U) - h(\mathbf{X} + \mathbf{Z}_0 | U) \right]$$

where the maximization is over all $(U, \mathbf{X})$ independent of $(\mathbf{Z}_0, \ldots, \mathbf{Z}_K)$ such that $\mathsf{E}[\mathbf{X}\mathbf{X}^{\mathsf{T}}] \preceq \mathbf{S}$. Note that when $K = 1$, this is a special case of [16, Theorem 8] with $\mu = 1$.

(a) Communication scenario 1
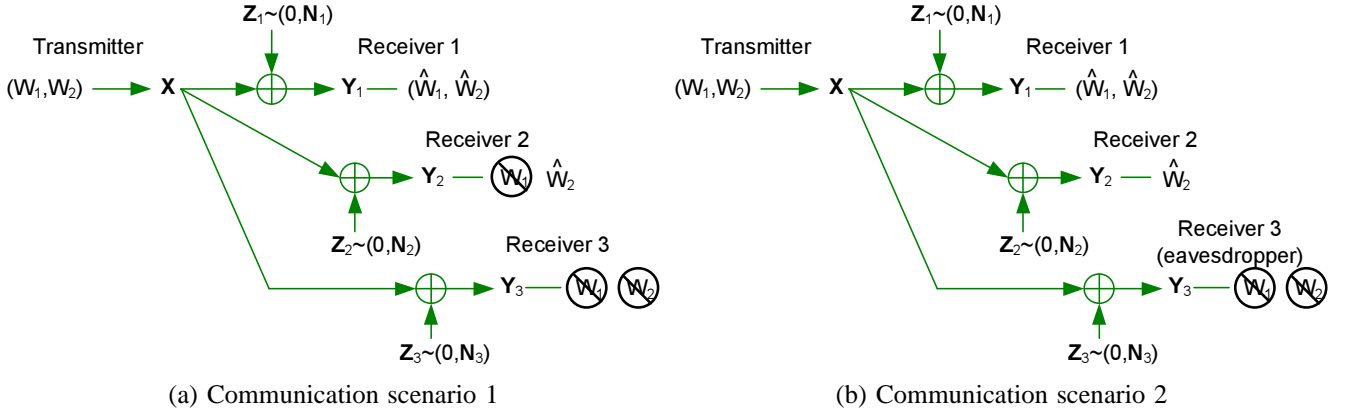
(b) Communication scenario 2

Fig. 1.   Degraded vector Gaussian broadcast channel with layered confidential messages

B. Applications on the Degraded Vector Gaussian Broadcast Channel with Layered Confidential Messages

Consider the following vector Gaussian broadcast channel with three receivers:

$$\mathbf{Y}_k[t] = \mathbf{X}[t] + \mathbf{Z}_k[t], \quad k = 1, 2, 3 \tag{8}$$

where $\{\mathbf{Z}_k[t]\}_t$, $k = 1, 2, 3$, are independent and identically distributed additive vector Gaussian noise processes with zero means and positive definite covariance matrices $\mathbf{N}_k$, respectively. The channel input $\{\mathbf{X}[t]\}_t$ is subject to a matrix constraint:

$$\frac{1}{n}\sum_{t=1}^{n}\mathbf{X}[t]\mathbf{X}^{\mathsf{T}}[t] \preceq \mathbf{S} \tag{9}$$

where $\mathbf{S}$ is a positive semidefinite matrix, and n is the block length. We assume that the noise covariance matrices are ordered as

$$\mathbf{N}_1 \preceq \mathbf{N}_2 \preceq \mathbf{N}_3, \tag{10}$$

i.e., the received signal $\mathbf{Y}_3[t]$ is (stochastically) degraded with respect to $\mathbf{Y}_2[t]$, which is further degraded with respect to $\mathbf{Y}_1[t]$.

We consider two different communication scenarios, both with two independent messages W_1 and W_2. In the first scenario (see Fig. 1-(a)), message W_1 is intended for receiver 1 but needs to be kept secret from receivers 2 and 3, and message W_2 is intended for receivers 1 and 2 but needs to be kept confidential from receiver 3. In the second scenario (see Fig. 1-(b)), message W_1 is intended for receivers 1 but needs to be kept secret from receiver receiver 3, and message W_2 is intended for receivers 1 but needs to be kept secret from receiver 3. The confidentiality of the messages at the unintended receivers is measured using the normalized information-theoretic

criteria [20], [21]:

$$\frac{1}{n}I(W_1; \mathbf{Y}_2^n) \to 0, \quad \frac{1}{n}I(W_1; \mathbf{Y}_3^n) \to 0, \quad \text{and} \quad \frac{1}{n}I(W_2; \mathbf{Y}_3^n) \to 0 \tag{11}$$

for the first scenario and

$$\frac{1}{n}I(W_1; \mathbf{Y}_3^n) \to 0, \quad \text{and} \quad \frac{1}{n}I(W_2; \mathbf{Y}_3^n) \to 0 \tag{12}$$

for the second scenario. Here, the limits are taken as the block length $n \to \infty$. The goal is to characterize the entire secrecy rate region $\mathcal{C}_s = \{(R_1, R_2)\}$ that can be achieved by any coding scheme.

To characterize the secrecy capacity regions, we will first consider the discrete memoryless version of the problem with transition probability $p(y_1, y_2, y_3|x)$ and degradedness order

$$X \to Y_1 \to Y_2 \to Y_3. \tag{13}$$

We have the following single-letter characterizations of the secrecy capacity regions.

Theorem 3: The secrecy capacity region of the discrete memoryless broadcast channel $p(y_1, y_2, y_3|x)$ with confidential messages W_1 (intended for receiver 1 but needs to be kept secret from receivers 2 and 3) and W_2 (intended for receivers 1 and 2 but needs to be kept secret from receiver 3) under the degradedness order (13) is given by the set of nonnegative rate pairs (R_1, R_2) such that

$$R_1 \le I(X; Y_1|U) - I(X; Y_2|U)$$

$$\text{and} \qquad R_2 \le I(U; Y_2) - I(U; Y_3) \tag{14}$$

for some jointly distributed (U, X) satisfying the Markov relation

$$U \to X \to (Y_1, Y_2, Y_3).$$

Theorem 4 ([22, Theorem 2]): The secrecy capacity region of the discrete memoryless broadcast channel $p(y_1, y_2, y_3|x)$ with confidential messages W_1 (intended for receiver 1 but needs to be kept secret from receiver 3) and W_2 (intended for receivers 1 and 2 but needs to be kept secret from receiver 3) under the degradedness order (13) is given by the set of nonnegative rate pairs (R_1, R_2) such that

$$R_1 \le I(X; Y_1|U) - I(X; Y_3|U)$$

$$\text{and} \qquad R_2 \le I(U; Y_2) - I(U; Y_3) \tag{15}$$

for some jointly distributed (U, X) satisfying the Markov relation

$$U \to X \to (Y_1, Y_2, Y_3).$$

A proof of Theorem 4 can be found in [22]. Theorem 3 can be proved in a similar fashion; for completeness, a proof is included in Appendix I. For the vector Gaussian broadcast channel (8) under the degradedness order (10), the single-letter expressions (14) and (15) can be further evaluated using the extremal entropy inequality (7). The results are summarized in the following theorems.

Theorem 5: The secrecy capacity region of the vector Gaussian broadcast channel (8) with confidential messages W_1 (intended for receiver 1 but needs to be kept secret from receivers 2 and 3) and W_2 (intended for receiver 1 and 2 but needs to be kept secret from receiver 3) and degradedness order (10) under the matrix constraint (9) is given by the set of nonnegative secrecy rate pairs (R_1, R_2) such that

$$R_1 \leq \frac{1}{2} \log \left| \frac{\mathbf{B} + \mathbf{N}_1}{\mathbf{N}_1} \right| - \frac{1}{2} \log \left| \frac{\mathbf{B} + \mathbf{N}_2}{\mathbf{N}_2} \right|$$

and

$$R_2 \leq \frac{1}{2} \log \left| \frac{\mathbf{S} + \mathbf{N}_2}{\mathbf{B} + \mathbf{N}_2} \right| - \frac{1}{2} \log \left| \frac{\mathbf{S} + \mathbf{N}_3}{\mathbf{B} + \mathbf{N}_3} \right| \tag{16}$$

for some $0 \preceq \mathbf{B} \preceq \mathbf{S}$.

Theorem 6: The secrecy capacity region of the vector Gaussian broadcast channel (8) with confidential messages W_1 (intended for receiver 1 but needs to be kept secret from receiver 3) and W_2 (intended for receivers 1 and 2 but needs to be kept secret from receiver 3) and degradedness order (10) under the matrix constraint (9) is given by the set of nonnegative secrecy rate pairs (R_1, R_2) such that

$$R_1 \leq \frac{1}{2} \log \left| \frac{\mathbf{B} + \mathbf{N}_1}{\mathbf{N}_1} \right| - \frac{1}{2} \log \left| \frac{\mathbf{B} + \mathbf{N}_3}{\mathbf{N}_3} \right|$$

and

$$R_2 \leq \frac{1}{2} \log \left| \frac{\mathbf{S} + \mathbf{N}_2}{\mathbf{B} + \mathbf{N}_2} \right| - \frac{1}{2} \log \left| \frac{\mathbf{S} + \mathbf{N}_3}{\mathbf{B} + \mathbf{N}_3} \right| \tag{17}$$

for some $0 \preceq \mathbf{B} \preceq \mathbf{S}$.

III. Proof of Theorem 1

In this section, we prove the generalized Costa EPI (3) as stated in Theorem 1. We first examine the equality condition. Note that when $\mathbf{X}$ is Gaussian, the generalized Costa EPI (3) becomes the matrix inequality:

$$|\mathbf{B} + \mathbf{A}^{\frac{1}{2}} \mathbf{N} \mathbf{A}^{\frac{1}{2}}|^{\frac{1}{n}} \geq |\mathbf{B} - \mathbf{A}\mathbf{B}|^{\frac{1}{n}} + |\mathbf{A}\mathbf{B} + \mathbf{A}\mathbf{N}|^{\frac{1}{n}}.$$

Suppose that $\mathbf{B} - \mathbf{A}\mathbf{B}$ and $\mathbf{B} + \mathbf{A}^{\frac{1}{2}} \mathbf{N} \mathbf{A}^{\frac{1}{2}}$ are proportional, i.e., there exists a real scalar c such that

$$\mathbf{B} + \mathbf{A}^{\frac{1}{2}} \mathbf{N} \mathbf{A}^{\frac{1}{2}} = c(\mathbf{B} - \mathbf{A}\mathbf{B}).$$

Since both matrices $\mathbf{A}$ and $\mathbf{B}$ are symmetric, this implies that $\mathbf{AB}$ is also symmetric, i.e.,

$$\mathbf{AB} = \mathbf{B}^\mathsf{T}\mathbf{A}^\mathsf{T} = \mathbf{BA}.$$

Therefore, $\mathbf{A}$ and $\mathbf{B}$ must have the *same* eigenvector matrix [23] and hence

$$\mathbf{AB} = \mathbf{A}^{\frac{1}{2}}\mathbf{BA}^{\frac{1}{2}}.$$

It follows that

$$\mathbf{A}^{\frac{1}{2}}\mathbf{BA}^{\frac{1}{2}} + \mathbf{A}^{\frac{1}{2}}\mathbf{NA}^{\frac{1}{2}} = \mathbf{B} + \mathbf{A}^{\frac{1}{2}}\mathbf{NA}^{\frac{1}{2}} - (\mathbf{B} - \mathbf{AB})$$
$$= (c-1)(\mathbf{B} - \mathbf{AB})$$

i.e., $\mathbf{A}^{\frac{1}{2}}\mathbf{BA}^{\frac{1}{2}} + \mathbf{A}^{\frac{1}{2}}\mathbf{NA}^{\frac{1}{2}}$ and $\mathbf{B} - \mathbf{AB}$ are proportional. Therefore,

$$|\mathbf{B} + \mathbf{A}^{\frac{1}{2}}\mathbf{NA}^{\frac{1}{2}}|^{\frac{1}{n}} = |\mathbf{B} - \mathbf{AB} + (\mathbf{A}^{\frac{1}{2}}\mathbf{BA}^{\frac{1}{2}} + \mathbf{A}^{\frac{1}{2}}\mathbf{NA}^{\frac{1}{2}})|^{\frac{1}{n}}$$
$$= |\mathbf{B} - \mathbf{AB}|^{\frac{1}{n}} + |\mathbf{A}^{\frac{1}{2}}\mathbf{BA}^{\frac{1}{2}} + \mathbf{A}^{\frac{1}{2}}\mathbf{NA}^{\frac{1}{2}}|^{\frac{1}{n}}$$
$$= |\mathbf{B} - \mathbf{AB}|^{\frac{1}{n}} + |\mathbf{AB} + \mathbf{AN}|^{\frac{1}{n}}.$$

This proved the desired equality condition.

We now turn to the proof of the inequality. First consider the special case when $|\mathbf{A}| = 0$. Since

$$h(\mathbf{X} + \mathbf{A}^{\frac{1}{2}}\mathbf{Z}) - h(\mathbf{X}) = I(\mathbf{A}^{\frac{1}{2}}\mathbf{Z}; \mathbf{X} + \mathbf{A}^{\frac{1}{2}}\mathbf{Z}) \geq 0,$$

we have

$$\exp\left[\frac{2}{n}h(\mathbf{X} + \mathbf{A}^{\frac{1}{2}}\mathbf{Z})\right] \geq \exp\left[\frac{2}{n}h(\mathbf{X})\right]$$
$$\geq |\mathbf{I} - \mathbf{A}|^{\frac{1}{n}}\exp\left[\frac{2}{n}h(\mathbf{X})\right]$$

where the last inequality follows from the assumption that $0 \preceq \mathbf{A} \preceq \mathbf{I}$ and hence $0 \leq |\mathbf{I} - \mathbf{A}| \leq 1$.

