On the Error Region for Channel Estimation Based Physical Layer Authentication over Rayleigh Fading

Augusto Ferrante, Nicola Laurenti, Chiara Masiero, Michele Pavon, and Stefano Tomasin, *Senior Member, IEEE.*

***Abstract*—For a physical layer message authentication procedure based on the comparison of channel estimates obtained from the received messages, we focus on an outer bound on the type I/II error probability region. Channel estimates are modelled as multivariate Gaussian vectors, and we assume that the attacker has only some side information on the channel estimate, which he does not know directly. We derive the attacking strategy that provides the tightest bound on the error region, given the statistics of the side information. This turns out to be a zero mean, circularly symmetric Gaussian density whose covariance matrices can be obtained by solving a constrained optimization problem. We propose an iterative algorithm for its solution: starting from the closed form solution of a relaxed problem, we obtain, by projection, an initial feasible solution; then, by an iterative procedure, we look for the fixed point solution of the problem. Numerical results show that for cases of interest the iterative approach converges, and perturbation analysis shows that the found solution is a local minimum.**

***Index Terms*—Authentication, Physical layer security, Rayleigh fading channels, Hypothesis testing**

I. Introduction

Physical layer security provides an effective defense mechanism which is complementary to higher layer security techniques. Indeed, it has the potential of resisting the attacks based on computational capabilities that may be feasible in the near future, e.g., by quantum computing. Moreover, security implemented at the physical layer is usually based on information theoretic arguments [1]. It therefore entails analytically predictable performance irrespective of the attacker capabilities and has recently been applied to widely used communication systems [2], [3]. One of the most desirable mechanisms of physical layer security is the authentication of the message source. This key task can be conveniently recast into a hypothesis testing problem [4], [5], namely to decide between hypothesis $\mathcal{H}_0$ that the message was effectively sent by the legitimate source, and hypothesis $\mathcal{H}_1$ that it was forged by an attacker.

Physical layer authentication has been addressed by considering either device-specific non-ideal transmission parameters extracted from the received signal [6], or channel characteristics to identify the link between a specific source and the receiver [7]–[9] (see the introduction of [9] for a survey on this topic). In this paper we focus on the latter case, which finds application in many wideband wireless systems, where even small changes in the position of the transmitter have a significant impact on the channel. In particular, we consider a scenario in which the test is performed in two phases. In the first phase, the receiver gets an authenticated noisy estimate x of the channel with respect to the legitimate transmitter. In the second phase, upon reception of a message, the receiver gets a new estimate u of the channel and compares it with x. Then, he must decide whether u is an estimate of the legitimate channel or the channel forged by an eavesdropper. This approach is used e.g., in [8], [9] and is quite general for channel-based authentication.

The performance of a binary hypothesis testing scheme is measured by the probability of type I (false alarm), and type II (missed detection) errors. Therefore, theoretical bounds on the achievable error probability region are of great importance to establish the effectiveness of practical schemes. For instance, [4] considered the traditional authentication scenario in which the legitimate parties can make use of a shared cryptographic key which is kept perfectly secret to the attackers. There, an outer bound on the achievable error region was derived, which holds irrespectively of the decision rule implemented by the receiver. Then, by fixing the false alarm probability, the outer bound is turned into a lower bound on the missed detection probability. An analogous approach was used in [11] and [12] within the different contexts of steganography and fingerprinting, respectively. Similarly, in [5], such lower bound is paired with an asymptotic upper bound, and both are derived also in the case that the legitimate parties share correlated sequences, instead of an identical key.

In the above cases, since the attacker has no information on the shared sequences, the optimal attacking strategy with respect to the outer bound is to present forged signals that, albeit independent of the legitimate shared key, are generated from the same marginal distribution as the legitimate signals. In our framework, on the contrary, the legitimate authentication signal is the actual realization of a fading wireless channel. Thus the attacker has some side information given by the channel estimates z he performs, which are in general correlated with the legitimate channel. We model channel estimates as correlated multivariate Gaussian vectors, which is a usual assumption in wireless transmissions, including those using orthogonal frequency division multiplexing (OFDM) or

A. Ferrante, N. Laurenti and S. Tomasin are with the Department of Information Engineering, University of Padua, Padua, Italy. Email: {first_name.last_name}@dei.unipd.it.

C. Masiero was with the Department of Information Engineering, University of Padua, Padua, Italy. Email: {first_name.last_name}@dei.unipd.it.

M. Pavon is with the Department of Mathematics, University of Padua, Italy. Email:pavon@math.unidp.it.

This work was supported in part by the Italian Ministry of Education and Research (MIUR) project ESCAPADE (Grant RBFR105NLC) under the "FIRB-Futuro in Ricerca 2010" funding program.

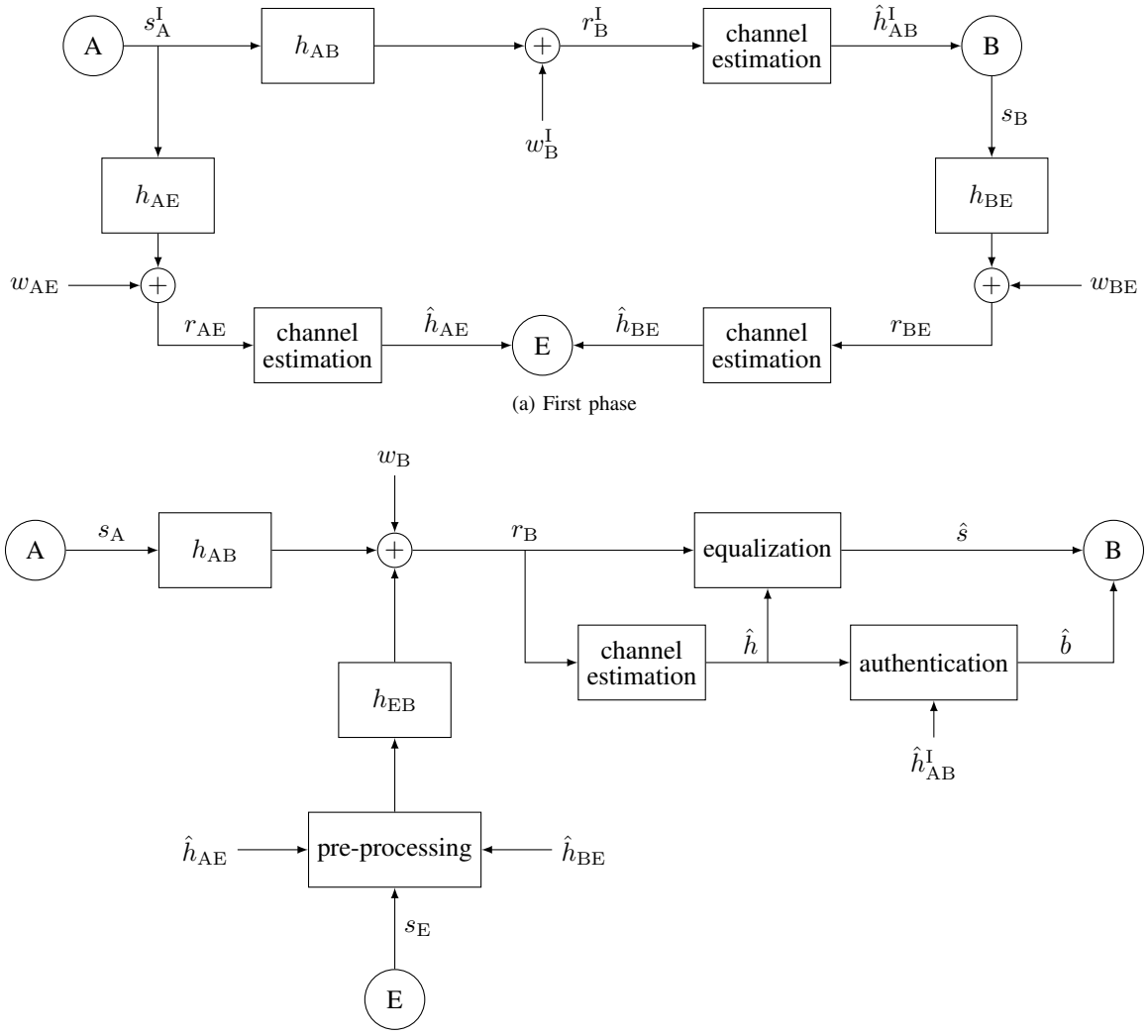

Fig. 1. Transmission channel scenario for the physical layer authentication problem.

multiple input multiple output (MIMO).

The contribution of our paper is thus threefold: 1) we derive an outer bound on the error probability region in terms of the attacker strategy; 2) we prove the existence of a strategy v, jointly Gaussian with z, that yields the tightest bound, and characterize its covariance through the solution of a system of two matrix equations; 3) we give an efficient technique for the numerical evaluation of the optimal attack strategy and the corresponding bound.

The paper is organized as follows. Section II introduces the problem formally, so that the theoretical results can be derived in Section III. Based on those results, in Section IV, we propose an efficient algorithm for the numerical evaluation of the optimal attack strategy. Then, in Section V we give examples of numerical results, and eventually we draw conclusions in Section VI.

In our notation, if $a \in \mathbb{C}^n$ and $b \in \mathbb{C}^m$ are random random vectors with random entries, K_{ab} denotes the $n \times m$ covariance matrix of vectors a and b, whereas $K_{\left[\begin{smallmatrix} a \\ b \end{smallmatrix}\right]}$ stands for the $(n+m)\times(n+m)$ covariance matrix of the vector $\left[\begin{smallmatrix} a \\ b \end{smallmatrix}\right]$. Symbol A^* denotes the complex conjugate transpose of matrix A, while $A^\dagger$ denotes the Moore-Penrose pseudo inverse of matrix A. Symbol I_n denotes the identity matrix of size $n \times n$. $A \otimes B$ denotes Kronecker product of matrices A and B.

