A Perspective on the MIMO Wiretap Channel

Oggier, Frederique; Hassibi, Babak

2015

Oggier, F., & Hassibi, B. (2015). A Perspective on the MIMO Wiretap Channel. Proceedings of the IEEE, 103(10), 1874-1882.

<https://hdl.handle.net/10356/79243>

<https://doi.org/10.1109/JPROC.2015.2468077>

© 2015 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. The published version is available at: [<http://dx.doi.org/10.1109/JPROC.2015.2468077>].

Downloaded on 26 Apr 2023 21:24:30 SGT

A Perspective on the MIMO Wiretap Channel

Frédérique Oggier and Babak Hassibi

Abstract

A wiretap channel is a communication channel between a transmitter Alice and a legitimate receiver Bob, in the presence of an eavesdropper Eve. The goal of communication is to achieve reliability between Alice and Bob, but also confidentiality despite Eve's presence. Wiretap channels are declined in all kinds of flavors, depending on the underlying channels used by the three players: discrete memoryless channels, additive Gaussian noise channels, or fading channels, to name a few. In this survey, we focus on the case where the three players use multiple antenna channels with Gaussian noise to communicate. After summarizing known results for multiple input multiple output (MIMO) channels, both in terms of achievable reliable data rate (capacity) and code design, we introduce the MIMO wiretap channel. We then state the MIMO wiretap capacity, summarize the idea of the proof(s) behind this result, and comment on the insights given by the proofs on the physical meaning of the secrecy capacity. We finally discuss design criteria for MIMO wiretap codes.

Index Terms

Channel capacity; Confidentiality; Error probability; MIMO channels; Wiretap channels.

I. INTRODUCTION

Wireless communication systems with multiple antennas have evolved, over the last 15 years, from research products with a great potential, to a point-to-point technology commonly deployed and used. They are often referred, in technical jargon, to MIMO systems, where MIMO stands

F. Oggier is with the Division of Mathematical Sciences, Nanyang Technological University, Singapore. B. Hassibi is with the Department of Electrical Engineering, California Institute of Technology, Pasadena, USA. Emails: frederique@ntu.edu.sg, hassibi@caltech.edu B. Hassibi's work for this paper was supported in part by the National Science Foundation under grants CNS-0932428, CCF-1018927, CCF-1423663 and CCF-1409204, by a grant from Qualcomm Inc., by NASA's Jet Propulsion Laboratory through the President and Directors Fund, by King Abdulaziz University, and by King Abdullah University of Science and Technology.

for multiple inputs, multiple outputs. While the earliest years of MIMO research focused on understanding the increase in communication rate inherited from the increase in the number of antennas, and on coding strategies to reap the promised benefit in terms of data rate, attention has turned more recently towards scenarios beyond point-to-point communications, involving many users, and security aspects of MIMO communications.

In this survey paper, we zoom on one aspect of point-to-point MIMO security, that of confidentiality, in the information theoretic context of so-called *wiretap channels*. Information theoretic security refers to a form of security which relies on randomness and uncertainty, and uses tools from information theory for security analysis. It contrasts with the more widely known computational security, a form of security which relies on the assumption that the computational power of the adversary involved is bounded. Therefore computational security protocols typically involve functions that are hard to compute. On the contrary, an adversary is allowed to be computationally unbounded in an information theoretic security setting, where security then comes from randomness, and typically from communication over noisy channels. Since the communication channels among the players have a crucial role to play in the context of information theoretic security, so does the knowledge that a player has about the channel of the other players.

The aim of this survey paper is to present the *MIMO wiretap channel*, a channel which models communication between a transmitter, Alice, and a receiver, Bob, in the presence of an eavesdropper, Eve, all using several antennas. The goal is for the transmitter to come up with a strategy that guarantees reliability for Bob, but also confidentiality despite Eve's presence. Analysis of a wiretap channel consists of understanding the maximum amount of information that Alice and Bob can reliably exchange (that is, the channel *capacity*), subject to the constraint that the amount of information leaked to Eve is bounded by a chosen threshold. For this, it is necessary to first understand the scenario where Alice and Bob communicate over a MIMO channel, without worrying about a potential eavesdropper. We thus start this paper in Section II by explaining standard results about MIMO channels, both in terms of capacity - the maximum amount of information transmitted reliably from Alice to Bob - and pairwise probability of

Fig. 1. A multiple input multiple output (MIMO) channel: a transmitter communicates with a receiver, both are equipped with several antennas, n_t for the transmitter, and n_r for the receiver. The receiver is assumed to perfectly know the channel.

error, the probability that Bob erroneously decodes the message sent by Alice. This gives us the foundations to understand the MIMO wiretap channel. We will present two points of view on the MIMO wiretap channel, following the two points of view of a regular MIMO channel: in Section III, we state a capacity result for the MIMO wiretap channel, while in Section IV, we review an attempt to study MIMO wiretap channels from an error probability point of view. Together with the MIMO wiretap capacity, we provide some insights on the proof(s) that establish it, and the corresponding intuitions in terms of physical interpretation of the MIMO wiretap channel. Some open questions are suggested to conclude.

II. MIMO CHANNELS: CAPACITY AND PAIRWISE ERROR PROBABILITY

There are two typical points of view to understand MIMO communication: one is in terms of capacity, that establishes the limits of reliable communication, while the other is coding centric, and consists of finding a code design criterion to obtain reliable coding strategies. We will review both next.

A. On MIMO Capacity

Consider a transmitter with n_t transmit antennas, a receiver with n receive antennas, and a MIMO channel which can be represented by an $n \times n_t$ channel matrix $\mathbf{H}$, whose entries are i.i.d. complex Gaussian with zero mean and unit variance, and capture the channel between each transmit and receiver antenna (see Fig. 1). The $n \times 1$ received signal $\mathbf{Y}$ is modeled by

$$\mathbf{Y} = \mathbf{H}\mathbf{X} + \mathbf{V} \quad (1)$$

where $\mathbf{X}$ is the $n_t \times 1$ transmitted vector, and $\mathbf{V}$ is the $n \times 1$ additive Gaussian noise. To be precise, the noise vector is assumed to be a white circularly symmetric complex Gaussian vector, normalized so that its covariance matrix is the identity matrix $\mathbf{I}_N$. We assume that the transmitted signal $\mathbf{X}$ has covariance matrix $K_X = \mathbb{E}[\mathbf{X}\mathbf{X}^*] \succeq \mathbf{0}_{n_t}$ which satisfies the power constraint $\text{Tr}(K_X) = P$.

