Decomposing the MIMO Wiretap Channel

Anatoly Khina
EE—Systems Dept., TAU
Tel Aviv, Israel
Email: anatolyk@eng.tau.ac.il

Yuval Kochman
School of CSE, HUJI
Jerusalem, Israel
Email: yuvalko@cs.huji.ac.il

Ashish Khisti
ECE Dept., U. of Toronto
Toronto, ON M5S 3G4, Canada
Email: akhisti@comm.utoronto.ca

***Abstract*—The problem of sending a secret message over the multiple-input multiple-output (MIMO) wiretap Gaussian channel is studied. While the capacity of this channel is known, it is not clear how to construct optimal coding schemes that achieve this capacity. In this work we show how to use linear operations along with successive interference cancellation in order to reduce the problem to that of designing optimal codes for the single-antenna additive-noise Gaussian wiretap channel. Much like popular communication techniques in the absence of an eavesdropper, the data is carried over parallel streams. The design approach is flexible enough to allow for using the same scalar wiretap code over all streams, or alternatively to use different scalar wiretap codes over parallel sub-channels without successive interference cancellation. This approach is applicable to more involved secrecy settings, by adjusting the linear operations performed by the encoder, and by jointly processing several channel uses.**

***Index Terms*—wiretap channel, MIMO channel, successive interference cancellation, matrix decomposition**

I. Introduction

The wiretap channel, introduced by Wyner [1], is composed of a sender ("Alice") who wishes to convey data to a legitimate user ("Bob"), such that the eavesdropper ("Eve") cannot recover any information of this data. The capacity of this channel [1] equals to a mutual-information difference, and was extended to the Gaussian case in [2]. Let the channels from Alice to Bob and Eve be given by

$$y_B = h_B x + z_B$$
$$y_E = h_E x + z_E\,,$$

where h_B and h_E are complex scalar gains, z_B and z_E are mutually-independent circularly-symmetric standard additive Gaussian noises and the transmission is subject to a unit power constraint. Then, the capacity is given by

$$C_S(h_B, h_E) = \left[\log\left(1+|h_B|^2\right) - \log\left(1+|h_E|^2\right)\right]_+, \quad (1)$$

where $[x]_+ \triangleq \max\{0, x\}$.

The vector extension of this result, the multiple-input multiple-output (MIMO) Gaussian wiretap channel or the multiple-input multiple-output multiple-eavesdropper (MIMOME) channel [3]–[5], is given by

$$\boldsymbol{y}_B = \mathbf{H}_B\boldsymbol{x} + \boldsymbol{z}_B \quad (2a)$$
$$\boldsymbol{y}_E = \mathbf{H}_E\boldsymbol{x} + \boldsymbol{z}_E, \quad (2b)$$

where $\boldsymbol{x}$, $\boldsymbol{y}_B$ and $\boldsymbol{y}_E$ are complex-valued vectors with dimensions of the number of antennas in the terminals of Alice, Bob and Eve, denoted by N_A, N_B, and N_E, respectively. The channel matrices $\mathbf{H}_B$ and $\mathbf{H}_E$ have the corresponding dimensions. The additive noise vectors $\boldsymbol{z}_B$ and $\boldsymbol{z}_E$ are mutually independent, i.i.d., circularly-symmetric Gaussian with unit element variance. Finally, the transmission is subject to a total (over all antennas) unit power constraint. The capacity of this channel is given by:

$$C_S(\mathbf{H}_B, \mathbf{H}_E) = \max_{\mathbf{K}:\,\mathrm{trace}\{\mathbf{K}\}\le 1} I_S(\mathbf{H}_B, \mathbf{H}_E, \mathbf{K}) \quad (3)$$
$$\triangleq \max_{\mathbf{K}:\,\mathrm{trace}\{\mathbf{K}\}\le 1} I(\mathbf{H}_B, \mathbf{K}) - I(\mathbf{H}_E, \mathbf{K})\,, \quad (4)$$

where

$$I(\mathbf{H}, \mathbf{K}) = \log\left|\mathbf{I} + \mathbf{H}\mathbf{K}\mathbf{H}^\dagger\right|\,.$$

Thus, capacity is given by the difference of mutual informations to Bob and Eve, optimized over all Gaussian channel inputs that satisfy the power constraint.

Although capacity is well understood, it is less clear how to construct codes for wiretap channels. For the scalar Gaussian case, various approaches have been suggested, see, e.g., [6]–[10] and references therein. However, assuming that we have such a code for the scalar case, how do we extend it to the vector case? Do we need to construct different codes for every dimension and every covariance matrix $\mathbf{K}$?

A similar question is encountered already in the context of communication over the MIMO channel without secrecy requirements (2a). For any input covariance matrix, the mutual information is given by $I(\mathbf{H}_B, \mathbf{K})$, and the capacity is the maximum over these matrices. The mutual information can be decomposed as

$$I(\mathbf{H}_B, \mathbf{K}) = \sum_{i=1}^{N_A} \log \lambda_i^2$$

where $\{\lambda_i\}$ are the *singular values*[1] of the matrix $\mathbf{G}_B = G(\mathbf{H}_B, \mathbf{K})$, where[2]

$$G(\mathbf{H}, \mathbf{K}) \triangleq \begin{pmatrix} \mathbf{H}\mathbf{K}^{1/2} \\ \mathbf{I} \end{pmatrix}\,. \quad (5)$$

[1]Throughout the paper, the singular values of a matrix are indexed in non-decreasing order.

[2]$\mathbf{K}^{1/2}$ is the principal square root of the Hermitian positive-semidefinite matrix $\mathbf{K}$, which may be found via orthogonal diagonalization or the Cholesky decomposition. However, it may be replaced, w.l.o.g., by any matrix $\mathbf{B}$ satisfying: $\mathbf{B}\mathbf{B}^\dagger = \mathbf{K}$.

This decomposition has an operational meaning: one may construct a singular-value decomposition (SVD) based scheme [11] with data streams of rates $R_i = \log \lambda_i^2$ which are encoded and decoded separately, except for some common "signal processing" part, independent of the code selection. Thus, the coding problem is reduced to a scalar one, and the gap to capacity depends on that of the scalar code.