II. Problem Statement

We consider the physical layer channel authentication scheme depicted in Fig. 1, where agents Alice (A) and Eve (E) transmit messages to Bob (B), and Bob aims at authenticating messages from Alice, i.e., reliably detecting whether she sent them or not. The authentication is performed via a two phase procedure [9]:

a) First Phase: In this phase, illustrated in Fig. 1.a, Alice (denoted by A in the figure) transmits one or more training messages, denoted by $s_{\mathrm{A}}^{\mathrm{I}}$, whose authenticity is guaranteed by higher layer techniques, to Bob (denoted by B), through the channel h_{AB}. From the received signal $r_{\mathrm{B}}^{\mathrm{I}}$, by conventional channel estimation techniques Bob gets a noisy estimate $\hat{h}_{\mathrm{AB}}^{\mathrm{I}}$ of the channel with respect to Alice and replies with an acknowledgement message s_{B}. At the same time Eve (denoted

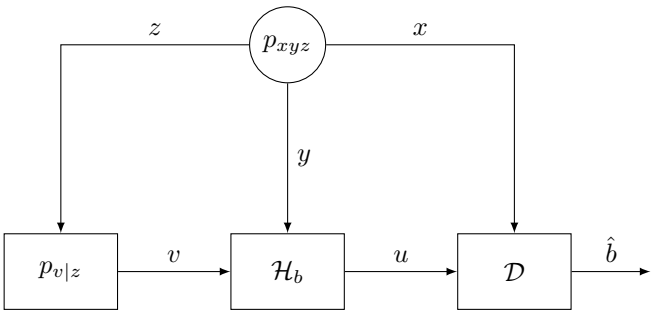

Fig. 2. Abstract model for physical layer authentication cast as an hypothesis testing problem with channel estimates as the available observations.

by E in the figure) overhears $s_{\rm A}^{\rm I}$ and $s_{\rm B}$ through her channels to Alice and Bob, respectively. Assuming that she also knows the content of these messages, by channel estimation from the received signals $r_{\rm AE}$ and $r_{\rm BE}$, Eve obtains (possibly noisy) estimates $\hat{h}_{\rm AE}, \hat{h}_{\rm BE}$ of the channels that link her to both agents.

b) Second Phase: Subsequently, as shown in Fig. 1.b, either Alice or Eve transmit message $s_{\rm A}$ or $s_{\rm E}$, respectively, which contains a training pattern. By this pattern Bob obtains a new noisy channel estimate $\hat{h}$ from $r_{\rm B}$. Authentication of the message is then performed by comparing estimate $\hat{h}$ with the template $\hat{h}_{\rm AB}^{\rm I}$. If the decision process $\mathcal{D}$ establishes that the two estimates are likely to be originated from the same channel realization, the message is deemed as coming from Alice and the binary flag $\hat{b}$ is set to zero, otherwise the message is deemed as forged and $\hat{b}$ is set to one. When Alice transmits in this phase, the new estimate $\hat{h}_{\rm AB}^{\rm II}$ of the Alice-Bob channel will not be identical to $\hat{h}_{\rm AB}^{\rm I}$, in general, due to the independent noises that affect both estimates. On the other hand, when Eve transmits in this phase, she can perform a pre-processing of her own messages in order to induce an equivalent channel estimate by Bob, who is as close as possible to $\hat{h}_{\rm AB}^{\rm I}$.

Note that we are now focusing on the authentication of a single message, but the authentication process can also be applied to multiple messages. Indeed, assuming that for each transmitted message a training sequence is also transmitted in order to allow channel estimation, a new authentication procedure can be issued for each message, and each procedure will be independent of the others. Thus, focusing on a single message authentication provides a full description of the system.

From now on, for the sake of a more compact notation, we let $x = \hat{h}_{\rm AB}^{\rm I}$, $y = \hat{h}_{\rm AB}^{\rm II}$, $z = (\hat{h}_{\rm AE}, \hat{h}_{\rm BE})$, $u = \hat{h}$ and we refer to the abstract representation of the authentication scenario given in Fig. 2. There, the joint probability density function (pdf) p_{xyz} of the channel estimates is determined by the fading environment and the estimation techniques adopted by the agents, which are assumed to be known to all of them. In order to consider a worst case scenario, we assume that Eve is able to forge any equivalent channel estimate v on Bob, neglecting power constraints and/or channel characteristics that, in practice, may prevent this and restrict the set of possible attacks. As a side effect, this assumptions also allows to simplify our analysis. The attacker's forging strategy can exploit of the information carried by her observations z, and for the sake of generality, we consider that she can make use of a probabilistic strategy, which is thus characterized by the conditional pdf $p_{v|z}$.

Note that, although our framework considers a single forging attempt, it can be extended to a sequence of attempts $\{v_i\}$, $i = 1, 2, \ldots$, where the attacker strategy is represented by the family of conditional pdfs $\left\{p_{v_i|z,v_1,\ldots,v_{i-1}}\right\}$.

Given the channel estimate u, Bob decides between the two hypotheses

$$\mathcal{H}_0 \;:\; u = y \qquad \text{message is from Alice,} \tag{1}$$

$$\mathcal{H}_1 \;:\; u = v \qquad \text{message was forged.} \tag{2}$$

In Fig. 2, being in hypothesis $\mathcal{H}_0$ or $\mathcal{H}_1$ is obtained by setting $b = 0$ or 1, respectively. Correct authentication is achieved when $\hat{b} = b$.

Recall that all channels are described by zero-mean circular symmetric complex Gaussian vectors with correlated entries, as a suitable model for many scenarios (including MIMO/OFDM). Moreover, we assume that all transmissions are corrupted by additive white Gaussian noise with zero mean. Similarly, we assume that also the channel estimates are zero-mean circular symmetric complex Gaussian vectors with correlated entries.[1] In particular, x, and y are n-dimensional, complex, circular symmetric Gaussian random vectors, z is an m-dimensional, complex, circular symmetric Gaussian random vector. On the other hand v is an n-dimensional, complex, random vector whose probability density is not specified as it will be chosen by the attacker in order to obtain better mimetic features. Note that Eve collects in z the estimates of channels to both Alice and Bob. For example, in an OFDM system n is the number of subcarriers, while $m = 2n$, since z collects two n-size channel estimates. For a MIMO system Alice and Bob may be equipped with a number of antennas different from those of Eve, and in this case again $n \neq m$.

We denote the set of all possible conditional distributions (forging strategies) $p_{v|z}(\cdot|\cdot)$ as

$$\mathcal{Q} = \left\{ q(\cdot|\cdot) \,:\, \mathbb{C}^n \times \mathbb{C}^m \to \mathbb{R}, q(b|c) \geq 0, \int q(b|c)\,\mathrm{d}b = 1 \right\}. \tag{3}$$

The performance of the authentication system is assessed by the type I error probability α, i.e., the probability that Bob discards a message as forged by Eve while it is coming from Alice

$$\alpha = \mathrm{P}[\hat{b} = 1|\mathcal{H}_0]\,, \tag{4}$$

and the type II error probability β, i.e., the probability that Bob accepts a message coming from Eve as legitimate

$$\beta = \mathrm{P}[\hat{b} = 0|\mathcal{H}_1]\,. \tag{5}$$

The aim of a clever design for the authentication scheme is to make both error probabilities α and β as small as possible.

[1]This is justified by the fact that, in order to be effective, estimates should have a distribution that is close to that of the target variable. Furthermore, under mild assumptions on the SNR and with a sufficient amount of data, estimates obtained e.g., with an maximum likelihood (ML) estimation, are asymptotically unbiased, efficient and Gaussian distributed themselves [13, §7.8].

Since it is trivial to achieve $\alpha+\beta=1$ with a random decision strategy that outputs $\hat{b}=1$ with probability α, independently of the observation u, we are only interested in values of α, β in the region

$$\mathcal{R}^0 = \{(\alpha,\beta) \,:\, \alpha \geq 0, \beta \geq 0, \alpha+\beta \leq 1\} \,. \tag{6}$$

A. Outer Bound on the Error Region for a Given Attacking Strategy

A first bound on the error region for a given attacking strategy can be obtained by applying the fundamental data processing inequality for the Kullback-Leibler (KL) divergence $\mathbb{D}\left(\cdot \,||\, \cdot\right)$ [14] to our binary hypothesis decision scheme $\mathcal{D}$. In fact, from [4], [11] we have[2]

$$\mathbb{D}\left(p_{\hat{b}|\mathcal{H}_1} \,||\, p_{\hat{b}|\mathcal{H}_0}\right) \leq \mathbb{D}\left(p_{xu|\mathcal{H}_1} \,||\, p_{xu|\mathcal{H}_0}\right) \,. \tag{7}$$

First we observe that $p_{\hat{b}|\mathcal{H}_0}(1) = \alpha$, $p_{\hat{b}|\mathcal{H}_0}(0) = 1-\alpha$, and similarly $p_{\hat{b}|\mathcal{H}_1}(0) = \beta$, $p_{\hat{b}|\mathcal{H}_1}(1) = 1-\beta$. Therefore, introducing the function[3]

$$f\left(\varphi,\psi\right) = \varphi \log \frac{\varphi}{1-\psi} + (1-\varphi)\log\frac{1-\varphi}{\psi} \,, \quad \varphi,\psi \in [0,1] \tag{8}$$

we can rewrite (7) as

$$f(\beta,\alpha) \leq \mathbb{D}\left(p_{xu|\mathcal{H}_1} \,||\, p_{xu|\mathcal{H}_0}\right) \,. \tag{9}$$

Since the observation z encloses all the information the attacker can exploit in order to deceive the receiver, we can assume that the forging strategy v is *conditionally independent* of the secure template x, given z. Then the divergence on the right side of (9) can be explicitly written for a given attacking strategy $p_{v|z} \in \mathcal{Q}$ as

$$\mathbb{D}\left(p_{xu|\mathcal{H}_1} \,||\, p_{xu|\mathcal{H}_0}\right) = \mathbb{D}\left(p_{xv} \,||\, p_{xy}\right) = D(p_{v|z}) \tag{10}$$

where

$$D(q) := \int\!\!\int \left[\int p_{xz}(a,c) q(b|c)\,\mathrm{d}c\right] \times \left[\log\left(\int p_{xz}(a,c)q(b|c)\,\mathrm{d}c\right) - \log p_{xy}(a,b)\right] \mathrm{d}a\,\mathrm{d}b \,. \tag{11}$$

For a given $f_0 \geq 0$, set

$$\mathcal{R}\left(f_0\right) := \left\{(\alpha,\beta) \in \mathcal{R}^0 \,:\, f(\beta,\alpha) \leq f_0\right\} . \tag{12}$$

Then (9) can be rewritten as an outer bound on the achievable (α,β) pairs, that is

$$(\alpha,\beta) \in \mathcal{R}\left(D(p_{v|z})\right) \,. \tag{13}$$

[2]Note that the symmetric bound $\mathbb{D}\left(p_{\hat{b}|\mathcal{H}_0} \,||\, p_{\hat{b}|\mathcal{H}_1}\right) \leq \mathbb{D}\left(p_{xu|\mathcal{H}_0} \,||\, p_{xu|\mathcal{H}_1}\right)$ holds as well (see also [9]).

