The MIMO Wiretap Channel Decomposed

Anatoly Khina, Yuval Kochman, and Ashish Khisti

***Abstract*—The problem of sending a secret message over the Gaussian multiple-input multiple-output (MIMO) wiretap 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 use linear operations along with successive interference cancellation to attain effective parallel single-antenna wiretap channels. By using independent scalar Gaussian wiretap codebooks over the resulting parallel channels, the capacity of the MIMO wiretap channel is achieved. The derivation of the schemes is based upon joint triangularization of the channel matrices. We find that the same technique can be used to re-derive capacity expressions for the MIMO wiretap channel in a way that is simple and closely connected to a transmission scheme. This technique allows to extend the previously proven strong security for scalar Gaussian channels to the MIMO case. We further consider the problem of transmitting confidential messages over a two-user broadcast MIMO channel. For that problem, we find that derivation of both the capacity and a transmission scheme is a direct corollary of the proposed analysis for the MIMO wiretap channel.**

***Index Terms*—Wiretap channel, MIMO channel, confidential broadcast, successive interference cancellation, dirty-paper coding, 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 (almost) any information of these data. The capacity of this channel [1], [2] equals to a mutual-information difference, and was extended to the Gaussian case in [3]. Let the channels from Alice to Bob and Eve be given by

$$\begin{aligned} y_B &= h_B x + z_B, \\ y_E &= h_E x + z_E, \end{aligned}$$

where h_B and h_E are complex scalar gains, z_B and z_E are mutually-independent circularly-symmetric Gaussian zero mean unit variance noises and the transmission is subject to a unit power constraint. Then, the capacity is achieved by a Gaussian input:

$$\begin{aligned} C_S(h_B, h_E) &= I\left(x; y_B\right) - I\left(x; y_E\right) && \text{(1a)} \\ &= \left[\log\left(1 + |h_B|^2\right) - \log\left(1 + |h_E|^2\right)\right]_+, && \text{(1b)} \end{aligned}$$

where $[a]_+ \triangleq \max\{0, a\}$ is the positive-part operation.

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 [4]–[6], is given by

$$\boldsymbol{y}_B = \mathbf{H}_B \boldsymbol{x} + \boldsymbol{z}_B, \tag{2a}$$

$$\boldsymbol{y}_E = \mathbf{H}_E \boldsymbol{x} + \boldsymbol{z}_E, \tag{2b}$$

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 zero mean unit element variance.

The secrecy capacity of this scenario for the case where the input is subject to an average *covariance constraint*[1]

$$\mathbf{K} \triangleq E\left[\boldsymbol{x}\boldsymbol{x}^\dagger\right] \preceq \bar{\mathbf{K}}, \tag{3}$$

and the case where the input is subject to a total (over all antennas) power constraint P:

$$\text{trace}(\mathbf{K}) \leq P,$$

was established in [6] and [4]–[6], respectively. Under a covariance constraint, this capacity is given by the difference of mutual informations to Bob and Eve, optimized over all Gaussian channel inputs that satisfy the respective input constraint:

$$C_S(\mathbf{H}_B, \mathbf{H}_E, \bar{\mathbf{K}}) = \max_{\mathbf{K} \preceq \bar{\mathbf{K}}} I_S(\mathbf{H}_B, \mathbf{H}_E, \mathbf{K}), \tag{4}$$

where

$$I_S(\mathbf{H}_B, \mathbf{H}_C, \mathbf{K}) \triangleq I(\mathbf{H}_B, \mathbf{K}) - I(\mathbf{H}_E, \mathbf{K}), \tag{5}$$

and

$$I(\mathbf{H}, \mathbf{K}) \triangleq \log\left|\mathbf{I} + \mathbf{H}\mathbf{K}\mathbf{H}^\dagger\right| \tag{6}$$

is the Gaussian vector mutual information (MI), and $|\mathbf{A}|$ denotes the determinant of $\mathbf{A}$. Later, Bustin *et al.* [7] provided an explicit solution to the maximization problem under the covariance constraint (4). A closed-form solution for the wiretap capacity under a total power constraint is yet to be found, although a numerical algorithm that approaches the global optimum was recently proposed [8]. We note that the capacity under a total power constraint can be written as the union of achievable regions under a covariance constraint (see [9, Lemma 1]):

$$C_S(\mathbf{H}_B, \mathbf{H}_E, P) = \max_{\bar{\mathbf{K}}:\, \text{trace}\{\bar{\mathbf{K}}\} = P} C_S(\mathbf{H}_B, \mathbf{H}_E, \bar{\mathbf{K}}). \tag{7}$$

Hence, we shall concentrate on the covariance constrained setting in this paper.

The material in this paper was presented in part at the *2014 IEEE International Symposium of Information Theory (ISIT)*, Honolulu, HI, USA, and at the *2015 IEEE ISIT*, Hong Kong.

[1] $\mathbf{A} \succeq \mathbf{0}$ denotes that $\mathbf{A}$ is a positive semidefinite matrix. $\mathbf{A} \preceq \mathbf{B}$ means that $(\mathbf{A} - \mathbf{B}) \succeq \mathbf{0}$.

The confidential broadcast channel offers a natural extension to the wiretap channel setting. In the confidential broadcast setting, Alice wishes to convey different data to two users ("Bob" and "Charlie"), such that (almost) no information can be recovered by one user about the data intended for the other user. That is, for the data that are intended for Bob, Charlie acts as the eavesdropper ("Eve" in the wiretap setting), whereas for the data intended for Charlie, Bob takes the role of Eve.

The capacity region of the Gaussian MIMO confidential broadcast channel, a scenario considered first in [10], was determined by Liu *et al.* [11] to be rectangular under the covariance constraint (3). Namely, it is given by all rate pairs (R_B, R_C) satisfying

$$R_B \le C_S\left(\mathbf{H}_B, \mathbf{H}_C, \bar{\mathbf{K}}\right), \tag{8a}$$
$$R_C \le C_S\left(\mathbf{H}_C, \mathbf{H}_B, \bar{\mathbf{K}}\right), \tag{8b}$$

where $\mathbf{H}_C$ is the channel matrix to Charlie replacing $\mathbf{H}_E$ in (2b), and $C_S(\mathbf{H}_B, \mathbf{H}_C, \bar{\mathbf{K}})$ is the capacity of the MIMO wiretap channel defined in (4). The converse is immediate, as both users achieve their maximal possible secrecy rates simultaneously; it is the direct part that is quite striking.

Although capacity is well understood, it is less clear how to construct codes for wiretap and confidential broadcast channels. For the scalar Gaussian case, various approaches have been suggested, see, e.g., [12]–[18] and references therein. However, assuming that we have such a code for the scalar case, it is not clear how to construct a capacity-achieving scheme for the MIMO setting.

In this work we present an approach that reduces these MIMO secrecy problems to scalar Gaussian ones by means of matrix decompositions, specifically joint unitary triangularizations [19]. The decompositions yield a layered coding scheme, where the secrecy capacity is approached by means of a scalar wiretap code in each layer and successive interference cancellation (SIC) at the receiver. The contribution of such an approach to the MIMO wiretap channel can be compared to that of singular-value decomposition (SVD) based schemes [20], or Vertical Bell-Laboratories Space–Time (V-BLAST) and decision feedback equalization (GDFE) schemes [21]–[24], to MIMO communication without secrecy constraints.

Beyond the architectural merit, our approach yields two more fruits. First, it enables us to revisit the capacity results for the MIMO wiretap and confidential MIMO broadcast channels. In that respect, we establish the optimal covariance matrix for the MIMO wiretap channel as well as an expression for the secrecy capacity in terms of the generalized singular values of suitably defined matrices. This re-derives a result by Bustin *et al.* [7], which was based on elaborate information-theoretic considerations, using a direct linear-algebraic approach. Turning to the confidential broadcast channel, we are able to re-derive (8) almost as a corollary of the analysis applied to the MIMO wiretap channel, also explaining the role of dirty-paper coding in this setup.

Second, reducing the MIMO problem to a scalar one allows us to leverage recent advances in the secrecy analysis of the scalar Gaussian wiretap channel: whereas we concentrate in this paper on constructing *weak secrecy* schemes, namely schemes for which

$$I\left(\boldsymbol{x}^n; \boldsymbol{y}_B^n\right) \le n\epsilon, \tag{9}$$

we show that in fact a special matrix triangularization allows to achieve *strong secrecy* guarantees for the MIMO wiretap channel, i.e.,

$$I\left(\boldsymbol{x}^n; \boldsymbol{y}_B^n\right) \le \epsilon, \tag{10}$$

where both (9) and (10) hold for any $\epsilon > 0$ and large enough blocklength n.

An outline of this paper is as follows. We start by reviewing the relevant unitary matrix decompositions in Section II. These decompositions are used to re-derive the MIMO wiretap capacity expressions in Section III. We further recall how these decompositions allow to construct capacity-achieving schemes for the MIMO channel without secrecy in Section IV. We extend this framework to work for the MIMO wiretap setting in Section V. Layered dirty-paper coding (DPC) [25] variants of this scheme are discussed in Section VI and are also shown to be capacity achieving. Finally, these schemes are utilized, along with the results of Section III, to construct a simple proof of the capacity region of the confidential MIMO broadcast setting as well as providing a layered-DPC scheme that attains it in Section VII.

II. Unitary Matrix Triangularization

In this section we briefly review some important matrix decompositions which will be used in the sequel. In Section II-A we recall the generalized triangular decomposition (GTD), and some of its important special cases which include the SVD, QR decomposition, and geometric mean decomposition (GMD).[2] Joint unitary triangularizations of two matrices are discussed in Section II-B.

Throughout this paper, we shall only need to decompose full-rank matrices with equal or more rows than columns.

A. Single Matrix Triangularization

The following definitions are used in this section.

Definition 1 (Multiplicative majorization; see [27])**.** Let $\boldsymbol{x}$ and $\boldsymbol{y}$ be two N-dimensional vectors of positive elements. Denote by $\tilde{\boldsymbol{x}}$ and $\tilde{\boldsymbol{y}}$ the vectors composed of the entries of $\boldsymbol{x}$ and $\boldsymbol{y}$, respectively, ordered non-increasingly. We say that $\boldsymbol{x}$ majorizes $\boldsymbol{y}$ ($\boldsymbol{x} \succeq \boldsymbol{y}$) if they have equal products:

$$\prod_{j=1}^{N} x_j = \prod_{j=1}^{N} y_j\,,$$

and their (ordered) elements satisfy, for any $1 \le \ell < N$,

$$\prod_{j=1}^{\ell} \tilde{x}_j \ge \prod_{j=1}^{\ell} \tilde{y}_j\,.$$

Definition 2 (Singular values; see [28])**.** Let $\mathbf{A}$ be a full-rank matrix of dimensions $M \times N$, where $M \ge N$. Then, the

[2] See [26] for a geometrical interpretation of these decompositions.

singular values (SVs) of $\mathbf{A}$ are the positive solutions σ of the equation

$$\left|\mathbf{A}^\dagger\mathbf{A} - \sigma^2\mathbf{I}\right| = 0.$$

Let the SV vector $\boldsymbol{\sigma}(\mathbf{A})$ be composed of all SVs (including their algebraic multiplicity), ordered non-increasingly.

The following is a straightforward extension of the definition of triangular matrices to non-square ones.

Definition 3 (Generalized Upper-Triangular Matrix)**.** An $M \times N$ matrix is said to be generalized upper triangular if

$$T_{i,j} = 0\,, \quad \forall i > j\,; \quad i = 1, \ldots, M\,; \quad j = 1, \ldots, N.$$

We use these definitions to characterize the set of all possible diagonals achievable via unitary triangularization, as follows.

Theorem 1 (Generalized Triangular Decomposition)**.** *Let* $\mathbf{A}$ *be a full-rank matrix of dimensions* $M \times N$*, where* $M \geq N$*, and* $\boldsymbol{t}$ *be an* N*-dimensional vector of positive elements. A GTD of the matrix* $\mathbf{A}$ *is given by*

$$\mathbf{A} = \mathbf{U}\mathbf{T}\mathbf{V}^\dagger, \tag{11}$$

where $\mathbf{U}$ *and* $\mathbf{V}$ *are unitary matrices of dimensions* $M \times M$ *and* $N \times N$*, respectively, and* $\mathbf{T}$ *is a generalized upper-triangular matrix with a prescribed set of diagonal values* $\boldsymbol{t}$*, i.e.,*

$$\begin{aligned} T_{ii} &= t_i\,, & i &= 1, \ldots, N\,, \\ T_{ij} &= 0\,, & \forall i &> j\,. \end{aligned}$$

Such a decomposition exists if and only if the vector $\boldsymbol{t}$ *is majorized by* $\boldsymbol{\sigma}(\mathbf{A})$*:*

$$\boldsymbol{\sigma}(\mathbf{A}) \succeq \boldsymbol{t}\,.$$

In other words, the singular values are an extremal case for the diagonal of all possible unitary triangularizations.

The necessity of the majorization condition was proven by Weyl [29]. Horn further showed that for any $\boldsymbol{r}$ that is majorized by $\boldsymbol{\sigma}$, there exists an upper triangular matrix with diagonal $\boldsymbol{r}$ and SV vector $\boldsymbol{\sigma}$ [30]. The sufficiency of the majorization condition as it appears in Theorem 1 was proved in [31]–[33], where also explicit constructions of the decomposition were introduced.

We now recall three important special cases of the GTD.

1) SVD (See, e.g., [28]): Here the resulting matrix $\mathbf{T}$ in (11) is a *diagonal* matrix, and its diagonal elements are equal to the singular values of the decomposed matrix $\mathbf{A}$.

2) QR Decomposition (See, e.g., [28]): In this decomposition, the matrix $\mathbf{V}$ in (11) equals to the identity matrix and hence does not depend on the matrix $\mathbf{A}$. This decomposition can be constructed by performing Gram–Schmidt orthonormalization on the (ordered) columns of the matrix $\mathbf{A}$.

3) GMD (See [31], [34], [35]): The diagonal elements of $\mathbf{T}$ in this decomposition are all equal to the geometric mean of its singular values $\boldsymbol{\sigma}(\mathbf{A})$, which is real and positive. Note that this decomposition always exists if $\mathbf{A}$ is full rank (since the vector of the SVs of $\mathbf{A}$ necessarily majorizes the vector of the diagonal elements of $\mathbf{T}$), but is not unique.

B. Joint Matrix Triangularization

The existence condition for a joint unitary triangularization of two matrices is similar to that of the GTD in Theorem 1, where the singular values are replaced by the generalized singular values (GSVs), and the diagonal of $\mathbf{T}$ is replaced by the ratio of the diagonals of the resulting generalized triangular matrices. These quantities are defined below.

Definition 4 (Generalized singular values [28], [36])**.** For any (ordered) matrix pair $(\mathbf{A}_1, \mathbf{A}_2)$, the GSVs are the non-negative solutions μ of the equation

$$\left|\mathbf{A}_1^\dagger\mathbf{A}_1 - \mu^2\mathbf{A}_2^\dagger\mathbf{A}_2\right| = 0.$$

Let the GSV vector $\boldsymbol{\mu}(\mathbf{A}_1, \mathbf{A}_2)$ be composed of all GSVs (including their algebraic multiplicity), ordered non-increasingly.

A characterization of the possible joint unitary triangularizations of two matrices with prescribed diagonal ratios is provided in the following theorem.