Next, consider the general case when $|\mathbf{A}| > 0$. The proof is rather long so we divide it into several steps.

Step 1–Constructing a monotone path. To prove the generalized Costa EPI (3), we can equivalently show that

$$\exp\left[\frac{2}{n}h(\mathbf{X} + \mathbf{Z})\right] \leq |\mathbf{A}|^{-\frac{1}{n}}\exp\left[\frac{2}{n}h(\mathbf{X} + \mathbf{A}^{\frac{1}{2}}\mathbf{Z})\right] - \left(\frac{|\mathbf{I} - \mathbf{A}|}{|\mathbf{A}|}\right)^{\frac{1}{n}}\exp\left[\frac{2}{n}h(\mathbf{X})\right]. \tag{18}$$

Since $\mathbf{X}$ and $\mathbf{Z}$ are independent, we have

$$
\begin{aligned}
h(\mathbf{X} + \mathbf{A}^{\frac{1}{2}}\mathbf{Z}) - h(\mathbf{X}) &= h(\mathbf{A}^{-\frac{1}{2}}\mathbf{X} + \mathbf{Z}) - h(\mathbf{A}^{-\frac{1}{2}}\mathbf{X}) \\
&= h(\mathbf{A}^{-\frac{1}{2}}\mathbf{X} + \mathbf{Z}) - h(\mathbf{A}^{-\frac{1}{2}}\mathbf{X}|\mathbf{Z}) \\
&= I(\mathbf{Z}; \mathbf{A}^{-\frac{1}{2}}\mathbf{X} + \mathbf{Z})
\end{aligned}
\tag{19}
$$

and

$$
h(\mathbf{X} + \mathbf{Z}) - h(\mathbf{X}) = I(\mathbf{Z}; \mathbf{X} + \mathbf{Z}).
\tag{20}
$$

Divide both sides of (18) by $\exp\left[\frac{2}{n}h(\mathbf{X})\right]$ and use (19) and (20). Then, (18) can be equivalently written as

$$
\exp\left[\frac{2}{n}I(\mathbf{Z}; \mathbf{X} + \mathbf{Z})\right] \leq |\mathbf{A}|^{-\frac{1}{n}}\left\{\exp\left[\frac{2}{n}I(\mathbf{Z}; \mathbf{A}^{-\frac{1}{2}}\mathbf{X} + \mathbf{Z})\right] - |\mathbf{I} - \mathbf{A}|^{\frac{1}{n}}\right\}.
\tag{21}
$$

Let

$$
F(\mathbf{D}) := |\mathbf{D}|^{\frac{2}{n}}\left\{\exp\left[\frac{2}{n}I(\mathbf{Z}; \mathbf{D}\mathbf{X} + \mathbf{Z})\right] - |\mathbf{I} - \mathbf{D}^{-2}|^{\frac{1}{n}}\right\}.
\tag{22}
$$

With this definition, (21) can be equivalently written as

$$
F(\mathbf{I}) \leq F(\mathbf{A}^{-\frac{1}{2}}).
\tag{23}
$$

To show the inequality (23), it is sufficient to construct a family of $n \times n$ positive definite matrices $\{\mathbf{D}(\gamma)\}_\gamma$ connecting $\mathbf{I}$ and $\mathbf{A}^{-\frac{1}{2}}$ such that $F(\mathbf{D}(\gamma))$ is monotone along the path. Unlike the scalar case where there is only one path connecting 1 to $1/\sqrt{a}$, in the matrix case there are infinitely many paths connecting $\mathbf{I}$ and $\mathbf{A}^{-\frac{1}{2}}$. Here, we consider the special choice

$$
\mathbf{D}(\gamma) := \left[\mathbf{I} + \gamma(\mathbf{A}^{-1} - \mathbf{I})\right]^{\frac{1}{2}}
\tag{24}
$$

and show that

$$
\frac{\partial F}{\partial \gamma} \geq 0, \quad \forall \gamma \in [0, 1].
\tag{25}
$$

along this particular path.

Step 2–Calculating the derivative $\frac{\partial F}{\partial \gamma}$. Following [14, Theorem 5], we have

$$
I(\mathbf{Z}; \mathbf{D}\mathbf{X} + \mathbf{Z}) = I(\mathbf{X}; \mathbf{D}\mathbf{X} + \mathbf{Z}) + h(\mathbf{Z}) - h(\mathbf{X}) - \log|\mathbf{D}|
$$

and

$$
\mathsf{Cov}(\mathbf{X}|\mathbf{D}\mathbf{X} + \mathbf{Z}) = \mathbf{D}^{-1}\mathsf{Cov}(\mathbf{Z}|\mathbf{D}\mathbf{X} + \mathbf{Z})\mathbf{D}^{-\mathsf{T}}.
$$

Let $\mathbf{N} := \mathsf{Cov}(\mathbf{Z})$ and note that $\mathbf{D}$ is symmetric. We have

$$
\begin{aligned}
\frac{\partial}{\partial \mathbf{D}} I(\mathbf{Z}; \mathbf{D}\mathbf{X} + \mathbf{Z}) &= \frac{\partial}{\partial \mathbf{D}} I(\mathbf{X}; \mathbf{D}\mathbf{X} + \mathbf{Z}) - \mathbf{D}^{-1} \\
&= \mathbf{N}^{-1} \mathbf{D} \, \mathsf{Cov}(\mathbf{X} | \mathbf{D}\mathbf{X} + \mathbf{Z}) - \mathbf{D}^{-1} \\
&= \left(\mathbf{N}^{-1} \mathsf{Cov}(\mathbf{Z} | \mathbf{D}\mathbf{X} + \mathbf{Z}) - \mathbf{I} \right) \mathbf{D}^{-1}
\end{aligned}
\tag{26}
$$

where the second equality follows from the fundamental relationship between the derivative of mutual information and MMSE estimate in linear vector Gaussian channels as stated in [18, Theorem 2].

From (26), the derivative $\frac{\partial F}{\partial \mathbf{D}}$ can be calculated as

$$
\begin{aligned}
\frac{\partial F}{\partial \mathbf{D}} =& \frac{2}{n} |\mathbf{D}|^{\frac{2}{n}} \mathbf{D}^{-1} \left\{ \exp\left[\frac{2}{n} I(\mathbf{Z}; \mathbf{D}\mathbf{X} + \mathbf{Z}) \right] - |\mathbf{I} - \mathbf{D}^{-2}|^{\frac{1}{n}} \right\} + \\
& |\mathbf{D}|^{\frac{2}{n}} \left\{ \frac{2}{n} \exp\left[\frac{2}{n} I(\mathbf{Z}; \mathbf{D}\mathbf{X} + \mathbf{Z}) \right] \frac{\partial I(\mathbf{Z}; \mathbf{D}\mathbf{X} + \mathbf{Z})}{\partial \mathbf{D}} - \frac{2}{n} |\mathbf{I} - \mathbf{D}^{-2}|^{\frac{1}{n}} (\mathbf{I} - \mathbf{D}^{-2})^{-1} \mathbf{D}^{-3} \right\} \\
=& \frac{2}{n} |\mathbf{D}|^{\frac{2}{n}} \left\{ \left\{ \exp\left[\frac{2}{n} I(\mathbf{Z}; \mathbf{D}\mathbf{X} + \mathbf{Z}) \right] - |\mathbf{I} - \mathbf{D}^{-2}|^{\frac{1}{n}} \right\} \mathbf{I} + \right. \\
& \left. \exp\left[\frac{2}{n} I(\mathbf{Z}; \mathbf{D}\mathbf{X} + \mathbf{Z}) \right] (\mathbf{N}^{-1} \mathsf{Cov}(\mathbf{Z} | \mathbf{D}\mathbf{X} + \mathbf{Z}) - \mathbf{I}) - |\mathbf{I} - \mathbf{D}^{-2}|^{\frac{1}{n}} (\mathbf{D}^2 - \mathbf{I})^{-1} \right\} \mathbf{D}^{-1} \\
=& \frac{2}{n} |\mathbf{D}|^{\frac{2}{n}} \left\{ \exp\left[\frac{2}{n} I(\mathbf{Z}; \mathbf{D}\mathbf{X} + \mathbf{Z}) \right] \mathbf{N}^{-1} \mathsf{Cov}(\mathbf{Z} | \mathbf{D}\mathbf{X} + \mathbf{Z}) - |\mathbf{I} - \mathbf{D}^{-2}|^{\frac{1}{n}} \left[\mathbf{I} + (\mathbf{D}^2 - \mathbf{I})^{-1} \right] \right\} \mathbf{D}^{-1}.
\end{aligned}
\tag{27}
$$

The derivative $\frac{\partial \mathbf{D}}{\partial \gamma}$ can be calculated as

$$
\begin{aligned}
\frac{\partial \mathbf{D}}{\partial \gamma} &= \frac{1}{2} \left[\mathbf{I} + \gamma(\mathbf{A}^{-1} - \mathbf{I}) \right]^{-\frac{1}{2}} (\mathbf{A}^{-1} - \mathbf{I}) \\
&= \frac{1}{2\gamma} \mathbf{D}^{-1} (\mathbf{D}^2 - \mathbf{I}) \\
&= \frac{1}{2\gamma} \mathbf{D} (\mathbf{I} - \mathbf{D}^{-2}).
\end{aligned}
\tag{28}
$$

By (27), (28) and the chain rule of differentiation [24, Chapter 17.5],

$$
\begin{aligned}
\frac{\partial F}{\partial \gamma} &= \mathsf{Tr} \left\{ \frac{\partial F}{\partial \mathbf{D}} \frac{\partial \mathbf{D}}{\partial \gamma} \right\} \\
&= \frac{|\mathbf{D}|^{\frac{2}{n}}}{n} \mathsf{Tr} \left\{ \left[\exp\left[\frac{2}{n} I(\mathbf{Z}; \mathbf{D}\mathbf{X} + \mathbf{Z}) \right] \mathbf{N}^{-1} \mathsf{Cov}(\mathbf{Z} | \mathbf{D}\mathbf{X} + \mathbf{Z}) - |\mathbf{I} - \mathbf{D}^{-2}|^{\frac{1}{n}} \left[\mathbf{I} + (\mathbf{D}^2 - \mathbf{I})^{-1} \right] \right] \frac{\mathbf{I} - \mathbf{D}^{-2}}{\gamma} \right\} \\
&= \frac{|\mathbf{D}|^{\frac{2}{n}}}{n\gamma} \mathsf{Tr} \left\{ \exp\left[\frac{2}{n} I(\mathbf{Z}; \mathbf{D}\mathbf{X} + \mathbf{Z}) \right] \mathbf{N}^{-1} \mathsf{Cov}(\mathbf{Z} | \mathbf{D}\mathbf{X} + \mathbf{Z})(\mathbf{I} - \mathbf{D}^{-2}) - |\mathbf{I} - \mathbf{D}^{-2}|^{\frac{1}{n}} \mathbf{I} \right\} \\
&= \frac{|\mathbf{D}|^{\frac{2}{n}}}{n\gamma} \left\{ \exp\left[\frac{2}{n} I(\mathbf{Z}; \mathbf{D}\mathbf{X} + \mathbf{Z}) \right] \mathsf{Tr} \left\{ \mathbf{N}^{-1} \mathsf{Cov}(\mathbf{Z} | \mathbf{D}\mathbf{X} + \mathbf{Z})(\mathbf{I} - \mathbf{D}^{-2}) \right\} - n |\mathbf{I} - \mathbf{D}^{-2}|^{\frac{1}{n}} \right\}.
\end{aligned}
\tag{29}
$$

Step 3–Proving $\frac{\partial F}{\partial \gamma} \geq 0$. The mutual information $I(\mathbf{Z}; \mathbf{DX} + \mathbf{Z})$ can be bounded from below as follows:

$$
\begin{aligned}
I(\mathbf{Z}; \mathbf{DX} + \mathbf{Z}) &\geq I(\mathbf{Z}; \mathsf{E}[\mathbf{Z}|\mathbf{DX} + \mathbf{Z}]) \\
&= h(\mathbf{Z}) - h(\mathbf{Z}|\mathsf{E}[\mathbf{Z}|\mathbf{DX} + \mathbf{Z}]) \\
&= \frac{1}{2} \log(2\pi e)^n |\mathbf{N}| - h(\mathbf{Z} - \mathsf{E}[\mathbf{Z}|\mathbf{DX} + \mathbf{Z}]|\mathsf{E}[\mathbf{Z}|\mathbf{DX} + \mathbf{Z}]) \\
&\geq \frac{1}{2} \log(2\pi e)^n |\mathbf{N}| - h(\mathbf{Z} - \mathsf{E}[\mathbf{Z}|\mathbf{DX} + \mathbf{Z}]) \\
&\geq \frac{1}{2} \log(2\pi e)^n |\mathbf{N}| - \frac{1}{2} \log(2\pi e)^n |\mathsf{Cov}(\mathbf{Z}|\mathbf{DX} + \mathbf{Z})| \\
&= \frac{1}{2} \log \frac{|\mathbf{N}|}{|\mathsf{Cov}(\mathbf{Z}|\mathbf{DX} + \mathbf{Z})|}.
\end{aligned}
\tag{30}
$$

Here, the first inequality follows from the Markov relation

$$
\mathbf{Z} \to \mathbf{DX} + \mathbf{Z} \to \mathsf{E}[\mathbf{Z}|\mathbf{DX} + \mathbf{Z}]
$$

and the chain rule of mutual information [25, Chapter 2.8]; the second inequality follows from the fact that conditioning reduces differential entropy [25, Chapter 9.6]; and the third inequality follows from the well-known fact that Gaussian maximizes differential entropy for a given covariance matrix [25, Chapter 9.6]. By (30),

$$
\begin{aligned}
|\mathbf{I} - \mathbf{D}^{-2}|^{\frac{1}{n}} \exp\left[-\frac{2}{n} I(\mathbf{Z}; \mathbf{DX} + \mathbf{Z})\right] &\leq |\mathbf{N}^{-1}\mathsf{Cov}(\mathbf{Z}|\mathbf{DX} + \mathbf{Z})(\mathbf{I} - \mathbf{D}^{-2})|^{\frac{1}{n}} \\
&\leq \frac{1}{n} \mathsf{Tr}\left\{\mathbf{N}^{-1}\mathsf{Cov}(\mathbf{Z}|\mathbf{DX} + \mathbf{Z})(\mathbf{I} - \mathbf{D}^{-2})\right\}
\end{aligned}
\tag{31}
$$

where the last inequality follows from the well-known inequality of arithmetic and geometric means [26, p. 136].