[3]Notice that $f(\varphi,\psi)$ is the KL divergence between two Bernoulli probability distributions of parameters φ and $1-\psi$, respectively.

B. Tightening the Outer Bound over All Attacking Strategies

Each outer bound in (13) is clearly looser[4] than

$$\mathcal{R}_\cap := \bigcap_{q\in\mathcal{Q}} \mathcal{R}\left(D(q)\right) = \mathcal{R}\left(D^\star\right) \tag{14}$$

where

$$D^\star = \inf_{q\in Q} D(q) \,. \tag{15}$$

Note that region (14) is not strictly speaking an outer bound of the type (13), since the infimum (15) may not be achievable, in general. In the case that it is achievable, (14) represents a worst case performance for the authentication system, over all possible attacking strategies. On the other hand, for the attacker, approaching (15) represents the possibility to effectively carry out an impersonation attack.

The main goal of this paper is to evaluate the tightest bound (14). We provide an attacking strategy that achieves (15), under the assumption that the observation z encodes all the information about the secure template x, which is available to the attacker. From what we have shown, this is equivalent to the following constrained optimization problem:

Problem 1: Given the zero-mean, circular symmetric, jointly Gaussian random vectors x, y, z with joint covariance matrix

$$K_{\left[\begin{smallmatrix}x\\y\\z\end{smallmatrix}\right]} := \begin{bmatrix} K_{xx} & K_{xy} & K_{xz} \\ K_{xy}^* & K_{yy} & K_{yz} \\ K_{xz}^* & K_{yz}^* & K_{zz} \end{bmatrix} , \tag{16}$$

find a joint probability distribution $p_{xvz} \in L^1(\mathbb{C}^{2n+m})$ such that its marginal p_{xv} minimizes $\mathbb{D}\left(p_{xv} \,||\, p_{xy}\right)$ under the constraints:

1. The marginal distribution of x, z corresponding to p_{xvz} is equal to the given distribution p_{xz}.
2. The random vectors v and x are conditionally independent given z.

III. Main Results

In this section, we address Problem 1. In particular, we show that the problem is feasible, that it admits an optimal solution and that this solution is circularly symmetric complex Gaussian. Finally, we show how to reformulate this problem in terms of the solutions of two coupled matrix equations. The first issue to be considered is the *feasibility* of Problem 1, namely the existence of a distribution p_{xvz} satisfying the constraints and such that $\mathbb{D}\left(p_{xv} \,||\, p_{xy}\right)$ is finite.

Lemma 1: Problem 1 is feasible.

Proof: Let v be an n-dimensional, complex, zero-mean, circular symmetric Gaussian random vector (with arbitrary covariance) independent of x and of z. It is immediate to check that the corresponding p_{xvz} satisfies the constraints and is such that $\mathbb{D}\left(p_{xv} \,||\, p_{xy}\right)$ is finite. ■

[4]The second equality in (14) is straightforward. In fact, observe that $\mathcal{R}\left(f_0\right) \subset \mathcal{R}\left(f_1\right)$ if and only if $f_0 \leq f_1$. Then, for all $q \in \mathcal{Q}$, since $D^\star \leq D(q)$, we have $\mathcal{R}\left(D^\star\right) \subset \mathcal{R}\left(D(q)\right)$, and hence $\mathcal{R}\left(D^\star\right) \subset \mathcal{R}_\cap$.

Conversely, in order to prove that $\mathcal{R}_\cap \subset \mathcal{R}\left(D^\star\right)$ we show that $(\alpha,\beta) \notin \mathcal{R}\left(D^\star\right) \Rightarrow (\alpha,\beta) \notin \mathcal{R}_\cap$. In fact let $(\alpha,\beta) \notin \mathcal{R}\left(D^\star\right)$, that is $f(\alpha,\beta) > D^\star$. Then, by (15) there exist some $q_0 \in \mathcal{Q}$ such that $f(\alpha,\beta) > D(q_0) > D^\star$, and hence $(\alpha,\beta) \notin \mathcal{R}\left(D(q_0)\right)$ and $(\alpha,\beta) \notin \mathcal{R}_\cap$. Thus, $\mathcal{R}_\cap \subset \mathcal{R}\left(D^\star\right)$ is proved.

Lemma 2: Let x and z be jointly Gaussian. For any attacking strategy p_{xv} having finite second moment and in which v and x are conditionally independent given z, v and x are also conditionally orthogonal given z, that is

$$\mathbb{E}\left[(x-\bar{\mathbb{E}}[x|z])(v-\bar{\mathbb{E}}[v|z])\right]=0, \tag{17}$$

where $\bar{\mathbb{E}}[\cdot|z]$ stands for the linear minimum mean square error estimator [13, §15.8] of "$\cdot$" given z.

Proof: We have

$$\begin{aligned}
&\mathbb{E}\left[(x-\bar{\mathbb{E}}[x|z])(v-\bar{\mathbb{E}}[v|z])\right] && \text{(18a)}\\
&\quad=\mathbb{E}_z\left[\mathbb{E}_{xv|z}\left[(x-\bar{\mathbb{E}}[x|z])(v-\bar{\mathbb{E}}[v|z])|z\right]\right] && \text{(18b)}\\
&\quad=\mathbb{E}_z\left[\mathbb{E}_{x|z}\left[(x-\bar{\mathbb{E}}[x|z])|z\right]\mathbb{E}_{v|z}\left[(v-\bar{\mathbb{E}}[v|z])|z\right]\right] && \text{(18c)}\\
&\quad=\mathbb{E}_z\left[\left(\mathbb{E}_{x|z}[x|z]-\bar{\mathbb{E}}[x|z]\right)\left(\mathbb{E}_{v|z}[v|z]-\bar{\mathbb{E}}[v|z]\right)\right], && \text{(18d)}
\end{aligned}$$

where $\mathbb{E}_x[\cdot]$ denotes the expectation taken with respect to vector x, and (18b) and (18c) follow from the Total Expectation Theorem and the definition of conditional independence, respectively. Since x and z are jointly Gaussian, we have that $\mathbb{E}[x|z]=\bar{\mathbb{E}}[x|z]$. Thus, we can conclude that the right-hand side of (18d) is equal to zero. ∎

In general, conditional independence does not imply conditional orthogonality, although for jointly Gaussian variables they are equivalent. However, we have proved that conditional independence of x and v given z implies that x and v are conditionally orthogonal given z, thanks to x and z being jointly Gaussian.

Notice that, since the attacker knows the joint probability density p_{xyz}, the corner blocks of $K^{-1}_{\left[\begin{smallmatrix}x\\v\\z\end{smallmatrix}\right]}$ are known [see also (16)]. For the sake of simplicity, we introduce the following symbols

$$X:=K_{vv},\quad Y:=K_{xv},\quad Z:=K_{vz}. \tag{19}$$

Hence, we can write

$$K^{-1}_{\left[\begin{smallmatrix}x\\v\\z\end{smallmatrix}\right]}=\begin{bmatrix}K_{xx} & Y & K_{xz}\\ Y^* & X & Z\\ K_{xz}^* & Z^* & K_{zz}\end{bmatrix}. \tag{20}$$

Recall that the conditional orthogonality of x and v given z is equivalent to the following zero-block pattern in its inverse[5]

$$K^{-1}_{\left[\begin{smallmatrix}x\\v\\z\end{smallmatrix}\right]}=\begin{bmatrix}* & 0 & *\\ 0 & * & *\\ * & * & *\end{bmatrix}. \tag{21}$$

In this way we have expressed the second constraint of Problem 1 in terms of the structure of the inverse of the covariance matrix. We can therefore enforce this constraint by resorting to a "maximum entropy" completion as described in [16] – see also [17] for a more general result and [18] for an application of this technique.

Lemma 3: If q_{G} is a circular symmetric Gaussian distribution, then, among all distributions p that share the same mean vector μ and covariance matrix K, the one that minimizes $\mathbb{D}(p\,||\,q_{\mathrm{G}})$ is circular symmetric and Gaussian.

Proof: Let p_G be a *circular symmetric Gaussian* probability density on $\mathbb{C}^n$ and let $p\neq p_G$ be any density having the same first and second moment as p_G. We denote by $H(p)$ the differential entropy of the density p, i.e., $H(p):=-\int p(a)\log p(a)\,\mathrm{d}a$. Then (see [19, Theorem 2]), we have the inequality

$$H(p)<H(p_G). \tag{22}$$

Now let q_G be any proper Gaussian density on $\mathbb{C}^n$. Under the same hypotheses, we have

$$\int \log q_G(x)p(x)\mathrm{d}x=\int \log q_G(x)p_G(x)\mathrm{d}x, \tag{23}$$

because $\log q_G(x)$ is a quadratic function of x. In view of (22) and (23), we now have

$$\begin{aligned}
\mathbb{D}(p\|q_G)&=\int\log\frac{p(x)}{q_G(x)}p(x)\mathrm{d}x\\
&=-H(p)-\int\log q_G(x)p(x)\mathrm{d}x\\
&=-H(p)-\int\log q_G(x)p_G(x)\mathrm{d}x\\
&>-H(p_G)-\int\log q_G(x)p_G(x)\mathrm{d}x=\mathbb{D}(p_G\|q_G),
\end{aligned} \tag{24}$$

Thus, the solution of any minimum entropy problem with circular symmetric Gaussian prior has to be circular symmetric and Gaussian. ∎

Lemma 4: If the second moment of p_{xv} is not finite, then $\mathbb{D}(p_{xv}\,||\,p_{xy})=\infty$.

