Joint Precoder and Artificial Noise Design for MIMO Wiretap Channels With Finite-Alphabet Inputs Based on the Cut-Off Rate

Sina Rezaei Aghdam, *Student Member, IEEE*, and Tolga M. Duman, *Fellow, IEEE*

***Abstract*— We consider precoder and artificial noise (AN) design for multi-antenna wiretap channels under the finite-alphabet input assumption. We assume that the transmitter has access to the channel coefficients of the legitimate receiver and knows the statistics of the eavesdropper's channel. Accordingly, we propose a secrecy rate maximization algorithm using a gradient descent-based optimization of the precoder matrix and an exhaustive search over the power levels allocated to the AN. We also propose algorithms to reduce the complexities of direct ergodic secrecy rate maximization by: 1) maximizing a cut-off rate-based approximation for the ergodic secrecy rate, simplifying the mutual information expression, which lacks a closed-form and 2) diagonalizing the channels toward the legitimate receiver and the eavesdropper, which allows for employing a per-group precoding-based technique. Our numerical results reveal that jointly optimizing the precoder and the AN outperforms the existing solutions in the literature, which rely on the precoder optimization only. We also demonstrate that the proposed low complexity alternatives result in a small loss in performance while offering a significant reduction in computational complexity.**

***Index Terms*— Physical layer security, finite-alphabet inputs, precoding, artificial noise, cut-off rate, MIMO communications.**

I. Introduction

SECURITY is an increasingly important issue in wireless networks. With the ever-growing demand for the privacy-sensitive wireless services, researchers are getting more and more interested in finding techniques which provide additional confidentiality guarantees. Securing the communication at the physical layer is an alternative or a complement to the conventional higher network-layer solutions, such as encryption. The basic principle of physical layer security is to exploit the randomness of the communication channels to allow a transmitter deliver its message to an intended receiver reliably while guaranteeing that a third party cannot infer any information about it [2].

Among the studies in the area of physical layer security, multi-antenna wiretap channels have been of particular interest [3] as exploiting multiple antennas for transmission has been identified as one of the key enablers for achieving secrecy. The increased dimensionality can be utilized by applying secrecy achieving strategies such as generalized singular value decomposition (GSVD)-based precoding [4], [5] and artificial noise (AN) injection [6], [7].

The secrecy capacity achieving input distribution over a Gaussian MIMO wiretap channel is proved to be Gaussian [4], [5]. While the optimal input covariance matrix is not available in closed form for the general case, effective numerical algorithms have been developed in [8] and [9] for its computation. As a result, much of the literature on physical layer security focuses on the Gaussian input assumption. However, an important scenario which is necessary to be studied when moving towards a practical implementation is the case where the channel inputs are drawn from discrete constellations. In this regard, single-antenna wiretap channels with discrete inputs have been studied in [10] by assuming an AWGN channel to the legitimate receiver and a fast fading channel to the eavesdropper. The case of MIMO wiretap channels with discrete inputs under quasi-static fading conditions has been investigated in [11]–[13]. While Bashar *et al.* [11] propose a GSVD-based precoding with the aid of the perfect channel state information (CSI) corresponding to both channels, the works in [12] and [13] propose AN-aided strategies for scenarios where the instantaneous CSI of the eavesdropper is not available at the transmitter. Bashar *et al.* [12] employ naive beamforming along with AN injection while considering single-antenna receivers. The strategy proposed in [13], on the other hand, relies on iterative maximization of an approximation to the instantaneous secrecy rate. In both of these studies, it has been shown that for maximizing secrecy rates at higher SNRs, it is desirable to allocate only a fraction of the total power for signal transmission and use the remaining power for AN injection.

In this work, we demonstrate that jointly optimizing the precoder matrix and the portion of power allocated to AN can outperform the solutions which rely on optimizing the precoder only. We introduce an iterative algorithm for direct maximization of the instantaneous secrecy rate which relies on a gradient descent based optimization of the precoder along with an exhaustive search for the optimal AN level. Noting that this approach possesses a high computational complexity due to the need for several evaluations of the mutual information expression (which lacks closed-form), we formulate a cut-off rate based approximation and use it as the precoder design

Manuscript received September 27, 2016; revised January 26, 2017; accepted March 20, 2017. Date of publication March 31, 2017; date of current version June 8, 2017. This work was supported by the Scientific and Technical Research Council of Turkey (TUBITAK) under Grant 113E223. This paper was presented in part at the IEEE International Symposium on Information Theory, Barcelona, Spain, July 2016 [1]. The associate editor coordinating the review of this paper and approving it for publication was L. Le. *(Corresponding Author: Tolga M. Duman.)*

The authors are with the Department of Electrical Engineering, Bilkent University, TR-06800 Ankara, Turkey (e-mail: aghdam@ee.bilkent.edu.tr; duman@ee.bilkent.edu.tr).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TWC.2017.2690279

metric. Once the precoder and AN using the cut-off rate based metric are obtained, an achievable secrecy rate expression is evaluated. Our numerical examples demonstrate that this scheme results in only a small loss in the achieved rates with respect to the direct maximization approach while requiring a significantly lower computational effort.

Employing AN along with the information signal is highly beneficial and can offer a significant enhancement in the achievable secrecy rates under the finite-alphabet input constraint. However, transmitting AN in the null space of the main channel (as done in [12] and [13]) is not applicable when the legitimate receiver is equipped with an equal or a larger number of antennas than the transmitter (i.e., when the null space dimensionality is zero). Hence, we also introduce a generalized AN aided transmission scheme in which the noise is injected in a direction which has a minimal effect on the received signal at the legitimate receiver.

Finally, for MIMO wiretap channels with large number of transmit antennas, we propose a per-group precoding scheme, as a further reduced complexity solution. Inspired by the idea proposed in [14], we divide the problem of precoder and AN design into a number of sub-problems via two-by-two grouping of the transmit antennas. Then, we obtain the optimal precoder and AN for each group independently. We demonstrate via examples that this scheme provides comparable secrecy rates to those achieved by the unconstrained precoders in spite of a drastic reduction in complexity.

The paper is organized as follows. Section II describes the system model under consideration. In Section III, we formulate the joint signal and AN design for directly maximizing the ergodic secrecy rates and we introduce the cut-off rate based optimization scheme. Section IV presents the generalized AN aided precoding. The per-group precoding scheme is explained in Section V. Section VI provides several numerical examples which demonstrate the efficacy of the proposed transmit signal design schemes, and finally, Section VII concludes the paper.

Notation: Throughout the paper, vectors and matrices are denoted with lowercase and uppercase bold letters, respectively. The expectation of a random variable X is represented by $\mathbb{E}_X\{.\}$ and $(.)^H$, $(.)^T$ and $\|\ .\ \|_F$ denote Hermitian, transpose and Frobenius norm operations, respectively.

II. System Model and Preliminaries

Consider a general MIMO wiretap channel. The transmitter, Alice, the legitimate receiver, Bob, and the eavesdropper, Eve, are assumed to be equipped with N_t, N_{r_b} and N_{r_e} antennas, respectively. The received vectors at Bob and Eve are given by

$$\mathbf{y} = \mathbf{H}_b\mathbf{x} + \mathbf{n}_y, \tag{1}$$

$$\mathbf{z} = \mathbf{H}_e\mathbf{x} + \mathbf{n}_z, \tag{2}$$

where $\mathbf{H}_b$ and $\mathbf{H}_e$ are the $N_{r_b} \times N_t$ and $N_{r_e} \times N_t$ channel matrices corresponding to the legitimate receiver's channel and the eavesdropper's channel, respectively. The elements of the channel matrix $\mathbf{H}_b$ are unit variance independent and identically distributed (i.i.d.) circularly symmetric complex Gaussian, i.e., $\mathcal{CN}(0, 1)$. We model the eavesdropper's channel as a doubly correlated fading MIMO channel, namely,

$$\mathbf{H}_e = \mathbf{\Psi}_r^{1/2}\hat{\mathbf{H}}_e\mathbf{\Psi}_t^{1/2}, \tag{3}$$

where $\mathbf{\Psi}_r$ and $\mathbf{\Psi}_t$ are the receive and transmit correlation matrices. $\hat{\mathbf{H}}_e$ is a complex matrix with i.i.d. zero mean unit variance circularly symmetric complex Gaussian entries. $\mathbf{n}_y$ and $\mathbf{n}_z$ are i.i.d. and they follow circularly symmetric complex Gaussian distributions, $\mathcal{CN}(0, \sigma^2_{\mathbf{n}_y})$ and $\mathcal{CN}(0, \sigma^2_{\mathbf{n}_z})$, respectively. Furthermore, $\mathbf{H}_b$, $\mathbf{H}_e$, $\mathbf{n}_y$ and $\mathbf{n}_z$ are independent. It is assumed that the fading process is ergodic. The legitimate receiver and the eavesdropper know their own channels perfectly. The transmitter knows the instantaneous channel of the legitimate receiver and only the statistics of the eavesdropper's CSI. In other words, the transmitter knows the correlation matrices $\mathbf{\Psi}_r$ and $\mathbf{\Psi}_t$ and the noise variance at the eavesdropper.

With the objective of maximizing the secrecy rates, the precoded signal is constructed as

$$\mathbf{x} = \mathbf{P}_D\mathbf{s} + \mathbf{P}_{AN}\mathbf{u}, \tag{4}$$

where $\mathbf{P}_D \in \mathbb{C}^{N_t \times N_t}$ is the data precoder matrix and $\mathbf{s} \in \mathbb{C}^{N_t \times 1}$ is the transmitted signal vector with zero mean and identity covariance matrix. Each element of $\mathbf{s}$ is drawn equiprobably from an M-ary signal constellation such as quadrature amplitude modulation (QAM) or phase shift keying (PSK). $\mathbf{P}_{AN}$ denotes the AN precoder matrix. In the remainder of the paper, we consider two scenarios for injection of the AN. First, we consider scenarios with $N_t > N_{r_b}$ where AN is injected along the null space of the main channel with $\mathbf{P}_{AN} = \frac{\alpha_{AN}}{\sqrt{N_t - N_{r_b}}}\mathbf{V}_b$ where $\mathbf{V}_b \in \mathbb{C}^{N_t \times (N_t - N_{r_b})}$ stands for an orthonormal basis for the null space of $\mathbf{H}_b$ and $\mathbf{u}$ denotes the noise term which follows $\mathcal{CN}(0, \mathbf{I}_{N_t - N_{r_b}})$. The portion of the power assigned to the AN is determined by the coefficient α_{AN}. For the scenarios with $N_t \leq N_{r_b}$, we consider a generalized AN similar to [15] where $\mathbf{u}$ follows $\mathcal{CN}(0, \mathbf{I}_{N_t})$ and the covariance of the AN signal $\mathbf{P}_{AN}\mathbf{u}$ is more general. We impose the power constraint

$$\mathrm{tr}(\mathbf{P}_D\mathbf{P}_D^H) + \mathrm{tr}(\mathbf{P}_{AN}\mathbf{P}_{AN}^H) \leq N_t. \tag{5}$$

The main user's and the eavesdropper's channels are both block fading. We assume that the channel gains are fixed during each coherence interval and they change independently from one coherence interval to the next. Furthermore, each coherence interval is large enough so that random coding arguments can be invoked, therefore an achievable ergodic secrecy rate can be calculated using [16]

$$\bar{R}_s = \mathbb{E}_{\mathbf{H}_b,\mathbf{H}_e}\Big\{\big(I(\mathbf{s};\mathbf{y}|\mathbf{H}_b) - I(\mathbf{s};\mathbf{z}|\mathbf{H}_e)\big)^+\Big\}, \tag{6}$$

where $I(\mathbf{s};\mathbf{y}|\mathbf{H}_b)$ and $I(\mathbf{s};\mathbf{z}|\mathbf{H}_e)$ are the instantaneous mutual information terms over the main channel and the eavesdropper's channel, respectively.[1] Irrespective of the precoder design approach, throughout the paper, we assume that the precoding is adopted along with the random coding and the

[1] This notation is different from the standard notation [17] (where $I(\mathbf{s};\mathbf{y}|\mathbf{H})$ stands for the mutual information averaged over $\mathbf{H}$). Here, $I(\mathbf{s};\mathbf{y}|\mathbf{H})$ refers to the instantaneous mutual information conditioned on the channel matrix $\mathbf{H}$.

rate adaptation schemes proposed in [16]. Hence, the ergodic secrecy rates are evaluated using (6) the achievability proof of which follows from the proof given in Appendix B in [16].

