Abstract—We show that for product sources, rate splitting is optimal for secret key agreement using limited one-way communication between two terminals. This yields an alternative information-theoretic-converse-style proof of the tensorization property of a strong data processing inequality originally studied by Erkip and Cover and amended recently by Anantharam et al. We derive a water-filling solution of the communication-rate–key-rate tradeoff for a wide class of discrete memoryless vector Gaussian sources which subsumes the case without an eavesdropper. Moreover, we derive an explicit formula for the maximum secret key per bit of communication for all discrete memoryless vector Gaussian sources using a tensorization property and a variation on the enhanced channel technique of Weingarten et al. Finally, a one-shot information spectrum achievability bound for key generation is proved from which we characterize the communication-rate–key-rate tradeoff for stationary Gaussian processes.

Index Terms—Random number generation, source coding, Gaussian processes, Correlation coefficient, Decorrelation, Fourier transforms, MIMO.

I. INTRODUCTION

A and B allowing public communication from A to B.

Figure 1. The basic model for secret key agreement between two terminals A and B allowing public communication from A to B.

An important scenario for secret key agreement (a.k.a. key generation) arises when terminals at distant locations have access to correlated sources and are allowed to communicate publicly in order to decide on a key which is kept unknown to an eavesdropper.

This paper was presented in part at 2014 IEEE International Symposium on Information Theory (ISIT). Copyright © 2014 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

The fundamental limit on the amount of secret key that can be generated from discrete memoryless sources was studied in [1,2], where single-letter solutions were derived for the class of protocols allowing limited one-way communication from one terminal to the other. However, for many models of interest in practice, the key capacity remains unknown, since the optimizations over auxiliary random variables in those single-letter formulas are usually hard to solve.

In [3] the fundamental limit was extended to sources with continuous alphabets; and it was shown that for vector Gaussian sources, which are natural models of multiple input multiple output (MIMO) systems, one auxiliary random variable suffices to characterize the rate region, instead of two in the general case [1], and it is enough to consider auxiliary random vectors that are jointly Gaussian with the sources. This observation is formally stated in Fact 1 ahead, the proof of which in [3] was based on the enhancement technique introduced by Weingarten et al. [4]. Consequently, the capacity region for vector Gaussian sources was posed as a (generally non-convex) matrix optimization problem. Still, an explicit formula for the key capacity was not derived except for scalar Gaussian sources.

In this paper we provide an explicit formula for the key capacity of vector Gaussian sources by considering a more general setup: the key capacity of arbitrary product sources. Specifically, suppose terminals A and B and an eavesdropper observe discrete memoryless vector sources \( X = (X_i)_{i=1}^L \), \( Y = (Y_i)_{i=1}^L \) and \( Z = (Z_i)_{i=1}^L \) respectively, where

\[
P_{XY} = \prod_{i=1}^L P_{X,Y_i} \quad (1)
\]

\[
P_{XZ} = \prod_{i=1}^L P_{X,Z_i} \quad (2)
\]

We call \((X, Y, Z)\) a product source because of the structure of its joint probability distribution. An example of product sources is illustrated in Figure 2.\(^1\) The maximal rate of secret key achievable as a function of public communication rate \( r \) from A to B is denoted as \( R(r) \). We show that

\[
R(r) = \max_{\sum_{i=1}^L r_i \leq r} \sum_{i=1}^L R_i(r_i) \quad (3)
\]

\(^1\)Actually Figure 2 only illustrates an unnecessarily special case of (1) and (2) where \( P_{XYZ} = \prod_{i=1}^L P_{X,Y_i,Z_i} \); c.f. Section II-B.
where $R_i(r_i)$ is the key-communication function corresponding to the $i$-th source triple: $(X_i, Y_i, Z_i)$. This is analogous to a result due to Shannon [5] on the rate distortion function of a product source with a separable distortion measure, which is obtained by summing the rates and distortions of points in the individual rate-distortion curves with the same slope.

In the case of jointly vector Gaussian sources without an eavesdropper (or with an eavesdropper but under a certain commutative condition on the covariance matrices), one can always apply separate invertible linear transforms on the vectors observed at A and B so that the source distribution is of the form in (1) and (2), thus deriving an explicit formula of $R(r)$ utilizing corresponding results of scalar Gaussian sources. The solution displays a "water filling" behavior similar to the rate distortion function of vector Gaussian sources (e.g. 6).

When the eavesdropper is present, the key-communication function is not always explicitly derived for vector Gaussian sources since the aforementioned commutative condition does not always hold. This motivates us to consider the maximum amount of secret key obtainable per bit of communication, denoted by $\eta_{Z}(X; Y)$. For vector Gaussian sources $\eta_{Z}(X; Y)$ can always be explicitly found; and in order to upper bound $\eta_{Z}(X; Y)$ we use an idea similar to but different than the enhanced channel introduced in 4. Analogous to $\eta_{Z}(X; Y)$ is the notion of channel capacity per unit cost, introduced in 2. As in the case of channel capacity per unit cost 2, a general formula for $\eta_{Z}(X; Y)$ can be obtained which is usually easier to compute both numerically and analytically. Some other general properties of $\eta_{Z}(X; Y)$ are discussed, including a formula of this quantity for product sources.

There is a curious connection between our results for product sources and the tensorization property of a strong data processing inequality originally studied by Erkip and Cover 8 and amended recently by Anantharam et al. 9. Suppose $P_{XY} = P_{XY|X}$ is given, and

$$s^*(X; Y) = \sup_{U - X - Y, I(U; X)\neq 0} \frac{I(U; Y)}{I(U; X)}. \quad (4)$$

In 8 it was mistakenly claimed that

$$s^*(X; Y) = \rho_m^2(X; Y) \quad (5)$$

where $\rho_m^2(X; Y)$ denotes the maximal correlation coefficient 10. In fact, 9 shows that (5) does not hold in general and gives a general but less explicit expression:

$$s^*(X; Y) = \sup_{Q_X \neq P_X} \frac{D(Q_Y \| P_Y)}{D(Q_X \| P_X)} \quad (6)$$

where $Q_X \rightarrow P_{Y|X} \rightarrow Q_Y$. Although $\rho_m^2(X; Y)$ and $s^*(X; Y)$ tensorize and do agree for some simple distributions of $P_{XY}$ such as Gaussian and binary with equiprobable marginals, it was already shown in 11 that they are not equal in general. Moreover, they are both closely linked to the problem of key generation 10,12. To add one more connection between $s^*(X; Y)$ and key generation, we demonstrate that (5) implies the tensorization property of $s^*(X; Y)$. The tensorization property of $s^*(X; Y)$ turns out to be the key to many of its applications, c.f. 11,15,16. In particular, it was shown in 11 via the tensorization of hypercontractivity of Markov operators.

Related to (memoryless) product Gaussian sources are (scalar) stationary Gaussian processes which generally have memory, since intuitively one can consider the spectral representation of stationary Gaussian processes and apply the insights from the above results concerning product sources. However there are several technical difficulties in turning this intuition into a formal proof; for example the known achievability bounds for the model under our consideration are mostly confined to memoryless sources. Thus as the first step of our proof we derive an original one-shot achievability bound via resolvability for general sources. It is relatively well known that resolvability can be applied to wiretap channels (see 17 and the references therein), and wiretap channel codes can be employed in the encoding schemes in key agreement (an idea due to 18; see also 19 Section 22.4.3). Based on these connections, a recent paper 20 derived upper and lower bounds on the key capacity for sources with memory. However those bounds may be loose, and they are still asymptotic (expressed in terms of probabilistic $\limsup$ of random variables) rather than one-shot. Moreover the setting therein is a special case of ours where the public communication rate is unlimited, and the proof technique involving modulo sums only applies to discrete sources, therefore those results are still not quite useful for resolving the achievable region for stationary Gaussian sources. In contrast, our achievability bound overcomes those issues by employing a different encoding scheme called likelihood encoder proposed recently in 21. We then apply certain asymptotic approximation theorems for Toeplitz matrices when specializing to Gaussian processes.

**Organization.** The formal definition of the key generation problem with limited one-way communication, as well as the setup of product sources and stationary Gaussian sources, are presented in Section 11. The main results are given in

---

**Figure 2.** An illustration of the product sources in (1) and (2).

---

2This notation defines a measure $Q_Y$ via $Q_Y(A) := \int P_{Y|X}(A) dQ_X(x)$ for any measurable $A \subseteq Y$.

3For the reason we just discussed, the $\rho^2_m(X; Y)$ in the expressions of efficiency functions in 12 should be replaced by $s^*(X; Y)$.

4Following our ISIT presentation of this work 13, Beigi and Gohari 14 extended such an idea and introduced several new tensorizing measures of correlation from the operational perspectives of coding theorems.
Section II. Preliminaries

A. Key Generation with One-Way Communication: Basic Setup

Throughout this paper, random variables (but not including deterministic constants) are denoted by upper-case letters, and vectors and matrices are denoted in bold face.

Consider the source model illustrated in Figure 1. Stationary sources of blocklength $n$ have the joint distribution $P_{X^Y^Z^n}$, where $X_i^n$ is a short hand notation for $(X_i, \ldots, X_j)^n$ and $X^n := X_1^n$. Upon receiving $X^n$, terminal A computes an integer $K \in \mathcal{K}$ and a message $W \in \mathcal{W}$, possibly stochastically degraded, according to $P_{K|X^n}$. The message $W$ is then sent through a noiseless public channel to terminal B, and B computes the key $\hat{K} = (W(X^n), Y^n) \in \mathcal{K}$ based on its available information. The probability of error and the measure of security are defined by

$$\begin{align*}
\epsilon_n &= \mathbb{P}[K \neq \hat{K}], \\
\nu_n &= \log |\mathcal{K}| - H(K|W, Z^n).
\end{align*}$$

A rate pair $(R, r)$ is said to be achievable if a sequence of schemes can be designed to satisfy the following conditions on the probability of disagreement and security:

$$\begin{align*}
\limsup_{n \to \infty} \frac{1}{n} \log |\mathcal{W}| &\leq r, \\
\liminf_{n \to \infty} \frac{-1}{n} \log |\mathcal{K}| &\geq R, \\
\lim_{n \to \infty} \epsilon_n = 0, \\
\lim_{n \to \infty} \nu_n = 0.
\end{align*}$$

In the remainder of Section II-A we focus on the case of stationary memoryless sources with per-symbol distribution $P_{X^nY^nZ^n}$. The achievable rate region is defined as

$$\mathcal{R}(X, Y, Z) := \{(R, r) : (R, r) \text{ is achievable}\},$$

and the key-communication function

$$R(r) := \sup\{R: (R, r) \in \mathcal{R}(X, Y, Z)\}$$

characterizes the maximal possible key rate given a certain public communication rate.

From (1), the region $\mathcal{R}(X, Y, Z)$ is the union of

$$[0, I(V; Y|U) - I(V; Z|U)] \times [I(U; V; X) - I(U; V; Y), \infty)$$

over all $U, V$ such that $(U, V) \neq X - (Y, Z)$.

For key generation with one-way communication under our consideration, only $P_{XY}$ and $P_{XZ}$ affect the achievable key-communication rates. Although beyond those joint distributions we do not need further information about the source, it is customary to say that $P_{XYZ}$ is stochastically degraded if $X \neq Y$ Z form a Markov chain under a joint distribution whose pairwise distributions are $P_{XY}$ and $P_{XZ}$. In this case, the above region can be simplified to the union of

$$[0, I(V; Y) - I(V; Z)] \times [I(V; X) - I(V; Y), \infty)$$

over all $V$ such that $V \neq X - (Y, Z)$.

For jointly Gaussian vectors $(X, Y, Z)$ it is generally not true that $P_{XYZ}$ is stochastically degraded. Thus it might seem remarkable that still only one auxiliary random variable is needed; and it can be chosen to be jointly Gaussian with the source vectors, as summarized below:

Fact 1 (3). Suppose $X^L, Y^L$, and $Z^L$ are jointly Gaussian vectors of length $L$, and $U$ and $V$ are random variables such that $(U, V) \neq X^L - (Y^L, Z^L)$ form a Markov chain. Then there exists a random vector $\bar{U}^L$ in $\mathbb{R}^L$ such that $\bar{U}^L, X^L$ are jointly Gaussian, $\bar{U}^L - X^L \sim (Y^L, Z^L)$, and

$$\begin{align*}
I(\bar{U}^L; X^L) - I(\bar{U}^L; Y^L) &\leq I(U; V; X^L) - I(U; V; Y^L), \\
I(\bar{U}^L; Y^L) - I(\bar{U}^L; Z^L) &\geq I(V; Y^L|U) - I(V; Z^L|U).
\end{align*}$$

As a consequence of Fact 1 the region $\mathcal{R}(X, Y, Z)$ is the union of

$$[0, I(U; Y) - I(U; Z)] \times [I(U; X) - I(U; Y), \infty)$$

over all $U$ such that $U \neq X - (Y, Z)$ and $U, X$ are jointly Gaussian. Note that $(U, X, Y, Z)$ are necessarily jointly Gaussian as well because of the Markov chain condition.

B. Key Generation from Product Sources

A product source is just a particular stationary memoryless source in which $P_{XYZ}$ has the structure of (1) and (2). Hence the setup for a product source model is the same as the stationary case of Part II-A with the exception that $X, Y$ and $Z$ are replaced with $L$-vectors $X, Y$ and $Z$.

We remind the reader that $\mathcal{R}(X, Y, Z)$ in II-A depends only on $P_{XY}$ and $P_{XZ}$, hence we do not need to define a product source with the more stringent condition of $P_{XYZ} = \prod_{i=1}^L P_{X_i Y_i Z_i}$. 

---

5Here we allow stochastic encoders to be consistent with the achievable scheme in Section IV even though in the literature $K$ and $W$ have often been defined as functions of $X^n$. 

---
### III. MAIN RESULTS

#### A. Secret Key Generation from Product Sources

Suppose we know the function $R_i(r)$ for each “factor” in the product source; what can we say about $R(r)$ for the whole source? As Theorem 3 elucidates, the rate splitting approach in which we produce keys separately for each factor source (with appropriately selected rates) achieves the optimal key rate. This is analogous to a result in rate distortion theory [5] as remarked in the introduction.

**Theorem 1.** In the problem of key generation from product sources satisfying (1) and (4), the maximum key rate satisfies

$$R(r) = \max_{\sum_{i=1}^{L} r_i \leq r} \sum_{i=1}^{L} R_i(r_i), \quad (20)$$

where $R_i(r_i)$ is the key-communication function corresponding to the $i$’th source triple $(X_i, Y_i, Z_i)$. Further, if $R_i(\cdot)$ is differentiable and $(r_1^*, \ldots, r_L^*)$ achieves the maximum in (20), then for each $i$, either $R_i'(r_i^*) = \mu$ for some constant $\mu$ or $r_i^* = 0$.

**Remark 1.** The result of (20) can be equivalently expressed as $R(X, Y, Z) = \sum_{i=1}^{L} R_i(X_i, Y_i, Z_i)$, where the summation is the Minkowski sum of sets in the Euclidean space.

**Proof:** Each key rate of $R_i(r_i^*)$ can be approached by a scheme that operates on the $i$’th source triple separately using a communication rate of $r_i^*$. From (2), the combination of these schemes forms a legitimate scheme for the product source, since the keys generated by those schemes are independent and their combination is asymptotically independent of $W$ and $Z^n$. Thus $\geq$ holds in (20).