Theorem 2 (Joint unitary triangularization [19])**.** *Let* $\mathbf{A}_1$ *and* $\mathbf{A}_2$ *be two full-rank matrices of dimensions* $M_1 \times N$ *and* $M_2 \times N$*, respectively, where* $M_1, M_2 \geq N$*, and* $\boldsymbol{t}$ *be an* N*-dimensional vector of positive elements. A joint unitary triangularization of the matrices* $\mathbf{A}_1$ *and* $\mathbf{A}_2$ *is given by*

$$\mathbf{A}_1 = \mathbf{U}_1\mathbf{T}_1\mathbf{V}^\dagger, \tag{12a}$$

$$\mathbf{A}_2 = \mathbf{U}_2\mathbf{T}_2\mathbf{V}^\dagger, \tag{12b}$$

where $\mathbf{U}_1$*,* $\mathbf{U}_2$ *and* $\mathbf{V}$ *are unitary matrices of dimensions* $M_1 \times M_1$*,* $M_2 \times M_2$ *and* $N \times N$*, respectively, and* $\mathbf{T}_1$ *and* $\mathbf{T}_2$ *are generalized upper-triangular matrices (recall Definition 3) with a prescribed set of diagonal ratios* $\boldsymbol{t}$*, i.e.,*

$$\begin{aligned} \frac{T_{1;ii}}{T_{2;ii}} &= t_i\,, & & & i &= 1, \ldots, N\,, \\ T_{k;i,j} &= 0\,, & k &= 1, 2\,, & \forall i &> j\,. \end{aligned}$$

Such a joint decomposition exists if and only if the vector $\boldsymbol{t}$ *is majorized by the GSV vector* $\boldsymbol{\mu}(\mathbf{A}_1, \mathbf{A}_2)$*:*

$$\boldsymbol{\mu}(\mathbf{A}_1, \mathbf{A}_2) \succeq \boldsymbol{t}\,. \tag{13}$$

In other words, the GSVs are an extremal case for the diagonal ratios of all possible joint unitary triangularizations. The joint unitary decomposition that corresponds to these extremal values is the GSVD.

Following the exposition in [37], [38], we next review the two forms of the GSVD — diagonal and triangular. The diagonal representation of the GSVD is better known. For a matrix pair $(\mathbf{A}_1, \mathbf{A}_2)$ it is given by [28], [36]:

$$\mathbf{A}_1 = \mathbf{U}_1\mathbf{D}_1\mathbf{X}^\dagger, \tag{14a}$$

$$\mathbf{A}_2 = \mathbf{U}_2\mathbf{D}_2\mathbf{X}^\dagger, \tag{14b}$$

where $\mathbf{U}_1$ and $\mathbf{U}_2$ are unitary, $\mathbf{X}$ is invertible, and $\mathbf{D}_1$ and $\mathbf{D}_2$ are generalized diagonal matrices (viz., $D_{k;i,j} = 0$ for $i \neq j$, where $D_{k;i,j}$ is the (i, j) entry of D_k) with positive diagonal values satisfying:

$$\mathbf{D}_1^\dagger\mathbf{D}_1 + \mathbf{D}_2^\dagger\mathbf{D}_2 = \mathbf{I}, \tag{15}$$

the ratios of which are equal to the GSVs:

$$\frac{D_{1;ii}}{D_{2;ii}} = \mu_i\left(\mathbf{A}_1, \mathbf{A}_2\right), \qquad i = 1, \ldots, N,$$

and are assumed, w.l.o.g., to be ordered non-increasingly. To obtain the triangular form of the GSVD, apply a QL decomposition[3] to $\mathbf{X}$, to attain:

$$\begin{aligned} \mathbf{A}_1 &= \mathbf{U}_1\mathbf{D}_1\mathbf{T}\mathbf{V}^\dagger \\ &\triangleq \mathbf{U}_1\mathbf{T}_1\mathbf{V}^\dagger, \qquad (16a) \\ \mathbf{A}_2 &= \mathbf{U}_2\mathbf{D}_2\mathbf{T}\mathbf{V}^\dagger \\ &\triangleq \mathbf{U}_2\mathbf{T}_2\mathbf{V}^\dagger, \qquad (16b) \end{aligned}$$

where $\mathbf{T}$ is upper triangular and $\mathbf{V}$ is unitary. By denoting $\mathbf{T}_1 \triangleq \mathbf{D}_1\mathbf{T}$ and $\mathbf{T}_2 \triangleq \mathbf{D}_2\mathbf{T}$, we attain the triangular form of the GSVD, which is, in turn, a special case of (12).

III. The MIMO Wiretap Capacity Revisited

In this section we re-derive the explicit capacity expression of Bustin *et al.* [7] for the MIMO wiretap channel under a covariance constraint (3) in terms of the GSVD. While we do not establish a new capacity result, our approach of simultaneous unitary triangularization will lead to a simplified representation of the optimal covariance matrix as well as layered coding schemes, as will be discussed in the subsequent sections.

The following augmented matrix structure, which serves as the MIMO channel analogue of the minimum mean square error (MMSE) variant of decision feedback equalization for linear time-invariant systems [39], will be instrumental throughout this work.

Definition 5 (Effective MMSE channel matrix)**.** Let $\mathbf{H}$ be a channel matrix of dimensions $N_B \times N_A$ and let $\mathbf{K}$ be the $N_A \times N_A$ input covariance matrix used over this channel. Then, the corresponding *effective MMSE channel matrix* is the $(N_A + N_B) \times N_A$ matrix

$$\mathbf{G}\left(\mathbf{H}, \mathbf{K}\right) \triangleq \begin{pmatrix} \mathbf{H}\mathbf{K}^{1/2} \\ \mathbf{I} \end{pmatrix}, \qquad (17)$$

where $\mathbf{I}$ is the identity matrix of dimension N_A and $\mathbf{K}^{1/2}$ is any matrix $\mathbf{B}$ satisfying $\mathbf{B}\mathbf{B}^\dagger = \mathbf{K}$.[4]

This definition naturally lends itself to an MMSE (capacity-achieving) variant of the V-BLAST/GDFE scheme [24], as will be described in Section IV. See also [40], [19], [26] for further explanations.

Construct the effective MMSE matrices $\mathbf{G}_B = \mathbf{G}(\mathbf{H}_B, \mathbf{K})$ and $\mathbf{G}_E = \mathbf{G}(\mathbf{H}_E, \mathbf{K})$, where $\mathbf{K}$ is subject to the constraining matrix $\bar{\mathbf{K}}$ (3): $\mathbf{K} \preceq \bar{\mathbf{K}}$.

Now, apply some joint unitary triangularization (11):

$$\mathbf{G}_B = \mathbf{U}_B\mathbf{T}_B\mathbf{V}_A^\dagger, \qquad (18a)$$

$$\mathbf{G}_E = \mathbf{U}_E\mathbf{T}_E\mathbf{V}_A^\dagger, \qquad (18b)$$

where $\mathbf{U}_B$, $\mathbf{U}_E$ and $\mathbf{V}_A$ are unitary, and $\mathbf{T}_B$ and $\mathbf{T}_E$ are generalized upper triangular (recall Definition 3).

Let $\{b_i\}$ and $\{e_i\}$ denote the diagonal values of $\mathbf{T}_B$ and $\mathbf{T}_E$, respectively, where, as explained in Section II-B, these values can be designed by varying $\mathbf{V}_A$. Using the fact that the absolute value of a determinant of a unitary matrix is equal to 1, and the fact that the determinant of a triangular matrix is equal to the product of its diagonal values, the Gaussian MI (6) can be expressed as:

$$\begin{aligned} I(\mathbf{H}_B, \mathbf{K}) &= \log\left|\mathbf{G}_B^\dagger\mathbf{G}_B\right| \qquad (19a) \\ &= \sum \log b_i^2, \qquad (19b) \end{aligned}$$

and similarly for Eve:

$$\begin{aligned} I(\mathbf{H}_E, \mathbf{K}) &= \log\left|\mathbf{G}_E^\dagger\mathbf{G}_E\right| \\ &= \sum \log e_i^2. \end{aligned}$$

Hence, their difference (5) is given by

$$I_S(\mathbf{H}_B, \mathbf{H}_E, \mathbf{K}) = \sum_{i=1}^{N_A} \log \frac{b_i^2}{e_i^2}. \qquad (20)$$

Note that the expression in (20) holds for any unitary matrix $\mathbf{V}_A$ in (18). Indeed, as we shall see later, this flexibility in choosing $\mathbf{V}_A$ can lead to different design tradeoffs in our layered coding schemes. Nevertheless, to derive an explicit capacity expression we specialize $\mathbf{V}_A$ to be the right unitary matrix of the GSVD (16), until the end of the section. The corresponding GSVs are hence equal to

$$\begin{aligned} \mu_i\left(\mathbf{H}_B, \mathbf{H}_E, \mathbf{K}\right) &\triangleq \mu_i\left(\mathbf{G}_B, \mathbf{G}_E\right) \\ &= \frac{b_i}{e_i}, \end{aligned}$$

where we use the notation $\mu_i\left(\mathbf{H}_B, \mathbf{H}_E, \mathbf{K}\right)$ to emphasize the dependence in $\mathbf{K}$. Without loss of generality, we assume that the GSV vector is non-increasing.

In terms of the GSVs, we can rewrite (4) as:

$$C_S(\mathbf{H}_B, \mathbf{H}_E, \bar{\mathbf{K}}) = \max_{\mathbf{K} \preceq \bar{\mathbf{K}}} \sum_{i=1}^{N_A} \log \mu_i^2\left(\mathbf{H}_B, \mathbf{H}_E, \mathbf{K}\right).$$

Indeed, in these terms the MIMO wiretap capacity can be expressed as follows.

Theorem 3 (MIMO wiretap capacity under a covariance constraint [7])**.** *The secrecy capacity under a covariance matrix constraint $\bar{\mathbf{K}}$ is given by*

$$\begin{aligned} C_S(\mathbf{H}_B, \mathbf{H}_E, \bar{\mathbf{K}}) &= \sum_{i=1}^{N_A} \left[\log \mu_i^2\left(\mathbf{H}_B, \mathbf{H}_E, \bar{\mathbf{K}}\right)\right]_+ \qquad (21a) \\ &= \sum_{i=1}^{L_B} \log \mu_i^2\left(\mathbf{H}_B, \mathbf{H}_E, \bar{\mathbf{K}}\right). \qquad (21b) \end{aligned}$$

This explicit capacity expression along with the optimal covariance matrix $\mathbf{K} \preceq \bar{\mathbf{K}}$ were established by Bustin *et al.* [7] using the channel enhancement technique along with vector extensions of the mutual information–minimum mean-square error (I–MMSE) relation. We present an alternative proof of

[3]This decomposition is similar to the QR decomposition, only instead of an upper-triangular matrix, the resulting matrix is lower triangular. This can be achieved, e.g., by applying Gram–Schmidt triangularization to the columns of a matrix, from last to first.

[4]Such a $\mathbf{B}$ can always be constructed, e.g., using the Cholesky decomposition or unitary diagonalization.

this result using a direct approach: once the optimization problem (4) is stated, it can be solved by linear algebra and elementary calculus only. The key to our proof is the following lemma.

Lemma 1. *Let* $\bar{\mathbf{K}}$ *and* $\mathbf{K}$ *be two matrices satisfying* $\mathbf{0} \preceq \mathbf{K} \preceq \bar{\mathbf{K}}$. *Then for all* $i = 1, \ldots, N_A$,

$$\left|\log \mu_i(\mathbf{H}_B, \mathbf{H}_E, \bar{\mathbf{K}})\right| \geq \left|\log \mu_i(\mathbf{H}_B, \mathbf{H}_E, \mathbf{K})\right|.$$

That is, as we "decrease" the input covariance, the GSVs move towards $\mu_i = 1$. The proof, which appears in Appendix A, uses standard matrix calculus to show that the differential of the i-th GSV, $d\mu_i$, with respect to a change in the covariance matrix $d\mathbf{K}$, is given by

$$d\mu_i = \left(\mu_i^2 - 1\right) \cdot \gamma_i(d\mathbf{K}),$$

where $\gamma_i(d\mathbf{K}) \geq 0$ for $d\mathbf{K} \succeq \mathbf{0}$. Or to put it differently, $d\mu_i > 0$ for $\mu_i > 1$, and $d\boldsymbol{\mu}_i < 0$ for $\mu_i < 1$.

By Lemma 1, clearly Theorem 3 gives an upper bound on the capacity. To see that it is achievable, consider the matrix:

$$\mathbf{K} = \bar{\mathbf{K}}^{1/2} \mathbf{V}_A \mathbf{I}_B \mathbf{V}_A^{\dagger} \bar{\mathbf{K}}^{1/2\dagger}, \tag{22}$$

where $\mathbf{V}_A$ is the right unitary matrix of the triangular form of the GSVD (16), $\mathbf{I}_B$ is a diagonal matrix whose first L_B diagonal values (corresponding to GSVs that are greater than 1) are equal to 1, and the remaining L_E — to 0. Trivially, $\mathbf{K} \preceq \bar{\mathbf{K}}$. The choice of $\mathbf{K}$ effectively truncates the GSVs of $\bar{\mathbf{K}}$:

$$\log \mu_i^2(\mathbf{H}_B, \mathbf{H}_C, \mathbf{K}) = \left[\log \mu_i^2\left(\mathbf{H}_B, \mathbf{H}_C, \bar{\mathbf{K}}\right)\right]_+.$$

This is formally proved in Appendix B.

Remark 1. The optimal covariance matrix $\mathbf{K}$ (22) is denoted by $\mathbf{K}_x^*$ in [7], where it is given in terms of the diagonal form of the GSVD (14):[5]

$$\mathbf{K} = \bar{\mathbf{K}}^{1/2} \mathbf{Y} \begin{bmatrix} \left(\mathbf{Y}_B^{\dagger} \mathbf{Y}_B\right)^{-1} & \mathbf{0}_{L_B \times L_E} \\ \mathbf{0}_{L_E \times L_B} & \mathbf{0}_{L_E \times L_E} \end{bmatrix} \mathbf{Y}^{\dagger} \bar{\mathbf{K}}^{\dagger/2}, \tag{24}$$

where $\mathbf{Y} = \mathbf{X}^{-\dagger}$ and $\mathbf{X}$ is the right invertible matrix of (14), $\mathbf{Y}_B$ is the sub-matrix composed of the first L_B columns of $\mathbf{Y}$, and $\mathbf{0}_{m \times n}$ denotes the all-zero matrix of dimensions $m \times n$. Comparing (22) and (24), it is evident that using the triangular form of the GSVD indeed simplifies the representation over using the diagonal one.

Remark 2. One may wonder why, of all possible choices of $\mathbf{V}_A$, the capacity is given in terms of the GSVD. An intuitive reason is as follows. By the majorization condition (13), the GSV vector is extremal among all possible diagonals. In particular, for any $\mathbf{V}_A$,

$$\sum_{i=1}^{N_A} \left[\log \mu_i^2\right]_+ \geq \sum_{i=1}^{N_A} \left[\log \frac{b_i^2}{e_i^2}\right]_+.$$

Thus, the sum (21a) is larger than the sum over diagonal ratios induced by other triangular decompositions.

[5] In [7] a specific choice of $\mathbf{K}^{1/2}$ was used: the matrix $\mathbf{B}$ that satisfies $\mathbf{BB} = \mathbf{K}$.

