Optimal Signal Processing for Common Randomness Generation over MIMO Gaussian Channels with Applications in Identification

Rami Ezzine, Wafa Labidi, Christian Deppe, Holger Boche

1 Technical University of Munich, TUM School of Computation, Information and Technology, Munich, Germany
2 Technical University of Braunschweig, Institute for Communications Technology, Braunschweig, Germany
3 6G-life, 6G research hub, Germany
4 Munich Center for Quantum Science and Technology, Munich, Germany
5 Munich Quantum Valley, Munich, Germany
rami.ezzine@tum.de, wafa.labidi@tum.de, christian.deppe@tu-braunschweig.de, boche@tum.de

In memory of Ning Cai

Abstract. Common randomness (CR), as a resource, is not commonly exploited in existing practical communication systems. In the CR generation framework, both the sender and receiver aim to generate a common random variable observable to both, ideally with low error probability. The availability of this CR allows us to implement correlated random protocols that can lead to faster and more efficient algorithms. Previous work focused on CR generation over perfect channels with limited capacity. In our work, we consider the problem of CR generation from independent and identically distributed (i.i.d.) samples of a correlated finite source with one-way communication over a Gaussian channel. We first derive the CR capacity for single-input single-output (SISO) Gaussian channels. This result is then used for the derivation of the CR capacity in the multiple-input multiple-output (MIMO) case. CR plays a key role in the identification scheme since it may allow a significant increase in the identification capacity of channels. In the identification framework, the decoder is interested in knowing whether a specific message of special interest to him has been sent or not, rather than knowing what the received message is. In many new applications, such as several machine-to-machine and human-to-machine systems and the tactile internet, this post-Shannon scheme is more efficient than classical transmission. In our work, we also consider a CR-assisted secure identification scheme and develop a lower bound on the corresponding secure identification capacity.

Keywords: Common randomness · Gaussian channels · secure identification.
1 Introduction

In the common randomness (CR) generation framework, the communicating parties, often referred to as terminals, aim to generate a common random variable observable to both, ideally with low error probability [1]. The availability of this shared randomness enables the implementation of correlated random protocols, which can result in faster and more efficient algorithms [2] [3].

CR is considered a highly promising resource for future communication systems due to its essential role in various communication tasks. For instance, CR plays a key role in the identification scheme, an approach in communications developed by Ahlswede and Dueck [4]. Interestingly, CR can significantly increase the identification capacity of channels. As a result, an enormous performance gain can be achieved by taking advantage of this resource. In the identification framework, the encoder sends an identification message over the channel. In contrast to transmission [5], the decoder is now interested in knowing whether a specific message of special interest to him was sent or not, rather than knowing what the received message is. In many new applications such as several machine-to-machine and human-to-machine systems [6], industry 4.0 [7], 6G communication systems [8] [9] and digital watermarking [10] [12], it appears that the identification scheme is more efficient than the classical transmission scheme.

CR is perhaps more evident in cryptography. In fact, it is used in the secret key generation problem [13]. Note that the key generation problem is an example of common randomness generation where secure communication between sender and receiver is ensured. It is worth mentioning that an interesting scenario in this context is the use of WiFi to exploit common randomness as a key, as introduced in [14]. In our work, however, we will not impose any secrecy constraints.

Additionally, CR plays an important role in modular coding schemes for secure communication. As discussed in [15], modular schemes for semantic security have been designed to integrate with arbitrary error-correcting codes, thereby establishing semantic security. Often, in seeded modular coding scenarios, legitimate parties possess CR as an additional resource, which can be used as a seed [16].

Furthermore, it was demonstrated in [17] that exploiting CR as a resource facilitates state estimation with error-free reconstruction of the state distribution in joint sensing and communication applications. Moreover, it was established that the presence of this resource is crucial for the perfect reconstruction of the state distribution. This characteristic of CR is highly intriguing for joint sensing and communication applications in 6G [8].

The Post-Shannon resource of CR can also be leveraged to achieve inherent resilience for the tactile internet and quantum communication systems. Specifically, when legitimate parties have access to a common random source as an additional coordination resource, communication becomes resilient against denial-of-service (DOS) attacks by jammers. Remarkably, only a few bits of CR are needed to counteract the jamming attack [18]. Incorporating resilience by design is crucial for ensuring trustworthiness in 6G [19]. In [20], CR was named as an important additional resource for future 6G systems due to the aforementioned
potential for a wide range of applications. The first network operators are already starting to set up a research network infrastructure for the generation and distribution of CR.

Previous work in [1] focused on the problem of CR generation from finite sources with unidirectional communication over perfect rate-limited channels. In our work, we consider the case when the terminals communicate over single-input single-output (SISO) as well as multiple-input multiple-output (MIMO) Gaussian channels. Gaussian channels are well-known for their practical relevance in many communication situations, e.g., satellite and deep space communication links [21], wired and wireless communications, etc. We characterize the CR capacity for our specified model. The latter is defined as the maximum rate of CR one can achieve using the resources available in the model.

In our work, we also address the problem of secure identification over Gaussian wiretap channels (GWCs) with common randomness (CR) being available as a resource. Secure identification has been extensively studied for discrete alphabets [6, 22, 23] over recent decades due to its important potential use in many future scenarios. Indeed, for discrete channels, it was proved in [22] that secure identification is robust under channel uncertainty and against jamming attacks. It has been demonstrated that, in contrast to secure transmission, the identification capacity of the discrete wiretap channel coincides with the capacity of the main channel. This holds true only if the secrecy capacity elaborated in [24] is strictly positive. Recently, the results were extended to the Gaussian case in [25]. However, as far as we know, there has been limited research on the secure CR-assisted identification capacity for GWCs. The wiretap channel is a basic model considered by Wyner [26] in information-theoretic security. The wiretapper, in contrast to the discrete case, is now not limited anymore and has an infinite alphabet. Moreover, we assume the wiretapper has access to the correlated source signals. This is advantageous for him because he has no limitations on the hardware resolution. In our coding scheme, the sender wants to send a secure identification message to the legitimate receiver so that the receiver is able to identify his message. Both the sender and the receiver share an extra resource of randomness. Meanwhile, the unauthorized party attempts to identify an unknown message.

The main contributions of this work consist of deriving a single-letter formula for the CR capacity for the standard two-source model with one-way communication over SISO and MIMO Gaussian channels, as well as using the obtained results on CR capacity to provide an achievable rate for correlation-assisted secure identification over Gaussian wiretap channels.

The paper is organized as follows: In Section 2, we provide the definition of an achievable CR rate for a model including two correlated sources with one-way communication over a SISO and MIMO Gaussian channel, respectively. Additionally, we introduce the main definitions of CR-assisted identification and secure identification. In Section 3, we propose a single-letter characterization of the CR capacity for the SISO Gaussian case. We use this result to completely solve the Gaussian MIMO case by establishing the corresponding CR capacity.
In Section 4, we derive a lower bound on the secure CR-assisted identification capacity of GWCs. Section 5 encompasses concluding remarks and proposes potential future research directions in this field. Auxiliary proofs are collected in the appendix.

2 Preliminaries

In this section, we introduce the different scenarios and channel models investigated for CR generation. Additionally, we provide some basic definitions regarding CR-assisted identification over Gaussian channels and establish the notation that will be used.

2.1 Notation

\( \mathbb{C} \) denotes the set of complex numbers; \( H(\cdot) \) and \( I(\cdot; \cdot) \) are the entropy and mutual information, respectively; \( h(\cdot) \) denotes the differential entropy; all information quantities are taken to base 2; \( \|a\|_2 \) denotes the L_2 norm of a vector \( a \); \( A^H \) stands for the Hermitian transpose of the matrix \( A \), \( |A| \) stands for the cardinality of the set \( A \), \( \log \) is taken to base 2 and \( \ln \) stands for the natural logarithm. \( T^n_{P_X} \) denotes the set of typical sequences of length \( n \) and of type \( P_X \) and \( T^n_{P_Y | X}(x^n) \) denotes the set of sequences \( y^n \) of length \( n \) having conditional type \( P_Y | X \) given the sequence \( x^n \) of length \( n \).

2.2 Common Randomness Generation: Two Correlated Sources with One-Way Communication over a Gaussian Channel

A discrete memoryless multiple source DMMS \( P_{XY} \in \mathcal{P}(X \times Y) \) with two components, with generic variables \( X \) and \( Y \) on alphabets \( \mathcal{X} \) and \( \mathcal{Y} \), correspondingly, is given. The \( n \)-lengths source outputs are observable at Terminals \( A \) and \( B \), respectively.

Terminal \( A \) generates a random variable \( K = \Phi(X^n) \) with alphabet \( \mathcal{K} \) and a random sequence \( T^n = A(X^n) \). \( T^n \) is sent over a Gaussian channel with input constraint. Let \( Z^n \) be the channel output. Terminal \( B \) generates a random variable \( L \) with the same alphabet \( \mathcal{K} \) as a function of \( Y^n \) and \( Z^n \), i.e., \( L = \Psi(Y^n, Z^n) \). Here, \( \Phi, A, \Psi \) refer to functions/signal processing algorithms.

A pair of random variables \( (K, L) \) is permissible if \( K \) and \( L \) are functions of the resources available at Terminal \( A \) and Terminal \( B \), respectively i.e.,

\[
K = \Phi(X^n), \quad L = \Psi(Y^n, Z^n).
\] (1)

Remark 1. In the case of a communication over a MIMO channel with channel output \( Z^n \), a pair of random variables \( (K, L) \) is permissible if \( K \) and \( L \) are functions of the resources available at Terminal \( A \) and Terminal \( B \), respectively, i.e.,

\[
K = \Phi(X^n), \quad L = \Psi(Y^n, Z^n).
\] (2)
Definition 1. A number \( H \) is called an achievable CR rate if for sufficiently large \( n \) and every \( \alpha > 0 \), \( \delta > 0 \), there exists a permissible pair of random variables \((K, L)\) such that
\[
\mathbb{P}[K \neq L] \leq \alpha \tag{3}
\]
\[
\frac{1}{n} H(K) > H - \delta. \tag{4}
\]

Definition 2. The CR capacity is the maximum achievable CR rate [2].