Finally, substituting (31) into (29) establishes the fact that $\frac{\partial F}{\partial \gamma} \geq 0$ for all $\gamma \in [0, 1]$. In particular, we have $F(\mathbf{D}(1)) \geq F(\mathbf{D}(0))$. This proved the desired inequality (21) and hence the generalized Costa EPI (3).

IV. PROOF OF THEOREM 2

In this section, we prove the extremal entropy inequality (7) as stated in Theorem 2. We will first state a series of corollaries of Theorem 1, as intermediate results leading to Theorem 2. Based on the final corollary, we will prove Theorem 2 using an *enhancement* argument.

Corollary 1: Let $\mathbf{Z}$ be a Gaussian random n-vector with a positive definite covariance matrix, and let $\mathbf{A}$ be an $n \times n$ positive real symmetric matrix such that $0 \preceq \mathbf{A} \preceq \mathbf{I}$. Then

$$
\exp\left[\frac{2}{n} h(\mathbf{X} + \mathbf{A}^{\frac{1}{2}}\mathbf{Z}|U)\right] \geq |\mathbf{I} - \mathbf{A}|^{\frac{1}{n}} \exp\left[\frac{2}{n} h(\mathbf{X}|U)\right] + |\mathbf{A}|^{\frac{1}{n}} \exp\left[\frac{2}{n} h(\mathbf{X} + \mathbf{Z}|U)\right]
\tag{32}
$$

for any $(\mathbf{X}, U)$ independent of $\mathbf{Z}$.

Corollary 2: Let $\mathbf{Z}_1$, $\mathbf{Z}_2$ and $\mathbf{Z}_3$ be Gaussian random n-vectors with positive definite covariance matrices $\mathbf{N}_1$, $\mathbf{N}_2$ and $\mathbf{N}_3$, respectively. Assume that $\mathbf{N}_1 \preceq \mathbf{N}_3$. If there exists an $n \times n$ positive semidefinite matrix $\mathbf{B}^*$ such that

$$(\mathbf{B}^* + \mathbf{N}_1)^{-1} + \mu(\mathbf{B}^* + \mathbf{N}_3)^{-1} = (1 + \mu)(\mathbf{B}^* + \mathbf{N}_2)^{-1} \tag{33}$$

for some real scalar $\mu \geq 0$, then

$$h(\mathbf{X} + \mathbf{Z}_1|U) + \mu h(\mathbf{X} + \mathbf{Z}_3|U) - (1 + \mu)h(\mathbf{X} + \mathbf{Z}_2|U)$$
$$\leq \frac{1}{2} \log |\mathbf{B}^* + \mathbf{N}_1| + \frac{\mu}{2} \log |\mathbf{B}^* + \mathbf{N}_3| - \frac{1 + \mu}{2} \log |\mathbf{B}^* + \mathbf{N}_2| \tag{34}$$

for any $(\mathbf{X}, U)$ independent of $(\mathbf{Z}_1, \mathbf{Z}_2, \mathbf{Z}_3)$.

Corollary 3: Let $\mathbf{Z}_k$, $k = 0, \ldots, K$, be a collection of $K + 1$ Gaussian random n-vectors with respective positive definite covariance matrices $\mathbf{N}_k$. Assume that $\mathbf{N}_1 \preceq \ldots \preceq \mathbf{N}_K$. If there exists an $n \times n$ positive semidefinite matrix $\mathbf{B}^*$ such that

$$\sum_{k=1}^{K} \mu_k(\mathbf{B}^* + \mathbf{N}_k)^{-1} = (\mathbf{B}^* + \mathbf{N}_0)^{-1} \tag{35}$$

for some $\mu_k \geq 0$ with $\sum_{k=1}^{K} \mu_k = 1$, then

$$\sum_{k=1}^{K} \mu_k h(\mathbf{X} + \mathbf{Z}_k|U) - h(\mathbf{X} + \mathbf{Z}_0|U) \leq \sum_{k=1}^{K} \frac{\mu_k}{2} \log |\mathbf{B}^* + \mathbf{N}_k| - \frac{1}{2} \log |\mathbf{B}^* + \mathbf{N}_0| \tag{36}$$

for any $(\mathbf{X}, U)$ independent of $(\mathbf{Z}_0, \ldots, \mathbf{Z}_K)$.

A proof of Corollaries 1, 2 and 3 can be found in Appendices II, III and IV, respectively. We are now ready to prove Theorem 2. Note that the special case with $\mathbf{M}_1 = \mathbf{M}_2 = 0$ was proved in Corollary 3. To extend the result of Corollary 3 to nonzero $\mathbf{M}_1$ and $\mathbf{M}_2$, we will consider an enhancement argument, which was first introduced by Weingarten, Steinberg and Shamai in [7].

Let $\widetilde{\mathbf{N}}_1$ and $\widetilde{\mathbf{N}}_0$ be $n \times n$ real symmetric matrices such that:

$$\mu_1(\mathbf{B}^* + \widetilde{\mathbf{N}}_1)^{-1} = \mu_1(\mathbf{B}^* + \mathbf{N}_1)^{-1} + \mathbf{M}_1 \tag{37}$$

$$\text{and} \quad (\mathbf{B}^* + \widetilde{\mathbf{N}}_0)^{-1} = (\mathbf{B}^* + \mathbf{N}_0)^{-1} + \mathbf{M}_2. \tag{38}$$

As shown in [7, Lemma 11 and 12], $\widetilde{\mathbf{N}}_1$ and $\widetilde{\mathbf{N}}_0$ satisfy the following properties:

$$0 \prec \widetilde{\mathbf{N}}_1 = \left(\mathbf{N}_1^{-1} + \mu_1^{-1}\mathbf{M}_1\right)^{-1} \preceq \mathbf{N}_1, \tag{39}$$

$$\widetilde{\mathbf{N}}_1 \preceq \widetilde{\mathbf{N}}_0 \preceq \mathbf{N}_0, \tag{40}$$

$$\left| \frac{\mathbf{B}^* + \widetilde{\mathbf{N}}_1}{\widetilde{\mathbf{N}}_1} \right| = \left| \frac{\mathbf{B}^* + \mathbf{N}_1}{\mathbf{N}_1} \right| \tag{41}$$

and

$$\left| \frac{\mathbf{S} + \widetilde{\mathbf{N}}_0}{\mathbf{B}^* + \widetilde{\mathbf{N}}_0} \right| = \left| \frac{\mathbf{S} + \mathbf{N}_2}{\mathbf{B}^* + \mathbf{N}_2} \right|. \tag{42}$$

Let $\widetilde{\mathbf{Z}}_0$ and $\widetilde{\mathbf{Z}}_1$ be two Gaussian n-vectors with covariance matrices $\widetilde{\mathbf{N}}_0$ and $\widetilde{\mathbf{N}}_1$, respectively. Note from (39) that $\widetilde{\mathbf{N}}_1 \preceq \mathbf{N}_1 \preceq \mathbf{N}_2 \preceq \ldots \preceq \mathbf{N}_K$. Moreover, substitute (37) and (38) into (4) and we have

$$\mu_1(\mathbf{B}^* + \widetilde{\mathbf{N}}_1)^{-1} + \sum_{k=2}^{K} \mu_k (\mathbf{B}^* + \mathbf{N}_k)^{-1} = (\mathbf{B}^* + \widetilde{\mathbf{N}}_0)^{-1}. \tag{43}$$

Thus, by Corollary 3

$$\mu_1 h(\mathbf{X} + \widetilde{\mathbf{Z}}_1 | U) + \sum_{k=2}^{K} \mu_k h(\mathbf{X} + \mathbf{Z}_k | U) - h(\mathbf{X} + \widetilde{\mathbf{Z}}_0 | U)$$
$$\leq \frac{\mu_1}{2}(\mathbf{B}^* + \widetilde{\mathbf{N}}_1)^{-1} + \sum_{k=2}^{K} \frac{\mu_k}{2} \log |\mathbf{B}^* + \mathbf{N}_k| - \frac{1}{2} \log |\mathbf{B}^* + \widetilde{\mathbf{N}}_0| \tag{44}$$

for any $(\mathbf{X}, U)$ independent of $(\widetilde{\mathbf{Z}}_0, \widetilde{\mathbf{Z}}_1, \mathbf{Z}_2, \ldots, \mathbf{Z}_K)$.

On the other hand, note from (39) that $\widetilde{\mathbf{N}}_1 \preceq \mathbf{N}_1$. We have

$$I(\mathbf{X}; \mathbf{X} + \mathbf{Z}_1 | U) \leq I(\mathbf{X}; \mathbf{X} + \widetilde{\mathbf{Z}}_1 | U)$$

for any $(\mathbf{X}, U)$ independent of $(\mathbf{Z}_1, \widetilde{\mathbf{Z}}_1)$. Thus,

$$h(\mathbf{X} + \widetilde{\mathbf{Z}}_1 | U) - h(\mathbf{X} + \mathbf{Z}_1 | U) \geq h(\widetilde{\mathbf{Z}}_1) - h(\mathbf{Z}_1)$$
$$= \frac{1}{2} \log \left| \frac{\widetilde{\mathbf{N}}_1}{\mathbf{N}_1} \right|$$
$$= \frac{1}{2} \log \left| \frac{\mathbf{B}^* + \widetilde{\mathbf{N}}_1}{\mathbf{B}^* + \mathbf{N}_1} \right| \tag{45}$$

where the last equality follows from (41).

Also note from (40) that $\widetilde{\mathbf{N}}_0 \preceq \mathbf{N}_0$. Let $\hat{\mathbf{Z}}_0$ be a Gaussian n-vector with covariance matrix $\mathbf{N}_0 - \widetilde{\mathbf{N}}_0$ and independent of $(\widetilde{\mathbf{Z}}_0, \mathbf{X}, U)$. We have

$$
\begin{aligned}
h(\mathbf{X} + \mathbf{Z}_0|U) - h(\mathbf{X} + \widetilde{\mathbf{Z}}_0|U) &= h(\mathbf{X} + \widetilde{\mathbf{Z}}_0 + \hat{\mathbf{Z}}_0|U) - h(\mathbf{X} + \widetilde{\mathbf{Z}}_0|U) \\
&= I(\hat{\mathbf{Z}}_0; \mathbf{X} + \widetilde{\mathbf{Z}}_0 + \hat{\mathbf{Z}}_0|U) \\
&\geq I(\hat{\mathbf{Z}}_0; \mathbf{X} + \widetilde{\mathbf{Z}}_0 + \hat{\mathbf{Z}}_0) \\
&\geq \frac{1}{2} \log \left| \frac{\mathsf{Cov}(\mathbf{X}) + \mathbf{N}_0}{\mathsf{Cov}(\mathbf{X}) + \widetilde{\mathbf{N}}_0} \right| \\
&\geq \frac{1}{2} \log \left| \frac{\mathbf{S} + \mathbf{N}_0}{\mathbf{S} + \widetilde{\mathbf{N}}_0} \right| &(46) \\
&= \frac{1}{2} \log \left| \frac{\mathbf{B}^* + \mathbf{N}_0}{\mathbf{B}^* + \widetilde{\mathbf{N}}_0} \right| &(47)
\end{aligned}
$$

for any $(\mathbf{X}, U)$ independent of $(\mathbf{Z}_0, \widetilde{\mathbf{Z}}_0)$ such that $\mathsf{E}[\mathbf{X}\mathbf{X}^\mathsf{T}] \preceq \mathbf{S}$. Here, the first inequality follows from the independence of $\hat{\mathbf{Z}}_0$ and U; the second inequality follows from the worst noise result [27, Lemma II.2]; the third inequality follows from the fact that $\widetilde{\mathbf{N}}_0 \preceq \mathbf{N}_0$ and $\mathsf{Cov}(\mathbf{X}) \preceq \mathsf{E}[\mathbf{X}\mathbf{X}^\mathsf{T}] \preceq \mathbf{S}$; and the last inequality follows from (42).