Remark 3. Using (7), the capacity of the MIMO wiretap channel under a power constraint P can be written as

$$C_S(\mathbf{H}_B, \mathbf{H}_C, P) = \max_{\mathbf{K}:\mathrm{trace}\{\mathbf{K}\}=P} \sum_{i=1}^{N_A} \left[\log \mu_i^2(\mathbf{H}_B, \mathbf{H}_C, \mathbf{K})\right]_+.$$

Remark 4. For the optimal $\mathbf{K}$ (22), all the GSVs are greater or equal to 1. To the contrary, assume that some are strictly smaller than 1; then, we can use a matrix $\mathbf{K}$ with the appropriate directions "nullified". Such a "truncated" matrix will satisfy the covariance constraint while improving the achievable secrecy rate of the scheme, in contradiction to the assumption. *A fortiori*, under a power constraint, the power saved by such a truncation can be allocated to "useful" directions.

IV. 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 [19], [26], [40].

In this work we shall assume all the scalar codes to be Gaussian, as defined next.

Definition 6 (Gaussian codebook)**.** A Gaussian codebook of length n, rate R and power $P - \epsilon$, where $\epsilon > 0$, consists of $\lceil 2^{nR} \rceil$ codewords of length n, denoted by $x^n(1), x^n(2), \ldots, x^n\left(\lceil 2^{nR} \rceil\right)$. The entries of all the codewords, $\{x_t(i) | t = 1, \ldots, n; i = 1, \ldots, \lceil 2^{nR} \rceil\}$, are i.i.d. with respect to a Gaussian distribution with zero mean and variance $P - \epsilon$.

Remark 5. In the sequel, with a slight abuse of notation, we shall refer to such codes as Gaussian codes of power P (where ϵ will serve as an implicit design parameter).

Consider the channel (2a). Construct the effective MMSE matrix $\mathbf{G}_B = \mathbf{G}(\mathbf{H}_B, \mathbf{K})$ as in Definition 5, and choose some unitary matrix $\mathbf{V}_A$.

Apply the GTD (11) to $\mathbf{G}_B$ with $\mathbf{V}_A$ as the right matrix:

$$\mathbf{G}_B = \mathbf{U}_B \mathbf{T}_B \mathbf{V}_A^{\dagger}. \tag{25}$$

Now let $\tilde{\boldsymbol{x}}$ be a vector of standard Gaussian variables, and set

$$\boldsymbol{x} = \mathbf{K}^{1/2} \mathbf{V}_A \tilde{\boldsymbol{x}}. \tag{26}$$

Denote by $\tilde{\mathbf{U}}_B$ the $N_B \times N_A$ upper-left sub-matrix of $\mathbf{U}_B$, and define

$$\tilde{\mathbf{T}}_B = \tilde{\mathbf{U}}_B^{\dagger} \mathbf{H}_B \mathbf{K}^{1/2} \mathbf{V}_A. \tag{27}$$

The following lemma, whose proof can be found in [24], [40, Lemma III.3], [41, Appendix I], provides the connection between the elements of $\mathbf{T}_B$ and $\tilde{\mathbf{T}}_B$.

Lemma 2. *Denote by* $[\mathbf{T}_B]$ *the* $N_A \times N_A$ *upper-triangular sub-matrix composed of the first* N_A *rows of* $\mathbf{T}_B$ (25).[6] *Then,* $\tilde{\mathbf{T}}_B$ (27) *is equal to*

$$\tilde{\mathbf{T}}_B = [\mathbf{T}_B] - [\mathbf{T}_B]^{-\dagger}.$$

[6] Since $\mathbf{T}_B$ is full rank, $[\mathbf{T}_B]$ is full rank too, and hence also invertible. Further, its diagonal elements are greater or equal to 1 due to the block $\mathbf{I}$ in the construction of $\mathbf{G}_B$.

In particular,

$$\tilde{T}_{B;i,j} = \begin{cases} T_{B;i,j} & i < j \\ T_{B;i,j} - 1/T_{B;i,j} & i = j \end{cases} \tag{28}$$

where $T_{B;i,j}$ and $\tilde{T}_{B;i,j}$ are the (i,j) entries of the matrices $\mathbf{T}_B$ and $\tilde{\mathbf{T}}_B$, respectively.

Let

$$\tilde{\boldsymbol{y}}_B = \tilde{\mathbf{U}}_B^\dagger \boldsymbol{y}_B \tag{29a}$$
$$= \tilde{\mathbf{U}}_B^\dagger \mathbf{H}_B \mathbf{K}^{1/2} \mathbf{V}_A \tilde{\boldsymbol{x}} + \tilde{\mathbf{U}}_B^\dagger \boldsymbol{z}_B \tag{29b}$$
$$= \tilde{\mathbf{T}}_B \tilde{\boldsymbol{x}} + \tilde{\boldsymbol{z}}_B \,. \tag{29c}$$

Since $\tilde{\mathbf{U}}_B$ is not unitary, the statistics of $\tilde{\boldsymbol{z}}_B \triangleq \tilde{\mathbf{U}}_B^\dagger \boldsymbol{z}_B$ differ from those of $\boldsymbol{z}_B$, and its covariance matrix is given by $\mathbf{K}_{\tilde{\boldsymbol{z}}_B} \triangleq \tilde{\mathbf{U}}_B \tilde{\mathbf{U}}_B^\dagger$. Now, for $i = 1, \ldots, N_A$, define [recall (28)]

$$y'_{B;i} = \tilde{y}_{B;i} - \sum_{\ell=i+1}^{N_A} T_{B;i,\ell} \tilde{x}_\ell \tag{30a}$$
$$= \tilde{T}_{B;i,i} \tilde{x}_i + \sum_{\ell=1}^{i-1} \tilde{T}_{B;i,\ell} \tilde{x}_\ell + \tilde{z}_{B;i} \tag{30b}$$
$$\triangleq \tilde{T}_{B;i,i} \tilde{x}_i + z_{B;i}^{\text{eff}} \,, \tag{30c}$$

$\tilde{z}_{B;i}$ and $z_{B;i}^{\text{eff}}$ are the i-th entries of the vectors $\tilde{\boldsymbol{z}}_B$ and $\boldsymbol{z}_B^{\text{eff}}$, respectively, and $z_{B;i}^{\text{eff}} \triangleq \sum_{\ell=1}^{i-1} \tilde{T}_{B;i,\ell} \tilde{x}_\ell + \tilde{z}_{B;i}$ is the resulting total effective noise vector.

In this scalar channel from $\tilde{x}_i$ to $y'_{B;i}$, resulting after the subtraction of the previously recovered symbols $\{\tilde{x}_\ell | \ell > i\}$, we view the remaining symbols $\{\tilde{x}_\ell | \ell < i\}$ as "interference", $\tilde{z}_{B;i}$ — as "noise", and their sum $z_{B;i}^{\text{eff}}$ — as "effective noise". The resulting signal-to-interference-and-noise ratio (SINR) is given by:

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

where $K_{\boldsymbol{z}_B^{\text{eff}};i,j}$ and $K_{\tilde{\boldsymbol{z}}_B;i,j}$ denote the (i,j) entries of $\mathbf{K}_{\tilde{\boldsymbol{z}}_B}$ and $\mathbf{K}_{\boldsymbol{z}_B^{\text{eff}}}$, respectively. The following key result achieves the mutual information [24], [40, Lemma III.3], [41, Appendix I] and is based on Lemma 2.[7]

$$I\left(\tilde{x}_i; \boldsymbol{y}_B \middle| \tilde{x}_{i+1}^{N_A}\right) = I\left(\tilde{x}_i; y'_{B;i}\right) \tag{31a}$$
$$= \log(1 + \text{SINR}_{B;i}) \tag{31b}$$
$$= \log\left(b_i^2\right), \tag{31c}$$

where $\{b_i\}$ are the diagonal values of $\mathbf{T}_B$ (25) [mind the difference from the diagonal values of $\tilde{\mathbf{T}}_B$ (28)], which satisfy

$$b_i^2 = 1 + \text{SINR}_{B;i} \tag{32}$$

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

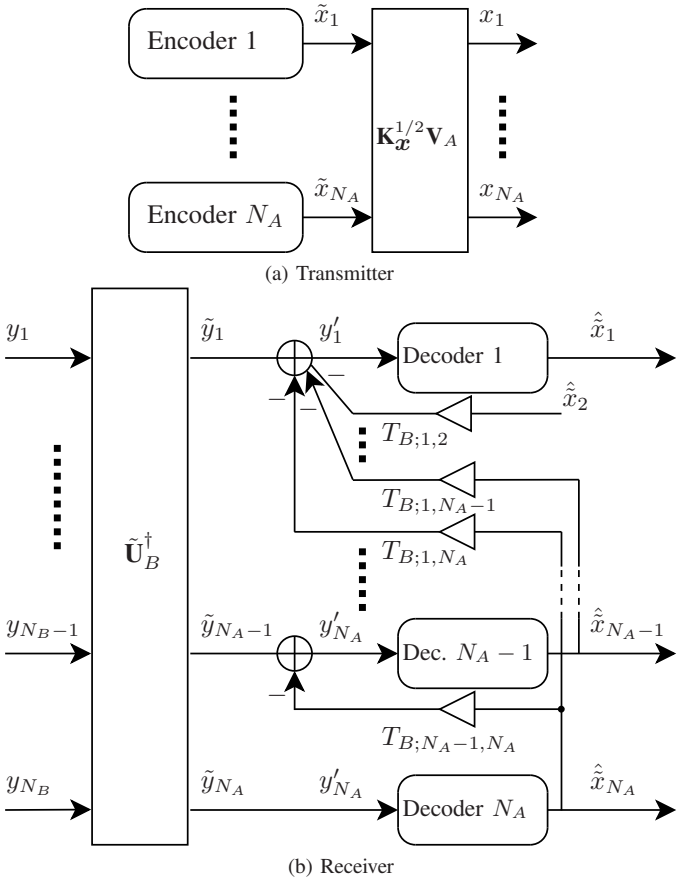


Fig. 1: Layered-SIC scheme. $\hat{\tilde{x}}_\ell$ denotes the decoded symbol $\hat{x}_\ell$ at the receiver.

and

$$\sum_{i=1}^{N_A} \log\left(b_i^2\right) = \sum_{i=1}^{N_A} \log\left(1 + \text{SINR}_{B;i}\right) = I(\mathbf{H}_B, \mathbf{K}),$$

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

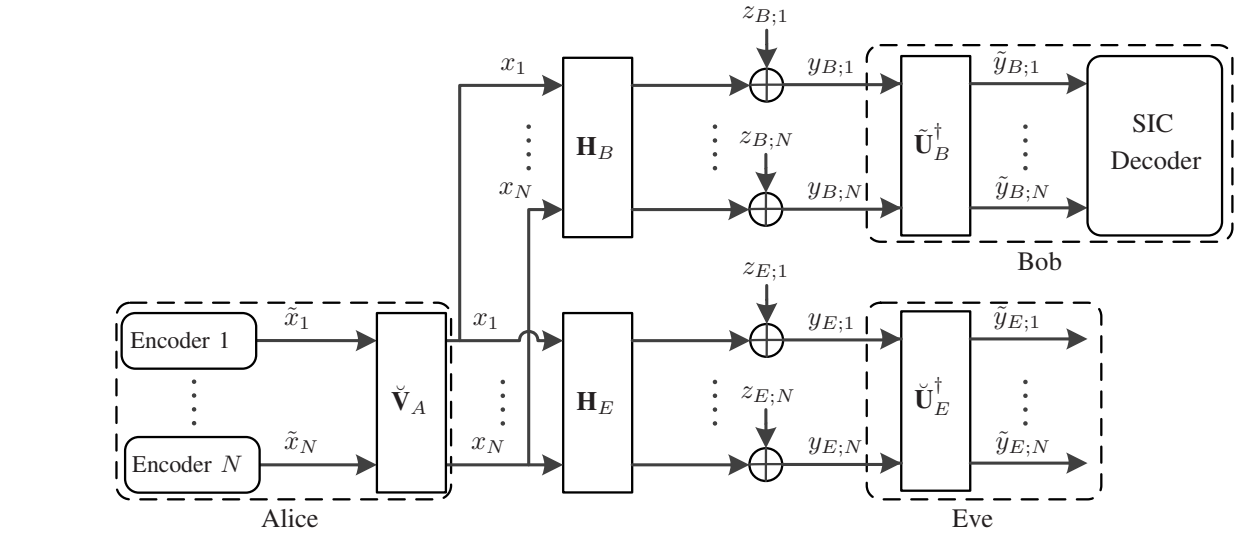
The analysis above immediately gives rise to the following scheme, depicted also in Fig. 1, which is, in turn, a variant of the renowned V-BLAST/GDFE scheme [21]–[24].

Scheme (Layered-SIC)**.**

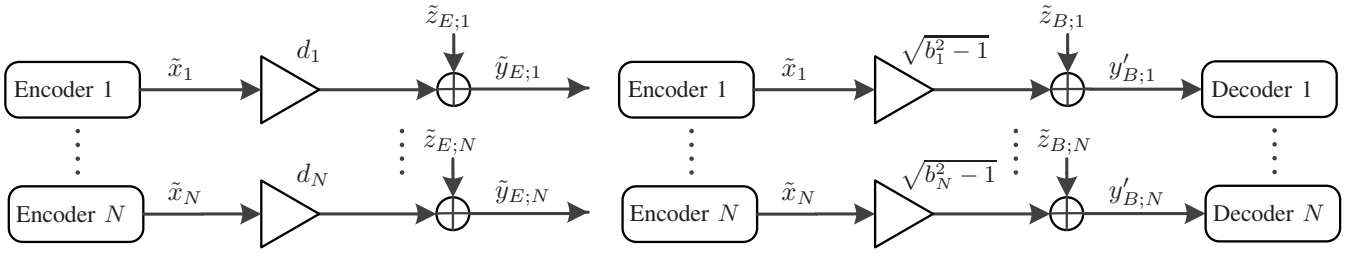
Offline:

- Select an admissible $N_A \times N_A$ input covariance matrix $\mathbf{K}$ that satisfies the input constraint.[8]
- Construct the effective MMSE matrix (17): $\mathbf{G}_B = \mathbf{G}(\mathbf{H}_B, \mathbf{K})$.
- Select a unitary triangularization (11) and apply it to the matrix $\mathbf{G}_B$, as in (25), to obtain the unitary matrices $\mathbf{U}_B$ and $\mathbf{V}_A$, and the generalized upper-triangular matrix $\mathbf{T}_B$.
- Denote the N_A diagonal elements of $\mathbf{T}_B$ by $\{b_i\}$.
- Denote by $\tilde{\mathbf{U}}_B$ the $N_B \times N_A$ upper-left sub-matrix of $\mathbf{U}_B$, and construct the corresponding matrix $\tilde{\mathbf{T}}_B$ according to (27): $\tilde{\mathbf{T}}_B = \tilde{\mathbf{U}}_B^\dagger \mathbf{H}_B \mathbf{K}^{1/2} \mathbf{V}_A$.

[8]More generally, any number $N \geq \text{rank}\{\mathbf{K}\}$ of scalar codebooks can be used; see [40], [19] for details.



(a) Full scheme and channel



(b) Equivalent channel to Eve. $\tilde{z}_{E;1}, \ldots, \tilde{z}_{E;N}$ are independent unit power AWGNs.