2.3 Gaussian Channel Model

![Diagram of Gaussian channel model](image)

Fig. 1: Standard Two-source model with one-way Communication over a SISO Gaussian channel

We consider first the SISO Gaussian channel, as depicted in Fig. 1. Terminal A encodes \( X^n \) into a sequence \( T^n \) satisfying
\[
\mathbb{E}[|T_i|^2] \leq P. \tag{5}
\]

It follows from (5) that each input sequence \( t^n \) lies in the new constrained input set \( \mathcal{T}_{n,P} \) defined as follows:
\[
\mathcal{T}_{n,P} = \{ t^n \in \mathcal{T}^n \subset \mathbb{C}^n \text{ realization of } T^n \text{ that is subject to } \mathbb{E}[|T_i|^2] \leq P, \quad i = 1, \ldots, n \}. \tag{6}
\]

The sequence \( T^n \) is sent over a Gaussian channel with an input constraint as in (5), and \( Z^n \) is defined as the channel output, where it holds that
\[
Z_i = T_i + \xi_i, \quad i = 1 \ldots n,
\]
where $\xi_i \sim \mathcal{N}_C(0, \sigma^2)$. For simplicity, we drop the index $i$. The channel has capacity

$$C(P) = \max_{P_T} I(T; Z). \quad (7)$$

### 2.4 MIMO Gaussian Channel Model

We consider second the MIMO Gaussian channel, as depicted in Fig. 2. Terminal $A$ encodes $X^n$ into a sequence $T^n \in \mathbb{C}^{N_T \times n}$, such that

$$E[\|T_i\|_2^2] \leq P, \quad i = 1, \ldots, n. \quad (8)$$

It follows that each input sequence $t^n$ lies in the input set $\mathcal{T}_{N_T \times n, P}$, defined as follows:

$$\mathcal{T}_{N_T \times n, P} = \{ t^n \in \mathbb{C}^{N_T \times n} \text{ realization of } T^n \text{ such that } E[\|T_i\|_2^2] \leq P \quad i = 1, \ldots, n \}. \quad (9)$$

We consider the following channel model with $N_T$ transmit antennas and $N_R$ receive antennas:

$$Z_i = HT_i + \xi_i, \quad \forall i = 1, \ldots, n, \quad (10)$$

where $n$, as previously mentioned, is the number of channel uses, as shown in Fig. 2. For simplicity, we drop the index $i$. The input vector $T \in \mathbb{C}^{N_T}$ contains the $N_T$ scalar transmitted signals and fulfills the following power constraint:

$$E[\|T\|_2^2] \leq P.$$
The output vector $Z \in \mathbb{C}^{N_R}$ comprises the scalar received signals of the $N_R$ channel outputs. The channel matrix

$$H = \begin{pmatrix} h_{11} & \cdots & h_{1N_T} \\ \vdots & \ddots & \vdots \\ h_{N_R1} & \cdots & h_{N_RN_T} \end{pmatrix} \in \mathbb{C}^{N_R \times N_T}$$

is a full-rank deterministic matrix. The entry $h_{ij}$ represents the channel gain from transmit antenna $j$ to receive antenna $i$. The vector $\xi \in \mathbb{C}^{N_R}$ is the circularly symmetric Gaussian noise, $\xi \sim \mathcal{CN}(0_{N_R}, \sigma^2 I_{N_R})$.

The MIMO channel has the capacity

$$C(P, N_T \times N_R) = \max_{\mathbb{E}[\|T\|^2] \leq P} I(T; Z). \quad (11)$$

### 2.5 CR-assisted Identification over Gaussian Channels

In 1989, Ahlswede and Dueck [4] proposed the identification scheme which is conceptually different from the classical transmission scheme of Shannon. In transmission, the encoder transmits a message over a channel $W$, and at the receiver side, the decoder wants to estimate this message based on the channel observation. However, this is not the case for identification. Indeed, in the identification scheme, the encoder sends an identification message (also called an identity) $M \in \mathcal{N}$ over the channel and the decoder is not interested in what the received message is, but he wants to check whether a specific message $\hat{M} \in \mathcal{N}$ has been sent or not. Naturally, the sender has no knowledge of this specific message, otherwise it would be a trivial problem. The identification problem can be regarded as solving many hypothesis testing problems occurring simultaneously.

There are many interesting applications of the identification scheme, such as in industry 4.0, online sales, and the healthcare field [6]. For instance, in product engineering, sensors are used to control the sequence of production. The sensor data is collected and processed by a central unit. Here, the receiver is interested in checking whether or not an error occurs in the sequence of production rather than determining the accurate sensor measurements. For insight into the explicit construction of identification codes, we refer the reader to [27]. Furthermore, a special identification code construction using tag codes with two concatenated Reed-Solomon codes is implemented in [28]. In our work, we are particularly interested in studying the problem of CR-assisted identification, in which the transmitter and the receiver have access to a correlated source $P_{XY} \in \mathcal{P} (X \times Y)$ as visualized in Fig. 3. Unlike in [4], we do not assume the existence of local randomness. In what follows, $x^n$ and $y^n$ are realizations of $X^n$ and $Y^n$, respectively.

**Definition 3.** A CR-assisted $(n, N, \lambda_1, \lambda_2)$ identification code for the Gaussian channel $W$ is a family of pairs $\{(u_i, D_i(y^n)) : i = 1, \ldots, N\}$, with

$$u_i = \Phi(x^n) \in \mathcal{T}_{n, P}, \quad D_i(y^n) \subset \mathcal{Z}^n, \quad \forall \ i \in \{1, \ldots, N\},$$
such that for all $i, j \in \{1, \ldots, N\}, i \neq j$ and $\lambda_1 + \lambda_2 < 1$, the errors of the first and second kind satisfy

\begin{align}
W^n((D_i(y^n))^c | u_i) &\leq \lambda_1, \quad \text{(12)} \\
W^n(D_i(y^n) | u_j) &\leq \lambda_2 \quad \forall i \neq j.
\end{align}

**Definition 4.** $C_{ID}^c(P)$ the CR-assisted identification capacity of the channel $W$ is defined as follows:

$$C_{ID}^c(P) = \max \left\{ R : \forall \lambda > 0, \exists n(\lambda) \text{ s.t. } n \geq n(\lambda) \text{ such that } N(n, \lambda) \geq 2^{nR} \right\},$$

where $N(n, \lambda)$ is the maximal cardinality such that a $(n, N, \lambda_1, \lambda_2)$ CR-assisted identification code for the channel $W$ exists.

**Definition 5.** For the MIMO channel described in (10), a CR-assisted $(n, N, \lambda_1, \lambda_2)$ identification-code is a family of pairs $\{(u_i, D_i(y^n)) : i = 1, \ldots, N\}$, such that for some $\lambda_1 + \lambda_2 < 1$ and for all $i \in \{1, \ldots, N\}$, we have

$$u_i \in \mathcal{T}_{N_T \times n, P},$$

$$D_i(y^n) \subset Z^n = \{ z^n = (z_1, z_2, \ldots, z_n) \in \mathbb{C}^{N_R \times n} \},$$

and with errors of the first and second kind that satisfy

\begin{align}
W^n((D_i(y^n))^c | u_i) &\leq \lambda_1 \\
W^n(D_i(y^n) | u_j) &\leq \lambda_2, \forall i \neq j.
\end{align}

**Remark 2.** The definitions of identification codes for the single-user MIMO channel are similar to the SISO case, except for the dimension of input and output sets. Indeed, at each time instant $i \in \{1, \ldots, n\}$, we send $N_T$ scalar signals and receive $N_R$ signals. Thus, compared to the SISO case, the input and output sets contain matrices instead of vectors.
2.6 CR-assisted Secure Identification over Gaussian Wiretap Channels

We focus now on CR-assisted secure identification depicted in Fig. 4. We consider the following standard model of the Gaussian wiretap channel (GWC):

\[ W_g: Z_i = T_i + \xi_i, \quad \forall i \in \{1, \ldots, m\}, \]
\[ V_{g'}: Z'_i = T_i + \phi_i, \quad \forall i \in \{1, \ldots, m\}, \]  

(14)

where \( T_m = (T_1, T_2, \ldots, T_m) \) corresponds to the channel input sequence, and where \( Z_m = (Z_1, Z_2, \ldots, Z_m) \) and \( Z'_m = (Z'_1, Z'_2, \ldots, Z'_m) \) are Bob and Eve’s observations, respectively. \( \xi_i \)s are i.i.d. and each \( \xi_i \) is drawn from a normal distribution denoted by \( g \) with zero-mean and variance \( \sigma^2 \). The channel input fulfills the following power constraint:

\[ \mathbb{E}[|T_i|^2] \leq P \quad i = 1 \ldots m. \]  

(15)

The input set is \( T_{m,P} \), defined in (6). The output sets are infinite \( Z = Z' = \mathbb{C} \). We denote the GWC by the pair \((W_g, V_{g'})\), where \( W_g \) and \( V_{g'} \) define the Gaussian channels to the legitimate receiver and the wiretapper, with capacities \( C(g, P) \) and \( C(g', P) \), respectively.

**Definition 6.** A CR-assisted \((m, N, \lambda_1, \lambda_2)\) identification code for the GWC \((W_g, V_{g'})\) is a family of pairs \( \{(u_i, D_i(y^n)), \quad i = 1, \ldots, N\} \), with

\[ u_i = \Phi(x^n) \in T_{m,P}, \quad D_i(y^n) \subset Z^m, \quad \forall i \in \{1, \ldots, N\}, \]

such that for all \( i, j \in \{1, \ldots, N\}, i \neq j \) and some \( E = E(x^n, y^n) \in Z^m \) and \( \lambda_1, \lambda_2 \leq \frac{1}{2} \), the errors of first and second kind satisfy, respectively:

\[ W^m_g((D_i(y^n))^{c}|u_i) \leq \lambda_1, \]  

(16)

\[ W^m_g(D_i(y^n)|u_j) \leq \lambda_2, \]  

(17)

and for \( \lambda \leq \frac{1}{2} \), it holds that

\[ V^m_{g'}(E|u_j) + V^m_{g'}(E^{c}|u_i) \geq 1 - \lambda. \]  

(18)
Remark 3. The last line means that the wiretapper cannot identify the identification message $i$.

Definition 7. $C_{SID}^s(g, g', P)$, the secure CR-assisted identification capacity of the channel $(W_g, V_{g'})$, is defined as follows:

$$C_{SID}^s(g, g', P) = \max \left\{ R : \forall \lambda > 0, \exists n(\lambda) \text{ s.t. for } n \geq n(\lambda) N_S(n, \lambda) \geq 2^{2\pi n} \right\},$$

where $N_S(n, \lambda)$ is the maximal cardinality such that a $(n, N, \lambda_1, \lambda_2)$ CR-assisted identification wiretap code for the channel $(W_g, V_{g'})$ exists.