Finally, put together (44), (45) and (47) and we may obtain

$$
\begin{aligned}
&\sum_{k=1}^{K} \mu_k h(\mathbf{X} + \mathbf{Z}_k|U) - h(\mathbf{X} + \mathbf{Z}_0|U) \\
&= \left[\mu_1 h(\mathbf{X} + \widetilde{\mathbf{Z}}_1|U) + \sum_{k=2}^{K} \mu_k h(\mathbf{X} + \mathbf{Z}_k|U) - h(\mathbf{X} + \widetilde{\mathbf{Z}}_0|U) \right] - \\
&\quad \mu_1 \left[h(\mathbf{X} + \widetilde{\mathbf{Z}}_1|U) - h(\mathbf{X} + \mathbf{Z}_1|U) \right] - \left[h(\mathbf{X} + \mathbf{Z}_0|U) - h(\mathbf{X} + \widetilde{\mathbf{Z}}_0|U) \right] \\
&\leq \left[\frac{\mu_1}{2} (\mathbf{B}^* + \widetilde{\mathbf{N}}_1)^{-1} + \sum_{k=2}^{K} \frac{\mu_k}{2} \log |\mathbf{B}^* + \mathbf{N}_k| - \frac{1}{2} \log |\mathbf{B}^* + \widetilde{\mathbf{N}}_0| \right] - \\
&\quad \frac{\mu_1}{2} \log \left| \frac{\mathbf{B}^* + \widetilde{\mathbf{N}}_1}{\mathbf{B}^* + \mathbf{N}_1} \right| - \frac{1}{2} \log \left| \frac{\mathbf{B}^* + \mathbf{N}_0}{\mathbf{B}^* + \widetilde{\mathbf{N}}_0} \right| \\
&= \sum_{k=1}^{K} \frac{\mu_k}{2} \log |\mathbf{B}^* + \mathbf{N}_k| - \frac{1}{2} \log |\mathbf{B}^* + \mathbf{N}_0|
\end{aligned}
$$

for any $(\mathbf{X}, U)$ independent of $(\mathbf{Z}_0, \mathbf{Z}_1, \ldots, \mathbf{Z}_K)$ such that $\mathsf{E}[\mathbf{X}\mathbf{X}^\mathsf{T}] \preceq \mathbf{S}$. This completes the proof of Theorem 2.

V. PROOF OF THEOREM 5

In this section, we prove Theorem 5. Note that the achievability of the secrecy rate region (16) can be obtained from the secrecy rate region (14) by letting $\mathbf{U}$ and $\mathbf{V}$ be two independent Gaussian vectors with zero means and covariance matrices $\mathbf{S} - \mathbf{B}$ and $\mathbf{B}$, respectively and $\mathbf{X} = \mathbf{U} + \mathbf{V}$. We therefore concentrate on the converse part of the theorem.

To show that (16) is indeed the secrecy capacity region of the vector Gaussian broadcast channel (8), we will consider proof by contradiction. Assume that (R_1^o, R_2^o) is an achievable secrecy rate pair that lies *outside* the secrecy rate region (16). Note that $\mathbf{N}_1 \preceq \mathbf{N}_2$. From [28, Theorem 1], we can bound R_1^o by

$$R_1^o \leq \frac{1}{2} \log \left| \frac{\mathbf{S} + \mathbf{N}_1}{\mathbf{N}_1} \right| - \frac{1}{2} \log \left| \frac{\mathbf{S} + \mathbf{N}_2}{\mathbf{N}_2} \right| = R_1^{max}.$$

Note that when $R_2^o = 0$, R_1^{max} is achievable by letting $\mathbf{B} = \mathbf{S}$ in (14). Thus, we may assume that $R_2^o > 0$ and write $R_1^o = R_1^* + \delta$ for some $\delta > 0$ where R_1^* is given by

$$\max_{\mathbf{B}} \quad \left[\frac{1}{2} \log \left| \frac{\mathbf{B} + \mathbf{N}_1}{\mathbf{N}_1} \right| - \frac{1}{2} \log \left| \frac{\mathbf{B} + \mathbf{N}_2}{\mathbf{N}_2} \right| \right]$$

$$\text{subject to:} \quad 0 \preceq \mathbf{B} \preceq \mathbf{S}$$

$$\frac{1}{2} \log \left| \frac{\mathbf{S} + \mathbf{N}_2}{\mathbf{B} + \mathbf{N}_2} \right| - \frac{1}{2} \log \left| \frac{\mathbf{S} + \mathbf{N}_3}{\mathbf{B} + \mathbf{N}_3} \right| \geq R_2^o.$$

Let $\mathbf{B}^*$ be an optimal solution to the above optimization program. Then, $\mathbf{B}^*$ must satisfy the following KKT conditions[1]:

$$(\mathbf{B}^* + \mathbf{N}_1)^{-1} + \mu(\mathbf{B}^* + \mathbf{N}_3)^{-1} + \mathbf{M}_1 = (1 + \mu)(\mathbf{B}^* + \mathbf{N}_2)^{-1} + \mathbf{M}_2 \tag{48}$$

$$\mathbf{B}^* \mathbf{M}_1 = 0 \tag{49}$$

$$\text{and} \quad (\mathbf{S} - \mathbf{B}^*)\mathbf{M}_2 = 0 \tag{50}$$

where $\mathbf{M}_1$ and $\mathbf{M}_2$ are $n \times n$ positive semidefinite matrices, and μ is a nonnegative real scalar such that $\mu > 0$ if and only if

$$\frac{1}{2} \log \left| \frac{\mathbf{S} + \mathbf{N}_2}{\mathbf{B}^* + \mathbf{N}_2} \right| - \frac{1}{2} \log \left| \frac{\mathbf{S} + \mathbf{N}_3}{\mathbf{B}^* + \mathbf{N}_3} \right| = R_2^o.$$

Thus,

$$R_1^o + \mu R_2^o = \left[\frac{1}{2} \log \left| \frac{\mathbf{B}^* + \mathbf{N}_1}{\mathbf{N}_1} \right| - \frac{1}{2} \log \left| \frac{\mathbf{B}^* + \mathbf{N}_2}{\mathbf{N}_2} \right| \right] + \mu \left[\frac{1}{2} \log \left| \frac{\mathbf{S} + \mathbf{N}_2}{\mathbf{B}^* + \mathbf{N}_2} \right| - \frac{1}{2} \log \left| \frac{\mathbf{S} + \mathbf{N}_3}{\mathbf{B}^* + \mathbf{N}_3} \right| \right] + \delta. \tag{51}$$

---

[1]As this optimization program is not convex, a set of constraint qualifications (CQs) should be checked to make sure that the KKT conditions indeed hold. The CQs stated in Appendix IV of [7] hold in a trivial manner for this program.

On the other hand, by the converse part of Theorem 3

$$R_1^o + \mu R_2^o \leq [I(\mathbf{X}; \mathbf{X} + \mathbf{Z}_1|U) - I(\mathbf{X}; \mathbf{X} + \mathbf{Z}_2|U)] + \mu[I(U; \mathbf{X} + \mathbf{Z}_2) - I(U; \mathbf{X} + \mathbf{Z}_3)]$$

$$= [h(\mathbf{Z}_2) - h(\mathbf{Z}_1)] - \mu[h(\mathbf{X} + \mathbf{Z}_3) - h(\mathbf{X} + \mathbf{Z}_2)]+$$

$$[h(\mathbf{X} + \mathbf{Z}_1|U) + \mu h(\mathbf{X} + \mathbf{Z}_3|U) - (1 + \mu)h(\mathbf{X} + \mathbf{Z}_2|U)]$$

$$= \frac{1}{2} \log \left| \frac{\mathbf{N}_2}{\mathbf{N}_1} \right| - \mu[h(\mathbf{X} + \mathbf{Z}_3) - h(\mathbf{X} + \mathbf{Z}_2)]+$$

$$[h(\mathbf{X} + \mathbf{Z}_1|U) + \mu h(\mathbf{X} + \mathbf{Z}_3|U) - (1 + \mu)h(\mathbf{X} + \mathbf{Z}_2|U)] \tag{52}$$

for some jointly distributed $(U, \mathbf{X})$ independent of $(\mathbf{Z}_1, \mathbf{Z}_2, \mathbf{Z}_3)$. Note that $\mathbf{N}_2 \preceq \mathbf{N}_3$. Similar to (46), we may obtain

$$h(\mathbf{X} + \mathbf{Z}_3) - h(\mathbf{X} + \mathbf{Z}_2) \geq \frac{1}{2} \log \left| \frac{\mathbf{S} + \mathbf{N}_3}{\mathbf{S} + \mathbf{N}_2} \right|. \tag{53}$$

Moreover, by letting

$$\mu_1 = \frac{1}{1 + \mu}, \quad \mu_3 = \frac{\mu}{1 + \mu}, \quad \tilde{\mathbf{M}}_1 = \frac{\mathbf{M}_1}{1 + \mu}, \quad \text{and} \quad \tilde{\mathbf{M}}_2 = \frac{\mathbf{M}_2}{1 + \mu}$$

we can rewrite the KKT conditions (48)–(50) as

$$\mu_1(\mathbf{B}^* + \mathbf{N}_1)^{-1} + \mu_3(\mathbf{B}^* + \mathbf{N}_3)^{-1} + \tilde{\mathbf{M}}_1 = (\mathbf{B}^* + \mathbf{N}_2)^{-1} + \tilde{\mathbf{M}}_2$$

$$\mathbf{B}^* \tilde{\mathbf{M}}_1 = 0$$

$$\text{and} \quad (\mathbf{S} - \mathbf{B}^*)\tilde{\mathbf{M}}_2 = 0.$$

Thus, by Theorem 2

$$h(\mathbf{X} + \mathbf{Z}_1|U) + \mu h(\mathbf{X} + \mathbf{Z}_3|U) - (1 + \mu)h(\mathbf{X} + \mathbf{Z}_2|U)$$

$$\leq \frac{1}{2} \log |\mathbf{B}^* + \mathbf{N}_1| + \frac{\mu}{2} \log |\mathbf{B}^* + \mathbf{N}_3| - \frac{1 + \mu}{2} \log |\mathbf{B}^* + \mathbf{N}_2|. \tag{54}$$

Substituting (53) and (54) into (52), we have

$$R_1^o + \mu R_2^o \leq \frac{1}{2} \log \left| \frac{\mathbf{N}_2}{\mathbf{N}_1} \right| - \frac{\mu}{2} \log \left| \frac{\mathbf{S} + \mathbf{N}_3}{\mathbf{S} + \mathbf{N}_2} \right| +$$

$$\left[\frac{1}{2} \log |\mathbf{B}^* + \mathbf{N}_1| + \frac{\mu}{2} \log |\mathbf{B}^* + \mathbf{N}_3| - \frac{1 + \mu}{2} \log |\mathbf{B}^* + \mathbf{N}_2| \right]$$

$$= \left[\frac{1}{2} \log \left| \frac{\mathbf{B}^* + \mathbf{N}_1}{\mathbf{N}_1} \right| - \frac{1}{2} \log \left| \frac{\mathbf{B}^* + \mathbf{N}_2}{\mathbf{N}_2} \right| \right] + \mu \left[\frac{1}{2} \log \left| \frac{\mathbf{S} + \mathbf{N}_2}{\mathbf{B}^* + \mathbf{N}_2} \right| - \frac{1}{2} \log \left| \frac{\mathbf{S} + \mathbf{N}_3}{\mathbf{B}^* + \mathbf{N}_3} \right| \right]. \tag{55}$$

Thus, we have obtained a contradiction between (51) and (55). As a result, all the achievable rate pairs must be

inside the secrecy rate region (16). This completes the proof of the theorem.

VI. PROOF OF THEOREM 6

In this section, we prove Theorem 6 following similar steps as those used in the proof for Theorem 5. The achievability of the secrecy rate region (17) can be obtained from the secrecy rate region (15) by letting $\mathbf{U}$ and $\mathbf{V}$ be two independent Gaussian vectors with zero means and covariance matrices $\mathbf{S} - \mathbf{B}$ and $\mathbf{B}$, respectively and $\mathbf{X} = \mathbf{U} + \mathbf{V}$. We therefore concentrate on the converse part of the theorem.

To show that (17) is indeed the secrecy capacity region of the vector Gaussian broadcast channel (8), we will use proof by contradiction. Assume that (R_1^o, R_2^o) is an achievable secrecy rate pair that lies *outside* the secrecy rate region (17). Note that $\mathbf{N}_1 \preceq \mathbf{N}_3$. From [28, Theorem 1], we can bound R_1^o by

$$R_1^o \leq \frac{1}{2} \log \left| \frac{\mathbf{S} + \mathbf{N}_1}{\mathbf{N}_1} \right| - \frac{1}{2} \log \left| \frac{\mathbf{S} + \mathbf{N}_3}{\mathbf{N}_3} \right| = R_1^{max}.$$

Note that when $R_2^o = 0$, R_1^{max} is achievable by letting $\mathbf{B} = \mathbf{S}$ in (15). Thus, we may assume that $R_2^o > 0$ and write $R_1^o = R_1^* + \delta$ for some $\delta > 0$ where R_1^* is given by

$$\max_{\mathbf{B}} \quad \left[\frac{1}{2} \log \left| \frac{\mathbf{B} + \mathbf{N}_1}{\mathbf{N}_1} \right| - \frac{1}{2} \log \left| \frac{\mathbf{B} + \mathbf{N}_3}{\mathbf{N}_3} \right| \right]$$

$$\text{subject to:} \quad 0 \preceq \mathbf{B} \preceq \mathbf{S}$$

$$\frac{1}{2} \log \left| \frac{\mathbf{S} + \mathbf{N}_2}{\mathbf{B} + \mathbf{N}_2} \right| - \frac{1}{2} \log \left| \frac{\mathbf{S} + \mathbf{N}_3}{\mathbf{B} + \mathbf{N}_3} \right| \geq R_2^o.$$

Let $\mathbf{B}^*$ be an optimal solution to the above optimization program. Then, $\mathbf{B}^*$ must satisfy the following KKT conditions:

$$(\mathbf{B}^* + \mathbf{N}_1)^{-1} + (\mu - 1)(\mathbf{B}^* + \mathbf{N}_3)^{-1} + \mathbf{M}_1 = \mu(\mathbf{B}^* + \mathbf{N}_2)^{-1} + \mathbf{M}_2 \tag{56}$$

$$\mathbf{B}^* \mathbf{M}_1 = 0 \tag{57}$$

$$\text{and} \quad (\mathbf{S} - \mathbf{B}^*)\mathbf{M}_2 = 0 \tag{58}$$

where $\mathbf{M}_1$ and $\mathbf{M}_2$ are $n \times n$ positive semidefinite matrices, and μ is a nonnegative real scalar such that $\mu \geq 1$.[2] Therefore,

$$R_2^o = \frac{1}{2} \log \left| \frac{\mathbf{S} + \mathbf{N}_2}{\mathbf{B}^* + \mathbf{N}_2} \right| - \frac{1}{2} \log \left| \frac{\mathbf{S} + \mathbf{N}_3}{\mathbf{B}^* + \mathbf{N}_3} \right|$$

and

$$R_1^o + \mu R_2^o = \left[\frac{1}{2} \log \left| \frac{\mathbf{B}^* + \mathbf{N}_1}{\mathbf{N}_1} \right| - \frac{1}{2} \log \left| \frac{\mathbf{B}^* + \mathbf{N}_3}{\mathbf{N}_3} \right| \right] + \mu \left[\frac{1}{2} \log \left| \frac{\mathbf{S} + \mathbf{N}_2}{\mathbf{B}^* + \mathbf{N}_2} \right| - \frac{1}{2} \log \left| \frac{\mathbf{S} + \mathbf{N}_3}{\mathbf{B}^* + \mathbf{N}_3} \right| \right] + \delta. \tag{59}$$

[2] If $\mu < 1$, it is easy to see that $\mathbf{B}^* = \mathbf{S}$ is an optimal solution and hence contradicts the assumption that $R_2^o > 0$.