III. Joint Precoder and Artificial Noise Design

In this section, we propose a precoder and AN design algorithm for scenarios with $N_t > N_{r_b}$ where AN is injected in null-space of the main channel. Noting that $\mathbf{P}_{AN} = \frac{\alpha_{AN}}{\sqrt{N_t - N_{r_b}}}\mathbf{V}_b$, with the aid of the instantaneous knowledge of the main channel $\mathbf{H}_b$ and the statistical knowledge of the eavesdropper's channel, we seek to find the optimal $\mathbf{P}_D$ and α_{AN} which maximize the instantaneous secrecy rate given by

$$R_s = \mathbb{E}_{\mathbf{H}_e}\Big\{\big(I(\mathbf{s};\mathbf{y}|\mathbf{H}_b) - I(\mathbf{s};\mathbf{z}|\mathbf{H}_e)\big)^+\Big\}. \tag{7}$$

Noting that this is not tractable, we formulate a related optimization problem by considering a lower-bound on R_s (similar to what is done in [12]), given by

$$R_{s,l} = I(\mathbf{s};\mathbf{y}|\mathbf{H}_b) - \mathbb{E}_{\mathbf{H}_e} I(\mathbf{s};\mathbf{z}|\mathbf{H}_e). \tag{8}$$

Therefore, we solve the following problem:

$$\max_{\mathbf{P}_D,\ \alpha_{AN}} \quad R_{s,l} \tag{9}$$

$$\text{s.t. } \mathrm{tr}(\mathbf{P}_D\mathbf{P}_D^H) + \alpha_{AN}^2 \le N_t. \tag{10}$$

As we will see later, the maximization of this lower-bound is tractable and its solution serves to increase R_s as evidenced by extensive simulations.

We introduce two approaches: the first one relies on directly maximizing $R_{s,l}$ in (8) while the second one utilizes a cut-off rate based approximation of $R_{s,l}$. After obtaining the optimal precoder matrix $\mathbf{P}_D$ and the AN level α_{AN} from either of these schemes, we evaluate the ergodic secrecy rates using (6) to demonstrate the efficacy of the proposed solutions.

A. Direct Maximization of $R_{s,l}$

Due to the nonconvexity of the problem in (9)-(10), obtaining a globally optimal closed-form solution is intractable. However, it is possible to implement numerical algorithms which iteratively search for local maxima of the objective function. To solve for the precoder and AN, we first optimize over $\mathbf{P}_D$ for a fixed α_{AN}. In this case, the optimization problem becomes

$$\max_{\mathbf{P}_D} \quad \big(I(\mathbf{s};\mathbf{y}|\mathbf{H}_b) - \mathbb{E}_{\mathbf{H}_e} I(\mathbf{s};\mathbf{z}|\mathbf{H}_e)\big) \tag{11}$$

$$\text{s.t. } \mathrm{tr}(\mathbf{P}_D\mathbf{P}_D^H) \le N_t - \alpha_{AN}^2, \tag{12}$$

where the instantaneous mutual information over the main channel is given by [18]

$$I(\mathbf{s};\mathbf{y}|\mathbf{H}_b) = N_t \log M - \frac{1}{M^{N_t}} \times \sum_{i=1}^{M^{N_t}} \mathbb{E}_{\mathbf{n}_y} \log \sum_{j=1}^{M^{N_t}} \exp\bigg(-\frac{\|\mathbf{H}_b\mathbf{P}_D\mathbf{d}_{ij} + \mathbf{n}_y\|^2 - \|\mathbf{n}_y\|^2}{\sigma_{\mathbf{n}_y}^2}\bigg), \tag{13}$$

with $\mathbf{d}_{ij} = \mathbf{s}_i - \mathbf{s}_j$, where $\mathbf{s}_i$ is one of the M^{N_t} possible input vectors for $\mathbf{s} \in \mathbb{C}^{N_t \times 1}$. To compute the second term in (11), we note that the received vector at the eavesdropper is given by

$$\mathbf{z} = \mathbf{H}_e\mathbf{P}_D\mathbf{s} + \mathbf{n}_z', \tag{14}$$

where $\mathbf{n}_z'$ is the summation of the AN and the thermal noise. When $N_t > N_{r_b}$, the channel input is as given in (4), and we have $\mathbf{n}_z' = \alpha_{AN}\mathbf{W}\mathbf{u} + \mathbf{n}_z$ with $\mathbf{W} = \frac{1}{\sqrt{N_t - N_{r_b}}}\mathbf{H}_e\mathbf{V}_b$.

We consider two cases separately. When $N_{r_e} = 1$, $\mathbf{n}_z'$ is a Gaussian random variable and the average mutual information can be written as

$$\mathbb{E}_{\mathbf{H}_e} I(\mathbf{s};\mathbf{z}|\mathbf{H}_e) = N_t \log M - \frac{1}{M^{N_t}} \times \sum_{m=1}^{M^{N_t}} \mathbb{E}_{\mathbf{H}_e,\mathbf{n}_z'} \log \sum_{k=1}^{M^{N_t}} \exp\bigg(-\frac{\|\mathbf{H}_e\mathbf{P}_D\mathbf{d}_{mk} + \mathbf{n}_z'\|^2 - \|\mathbf{n}_z'\|^2}{\sigma_{\mathbf{n}_z'}^2}\bigg), \tag{15}$$

where $\sigma_{\mathbf{n}_z'}^2 = \sigma_{\mathbf{n}_z}^2 + \alpha_{AN}^2\mathbf{w}\mathbf{w}^H$ with $\mathbf{w} = \frac{1}{\sqrt{N_t - N_{r_b}}}\mathbf{h}_e\mathbf{V}_b$. If $N_{r_e} > 1$, $\mathbf{n}_z'$ becomes a zero-mean colored Gaussian noise vector with covariance matrix $\mathbf{K}_{\mathbf{n}_z'} = \mathbf{W}\mathbf{W}^H + \sigma_{\mathbf{n}_z}^2\mathbf{I}_{N_{r_e}}$. Therefore, in order to evaluate $\mathbb{E}_{\mathbf{H}_e} I(\mathbf{s};\mathbf{z}|\mathbf{H}_e)$, one can first whiten the noise term by pre-multiplying the received vector in (14) by $\mathbf{K}_{\mathbf{n}_z'}^{-\frac{1}{2}}$ resulting in

$$\mathbf{z}' = \mathbf{K}_{\mathbf{n}_z'}^{-\frac{1}{2}}\mathbf{H}_e\mathbf{P}_D\mathbf{s} + \mathbf{n}_z'', \tag{16}$$

where $\mathbf{n}_z''$ is a zero-mean additive white Gaussian noise vector with $\mathbf{K}_{\mathbf{n}_z''} = \mathbf{I}_{N_{r_e}}$, and using the expression

$$\mathbb{E}_{\mathbf{H}_e} I(\mathbf{s};\mathbf{z}'|\mathbf{H}_e) = N_t \log M - \frac{1}{M^{N_t}} \times \sum_{m=1}^{M^{N_t}} \mathbb{E}_{\mathbf{H}_e,\mathbf{n}_z''} \log \sum_{k=1}^{M^{N_t}} \exp\big(-\|\mathbf{K}_{\mathbf{n}_z'}^{-\frac{1}{2}}\mathbf{H}_e\mathbf{P}_D\mathbf{d}_{mk} + \mathbf{n}_z''\|^2 + \|\mathbf{n}_z''\|^2\big), \tag{17}$$

which is equivalent to $\mathbb{E}_{\mathbf{H}_e} I(\mathbf{s};\mathbf{z}|\mathbf{H}_e)$ as the transformation is one-to-one. The necessary conditions for optimality of the precoder matrix for maximization of $R_{s,l}$ with perfect main channel CSI and statistical CSI of the eavesdropper can be obtained as follows.

Proposition 1: The solution of the optimization problem (11)-(12) satisfies the following optimality criteria:

$$\frac{\log_2 e}{\sigma_{\boldsymbol{n}_y}^2}\big(\boldsymbol{H}_b^H\boldsymbol{H}_b\boldsymbol{P}_D\boldsymbol{\Delta}_b(\boldsymbol{P}_D)\big) - \frac{\log_2 e}{\sigma_{\boldsymbol{n}_z'}^2}\mathbb{E}_{\boldsymbol{H}_e}\Big\{\big(\boldsymbol{H}_e^H\boldsymbol{H}_e\boldsymbol{P}_D\boldsymbol{\Delta}_e(\boldsymbol{P}_D)\big)\Big\} = \theta\boldsymbol{P}_D, \tag{18}$$

$$\theta\big(tr(\boldsymbol{P}_D\boldsymbol{P}_D^H) + \alpha_{AN}^2 - N_t\big) = 0, \tag{19}$$

$$\theta \ge 0, \tag{20}$$

$$tr(\boldsymbol{P}_D\boldsymbol{P}_D^H) + \alpha_{AN}^2 - N_t \le 0, \tag{21}$$

where θ is the Lagrange multiplier corresponding to the constraint in (12) and $\boldsymbol{\Delta}_b(\boldsymbol{P}_D)$ and $\boldsymbol{\Delta}_e(\boldsymbol{P}_D)$ are the receive