We can apply a similar decomposition to the mutual-information difference (4). Namely [12]

$$I_S\left(\mathbf{H}_B, \mathbf{H}_E, \mathbf{K}\right) = \sum_{i=1}^{N_A} \left[\log \gamma_i^2\right]_+ , \tag{6}$$

where $\{\gamma_i\}$ are the ratios of singular values of $\mathbf{G}_B$ and $\mathbf{G}_E = G(\mathbf{H}_E, \mathbf{K})$ (5).[3] But does this decomposition imply a scheme? This is not trivial, since (as we show in the sequel) the resulting eavesdropper measurement vector $\boldsymbol{y}_B$ contains dependent components, each one in general being a function of all of the data streams. In principle, then, Eve could obtain information at least about some of the sub-streams. Indeed, facing that problem, prior work [3] suggested a scheme where $\boldsymbol{y}_E$ contains independent functions of the different sub-streams; however, the scheme is optimal only in the high signal-to-noise ratio (SNR) limit.

In this work we present schemes based on per-stream wiretap codes that allow to achieve the capacity of the MIMO wiretap channel. We start by reviewing the capacity-achieving schemes for communication over single-user MIMO channels, in Section II. In Section III we show how to modify the scheme of [3] such that Eve receives the streams independently, yet optimality is maintained at general SNR. However, maintaining orthogonality may not be desirable, as it consumes all of the degrees of freedom offered by the channel. For this, we prove an achievable secrecy region of superposition coding over any memoryless channel in Section IV, which allows more flexibility for the Gaussian MIMO setting, and is used in Section V for the MIMO Gaussian case, proving the optimality of a general scheme which presents to Eve non-orthogonal sub-channels. Finally, in Section VI we discuss various secrecy network settings allowed by this general framework.

II. Scalar Transmission over MIMO Channels

In this section we briefly review the connection between matrix decompositions and scalar transmission schemes, without secrecy requirements. For a more thorough account, the reader is referred to [13], [14]. Consider the channel (2a). Construct the augmented matrix $\mathbf{G}_B$ (5) and choose some unitary matrix $\mathbf{V}$ (the considerations for choosing $\mathbf{V}$ will become clear later). Apply the QR decomposition:

$$\mathbf{G}_B\mathbf{V} = \mathbf{U}_B\mathbf{T}_B$$

where $\mathbf{U}_B$ is unitary and $\mathbf{T}_B$ is upper-triangular. Now let $\tilde{\boldsymbol{x}}$ be a vector of standard Gaussian variables, and set

$$\boldsymbol{x} = \mathbf{K}^{1/2}\mathbf{V}\tilde{\boldsymbol{x}} \,. \tag{7}$$

[3]It may look surprising that the *individual* SVDs of $\mathbf{G}_B$ and $\mathbf{G}_E$ are aligned. In fact, a *joint* decomposition of $\mathbf{H}_B$ and $\mathbf{H}_E$ is implicit in the optimal choice of $\mathbf{K}$, which appears in $\mathbf{G}_B$ and $\mathbf{G}_E$. See Remark 3 in Section III.

Denote by $\tilde{\mathbf{U}}_B$ the sub-matrix consisting of the upper-left $N_B \times N_A$ block of $\mathbf{U}_B$, define $\tilde{\mathbf{T}} = \tilde{\mathbf{U}}_B^\dagger \mathbf{K}^{1/2}\mathbf{V}$, and let

$$\tilde{\boldsymbol{y}}_B = \tilde{\mathbf{U}}_B^\dagger \boldsymbol{y}_B = \tilde{\mathbf{U}}_B^\dagger \mathbf{K}^{1/2}\mathbf{V}\tilde{\boldsymbol{x}} + \tilde{\mathbf{U}}_B^\dagger \boldsymbol{z}_B = \tilde{\mathbf{T}}\tilde{\boldsymbol{x}} + \tilde{\boldsymbol{z}}_B \,. \tag{8}$$

Since $\tilde{\mathbf{U}}_B$ is not unitary, the statistics of $\tilde{\boldsymbol{z}} \triangleq \tilde{\mathbf{U}}_B^\dagger \boldsymbol{z}$ differ from those of $\boldsymbol{z}$, and its covariance matrix is given by $\mathbf{K}_{\tilde{\boldsymbol{z}}} \triangleq \tilde{\mathbf{U}}_B \tilde{\mathbf{U}}_B^\dagger$. Now, for $i = 1, \ldots, N_A$, define

$$\begin{aligned} y'_{B;i} &= \tilde{y}_{B;i} - \sum_{\ell=i+1}^{N_A} \tilde{T}_{i,\ell}\tilde{x}_\ell \\ &= \tilde{T}_{i,i}\tilde{x}_i + \sum_{\ell=1}^{i-1} \tilde{T}_{i,\ell}\tilde{x}_\ell + \tilde{z}_i \triangleq \tilde{T}_{i,i}\tilde{x}_i + z_i^{\text{eff}} \,. \end{aligned} \tag{9}$$

In this scalar channel from $\tilde{x}_i$ to $y'_{B;i}$, we see other $\tilde{x}_\ell$ as "interference", $\tilde{z}_i$ — as "noise", and their sum z_i^{eff} — as "effective noise". The resulting signal-to-interference-and-noise ratio (SINR) is given by:

$$S_i \triangleq \frac{(\tilde{T}_{i,i})^2}{K_{\boldsymbol{z}^{\text{eff}};i,i}} \triangleq \frac{(\tilde{T}_{i,i})^2}{K_{\tilde{\boldsymbol{z}};i,i} + \sum_{\ell=1}^{i-1}(\tilde{T}_{i,\ell})^2} \,,$$

where $K_{\tilde{\boldsymbol{z}};i,j}$ denotes the (i,j) entry of $\mathbf{K}_{\tilde{\boldsymbol{z}}}$. The following key result achieves the mutual information [13, Lemma III.3], [15][4]

$$\begin{aligned} I\left(\tilde{x}_i; \boldsymbol{y}_B \middle| \tilde{x}_{i+1}^{N_A}\right) &= I\left(\tilde{x}_i; y'_{B;i}\right) \\ &= \log(1 + S_i) = \log(b_i^2) \,, \end{aligned} \tag{10}$$

where $\{b_i\}$ are the diagonal values of $\mathbf{T}_B$,[5] such that

$$\sum_{i=1}^{N_A} \log\left(b_i^2\right) = \sum_{i=1}^{N_A} \log\left(1 + S_i\right) = I(\mathbf{H}_B, \mathbf{K}) \,, \tag{11}$$

which equals the channel capacity for the optimal $\mathbf{K}$.

