Channel Covariance Conversion and Modelling Using Infinite Dimensional Hilbert Spaces

Lorenzo Miretti, *Student Member, IEEE,* Renato L. G. Cavalcante, *Member, IEEE,* and Slawomir Stańczak, *Senior Member, IEEE*

***Abstract*—This study considers the channel covariance conversion problem, which consists in estimating the spatial covariance matrix of a wireless channel by exploiting measurements obtained on a different carrier frequency and stationarity properties of the propagation environment across sufficiently close frequency bands. The first contribution given in this study is a modelling framework based on infinite dimensional Hilbert spaces that unifies a plethora of classical and novel covariance models with different degrees of complexity and generality, while still effectively capturing important properties of the propagation environment and of the antenna array. Given this framework, this study addresses the channel covariance conversion problem by proposing two simple yet effective algorithms based on set-theoretic methods that outperform existing model-based approaches both in terms of accuracy and complexity. In particular, the first algorithm is implementable as a simple matrix-vector multiplication. Moreover, in contrast to the aforementioned approaches, both algorithms can be applied to general propagation and array models such as dual-polarized antenna arrays, making them suitable for modern 5G and beyond systems.**

***Index Terms*—Massive MIMO, FDD, spatial covariance, Hilbert spaces, dual-polarized array**

I. Introduction

THE ability of base stations (BSs) to estimate in real-time channel second-order statistics in the form of spatial covariance matrix is of fundamental importance for modern multiple-input multiple-output (MIMO) wireless communication systems adopting massive MIMO technologies [1], [2]. The benefits of this ability have been identified under several contexts, including resource allocation, interference suppression, channel state information (CSI) aquisition, and user localization [3]–[8]. The key idea underlying most of the rich literature on covariance-aided transmission schemes is that practical massive MIMO channels are strongly correlated [5]–[7], [9], [10], and that this correlation can be exploited to represent the channels of multiple user equipment (UE) on different lower-dimensional subspaces.

Manuscript received November 27, 2020; revised April 9, 2021; accepted May 15, 2021. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Lin Bai. (*Corresponding author: Lorenzo Miretti.)*

L. Miretti is with the Department of Communication Systems, EURECOM, Sophia Antipolis 06904, France (e-mail: lorenzo.miretti@eurecom.fr).

R. L. G. Cavalcante and S. Stańczak are with the Department of Wireless Communications and Networks, Fraunhofer Institute for Telecommunications Heinrich-Hertz-Institut, Berlin 10587, Germany (e-mail: renato.cavalcante@hhi.fraunhofer.de, slawomir.stanczak@hhi.fraunhofer.de). S. Stańczak is also with the Department of Telecommunication Systems, Technical University of Berlin, Berlin, Germany.

Digital Object Identifier 10.1109/TSP.2021.3082461

The seminal study [6] shows that spatial covariance information is essential to make massive MIMO technologies practically appealing also for frequency division duplex (FDD) mode, which is severely limited by the channel feedback overhead under the classical yet unrealistic spatially uncorrelated fading assumption. More generally, statistical CSI is relevant to enhance or even replace instantaneous CSI in all system setups where the latter is much more difficult to obtain due to time and/or resources constraints. In fact, as a first approximation, statistical CSI is stable for a time frame T_{WSS} over which the propagation environment can be assumed stationary, which is typically several orders of magnitude larger than the channel coherence time [5], [6].

Nevertheless, efficient downlink (DL) covariance estimation for FDD systems is far from being trivial. For example, traditional closed-loop schemes based on DL training and covariance feedback from the UEs at intervals corresponding to T_{WSS} may still experience excessive overhead, especially for regimes with a large number of BS antennas and with nonnegligible mobility. To overcome this limitation, many authors have considered an alternative approach based on estimating the downlink (DL) covariance matrix from uplink (UL) channel measurements [11]–[16]. Although efficient estimation of large covariance matrices from limited training samples is a challenging problem in itself, we neglect this aspect in this work and focus on the so-called *covariance conversion* problem, defined as the problem of estimating the DL covariance matrix from the UL covariance matrix, which is assumed to be known with sufficiently high accuracy at the BS. Note that the UL covariance matrix is in practice much easier to estimate for massive MIMO systems since UL channel estimation consumes dramatically less resources [2].

Specifically, our contributions to the covariance conversion problem are the following:[1]

- To address the problem with a high degree of generality, we provide a framework based on the theory of infinite dimensional Hilbert spaces that allows us to represent in a unified way different channel spatial covariance models. The framework captures relevant aspects of the propagation environment and of the antenna array design. In particular, our framework covers novel covariance expressions for

[1]Part of the results given by this manuscript was presented at IEEE ICASSP 2018 [17] and at IEEE GlobalSIP 2018 [18]. This manuscript completes and extends [17], [18] by providing a unified presentation, details of derivations, extensions to different channel and system models, and updated comparison with existing literature.

dual-polarized antenna arrays, which, to the best of our knowledge, have not been available in the literature.
- Under the aforementioned modelling framework, we propose two general and efficient algorithms for covariance conversion with different accuracy and complexity. They exploit a statistical form of channel reciprocity for FDD systems called *angular reciprocity*, here specialized by assuming frequency invariance of the so called *angular power spectrum* (APS). This type of reciprocity is at the core of most related schemes addressing the covariance conversion problem, and it is detailed in the following sections.

A. Structure of the paper and preliminaries

This paper is organized as follows: In Section II, we present general models for channel covariance matrices and their unified representation in infinite dimensional Hilbert spaces. The derivations of these models from various popular channel models and system setups are provided in Section V. In Section III, we describe the proposed algorithms for covariance conversion. Specific implementations for uniform linear arrays (ULAs) and uniform planar arrays (UPAs) are discussed in Section IV. The performance of the proposed algorithms are then evaluated via numerical simulations in Section VI. Finally, the advantages of the proposed covariance conversion algorithms over competing approaches, including the benefits and potentials of the developed covariance modelling framework, are discussed in Section VII.

Hereafter, we use the following notation. The operators $(\cdot)^{\mathsf{T}}$ and $(\cdot)^{\mathsf{H}}$ denote respectively the transpose and Hermitian transpose of matrices and vectors, and $\|\cdot\|_{\mathrm{F}}$ is the Frobenius norm. By $\mathbb{R}_+$ we denote the set of nonnegative real numbers, and j is the imaginary unit. We use $\Re[\cdot]$ and $\Im[\cdot]$ to denote, respectively, the real and imaginary parts. The standard vectorization operator, i.e. the operator that stacks column-wise the elements of a matrix $\mathbf{A}$ into a column vector, is denoted by $\mathrm{vec}(\mathbf{A})$. The set of all square Lebesgue integrable functions over a measurable set $\mathcal{I} \subset \mathbb{R}$ is denoted by $L^2[\mathcal{I}]$. We denote by $(\mathcal{H}, \langle\cdot,\cdot\rangle)$ a Hilbert space with inner product $\langle\cdot,\cdot\rangle$ and induced norm $\|x\| := \sqrt{\langle x, x\rangle}$. Given a Hilbert space, we denote by $x^{(i)} \rightharpoonup x$ a sequence $\left(x^{(i)}\right)_{i\in\mathbb{N}}$ weakly convergent to a point x, i.e., a sequence satisfying $(\forall y \in \mathcal{H})\ \lim_{i\to\infty}\langle x^{(i)}, y\rangle = \langle x, y\rangle$. Furthermore, we recall the following relevant property (see, e.g., [19, Ex. 1.4-3]):

Property 1. *Let $(\mathcal{H}_1, \langle\cdot,\cdot\rangle_1)$ and $(\mathcal{H}_2, \langle\cdot,\cdot\rangle_2)$ be two Hilbert spaces. Then $(\mathcal{H}, \langle\cdot,\cdot\rangle)$, where $\mathcal{H} := \mathcal{H}_1 \times \mathcal{H}_2$ with sum and scalar multiplication $(\forall (x_1, x_2) \in \mathcal{H})$, $(\forall (y_1, y_2) \in \mathcal{H})$*

$$(x_1, x_2) + (y_1, y_2) := (x_1 + y_1, x_2 + y_2),$$
$$(\forall k \in \mathbb{R}) \quad k(x_1, x_2) := (kx_1, kx_2),$$

and where $(\forall (x_1, x_2) \in \mathcal{H})$, $(\forall (y_1, y_2) \in \mathcal{H})$

$$\langle (x_1, x_2), (y_1, y_2)\rangle := \langle x_1, y_1\rangle_1 + \langle x_2, y_2\rangle_2,$$

is also a Hilbert space.

Finally, whenever necessary, we use the notation $\mathbf{A}^f$, $f \in \{\mathrm{u}, \mathrm{d}\}$, to empathize the frequency dependency of a matrix $\mathbf{A}$, where u and d denote the UL and DL frequency, respectively.

II. Channel Covariance Models Using Infinite Dimensional Hilbert Spaces

The objective of this section is to provide a general approach to channel spatial covariance modelling based on the abstract framework of infinite dimensional Hilbert spaces [19]–[22]. Specifically, we first review and extend popular covariance models with different degrees of complexity and generality. Then, by leveraging on Hilbert space theory, we show the existence of a unified representation that rigorously describes the impacts of the propagation environment and of the antenna array in a compact form.

A. System model

We consider a MIMO wireless channel between a BS with N antennas and a UE with U antennas. We denote by $\mathbf{H}[t] \in \mathbb{C}^{N\times U}$ a sample[2] of the random channel matrix between the BS and an arbitrary user at time $t \in \mathbb{Z}$. We recall that this model is suitable for describing either narrow-band systems or wide-band multi-carrier systems, e.g., orthogonal frequency division multiplexing (OFDM) systems [23], [24].

We assume $\mathbf{H}[t]$ to be a random process correlated both in time and in the spatial domain, with spatial covariance matrix

$$(\forall t \in \mathbb{Z}) \quad \mathbf{R}[t] := \mathbb{E}\left[\mathrm{vec}(\mathbf{H}[t] - \bar{\mathbf{H}}[t])\mathrm{vec}(\mathbf{H}[t] - \bar{\mathbf{H}}[t])^{\mathsf{H}}\right],$$

where $\bar{\mathbf{H}}[t] := \mathbb{E}\left[\mathbf{H}[t]\right]$ is the channel mean. By the so called *windowed-WSS* assumption, we assume that, for a time window of length T_{WSS}, we can approximate the channel as a wide-sense stationary (WSS) process [24], which implies that $\mathbf{R}[t]$ and $\bar{\mathbf{H}}[t]$ can be assumed constant for T_{WSS} time slots. Furthermore, we define the coherence block interval L_c as the minimum lag l such that $\mathbf{H}[t]$ and $\mathbf{H}[t+l]$ can be considered independent. In narrow-band systems, L_c is usually set equal to the channel coherence time T_c, expressed in discrete time. In wide-band OFDM systems it is usually given by the product $L_c \approx T_c B_c$, where B_c denotes the coherence bandwidth. Moreover, we assume $T_{\mathrm{WSS}} \gg L_c$, which is a reasonable assumption for relatively low-mobility scenarios [5], [10], and is at the core of most literature on wireless networks [1], [2], [23], [24]. Throughout this work, we focus on realizations $\mathbf{R} := \mathbf{R}[t]$, where the time-index is omitted for brevity because of the time invariance of $\mathbf{R}[t]$ over a time window T_{WSS}, and we consider w.l.o.g. $\bar{\mathbf{H}}[t] = \mathbf{0}$ since $\mathbf{R}[t]$ depends only on the zero-mean fading process $\mathbf{H}[t] - \bar{\mathbf{H}}[t]$.