(c) Equivalent channel to Bob assuming correct past decision at the SIC decoder (which is achieved for a large enough blocklength n).

Fig. 2: Layered-SIC scheme for the MIMO wiretap channel. We assume here $N = N_A = N_B = N_E$, for ease of presentation.

- Construct N_A scalar *Gaussian codes* of length n and unit power that are good for SNRs $\{b_i^2 - 1\}$, i.e., codes of rates close to

$$\left\{R_i \middle| R_i = \log\left(b_i^2\right),\, i \in \{1, \ldots, N_A\}\right\}. \tag{33}$$

Alice: At each time instant $t = 1, \ldots, n$:

- Forms the vector $\tilde{\boldsymbol{x}}$ of length N_A, by taking one sample from each codebook.
- Attains the vector $\boldsymbol{x}$ by multiplying $\tilde{\boldsymbol{x}}$ by $\mathbf{V}_A$ and $\mathbf{K}^{1/2}$:

$$\boldsymbol{x} = \mathbf{K}^{1/2}\mathbf{V}_A\tilde{\boldsymbol{x}}. \tag{34}$$

- Transmits $\boldsymbol{x}$.

Bob:

- At each time instant $t = 1, \ldots, n$, receives $\boldsymbol{y}_B$ and forms $\tilde{\boldsymbol{y}}_B$ according to (29):

$$\begin{aligned}\tilde{\boldsymbol{y}}_B &= \tilde{\mathbf{U}}_B^\dagger \boldsymbol{y}_B \\ &= \tilde{\mathbf{T}}_B\tilde{\boldsymbol{x}} + \tilde{\boldsymbol{z}}_B.\end{aligned}$$

- Decodes the n-length codewords using SIC, from last ($i = N_A$) to first ($i = 1$): Assuming correct decoding of all codebooks $i+1, \ldots, N_A$, Bob forms $y'_{B;i}$ (30):

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

and recovers $\tilde{x}_i$.

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

Remark 6. The SIC procedure and the performance analysis of the scheme implicitly assume that the yet-undecoded codebooks can be considered as AWGN, and consequently that each codebook should be capacity achieving for an AWGN channel. This is indeed true for Gaussian codes (recall Definition 6) but not for any single-user scalar capacity-achieving codes as is discussed in Section VIII.

V. Multi-Stream Schemes for the MIMO Wiretap Channel

Equipped with the results presented in the previous sections, we describe how to construct multi-stream schemes that achieve the capacity of the MIMO wiretap channel.

We first describe a scheme in which the channel to Eve is effectively diagonalized, in Section V-A. This particular choice facilitates the proof of both weak and strong secrecy guarantees over this channel. We then extend this result in Section V-B, by proving that any joint triangularization (12) can be used to construct a multi-stream capacity-achieving scheme.

A. Orthogonalizing Eve's Channel

We now present a simple adaptation of the layered-SIC scheme of Section IV to the MIMO wiretap setting, depicted

also in Fig. 2, that achieves the secrecy capacity of the channel using scalar wiretap codes.

To this end, we note that the layered-SIC scheme is capacity-achieving (without secrecy constraints) for any choice of $\mathbf{V}_A$ in (25). In particular, we can choose this matrix to be the unitary matrix that diagonalizes Eve's effective channel matrix, namely, the right matrix of the SVD of Eve, denoted by $\breve{\mathbf{V}}_A$:

$$\mathbf{H}_E\mathbf{K}^{1/2} = \breve{\mathbf{U}}_E\breve{\mathbf{D}}_E\breve{\mathbf{V}}_A^\dagger. \tag{35}$$

Applying this $\breve{\mathbf{V}}_A$ to $\mathbf{H}_E$ (followed by $\mathbf{K}^{1/2}$) provides effective parallel scalar independent channels to Eve, of SNRs $\{d_i^2\}$, where $\{d_i\}$ are the diagonal values of $\mathbf{D}_E$, which constitute the singular values of $\mathbf{H}_E\mathbf{K}^{1/2}$.

The following simple lemma summarizes the connection between the SVDs of the effective channel matrix (35) and the effective MMSE channel matrix $\mathbf{G}_E(\mathbf{H}_E, \mathbf{K})$.

Lemma 3 (Connection to effective MMSE matrix)**.** *The SVD of the effective MMSE matrix $\mathbf{G}_E = \mathbf{G}(\mathbf{H}_E, \mathbf{K})$ (recall Definition 5) is given by*

$$\mathbf{G}_E = \mathbf{U}_E\mathbf{D}_E\mathbf{V}_A, \tag{36}$$

where $\mathbf{D}_E$ is a generalized diagonal matrix (viz., $D_{E;i,j} = 0$ for $i \neq j$); denote its diagonal elements by $\{e_i\}$.

The SVD of $\mathbf{G}_E$ (36) is connected to the SVD of $\mathbf{H}_E\mathbf{K}^{1/2}$ (35) as follows. Define $d_i = 0$ for $i > N_A$, and note that $e_i = 1$ 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) *$\breve{\mathbf{V}}_A = \mathbf{V}_A$, i.e., $\mathbf{G}_E$ and $\mathbf{H}_E\mathbf{K}^{1/2}$ are diagonalized by the same right matrix.*
2) *$1 + d_i^2 = e_i^2\,, \qquad i = 1, \ldots, N_A$.*
3) *$\tilde{\mathbf{U}}_E = \breve{\mathbf{U}}_E\Lambda_E$, where $\tilde{\mathbf{U}}_E$ is the $N_E \times N_A$ upper-left submatrix of $\mathbf{U}_E$.*

The respective decomposition of $\mathbf{G}_B$ is as in (25), where the diagonal values of the resulting generalized triangular matrix $\mathbf{T}_B$ are $\{b_i\}$.

Since Eve observes parallel independent channels, using scalar wiretap codes over these channels, that are matched to the SNRs to Eve, $\{d_i^2\}$, guarantees the secrecy of the scheme. Moreover, by using wiretap codes that work with respect to the SNRs to Bob of (32), the secrecy capacity is achieved. This is formally stated in the following theorem.

Theorem 4. *The layered-SIC scheme of Section IV achieves the secrecy capacity under a covariance constraint $C_S\left(\mathbf{H}_B, \mathbf{H}_E\,\bar{\mathbf{K}}\right)$ by using:*

- *The optimal input covariance matrix $\mathbf{K}$ of (22).*
- *Choosing $\mathbf{V}_A$ of the SVD of $\mathbf{H}_E\mathbf{K}^{1/2}$ (35).*
- *Scalar Gaussian capacity-achieving wiretap codes that are designed for the Bob–Eve SNR-pairs $\left\{\left(b_i^2 - 1, d_i^2\right)\right\}$.*

Proof: The proof easily follows by noting that the resulting channel to Eve is diagonal, i.e., parallel scalar AWGN channels. Hence, by using independent (wiretap) Gaussian codes, secrecy is guaranteed over the parallel channels. By combining the result of Section IV for SIC for MIMO channels without secrecy, correct decoding at Bob's end is guaranteed.

Codebook construction: N_A Gaussian codebooks $\{\mathcal{C}_k | k = 1, \ldots, N_A\}$ of length n are generated independently, as in Definition 6. Codebook $\mathcal{C}_k$ contains $\left\lceil 2^{nR_k}\right\rceil \times \left\lceil 2^{n\tilde{R}_k}\right\rceil$ codewords. Each codeword within $\mathcal{C}_k$ is assigned a unique index pair (m_k, f_k), where $m_k \in \left\{1, \ldots, \left\lceil 2^{nR_k}\right\rceil\right\}$ and $f_k \in \left\{1, \ldots, \left\lceil 2^{n\tilde{R}_k}\right\rceil\right\}$. With a slight abuse of notation, we shall refer to such codes as wiretap Gaussian codes of rate-pairs $\left\{\left(R_k, \tilde{R}_k\right)\right\}$.

Let $\epsilon > 0$. Then the rates are chosen as[9]

$$R_k = \log\frac{b_k^2}{1 + d_k^2} - 2\epsilon = \log\frac{b_k^2}{e_k^2} - 2\epsilon, \tag{37a}$$

$$\tilde{R}_k = \log(1 + d_k^2) + \epsilon = \log e_k^2 + \epsilon. \tag{37b}$$

Encoding (Alice): Constructs N_A codewords $\{\tilde{x}_k \in \mathcal{C}_k | k = 1, \ldots, N_A\}$ as follows. $\tilde{x}_k$ is chosen from $\mathcal{C}_k$ according to the sub-message m_k intended to Bob and a fictitious sub-message f_k which is chosen uniformly at random. The transmitted signal at every time instant, $\boldsymbol{x}$, is then constructed as in the layered-SIC scheme of Section IV.

Decoding (Bob): Bob performs SIC decoding as in the layered-SIC scheme of Section IV to recover $\{(m_k, f_k)\}$, and discards $\{f_k\}$. Since $R_k + \tilde{R}_k < \log b_k^2$ for every k, the decoding error probability of Bob can be made arbitrarily small by taking a large enough n.

Secrecy analysis (Eve): The resulting channel to Eve (35) (depicted also in Fig. 2b) is diagonal:

$$\tilde{\boldsymbol{y}}_E = \breve{\mathbf{D}}_E\tilde{\boldsymbol{x}} + \tilde{\boldsymbol{z}}_E,$$

where $\tilde{\boldsymbol{z}}_E$ is AWGN with zero mean and identity covariance matrix. That is, the effective channel to Eve comprises independent AWGN channels. Over the resulting scalar AWGN channels, wiretap Gaussian codes are known to attain strong secrecy [42], where $\tilde{R}_k$ is chosen to be (slightly) above the *channel resolvability*, i.e., $\tilde{R}_k = \log(1 + d_k^2) + \epsilon$ for $\epsilon > 0$. This is a stronger requirement, as opposed to the choice $\tilde{R}_k = \log(1 + d_k^2) - \epsilon$ for $\epsilon > 0$, which facilitates an easier proof of weak secrecy guarantees for this channel (see, e.g., [43, Ch. 22]).

Total rate: By using (20), (37a), the total rate is equal to

$$\begin{aligned} R &= \sum_{k=1}^{N_A} R_k \\ &= \sum_{k=1}^{N_A}\left(\log\frac{b_k^2}{e_k^2} - 2\epsilon\right) \\ &= I_S\left(\mathbf{H}_B, \mathbf{H}_E, \mathbf{K}\right) - 2N_A\epsilon. \end{aligned}$$

By choosing the optimal $\mathbf{K}$, and taking a large enough n, this rate can be made arbitrarily close to the secrecy capacity C_S while guaranteeing both weak and strong secrecy. ■

Remark 7. In the proofs to follow, with a slight abuse of notation, we shall state the sizes of the codebook without

[9]To establish weak secrecy, $\tilde{R}_k$ can be relaxed to $\tilde{R}_k = \log e_k^2 - \epsilon$. The choice in (37b) allows to establish strong secrecy, as is further explained in the sequel.

explicitly using the ceiling operation $\lceil\cdot\rceil$, as its effect becomes negligible for large values of n.

Remark 8. In the celebrated SVD-based scheme for MIMO channels of Telatar [20], the SVD is applied to the *physical* channel matrix $\mathbf{H} = \mathbf{U}\mathbf{D}\mathbf{V}_A^\dagger$. The transmitted signal is then formed according to (26), where the non-unitary matrix $\mathbf{K}^{1/2}$ (over the effective diagonal channel $\mathbf{D}$) is diagonal, with entries set by the water-filling solution. Thus, the SVD plays two roles: it serves both for reducing the coding task to that of coding over scalar channels and for constructing the optimal input covariance matrix.

In contrast, in (35) the SVD is applied to the *effective* channel matrix $\mathbf{H}_E\mathbf{K}^{1/2}$, which already includes the non-unitary "coloring" part $\mathbf{K}^{1/2}$. Thus, it is only used for reducing the coding task. This form is more general, in the sense that it allows for a choice of $\mathbf{K}$ that is not related to a diagonal decomposition of the channel, e.g., subject to individual power constraints, or where the target expression is different, e.g., an MI difference as in this work. Finally, note that the rate of (20) can be achieved using the proposed scheme, even if $\mathbf{K}$ is suboptimal (when exact calculation of the optimal $\mathbf{K}$ is hard).

B. General Multi-Stream Scheme

We next show that, in fact, secrecy capacity can be achieved using the layered-SIC scheme and scalar wiretap codes for any choice $\mathbf{V}_A$, and by this generalizing the result of Section V-A to transmission that is not necessarily orthogonal over Eve's channel. Specifically, we show that the secrecy capacity can be achieved using any joint triangularization of the effective MMSE channel matrices (18) (*any* unitary matrix $\mathbf{V}_A$ at the encoder). In the general case, Eve's resulting matrix is triangular and hence denoted by $\mathbf{T}_E$, as in (18b). The diagonal values of $\mathbf{T}_E$ are denoted by $\{e_i\}$. The resulting family of schemes includes two important special cases, discussed in Section V-C, in addition to the one introduced in Section V-A.

Theorem 5. *The layered-SIC scheme of Section IV achieves the secrecy capacity under a covariance constraint* $C_S\left(\mathbf{H}_B, \mathbf{H}_E\ \mathbf{K}\right)$ *by using:*

- *The optimal input covariance matrix* $\mathbf{K}$ *of* (22)*.*
- Any *joint unitary triangularization* (18)*.*
- *Scalar Gaussian capacity-achieving wiretap codes that are designed for the Bob–Eve SNR-pairs* $\left\{\left(b_i^2-1, e_i^2-1\right)\right\}$*, where* $\{b_i\}$ *and* $\{e_i\}$ *are defined as in Section III.*

We use the following result, proved in Appendix C, for the proof of this theorem, which extends beyond the Gaussian wiretap setting, for both the discrete and the continuous cases.

Proposition 1. *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* $\{\tilde{x}_i : i = 1, \ldots, N_A\}$ *and a scalar function* φ *such that*

$$x = \varphi\left(\tilde{x}_1, \ldots, \tilde{x}_{N_A}\right). \tag{38}$$

Then, for $\epsilon > 0$*, however small, and for any joint distribution* $p(\tilde{\mathsf{x}}_1, \ldots, \tilde{\mathsf{x}}_{N_A})$*, there exists a scheme which achieves weak secrecy, with the* k*-th codebook conveying a rate:*

$$R_k = I(\tilde{\mathsf{x}}_k; \mathsf{y}_B|\tilde{\mathsf{x}}_{k+1}^{N_A}) - I(\tilde{\mathsf{x}}_k; \mathsf{y}_E|\tilde{\mathsf{x}}_{k+1}^{N_A}) - \epsilon. \tag{39}$$

Remark 9. The secrecy-proof of this result uses a "genie-aided" argument: in the mutual information of the k-th codeword recovered by Eve, we provide all previous codewords $\{\tilde{\boldsymbol{x}}_\ell |\, \ell = k+1, \ldots, N_A\}$ as "genie", even though Eve cannot recover these messages. Bob, on the other hand, uses successive decoding to recover the messages. Thus, the allocation of rates $\{R_k\}$ in (39) guarantees that all the messages $(m_1, ..m_{N_A})$ remain jointly secured from the eavesdropper's channel output sequence.