The capacity of a MIMO channel is informally the maximum amount of information that can be communicated reliably between the transmitter and the receiver. It turns out that the capacity depends on the channel knowledge (called Channel State Information (CSI)) at both the transmitter and receiver.

When the channel is constant and known perfectly at the transmitter and the receiver, the capacity C is

$$C = \max_{\substack{K_X \succeq \mathbf{0}_{n_t} \\ \text{Tr}(K_X) = P}} \log \det(\mathbf{I}_n + \mathbf{H}K_X\mathbf{H}^*). \quad (2)$$

In this case, Telatar [1] showed that the MIMO channel can be converted into parallel single antenna channels through singular value decomposition (SVD) of the channel matrix, that is, by writing $\mathbf{H} = BDA^*$, with B and A two unitary matrices, and D a diagonal matrix with the singular values λ_i^2 , $i = 1, \dots, \min(n_t, n)$ on its diagonal. The SVD therefore yields $\min(n_t, n)$ parallel channels with gains corresponding to the singular values λ_i^2 . Waterfilling the transmit power over these parallel channels leads to the optimal power allocation P_i for the i th channel

eigenmode:

$$P_i = \max \left\{ \mu - \frac{1}{\lambda_i^2}, 0 \right\}, \quad 1 \leq i \leq \min(n_t, n),$$

where $\mu \in \mathbb{R}$ is such that $\sum_{i=1}^{\min(n_t, n)} P_i = P$.

Consider now the case of perfect CSI at the receiver, but only channel distribution information at the transmitter, which happens for example when the receiver does not feed back any information on the channel and the transmitter only has some a priori distribution given e.g. by geographical considerations. It is known from [1] and [2] that the capacity is given by

$$C = \max_{\substack{K_X \succeq \mathbf{0}_{n_t} \\ \text{Tr}(K_X) = P}} \mathbb{E}_H [\log \det(\mathbf{I}_n + \mathbf{H}K_X\mathbf{H}^*)],$$

and that the best strategy consists of transmitting independent complex circular Gaussian symbols along the eigenvectors of K_X . The powers allocated to each eigenvector are given by the eigenvalues of K_X , and the optimum covariance matrix K_X is a scaled identity matrix, meaning that the transmit power is divided equally among all the transmit antennas. In summary, the capacity C is thus

$$C = \mathbb{E}_H \left[\log \det \left(\mathbf{I}_n + \frac{P}{n_t} \mathbf{H}\mathbf{H}^* \right) \right].$$

It turns out that when n_t, n grow, the capacity grows linearly with $\min(n_t, n)$, which justifies the increase in the number of antennas to speed up the data rate. Many other different notions of capacity are known for MIMO systems, see e.g. [3] for a survey, or [4], [5] for more complete references on MIMO communications. After establishing the limit of communications over MIMO channels, we discuss coding schemes.

B. Pairwise Probability of Error and Diversity

We now look at space-time coding, a coding strategy for MIMO channels, which exploits not only the presence of multiple antennas (captured in the term “space”), but also a stable channel over a period T of time, called *coherence time*, reflected in the term “time”. We keep the same assumptions as in (1). Only now, under the assumption that $\mathbf{H}$ is constant over a period of

$T \geq 1$, we get that the received signal $\mathbf{Y}$ is an $n \times T$ matrix given by

$$\mathbf{Y} = \mathbf{H}\mathbf{X} + \mathbf{V} \quad (3)$$

where the transmitted signal $\mathbf{X}$ is now an $n_t \times T$ matrix, and $\mathbf{V}$ is the $n \times T$ additive white Gaussian noise, whose independent entries are complex Gaussian random variables with zero mean and variance σ^2 (in this subsection, the variance of the noise is not normalized, to follow the convention mostly used in the literature).

This time, the goal is to design a codebook, that is a family of $n_t \times T$ matrices, into which the information symbols to be transmitted will be mapped. To obtain a code design criterion to construct such a codebook, a classical approach [6] is to compute an upper bound on the pairwise probability of error, i.e., the probability that, when a codeword $\mathbf{X}$ is transmitted, an error is made while applying the decoding rule $\hat{\mathbf{X}} = \min_{\mathbf{X}} \|\mathbf{Y} - \mathbf{H}\mathbf{X}\|^2$ and $\hat{\mathbf{X}}$ is erroneously chosen as the transmitted codeword.

It was shown [6] that

$$P(\mathbf{X} \rightarrow \hat{\mathbf{X}}) = \mathbb{E} \left[\mathcal{Q} \left(\frac{\|\mathbf{H}(\mathbf{X} - \hat{\mathbf{X}})\|}{\sqrt{2\sigma^2}} \right) \right] \leq \det \left[\mathbf{I}_{n_t} + \frac{(\mathbf{X} - \hat{\mathbf{X}})(\mathbf{X} - \hat{\mathbf{X}})^*}{4\sigma^2} \right]^{-n}. \quad (4)$$

where $\mathcal{Q}$ is the Gaussian tail function and the average is taken over all channel realizations $\mathbf{H}$. The role of the Gaussian tail function is easily understood: there is an error whenever the noise is strong enough to carry the transmitted matrix $\mathbf{X}$ closer to $\hat{\mathbf{X}}$ than to $\mathbf{X}$ (or more precisely, closer to $\mathbf{H}\hat{\mathbf{X}}$ than to $\mathbf{H}\mathbf{X}$).

It is well known that this upper bound is not very accurate for fading channels. Nevertheless, for high signal-to-noise ratios (small σ^2), this yields reasonably simple code design criteria [6]: diversity and coding gain, as summarized next.