B. Angular reciprocity

In general, the channel $\mathbf{H}[t]$ depends on the carrier frequency, hence classical channel reciprocity where the UL channel $\mathbf{H}^{\mathrm{u}}[t]$ and the DL channel $\mathbf{H}^{\mathrm{d}}[t]$ satisfy $(\forall t \in \mathbb{Z})$ $\mathbf{H}^{\mathrm{u}}[t] = \mathbf{H}^{\mathrm{d}}[t]$ cannot be assumed for FDD systems. However, by representing the channel via *directional* models, a weaker form of channel reciprocity in the so-called *angular domain* can be assumed. For example, by considering single-antenna UEs (i.e. $U = 1$), 2D propagation, and unpolarized antenna

[2]With an abuse of notation, in this work we do not typographically differentiate random variables from their realizations. The interpretation that should be applied will be clear from the context.

arrays, a classical model for the UL and DL spatial covariance matrices $\mathbf{R}^{\mathrm{u}}$ and $\mathbf{R}^{\mathrm{d}}$ gives[3] [4], [5], [10], [25]–[27]

$$(\forall f \in \{\mathrm{u},\mathrm{d}\}) \quad \mathbf{R}^f = \int_\Omega \rho(\theta)\mathbf{a}^f(\theta)\mathbf{a}^f(\theta)^{\mathsf{H}} d\theta, \qquad (1)$$

where $\rho : \Omega \to \mathbb{R}^+$ is the angular power spectrum (APS) describing the received or transmitted power[4] in a given physical direction $\theta \in \Omega \subseteq [-\pi, \pi]$, and $\mathbf{a}^{\mathrm{u}} : \Omega \to \mathbb{C}^{N\times 1}$, $\mathbf{a}^{\mathrm{d}} : \Omega \to \mathbb{C}^{N\times 1}$ are the frequency dependent antenna array responses. The angular reciprocity is modeled by assuming that the APS, unlike the array response, is frequency invariant. This assumption is motivated by several measurement campaigns (see for example [24], [28]), where, for typical duplex gaps (10-100 MHz), the APS is shown to exhibit strong frequency correlation properties. In Section V we also justify this assumption analytically by considering popular channel models in the literature.

C. Directional covariance models for realistic systems

The covariance model (1), owing to its simplicity, is still very popular in the scientific literature, but the increasing complexity of modern wireless communication systems (e.g., the 4G and 5G architectures) has demanded channel models able to describe real propagation phenomena more accurately (see, for example, the relatively recent 3GPP technical report [29]).

Therefore, to address the limitations of (1), in the following we illustrate several directional models for the spatial covariance matrices that consider wave propagation in 3D environments, dual-polarized antenna arrays, and multi-antenna UEs. For generality and to avoid unnecessary technical digressions, in this section we do not specify the underlying model for $\mathbf{H}[t]$. However, the proposed models are formally justified later in Section V.

1) 3D propagation: We assume single-antenna UEs, and unpolarized antenna arrays. The model in (1) can be readily extended to include 3D propagation as follows:

$$(\forall f \in \{\mathrm{u},\mathrm{d}\}) \quad \mathbf{R}^f = \int_\Omega \rho(\boldsymbol{\theta})\mathbf{a}^f(\boldsymbol{\theta})\mathbf{a}^f(\boldsymbol{\theta})^{\mathsf{H}} d\boldsymbol{\theta}, \qquad (2)$$

where $\boldsymbol{\theta} \in \Omega \subseteq [-\pi, \pi] \times [0, \pi]$ represents the azimuth and the zenith in a spherical coordinate system. A detailed justification of the above expression for narrow-band and wide-band OFDM systems is given in Section V-A and Section V-B, respectively.

2) Dual-polarized antenna arrays: We assume single-antenna UEs and 3D propagation. To include dual-polarized antenna arrays, we model the UL and DL covariance matrices as follows: $(\forall f \in \{\mathrm{u},\mathrm{d}\})$

$$\mathbf{R}^f = \int_\Omega \rho_V(\boldsymbol{\theta})\mathbf{a}_V^f(\boldsymbol{\theta})\mathbf{a}_V^f(\boldsymbol{\theta})^{\mathsf{H}} d\boldsymbol{\theta} + \int_\Omega \rho_H(\boldsymbol{\theta})\mathbf{a}_H^f(\boldsymbol{\theta})\mathbf{a}_H^f(\boldsymbol{\theta})^{\mathsf{H}} d\boldsymbol{\theta}, \qquad (3)$$

where $\mathbf{A}^f : \Omega \to \mathbb{C}^{N\times 2} : \boldsymbol{\theta} \mapsto [\mathbf{a}_V^f(\boldsymbol{\theta}), \mathbf{a}_H^f(\boldsymbol{\theta})]$ is the dual polarized antenna array response of the BS, and where the functions $\rho_V, \rho_H : \Omega \to \mathbb{R}^+$, here denominated respectively as *vertical* angular power spectrum (V-APS) and *horizontal* angular power spectrum (H-APS), describe the received or transmitted angular power spectra for the vertically and horizontally polarized waves. Expression (3) is derived in Section V-C by considering a popular channel model for dual-polarized antenna arrays following [29]. Intuitively, this model assumes the two polarizations to fade independently, as discussed in detail in Section V-C. Furthermore, Section V-C also briefly discusses how a slightly more general model than (3) can be used to include the case of dependent fading across polarizations, which is not considered here for simplicity.

3) Multi-antenna UEs: We assume 3D propagation, and polarized antenna-arrays. Directional expressions for the full covariance matrix $\mathbf{R}^f$ are difficult to obtain. Nevertheless, in many applications the second-order statistics of the channel can be partially described by using the *receive* and *transmit* covariance matrices, which are given by [24]

$$\mathbf{R}^{\mathrm{u}}_{\mathrm{RX}} := \mathbb{E}\left[\mathbf{H}^{\mathrm{u}}[t]\mathbf{H}^{\mathrm{u}}[t]^{\mathsf{H}}\right], \quad \mathbf{R}^{\mathrm{u}}_{\mathrm{TX}} := \mathbb{E}\left[\mathbf{H}^{\mathrm{u}}[t]^{\mathsf{H}}\mathbf{H}^{\mathrm{u}}[t]\right],$$
$$\mathbf{R}^{\mathrm{d}}_{\mathrm{RX}} := \mathbb{E}\left[\mathbf{H}^{\mathrm{d}}[t]^{\mathsf{H}}\mathbf{H}^{\mathrm{d}}[t]\right], \quad \mathbf{R}^{\mathrm{d}}_{\mathrm{TX}} := \mathbb{E}\left[\mathbf{H}^{\mathrm{d}}[t]\mathbf{H}^{\mathrm{d}}[t]^{\mathsf{H}}\right].$$

By focusing on $\mathbf{R}^{\mathrm{u}}_{\mathrm{RX}}$ and on $\mathbf{R}^{\mathrm{d}}_{\mathrm{TX}}$, which intuitively describe the channel statistics "seen" from the BS point of view, we consider the following models:

$$\mathbf{R}^{\mathrm{u}}_{\mathrm{RX}} = \tilde{\mathbf{R}}^{\mathrm{u}}, \quad \mathbf{R}^{\mathrm{d}}_{\mathrm{TX}} = \tilde{\mathbf{R}}^{\mathrm{d}}, \qquad (4)$$

where $\tilde{\mathbf{R}}^f$ $(f \in \{\mathrm{u},\mathrm{d}\})$ is given by the right-hand side of either (2) for the unpolarized case or (3) for the dual-polarized case. The above expressions are justified in Section V-D. Note that similar expressions for $\mathbf{R}^{\mathrm{d}}_{\mathrm{RX}}$ and $\mathbf{R}^{\mathrm{u}}_{\mathrm{TX}}$ can be obtained by switching the role of BS and UE.

As a last important remark, we assume that *all* the functions defined in this section are square-integrable. More precisely, we assume $\rho, \rho_V, \rho_H \in L^2(\Omega)$. Similarly, by letting a to be an arbitrary entry of $\mathbf{a}^{\mathrm{u}}$, $\mathbf{a}^{\mathrm{d}}$, $\mathbf{A}^{\mathrm{u}}$, $\mathbf{A}^{\mathrm{d}}$, we assume $\Re[a], \Im[a] \in L^2(\Omega)$. Note that this assumption is reasonable, because in real-world systems these functions are typically continuous and bounded on a compact domain owing to their physical meaning in terms of transmit or received power for a given direction.

D. Unified representation using Hilbert spaces

The common characteristic of all spatial covariance models (1), (2), (3), and (4) illustrated in the previous sections is that they are given by integral expressions involving essentially two terms:

- A term that is frequency invariant, referred in general to as APS, which describes the distribution of the received or transmitted power in the angular domain Ω.
- A term that is frequency dependent, which describes the BS antenna array response in the angular domain Ω.

The frequency invariance property of the APS is motivated by channel reciprocity in the angular-domain, as discussed in Section II-B. Furthermore, we point out that these terms are also time invariant. In particular, the APS is assumed to be

[3]Throughout this work, integrals involving vectors and matrices should be understood coordinate-wise.

[4]Note that this physical interpretation refers only to the signal components originating the zero-mean fading process, i.e., it does not consider the channel mean, which typically corresponds to a line-of-sight component.

TABLE I: Mapping rules for channel covariance modeling using infinite dimensional Hilbert spaces.

Unpolarized	Dual-polarized
$\rho \in \mathcal{H} := L^2(\Omega)$	$\rho = (\rho_V, \rho_H) \in \mathcal{H} := L^2(\Omega) \times L^2(\Omega)$
$\langle x, y \rangle := \int_\Omega x(\boldsymbol{\theta}) y(\boldsymbol{\theta}) d\boldsymbol{\theta}$	$\langle x, y \rangle := \int_\Omega x_V(\boldsymbol{\theta}) y_V(\boldsymbol{\theta}) d\boldsymbol{\theta} + \int_\Omega x_H(\boldsymbol{\theta}) y_H(\boldsymbol{\theta}) d\boldsymbol{\theta}$
$g_m^f(\boldsymbol{\theta}) := T_m(\mathbf{a}^f(\boldsymbol{\theta})\mathbf{a}^f(\boldsymbol{\theta})^{\mathsf{H}})$	$g_m^f := \left(g_{m,V}^f, g_{m,H}^f\right)$ $g_{m,p}^f(\boldsymbol{\theta}) := T_m\left(\mathbf{a}_p^f(\boldsymbol{\theta})\mathbf{a}_p^f(\boldsymbol{\theta})^{\mathsf{H}}\right)$ $p = V, H$

constant over a time window T_{WSS} as described in Section II-A. The important feature of these models is that they are able to pinpoint some effects of the propagation environment, which is relatively robust to frequency changes, and of the antenna array. Interestingly, this general description can be formalized in a unified way. To this end, let us consider the equivalent vectorized real representation $\mathbf{r}^f := T(\mathbf{R}^f)$ of the matrices[5] $\mathbf{R}^f \in \mathbb{C}^{N\times N}$ given by the bijective map

$$T : \mathbb{C}^{N\times N} \to \mathbb{R}^M : \mathbf{X} \mapsto \mathrm{vec}\left([\Re[\mathbf{X}]\ \Im[\mathbf{X}]]\right),$$

where $M := 2N^2$. Let us further denote by $T_m(\mathbf{X})$ the mth element of $T(\mathbf{X})$. We can now unify the expressions for the different channel covariance models as follows.

Remark 1. *The expressions* (1)*,* (2)*,* (3)*, and* (4) *correspond to a system of linear equations in a suitable infinite dimensional Hilbert space* $(\mathcal{H}, \langle\cdot,\cdot\rangle)$ *of the form*

$$r_m^f = \langle \rho, g_m^f \rangle, \quad m = 1, \ldots, M, \quad f \in \{\mathrm{u}, \mathrm{d}\}, \tag{5}$$

where $r_m^f := T_m(\mathbf{R}^f)$*, and where* $\rho, g_m^f \in \mathcal{H}$*. The rules for mapping* (1)*,* (2)*,* (3)*, and* (4) *to the corresponding expressions in the chosen Hilbert space are given in Table I.*

Notice that since covariance matrices are Hermitian symmetric, the number of different equations in (5) are at most N^2. It is possible to modify the map T such that all the duplicated equations of (5) are removed, but we omit this trivial operation for notational simplicity.