Proof: We assume that $\mathbb{D}(p_{xv}\,||\,p_{xy})$ is finite and show that the second moment of p_{xv} is finite. Let us first recall the variational formula for the relative entropy [20, page 68]:

$$\begin{aligned}
\mathbb{D}(p_{xv}\,||\,p_{xy})=\sup_{\varphi\in\Phi}\Bigg\{&\int_{\mathbb{C}^{2n}}\varphi(a)p_{xv}(a)\mathrm{d}a\\
&-\log\left[\int_{\mathbb{C}^{2n}}\exp[\varphi(a)]p_{xy}(a)\mathrm{d}a\right]\Bigg\}
\end{aligned} \tag{25}$$

where Φ is the set of bounded functions. Observe now that, since p_{xy} is a Gaussian probability density, there exists $\varepsilon>0$ such that

$$L:=\mathbb{E}[\exp(\varepsilon\|[x,y]\|^2)]=\int_{\mathbb{C}^{2n}}\exp(\varepsilon\|a\|^2)p_{xy}(a)\mathrm{d}a \tag{26}$$

is finite. Let us now consider the following sequence of bounded functions:

$$\varphi_\ell(a):=\begin{cases}\varepsilon\|a\|^2, & \text{if } \|a\|^2\le\ell,\\ 0, & \text{if } \|a\|^2>\ell.\end{cases} \tag{27}$$

From (25) we get that for all $\ell=1,2,\ldots,$

$$\begin{aligned}
\mathbb{D}(p_{xv}\,||\,p_{xy})+\log\left[\int_{\mathbb{C}^{2n}}\exp[\varphi_\ell(a)]p_{xy}(a)\mathrm{d}a\right]\\
\ge\int_{\mathbb{C}^{2n}}\varphi_\ell(a)p_{xv}(a)\mathrm{d}a,
\end{aligned} \tag{28}$$

or, equivalently, from (27),

$$\begin{aligned}
\frac{1}{\varepsilon}\left\{\mathbb{D}(p_{xv}\,||\,p_{xy})+\log\left[\int_{\mathbb{C}^{2n}}\exp[\varphi_\ell(a)]p_{xy}(a)\mathrm{d}a\right]\right\}\\
\ge\int_{\Omega_\ell}\|a\|^2p_{xv}(a)\mathrm{d}a,
\end{aligned} \tag{29}$$

[5]A proof can be worked out in the same vein of [15, Section 2].

where $\Omega_\ell := \{a \in \mathbb{C}^{2n} : \ \|a\|^2 \leq \ell\}$. As $\ell \to \infty$, the left-hand side of (29) converges to $\frac{1}{\varepsilon}[\mathbb{D}\left(p_{xv} \,||\, p_{xy}\right) + L]$ while the right hand side converges to the trace of the second moment of p_{xv}. Such a trace is therefore finite and thus also the second moment of p_{xv} is finite. ∎

We are now ready to consider the *existence* problem. As in many optimization problems this is one of the most delicate issue.

Theorem 1: There exists an optimal solution $p^\star_{xv}$ of Problem 1.

Proof: Let $p^{(j)}_{xvz}$, $j = 1, 2, \ldots$, be a sequence of probability densities satisfying the constraints of Problem 1 and such that the corresponding marginals $p^{(j)}_{xv}$ satisfy

$$\lim_{j\to\infty} \mathbb{D}\left(p^{(j)}_{xv} \,||\, p_{xy}\right) = D^\star,$$

with $D^\star$ given by (15). In view of Lemma 4, we can assume that all $p^{(j)}_{xvz}$ have finite mean vector μ_j and covariance matrix $\bar{K}_j$. Let m_j and K_j be the mean and covariance of $p^{(j)}_{xv}$, i.e., m_j are the first $2n$ components of μ_j and K_j is the $2n \times 2n$ upper-left block of $\bar{K}_j$. Now, in view of Lemma 3,

$$\begin{aligned}\mathbb{D}\left(p^{(j)}_{xv} \,||\, p_{xy}\right) &\geq \mathbb{D}\left(p^{(j)}_{xv,G} \,||\, p_{xy}\right)\\ &= \mathrm{tr}[K^{-1}_{\left[\begin{smallmatrix}x\\y\end{smallmatrix}\right]} K_j] + m_j^* K^{-1}_{\left[\begin{smallmatrix}x\\y\end{smallmatrix}\right]} m_j\\ &\quad - \log\left[\frac{\det[K_j]}{\det[K_{\left[\begin{smallmatrix}x\\y\end{smallmatrix}\right]}]}\right] - 2n, \end{aligned} \tag{30}$$

where $p^{(j)}_{xv,G}$ is the Gaussian distribution having mean vector m_j and covariance matrix K_j. It is easy to check that the right-hand side of (30) diverges if at least one of $\|K_j\|$ and $\|m_j\|$ does. Hence, both $\|K_j\|$ and $\|m_j\|$ remain bounded, and also μ_j and $\bar{K}_j$ remain bounded. Therefore, there exists a subsequence $p^{(j_i)}_{xvz}$, $i = 1, 2, \ldots$, such that $\bar{K}_{j_i}$ and μ_{j_i} converge as $i \to \infty$. Let $\bar{K}^\star$ and $\mu^\star$ be their limits and let $K^\star$ and $m^\star$ be the corresponding limits of K_{j_i} and m_{j_i}. Notice now that each density of the corresponding sequence of Gaussian distributions with the same mean and variance $p^{(j_i)}_{xvz,G}$ satisfies the constraints of Problem 1. In fact, the marginal p_{xz} does not change and, in view of (21), the second constraint only depends on the variance matrix. Let $p^\star_{xvz,G}$ be the Gaussian distribution whose mean and variance are $\mu^\star$ and $\bar{K}^\star$, respectively. In view of the previous arguments, we can conclude that $p^{(j_i)}_{xvz,G}$ satisfies the constraints of Problem 1. Let $p^\star_{xv,G}$ be the corresponding marginal. We have

$$\begin{aligned} D^\star &= \lim_{i\to\infty} \mathbb{D}\left(p^{(j_i)}_{xv,G} \,||\, p_{xy}\right)\\ &\geq \lim_{i\to\infty} \mathbb{D}\left(p^{(j_i)}_{xv,G} \,||\, p_{xy}\right)\\ &= \lim_{i\to\infty} \mathrm{tr}[K^{-1}_{\left[\begin{smallmatrix}x\\y\end{smallmatrix}\right]} K_{j_i}] + m^*_{j_i} K^{-1}_{\left[\begin{smallmatrix}x\\y\end{smallmatrix}\right]} m_{j_i} - \log\frac{\det[K_{j_i}]}{\det[K_{\left[\begin{smallmatrix}x\\y\end{smallmatrix}\right]}]} - 2n\\ &= \mathrm{tr}[K^{-1}_{\left[\begin{smallmatrix}x\\y\end{smallmatrix}\right]} K^\star] + (m^\star)^* K^{-1}_{\left[\begin{smallmatrix}x\\y\end{smallmatrix}\right]} m^\star - \log\frac{\det[K^\star]}{\det[K_{\left[\begin{smallmatrix}x\\y\end{smallmatrix}\right]}]} - 2n\\ &= \mathbb{D}\left(p^\star_{xv,G} \,||\, p_{xy}\right). \end{aligned} \tag{31}$$

Thus $p^\star_{xvz,G}$ solves Problem 1. ∎

Notice that from (31) it is immediate to see that the optimal solution not only exists but is Gaussian distributed with zero mean.

Corollary 1: Let x and y be jointly Gaussian. Then the solution of Problem 1 is zero mean and Gaussian.

We are now ready to find the solution of our problem.

Theorem 2: The solution of Problem 1 is the zero mean circularly symmetric Gaussian density $p^\star_{xvz}$ whose covariance matrix is

$$K_{\left[\begin{smallmatrix}x\\v\\z\end{smallmatrix}\right]} = \begin{bmatrix} K_{xx} & K_{xz}K^{-1}_{zz}Z^* & K_{xz}\\ ZK^{-1}_{zz}K^*_{xz} & ZK^{-1}_{zz}Z^* + CC^* & Z\\ K^*_{xz} & Z^* & K_{zz}\end{bmatrix}, \tag{32}$$

where Z and C solve

$$\begin{cases} C^* = C^*(ZK^{-1}_{zz}BK^{-1}_{zz}Z^* + CC^*)^{-1}A\\ Z^* = K^*_{xz}K^{-1}_{xx}K_{xy}\\ \qquad + BK^{-1}_{zz}Z^*(ZK^{-1}_{zz}BK^{-1}_{zz}Z^* + CC^*)^{-1}A \end{cases} \tag{33}$$

with

$$A := K_{yy} - K^*_{xy}K^{-1}_{xx}K_{xy}\ , \tag{34a}$$

$$B := K_{zz} - K^*_{xz}K^{-1}_{xx}K_{xz}\ . \tag{34b}$$

Proof: See the Appendix. ∎

In view of (11) and (14), Theorem 2 provides the tightest bound to the error region (14). Indeed, let $K_{\left[\begin{smallmatrix}x\\v\end{smallmatrix}\right]}$ be a shorthand notation for the $2n \times 2n$ upper-left corner of (32). Then, $D^\star$ is given by

$$\begin{aligned} D^\star &= \mathbb{D}(p^\star_{xv} \| p_{xy})\\ &= -\log\det(K_{\left[\begin{smallmatrix}x\\v\end{smallmatrix}\right]}K^{-1}_{\left[\begin{smallmatrix}x\\y\end{smallmatrix}\right]})\\ &\quad + \mathrm{tr}\, K^{-1}_{\left[\begin{smallmatrix}x\\y\end{smallmatrix}\right]}\left(K_{\left[\begin{smallmatrix}x\\v\end{smallmatrix}\right]} - K_{\left[\begin{smallmatrix}x\\y\end{smallmatrix}\right]}\right). \end{aligned} \tag{35}$$