If the rank of the codeword difference matrix $\mathbf{X} - \hat{\mathbf{X}}$ is equal to n_t for $\mathbf{X} \neq \hat{\mathbf{X}}$, the code is said to have *full diversity* and obtaining codes with full diversity is the first design criterion for a codebook with good performance in terms of error probability. The physical interpretation of diversity is that we can exploit all the $n_t n$ independent channels available in the MIMO system

to achieve reliability. The diversity also represents the slope of the pairwise error probability in a log-log scale plot.

In the case of full diversity codes, the second criterion is to maximize the *coding gain*

$$\left(\min_{\mathbf{X} \neq \hat{\mathbf{X}}} \det(\mathbf{X} - \hat{\mathbf{X}})(\mathbf{X} - \hat{\mathbf{X}})^* \right)^{1/n_t} \quad (5)$$

In the case of linear codes, where $\mathbf{X} - \hat{\mathbf{X}}$ is always a codeword itself, then the coding gain simplifies to

$$\left(\min_{\mathbf{X} \neq \mathbf{0}_{n_t \times T}} \det(\mathbf{X}\mathbf{X}^*) \right)^{1/n_t}. \quad (6)$$

See e.g. [7], [8], for examples of space-time codewords satisfying full diversity. The problem of designing space-time codes with full diversity and high coding gain has been an active research area, and numerous good space-time codes have been proposed (see e.g. [4]). Those proposed in [7], [8] furthermore carry a high amount of information symbols, thanks to their underlying lattice structure (where by lattice structure, we mean that every space-time codeword can be written as linear combinations of basis matrices, weighted by the information symbols to be transmitted).

III. THE CAPACITY OF THE MIMO WIRETAP CHANNEL

Now that basic results on MIMO communications have been recalled, we move on to introduce wiretap channels. We then concentrate on the MIMO wiretap channel, its capacity and design criteria for MIMO wiretap codes.

A. Background on Wiretap Channels

A *wiretap channel* is a broadcast channel introduced by Wyner [9] to study confidential transmission. A wiretap channel comprises three players, two legitimate players, Alice and Bob, and an eavesdropper Eve. Alice wants to communicate a secret message of length k to Bob, she encodes it into a codeword of length $n \geq k$, which is sent through a noisy broadcast channel, both Bob and Eve receive a noisy version, Eve's version is assumed to be "noisier" than that of

Bob. Confidentiality is measured in terms of Shannon entropy, which measures the uncertainty contained in a random variable. Given a secret message W^k that Alice wants to send, and the message Z^n that Eve receives, the entropy $H(W^k|Z^n)$ is the amount of uncertainty that Eve experiences regarding the secret W^k knowing Z^n . If the entropy of W^k given Z^n is the same as the entropy of W^k , then for Eve to guess the secret W^k knowing Z^n is the same as just guessing W^k , which means perfect confidentiality - or *perfect secrecy*. A weaker definition of perfect secrecy consists of asking instead that $H(W^k|Z^n)/n$ tends to the entropy rate $H(W^k)/n$ when n grows.

Wyner considered discrete memoryless channels, and computed a two-dimensional region that characterizes the maximal rate of communication between Alice and Bob, given the rate of leakage tolerated for Eve. When n grows, the *secrecy capacity* C_S between Alice and Bob is the channel capacity, that is the maximal amount of data that can be transmitted reliably to Bob, subject to the constraint of perfect confidentiality. Wyner showed that the secrecy capacity C_S is actually the difference of the capacity of the two users: $C_S = C_M - C_E$, where C_M is the capacity of the main channel, that is of the channel from Alice to Bob, and C_E is Eve's capacity. A key assumption is that the channel of Eve is a degraded version of the channel of the legitimate receiver. Informally this means that the noise experienced by Eve is added to that of Bob.

The work of Wyner can be put in perspective with that of Shannon [10], which, in a sense pessimistically, showed 20 years earlier that if a communication channel is noiseless, perfect secrecy costs in pure randomness as many bits as there are bits in the secret message to be transmitted. Wyner instead proved that perfect secrecy can be achieved at a lower cost, assuming Eve's channel has some disadvantage with respect to Bob's.

The first generalization of wiretap channels to continuous channels was proposed by Leung-Yan-Cheong and Hellman [11], for Gaussian channels, under the same assumption of degraded channels (the noise variance σ_M^2 of Bob is smaller than σ_E^2 , that of Eve), and it was shown that the secrecy capacity C_S is the difference of the two capacities: $C_S = C_M - C_E$. In [12], it was shown that the secrecy capacity is also the difference of the two capacities in the case of a single

antenna fading channel, under the assumption of asymptotically long intervals during which the channel is assumed constant, when the transmitter either knows both Bob and Eve channels, or only the legitimate channel.

The secrecy capacity for discrete memoryless channels, Gaussian channels, and fading channels, is thus for these three cases the difference of Bob's and Eve's capacities. This is a very neat result, which could be interpreted intuitively as follows: Alice should send as message a combination of random symbols and symbols actually containing information. The amount of information is the secrecy capacity, and there should be as many random symbols as there is capacity for Eve's channel. Indeed, if Eve decodes only random symbols, then the information symbols are confidential! How this should actually be done is far from obvious.

Before discussing the secrecy capacity of the MIMO wiretap channel, we would like to note that a first study of the security of MIMO channels was proposed by Hero [13]. In a different context than the wiretap channel, he introduced the so-called constraints of low probability of detection, and low probability of intercept, considering the scenario where the transmitter Alice and the receiver Bob are both informed about their channel while the eavesdropper is only informed about his. Such assumptions are not present for wiretap channels, where the eavesdropper is knowledgeable about everything.