On the other hand, by the converse part of Theorem 4

$$R_1^o + \mu R_2^o \leq [I(\mathbf{X}; \mathbf{X} + \mathbf{Z}_1|U) - I(\mathbf{X}; \mathbf{X} + \mathbf{Z}_3|U)] + \mu[I(U; \mathbf{X} + \mathbf{Z}_2) - I(U; \mathbf{X} + \mathbf{Z}_3)]$$

$$= [h(\mathbf{Z}_3) - h(\mathbf{Z}_1)] - \mu[h(\mathbf{X} + \mathbf{Z}_3) - h(\mathbf{X} + \mathbf{Z}_2)]+$$

$$[h(\mathbf{X} + \mathbf{Z}_1|U) + (\mu - 1)h(\mathbf{X} + \mathbf{Z}_3|U) - \mu h(\mathbf{X} + \mathbf{Z}_2|U)]$$

$$\leq \frac{1}{2}\log\left|\frac{\mathbf{N}_3}{\mathbf{N}_1}\right| - \frac{\mu}{2}\log\left|\frac{\mathbf{S} + \mathbf{N}_3}{\mathbf{S} + \mathbf{N}_2}\right|+$$

$$[h(\mathbf{X} + \mathbf{Z}_1|U) + (\mu - 1)h(\mathbf{X} + \mathbf{Z}_3|U) - \mu h(\mathbf{X} + \mathbf{Z}_2|U)] \qquad (60)$$

for some jointly distributed $(U, \mathbf{X})$ independent of $(\mathbf{Z}_1, \mathbf{Z}_2, \mathbf{Z}_3)$, where the last inequality follows from (53).

Since $\mu \geq 1$, by letting

$$\mu_1 = \frac{1}{\mu}, \quad \mu_3 = \frac{\mu - 1}{\mu}, \quad \tilde{\mathbf{M}}_1 = \frac{\mathbf{M}_1}{\mu}, \quad \text{and } \tilde{\mathbf{M}}_2 = \frac{\mathbf{M}_2}{\mu}$$

we can rewrite the KKT conditions (56)–(58) as

$$\mu_1(\mathbf{B}^* + \mathbf{N}_1)^{-1} + \mu_3(\mathbf{B}^* + \mathbf{N}_3)^{-1} + \tilde{\mathbf{M}}_1 = (\mathbf{B}^* + \mathbf{N}_2)^{-1} + \tilde{\mathbf{M}}_2$$

$$\mathbf{B}^*\tilde{\mathbf{M}}_1 = 0$$

$$\text{and} \qquad (\mathbf{S} - \mathbf{B}^*)\tilde{\mathbf{M}}_2 = 0.$$

Thus, by Theorem 2

$$h(\mathbf{X} + \mathbf{Z}_1|U) + (\mu - 1)h(\mathbf{X} + \mathbf{Z}_3|U) - \mu h(\mathbf{X} + \mathbf{Z}_2|U)$$

$$\leq \frac{1}{2}\log|\mathbf{B}^* + \mathbf{N}_1| + \frac{1 - \mu}{2}\log|\mathbf{B}^* + \mathbf{N}_3| - \frac{\mu}{2}\log|\mathbf{B}^* + \mathbf{N}_2|. \qquad (61)$$

Substituting (54) into (60), we have

$$R_1^o + \mu R_2^o \leq \frac{1}{2}\log\left|\frac{\mathbf{N}_3}{\mathbf{N}_1}\right| - \frac{\mu}{2}\log\left|\frac{\mathbf{S} + \mathbf{N}_3}{\mathbf{S} + \mathbf{N}_2}\right|+$$

$$\left[\frac{1}{2}\log|\mathbf{B}^* + \mathbf{N}_1| + \frac{\mu - 1}{2}\log|\mathbf{B}^* + \mathbf{N}_3| - \frac{\mu}{2}\log|\mathbf{B}^* + \mathbf{N}_2|\right]$$

$$= \left[\frac{1}{2}\log\left|\frac{\mathbf{B}^* + \mathbf{N}_1}{\mathbf{N}_1}\right| - \frac{1}{2}\log\left|\frac{\mathbf{B}^* + \mathbf{N}_3}{\mathbf{N}_3}\right|\right] + \mu\left[\frac{1}{2}\log\left|\frac{\mathbf{S} + \mathbf{N}_2}{\mathbf{B}^* + \mathbf{N}_2}\right| - \frac{1}{2}\log\left|\frac{\mathbf{S} + \mathbf{N}_3}{\mathbf{B}^* + \mathbf{N}_3}\right|\right]. \qquad (62)$$

Thus, we have obtained a contradiction between (59) and (62). As a result, all the achievable rate pairs must be inside the secrecy rate region (17). This completes the proof of the theorem.

VII. CONCLUSIONS

This paper has considered an EPI of Costa and has established a natural generalization by replacing the scalar parameter in the original Costa EPI with a matrix one. The generalized Costa EPI has been proven using a perturbation approach via a fundamental relationship between the derivative of mutual information and the MMSE in linear vector Gaussian channels. This is an example of how the connections between information theory and statistics can be explored to provide new mathematical tools for information theory.

As an application, a new extremal entropy inequality has been derived from the generalized Costa EPI and then used to characterize the secrecy capacity regions of the degraded vector Gaussian broadcast channel problem with layered confidential messages. We expect that the generalized Costa EPI will also play important roles in solving some other Gaussian network communication problems.

APPENDIX I

PROOF OF THEOREM 3

A. Achievability

We first show that the secrecy rate region (14) is achievable. Following the idea of superposition coding for the degraded broadcast channel [3], we introduce an auxiliary codebook which can be distinguished by both receiver 1 and receiver 2. The codebook is generated using random binning [20], [21].

Fix $p(u)$ and $p(x|u)$ and let

$$R'_1 = I(X; Y_2|U) - \epsilon_1 \tag{63a}$$

and
$$R'_2 = I(U; Y_3) - \epsilon_1 \tag{63b}$$

for some $\epsilon_1 > 0$. Let

$$L_k = 2^{nR_k}, \quad J_k = 2^{nR'_k} \quad \text{and,} \quad T_k = L_k J_k \quad k = 1, 2.$$

Without loss of generality, L_k, L'_k and J_k are assumed to be integers.

Codebook generation: Generate T_2 independent codewords u^n of length n according to $\prod_{i=1}^n p(u_i)$ and label them as

$$u^n(w_2, j_2), \quad w_2 \in \{1, \ldots, L_2\}, \quad j_2 \in \{1, \ldots, J_2\}.$$

For each codeword $u^n(w_2, j_2)$, generate T_1 independent codewords x^n according to $\prod_{i=1}^n p(x_i|u_i)$ and label them as

$$x^n(w_1, j_1, w_2, j_2) = x^n\big(w_1, j_1, u^n(w_2, j_2)\big), \quad w_k \in \{1, \ldots, L_k\} \quad \text{and} \quad j_k \in \{1, \ldots, J_k\}.$$

Encoding: To send a message pair (w_1, w_2), the transmitter randomly chooses a pair (j_1, j_2) and sends the corresponding codeword $x^n(w_1, j_1, w_2, j_2)$ through the channel.

Decoding: Receiver 2 determines the unique w_2 such that

$$\left(u^n(w_2, j_2), y_2^n\right) \in \mathcal{A}_\epsilon^{(n)}(p_{U,Y_2})$$

where $\mathcal{A}_\epsilon^{(n)}(p_{U,Y_2})$ denotes the set of jointly typical sequences u^n and y_2^n with respect to $p(u, y_2)$. If there are none such or more than one such, an error is declared. Receiver 1 looks for the unique (w_1, w_2) such that

$$\left(u^n(w_2, j_2), x^n(w_1, j_1, w_2, j_2), y_1^n\right) \in \mathcal{A}_\epsilon^{(n)}(p_{U,X,Y_1})$$

where $\mathcal{A}_\epsilon^{(n)}(p_{U,X,Y_1})$ denotes the set of jointly typical sequences u^n, x^n and y_1^n with respect to $p(u, x, y_1)$. Otherwise, an error is declared.

Error probability analysis: By the symmetry of the codebook generation, the probability error does not depend on which codeword was sent. Hence, without loss of generality, we may assume that the transmitter sends the message pair $(w_1, w_2) = (1, 1)$ associated with the codeword $x^n(1, 1, 1, 1)$ and define the corresponding event

$$\mathcal{K}_1 := \{x^n(1, 1, 1, 1) \text{ was sent}\}.$$

First consider the decoding at receiver 2, for which we will show that receiver 2 is able to decode $u^n(w_2, j_2)$ with small probability of error if $R_2 + R_2' < I(U; Y_2)$. To prove this, define the event

$$\mathcal{E}_2(w_2, j_2) := \left\{\left(u^n(w_2, j_2), y_2^n\right) \in \mathcal{A}_\epsilon^{(n)}(p_{U,Y_2})\right\}.$$

Then, the probability of error at receiver 2 can be bounded from above as

$$P_{e,2}^{(n)} \leq \Pr\left\{\bigcap_{j_2} \mathcal{E}_2^c(1, j_2)\Big|\mathcal{K}_1\right\} + \sum_{w_2 \neq 1, j_2} \Pr\{\mathcal{E}_2(w_2, j_2)|\mathcal{K}_1\}$$

$$\leq \Pr\{\mathcal{E}_2^c(1, 1)|\mathcal{K}_1\} + \sum_{w_2 \neq 1, j_2} \Pr\{\mathcal{E}_2(w_2, j_2)|\mathcal{K}_1\}$$

where

$$\mathcal{E}_2^c(1, j_2) := \left\{\left(u^n(1, j_2), y_2^n\right) \notin \mathcal{A}_\epsilon^{(n)}(p_{U,Y_2})\right\}.$$

For large enough n and $R_2 + R_2' < I(U; Y_2)$, the joint asymptotic equipartition property (AEP) [25, Chapter 14.2]

implies

$$P_{e,2}^{(n)} \le \epsilon + T_2 2^{-n[I(U;Y_2)-\epsilon]}$$

$$= \epsilon + 2^{n(R_2+R_2')} 2^{-n[I(U;Y_2)-\epsilon]}$$

$$\le 2\epsilon. \tag{64}$$

Next, we will show that receiver 1 can successfully decode both u^n and x^n if

$$R_1 + R_1' < I(X;Y_1|U)$$

$$\text{and} \qquad R_2 + R_2' < I(U;Y_2). \tag{65}$$

Define the events

$$\mathcal{E}_{1,1}(w_1, j_1, w_2, j_2) := \left\{ \left(u^n(w_2, j_2), x^n(w_1, j_1, w_2, j_2), y_1^n \right) \in \mathcal{A}_\epsilon^{(n)}(p_{U,X,Y_1}) \right\}.$$

$$\text{and} \qquad \mathcal{E}_1(w_2, j_2) := \left\{ \left(u^n(w_2, j_2), y_1^n \right) \in \mathcal{A}_\epsilon^{(n)}(p_{U,Y_1}) \right\}$$

where $\mathcal{A}_\epsilon^{(n)}(p_{U,Y_1})$ denotes the set of jointly typical sequences u^n and y_1^n with respect to $p(u, y_1)$. Then, the probability of error

$$P_{e,1}^{(n)} \le \Pr\{\mathcal{E}_1^c(1,1)|\mathcal{K}_1\} + \sum_{w_2 \ne 1, \, j_2} \Pr\{\mathcal{E}_1(w_2, j_2)|\mathcal{K}_1\} + \sum_{w_1 \ne 1, j_1,} \Pr\{\mathcal{E}_{1,1}(w_1, j_1, 1, 1)|\mathcal{K}_1\}$$

where

$$\mathcal{E}_1^c(1,1) := \left\{ \left(u^n(1,1), y_1^n \right) \notin \mathcal{A}_\epsilon^{(n)}(p_{U,Y_1}) \right\}.$$

By the AEP [25, Chapter 14.2],

$$\Pr\{\mathcal{E}_1^c(1,1)|\mathcal{K}_1\} \le \epsilon,$$

$$\Pr\{\mathcal{E}_1(w_2, j_2)|\mathcal{K}_1\} \le 2^{-n[I(U;Y_1)-\epsilon]}, \quad \text{for } w_2 \ne 1,$$

$$\text{and} \qquad \Pr\{\mathcal{E}_{1,1}(w_1, j_1, 1, 1)|\mathcal{K}_1\} \le 2^{-n[I(X;Y_1|U)-\epsilon]}, \quad \text{for } w_1 \ne 1.$$

Since the channel is degraded, we have $I(U;Y_1) \ge I(U;Y_2)$. Hence, if n is large enough and the condition (65) holds, the probability of error at receiver 1 can be bounded from above as

$$P_{e,1}^{(n)} \le \epsilon + T_2 2^{-n[I(U;Y_1)-\epsilon]} + T_1 2^{-n[I(X;Y_1|U)-\epsilon]}$$

$$\le \epsilon + 2^{n(R_2+R_2')} 2^{-n[I(U;Y_2)-\epsilon]} + 2^{n(R_1+R_1')} 2^{-n[I(X;Y_1|U)-\epsilon]}$$

$$\le 3\epsilon. \tag{66}$$

Together, (64) and (66) illustrate that messages (w_1, w_2) can be decoded at receiver 1 with a total probability of error that goes to 0 as long as the rate pair (R_1, R_2) satisfies (14).