Algorithm 1 Gradient Descent for Maximizing $R_{s,l}$

Consider different values for $\alpha_{AN} \in [0\ \sqrt{N_t}]$ and for each value of α_{AN}, repeat:
Step 1: Initialize $\mathbf{P}_{D_1}$ with constraint $\text{tr}(\mathbf{P}_{D_1}\mathbf{P}_{D_1}^H) \leq N_t - \alpha_{AN}^2$. Set step size μ and min. tolerance μ_{min}
Step 2: Set $k = 1$, compute $R_{s_1} = R_{s,l}(\mathbf{P}_{D_1})$
Step 3: Compute $\nabla_{\mathbf{P}_D} R_{s,l}(\mathbf{P}_D)$
Step 4: If $\mu \geq \mu_{min}$ goto Step 5, otherwise Stop algorithm and return $\mathbf{P}_{D_k}$
Step 5: Calculate $\hat{\mathbf{P}}_{D_k} = \mathbf{P}_{D_k} + \mu \nabla_{\mathbf{P}_D} R_{s,l}(\mathbf{P}_{D_k})$ and if $\text{tr}(\hat{\mathbf{P}}_{D_k}\hat{\mathbf{P}}_{D_k}^H) > N_t - \alpha_{AN}^2$, normalize as $\hat{\mathbf{P}}_{D_k} = \sqrt{\frac{N_t - \alpha_{AN}^2}{\text{tr}(\hat{\mathbf{P}}_{D_k}\hat{\mathbf{P}}_{D_k}^H)}}\hat{\mathbf{P}}_{D_k}$
Step 6: Compute $\hat{R}_s = R_{s,l}(\hat{\mathbf{P}}_{D_k})$; If $\hat{R}_s \geq R_{s_k}$ update $R_{s_{k+1}} = \hat{R}_s$ and $\mathbf{P}_{D_{k+1}} = \hat{\mathbf{P}}_{D_k}$ & goto Step 7, otherwise, let $\mu = 0.5\mu$ and goto Step 4
Step 7: $k = k + 1$ goto Step 3
Select α_{AN} and the corresponding optimal $\mathbf{P}_D$ which result in the maximum $R_{s,l}$.

minimum mean square error (MMSE) matrices at the legitimate receiver and the eavesdropper, respectively, and are given by [20]

$$\boldsymbol{\Delta}_b(\boldsymbol{P}_D) = \mathbb{E}\big\{(\boldsymbol{s} - \mathbb{E}\{\boldsymbol{s}|\boldsymbol{y}\})(\boldsymbol{s} - \mathbb{E}\{\boldsymbol{s}|\boldsymbol{y}\})^H\big\}, \tag{22}$$

$$\boldsymbol{\Delta}_e(\boldsymbol{P}_D) = \mathbb{E}\big\{(\boldsymbol{s} - \mathbb{E}\{\boldsymbol{s}|\boldsymbol{z}\})(\boldsymbol{s} - \mathbb{E}\{\boldsymbol{s}|\boldsymbol{z}\})^H\big\}. \tag{23}$$

Proof: This is slight modification of the [13, Proposition 1] and the proof follows from a similar Karush-Kuhn-Tucker (KKT) analysis as given in Appendix A in [13]. ■

In order to solve the optimization problem in (11)-(12), a gradient descent algorithm [19] can be employed. In this scheme, the precoder is updated as

$$\mathbf{P}_D(k+1) = \big[\mathbf{P}_D(k) + \mu \nabla_{\mathbf{P}_D} R_{s,l}(k)\big]^{\dagger}_{\text{tr}(\mathbf{P}_D\mathbf{P}_D^H) \leq N_t - \alpha_{AN}^2}, \tag{24}$$

where k and μ are the iteration index and the step-size of the update, respectively, and $[.]^{\dagger}_{\text{tr}(\mathbf{P}_D\mathbf{P}_D^H) \leq N_t - \alpha_{AN}^2}$ stands for the normalization which guarantees the feasibility of the solution at each step. More specifically, for cases where the updated precoder matrix $\hat{\mathbf{P}}_{D_k}$ does not satisfy the constraint in (12), similar to [20], we adopt a normalization as

$$\hat{\mathbf{P}}_{D_k} = \sqrt{(N_t - \alpha_{AN}^2)/\text{tr}(\hat{\mathbf{P}}_{D_k}\hat{\mathbf{P}}_{D_k}^H)}\hat{\mathbf{P}}_{D_k}, \tag{25}$$

which projects the solution onto the feasible set. The optimality of the precoder matrix which is obtained as the solution of gradient descent search can be proved by showing that (18) holds for a fixed $\theta \geq 0$.

So as to obtain the optimal $(\alpha_{AN}, \mathbf{P}_D)$, namely, to solve (9)-(10), we repeat this gradient descent algorithm for different values of α_{AN} and select the best $(\alpha_{AN}, \mathbf{P}_D)$ pair as described in Algorithm 1. For each value of α_{AN}, the algorithm should be repeated with multiple initializations of $\mathbf{P}_D$ to increase the likelihood for the gradient descent algorithm to converge to the globally optimal solution.

Note that the implementation of Algorithm 1 requires evaluation of the gradient of $R_{s,l}$ which is given by

$$\nabla_{\mathbf{P}_D} R_{s,l}(\mathbf{P}_D) = \frac{\log_2 e}{\sigma_{\mathbf{n}_y}^2}\big(\mathbf{H}_b^H \mathbf{H}_b \mathbf{P}_D \boldsymbol{\Delta}_b(\mathbf{P}_D)\big) - \mathbb{E}_{\mathbf{H}_e}\left\{\frac{\log_2 e}{\sigma_{\mathbf{n}'_z}^2}\Big(\mathbf{H}_e^H \mathbf{H}_e \mathbf{P}_D \boldsymbol{\Delta}_e(\mathbf{P}_D)\Big)\right\}. \tag{26}$$

B. Cut-Off Rate Based Approximation for $R_{s,l}$

The instantaneous and average mutual information terms in (8) lack closed-form expressions and involve multiple integrals. Specifically, computation of $R_{s,l}$ requires $2N_{r_e}(N_t + 1) + 2N_{r_b}$ integrals to be evaluated. Alternatively, to estimate $I(\mathbf{s}; \mathbf{y}|\mathbf{H}_b)$ and $\mathbb{E}_{\mathbf{H}_e} I(\mathbf{s}; \mathbf{z}|\mathbf{H}_e)$, one can take advantage of Monte Carlo methods, which require averaging over sufficiently large number of noise and channel samples, making it a computationally complex task.

So as to lower the computational complexity associated with the transmit signal design algorithm, closed form approximations of the mutual information can be employed [22], [23]. To this end, we propose to employ a cut-off rate based metric given by

$$R'_{s,l} = R_0^{(B)} - \bar{R}_0{}^{(E)}, \tag{27}$$

where $R'_{s,l}$ is an approximation of the instantaneous secrecy rate, with $R_0^{(B)}$ being the instantaneous cut-off rate for Bob, which is a valid lower-bound on the mutual information, given by [24]

$$R_0^{(B)} = 2N_t \log M - \log \sum_{i=1}^{M^{N_t}} \sum_{j=1}^{M^{N_t}} \exp\left(-\frac{\mathbf{d}_{ij}^H \mathbf{P}_D^H \mathbf{H}_b^H \mathbf{H}_b \mathbf{P}_D \mathbf{d}_{ij}}{4\sigma_{\mathbf{n}_y}^2}\right), \tag{28}$$

and $\bar{R}_0{}^{(E)}$ is the average cut-off rate over the eavesdropper's channel. The details of the derivation of $R_0^{(B)}$ is given in Appendix VII. If $N_{r_e} = 1$,

$$\bar{R}_0{}^{(E)} = 2N_t \log M - \mathbb{E}_{\mathbf{H}_e} \log \sum_{m=1}^{M^{N_t}} \sum_{k=1}^{M^{N_t}} \exp\left(-\frac{\mathbf{d}_{mk}^H \mathbf{P}_D^H \mathbf{H}_e^H \mathbf{H}_e \mathbf{P}_D \mathbf{d}_{mk}}{4\sigma_{\mathbf{n}'_z}^2}\right), \tag{29}$$

where $\sigma_{\mathbf{n}'_z}^2 = \sigma_{\mathbf{n}_z}^2 + \alpha_{AN}^2 \mathbf{w}\mathbf{w}^H$. Similar to the average mutual information, $\bar{R}_0{}^{(E)}$ can also be evaluated for the scenarios with $N_{r_e} > 1$ after noise whitening as in (16) resulting in

$$\bar{R}_0{}^{(E)} = 2N_t \log M - \mathbb{E}_{\mathbf{H}_e} \log \sum_{m=1}^{M^{N_t}} \sum_{k=1}^{M^{N_t}} \exp\left(-\frac{\mathbf{d}_{mk}^H \mathbf{P}_D^H \mathbf{H}_e^H \mathbf{K}_{\mathbf{n}'_z}^{-1} \mathbf{H}_e \mathbf{P}_D \mathbf{d}_{mk}}{4}\right). \tag{30}$$

Note that $R'_{s,l}$ is not an achievable rate. However, as we will see later, it can be used as an effective design metric to obtain the precoder matrices. Once the solution for the transmit signal

TABLE I
THE NUMBER OF MATRIX MULTIPLICATION STEPS

$I(\mathbf{s};\mathbf{y}\|\mathbf{H}_b)$	$R_0^{(B)}$	$\mathbb{E}_{\mathbf{H}_e} I(\mathbf{s};\mathbf{z}\|\mathbf{H}_e)$	$\bar{R}_0{}^{(E)}$
$12M^{2N_t}N_{\text{samp}}$	$5M^{2N_t}$	$12M^{2N_t}N_{\text{samp}}^2$	$8M^{2N_t}N_{\text{samp}}$

with this metric is obtained, the achievable secrecy rates are evaluated using (6).

In order to demonstrate that employing the cut-off rate based design metric in (27) instead of directly maximizing (8) can significantly reduce the computational complexity, we compare the matrix multiplication steps required for evaluation of $R_{s,l}$ and $R'_{s,l}$ in Table I. We assume that N_{samp} is the number of sample points required for an accurate estimation of the expectation operators, $\mathbb{E}_{\mathbf{n}}$ and $\mathbb{E}_{\mathbf{H}_e}$. For instance, it can be observed through numerical experiments that a sufficiently accurate estimation of the average mutual information $\mathbb{E}_{\mathbf{H}_e} I(\mathbf{s};\mathbf{z}|\mathbf{H}_e)$ requires averaging over at least $N_{\text{samp}} = 500$ realizations of noise and channel coefficients. Accordingly, Table I reveals that the computational complexity associated with calculation of $R'_{s,l}$ is considerably smaller than that of $R_{s,l}$. This can also be shown by comparing the CPU times required for evaluation of these metrics. As an example, for a $4 \times 4 \times 4$ wiretap channel with $M = 2$ and $N_{\text{samp}} = 500$, computation of $R_{s,l}$ and $R'_{s,l}$ takes 962.3693 and 0.6598 seconds, respectively, over an Intel Core-i7-4770, 3.4 GHz processor.

In order to maximize $R'_{s,l}$, we jointly optimize $\mathbf{P}_D$ and α_{AN} using Algorithm 1 by replacing $R_{s,l}$ with $R'_{s,l}$. Gradient of $R'_{s,l}$ is given in (31), shown at the bottom of this page where

$$\kappa_b = 4\sigma^2_{\mathbf{n}'_y}(\ln(2)) \sum_{i'=1}^{M^{N_t}} \sum_{j'=1}^{M^{N_t}} \exp\left(- \frac{\mathbf{d}^H_{i'j'} \mathbf{P}_D^H \mathbf{H}_b^H \mathbf{H}_b \mathbf{P}_D \mathbf{d}_{i'j'}}{4\sigma^2_{\mathbf{n}'_y}} \right), \tag{32}$$

and

$$\kappa_e = 4\sigma^2_{\mathbf{n}'_z}(\ln(2)) \sum_{m'=1}^{M^{N_t}} \sum_{k'=1}^{M^{N_t}} \exp\left(- \frac{\mathbf{d}^H_{m'k'} \mathbf{P}_D^H \mathbf{H}_e^H \mathbf{H}_e \mathbf{P}_D \mathbf{d}_{m'k'}}{4\sigma^2_{\mathbf{n}'_z}} \right). \tag{33}$$

The details of this derivation are given in Appendix VII.