B. The MIMO Wiretap Channel

A MIMO wiretap channel is a broadcast channel, where the channels between the transmitter Alice, and a legitimate receiver, Bob, as well as between Alice and Eve, are MIMO channels, as described in (1). More precisely, we suppose that the transmitter Alice has n_t transmit antennas, while Bob and Eve have n_M and n_E receive antennas (see Fig. 2) respectively. The $n_M \times 1$ received signal $\mathbf{Y}$ by Bob, and the $n_E \times 1$ signal $\mathbf{Z}$ spied by Eve, are modeled by

$$\mathbf{Y} = \mathbf{H}_M \mathbf{X} + \mathbf{V}_M \tag{7}$$

$$\mathbf{Z} = \mathbf{H}_E \mathbf{X} + \mathbf{V}_E$$

Fig. 2. A MIMO wiretap channel: a legitimate transmitter communicates with a legitimate receiver in the presence of an eavesdropper. All three players are equipped with several antennas, n_t for Alice, n_M for Bob, and n_E for Eve.

where $\mathbf{X}$ is the $n_t \times 1$ transmitted vector as in (1). The noise vectors $\mathbf{V}_M$ and $\mathbf{V}_E$ are $n_M \times 1$, respectively $n_E \times 1$ additive Gaussian noise vectors. They are assumed to be white circularly symmetric complex Gaussian vectors, normalized so that their covariance matrices are the identity matrices $\mathbf{I}_{n_M}$ and $\mathbf{I}_{n_E}$.

The channel matrices $\mathbf{H}_M$ and $\mathbf{H}_E$ are respectively $n_M \times n_t$ and $n_E \times n_t$ fixed channel matrices such that $\mathbf{H}_M^* \mathbf{H}_M \succ \mathbf{0}_{n_t}$, $\mathbf{H}_E^* \mathbf{H}_E \succ \mathbf{0}_{n_t}$, that is, they are non-singular. They are assumed to be known at the transmitter. Recall from (2) that if we were only interested in communication between Alice and Bob, ignoring Eve, the capacity C_M (of the main channel) would be

$$C_M = \max_{\substack{K_X \succeq \mathbf{0}_{n_t} \\ \text{Tr}(K_X) = P}} \log \det(\mathbf{I}_{n_M} + \mathbf{H}_M K_X \mathbf{H}_M^*).$$

In the presence of Eve however, the goal of a wiretap transmission scheme is to prevent Eve to obtain any information about the secret message sent by Alice to Bob. Formally, as explained above, this means that the entropy of the secret knowing Eve's received signal should be the same as the entropy of the message itself. The *secrecy capacity* of the MIMO wiretap channel, namely the maximum amount of information that can be transmitted reliably between Alice and

Bob, under perfect secrecy was independently proven to be [14], [15], [16]

$$C_S = \max_{\substack{K_X \succeq \mathbf{0}_{n_t} \\ \text{Tr}(K_X)=P}} \log \det(\mathbf{I}_{n_M} + \mathbf{H}_M K_X \mathbf{H}_M^*) - \log \det(\mathbf{I}_{n_E} + \mathbf{H}_E K_X \mathbf{H}_E^*). \quad (8)$$

A particular case for two antennas was proven in [17].

We notice that the first term $\log \det(\mathbf{I}_{n_M} + \mathbf{H}_M K_X \mathbf{H}_M^*)$ corresponds to the expression which is maximized (over all K_X with $\text{Tr}(K_X) = P$) in the capacity of Bob's channel, while $\log \det(\mathbf{I}_{n_E} + \mathbf{H}_E K_X \mathbf{H}_E^*)$ would be similarly optimized to obtain Eve's channel capacity. It is worth nothing that the maximum over K_X of the difference in (8) is not equal to the difference between the maximum of the two terms. This result could have been expected based on the earlier known results, but for the remarkable fact that in the MIMO case, the channel is *not degraded*, unlike for the previous known cases. In other words, the result holds irrespectively of whether it is true that $\mathbf{H}_M^* \mathbf{H}_M \succeq \mathbf{H}_E^* \mathbf{H}_E$.

To prove a capacity result, a typical approach consists of two steps: find an upper bound on the capacity, that is, show that the capacity can never be larger than a given bound, and then explicit a transmission strategy (which may not be simple, since it may require to discuss code design) whose data rate achieves this upper bound.

That the perfect secrecy rate

$$R_S = \max_{\substack{K_X \succeq \mathbf{0}_{n_t} \\ \text{Tr}(K_X)=P}} \log \det(\mathbf{I}_{n_M} + \mathbf{H}_M K_X \mathbf{H}_M^*) - \log \det(\mathbf{I}_{n_E} + \mathbf{H}_E K_X \mathbf{H}_E^*)$$

is achievable is the simplest of the two steps needed to prove the secrecy capacity of the MIMO wiretap channel, and was already proven in [18]. This follows from the fact that once K_X is chosen, the difference between the resulting mutual informations to the legitimate user and eavesdropper can be secretly transmitted.

Perfect secrecy requests that when n grows, $H(W^k|Z^n)$ tends to $H(W^k)$, or, more weakly, that the secrecy rate is actually equal to the entropy rate of the transmitted message conditioned on Eve's knowledge. Since $H(W^k|Z^n)$ is upper bounded by a term that essentially depends on $I(X^n; Y^n|Z^n)$ (as done e.g. in [12]), the main work to upper bound R_S consists of upper

bounding $I(X; Y|Z)$, the mutual information between X and Y conditioned on the knowledge of Z .