This analysis immediately gives rise to the following scheme, which is, in turn, a variant of the renowned V-BLAST/GDFE scheme [15]–[17].

Scheme 1 (MIMO comm. without secrecy constraint):

Offline: construct good N_A scalar AWGN codes that are good for SNRs $\{S_i\}$.[6]

Alice: At each time instance:

- Forms $\tilde{\boldsymbol{x}}$, using one sample from each codebook
- Transmits $\boldsymbol{x}$ according to (7):

$$\boldsymbol{x} = \mathbf{K}^{1/2}\mathbf{V}\tilde{\boldsymbol{x}}$$

Bob:

- At each time instance forms $\tilde{\boldsymbol{y}}_B$ according to (8):

$$\tilde{\boldsymbol{y}}_B = \tilde{\mathbf{U}}_B^\dagger \boldsymbol{y}_B = \tilde{\mathbf{T}}\tilde{\boldsymbol{x}} + \tilde{\boldsymbol{z}}_B$$

[4]Note that, even though $\tilde{\boldsymbol{z}}$ has dependent components, the entries of the effective noise $\boldsymbol{z}^{\text{eff}}$, are independent.

[5]The diagonal of the triangular matrix resulting after applying the (canonical) QR decomposition is real and non-negative.

[6]More generally, any number $N \geq \text{rank}\{\mathbf{K}\}$ of scalar codebooks can be used; see [13], [14] for details.

- The codebooks are decoded using successive interference cancellation (SIC), from last ($i = N_A$) to first ($i = 1$). Assuming correct decoding of all codebooks $i+1, \ldots, N_A$, Bob forms $y'_{B;i}$ (9):

$$y'_{B;i} = \tilde{T}_{i,i}\tilde{x}_i + z_i^{\text{eff}}\,.$$

By the analysis above, the scheme is optimal in the sense that the sum of codebook rates can approach the channel capacity.

III. Orthogonalizing Eve's Channel

In this section we present a scheme that allows to approach the capacity of the MIMO wiretap channel (2) using any good scalar wiretap codes. The scheme is based upon the decomposition presented above, with the unitary $\mathbf{V}$ chosen according to the SVD of Eve's channel:

$$\mathbf{G}_E = \mathbf{U}_E\mathbf{D}_E\mathbf{V}^\dagger\,, \tag{12}$$

where $\mathbf{D}_E$ is generalized-diagonal of dimensions $(N_A+N_E)\times N_A$. Denote the diagonal values as $\{e_i\}$.

The scheme we use contains parts which are identical to parts in Scheme 1. We include these for completeness.

Scheme 2 (MIMO Communication over a wiretap channel):

Offline:

- Apply the QR decomposition to

$$\mathbf{G}_k\mathbf{V} = \mathbf{U}_k\mathbf{T}_k\,, \qquad k \in \{B, E\}\,,$$

  where $\{b_i\}$ and $\{e_i\}$ are the diagonal values of $\mathbf{T}_B$ and $\mathbf{T}_E$, respectively, and $\tilde{\mathbf{U}}_B$ is the upper-left $N_B \times N_A$ sub-matrix of $\mathbf{U}_B$
- Construct good scalar Gaussian wiretap codes, designed for Bob's SNRs $\{b_i^2 - 1\}$ and Eve's SNRs $\{e_i^2 - 1\}$

Alice: At each time instance:

- Forms $\tilde{\boldsymbol{x}}$, using one sample from each codebook
- Transmits $\boldsymbol{x}$ according to (7): $\boldsymbol{x} = \mathbf{K}^{1/2}\mathbf{V}\tilde{\boldsymbol{x}}$

Bob:

- At each time instance forms $\tilde{\boldsymbol{y}}_B$ according to (8)
- The codebooks are decoded using SIC, from last ($i = N_A$) to first ($i = 1$). Assuming correct decoding of all codebooks $i+1, \ldots, N_A$, Bob forms $y'_{B;i}$ (9).

The optimality of this scheme is stated in the following theorem.

Theorem 1: Scheme 2 achieves the secrecy capacity $C_S\left(\mathbf{H}_B, \mathbf{H}_E\right)$, by using the optimal input covariance matrix $\mathbf{K}$ of (3), $\mathbf{V}$ of the SVD of $\mathbf{G}_E$ (12), and *any* scalar Gaussian capacity-achieving wiretap codes that are designed for the Bob–Eve SNR-pairs $\left\{\left(b_i^2 - 1, e_i^2 - 1\right)\right\}$.

Remark 1: Strong (resp. weak) secrecy of the scalar codes guarantees strong (resp. weak) secrecy of the full scheme.

For the proof, we shall use the following connection with the SVD without the identity matrix:

$$\mathbf{H}_E\mathbf{K}^{1/2} = \bar{\mathbf{U}}_E\bar{\mathbf{D}}_E\bar{\mathbf{V}}^\dagger\,,$$

with the diagonal elements of $\bar{\mathbf{D}}_E$ denoted by $\{d_i\}$.

Lemma 1: Define $d_i = 0$ and $e_i = 1$ for $i > N_A$. Define further Λ_E as the generalized diagonal matrix of dimensions $N_E \times N_A$ whose diagonal is equal to $\left(\frac{d_1}{e_1}, \ldots, \frac{d_r}{e_r}\right)$, where $r = \min\{N_A, N_E\}$. Then,

1) $\bar{\mathbf{V}} = \mathbf{V}$
2) $1 + d_i^2 = e_i^2\,, \qquad i = 1, \ldots, N_A$
3) $\tilde{\mathbf{U}}_E = \bar{\mathbf{U}}_E\ \Lambda_E$

The proof is straightforward.