By establishing a connection between channel covariance models and the theory of Hilbert spaces, Remark 1 unfolds a rich body of modern signal processing tools that can be applied to wireless communications problems in a unified way. Guided by this observation, in Section III we derive novel solutions for the relevant application of channel covariance conversion. We point out that similar mathematical formalisms have been already applied to many other wireless communications problems (see, e.g., [22]). However, to the best of our knowledge, channel covariance models with a degree of generality such as in (3) have not been covered before.

III. Set-Theoretic Covariance Conversion

In this section we propose a novel technique to infer $\mathbf{R}^{\mathrm{d}}$ from the observed UL covariance $\mathbf{R}^{\mathrm{u}}$, given the directional models described in Section II. Throughout this section, we consider the unified representation of such models in a suitable infinite dimensional Hilbert space given by Remark 1. The main idea builds on channel reciprocity in the angular domain, and it can be summarized into two steps as follows:

1) Given $\mathbf{R}^{\mathrm{u}}$ we obtain an estimate $\hat{\rho}$ of the APS ρ from (5) and known properties of ρ.
2) We compute an estimate of $\mathbf{R}^{\mathrm{d}}$ by using (5) with ρ replaced by its estimate $\hat{\rho}$.

In particular, the APS estimation problem in the first step is addressed by formalizing it as a *convex feasibility* problem. We propose two versions of a set-theoretic approach that differ in the definition of the solution set, leading to two variants of the proposed algorithm with different accuracy-complexity trade-offs. In this work, analytic or experimental knowledge of the array responses and hence of the functions g_m^f in (5) is assumed; this knowledge is cell-independent and it holds for the entire lifetime of the antenna array. Note that, although related, APS estimation is different from classical parametric or nonparametric direction of arrival (DoA) estimation, where the goal is typically the localization within Ω of a finite set of radiating sources [30, Chapter 6].

A. Projection onto a linear variety (PLV)

The inverse problem of finding ρ given g_m^{u} and r_m^{u}, $m = 1, \ldots, M$ is typically ill-posed: Unless strong additional assumptions on ρ are available, it is generally impossible to guarantee its perfect recovery from a finite set of measurements. In this study, inspired by the set-theoretic paradigm [19], [31]–[33], we estimate ρ by solving

$$\text{find } \rho^* \in V := \cap_{m=1}^M V_m \neq \emptyset, \tag{6}$$

where $V_m := \{\rho \in \mathcal{H} : \langle \rho, g_m^{\mathrm{u}} \rangle = r_m^{\mathrm{u}}\}$ for $m = 1, \ldots, M$. Notice that non-emptiness of V is guaranteed[6] since it contains at least the true APS. The above problem is a feasibility problem involving simple hyperplanes. The set-theoretic paradigm relies on the notion of feasibility to produce solutions that are consistent to all information arising from input data [31]. All the candidate solutions of (6) are equivalent based only on the information given by $\mathbf{r}^{\mathrm{u}} = T(\mathbf{R}^{\mathrm{u}})$. However, to keep the resulting algorithm simple, we choose the unique minimum norm solution

$$\hat{\rho} \in \arg\min_{\rho \in V} \|\rho\|,$$

which corresponds to the orthogonal projection $P_V(0)$ of the zero vector onto the linear variety V. This projection has the following well-known closed-form expression [20, Chapter 3]:

$$\hat{\rho} = \sum_{m=1}^M \alpha_m g_m^{\mathrm{u}}, \tag{7}$$

where $\boldsymbol{\alpha} := [\alpha_1 \ldots \alpha_M]$ is any solution to the linear system

$$\mathbf{r}^{\mathrm{u}} = \mathbf{G}^{\mathrm{u}} \boldsymbol{\alpha}, \tag{8}$$

[5]Given the equivalence of their expressions, with abuse of notation in what follows we use $\mathbf{R}^{\mathrm{u}}$ and $\mathbf{R}^{\mathrm{d}}$ also to denote $\mathbf{R}_{\mathrm{RX}}^{\mathrm{u}}$ and $\mathbf{R}_{\mathrm{TX}}^{\mathrm{d}}$ respectively.

[6]This is not necessarily true for imperfect knowledge of $\mathbf{R}^{\mathrm{u}}$ (see Sect. III-C), where non-emptiness should be checked case by case. However, based on our experiments, this assumption is hardly violated in practice.

TABLE II: Definition of the cone $Z \in \mathcal{H}$ and associated projection $P_Z : \mathcal{H} \to \mathcal{H}$ for different choices of Hilbert space.

Unpolarized	Dual-polarized
$Z := \{x \in \mathcal{H} :\ (\forall \boldsymbol{\theta} \in \Omega)$ $x(\boldsymbol{\theta}) \geq 0$ a.e.$\}$	$Z := \{(x_V, x_H) \in \mathcal{H} :\ (\forall \boldsymbol{\theta} \in \Omega)$ $x_V(\boldsymbol{\theta}) \geq 0,\ x_H(\boldsymbol{\theta}) \geq 0$ a.e.$\}$
$P_Z(x) = \bar{x}$ $\bar{x}(\boldsymbol{\theta}) := \begin{cases} x(\boldsymbol{\theta}), \text{if } x(\boldsymbol{\theta}) \geq 0 \\ 0, \text{otherwise} \end{cases}$	$P_Z((x_V, x_H)) = (\bar{x}_V, \bar{x}_H)$ $\bar{x}_p(\boldsymbol{\theta}) := \begin{cases} x_p(\boldsymbol{\theta}), \text{if } x_p(\boldsymbol{\theta}) \geq 0 \\ 0, \text{otherwise} \end{cases}$ $p = V, H$

where

$$\mathbf{G}^{\mathrm{u}} = \begin{bmatrix} \langle g_1^{\mathrm{u}}, g_1^{\mathrm{u}} \rangle & \langle g_1^{\mathrm{u}}, g_2^{\mathrm{u}} \rangle & \dots & \langle g_1^{\mathrm{u}}, g_M^{\mathrm{u}} \rangle \\ \langle g_2^{\mathrm{u}}, g_1^{\mathrm{u}} \rangle & \langle g_2^{\mathrm{u}}, g_2^{\mathrm{u}} \rangle & \dots & \langle g_2^{\mathrm{u}}, g_M^{\mathrm{u}} \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle g_M^{\mathrm{u}}, g_1^{\mathrm{u}} \rangle & \langle g_M^{\mathrm{u}}, g_2^{\mathrm{u}} \rangle & \dots & \langle g_M^{\mathrm{u}}, g_M^{\mathrm{u}} \rangle \end{bmatrix}.$$

Note that the linear system in (8) is guaranteed to have at least one solution (from the projection theorem, all solutions give the unique projection in (7)).

We then obtain an estimate of $\mathbf{R}^{\mathrm{d}}$ by replacing ρ in (5) with its estimate $\hat{\rho}$ obtained in (7):

$$\hat{r}_m^{\mathrm{d}} = \langle \hat{\rho}, g_m^{\mathrm{d}} \rangle = \sum_{l=1}^{M} \alpha_l \langle g_l^{\mathrm{u}}, g_m^{\mathrm{d}} \rangle \quad m = 1 \dots M, \tag{9}$$

which can be rewritten in matrix form as

$$\hat{\mathbf{r}}^{\mathrm{d}} = \mathbf{Q}\boldsymbol{\alpha},$$

where $\hat{\mathbf{r}}^{\mathrm{d}}$ is an estimate of the vector $\mathbf{r}^{\mathrm{d}} = T(\mathbf{R}^{\mathrm{d}})$, $\boldsymbol{\alpha}$ is a solution to the linear system (8) given the UL measurements $\mathbf{r}^{\mathrm{u}} = \mathbf{G}^{\mathrm{u}}\boldsymbol{\alpha}$ as mentioned above, and

$$\mathbf{Q} = \begin{bmatrix} \langle g_1^{\mathrm{d}}, g_1^{\mathrm{u}} \rangle & \langle g_1^{\mathrm{d}}, g_2^{\mathrm{u}} \rangle & \dots & \langle g_1^{\mathrm{d}}, g_M^{\mathrm{u}} \rangle \\ \langle g_2^{\mathrm{d}}, g_1^{\mathrm{u}} \rangle & \langle g_2^{\mathrm{d}}, g_2^{\mathrm{u}} \rangle & \dots & \langle g_2^{\mathrm{d}}, g_M^{\mathrm{u}} \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle g_M^{\mathrm{d}}, g_1^{\mathrm{u}} \rangle & \langle g_M^{\mathrm{d}}, g_2^{\mathrm{u}} \rangle & \dots & \langle g_M^{\mathrm{d}}, g_M^{\mathrm{u}} \rangle \end{bmatrix}.$$

Finally, it can be easily seen that the above algorithm can be implemented via the following simple operation

$$\mathbf{r}^{\mathrm{d}} = \mathbf{F}\mathbf{r}^{\mathrm{u}}, \quad \mathbf{F} := \mathbf{Q}(\mathbf{G}^{\mathrm{u}})^{\dagger} \tag{10}$$

where $(\mathbf{G}^{\mathrm{u}})^{\dagger}$ is the Moore-Penrose pseudo-inverse of $\mathbf{G}^{\mathrm{u}}$. Note that both $\mathbf{G}^{\mathrm{u}}$ and $\mathbf{Q}$ (and hence $\mathbf{F}$) depend only on the array responses, and they can thus be computed only once for the entire system lifetime.

B. Exploiting non-negativity of the APS

In the following, we propose an extension of the previous algorithm by considering that, being a power spectrum, ρ is non-negative. More precisely, we solve

$$\text{find } \rho^* \in C := V \cap Z, \tag{11}$$

where V is the linear variety defined in (6) and Z is the closed convex cone of (pairs of) non-negative functions in $\mathcal{H}$. The formal definition of Z for the Hilbert spaces in Table I is given in Table II.

A solution to (11) can be found by applying one of the many existing iterative projection methods for convex feasibility problems available in literature. These methods typically produce a sequence $(\rho^{(i)})_{i \in \mathbb{N}} \subset \mathcal{H}$ such that $\rho^{(i)} \rightharpoonup \rho^*$, where ρ^* is some point in C. In particular, we use the following fast iterative method called *extrapolated alternating projection method (EAPM)* [34], which produces a sequence $(\rho^{(i)})_{i \in \mathbb{N}}$ via

$$\rho^{(i+1)} = \rho^{(i)} + \nu K_i \left[P_V(P_Z(\rho^{(i)})) - \rho^{(i)} \right] \quad (\forall i \in \mathbb{N}), \tag{12}$$

where $\nu \in (0, 2)$ is a step size, and K_i is the extrapolation parameter defined as

$$K_i = \begin{cases} \dfrac{\|P_Z(\rho^{(i)}) - \rho^{(i)}\|^2}{\|P_V(P_Z(\rho^{(i)})) - \rho^{(i)}\|^2}, & \text{if } \rho^{(i)} \notin Z \\ 1, & \text{if } \rho^{(i)} \in Z \end{cases}.$$

The initial condition $\rho^{(0)} \in V$ can be arbitrary, but here it is set to $\rho^{(0)} = P_V(0)$, the solution proposed in Section III-A. The projection $P_V : \mathcal{H} \to \mathcal{H}$ onto the set V is given by [20, Chapter 3]

$$(\forall x \in \mathcal{H}) \quad P_V(x) = x - \sum_{m=1}^{M} \beta_m g_m^{\mathrm{u}} + P_V(0),$$

with $\boldsymbol{\beta} := [\beta_1 \dots \beta_M]$ being a solution to the linear system $\mathbf{b} = \mathbf{G}^{\mathrm{u}}\boldsymbol{\beta}$ where the m-th element of $\mathbf{b}$ is given by $b_m = \langle x, g_m^{\mathrm{u}} \rangle$ and $\mathbf{G}^{\mathrm{u}}$ is the same as in (8). The projection $P_Z : \mathcal{H} \to \mathcal{H}$ for the given choice of Hilbert space is given in Table II (see also [19, p. 284]).