Equivocation calculation: To show that (11) holds, we consider the following lower bound on the equivocation:

$$
\begin{aligned}
H(W_1|Y_2^n) &\geq H(W_1|Y_2^n, U^n) \\
&= H(W_1, Y_2^n|U^n) - H(Y_2^n|U^n) \\
&= H(X^n, Y_2^n|U^n) - H(X^n|W_1, Y_2^n, U^n) - H(Y_2^n|U^n) \\
&= H(X^n|U^n) + H(Y_2^n|X^n, U^n) - H(X^n|W_1, Y_2^n, U^n) - H(Y_2^n|U^n) \\
&= H(X^n|U^n) - H(X^n|W_1, Y_2^n, U^n) - I(X^n; Y_2^n|U^n)
\end{aligned}
\tag{67}
$$

where the second equality is due to the fact that W_1 is independent of everything else given X^n.

According to the codebook generation, for a given $U^n = u^n$, X^n has T_1 possible values with equal probabilities. Hence,

$$
\begin{aligned}
H(X^n|U^n) &= n(R_1 + R_1') \\
&= n[R_1 + I(X; Y_2|U) - \epsilon_1]
\end{aligned}
\tag{68}
$$

where (68) follows from the definition of R_1' in (63a).

Next, we show that for any given $\epsilon_2 > 0$, $H(X^n|W_1, Y_2^n, U^n) \leq n\epsilon_2$ for large enough n. To calculate $H(X^n|W_1, Y_2^n, U^n)$, consider the following hypothetical scenario. Fix $W_1 = w_1$, and assume that the transmitter sends a codeword $x^n(w_1, j_1, u^n(w_2, j_2))$, $j_1 \in \{1, \ldots, J_1\}$. Assume that receiver 2 knows the sequence $U^n = u^n(w_2, j_2)$. Given index $W_1 = w_1$, receiver 2 decodes the codeword $x^n(w_1, j_1, u^n)$ (i.e., looks for the index j_1) based on the received sequence y_2. Let $\lambda(w_1)$ denote the average probability of error of decoding the index j_1 at receiver 2. By the AEP [25, Chapter 14.2], we have $\lambda(w_1) \leq \epsilon$ for sufficiently large n. By Fano's inequality [25, Chapter 2.11],

$$
\begin{aligned}
\frac{1}{n} H(X^n|W_1 = w_1, Y_2^n, U^n) &\leq \frac{1}{n} + \lambda(w_1) \frac{\log_2 J_1}{n} \\
&\leq \frac{1}{n} + \epsilon R_1' \\
&:= \epsilon_2.
\end{aligned}
$$

Consequently,

$$\frac{1}{n}H(X^n|W_1, Y_2^n, U^n) = \frac{1}{n}\sum_{w_1=1}^{L_1} \Pr(W_1 = w_1)H(X^n|W_1 = w_1, Y_2^n, U^n)$$

$$\leq \epsilon_2. \tag{69}$$

By the AEP [25, Chapter 14.2], for any ϵ_3

$$I(X^n; Y_2^n|U^n) \leq nI(X; Y_2|U) + n\epsilon_3 \tag{70}$$

for sufficiently large n. Substituting (68), (69) and (70) into (67), we have

$$\frac{1}{n}H(W_1|Y_2^n) \geq R_1 - (\epsilon_1 + \epsilon_2 + \epsilon_3).$$

Similarly, we can show that

$$H(W_2|Y_3^n) \geq H(U^n) - H(U^n|W_2, Y_3^n) - I(U^n; Y_3^n)$$

where

$$H(U^n) = n[R_2 + I(U; Y_3) - \epsilon_1]$$

$$H(U^n|W_2, Y_3^n) \leq n\epsilon_2'$$

and $$I(U^n; Y_3^n) \leq n[I(U; Y_3) + \epsilon_3'],$$

where ϵ_2' and ϵ_3' vanishes in the limit as $n \to \infty$. Hence,

$$\frac{1}{n}H(W_2|Y_3^n) \geq R_2 - (\epsilon_1 + \epsilon_2' + \epsilon_3').$$

Note that Y_3 is degraded with respect to Y_2. Therefore,

$$\begin{aligned} H(W_1|Y_3^n) &\geq H(W_1|Y_2^n, Y_3^n) \\ &= H(W_1|Y_2^n) \\ &\geq R_1 - (\epsilon_1 + \epsilon_2 + \epsilon_3). \end{aligned}$$

This proves the security condition (11) and hence the achievability part of the theorem.

B. The Converse

We first bound from above the secrecy rate R_1. The perfect secrecy condition (11) implies that for all $\epsilon > 0$,

$$H(W_1|Y_2^n) \geq H(W_1) - n\epsilon \tag{71a}$$

and
$$H(W_2|Y_3^n) \geq H(W_2) - n\epsilon. \tag{71b}$$

On the other hand, Fano's inequality [25, Chapter 2.11] implies that for any $\epsilon_0 > 0$,

$$H(W_1|Y_1^n) \leq \epsilon_0 \log\left(2^{nR_1} - 1\right) + h(\epsilon_0) := n\delta_1 \tag{72a}$$

and
$$H(W_2|Y_2^n) \leq \epsilon_0 \log\left(2^{nR_2} - 1\right) + h(\epsilon_0) := n\delta_2. \tag{72b}$$

Thus,

$$
\begin{aligned}
nR_1 &= H(W_1) \\
&\leq \left[H(W_1|Y_2^n) + n\epsilon\right] + \left[n\delta_1 - H(W_1|Y_1^n)\right] \\
&\leq H(W_1, W_2|Y_2^n) - H(W_1|Y_1^n, W_2) + n(\epsilon + \delta_1) \\
&\leq H(W_1|Y_2^n, W_2) - H(W_1|Y_1^n, W_2) + n(\epsilon + \delta_1 + \delta_2)
\end{aligned}
\tag{73}
$$

where the first inequality follows from (71a) and (72a), and the last inequality follows from (72b). Let $\delta = \epsilon + \delta_1 + \delta_2$. By the chain rule of the mutual information [25, Chapter 2.5],

$$
\begin{aligned}
n(R_1 - \delta) &\leq I(W_1; Y_1^n|W_2) - I(W_1; Y_2^n|W_2) \\
&= \sum_{i=1}^{n} \left[I(W_1; Y_{1,i}|W_2, Y_{1,i+1}^n) - I(W_1; Y_{2,i}|W_2, Y_2^{i-1})\right] \\
&= \sum_{i=1}^{n} \left[I(W_1; Y_{1,i}|W_2, Y_{1,i+1}^n, Y_2^{i-1}) - I(W_1; Y_{2,i}|W_2, Y_{1,i+1}^n, Y_2^{i-1})\right]
\end{aligned}
\tag{74}
$$

where the last equality follows from [21, Lemma 7]. Let

$$V_i := \left(Y_{1,i+1}^n, Y_2^{i-1}\right). \tag{75}$$

We can further bound (74) from above as

$$n(R_1 - \delta) \leq \sum_{i=1}^{n} \left[I(W_1, X_i; Y_{1,i}|W_2, V_i) - I(W_1, X_i; Y_{2,i}|W_2, V_i) \right]$$

$$- \sum_{i=1}^{n} \left[I(X_i; Y_{1,i}|W_1, W_2, V_i) - I(X_i; Y_{2,i}|W_1, W_2, V_i) \right]$$

$$\leq \sum_{i=1}^{n} \left[I(W_1, X_i; Y_{1,i}|W_2, V_i) - I(W_1, X_i; Y_{2,i}|W_2, V_i) \right]$$

$$= \sum_{i=1}^{n} \left[I(X_i; Y_{1,i}|W_2, V_i) - I(X_i; Y_{2,i}|W_2, V_i) \right] \tag{76}$$

where the second inequality follows from the Markov relation

$$(W_1, W_2, V_i) \to X_i \to Y_{1,i} \to Y_{2,i},$$

and the last equality is due to the fact that $Y_{1,i}$ and $Y_{2,i}$ are conditionally independent of everything else given X_i.

Next, we bound from above the secrecy rate R_2. By (71b) and (72b),

$$nR_2 = H(W_2)$$

$$\leq \left[H(W_2|Y_3^n) + n\epsilon \right] + \left[n\delta_2 - H(W_2|Y_2^n) \right]$$

$$= I(W_2; Y_2^n) - I(W_2; Y_3^n) + n(\epsilon + \delta_2)$$

$$= \sum_{i=1}^{n} \left[I(W_2; Y_{2,i}|Y_{2,i+1}^n) - I(W_2; Y_{3,i}|Y_3^{i-1}) \right] + n(\epsilon + \delta_2). \tag{77}$$

Let $\delta' := \epsilon + \delta_2$ and

$$V_i' := \left(Y_{2,i+1}^n, Y_3^{i-1} \right). \tag{78}$$

Applying [21, Lemma 7] again, we may obtain

$$n(R_2 - \delta') \leq \sum_{i=1}^{n} \left[I(W_2; Y_{2,i}|V_i') - I(W_2; Y_{3,i}|V_i') \right]$$

$$= \sum_{i=1}^{n} \left[I(W_2, V_i'; Y_{2,i}) - I(W_2, V_i'; Y_{3,i}) \right] - \sum_{i=1}^{n} \left[I(V_i'; Y_{2,i}) - I(V_i'; Y_{3,i}) \right]$$

$$\leq \sum_{i=1}^{n} \left[I(W_2, V_i'; Y_{2,i}) - I(W_2, V_i'; Y_{3,i}) \right] \tag{79}$$

where the last inequality follows from the Markov relation $V_i' \to Y_{1,i} \to Y_{2,i}$. Furthermore, by the definitions of V_i and V_i' in (75) and (78) respectively,

$$V_i' \to (W_2, V_i) \to (Y_{2,i}, Y_{3,i}). \tag{80}$$

By (79) and (80),

$$
n(R_2 - \delta') \leq \sum_{i=1}^{n} \left[I(W_2, V_i', V_i; Y_{2,i}) - I(W_2, V_i', V_i; Y_{3,i}) \right] - \sum_{i=1}^{n} \left[I(V_i; Y_{2,i}|W_2, V_i') - I(V_i; Y_{3,i}|W_2, V_i') \right]
$$

$$
= \sum_{i=1}^{n} \left[I(W_2, V_i; Y_{2,i}) - I(W_2, V_i; Y_{3,i}) \right] - \sum_{i=1}^{n} \left[I(V_i; Y_{2,i}|W_2, V_i') - I(V_i; Y_{3,i}|W_2, V_i') \right]. \tag{81}
$$

Note that $Y_{3,i}$ is conditionally independent of everything else given $Y_{2,i}$. Hence,

$$
I(V_i; Y_{3,i}|W_2, V_i') \leq I(V_i; Y_{2,i}, Y_{3,i}|W_2, V_i')
$$

$$
= I(V_i; Y_{2,i}|W_2, V_i') + I(V_i; Y_{3,i}|Y_{2,i}, W_2, V_i')
$$

$$
= I(V_i; Y_{2,i}|W_2, V_i'). \tag{82}
$$

Substituting (82) into (81), we have

$$
R_2 \leq \frac{1}{n} \sum_{i=1}^{n} \left[I(W_2, V_i; Y_{2,i}) - I(W_2, V_i; Y_{3,i}) \right] + \delta'. \tag{83}
$$

Finally, let

$$
U_i := (W_2, V_i). \tag{84}
$$

With this definition, (76) and (83) can be rewritten as

$$
R_1 \leq \frac{1}{n} \sum_{i=1}^{n} \left[I(X_i; Y_{1,i}|U_i) - I(X_i; Y_{2,i}|U_i) \right] + \delta.
$$

$$
\text{and} \qquad R_2 \leq \frac{1}{n} \sum_{i=1}^{n} \left[I(U_i; Y_{2,i}) - I(U_i; Y_{3,i}) \right] + \delta'. \tag{85}
$$

Following the standard single-letterization process (e.g., see [25, Chapter 14.3]), we have the desired converse result.