Remark 4. As correlation cannot increase the Shannon message-transmission capacity, it is not utilized in current communication systems. However, this is not the case for identification. We will demonstrate in Section 4 that for the identification task, we can achieve performance gains by taking advantage of CR.

3 Common Randomness Capacity

In this section, we propose a single-letter characterization of the CR capacity for the scenarios presented in the previous section and provide a rigorous proof of it.

3.1 SISO Case

We start with the first scenario depicted in Fig. 1 where the communication is over a SISO Gaussian channel with the power constraint defined in (5).

Proposition 1. For the model in Fig. 1, the CR capacity $C_{CR}(P)$ is equal to

$$C_{CR}(P) = \max_{U \in X \oplus Y} I(U; X). \quad (19)$$

Direct Proof: We extend the coding scheme provided in [1] to Gaussian channels. By continuity, it suffices to show that

$$\max_{U \in X \oplus Y} I(U; X)$$

is an achievable CR rate for every $R' < C(P)$. Let $U$ be a random variable satisfying $U \oplus X \oplus Y$ and $I(U; X) - I(U; Y) \leq R'$.

We are going to show that $H = I(U; X)$ is an achievable CR rate.
assume that the distribution of $U$ is a possible type for block-length $n$. For some
\[ \mu > 0, \]
we let
\[ N_1 = \left[ 2^{n[I(U;X) - I(U;Y) + 3\mu]} \right] \]
and
\[ N_2 = \left[ 2^{n[I(U;Y) - 2\mu]} \right]. \]

For each pair $(i,j)$ with $1 \leq i \leq N_1$ and $1 \leq j \leq N_2$, we define a random
sequence $U_{i,j} \in U^n$ of type $P_Y$. Let $M = U_{1,1}, \ldots, U_{N_1,N_2}$ be the joint random variable of all $U_{i,j}$s. We define $\Phi_M$ as follows: Let $\Phi_M(X^n) = U_{i,j}$, if $U_{i,j}$ is jointly $UX$-typical with $X^n$ (either one if there are several). If no such $U_{i,j}$ exists, then $\Phi_M(X^n)$ is set to a constant sequence $u_0$ different from all the $U_{i,j}$s, jointly $UX$-typical with none of the realizations of $X^n$ and known to both terminals.

We further define the following two sets which depend on $M$:
\[ S_1(M) = \{ (x^n, y^n) : (\Phi_M(x^n), x^n, y^n) \in T^n_{U,X,Y} \} \]
and
\[ S_2(M) = \{ (x^n, y^n) : (x^n, y^n) \in S_1(M) \text{ s.t. } U_{i,j} = \Phi_M(x^n) \text{ and } \exists U_{i,\ell} \neq U_{i,j} \text{ jointly } UY \text{-typical with } y^n \text{ (with the same first index } i) \}. \]

It is proved in [1] that
\[ E_M [P \{ (X^n, Y^n) \notin S_1(M) \} + P \{ (X^n, Y^n) \in S_2(M) \}] \leq \beta(n), \quad (20) \]
where $\beta(n) \leq \frac{\alpha}{4}$ for sufficiently large $n$. We choose a realization
\[ m = u_{1,1}, \ldots, u_{N_1,N_2} \]
satisfying
\[ P \{ (X^n, Y^n) \notin S_1(m) \} + P \{ (X^n, Y^n) \in S_2(m) \} \leq 2\beta(n). \quad (21) \]

From (20) and using Markov inequality, we know that such a realization exists. We denote $\Phi_m$ by $\Phi$. We assume that each $u_{i,j}, i = 1 \ldots N_1, j = 1 \ldots N_2$, is known to both terminals. This means that $N_1$ codebooks $C_i, 1 \leq i \leq N_1$, are known to both terminals, where each codebook contains $N_2$ sequences, $u_{i,j}, j = 1, \ldots, N_2$.

Let $x^n$ be any realization of $X^n$ and $y^n$ be any realization of $Y^n$. Let $f_1(x^n) = i$ if $\Phi(x^n) = u_{i,j}$. Otherwise, if $\Phi(x^n) = u_0$, then $f_1(x^n) = N_1 + 1$. Since $R' < C(P)$, we choose $\mu$ to be sufficiently small such that
\[ \frac{\log ||f_1||}{n} = \frac{\log(N_1 + 1)}{n} \leq C(P) - \mu', \quad (22) \]
for some \( \mu' > 0 \), the message \( i^* = f_1(x^n) \), with \( i^* \in \{1, \ldots, N_1 + 1\} \), is encoded to a sequence \( t^n \) using a code sequence \( (f_n^*)_{n=1}^{\infty} \), using a suitable \textit{forward error correcting code}, with rate \( \frac{\log \|f_n^*\|}{n} = \frac{\log \|f_1\|}{n} \) satisfying \((22)\) and with maximum error probability not exceeding \( \frac{\alpha}{2} \) for sufficiently large \( n \). Here, \( \|f_1\| \) refers to the cardinality of the set of messages \( \{i^* : i^* = 1, \ldots, N_1 + 1\} \). The sequence \( t^n \) is sent over the Gaussian channel. Let \( z^n \) be the corresponding channel output sequence. Terminal \( B \) decodes the message \( \tilde{i}^* \) from the knowledge of \( z^n \). Let \( \Psi(y^n, z^n) = u_{\tilde{i}^*,j} \) if \( u_{\tilde{i}^*,j} \) and \( y^n \) are jointly \( UY \)-typical. If there is no such \( u_{\tilde{i}^*,j} \) or there are several, we set \( \Psi(y^n, z^n) = u_0 \) (since \( K \) and \( L \) must have the same alphabet). Now, we are going to show that the requirements in (3) and (4) are satisfied. We define next for any \((i, j) \in \{1, \ldots, N_1\} \times \{1, \ldots, N_2\}\) the set \( S = \{x^n \in X^n \text{ s.t. } (u_{i,j}, x^n) \text{ jointly } UX \text{-typical}\} \).

Then, it holds that

\[
\begin{align*}
\mathbb{P}[K = u_{i,j}] &= \sum_{x^n \in S} \mathbb{P}[K = u_{i,j}|X^n = x^n] P_X^n(x^n) \\
&\quad + \sum_{x^n \in S^c} \mathbb{P}[K = u_{i,j}|X^n = x^n] P_X^n(x^n) \\
&\overset{(a)}{=} \sum_{x^n \in S} \mathbb{P}[K = u_{i,j}|X^n = x^n] P_X^n(x^n) \\
&\leq \sum_{x^n \in S} P_X^n(x^n) \\
&= P_X^n(\{x^n : (u_{i,j}, x^n) \text{ jointly } UX \text{-typical}\}) \\
&= 2^{-n I(U;X) - \kappa(n)},
\end{align*}
\]

for some \( \kappa(n) > 0 \) with \( \lim_{n \to \infty} \frac{\kappa(n)}{n} = 0 \), where \((a)\) follows because for \((u_{i,j}, x)\) being not jointly \( UX \)-typical, we have \( \mathbb{P}[K = u_{i,j}|X^n = x^n] = 0 \). This yields \( H(K) \geq n I(U;X) - \kappa'(n) \) for some \( \kappa'(n) > 0 \) with \( \lim_{n \to \infty} \frac{\kappa'(n)}{n} = 0 \). Therefore, for sufficiently large \( n \), it holds that

\[
\frac{H(K)}{n} > H - \delta.
\]

Thus, \((3)\) is satisfied. Now, it remains to prove that \((4)\) is satisfied. For this purpose, we define the following event:

\[ D_m = \text{"} \Phi(X^n) \text{ is equal to none of the } u_{i,j}'s \text{"}. \]
We denote its complement by $D^c_m$. We further define $I^* = f_1(X^n)$ to be the random message generated by Terminal $A$ and $\tilde{I}^*$ to be the random message decoded by Terminal $B$. We have

\[
\mathbb{P}[K \neq L] = \mathbb{P}[K \neq L | I^* = \tilde{I}^*] \mathbb{P}[I^* = \tilde{I}^*] \\
+ \mathbb{P}[K \neq L | I^* \neq \tilde{I}^*] \mathbb{P}[I^* \neq \tilde{I}^*] \\
\leq \mathbb{P}[K \neq L | I^* = \tilde{I}^*] + \mathbb{P}[I^* \neq \tilde{I}^*].
\]

Here,

\[
\mathbb{P}[K \neq L | I^* = \tilde{I}^*] = \mathbb{P}[K \neq L | I^* = \tilde{I}^*, D^c_m] \mathbb{P}[D^c_m | I^* = \tilde{I}^*] \\
+ \mathbb{P}[K \neq L | I^* \neq \tilde{I}^*, D^c_m] \mathbb{P}[D^c_m | I^* = \tilde{I}^*] \\
\overset{(a)}{=} \mathbb{P}[K \neq L | I^* = \tilde{I}^*, D^c_m] \mathbb{P}[D^c_m | I^* = \tilde{I}^*] \\
\leq \mathbb{P}[K \neq L | I^* = \tilde{I}^*]
\]

where (a) follows from $\mathbb{P}[K \neq L | I^* = \tilde{I}^*, D^c_m] = 0$, since conditioned on $I^* = \tilde{I}^*$ and $D^c_m$, we know that $K$ and $L$ are both equal to $u_0$. Thus, we obtain

\[
\mathbb{P}[K \neq L] \leq \mathbb{P}[K \neq L | I^* = \tilde{I}^*, D^c_m] + \mathbb{P}[I^* \neq \tilde{I}^*] \\
\leq \mathbb{P}[A(m) \cup B(m)] + \mathbb{P}[I^* \neq \tilde{I}^*] \\
\overset{(a)}{=} \mathbb{P}[A(m)] + \mathbb{P}[B(m)] + \mathbb{P}[I^* \neq \tilde{I}^*] \\
\overset{(b)}{=} 2\beta(n) + \mathbb{P}[I^* \neq \tilde{I}^*] \\
\leq \alpha,
\]

where (a) follows because the events $A(m)$ and $B(m)$ are independent, (b) follows from (21) and (c) follows because $2\beta(n) + \theta \leq \alpha$ for sufficiently large $n$. This proves (3). This completes the achievability proof.