Proof of Theorem 5: We specialize the general superposition coding framework of Proposition 1 to the linear encoder structure and independent Gaussian distributions of $(\tilde{\mathsf{x}}_1, \ldots, \tilde{\mathsf{x}}_{N_A})$. Use

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

in (38), where the vector $\tilde{\boldsymbol{x}}$ is composed of one symbol from each codebook: $\tilde{\boldsymbol{x}} = (\tilde{x}_1, \ldots, \tilde{x}_k)^T$.[10]

Each codebook is a scalar Gaussian wiretap codebook of average unit power. The achievable secrecy rate of codebook $k = 1, \ldots, N_A$ is given by (39):

$$\begin{aligned}R_k &= I\left(\tilde{\mathsf{x}}_k; \mathsf{y}_B \middle| \tilde{\mathsf{x}}_{k+1}^{N_A}\right) - I\left(\tilde{\mathsf{x}}_k; \mathsf{y}_E \middle| \tilde{\mathsf{x}}_{k+1}^{N_A}\right) - \epsilon && \text{(40a)}\\ &= I\left(\tilde{\mathsf{x}}_k; \mathsf{y}'_{B;k}\right) - I\left(\tilde{\mathsf{x}}_k; \mathsf{y}'_{E;k}\right) - \epsilon && \text{(40b)}\\ &= \log\left(b_k^2\right) - \log\left(e_k^2\right) - \epsilon && \text{(40c)}\\ &= \log\frac{b_k^2}{e_k^2} - \epsilon\,, && \text{(40d)}\end{aligned}$$

where (40c) and (40b) are due to (31a) and (31c), respectively. Thus, using the result of (20), we can achieve

$$\begin{aligned}R &= \sum_{k=1}^{N} R_k\\ &= \sum_{k=1}^{N}\left[\log\frac{b_k^2}{e_k^2}\right]_+ - \epsilon\\ &= I_S\left(\mathbf{H}_B, \mathbf{H}_E, \mathbf{K}\right),\end{aligned}$$

and for the optimal covariance matrix $\mathbf{K}$ the scheme approaches the secrecy capacity. ■

C. Important Special Cases

We now present "special" choices of $\mathbf{V}_A$ which provide various advantages.

1) Orthogonalizing Eve's channel: The scheme of Section V-A is a special case of proposed scheme in this subsection, since, as explained in Lemma 3, the unitary matrix $\mathbf{V}_A$ of the SVD of $\mathbf{H}_E\mathbf{K}^{1/2}$ is identical to that of the SVD of $\mathbf{G}_E$ (18b).

[10] Here, in contrast to Appendix C, boldface letters represent spatial vectors and time indices are suppressed.

2) Orthogonalizing Bob's channel — Avoiding SIC: Performing SIC adds complexity to the decoder, as well as introduces potential error propagation. We can avoid this by performing SVD with respect to Bob's channel, as opposed to Eve's channel, as done in Section V-A. That is, choose $\mathbf{V}_A$ such that

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

where $\mathbf{D}_B$ is diagonal. As happens with Eve in Section V-A, Bob obtains a diagonal equivalent channel, where each sub-stream can be decoded independently.

3) Avoiding individual bit-loading: When using (non-secret) communication schemes based on SVD or QR, as in the layered-SIC scheme, 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. By using the GMD, described in Section II-A, instead, a constant diagonal is achieved, which translates into equal SNRs for all parallel channels. This suggests, in turn, that bit-loading can be avoided altogether and that the codewords sent over the resulting sub-channels can be drawn from the same codebook.

A similar result can be achieved for the wiretap setting. To this end we require the usage of a modular scheme that transforms good AWGN codes of a rate close to $\log(b^2)$ for Bob into wiretap codes of rates close to $\{\log(b^2/e_i^2)\}$. This way, after applying the GMD to $\mathbf{G}_B$, the same AWGN codebook can be used over all sub-channels, where for each sub-channel a different transformation into a wiretap code is used, that depends on its effective SNR to Eve (e_i^2-1). Indeed, such a modular approach exists; see Section VIII.

Remark 10. It is possible to use the same wiretap code without assuming the modular wiretap code construction, by using a joint matrix decomposition that achieves constant diagonals for both triangular matrices simultaneously. A construction that essentially achieves this property was proposed in [26].

VI. Dirty-Paper Coding Based Schemes

In this section we construct the DPC counterparts of the layered-SIC scheme for Gaussian MIMO channels with and without secrecy constraints. In these variants the successive decoding process of the scalar codes is replaced with a successive encoding one; consequently, all (scalar) codebooks can be recovered in parallel and independently of each other. The latter makes these variants useful for more complex settings, such as the confidential MIMO broadcast setting treated in Section VII. We start by presenting the DPC-based schemes without secrecy constraints, in Section VI-A. We then construct a variant for the MIMO wiretap setting, in Section VI-B, which again achieves the secrecy capacity of the channel.

A. Without Secrecy Constraints

We now briefly review the DPC variant of the layered-SIC scheme, which is based in turn on [44], [45] (see also [40]).

Scheme (Layered-DPC)**.**

Offline:

- Select an admissible $N_A \times N_A$ input covariance matrix $\mathbf{K}$ that satisfies the input constraint.
- Construct the effective MMSE matrix (17): $\mathbf{G}_B = \mathbf{G}(\mathbf{H}_B, \mathbf{K})$.
- Select a unitary triangularization (11) and apply it to the matrix $\mathbf{G}_B$, as in (25), to obtain the unitary matrices $\mathbf{U}_B$ and $\mathbf{V}_A$, and the generalized upper-triangular matrix $\mathbf{T}_B$.
- Denote the N_A diagonal elements of $\mathbf{T}_B$ by $\{b_i\}$.
- Denote by $\tilde{\mathbf{U}}_B$ the $N_B \times N_A$ upper-left sub-matrix of $\mathbf{U}_B$, and construct the corresponding matrix $\tilde{\mathbf{T}}_B$ according to (27): $\tilde{\mathbf{T}}_B = \tilde{\mathbf{U}}_B^\dagger \mathbf{H}_B \mathbf{K}^{1/2} \mathbf{V}_A$.
- Construct N_A scalar *dirty-paper codes* [25] of length n — codes generated via random binning with respect to i.i.d. Gaussian distributions. Codebook i $(1 \le i \le N_A)$ is constructed for a channel with AWGN of unit power, SNR $(b_i^2 - 1)$, interference [recall (28)]

$$\sum_{\ell=i+1}^{N_A} T_{B;i,\ell} \tilde{x}_\ell \tag{41}$$

  which is available as side information at the transmitter, and rate R_i close to $\log(b_i^2)$ [recall (33)].

Alice: At each time instant $t = 1, \ldots, n$:

- Generates $\tilde{x}_i$ from last $(i = N_A)$ to first $(i = 1)$, where $\tilde{x}_i$ is generated according to the message to be conveyed and the interference (41).
- Forms $\tilde{\boldsymbol{x}}$ with entries $\{\tilde{x}_i\}$.
- Attains the vector $\boldsymbol{x}$ by multiplying $\tilde{\boldsymbol{x}}$ by $\mathbf{V}_A$ and $\mathbf{K}^{1/2}$ as in (34).
- Transmits $\boldsymbol{x}$.

Bob:

- At each time instant $t = 1, \ldots, n$, receives $\boldsymbol{y}_B$ and forms $\tilde{\boldsymbol{y}}_B$ according to (29):

$$\begin{aligned} \tilde{\boldsymbol{y}}_B &= \tilde{\mathbf{U}}_B^\dagger \boldsymbol{y}_B \\ &= \tilde{\mathbf{T}}_B \tilde{\boldsymbol{x}} + \tilde{\boldsymbol{z}}_B. \end{aligned}$$

- Decodes the codebooks using dirty-paper decoders, where $\tilde{x}_i$ is decoded from $\tilde{y}_{B;i}$.

By using good dirty-paper codes, capacity is achieved; see, e.g., [40].

We further note that codeword $\tilde{x}_i$ is recovered from $\tilde{y}_{B;i}$ regardless of whether the other codewords $\{\tilde{x}_j | j \ne i\}$ were recovered or not.

B. MIMO Wiretap Channel

By replacing the dirty-paper scalar codes in the layered-DPC scheme of VI-A with scalar dirty-paper wiretap codes [46], [47], a scheme that approaches the MIMO wiretap secrecy capacity can be constructed.

Theorem 6. *The layered-DPC scheme of Section VI-A achieves the secrecy capacity under a covariance constraint* $C_S\left(\mathbf{H}_B, \mathbf{H}_E \,|\, \bar{\mathbf{K}}\right)$ *by using:*

- *The optimal input covariance matrix* $\mathbf{K}$ *of* (22)*.*

- Any *joint unitary triangularization* (18).
- *Scalar Gaussian dirty-paper wiretap codes, where the i-th codebook ($i = 1, \ldots, N_A$) is designed for*
  - *Bob's SNR of $(b_i^2 - 1)$ and interference signal $\sum_{\ell=i+1}^{N_A} T_{B;i,\ell}\tilde{\mathsf{x}}_\ell$.*
  - *Eve's SNR of $(e_i^2 - 1)$.*
  - *Rate close to $R_i = \log(b_i^2/e_i^2)$.*

We next prove the existence of such codes and consequently also the result of Theorem 6.

Proof: The proof follows by a standard extension of the proof of Theorem 5 to the dirty-paper case [25], [46], [47].

Codebook construction: For each $k = 1, \ldots, N_A$, we generate a codebook $\mathcal{C}_k$ of $2^{n(R_k+\tilde{R}_k)}$ sub-codebooks, where n is length of the codewords. Each such sub-codebook is assigned a unique index pair (m_k, f_k), where $m_k \in \{1, 2, \ldots, 2^{nR_k}\}$ and $f_k \in \{1, 2, \ldots, 2^{n\tilde{R}_k}\}$, and contains $2^{n[R_k^U - (R_k+\tilde{R}_k)]}$ codewords. Each codeword within codebook k is generated independently in an i.i.d. manner with respect to a Gaussian distribution $p(\mathsf{u}_k)$ with parameters dictated by

$$\mathsf{u}_k = \tilde{T}_{B;k,k}\tilde{\mathsf{x}}_k + \alpha_k \sum_{\ell=k+1}^{N_A} \tilde{T}_{B;k,\ell}\tilde{\mathsf{x}}_\ell \,, \tag{42a}$$

$$\alpha_k \triangleq \frac{b_k^2 - 1}{b_k^2} \,, \tag{42b}$$

for zero mean unit power i.i.d. Gaussian random variables $\{\tilde{\mathsf{x}}_k | k = 1, \ldots, N_A\}$.

Note that since in this case the interference (available as side information to Alice) in sub-channel k is composed of messages $\{x_\ell | \ell = 1, \ldots, N_A\}$, the information carried by the sets $\{\tilde{\mathsf{x}}_\ell | \ell = 1, \ldots, N_A\}$ and $\{\mathsf{u}_\ell | \ell = 1, \ldots, N_A\}$ is the same.

Let $\epsilon > 0$. Then the rates are chosen as

$$\begin{aligned} R_k &\triangleq I\left(\mathsf{u}_k; \mathbf{y}_B\right) - I\left(\mathsf{u}_k; \mathbf{y}_E, \mathsf{u}_{k+1}^{N_A}\right) - \epsilon \\ &= \left[I\left(\mathsf{u}_k; \mathbf{y}_B\right) - I\left(\mathsf{u}_k; \mathsf{u}_{k+1}^{N_A}\right)\right] - I\left(\mathsf{u}_k; \mathbf{y}_E \middle| \mathsf{u}_{k+1}^{N_A}\right) - \epsilon \\ &= I\left(\tilde{\mathsf{x}}_k; \mathbf{y}_B \middle| \tilde{\mathsf{x}}_{k+1}^{N_A}\right) - I\left(\tilde{\mathsf{x}}_k; \mathbf{y}_E \middle| \tilde{\mathsf{x}}_{k+1}^{N_A}\right) - \epsilon \\ &= \log \frac{b_k^2}{e_k^2} - \epsilon, \end{aligned} \tag{43a}$$

$$\begin{aligned} \tilde{R}_k &\triangleq I\left(\mathsf{u}_k; \mathbf{y}_E \middle| \mathsf{u}_{k+1}^{N_A}\right) - \epsilon = I\left(\tilde{\mathsf{x}}_k; \mathbf{y}_E \middle| \tilde{\mathsf{x}}_{k+1}^{N_A}\right) - \epsilon \\ &= \log e_k^2 - \epsilon, \end{aligned} \tag{43b}$$

$$\begin{aligned} R_k^U &\triangleq I\left(\mathsf{u}_k; \mathbf{y}_B\right) - \epsilon \\ &= \log\left(b_k^2 + \sum_{\ell=k+1}^{N_A} |T_{B;k,\ell}|^2\right) - \epsilon \,. \end{aligned} \tag{43c}$$

Encoding (Alice): Encoding is carried in a successive manner, from last ($k = N_A$) to first ($k = 1$). Within codebook k, the index of the sub-codebook to be used is determined by the secret message m_k and a fictitious message f_k drawn uniformly over their respective ranges. The codeword $\boldsymbol{u}_k$, within sub-codebook (m_k, f_k) that is selected, is the one that is jointly typical with the side information $\sum_{\ell=k+1}^{N_A} \tilde{T}_{B;k,\ell}\tilde{x}_\ell$. If no such codeword $\boldsymbol{u}_k$ exists, then the first codeword is selected.

Decoding (Bob): Bob recovers (m_k, f_k) using standard dirty-paper decoding as in Section VI-A, and discards f_k. The error probability can be made arbitrarily small by taking a large enough n.

Secrecy analysis (Eve): As in the proof of Proposition 1, we provide $\{\mathsf{u}_\ell | \ell = k+1, \ldots, N_A\}$ as a genie for the secrecy analysis of u_k. By recalling that $\{\tilde{\mathsf{x}}_\ell | \ell = k+1, \ldots, N_A\}$ and $\{\mathsf{u}_\ell | \ell = k+1, \ldots, N_A\}$ carry the same information, and the linear relation in the definition of u_k (42a), the secrecy analysis reduces to the analysis in the proof of Proposition 1, as appears in Appendix C, specialized to the Gaussian case. ■

VII. Confidential Broadcast as a Consequence

In this section we consider the two-user MIMO confidential broadcast scenario. Namely, "Eve" is replaced with "Charlie" in (2b), and the corresponding noise, output and channel matrix are denoted by $\boldsymbol{z}_C$, $\boldsymbol{y}_C$ and $\mathbf{H}_C$, respectively.

We next show that, under the covariance matrix constraint, the rectangular capacity region (8), that was established in [11], can be attained as a natural extension of the capacity derivation for the MIMO wiretap channel and the layered DPC scheme proposed in Sections III and VI, respectively.

A. Capacity Region

We saw in Section III that in order to achieve the secrecy capacity where Charlie takes the role of Eve, the GSVD needs to be applied to $(\mathbf{G}_B, \mathbf{G}_C)$ and only the sub-channels corresponding to GSVs that are greater than 1 (corresponding to sub-channels with greater SNR to Bob than to Charlie) need to be used, and the rest — nullified.

However, we note that, if we were interested in confidential communication with Charlie rather than with Bob, we would get the same solution with the roles of $\mathbf{H}_B$ and $\mathbf{H}_C$ reversed. This, in turn, means inversion of the GSVs:

$$\log \mu_i(\mathbf{H}_C, \mathbf{H}_B, \bar{\mathbf{K}}) = -\log \mu_i(\mathbf{H}_B, \mathbf{H}_C, \bar{\mathbf{K}}).$$

In these terms, we can write the rectangular capacity-region of the confidential broadcast channel (8), established first in [11], as follows.