Consider the circular symmetric Gaussian density $p^\star_{xvz}$, with zero mean and covariance (32). Note that it is such that x and v are conditionally independent given z. Then, by marginalizing and conditioning, we can obtain an optimum attacking strategy $p^\star_{v|z}(\cdot|a)$ which achieves (14). It is given by the proper Gaussian density whose mean and variance are defined by

$$\mu_{v|z} := ZK^{-1}_{zz}z \tag{36}$$

$$K_{v|z} := K_{vv} - K_{vz}K^{-1}_{zz}K^*_{vz} = CC^*\,. \tag{37}$$

IV. Efficient Computation of the Tightest Bound

In view of Theorem 2, in order to provide the expression of the optimal solution $p^\star_{xvz}$, we have to compute matrices C and Z, which solve the system of nonlinear matrix equations (33). This appears however to be a highly non trivial task. Thus, we propose a two stage algorithm:

1) **Feasible (projected) Solution.** To begin with, we deal with an optimization problem which can be considered a relaxed version of Problem 1, since no positivity constraints on matrix $K_{\left[\begin{smallmatrix}x\\v\\z\end{smallmatrix}\right]}$ are imposed. This task turns out to be much simpler to achieve. Indeed, the solution can be computed in closed form. Then, we project the solution to the relaxed problem onto the feasible set,

i.e., the set of pairs (X, Z) which make $K_{\left[\begin{smallmatrix} x \\ v \\ z \end{smallmatrix}\right]}$ positive definite.

2) **Iterative Algorithm.** We use the projection as a starting point for an iterative update procedure whose fixed point satisfies (33).

Next we provide some details for each phase.

Feasible Solution. Minimizing (11) with no constraints on the positivity of $K_{\left[\begin{smallmatrix} x \\ v \\ z \end{smallmatrix}\right]}$ is equivalent to solving

Problem 2:

$$\begin{aligned} &\arg\min_{X,Z} J(K_{\left[\begin{smallmatrix} x \\ v \end{smallmatrix}\right]}(Z,X)) \\ &:= \Big\{ -\log\det(K_{\left[\begin{smallmatrix} x \\ v \end{smallmatrix}\right]}(Z,X)K^{-1}_{\left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right]}) \\ &\qquad + \mathrm{tr}(K_{\left[\begin{smallmatrix} x \\ v \end{smallmatrix}\right]}(Z,X)K^{-1}_{\left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right]}) \Big\} , \end{aligned} \tag{38}$$

where we have made the dependence of $K_{\left[\begin{smallmatrix} x \\ v \end{smallmatrix}\right]}$ on (Z, X) explicit. In the same vein of the proof of Theorem 2, we work out the optimality conditions that X and Z have to satisfy, based on the analysis of the first variation $D[J(K_{\left[\begin{smallmatrix} x \\ v \end{smallmatrix}\right]}); \delta K_{\left[\begin{smallmatrix} x \\ v \end{smallmatrix}\right]}]$. Some easy algebraic calculations lead us to the closed form of an optimal solution (Z, X):

$$\begin{cases} Z = K^*_{xy}K^{-1}_{xx}K_{xz}(K^*_{xz}K^{-1}_{xx}K_{xz})^\dagger K_{zz}, \\ X = K_{yy} - K^*_{xy}K_{xx}^{-\frac{1}{2}} \\ \qquad \left[I_n - K_{xx}^{-\frac{1}{2}}K_{xz}(K^*_{xz}K^{-1}_{xx}K_{xz})^\dagger K^*_{xz}K_{xx}^{-\frac{1}{2}}\right] K_{xx}^{-\frac{1}{2}}K_{xy} . \end{cases} \tag{39}$$

If the obtained X and Z are such that $X - ZK^{-1}_{zz}K^* \geq 0$, the algorithm terminates. Otherwise, a pair (C, Z) is obtained as follows. Let T be a unitary matrix such that $\Sigma_T := T^*(X - ZK^{-1}_{zz}K^*)T = \mathrm{diag}(d_1, d_2, \ldots, d_k, \delta_1, \delta_2, \ldots, \delta_h)$, where d_i are real positive and in decreasing order, and δ_i are negative or zero. Let $\Sigma'_T := \mathrm{diag}(d_1, d_2, \ldots, d_k, \varepsilon, \varepsilon, \ldots, \varepsilon)$, where ε is a "small" parameter, e.g., $\varepsilon := d_k/100$. Let $\Sigma' := T\Sigma'_T T^* > 0$ and C be such that $CC^* = \Sigma'$.

Iterative Algorithm. We use the pair (C, Z) as a starting point for the iterations

$$\begin{cases} C^*(k+1) = C^*(k)(Z(k)K^{-1}_{zz}BK^{-1}_{zz}Z^*(k) \\ \qquad\qquad + C(k)C^*(k))^{-1}A \\ Z^*(k+1) = K^*_{xz}K^{-1}_{xx}K_{xy}BK^{-1}_{zz}Z^*(k) \\ \qquad\qquad (Z(k)K^{-1}_{zz}BK^{-1}_{zz}Z^*(k) + C(k)C^*(z))^{-1}A \end{cases} \tag{40}$$

with A and B as defined in (34a)–(34b). By the iterative process we aim at finding a fixed point for (40), which provides the solution of Problem 1. The iterative process can be stopped either after a fixed number of iterations, or when the variation of the cost $D^\star(k)$ over one iteration is smaller than a given percentage.

V. Numerical Results

In this section we provide numerical evidence of the effectiveness of the proposed algorithm.

In order to assess the performance of the proposed algorithm for the computation of the tightest bound, we consider the case where $m = n$ and the covariance matrix is given by

$$K_{\left[\begin{smallmatrix} x \\ y \\ z \end{smallmatrix}\right]} = \begin{bmatrix} 1 & \sigma & \rho \\ \sigma^* & 1 & \tau \\ \rho^* & \tau^* & 1 \end{bmatrix} \otimes H_n , \tag{41}$$

where matrix H_n, represents the correlation between coefficients in the same channel estimate, while the parameter ρ dictates the correlation between channel estimates performed by Eve and the legitimate channel template, while σ represents the correlation between successive estimates of the same legitimate channel performed by Bob in the two phases. Note that this formulation simplifies the discussion of results as it allows to separate the correlation between terminals from that between channel coefficients.

For instance, by choosing $H_n = I_n$, this scenario corresponds to an OFDM transmission with uncorrelated channel frequency response. Besides being an asymptotic case widely considered in the literature, this is also a practical scenario, when channel estimation is performed on a subset of subcarriers with cardinality smaller than the number of channel taps, and the channel taps are independent Gaussian variables. As a possibly more relistic example, consider an impulse response with n independent channel taps having zero mean and exponentially decaying power delay profile with exponential parameter λ. In this case, the correlation matrix between channel coefficients in the frequency domain is given by the matrix H_n where the generic (p, q) entry is

$$\begin{aligned} [H_n]_{p,q} &= \sum_{\ell=0}^{n-1} e^{-2\pi\imath(p-q)\ell/n} e^{-\lambda\ell} \\ &= \frac{1 - e^{-\lambda n}}{1 - e^{-[\lambda + 2\pi\imath(p-q)/n]}} . \end{aligned} \tag{42}$$

Under the assumption (41), the submatrices in (16) are all multiples of H_n, and the closed form solution (39) simplifies to

$$\begin{cases} Z = \frac{\sigma^*}{\rho^*} H_n \\ X = H_n , \end{cases} \tag{43}$$

which is feasible if and only if $|\rho| \geq |\sigma|$, that is if the legitimate channel estimate in the first phase is more correlated with the attacker observations than with the legitimate channel estimate in the second phase, and in this case it provides $D^\star = 0$ and the optimal solution

$$\mu_{v|z} = \tfrac{\sigma^*}{\rho^*}\, z \tag{44}$$

$$K_{v|z} = \left(1 - \left|\tfrac{\sigma}{\rho}\right|^2\right) H_n . \tag{45}$$

This agrees with the intuition that if the attacker enjoys such advantage in the correlation coefficients he can forge a channel that is indistinguishable by the receiver from the legitimate one.