**Converse Proof:** Let $H$ be any achievable CR rate. So, for every $\alpha, \delta > 0$ and for sufficiently large $n$, there exists a permissible pair of random variables $(K, L)$ according to a fixed CR-generation protocol of block-length $n$ such that

\[
\mathbb{P}[K \neq L] \leq \alpha,
\]

and

\[
\frac{1}{n} H(K) > H - \delta.
\]

In our proof, we will use the following lemma:
Lemma 1. (Lemma 17.12 in [30]) For arbitrary random variables $R_1$ and $R_2$ and sequences of random variables $X^n$ and $Y^n$, it holds that

$$I(R_1; X^n | R_2) - I(R_1; Y^n | R_2)$$

$$= \sum_{i=1}^{n} I(R_1; X_i | X_1, \ldots, X_{i-1}, Y_{i+1}, \ldots, Y_n, R_2)$$

$$- \sum_{i=1}^{n} I(R_1; Y_i | X_1, \ldots, X_{i-1}, Y_{i+1}, \ldots, Y_n, R_2)$$

$$= n[I(R_1; X_J | V) - I(R_1; Y_J | V)],$$

where $V = (X_1, \ldots, X_{J-1}, Y_{J+1}, \ldots, Y_n, R_2, J)$, with $J$ being a random variable independent of $R_1$, $R_2$, $X^n$ and $Y^n$ and uniformly distributed on $\{1, \ldots, n\}$.

Let $J$ be a random variable uniformly distributed on $\{1, \ldots, n\}$ and independent of $K$, $X^n$ and $Y^n$. We further define $U = (K, X_1, \ldots, X_{J-1}, Y_{J+1}, \ldots, Y_n, J)$. It holds that $U \perp X_J \perp Y_J$. Notice that

$$\frac{1}{n} H(K) \overset{(a)}{=} \frac{1}{n} H(K) - \frac{1}{n} H(K|X^n)$$

$$= \frac{1}{n} I(K; X^n)$$

$$\overset{(b)}{=} \frac{1}{n} \sum_{i=1}^{n} I(K; X_i | X_1, \ldots, X_{i-1})$$

$$= I(K; X_J | X_1, \ldots, X_{J-1}, J)$$

$$\overset{(c)}{\leq} I(U; X_J),$$

where (a) follows because $K = \Phi(X^n)$ and (b) and (c) follow from the chain rule for mutual information. Applying Lemma 1 for $R_1 = K$, $R_2 = \emptyset$ with $V = (X_1, \ldots, X_{J-1}, Y_{J+1}, \ldots, Y_n, J)$ yields

$$\frac{1}{n} [I(K; X^n) - I(K; Y^n)]$$

$$= I(K; X_J | V) - I(K; Y_J | V)$$

$$\overset{(a)}{=} I(KV; X_J) - I(V; X_J) - I(KV; Y_J) + I(V; Y_J)$$

$$\overset{(b)}{=} I(U; X_J) - I(U; Y_J),$$

where (a) follows from the chain rule for mutual information and from the fact that $V$ is independent of $(X_J, Y_J)$ and (b) follows from $U = (K, V)$. It results
using (25) that

\[ I(U; X_j) - I(U; Y_j) = \frac{1}{n} [I(K; X^n) - I(K; Y^n)] \]

\[ = \frac{1}{n} H(K) - \frac{1}{n} I(K; Y^n) \]

\[ = \frac{1}{n} H(K | Y^n) \]

\[ = \frac{1}{n} H(K | Y^n, Z^n) + \frac{1}{n} I(K; Z^n | Y^n) \]

\[ \leq \frac{1}{n} H(K | L) + \frac{1}{n} I(K; Z^n | Y^n) \]

\[ (a) \quad \frac{1}{n} (1 + \log |K| \mathbb{P}[K \neq L]) + \frac{1}{n} I(K; Z^n | Y^n) \]

\[ (b) \quad \frac{\alpha}{n} \log |X| + \frac{1}{n} I(K; Z^n | Y^n). \]

where (a) follows from Fano’s inequality and (b) follows from (23) and from the fact |K| ≤ |X|^n. On the one hand, we have

\[ \frac{1}{n} I(K; Z^n | Y^n) \leq \frac{1}{n} I(X^n, K; Z^n | Y^n) \]

\[ \leq \frac{1}{n} I(T^n; Z^n | Y^n) \]

\[ = \frac{1}{n} h(Z^n | Y^n) + \frac{1}{n} h(Z^n | T^n, Y^n) \]

\[ \leq \frac{1}{n} h(Z^n | Y^n) + \frac{1}{n} I(Z^n; T^n) \]

\[ \leq \frac{1}{n} h(Z^n) + \frac{1}{n} I(Z^n; T^n) \]

\[ = \frac{1}{n} I(T^n; Z^n), \]

where (a) follows from the Data Processing Inequality because \( Y^n \triangleleft X^n K \triangleleft T^n \triangleleft Z^n \) forms a Markov chain, (b) follows because \( Y^n \triangleleft X^n K \triangleleft T^n \triangleleft Z^n \) forms a Markov chain, (c) follows because conditioning does not increase entropy,
On the other hand, we have:

\[
\frac{1}{n} I(T^n; Z^n) \\
\overset{(a)}{=} \frac{1}{n} \sum_{i=1}^{n} I(Z_i; T^n | Z^{i-1}) \\
= \frac{1}{n} \sum_{i=1}^{n} h(Z_i | Z^{i-1}) - h(Z_i | T^n, Z^{i-1}) \\
\overset{(b)}{=} \frac{1}{n} \sum_{i=1}^{n} h(Z_i | Z^{i-1}) - h(Z_i | T_i) \\
\overset{(c)}{\leq} \frac{1}{n} \sum_{i=1}^{n} h(Z_i) - h(Z_i | T_i) \\
= \frac{1}{n} \sum_{i=1}^{n} I(T_i; Z_i) \\
\overset{(d)}{\leq} C(P),
\]

where (a) follows from the chain rule for mutual information, (b) follows because

\[T_1, \ldots, T_{i-1}, T_{i+1}, \ldots, T_n, Z^{i-1} \circ T_i \circ Z_i \]

forms a Markov chain, (c) follows because conditioning does not increase entropy

and (d) follows from (7) and from the fact that \(T_i, i = 1, \ldots, n\) satisfies the power

constraint in (8). Thus, we obtain

\[I(U; X_J) - I(U; Y_J) \leq C(P) + \zeta(n, \alpha),\]

where \(\zeta(n, \alpha) = \frac{1}{n} + \alpha \log |\mathcal{X}|\). Since the joint distribution of \(X_J\) and \(Y_J\)

is equal to \(P_{X_Y}\), it follows that \(\frac{H(X)}{n}\) is upper-bounded by \(I(U; X)\) subject to

\(I(U; X) - I(U; Y) \leq C(P) + \zeta(n, \alpha)\) with \(U\) satisfying \(U \circ X \circ Y\). As a result,

it follows using (24) that any achievable CR rate satisfies

\[H \leq \max_{U \circ X \circ Y} I(U; X) + \delta.\]  \hspace{1cm} (26)

By taking the limit when \(n\) tends to infinity and then the infimum over all \(\alpha > 0, \delta > 0\), of the right-hand side of (26), it follows that

\[H \leq \max_{U \circ X \circ Y} I(U; X).\]

This completes the converse proof.
Remark 5. There exists $P_\star$ such that $C(P_\star) = H(X|Y)$, where

$$C_{CR}(P) = C_{CR}(P_\star) = H(X) \quad \forall P \geq P_\star.$$ 

Example 1. Consider the example of binary sources such that $|\mathcal{X}| = |\mathcal{Y}| = 2$ with $P_X(0) = P_X(1) = \frac{1}{2}$. We consider the following transition probability

$$P_{Y|X}(y|0) = \frac{1 - \mu}{\mu} \quad 0 \leq \mu \leq \frac{1}{2}$$

$$P_{Y|X}(y|1) = \frac{\mu}{1 - \mu} \quad 0 \leq \mu \leq \frac{1}{2}$$

with

$$P_{XY}(x, y) = P_{Y|X}(y|x)P_X(x) \quad (x, y) \in \{0, 1\}^2.$$ 

In this case, it holds that

$$C_{CR}(P_\star) = C_{CR}(P_\star, P_{XY}) = 1$$

and that

$$\frac{1}{2} \log(1 + \frac{P_\star}{\sigma^2}) = H(X|Y)$$

$$= (1 - \mu) \log(\frac{1}{1 - \mu}) + \mu \log(\frac{1}{\mu}).$$

We define

$$f(\mu) = (1 - \mu) \log(\frac{1}{1 - \mu}) + \mu \log(\frac{1}{\mu}).$$

As a result, $P_\star$ is chosen such that:

$$P_\star = \sigma^2(2^{2f(\mu)} - 1) \quad 0 \leq \mu \leq \frac{1}{2}.$$ 

In Fig. 5 the channel power $P_\star$ is plotted as a function of the parameter $\mu$, with a fixed noise variance of $\sigma^2 = 1$. As $\mu$ increases, the correlation between the binary sources decreases. Consequently, the optimal power $P_\star$, starting at which the common randomness capacity is the highest possible, also increases, as depicted in Fig. 5.
Fig. 5: Channel power $P_\star$ in function of the parameter $\mu$, for a noise variance $\sigma^2 = 1$.

**Optimization Problem:** In this section, we solve the constrained optimization problem presented in Proposition 1. We consider the same sources and transition probability as in Example 1 ($P_X$ and $P_{Y|X}$ are given as in Example 1). Assume that the random variable $U$ has alphabet $\mathcal{U}$, then by applying the Support Lemma [30], it holds that the cardinality of the set $\mathcal{U}$ satisfies the following constraint [1]

$$|U| \leq |\mathcal{X}| + 1.$$

For $U \circ X \circ Y$, it holds that

$$I(U; X) - I(U; Y) = I(U; X|Y),$$

where

$$I(U; X|Y) \leq H(X|Y).$$

We can write the optimization problem in (19) as follows:

$$C_{CR}(P) = \max_{U \circ X \circ Y} \min \left\{ I(U; X) \mid I(U; X|Y) \leq \min\{C(P), H(X|Y)\} \right\}.$$

$$|U| \leq |\mathcal{X}| + 1.$$
Let $\mathcal{U} = \{u_1, u_2, u_3\}$. We define $\theta$ as follows:

$$
\theta = \begin{pmatrix}
P_{UX}(u_1, 0) \\
P_{UX}(u_2, 0) \\
P_{UX}(u_3, 0) \\
P_{UX}(u_1, 1) \\
P_{UX}(u_2, 1) \\
P_{UX}(u_3, 1)
\end{pmatrix},
$$

with

$$
\begin{pmatrix}
1 & 1 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 1 & 1
\end{pmatrix} \theta = \begin{pmatrix}
P_X(0) \\
P_X(1)
\end{pmatrix}.
$$

We obtain the following equivalent constrained optimization problem:

$$
\min_{\theta \in \Theta: \forall i \in \{1, \ldots, 5\}, g_i(\theta) \leq 0} g_0(\theta),
$$

where $\Theta = \{\theta: \theta \geq 0 \text{ and } 1^T \theta = 1\}$ with $1 = (1, 1, 1, 1, 1, 1)$.