For finite-alphabet inputs with equal SNR values at the legitimate receiver and the eavesdropper, a lower fraction of the power should be allocated to data transmission at higher SNRs [10], [12], [25]. This is due to the fact that, under finite-alphabet input constraints, transmission at full power for high SNRs allows the eavesdropper to acquire the maximum number of bits per channel use which results in zero secrecy. Hence, it is possible to limit the search space of the optimization in Algorithm 1 according to the SNR values. For higher SNRs, it is reasonable to search among larger α_{AN} values, whereas, at low SNRs, the search should be carried out among α_{AN}'s near 0.

IV. GENERALIZED AN-AIDED PRECODING

In the previous section, we introduced a precoder and AN design algorithm in which the AN is transmitted in conjunction with the information signal, and is designed to be orthogonal to the intended receiver in such a way that only the eavesdropper suffers a degradation in the receiver performance. However, such AN injection is not applicable when the number of antennas at the legitimate receiver is greater than the number of transmit antennas (i.e., when the null space dimensionality is 0), hence, we need to seek an alternative approach.

For the cases with $N_t \leq N_{r_b}$, we employ a joint precoder and generalized AN design scheme. The notion of generalized AN has been proposed in [15]. Dissimilar to the conventional AN which is only allowed to be transmitted in the null-space of $\mathbf{H}_b$, generalized AN possesses a more flexible covariance matrix.

The received vectors at the legitimate receiver and the eavesdropper are given as

$$\mathbf{y} = \mathbf{H}_b \mathbf{P}_D \mathbf{s} + \mathbf{H}_b \mathbf{P}_{AN} \mathbf{u}' + \mathbf{n}_y, \tag{34}$$

$$\mathbf{z} = \mathbf{H}_e \mathbf{P}_D \mathbf{s} + \mathbf{H}_e \mathbf{P}_{AN} \mathbf{u}' + \mathbf{n}_z, \tag{35}$$

where $\mathbf{P}_{AN}$ is the $N_t \times N_t$ precoder matrix for the AN signal and $\mathbf{u}'$ follows $\mathcal{CN}(0, \mathbf{I}_{N_t})$. The objective is to obtain optimal $\mathbf{P}_D$ and $\mathbf{P}_{AN}$ by solving the following problem

$$\max_{\mathbf{P}_D,\ \mathbf{P}_{AN}} \quad R_{s,l} \tag{36}$$

$$\text{s.t. } \text{tr}(\mathbf{P}_D \mathbf{P}_D^H) + \text{tr}(\mathbf{P}_{AN} \mathbf{P}_{AN}^H) \leq N_t, \tag{37}$$

where $R_{s,l}$ is given in (8). Since $\mathbf{n}'_y = \mathbf{H}_b \mathbf{P}_{AN} \mathbf{u}' + \mathbf{n}_y$ is colored with covariance $\mathbf{K}_{\mathbf{n}'_y} = \mathbf{H}_b \mathbf{P}_{AN} \mathbf{P}_{AN}^H \mathbf{H}_b^H + \sigma^2_{\mathbf{n}_y} \mathbf{I}_{N_{r_b}}$, the mutual information over the main channel can be calculated after whitening the noise by pre-multiplying (34) by $\mathbf{K}_{\mathbf{n}'_y}^{-\frac{1}{2}}$, i.e., obtaining

$$\mathbf{y}'' = \mathbf{K}_{\mathbf{n}'_y}^{-\frac{1}{2}} \mathbf{H}_b \mathbf{P}_D \mathbf{s} + \mathbf{n}''_y, \tag{38}$$

$$\nabla_{\mathbf{P}_D} R'_{s,l}(\mathbf{P}_D) = \frac{1}{\kappa_b} \sum_{i=1}^{M^{N_t}} \sum_{j=1}^{M^{N_t}} (\mathbf{H}_b^H \mathbf{H}_b \mathbf{P}_D \mathbf{d}_{ij} \mathbf{d}_{ij}^H) \exp\left(- \frac{\mathbf{d}_{ij}^H \mathbf{P}_D^H \mathbf{H}_b^H \mathbf{H}_b \mathbf{P}_D \mathbf{d}_{ij}}{4\sigma^2_{\mathbf{n}_y}} \right) - \sum_{m=1}^{M^{N_t}} \sum_{k=1}^{M^{N_t}} \mathbb{E}_{\mathbf{H}_e} \frac{1}{\kappa_e} (\mathbf{H}_e^H \mathbf{H}_e \mathbf{P}_D \mathbf{d}_{mk} \mathbf{d}_{mk}^H)$$
$$\times \exp\left(- \frac{\mathbf{d}_{mk}^H \mathbf{P}_D^H \mathbf{H}_e^H \mathbf{H}_e \mathbf{P}_D \mathbf{d}_{mk}}{4\sigma^2_{\mathbf{n}'_z}} \right), \tag{31}$$

Algorithm 2 Alternating Optimization for Maximizing $R_{s,l}$

Initialize $\lambda_h > \lambda_l = 0$, $\mathbf{P}_D$, $\mathbf{P}_{AN}$ and the convergence criteria ϵ_L and ϵ_λ:
Step 1: update $\lambda = \frac{1}{2}(\lambda_l + \lambda_h)$
Step 2: repeat:
obtain optimal $\mathbf{P}_D$ with fixed $\mathbf{P}_{AN}$ using gradient descent optimization
obtain optimal $\mathbf{P}_{AN}$ with fixed $\mathbf{P}_D$ using gradient descent optimization
until: consecutive values of $L(\mathbf{P}_D, \mathbf{P}_{AN}, \lambda)$ differ by less than ϵ_L
Step 3: If $\text{tr}(\mathbf{P}_D\mathbf{P}_D^H) + \text{tr}(\mathbf{P}_{AN}\mathbf{P}_{AN}^H) < N_t$ then update $\lambda_h = \lambda$
If $\text{tr}(\mathbf{P}_D\mathbf{P}_D^H) + \text{tr}(\mathbf{P}_{AN}\mathbf{P}_{AN}^H) > N_t$ then update $\lambda_l = \lambda$
until: two consecutive values of λ differ by less than ϵ_λ.

where $\mathbf{n}_y''$ is a zero-mean additive white Gaussian noise with unit variance, resulting in

$$I(\mathbf{s}; \mathbf{y}''|\mathbf{H}_b) = N_t \log M - \frac{1}{M^{N_t}} \times \sum_{i=1}^{M^{N_t}} \mathbb{E}_{\mathbf{n}_y''} \log \sum_{j=1}^{M^{N_t}} \exp\left(-\|\mathbf{K}_{\mathbf{n}_y'}^{-\frac{1}{2}}\mathbf{H}_b\mathbf{P}_D\mathbf{d}_{ij} + \mathbf{n}_y''\|^2 + \|\mathbf{n}_y''\|^2\right). \tag{39}$$

The expression for $\mathbb{E}_{\mathbf{H}_e} I(\mathbf{s}; \mathbf{z}|\mathbf{H}_e)$ is given in (17) where $\mathbf{K}_{\mathbf{n}_z'} = \mathbf{H}_e\mathbf{P}_{AN}\mathbf{P}_{AN}^H\mathbf{H}_e^H + \sigma_{\mathbf{n}_z}^2\mathbf{I}_{N_{re}}$.

In order to solve (36)-(37), we compute the Lagrangian of the problem as

$$L(\mathbf{P}_D, \mathbf{P}_{AN}, \lambda) = R_{s,l} + \lambda(N_t - \text{tr}(\mathbf{P}_D\mathbf{P}_D^H) - \text{tr}(\mathbf{P}_{AN}\mathbf{P}_{AN}^H)), \tag{40}$$

where λ is the Lagrange dual variable associated with the constraint in (37). For a fixed dual variable λ, the dual function is defined as

$$D(\lambda) = \max_{\mathbf{P}_D,\ \mathbf{P}_{AN}} L(\mathbf{P}_D, \mathbf{P}_{AN}, \lambda). \tag{41}$$

Then, the dual optimization problem can be written as

$$\min_{\lambda>0} D(\lambda). \tag{42}$$

Noting that $D(\lambda)$ is a convex function in λ, we update the dual variable using the bisection method similar to [26]. That is to say, when the subgradient $\nabla D(\lambda) = N_t - \text{tr}(\mathbf{P}_D\mathbf{P}_D^H) - \text{tr}(\mathbf{P}_{AN}\mathbf{P}_{AN}^H)$ is positive, we decrease λ in the bisection method; otherwise, we increase it. Indeed, we can interpret λ in (40) as a price for power which should increase when the power constraint is exceeded, and it should decrease otherwise.

In order to maximize the Lagrangian for a fixed λ, we employ a coordinate descent algorithm which relies on updating $\mathbf{P}_D$ and $\mathbf{P}_{AN}$ in an alternating fashion as described in Algorithm 2. After obtaining the optimal λ, the corresponding $(\mathbf{P}_D, \mathbf{P}_{AN})$ pair is used as the precoders. Obtaining the optimal $\mathbf{P}_D$ with a fixed $\mathbf{P}_{AN}$, and conversely, obtaining the optimal $\mathbf{P}_{AN}$ with a fixed $\mathbf{P}_D$ is carried out with the aid of gradient descent type solutions. Particularly, with a fixed $\mathbf{P}_{AN}$, the optimal $\mathbf{P}_D$ is obtained by using a similar procedure as described in steps 1 through 7 of Algorithm 1. In this case, the data precoder is updated as

$$\mathbf{P}_D(k+1) = \left[\mathbf{P}_D(k) + \mu\nabla_{\mathbf{P}_D} R_{s,l}(k)\right]^{\dagger}_{\text{tr}(\mathbf{P}_D\mathbf{P}_D^H)\le N_t - \text{tr}(\mathbf{P}_{AN}\mathbf{P}_{AN}^H)}. \tag{43}$$

Similarly, with a fixed $\mathbf{P}_D$, the AN precoder matrix $\mathbf{P}_{AN}$ is updated as

$$\mathbf{P}_{AN}(k+1) = \left[\mathbf{P}_{AN}(k) + \mu\nabla_{\mathbf{P}_{AN}} R_{s,l}(k)\right]^{\dagger}_{\text{tr}(\mathbf{P}_{AN}\mathbf{P}_{AN}^H)\le N_t - \text{tr}(\mathbf{P}_D\mathbf{P}_D^H)}. \tag{44}$$

We note that these steps are considerably simplified by replacing the mutual information with the cut-off rate expression as an approximate solution. The cut-off rate based approximation of the instantaneous secrecy rates can be calculated after whitening the additive noise terms in (34) and (35). That is to say, $R_0^{(B)}$ and $\bar{R}_0^{(E)}$ are calculated for the equivalent channels in (38) and (16), respectively. The gradient of the cut-off rate based approximation, $R_{s,l}'$ with respect to $\mathbf{P}_D$ is attained by replacing $\mathbf{H}_b$ and $\mathbf{H}_e$ in (31)-(33) by $\mathbf{K}_{\mathbf{n}_y'}^{-\frac{1}{2}}\mathbf{H}_b$ and $\mathbf{K}_{\mathbf{n}_z'}^{-\frac{1}{2}}\mathbf{H}_e$, respectively.