Now, by proceeding along the same lines as in Section III-A, an estimate of $\mathbf{R}^{\mathrm{d}}$ can be obtained by letting

$$\hat{r}_m^{\mathrm{d}} = \langle \hat{\rho}, g_m^{\mathrm{d}} \rangle \quad m = 1, \dots, M.$$

where $\hat{\rho}$ results from (12).

We conclude this section by pointing out that, in contrast to the PLV algorithm given in Section III-A, each iteration of EAPM typically requires the *online* computation of inner products in the form of integrals. A relatively low-complexity implementation of EAPM is obtained for example by numerically approximating such integrals as

$$\int_{\Omega} x(\boldsymbol{\theta}) y(\boldsymbol{\theta}) d\boldsymbol{\theta} \approx \sum_{d=1}^{D} x(\boldsymbol{\theta}_d) y(\boldsymbol{\theta}_d) \Delta_d,$$

where $\{\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_D\} \subset \Omega$ are the elements of a discrete grid of D samples of the spherical coordinate system Ω, and where $\Delta_d \in \mathbb{R}^+$, $d = 1, \dots, D$, are some given weighting coefficients. We remark that the above approximation can be equivalently interpreted as working in a finite-dimensional Hilbert space equipped with the standard Euclidean inner product, and where we replace ρ and g_n^f by their sampled versions as similarly done for the NNLS method of [13]. Both the discussed numerical implementation of EAPM and the NNLS method produce a solution to a discrete approximation of (11). Another appealing projection-based alternative is the variant of the Douglas-Rachford method studied in [35], where the authors have recently proved convergence in a finite number of steps for problems as the discrete approximation of (11).

C. Imperfect uplink covariance knowledge

The algorithms presented so far assume the perfect knowledge of $\mathbf{R}^{\mathrm{u}}$. In this section instead we consider a practical scenario in which the BS has access only to the UL sample covariance

$$\bar{\mathbf{C}}^{\mathrm{u}} := \frac{1}{N_s}\sum_{n=1}^{N_s}\hat{\mathbf{H}}^{\mathrm{u}}[n](\hat{\mathbf{H}}^{\mathrm{u}}[n])^{\mathsf{H}}$$

computed from a limited number N_s of channel estimates given by $\hat{\mathbf{H}}^{\mathrm{u}}[n] = \mathbf{H}^{\mathrm{u}}[nL_c]+\mathbf{Z}[n]$, where $\mathbf{Z}[n]$ is a temporally white stationary noise process independent of $\mathbf{H}^{\mathrm{u}}[nL_c]$ and with $\mathbb{E}\left[\mathbf{Z}[n]\mathbf{Z}[n]^{\mathsf{H}}\right] =: \mathbf{\Sigma}_z$. It is important to underline that the samples $\mathbf{H}^{\mathrm{u}}[nL_c]$ are taken with a spacing equal to the coherence time L_c, so that they can be considered independent. This imperfect knowledge leads to a performance degradation of the proposed algorithms. However, in the following, we discuss a correction procedure that can be applied to $\bar{\mathbf{C}}^{\mathrm{u}}$ in order to mitigate these effects.

Let $(\mathcal{H}_S, \langle\cdot,\cdot\rangle)$ be the complex Hilbert space of all $N \times N$ Hermitian matrices whose inner product is defined by $\langle \mathbf{A}, \mathbf{B}\rangle = \mathrm{trace}(\mathbf{B}^{\mathsf{H}}\mathbf{A})$, and let $\mathcal{S}_+$ be the subset of $\mathcal{H}_S$ composed by positive semi-definite (PSD) matrices. Similarly to [36], we use the matrix $\bar{\mathbf{R}}^{\mathrm{u}} := \bar{\mathbf{C}}^{\mathrm{u}} - \mathbf{\Sigma}_z$ to obtain an estimate of $\mathbf{R}^{\mathrm{u}}$ by projecting it onto $\mathcal{S}_+$ as follows:

$$P_{\mathcal{S}_+}(\bar{\mathbf{R}}^{\mathrm{u}}) = \mathbf{U}\mathbf{\Delta}_+\mathbf{U}^{\mathsf{H}},$$

with $\mathbf{U}$ and $\mathbf{\Delta}_+$ obtained from the eigen-decomposition $\bar{\mathbf{R}}^{\mathrm{u}} = \mathbf{U}\mathbf{\Delta}\mathbf{U}^{H}$, and by defining $\mathbf{\Delta}_+ := \max(\mathbf{\Delta}, \mathbf{0})$, which is a short-hand for an element-wise standard $\max(\cdot,\cdot)$ operator over real numbers. Note that this procedure is presented in [36] as the maximum-likelihood (ML) PSD estimate for Gaussian distributed channel vectors. In this work, this operation is applied without assuming Gaussian distributed channels. In fact, regardless of the channel distribution, if $\bar{\mathbf{R}}^{\mathrm{u}} \notin \mathbb{S}_+$ the firmly nonexpansiveness property of projections guarantees that this correction procedure produces a PSD estimate which is strictly closer to $\mathbf{R}^{\mathrm{u}}$ w.r.t. the Hilbertian metric $d(\mathbf{A},\mathbf{B}) := \|\mathbf{A}-\mathbf{B}\|_{\mathrm{F}} = \langle \mathbf{A}-\mathbf{B}, \mathbf{A}-\mathbf{B}\rangle^{\frac{1}{2}}$ induced by the chosen inner product.

Moreover, according to the specific array geometry, the covariance matrix often shows an additional structure on top of the positive semi-definiteness. Thus, it is reasonable to further process $P_{\mathcal{S}_+}(\bar{\mathbf{R}}^{\mathrm{u}})$ to restore the envisioned structure. The description of this procedure is left for the following sections, because it depends on the particular array model.

IV. Implementation Aspects for ULA and UPA

In this section we discuss practical implementation aspects of the proposed schemes when specialized to uniform linear arrays (ULAs), and uniform planar arrays (UPAs) with cross-polarized antenna pairs. The former model is very popular in the scientific literature due to its simplicity, while the latter is more relevant for realistic massive MIMO systems.

A. Implementation for uniform linear arrays

In this section we discuss the implementation of the proposed schemes for ULA with N antennas at the BS. Due to their geometrical properties which do not allow to distinguish waves in a 3D coordinate system, we consider a 2D propagation environment $\Omega \in [-\pi, \pi]$. Recall that the array response of a ULA is given by

$$\mathbf{a}(\theta) = \frac{1}{\sqrt{N}}\left[1 \;\; e^{j2\pi\frac{d}{\lambda}\sin\theta} \;\; \dots \;\; e^{j2\pi\frac{d}{\lambda}(N-1)\sin\theta}\right]^T,$$

where $d \in \mathbb{R}$ and $\lambda \in \mathbb{R}$ denote, respectively, the inter-antenna spacing and the carrier wavelength. Moreover, since ULAs are not able to distinguish among a DoA θ and its reciprocal $\theta + \pi$, we assume that the APS is confined to the interval $\Omega = [-\pi/2, \pi/2]$, which implies $\langle f, g\rangle = \int_{-\pi/2}^{\pi/2} f(\theta)g(\theta)d\theta$. This assumption is supported by the fact that real systems often work with a similar or even narrower cell sectorization. Furthermore, for ULAs, channel covariance matrices are Hermitian Toeplitz, so they can be completely represented by their first columns. For this reason, we can remove redundant equations in Proposition 1 by redefining $T(\mathbf{X}) := \mathrm{vec}([\Re[\mathbf{x}_1]\ \Im[\mathbf{x}_1]])$, where $\mathbf{x}_1$ denotes the first column of $\mathbf{X}$.

1) Analytical expressions for $\mathbf{G}^{\mathrm{u}}$ and $\mathbf{Q}$: In the next proposition we show that for ULAs the matrix $\mathbf{G}^{\mathrm{u}}$ and $\mathbf{Q}$ can be computed in closed-form.

Proposition 1. *For ULAs, the matrices $\mathbf{G}^{\mathrm{u}}$ and $\mathbf{Q}$ defined in Sect. III-A take the following analytical form in terms of the Bessel function of the first kind, zero order $J_0 : \mathbb{R} \to \mathbb{R}$:*

$$\mathbf{G}^{\mathrm{u}} = \frac{\pi}{2N^2}\begin{bmatrix}\mathbf{G}_{\Re} & \mathbf{0}\\ \mathbf{0} & \mathbf{G}_{\Im}\end{bmatrix} \quad \mathbf{Q} = \frac{\pi}{2N^2}\begin{bmatrix}\mathbf{Q}_{\Re} & \mathbf{0}\\ \mathbf{0} & \mathbf{Q}_{\Im}\end{bmatrix},$$

where the elements corresponding to the (n,m)-entries of $\mathbf{G}_{\Re}$, $\mathbf{G}_{\Im}$, $\mathbf{Q}_{\Re}$,$\mathbf{Q}_{\Im} \in \mathbb{R}^{N\times N}$ are given by

$$\mathbf{G}_{\Re,nm} = J_0(x_{nm}) + J_0(y_{nm}),\ \mathbf{Q}_{\Re,nm} = J_0(p_{nm}) + J_0(q_{nm}),$$
$$\mathbf{G}_{\Im,nm} = J_0(x_{nm}) - J_0(y_{nm}),\ \mathbf{Q}_{\Im,nm} = J_0(p_{nm}) - J_0(q_{nm}),$$

and where, for $n, m = 1, \dots, N$,

$$x_{nm} = 2\pi\frac{d}{\lambda^{\mathrm{u}}}(n-m), \quad p_{nm} = 2\pi d\left(\frac{n-1}{\lambda^{\mathrm{d}}} - \frac{m-1}{\lambda^{\mathrm{u}}}\right),$$
$$y_{nm} = 2\pi\frac{d}{\lambda^{\mathrm{u}}}(n+m-2), \quad q_{nm} = 2\pi d\left(\frac{n-1}{\lambda^{\mathrm{d}}} + \frac{m-1}{\lambda^{\mathrm{u}}}\right).$$

Proof: The chosen vectorization operator gives

$$g^f_m(\theta) = \frac{1}{N}\cos\left(2\pi\frac{d}{\lambda^f}(m-1)\sin\theta\right),\ 1 \le m \le N,$$
$$g^f_m(\theta) = \frac{1}{N}\sin\left(2\pi\frac{d}{\lambda^f}(m-N-1)\sin\theta\right),\ N < m \le 2N.$$

The proof follows by applying standard *product-to-sum* trigonometric identities, and by making use of the integral representation of the Bessel function

$$J_0(x) = \frac{1}{\pi}\int_0^{\pi}\cos(x\sin(\theta))d\theta.$$

■

2) Improving the estimation of the UL covariance matrix: The direct feeding of either $\bar{\mathbf{R}}^{\mathrm{u}}$ or the $P_{\mathcal{S}_+}(\bar{\mathbf{R}}^{\mathrm{u}})$ defined in Section III-C as input to the proposed algorithms may result in poor performance because the Toeplitz assumption imposed by the ULA is not satisfied.