The objective function is

$$
g_0(\theta) = -I(U; X).
$$

In addition, the constraint functions are expressed as follows:

$$
g_1(\theta) = I(U; X|Y) - \min\{C(P), H(X|Y)\}
$$

$$
g_2(\theta) = \sum_{u \in \mathcal{U}} P_{UX}(u, 0) - P_X(0)
$$

$$
g_3(\theta) = -\sum_{u \in \mathcal{U}} P_{UX}(u, 0) + P_X(0)
$$

$$
g_4(\theta) = \sum_{u \in \mathcal{U}} P_{UX}(u, 1) - P_X(1)
$$

$$
g_5(\theta) = -\sum_{u \in \mathcal{U}} P_{UX}(u, 1) + P_X(1).
$$

**Remark 6.** The optimization problem is non-convex since the objective function $g_0(\theta)$ is non-convex. The non-convexity of $g_0(\theta)$ is shown in the appendix.

To solve the optimization problem, we define the Lagrangian function $\mathcal{L} : \Theta \times \Lambda \rightarrow \mathbb{R}$

$$
\mathcal{L}(\theta, \lambda) = g_0(\theta) + \sum_{i=1}^{5} \lambda_i g_i(\theta),
$$

(27)

where $\Lambda = \{\lambda = (\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5)^T \in \mathbb{R}^5 : \lambda \geq 0\}$. 
Optimizing the Lagrangian is interpreted as playing a two-player zero-sum game: the first player chooses $\theta$ that minimizes $L(\theta, \lambda)$ and the second player chooses $\lambda$ that maximizes it. A pure Nash equilibrium might in general not exist. However, a mixed Nash Equilibrium does exist \[31\]. In what follows, the relationship between an approximate mixed Nash equilibrium of the Lagrangian game and a nearly-optimal nearly-feasible solution to (27) is characterized.

**Theorem 1.** \[31\] Let $\theta^{(1)} \ldots \theta^{(T)} \in \Theta$ and $\lambda^{(1)} \ldots \lambda^{(T)} \in \Lambda$ be sequences of vectors that satisfy an approximate mixed Nash equilibrium, i.e.

$$
\max_{\lambda \in \Lambda} \frac{1}{T} \sum_{t=1}^{T} L\left(\theta^{(t)}, \lambda^{(t)}\right) - \inf_{\theta^* \in \Theta} \frac{1}{T} \sum_{t=1}^{T} L\left(\theta^*, \lambda^{(t)}\right) \leq \epsilon.
$$

Define $\bar{\theta}$ such that $\bar{\theta} = \theta^{(t)}$ with probability $\frac{1}{T}$. Then it holds that $\bar{\theta}$ is nearly-optimal in expectation, i.e.,

$$
E_{\theta} \left[ g_0(\bar{\theta}) \right] \leq \inf_{\theta^* \in \Theta : g_i(\theta^*) \leq 0} g_0(\theta^*) + \epsilon.
$$

The following algorithm optimizes the Lagrangian function in (27) in the non-convex setting.

**Algorithm 1** Lagrangian-formulation Optimization in the non-convex setting \[31\]

1. Initialize $\lambda^{(1)} = 0$
2. for $t \in [T]$
   (a) $\theta^{(t)} = O_\rho \left( L\left( \cdot, \lambda^{(t)} \right) \right)$
   (b) $\Delta^{(t)}_\lambda$ gradient of $L\left( \theta^{(t)}, \lambda^{(t)} \right)$ w.r.t to $\lambda$
   (c) Update $\lambda^{(t+1)} = \Pi_{\Lambda} \left( \lambda^{(t)} + \frac{\eta_\lambda}{\sqrt{\mathbb{E}_{\lambda, i} \Delta^{(t)}_\lambda}} \Delta^{(t)}_\lambda \right)$
end
3. Return $\theta^{(1)} \ldots \theta^{(T)} \quad \lambda^{(1)} \ldots \lambda^{(T)}$

The step a) consists of computing the $\rho$-approximate Bayesian optimization oracle which is defined as follows:

**Definition 8.** \[31\] A $\rho$- approximate Bayesian optimization oracle is a function $O_\rho : (\Theta \rightarrow \mathbb{R}) \rightarrow \Theta$ for which:

$$
f(O_\rho(f)) \leq \inf_{\theta^* \in \Theta} f(\theta^*) + \rho.
$$

The gradient in step b) is expressed as follows:

$$
\Delta_\lambda = (g_1(\theta), g_2(\theta), g_3(\theta), g_4(\theta), g_5(\theta))^T.
$$
In step c), we perform first an AdaGrad \cite{32} update, where $\eta_\lambda$ stands for the initial learning rate and $\tau_\lambda$ is a smoothing term. In addition, $G_{\lambda,t}$ is a diagonal matrix that contains the sum of the squares of the past gradients with respect to all parameters $\lambda$ along its diagonal. Second, we perform a projection onto $\Lambda$ such that:

$$\Pi_{\Lambda}(z) = \max(0, z).$$

We obtain $T$ candidate solutions $\theta^{(1)} \ldots \theta^{(T)}$. The goal is to yield a uniform distribution over these $T$ candidates that is approximately feasible according to Theorem 1.

**Remark 7.** The approach proposed above is idealized \cite{31}. In practice, we opt for the typical approach: pretending that our problem is convex and using a first-order stochastic algorithm such as AdaGrad. This is illustrated in Algorithm 2.

**Algorithm 2** Lagrangian-formulation Optimization in the convex setting \cite{31}

1. Initialize $\theta^{(1)} \in \Theta$, $\lambda^{(1)} = 0$
2. for $t \in [T]$
   (a) $\Delta_{\theta}^{(t)}$ sub-gradient of $L(\theta^{(t)}, \lambda^{(t)})$ w.r.t to $\theta$
   (b) $\Delta_{\lambda}^{(t)}$ gradient of $L(\theta^{(t)}, \lambda^{(t)})$ w.r.t to $\lambda$
   (c) Update $\theta^{(t+1)} = \Pi_{\Theta}(\theta^{(t)} - \frac{\eta_\theta}{\sqrt{G_{\theta,t} + \tau_\theta}} \Delta_{\theta}^{(t)})$
   (d) Update $\lambda^{(t+1)} = \Pi_{\Lambda}(\lambda^{(t)} + \frac{\eta_\lambda}{\sqrt{G_{\lambda,t} + \tau_\lambda}} \Delta_{\lambda}^{(t)})$
end
3. Return $\theta^{(1)} \ldots \theta^{(T)}$, $\lambda^{(1)} \ldots \lambda^{(T)}$

In step a), we compute the sub-gradient $\Delta_{\theta}$, where it holds that

$$\forall u \in U :$$

$$\frac{\partial L(\theta, \lambda)}{\partial P_{UX}(u, 0)} = -\log \left( \frac{P_{UX}(u, 0)}{\sum_{x' \in \mathcal{X}} P_{UX}(u, x')} \right)$$

$$+ \lambda_1 \sum_{y \in \mathcal{Y}} P_{Y|X}(y|0) \log \left( \frac{P_{Y|X}(y|0) P_{UX}(u, 0)}{\sum_{x' \in \mathcal{X}} P_{Y|X}(y|x') P_{UX}(u, x')} \right)$$

$$+ \lambda_2 - \lambda_3$$
\[
\frac{\partial L (\theta, \lambda)}{\partial P_{UX} (u, 1)} = -\log \left( \frac{P_{UX} (u, 1)}{\sum_{x' \in X} P_{UX} (u, x')} \right) + \lambda_1 \sum_{y \in Y} P_{Y|X} (y|1) \log \left( \frac{P_{Y|X} (y|1) P_{UX} (u, 1)}{\sum_{x' \in X} P_{Y|X} (y|x) P_{UX} (u, x')} \right) + \lambda_4 - \lambda_5.
\]

It is worth-mentioning here that \( \frac{\partial g_0 (\theta)}{\partial P_{UX} (u,x)} \) is computed in (29) and that \( \frac{\partial g_1 (\theta)}{\partial P_{UX} (u,x)} \) is computed for \( U \circ - X \circ - Y \) in (34) (see Appendix).

In step c) and step d), AdaGrad updates are performed, where \( \eta_\lambda \) and \( \eta_\theta \) correspond to the initial learning rates and \( \tau_\lambda \) and \( \tau_\theta \) are smoothing terms. Furthermore, \( G_{\lambda,t} \) and \( G_{\theta,t} \) are diagonal matrices that contain the sum of the squares of the past gradients with respect to all parameters \( \lambda \) and \( \theta \) respectively, along their diagonal.

The projection \( \Pi_{\Theta} (z) \) onto \( \Theta \) in step c) corresponds to the Euclidean projection onto the probability simplex which is computed using the following algorithm:

**Algorithm 3** Euclidean projection of a vector onto the probability simplex

1. **Input:** \( z \in \mathbb{R}^D \)
   (a) Sort \( z \) into \( w : w_1 \geq w_2 \geq \ldots \geq w_D \)
   (b) Find \( \gamma = \max \left( 1 \leq j \leq D : w_j + 1 - \frac{\sum_{i=1}^{D} w_i}{j} \right) \)
   (c) Define \( \kappa = \frac{1}{2} (1 - \sum_{i=1}^{D} w_i) \)
2. **Output:** \( \theta \) s.t \( \theta_i = \max \{ z_i + \kappa, 0 \} \) \( i = 1 \ldots D \)

**Simulation Results:** In this section, we present our numerical results. We study the CR capacity for different channel input powers as well as for different values of the parameter \( \mu \). We fix \( T = 5000 \). Algorithm [2] is implemented for given \( P \) and given \( \mu \) and for different values of the initial learning rates \( \eta_\theta \) and \( \eta_\lambda \). At the end, we consider the pair \( (\eta_\theta, \eta_\lambda) \) for which \( \theta^{(1)} \ldots \theta^{(T)} \in \Theta \) and \( \lambda^{(1)} \ldots \lambda^{(T)} \in \Lambda \) yield the smallest \( \epsilon \) in Theorem 1. We vary first the parameter \( \mu \) in \( \{0, 0.1, 0.2, 0.3, 0.4, 0.5\} \) and plot for each \( \mu \) the common randomness capacity as a function of the power, as depicted in Fig. 6. Next, we generate a three-dimensional plot of the common randomness capacity as a function of both the
Optimal Signal Processing for CR Generation over Gaussian Channels

Fig. 6: CR capacity in function of the power for a noise variance $\sigma^2 = 1$ and for different values of $\mu$.