On the other hand, in the less pessimistic and supposedly more realistic condition that $|\rho| < |\sigma|$, the iterative algorithm to compute the optimal attack can be written as

$$C(k) = c(k)H_n \quad , \qquad Z(k) = \zeta(k)H_n$$

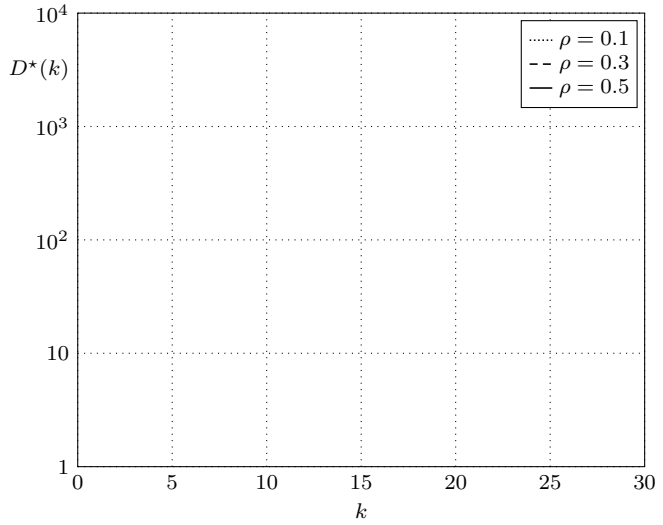

Fig. 3. Cost of the solution $D^\star(k)$ computed by the iterative algorithm as a function of the iteration number k, with $n = m = 64$, for $\sigma = 0.7$ and $\rho = 0.1, 0.3, 0.5$.

with the initial scalar values

$$c(0) = \epsilon \quad , \qquad \zeta(0) = \frac{\sigma^*}{\rho^*}$$

and the iteration equations

$$\begin{cases} c(k+1) = \dfrac{c(k)(1-|\rho|^2)}{|\zeta(k)|^2(1-|\sigma|^2)+|c(k)|^2} \\ \zeta(k+1) = \rho\sigma^* + \dfrac{\zeta(k)(1-|\sigma|^2)(1-|\rho|^2)}{|\zeta(k)|^2(1-|\sigma|^2)+|c(k)|^2} \end{cases} \tag{46}$$

First we assess the performance of the iterative algorithm. Fig. 3 shows the cost of the optimum solution $D^\star(k)$ as a function of the number of iterations for the iterative algorithm, with $n = m = 64$, and various values of ρ. We observe that the iterative algorithm always converges to a fixed point of (40) and that a solution with good accuracy is achieved in less than 100 iterations. Thus, in the following we consider this value for the maximum number of iterations.

Fig. 4 shows the outer bound on the type I/II error probability region (α, β) for various values of the correlation parameter ρ, and for $n = m = 64$, as obtained from the proposed iterative approach. As expected, we observe that for increasing values of ρ, the region of achievable values of α and β gets narrower. In particular, for the considered scenario ($n = 16$, $\sigma = 0.7$), we see that it is impossible to bring both type I and type II error probabilities below 10^{-1} already for $\rho = 0.5$.

In Fig. 5 we report the results obtained for both the initial feasible solution (projection of the solution of (39)) and final solution of the iterative algorithm, as a function of ρ, for two different values of σ.

We note that for low values of ρ the iterative algorithm remarkably lowers the value of the cost function from the initial feasible solution, thus motivating its use, although it comes at the cost of more computations. On the other hand, as ρ approaches σ the initial feasible solution gets closer to the unconstrained solution, and hence it is already close to the final optimum.

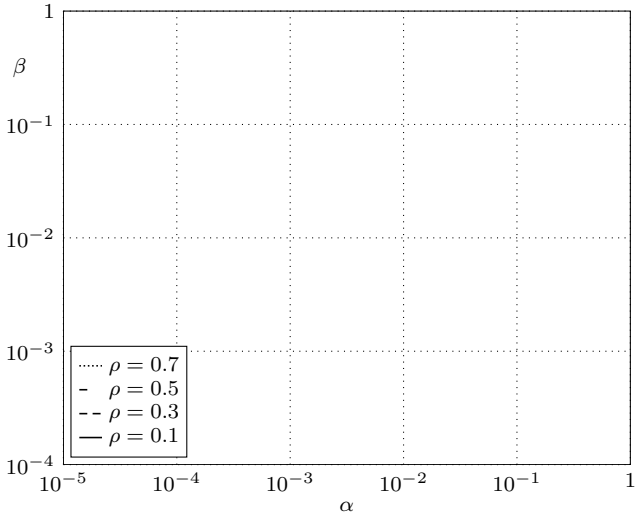

Fig. 4. Bound on the region of type II (β) vs. type I (α) error probability for various values of the correlation parameter ρ, with $n = 16$, $\sigma = 0.7$.

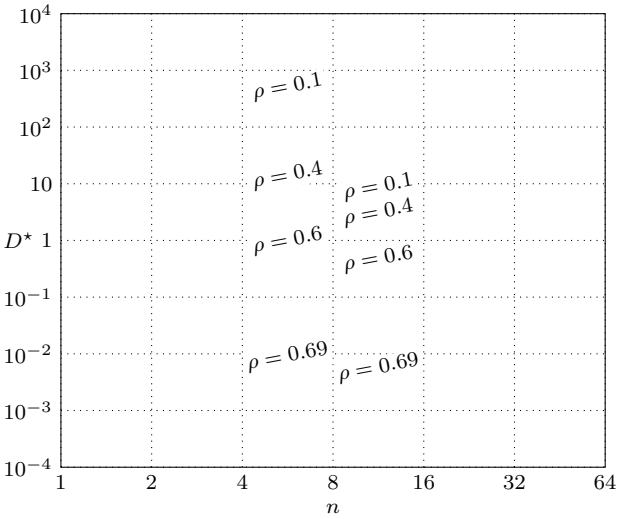

Fig. 5. Cost function $D^\star$ as a function of n, for several values of ρ and $\sigma = 0.7$. The cost $D^\star(0)$ of the initial projected solution is shown in dashed lines, while that of the final solution of the iterative algorithm is in solid lines.

Moreover, it is observed that the cost function $D^\star$ is independent of the particular choice of the matrix M_n, as long as it is positive definite, depending linearly only on its size n. For the considered case of OFDM transmission, this means that more dispersive channels having independent taps have the potential to provide a better authentication system. This phenomenon was already seen, e.g., in [9].

A. Randomly Correlated Channels

We now consider channels with random correlation.

In the model (41)–(42) we choose σ uniformly distributed in $[0.8, 0.9]$ and ρ uniformly distributed in $[0, 0.6]$.

Even in this case we have verified that setting the maximum number of iteration to 100 is enough for the convergence of the

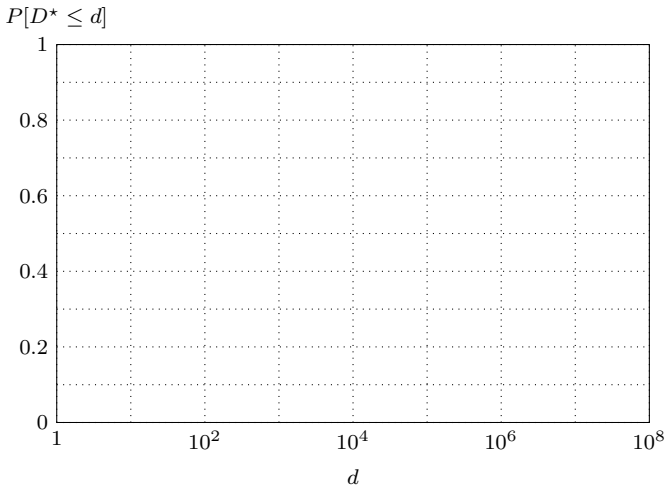

Fig. 6. CDF of the cost function $D^\star$ in the random correlation model with σ uniform in $[0.8, 0.9]$, ρ uniform in $[0, 0.6]$ and $n = 64$, for the initial projected solution (dashed lines) and the final solution of the iterative algorithm (solid lines).

iterative algorithm. Fig. 6 shows the cumulative distribution function (CDF) of $D^\star$ for $n = 64$, at the convergence of the iterative algorithm, as well as at the initial feasible solution obtained by projection.

We have also carried out a perturbation analysis. In particular, we evaluate the effects of small perturbations of Z and C generated as Gaussian random variables with norm $0.01\|Z\|$ and $0.01\|C\|$, respectively. Results not reported here show that it provides a negligible variation with respect to the solution of the iterative approach. This supports the conclusion that the iterative approach reaches a minimum for $\mathbb{J}(K_{\left[\begin{smallmatrix} x \\ v \end{smallmatrix}\right]})$. We also applied the iterative algorithm starting from the perturbed solutions which led to cost improvements. Again, this procedure achieves very small variation with a relative increase of the cost function by 10^{-4}.

VI. Conclusions

We have considered the problem of deriving a universal performance bound, for a message source authentication scheme based on channel estimates in a wireless fading scenario, where an attacker may have correlated observations available. We have formulated an outer bound to the region of achievable false alarm and missed detection probabilities, which is universal across all possible decision rules by the receiver.

Under the assumption that the channels are represented by multivariate complex Gaussian variables, we have proved that the tightest bound corresponds to a forging strategy that produces a zero mean signal which is jointly Gaussian with the attacker observations. Furthermore, we have derived a characterization of their joint covariance matrix through the solution of a system of two nonlinear matrix equations. Based upon this characterization, we have also devised an efficient iterative algorithm for its computation: the solution to the matrix system appears as a fixed point of the iteration.

From the numerical results, we conjecture that the proposed iterative approach for the best attacking strategy converges in general, although determining its convergence seems a highly difficult problem. Moreover, from the perturbation analysis, we deduce that the limit point is a local minimum. We have therefore provided an effective method for the attacking strategy that yields the tightest bound on the error region of any message authentication procedure.