Theorem 7. *The capacity region of the confidential MIMO broadcast channel under an input covariance constraint $\bar{\mathbf{K}}$ is given by all rates (R_B, R_C) satisfying:*

$$R_B \le \sum_{i=1}^{N_A} \left[\log \mu_i^2\left(\mathbf{H}_B, \mathbf{H}_C, \bar{\mathbf{K}}\right)\right]_+ , \tag{44a}$$

$$R_C \le \sum_{i=1}^{N_A} \left[-\log \mu_i^2\left(\mathbf{H}_B, \mathbf{H}_C, \bar{\mathbf{K}}\right)\right]_+ . \tag{44b}$$

Remark 11. Similarly to the MIMO wiretap channel, the capacity region under a power constraint P is just the union of all (rectangular) regions under a covariance constraint with small enough trace.

The converse part of this result is trivial by Theorem 3, since both users attain their individual secrecy capacities. For the direct part, it is tempting to think that since different GSVs are nullified for Bob and for Charlie, Alice can achieve

their optimal rates simultaneously by communicating over orthogonal "subspaces". However, since the matrices $\mathbf{T}_B$ and $\mathbf{T}_C$ are not diagonal, these "subspaces" are not orthogonal, and some more care is needed.

To this end, in the next section we put into force the layered-DPC scheme of Section VI, which allows to recover the sub-message transmitted over each sub-channel independently, without the recovery of other sub-messages (in contrast to the layered-SIC scheme). This property is required by at least one of the users — Bob or Charlie — as each of them recovers only a subset of all the transmitted sub-messages. The derivation of the scheme thus provides a constructive proof for the direct part of Theorem 7, which is an alternative to the proof in [11].

B. Capacity Achieving Schemes

In view of Theorem 2 and the schemes developed for the MIMO wiretap channel, the result of Section III has a rather intuitive interpretation: $\mathbf{V}_A$ of the GSVD is the precoding matrix that designs the ratios between $\{b_i\}$ and $\{c_i\}$ to be as large as possible ($\{c_i\}$ replacing $\{e_i\}$), which corresponds to maximizing the achievable secrecy rate to Bob. In order to achieve Bob's secrecy capacity, only the sub-channels for which the secrecy rate is positive ($b_i > c_i$) need to be utilized. Allocating the remaining sub-channels to Charlie, on the other hand, attains Charlie's optimal covariance matrix.

Combining the two gives rise to the following scheme, which is a straightforward adaptation of the layered-DPC scheme of Section VI for the wiretap channel.

Scheme (Confidential broadcast via layered-DPC)**.**

Offline:

- Construct the effective MMSE matrix (17): $\bar{\mathbf{G}}_B \triangleq \mathbf{G}(\mathbf{H}_B, \bar{\mathbf{K}})$ and $\bar{\mathbf{G}}_C \triangleq \mathbf{G}(\mathbf{H}_C, \bar{\mathbf{K}})$, where $\bar{\mathbf{K}}$ is the constraining matrix.
- Apply the triangular form of the GSVD (16) to $(\bar{\mathbf{G}}_B, \bar{\mathbf{G}}_C)$ as in (25), to obtain the unitary matrices $\mathbf{U}_B$, $\mathbf{U}_C$ and $\mathbf{V}_A$, and the generalized upper-triangular matrices $\mathbf{T}_B$ and $\mathbf{T}_C$.
- Denote the diagonal elements of $\mathbf{T}_B$ and of $\mathbf{T}_C$ by $\{b_i\}$ and $\{c_i\}$, respectively.
- Denote further the (first) number of indices for which $b_i > c_i$ by L_B. The remaining $L_C = N_A - L_B$ indices satisfy $c_i \geq b_i$.
- Denote by $\tilde{\mathbf{U}}_B$ the upper-left $N_B \times L_B$ sub-matrix of $\mathbf{U}_B$, and by $\tilde{\mathbf{U}}_C$ — the upper-right $N_C \times L_C$ sub-matrix of $\mathbf{U}_C$.
- Construct $\tilde{\mathbf{T}}_B$ and $\tilde{\mathbf{T}}_C$ as in (27):

$$\tilde{\mathbf{T}}_B = \tilde{\mathbf{U}}_B^\dagger \mathbf{H}_B \mathbf{K}^{1/2} \mathbf{V}_A \,,$$
$$\tilde{\mathbf{T}}_C = \tilde{\mathbf{U}}_C^\dagger \mathbf{H}_C \mathbf{K}^{1/2} \mathbf{V}_A \,.$$

- Construct N_A good scalar dirty-paper wiretap codes of unit power and length n, denoted by $\{\tilde{x}_i | i = 1, \ldots, N_A\}$ (with the time index omitted to simplify notation), generated via random binning with respect to i.i.d. Gaussian distributions, as follows.
  - The first L_B codes are intended for Bob: Codebook $\tilde{x}_i$ ($1 \leq i \leq L_B$) of a rate close to $R_i = \log\left(b_i^2/c_i^2\right)$ is constructed for an AWGN channel to Bob of SNR $b_i^2 - 1$, and interference:

$$\sum_{\ell=i+1}^{N_A} T_{B;i,\ell} \tilde{x}_\ell \,,$$

    and for an AWGN channel to Charlie of SNR $c_i^2 - 1$.
  - The remaining L_C codes are intended for Charlie: Codebook $\tilde{x}_i$ ($L_B + 1 \leq i \leq N_A$) of a rate close to $R_i = \log\left(c_i^2/b_i^2\right)$ is constructed for an AWGN channel to Charlie of SNR $c_i^2 - 1$ and interference:

$$\sum_{\ell=i+1}^{N_A} T_{C;i,\ell} \tilde{x}_\ell \,,$$

    and for an AWGN channel to Bob of SNR $b_i^2 - 1$.

Alice: At each time instant $t = 1, \ldots, n$:

- Generates $\tilde{x}_i$ from last ($i = N_A$) to first ($i = 1$), where $\tilde{x}_i$ is generated according to the message to be conveyed and the signals $\{\tilde{x}_\ell | \ell = i+1, \ldots, N_A\}$.
- Forms $\tilde{\boldsymbol{x}}$ with entries $\{\tilde{x}_i\}$.
- Attains the vector $\boldsymbol{x}$ by multiplying $\tilde{\boldsymbol{x}}$ by $\mathbf{V}_A$ and $\mathbf{K}^{1/2}$ as in (34).
- Transmits $\boldsymbol{x}$.

Bob:

- At each time instant $t = 1, \ldots, n$, receives $\boldsymbol{y}_B$ and forms $\tilde{\boldsymbol{y}}_B$ according to (29):

$$\begin{aligned}\tilde{\boldsymbol{y}}_B &= \tilde{\mathbf{U}}_B^\dagger \boldsymbol{y}_B \\ &= \tilde{\mathbf{T}}_B \tilde{\boldsymbol{x}} + \tilde{\boldsymbol{z}}_B .\end{aligned}$$

- Decodes codebooks $i = 1, \ldots, L_B$ using dirty-paper decoders, where $\tilde{x}_i$ is decoded from $\tilde{y}_{B;i}$.

Charlie:

- At each time instant forms

$$\begin{aligned}\tilde{\boldsymbol{y}}_C &= \tilde{\mathbf{U}}_C^\dagger \boldsymbol{y}_C \\ &= \tilde{\mathbf{T}}_C \tilde{\boldsymbol{x}} + \tilde{\boldsymbol{z}}_C .\end{aligned}$$

- Decodes codebooks $i = L_B + 1, \ldots, N_A$ using dirty-paper decoders, where $\tilde{x}_i$ is decoded from $\tilde{y}_{C;(i-L_B)}$.

The following theorem proves that this scheme allows both users to attain their respective secrecy capacities *simultaneously*, providing a proof for Theorem 7.

Theorem 8. *The layered-DPC confidential broadcast scheme achieves the secrecy capacity region under a covariance constraint* (44) *by:*

- *Using scalar Gaussian dirty-paper wiretap codes intended for Bob, as follows, where the i-th codebook ($i = 1, \ldots, L_B$) is designed for:*
  - *Bob's SNR of $(b_i^2 - 1)$ and interference signal $\sum_{\ell=i+1}^{N_A} T_{B;i,\ell} \tilde{x}_\ell$.*
  - *Charlie's SNR of $(c_i^2 - 1)$.*
  - *Rate close to $R_i = \log(b_i^2/c_i^2)$.*
- *Using scalar Gaussian DPC wiretap codes intended for Charlie, as follows, where the i-th codebook ($i = L_B + 1, \ldots, N_A$) is designed for:*

- *Charlie's SNR of* $(c_i^2 - 1)$ *and interference* $\sum_{\ell=i+1}^{N_A} T_{C;i,\ell}\tilde{x}_\ell$.
- *Bob's SNR of* $(b_i^2 - 1)$.
- *Rate close to* $R_i = \log(c_i^2/b_i^2)$.

Proof sketch: We start by noting that since the capacity region is rectangular, it suffices to show how to approach the corner point of this region. The proof relies on the fact that in the layered-DPC scheme for the MIMO wiretap channel of Section VI, each sub-codebook is recovered independently, regardless of the other sub-codebooks. Hence, the proof of the decodability and secrecy analysis for Charlie are the same as in the proof of Theorem 6 (with Charlie being the "legitimate" user). In the treatment for Bob, a small variation is needed: the interference over sub-channel i ($1 \le i \le L_B$) is composed of both, messages intended for Charlie, $\tilde{x}_{L_B+1}^{N_A}$, and messages intended for Bob, $\tilde{x}_{i+1}^{L_B}$. Thus, the DPC for Bob is carried with respect to both of these interferences, and the decodability and secrecy analysis follow as in the proof of Theorem 6. ■

Remark 12 (Replacing DPC with SIC). DPC was used in the layered-DPC scheme for both users. However, in the proposed scheme one may use SIC instead of DPC for Charlie, as is done in the layered-SIC scheme for the MIMO wiretap problem. Alternatively, by using lower-triangular matrices instead of upper-triangular ones in (18) (which corresponds to switching roles between Bob and Charlie in the construction of the scheme), one can use SIC for Bob and DPC for Charlie. This phenomenon was also observed by Liu *et al.* [11]. Unfortunately, this scheme does not allow, in general, to avoid DPC for both of the users.

Remark 13 (Other choices of precoding matrices). In Section V-C, different choices of $\mathbf{V}_A$ were proposed for the MIMO wiretap problem: diagonalizing either $\mathbf{T}_B$ or $\mathbf{T}_C$, which corresponds to avoiding SIC by Bob or guaranteeing strong secrecy, respectively; or, by balancing all the SNRs of the sub-channels to Bob, which allows using the same codebook over all sub-channels and avoiding bit-loading / rate allocation. The analog in the case of confidential broadcast can be achieved by applying block diagonal unitary operations, in addition to the matrix $\mathbf{V}_A$ that is dictated by the GSVD, where the blocks correspond to the sub-channels that are allocated to Bob and to Charlie, of dimensions $L_B \times L_B$ and $L_C \times L_C$, respectively. However, whereas we can avoid SIC and DPC at Bob's end in the layered confidential broadcast scheme by diagonalizing his channel, we cannot achieve this result for both Charlie and Bob simultaneously, as DPC needs to be employed for at least one of the users.

VIII. Discussion: From Random Ensembles to Specific Codes

In this work, we have demonstrated how scalar codes can be used for some MIMO secrecy scenarios. Throughout the work, we have assumed that these scalar codes are taken from a random Gaussian ensemble, suitable in an appropriate sense (with or without secrecy constraints, with or without side information). One may be interested in a stronger result, where *any* scalar codes that are good in the appropriate sense can be used, without worrying about the way they were created. Further, it is desirable to construct MIMO secrecy schemes using *any standard* (non-secrecy) scalar codes that are good for communication over the (non-secrecy) AWGN channel. To that end, one may hope to combine the approach of the current work with procedures that construct scalar wiretap codes from non-secrecy ones, such as [12] (which is based upon similar techniques for discrete wiretap channels proposed in [48], [49]). Unfortunately, as we report in [50], there are some obstacles.

Surprisingly, the problem lies already in the use of scalar codes for MIMO communications without secrecy constraints. Recall the V-BLAST/GDFE schemes presented in Section IV and depicted in Fig. 1. Such schemes are widely accepted in the literature as capacity achieving, without proposing any treatment or analysis for specific codes. In practice, such schemes are used in conjunction with arbitrary scalar codebooks, e.g., one-dimensional constellations with some error-correction code [27]; however, the combination does not necessarily approach capacity even if the individual codes do. Indeed, for some specific channel matrices, the scheme might perform very poorly. To see this, consider (30). This is a multiple-access channel (MAC) from the inputs $\tilde{x}_1, \ldots, \tilde{x}_i$ to the output $y'_{B;i}$. The SIC decoder treating all inputs as noise is equivalent to a stage of a successive-decoding procedure for the MAC. For the MAC, in turn, not any collection of good AWGN codes achieves capacity (see, e.g., [51]). For example, assume that a MAC is given by

$$y_B = x_1 + x_2 + z.$$

Now further assume that the two codebooks are nested lattices. In that case (up to shaping), any possible point of $x_1 + x_2$ is also a point of the higher-rate code, thus one codebook cannot be decoded without the other. The problem is not restricted to integer coefficient ratios but affects performance for coefficients close to any "simple" ratio; see, e.g., [52, Section III].

Returning back to the multi-stream schemes for the MIMO wiretap setup of Section V, the decoder of Bob will also incur the same difficulty discussed above when generalizing to arbitrary scalar codes. Furthermore, the same issue arises in our secrecy analyses (except when Eve's channel is orthogonalized, as in Section V-A): We successively provide Eve with previous messages as a "genie" side information. As a result the proof hinges on Eve's disability to perform a successive decoding process in the presence of interference from yet undecoded messages. Here also this interference is taken to be Gaussian and alignment might help Eve.

To conclude, of the two ingredients needed for adjusting *any* codes that are good for communication over scalar AWGN channels to the MIMO wiretap channel, the secrecy part can be treated by the procedure of [12]. The remaining problem is similar to the one in SIC without secrecy constraints. Indeed, obtaining good scalar Gaussian codes that approach capacity under SIC (without secrecy) from arbitrary scalar Gaussian codes remains an interesting open problem.

APPENDIX A
PROOF OF LEMMA 1

The following proposition will be used in the proof of Lemma 1.