Proof of Theorem 1: Recalling (1), the total rate can approach

$$R = \sum_{i=1}^{N_A} C_S\left(\sqrt{b_i^2 - 1}, \sqrt{e_i^2 - 1}\right) = \sum_{i=1}^{N_A}\left[\log\frac{b_i^2}{e_i^2}\right]_+. \tag{13}$$

Due to (11), this is at least the mutual-information difference (4), thus capacity can be approached with the optimal $\mathbf{K}$.

Bob can decode just as he did without secrecy; it remains to bound the mutual information that Eve can gain. Let $\tilde{\boldsymbol{y}}_E = \tilde{\mathbf{U}}_E^\dagger\boldsymbol{y}_E$ as in (8). Using Lemma 1, we now show that there is no loss in information when applying $\tilde{\mathbf{U}}_E^\dagger = \Lambda_E^\dagger U_E^\dagger$, that is, $I(\tilde{\boldsymbol{x}}; \tilde{\boldsymbol{y}}_E) = (\tilde{\boldsymbol{x}}; \boldsymbol{y}_E)$. For this, apply first $\bar{U}_E^\dagger$ to $\boldsymbol{y}_E$:

$$\begin{aligned}\bar{\mathbf{U}}^\dagger\boldsymbol{y}_E &= \bar{\mathbf{U}}_E^\dagger(\mathbf{H}_E\mathbf{K}^{1/2}\bar{\mathbf{V}}\tilde{\boldsymbol{x}} + \boldsymbol{z}_E)\\ &= \bar{\mathbf{D}}_E\tilde{\boldsymbol{x}} + \bar{\mathbf{U}}_E\boldsymbol{z}_E = \bar{\mathbf{D}}_E\tilde{\boldsymbol{x}} + \bar{\boldsymbol{z}}_E\,,\end{aligned} \tag{14}$$

where $\bar{\boldsymbol{z}}_E \triangleq \bar{\mathbf{U}}_E\boldsymbol{z}_E$ has the same statistics as $\boldsymbol{z}_E$. The resulting channel is diagonal with i.i.d. noise. Since, $\bar{\mathbf{U}}_E^\dagger$ is invertible, its application incurs no loss in information.

Next multiply (14) by the generalized diagonal matrix $\Lambda_E^\dagger$:

$$\tilde{\boldsymbol{y}}_E = \tilde{\mathbf{U}}_E\boldsymbol{y}_E = \Lambda_E^\dagger\bar{\mathbf{U}}_E^\dagger\boldsymbol{y}_E = \Lambda_E^\dagger\bar{\mathbf{D}}_E\tilde{\boldsymbol{x}} + \Lambda_E^\dagger\bar{\boldsymbol{z}}_E\,.$$

The resulting $N_A \times N_A$ diagonal matrix $\Lambda_E^\dagger\bar{\mathbf{D}}_E$ has the same rank as $\bar{\mathbf{D}}_E$, that is, it has no effect on the SNR (or information) of any of the parallel channels, as desired.

Hence, applying $\tilde{\mathbf{U}}_E^\dagger$ results in parallel independent Gaussian scalar channels, with no loss of information. The resulting parallel (orthogonal) AWGN channels have SNRs $d_i^2 = e_i^2 - 1$, as we assumed in constructing the scalar wiretap codes. Thus, secrecy is guaranteed. ∎

Remark 2: For the optimal $\mathbf{K}$, either $b_i > e_i$ or $b_i = e_i = 1$, for all i, since otherwise $\mathbf{K}$ can be improved by allocating power to sub-channels for which $b_i > e_i$ — in contradiction to the optimality of $\mathbf{K}$. Hence, for the optimal choice of $\mathbf{K}$, the limiting operation in (13) is inactive.

Remark 3: In the celebrated SVD-based scheme for MIMO channels of [11], the SVD plays a very different role than in (12). It applies the SVD to the *physical* channel matrix $\mathbf{H} = \mathbf{U}\mathbf{D}\mathbf{V}^\dagger$. The transmitted signal is then formed as $\boldsymbol{x} = \mathbf{V}\boldsymbol{\Phi}\tilde{\boldsymbol{x}}$, where $\boldsymbol{\Phi}$ is a water-filling (non-unitary) matrix and $\tilde{\boldsymbol{x}}$ is a vector whose entries comprise the channel codebooks. In contrast, in (12) the SVD is applied to the augmented channel matrix $\mathbf{G}_e$, which already includes the non-unitary $\mathbf{K}^{1/2}$. Thus, there is an order reversal. However, we did not specify how to construct the optimal $\mathbf{K}$. Indeed, for Scheme 1 (without secrecy constraints) it can be obtained by SVD of the channel matrix. However, even without secrecy constrants, if the sum-power consraint is replaced by, e.g., individual power consraints, it is no longer possible to obtain the optimal $\mathbf{K}$ by

decomposition. Finally note that the rate of (6) can be achieved using Scheme 14, even if $\mathbf{K}$ is suboptimal.

IV. Superposition Coding for the Wiretap Channel

In this section we generalize our view beyond the Gaussian setting. We consider the problem of using superposition coding over a memoryless wiretap channel and establish an achievable rate region. This result is used in Section V to prove the optimality of a class of schemes for the MIMO wiretap channel. In this section, with a slight abuse of notation, we denote by boldface letters n-length sequences, with n being the block length (in contrast to the other parts of the paper, where boldface letters denote spatial vectors).