By optimizing the matrix $\mathbf{P}_{AN}$, the transmitter can highly suppress the useful signal at the eavesdropper whereas the degradation over the main channel is kept at a limited level as will be further illustrated via examples. It should be noted that, since the problem in (36)-(37) is non-convex, there is no guarantee that there is no duality gap, however, as will be demonstrated using numerical examples in Section VI, the approach is highly effective.

V. Per-Group Precoding for Large MIMO Wiretap Channels

The proposed transmit signal design algorithm in Section III-B possesses a lower complexity with respect to directly maximizing the secrecy rates due to the elimination of the averaging over channel and noise samples. However, this may still be a complex task in evaluation of the mutual information and in obtaining the optimal precoder, especially when N_t is large. With this motivation, we now provide a transmit signal design algorithm which reduces the search space for the optimal precoder and the AN, and hence the complexity.

The basic idea behind the proposed scheme is to decouple the data streams and the AN over parallel equivalent subchannels towards the legitimate receiver and the eavesdropper, and accordingly reduce the dimensionality of the search space for the transmit signal optimization. More specifically, such a decoupling will allow us to group subchannels in pairs and design the precoder and the AN for each pair separately. We note that the idea of per-group precoding has been recently proposed for capacity maximization in MIMO channels [14], [28] and for sum rate maximization of multiple access channels [29]. Here, we extend it to the case of MIMO wiretap channels.

In order to obtain a decoupled structure, we take advantage of GSVD of the pair $(\mathbf{H}_b, \mathbf{\Psi}_t^{1/2})$ which results in [27]

$$\mathbf{H}_b = \mathbf{U}_b \mathbf{\Sigma}_b [\mathbf{\Omega}^{-1}\ \mathbf{0}_{k\times N_t-k}] \mathbf{Q}^H, \tag{45}$$

$$\mathbf{\Psi}_t^{1/2} = \mathbf{U}_{\Psi_t} \mathbf{\Sigma}_{\Psi_t} [\mathbf{\Omega}^{-1}\ \mathbf{0}_{k\times N_t-k}] \mathbf{Q}^H, \tag{46}$$

where $\mathbf{U}_b \in \mathbb{C}^{N_{r_b}\times N_{r_b}}$, $\mathbf{U}_{\Psi_t} \in \mathbb{C}^{N_t\times N_t}$ and $\mathbf{Q} \in \mathbb{C}^{N_t\times N_t}$ are unitary matrices and $\mathbf{\Omega} \in \mathbb{C}^{k\times k}$ is a nonsingular matrix where $k = \text{rank}([\mathbf{H}_b^H \quad \mathbf{\Psi}_t^H]^H)$. $\mathbf{\Sigma}_b$ and $\mathbf{\Sigma}_{\Psi_t}$ are $N_{r_b} \times k$ and $N_t \times k$ matrices as [2]

$$\mathbf{\Sigma}_b = \begin{array}{c} \\ N_{r_b}-p-o \\ o \\ p \end{array} \begin{array}{c} \begin{array}{ccc} k-p-o & o & p \end{array} \\ \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_b & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} \end{bmatrix} \end{array}, \tag{47}$$

$$\mathbf{\Sigma}_{\Psi_t} = \begin{array}{c} \\ N_t-p-o \\ o \\ p \end{array} \begin{array}{c} \begin{array}{ccc} k-p-o & o & p \end{array} \\ \begin{bmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_{\Psi_t} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \end{array}, \tag{48}$$

where $p = \dim\big(\text{null}(\mathbf{H}_b)^\perp \cap \text{null}(\mathbf{\Psi}_t^{1/2})\big)$ and $o = \dim\big(\text{null}(\mathbf{H}_b) \cap \text{null}(\mathbf{\Psi}_t^{1/2})^\perp\big)$. $\mathbf{D}_b$ and $\mathbf{D}_{\Psi_t}$ are diagonal matrices with real and strictly positive entries. Also, by applying eigenvalue decomposition on $\mathbf{\Psi}_r^{1/2}$, we have

$$\mathbf{\Psi}_r^{1/2} = \mathbf{U}_{\Psi_r} \mathbf{\Sigma}_{\Psi_r} \mathbf{U}_{\Psi_r}^H, \tag{49}$$

where $\mathbf{U}_{\Psi_r}$ is a unitary matrix whose columns are eigenvectors of $\mathbf{\Psi}_r^{1/2}$, and $\mathbf{\Sigma}_{\Psi_r}$ represents a diagonal matrix whose diagonal entries are the eigenvalues of $\mathbf{\Psi}_r^{1/2}$. We now construct the channel input in (4) with the data and AN precoder matrices of the following form

$$\mathbf{P}_D = \mathbf{Q}\mathbf{B}\mathbf{P}_{Dg}, \tag{50}$$

$$\mathbf{P}_{AN} = \mathbf{Q}\mathbf{B}\mathbf{P}_{ANg}, \tag{51}$$

where

$$\mathbf{B} = \begin{bmatrix} \mathbf{\Omega}_{k\times k} & \mathbf{0}_{l\times l} \\ \mathbf{0}_{l\times k} & \mathbf{0}_{l\times k} \end{bmatrix}, \tag{52}$$

where $l = N_t - k$. Hence, the received signal vector at the legitimate receiver is given by

$$\mathbf{y} = \mathbf{U}_b \mathbf{\Sigma}_b \Big[\mathbf{I}_{k\times k} \quad \mathbf{0}_{k\times l}\Big] \big(\mathbf{P}_{Dg}\mathbf{s} + \mathbf{P}_{ANg}\mathbf{u}'\big) + \mathbf{n}_y, \tag{53}$$

By pre-multiplying (53) by $\mathbf{U}_b^H$ we obtain the following equivalent model

$$\tilde{\mathbf{y}} = \mathbf{\Sigma}_b \Big[\mathbf{I}_{k\times k} \quad \mathbf{0}_{k\times l}\Big] \big(\mathbf{P}_{Dg}\mathbf{s} + \mathbf{P}_{ANg}\mathbf{u}'\big) + \tilde{\mathbf{n}}_y, \tag{54}$$

where $\tilde{\mathbf{n}}_y = \mathbf{U}_b^H \mathbf{n}_y$ which has the same statistics as $\mathbf{n}_y$. Clearly, this strategy converts the main channel to a diagonal MIMO channel. Similarly, we obtain an equivalent diagonal channel towards Eve. To do this, consider the received signal at the eavesdropper

$$\mathbf{z} = \mathbf{U}_{\Psi_r} \mathbf{\Sigma}_{\Psi_r} \mathbf{U}_{\Psi_r}^H \hat{\mathbf{H}}_e \mathbf{U}_{\Psi_t} \mathbf{\Sigma}_{\Psi_t} \Big[\mathbf{I}_{k\times k} \quad \mathbf{0}_{k\times l}\Big] \big(\mathbf{P}_{Dg}\mathbf{s} + \mathbf{P}_{ANg}\mathbf{u}'\big) + \mathbf{n}_z'. \tag{55}$$

Pre-multiplying (55) by $\mathbf{U}_{\Psi_r}^H$ results in the following equivalent input-output relationship

$$\tilde{\mathbf{z}} = \mathbf{\Sigma}_{\Psi_r} \tilde{\mathbf{H}}_e \mathbf{\Sigma}_{\Psi_t} \Big[\mathbf{I}_{k\times k} \quad \mathbf{0}_{k\times l}\Big] \big(\mathbf{P}_{Dg}\mathbf{s} + \mathbf{P}_{ANg}\mathbf{u}'\big) + \tilde{\mathbf{n}}_z, \tag{56}$$

[2]The number of columns and rows of all the sub-matrices are shown explicitly in (47) and (48).

where $\tilde{\mathbf{n}}_z = \mathbf{U}_{\Psi_r}^H \mathbf{n}_z$ and $\tilde{\mathbf{H}}_e = \mathbf{U}_{\Psi_r}^H \hat{\mathbf{H}}_e \mathbf{U}_{\Psi_t}$ have same statistics as $\mathbf{n}_z$ and $\hat{\mathbf{H}}_e$, respectively [22].

Proposition 2 (Taken from [14]): For large-dimensional set-ups, the mutual information corresponding to the virtual channel input-output relationship in (56) is approximated as

$$\mathbb{E}_{\tilde{\boldsymbol{H}}_e} I(\boldsymbol{s}; \tilde{\boldsymbol{z}}|\tilde{\boldsymbol{H}}_e) \approx I(\boldsymbol{x}_{eq}; \boldsymbol{z}_{eq}|\sqrt{\boldsymbol{\Xi}}) + \log_2 \det(\boldsymbol{I}_{N_{r_e}} + \boldsymbol{R}_{eq}) - \gamma_{eq}\phi_{eq} \log e, \tag{57}$$

where $I(\boldsymbol{x}_{eq}; \boldsymbol{z}_{eq}|\sqrt{\boldsymbol{\Xi}})$ stands for the ergodic mutual information corresponding to the diagonal MIMO relationship

$$\boldsymbol{z}_{eq} = \boldsymbol{\Xi}^{1/2}\boldsymbol{x}_{eq} + \tilde{\boldsymbol{n}}_z, \tag{58}$$

with $\boldsymbol{x}_{eq} = \boldsymbol{P}_{Dg}\boldsymbol{s} + \boldsymbol{P}_{ANg}\boldsymbol{u}'$ and $\boldsymbol{\Xi}^{1/2}$ is a diagonal matrix which is a function of three auxiliary variables, $\boldsymbol{R}_{eq}$, γ_{eq} and ϕ_{eq} which are the solutions of the following coupled equations:

$$\boldsymbol{\Xi} = \gamma_{eq}\boldsymbol{\Sigma}_{\Psi_t}^2, \quad \boldsymbol{R}_{eq} = \phi_{eq}\boldsymbol{\Sigma}_{\Psi_r}^2,$$
$$\gamma_{eq} = tr((\boldsymbol{I}_{N_{r_e}} + \boldsymbol{R}_{eq})^{-1}\boldsymbol{\Sigma}_{\Psi_r}^2), \quad \phi_{eq} = tr(\boldsymbol{\Gamma}_{eq}\boldsymbol{\Sigma}_{\Psi_t}^2), \tag{59}$$

where $\boldsymbol{\Gamma}_{eq} = \mathbb{E}\big\{(\boldsymbol{s} - \mathbb{E}\{\boldsymbol{s}|\boldsymbol{z}_{eq}\})(\boldsymbol{s} - \mathbb{E}\{\boldsymbol{s}|\boldsymbol{z}_{eq}\})^H\big\}$ is the MMSE matrix corresponding to channel (58).