To mitigate this problem, a possible approach is to feed as input the projection of $\bar{\mathbf{R}}^{\mathrm{u}}$ onto the set $\mathcal{T}_+ := \mathcal{S}_+ \cap \mathcal{T}$, where $\mathcal{T} \subset \mathcal{H}_S$ is the set of Toeplitz matrices, given by

$$\hat{\mathbf{R}}^{\mathrm{u}} \in \arg\min_{\mathbf{X}\in\mathcal{T}_+} \|\mathbf{X} - \bar{\mathbf{R}}^{\mathrm{u}}\|_{\mathrm{F}}. \tag{13}$$

Since the projections on $\mathcal{S}_+$ and $\mathcal{T}$ are known [37] and easy to compute, it is possible to compute $\hat{\mathbf{R}}^{\mathrm{u}}$ by applying standard methods such as the Dykstra's or Haugazeau's algorithm [21, Chapter 20]. In this work, we use the approach described in [37], which solves the following relaxation of (13)

$$\text{find } \mathbf{X}^* \in \mathcal{T}_+ \cap \mathcal{D},\ \mathcal{D} := \{\mathbf{X} \in \mathcal{H}_{\mathcal{S}} : \|\mathbf{X} - \bar{\mathbf{R}}^{\mathrm{u}}\|_{\mathrm{F}} \leq \delta\}, \tag{14}$$

where δ is a tunable error tolerance, by using an alternating projection method producing a sequence convergent to a point in $\mathcal{T}_+ \cap \mathcal{D}$. Note that δ should be chosen such that $\mathcal{T}_+ \cap \mathcal{D}$ is non-empty.

B. Implementation for uniform planar arrays with cross-polarized antennas

We now consider a UPA with cross-polarized antennas, defined as a rectangular grid of identical and equispaced antenna elements, each of them composed of a pair of two vertically polarized antennas with a polarization slant of $\pm 45°$. We denote by N_V and N_H respectively the number of vertical and horizontal elements, and by d the horizontal and vertical inter-antenna spacing. We further denote by $x(u, v, 1)$ the antenna in position (u, v), $u = 1, \ldots, N_V$ and $v = 1, \ldots, N_H$, with $+45°$ polarization slant, and by $x(u, v, 2)$ the co-located antenna with $-45°$ polarization slant. By arranging every column $\mathbf{h}$ of the channel matrix $\mathbf{H}$ as $\mathbf{h} := \begin{bmatrix}\mathbf{h}_1^{\mathsf{T}} & \mathbf{h}_2^{\mathsf{T}}\end{bmatrix}^{\mathsf{T}}$, where the channel coefficient for antenna $x(u, v, k)$ corresponds to the nth element of the vector $\mathbf{h}_k \in \mathbb{C}^{N_V N_H}$, with $n = (u-1)N_H + v$, the array response is given by

$$(\forall p \in \{V, H\}), (\forall \boldsymbol{\theta} \in \Omega) \quad \mathbf{a}_p(\boldsymbol{\theta}) := \begin{bmatrix}\mathbf{a}_{p,1}^{\mathsf{T}}(\boldsymbol{\theta}) & \mathbf{a}_{p,2}^{\mathsf{T}}(\boldsymbol{\theta})\end{bmatrix}^{\mathsf{T}},$$
$$\mathbf{a}_{p,1}(\boldsymbol{\theta}) := a_{p,1}(\boldsymbol{\theta}) e^{j\boldsymbol{\Psi}(\boldsymbol{\theta})}, \quad \mathbf{a}_{p,2}(\boldsymbol{\theta}) := a_{p,2}(\boldsymbol{\theta}) e^{j\boldsymbol{\Psi}(\boldsymbol{\theta})},$$

where $a_{V,1}(\boldsymbol{\theta}) \in \mathbb{R}_+$ and $a_{H,1}(\boldsymbol{\theta}) \in \mathbb{R}_+$ are the vertical and horizontal radiation pattern of the $+45°$ polarized antennas, $a_{V,2}(\boldsymbol{\theta}) \in \mathbb{R}_+$, $a_{H,2}(\boldsymbol{\theta}) \in \mathbb{R}_+$ the vertical and horizontal radiation patters for the $-45°$ polarized antennas, and where we used the shorthand

$$e^{j\boldsymbol{\Psi}(\boldsymbol{\theta})} := \begin{bmatrix}e^{j\Psi_1(\boldsymbol{\theta})} & e^{j\Psi_2(\boldsymbol{\theta})} & \ldots & e^{j\Psi_{N_V N_H}(\boldsymbol{\theta})}\end{bmatrix}^{\mathsf{T}},$$

where $\Psi_n(\boldsymbol{\theta}) = \Psi_n(\theta_1, \theta_2)$ is the geometry-only dependent phase term of an antenna in position (u, v), given by

$$\Psi_{(u-1)N_H+v}(\boldsymbol{\theta}) = \\ = 2\pi\frac{d}{\lambda}\left[(u-1)\cos(\theta_1) + (v-1)\sin(\theta_1)\sin(\theta_2)\right].$$

For this antenna array, in contrast to the ULA, the presence of realistic radiation patterns makes analytical expressions for $\mathbf{G}^{\mathrm{u}}$ and $\mathbf{Q}$ difficult to derive. However, we can still identify the following covariance structure:

Proposition 2. *By assuming without loss of generality that $N_V \geq N_H$, the covariance matrix for the considered UPA takes on the following block structure:*

$$\mathbf{R} = \begin{bmatrix}\mathbf{B}_1 & \mathbf{B}_2^{\mathsf{H}} \\ \mathbf{B}_2 & \mathbf{B}_3\end{bmatrix} \in \mathbf{C}^{2N_V N_H \times 2N_V N_H},$$

where every macro-block $\mathbf{B}_l \in \mathbf{C}^{N_V N_H \times N_V N_H}$, $l = 1, 2, 3$, is Hermitian and it has the following block structure:

$$\mathbf{B}_l = \begin{bmatrix} \mathbf{B}_{l,1} & & & & \\ \mathbf{B}_{l,2} & \mathbf{B}_{l,1} & & & \\ \mathbf{B}_{l,3} & \mathbf{B}_{l,2} & \mathbf{B}_{l,1} & & \\ \vdots & \vdots & \vdots & \ddots & \\ \mathbf{B}_{l,N_V} & \ldots & \mathbf{B}_{l,3} & \mathbf{B}_{l,2} & \mathbf{B}_{l,1} \end{bmatrix},$$

where every block $\mathbf{B}_{l,i} \in \mathbf{C}^{N_H \times N_H}$, $i = 1, \ldots N_V$ has identical diagonal entries b_{li}, and every block $\mathbf{B}_{l,1}$ is Hermitian Toeplitz.

The proof is omitted here but we point out that it follows directly from the definition of the matrices $\mathbf{a}_V(\boldsymbol{\theta})\mathbf{a}_V(\boldsymbol{\theta})^{\mathsf{H}}$ and $\mathbf{a}_H(\boldsymbol{\theta})\mathbf{a}_H(\boldsymbol{\theta})^{\mathsf{H}}$ of (3). In light of Proposition 2, we can specialize the proposed algorithms as follows.

1) Efficient covariance vectorization: The covariance matrix $\mathbf{R}$ can be fully described by using the blocks $\mathbf{B}_{l,i} \in \mathbb{C}^{N_H \times N_H}$, $i = 1, \ldots N_V$, $l = 1, 2, 3$. Furthermore, the blocks $\mathbf{B}_{l,1}$ can be represented by only N_H complex numbers because of the Hermitian Toeplitz structure, while the blocks $\mathbf{B}_{l,i}$, $i \neq 1$, can be represented by $N_H^2 - (N_H - 1)$ complex numbers because only one diagonal entry is sufficient. Overall, this means that it is possible to redefine the map T so that we can completely reconstruct $\mathbf{R}$ by using $M = 6(N_H + (N_V - 1)(N_H^2 - N_H + 1))$ real numbers, instead of $M = 2(N_V N_H)^2$. As a result, the algorithm has substantially lower complexity compared to the structure-unaware algorithm presented in Section III-A.

2) Improving the estimation of the UL covariance matrix: Similarly to Section IV-A2, the estimation $\hat{\mathbf{R}}^{\mathrm{u}}$ of the covariance matrix $\mathbf{R}^{\mathrm{u}}$ obtained from the sample covariance as described in Section III-C can be further improved by exploiting its particular structure. In this work, we consider modified versions of (13) and (14) obtained by replacing $\mathcal{T}$ with the subspace $\mathcal{W} \subset \mathcal{H}_{\mathcal{S}}$ of matrices with elements constrained to be equal according to Proposition 2. Again, these problems can be efficiently solved by using the same methods as in Section IV-A2. In fact, it can be easily shown that the projection $P_{\mathcal{W}}(\mathbf{X}_0) \in \arg\min_{\mathbf{X}\in\mathcal{W}} \|\mathbf{X} - \mathbf{X}_0\|_{\mathrm{F}}$ is simply obtained by replacing every element of $\mathbf{X}_0$ with the arithmetic average of all its elements that are constrained to be equal to that element.

Remark 2 (Alternative covariance model)**.** *Arrays with pairs of co-located cross-polarized antennas have been also considered in the recent study [38], where, for 2D propagation,*

$N_V = 1$, and single antenna UEs, the authors propose the following model:

$$\begin{aligned}\mathbf{R} &= \mathbb{E}[\mathbf{h}\mathbf{h}^{\mathsf{H}}] = \begin{bmatrix}\mathbf{R}_1 & \mathbf{R}_{12}\\ \mathbf{R}_{12}^{\mathsf{H}} & \mathbf{R}_2\end{bmatrix},\\ \mathbf{R}_k &= \textstyle\int_\Omega \gamma_k(\theta) e^{j\mathbf{\Psi}(\theta)}(e^{j\mathbf{\Psi}(\theta)})^{\mathsf{H}} d\theta, \quad k = 1, 2,\\ \mathbf{R}_{12} &= \textstyle\int_\Omega \gamma_{12}(\theta) e^{j\mathbf{\Psi}(\theta)}(e^{j\mathbf{\Psi}(\theta)})^{\mathsf{H}} d\theta,\end{aligned} \tag{15}$$

where $\gamma_k : \Omega \to \mathbb{R}_+$ is again a power spectrum, and where $\gamma_{12} : \Omega \to \mathbb{C}$ is a cross-correlation function. Similar expressions can be directly obtained from (3) *by identifying $\gamma_k(\theta) := \rho_V(\theta) a_{V,k}^2(\theta) + \rho_H(\theta) a_{H,k}^2(\theta)$ and $\gamma_{12}(\theta) := \rho_V(\theta) a_{V,1}(\theta) a_{V,2}(\theta) + \rho_H(\theta) a_{H,1}(\theta) a_{H,2}(\theta)$. These identifications can be interpreted as unpolarized models similar to* (1)*, where only the phase response is made explicit, and where the radiation patterns are embedded into equivalent power spectra. The above identifications consider γ_{12} to be real-valued and non-negative, and provide a clear physical meaning to the cross-correlation between $\mathbf{h}_1$ and $\mathbf{h}_2$, which intuitively differ only in terms of (real-valued and non-negative) radiation patterns. Furthermore, unlinke* (15)*, the proposed model* (3) *fully separates the effect of the impinging wave and of the antenna array, since the radiation pattern is made explicit. Note that such property is of interest for all problems (as the covariance conversion problem considered in here) where the full array response is known, and hence $\mathbf{R}$ can be determined by estimating (ρ_V, ρ_H) rather than $(\gamma_1, \gamma_2, \gamma_{12})$.*

V. Derivation of Directional Channel Covariance Models for Realistic Systems

The goal of this section is to provide formal justifications for the spatial channel covariance expressions presented in Section II. To widen as much as possible the domain of application of the proposed algorithms, we carry a comprehensive analysis of several popular directional channel models. These models are presented in a bottom-up fashion, by focusing on different aspects with increasing complexity. In particular, we first review how robust these models are to different channel modelling philosophies and to different system designs (narrowband or wideband OFDM). Then, we derive similar expressions for dual-polarized antenna arrays and multiple antennas at the UE, which, to the best of the authors knowledge, are not available in the literature.