By (16) the achievable region $\mathcal{R}(X^L, Y^L, Z^L)$ is the union of

$$[0, I(V; Y^L|U) - I(V; Z^L|U)] \times [I(U; V; X^L) - I(U, V; Y^L), \infty) \quad (21)$$

over all $(U, V)$ such that $(U, V) - X^L - (Y^L, Z^L)$. The achievable region with rate splitting is the union of

$$\left[0, \sum_{i=1}^{L} I(V_i; Y_i|U_i) - \sum_{i=1}^{L} I(V_i; Z_i|U_i) \right] \times \left[\sum_{i=1}^{L} I(U_i, V_i; X_i) - \sum_{i=1}^{L} I(U_i, V_i; Y_i), \infty\right) \quad (22)$$

over all $(U_i, V_i)$ such that $(U_i, V_i) - X_i - (Y_i, Z_i)$, which contains the union of (21), according to Lemma 6 in Appendix A. Hence we also have $\leq$ in (20).

The last claim in the theorem for differentiable $R_i(\cdot)$ can be verified from the KKT condition and the fact that $R_i(\cdot)$ is a concave function for each $i$.

From Theorem 3 we derive the communication-rate–key-rate tradeoff for product Gaussian sources. The solution displays a “water-filling” behaviour which is reminiscent of the rate-distortion function for Gaussian vectors [6].

---

6From this argument we see that $\geq$ in (29) only requires (4) but not (1). In words, a rate-splitting key-agreement scheme designed for product sources will be reliable and secure even if the $Y$ vector is correlated. This can only correlate the decoding errors, which are negligible anyway.

**Theorem 2.** If $(X^L, Y^L, Z^L)$ are product Gaussian sources, then the achievable communication and key rates are parameterized by $\mu > 0$ as

$$r = \frac{1}{2} \sum_{i: \beta_i > \mu} \log \frac{\beta_i(\mu + 1)}{(\beta_i + 1)\mu}, \quad (23)$$

$$R = \frac{1}{2} \sum_{i: \beta_i > \mu} \log \frac{\beta_i + 1}{\mu + 1}, \quad (24)$$

where

$$\beta_i = \frac{\rho_{X_i, Y_i}^2 - \rho_{X_i, Z_i}^2}{1 - \rho_{X_i, Y_i}^2}. \quad (25)$$

**Remark 2.** The usefulness of the $i$-th component of the product source is completely captured by $\beta_i$. In (24) and (25) the $i$’th term enters the summations if and only if $\beta_i$ is large enough; in other words, only the components that are strong enough are “on”. This is similar to water-filling over Gaussian channels (avoiding low SNR channels) and rate-distortion (neglecting to compress weak source components).

**Remark 3.** In Theorem 2 if we drop assumption that $Y^L$ is Gaussian, i.e., only assume that $(X^L, Y^L, Z^L)$ is a product source where $X^L, Z^L$ are jointly Gaussian, and $(X^L, Y^L)$ has the same first and second order statistics as $(X^L, Y^L)$. If $U$ is a Gaussian auxiliary random variable as in (19), then $(U, Y^L)$ has the same first and second order statistics as $(U, Y^L)$. For arbitrary $P \ll Q$, define the relative information

$$i_{P||Q}(x) = \log \frac{dP}{dQ}(x) \quad (26)$$

as the logarithm of the Radon-Nikodym derivative. Then we have

$$I(U; Y^L) - I(U; Y^L_G)$$

$$= D(P_{U^L | Y^L} || P_U \times P_{Y^L}) - E \left[ \log \frac{P_{U^L | Y^L_G} || P_U \times P_{Y^L_G}}{(U^L, Y^L_G)} \right]$$

$$= D(P_{U^L | Y^L} || P_U \times P_{Y^L}) - E \left[ \log \frac{P_{U^L G} \times P_{Y^L}}{P_{U^L, Y^L_G}} \right]$$

$$= D(P_{U^L, Y^L} || P_{U^L G}) - D(P_{Y^L} || P_{Y^L_G})$$

$$\geq 0 \quad (27)$$

where (25) is because $i_{P_{U^L G} \times P_{Y^L_G}}(u, y^L)$ is only a second order polynomial of $(u, y^L)$. Hence the secret key can be generated more efficiently than in the Gaussian case:

$$I(U; X^L) - I(U; Y^L_G) \leq I(U; X^L) - I(U; Y^L) \quad (31)$$

$$I(U; Y^L) - I(U; Z^L) \leq I(U; Y^L_G) - I(U; Z^L) \quad (32)$$

For a positive-semidefinite matrix $\Sigma$, let $\Sigma^{-1/2}$ be a positive definite matrix such that $\Sigma^{-1/2} \Sigma \Sigma^{-1/2} = I_r$, where $I_r$ denotes the identity matrix of dimension $r = \text{rank}(\Sigma)$. Also write $\Sigma^{-1} = (\Sigma^{-1/2})^2$, which is the matrix inverse when $\Sigma$ is invertible. Note that under this definition $\Sigma^{-1/2}$ (and therefore
\( \Sigma^{-1} \) may not be unique. The following fact about Gaussian distributions is useful. The proof is based on the singular value decomposition and is deferred to Appendix C.

**Lemma 1.** For a set of vector random variables \( X, Y, Z \), there exist invertible linear transforms \( X \mapsto X', Y \mapsto Y', Z \mapsto Z \) such that all the five covariance matrices \( \Sigma_X, \Sigma_Y, \Sigma_Z, \Sigma_{XY}, \Sigma_{XZ} \) are diagonalized if and only if \( G \) commutes with \( H \) where

\[
G := \Sigma_X^{-1/2}\Sigma_{XY}\Sigma_Y^{-1}\Sigma_{YX}\Sigma_X^{-1/2}, \quad (33)
\]

\[
H := \Sigma_X^{-1/2}\Sigma_{XZ}\Sigma_Z^{-1}\Sigma_{ZX}\Sigma_X^{-1/2}. \quad (34)
\]

**Remark 4.** The linear transforms being invertible ensures \( \mathcal{R}(X, Y, Z) = \mathcal{R}(\tilde{X}, \tilde{Y}, \tilde{Z}) \).

**Remark 5.** For Hermitian matrices \( A \) and \( B \) of the same dimensions, we write \( A \preceq B \) if \( B - A \) is positive-semidefinite. From the positive definiteness of covariance matrices it’s straightforward to show that \( 0 \preceq G \preceq I \) and \( 0 \preceq H \preceq I \), where \( G \) and \( H \) are defined in (33) and (34). Indeed \( G \) and \( H \) take the roles of \( \rho_{XY}^2 \) and \( \rho_{XZ}^2 \) in the scalar Gaussian case.

**Remark 6.** If \( X^n, Y^n \) and \( Z^n \) are drawn from jointly stationary Gaussian processes, then the commutativity assumption in the lemma is satisfied approximately for \( n \) large. This is due to the commutativity of convolution.

**Corollary 1.** If \( X \) and \( Y \) are jointly Gaussian vectors, then there exist invertible linear transforms \( X \mapsto X' \) and \( Y \mapsto Y' \) such that \( \Sigma_{X'}, \Sigma_{Y'} \) and \( \Sigma_{X'Y'} \) are diagonalized.

Thanks to Corollary 1, the task of finding the key capacity of arbitrarily correlated Gaussian vector sources in the absence of an eavesdropper is reduced to the case of product Gaussian sources \( (X^L, Y^L) \) satisfying (1) and (2). Note that assuming \( X \) and \( Y \) have the same length does not lose generality since one can always pad zero coordinates to \( X \) and \( Y \) so that they have the same length. In the presence of an eavesdropper, it is not always possible to reduce the problem to the case of product sources, since the commutativity condition in Lemma 1 is not always fulfilled; we discuss its practical relevance later in Section 3.

**Proof of Theorem 2** Reference [3] derived an explicit formula for the achievable key rate in the case of scalar Gaussian sources, which, in our notations, can be expressed as

\[
R(r) = \frac{1}{2} \log \left( 1 + \frac{\beta}{\beta^+} \exp(-2r) \right) \tag{35}
\]

where \( \beta := \frac{\rho_{XZ}^2}{\rho_{XY}^2} \) and \( \beta^+ := \max\{\beta, 0\} \). The bases of log and exp in (35) depend on the unit of the information rates (e.g. bits or nats).

Now consider the product sources, and suppose that \( (r_1^*, \ldots, r_L^*) \) achieves the maximum in (20). According to Theorem 3 either \( R_i^*(r_i^*) = \frac{\beta_i^+ \exp(-2r_i^*)}{1 + \beta_i^+ \exp(-2r_i^*)} = \mu \) or \( r_i^* = 0 \) for each \( i \), where \( \mu \) is some constant. For fixed \( \mu \), this means

\[
r_i^* = \max \left\{ 0, \frac{1}{2} \log \frac{(1 + \beta_i)(1 + \beta_i^+)}{(\mu + 1 + \beta_i^+)} \right\}. \tag{36}
\]

Equivalently, we can write

\[
r_i^* = \frac{1}{2} \log \frac{\beta_i^+(m_i + 1)}{(\beta_i^+ + 1)m_i}, \tag{37}
\]

where \( m_i := \min\{\mu, \beta_i^+\} \). The claim then follows by substituting the value of \( r_i^* \) into (35) and applying (20).

**B. Secret Key per Bit of Communication**

Fix \( P_{XYZ} \). The secret key per bit of communication is defined as

\[
\eta_Z(X;Y) := \sup_{r > 0} \frac{R(r)}{r}. \tag{38}
\]

From the convexity of the achievable rate region one immediately sees that \( \eta_Z(X;Y) = \lim_{r \to 0} \frac{R(r)}{r} \).

Define

\[
s_Z^2(X;Y) := \sup_{U,V} \frac{I(V;Y|U) - I(V;Z|U)}{I(V;X|U) - I(V;Z|U) + I(U;X) - I(U;Y)}, \tag{39}
\]

where the supremum is over all \( (U, V) \) such that \( (U, V) \) and \( (X - Y, Z) \) form a Markov chain and that the denominator in (39) does not vanish. Note that the denominator is always nonnegative; if it vanishes for all \( U, V \), then so does the numerator and we set \( s_Z^2(X;Y) = 0 \). From (39) and (16) it is immediate to see how \( s_Z^2(X;Y) \) is related to \( \eta_Z(X;Y) \).

In the special case of no eavesdropper, this is related to the result in [12], which uses the incorrect constant \( \rho_{XZ}^2(X;Y) \) as we mentioned earlier.

**Theorem 3.** Secret key per bit of communication is linked to \( s_Z^2(X;Y) \) by

\[
\eta_Z(X;Y) = \frac{s_Z^2(X;Y)}{1 - s_Z^2(X;Y)}. \tag{40}
\]

**Proof:** The characterization of \( \mathcal{R}(X, Y, Z) \) is given in (16). Thus \( \sup_{r > 0} \frac{R(r)}{r} = \frac{s_Z^2(X;Y)}{1 - s_Z^2(X;Y)} \) follows immediately from the definition of \( s_Z^2(X;Y) \). The claim of \( \lim_{r \to 0} \frac{R(r)}{r} = \sup_{r > 0} \frac{R(r)}{r} \) follows from the convexity of the achievable rate region.

The following results provide some basic properties of \( s_Z^2(X;Y) \). The rationale for defining \( s_Z^2(X;Y) \) can be explained by Theorem 3 and 6 in Theorem 4.

**Theorem 4** (Properties of \( s_Z^2(X;Y) \)).

1. For any \( P_{XYZ} \),

\[
0 \leq s_Z^2(X;Y) \leq 1. \tag{41}
\]

2. For product sources \( (X^L, Y^L, Z^L) \) as in (1) and (2),

\[
s_Z^2(X^L;Y^L) = \max_{1 \leq i \leq L} s_Z^2_i(X_i;Y_i). \tag{42}
\]

3. For arbitrary \( P_{XYZ} = P_XP_Y|Z \),

\[
s_Z^2(X;Y) \leq \frac{I(\tilde{V};\tilde{Y}) - I(\tilde{V};\tilde{Z}) + D(Q_X\|P_X) - D(Q_Y\|P_Y)}{I(\tilde{V};\tilde{Z}) + D(Q_X\|P_X) - D(Q_Y\|P_Y)}. \tag{43}
\]
where \( \hat{V}, \hat{X}, \hat{Y}, \hat{Z} \) have the joint distribution

\[
P_{Y \hat{X} \hat{Y} \hat{Z}}(v, x, y, z) = Q_{V, X}(v, x)P_{Y | Z}(y | z|x). \tag{44}
\]

The supremum is over all \( Q_{V, X} \) such that the above denominator does not vanish.

Computation can be further simplified when the source has certain structures:

4) If \( P_{XYZ} \) is stochastically degraded,

\[
s^*_Z(X; Y) = \sup_{Q_X} \frac{D(Q_Y \| P_Y) - D(Q_Z \| P_Z)}{D(Q_X \| P_X) - D(Q_Z \| P_Z)} \tag{45}
\]

where \( Q_{XYZ} = Q_X P_{Y\mid Z} | X \).

5) As a special case of (45), if \( X = (X', Z), Y = (Y', Z) \), then

\[
s^*_Z(X; Y) = \text{ess sup}_{z \in \mathcal{Z}} s^*(X'; Y' | Z = z), \tag{46}
\]

where \( \text{ess sup} \) denotes the essential supremum of a real valued function.

6) If \( Z \) is constant, we recover \( s^*_Z(X; Y) = s^*(X; Y) \), which is the best constant for the strong data processing inequality defined in (6).

**Proof:** See Appendix B.

**Remark 7.** The interpretation of the tensorization of \( s^*_Z(X; Y) \) in (44) is that, with small allowable public communication, it is always efficient to only use the best component of the product sources. Alternatively, the fact that rate splitting is optimal for product sources implies the tensorization property of \( s^*_Z(X; Y) \).

**Remark 8.** If the source is stochastically degraded, then \( s_Z(X; Y) \) can be computed from (45) which only requires optimizing over an auxiliary distribution \( Q_X \), instead of the optimization over a family of distributions \( P_{U|X} \) when computing the rate region via (19). Similarly for non-degraded sources, (45) only involves optimizing over \( Q_{V, X} \) whereas the region rate region (16) requires optimizing over \( P_{UV|X} \). Thus, in either case, the optimization problem may be considerably reduced if one is only interested in \( \eta_Z(X; Y) \) instead of the whole rate region.

**Example 1 (Symmetric Bernoulli Source).** Suppose \( X, Y \) and \( Z \) are symmetric Bernoulli random variables, with \( \xi_{XY} := \mathbb{P}[X \neq Y] \) and \( \xi_{XZ} := \mathbb{P}[X \neq Z] \) satisfying \( \xi_{XY} \leq \xi_{XZ} < \frac{1}{2} \).

The achievable region \( \mathcal{R}(X, Y, Z) \) was derived in [22], from which one can obtain

\[
\eta_Z(X; Y) = \frac{(1 - 2\xi_{XY})^2 - (1 - 2\xi_{XZ})^2}{1 - (1 - 2\xi_{XY})^2}. \tag{47}
\]

Since \( X, Y, \) and \( Z \) are stochastically degraded, we can assume without loss of generality that \( X - Y - Z \) form a Markov chain, and use (40) and (45) to obtain (47). In this case (45) is supremized as \( Q_X \) approaches the equiprobable distribution on \( \{0, 1\} \).