Theorem 2: Let $p(y_B|x)$ and $p(y_E|x)$ be the transition distributions for the legitimate user ("Bob") and the eavesdropper ("Eve"), respectively, of a memoryless wiretap channel, where x is the transmitted signal, and y_B and y_E are the channel outputs to Bob and Eve, respectively. Let a superposition coding scheme be defined by codes $\{\tilde{x}_i : i = 1, \ldots, N_A\}$ of the respective rates and a scalar function φ such that

$$x = \varphi\left(\tilde{x}_1, \ldots, \tilde{x}_{N_A}\right) . \tag{15}$$

Then, for $\epsilon > 0$, however small, and for any product distribution $\prod_{k=1}^{N_A} p_{\tilde{\mathsf{x}}_k}(\cdot)$, there exists a scheme which achieves weak secrecy, with the k-th codebook conveying a rate:

$$R_k = I(\tilde{\mathsf{x}}_k; \mathsf{y}_B|\tilde{\mathsf{x}}_{k+1}^{N_A}) - I(\tilde{\mathsf{x}}_k; \mathsf{y}_E|\tilde{\mathsf{x}}_{k+1}^{N_A}) - \epsilon\,. \tag{16}$$

Proof: Denote

$$\tilde{R}_k \triangleq I(\tilde{\mathsf{x}}_k; \mathsf{y}_E|\tilde{\mathsf{x}}_{k+1}^{N_A}) - \epsilon\,. \tag{17}$$

For each $k = 1, \ldots, N_A$, generate a codebook $\mathcal{C}_k$ of $2^{n(R_k+\tilde{R}_k)}$ codewords, where the codebooks are independently generated with i.i.d. with distributions $p_{\tilde{\mathsf{x}}_1}(\cdot), \ldots, p_{\tilde{\mathsf{x}}_{N_A}}(\cdot)$. Within each codebook, each codeword is assigned a unique index (m_k, f_k) where $m_k \in \{1, 2, \ldots, 2^{nR_k}\}$ and $f_k \in \{1, 2, \ldots, 2^{n\tilde{R}_k}\}$. Each codeword is selected according to the secret message m_k and a fictitious message f_k drawn uniformly over its range. The transmitted codeword is therefore $\boldsymbol{x} = \varphi\left(\tilde{\boldsymbol{x}}_1(m_1, f_1), \ldots, \tilde{\boldsymbol{x}}_{N_A}(m_{N_A}, f_{N_A})\right)$. Bob's decoding is based on successive decoding starting from the last message ($k = N_A$) and proceeding to the first ($k = 1$).

Since

$$\begin{aligned} R_k + \tilde{R}_k &= I\left(\tilde{\mathsf{x}}_k; \mathsf{y}_B\middle|\tilde{\mathsf{x}}_{k+1}^{N_A}\right) - 2\epsilon && \text{(18a)} \\ &< I\left(\tilde{\mathsf{x}}_k; \mathsf{y}_B\middle|\tilde{\mathsf{x}}_{k+1}^{N_A}\right), && \text{(18b)} \end{aligned}$$

the decoding of each combined message (m_k, f_k) succeeds with arbitrarily high probability, as $n \to \infty$.

In order to satisfy the secrecy constraint, the following condition must hold, for any $\tilde{\epsilon} > 0$ and large enough n:

$$\frac{1}{n} H\left(m_1, \ldots, m_{N_A}|\boldsymbol{y}_E, \mathcal{C}\right) \geq \frac{1}{n} H(m_1, \ldots, m_{N_A}) - \tilde{\epsilon}\,,$$

where $\mathcal{C} = \{\mathcal{C}_1, \ldots, \mathcal{C}_{N_A}\}$ denotes the overall collection of the N_A codebooks.

It suffices to show that for any $\epsilon' > 0$, and large enough n,

$$\frac{1}{n} H(m_k|\boldsymbol{y}_E, m_{k+1}^{N_A}, \mathcal{C}) \geq \frac{1}{n} H(m_k) - \epsilon'$$

is satisfied for each k.

Note that

$$\begin{aligned} &H\left(m_k\middle|\boldsymbol{y}_E, m_{k+1}^{N_A}, \mathcal{C}\right) \geq H\left(m_k\middle|\boldsymbol{y}_E, \tilde{\boldsymbol{x}}_{k+1}^{N_A}, \mathcal{C}\right) \\ &= H\left(m_k, \tilde{\boldsymbol{x}}_k\middle|\boldsymbol{y}_E, \tilde{\boldsymbol{x}}_{k+1}^{N_A}, \mathcal{C}\right) - H\left(\tilde{\boldsymbol{x}}_k\middle|m_k, \boldsymbol{y}_E, \tilde{\boldsymbol{x}}_{k+1}^{N_A}, \mathcal{C}\right) \\ &= H\left(\tilde{\boldsymbol{x}}_k\middle|\boldsymbol{y}_E, \tilde{\boldsymbol{x}}_{k+1}^{N_A}, \mathcal{C}\right) - H\left(f_k\middle|m_k, \boldsymbol{y}_E, \tilde{\boldsymbol{x}}_{k+1}^{N_A}, \mathcal{C}\right) . \end{aligned}$$

Due to (17), in our construction the eavesdropper can decode f_k with probability going to 1, given $\left(m_k, \boldsymbol{y}_E, \tilde{\boldsymbol{x}}_{k+1}^{N_A}, \mathcal{C}\right)$, and hence the second term is vanishingly small. Thus, we are left with

$$\begin{aligned} &H\left(m_k\middle|\boldsymbol{y}_E, m_{k+1}^{N_A}, \mathcal{C}\right) \geq H\left(\tilde{\boldsymbol{x}}_k\middle|\boldsymbol{y}_E, \tilde{\boldsymbol{x}}_{k+1}^{N_A}, \mathcal{C}\right) - n\epsilon'_n \\ &= H\left(\tilde{\boldsymbol{x}}_1^k\middle|\tilde{\boldsymbol{x}}_{k+1}^{N_A}, \boldsymbol{y}_E, \mathcal{C}\right) - H\left(\tilde{\boldsymbol{x}}_1^{k-1}\middle|\tilde{\boldsymbol{x}}_k^{N_A}, \boldsymbol{y}_E, \mathcal{C}\right) - n\epsilon'_n\,. \end{aligned}$$