In [14], an upper bound was computed using the following steps. First

$$I(X; Y|Z) \leq \max_{K_X \succeq \mathbf{0}_{n_t}} \tilde{I}(X; Y|Z)$$

where

$$\tilde{I}(X; Y|Z) = \log \det \left(\mathbf{I}_{n_t} + [\mathbf{H}_M^*, \mathbf{H}_E^*] \begin{bmatrix} \mathbf{I}_{n_M} & A \\ A^* & \mathbf{I}_{n_E} \end{bmatrix} \begin{bmatrix} \mathbf{H}_M \\ \mathbf{H}_E \end{bmatrix} K_X \right) - \log \det(\mathbf{I}_{n_E} + \mathbf{H}_E K_X \mathbf{H}_E^*)$$

and A is an $n_M \times n_E$ matrix representing the correlation between $\mathbf{V}_M$ and $\mathbf{V}_E$. This upper bound is obtained by assuming that the legitimate receiver knows both its channel and that of the eavesdropper, in which case the optimal input distribution is luckily Gaussian. The trick then is that the secrecy capacity does not depend on A , however, since $\tilde{I}(X; Y|Z)$ is actually concave in K_X and convex in A , we may use A to tighten the upper bound as follows:

$$I(X; Y|Z) \leq \min_A \max_{K_X \succeq \mathbf{0}_{n_t}} \tilde{I}(X; Y|Z) = \max_{K_X \succeq \mathbf{0}_{n_t}} \min_A \tilde{I}(X; Y|Z).$$

It turns out that the optimization problem $\min_A \tilde{I}(X; Y|Z)$ can be solved in closed form expression, and the optimal value $\tilde{A}$ which provides a local minimum can be computed. The knowledge of $\tilde{A}$ brings an important piece of information: the optimal covariance matrix for our original problem $\max_{K_X \succeq \mathbf{0}_{n_t}} \tilde{I}(X; Y|Z)$ is a low rank matrix, or rank r , which in turn allows us to conclude, solving a Riccati equation, that

$$\tilde{A} = (\mathbf{H}_E(\mathbf{H}_M^* \mathbf{H}_M)^{-1} \mathbf{H}_M^* B \mathbf{H}_M U_X V, \mathbf{H}_E(\mathbf{H}_E^* \mathbf{H}_E)^{-1} \mathbf{H}_M^* W)(B \mathbf{H}_M U_X V, W)^{-1}$$

where $K_X = U_X U_X^*$ and $B = (\mathbf{H}_M K_X \mathbf{H}_M^* + \mathbf{I})^{-1}$. We notice some freedom in the $n_M \times (r_M - r)$ matrix W and the $r \times r$ matrix V . It is enough that V, W are such that $\mathbf{I} - \tilde{A} \tilde{A}^* \succ \mathbf{0}$, under which condition the proof of the secrecy capacity is concluded by showing that this optimal matrix $\tilde{A}$ makes the converse match the achievability. It is intuitively clear that the optimal input covariance matrix $\tilde{K}_X$ has to be low rank, it means that no signal is sent in some directions

favorable to the eavesdropper.

In [15], inspired by the secrecy capacity of non-degraded channels for discrete memoryless channels [19], the starting point for upper bounding the secrecy capacity is

$$C_S \leq \min_{K_V \succeq \mathbf{0}} \max_{K_X \succeq \mathbf{0}} I(X; Y|Z)$$

where

$$K_V = \begin{bmatrix} \mathbf{I}_{n_M} & A \\ A^* & \mathbf{I}_{n_E} \end{bmatrix}$$

and A is as above the correlation matrix between the noises $\mathbf{V}_M$ and $\mathbf{V}_E$. The first step of the proof of [15] consists of proving that the upper bound is an optimization problem which is convex in K_V and concave in K_X , and has a finite optimal solution $(\tilde{K}_X, \tilde{K}_V)$. Now the problem reduces to compute this solution, which once is found, can be put in the upper bound to obtain the desired result. In the process of computing $(\tilde{K}_X, \tilde{K}_V)$, an interesting property is derived:

$$\tilde{A}^* \mathbf{H}_M \tilde{U}_X = \mathbf{H}_E \tilde{U}_X$$

for all full column rank $\tilde{U}_X$ such that $\tilde{U}_X \tilde{U}_X^* = \tilde{K}_X$, and $\tilde{A}$ is the optimal correlation between $\mathbf{V}_M$ and $\mathbf{V}_E$ corresponding to $\tilde{K}_V$. The consequence of this result is that Bob can simulate Eve's channel by computing

$$\begin{aligned} \mathbf{Z}' &= \tilde{A}^* \mathbf{Y} + \mathbf{W} \\ &= \tilde{A}^* \mathbf{H}_M \tilde{U}_X \mathbf{X}' + \tilde{A}^* \mathbf{V}_M + \mathbf{W} \\ &= \mathbf{H}_E \tilde{U}_X \mathbf{X}' + \mathbf{V}'_E \\ &= \mathbf{H}_E \mathbf{X} + \mathbf{V}'_E \end{aligned}$$

where $\mathbf{W}$ has covariance $\mathbf{I} - \tilde{A}^* \tilde{A}$ which is independent of $\mathbf{Y}$, $\mathbf{X} = \tilde{U}_X \mathbf{X}'$ has covariance $\mathbf{I}$, and $\mathbf{V}'_E$ has covariance $\mathbf{I}$. The effective channel to the eavesdropper turns out to be a degraded version of that to the intended receiver. This means that no information should be sent where the eavesdropper has a more advantageous signal than Bob.

The commonalities between the two proofs in [14], [15] are technically the use of the cross correlation matrix A for the purpose of tightening the upper bound, and the recognition that the upper bound is a convex-concave problem with an optimal finite solution, and from an intuitive point of view, the understanding that the optimal input covariance matrix consists of transmitting no information along any direction where the eavesdropper observes a stronger signal than the legitimate receiver Bob. This also explains why the upper bound obtained by assuming that Bob also knows Eve's channel turns out to be the capacity (the rate cannot be improved by giving Bob Eve's channel).

The proof of [16] is of different nature. It does start with the secrecy capacity of non-degraded channels for discrete memoryless channels [19], and the goal of [16] is to evaluate this expression, which holds for continuous alphabets under a power constraint, for the case of MIMO wiretap channels. In order to do so, the key idea is to exploit a relationship between the derivative of mutual information, and the minimum mean-square error (MMSE). In doing so, an interesting result is proven which relates to the interpretation of the optimal covariance matrix, namely, that it is possible to maximize the secrecy capacity by considering an input covariance matrix K_X that does not take into account the subchannels for which the eavesdropper has an advantage over the legitimate recipient Bob.