**Example 2 (Scalar Gaussian Source).** Setting \( L = 1 \) in Theorem 2 gives

\[
\eta_Z(X; Y) = \beta \tag{48}
\]

where

\[
\beta := \frac{\rho_{XY}^2 - \rho_{XZ}^2}{1 - \rho_{XY}^2} \tag{49}
\]

for jointly Gaussian random variables \( X, Y \) and \( Z \) satisfying \( \rho_{XZ} \leq \rho_{XY} < 1 \). We remark that the (less trivial) direction of \( \eta_Z(X; Y) \leq \beta \) can also be expected from Example 1 (whereas the proof of this direction using Theorem 2 essentially relies on entropy power inequality buried in Fact 1); see Appendix D.

**Example 3 (Product Gaussian Source).** If \( X, Y \) and \( Z \) are as in Theorem 2 then

\[
\eta_Z(X; Y) = \max_{1 \leq i \leq L} \beta_i^+, \tag{50}
\]

where \( \beta_i \) is as in Theorem 2.

In addition to the potential dimension reduction in numerical evaluations (see Remark 8), another important motivation for considering \( \eta_Z(X; Y) \) is that there exist source distributions for which \( \eta_Z(X; Y) \) can be computed analytically even though \( \mathcal{R}(X, Y, Z) \) is not completely known, as epitomized by the case of vector Gaussian sources in Theorem 5 below. Note that Theorem 5 holds even when the commutativity in Lemma 11 fails. The achievability (lower bound) part of Theorem 5 is accomplished by choosing an appropriate sequence of \( Q_{V, X} \) in (43) followed by routine computations; the converse part requires slightly more ingenuity: we construct a new source distribution \( P_{XYZ} \) satisfying \( \mathcal{R}(X, Y, Z) \subseteq \mathcal{R}(X, Y, \hat{Z}) \), but for which the commutativity in Lemma 11 is fulfilled and \( \eta_Z(X; Y) = \eta_Z(X; Y) \). Details of the proof are relegated to Appendix E.

**Theorem 5.** If \( X, Y \) and \( Z \) in the key generation model are jointly Gaussian vectors, then

\[
\eta_Z(X; Y) = \lambda_{\max}^{+}((G - H)(I - G)^{-1}), \tag{51}
\]

where \( \lambda_{\max}(\cdot) \) and \( \lambda_{\min}(\cdot) \) denote the largest and smallest eigenvalues of a matrix, and recall the notation \( \lambda_{\max}^{+} := \max\{0, \lambda_{\max}\} \).

C. Key-Communication Function for Stationary Gaussian Processes

We now derive the key-rate–communication-rate tradeoff for stationary Gaussian processes \( (X, Y, Z) \). In contrast to the setting of product sources since in this section we deal with sources with memory. However as mentioned in Remark 6, one can still apply Lemma 1 and in fact the linear transforms can be easily found. Let us discuss the intuitions before diving into the formal proof. As a first attempt, it is tempting to pick the Fourier transform as the invertible linear transforms in Lemma 1 since it diagonalizes circulant matrices [23]. However this is not an allowable choice, since the linear transforms in Lemma 1 are real, thereby excluding the Fourier transform. In general, complex linear transforms are not useful.
for the conversion to product sources, since complex Gaussian variables may not be independent even if their correlation coefficient is zero.

The Fourier transform, however, is not too far from the correct choice. If a circulant matrix is symmetric, we can also diagonalize it with the sine/cosine orthogonal matrix (to be defined soon). In general, the cross-correlations $R_{XY}$ and $R_{XZ}$ are not symmetric, so the trick is to first pass $Y$ through a filter whose impulse response is $R_{XY}$, the correlation function between $X$ and $Y$, resulting in a new process $\hat{Y}$. Similarly, we construct $\hat{Z}$ by convolving with $R_{XZ}$ yielding

$$R_{X\hat{Y}} = R_{XY} * R_{Y\hat{Y}},$$

$$R_{X\hat{Z}} = R_{XZ} * R_{Z\hat{Z}},$$

which are symmetric functions. Set $\hat{X} = Q^T X^n$, $\hat{Y} = Q^T Y^n$, $\hat{Z} = Q^T Z^n$ where $Q$ is the sine/cosine orthogonal matrix, i.e., for $1 \leq k, l \leq n$,

$$Q_{kl} := \begin{cases} \cos(\frac{2\pi}{n} \frac{1}{2} k) & l \text{ is odd}; \\ \sin(\frac{2\pi}{n} \frac{1}{2} l) & l \text{ is even}. \end{cases}$$

Then the covariance matrices $\Sigma_{\hat{X}}, \Sigma_{\hat{Y}}, \Sigma_{\hat{Z}}, \Sigma_{\hat{XY}}, \Sigma_{\hat{XZ}}$ will be asymptotically diagonal as their dimension grows.

In summary, the original Gaussian sources are converted to sources satisfying the product assumption (1) and (2) in the spectral representation, and the correlation coefficients corresponding to frequency $\omega$ (which relates to the factor $\frac{2\pi}{n} \frac{1}{2}$ in (53)) are

$$\rho_{XY}(\omega) := \frac{|S_{XY}(\omega)|}{\sqrt{S_X(\omega)S_Y(\omega)}},$$

$$\rho_{XZ}(\omega) := \frac{|S_{XZ}(\omega)|}{\sqrt{S_X(\omega)S_Z(\omega)}},$$

where $S_X, S_Y, S_Z, S_{XY}, S_{XZ}$ denote the spectral densities and joint spectral densities. From (55), (56) and Theorem 2 we can anticipate the expression in the next result. To prove it rigorously we impose a technical condition that requires all correlations and cross-correlations to be absolutely summable (that is, the corresponding spectrum functions are in the “Wiener class” [23]). We do not believe this condition to be crucial for the validity of the result.

**Theorem 6.** Suppose $X$, $Y$ and $Z$ are Wiener class stationary Gaussian processes, and

$$\beta(\omega) := \frac{|S_{XY}(\omega)|^2 S_Z(\omega) - |S_{XZ}(\omega)|^2 S_Y(\omega)}{S_X(\omega)S_Y(\omega)S_Z(\omega) - |S_{XY}(\omega)|^2 S_Z(\omega)}$$

is well-defined, that is, excluding the 0/0 case. Then the achievable communication and key rates are parameterized by $\mu > 0$ as

$$r = \frac{1}{4\pi} \int_{|\beta(\omega)| > \mu} \log \beta(\omega)(\mu + 1) d\omega,$$

$$R = \frac{1}{4\pi} \int_{|\beta(\omega)| > \mu} \log \beta(\omega) + 1 d\omega.$$  

**Remark 9.**

$$\eta_{XY}(\omega) = \sup_{x, y, z \in [0, 2\pi]} (\beta^+(\omega)).$$

**Remark 10.** From (65) and (66) we can verify that

$$\beta(\omega) = \frac{\rho_{XY}^2(\omega) - \rho_{XZ}^2(\omega)}{1 - \rho_{XY}^2(\omega)},$$

which is the counterpart of $\beta_i$ in Theorem 2.

The achievability proof of Theorem 6 is given in Sections IV and V and the converse is relegated to Appendix H.

**IV. ACHIEVABILITY OF ONE-SHOT KEY GENERATION**

The single-letter expressions of (16) or (19) only apply to discrete memoryless sources. In order to allow memory, and in particular to prove the achievability part of Theorem 6 we derive a one-shot achievability result in this section. The proof relies on a stochastic encoding scheme called likelihood encoder [21]. The idea is to introduce an idealized distribution which is easier to work with, and which approximates the true distribution in total variation distance under certain rate conditions according to soft covering lemma/resolvability [24].

**Notation 1.** Given $P_{XY}$, denote the information density by

$$\nu_{X,Y}(x; y) := \log \frac{dP_{XY}}{d(P_X \times P_Y)}(x, y).$$

**Theorem 7.** Suppose the sources are distributed according to $P_{XYZ}$, the integers $M, M_1, M_2 > 0$, and $P_{U|X}$ is a conditional distribution on an arbitrary alphabet $U$. Then there is a scheme such that $|W| = M$, $|K| = M_1$, and that

$$\mathbb{P}(K \neq K^*) \leq \epsilon^*,$$

$$\log M_1 - H(K|WZ) \leq \inf_{0 < \delta \leq e^{-1} M_2^3} \left\{ (T^* + 8\delta) \log \frac{M^2}{\delta} \right\},$$

where $\epsilon^*$ and $T^*$ are defined in (111) and (112).

**Proof:** Fix the joint distribution of the sources $P_{XYZ}$. Let $P_{U|X}$ be a distribution induced by the likelihood encoder [21].

$$P_{WKL|XYZ}(w, k, l|x, y, z) = \frac{1}{Z_x} P_{X|U}(x|U(w, k, l))$$

according to $P_U$. Let $P_{WKL|XYZ}$ be the distribution induced by the stochastic encoder in (66) outputs the indices
for each bounds on the total variations with respect to the codebook distance; see for example [24].

where subsets probability measures distance and 

Note that is an equiprobable distribution, hence by the construction of the likelihood encoder we have

We now digress into a brief review of the total variation distance. By definition, the total variation distance between probability measures and , we have the following according to Theorem VII.1 in [24], we have the following

(70)

Below are some of the relevant properties of total variational distance; see for example [24].

Property 1. 1) Triangle inequality: if , and are distributions on the same sample space, then

(71)

2) If and are joint distributions on , then

where the equality holds when .

According to Theorem VII.1 in [24], we have the following bounds on the total variations with respect to the codebook :

(72)

for each , where

(73)

(74)

and is computed with the joint distribution .

From we have

(75)

(76)

and is computed with the joint distribution . Therefore by [68],

(77)

and

(78)

Equations (78), (85), (86) and the triangle inequality imply that

(79)

Lemma 2. For any ,

(80)

Proof:

(81)

where the last step used the inequality in [25].

Thanks to the lemma, for any we have

(82)

and is computed with the joint distribution . Therefore by [68],

(83)

(84)

and

(85)

(86)

(87)

(88)

(89)

(90)

(91)

(92)

(93)

(94)

(95)

(96)

(97)

(98)

(99)

(100)
where we used Jensen’s inequality in (99) and \( x \log \frac{1}{x} \leq (x + \delta) \log \frac{1}{\delta} + \frac{1}{2} \delta \) for all \( x > 0 \) and \( 0 < \delta < e^{-1} \lambda \) in (100). Averaging over the codebook and applying (83), we obtain

\[
E_C I(K; Z, W) \leq 2 \log \frac{M^2}{\delta} (T_1 + T_2 + 2T_3 + 3\delta) .
\]

Similarly from (85) we have \( E_C |P_K - Q_K| \leq T_3 \), hence

\[
E_C [\log K - H(K)] = E_C D(P_K || Q_K) \leq 2 \log \frac{M^2}{\delta} (T_3 + \delta) .
\]

Thus for the security constraint, we have

\[
E_C [\log K - H(K|WZ)] \leq 2 \log \frac{M^2}{\delta} (T_1 + T_2 + 3T_3 + 3\delta) ,
\]

which follows from (101) and (102).

For the key agreement constraint, choose a good channel decoder \( P_K^{|WY} \), and let

\[
P_{WKLYZ} = P_{WKXYZ} P_K^{|WY} ,
\]

\[
Q_{WKLYZ} = Q_{WKXYZ} P_K^{|WY} .
\]

Then using a single-shot version of Shannon’s achievable bound (26) for discrete memoryless channels, the error probability of the channel decoder can be bounded as \( E_C P_Q(\hat{K} \neq K) \leq \epsilon \) where we have defined

\[
\epsilon := \inf \{ \gamma > 0 \mid P_{U,Y | U}(\gamma) \leq \log(M_1 M_2 - 1) + \gamma \} ,
\]

and \( P_{U,Y | U} \) is computed with the joint distribution \( P_{U,Y} \). Then, the probability of decoding \( \hat{K} \) erroneously under the true distribution is bounded as

\[
E_C P_P(\hat{K} \neq K) \leq E_C P_Q(\hat{K} \neq K) + |P_X - Q_X| \leq T_3 + \epsilon ,
\]

where \( P_P \) and \( P_Q \) denote the probabilities under the distributions \( P_{XYWK} \) and \( Q_{XYWK} \), respectively. In (107) we used \( P_{WKXY|X} = Q_{WKXY|X} \), which follows from \( P_{WKXY|X} = Q_{WKXY|X} \) in (69), and that \( K \) and \( \hat{K} \) are functions of \( X \) and \( (W, Y) \), respectively. By Markov’s inequality,

\[
\frac{1}{2} \epsilon \leq \frac{1}{2}
\]

Similarly from (103),

\[
E_C \left[ \log K - H(K|WZ) > 4(T_1 + T_2 + 3T_3 + 3\delta) \log \frac{M^2}{\delta} \right] \leq \frac{1}{2} .
\]

Hence there exists a codebook which satisfies the properties in Theorem 7 where

\[
\epsilon^* := 2(T_3 + \epsilon) ,
\]

\[
T^* := 4(T_1 + T_2 + 3T_3) ;
\]

and \( T_1, T_2, T_3, \epsilon \) are as in (75), (76), (82) and (106).

V. APPROXIMATION OF GAUSSIAN PROCESSES AND ACHIEVABILITY OF THEOREM 6

In this section we apply Theorem 7 to stationary Gaussian processes to finish the achievability part of Theorem 6. The derivation is essentially based on the asymptotic distribution of the eigenvalues of Toeplitz matrices, a brief review of which is given in Appendix I.

We now introduce notations for Toeplitz matrices and circulant matrices. Given a continuous function \( f \) on \([0, 2\pi]\), define for \( k = 0, 1, \ldots, n - 1 \),

\[
t_k := \frac{1}{2\pi} \int_0^{2\pi} f(\omega) e^{i k \omega} d\omega ,
\]

\[
\epsilon_k^{(n)} := \frac{1}{n} \sum_{m=\infty}^{n-1} t_{k + mn} .
\]

Note that from (114), an equivalent way of defining \( \epsilon_k^{(n)} \) is

\[
\epsilon_k^{(n)} := \frac{1}{n} \sum_{j=0}^{n-1} f(2\pi j/n) e^{2\pi i j k/n} .
\]

If \( \{t_k\} \) has fast decay, then \( \{\epsilon_k^{(n)}\} \) approximates \( \{t_k\} \) for large \( n \). The advantage of \( \{\epsilon_k^{(n)}\} \) over \( \{t_k\} \) is that the former is a periodic sequence. For \( 0 \leq i, j \leq n - 1 \), define

\[
[T_n(f)]_{i,j} := t_{i-j} ,
\]

\[
[C_n(f)]_{i,j} := \epsilon_{i-j} .
\]

Then it is clear that (117) is a circulant matrix.