Since the two equivocations are the same quantity up to an index shift, it suffices to show that for $\delta_1 > 0$ and $\delta_2 > 0$ that vanish with ϵ and large enough n,

$$\begin{aligned} &\sum_{\ell=1}^{k}\left[I\left(\tilde{\mathsf{x}}_\ell; \mathsf{y}_B\middle|\tilde{\mathsf{x}}_{\ell+1}^{N_A}\right) - I\left(\tilde{\mathsf{x}}_\ell; \mathsf{y}_E\middle|\tilde{\mathsf{x}}_{\ell+1}^{N_A}\right)\right] - \delta_1 && \text{(19a)} \\ &\leq \frac{1}{n} H\left(\tilde{\boldsymbol{x}}_1^k\middle|\tilde{\boldsymbol{x}}_{k+1}^{N_A}, \boldsymbol{y}_E, \mathcal{C}\right) && \text{(19b)} \\ &\leq \sum_{\ell=1}^{k} I\left(\tilde{\mathsf{x}}_\ell; \mathsf{y}_B\middle|\tilde{\mathsf{x}}_{\ell+1}^{N_A}\right) - I\left(\tilde{\mathsf{x}}_\ell; \mathsf{y}_E\middle|\tilde{\mathsf{x}}_{\ell+1}^{N_A}\right) + \delta_2\,. && \text{(19c)} \end{aligned}$$

To establish (19b) we use the fact that the sequences $\tilde{\boldsymbol{x}}_\ell$ are selected independently so that, for large enough n,

$$\begin{aligned} &H\left(\tilde{\boldsymbol{x}}_1^k\middle|\tilde{\boldsymbol{x}}_{k+1}^{N_A}, \boldsymbol{y}_E, \mathcal{C}\right) && \text{(20a)} \\ &= \left[\sum_{\ell=1}^{k} H\left(\tilde{\boldsymbol{x}}_\ell|\mathcal{C}\right)\right] - I\left(\tilde{\boldsymbol{x}}_1^k; \boldsymbol{y}_E\middle|\tilde{\boldsymbol{x}}_{k+1}^{N_A}, \mathcal{C}\right) && \text{(20b)} \\ &\geq \left[\sum_{\ell=1}^{k} H\left(\tilde{\boldsymbol{x}}_\ell|\mathcal{C}\right)\right] - nI\left(\tilde{\mathsf{x}}_1^k; \mathsf{y}_E\middle|\tilde{\mathsf{x}}_{k+1}^{N_A}\right) && \text{(20c)} \\ &\geq n\sum_{\ell=1}^{k}\left[I\left(\tilde{\mathsf{x}}_\ell; \mathsf{y}_B\middle|\tilde{\mathsf{x}}_{\ell+1}^{N_A}\right) - I\left(\tilde{\mathsf{x}}_\ell; \mathsf{y}_E\middle|\tilde{\mathsf{x}}_{\ell+1}^{N_A}\right) - 3\epsilon\right] && \text{(20d)} \end{aligned}$$

where to establish (20c) we use the fact that the channel is memoryless, and (20d) follows from (18a).

To establish (19c), we use [18, Lemma 1], by substituting:

- $S = \sum_{\ell=1}^{k}\left(R_\ell + \tilde{R}_\ell\right)$
- $\mathsf{u} = \tilde{\mathsf{x}}_{k+1}^{N_A}$
- $\mathsf{v} = \tilde{\mathsf{x}}_1^k$
- $\mathsf{z} = \mathsf{y}_E$
- $L \triangleq (m_1^k, f_1^k) \in [1, 2^{nS}]$

The conditions for the lemma hold since

$$H\left(\tilde{\mathsf{x}}_1^k\middle|\tilde{\mathsf{x}}_{k+1}^{N_A}, \mathsf{y}_E, \mathcal{C}\right) = H\left(L\middle|\tilde{\mathsf{x}}_{k+1}^{N_A}, \mathsf{y}_E, \mathcal{C}\right) ,$$

and

$$S = \sum_{\ell=1}^{k} \left(R_\ell + \tilde{R}_\ell \right) \tag{21a}$$

$$= \left[\sum_{\ell=1}^{k} I\left(\tilde{\mathsf{x}}_\ell ; \mathsf{y}_B \middle| \tilde{\mathsf{x}}_{\ell+1}^{N_A} \right) \right] - 2\epsilon \tag{21b}$$

$$> \left[\sum_{\ell=1}^{k} I\left(\tilde{\mathsf{x}}_\ell ; \mathsf{y}_E \middle| \tilde{\mathsf{x}}_{\ell+1}^{N_A} \right) \right] + \delta \tag{21c}$$

$$= I\left(\tilde{\mathsf{x}}_1^k ; \mathsf{y}_E \middle| \tilde{\mathsf{x}}_{k+1}^{N_A} \right) + \delta \tag{21d}$$

where (21c) follows from the fact that the communication rate R_ℓ of each sub-channel must be positive (and ϵ and δ are small enough, and n is sufficiently large), else it is not used.

Since we have shown (19b) and (19c), the secrecy analysis is now complete. ∎

V. General Multi-Stream Scheme

In this section we specialize the results of Section IV to the wiretap MIMO channel. This allows us to generalize Scheme 2 to transmission that is not necessarily orthogonal over Eve's channel. Specifically, in Section V-A we show that the secrecy capacity can be achieved using *any* unitary matrix $\mathbf{V}$ at the encoder. The resulting family of schemes includes two important special cases, discussed in Section V-B.

A. Secrecy Constraint Proof

Theorem 3: For the optimal covariance matrix $\mathbf{K}$ and any unitary $\mathbf{V}$, there exist scalar codebooks s.t. Scheme 2 approaches the secrecy capacity $C_S(\mathbf{H}_B, \mathbf{H}_E)$.