IV. AN ERROR PROBABILITY APPROACH

We are next interested in deriving a code design criterion for MIMO wiretap codes. The MIMO wiretap channel considered is then as that of (7), except that we assume a coherence time T typically larger than 1. Keeping the notations of (7), the noise vectors $\mathbf{V}_M$ and $\mathbf{V}_E$ are $n_M \times T$, respectively $n_E \times T$ additive Gaussian noise vectors, both with coefficients that have zero mean, and respective variance σ_M^2 and σ_E^2 . The fading coefficients of $\mathbf{H}_M$ and $\mathbf{H}_E$ are complex Gaussian i.i.d. random variables, and in particular $\mathbf{H}_E$ has covariance matrix $\sigma_{\mathbf{H}_E}^2 \mathbf{I}_{n_E}$.

We know what are the code designs for space-time codes to achieve a low probability of error for Bob, as was shortly reviewed in Section II, and keeping the convention of Section II, we do not normalize the covariance matrices for deriving code design criteria. Therefore the question

is how to achieve confidentiality. To explore an answer, we next describe an encoding scheme called coset coding, which is a well accepted encoding to provide confusion in the presence of an eavesdropper, whose original version is already present in Wyner’s work [9]. We then discuss a translation of the pairwise error probability approach to bound Eve’s probability of successfully decoding a confidential message. The question of designing wiretap MIMO codes is still an area of research, and other techniques have been studied, e.g. in [20], where the authors considered combining linear operations along with successive interference cancellation to reduce the problem of designing MIMO wiretap codes to that of designing wiretap codes for the Gaussian wiretap channel. We would like to emphasize here that the construction of Gaussian wiretap codes is itself difficult (see e.g. [21]). In fact, finding capacity achieving wiretap codes is in general a highly non-trivial problem (see [22] for recent results using polar codes).

A. Wiretap Coset Coding

Any wiretap coding scheme relies on the injection of randomness at the transmitter, whose goal is to increase Eve’s confusion, that is to amplify the noise experienced by Eve. This is practically done with coset encoding, as proposed in the original work by Wyner [9]. The idea of coset encoding is as follows: partition the set of codewords into subsets (called cosets), label each coset by information symbols, and then pick a codeword uniformly at random within a coset for the actual transmission. The term “coset” more precisely fits the cases where codewords come from a linear code, or where lattice points are sent: a linear code and a lattice have a group structure, thus a subcode and a sublattice are subgroups, and we recover the usual meaning of cosets of a subgroup (translates of a subgroup that partition the group).

When the channel is continuous, and lattice codes are used, coset coding becomes lattice coset coding. This fits the case of MIMO channels as well, since most well known families of space-time codes indeed have an underlying lattice structure (e.g. [7], [8]). To perform lattice coset coding, Alice chooses a lattice Λ_M , which refers to the underlying lattice structure of the space-time code used to transmit to Bob. Therefore a space-time code $\mathbf{X}$ is seen as a lattice point. Alice partitions Λ_M into a union of disjoint cosets $\Lambda_E + C$, where Λ_E is a sublattice

of Λ_M that serves for Eve's confusion and C is a vector which encodes her data. Alice then randomly chooses a random vector $R \in \Lambda_E$ so that the transmitted lattice point $\mathbf{X} \in \Lambda_M$ is finally $\mathbf{X} = R + C \in \Lambda_E + C$.

B. Eve's Probability of Correct Decision

In this subsection, we mimic the steps involved in Bob's pairwise probability of error, for Eve's probability, assuming coset encoding at the transmitter. We would like to emphasize that the goal of these computations is to show the result of applying a standard pairwise probability of error to attempt to characterize a confidentiality code design criterion. However, it remains an open question whether the obtained criterion can be linked to the original characterization of secrecy.

We have that Eve's probability $P_{c,e,\mathbf{H}_E}$ of correctly decoding a space-time code $\mathbf{X}$ sent by Alice is [23]

$$P_{c,e,\mathbf{H}_E} \leq \frac{\text{vol}(\Lambda_M)}{(2\pi\sigma_E^2)^{n_t T}} \det(\mathbf{H}_E \mathbf{H}_E^*)^T \sum_{\mathbf{X} \in \Lambda_E} e^{-\|\mathbf{H}_E \mathbf{X}\|_F^2 / 2\sigma_E^2}$$

where $\text{vol}(\Lambda_M)$ is the volume of the lattice Λ_M . We notice the sum in the upper bound, which is the result of Alice using coset coding. Therefore, averaging over the possible channel realizations $\mathbf{H}_E$ we get that Eve's average probability $\bar{P}_{c,e} = \mathbb{E}_{\mathbf{H}_E} [P_{c,e,\mathbf{H}_E}]$ of correct decision is

$$\bar{P}_{c,e} \leq \frac{\text{vol}(\Lambda_M)}{(2\pi\sigma_E^2)^{T n_t} (2\pi\sigma_{\mathbf{H}_E}^2)^{n_E n_t}} \sum_{\mathbf{X} \in \Lambda_E} \int_{\mathbb{C}^{n_E \times n_t}} \det(\mathbf{H}_E \mathbf{H}_E^*)^T e^{-\text{Tr}\left(\mathbf{H}_E^* \mathbf{H}_E \left[\frac{1}{2\sigma_{\mathbf{H}_E}^2} \mathbf{I}_{n_t} + \frac{1}{2\sigma_E^2} \mathbf{X} \mathbf{X}^*\right]\right)} d\mathbf{H}_E$$

which, after integration, becomes

$$\boxed{\bar{P}_{c,e} \leq C_{\text{MIMO}} \gamma_E^{T n_t} \sum_{\mathbf{X} \in \Lambda_E} \det(\mathbf{I}_{n_t} + \gamma_E \mathbf{X} \mathbf{X}^*)^{-n_E - T}} \quad (9)$$

where we let $\gamma_E = \frac{\sigma_{\mathbf{H}_E}^2}{\sigma_E^2}$ denote Eve's signal-to-noise ration (SNR), Γ_{n_t} denotes the Gamma function, and

$$C_{\text{MIMO}} = \frac{\text{vol}(\Lambda_M) \Gamma_{n_t}(n_E + T)}{\pi^{n_t T} \Gamma_{n_t}(n_E)}.$$