A. Narrow-band systems

1) Discrete scattering: Let us consider a narrow-band system in a 2D (azimuth-only) environment. A classical expression for a realization of the channel $\mathbf{h}^f$ at an arbitrary time and frequency $f \in \{\mathrm{u}, \mathrm{d}\}$ is given by [5], [24]:

$$\mathbf{h}^f = \frac{1}{\sqrt{N_p}} \sum_{c=1}^{N_c} \sum_{i=1}^{N_p} \sqrt{\alpha_c} e^{j\varphi_{ic}} \mathbf{a}^f(\theta_{ic}) \tag{16}$$

where the channel vector $\mathbf{h}^f \in \mathbb{C}^{N\times 1}$ is expressed in terms of its multipath components, possibly clustered according to the popular *geometry-based* stochastic channel model (GSCM), which generalizes the so called *one-ring* model used to justify the traditional i.i.d. Rayleigh fading assumption [24]. The details are listed in the following.

- N_c denotes the number of clusters and N_p the associated number of subpaths.
- $\theta_{ic} \in \Omega \subseteq [-\pi, \pi]$ is the direction of arrival (DoA) or the direction of departure (DoD) of subpath i belonging to cluster c, respectively for the UL and for the DL. It is assumed to be drawn from a frequency invariant probability density function $f_c \in L^2(\Omega)$.
- $\alpha_c > 0$ is the power of all the subpaths of cluster c.
- $\mathbf{a}^f \in \mathbb{C}^N$ is the frequency dependent BS array response, with elements a_n^f satisfying $\Re[a_n^f], \Im[a_n^f] \in L^2(\Omega)$, $\forall n$.
- φ_{ic} are the phase shift terms of each path, i.i.d. uniformly distributed in $[-\pi, \pi]$.

Remark 3. *We point out that frequency invariance of the distribution f_c is a milder assumption than the frequency invariance of the realizations θ_{ic} as sometimes assumed in the literature relying on angular reciprocity.*

The GSCM is particularly suitable for outdoor environments, where the clusters have the physical meaning of macro-objects responsible for the main reflections in the cell. For this reason, a model very similar to (16) is implemented in many simulators; e.g. the ones compliant with the 3GPP technical document [39]. The main difference between (16) and the model proposed by 3GPP is that, in the latter, only a single DoA/DoD for each cluster is drawn statistically, while the the remaining subpaths are obtained deterministically from tables. This is done mainly to reduce the complexity of the simulation. In this work, instead, the proposed model is kept more general to not confine subpaths angles into a pre-defined grid, which is likely to be unrealistic. Moreover, we stress that in this work we do not specify any type of statistical dependence between UL and DL realizations of the random variables φ_{ic} and θ_{ic}.

Furthermore, we point out that (16), in contrast to [39], does not take into account the time dependent phase term $e^{j2\pi\nu_{ic}t}$, where t is the time and ν_{ic} is the Doppler shift of subpath i of cluster c, which models the short-term time evolution of the channel. However, as the focus of this work is on long-term channel statistics, we consider only a long-term time evolution model, given in a statistical sense. More precisely, we model the time evolution of the channel as follows. The fast time-varying parameters φ_{ic} and θ_{ic} are drawn independently and kept fixed at intervals corresponding to the coherence time T_c (block-fading assumption); the slow time-varying parameters α_c and f_c are assumed constant over a WSS window $T_{\mathrm{WSS}} \gg T_c$. This model reflects the classical *windowed* WSS assumption described in Section II-A.

The next proposition shows the expression for the spatial channel covariance under the above channel model.

Proposition 3. *For $f \in \{\mathrm{u}, \mathrm{d}\}$, the covariance matrix $\mathbf{R}^f := \mathbb{E}[\mathbf{h}^f(\mathbf{h}^f)^{\mathsf{H}}]$ for the channel model in* (16)*, is given by*

$$\mathbf{R}^f = \int_\Omega \rho(\theta) \mathbf{a}^f(\theta) \mathbf{a}^f(\theta)^{\mathsf{H}} d\theta,$$

where the function $\rho : \Omega \to \mathbb{R}^+$ is the angular power

spectrum[7]*, and it is given by*

$$\rho(\theta) := \sum_{c=1}^{N_c} f_c(\theta)\alpha_c. \tag{17}$$

Proof: The proof follows as a special case of Proposition 5 or 6 shown later and hence it is omitted. ∎

Remark 4. *Note that assuming discrete angles θ_{ic} drawn from a continuous probability density function f_c is different than assuming discrete spectral components as sometimes done in related literature. In this work, we do not consider discrete spectral components, although we do allow for APS components $\alpha_c f_c$ with arbitrarily small yet non-zero support.*

2) Continuous scattering: Most of simulators and algorithms available in the massive MIMO literature and also those for problems related to, for example, DoA estimation [30] rely on discrete scattering models similar to the one presented in Section V-A1. However some authors (see, for example, [10]) consider generalizations of the discrete scattering model described in Section V-A1 by assuming the channel vector to be formed by a superposition of a continuum of array responses $\mathbf{a}^f(\theta)$, weighted by a continuous function $\sqrt{\rho}$. More precisely, these generalizations consider

$$(\forall f \in \{\mathrm{u},\mathrm{d}\}) \quad \mathbf{h}^f = \int_\Omega \sqrt{\rho(\theta)} z(\theta)\mathbf{a}^f(\theta)d\theta, \tag{18}$$

where $\rho : \Omega \to \mathbb{R}^+$ is the APS describing the received or transmitted power per unit angle, and where $z(\theta)$ is a white unitary power complex random process in the angular domain modelling small-scale fading. Angular reciprocity is here modeled assuming equal APS for UL and DL. The time evolution of the channel is modelled similarly to Section V-A1, i.e., $z(\theta)$ is drawn independently and kept fixed at intervals corresponding to the coherence time T_c, while the APS is assumed constant over a WSS window $T_{WSS} \gg T_c$.

The next proposition shows that the expressions for the covariance matrix for the channel model considered here is similar to the one considering the discrete scattering model in Section V-A1.

Proposition 4. *[10] For $f \in \{\mathrm{u},\mathrm{d}\}$, the covariance matrix $\mathbf{R}^f := \mathbb{E}[\mathbf{h}^f(\mathbf{h}^f)^{\mathsf{H}}]$ for the channel model in* (18) *is given by*

$$\mathbf{R}^f = \int_\Omega \rho(\theta)\mathbf{a}^f(\theta)\mathbf{a}^f(\theta)^{\mathsf{H}}d\theta. \tag{19}$$

B. Wide-band OFDM systems

Let us consider a wide-band channel in an under-spread environment; i.e. with delay spread $T_s \ll T_c$, an assumption that is typically done while designing an OFDM system [23, Chapter 3.4]. Let us extend the narrow-band GSCM discrete scattering model (16) by using the same approach proposed in [24, Chapter 6] and in the 3GPP technical document [39], denoted as *tapped delay line*. After sampling, and using $l \in \mathbb{N}$ to denote the discrete time index of the lth tap, the sampled impulse response in the delay domain, for frequency $f \in \{\mathrm{u},\mathrm{d}\}$, is given by:

$$\mathbf{h}^f[l] = \sum_{c=1}^{N_c} \mathbf{h}_c^f \delta[l - l_c], \quad \mathbf{h}_c^f := \sqrt{\frac{\alpha_c}{N_p}} \sum_{i=1}^{N_p} e^{j\varphi_{ic}}\mathbf{a}^f(\theta_{ic}), \tag{20}$$

where $\delta[l]$ is the discrete unit sample function, $l_c \in \mathbb{N}$ denotes the discrete time delay of all the subpaths belonging to cluster c, which are assumed to be unresolvable in the delay domain after sampling, and where all the other parameters and random variables are defined as in Section V-A1. We assume l_c to be equal for UL and DL, and constant over T_{WSS}.

The channel vector in the sub-carrier domain of an OFDM system is then given by [23, Chapter 3.4]:

$$(\forall f \in \{\mathrm{u},\mathrm{d}\}) \quad \tilde{\mathbf{h}}^f[k] = \sum_{l=0}^{L-1} \mathbf{h}^f[l] e^{-j\frac{2\pi kl}{N_s}}, \tag{21}$$

where L is the impulse response length, N_s is the chosen OFDM block length, and $k = 0, \ldots, (N_s - 1)$ is the sub-carrier index.

Proposition 5. *For $f \in \{\mathrm{u},\mathrm{d}\}$, the covariance matrix in the sub-carrier domain $\tilde{\mathbf{R}}^f[k] := \mathbb{E}[\tilde{\mathbf{h}}[k]^f(\tilde{\mathbf{h}}^f[k])^{\mathsf{H}}]$ for the channel model in* (21) *is given by*

$$\tilde{\mathbf{R}}^f[k] = \int_\Omega \rho(\theta)\mathbf{a}^f(\theta)\mathbf{a}^f(\theta)^{\mathsf{H}}d\theta, \quad \forall k,$$

where the APS $\rho : \Omega \to \mathbb{R}^+$ is given by (17)*.*

Proof: Let us drop for simplicity the UL/DL superscript. By recalling the channel model in (20), we obtain

$$\begin{aligned}
\mathbf{R}[l,l'] &:= \mathbb{E}\left[\mathbf{h}[l]\mathbf{h}^{\mathsf{H}}[l']\right] \\
&= \sum_{c=1}^{N_c}\sum_{c'=1}^{N_c} \mathbb{E}\left[\mathbf{h}_c\mathbf{h}_{c'}^{\mathsf{H}}\right] \delta[l - l_c]\delta[l' - l_{c'}] \\
&\overset{(a)}{=} \sum_{c=1}^{N_c} \mathbb{E}\left[\mathbf{h}_c\mathbf{h}_c^{\mathsf{H}}\right] \delta[l - l_c]\delta[l' - l_c] \\
&= \sum_{c=1}^{N_c} \mathbb{E}\left[\mathbf{h}_c\mathbf{h}_c^{\mathsf{H}}\right] \delta[l - l_c]\delta[l - l'] \\
&\overset{(b)}{=} \sum_{c=1}^{N_c} \left(\int_\Omega \alpha_c f_c(\theta)\mathbf{a}(\theta)\mathbf{a}^H(\theta)d\theta\right) \delta[l - l_c]\delta[l - l'],
\end{aligned}$$

where equalities (a) and (b) follow from the property $\mathbb{E}\left[e^{j\varphi_{ic}}e^{-j\varphi_{i'c'}}\right] = \delta[i - i']\delta[c - c']$ due to the uncorrelated phases of the multipath components, and by further simple manipulations. By considering now the channel in the sub-carrier domain defined in (21), we obtain

$$\begin{aligned}
\tilde{\mathbf{R}}[k] &= \sum_{l=0}^{L-1}\sum_{l'=0}^{L-1} \mathbf{R}[l,l']e^{-j\frac{2\pi}{N_s}k(l-l')} \\
&= \sum_{l=0}^{L-1}\sum_{c=1}^{N_c} \left(\int_\Omega \alpha_c f_c(\theta)\mathbf{a}(\theta)\mathbf{a}^H(\theta)d\theta\right) \delta[l - l_c] \\
&= \int_\Omega \left(\sum_{c=1}^{N_c} \alpha_c f_c(\theta)\right) \mathbf{a}(\theta)\mathbf{a}^{\mathsf{H}}(\theta)d\theta.
\end{aligned}$$

∎

[7]The name angular power spectrum is justified by its physical interpretation as a power density $\int_\Omega \rho(\theta)d\theta = \sum \alpha_c$.

Remark 5. *A similar expression as in Proposition 5 could be also derived for a continuous scattering modelling philosophy, by considering an extension of* (18) *given by*

$$\mathbf{h}^f[l] = \int_\Omega \sqrt{\gamma(\theta,l)} z(\theta,l)\mathbf{a}^f(\theta)d\theta,$$

where $\gamma : \Omega \times \mathbb{N} \to \mathbb{R}^+$ describes the received or transmitted power per unit angle on the l-th tap, and where $z(\theta,l)$ is a unitary power complex random process, jointly white in the angular and delay domain.