Proof: We specialize the general superposition coding framework of Theorem 2 to the linear encoder structure. Use

$$\boldsymbol{x} = \varphi\left(\tilde{x}_1, \ldots, \tilde{x}_{N_A}\right) = \mathbf{K}^{1/2}\mathbf{V}\tilde{\boldsymbol{x}}\,,$$

in (15), where the vector $\tilde{\boldsymbol{x}}$ is composed of one symbol from each codebook: $\tilde{\boldsymbol{x}} = (\tilde{x}_1, \ldots, \tilde{x}_k)^T$.[7]

Each codebook is a scalar Gaussian wiretap codebook of average power 1. The achievable secrecy rate of codebook $k = 1, \ldots, N_A$ is given by (16):

$$R_k = I(\tilde{\mathsf{x}}_k ; \mathsf{y}_B | \tilde{\mathsf{x}}_{k+1}^{N_A}) - I(\tilde{\mathsf{x}}_k ; \mathsf{y}_E | \tilde{\mathsf{x}}_{k+1}^{N_A}) - \epsilon \tag{22a}$$

$$= \log\left(b_k^2\right) - \log\left(e_k^2\right) - \epsilon = \log\left(\frac{b_k^2}{e_k^2}\right) - \epsilon \tag{22b}$$

$$= I\left(\tilde{\mathsf{x}}_k ; \tilde{\mathsf{y}}_{B;k}\right) - I\left(\tilde{\mathsf{x}}_k ; \tilde{\mathsf{y}}_{E;k}\right) - \epsilon\,, \tag{22c}$$

where (22b) is due to (10).

Thus, using the result of (11), we can approach

$$R = \sum_{k=1}^{N} R_k = \sum_{k=1}^{N} \left[\log \frac{b_k^2}{e_k^2} \right]_+ - \epsilon$$

and for the optimal $\mathbf{K}$, the secrecy capacity (3) can be approached. ∎

[7]Here, in contrast to Section IV, boldface letters represent spatial vectors and time indices are suppressed.

Remark 4: Note that in our coding scheme, it is sufficient for Bob to use a SIC decoder as explained before. On the other hand, our allocation of rates $\{R_k\}$ in (22b) guarantees that all the messages $(m_1, ..m_{N_A})$ remain jointly secure from the eavesdroppers channel output sequence as stated in Theorem 2. In particular our result does not restrict Eve to implement a SIC decoder for guaranteeing secrecy.

Remark 5: Even though we considered a sum-power constraint, this scheme is readily applicable for other input constraints as well, e.g., an input covariance constraint or individual power constraints, by using the optimal input covariance matrix $\mathbf{K}$ resulting from the optimization for the specific input constraint.

B. Important Special Cases

We now present "special" choices of $\mathbf{V}$. Beyond performing SVD with respect to Eve's channel as done in Section III which yields an easy secrecy proof, the following choices yield practical advantages.

1) Avoiding SIC: Performing SIC adds complexity to the decoder, as well as introduces potential error propagation. We can avoid this by performing SVD w.r.t. Bob's channel, as opposed to Eve's channel as done in Section III. That is, choose $\mathbf{V}$ s.t.

$$\mathbf{G}_B = \mathbf{U}_B \mathbf{D}_B \mathbf{V}^\dagger\,.$$

As happens with Eve in Section III, Bob obtains a diagonal equivalent channel, where each sub-stream can be decoded independently.

2) Avoiding individual bit-loading: When using (non-secret) communication schemes based on SVD or QR, as in Scheme 1, the effective sub-channel gains $\{b_i\}$ are different in general. This requires, in turn, a bit-loading mechanism and the design of codes of different rates, matching these gains. An elegant way to avoid this was proposed in [13], [19]. Instead of designing a diagonal matrix, using unitary operations, a triangular form with a constant diagonal is attained, the constant value on the diagonal being equal to the geometric mean of the singular values. Thus, this decomposition is called "geometric mean decomposition" or GMD in [13] (QRS in [19]). The constant diagonal suggests that bit-loading can be avoided altogether and that the codewords sent over the resulting sub-channels can be drawn from the same codebook.

In the wiretap setting, however, using the same codebook over all the sub-channels requires a joint unitary triangularization of two matrices ($\mathbf{H}_B$ and $\mathbf{H}_E$) that achieves *constant* diagonals for both, using the same unitary matrix $\mathbf{V}$ on the right. As is shown in [20], such a decomposition, called 2-GMD, does not exist in general. Nevertheless, a nearly-optimal scheme, that processes several adjacent channel uses together (reminiscent of space–time coding structures), allows to approach this result. If N_0 channel uses are processed together, the rate efficiency of the scheme is at least $(N_0 - N_A + 1)/N_A$.

Remark 6: The 2-GMD can be extended to any number of matrices L. The resulting decomposition, named L-GMD [20], will be used in the next section.

VI. Application to other Secrecy Settings

Scheme 2 allows to construct schemes for more complex secret communication scenarios:

1) **Broadcast channel with confidential messages:** The Gaussian MIMO wiretap channel can be considered as a special case ("corner point") of the Gaussian MIMO broadcast with confidential messages. The capacity region of this scenario has been established in [21], where, for the achievability proof, secret dirty-paper coding (DPC) was utilized. Indeed, a DPC variant of Scheme 2 can be constructed. However, the whole capacity region of the confidential-message broadcast can be also be achieved using only SIC. Note that this is in contrast to the (non-confidential) private message counterpart, where all known proofs use DPC.
2) **Doubly compound wiretap channel:** The compound legitimate-user compound eavesdropper ("doubly compound") Gaussian MIMO wiretap channel (see [22], [23] and references therein) can be treated in a similar fashion to the scenario discussed in this paper, by applying the L-GMD decomposition to all the legitimate-user and eavesdropper augmented channel matrices (with the appropriate input covariance matrix $\mathbf{K}$). The rate of the (same) scalar wiretap code is dictated by the lowest (constant) channel gain (constant diagonal value) of all the legitimate users and the largest channel gain of all the eavesdroppers.
3) **Security embedding:** The problem of secured embedding or two-level security wiretap channel was proposed by Ly et al. [24]. This scenario is composed of a legitimate user and two eavesdroppers, where the first ("strong") eavesdropper is allowed to gain information about one of the messages (but not the second eavesdropper), but none of the two eavesdroppers is allowed to recover any information about the second message. The capacity of this problem, for the Gaussian parallel-channels case has been determined in [24]. For the Gaussian MIMO variant of this problem (as well as the special case of parallel channels), similarly to the proposed treatment for the compound setting, we propose to use the 3-GMD for the three matrices[8] for the transmission of the message which has to remain confidential from both of the eavesdroppers, and superimposing (any variant of) Scheme 2 for the legitimate user and the "weak" eavesdropper.
4) **Common message:** All of the schemes above can be generalized, in a straightforward manner, to the case where an additional common message needs to be conveyed to all the users. This is materialized by treating the confidential messages as noise, and superimposing an additional common message which is implemented in a similar fashion to Scheme 2, as is explained in [14], [20].