APPENDIX II

PROOF OF COROLLARY 1

Fix $U = u$. By the generalized Costa EPI (3), we have

$$
h(\mathbf{X} + \mathbf{A}^{\frac{1}{2}}\mathbf{Z}|U = u) \geq \frac{n}{2} \log \left\{ |\mathbf{I} - \mathbf{A}|^{\frac{1}{n}} \exp\left[\frac{2}{n} h(\mathbf{X}|U = u) \right] + |\mathbf{A}|^{\frac{1}{n}} \exp\left[\frac{2}{n} h(\mathbf{X} + \mathbf{Z}|U = u) \right] \right\}. \tag{86}
$$

Taking expectation over U on both sides of (86), we may obtain

$$
\begin{aligned}
h(\mathbf{X} + \mathbf{A}^{\frac{1}{2}}\mathbf{Z}|U) &\geq \frac{n}{2}\mathsf{E}\left[\log\left\{|\mathbf{I} - \mathbf{A}|^{\frac{1}{n}}\exp\left[\frac{2}{n}h(\mathbf{X}|U=u)\right] + |\mathbf{A}|^{\frac{1}{n}}\exp\left[\frac{2}{n}h(\mathbf{X}+\mathbf{Z}|U=u)\right]\right\}\right] \\
&\geq \frac{n}{2}\log\left\{|\mathbf{I} - \mathbf{A}|^{\frac{1}{n}}\exp\left[\frac{2}{n}\mathsf{E}\left[h(\mathbf{X}|U=u)\right]\right] + |\mathbf{A}|^{\frac{1}{n}}\exp\left[\frac{2}{n}\mathsf{E}\left[h(\mathbf{X}+\mathbf{Z}|U=u)\right]\right]\right\} \\
&= \frac{n}{2}\log\left\{|\mathbf{I} - \mathbf{A}|^{\frac{1}{n}}\exp\left[\frac{2}{n}h(\mathbf{X}|U)\right] + |\mathbf{A}|^{\frac{1}{n}}\exp\left[\frac{2}{n}h(\mathbf{X}+\mathbf{Z}|U)\right]\right\}
\end{aligned}
\tag{87}
$$

where the second inequality follows from Jensen's inequality [25, Chapter 2.6] and the convexity of $\log\left(a_1 e^{x_1} + a_2 e^{x_2}\right)$ in (x_1, x_2) for $a_1, a_2 \geq 0$. Taking logarithm on both sides of (87) proves the desired inequality (32).

APPENDIX III

PROOF OF COROLLARY 2

Note that when $\mu = 0$, (33) implies that $\mathbf{N}_1 = \mathbf{N}_2$. Thus, both sides of (34) are equal to zero and the inequality holds trivially with an equality. For the rest of the proof, we will assume that $\mu > 0$. The proof is rather long so we divide it into several steps.

Step 1–Generalized eigenvalue decomposition. We start by applying generalized eigenvalue decomposition [23] to the positive define matrices $\mathbf{B}^* + \mathbf{N}_1$ and $\mathbf{B}^* + \mathbf{N}_2$. There exists an *invertible* generalized eigenvector matrix $\mathbf{V}$ such that

$$
\mathbf{V}^\mathsf{T}(\mathbf{B}^* + \mathbf{N}_1)\mathbf{V} = \mathbf{\Lambda}_1
\tag{88}
$$

$$
\text{and} \quad \mathbf{V}^\mathsf{T}(\mathbf{B}^* + \mathbf{N}_2)\mathbf{V} = \mathbf{\Lambda}_2
\tag{89}
$$

where $\mathbf{\Lambda}_1$ and $\mathbf{\Lambda}_2$ are positive definite *diagonal* matrices. Let

$$
\mathbf{\Lambda}_3 := \mathbf{V}^\mathsf{T}(\mathbf{B}^* + \mathbf{N}_3)\mathbf{V}
\tag{90}
$$

be an $n \times n$ positive definite matrix. By (33),

$$
\mathbf{\Lambda}_1^{-1} + \mu\mathbf{\Lambda}_3^{-1} = (1+\mu)\mathbf{\Lambda}_2^{-1}.
\tag{91}
$$

Thus, $\mathbf{\Lambda}_3$ is also diagonal. Moreover, since $\mathbf{N}_1 \preceq \mathbf{N}_3$,

$$
\mathbf{\Lambda}_1 - \mathbf{\Lambda}_3 = \mathbf{V}^\mathsf{T}(\mathbf{N}_1 - \mathbf{N}_3)\mathbf{V} \preceq 0.
$$

and hence

$$
\mathbf{\Lambda}_1 \preceq \mathbf{\Lambda}_3.
\tag{92}
$$

Step 2–Choosing matrix parameter $\mathbf{A}$. Let $\tilde{\boldsymbol{\Lambda}}_3 = \boldsymbol{\Lambda}_3 + \epsilon\mathbf{I}$ for some $\epsilon > 0$, and let $\tilde{\boldsymbol{\Lambda}}_2$ be an $n \times n$ matrix such that

$$\boldsymbol{\Lambda}_1^{-1} + \mu\tilde{\boldsymbol{\Lambda}}_3^{-1} = (1+\mu)\tilde{\boldsymbol{\Lambda}}_2^{-1}. \tag{93}$$

Clearly, $\tilde{\boldsymbol{\Lambda}}_2$ is diagonal. Moreover, by (92)

$$\boldsymbol{\Lambda}_1 \prec \tilde{\boldsymbol{\Lambda}}_3. \tag{94}$$

Note that $\mu > 0$ so by (93) and (94)

$$\boldsymbol{\Lambda}_1 \prec \tilde{\boldsymbol{\Lambda}}_2 \prec \tilde{\boldsymbol{\Lambda}}_3. \tag{95}$$

Comparing (91) and (93) and using the fact that $\boldsymbol{\Lambda}_3 \prec \tilde{\boldsymbol{\Lambda}}_3$, we have

$$\boldsymbol{\Lambda}_2 \prec \tilde{\boldsymbol{\Lambda}}_2. \tag{96}$$

Now let

$$\mathbf{Y}_1 := \mathbf{V}^\mathsf{T}(\mathbf{X} + \mathbf{Z}_1)$$

$$\mathbf{Y}_2 := \mathbf{V}^\mathsf{T}(\mathbf{X} + \widetilde{\mathbf{Z}}_2)$$

$$\text{and} \qquad \mathbf{Y}_3 := \mathbf{V}^\mathsf{T}(\mathbf{X} + \widetilde{\mathbf{Z}}_3)$$

where $\widetilde{\mathbf{Z}}_2$ and $\widetilde{\mathbf{Z}}_3$ are Gaussian n-vectors with covariance matrices

$$\begin{aligned}
\widetilde{\mathbf{N}}_2 &= \mathbf{V}^{-\mathsf{T}}\tilde{\boldsymbol{\Lambda}}_2\mathbf{V}^{-1} - \mathbf{B}^* \\
&\succ \mathbf{V}^{-\mathsf{T}}\boldsymbol{\Lambda}_2\mathbf{V}^{-1} - \mathbf{B}^* \\
&= (\mathbf{B}^* + \mathbf{N}_2) - \mathbf{B}^* \\
&= \mathbf{N}_2
\end{aligned}$$

and

$$\begin{aligned}
\widetilde{\mathbf{N}}_3 &= \mathbf{V}^{-\mathsf{T}}\tilde{\boldsymbol{\Lambda}}_3\mathbf{V}^{-1} - \mathbf{B}^* \\
&= \mathbf{V}^{-\mathsf{T}}(\boldsymbol{\Lambda}_3 + \epsilon\mathbf{I})\mathbf{V}^{-1} - \mathbf{B}^* \\
&= (\mathbf{B}^* + \mathbf{N}_3 + \epsilon\mathbf{V}^{-\mathsf{T}}\mathbf{V}^{-1}) - \mathbf{B}^* \\
&= \mathbf{N}_3 + \epsilon\mathbf{V}^{-\mathsf{T}}\mathbf{V}^{-1}
\end{aligned}$$

respectively and are independent of $\mathbf{X}$. The covariance matrices of $\mathbf{Y}_k$, $k = 1, 2, 3$, can be calculated as $\mathbf{V}^\mathsf{T}[\mathsf{Cov}(\mathbf{X}) - \mathbf{B}^*]\mathbf{V} + \mathbf{\Lambda}_1$, $\mathbf{V}^\mathsf{T}[\mathsf{Cov}(\mathbf{X}) - \mathbf{B}^*]\mathbf{V} + \tilde{\mathbf{\Lambda}}_2$ and $\mathbf{V}^\mathsf{T}[\mathsf{Cov}(\mathbf{X}) - \mathbf{B}^*]\mathbf{V} + \tilde{\mathbf{\Lambda}}_3$, respectively. Thus, $\mathbf{Y}_2$ and $\mathbf{Y}_3$ can be equivalently written as

$$\mathbf{Y}_3 = \mathbf{Y}_1 + \mathbf{Z}$$

$$\text{and} \qquad \mathbf{Y}_2 = \mathbf{Y}_1 + \mathbf{A}^{\frac{1}{2}}\mathbf{Z}$$

where $\mathbf{Z}$ is a Gaussian n-vector with covariance matrix $\tilde{\mathbf{\Lambda}}_3 - \mathbf{\Lambda}_1 \succ 0$ and is independent of $\mathbf{Y}_1$, and

$$\mathbf{A} := (\tilde{\mathbf{\Lambda}}_2 - \mathbf{\Lambda}_1)(\tilde{\mathbf{\Lambda}}_3 - \mathbf{\Lambda}_1)^{-1}. \tag{97}$$

Clearly, $\mathbf{A}$ is diagonal. Moreover, by (95) $0 \prec \mathbf{A} \prec \mathbf{I}$.

Step 3–Applying generalized Costa's EPI. By the generalized Costa EPI (3),

$$h(\mathbf{Y}_2|U) \geq \frac{n}{2} \log \left\{ |\mathbf{I} - \mathbf{A}|^{\frac{1}{n}} \exp\left[\frac{2}{n} h(\mathbf{Y}_1|U)\right] + |\mathbf{A}|^{\frac{1}{n}} \exp\left[\frac{2}{n} h(\mathbf{Y}_3|U)\right] \right\}.$$

Thus,

$$h(\mathbf{Y}_1|U) + \mu h(\mathbf{Y}_3|U) - (1 + \mu)h(\mathbf{Y}_2|U)$$
$$\leq h(\mathbf{Y}_1|U) + \mu h(\mathbf{Y}_3|U) - \frac{(1 + \mu)n}{2} \log \left\{ |\mathbf{I} - \mathbf{A}|^{\frac{1}{n}} \exp\left[\frac{2}{n} h(\mathbf{Y}_1|U)\right] + |\mathbf{A}|^{\frac{1}{n}} \exp\left[\frac{2}{n} h(\mathbf{Y}_3|U)\right] \right\}. \tag{98}$$

Now we consider the function

$$f(b, c) = b + \mu c - \frac{(1 + \mu)n}{2} \log \left[|\mathbf{I} - \mathbf{A}|^{\frac{1}{n}} \exp\left(\frac{2b}{n}\right) + |\mathbf{A}|^{\frac{1}{n}} \exp\left(\frac{2c}{n}\right) \right].$$

Note that

$$\nabla f(b, c) = \begin{bmatrix} 1 - (1 + \mu)\dfrac{|\mathbf{I} - \mathbf{A}|^{\frac{1}{n}} \exp(2b/n)}{|\mathbf{I} - \mathbf{A}|^{\frac{1}{n}} \exp(2b/n) + |\mathbf{A}|^{\frac{1}{n}} \exp(2c/n)} \\ \mu - (1 + \mu)\dfrac{|\mathbf{A}|^{\frac{1}{n}} \exp(2c/n)}{|\mathbf{I} - \mathbf{A}|^{\frac{1}{n}} \exp(2b/n) + |\mathbf{A}|^{\frac{1}{n}} \exp(2c/n)} \end{bmatrix}$$

and

$$\nabla^2 f(b, c) = -\frac{2(1 + \mu)}{n} \frac{|\mathbf{A}|^{\frac{1}{n}} |\mathbf{I} - \mathbf{A}|^{\frac{1}{n}} \exp[(2b + 2c)/n]}{\left[|\mathbf{I} - \mathbf{A}|^{\frac{1}{n}} \exp(2b/n) + |\mathbf{A}|^{\frac{1}{n}} \exp(2c/n) \right]^2} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \preceq 0.$$

So $f(b, c)$ is concave in (b, c). By setting $\nabla f(b, c) = 0$, the global maximum is achieved when

$$c = b + \frac{n}{2} \log \left[\mu \left(\frac{|\mathbf{I} - \mathbf{A}|}{|\mathbf{A}|} \right)^{\frac{1}{n}} \right]$$

and the maximum is given by

$$\frac{\mu n}{2} \log \left[\mu \left(\frac{|\mathbf{I} - \mathbf{A}|}{|\mathbf{A}|} \right)^{\frac{1}{n}} \right] - \frac{(1+\mu)n}{2} \log \left[(1+\mu)|\mathbf{I} - \mathbf{A}|^{\frac{1}{n}} \right].$$

Hence,

$$h(\mathbf{Y}_1|U) + \mu h(\mathbf{Y}_3|U) - (1+\mu)h(\mathbf{Y}_2|U)$$
$$\leq \frac{\mu n}{2} \log \left[\mu \left(\frac{|\mathbf{I} - \mathbf{A}|}{|\mathbf{A}|} \right)^{\frac{1}{n}} \right] - \frac{(1+\mu)n}{2} \log \left[(1+\mu)|\mathbf{I} - \mathbf{A}|^{\frac{1}{n}} \right]. \tag{99}$$