Proposition 2. *Let $\mathbf{A}_1$ and $\mathbf{A}_2$ be $m_1 \times n$ and $m_2 \times n$ full-rank matrices, respectively, where $m_1 \geq n$ and $m_2 \geq n$. Consider the generalized eigenvalue (GEV) problem:*

$$\mathbf{A}_1^\dagger \mathbf{A}_1 \boldsymbol{y} = \lambda \mathbf{A}_2^\dagger \mathbf{A}_2 \boldsymbol{y}\,.$$

Then, the generalized eigenvalues of $(\mathbf{A}_1^\dagger \mathbf{A}_1, \mathbf{A}_2^\dagger \mathbf{A}_2)$, $\{\lambda_i\}$, are the GSVs of $(\mathbf{A}_1, \mathbf{A}_2)$, $\{\mu_i\}$, and the generalized eigenvectors are the corresponding columns of

$$\mathbf{Y} = \mathbf{X}^{-\dagger}.$$

Furthermore, the differential of the GEV λ in terms of the differentials of $\mathbf{A}_1^\dagger \mathbf{A}_1$ and of $\mathbf{A}_2^\dagger \mathbf{A}_2$ is given by

$$d\lambda = \frac{\boldsymbol{y}^\dagger \Big(d(\mathbf{A}_1^\dagger \mathbf{A}_1) - \lambda d(\mathbf{A}_2^\dagger \mathbf{A}_2) \Big) \boldsymbol{y}}{\boldsymbol{y}^\dagger \mathbf{A}_1^\dagger \mathbf{A}_1 \boldsymbol{y}}\,. \tag{45}$$

Proof: The first part of the proposition easily follows from

$$\mathbf{G}_B^\dagger \mathbf{G}_B \mathbf{Y} = \mathbf{X} \mathbf{D}_B^2\,,$$
$$\mathbf{G}_E^\dagger \mathbf{G}_E \mathbf{Y} = \mathbf{X} \mathbf{D}_E^2\,.$$

The proof of the differential identity (45) can be derived by standard eigenvalue perturbation analysis; see, e.g., [53]. ∎

Consider now the diagonal variant of the GSVD of $\mathbf{G}_B = \mathbf{G}(\mathbf{H}_B, \mathbf{K})$ and $\mathbf{G}_E = \mathbf{G}(\mathbf{H}_E, \mathbf{K})$ (14):

$$\mathbf{G}_B = \mathbf{U}_B \mathbf{D}_B \mathbf{X}^\dagger, \tag{46a}$$
$$\mathbf{G}_E = \mathbf{U}_E \mathbf{D}_E \mathbf{X}^\dagger, \tag{46b}$$

and denote the squared GSV vector by $\boldsymbol{\lambda}$, i.e., the vector whose entries satisfy:

$$\lambda_i \triangleq \mu_i^2\,.$$

Note further that $0 < \mu_i, \lambda_i < \infty$, since $\mathbf{G}_B$ and $\mathbf{G}_E$ are of full rank [recall (17)].

Following (20), the MI difference in terms of $\{\lambda_i\}$ is equal to

$$I_S(\mathbf{H}_B, \mathbf{H}_E, \mathbf{K}) = \sum \log \lambda_i\,.$$

By applying the result of Proposition 2 to the effective channel matrices of (46), we obtain the following lemma.

Lemma 4. *The differential of the GSV λ_i ($i = 1, \ldots, N_A$), in terms of the differential of the covariance matrix $\mathbf{K}$, is given by*

$$e_i^2 d\lambda_i = (\lambda_i - 1) \boldsymbol{y}_i^\dagger \mathbf{B}^{-1} (d\mathbf{K}) \mathbf{B}^{-\dagger} \boldsymbol{y}_i\,,$$

where $\mathbf{B} = \mathbf{K}^{1/2}$, $\boldsymbol{e}$ is the diagonal of $\mathbf{D}_E$, and $\boldsymbol{y}_i$ is the corresponding generalized eigenvector corresponding to λ_i.

Proof: Perturbing $\mathbf{K}$ results in the following differentials of $\mathbf{G}_B^\dagger \mathbf{G}_B$ and $\mathbf{G}_E^\dagger \mathbf{G}_E$ (46) :

$$2d(\mathbf{G}_B^\dagger \mathbf{G}_B) = \mathbf{B}^{-1}(d\mathbf{K}) \mathbf{H}_B^\dagger \mathbf{H}_B \mathbf{B} + \mathbf{B}^\dagger \mathbf{H}_B^\dagger \mathbf{H}_B (d\mathbf{K}) \mathbf{B}^{-\dagger}, \tag{47a}$$

$$2d(\mathbf{G}_E^\dagger \mathbf{G}_E) = \mathbf{B}^{-1}(d\mathbf{K}) \mathbf{H}_E^\dagger \mathbf{H}_E \mathbf{B} + \mathbf{B}^\dagger \mathbf{H}_E^\dagger \mathbf{H}_E (d\mathbf{K}) \mathbf{B}^{-\dagger}. \tag{47b}$$

Substituting (47) in (45), gives rise to

$$\begin{aligned} 2e_i^2 d\lambda_i &= \boldsymbol{y}_i^\dagger \Big(\mathbf{B}^{-1}(d\mathbf{K})(\mathbf{H}_B^\dagger \mathbf{H}_B - \lambda_i \mathbf{H}_E^\dagger \mathbf{H}_E)\mathbf{B} \\ &\quad + \mathbf{B}^\dagger (\mathbf{H}_B^\dagger \mathbf{H}_B - \lambda_i \mathbf{H}_E^\dagger \mathbf{H}_E)(d\mathbf{K}) \mathbf{B}^{-\dagger} \Big) \boldsymbol{y}_i \\ &= \boldsymbol{y}_i^\dagger \Big(\mathbf{B}^{-1}(d\mathbf{K}) \mathbf{B}^{-\dagger} \mathbf{B}^\dagger (\mathbf{H}_B^\dagger \mathbf{H}_B - \lambda_i \mathbf{H}_E^\dagger \mathbf{H}_E)\mathbf{B} \\ &\quad + \mathbf{B}^\dagger (\mathbf{H}_B^\dagger \mathbf{H}_B - \lambda_i \mathbf{H}_E^\dagger \mathbf{H}_E) \mathbf{B} \mathbf{B}^{-1} (d\mathbf{K}) \mathbf{B}^{-\dagger} \Big) \boldsymbol{y}_i \\ &= 2(\lambda_i - 1) \boldsymbol{y}_i^\dagger \mathbf{B}^{-1} (d\mathbf{K}) \mathbf{B}^{-\dagger} \boldsymbol{y}_i\,, \end{aligned}$$

as desired. ∎

Corollary 1. *If $d\mathbf{K}$ is positive semidefinite, then the sign of $d\lambda_i$ equals the sign of $\lambda_i - 1$.*

The result of Lemma 1 follows immediately from this corollary.

APPENDIX B
TRUNCATION OF GENERALIZED SINGULAR VALUES

Apply the triangular variant of the GSVD (16) to the matrices $\mathbf{G}_B = \mathbf{G}(\mathbf{H}_B, \mathbf{K})$ and $\mathbf{G}_E = \mathbf{G}(\mathbf{H}_E, \mathbf{K})$, as in (17) and (18):

$$\mathbf{G}_B \triangleq \begin{pmatrix} \mathbf{H}_B \mathbf{K}^{1/2} \\ \mathbf{I} \end{pmatrix} = \mathbf{U}_B \mathbf{D}_B \mathbf{T} \mathbf{V}_A^\dagger\,, \tag{48a}$$

$$\mathbf{G}_E \triangleq \begin{pmatrix} \mathbf{H}_E \mathbf{K}^{1/2} \\ \mathbf{I} \end{pmatrix} = \mathbf{U}_E \mathbf{D}_E \mathbf{T} \mathbf{V}_A^\dagger\,. \tag{48b}$$

Using any unitary matrix $\mathbf{Q}$ instead of $\mathbf{I}$ in the definition of $\mathbf{G}_B$ and $\mathbf{G}_E$, has no effect on the resulting matrices $\mathbf{V}_A$, $\mathbf{T}$, $\mathbf{D}_B$ and $\mathbf{D}_E$:

$$\begin{pmatrix} \mathbf{H}_B \mathbf{K}^{1/2} \\ \mathbf{Q} \end{pmatrix} = \mathbf{U}_B^{\mathbf{Q}} \mathbf{D}_B \mathbf{T} \mathbf{V}_A^\dagger\,,$$
$$\begin{pmatrix} \mathbf{H}_E \mathbf{K}^{1/2} \\ \mathbf{Q} \end{pmatrix} = \mathbf{U}_E^{\mathbf{Q}} \mathbf{D}_E \mathbf{T} \mathbf{V}_A^\dagger\,.$$

Furthermore, the upper-left $N_B \times N_A$ and $N_B \times N_E$ of the resulting left unitary matrices $\mathbf{U}_B^{\mathbf{Q}}$ and $\mathbf{U}_E^{\mathbf{Q}}$, respectively, are equal to those of $\mathbf{U}_B$ and $\mathbf{U}_E$ of (48).

Using the last observation with $\mathbf{Q} = \mathbf{V}_A^\dagger$ and (48) for the matrices

$$\mathbf{G}_B^{\mathbf{V}} \triangleq \begin{pmatrix} \mathbf{H}_B \mathbf{K}^{1/2} \mathbf{V}_A \\ \mathbf{I} \end{pmatrix} = \begin{pmatrix} \mathbf{H}_B \mathbf{K}^{1/2} \\ \mathbf{V}_A^\dagger \end{pmatrix} \mathbf{V}_A\,,$$
$$\mathbf{G}_E^{\mathbf{V}} \triangleq \begin{pmatrix} \mathbf{H}_E \mathbf{K}^{1/2} \mathbf{V}_A \\ \mathbf{I} \end{pmatrix} = \begin{pmatrix} \mathbf{H}_E \mathbf{K}^{1/2} \\ \mathbf{V}_A^\dagger \end{pmatrix} \mathbf{V}_A\,,$$

gives rise to the GSVD of $\mathbf{G}_B^{\mathbf{V}}$ and $\mathbf{G}_E^{\mathbf{V}}$:

$$\mathbf{G}_B^{\mathbf{V}} \triangleq \mathbf{U}_B^{\mathbf{V}} \mathbf{D}_B \mathbf{T}\,, \tag{49a}$$

$$\mathbf{G}_E^{\mathbf{V}} \triangleq \mathbf{U}_E^{\mathbf{V}} \mathbf{D}_E \mathbf{T}\,, \tag{49b}$$

where $\mathbf{U}_B^{\mathbf{V}}$ and $\mathbf{U}_E^{\mathbf{V}}$ are unitary (and their $N_B \times N_A$ and $N_E \times N_A$ upper-left sub-matrices are equal to those of $\mathbf{U}_B$ and $\mathbf{U}_E$, respectively).

That is, the GSVD of $\mathbf{G}_B^{\mathbf{V}}$ and $\mathbf{G}_E^{\mathbf{V}}$ is achieved by applying a QR decomposition to each of them.

The representation in (49) allows us to incorporate a truncation operation:

$$\begin{aligned}\mathbf{G}'_B &\triangleq \begin{pmatrix}\mathbf{H}_B\mathbf{K}^{1/2}\mathbf{V}_A\mathbf{I}_B\\ \mathbf{I}\end{pmatrix}\\ &= \mathbf{U}'_B\mathbf{D}'_B\mathbf{T}' \qquad (50a)\\ \mathbf{G}'_E &\triangleq \begin{pmatrix}\mathbf{H}_E\mathbf{K}^{1/2}\mathbf{V}_A\mathbf{I}_B\\ \mathbf{I}\end{pmatrix}\\ &= \mathbf{U}'_E\mathbf{D}'_E\mathbf{T}', \qquad (50b)\end{aligned}$$

where $\mathbf{U}'_B$ and $\mathbf{U}'_E$ are unitary having the same first L_B columns as $\mathbf{U}^{\mathrm{V}}_B$ and $\mathbf{U}^{\mathrm{V}}_E$, respectively; $\mathbf{T}'$, $\mathbf{D}'_B$ and $\mathbf{D}'_E$ have the same first L_B columns as $\mathbf{T}$, $\mathbf{D}_B$ and $\mathbf{D}_E$, respectively, whereas the remaining $L_E = N_A - L_B$ columns are all zero except for the diagonal elements, which are equal to 1:

$$\begin{aligned}D'_{B;i,j} = D'_{E;i,j} = T'_{i,j} &= 1, \qquad i = j, j > L_B\,;\\ D'_{B;i,j} = D'_{E;i,j} = T'_{i,j} &= 0, \qquad i \neq j, j > L_B\,.\end{aligned}$$

The latter is easily seen by noting that the QR decomposition carries out a Gram–Schmidt process over the columns of the decomposed matrices, and hence the first L_B columns remain the same after applying $\mathbf{I}_B$, whereas the structure of the remaining columns is trivial due to the nullification of the last L_E columns of $\mathbf{H}_B\mathbf{K}^{1/2}\mathbf{V}_A$.

We note that (50) is the GSVD of $\mathbf{G}'_B$ and $\mathbf{G}'_E$ up to the normalization property (15), which has no effect on the GSVs and can be achieved by a multiplication by an $N_A \times N_A$ diagonal matrix with its first L_B entries equal to 1 and the remaining entries — to $1/\sqrt{2}$.

The desired result is established by noting that $\mathbf{K}^{1/2} = \bar{\mathbf{K}}^{1/2}\mathbf{V}_A\mathbf{I}_B$, and that the first L_B GSVs of $(\mathbf{G}'_B, \mathbf{G}'_E)$ are equal to the first L_B GSVs of $(\mathbf{G}_B, \mathbf{G}_E)$ (the GSVs that are greater than 1) and the remaining GSVs of $(\mathbf{G}'_B, \mathbf{G}'_E)$ are equal to 1.

Appendix C
Proof of Proposition 1

In this appendix, 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).

Proof of Proposition 1: Denote

$$\tilde{R}_k \triangleq I(\tilde{\mathsf{x}}_k; \mathsf{y}_E|\tilde{\mathsf{x}}_{k+1}^{N_A}) - \epsilon. \qquad (51)$$

The codebooks are generated sequentially, from last ($k = N_A$) to first ($k = 1$), as follows. For $k = N_A$, construct the codebook $\mathcal{C}_{N_A}$ of $2^{n\left(R_{N_A}+\tilde{R}_{N_A}\right)}$ codewords, that are generated independently with i.i.d. entries with respect to $p\left(\tilde{\mathsf{x}}_{N_A}\right)$. For $k \in \{1, \ldots, N_A - 1\}$, for each (already generated) codeword set $(\tilde{\boldsymbol{x}}_{k+1}, \ldots, \tilde{\boldsymbol{x}}_{N_A}) \in \mathcal{C}_{k+1} \times \cdots \times \mathcal{C}_{N_A}$, generate a codebook of $2^{n\left(R_k+\tilde{R}_k\right)}$ codewords with respect to $\prod_{i=1}^n p\left(\tilde{\mathsf{x}}_k|\tilde{x}_{k+1}(i), \ldots, \tilde{x}_{N_A}(i)\right)$, where $\tilde{x}_\ell(i)$ is the i-the letter of the codeword $\tilde{\boldsymbol{x}}_\ell$. Within each codebook, each codeword is assigned a unique index pair (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

$$\begin{aligned}R_k + \tilde{R}_k &= I\left(\tilde{\mathsf{x}}_k; \mathsf{y}_B\middle|\tilde{\mathsf{x}}_{k+1}^{N_A}\right) - 2\epsilon \qquad (52a)\\ &< I\left(\tilde{\mathsf{x}}_k; \mathsf{y}_B\middle|\tilde{\mathsf{x}}_{k+1}^{N_A}\right), \qquad (52b)\end{aligned}$$