Using this result, we focus on the received vectors, $\tilde{\mathbf{y}}$ and $\mathbf{z}_{eq}$ given in (54) and (58), respectively, and design precoders which can partition the multi-antenna wiretap channel into a number of independent groups. More specifically, by employing precoders in (50) with $\mathbf{P}_{D,g}$ and $\mathbf{P}_{AN,g}$ taking the following form

$$\begin{bmatrix} P_{11} & P_{12} & 0 & 0 & \dots & 0 \\ P_{21} & P_{22} & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & \dots & P_{(N_t-1)(N_t-1)} & P_{(N_t-1)N_t} \\ 0 & 0 & \dots & \dots & P_{N_t\ (N_t-1)} & P_{N_tN_t} \end{bmatrix}, \tag{60}$$

we obtain two-by-two paired transmitted data streams. With this structure, the m^{th} data stream is transmitted along the $(2m-1)^{th}$ and $(2m)^{th}$ diagonal entries of $\mathbf{\Sigma}_b$ and $\boldsymbol{\Xi}^{1/2}$. Hence, we have

$$\tilde{\mathbf{y}}_m = \mathbf{\Sigma}_b \Big[\mathbf{I}_{k\times k} \quad \mathbf{0}_{k\times l}\Big] \big(\mathbf{P}_{Dg_m}\mathbf{s} + \mathbf{P}_{ANg_m}\mathbf{u}'\big) + \tilde{\mathbf{n}}_{y_m}, \tag{61}$$

$$\mathbf{z}_{eq_m} = \boldsymbol{\Xi}^{1/2}\big(\mathbf{P}_{Dg}\mathbf{s} + \mathbf{P}_{ANg}\mathbf{u}'\big) + \tilde{\mathbf{n}}_{z_m}, \tag{62}$$

$$\mathbf{P}_{Dg_m} = \begin{bmatrix} P_{D,(2m-1)(2m-1)} & P_{D,(2m-1)(2m)} \\ P_{D,(2m)(2m-1)} & P_{D,(2m)(2m)} \end{bmatrix}, \tag{63}$$

$$\mathbf{P}_{ANg_m} = \begin{bmatrix} P_{AN,(2m-1)(2m-1)} & P_{AN,(2m-1)(2m)} \\ P_{AN,(2m)(2m-1)} & P_{AN,(2m)(2m)} \end{bmatrix}, \tag{64}$$

where $m = 1, 2, \dots, \frac{N_t}{2}$. Finally, we note that the transmit signal design algorithm proposed in Section III can be applied to each group, separately. This is to say, instead of the original optimization problem in (9)-(10), the following $N_t/2$ subproblems can be solved.

$$\max_{\mathbf{P}_{Dgm},\ \mathbf{P}_{ANgm}} \quad \tilde{R}_{s_m}, \quad m = 1, 2, \dots, \frac{N_t}{2} \tag{65}$$

$$\text{s.t. } \text{tr}(\mathbf{P}'_{Dg_m}\mathbf{P}'^H_{Dg_m}) + \text{tr}(\mathbf{P}'_{ANg_m}\mathbf{P}'^H_{ANg_m}) \le 2, \tag{66}$$

where $\mathbf{P}'_{Dg_m}$ is an $N_t \times N_t$ matrix as

$$\mathbf{P}'_{Dg_m} = \mathbf{QB}\begin{bmatrix} \mathbf{P}_{Dg_m} & \mathbf{0}_{2\times\ (N_t-2)} \\ \mathbf{0}_{(N_t-2)\times 2} & \mathbf{0}_{(N_t-2)\times\ (N_t-2)} \end{bmatrix}, \quad (67)$$

$$\mathbf{P}'_{ANg_m} = \mathbf{QB}\begin{bmatrix} \mathbf{P}_{ANg_m} & \mathbf{0}_{2\times\ (N_t-2)} \\ \mathbf{0}_{(N_t-2)\times 2} & \mathbf{0}_{(N_t-2)\times\ (N_t-2)} \end{bmatrix}, \quad (68)$$

and $\tilde{R}_{s_m}$ is the difference between the instantaneous mutual information of the m^{th} group in (61) and the approximation in (57) for the m^{th} group in (62). After obtaining $\mathbf{P}_{D,g}$ and $\mathbf{P}_{AN,g}$ we construct the precoder matrices $\mathbf{P}_D$ and $\mathbf{P}_{AN}$ using (50) and (51). If these matrices do not satisfy the power constraint $\mathrm{tr}(\mathbf{P}_D\mathbf{P}_D^H) + \mathrm{tr}(\mathbf{P}_{AN}\mathbf{P}_{AN}^H) \leq N_t$, we adopt normalizations similar to (25) so that $\mathrm{tr}(\mathbf{P}_D\mathbf{P}_D^H) + \mathrm{tr}(\mathbf{P}_{AN}\mathbf{P}_{AN}^H) = N_t$.

Taking advantage of the per-group precoding scheme considerably reduces the computational complexity associated with the evaluation of mutual information or the cut-off rate, and accordingly, it simplifies the transmit signal design. For example, consider a $4 \times 4 \times 4$ MIMO wiretap channel with QPSK inputs. Using the equivalent channels $\tilde{\mathbf{y}}$ and $\mathbf{z}_{eq}$ with precoders in the form of (60) reduces the computational complexity (roughly) by a factor of $\frac{4^{2\times 4}}{2\times 4^{2\times 2}} = 128$ [14].

Finally, we emphasize that while we only consider the case of two-by-two pairing of the data streams in (60), N_g-by-N_g coupling of the data streams with $N_g > 2$ is also possible and it is expected to improve the precoder performance (with an increase in complexity). Namely, there is a trade-off between performance and complexity where $N_g = N_t$ coincides with the case of complete search adopted in Sections III and IV.

VI. Numerical Examples

In order to demonstrate the efficacy of the proposed signal design schemes, we provide several numerical examples. Throughout the simulations, equal noise levels are assumed at the legitimate receiver and the eavesdropper. The numerical results are provided for the scenarios with constant and fading main channels, separately. We set $\mu = 0.5$ and $\mu_{min} = 0.01$ in implementation of Algorithm 1, and we consider $\epsilon_L = \epsilon_\lambda = 0.01$ in execution of Algorithm 2.

A. Constant Main Channel

In this example, we assume that the main channel is fixed throughout the whole transmission period and is given as

$$\mathbf{H}_b = \begin{bmatrix} 0.5128 - 0.3239j & -0.8903 - 0.0318j \end{bmatrix}. \quad (69)$$

We consider 500 realizations of $\mathbf{H}_e$ for evaluation of the average mutual information for the eavesdropper and calculate the secrecy rate using (7). The eavesdropper's channel is assumed to be correlated according to (3) where $\mathbf{\Psi}_t$ and $\mathbf{\Psi}_r$ have exponentially decaying entries, i.e.,

$$[\mathbf{\Psi}_t]_{ij} = \rho_t^{|i-j|}, \quad \text{and } [\mathbf{\Psi}_r]_{ij} = \rho_r^{|i-j|}, \quad (70)$$

with $\rho_t = 0.9$ and $\rho_r = 1$.

Fig. 1 compares the ergodic secrecy rates for the three different transmit signal design algorithms. In implementation of Algorithm 1, we perform the search among 5 different values of α_{AN} which are selected according to the SNR values and

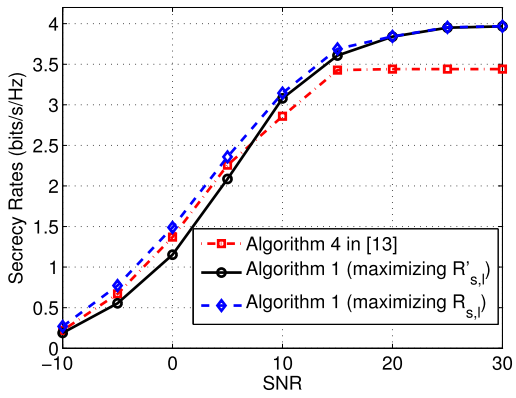

Fig. 1. Secrecy rates with QPSK inputs for a wiretap channel with $(N_t, N_{r_b}, N_{r_e}) = (2, 1, 1)$ with the main channel given in (69) and the eavesdropper channel with $\rho_t = 0.9$ and $\rho_r = 1$.

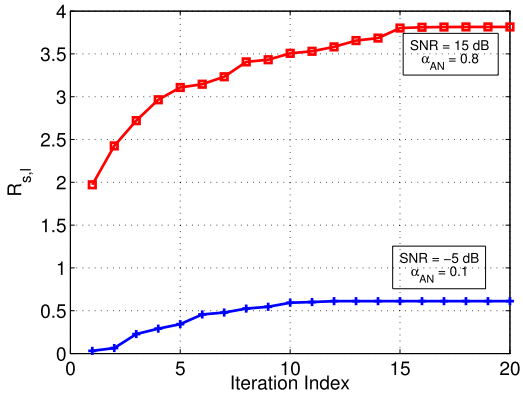

Fig. 2. Convergence of Algorithm 1 for the same setting as Fig. 1.

for each value of α_{AN}, we repeat the algorithm for 4 different initialization of $\mathbf{P}_D$. For a fair comparison, we run the algorithm proposed in [13] with 20 initializations. We observe that the maximization of $R_{s,l}$ while jointly optimizing the precoder matrix and the power allocated to the AN, i.e., employing Algorithm 1, yields higher secrecy rates compared to the scheme given in [13] which relies on the precoder optimization only. Furthermore, it can be inferred from Fig. 1 that the maximization of the cut-off rate based design metric $R'_{s,l}$ using Algorithm 1 incurs a relatively small loss with respect to the scheme proposed in [13] in low and moderate SNR values. While it is only approximate, the proposed algorithm even outperforms the algorithm in [13] in high SNRs due to the joint optimization of the precoder and AN. We also note that, the cut-off rate based optimization undergoes a loss of 28% and 2% in the achievable secrecy rates at SNR $= -5$ dB and at SNR $= 10$ dB, respectively, and it achieves almost the same performance as the direct maximization method at high SNRs. This comparable performance is achieved with a much lower computational complexity, e.g., for this example, the CPU times is reduced roughly by a factor of 1000.

Fig. 2 illustrates the convergence behavior of Algorithm 1 for given values of α_{AN} in different SNRs. Values of $R_{s,l}$ is depicted in each iteration and it can be observed that the proposed algorithm needs only a few iterations to converge. The final output of Algorithm 1 in these examples

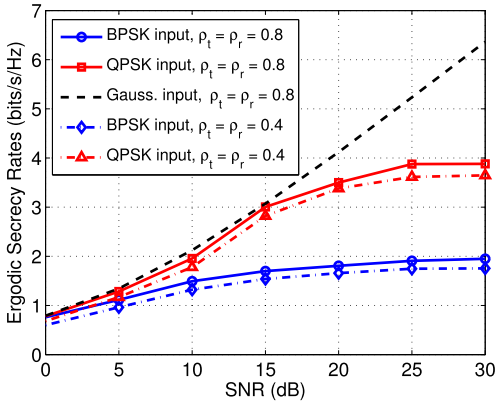

Fig. 3. Achievable secrecy rates for fading main channel, with different values of M ($N_t = 2$ and $N_{r_e} = N_{r_b} = 1$).