Fig. 7: CR capacity in function of the power and the parameter $\mu$ for a noise variance $\sigma^2 = 1$.

power and the parameter $\mu$. This is illustrated in Fig. 7. We consider first the specific case when $\mu = 0$. Clearly, for $\mu = 0$, it holds that $Y = X$. Therefore, no communication over the channel is required to achieve the maximal amount
of common randomness equal to $H(X)$. This is numerically verified in Fig. 6 and Fig. 7 where for $\mu = 0$ and $P = 0$ the common randomness capacity is equal to $H(X) = 1$. However, for $0 < \mu \leq \frac{1}{2}$, a communication over the channel is necessary to generate common randomness between the two terminals ($C_{CR}(0) = 0$ for $0 < \mu \leq \frac{1}{2}$). This is due to the fact that $X$ and $Y$ have an indecomposable joint distribution [1] for $0 < \mu \leq \frac{1}{2}$. Furthermore, the higher the parameter $\mu$ is, the less correlated the sources are. As a result, we need to investigate more power in order to achieve the same amount of common randomness obtained for lower values of $\mu$. This is clearly observable in Fig. 6. For instance, for $\mu = 0.2$, the common randomness capacity is equal to 0.86 for $P \approx 1.33$, whereas, for $\mu = 0.5$, the same amount of common randomness is achieved for $P \approx 2.34$.

3.2 MIMO Case

We now focus on the second scenario depicted in Fig. 2 where the communication is over a MIMO Gaussian channel with the power constraint defined in (8).

**Proposition 2.** For the model in Fig. 2, the CR capacity $C_{CR}(P, N_T \times N_R)$ is equal:

$$C_{CR}(P, N_T \times N_R) = \max_{U \circ X \circ Y \atop I(U;X) - I(U;Y) \leq C(P,N_T \times N_R)} I(U; X).$$

![Diagram](image)

**Fig. 8:** Decomposition of the MIMO channel into $N_{\text{min}}$ parallel channels through SVD.

**Direct Proof**

**Proof.** The capacity $C(P, N_T \times N_R)$ can be computed by converting the MIMO channel into parallel, independent and scalar Gaussian sub-channels. This conversion is based on the following singular value decomposition (SVD) of the
channel matrix $H$:

$$H = U \Lambda V^H,$$

where $U \in \mathbb{C}^{N_R \times N_R}$ and $V \in \mathbb{C}^{N_T \times N_T}$ are unitary matrices. $\Lambda \in \mathbb{C}^{N_R \times N_T}$ is a diagonal matrix, whose diagonal elements $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{N_{\min}}$ are the ordered singular values of the channel matrix $H$. We denote with $N_{\min}$ the rank of $H$, $N_{\min} := \min(N_T, N_R)$. If we multiply (10) with the unitary matrix $U^H$, we then obtain

$$U^H Z = U^H U \Lambda V^H T + U^H \xi.$$

It can easily be checked that $\tilde{\xi}$ has the same distribution as $\xi$, i.e., $\tilde{\xi} \sim \mathcal{N}(0, \sigma^2 I_{N_R})$, and we have

$$\mathbb{E}[T^H T] = \mathbb{E}[T^H V V^H T] = \mathbb{E}[T^H T].$$

We obtain the $N_{\min}$ independent scalar Gaussian channels depicted in Fig. 8

$$\tilde{Z}_\ell = \lambda_\ell \tilde{T}_\ell + \tilde{\xi}_\ell \quad \ell = 1 \ldots N_{\min}.$$ 

The SVD can be interpreted as a pre-processing (multiplication with $V$) and a post-processing (multiplication with $U^H$). The optimization problem in (11) is reduced to

$$C(P, N_T \times N_R) = \max_{\tilde{P}_1 \ldots \tilde{P}_{N_{\min}}} \sum_{\ell=1}^{N_{\min}} \log \left( 1 + \frac{\lambda_\ell^2}{\sigma^2} \tilde{P}_\ell \right),$$

s.t. $\sum_{\ell=1}^{N_{\min}} \tilde{P}_\ell \leq P$ and $\tilde{P}_\ell \geq 0 \quad \ell = 1 \ldots N_{\min}$.

It holds that

$$C(P, N_T \times N_R) = \sum_{\ell=1}^{N_{\min}} C(\tilde{P}_\ell),$$

where the capacity of each sub-channel is expressed as

$$C(\tilde{P}_\ell) = \log \left( 1 + \frac{\lambda_\ell^2}{\sigma^2} \tilde{P}_\ell \right) \quad \ell = 1 \ldots N_{\min}.$$ 

The power $\tilde{P}_\ell$ is called the waterfilling rule and it is expressed as follows:

$$\tilde{P}_\ell = \max \left( 0, \kappa - \frac{\sigma^2}{\lambda_\ell} \right) \quad \ell = 1 \ldots N_{\min},$$

where $\kappa$ is the waterfilling level.
By Proposition 1, we have for $\ell = 1 \ldots N_{\min}$
\[
C_{CR}(\hat{P}_\ell) = \max_{U_\ell \in X \times Y} I(U_\ell; X) \quad \text{s.t.} \quad I(U_\ell; X) - I(U_\ell; Y) \leq C(\hat{P}_\ell)
\]

Since we may lose information through processing, it holds that
\[
C_{CR}(P, N_T \times N_R) \geq \sum_{\ell=1}^{N_{\min}} C_{CR}(\hat{P}_\ell)
\]

\[
= \sum_{\ell=1}^{N_{\min}} \max_{U_\ell \in X \times Y} I(U_\ell; X) \quad \text{s.t.} \quad I(U_\ell; X) - I(U_\ell; Y) \leq C(\hat{P}_\ell)
\]

\[
= \max_{U_1 \ldots U_{N_{\min}}} \max_{\ell=1 \ldots N_{\min}} I(U_\ell; X) \quad \text{s.t.} \quad I(U_\ell; X) - I(U_\ell; Y) \leq C(\hat{P}_\ell) \quad \ell = 1 \ldots N_{\min}
\]

\[
\geq \sum_{\ell=1}^{N_{\min}} \max_{U_\ell \in X \times Y} I(U_\ell; X) \quad \text{s.t.} \quad U_\ell \text{ independent of } (X, Y), \quad \ell = 1 \ldots N_{\min}
\]

\[
\geq \max_{U_1 \ldots U_{N_{\min}}} \max_{\ell=1 \ldots N_{\min}} I(U_\ell; X) - \sum_{\ell=1}^{N_{\min}} I(U_\ell; Y) \leq C(\hat{P}_\ell)
\]

\[
\overset{(a)}{=} \max_{V} I(V; X) \quad \text{s.t.} \quad I(V; X) - I(V; Y) \leq \sum_{\ell=1}^{N_{\min}} C(\hat{P}_\ell)
\]

\[
= \max_{V} I(V; X),
\]

where (a) follows from defining $V$ such that $U_\ell \in V \in X \times Y$, $\ell = 1 \ldots N_{\min}$, with $I(V; X) = \sum_{\ell=1}^{N_{\min}} I(U_\ell; X)$ and $I(V; Y) = \sum_{\ell=1}^{N_{\min}} I(U_\ell; Y)$. Here, $U_\ell$, $\ell = 1 \ldots N_{\min}$, satisfy the following constraints:

1. Each $U_\ell$, $\ell = 1 \ldots N_{\min}$ is independent of $(X, Y)$.
2. $I(U_\ell; X) - I(U_\ell; Y) \leq C(\hat{P}_\ell)$ for $\ell = 1 \ldots N_{\min}$.
3. $U_\ell$, $\ell = 1 \ldots N_{\min}$, are pairwise independent.

The existence of such $V$ is proved in the appendix.
Converse Part: The converse proof for the MIMO case is analogous to the converse proof for the SISO case.

Remark 8. The signal processing presented in Section 3.2 is optimal in the sense that with this processing, one can demonstrate the achievability of the common randomness capacity over the MIMO Gaussian channel.

Remark 9. Since $C(P, N_T \times N_R) \geq C(P)$, it follows from Proposition 1 and Proposition 2 that:

$$C_{CR}(P, N_T \times N_R) \geq C_{CR}(P).$$

Intuitively, because the MIMO channel has a higher capacity than the SISO channel, the amount of information that can be reliably transmitted is greater. Consequently, by communicating over the MIMO channel, Terminals A and B can generate a greater amount of common randomness.

4 Application of Common Randomness: Secure Identification

In this section, we explore a significant application of CR generation: the identification paradigm. Unlike transmission, it appears that the resource CR can enhance the identification capacity of channels. We introduce a coding scheme for CR-assisted secure identification and prove a lower bound on the secure identification capacity within this setting, as illustrated in Proposition 3.

Proposition 3. Let $C_{SID}(g, g', P)$ and $C_S(g, g', P)$, respectively, be the secure identification capacity and the secrecy capacity for the model in Fig. 4, respectively. It holds that if $C_S(g, g', P) > 0$ then

$$C_{SID}(g, g', P) \geq \max_{U \sim X \rightarrow Y} I(U; X).$$

Remark 10. Consider in particular the case when $C(g, P) \geq H(X|Y)$. Then, it follows from Proposition 3 that $C_{SID}(g, g', P) \geq H(X)$ as long as $C_S(g, g', P) > 0$.