References

[1] M. Bloch and J. Barros, *Physical-Layer Security. From Information Theory to Security Engineering*. Cambridge University Press, 2011.
[2] F. Renna, N. Laurenti, H. V. Poor, "Physical-layer secrecy for OFDM transmissions over fading channels," *IEEE Trans. Inf. Forens. and Security*, vol. 7, no. 4, pp. 1354–1367, Aug. 2012.
[3] S. Tomasin "Resource allocation for secret transmissions over MIMOME fading channels, in Proc. *IEEE Global Conf. on Commun. (GLOBECOM)*, Atlanta, Georgia, Dic. 2013.
[4] U.M. Maurer, "Authentication theory and hypothesis testing," *IEEE Trans. Inf. Theory*, vol. 46, Jul. 2000, pp. 1350–1356.
[5] L. Lai, H. El Gamal, and H. V. Poor "Authentication over noisy channels", *IEEE Trans. Inf. Theory*, vol. 55, no. 2, pp. 906–916, Feb. 2009.
[6] T. Daniels, M. Mina, and S.F. Russell, "A signal fingerprinting paradigm for general physical layer and sensor network security and assurance," *IEEE SECURECOMM*, pp. 1-3, Athens (Greece), Sep. 2005.
[7] D.B. Faria and D.R. Cheriton, "Detecting identity-based attacks in wireless networks using signalprints," *ACM WiSe*, pp. 43–52, Los Angeles (CA), Sep. 2006.
[8] L. Xiao, L.J. Greenstein, L. Fellow, N.B. Mandayam, and W. Trappe, "Channel-based spoofing detection in frequency-selective Rayleigh channels," *IEEE Trans. Wireless Commun.*, vol. 8, 2009, pp. 5948–5956.
[9] P. Baracca, N. Laurenti, and S. Tomasin, "Physical layer authentication over MIMO fading wiretap channels," *IEEE Trans. Wireless Commun.*, vol. 11, 2012, pp. 2564–2573
[10] 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. Surv. Tutorials*, vol. 16, no. 3, pp. 15501573, Jan. 2014.
[11] C. Cachin, "An information-theoretic model for steganography," in Proc. *Int. Workshop on Inf. Hiding (IH'98)*, Portland, OR, April 14–17, 1998, vol. LNCS-1525, pp. 306–318.
[12] M. Barni, and B. Tondi, "The source identification game: An information-theoretic perspective," *IEEE Trans. on Inform. Forens. Security*, vol. 8, no. 3, pp. 450–463, Mar. 2013.
[13] S. M. Kay, *Fundamentals of statistical signal processing. Estimation theory*, Prentice Hall, 1993.
[14] S. Kullback, *Information Theory and Statistics*, Dover Publications, NY, 1967.
[15] T. P. Speed, and H. T. Kiiveri, "Gaussian Markov distributions over finite graphs", *Annals of Statistics*, vol. 14, no. 1, pp.138–150, Mar. 1986.
[16] A. P. Dempster, "Covariance selection," *Biometrics*, vol. 28, 1972, pp. 157–175.
[17] A. Ferrante and M. Pavon, "Matrix completion *à la* Dempster by the principle of parsimony," *IEEE Trans. Inf. Theory*, vol. 57, 2011, pp. 3925–3931.
[18] F. Carli, A. Ferrante, M. Pavon, and G. Picci, "A maximum entropy solution of the covariance extension problem for reciprocal processes," *IEEE Trans. Automatic Control*, vol. 56, 2011, pp. 1999–2012.
[19] F. D. Neeser, and J. L. Massey, "Proper complex random processes with applications to information theory", *IEEE Trans. Inf. Theory*, vol. 39, no. 4, pp.1293–1302, Jul. 1993.
[20] J.-D. Deuschel, and D. W. Stroock, *Large deviations.* Academic Press, New York, 1989.

Appendix

In this Appendix we provide the proof of Theorem 2.

We have already shown that the optimal solution is a zero-mean Gaussian distribution having covariance matrix (20) where

$$K_{\left[\begin{smallmatrix} x \\ z \end{smallmatrix}\right]} := \begin{bmatrix} K_{xx} & K_{xz} \\ K_{xz}^* & K_{zz} \end{bmatrix} > 0$$

is given. Clearly in this way the first constraint of Problem 1 is automatically satisfied for any X, Y, Z. We now show that the second constraint is equivalent to impose

$$Y = K_{xz}K_{zz}^{-1}Z^*.$$

Indeed, in view of Lemma 2, x and v are conditionally orthogonal given z, so that the inverse of K_{xvz} must exhibit the zero-block pattern (21). Based on this information, we can compute Y as a function of Z and X by employing the block-matrix inversion formula

$$M_1 = \begin{bmatrix} A_1 & B_1 \\ C_1 & D_1 \end{bmatrix} \Rightarrow M_1^{-1} = \begin{bmatrix} (M_1^{-1})_{11} & (M_1^{-1})_{12} \\ (M_1^{-1})_{21} & (M_1^{-1})_{22} \end{bmatrix}. \tag{47}$$

where

$$(M_1^{-1})_{11} = (A_1 - B_1D_1^{-1}C_1)^{-1} \tag{48}$$
$$(M_1^{-1})_{12} = -A_1^{-1}B_1(D_1 - C_1A_1^{-1}B_1)^{-1} \tag{49}$$
$$(M_1^{-1})_{21} = -D_1^{-1}C_1(A_1 - B_1D_1^{-1}C_1)^{-1} \tag{50}$$
$$(M_1^{-1})_{22} = (D_1 - C_1A_1^{-1}B_1)^{-1} \tag{51}$$

We partition $K_{\left[\begin{smallmatrix} x \\ v \\ z \end{smallmatrix}\right]}$ as

$$K_{\left[\begin{smallmatrix} x \\ v \\ z \end{smallmatrix}\right]} = \begin{bmatrix} A_1 & B_1 \\ C_1 & D_1 \end{bmatrix}, \tag{52}$$

where

$$A_1 := K_{xx}, \quad B_1 := \begin{bmatrix} Y & K_{xz} \end{bmatrix},$$
$$C_1 := \begin{bmatrix} Y^* \\ K_{xz}^* \end{bmatrix}, \quad D_1 := \begin{bmatrix} X & Z \\ Z^* & K_{zz} \end{bmatrix}.$$

Therefore, the block in position $(1,2)$ of K_{xvz}^{-1} (with respect to the partition (52)) is given by

$$\begin{aligned} &-A_1^{-1}B_1(D_1 - C_1A_1^{-1}B_1)^{-1} = -K_{xx}^{-1}\begin{bmatrix} Y & K_{xz} \end{bmatrix} \times \\ &\left(\begin{bmatrix} X & Z \\ Z^* & K_{zz} \end{bmatrix} - \begin{bmatrix} Y^* \\ K_{xz}^* \end{bmatrix} K_{xx}^{-1} \begin{bmatrix} Y & K_{xz} \end{bmatrix}\right)^{-1} \\ &= -K_{xx}^{-1}\begin{bmatrix} Y & K_{xz} \end{bmatrix} \times \\ &\left(\underbrace{\begin{bmatrix} X - Y^*K_{xx}^{-1}Y & Z - Y^*K_{xx}^{-1}K_{xz} \\ Z^* - K_{xz}^*K_{xx}^{-1}Y & K_{zz} - K_{xz}^*K_{xx}^{-1}K_{xz} \end{bmatrix}}_{:=M_2}\right)^{-1}. \end{aligned}$$

In order to impose the zero-block pattern (21) to the inverse, we make the block in position $(1,1)$ in $-A_1^{-1}B_1(D_1 - C_1A_1^{-1}B_1)^{-1}$ vanish. Note that we need to explicitly compute only the elements in the first column block of ${M_2}^{-1}$. Let

$$\begin{aligned} \begin{bmatrix} A_2 & B_2 \\ C_2 & D_2 \end{bmatrix} &:= \begin{bmatrix} X - Y^*K_{xx}^{-1}Y & Z - Y^*K_{xx}^{-1}K_{xz} \\ Z^* - K_{xz}^*K_{xx}^{-1}Y & K_{zz} - K_{xz}^*K_{xx}^{-1}K_{xz} \end{bmatrix} \\ &= M_2 \end{aligned} \tag{53}$$

Thus, in view of the matrix inversion lemma, the first block column in ${M_2}^{-1}$ is given by

$$\begin{bmatrix} (A_2 - B_2D_2^{-1}C_2)^{-1} \\ -D_2^{-1}C_2(A_2 - B_2D_2^{-1}C_2)^{-1} \end{bmatrix}.$$

Therefore, orthogonality of x and v given z implies

$$\begin{aligned} 0 &= -K_{xx}^{-1}\begin{bmatrix} Y & K_{xz} \end{bmatrix} \times \\ &\quad \begin{bmatrix} (A_2 - B_2D_2^{-1}C_2)^{-1} & 0 \\ 0 & -D_2^{-1}C_2(A_2 - B_2D_2^{-1}C_2)^{-1} \end{bmatrix} \\ &= -K_{xx}^{-1}Y(A_2 - B_2D_2^{-1}C_2)^{-1} \\ &\quad + K_{xx}^{-1}K_{xz}D_2^{-1}C_2(A_2 - B_2D_2^{-1}C_2)^{-1} \\ &= Y - K_{xz}D_2^{-1}C_2 \ , \end{aligned} \tag{54}$$

so that

$$\begin{aligned} Y &= K_{xz}\left(K_{zz} - K_{xz}^*K_{xx}^{-1}K_{xz}\right)^{-1}\left(Z^* - K_{xz}^*K_{xx}^{-1}Y\right) \\ &= \Big[\Big(I + K_{xz}\left(K_{zz} - K_{xz}^*K_{xx}^{-1}K_{xz}\right)^{-1} \\ &\quad \times K_{xz}^*K_{xx}^{-1}\Big)\Big]^{-1} K_{xz}\left(K_{zz} - K_{xz}^*K_{xx}^{-1}K_{xz}\right)^{-1} Z^* \\ &= K_{xz}K_{zz}^{-1}Z^* \ . \end{aligned}$$