In order to design a good lattice code for the MIMO wiretap channel with respect to confi-

dentiality, we try to derive a code design criterion from (9), similarly to the way diversity and coding gain were obtained for reliability:

$$\bar{P}_{c,e} \leq C_{\text{MIMO}} \gamma_E^{T n_t} \left[1 + \sum_{\mathbf{X} \in \Lambda_E \setminus \{0\}} \det(\mathbf{I}_{n_t} + \gamma_E \mathbf{X} \mathbf{X}^*)^{-n_E - T} \right].$$

We suppose that the space-time code used to transmit data to Bob is fully diverse, since a wiretap code needs to ensure reliability for Bob, namely, if $\mathbf{X} \neq 0$ and $T \geq n_t$ then, the rank $\mathbf{X}$ is n_t . To minimize Eve's average probability of correct decoding, a first simplified design criterion is then

$$\min_{\Lambda_E} \sum_{\mathbf{X} \in \Lambda_E \setminus \{0\}} \frac{1}{\det(\mathbf{X} \mathbf{X}^*)^{n_E + T}}. \quad (10)$$

We note that the tightness of the proposed bound depends on the values of γ_E . Furthermore, this bound was computed using an infinite lattice Λ_E , which often gives an infinite sum. However, it is possible to obtain the same bound, but with a sum over $\mathbf{X}$ inside a finite subset carved from Λ_E .

The interesting point of this approach is that, similarly to what happened for reliability, the end result of the upper bound computation is a code design which is tractable. Such a code design is not obvious from the information theoretic analysis. For the case of Gaussian wiretap codes, a similar pairwise probability of error design criterion was proposed in terms of theta series of the lattice used to design a lattice wiretap code [24], and it was shown that the same theta series expression turns out to be a bound on the mutual information between Alice's confidential message, and Eve's intercepted one [25]. However, whether the design criterion presented here for MIMO wiretap codes actually relates (if at all) to the original criterion for secrecy proposed by Wyner is open, in that the work of [25] has not (yet) been generalized to fading channels. The connection provided in the Gaussian case suggests that the probability error criterion could be related to Wyner's one, otherwise, a high probability of error alone for Eve is not enough in itself to guarantee which amount of information is leaked. On the other hand, as mentioned in the beginning of Section IV, MIMO wiretap codes achieving the secrecy capacity and satisfying

Wyner's original criterion for secrecy have been proposed in [20].

V. OPEN QUESTIONS AND PERSPECTIVES

The MIMO wiretap channel is a communication channel, that establishes an information theoretic framework to study the limits of reliable and confidential communication between two legitimate users Alice and Bob, in the presence of an eavesdropper Eve. It is a generalization of the wiretap channel introduced by Wyner for discrete memoryless channels in the seventies. The secrecy capacity of MIMO wiretap channels was established a few years ago, under the assumption that Alice knows both Bob and Eve's channels.

Since then, different generalizations have been proposed, by going beyond the point-to-point setting (e.g., by introducing relays), or by varying the wiretapper model (e.g., by considering an honest but curious adversary). There are also obvious questions, aligned with those addressed in the regular MIMO setting, namely, how do relaxations on the channel knowledge affect the channel capacity? In the same spirit, MIMO research has in the meantime evolved, giving rise to evolutionary MIMO techniques, comprising cutting-edge and future MIMO techniques such as coordinated (multipoint) MIMO, massive MIMO, and mmWave MIMO. Evolutionary MIMO wiretap channels are open research questions (see [26] for one view point on physical layer security in the context of massive MIMO channels).

The error probability point of view on MIMO wiretap channels is more recent, and opens many questions, starting from that of relating the code design criteria extracted from the error probability to the mutual information between the secret message and Eve's intercepted message. Progresses in that direction are available for Gaussian channels [25], and would be desirable for MIMO channels as well. The actual design of good MIMO wiretap codes remains mainly open as well.

Addressing the connection between mutual information and error probability for wiretap channels arguably resembles opening the Pandora box: as of now, the topic of security using wiretap codes remains fairly disputed among researchers. While there is no question about the establishment of the MIMO wiretap capacity, how useful (MIMO) wiretap codes really are,

whether they can be used in practice, whether a high error probability for Eve is good enough a security guarantee, remain questions on which researchers still are mostly divided. From this point of view, the most challenging open question of all could be to actually design MIMO wiretap codes that can be implemented, and whose security can then be really experimented.