Proof. Given a DMMS $P_{XY}$, Alice observes the outputs $X^n$ and Bob observes the outputs $Y^n$. Alice generates a random variable $K$ with alphabet $K = \{1 \ldots M\}'$, such that $K = \Phi(X^n)$. To send a message $i$, we prepare a set of coloring-functions or mappings $E_i$ known by the sender and the receiver.

$$E_i: K \rightarrow \{1, \ldots, M''\}$$

: $K$ coloring $\rightarrow E_i(K)$ color
$X^n$ is encoded to a sequence $T^n$ using an error correcting code, then $E_i(K)$ is encoded to a sequence $T^{[\sqrt{n}]}$ using a wiretap code as proposed in [37]. The sequence $T^m$, obtained by concatenating $T^n$ and $T^{[\sqrt{n}]}$, where $m = n + \lceil \sqrt{n} \rceil$ as depicted in Fig. 9, fulfills the power constraint:

$$E[T^2_i] \leq P \quad \forall i = 1 \ldots m.$$  

$T^m$ is sent over the Wiretap channel.

Bob generates $L = \Psi(Y^n, Z^n)$ such that $Pr[K \neq L]$ is low. Since we have used a wiretap code in the second part, Bob, interested in $i'$, can identify whether the message of interest was sent or not. We choose the rate of the first code to be approximately equal to the capacity of the channel to the legitimate receiver so that Bob can identify the message at a rate approximately equal to $C(g, P)$, the transmission capacity of the main channel, without paying a price for the identification task. Although Eve can decode with low error probability the sequence $T^n$, she cannot, with this setting, identify the color, i.e., the second fundamental part of the sent codeword, even if she knows the correlation between the sources. Thus, the wiretapper cannot identify the message $i$. For more details regarding the proof, we refer the reader to [25]. Let us denote the CR capacity for this model by $C_{CR}(g, P)$. Then, by applying the Transformator-Lemma [38] [39], it holds that

$$C_{SID}^e(g, g', P) \geq C_{CR}(g, P) \quad \text{if} \quad C_S(g, g', P) > 0.$$  

In addition, it holds by Proposition 1 that

$$C_{CR}(g, P) = \max_{U \in \mathcal{X} \otimes \mathcal{Y}} I(U; X),$$

$$I(U; X) - I(U; Y) \leq C(g, P).$$

Remark 11. It has recently been proven in [25] that the secure identification capacity with randomized encoding is equal to the capacity of the channel to
the legitimate receiver $C(g, P)$, provided that the secrecy capacity is strictly positive. As long as $P$ is chosen to satisfy $C(g, P) \leq H(X)$, the lower bound in Proposition 3 may exceed the capacity of the main channel. Let us reconsider the example of binary sources presented in Example 1. It is observed in Fig. 10 that we can achieve a performance gain of at least 0.278 for $P \approx 1.72$, $\mu = 0.2$, and $\sigma^2 = 1$.

![Fig. 10: Comparison of the lower bound on CR-assisted secure identification to the capacity of secure identification with randomized encoding for a noise variance $\sigma^2 = 1$, for the correlated sources presented in Example 1 with $\mu = 0.2$.](image)

Remark 12. Clearly, we can proceed analogously to derive a lower bound for CR-assisted secure identification capacity over MIMO GWCs.

5 Conclusions

We studied the problem of CR generation over single-user SISO and MIMO Gaussian channels due to their practical relevance in various communication
scenarios, such as satellite and deep space communication links, wired and wireless communications, etc. We provided a single-letter characterization of the CR capacity for both scenarios along with rigorous proofs. Additionally, we demonstrated that through CR generation, significant performance gains could be achieved in Post-Shannon communication tasks, which could be advantageous in numerous new applications, including machine-to-machine and human-to-machine systems, as well as the tactile internet. Specifically, we proposed a coding scheme for secure identification over the GWC with CR available as a resource and established a lower bound on the secure identification capacity within this framework. This lower bound may exceed the transmission capacity of the main channel, which is equal to the secure identification capacity in the case of randomized encoding, provided the secrecy capacity is strictly positive. As a future work, we suggest investigating the impact of antenna correlation on the CR capacity of MIMO Gaussian channels. Subsequent research could focus on providing a single-letter characterization of the CR-assisted secure identification capacity of the GWC and exploring CR-assisted identification for continuous-time channels.

A Appendix

A.1 Proof of the Existence of a random variable $V$ as defined in [28]

Proof. We want to show that a random variable $V$ exists such that $U_\ell \circ V \circ X \circ Y$, $\ell = 1 \ldots N_{\text{min}}$ with $I(V; X) = \sum_{\ell=1}^{N_{\text{min}}} I(U_\ell; X)$ and $I(V; Y) = \sum_{\ell=1}^{N_{\text{min}}} I(U_\ell; Y)$, where $U_\ell, \ell = 1 \ldots N_{\text{min}}$, satisfy the following constraints:

1. Each $U_\ell, \ell = 1 \ldots N_{\text{min}}$, is independent of $(X, Y)$.
2. $I(U_\ell; X) - I(U_\ell; Y) \leq C(\hat{P}_\ell)$, $\ell = 1 \ldots N_{\text{min}}$
3. $U_\ell, \ell = 1 \ldots N_{\text{min}}$, are pairwise independent

It suffices to consider $V = U_1 \ldots U_{N_{\text{min}}}$. Then it holds that

$$I(V; X) = I(U_1 \ldots U_{N_{\text{min}}}; X)$$
$$= \sum_{\ell=1}^{N_{\text{min}}} I(U_\ell; X | U_1 \ldots U_{\ell-1})$$
$$\overset{(a)}{=} \sum_{\ell=1}^{N_{\text{min}}} I(U_\ell; X),$$

where $(a)$ follows because $U_\ell, \ell = 1 \ldots N_{\text{min}}$, are pairwise independent and because each $U_\ell, \ell = 1 \ldots N_{\text{min}}$, is independent of $X$. Analogously, it holds that $I(V; Y) = \sum_{\ell=1}^{N_{\text{min}}} I(U_\ell; Y)$.

The Markov chain $U_\ell \circ V \circ X \circ Y$, $\ell = 1 \ldots N_{\text{min}}$, is satisfied since for
\( \ell = 1 \ldots N_{\text{min}} \), we have
\[
\mathbb{P}[Y = y | X = x, V = v, U_{\ell} = u_{\ell}]
\]
\( ^{(a)} \) \( = \mathbb{P}[Y = y | X = x, V = v] \)
\( = \mathbb{P}[Y = y | X = x, U = u_1, U_2 = u_2, \ldots, U_{N_{\text{min}}} = u_{N_{\text{min}}} ]\)
\( ^{(b)} \) \( = \mathbb{P}[Y = y | X = x], \)

where (a) follows because \( V = U_1 \ldots U_{N_{\text{min}}} \) and where (b) follows because each \( U_{\ell} \) is independent of \( (X, Y) \).

**A.2 Proof of the Non-Convexity of \( g_0(\theta) \)**

It holds that
\[
g_0(\theta) = -I(U; X)
\]
\( = -H(X) - H(U) + H(XU), \)

which yields
\[
\frac{\partial g_0(\theta)}{\partial P_{UX}(u, x)} = \frac{\partial H(XU)}{\partial P_{UX}(u, x)} \frac{\partial H(U)}{\partial P_{UX}(u, x)}.
\]

On one side, we have
\[
H(U) = -\sum_{u' \in \mathcal{U}} P_{U}(u') \log(P_{U}(u'))
\]
\( = -\sum_{u' \in \mathcal{U}} \left( \sum_{x' \in \mathcal{X}} P_{UX}(u', x') \right) \log \left( \sum_{x' \in \mathcal{X}} P_{UX}(u', x') \right) \)
yielding
\[
\frac{\partial H(U)}{\partial P_{UX}(u, x)} = - \left[ \log \left( \sum_{x' \in \mathcal{X}} P_{UX}(u, x') \right) + \frac{1}{\ln(2)} \right].
\]

On the other side, we have
\[
H(UX) = -\sum_{u', x' \in \mathcal{X}} P_{UX}(u', x') \log(P_{UX}(u', x'))
\)
yielding
\[
\frac{\partial H(UX)}{\partial P_{UX}(u, x)} = - \left[ \log(P_{UX}(u, x)) + \frac{1}{\ln(2)} \right].
\]

Thus, we obtain
\[
\frac{\partial g_0(\theta)}{\partial P_{UX}(u, x)} = \log \left( \frac{\sum_{x' \in \mathcal{X}} P_{UX}(u, x')}{P_{UX}(u, x)} \right), \quad (29)
\]
As a result,
\[
\frac{\partial^2 g_0(\theta)}{\partial^2 P_{UX}(u,x)} = \frac{P_{UX}(u,x) - \sum_{x'\in X} P_{UX}(u,x')}{P_{UX}(u,x) \sum_{x'\in X} P_{UX}(u,x')} \leq 0
\]

where
\[
\sum_{u',x'\in X} \frac{\partial^2 g_0(\theta)}{\partial^2 P_{UX}(u',x')} < 0.
\]

This implies that the Hessian matrix of \( g_0 \) is not positive semi-definite, which proves the non-convexity of \( g_0 \) in \( \theta \).

### A.3 Computation of \( \frac{\partial g_1(\theta)}{\partial P_{UX}(u,x)} \) for \( U \circ X \circ Y \)

It holds that
\[
g_1(\theta) = I(U;X|Y) - \min\{C(P), H(X|Y)\} = H(X|Y) + H(U|Y) - H(U|X|Y) - \min\{C(P), H(X|Y)\}.
\]

Thus, we have
\[
\frac{\partial g_1(\theta)}{\partial P_{UX}(u,x)} = \frac{\partial H(U|Y)}{\partial P_{UX}(u,x)} - \frac{\partial H(U|X|Y)}{\partial P_{UX}(u,x)}
\]

The Markov chain \( U \circ X \circ Y \) implies that

\( \forall x \in X, \forall u \in U \) and \( \forall y \in Y \):
\[
P_{UX|Y}(u,x|y) = \frac{P_{Y|X}(y|x) P_{UX}(u,x)}{P_Y(y)},
\]

yielding
\[
\frac{\partial P_{UX|Y}(u,x|y)}{\partial P_{UX}(u,x)} = \frac{P_{Y|X}(y|x)}{P_Y(y)}
\]

and that \( \forall u \in U \) and \( \forall y \in Y \):
\[
P_{U|Y}(u|y) = \sum_{x' \in X} P_{Y|X}(y|x') P_{UX}(u,x'),
\]

yielding
\[
\frac{\partial P_{U|Y}(u|y)}{\partial P_{UX}(u,x)} = \frac{P_{Y|X}(y|x)}{P_Y(y)}.
\]

Furthermore, we have
\[
H(U|X|Y) = -\sum_{y \in Y} P_Y(y) \sum_{u',x' \in X} P_{UX|Y}(u',x'|y) \log(P_{UX|Y}(u',x'|y)),
\]
which yields
\[ \frac{\partial H(UX|Y)}{\partial P_{UX}(u,x)} \]
\[ \overset{(a)}{=} - \sum_{y \in \mathcal{Y}} P_{Y|X}(y|x) \left[ \log \left( P_{UX|Y}(u,x|y) \right) + \frac{1}{\ln(2)} \right], \]
where (a) follows from using the sum and product rule of derivatives and from (31). Similarly, it holds that
\[ H(U|Y) = - \sum_{y \in \mathcal{Y}} P_Y(y) \sum_{u' \in \mathcal{U}} P_{U|Y}(u'|y) \log \left( P_{U|Y}(u'|y) \right). \]
Thus, we obtain
\[ \frac{\partial H(U|Y)}{\partial P_{UX}(u,x)} \]
\[ \overset{(a)}{=} - \sum_{y \in \mathcal{Y}} P_{Y|X}(y|x) \left[ \log \left( P_{U|Y}(u|y) \right) + \frac{1}{\ln(2)} \right], \]
where (a) follows from using the sum and product rule of derivatives and from (33).