In this way, we have parametrized all matrices $K_{\left[\begin{smallmatrix} x \\ v \\ z \end{smallmatrix}\right]}$ whose inverse has the specified structure. At this point, we could minimize the divergence $\mathbb{D}(p_{xv}\|p_{xy})$ over Z and X. This turns out to be an easy problem that can be solved in closed form. This, however, is not the solution[6] to our original problem since there is yet another (hidden) constraint that we need to impose, namely that matrix

$$K_{\left[\begin{smallmatrix} x \\ v \\ z \end{smallmatrix}\right]} = \begin{bmatrix} K_{xx} & K_{xz}K_{zz}^{-1}Z^* & K_{xz} \\ (K_{xz}K_{zz}^{-1}Z^*)^* & X & Z \\ K_{xz}^* & Z^* & K_{zz} \end{bmatrix} \tag{55}$$

is a *bona fide* covariance matrix, i.e., it is positive semidefinite. Since $K_{\left[\begin{smallmatrix} x \\ z \end{smallmatrix}\right]}$ is positive definite, this constraint is equivalent to

$$X - \begin{bmatrix} (K_{xz}K_{zz}^{-1}Z^*)^* & Z \end{bmatrix} \begin{bmatrix} K_{xx} & K_{xz} \\ K_{xz}^* & K_{zz} \end{bmatrix}^{-1} \begin{bmatrix} K_{xz}K_{zz}^{-1}Z^* \\ Z^* \end{bmatrix} \geq 0$$

which, with simple algebraic manipulations, is seen to be equivalent to

$$X - ZK_{zz}^{-1}Z^* \geq 0. \tag{56}$$

The positivity constraint is then automatically satisfied if we re-parametrize the unknown matrix X in term of a new matrix C in the form

$$X = ZK_{zz}^{-1}Z^* + CC^*. \tag{57}$$

The optimal solution can now be easily obtained by solving the following *unconstrained* optimization problem

$$\arg\min_{C,Z} \mathbb{D}(p_{xv}\|p_{xy}). \tag{58}$$

Since

$$K_{\left[\begin{smallmatrix} x \\ v \end{smallmatrix}\right]} := \begin{bmatrix} K_{xx} & K_{xz}K_{zz}^{-1}Z^* \\ Z(K_{xz}K_{zz}^{-1})^* & ZK_{zz}^{-1}Z^* + CC^* \end{bmatrix}, \tag{59}$$

$$K_{\left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right]} := \begin{bmatrix} K_{xx} & K_{xy} \\ K_{xy}^* & K_{yy} \end{bmatrix}, \tag{60}$$

[6]Here we mention this simplified optimization problem because, as discussed later, it turns out to be very useful as the first step of an efficient numerical procedure that computes the solution of our original problem.

solving (58) is equivalent to compute

$$\arg\min_{Z,C}\left\{-\log\det(K_{\left[\begin{smallmatrix}x\\v\end{smallmatrix}\right]}K_{\left[\begin{smallmatrix}x\\y\end{smallmatrix}\right]}^{-1})+\operatorname{tr}K_{\left[\begin{smallmatrix}x\\y\end{smallmatrix}\right]}^{-1}K_{\left[\begin{smallmatrix}x\\v\end{smallmatrix}\right]}\right\}. \quad (61)$$

We are then led to the formulation of Problem 1. Let

$$J\left(K_{\left[\begin{smallmatrix}x\\v\end{smallmatrix}\right]}\right) := -\log\det(K_{\left[\begin{smallmatrix}x\\v\end{smallmatrix}\right]}K_{\left[\begin{smallmatrix}x\\y\end{smallmatrix}\right]}^{-1}) + \operatorname{tr}K_{\left[\begin{smallmatrix}x\\y\end{smallmatrix}\right]}^{-1}K_{\left[\begin{smallmatrix}x\\v\end{smallmatrix}\right]}. \quad (62)$$

Introducing the partitioning in square blocks

$$K_{\left[\begin{smallmatrix}x\\y\end{smallmatrix}\right]}^{-1} - K_{\left[\begin{smallmatrix}x\\v\end{smallmatrix}\right]}^{-1} = \begin{bmatrix}\Delta_{11} & \Delta_{12}\\ \Delta_{21} & \Delta_{22}\end{bmatrix} \quad (63)$$

and the notation

$$F := \delta Z K_{zz}^{-1} Z^* + Z K_{zz}^{-1}\delta Z^* + \delta C C^* + C\delta C^* \quad (64)$$

$$G := \Delta_{21}K_{xz}K_{zz}^{-1}\delta Z^* + \Delta_{22}\left[\delta Z K_{zz}^{-1} Z^* + Z K_{zz}^{-1}\delta Z^* + \delta C C^* + C\delta C^*\right]. \quad (65)$$

the first variation of $J(K_{\left[\begin{smallmatrix}x\\v\end{smallmatrix}\right]})$ is

$$\begin{aligned} &D[J(K_{\left[\begin{smallmatrix}x\\v\end{smallmatrix}\right]});\delta K_{\left[\begin{smallmatrix}x\\v\end{smallmatrix}\right]}]\\ &= \operatorname{tr}\left[(K_{\left[\begin{smallmatrix}x\\y\end{smallmatrix}\right]}^{-1} - K_{\left[\begin{smallmatrix}x\\v\end{smallmatrix}\right]}^{-1})\,\delta K_{\left[\begin{smallmatrix}x\\v\end{smallmatrix}\right]}\right]\\ &= \operatorname{tr}\left[\begin{bmatrix}\Delta_{11} & \Delta_{12}\\ \Delta_{21} & \Delta_{22}\end{bmatrix}\begin{bmatrix}0 & K_{xz}K_{zz}^{-1}\delta Z^*\\ \delta Z(K_{xz}K_{zz}^{-1})^* & F\end{bmatrix}\right]\\ &= \operatorname{tr}\begin{bmatrix}\Delta_{12}\delta Z(K_{xz}K_{zz}^{-1})^* & *\\ * & G\end{bmatrix}. \end{aligned} \quad (66)$$

By the properties of the trace and the Hermitian symmetry, we get that the first variation vanishes if and only if

$$\operatorname{tr}\left[\left((K_{xz}K_{zz}^{-1})^*\Delta_{12} + Z^*K_{zz}^{-1}\Delta_{22}\right)\delta Z + C^*\Delta_{22}\delta C\right] = 0. \quad (67)$$

This holds for all δZ, δC if and only if

$$\begin{cases}(K_{xz}K_{zz}^{-1})^*\Delta_{12} + K_{zz}^{-1}Z^*\Delta_{22} = 0\\ C^*\Delta_{22} = 0\end{cases} \quad (68)$$

The first equation in (68) can be simplified so that it reads

$$K_{xz}\Delta_{12} + Z^*\Delta_{22} = 0. \quad (69)$$

The matrix inversion lemma allows to compute an explicit expression for matrix Δ

$$\begin{aligned}\Delta_{12} &= -K_{xx}^{-1}K_{xy}(K_{yy} - K_{xy}^*K_{xx}^{-1}K_{xy})^{-1} +\\ &\quad K_{xx}^{-1}K_{xz}K_{zz}^{-1}Z^*\times\\ &\quad \left[ZK_{zz}^{-1}(K_{zz} - K_{xz}^*K_{xx}^{-1}K_{xz})K_{zz}^{-1}Z^* + CC^*\right]^{-1},\\ \Delta_{22} &= \left(K_{yy} - K_{yz}K_{xx}^{-1}K_{xy}\right)^{-1} -\\ &\quad \left[ZK_{zz}^{-1}(K_{zz} - K_{xz}^*K_{xx}^{-1}K_{xz})K_{zz}^{-1}Z^* + CC^*\right]^{-1}.\end{aligned}$$

Now, using (34), we can write

$$\begin{aligned}\Delta_{12} &= -K_{xx}^{-1}K_{xy}A^{-1} + K_{xx}^{-1}K_{xz}K_{zz}^{-1}Z^*\times\\ &\quad \left[ZK_{zz}^{-1}BK_{zz}^{-1}Z^* + CC^*\right]^{-1},\\ \Delta_{22} &= A^{-1} - \left(ZK_{zz}^{-1}BK_{zz}^{-1}Z^* + CC^*\right)^{-1}.\end{aligned}$$

Therefore, after some manipulation, we conclude that the optimum solution is provided by C and Z solving (33).

Augusto Ferrante was born in Piove di Sacco, Italy, on August 5, 1967. He received the "Laurea" degree, *cum laude*, in Electrical Engineering in 1991 and the Ph.D. degree in Control Systems Engineering in 1995, both from the University of Padova.

He has been on the faculty of the Colleges of Engineering of the University of Udine, and of the "Politecnico di Milano". He is presently Professor in the "Department of Information Engineering" of the University of Padova.

His research interests are in the areas of linear systems, spectral estimation, optimal control and optimal filtering, quantum control, and stochastic realization.

Nicola Laurenti received his Laurea Degree in Electrical Engineering in 1995 and his PhD in Electronic and Telecommunication Engineering in 1999 both from University of Padua, Italy. Since 2001 he has been an Assistant Professor at the Department of Information Engineering of University of Padua. In 2008-09 he was a Visiting Scholar at the Coordinated Science Laboratory of the University of Illinois at Urbana-Champaign. In 1992-93 he was an exchange student at the University of California at Berkeley. His research interests mainly focus on network security at lower layers (physical, data link and network), information theoretic security and quantum key distribution, but also include other aspects of digital communications, especially multicarrier modulation and ultra wide band transmission.

Chiara Masiero was born in Cittadella, Italy, on June 18, 1986. She received the M. Sc. Degree in Automation Engineering in 2010, and the Ph. D. Degree in 2014, both from the University of Padova, Italy. She was a visiting Ph.D. Student at Shanghai Jiao Tong University from September 2012 to February 2013. Her main research interests include system identification and multivariate moment problems, with applications to spectral estimation and wireless communication security. Currently she works as Business Intelligence consultant at Icon-sulting S.p.A.

Michele Pavon was born in Venice, Italy, on October 12, 1950. He received the Laurea degree from the University of Padova, Padova, Italy, in 1974, and the Ph.D. degree from the University of Kentucky, Lexington, in 1979, both in mathematics. After service in the Italian Army, he was on the research staff of LADSEB-CNR, Padua, Italy, for six years.Since July 1986, he has been a Professor at the School of Engineering, the University of Padova. He has visited several institutions in Europe, Northern America and Asia. His present research interests include maximum entropy and optimal transport problems.

Stefano Tomasin (S'99 – M'03 – SM'11) received the Ph.D. degree in Telecommunications Engineering from the University of Padua, Italy, in 2003. He has been doing internships at the IBM Research Laboratory in Switzerland and Philips Research in the Netherlands. In 2005 Dr. Tomasin joined University of Padua as Assistant Professor. Since then he has been visiting for extensive periods Qualcomm in California and the Polytechnic Institute of NYU in New York. His current research interests include physical layer security, signal processing and scheduling for wireless communications, optimization techniques for smart grids. Since 2011 Dr. Tomasin is Editor of both the IEEE Transactions of Vehicular Technologies and the EURASIP Journal of Wireless Communications and Networking.