Using the above notations, the covariance matrix of the vector \( (X^n, Y^n, Z^n) \) which are samples from \( (X, Y, Z) \) can be expressed as

\[
T_n := \begin{pmatrix}
T_n(S_X) & T_n(S_{XY}) & T_n(S_{XZ}) \\
T_n(S_{XY}) & T_n(S_Y) & T_n(S_{YZ}) \\
T_n(S_{XZ}) & T_n(S_{YZ}) & T_n(S_Z)
\end{pmatrix} ,
\]

Now define a positive-semidefinite matrix composed of circulant blocks

\[
C_n := \begin{pmatrix}
C_n(S_X) & C_n(S_{XY}) & C_n(S_{XZ}) \\
C_n(S_{XY}) & C_n(S_Y) & C_n(S_{YZ}) \\
C_n(S_{XZ}) & C_n(S_{YZ}) & C_n(S_Z)
\end{pmatrix} .
\]

We assume that all the spectrums belong to the Wiener class. Then from Fact 4 in Appendix I we have

\[
T_n \sim C_n
\]

since the corresponding blocks in \( T_n \) and \( C_n \) are asymptotically equivalent. We shall use \( C_n \) as a proxy for \( T_n \) in the subsequent analysis. Let \( (X^n, Y^n, Z^n) \) be a zero mean Gaussian vector with covariance matrix \( C_n \). Suppose \( Q \) is the sin/cosine orthogonal matrix (see (54)). Define

\[
\tilde{X} = Q^\top X ,
\]

\[
\tilde{Y} = Q^\top C_n \begin{pmatrix}
S_{XY} \\
S_{XY}
\end{pmatrix} Y ,
\]

\[
\tilde{Z} = Q^\top C_n \begin{pmatrix}
S_{XZ} \\
S_{XZ}
\end{pmatrix} Z .
\]
Here $S_{XY}(\omega)$ can be arbitrarily set to 1 if $S_{XY}(\omega) = 0$. This ensures that $C_n(S_{XY}(\omega))$ is an invertible, and in particular, unitary matrix. Note that the simplified discussion in III.C corresponds to replacing $C_n(S_{XY}(\omega))$ in (123) with $C_n(S_{XY}(\omega))$, which may be singular. One can verify that $(\hat{X}, \hat{Y}, \hat{Z})$ has the product structure of (11) and (12). Next we shall specify an auxiliary distribution $P_{\tilde{U} | \tilde{X}}$. We first design the correlation coefficients $\rho_{UX} : [0, 2\pi) \to [0, 1]$ as

$$\rho_{UX}(\omega) = \begin{cases} \frac{1}{\sqrt{1 + \rho_{XY}(\omega)}}, & \beta(\omega) > \mu \\ 0, & \beta(\omega) \leq \mu \end{cases}$$

for $\omega \in [0, 2\pi)$, where $\rho_{XX}(\omega)$ and $\rho_{XX}(\omega)$ are as in (55) and (56). The definition (125) ensures that $\rho_{UX}$ satisfies

$$\log (\beta(\omega) + 1) \mu + 1 = \log \left( \frac{1}{1 - \rho_{UX}(\omega)} \right) - \log \left( \frac{1}{1 - \rho_{UX}(\omega) \rho_{XY}(\omega)} \right),$$

and

$$\log \left( \frac{1}{\rho_{UX}(\omega)} \right) \rho_{XY}(\omega) = \log \left( \frac{1}{1 - \rho_{UX}(\omega) \rho_{XY}(\omega)} \right).$$

The intuition for $\rho_{UX}$ is as follows: suppose $U$ is a Gaussian process jointly stationary with $X$ and $Y$. Then we can verify that $\hat{U}$ is zero mean jointly Gaussian with $\hat{X}$ and $\hat{Y}$, satisfying

$$\rho_{UX}(\omega) = \rho_{UX}(\omega), \quad \text{for all } \rho \in [0, 1],$$

for all $\rho \in [0, 1]$. This ensures that $\eta := u_{i}(X, U)$, in which $U, X$ are jointly Gaussian with correlation coefficient $\rho$.

**Proof:** See Appendix E

Now return to the proof of Theorem 6 Define

$$\delta := 1 - \sup_{0 \leq \omega < 2\pi} \rho_{UX}(\omega).$$

From the assumption of Theorem 6 we know that $S_{X}(\omega), S_{Z}(\omega)$ and $S_{Z}(\omega)$ do not vanish for any $\omega \in [0, 2\pi)$, since otherwise $\beta(\omega)$ will be a fraction of the type $\frac{1}{2}$ for some $\omega$. This in turn implies that

$$\min_{0 \leq \omega < 2\pi} S_{Y}(\omega) > 0,$$

and

$$\min_{0 \leq \omega < 2\pi} S_{Z}(\omega) > 0;$$

since $S_{X}(\omega), S_{Y}(\omega)$ and $S_{Z}(\omega)$ are continuous functions on the compact set $[0, 2\pi)$. We shall make an additional assumption that

$$\sup_{0 \leq \omega < 2\pi} \rho_{XY}(\omega) < 1.$$

Fortunately, the proof does not lose any generality due to the assumptions of (135):

**Lemma 4.** If Theorem 6 holds for sources satisfying (135), then it must also hold without those assumptions.

**Proof:** Assume that Theorem 6 is proved under the assumptions (135). For general source $(X, Y, Z)$ and $\lambda \in [0, 1)$, we can degrade $Y$ by $Y^\lambda := Y + \lambda N$, where $N$ is a stationary white Gaussian processes such that $\lambda$ and $(X, Y, Z)$ are independent. Let $\beta(\omega)$ be as defined in Theorem 6 but for the new source $(X, Y^\lambda, Z)$, and define

$$r^\lambda := \frac{1}{4\pi} \int_{\beta(\omega) > \mu} \log \frac{\beta(\omega)(\mu + 1)}{\beta(\omega) + 1} \mu d\omega,$$

and

$$R^\lambda := \frac{1}{4\pi} \int_{\mu < \beta(\omega) < \min_{\omega \in [0, \pi]} \beta(\omega)} \log \frac{\beta(\omega) + 1}{\mu + 1} d\omega.$$
where \( [129] \) used the monotonically increasing property of the rational function on the right hand side of \( [138] \) in \( \rho_{XY}^2(\omega) \in [0, (1+\mu)\rho_{XY}^2(\omega) - \mu) \). This means that \( \delta > 0 \) in \( [139] \), which will be essential to applying Lemma \([3]\).

For Wiener class Gaussian processes, the spectral function is continuous. Hence from \([130], [131] \) and the definition of Riemann integral we have

\[
\frac{1}{n} \sum_{i=1}^{n} \mathbb{E} \eta_i(n) \to \frac{1}{4\pi} \int \log \left( 1 - \rho_{UX}^2(\omega) \right) d\omega \quad (141)
\]

\[
= I(U; X). \quad (142)
\]

Now fix \( B > I(U; X) \). Define \( P_{U|X} = P_{\hat{U}|\hat{X}} \). According to Corollary \([1]\) there exist non-degenerate linear transforms on \( U \) and \( X \) to obtain \( \hat{U} \) and \( \hat{X} \) such that \( P_{\hat{U}\hat{X}} = \prod_{i=1}^{n} P_{\hat{U}_i\hat{X}_i} \). Let \( \hat{p}_i(n), i = 1, \ldots, n \) be the correlation coefficients between \( U_i \) and \( \hat{X}_i \). From the proof of Lemma \([1]\) one can verify that \( \hat{p}_i(n)^2, i = 1, \ldots, n \) are eigenvalues of \( I - \hat{\Sigma}_X^{-\frac{1}{2}} \hat{\Sigma}_X \sum_{U} \hat{\Sigma}_X^{-\frac{1}{2}} \), and \( p_i(n)^2, i = 1, \ldots, n \) are eigenvalues of \( I - \Sigma_X^{-\frac{1}{2}} \Sigma_X \sum_{U} \Sigma_X^{-\frac{1}{2}} \). However these two matrices are asymptotically equivalent, and their largest eigenvalues are uniformly upper bounded away from one, which follows immediately from Fact \([5]\) and the following result.

**Lemma 5.** Under the assumptions \([134] \) and \([135] \), we have

(a) \[
\Sigma_X \sim \Sigma_{\hat{X}}, \quad \Sigma_Y \sim \Sigma_{\hat{Y}}, \quad \Sigma_Z \sim \Sigma_{\hat{Z}}. \quad (143)
\]

Moreover, the smallest eigenvalues of these matrices are uniformly bounded (meaning that the bound is independent of \( n \)) away from zero, and their largest eigenvalues are also uniformly upper bounded.

(b) \[
\Sigma_{X|U} \sim \Sigma_{\hat{X}|\hat{U}}, \quad \Sigma_{Y|U} \sim \Sigma_{\hat{Y}|\hat{U}}, \quad \Sigma_{Z|U} \sim \Sigma_{\hat{Z}|\hat{U}}. \quad (144)
\]

Moreover, the smallest eigenvalues of these matrices are uniformly bounded away from zero.

**Proof:** See Appendix \([C]\).

Therefore \( \{\hat{p}_i(n)^2\} \) is asymptotically equally distributed as \( \{(\rho_i(n)^2)\} \) on \( [0, 1 - \delta_0] \) for some \( \delta_0 > 0 \) according to Fact \([6]\).

It follows that for any continuous function \( F \) on \( [0, 1 - \delta_0] \),

\[
\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} F(\hat{p}_i(n)^2) = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} F(\rho_i(n)^2). \quad (145)
\]

Define \( \eta_i(n) = \frac{\alpha_i}{\hat{U}_i - \hat{U}_{i+1}} \). Then for some \( \epsilon > 0 \), there exists \( t > 0 \) such that for all \( n \),

\[
\frac{1}{n} \ln \mathbb{P}(iX, U(X; U) \geq nB) = \frac{1}{n} \ln \mathbb{P} \left( \frac{1}{n} \sum_{i=1}^{n} \eta_i(n) \geq B \right) \quad (146)
\]

\[
\leq \frac{1}{n} \sum_{i=1}^{n} \ln \mathbb{E} \eta_i(n) - tB \quad (147)
\]

\[
\leq t(1 + \epsilon) \frac{1}{n} \sum_{i=1}^{n} \mathbb{E} \eta_i(n) + et - tB \quad (148)
\]

where \( [147] \) is from Markov’s inequality (or the Chernoff bound) and \( [148] \) uses Lemma \([3]\) and the fact that \( |\hat{p}_i(n)| < \sqrt{1 - \delta_0} \). Now let \( F: \ x \mapsto \frac{1}{2} \log \left( \frac{1}{1-x} \right) \). From \([143]\) and \([142]\), there exists \( n_0 > 0 \) such that for \( n > n_0 \),

\[
\frac{1}{n} \sum_{i=1}^{n} \mathbb{E} \eta_i(n) < I(U; X) + \frac{\epsilon}{1 + \epsilon}. \quad (149)
\]

Then \( [148] \) and \([149]\) imply that for \( n > n_0 \),

\[
\frac{1}{n} \ln \mathbb{P}(iX, U(X; U) \geq nB) < t[(1 + \epsilon)I(U; X) + 2\epsilon - B] \quad (150)
\]

\[
< -\epsilon. \quad (151)
\]

To finish the achievability proof, we need to show that the bounds in Theorem \([7]\) converge to zero for rate pairs in the interior of \( \mathcal{R}(X, Y, Z) \). An inspection of the bounds in Theorem \([7]\) reveals that it suffices to show that \( n \to \infty \)

1. \( \mathbb{P}(iU, X(X; U) > nB) \) converges to 0 exponentially fast;
2. \( \mathbb{P}(iU, Y(X; U) < nC) \) converges to 0;
3. \( \mathbb{P}(iU, Z(X; U) > nD) \) converges to 0 exponentially fast, for \( P_{UXYZ} := P_{\hat{U}\hat{X}\hat{Y}\hat{Z}} \) and any \( B > I(U; X), C < I(U; Y) \) and \( D > I(U; Z) \). Speed of convergence is imposed in 1) and 3), so that upon choosing \( \delta \) to be exponentially decreasing in \( n \), the term

\[
T^* + 8\delta = 4(T_1 + T_2 + 3T_3 + 2\delta) \quad (152)
\]

in \([64]\) is also exponentially decreasing in \( n \), thus annihilating the term \( \log \frac{M}{n} \) in \([64]\), which grows linearly in \( n \). From \([151]\) we see the validity of property 1).

The proof of 3) follows the same steps as that of 1). Similar to \([142]\), we have

\[
\frac{1}{n} \ln \mathbb{P}(iU; Z(X; U) \geq nD) \leq t(1 + \epsilon) \frac{1}{n} I(U; Z) + et - tD. \quad (156)
\]

Then \([155]\) and \([156]\) will imply 3) once

\[
\lim_{n \to \infty} \frac{1}{n} I(U; Z) - I(U; \hat{Z}) = 0 \quad (157)
\]

is established. Now suppose \( U \to X \) and \( Z \to Z \) are the diagonalizing linear transforms in Lemma \([1]\). Then it suffices to show that \( \{p_{\hat{U}\hat{Z}}^2\}_{i=1}^{n} \) and \( \{\hat{p}_{\hat{U}\hat{Z}}\}_{i=1}^{n} \) are asymptotically equally distributed on \( [0, 1 - \delta_0] \). Indeed, we first note that \( \max_{1 \leq i \leq n} |p_{\hat{U}\hat{Z}}| \) is the maximal correlation coefficient between \( U \) and \( Z \), and \( \max_{1 \leq i \leq n} p_{\hat{U}_i\hat{Z}_i} \) is the maximal correlation coefficient between \( \hat{U} \) and \( \hat{Z} \). Hence we have shown
that \( \rho_{U,i}^2 \) and \( \rho_{Z,i}^2 \) are bounded in \([0, 1 - \delta_0]\). To show their asymptotic equidistribution, it remains to prove that

\[
I - \Sigma_z^{-1} \Sigma_{z|u} \Sigma_z^{-1} \sim I - \Sigma_z^{-1} \Sigma_{z|u} \Sigma_z^{-1} \tag{158}
\]

which follows immediately from Lemma 5 and Fact 5.

The proof of 2) is simpler: without an requirement on the speed of convergence, we can just use a coarse upper bounded via Chebyshev’s inequality:

\[
P(\|u; y\| < nC) \leq \frac{\Var(\|u; y\|)}{n^2(\frac{1}{n} I(u; y) - C)^2} \tag{159}
\]

The roles of \( Y \) and \( Z \) are identical to the counterparts of (155) and (157) respectively, so we have

\[
\lim_{n \to \infty} \frac{1}{n} I(\|u; y\|) = I(\|u; y\|). \tag{160}
\]

Suppose \( U \to U \) and \( Z \to Y \) are the diagonalizing linear transforms in Lemma 1. Then as before \( \max_{1 \leq i \leq n} \rho_{U,i}^2 \leq 1 - \delta_0 \) which is uniformly upper bounded for all \( n \). Hence there exists a uniform upper bound \( \Var(\|u; y\|) = \sum_{1 \leq i \leq n} (\rho_{u,i}^2)^2 \) for some \( V > 0 \) independent of \( n \). Then \( \Var(\|u; y\|) \leq n \cdot V \), and so condition 2) is true by virtue of (159) and (160). The achievability proof for Theorem 6 is completed.

Remark 11. Although the assumption that \( \beta(\omega) \) is well defined for each \( \omega \in [0, 2\pi] \) in Theorem 6 is fairly reasonable, it is still possible that \( \beta(\omega) \) is not defined for a set of frequencies of measure zero yet the Lebesgue integrals in (58) and (59) still make sense. In such a case, we no longer have the convenient conditions in (134). However, if only the first two conditions in (134) are unfulfilled and \( \min_{0 \leq \omega \leq 2\pi} s_{2z}(\omega) > 0 \) remains true, we can still prove Theorem 6 by showing the achievable rate for some degraded X and Y first and then applying the closure property of the achievable region, which is similar to the argument in Lemma 4. Nonetheless, our proof cannot be easily extended to the case where \( \min_{0 \leq \omega \leq 2\pi} s_{2z}(\omega) = 0 \), since degrading the eavesdropper’s observation can only augment the achievable region.