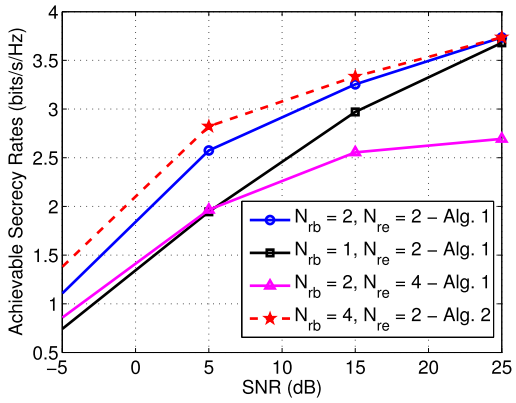

Fig. 4. Ergodic Secrecy Rates with different number of antennas at the receiver ends ($N_t = 4$, BPSK inputs).

are as follows. At SNR $= -5$dB, with $\alpha_{AN} = 0.1$,

$$\mathbf{P}_{D_1} = \begin{bmatrix} -0.1438 - 0.8947j & 0.0509 + 0.1416j \\ 0.4118 + 0.9718j & -0.0950 - 0.1522j \end{bmatrix}, \quad (71)$$

and at SNR $= 15$dB, the algorithm converges to

$$\mathbf{P}_{D_2} = \begin{bmatrix} 0.3124 - 0.2078j & -0.6076 + 0.4017j \\ -0.2857 + 0.2432j & 0.5536 - 0.4914j \end{bmatrix}, \quad (72)$$

with $\alpha_{AN} = 0.8$. The (local) optimality of these results is verified by showing that (18) holds as

$$\nabla R_{s,l}(\mathbf{P}_{D_1}) \simeq \theta_1 \mathbf{P}_{D_1}, \qquad \nabla R_{s,l}(\mathbf{P}_{D_2}) \simeq \theta_2 \mathbf{P}_{D_2}, \quad (73)$$

with $\theta_1 = 0.15$ and $\theta_2 = 0.22$, respectively.

B. Fading Main Channel

We now consider fading channels towards the legitimate receiver and the eavesdropper. In particular, we assume that the channel gains over both links change independently from one coherence interval to the next and accordingly, optimal $\mathbf{P}_D$ and α_{AN} (or $\mathbf{P}_{AN}$ when employing Algorithm 2) are obtained for each realization of $\mathbf{H}_b$ with the aid of the cut-off rate based approximations and the secrecy rate is averaged over 500 realizations according to (6). The eavesdropper's channel is assumed to be correlated as in (3) where the transmit and receive correlations follow (70).

Fig. 3 compares the secrecy rates achieved by transmissions with different channel inputs under two different correlation scenarios for the eavesdropper's channel. We observe that, when the SNR is sufficiently high, the proposed transmit signal design scheme provides achievable secrecy rates close to $N_t \log M$, i.e., the maximum rate which can be attained by the legitimate receiver assuming finite-alphabet inputs. As expected, higher secrecy rates are attained when the eavesdropper's channel is highly correlated. It is also observed that the achievable secrecy rates increase with M for a fixed number of transmit and receive antennas. Furthermore, Fig. 3 reveals an important difference between the secrecy behavior of the Gaussian vs. finite alphabet inputs. That is, while the achievable secrecy rates with Gaussian inputs increase monotonically with increasing SNR, it saturates for the latter scenario.

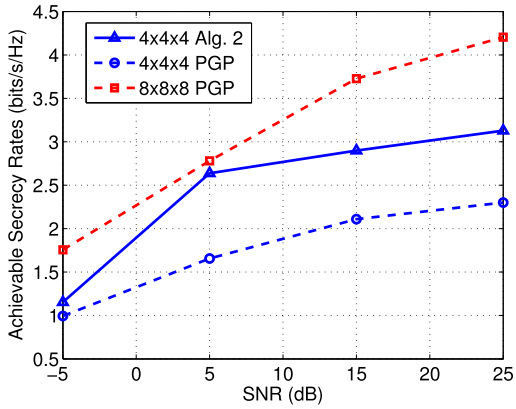

Fig. 5. Ergodic secrecy rates for BPSK inputs using Algorithm 2 with and without per-group precoding.

The effect of different number of receive antennas on the achievable secrecy rates is demonstrated in Fig. 4. The eavesdropper's channel is correlated with correlation matrices given in (70) with $\rho_t = 0.8$ and $\rho_r = 0.8$. Clearly, the proposed transmit signal design scheme is capable of providing positive secrecy rates even for the cases where the eavesdropper is equipped with a larger number of antennas than the legitimate receiver. Furthermore, it can be observed that increasing the number of receive antennas at the legitimate receiver results in increased achievable secrecy rates for fixed N_{r_e}. It should be noted that, for the case of $N_{r_b} = 4$, the generalized AN-aided precoding provides higher secrecy rates with respect to the cases with $N_{r_b} < 4$ in spite of a leakage of AN at the legitimate receiver. We attribute this to the fact that the more flexible covariance matrix of the generalized AN is more effective than the conventional AN in terms of suppressing the reception at the eavesdropper. Furthermore, thanks to the optimization in Algorithm 2 the leakage of AN at the legitimate receiver is small.

Fig. 5 compares the ergodic secrecy rates achieved with the proposed transmit signal design algorithms with and without per-group precoding. It can be inferred from this figure that the secrecy rates achieved by the solution of the relaxed problem undergoes a degradation with respect to the solution without per-group precoding. However, the considerably lower complexity of the per-group precoding technique makes possible obtaining suboptimal data and AN precoding matrices

for large set-ups which is intractable by directly employing Algorithms 1 and 2.

VII. Conclusions

In this paper, we have proposed iterative joint precoder and AN design schemes for maximization of ergodic secrecy rates for MIMO wiretap channels with finite-alphabet inputs with perfect and statistical CSI corresponding to the main channel and the eavesdropper's channel, respectively. We show that maximizing a cut-off rate based approximation of the instantaneous secrecy rate is a promising low-complexity alternative to the direct maximization approach. We have also studied a generalized AN-aided precoding scheme for the scenarios where injection of AN over the null-space of the main channel is not applicable. Our findings demonstrate that the problem of precoder and AN design is considerably simplified for large MIMO wiretap channels by two-by-two pairing of the transmit antennas and obtaining the optimal solution for each pair, separately. Examplary numerical results clearly show that the proposed transmit signal design methods provide positive secrecy rates in a variety of scenarios and yield an enhanced secrecy performance compared to the existing solutions in spite of their significantly lower computational complexities.

Appendix A
Derivation of the Cut-Off Rate Expression in (28)

The cut-off rate expression in (28) can be derived using the formula given in [24, eq. (4.3.34)], as

$$R_0^{(B)} = -\log \sum_{i=1}^{M^{N_t}} \sum_{j=1}^{M^{N_t}} \frac{1}{M^{2N_t}} \int p(\mathbf{y}|\mathbf{s}_i, \mathbf{H}_b)^{1/2} \times p(\mathbf{y}|\mathbf{s}_j, \mathbf{H}_b)^{1/2} d\mathbf{y}. \quad (74)$$

By substituting $p(\mathbf{y}|\mathbf{s}_i, \mathbf{H}_b)$ and $p(\mathbf{y}|\mathbf{s}_j, \mathbf{H}_b)$ in (74). Given $\mathbf{s}_i$ and $\mathbf{H}_b$ and for a fixed $\mathbf{P}_D$, $\mathbf{y}$ is a complex Gaussian random variable with zero-mean and variance $\mathbf{H}_b\mathbf{P}_D\mathbf{s}_i$. Hence, the conditional probability density function can be obtained as

$$p(\mathbf{y}|\mathbf{s}_i, \mathbf{H}_b) = \frac{1}{\pi^{N_{r_b}} \sigma_{\mathbf{n}_y}^{2N_{r_b}}} \exp\Big(- \frac{\|\mathbf{y} - \mathbf{H}_b\mathbf{P}_D\mathbf{s}_i\|^2}{\sigma_{\mathbf{n}_y}^2}\Big). \quad (75)$$

By plugging $p(\mathbf{y}|\mathbf{s}_i, \mathbf{H}_b)$ and $p(\mathbf{y}|\mathbf{s}_j, \mathbf{H}_b)$ into (28), we get (76). The integrand I_1 can be simplified as (77). Then, completion of the square in the exponent and substituting I_1 into (76) yields (78) (equations (76), (77) and (78) can be found at the bottom of this page).

The integral I_2 is equal to 1 since the integrand is a multi-variate Gaussian probability density function. Therefore, the final result simplifies to

$$R_0^{(B)} = 2N_t \log M - \log \sum_{i=1}^{M^{N_t}} \sum_{j=1}^{M^{N_t}} \exp\Big(- \frac{\|\mathbf{H}_b\mathbf{P}_D\mathbf{d}_{ij}\|^2}{4\sigma_{\mathbf{n}_y}^2}\Big), \quad (79)$$

where $\mathbf{d}_{ij} = \mathbf{s}_i - \mathbf{s}_j$ concluding the derivation of (28).

Appendix B
Derivation of $\nabla_{\mathbf{P}_D} R'_{s,l}$

We apply the matrix differentiation technique in [31] to derive the gradient of $R'_{s,l}$. As a first step, we evaluate the derivative of logarithm of sum of exponentials. Accordingly $\nabla R'_{s,l}$ can be written as

$$\begin{aligned}\nabla_{\mathbf{P}_D} R'_{s,l}(\mathbf{P}_D) &= \frac{1}{\kappa_b} \sum_{i=1}^{M^{N_t}} \sum_{j=1}^{M^{N_t}} \big(\nabla_{\mathbf{P}_D} \Upsilon_{b,ij}\big) \exp\Big(- \frac{\mathbf{d}_{ij}^H \mathbf{P}_D^H \mathbf{H}_b^H \mathbf{H}_b \mathbf{P}_D \mathbf{d}_{ij}}{4\sigma_{\mathbf{n}_y}^2}\Big) \\ &\quad - \sum_{m=1}^{M^{N_t}} \sum_{k=1}^{M^{N_t}} \mathbb{E}_{\mathbf{H}_e} \frac{1}{\kappa_e} \big(\nabla_{\mathbf{P}_D} \Upsilon_{e,mk}\big) \\ &\quad \times \exp\Big(- \frac{\mathbf{d}_{mk}^H \mathbf{P}_D^H \mathbf{H}_e^H \mathbf{H}_e \mathbf{P}_D \mathbf{d}_{mk}}{4\sigma_{\mathbf{n}'_z}^2}\Big). \end{aligned} \quad (80)$$

where κ_b and κ_e are as defined in (32) and (33), and also, $\Upsilon_{b,ij} = \mathbf{d}_{ij}^H \mathbf{P}_D^H \mathbf{H}_b^H \mathbf{H}_b \mathbf{P}_D \mathbf{d}_{ij}$ and $\Upsilon_{e,mk} = \mathbf{d}_{mk}^H \mathbf{P}_D^H \mathbf{H}_e^H \mathbf{H}_e \mathbf{P}_D \mathbf{d}_{mk}$. Using the definition of the complex gradient vector

$$[\nabla_{\mathbf{G}} f]_{ij} = \frac{\partial f}{\partial [\mathbf{G}^*]_{ij}}, \quad (81)$$

where the complex derivative of scalar function f is defined as $\frac{\partial f}{\partial g^*} = \frac{\partial Re\{f\}}{\partial g^*} + j\frac{\partial Im\{f\}}{\partial g^*}$, we obtain

$$\nabla_{\mathbf{P}_D} \Upsilon_{b,ij} = \mathbf{H}_b^H \mathbf{H}_b \mathbf{P}_D \mathbf{d}_{ij} \mathbf{d}_{ij}^H, \quad (82)$$

$$\nabla_{\mathbf{P}_D} \Upsilon_{e,mk} = \mathbf{H}_e^H \mathbf{H}_e \mathbf{P}_D \mathbf{d}_{mk} \mathbf{d}_{mk}^H. \quad (83)$$