References

[1] A. D. Wyner, "The wiretap channel," *IEEE Trans. Info. Theory*, vol. 54, pp. 1355–1387, 1975.

[2] S. K. Leung-Yan-Cheong and M. E. Hellman, "The Gaussian wiretap channel," *IEEE Trans. Info. Theory*, vol. 24, pp. 451–456, 1978.

[3] A. Khisti and G. W. Wornell, "Secure transmission with multiple antennas—part II: The MIMOME wiretap channel," *IEEE Trans. Info. Theory*, vol. 56, pp. 5515–5532, 2010.

[4] F. Oggier and B. Hassibi, "The secrecy capacity of the MIMO wiretap channel," *IEEE Trans. Info. Theory*, vol. 57, pp. 4961–4972, 2011.

[5] T. Liu and S. Shamai, "A note on the secrecy capacity of the multiple-antenna wiretap channel," *IEEE Trans. Info. Theory*, vol. 55, pp. 2547–2553, 2009.

[6] H. Tyagi and A. Vardy, "Explicit capacity-achieving coding scheme for the Gaussian wiretap channel," in *Proc. IEEE Int. Symp. on Info. Theory (ISIT)*, Honolulu, HI, USA, June/July 2014.

[7] D. Klinc, H. Jeongseok, S. W. McLaughlin, J. Barros, and B.-J. Kwak, "LDPC codes for the Gaussian wiretap channel," *IEEE Trans. Info. Theory*, vol. 6, pp. 532–540, 2011.

[8] F. Oggier, P. Solé, and J.-C. Belfiore, "Lattice codes for the wiretap Gaussian channel: Construction and analysis," *IEEE Trans. Info. Theory*, Submitted, Jan. 2013. Available online at http://arxiv.org/abs/0708.4219.

[9] H. Mahdavifar and A. Vardy, "Achieving the secrecy capacity of wiretap channels using polar codes," *IEEE Trans. Info. Theory*, vol. 57, pp. 6428–6443, 2011.

[10] M. Andersson, "Coding for the wiretap channel," Ph.D. dissertation, School of Electrical Engineering (EES), Royal Institute of Technology (KTH), Stockholm, Sweden, 2011.

[11] E. Telatar, "Capacity of the multiple antenna Gaussian channel," *Europ. Trans. Telecommun.*, vol. 10, pp. 585–595, Nov. 1999.

[12] R. Bustin, R. Liu, H. V. Poor, and S. Shamai, "An MMSE approach to the secrecy capacity of the MIMO Gaussian wiretap channel," *EURASIP Journal on Wireless Comm. and Networking. Special Issue on Wireless Physical Security*, 2009.

[13] Y. Jiang, W. Hager, and J. Li, "Uniform channel decomposition for MIMO communications," *IEEE Trans. Sig. Proc.*, vol. 53, pp. 4283–4294, 2005.

[14] A. Khina, Y. Kochman, and U. Erez, "Joint unitary triangularization for MIMO networks," *IEEE Trans. Sig. Proc.*, vol. 60, pp. 326–336, 2012.

[15] B. Hassibi, "An efficient square-root algorithm for BLAST," in *2000 International Conference on Acoustics Speech and Signal Processing (ICASSP'2000)*, vol. 2, Istanbul, Turkey, 2000, pp. 737–740.

[16] P. W. Wolniansky, G. J. Foschini, G. D. Golden, and R. A. Valenzuela, "V-BLAST: an architecture for realizing very high data rates over the rich-scattering wireless channel," in *ISSSE 1998, URSI International Symposium*, pp. 295–300.

[17] J. M. Cioffi and G. D. Forney Jr., "Generalized decision-feedback equalization for packet transmission with ISI and Gaussian noise," in *Comm., Comp., Cont. and Sig. Proc.* Springer US, 1997, pp. 79–127.

[18] Y.-K. Chia and A. E. Gamal, "Three-receiver broadcast channels with common and confidential messages," *IEEE Trans. Info. Theory*, vol. 58, pp. 2748–2765, 2012.

[19] J.-K. Zhang, A. Kavčić, and K. M. Wong, "Equal-diagonal QR decomposition and its application to precoder design for successive-cancellation detection," *IEEE Trans. Info. Theory*, vol. 51, pp. 154–172, 2005.

[20] A. Khina, A. Hitron, I. Livni, and U. Erez, "Joint unitary triangularization for Gaussian multi-user MIMO networks," Tech. Rep., June 2013. [Online]. Available: http://arxiv.org/abs/1306.4350

[21] R. Liu, T. Liu, H. V. Poor, and S. Shamai, "Multiple-input multiple-output Gaussian broadcast channels with confidential messages," *IEEE Trans. Info. Theory*, vol. 56, pp. 4215–4227, 2010.

[22] Y. Liang, G. Kramer, H. V. Poor, and S. Shamai, "Compound wiretap channels," *EURASIP Journal on Wireless Comm. and Networking*, 2009.

[23] E. Ekrem and S. Ulukus, "On Gaussian MIMO compound wiretap channels," in *44th Annual Conf. on Info. Sciences and Systems (CISS)*, March 2010.

[24] H. D. Ly, T. Liu, and T. Blankenship, "Security embedding codes," *IEEE Trans. Info. Theory*, vol. 7, pp. 148–159, 2012.

[8] With an appropriate generalization of Theorem 2 for this case.