VI. DISCUSSION

As remarked earlier, Theorem 6 is analogous to a rate distortion theorem for product sources under additive distortion measure; in fact one can show a similar result for channel capacity with additive cost constraints. Related phenomena in information theory also include the additivity of channel capacity (without input constraints) and Wyner’s common information [27]. In those cases, the achievable rate region of the product source/channel is the Minkowski sum of the achievable region of the factor sources/channels. The evidence points to the principle that rate splitting is optimal for product resources asymptotically in most information theoretic problems admitting single-letter solutions. Indeed, the algebraic manipulations in the converse proofs usually rely only on the independence of \( \{X_i\} \), and do not require them to be identically distributed. Hence the main element in proving such a result about rate splitting (e.g. Lemma 6 in the appendix) is usually related to the converse proof of the corresponding coding theorem. However, there are a number of examples where the achievable regions fail to satisfy such an additive property (c.f. a relay broadcast channel discussed in Remark 17), although the exact region is not known. Moreover, this rule also fails quite often for coding problems of combinatorial nature. For example, the additivity of zero error capacity was a famous conjecture [5][31] which has now been disproved [32].

It is also interesting to consider the constant

\[ s_+^*(X; Y) := \inf_{U - X - Y} I(U; X) = \inf_{Q_X \neq P_X} D(Q_X || P_X) / D(Q_Y || P_Y) \]

where \( Q_X \to P_Y | X \to Q_Y \). Interestingly, \( s_+^*(X; Y) \) does not tensorize, and in fact it usually vanishes exponentially in \( L \) for i.i.d. \( \{X_i, Y_i\}_{i=1}^L \). Indeed if \( H(X|Y) > 0 \), we can choose \( R \in (I(X; Y), H(X)) \). Set \( P_{Y^L|X^L} = \prod_{i=1}^L P_{Y_i|X_i} \) and \( P_{X^L} = \prod_{i=1}^L P_{X_i} \). By resolvability/soft covering lemma and its strong converse [27][33][24], we can choose a \([2^R]\)-type distribution \( Q_{X^L} \) and set \( Q_{X^L} \to P_{Y^L|X^L} \to Q_{Y^L} \) such that \( D(Q_{Y^L} || P_{Y^L}) \) converges to zero exponentially as \( n \to \infty \) whereas \( D(Q_{X^L} || P_{X^L}) \) is bounded away from zero, from which the exponential decay of \( D(Q_{X^L} || P_{X^L}) \) follows. This implies, among other things, that no information theoretic problem can have a single-letter solution of the form \( I(U; Y), \infty) \times [0, I(U; X)] \).

VII. ACKNOWLEDGMENTS

We are pleased to acknowledge Sanket Satpathy for suggesting the Minkowski sum interpretation in Theorem 3 and Shun Watanabe for pointing out the last two examples in footnote 8. This work was supported by NSF under Grants CCF-1350595, CCF-1116013, CCF-1319299, CCF-1319304, and the Air Force office of Scientific Research under Grant FA9550-15-1-0180, FA9550-12-1-0196.

APPENDIX A

A KEY OBSERVATION FOR PRODUCT SOURCES

The following observation is central to the proof of both tensorization property of \( s_2^*(X; Y) \) and the optimality of rate splitting in Theorem 3. It thus manifests how the two problems are inherently connected.

Lemma 6. Suppose that \( \{X_i, Y_i, Z_i\}_{i=1}^L \) possess the product structure of (2) and (3), and \( (U, V) \) are r.v.’s such that \( (U, V) - X^L - (Y^L, Z^L) \). Then there exist \( U^L \) and \( V^L \) such

---

8Exceptions to this principle do exist, for example the key generation with an omniscient helper problem [28], the mismatched broadcast channel with a common message [19, Remark 9.6], and lossy compression with mismatched side-information [29].

9In [33] a probability distribution \( P \) is called \( M \)-type if \( P(a)M \) is an integer for each \( a \in A \).
that \((U_i, V_i) - X_i - (Y_i, Z_i)\) for \(i = 1, \ldots, L\) and

\[
I(U, V; X^L) - I(U, V; Y^L) \geq \sum_{i=1}^L [I(U_i, V_i; X_i) - I(U_i, V_i; Y_i)],
\]

\(\text{(162)}\)

holds, which is a standard identity in multiuser information theory (see for example [34] Lemma 4.1).

Next, observe that

\[
I(U, V; X^L) - I(U, V; Y^L)
\]

\[
= \sum_{i=1}^L [I(U_i, V_i; X_i|X_{i+1}^L, Y_{i-1}) - I(U_i, V_i; Y_i|X_{i+1}^L, Y_{i-1})]
\]

\(\text{(166)}\)

\[
= \sum_{i=1}^L [I(U_i, V_i; X_{i+1}^L, Y_{i-1}; X_i) - I(U_i, V_i; X_{i+1}^L, Y_{i-1}; Y_i)]
\]

\(\text{(167)}\)

\[
= \sum_{i=1}^L [I(U_i, V_i; X_{i+1}^L, Y_{i-1}, Z_{i+1}^L; X_i) - I(U_i, V_i; X_{i+1}^L, Y_{i-1}, Z_{i+1}^L; Y_i)]
\]

\(\text{(168)}\)

\[
= \sum_{i=1}^L [I(U_i, V_i; Y_{i-1}^L; X_i) - I(U_i, V_i; Y_{i-1}^L, Z_{i+1}^L; X_i)]
\]

\(\text{(169)}\)

\[
\geq \sum_{i=1}^L [I(U_i, V_i; Y_{i-1}^L, Z_{i+1}^L; X_i) - I(U_i, V_i; Y_{i-1}^L, Z_{i+1}^L; Y_i)]
\]

\(\text{(170)}\)

\[
= \sum_{i=1}^L [I(\bar{U}_i; V_i; X_i) - I(\bar{U}_i; V_i; Y_i)],
\]

\(\text{(171)}\)

where \(\text{(166)}\) is again an application of [34] Lemma 4.1, and \(\text{(167)}\) is from the independence \((X_i, Y_i) \perp (X_{i+1}^L, Y_{i-1})\). Equality \(\text{(168)}\) follows from the Markov condition \(Z_{i+1}^L = (U_i, V_i, X_{i+1}^L, Y_{i-1})\) \(\setminus (X_i, Y_i)\). Inequality \(\text{(170)}\) is because of \(I(X_{i+1}^L; X_i|U_i, V_i, Y_{i-1}, Z_{i+1}^L) = I(X_{i+1}^L; X_i|U_i, V_i, Y_{i-1}, Z_{i+1}^L)\) due to the Markov condition \(X_{i+1}^L = (U_i, V_i, Y_{i-1}, Z_{i+1}^L, X_i) - Y_i\).

Finally, for each \(i\) let \(U_i\) be a r.v. such that

\[
P_{U_i|V_i, X_i, Y_i, Z_i}(u_i|v_i, x_i, y_i, z_i) = P_{U_i|V_i, X_i}(u_i|v_i, x_i).
\]

\(\text{(172)}\)

Then \((U_i, V_i, X_i, Z_i)\) and \((\bar{U}_i, V_i, X_i, Z_i)\) have the same distribution, hence

\[
I(V_i; Z_i|U_i) = I(V_i; \bar{Z}_i|\bar{U}_i).
\]

\(\text{(173)}\)

By the same token, we also have

\[
I(V_i; Y_i|U_i) = I(V_i; \bar{Y}_i|\bar{U}_i),
\]

\(\text{(174)}\)

\[
I(U_i, V_i; Y_i) = I(\bar{U}_i, \bar{V}_i; \bar{Y}_i),
\]

\(\text{(175)}\)

\[
I(U_i, V_i; X_i) = I(\bar{U}_i, \bar{V}_i; X_i),
\]

\(\text{(176)}\)

\[
I(V_i; Z_i|U_i) = I(V_i; \bar{Z}_i|\bar{U}_i).
\]

\(\text{(177)}\)

Therefore we see that \(\text{(165)}, \text{(171)}\) imply the desired result, once we make the substitutions with \(\text{(173)}\) \(\text{-} \text{(177)}\). \(\blacksquare\)

In the case where \(Z_i^L\) does not exist, the tensorization property of \(s^*(X; Y)\) and Theorem 3 can also be proved using the following result.

**Lemma 7.** Suppose that \(\{(X_i, Y_i)\}_{i=1}^L\) possess the product structure of \(\text{(1)}\) \(\text{and} (2)\), and \(U\) is a r.v. such that \(U - X^L - Y^L\).
Then there exist $U^L$ such that $U_i - X_i - Y_i$ for $i = 1, \ldots, L$ and

$$I(U; X^L) = \sum_{i=1}^L I(U_i; X_i), \quad (178)$$

$$I(U; Y^L) \leq \sum_{i=1}^L I(U_i; Y_i). \quad (179)$$

**Proof:** By induction, it suffices to prove the case of $L = 2$. Let $U_1 := U$ and $U_2 := (U, X_1)$. We have:

$$I(U; X^2) = I(U; X_1) + I(U; X_2|X_1) = I(U; X_1) + [I(U; X_2|X_1) + I(X_1; X_2)]$$

$$= I(U; X_1) + I(U, X_1; X_2),$$

$$I(U; Y^2) = I(U; Y_1) + I(U; Y_2|Y_1) = I(U; Y_1) + [I(U; Y_2|Y_1) + I(Y_1; Y_2)]$$

$$= I(U; Y_1) + I(U, Y_1; Y_2) \leq I(U; Y_1) + I(U, X_1; Y_2)$$

$$= I(U; Y_1) + I(U, X_1; Y_2). \quad (180)$$

where the last equality is from the Markov chain $Y_1 - \langle U, X_1 \rangle - Y_2$.

Note that setting $Z^L$ in Lemma 6 to be a constant will imply the existence of $U^L$ satisfying $U_i - X_i - Y_i$ and the inequalities

$$I(U; X^L) \geq \sum_{i=1}^L I(U_i; X_i), \quad (181)$$

$$I(U; Y^L) = \sum_{i=1}^L I(U_i; Y_i), \quad (182)$$

which are different from (178) and (179). Hence Lemma 7 is not a special case of Lemma 6.

**Appendix B**

**Proof of Theorem 4**

1) From the data processing inequality the denominator in (39) is nonnegative, and $s_Z^*(X; Y) \leq 1$. If there exists a distribution $U$ such that $I(U; X) - I(U; Y) > 0$, we can choose $V$ independent of $U, X, Y$ so that the numerator vanishes whereas the denominator is positive, which shows that $s_Z^*(X; Y) \geq 0$. Otherwise if $I(U; X) - I(U; Y) = 0$ for all $U$, the numerator will always be nonnegative:

$$I(V; Y|U) - I(V; Z|U) = I(U; V; Y) - I(U; V; Z) - I(U; Y) + I(U; Z)$$

$$= I(U; V; X) - I(U, V; Z) - I(U; X) + I(U; Z). \quad (183)$$

Hence $s_Z^*(X; Y) \geq 0$ always holds.

Of course, from an operational viewpoint $0 \leq s_Z^*(X; Y) \leq 1$ must be true because of Part 3) as well.

2) We only show that

$$s_Z^*(X^L, Y^L) \leq \max_{1 \leq i \leq n} s_Z^*(X_i; Y_i) \quad (184)$$

since the other direction is trivial. For any $U, V$ such that $(U, V) - X^L - (Y^L, Z^L)$ and both

$$I(U, V; X^L) - I(U, V; Y^L) > 0 \quad (185)$$

and

$$I(V; Y^L|U) - I(V; Z^L|U) > 0, \quad (186)$$

let $U^L, V^L$ be as in Lemma 6 in the appendix. That is, $U^L, V^L$ are such that $(U_i, V_i) - X_i - (Y_i, Z_i)$ for each $i$ and both

$$I(U_i, V_i; X_i) - I(U_i, V_i; Y_i) \geq \sum_{i=1}^L \left[ I(U_i, V_i; X_i) - I(U_i, V_i; Y_i) \right] \quad (187)$$

and

$$I(V; Y^L|U) - I(V; Z^L|U) = \sum_{i=1}^L [I(V_i; Y_i|U_i) - I(V_i; Z_i|U_i)] \quad (188)$$

hold. Then

$$\frac{I(V; Y^L|U) - I(V; Z^L|U)}{I(U, V; X^L) - I(U, V; Y^L)} \leq \frac{\sum_{i=1}^L [I(V_i; Y_i|U_i) - I(V_i; Z_i|U_i)]}{\sum_{i=1}^L [I(U_i, V_i; X_i) - I(U_i, V_i; Y_i) - I(V_i; Y_i|U_i) - I(V_i; Z_i|U_i)] \quad (189)$$

$$\leq \max_{1 \leq i \leq L} I(U_i, V_i; X_i) - I(U_i, V_i; Y_i) \quad (190)$$

where $I$ is the set of indices such that $I(U_i, V_i; X_i) - I(U_i, V_i; Y_i) \neq 0$, and the supremum are over all $U_i, V_i$ such that $(U_i, V_i) - X_i - (Y_i, Z_i)$ and $I(U_i, V_i; X_i) - I(U_i, V_i; Y_i) \neq 0$. Supremizing with respect to $U, V$ on the right hand side of (189) shows the tensorization property of $\frac{s_Z^*(X; Y)}{1-s_Z^*(X; Y)}$, which is equivalent to the tensorization property of $s_Z^*(X; Y)$.

3) One direction of the inequality is trivial: if $U$ and $V$ are such that $I(V; Y|U) - I(V; Z|U) \geq 0$, we have (see 191-193). For the other direction, to construct a distribution on $(U, V, X, Y, Z)$ from $Q$, we use a binary $U$ biased heavily toward zero. When $U = 1$, the distribution is as specified by $Q$. When $U = 0, V$ is independent of $(X, Y, Z)$, and the marginal distribution on $X$ balanced slightly to counteract $Q$, so that on average the distribution on $X$ is the source distribution. Even though this distribution is only rarely behaving according to $Q$ (i.e. only when $U = 1$, which has low probability), we will see that the quantity of interest only depends on $Q$. Formally, for any $Q_{V,X}$, consider

$$Q_{V,X}^{(1)} := Q_{V,X}, \quad (194)$$

$$Q_{V,X}^{(0)} := P_{V} \cdot \frac{P_X - \alpha Q_X^{(1)}}{1 - \alpha}, \quad (195)$$

$$P_{Q,UV}^2(x, u, v) := (1 - \alpha)Q_{V,X}^{(0)}(v, x)_{1_{u=0}} + \alpha Q_{V,X}^{(1)}(v, x)_{1_{u=1}}, \quad (196)$$
where $P_V$ is an arbitrary probability distribution on $\mathcal{V}$. Then clearly $P^\alpha_X = P_X$ for each $0 < \alpha < 1$. Finally, define

$$P^\alpha_{X|UV} = P^\alpha_{X|U}P_{Y|X}. \quad (197)$$

In (43) we have implicitly assumed that $D(Q_X \| P_X)$ is well defined and so the support of $Q_X$ is a subset of the support of $P_X$. Thus (195) is a well-defined distribution for $\alpha > 0$ small enough. Then, we can verify that $P^\alpha_{X|UV} = P^\alpha_{X|V}$ and the Markov chain $(U, V) - X - (Y, Z)$ with respect to $P^\alpha$. Next observe that (see (198)-(200)) as $\alpha \downarrow 0$, where $(U, V, X, Y, Z)$ has the joint distribution $P_{UVX|YZ} := P^\alpha_{UVX|YZ}$, and the distribution of $(V, X, Y, Z)$ is as in (43). Equation (198) is from the independence between $V$ and $(X, Y, Z)$ under $U = 0$. To justify (200), recall the property of relative entropy that if $P^\alpha_X := \lambda P_X + (1 - \lambda) P_0$ is a distribution for sufficiently small $\lambda > 0$, then $D(P^\alpha_X \| P_0) = o(\lambda)$. This smoothness condition implies

$$D(P_{X|U=0} \| P_X) = o(\alpha), \quad (201)$$

$$D(P_{Y|U=0} \| P_Y) = o(\alpha). \quad (202)$$

Therefore (200) is true, and the $\geq$ part of (43) holds.