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 (51), 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,

$$\begin{aligned}&\sum_{\ell=1}^k \left[I\left(\tilde{\mathsf{x}}_\ell; \mathsf{y}_B\middle|\tilde{\mathsf{x}}_{\ell+1}^{N_A}\right) - I\left(\tilde{\mathsf{x}}_\ell; \mathsf{y}_E\middle|\tilde{\mathsf{x}}_{\ell+1}^{N_A}\right)\right] - \delta_1 \qquad (53a)\\ &\leq \frac{1}{n}H\left(\tilde{\boldsymbol{x}}_1^k\middle|\tilde{\boldsymbol{x}}_{k+1}^{N_A}, \boldsymbol{y}_E, \mathcal{C}\right) \qquad (53b)\\ &\leq \sum_{\ell=1}^k I\left(\tilde{\mathsf{x}}_\ell; \mathsf{y}_B\middle|\tilde{\mathsf{x}}_{\ell+1}^{N_A}\right) - I\left(\tilde{\mathsf{x}}_\ell; \mathsf{y}_E\middle|\tilde{\mathsf{x}}_{\ell+1}^{N_A}\right) + \delta_2\,. \qquad (53c)\end{aligned}$$

To establish (53b) we use the fact that the sequences $\tilde{\boldsymbol{x}}_\ell$ are selected independently given $\tilde{\boldsymbol{x}}_{\ell+1}^{N_A}$, so that, for large enough n, the following chain of inequalities holds

$$\begin{aligned}&H\left(\tilde{\boldsymbol{x}}_1^k\middle|\tilde{\boldsymbol{x}}_{k+1}^{N_A}, \boldsymbol{y}_E, \mathcal{C}\right) \qquad (54a)\\ &= H\left(\tilde{\boldsymbol{x}}_1^k\middle|\tilde{\boldsymbol{x}}_{k+1}^{N_A}, \mathcal{C}\right) - I\left(\tilde{\boldsymbol{x}}_1^k; \boldsymbol{y}_E\middle|\tilde{\boldsymbol{x}}_{k+1}^{N_A}, \mathcal{C}\right) \qquad (54b)\\ &= \sum_{\ell=1}^k \left[H\left(\tilde{\boldsymbol{x}}_\ell\middle|\tilde{\boldsymbol{x}}_{\ell+1}^{N_A}, \mathcal{C}\right) - I\left(\tilde{\boldsymbol{x}}_\ell; \boldsymbol{y}_E\middle|\tilde{\boldsymbol{x}}_{\ell+1}^{N_A}, \mathcal{C}\right)\right] \qquad (54c)\end{aligned}$$

$$= \sum_{\ell=1}^{k} \left[nI\left(\tilde{\mathsf{x}}_\ell; \mathsf{y}_B \middle| \tilde{\mathsf{x}}_{\ell+1}^{N_A}\right) - 2\epsilon - I\left(\tilde{\boldsymbol{x}}_\ell; \boldsymbol{y}_E \middle| \tilde{\boldsymbol{x}}_{\ell+1}^{N_A}, \mathcal{C}\right)\right] \quad (54d)$$

$$\geq n \sum_{\ell=1}^{k} \left[I\left(\tilde{\mathsf{x}}_\ell; \mathsf{y}_B \middle| \tilde{\mathsf{x}}_{\ell+1}^{N_A}\right) - I\left(\tilde{\mathsf{x}}_\ell; \mathsf{y}_E \middle| \tilde{\mathsf{x}}_{\ell+1}^{N_A}\right) - 3\epsilon \right], \quad (54e)$$

where (54d) follows from (52a), and to establish (54e) we use the fact that the channel is memoryless along with standard typicality arguments [54].

To establish (53c), we use [55, 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

$$S = \sum_{\ell=1}^{k} \left(R_\ell + \tilde{R}_\ell\right) \quad (55a)$$

$$= \left[\sum_{\ell=1}^{k} I\left(\tilde{\mathsf{x}}_\ell; \mathsf{y}_B \middle| \tilde{\mathsf{x}}_{\ell+1}^{N_A}\right)\right] - 2\epsilon \quad (55b)$$

$$> \left[\sum_{\ell=1}^{k} I\left(\tilde{\mathsf{x}}_\ell; \mathsf{y}_E \middle| \tilde{\mathsf{x}}_{\ell+1}^{N_A}\right)\right] + \delta \quad (55c)$$

$$= I\left(\tilde{\mathsf{x}}_1^k; \mathsf{y}_E \middle| \tilde{\mathsf{x}}_{k+1}^{N_A}\right) + \delta, \quad (55d)$$

where (55c) 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 proved (53b) and (53c), the secrecy analysis is now complete. ■

Remark 14. For the special case of mutually independent $(\tilde{\mathsf{x}}_1, \ldots, \tilde{\mathsf{x}}_{N_A})$, there is no need to generate a different codebook $\mathcal{C}_k$ for each selection of preceding codewords $(\tilde{\boldsymbol{x}}_{k+1}, \ldots, \tilde{\boldsymbol{x}}_{N_A})$, and the same codebook can be applied regardless of the other codewords.

ACKNOWLEDGMENT

The authors thank Ziv Goldfeld for proposing to extend the result of Proposition 1 from independent codes $x_1, \ldots, x_{N_A}$ to dependent ones, and Ronit Bustin for helpful discussions and for pointing their attention to the work of Baccelli *et al.* [51].

REFERENCES

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

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

[3] S. K. Leung-Yan-Cheong and M. E. Hellman, "The Gaussian wire-tap channel," *IEEE Trans. Inf. Theory*, vol. 24, no. 4, pp. 451–456, Jul. 1978.

[4] A. Khisti and G. W. Wornell, "Secure transmission with multiple antennas—part II: The MIMOME wiretap channel," *IEEE Trans. Inf. Theory*, vol. 56, no. 11, pp. 5515–5532, Nov. 2010.

[5] F. Oggier and B. Hassibi, "The secrecy capacity of the MIMO wiretap channel," *IEEE Trans. Inf. Theory*, vol. 57, no. 8, pp. 4961–4972, Aug. 2011.

[6] T. Liu and S. Shamai, "A note on the secrecy capacity of the multiple-antenna wiretap channel," *IEEE Trans. Inf. Theory*, vol. 55, no. 6, pp. 2547–2553, June 2009.

[7] 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*, July 2009.

[8] S. Loyka and C. D. Charalambous, "An algorithm for global maximization of secrecy rates in Gaussian MIMO wiretap channels," *IEEE Trans. Comm.*, vol. 63, no. 6, pp. 2288–2299, June 2015.

[9] H. Weingarten, Y. Steinberg, and S. Shamai, "The capacity region of the Gaussian multiple-input multiple-output broadcast channel," *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 3936–3964, Sep. 2006.

[10] R. Liu and H. V. Poor, "Secrecy capacity region of a multiple-antenna Gaussian broadcast channel with confidential messages," *IEEE Trans. Inf. Theory*, vol. 55, no. 3, pp. 1235–1249, Mar. 2009.

[11] R. Liu, T. Liu, H. V. Poor, and S. Shamai, "Multiple-input multiple-output Gaussian broadcast channels with confidential messages," *IEEE Trans. Inf. Theory*, vol. 56, no. 9, pp. 4215–4227, Sep. 2010.

[12] 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, pp. 956–960.

[13] A. Thangaraj, S. Dihidar, A. R. Calderbank, S. W. McLaughlin, and J.-M. Merolla, "Applications of LDPC codes to the wiretap channel," *IEEE Trans. Inf. Theory*, vol. 53, no. 8, pp. 2933–2945, Aug. 2007.

[14] D. Klinc, H. Jeongseok, S. W. McLaughlin, J. Barros, and B.-J. Kwak, "LDPC codes for the Gaussian wiretap channel," *IEEE Trans. Info. Forensics and Security*, vol. 6, no. 3, pp. 532–540, Sep 2011.

[15] F. Oggier, P. Solé, and J.-C. Belfiore, "Lattice codes for the wiretap Gaussian channel: Construction and analysis," *IEEE Trans. Inf. Theory*, vol. 62, no. 10, pp. 5690–5708, Oct. 2016.

[16] H. Mahdavifar and A. Vardy, "Achieving the secrecy capacity of wiretap channels using polar codes," *IEEE Trans. Inf. Theory*, vol. 57, no. 10, pp. 6428–6443, Oct. 2011.

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

[18] Y. Yan, L. Liu, and C. Ling, "Polar lattices for strong secrecy over the mod-λ gaussian wiretap channel," in *Proc. IEEE Int. Symp. on Inf. Theory (ISIT)*, Honolulu, HI, USA, June/July 2014, pp. 961–965.

[19] A. Khina, Y. Kochman, and U. Erez, "Joint unitary triangularization for MIMO networks," *IEEE Trans. Sig. Proc.*, vol. 60, no. 1, pp. 326–336, Jan. 2012.

[20] E. Telatar, "Capacity of the multiple antenna Gaussian channel," *Europ. Trans. Telecommun.*, vol. 10, no. 6, pp. 585–595, Nov. 1999.

[21] G. Foschini, "Layered space–time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Sys. Tech. Jour.*, vol. 1, no. 2, pp. 41–59, 1996.

[22] 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 *Proc. URSI Int. Symp. Sig., Sys., Elect. (ISSSE)*, Sep./Oct. 1998, pp. 295–300.

[23] 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.* US: Springer, 1997, pp. 79–127.

[24] B. Hassibi, "An efficient square-root algorithm for BLAST," in *Proc. IEEE Int. Conf. Acoust. Speech and Sig. Proc. (ICASSP)*, vol. 2, Istanbul, Turkey, June 2000, pp. 737–740.

[25] M. H. M. Costa, "Writing on dirty paper," *IEEE Trans. Inf. Theory*, vol. 29, no. 3, pp. 439–441, May 1983.

[26] A. Khina, I. Livni, A. Hitron, and U. Erez, "Joint unitary triangularization for Gaussian multi-user MIMO networks," *IEEE Trans. Inf. Theory*, vol. 61, no. 5, pp. 2662–2692, May 2015.

[27] D. P. Palomar and Y. Jiang, "MIMO transceiver design via majorization theory," *Found. Trends Comm. Info. Theory*, vol. 3, no. 4, pp. 331–551, Nov. 2006.

[28] G. H. Golub and C. F. Van Loan, *Matrix Computations, 3rd ed.* Baltimore: Johns Hopkins University Press, 1996.

[29] H. Weyl, "Inequalities between two kinds of eigenvalues of a linear transformation," in *Proc. Nat. Acad. Sci. USA, 35*, no. 7, May 1949, pp. 408–411.

[30] A. Horn, "On the eigenvalues of a matrix with prescribed singular values," in *Proc. Amer. Math. Soc.*, vol. 5, no. 1, Feb. 1954, pp. 4–7.

[31] P. Kosowski and A. Smoktunowicz, "On constructing unit triangular matrices with prescribed singular values," *Computing*, vol. 64, no. 3, pp. 279–285, May 2000.

[32] Y. Jiang, W. Hager, and J. Li, "The generalized triangular decompostion," *Math. of Comput.*, vol. 77, no. 262, pp. 1037–1056, Oct. 2008.

[33] J.-K. Zhang and K. M. Wong, "Fast QRS decomposition of matrix and its applications to numerical optimization," Dpt. of Elect. and Comp. Engineering, McMaster University, Tech. Rep. [Online]. Available: http://www.ece.mcmaster.ca/~jkzhang/papers/sam_qrs.pdf

[34] 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. Inf. Theory*, vol. 51, no. 1, pp. 154–172, Jan. 2005.

[35] Y. Jiang, W. Hager, and J. Li, "The geometric mean decompostion," *Lin. Algebra and Its Apps.*, vol. 396, pp. 373–384, Feb. 2005.

[36] C. F. Van Loan, "Generalizing the singular value decomposition," *SIAM J. Numer.*, vol. 13, no. 1, pp. 76–83, Mar. 1976.

[37] Z. Bai, "The CSD, GSVD, their applications and computations," *IMA Preprint 958, University of Minnesota, Minneapolis*, April 1992.

[38] C. C. Paige and M. A. Saunders, "Towards a generalized singular value decomposition," *SIAM J. Numer.,* Anal. 18, no. 3, pp. 398–405, Jun. 1981.

[39] J. M. Cioffi, G. P. Dudevoir, M. V. Eyuboglu, and G. D. Forney Jr., "MMSE decision-feedback equalizers and coding — Part I: Equalization results," *IEEE Trans. Comm.*, vol. 43, no. 10, pp. 2582–2594, Oct. 1995.

[40] Y. Jiang, W. Hager, and J. Li, "Uniform channel decomposition for MIMO communications," *IEEE Trans. Sig. Proc.*, vol. 53, no. 11, pp. 4283–4294, Nov. 2005.

[41] H. El Gamal, G. Caire, and M. O. Damen, "Lattice coding and decoding achieve the optimal diversity–multiplexing tradeoff of MIMO channels," *IEEE Trans. Inf. Theory*, vol. 50, pp. 968–985, 2004.

[42] M. R. Bloch and J. N. Laneman, "Strong secrecy from channel resolvability," *IEEE Trans. Inf. Theory*, vol. 59, no. 12, pp. 8077–8098, Dec. 2013.

[43] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge University Press, 2011.

[44] J. M. Cioffi and G. Ginis, "A multi-user precoding scheme achieving crosstalk cancellation with application to DSL systems," in *Proc. Asilomar Conf. Sig., Sys and Comp.*, vol. 2, Pacific Grove, CA, USA, Oct./Nov. 2000, pp. 1627–1631.

[45] G. Caire and S. Shamai, "On the achievable throughput of a multiantenna Gaussian broadcast channel," *IEEE Trans. Inf. Theory*, vol. 49, no. 7, pp. 1649–1706, July 2003.

[46] C. Mitrpant, A. J. Han Vinck, and Y. Luo, "An achievable region for the Gaussian wiretap channel with side information," *IEEE Trans. Inf. Theory*, vol. 52, no. 5, pp. 2181–2190, May 2006.

[47] Y. Chen and A. J. Han Vinck, "Wiretap channel with side information," *IEEE Trans. Inf. Theory*, vol. 54, no. 1, pp. 395–402, Jan. 2008.

[48] M. Bellare, S. Tessaro, and A. Vardy, "Semantic security for the wiretap channel," in *Proc. CRYPTO, LNCS*, vol. 7417, 2012, pp. 294–311.

[49] M. Hayashi and R. Matsumoto, "Construction of wiretap codes from ordinary channel codes," in *Proc. Int. Symp. Info. Theory (ISIT), Austin, TX*, June 2010, pp. 2538–2542.

[50] A. Khina, Y. Kochman, and A. Khisti, "From ordinary AWGN codes to optimal MIMO wiretap schemes," in *Proc. IEEE Info. Theory Workshop (ITW)*, Hobart, Tas, Australia, Oct./Nov. 2014, pp. 632–636.

[51] F. Baccelli, A. El Gamal, and D. N. C. Tse, "Interference networks with point-to-point codes," *IEEE Trans. Inf. Theory*, vol. 57, no. 5, pp. 2582–2596, May 2011.

[52] O. Ordentlich and U. Erez, "On the robustness of lattice interference alignment," *IEEE Trans. Inf. Theory*, vol. 59, no. 5, pp. 2735–2759, May 2013.

[53] J. de Leeuw, "Derivatives of generalized eigen systems with applications," Preprint Series 528, Department of Statistics, UCLA, Sep. 2007.

[54] T. M. Cover and J. A. Thomas, *Elements of Information Theory, Second Edition*. New York: Wiley, 2006.

[55] Y.-K. Chia and A. El Gamal, "Three-receiver broadcast channels with common and confidential messages," *IEEE Trans. Inf. Theory*, vol. 58, no. 5, pp. 2748–2765, May 2012.