REFERENCES

- [1] E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecomm. ETT*, vol. 10, no. 6, Nov. 1999.
- [2] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Commun.*, Kluwer Academic Press, no. 6, pp. 311-335, 1998.
- [3] A. Goldsmith, S. A. Jafar, N. Jindal, S. Vishwanath, "Capacity Limits of MIMO Channels", *IEEE Journal on Selected Areas in Communications*, Vol. 21, No. 5, June 2003.
- [4] A. Paulraj, R. Nabar, D. Gore, "Introduction to Space-Time Wireless Communications", *Cambridge University Press*, 2008.
- [5] E. Biglieri, R. Calderbank, A. Constantinides, A. Goldsmith, A. Paulraj, H. V. Poor, "MIMO Wireless Communications", *Cambridge University Press*, 2010.
- [6] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Spacetime codes for high data rate wireless communication: Performance analysis and code construction", *IEEE Transactions on Information Theory*, vol. 44, no. 2, March 1998.
- [7] B. Hassibi and B. M. Hochwald, "High-rate codes that are linear in space and time," *IEEE Trans. on Inform. Theory*, vol. 48, no. 7, July 2002.
- [8] F. Oggier, J.-C. Belfiore and E. Viterbo, "Cyclic Division Algebras: A Tool for SpaceTime Coding", *Foundations and Trends in Communications and Information Theory*, Vol. 4, No. 1, Nov. 2007.
- [9] A.D. Wyner, "The wire-tap channel," *Bell. Syst. Tech. J.*, vol. 54, October 1975.
- [10] C. Shannon, "Communication Theory of Secrecy Systems", *Bell System Technical Journal*, vol. 28(4), page 656-715, 1949.
- [11] S.K. Leung-Yan-Cheong, M.E. Hellman, "The Gaussian Wire-Tap Channel", *IEEE Trans. on Information Theory*, vol. 24, July 1978.
- [12] P. Gopala, L. Lai, and H. El Gamal, "On the Secrecy Capacity of Fading Channels," *IEEE Transactions on Information Theory*, Vol. 54, No. 10, Oct. 2008.
- [13] A. O. Hero, "Secure Space-Time Communication," *IEEE Trans. on Info Theory*, Vol. 49, No. 12, Dec. 2003.
- [14] F. Oggier, B. Hassibi, "The Secrecy Capacity of the MIMO Wiretap Channel", *IEEE Trans. on Information Theory*, vol. 57, no. 8, Aug. 2011.
- [15] A. Khisti, G.W. Wornell, "Secure Transmission With Multiple AntennasPart II: The MIMOME Wiretap Channel", *IEEE Trans. on Information Theory*, vol. 56, no. 11, Nov. 2010.
- [16] R. Bustin, R. Liu, H.V. Poor, S. Shamai (Shitz), "An MMSE Approach to the Secrecy Capacity of the MIMO Gaussian Wiretap Channel", *EURASIP Journal on Wireless Communications and Networking*, Volume 2009, Article ID 370970.
- [17] S. Shafiee, N. Liu, and S. Ulukus, "Towards the secrecy capacity of the Gaussian MIMO wire-tap channel: The 2-2-1 channel," *IEEE Trans. Inform. Theory*, vol. 55, no. 9, pp. 4033-4039, Sep. 2009.

- [18] Z. Li, W. Trappe, R. Yates, “Secret communication via multi-antenna transmission”, in the proceedings of *Conference on Information Sciences and Systems (CISS)*, March 2007.
- [19] I. Csiszár and J. Körner, “Broadcast channels with confidential messages”, *IEEE Trans. Inform. Theory*, vol. 24, pp. 339-348, 1978.
- [20] A. Khina, Y. Kochman, A. Khisti, “Decomposing the MIMO wiretap channel”, *IEEE International Symposium on Information Theory (ISIT)*, 2014.
- [21] H. Tyagi, A. Vardy, “Explicit Capacity-Achieving Coding Scheme For The Gaussian Wiretap Channel”, *IEEE International Symposium on Information Theory (ISIT)*, 2014.
- [22] H. Mahdavi, A. Vardy, “Achieving the Secrecy Capacity of Wiretap Channels Using Polar Codes”, *IEEE Trans. on Information Theory*, vol. 57, no 10, Oct. 2011.
- [23] J.-C. Belfiore, F. Oggier, “An Error Probability Approach to MIMO Wiretap Channels”, *IEEE Trans. on Communications*, vol. 61, no. 8, June 2013.
- [24] J.-C. Belfiore, F. Oggier, “Secrecy Gain: a Wiretap Lattice Code Design”, *International Symposium on Information Theory and Applications (ISITA)*, 2010.
- [25] C. Ling, L. Luzzi, J.-C. Belfiore, D. Stehlé “Semantically Secure Lattice Codes for the Gaussian Wiretap Channel”, <http://arxiv.org/abs/1210.6673>.
- [26] T. Dean, A. Goldsmith, “Physical-layer cryptography through massive MIMO”, in the proceedings of *IEEE Information Theory Workshop (ITW)* 2013, Sevilla.

Frédérique Oggier obtained a Ph.D. thesis in Mathematics (2005) from the Swiss Federal Institute of Technology, Lausanne (EPFL). She is currently an Associate Professor at the School of Physical and Mathematical Sciences, Nanyang Technological University (NTU), Singapore. Her main research interests are in applied algebra to coding problems arising in wireless communications, distributed networked storage as well as information theoretic security.

Babak Hassibi was born in Tehran, Iran, in 1967. He received the B.S. degree from the University of Tehran in 1989, and the M.S. and Ph.D. degrees from Stanford University in 1993 and 1996, respectively, all in electrical engineering.

He has been with the California Institute of Technology since January 2001, where he is currently the Gordon M Binder/Amgen Professor Of Electrical Engineering. From 2008-2015 he was Executive Officer of Electrical Engineering, as well as Associate Director of Information Science and Technology. From October 1996 to October 1998 he was a research associate at the Information Systems Laboratory, Stanford University, and from November 1998 to December 2000 he was a Member of the Technical Staff in the Mathematical Sciences Research Center at Bell Laboratories, Murray Hill, NJ. He has also held short-term appointments at Ricoh California Research Center, the Indian Institute of Science, and Linkoping University, Sweden. His research interests include wireless communications and networks, robust estimation and control, adaptive signal processing and linear algebra. He is the coauthor of the books (both with A.H. Sayed and T. Kailath) *Indefinite Quadratic Estimation and Control: A Unified Approach to H^2 and H^∞ Theories* (New York: SIAM, 1999) and *Linear Estimation* (Englewood Cliffs, NJ: Prentice Hall, 2000). He is a recipient of an Alborz Foundation Fellowship, the 1999 O. Hugo Schuck best paper award of the American Automatic Control Council (with H. Hindi and S.P. Boyd), the 2002 National Science Foundation Career Award, the 2002 Okawa Foundation Research Grant for Information and Telecommunications, the 2003 David and Lucille Packard Fellowship for Science and Engineering, the 2003 Presidential Early Career Award for Scientists and Engineers (PECASE), and the 2009 Al-Marai Award for Innovative Research in Communications, and was a participant in the 2004 National Academy of Engineering “Frontiers in Engineering” program.

He has been a Guest Editor for the IEEE Transactions on Information Theory special issue on “space-time transmission, reception, coding and signal processing” was an Associate Editor for Communications of the IEEE Transactions on Information Theory during 2004-2006, and is currently an Editor for the Journal “Foundations and Trends in Information and Communication” and for the IEEE Transactions on Network Science and Engineering. He is an IEEE Information Theory Society Distinguished Lecturer for 2016-2017.