4) In the case of degraded sources $X - Y - Z$, we can write (see (203)-(206)) where the first inequality is from $-D(P_Y \| P_Y) + D(P_Z \| P_Z) \leq 0$, and the second inequality used the fact that $D(P_{Y|V=\alpha} \| P_Y) - D(P_{Z|V=\alpha} \| P_Z) \geq 0$. This establishes the “$\leq$” part of (45). Conversely, for any $Q_X$, define

$$Q^{(1)}_X = Q_X, \quad (207)$$

$$Q^{(0)}_X = \frac{P_X - \alpha Q^{(1)}_X}{1 - \alpha}, \quad (208)$$

$$Q^V_{X|V}(v, x) = \alpha Q^{(1)}_{X|V}(v, x)1_{v=1} + (1 - \alpha)Q^{(0)}_{X|V}(v, x)1_{v=0}. \quad (209)$$

Let $P_{V|X} = Q^\alpha_{X}P_{Y|X}$. Notice that $P_{X|UV} = P_{X|UV}$. Then

$$I(V; Y|U) - I(V; Z|U)$$

$$= \int [I(V; Y|U = u) - I(V; Z|U = u)]dP_U(u)$$

$$\leq \sup_u I(V; X|U = u) - I(V; Z|U = u) + D(P_X|U=\alpha \| P_X) - D(P_Y|U=\alpha \| P_Y)dP_U(u) \quad (191)$$

$$= \int [I(V; Y|U = u) - I(V; Z|U = u) + D(P_X|U=\alpha \| P_X) - D(P_Y|U=\alpha \| P_Y)]dP_U(u) \quad (192)$$

$$\leq \sup Q^V_X I(V; X) - I(V; Z) + D(Q_X \| P_X) - D(Q_Y \| P_Y). \quad (193)$$

This establishes (45).

6) If the sources are of the form $X = (X', Z)$, $Y = (Y', Z)$, then we have

$$I(V; Y|U) - I(V; Z|U)$$

$$= \int [I(V; Y|U = u) - I(V; Z|U = u)]dP_U(u)$$

$$\leq \sup u I(V; X|U = u) - I(V; Z|U = u) + D(P_X|U=\alpha \| P_X) - D(P_Y|U=\alpha \| P_Y)dP_U(u) \quad (191)$$

$$= \int [I(V; Y|U = u) - I(V; Z|U = u) + D(P_X|U=\alpha \| P_X) - D(P_Y|U=\alpha \| P_Y)]dP_U(u) \quad (192)$$

$$\leq \sup Q^V_X I(V; X) - I(V; Z) + D(Q_X \| P_X) - D(Q_Y \| P_Y). \quad (193)$$

This establishes (46).
APPENDIX C

PROOF OF LEMMA [1]

With the invertible linear transform \( \bar{X} := \Sigma_{\bar{X}}^{1/2} X \), we have

\[
\Sigma_{\bar{X}} = \left( \begin{array}{ll} I_{r_x} & 0 \\ 0 & 0 \end{array} \right),
\]

where \( r_x = \text{rank}(\Sigma_X) \). Similar structures are also present in \( \Sigma_Y \) and \( \Sigma_Z \). By positive-semidefiniteness of the covariance matrix, we have the form

\[
\Sigma_{\bar{X},Y} = \left( \begin{array}{ll} A_{x,y} & 0 \\ 0 & 0 \end{array} \right),
\]

\[
\Sigma_{\bar{X},Z} = \left( \begin{array}{ll} A_{x,z} & 0 \\ 0 & 0 \end{array} \right),
\]

where \( A_{x,y} \) and \( A_{x,z} \) are \( r_x \times r_y \) and \( r_x \times r_z \) matrices, respectively. However, we also have

\[
\Sigma_{\bar{X},Y} = \Sigma_{\bar{X}}^{1/2} \Sigma_{X,Y} \Sigma_{\bar{Y}}^{1/2},
\]

\[
\Sigma_{\bar{X},Z} = \Sigma_{\bar{X}}^{1/2} \Sigma_{X,Z} \Sigma_{\bar{Z}}^{1/2},
\]

Hence if \( G \) and \( H \) as defined in [33] and [34] commute, then so do \( A_{x,y}A_{y,x} \) and \( A_{x,z}A_{z,x} \). Since commuting matrices are simultaneously diagonalizable [35], that is, there exists an orthogonal matrix \( Q_x \) such that \( Q_x A_{x,y} A_{y,x} Q_y^T \) and \( Q_x A_{x,z} A_{z,x} Q_z^T \) are diagonal. This in turn implies the existence of \( Q_y \) and \( Q_z \) such that \( Q_y A_{y,x} Q_x^T \) and \( Q_z A_{z,x} Q_z^T \) are diagonal. Therefore, after the transforms

\[
X \mapsto \bar{X} := \left( \begin{array}{ll} Q_x & 0 \\ 0 & I_{n-r_x} \end{array} \right) \Sigma_{X}^{1/2} X,
\]

\[
Y \mapsto \bar{Y} := \left( \begin{array}{ll} Q_y & 0 \\ 0 & I_{n-r_y} \end{array} \right) \Sigma_{Y}^{1/2} Y,
\]

\[
Z \mapsto \bar{Z} := \left( \begin{array}{ll} Q_z & 0 \\ 0 & I_{n-r_z} \end{array} \right) \Sigma_{Z}^{1/2} Z,
\]

the matrices \( \Sigma_{\bar{X}}, \Sigma_{\bar{Y}}, \Sigma_{\bar{Z}}, \Sigma_{\bar{X}Y}, \Sigma_{\bar{X}Z} \) and \( \Sigma_{\bar{X}\bar{Z}} \) are diagonal.

Conversely, if the asserted transforms exist, then there must exist orthogonal matrices \( Q_y \) and \( Q_z \) such that \( Q_y A_{y,x} Q_x^T \) and \( Q_z A_{z,x} Q_z^T \) are diagonal. Hence \( A_{x,y} A_{y,x} \) and \( A_{x,z} A_{z,x} \) commute, and so do \( G \) and \( H \).

APPENDIX D

CONNECTION BETWEEN GAUSSIAN AND BERNOULLI SOURCES IN EXAMPLE [2]

Suppose \( P_UVV^\top = \prod_{i=1}^L P_{U_i V_i W_i} \), and \( U_i, V_i \) and \( W_i \) are symmetric Bernoulli random variables such that

\[
1 - 2P[U_i \neq V_i] = \rho_{XY}, \quad 1 - 2P[U_i \neq W_i] = \rho_{XZ}.
\]

Define

\[
\bar{X}_L := \frac{1}{L} \sum_{i=1}^L U_i, \quad \bar{Y}_L := \frac{1}{L} \sum_{i=1}^L V_i, \quad \bar{Z}_L := \frac{1}{L} \sum_{i=1}^L W_i,
\]

where the additions are on \( \mathbb{R} \). Assuming without loss of generality that \( X, Y \) and \( Z \) have unit variances, then by central limit theorem \( P_{\bar{X}_L \bar{Y}_L} \) and \( P_{\bar{X}_L \bar{Z}_L} \) converge to \( P_{XY} \) and \( P_{XZ} \) as \( L \to \infty \), hence we expect (without a formal proof here) that \( \eta_{Z}(X;Y) = \lim_{L\to\infty} \eta_{\bar{Z}_L}(\bar{X}_L;\bar{Y}_L) \). Observe that

\[
\eta_{\bar{Z}_L}(\bar{X}_L;\bar{Y}_L) = \eta_W(\bar{X}_L;\bar{Y}_L) \leq \eta_W(U;V) \leq \eta_W(U_1;V_1) = \eta_{W_1}(U_1;V_1)
\]

where \( \bar{Z}_L \) is because \( \bar{Z}_L \) is a sufficient statistic of \( W \) for \( (\bar{X}_L, \bar{Y}_L) \); \( \bar{Z}_L \) is because processing \( U \) and \( V \) reduces key capacity; and \( \bar{Z}_L \) uses the tensorization property [42]. Then from [47] we see \( \eta_{Z}(X;Y) \leq \rho_{XY}^2 - \rho_{XY}^2 \rho_{XZ}^2 \). Note that this central limit argument is similar to a celebrated proof of Gaussian hypercontractivity using Boolean hypercontractivity.
due to Leonard Gross [36], which illustrates the interesting
correlation between Gaussian and symmetric Bernoulli distri-
butions.

**APPENDIX E**

**PROOF OF THEOREM 5**

Recall the following facts from linear algebra (see for example [37]):

**Fact 2.** If \( A \) and \( B \) are matrices of the same dimension, then \( AB^\top \) and \( B^\top A \) have the same nonzero eigenvalues.

**Fact 3.** If \( A \) is a square matrix, then
\[
|I + \epsilon A| = I + \epsilon \text{tr}(A) + O(\epsilon^2).
\] (231)

Now we are in the position of proving Theorem 5. Let \( s := \lambda_{\max}((G - H)(I - H)^{-1/2}) \). We first show that \( s^2_\Sigma(X, Y) \geq s^+ \). Since \( s^2_\Sigma(X, Y) \) is nonnegative we only need to focus on the case of \( s \geq 0 \). By restricting \( Q_{VX} \) in (233) to have the marginal distribution \( P_X \) on \( X \), we find
\[
s^2_\Sigma(X, Y) \geq \sup_{P_{V|X}} \frac{I(V; Y) - I(V; Z)}{I(V; X) - I(V; Z)}.
\] (232)

We remark that using Fact 1 one can actually show that (232) holds with equality, although we shall not use the "\( \leq \)" direction.

Let
\[
(I - H)^{-1/2} (G - H)^{1/2} (I - H)^{-1/2} = QAQ^\top
\] (233)
be the eigendecomposition of \( (I - H)^{-1/2} (G - H)^{1/2} (I - H)^{-1/2} \), where \( Q \) is an orthogonal matrix and \( A \) is a diagonal matrix. Here we can take the square root of \( I - H \) because it is a positive-semidefinite matrix according to Remark 5. By Fact 4, \( s \) is the largest eigenvalue of \( (I - H)^{-1/2} (G - H)^{1/2} (I - H)^{-1/2} \), hence we can assume without loss of generality that \( \Lambda_{1,1} = s \). For each \( \epsilon > 0 \) define the \( L \times L \) matrices
\[
D_\epsilon = \begin{pmatrix}
\epsilon & 0 \\
0 & 0
\end{pmatrix}
\] (234)
and
\[
\Delta_\epsilon = (I - H)^{-1/2} Q D_\epsilon Q^\top (I - H)^{-1/2}.
\] (235)
Choose \( V_\epsilon \) to be a random \( L \)-vector such that \( V_\epsilon \) and \( X \) are jointly Gaussian, \( V_\epsilon - X - (Y, Z) \), and
\[
\Sigma_{X|V_\epsilon} = \Sigma_{X}^{1/2} (I - \Delta_\epsilon) \Sigma_{X}^{1/2}.
\] (236)
This determines the joint distribution (up to a shift and a linear transform of \( V_\epsilon \), which are irrelevant), since the unconditional covariance of \( X \) is given in the problem statement. Then, observe that (see (237)-(240)); where the last step uses (236).

Hence,
\[
I(V_\epsilon; Y) = \frac{1}{2} \log \frac{|\Sigma_Y|}{|\Sigma_{Y|V_\epsilon}|} = -\frac{1}{2} \log |I - \Sigma_Y^{-1/2} \Sigma_{YX} \Sigma_X^{-1/2} \Delta_\epsilon \Sigma_X^{-1/2} \Sigma_{XY} \Sigma_Y^{-1/2}|
\] (241)
\[
= -\frac{1}{2} \log |I - G \Delta_\epsilon|
\] (242)
\[
= \log e^{-2 \text{tr}(G \Delta_\epsilon) + O(\epsilon^2)}
\] (243)
where (243) uses Fact 2 (or the Sylvester determinant identity) and (244) uses Fact 3. By the same token, we have shown
\[
I(V_\epsilon; X) = \frac{\log e^{-2 \text{tr}(\Delta_\epsilon) + O(\epsilon^2)}}{2}
\] (245)
and
\[
I(V_\epsilon; Z) = \frac{\log e^{-2 \text{tr}(H \Delta_\epsilon) + O(\epsilon^2)}}{2}.
\] (246)

Therefore,
\[
\lim_{\epsilon \downarrow 0} \frac{I(V_\epsilon; Y) - I(V_\epsilon; Z)}{\epsilon} = \frac{\text{tr}(G - H)}{\text{tr}(I - H)}
\] (247)
\[
= \frac{\text{tr} \left( (G - H)(I - H)^{-1/2} Q D_\epsilon Q^\top (I - H)^{-1/2} \right)}{\text{tr}(D_\epsilon)}
\] (248)
\[
= \frac{\text{tr}(A D_\epsilon)}{\text{tr}(D_\epsilon)}
\] (249)
\[
= \Lambda_{1,1}
\] (250)
\[
= s,
\] (251)
Hence by (232) we have shown that \( s^2_\Sigma(X, Y) \geq s = s^+ \).