Step 4–Calculating $\log |\mathbf{A}|$ *and* $\log |\mathbf{I} - \mathbf{A}|$. Note that (93) can be rewritten as

$$\mu(\mathbf{\Lambda}_1^{-1} - \tilde{\mathbf{\Lambda}}_3^{-1}) = (1+\mu)(\mathbf{\Lambda}_1^{-1} - \tilde{\mathbf{\Lambda}}_2^{-1})$$

which gives

$$\left| \frac{\tilde{\mathbf{\Lambda}}_2 - \mathbf{\Lambda}_1}{\tilde{\mathbf{\Lambda}}_3 - \mathbf{\Lambda}_1} \right| = \left(\frac{\mu}{1+\mu} \right)^n \left| \frac{\tilde{\mathbf{\Lambda}}_2}{\tilde{\mathbf{\Lambda}}_3} \right|. \tag{100}$$

Similarly, we have

$$(\mathbf{\Lambda}_1^{-1} - \tilde{\mathbf{\Lambda}}_3^{-1}) = (1+\mu)(\tilde{\mathbf{\Lambda}}_2^{-1} - \tilde{\mathbf{\Lambda}}_3^{-1})$$

and hence

$$\left| \frac{\tilde{\mathbf{\Lambda}}_3 - \tilde{\mathbf{\Lambda}}_2}{\tilde{\mathbf{\Lambda}}_3 - \mathbf{\Lambda}_1} \right| = \left(\frac{1}{1+\mu} \right)^n \left| \frac{\tilde{\mathbf{\Lambda}}_2}{\mathbf{\Lambda}_1} \right|. \tag{101}$$

According to the definition of $\mathbf{A}$ in (97),

$$\log |\mathbf{A}| = \log \left| \frac{\tilde{\mathbf{\Lambda}}_2 - \mathbf{\Lambda}_1}{\tilde{\mathbf{\Lambda}}_3 - \mathbf{\Lambda}_1} \right|$$
$$= \log \left[\left(\frac{\mu}{1+\mu} \right)^n \left| \frac{\tilde{\mathbf{\Lambda}}_2}{\tilde{\mathbf{\Lambda}}_3} \right| \right] \tag{102}$$

and

$$\log |\mathbf{I} - \mathbf{A}| = \log \left| \frac{\tilde{\mathbf{\Lambda}}_3 - \tilde{\mathbf{\Lambda}}_2}{\tilde{\mathbf{\Lambda}}_3 - \mathbf{\Lambda}_1} \right|$$
$$= \log \left[\left(\frac{1}{1+\mu} \right)^n \left| \frac{\tilde{\mathbf{\Lambda}}_2}{\mathbf{\Lambda}_1} \right| \right] \tag{103}$$

where (102) and (103) follow (100) and (101), respectively. Substituting (102) and (103) into (99), we have

$$h(\mathbf{Y}_1|U) + \mu h(\mathbf{Y}_3|U) - (1+\mu)h(\mathbf{Y}_2|U) \leq \frac{1}{2} \log |\mathbf{\Lambda}_1| + \frac{\mu}{2} \log |\tilde{\mathbf{\Lambda}}_3| - \frac{1+\mu}{2} \log |\tilde{\mathbf{\Lambda}}_2|. \tag{104}$$

Step 5–Letting $\epsilon \downarrow 0$. Note that $\tilde{\mathbf{\Lambda}}_3 = \mathbf{\Lambda}_3 + \epsilon\mathbf{I} \to \mathbf{\Lambda}_3$ and $\widetilde{\mathbf{N}}_3 = \mathbf{N}_3 + \epsilon\mathbf{V}^{-\mathsf{T}}\mathbf{V}^{-1} \to \mathbf{N}_3$ in the limit as $\epsilon \downarrow 0$. Moreover, by (93) we have $\tilde{\mathbf{\Lambda}}_2 \to \mathbf{\Lambda}_2$ and hence

$$
\begin{aligned}
\widetilde{\mathbf{N}}_2 &= \mathbf{V}^{-\mathsf{T}}\tilde{\mathbf{\Lambda}}_2\mathbf{V}^{-1} - \mathbf{B}^* \\
&\to \mathbf{V}^{-\mathsf{T}}\mathbf{\Lambda}_2\mathbf{V}^{-1} - \mathbf{B}^* \\
&= (\mathbf{B}^* + \mathbf{N}_2) - \mathbf{B}^* \\
&= \mathbf{N}_2.
\end{aligned}
$$

Letting $\epsilon \downarrow 0$ on both sides of (104), we have

$$
h(\mathbf{V}^{\mathsf{T}}(\mathbf{X} + \mathbf{N}_1)|U) + \mu h(\mathbf{V}^{\mathsf{T}}(\mathbf{X} + \mathbf{N}_3)|U) - (1+\mu)h(\mathbf{V}^{\mathsf{T}}(\mathbf{X} + \mathbf{N}_2)|U)
$$
$$
\leq \frac{1}{2}\log|\mathbf{\Lambda}_1| + \frac{\mu}{2}\log|\mathbf{\Lambda}_3| - \frac{1+\mu}{2}\log|\mathbf{\Lambda}_2|. \tag{105}
$$

Using the fact that

$$
h(\mathbf{V}^{\mathsf{T}}(\mathbf{X} + \mathbf{N}_1)|U) = h(\mathbf{X} + \mathbf{N}_1|U) + \log|\mathbf{V}|
$$

and

$$
\begin{aligned}
\log|\mathbf{\Lambda}_k| &= \log|\mathbf{V}^{\mathsf{T}}(\mathbf{B}^* + \mathbf{N}_k)\mathbf{V}| \\
&= \log|\mathbf{B}^* + \mathbf{N}_k| + 2\log|\mathbf{V}|
\end{aligned}
$$

for $k = 1, 2, 3$, the desired inequality (34) can be obtained from (105). This completes the proof of the corollary.

APPENDIX IV

PROOF OF COROLLARY 3

Here, we prove Corollary 3 using mathematical induction. Note that when $K = 1$, (35) implies that $\mathbf{N}_1 = \mathbf{N}_0$. Thus, the inequality (36) holds trivially with equality for any $(U, \mathbf{X})$ independent of $(\mathbf{Z}_0, \mathbf{Z}_1)$.

Assume that the inequality (36) holds for $K = Q - 1$. Let $\mathbf{N}$ be an $n \times n$ symmetric matrix such that

$$
(\mathbf{B}^* + \mathbf{N})^{-1} = \sum_{k=1}^{Q-1} \mu'_k(\mathbf{B}^* + \mathbf{N}_k)^{-1} \tag{106}
$$

where

$$
\mu'_k := \frac{\mu_k}{\sum_{j=1}^{Q-1}\mu_j}, \quad j = 1, \ldots, Q.
$$

By the assumption $\mathbf{N}_1 \preceq \ldots \preceq \mathbf{N}_{Q-1}$, we have from (106)

$$\mathbf{N}_1 \preceq \mathbf{N} \preceq \mathbf{N}_{Q-1}. \tag{107}$$

Let $\mathbf{Z}$ be a Gaussian random n-vector with covariance matrix $\mathbf{N}$ and independent of $(U, \mathbf{X})$. By the induction assumption and (106),

$$\sum_{k=1}^{Q-1} \mu'_k h(\mathbf{X} + \mathbf{Z}_k | U) - h(\mathbf{X} + \mathbf{Z} | U) \leq \sum_{k=1}^{Q-1} \frac{\mu'_k}{2} \log |\mathbf{B} + \mathbf{N}_k| - \frac{1}{2} \log |\mathbf{B} + \mathbf{N}|. \tag{108}$$

On the other hand, substitute (106) into (35) and we have

$$(\mathbf{B} + \mathbf{N})^{-1} + \mu'_Q (\mathbf{B} + \mathbf{N}_Q)^{-1} = (1 + \mu'_Q)(\mathbf{B} + \mathbf{N}_0)^{-1}.$$

Note from (107) that $\mathbf{N} \preceq \mathbf{N}_{Q-1} \preceq \mathbf{N}_Q$. Thus, by Corollary 2

$$\begin{aligned} h(\mathbf{X} + \mathbf{Z}|U) + \mu'_Q h(\mathbf{X} + \mathbf{Z}_Q|U) - (1 + \mu'_Q) h(\mathbf{X} + \mathbf{Z}_0|U) \\ \leq \frac{1}{2} \log |\mathbf{B} + \mathbf{N}| + \frac{\mu'_Q}{2} \log |\mathbf{B} + \mathbf{N}_Q| - \frac{1 + \mu'_Q}{2} \log |\mathbf{B} + \mathbf{N}_0|. \end{aligned} \tag{109}$$

Putting together (108) and (109), we have

$$\sum_{j=1}^{Q} \mu_j h(\mathbf{X} + \mathbf{Z}_j | U) - h(\mathbf{X} + \mathbf{Z}_0 | U) \leq \sum_{j=1}^{Q} \frac{\mu_j}{2} \log |\mathbf{B} + \mathbf{N}_j| - \frac{1}{2} \log |\mathbf{B} + \mathbf{N}_0|.$$

This proved the induction step and hence the corollary.

REFERENCES

[1] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423 and 623–656, Jul. and Oct. 1948.

[2] A. J. Stam, "Some inequalities satisfied by the quantities of information of Fisher and Shannon," *Inform. Control*, vol. 2, pp. 101–112, Jun. 1959.

[3] P. P. Bergmans, "Random coding theorem for broadcast channels with degraded components," *IEEE Trans. Inf. Theory*, vol. 19, pp. 197–207, Mar. 1973.

[4] S. K. Leung-Yan-Cheong and M. E. Hellman, "The Gaussian wire-tap channel," *IEEE Trans. Inf. Theory*, vol. 24, no. 4, pp. 51–456, Jul. 1978.

[5] L. Ozarow, "On a source coding problem with two channels and three receivers," *Bell Syst. Tech. J.*, vol. 59, no. 10, pp. 1909–1921, Dec. 1980.

[6] Y. Oohama, "The rate-distortion function for the quadratic Gaussian CEO problem," *IEEE Trans. Inf. Theory*, vol. 44, no. 3, pp. 1057–1070, May 1998.

[7] H. Weingarten, Y. Steinberg, and S. Shamai (Shitz), "The capacity region of the Gaussian multiple-input multiple-output broadcast channel," *IEEE Trans. Inf. Theory*, vol. 52, pp. 3936–3964, Sep. 2006.

[8] M. H. M. Costa, "A new entropy power inequality," *IEEE Trans. Inf. Theory*, vol. 31, pp. 751–760, Nov. 1985.

[9] ——, "On the Gaussian interference channel," *IEEE Trans. Inf. Theory*, vol. 31, pp. 607–615, Sep. 1985.

[10] A. Lapidoth and S. M. Moser, "Capacity bounds via duality with applications to multiple-antenna systems on flat-fading channels," *IEEE Trans. Inf. Theory*, vol. 49, pp. 2426–2467, Oct. 2003.

[11] A. Dembo, T. M. Cover, and J. A. Thomas, "Information theoretic inequalities," *IEEE Trans. Inf. Theory*, vol. 37, pp. 1501–1518, Nov. 1991.

[12] D. Guo, S. Shamai (Shitz), and S. Verdú, "Mutual information and minimum mean-square error in Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1261–1282, Apr. 2005.

[13] A. Dembo, "Simple proof on the concavity of the entropy power with respect to added Gaussian noise," *IEEE Trans. Inf. Theory*, vol. 35, pp. 887–888, Jul. 1989.

[14] D. Guo, S. Shamai (Shitz), and S. Verdú, "Proof of entropy power inequalities via MMSE," in *Proc. IEEE Int. Symp. Information Theory*, Seattle, WA, July 9-14, 2006.

[15] M. Payaró and D. P. Palomar, "Hessian matrix and concavity properties of mutual information and entropy in linear vector Gaussian channels," *IEEE Trans. Inf. Theory*, submitted for publication.

[16] T. Liu and P. Viswanath, "An extremal inequality motivated by multiterminal information-theoretic problems," *IEEE Trans. Inf. Theory*, vol. 53, pp. 1839–1851, May 2007.

[17] K. Marton, "A coding theorem for the discrete memoryless broadcast channel," *IEEE Trans. Inf. Theory*, vol. 25, pp. 306–311, May 1979.

[18] D. P. Palomar and S. Verdú, "Gradient of mutual information in linear vector Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 52, pp. 141–154, Jan. 2006.

[19] D. P. Bertsekas, A. Nedic, and A. E. Ozdaglar, *Convex Analysis and Optimization*. Belmont, MA: Athena Scientific, 2003.

[20] A. D. Wyner, "The wire-tap channel," *Bell Syst. Tech. J.*, vol. 54, no. 8, pp. 1355–1387, Oct. 1975.

[21] I. Csiszár and J. Körner, "Broadcast channels with confidential messages," *IEEE Trans. Inf. Theory*, vol. 24, no. 3, pp. 339–348, May 1978.

[22] G. Bagherikaram, A. S. Motahari, and A. K. Khandani, "Secure broadcasting: The secrecy rate region," *IEEE Trans. Inf. Theory*, submitted, Dec. 2008.

[23] G. Strang, *Linear Algebra and Its Applications*. Wellesley, MA: Wellesley-Cambridge Press, 1998.

[24] G. A. F. Seber, *A Matrix Handbook for Statisticians*. New York: John Wiley & Sons, Inc., 2008.

[25] T. Cover and J. Thomas, *Elements of Information Theory*. New York: John Wiley & Sons, Inc., 1991.

[26] I. M. Gel'fand and A. Shen, *Algebra*, 3rd ed. Basel, Switzerland: Birkhauser Verlag, 1993.

[27] S. N. Diggavi and T. M. Cover, "The worst additive noise under a covariance constraint," *IEEE Trans. Inf. Theory*, vol. 47, pp. 3072–3081, Nov. 2001.

[28] T. Liu and S. Shamai (Shitz), "A note on the secrecy capacity of the multiantenna wiretap channel," *IEEE Trans. Inf. Theory*, to appear.