C. Realistic propagation and antennas

In this section we extend the simple channel model for narrow-band systems and discrete scattering of Section V-A1 to a more realistic model which takes into account 3D propagation and dual-polarized antenna arrays, by following closely [29, Eq. (7.3-22)]. Specifically, we assume the channel vector $\mathbf{h}^f$ at an arbitrary time and frequency $f \in \{\mathrm{u},\mathrm{d}\}$ to be given by

$$\mathbf{h}^f := \sum_{c=1}^{N_c} \mathbf{h}_c^f, \quad \mathbf{h}_c^f := \sqrt{\frac{\alpha_c}{N_p}} \sum_{i=1}^{N_p} \mathbf{A}^f(\boldsymbol{\theta}_{ic})\mathbf{M}_{ic}\mathbf{B}(\boldsymbol{\phi}_{ic})^{\mathsf{H}}, \tag{22}$$

where

- N_c denotes the number of clusters and N_p the associated number of subpaths.
- $\boldsymbol{\theta}_{ic}$ and $\boldsymbol{\phi}_{ic}$ are either the DoD and DoA of subpath i of cluster c for the DL case, or the DoA and DoD of subpath i of cluster c for the UL case. The directions $\boldsymbol{\theta}_{ic}$ and $\boldsymbol{\phi}_{ic}$ are defined as tuples taking values in the set $\Omega \in [-\pi,\pi] \times [0,\pi]$, representing the azimuth and the zenith of a spherical coordinate system. They are drawn independently from a frequency invariant joint distribution $f_c \in L^2[\Omega] \times L^2[\Omega]$.
- $\mathbf{A}^f : \Omega \to \mathbb{C}^{N\times 2}$ is the frequency dependent dual-polarized BS antenna array response. The columns of $\mathbf{A}^f$ are denoted by $[\mathbf{a}_V^f, \mathbf{a}_H^f] := \mathbf{A}^f$ and they describe respectively the array responses for the vertical and for the horizontal polarization.
- $\alpha_c > 0$ is the power of all subpaths of cluster c.
- $\mathbf{B} : \Omega \to \mathbb{R}_+^{1\times 2}$ is the dual-polarized antenna radiation pattern of the UE. It is assumed to be frequency independent. The columns of $\mathbf{B}$ are denoted with $[b_V, b_H] := \mathbf{B}$, and they describe respectively the radiation patterns for the vertical and for the horizontal polarization. Note that the models in Section V-A and Section V-B consider omni-directional antennas at the UE, thus ignoring its angular response.
- The random matrix

$$\mathbf{M}_{ic} := \begin{bmatrix} e^{j\varphi_{VV,ic}} & \frac{1}{\sqrt{K_{ic}}} e^{j\varphi_{VH,ic}} \\ \frac{1}{\sqrt{K_{ic}}} e^{j\varphi_{HV,ic}} & e^{j\varphi_{HH,ic}} \end{bmatrix},$$

model the fading of the vertical and horizontal polarization as well as of the cross-polarization terms originated by the polarization changes that the electromagnetic waves undergo during the propagation. The random phases $\{\varphi_{VV,ic}, \varphi_{VH,ic}, \varphi_{HV,ic}, \varphi_{HH,ic}\} =: \boldsymbol{\varphi}$ are assumed i.i.d. uniformly distributed in $[-\pi,\pi]$. The parameters K_{ic} are the cross-polarization power ratios (XPRs), and they are assumed to be i.i.d. random variables. This polarization model is identical to the one suggested by the 3GPP technical document [29] and by [24, Chapter 7], where the two polarizations are assumed to fade independently.

We model the time evolution of the channel as follows: the fast time-varying parameters $\boldsymbol{\varphi}$, $\boldsymbol{\theta}_{ic}$, $\boldsymbol{\phi}_{ic}$, and K_{ic} are drawn independently and kept fixed at intervals corresponding to the coherence time T_c; slowly-varying parameters α_c and f_c are assumed constant over a WSS window $T_{\mathrm{WSS}} \gg T_c$.

Proposition 6. *For $f \in \{\mathrm{u},\mathrm{d}\}$, the covariance matrix $\mathbf{R}^f := \mathbb{E}[\mathbf{h}^f(\mathbf{h}^f)^{\mathsf{H}}]$ for the channel model in (22) is given by*

$$\mathbf{R}^f = \int_\Omega \rho_V(\boldsymbol{\theta})\mathbf{a}_V^f(\boldsymbol{\theta})\mathbf{a}_V^f(\boldsymbol{\theta})^{\mathsf{H}}d\boldsymbol{\theta} + \int_\Omega \rho_H(\boldsymbol{\theta})\mathbf{a}_H^f(\boldsymbol{\theta})\mathbf{a}_H^f(\boldsymbol{\theta})^{\mathsf{H}}d\boldsymbol{\theta} \tag{23}$$

where the functions $\rho_V, \rho_H : \Omega \to \mathbb{R}^+$, here denominated respectively as "vertical polarization angular power spectrum" (V-APS) and "horizontal polarization angular power spectrum" (H-APS) are given by

$$\rho_V(\boldsymbol{\theta}) := \sum_{c=1}^{N_c} \alpha_c \int_\Omega f_c(\boldsymbol{\theta},\boldsymbol{\phi}) \left(b_V^2(\boldsymbol{\phi}) + \frac{1}{K} b_H^2(\boldsymbol{\phi}) \right) d\boldsymbol{\phi},$$

$$\rho_H(\boldsymbol{\theta}) := \sum_{c=1}^{N_c} \alpha_c \int_\Omega f_c(\boldsymbol{\theta},\boldsymbol{\phi}) \left(b_H^2(\boldsymbol{\phi}) + \frac{1}{K} b_V^2(\boldsymbol{\phi}) \right) d\boldsymbol{\phi},$$

where $1/K := \mathbb{E}[1/K_{ic}]$ describes the average effect of the XPRs K_{ic}.

Proof: Let us drop for simplicity the UL/DL superscripts. By recalling the channel model in (22), and by computing expectations over the random phases $\boldsymbol{\varphi}$ conditioned on the random angles $\boldsymbol{\theta} := \{\boldsymbol{\theta}_{ic}\}$, $\boldsymbol{\phi} := \{\boldsymbol{\phi}_{ic}\}$, and on the random XPRs $\mathbf{K} := \{K_{ic}\}$, we obtain

$$\begin{aligned} \mathbf{R}'_{cc'} &:= \mathbb{E}[\mathbf{h}_c\mathbf{h}_{c'}^{\mathsf{H}} | \boldsymbol{\theta},\boldsymbol{\phi},\mathbf{K}] \\ &= \frac{\sqrt{\alpha_c\alpha_{c'}}}{N_p} \sum_{i=1}^{N_p}\sum_{i'=1}^{N_p} \mathbf{A}(\boldsymbol{\theta}_{ic})\mathbf{X}_{ic,i'c'}\mathbf{A}(\boldsymbol{\theta}_{i'c'})^{\mathsf{H}}, \end{aligned}$$

where

$$\begin{aligned} \mathbf{X}_{ic,i'c'} :=& \mathbb{E}\left[\mathbf{M}_{ic}\mathbf{B}(\boldsymbol{\phi}_{ic})^{\mathsf{H}}\mathbf{B}(\boldsymbol{\phi}_{i'c'})\mathbf{M}_{i'c'}^{\mathsf{H}} \middle| \boldsymbol{\theta},\boldsymbol{\phi},\mathbf{K}\right] \\ =& \begin{bmatrix} b_V^2(\boldsymbol{\phi}_{ic}) + \frac{b_H(\boldsymbol{\phi}_{ic})^2}{K_{ic}} & 0 \\ 0 & b_H^2(\boldsymbol{\phi}_{ic}) + \frac{b_V(\boldsymbol{\phi}_{ic})^2}{K_{ic}} \end{bmatrix} \\ & \times \delta[i-i']\delta[c-c']. \end{aligned}$$

The expression for $\mathbf{X}_{ic,i'c'}$ is due to the phases $\boldsymbol{\varphi}$ being i.i.d. uniform, and its derivation is similar to that in Section V-B. Since there is no correlation among different multipath components, by averaging over the XPRs $\mathbf{K}$ conditioned over the random angles $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$, we obtain

$$\begin{aligned} \mathbf{R}''_{cc} &:= \mathbb{E}[\mathbf{R}'_{cc} | \boldsymbol{\theta},\boldsymbol{\phi}] \\ &= \frac{\alpha_c}{N_p} \sum_{i=1}^{N_p}\sum_{i'=1}^{N_p} \mathbf{A}(\boldsymbol{\theta}_{ic})\mathbb{E}[\mathbf{X}_{ic,i'c} | \boldsymbol{\phi}]\mathbf{A}(\boldsymbol{\theta}_{ic})^{\mathsf{H}} \end{aligned}$$

$$= \frac{\alpha_c}{N_p} \sum_{i=1}^{N_p} \mathbf{A}(\boldsymbol{\theta}_{ic}) \mathbb{E}[\mathbf{X}_{ic,ic}|\boldsymbol{\phi}] \mathbf{A}(\boldsymbol{\theta}_{ic})^{\mathsf{H}}$$
$$= \frac{\alpha_c}{N_p} \sum_{i=1}^{N_p} \left(b_V^2(\boldsymbol{\phi}_{ic}) + \frac{b_H^2(\boldsymbol{\phi}_{ic})}{K} \right) \mathbf{a}_V(\boldsymbol{\theta}_{ic}) \mathbf{a}_V(\boldsymbol{\theta}_{ic})^{\mathsf{H}}$$
$$+ \frac{\alpha_c}{N_p} \sum_{i=1}^{N_p} \left(b_H^2(\boldsymbol{\phi}_{ic}) + \frac{b_V^2(\boldsymbol{\phi}_{ic})}{K} \right) \mathbf{a}_H(\boldsymbol{\theta}_{ic}) \mathbf{a}_H(\boldsymbol{\theta}_{ic})^{\mathsf{H}},$$

where we define $1/K := \mathbb{E}[1/K_{ic}]$. By letting now

$$\tilde{b}_V^2(\boldsymbol{\phi}) := b_V^2(\boldsymbol{\phi}) + \frac{b_H^2(\boldsymbol{\phi})}{K}, \; \tilde{b}_H^2(\boldsymbol{\phi}) := b_H^2(\boldsymbol{\phi}) + \frac{b_V^2(\boldsymbol{\phi})}{K},$$

and again since there is no correlation among different multi-path components, we obtain

$$\mathbf{R} = \sum_{c=1}^{N_c} \mathbb{E}\left[\mathbf{R}''_{cc}\right]$$
$$= \sum_{c=1}^{N_c} \Bigg[\int_\Omega \int_\Omega \alpha_c f_c(\boldsymbol{\theta}, \boldsymbol{\phi}) \tilde{b}_V^2(\boldsymbol{\phi}) \mathbf{a}_V(\boldsymbol{\theta}) \mathbf{a}_V(\boldsymbol{\theta})^{\mathsf{H}} d\boldsymbol{\theta} d\boldsymbol{\phi}$$
$$+ \int_\Omega \int_\Omega \alpha_c f_c(\boldsymbol{\theta}, \boldsymbol{\phi}) \tilde{b}_H^2(\boldsymbol{\phi}) \mathbf{a}_H(\boldsymbol{\theta}) \mathbf{a}_H(\boldsymbol{\theta})^{\mathsf{H}} d\boldsymbol{\theta} d\boldsymbol{\phi} \Bigg]$$
$$= \int_\Omega \left[\sum_{c=1}^{N_c} \alpha_c \int_\Omega f_c(\boldsymbol{\theta}, \boldsymbol{\phi}) \tilde{b}_V^2(\boldsymbol{\phi}) d\boldsymbol{\phi} \right] \mathbf{a}_V(\boldsymbol{\theta}) \mathbf{a}_V(\boldsymbol{\theta})^{\mathsf{H}} d\boldsymbol{\theta}$$
$$+ \int_\Omega \left[\sum_{c=1}^{N_c} \alpha_c \int_\Omega f_c(\boldsymbol{\theta}, \boldsymbol{\phi}) \tilde{b}_H^2(\boldsymbol{\phi}) d\boldsymbol{\phi} \right] \mathbf{a}_H(\boldsymbol{\theta}) \mathbf{a}_H(\boldsymbol{\theta})^{\mathsf{H}} d\boldsymbol{\theta},$$

and the expression (23) is proved. ∎

Remark 6. *If the two polarizations do not fade independently, the resulting covariance model have additional terms of the type $\int_\Omega \rho_{VH}(\boldsymbol{\theta}) \mathbf{a}_V(\boldsymbol{\theta}) \mathbf{a}_H(\boldsymbol{\theta})^{\mathsf{H}} d\boldsymbol{\theta}$ modelling cross-correlation among the two polarizations. The resulting expression is still representable in an infinite dimensional Hilbert space similarly to Section II-D. However, for simplicity, in this work we do not consider this scenario.*