Conversely, to show \( s^2_\Sigma(X, Y) \leq s^+ \), we may assume without loss of generality that \( s < 1 \) since Remark 5 implies that \( s \leq 1 \) and when \( s = 1 \) the claim is trivially true. We have remarked that \( s \) is the largest eigenvalue of \( (I - H)^{-1/2} (G - H)(I - H)^{-1/2} \), hence
\[
(I - H)^{-1/2} (G - H)(I - H)^{-1/2} \leq s I,
\] (252)
which implies
\[
I - G \geq (1 - s)(I - H).
\] (253)

Now define \( \tilde{H} := I - \frac{1}{1-s}(I - G) \), then
\[
I - G = (1-s)(I - \tilde{H}).
\] (254)
From (233) and (254) it is clear that
\[
\tilde{H} \preceq H.
\] (255)
By (255), we can find a Gaussian \( L \)-vector \( W \) independent of \((X, Y, Z)\) and define
\[
\tilde{Z} = Z + W
\] (256)
such that
\[
\tilde{H} = \Sigma_X^{-1/2} \Sigma_{XZ} \Sigma_Z^{-1} \Sigma_{ZZ} \Sigma_X^{-1/2}.
\] (257)
Since \( X \perp W \), we see that
\[
\hat{H} = \Sigma_X^{-1/2} \Sigma_X Z^{-1} \Sigma_{ZX}^{-1/2},
\]
which agrees with the definition (33), i.e. \( \hat{H} \) is the corresponding matrix for the source \((X, Y, Z)\). A noisier observation for the eavesdropper is advantageous for key generation, hence \( \eta_Z(X; Y) \geq \eta_Z(X; Y), \) and so \( s_Z^*(X; Y) \geq s_Z^*(X; Y). \)
Moreover from (253) we see that \( \hat{H} \) commutes with G, so that we can apply Lemma [1] to find invertible linear transforms \( X \mapsto \bar{X}, Y \mapsto \bar{Y}, Z \mapsto \bar{Z} \) such that \((\bar{X}, \bar{Y}, \bar{Z})\) is a product source in the sense of [1] and (2). Furthermore, from the proof of Lemma [1] one sees that
\[
1 - \rho_{\bar{X}, \bar{Y}_{i}} = (1 - s)(1 - \rho_{\bar{X}, \bar{Z}_{i}})
\]
(259)
for \( i = 1, \ldots, L \). Hence by (42),
\[
s_{\bar{Z}}^*(X; Y) = s_{\bar{Z}}^*(\bar{X}; \bar{Y})
= s_{\bar{Z}}^*(\bar{X}_{i}; \bar{Y}_{i})
= \left( \rho_{\bar{X}, \bar{Y}_{i}} - \rho_{\bar{X}, \bar{Z}_{i}} \right)^{+}
= s_{\bar{Z}}^+(X; Y)
\]
(260)
(261)
(262)
and we can conclude that
\[
s_{\bar{Z}}^*(X; Y) \leq s_{\bar{Z}}^+(X; Y) \leq s^+.
\]
(263)
In summary we have shown that \( s_{\bar{Z}}^*(X; Y) = s^+ \), or equivalently
\[
\eta_Z(X; Y) = s_{\bar{Z}}^+(X; Y) = \lambda^{+}((G - \hat{H})(I - G)^{-1})
\]
(265)
as desired.

**APPENDIX F**

**PROOF OF LEMMA [3]**

From Jensen’s inequality we have
\[
\ln \mathbb{E} e^{t \eta} \geq \mathbb{E} \ln e^{t \eta} = t \mathbb{E} \eta,
\]
(266)
The proof of the other part of the bound in (132) is essentially based on uniform integrality of \( \{e^{t \eta}\}_{t \in [0, \delta]} \). Without loss of generality we can assume that \( U, X \) are zero mean with unit variance. Also it suffices to consider only the case of \( \rho \geq 0 \) since otherwise the correlation coefficient between \( -U \) and \( X \) is \( -\rho < 0 \) but the distribution of \( -U; X, (\bar{U}; X) \) is the same as that of \( \eta \). Now \( N := \frac{X - \rho U}{\sqrt{1 - \rho^2}} \) is zero mean, with unit variance, and independent of \( U \). Note that
\[
|\eta| = \left| \frac{1}{2} \log \frac{1}{1 - \rho^2} - \frac{1}{2} \log e \left( \frac{\rho^2 U^2 + \rho^2 X^2}{1 - \rho^2} - 2 \rho U X \right) \right|
\]
(267)
\[
\leq \frac{1}{2} \log \frac{1}{1 - \rho^2} + \frac{1}{2} \log e \left( \frac{\rho U^2 + \rho^2 X^2}{1 - \rho^2} \right)
\]
(268)
\[
\leq \frac{1}{2} \log \frac{1}{1 - \rho^2} + \frac{1}{2} \log e \left( \frac{\rho U^2 + X^2}{1 - \rho} \right)
\]
(269)
\[
\leq \frac{1}{2} \log \frac{1}{1 - \rho^2} + \frac{\rho \log e}{2 \delta} (U^2 + X^2)
\]
(270)
\[
= \frac{1}{2} \log \frac{1}{1 - \rho^2} + \frac{\rho \log e}{2 \delta} [U^2 + (\sqrt{1 - \rho^2} N + \rho U)^2]
\]
(271)
\[
\leq \frac{1}{2} \log \frac{1}{1 - \rho^2} + \frac{3 \log e}{2 \delta} (U^2 + N^2)
\]
(272)
\[
\leq \frac{1}{2} \log \frac{1}{1 - (1 - \delta)^2} + \frac{3 \log e}{2 \delta} (U^2 + N^2).
\]
(273)
It is easy to show that for any \( \lambda < \frac{1}{4} \), \( \mathbb{E} e^{2 \lambda U^2} = \mathbb{E} e^{2 \lambda N^2} \) is finite, and hence
\[
\mathbb{E} e^{\lambda U^2} e^{\lambda N^2} \leq \sqrt{\mathbb{E} e^{2 \lambda U^2} \mathbb{E} e^{2 \lambda N^2}}
\]
(274)
(275)
Let \( \xi \) be the random variable in (273), whose distribution does not depend on \( \rho \). By (275), \( \mathbb{E} e^{\xi t} < \infty \) for all \( 0 < t < \frac{\delta}{6 \log e} \).
Now for each \( \delta < 0 \),
\[
\lim_{\Delta \to \infty} \sup_{\rho \in [0, 1 - \delta], t \in [0, \frac{1}{t \log e}]} \mathbb{E} 1_{|\eta| \geq \Delta e^{t \eta}}
\]
(267)
\[
\leq \lim_{\Delta \to \infty} \sup_{\rho \in [0, 1 - \delta], t \in [0, \frac{1}{t \log e}]} \mathbb{E} 1_{|\eta| \geq \Delta e^{t \eta}}
\]
(276)
\[
\leq \lim_{\Delta \to \infty} \sup_{t \in [0, \frac{1}{t \log e}]} \mathbb{E} 1_{\xi \geq \Delta e^{t \xi}}
\leq \lim_{\Delta \to \infty} \sup_{t \in [0, \frac{1}{t \log e}]} \mathbb{E} 1_{\xi \geq \Delta e^{t \xi}}
\leq \lim_{\Delta \to \infty} \mathbb{E} 1_{\xi \geq \Delta} \exp \left( \frac{\delta \xi}{7 \log e} \right)
= 0
\]
(278)
(279)
where the last step follows from bounded convergence theorem (or dominated convergence theorem). Then there exists \( \Delta_0 > 0 \) large enough such that
\[
\sup_{\rho \in [0, 1 - \delta], t \in [0, \frac{1}{t \log e}]} \mathbb{E} 1_{|\eta| \geq \Delta e^{t \eta}} < \frac{\xi}{4},
\]
(280)
\[
P[|\xi| < \Delta_0] > \frac{1}{2}
\]
(281)
\[
\Sigma_{Y|V_1} = \Sigma_{Y|X} + \mathbb{E} \left[ (E[Y|X] - E[Y|V_1])(E[Y|X] - E[Y|V_1])^T \right] V_1
\]
(237)
\[
\Sigma_{Y|X} + \mathbb{E} [\Sigma_{YX} \Sigma_X^{-1} (X - E[X|V_1])(X - E[X|V_1])^T \Sigma_X^{-1} \Sigma_{XY} V_1]
\]
(238)
\[
\Sigma_{Y|X} + \Sigma_{YX} \Sigma_X^{-1} \Sigma_{X|V_1} \Sigma_X^{-1} \Sigma_{XY}
\]
(239)
\[
\Sigma_Y - \Sigma_{YX} \Sigma_X^{-1} \Delta \Sigma_X^{-1} \Sigma_{XY}
\]
(240)
for $\Delta \geq \Delta_0$. Now observe that from (272), there exists a r.v. 
$\zeta = C_1(U^2 + N^2) + C_2$ such that $|\eta| < \rho \zeta$ whenever $\rho < \frac{1}{2}$, 
where $C_1, C_2$ are constants depending only on $\delta$. Then,

$$\sup_{t < \frac{1}{2}} \sup_{\rho < \delta_0} \frac{\ln Ee^{\eta t}}{t} \leq \sup_{t < \frac{1}{2}} \sup_{\rho < \delta_0} \frac{\ln Ee^{\rho t \zeta}}{t}$$

(282) \hspace{1cm} \text{(b)}

(283)

$$\leq \sup_{t < \frac{1}{2}} \frac{\ln Ee^{\delta_0 t \zeta}}{t} \leq 2 \ln E \frac{\delta_0}{\delta}$$

(284)

$$\rightarrow 0, \quad \delta_0 \rightarrow 0,$$

(285)

where (284) follows from convexity of the cumulant generating function. Thus, we can pick $\delta_0$ small enough such that

$$\sup_{t < \frac{1}{2}} \sup_{\rho < \delta_0} \frac{\ln Ee^{\eta t}}{t} \leq \epsilon.$$ 

(286)

On the other hand, for $\rho \geq \delta_0$ we have

$$\sup_{\rho \leq \delta_0 < 1-\delta} \ln \frac{Ee^{\eta t}}{t}$$

(287)

$$\leq \sup_{\rho \leq \delta_0 < 1-\delta} \frac{\ln (E1_{|\eta|\geq \Delta_0} e^{\eta t} + E1_{|\eta|< \Delta_0} e^{\eta t})}{t}$$

(288)

$$\leq \sup_{\rho \leq \delta_0 < 1-\delta} \frac{\ln (E1_{|\eta|\geq \Delta_0} e^{\eta t} + E1_{|\eta|< \Delta_0} e^{\eta t})}{tE\eta} \left(1 + \frac{\epsilon/4}{e^{-\Delta_0 P(|\eta| < \Delta_0)}}\right)$$

(289)

$$\leq \sup_{\rho \leq \delta_0 < 1-\delta} \left\{ \frac{\ln (E1_{|\eta|\geq \Delta_0} e^{\eta t} + E1_{|\eta|< \Delta_0} e^{\eta t})}{tE\eta} \right\}$$

(290)

$$\leq \sup_{\rho \leq \delta_0 < 1-\delta} \left\{ \frac{(1 + \frac{\epsilon}{4})}{e^{t\Delta_0}} \right\}$$

(291)

$$\rightarrow 1 + \epsilon/2, \quad t \rightarrow 0.$$ 

(292)

where we have defined $\beta_t := \max\{\frac{e^{t\Delta_0}}{1 + t\Delta_0}, \frac{e^{-t\Delta_0}}{1 - t\Delta_0}\}$, and used the fact that $\beta_t = 1 + O(t^2)$ when $t \rightarrow 0$. Finally, in view of

(286) and (289), there exist $t < \frac{1}{2}$ small enough such that for each $\rho \in [0, 1-\delta]$, either $\ln Ee^{\eta t} < t\epsilon$ or $\ln Ee^{\eta t} < (1 + \epsilon)E\eta$ hold.

**APPENDIX G**

**PROOF OF LEMMA 5**

(a) The asymptotic equivalences have been remarked earlier in (121), so we only have to bound the eigenvalues. From (23) Lemma 4.1] we have

$$0 < \min_{\omega \in (0, 2\pi)} S_X(\omega) \leq \lambda_{\text{min}}(\Sigma_X) \leq \lambda_{\text{max}}(\Sigma_X) \leq \max_{\omega \in (0, 2\pi)} S_X(\omega);$$

(293)

from (115) the eigenvalues of $\Sigma_X$ are

$$\{S_X(2\pi k/n)\}_{k=1}^n,$$

which are also bounded between $\min_{\omega \in [0, 2\pi]} S_X(\omega)$ and $\max_{\omega \in [0, 2\pi]} S_X(\omega)$. Similarly, the eigenvalues of $\Sigma_\hat{Y}$ and $\Sigma_Y$ are bounded between $\min_{\omega \in [0, 2\pi]} S_Y(\omega)$ and $\max_{\omega \in [0, 2\pi]} S_Y(\omega)$; and the eigenvalues of $\Sigma_\hat{X}$ and $\Sigma_X$ are bounded between $\min_{\omega \in [0, 2\pi]} S_\hat{X}(\omega)$ and $\max_{\omega \in [0, 2\pi]} S_\hat{X}(\omega)$.

We first show that $\Sigma_{X|U} \sim \Sigma_{\hat{X}|\hat{U}}$. Let $R$ be the diagonal matrix whose $(i, i)$ entry is $\rho_i^{(n)}$. Clearly both $\Sigma_{X|U}$ and $\Sigma_{\hat{X}|\hat{U}}$ depend only on $R$ and $\Sigma_X$, and do not depend on the scaling of $\hat{U}$. However, to compute $\Sigma_{\hat{X}|\hat{U}}$, it is convenient to specify $P_{\hat{U}|\hat{X}}$ via the following random transformation:

$$\hat{U} = (I - R^2)^{\frac{1}{2}} W + RX,$$

(294)

where $W$ is a zero mean Gaussian vector with covariance matrix $\Sigma_X$ and independent of $X$. Then the conditional covariance matrices can be expressed as

$$\Sigma_{\hat{X}|\hat{U}} = \Sigma_X - \Sigma_{\hat{X}X} \Sigma_{\hat{U}}^{-1} \Sigma_{\hat{U}X}$$

(295)

$$= \Sigma_X - \Sigma_{\hat{X}X} QR [(I - R^2)^{\frac{1}{2}} \Sigma_X (I - R^2)^{\frac{1}{2}} + RQ^\top \Sigma_X QR^-1 RQ^{-1} \Sigma_X.$$ 

(296)

and

$$\Sigma_{X|U} = \Sigma_X - \Sigma_{XU} \Sigma_U^{-1} \Sigma_{UX}$$

(297)

$$= \Sigma_X - \Sigma_{XU} QR [(I - R^2)^{\frac{1}{2}} \Sigma_X (I - R^2)^{\frac{1}{2}} + RQ^\top \Sigma_X QR^-1 RQ^{-1} \Sigma_X.$$ 

(298)

It is easy to see that the smallest eigenvalue of $(I - R^2)^{\frac{1}{2}} \Sigma_X (I - R^2)^{\frac{1}{2}}$ is lower bounded by $\min_{\omega \in [0, 2\pi]} S_X(\omega)$ of $\{\lambda_{\text{max}}(\Sigma_X \omega)\}$ which is positive due to (140). Therefore, Fact5 and Part (a) imply the asymptotic equivalence $\Sigma_{X|U} \sim \Sigma_{\hat{X}|\hat{U}}$. Next, from the Markov chains $U - X - Y$ and $\hat{U} - \hat{X} - \hat{Y}$, we can show that (similar to the derivations in 240)

$$\Sigma_{Y|U} = \Sigma_{Y|X} + \Sigma_{XY} \Sigma_\hat{X}^{-1} \Sigma_{X|U} \Sigma_X^{-1} \Sigma_{XY}$$

$$= \Sigma_Y - \Sigma_{XY} \Sigma_X^{-1} \Sigma_{XY}$$

(299)

and

$$\Sigma_{Y|0} = \Sigma_Y - \Sigma_{XY} \Sigma_X^{-1} \Sigma_{XY}$$

(300)

Therefore (121), $\Sigma_{X|U} \sim \Sigma_{\hat{X}|\hat{U}}$, and Part (a) immediately establish the relation $\Sigma_{Y|U} \sim \Sigma_{\hat{Y}|\hat{U}}$. Note that (299) can be written as

$$\Sigma_{Y|U} = (I - A(I - \Sigma_X^{-\frac{1}{2}} \Sigma_{X|U} \Sigma_X^{-\frac{1}{2}}))A^\top \Sigma_Y$$

(301)

where we have defined $A = \Sigma_X^{-\frac{1}{2}} \Sigma_{XY} \Sigma_X^{-\frac{1}{2}}$. From the result of Part (a) we see that

$$\lambda_{\text{max}}(I - \Sigma_X^{-\frac{1}{2}} \Sigma_{X|U} \Sigma_X^{-\frac{1}{2}}) < 1 - \delta$$

(302)

for some $\delta > 0$ which is independent of $n$. However the positive-semidefiniteness of the covariance
matrix of \((X^T, Y^T)\) implies that the largest singular value \(\sigma_{\text{max}}(A) \leq 1\), which in turn gives
\[
\lambda_{\text{max}}(A(I - \Sigma_X^{-\frac{1}{2}} \Sigma_X(U \Sigma_X^{-\frac{1}{2}}) A^T) < 1 - \delta. \tag{303}
\]

Therefore we have the uniform lower bound
\[
\lambda_{\text{min}}(\Sigma_Y(U)) \geq \min_{\omega \in [0, 2\pi]} S_Y(\omega) \quad \tag{304}
\]
\[
(1 - \lambda_{\text{max}}(A(I - \Sigma_X^{-\frac{1}{2}} \Sigma_X(U \Sigma_X^{-\frac{1}{2}}) A^T)) > \delta \min_{\omega \in [0, 2\pi]} S_Y(\omega), \quad \forall n > 0. \tag{306}
\]

A similar uniform lower bound can be obtained for \(\Sigma_{\tilde{Y} \tilde{Y}}\). The relation \(\Sigma_{2YU} = \Sigma_{2ZU}\) and the uniform lower boundedness of their eigenvalues can be shown in the exactly same way since the roles of \(Y\) and \(Z\) are equal for this problem.