By replacing (82) and (83) in (80), (31) follows.

$$R_0^{(B)} = 2N_t \log M - \log \sum_{i=1}^{M^{N_t}} \sum_{j=1}^{M^{N_t}} \underbrace{\int \Big[\frac{1}{\pi^{N_{r_b}} \sigma_{\mathbf{n}_y}^{2N_{r_b}}} \exp\Big(- \frac{\|\mathbf{y} - \mathbf{H}_b\mathbf{P}_D\mathbf{s}_i\|^2}{\sigma_{\mathbf{n}_y}^2}\Big)\Big]^{\frac{1}{2}} \Big[\frac{1}{\pi^{N_{r_b}} \sigma_{\mathbf{n}_y}^{2N_{r_b}}} \exp\Big(- \frac{\|\mathbf{y} - \mathbf{H}_b\mathbf{P}_D\mathbf{s}_j\|^2}{\sigma_{\mathbf{n}_y}^2}\Big)\Big]^{\frac{1}{2}} d\mathbf{y}}_{I_1}. \quad (76)$$

$$I_1 = \frac{1}{\pi^{N_{r_b}} \sigma_{\mathbf{n}_y}^{2N_{r_b}}} \exp\Big(- \frac{\|\mathbf{y}\|^2 + \frac{1}{2}\|\mathbf{H}_b\mathbf{P}_D\mathbf{s}_i\|^2 + \frac{1}{2}\|\mathbf{H}_b\mathbf{P}_D\mathbf{s}_j\|^2 - Re\{(\mathbf{H}_b\mathbf{P}_D\mathbf{S}_i)^*\mathbf{y}\} - Re\{(\mathbf{H}_b\mathbf{P}_D\mathbf{S}_j)^*\mathbf{y}\}}{\sigma_{\mathbf{n}_y}^2}\Big). \quad (77)$$

$$R_0^{(B)} = 2N_t \log M - \log \sum_{i=1}^{M^{N_t}} \sum_{j=1}^{M^{N_t}} \underbrace{\int \Big[\frac{1}{\pi^{N_{r_b}} \sigma_{\mathbf{n}_y}^{2N_{r_b}}} \exp\Big(- \frac{\|\mathbf{y} - \mathbf{H}_b\mathbf{P}_D(\frac{\mathbf{s}_i+\mathbf{s}_j}{2})\|^2}{\sigma_{\mathbf{n}_y}^2}\Big)\Big] d\mathbf{y}}_{I_2} \exp\Big(- \frac{\|\mathbf{H}_b\mathbf{P}_D(\frac{\mathbf{s}_i-\mathbf{s}_j}{2})\|^2}{\sigma_{\mathbf{n}_y}^2}\Big). \quad (78)$$

REFERENCES

[1] S. R. Aghdam and T. M. Duman, "Low complexity precoding for MIMOME wiretap channels based on cut-off rate," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Barcelona, Spain, Jul. 2016, pp. 2988–2992.

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

[3] A. Mukherjee, S. A. A. Fakoorian, J. Huang, and A. L. Swindlehurst, "Principles of physical layer security in multiuser wireless networks: A survey," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 3, pp. 1550–1573, 3rd Quart. 2014.

[4] A. Khisti and G. W. Wornell, "Secure transmission with multiple antennas I: The MISOME wiretap channel," *IEEE Trans. Inf. Theory*, vol. 56, no. 7, pp. 3088–3104, Jul. 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] R. Negi and S. Goel, "Secret communication using artificial noise," in *Proc. IEEE Veh. Technol. Conf.*, vol. 3. Dallas, TX, USA, Sep. 2005, pp. 1906–1910.

[7] S. Goel and R. Negi, "Guaranteeing secrecy using artificial noise," *IEEE Trans. Wireless Commun.*, vol. 7, no. 6, pp. 2180–2189, Jun. 2008.

[8] Q. Li, M. Hong, H. T. Wai, Y. F. Liu, W. K. Ma, and Z. Q. Luo, "Transmit solutions for MIMO wiretap channels using alternating optimization," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 9, pp. 1714–1727, Sep. 2013.

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

[10] Z. Li, R. Yates, and W. Trappe, "Achieving secret communication for fast Rayleigh fading channels," *IEEE Trans. Wireless Commun.*, vol. 9, no. 9, pp. 2792–2799, Sep. 2010.

[11] S. Bashar, Z. Ding, and C. Xiao, "On secrecy rate analysis of MIMO wiretap channels driven by finite-alphabet input," *IEEE Trans. Commun.*, vol. 60, no. 12, pp. 3816–3825, Dec. 2012.

[12] S. Bashar, Z. Ding, and C. Xiao, "On the secrecy rate of multi-antenna wiretap channel under finite-alphabet input," *IEEE Commun. Lett.*, vol. 15, no. 5, pp. 527–529, May 2011.

[13] Y. Wu, C. Xiao, Z. Ding, X. Gao, and S. Jin, "Linear precoding for finite-alphabet signaling over MIMOME wiretap channels," *IEEE Trans. Veh. Technol.*, vol. 61, no. 6, pp. 2599–2612, Jul. 2012.

[14] Y. Wu, C.-K. Wen, D. W. K. Ng, R. Schober, and A. Lozano, "Low-complexity MIMO precoding with discrete signals and statistical CSI," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Kuala Lumpur, Malaysia, May 2016, pp. 1–6.

[15] P.-H. Lin, S.-H. Lai, S.-C. Lin, and H.-J. Su, "On secrecy rate of the generalized artificial-noise assisted secure beamforming for wiretap channels," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 9, pp. 1728–1740, Sep. 2013.

[16] P. K. Gopala, L. Lai, and H. El Gamal, "On the secrecy capacity of fading channels," *IEEE Trans. Inf. Theory*, vol. 54, no. 10, pp. 4687–4698, Oct. 2008.

[17] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. New York, NY, USA: Wiley, 2006.

[18] C. Xiao and Y. R. Zheng, "On the mutual information and power allocation for vector Gaussian channels with finite discrete inputs," in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, New Orleans, LA, USA, Dec. 2008, pp. 1–5.

[19] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[20] D. P. Palomar and S. Verdu, "Gradient of mutual information in linear vector Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 141–154, Jan. 2006.

[21] S. R. Aghdam and T. M. Duman, "Physical layer security for space shift keying transmission with precoding," *IEEE Wireless Commun. Lett.*, vol. 5, no. 2, pp. 180–183, Apr. 2016.

[22] W. Zeng, C. Xiao, M. Wang, and J. Lu, "Linear precoding for finite-alphabet inputs over MIMO fading channels with statistical CSI," *IEEE Trans. Signal Process.*, vol. 60, no. 7, pp. 3134–3148, Jul. 2012.

[23] A. Yadav, M. Juntti, and J. Lilleberg, "Linear precoder design for doubly correlated partially coherent fading MIMO channels," *IEEE Trans. Wireless Commun.*, vol. 13, no. 7, pp. 3621–3635, Jul. 2014.

[24] S. G. Wilson, *Digital Modulation and Coding*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1995.

[25] T. Y. Liu, S. C. Lin, and Y. W. P. Hong, "On the role of artificial noise in training and data transmission for secret communications," *IEEE Trans. Inf. Forensics Security*, vol. 12, no. 3, pp. 516–531, Mar. 2017.

[26] H. Qin, "Power allocation and time-domain artificial noise design for wiretap OFDM with discrete inputs," *IEEE Trans. Wireless Commun.*, vol. 12, no. 6, pp. 2717–2729, Jun. 2013.

[27] 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.

[28] Y. Wu, C.-K. Wen, D. W. K. Ng, R. Schober, and A. Lozano. (Jun. 2016). "Low-complexity MIMO precoding for finite-alphabet signals." [Online]. Available: https://arxiv.org/abs/1606.03380

[29] Y. Wu, C.-K. Wen, C. Xiao, X. Gao, and R. Schober, "Linear precoding for the MIMO multiple access channel with finite alphabet inputs and statistical CSI," *IEEE Trans. Wireless Commun.*, vol. 14, no. 2, pp. 983–997, Feb. 2015.

[30] Z. Rezki, B. Alomair, and M. S. Alouini, "On the secrecy capacity of the MISO wiretap channel under imperfect channel estimation," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2014, pp. 1602–1607.

[31] A. Hjørungnes, *Complex-Valued Matrix Derivatives*. Cambridge, U.K.: Cambridge Univ. Press, 2011.

Sina Rezaei Aghdam (S'15) received the B.S. and M.S. degrees in electrical engineering from the Amirkabir University of Technology (Tehran Polytechnic) in 2011 and 2013, respectively. He is currently pursuing the Ph.D. degree in electrical engineering. He has been a Research Assistant with the Communication Theory and Applications Research Laboratory, Bilkent University, since 2013. His current research focuses on wireless communications and information theory.

Tolga M. Duman (S'95–M'98–SM'03–F'11) received the B.S. degree from Bilkent University, Ankara, Turkey, in 1993, and the M.S. and Ph.D. degrees from Northeastern University, Boston, MA, USA, in 1995 and 1998, respectively, all in electrical engineering. He has been with the Electrical Engineering Department, Arizona State University, as an Assistant Professor from 1998 to 2004, an Associate Professor from 2004 to 2008, and a Professor from 2008 to 2015. He is currently a Professor with the Electrical and Electronics Engineering Department, Bilkent University, and an Adjunct Professor with the School of ECE, Arizona State University. His current research interests are in systems, with particular focus on communication and signal processing, including wireless and mobile communications, coding/modulation, coding for wireless communications, data storage systems, and underwater acoustic communications.

Dr. Duman was a recipient of the National Science Foundation CAREER Award and the IEEE Third Millennium Medal. He served as an Editor of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS from 2003 to 2008, the IEEE COMMUNICATIONS SURVEYS AND TUTORIALS from 2002 to 2007, the IEEE TRANSACTIONS ON COMMUNICATIONS from 2007 to 2012, and the *Physical Communication* (Elsevier) from 2010 to 2016. He has been the Coding and Communication Theory Area Editor of the IEEE TRANSACTION ON COMMUNICATIONS since 2011, an Editor of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATION since 2016, and the Editor-in-Chief of the *Physical Communication* (Elsevier) since 2016.