As a result, we have
\[ \frac{\partial g_1(\theta)}{\partial P_{UX}(u,x)} \]
\[ = \sum_{y \in \mathcal{Y}} P_{Y|X}(y|x) \log \left( \frac{P_{UX|Y}(u,x|y)}{P_{U|Y}(u|y)} \right), \]
\[ \overset{(a)}{=} \sum_{y \in \mathcal{Y}} P_{Y|X}(y|x) \log \left( \frac{P_{Y|X}(y|x) P_{UX}(u,x)}{\sum_{x' \in \mathcal{X}} P_{Y|X}(y|x') P_{UX}(u,x')} \right), \quad (34) \]
where (a) follows from (30) and (32).

Acknowledgments

The authors acknowledge the financial support by the Federal Ministry of Education and Research of Germany in the program of “Souverän. Digital. Vernetzt.”. Joint project 6G-life, project identification number: 16KISK002. Holger Boche and Christian Deppe further gratefully acknowledge the financial support by the BMBF Quantum Programm QD-CamNetz, Grant 16KISQ077, QuaPhySI, Grant 16KIS1598K, and QUIET, Grant 16KISQ093. Christian Deppe was supported by the Bundesministerium für Bildung und Forschung (BMBF) through Grant 16KIS1005. Rami Ezzine and Wafa Labidi were supported by the BMBF through Grant 16KIS1003K.
References

1. R. Ahlswede and I. Csiszár, “Common randomness in information theory and cryptography. II. CR capacity,” IEEE Transactions on Information Theory, vol. 44, no. 1, pp. 225–240, 1998.
2. M. Sudan, H. Tyagi, and S. Watanabe, “Communication for generating correlation: A unifying survey,” IEEE Transactions on Information Theory, vol. 66, no. 1, pp. 5–37, 2020.
3. M. Naor, M. Parte, and E. Yogev, “The power of distributed verifiers in interactive proofs,” in Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms. SIAM, 2020, pp. 1096–115.
4. R. Ahlswede and G. Dueck, “Identification via channels,” IEEE Transactions on Information Theory, vol. 35, no. 1, pp. 15–29, 1989.
5. C. E. Shannon, “A mathematical theory of communication,” Bell System Technical Journal, vol. 27, pp. 379–423, 623–656, 1948.
6. H. Boche and C. Deppe, “Secure identification for wiretap channels; robustness, super-additivity and continuity,” IEEE Transactions on Information Forensics and Security, vol. 13, no. 7, pp. 1641–1655, 2018.
7. Y. Lu, “Industry 4.0: A survey on technologies, applications and open research issues,” Journal of Industrial Information Integration, vol. 6, no. 1, pp. 1–10, 2017.
8. G. P. Fettweis and H. Boche, “6G: The personal tactile internet—and open questions for information theory,” IEEE BITS the Information Theory Magazine, vol. 1, no. 1, pp. 71–82, 2021.
9. J. Cabrera, H. Boche, C. Deppe, R. Schaefer, C. Scheunert, and F. Fitzek, “6G and the post-shannon theory,” in 2022, E. Bertin, N. Crespi, and T. Magedanz, Eds., 2022.
10. P. Moulin, “The role of information theory in watermarking and its application to image watermarking,” Signal Processing, vol. 81, no. 6, pp. 1121–1139, 2001.
11. R. Ahlswede and N. Cai, “Watermarking identification codes with related topics on common randomness,” in Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 107–153.
12. Y. Steinberg and N. Merhav, “Identification in the presence of side information with application to watermarking,” IEEE Transactions on Information Theory, vol. 47, no. 4, pp. 1410–1422, 2001.
13. R. Ahlswede and I. Csiszár, “Common randomness in information theory and cryptography. I. secret sharing,” IEEE Transactions on Information Theory, vol. 39, no. 4, pp. 1121–1132, 1993.
14. S. Mathur, W. Trappe, N. Mandayam, C. Ye, and A. Reznik, “Radio-telepathy: Extracting a secret key from an unauthenticated wireless channel,” in Proceedings of the 14th ACM International Conference on Mobile Computing and Networking, ser. MobiCom ’08, 2008, p. 128–139.
15. M. Bellare, S. Tessaro, and A. Vardy, “Semantic security for the wiretap channel,” in Advances in Cryptology - CRYPTO 2012, ser. Lecture Notes in Computer Science, R. Safavi-Naini and R. Canetti, Eds. Berlin, Heidelberg: Springer, 2012, vol. 7417, pp. 294–311.
16. M. Wiese and H. Boche, “Semantic security via seeded modular coding schemes and ramanujan graphs,” IEEE Transactions on Information Theory, vol. 67, no. 1, pp. 52–80, 2021.
17. G. P. Fettweis and H. Boche, “On 6G and trustworthiness,” Communications of the ACM, vol. 65, no. 4, pp. 48–49, 2022.
18. H. Boche, R. F. Schaefer, and H. V. Poor, “Denial-of-service attacks on communication systems: Detectability and jammer knowledge,” IEEE Transactions on Signal Processing, vol. 68, pp. 3754–3768, 2020.

19. Y. Chen, T. Oechtering, H. Boche, M. Skoglund, and Y. Luo, “Distribution-preserving integrated sensing and communication with secure reconstruction,” in IEEE International Symposium on Information Theory, 2024.

20. P. Schwenteck, G. T. Nguyen, H. Boche, W. Kellerer, and F. H. P. Fitzek, “6G perspective of mobile network operators, manufacturers, and verticals,” IEEE Networking Letters, vol. 5, no. 3, pp. 169–172, 2023.

21. D. D. N. Bevan, V. T. Ermolayev, A. G. Flaksman, I. M. Averin, and P. M. Grant, “Gaussian channel model for macrocellular mobile propagation,” in 13th European Signal Processing Conference, 2005, pp. 1–4.

22. H. Boche and C. Deppe, “Secure identification under passive eavesdroppers and active jamming attacks,” IEEE Transactions on Information Forensics and Security, vol. 14, no. 2, pp. 472–485, 2019.

23. S. Baur, C. Deppe, and H. Boche, “Secure storage for identification; random resources and privacy leakage,” IEEE Transactions on Information Forensics and Security, vol. 14, no. 8, pp. 2013–2027, 2019.

24. I. Csiszár and J. Korner, “Broadcast channels with confidential messages,” IEEE Transactions on Information Theory, vol. 24, no. 3, pp. 339–348, 1978.

25. W. Labidi, C. Deppe, and H. Boche, “Secure identification for Gaussian channels,” in ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2020, pp. 2872–2876.

26. A. D. Wyner, “The wire-tap channel,” The Bell System Technical Journal, vol. 54, no. 8, pp. 1355–1387, 1975.

27. S. Verdu and V. K. Wei, “Explicit construction of optimal constant-weight codes for identification via channels,” IEEE Transactions on Information Theory, vol. 39, no. 1, pp. 30–36, 1993.

28. S. Derebeyoğlu, C. Deppe, and R. Ferrara, “Performance analysis of identification codes,” Entropy, vol. 22, no. 10, p. 1067, 2020.

29. I. Bjelaković, H. Boche, and J. Sommerfeld, “Capacity results for compound wire-tap channels,” in 2011 IEEE Information Theory Workshop, 2011, pp. 60–64.

30. I. Csiszár and J. Körner, Information Theory: Coding Theorems for Discrete Memoryless Systems, 2nd ed. Cambridge University Press, 2011.

31. A. Cotter, H. Jiang, and K. Sridharan, “Two-player games for efficient non-convex constrained optimization,” in Proceedings of Machine Learning Research, vol. 98, 2019, pp. 1–33.

32. J. Duchi, E. Hazan, and Y. Singer, “Adaptive subgradient methods for online learning and stochastic optimization,” Journal of Machine Learning Research, vol. 12, no. 61, pp. 2121–2159, 2011.

33. W. Wang and M. A. Carreira-Perpinán, “Projection onto the probability simplex: An efficient algorithm with a simple proof, and an application,” arXiv preprint arXiv:1309.1541, 2013.

34. E. Telatar, “Capacity of multi-antenna Gaussian channels,” European Transactions on Telecommunications, vol. 10, no. 6, pp. 585–595, 1999.

35. D. Tse and P. Viswanath, Fundamentals of Wireless Communication. New York, NY, USA: Cambridge University Press, 2005.

36. W. Yu, W. Rhee, S. Boyd, and J. Cioffi, “Iterative water-filling for Gaussian vector multiple-access channels,” IEEE Transactions on Information Theory, vol. 50, no. 1, pp. 145–152, 2004.
37. R. Ahlswede and Z. Zhang, “New directions in the theory of identification via channels,” *IEEE Transactions on Information Theory*, vol. 41, no. 4, pp. 1040–1050, 1995.

38. R. Ahlswede, “General theory of information transfer: Updated,” *Discrete Applied Mathematics*, vol. 156, pp. 1348–1388, 2008.

39. R. Ahlswede and G. Dueck, “Identification in the presence of feedback-a discovery of new capacity formulas,” *IEEE Transactions on Information Theory*, vol. 35, no. 1, pp. 30–36, 1989.