**APPENDIX H**

**CONVERSE OF THEOREM [6]**

The first step towards the converse proof is to bound the key rate and the transmission rate with multi-letter expressions. This part is similar to the initial steps in the converse proof of key capacity of memoryless sources, c.f. [1].

Consider
\[
\log |K| = H(K|W; Z^n) + \nu_n \quad \tag{307}
\]
\[
\leq H(K) + \nu_n \quad \tag{308}
\]
\[
\leq H(K) - H(K|Y^n, W) + n\gamma_n + \nu_n \quad \tag{309}
\]
\[
= I(K; Y^n, W) + n\gamma_n + \nu_n \quad \tag{310}
\]
\[
\leq I(K; Y^n, W) - I(K; Z^n, W) + n\gamma_n + 2\nu_n \quad \tag{311}
\]
\[
= I(K; Y^n|W) - I(K; Z^n|W) + n\gamma_n + 2\nu_n, \quad \tag{312}
\]

where (308) and (311) are from the definition of \(\nu_n\) and (309) is from Fano’s inequality, with \(\gamma_n := \frac{1}{2}\epsilon_n \log |K| + h(\epsilon_n)\).

As for the transmission rate, note that
\[
\log |W| \geq H(W) \quad \tag{313}
\]
\[
\geq H(W|Y^n) - H(W, K|X^n) \quad \tag{314}
\]
\[
\geq H(W|Y^n) + H(K|W, Y^n) - n\gamma_n - H(W, K|X^n) \quad \tag{315}
\]
\[
= H(K, W|Y^n) - n\gamma_n - H(W, K|X^n) \quad \tag{316}
\]
\[
= I(K, W; X^n) - I(K, W; Y^n) - n\gamma_n, \quad \tag{317}
\]

where (315) used Fano’s inequality.

Now suppose \((R, r)\) is achievable, where \(r > 0, R > 0\). We identify \(K\) and \(W\) in (312), (317) with \(V, U\) respectively, and then apply Fact [1]. Also notice that \(\lim_{n \to 0} \gamma_n = \lim_{n \to \infty} \gamma_n = 0\). These imply the existence of a sequence of conditional Gaussian distributions \(P_{U^n|X}\) such that
\[
r \geq \lim_{n \to \infty} \frac{1}{n} \left[ I(X^n; U^n) - I(Y^n; U^n) \right]; \quad \tag{318}
\]
\[
R \leq \lim_{n \to \infty} \frac{1}{n} \left[ I(Y^n; U^n) - I(Z^n; U^n) \right]. \quad \tag{319}
\]

As in Section [V] let \(\hat{X}, \hat{Y}\) and \(\hat{Z}\) be jointly Gaussian vectors with circulant covariance matrices defined in (120); and \(\bar{X}, \bar{Y}, \bar{Z}\) be the result of applying the linear transforms in (122)-(124). Then
\[
r \geq \lim_{n \to \infty} \frac{1}{n} \left[ I(\hat{X}; \bar{U}) - I(\hat{Y}; \bar{U}) \right]; \quad \tag{320}
\]
\[
R \leq \lim_{n \to \infty} \frac{1}{n} \left[ I(\hat{Y}; \bar{U}) - I(\hat{Z}; \bar{U}) \right]. \quad \tag{321}
\]

For \(x > 0\) define the decreasing functions:
\[
f(x) = \frac{1}{4\pi} \int_{\beta(\omega) > x} \log (\beta(\omega) + 1) \, d\omega, \quad \tag{322}
\]
\[
g(x) = \frac{1}{4\pi} \int_{\beta(\omega) > x} \log (\beta(\omega) + 1) \, d\omega, \quad \tag{323}
\]
\[
f_n(x) := \frac{1}{2n} \sum_{i; \beta_i(n) > x} \log (\beta_i(n) + 1) \quad \tag{324}
\]
\[
g_n(x) := \frac{1}{2n} \sum_{i; \beta_i(n) > x} \log (\beta_i(n) + 1) \quad \tag{325}
\]

where
\[
\beta_i(n) := \frac{\beta_{i,1}^2(n)}{1 - \beta_{i,1}^2(n)} \quad \tag{326}
\]

The empirical distribution of \(\{\beta_i(n)\}_{i=1}^n\) converges weakly to the distribution of \(\beta(W)\) when \(W\) is uniformly distributed on \([0, 2\pi]\), which means that for any \(x > 0\) it holds that
\[
\lim_{n \to \infty} f_n(x) = f(x), \quad \tag{327}
\]
\[
\lim_{n \to \infty} g_n(x) = g(x). \quad \tag{328}
\]

By Theorem [2] there is a sequence \(\{\mu_n\}\) such that \(I(\hat{X}; \bar{U}) - I(\hat{Y}; \bar{U}) \geq f_n(\mu_n)\) and \(I(\hat{Y}; \bar{U}) - I(\hat{Z}; \bar{U}) \leq g_n(\mu_n)\), and so
\[
r \geq \limsup_{n \to \infty} f_n(\mu_n), \quad \tag{329}
\]
\[
R \leq \liminf_{n \to \infty} g_n(\mu_n). \quad \tag{330}
\]

Define \(\mu := \limsup_{n \to \infty} \mu_n\). We observe that \(\mu_n\) is bounded away from 0 and \(+\infty\); suppose on the contrary that it is not bounded away from 0. Choose \(\epsilon > 0\) small enough such that \(f(\epsilon) - \epsilon > r\) (which is possible since \(\lim_{\epsilon \to 0} f(\epsilon) = +\infty\) by monotone convergence theorem), and there is a subsequence \(\{\mu_n\}_{i=1}^\infty\) such that \(\mu_n < \epsilon\) for all \(k\). From monotonicity of \(f_n\) we see that
\[
f_{nk}(\mu_{nk}) \geq f_n(\epsilon) \to f(\epsilon), \quad k \to \infty. \quad \tag{331}
\]

This implies that \(f_{nk}(\mu_{nk}) > f(\epsilon) - \epsilon > r\) when \(k\) is sufficiently large, which contradicts (329). Similarly we can also show that \(\mu_n\) is upper bounded: if otherwise, we pick \(M > 0\) such that \(g(M) < \frac{R}{2}\) (which is possible since \(\lim_{\epsilon \to +\infty} g(x) = 0\) by monotone convergence theorem), and choose a subsequence \(\{\mu_{nk}\}_{i=1}^\infty\) such that \(\mu_{nk} > M\) for all \(k\). Then from monotonicity of \(g_n\) we see that
\[
g_{nk}(\mu_{nk}) \leq g_n(\mu) \to g(M) < \frac{R}{2}, \quad k \to \infty. \quad \tag{332}
\]
This implies that \( g_n(\mu_{i_k}) \leq \frac{1}{2} R \) for \( k \) large enough, which contradicts (330). Thus, we may assume that \( c < \mu_n < d \), for some \( 0 < c < d \).

By differentiation it’s easy to see that \( f_n = \frac{\log e}{2c(1+c)} \) Lipschitz on \([c, d]\). Now choose a new subsequence \( \{\mu_{i_k}\}_{k=1}^\infty \) which converges to \( \mu \). We have

\[
|f_{i_k}(\mu_{i_k}) - f(\mu)| \leq |f_{i_k}(\mu_{i_k}) - f_{i_k}(\mu)| + |f_{i_k}(\mu) - f(\mu)|
\]

\[
\leq \frac{\log e}{2c(1+c)} |\mu_{i_k} - \mu| + |f_{i_k}(\mu) - f(\mu)|
\]

\[
\rightarrow 0, \quad k \rightarrow \infty.
\]

(333)

where (335) used (327). Hence

\[
\lim \inf_{n \to \infty} f_n(\mu_n) \leq \lim_{k \to \infty} f_{i_k}(\mu_{i_k}) = f(\mu). \tag{336}
\]

Similarly to (333)-(336), we can also show that

\[
\lim \sup_{n \to \infty} g_n(\mu_n) \leq \lim_{k \to \infty} g_{i_k}(\mu_{i_k}) = g(\mu). \tag{337}
\]

The proof is accomplished by combining (329), (330), (336) and (337).

APPENDIX I

REVIEW OF RESULTS ON TOEPLITZ APPROXIMATION

The asymptotic distribution of the eigenvalues of Toeplitz matrices can be described in terms of the “equal distribution” introduced by H. Weyl [38].

Definition 1. [39] For each \( n \) consider two sets of \( n \) real numbers \( \{a_i^{(n)}\}_{i=1}^n \) and \( \{b_i^{(n)}\}_{i=1}^n \) satisfying

\[ A < a_i^{(n)} < B, \quad A < b_i^{(n)} < B, \quad \forall 1 \leq i \leq n, \quad n \geq 1. \tag{338} \]

for some \( A, B > 0 \). The sequences \( \{a_i^{(n)}\}_{i=1}^n \) and \( \{b_i^{(n)}\}_{i=1}^n \) are said to be asymptotically equally distributed in \([A, B]\) if for any continuous function \( F: [A, B] \to \mathbb{R} \), it holds that

\[
\lim_{n \to \infty} \frac{\sum_{i=1}^n [F(a_i^{(n)}) - F(b_i^{(n)})]}{n} = 0. \tag{339}
\]

Denote by \( \mathbb{P}_{a^{(n)}} \) the empirical distribution of \( \{a^{(n)}\} \). Then (339) can be expressed as

\[
\lim_{n \to \infty} \mathbb{E}[F(X_a) - F(X_b)] = 0, \tag{340}
\]

where the random variables \( X_a \) and \( X_b \) are distributed according to \( \mathbb{P}_{a^{(n)}} \) and \( \mathbb{P}_{b^{(n)}} \), respectively.

Definition 2. Consider the sets \( \{a_i^{(n)}\}_{i=1}^n \) of real numbers from \([A, B]\). The empirical distribution \( \mathbb{P}_{a^{(n)}} \) is said to converge weakly to a measure \( \mu \) on \([A, B]\) if for any continuous function \( F: [A, B] \to \mathbb{R} \),

\[
\lim_{n \to \infty} \mathbb{E}[F(X_a)] = \mathbb{E}[F(X_\mu)], \tag{341}
\]

where the random variables \( X_a \) and \( X_\mu \) are distributed according to \( \mathbb{P}_{a^{(n)}} \) and \( \mu \), respectively.

Definition 3. [23] We say \( A_n \) and \( B_n \) are asymptotically equivalent (denoted as \( A_n \sim B_n \)) for two sequences of matrices \( \{A_n\} \) and \( \{B_n\} \) if

1) \( A_n \) and \( B_n \) are uniformly bounded in \( \ell_2 \) operator norm, i.e. for some \( M > 0 \),

\[
||A_n||, ||B_n|| \leq M < \infty, \quad n = 1, 2, \ldots \tag{342}
\]

2) \( A_n - B_n \) converges to zero in the weak norm:

\[
\lim_{n \to \infty} ||A_n - B_n|| = 0, \tag{343}
\]

where \( |A| := (\frac{1}{n} \text{tr}(A^* A))^{1/2} \).

The key result we use in Section V is then expressed as:

Fact 4. [23] Lemma 4.6 If \( f \) is in the Wiener class then \( T_n(f) \sim C_n(f) \), where the notations \( T_n \) and \( C_n \) are as in (110) and (117).

The following property will be useful later in proving asymptotic equivalence of Toeplitz matrices. The claim about square root matrices follows from [40, Theorem 1] by particularizing the continuous function therein to the square root function, while all other claims are from [23, Theorem 2.1].

Fact 5. Sums and products of asymptotically equivalent matrices are asymptotically equivalent. If the smallest singular values of asymptotically equivalent matrices are uniformly lower bounded, then their inverses are also asymptotically equivalent. Moreover, square roots of asymptotically equivalent positive-semidefinite matrices are asymptotically equivalent.

The relevance of asymptotically equivalent matrices to coding theorems lies in the following fact:

Fact 6. [23] Theorem 2.4] Let \( A_n \) and \( B_n \) be asymptotically equivalent sequences of Hermitian matrices with eigenvalues inside the interval \([m, M]\). Then the eigenvalues of \( A_n \) and \( B_n \) are asymptotically equally distributed on \([m, M]\).

REFERENCES

[1] I. Csiszár and P. Narayan, “Common randomness and secret key generation with a helper,” IEEE Transactions on Information Theory, vol. 46, no. 2, pp. 344–366, Feb. 2000.
[2] R. Ahlswede and I. Csiszár, “Common randomness in information theory and cryptography. Part II. CR capacity,” IEEE Transactions on Information Theory, vol. 44, no. 1, pp. 225–240, Jan. 1998.
[3] S. Watanabe and Y. Oohama, “Secret key agreement from vector Gaussian sources by rate limited public communication,” in Proceedings of 2010 IEEE International Symposium on Information Theory, pp. 2597–2601.
[4] H. Weingarten, Y. Steinberg, and S. Shamai, “The capacity region of the Gaussian MIMO broadcast channel,” IEEE Transactions on Information Theory, vol. 52, no. 9, pp. 3936–3964, Sept. 2004.
[5] C. E. Shannon, “Coding theorems for a discrete source with a fidelity criterion,” IEEE Transactions on Information Theory, vol. 26, no. 7, pp. 142–163, 1959.
[6] T. M. Cover and J. A. Thomas, Elements of Information Theory, Second Edition. John Wiley & Sons, 2012.
[7] S. Verdú, “On channel capacity per unit cost,” IEEE Transactions on Information Theory, vol. 36, no. 5, pp. 1019–1030, May 1990.
[8] E. Erkip and T. M. Cover, “The efficiency of investment information,” IEEE Transactions on Information Theory, vol. 44, no. 3, pp. 1026–1040, Mar. 1998.
[9] V. Anantharam, A. Gohari, S. Kamath, and C. Nair, “On maximal correlation, hypercontractivity, and the data processing inequality studied by Erkip and Cover,” arXiv preprint arXiv:1304.6133, 2013.
[10] H. S. Witsenhausen, “On sequences of pairs of dependent random variables,” SIAM Journal on Applied Mathematics, vol. 28, no. 1, pp. 100–113.