Remark 7. *By following a similar approach as in Sections V-A2 and V-B, expressions of the type in* (23) *can be derived also by considering a continuous scattering model and/or OFDM. The details are omitted due to space limitations.*

D. Multi-antenna UEs

Let us consider the directional channel model (22) extended to the multi-antenna UE case as follows [29, Eq. (7.3-22)]:

$$\mathbf{H}^f = \sum_{c=1}^{N_c} \mathbf{H}_c^f, \quad \mathbf{H}_c^f := \sqrt{\frac{\alpha_c}{N_p}} \sum_{i=1}^{N_p} \mathbf{A}^f(\boldsymbol{\theta}_{ic}) \mathbf{M}_{ic} \mathbf{B}^f(\boldsymbol{\phi}_{ic})^{\mathsf{H}}, \tag{24}$$

$\forall f \in \{\mathrm{u}, \mathrm{d}\}$, where all the quantities are identical to (22), except for the effect of the UE that is now modeled by using the frequency dependent dual-polarized antenna array responses $\mathbf{B}^f : \Omega \to \mathbb{C}^{U \times 2}$. The columns of $\mathbf{B}^f$ are denoted by $[\mathbf{b}_V^f, \mathbf{b}_H^f]$, and they describe respectively the array responses for the vertical and for the horizontal polarization. We also define $\|\mathbf{b}_V(\boldsymbol{\phi})\|^2 := \|\mathbf{b}_V^f(\boldsymbol{\phi})\|^2$ and $\|\mathbf{b}_H(\boldsymbol{\phi})\|^2 := \|\mathbf{b}_H^f(\boldsymbol{\phi})\|^2$, which are assumed to be frequency invariant because they do not depend on the phase response of the array, but just on the magnitude (i.e. the radiation pattern) for each antenna element, which is generally assumed to be frequency independent.

Proposition 7. *The UL receive covariance matrix $\mathbf{R}_{\mathrm{RX}}^{\mathrm{u}} := \mathbb{E}[\mathbf{H}^{\mathrm{u}}(\mathbf{H}^{\mathrm{u}})^{\mathsf{H}}]$ and the DL transmit covariance matrix $\mathbf{R}_{\mathrm{TX}}^{\mathrm{d}} := \mathbb{E}[\mathbf{H}^{\mathrm{d}}(\mathbf{H}^{\mathrm{d}})^{\mathsf{H}}]$ for the channel model in (24) are given by*

$$\mathbf{R}_{\mathrm{RX}}^{\mathrm{u}} = \int_\Omega \rho_V(\boldsymbol{\theta}) \mathbf{a}_V^{\mathrm{u}}(\boldsymbol{\theta}) \mathbf{a}_V^{\mathrm{u}}(\boldsymbol{\theta})^{\mathsf{H}} d\boldsymbol{\theta} + \int_\Omega \rho_H(\boldsymbol{\theta}) \mathbf{a}_H^{\mathrm{u}}(\boldsymbol{\theta}) \mathbf{a}_H^{\mathrm{u}}(\boldsymbol{\theta})^{\mathsf{H}} d\boldsymbol{\theta},$$
$$\mathbf{R}_{\mathrm{TX}}^{\mathrm{d}} = \int_\Omega \rho_V(\boldsymbol{\theta}) \mathbf{a}_V^{\mathrm{d}}(\boldsymbol{\theta}) \mathbf{a}_V^{\mathrm{d}}(\boldsymbol{\theta})^{\mathsf{H}} d\boldsymbol{\theta} + \int_\Omega \rho_H(\boldsymbol{\theta}) \mathbf{a}_H^{\mathrm{d}}(\boldsymbol{\theta}) \mathbf{a}_H^{\mathrm{d}}(\boldsymbol{\theta})^{\mathsf{H}} d\boldsymbol{\theta},$$

where the functions $\rho_V, \rho_H : \Omega \to \mathbb{R}^+$ are given by

$$\rho_V(\boldsymbol{\theta}) := \sum_{c=1}^{N_c} \alpha_c \int_\Omega f_c(\boldsymbol{\theta}, \boldsymbol{\phi}) \left(\|\mathbf{b}_V(\boldsymbol{\phi})\|^2 + \frac{1}{K} \|\mathbf{b}_H(\boldsymbol{\phi})\|^2 \right) d\boldsymbol{\phi},$$
$$\rho_H(\boldsymbol{\theta}) := \sum_{c=1}^{N_c} \alpha_c \int_\Omega f_c(\boldsymbol{\theta}, \boldsymbol{\phi}) \left(\|\mathbf{b}_H(\boldsymbol{\phi})\|^2 + \frac{1}{K} \|\mathbf{b}_V(\boldsymbol{\phi})\|^2 \right) d\boldsymbol{\phi},$$

and where $1/K := \mathbb{E}[1/K_{ic}]$.

Proof: (sketch) Let us drop the UL and DL superscripts for simplicity. By focusing on the expression for the matrix $\mathbb{E}[\mathbf{H}\mathbf{H}^{\mathsf{H}}]$, the proof is identical to the one presented in Section V-C, by replacing $b_V(\boldsymbol{\phi})^2$ and $b_H(\boldsymbol{\phi})^2$ with $\|\mathbf{b}_V(\boldsymbol{\phi})\|^2$ and $\|\mathbf{b}_H(\boldsymbol{\phi})\|^2$, respectively. ∎

VI. Performance Evaluation

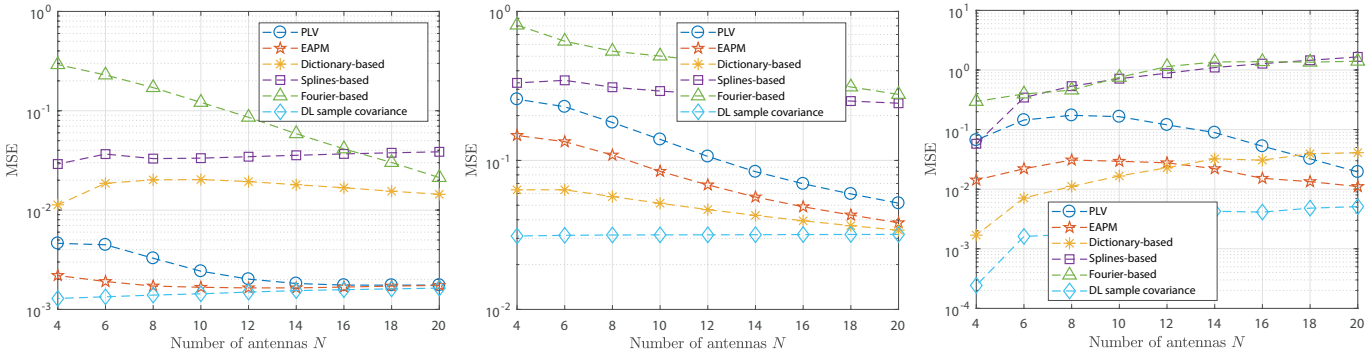
A. Comparison with state-of-the-art techniques

For simplicity, in this first numerical evaluation, we assume the following correlated Rayleigh channel model:

$$\mathbf{h}^{\mathrm{u}}[t] \sim \mathcal{CN}(\mathbf{0}, \mathbf{R}^{\mathrm{u}}), \quad \mathbf{h}^{\mathrm{d}}[t] \sim \mathcal{CN}(\mathbf{0}, \mathbf{R}^{\mathrm{d}}),$$

with spatial covariance matrices $\mathbf{R}^{\mathrm{u}}$ and $\mathbf{R}^{\mathrm{d}}$ given by (1). We simulate a simple 2D model for the APS inspired by the GSCM channel model described in Section V-A1, where ρ is assumed to be composed by a weighted superposition of probability density functions $\rho(\theta) = \sum_{q=1}^{Q} f_q(\theta) \alpha_q$. As an example, in the following we assume Gaussian distributions $f_q \sim \mathcal{N}\left(\phi_q, \Delta_q^2\right)$ with ϕ_q uniformly drawn from $[-\pi/3, \pi/3]$ and standard deviation (also called *angular spread*) Δ_q uniformly drawn from $[3°, 8°]$, weights α_q uniformly drawn from $[0, 1]$ and further normalized such that $\sum_{q=1}^{Q} \alpha_q = 1$, and Q uniformly drawn from $\{1, 2, 3, 4, 5\}$. These statistical quantities are introduced to emulate the effect of different scattering patterns corresponding to random user locations. A ULA is assumed for the BS operating at UL/DL carrier wavelengths of $\lambda = 3 \cdot 10^8 / f_c$ with $f_c = 1.8$ Ghz and 1.9 Ghz respectively. The antenna spacing d is set to half UL wavelength. The BS is assumed to have access only to a UL sample covariance matrix computed from $N_s = 1000$ noisy channel estimates as described in Sections III-C, with $\boldsymbol{\Sigma}_z = \sigma_z^2 \mathbf{I}$ and noise power computed from a given $\mathrm{SNR}_{\mathrm{est}} := \frac{1}{N \sigma_z^2}$.

The performance of the two algorithms defined in Sect. III-A and III-B are compared with the algorithms proposed in [11], [12], and [15], referred, respectively, to *splines*-based, *Fourier*-based, and *dictionary*-based. The DL sample



(a) Normalized Euclidean distance (b) Affine invariant distance (c) Principal subspaces distance

Fig. 1: Comparison of different DL covariance estimators vs number of BS antennas N.

covariance, obtained with the same number of samples and SNR as for the UL, is used as a baseline. For fairness, all the sample covariances used in this comparison are corrected with the Toeplitzation procedure outlined in Sect. IV-A2. The accuracy of an estimate $\hat{\mathbf{R}}$ of $\mathbf{R}$ is evaluated in terms of the mean square error MSE $:= \mathbb{E}[e^2(\mathbf{R}, \hat{\mathbf{R}})]$, where $e(\cdot, \cdot)$ is a given error metric. In particular, we consider:

- The normalized Euclidean distance $e(\mathbf{R}, \hat{\mathbf{R}}) := \|\mathbf{R} - \hat{\mathbf{R}}\|_F / \|\mathbf{R}\|_F$.
- [14], [40] The affine invariant distance in the Riemannian space of PSD matrices $e(\mathbf{R}, \hat{\mathbf{R}}) := \| \log(\mathbf{R}^{\frac{1}{2}} \hat{\mathbf{R}}^{-1} \mathbf{R}^{\frac{1}{2}}) \|_F$.
- [40] The Grassmanian distance between the principal subspaces $\mathbf{U}_p$,$\hat{\mathbf{U}}_p$ defined from $\mathbf{R}$, $\hat{\mathbf{R}}$ by considering their eigenvectors corresponding to the minimum number p of largest eigenvalues λ_n satisfying $\sum_{n=1}^{p} \lambda_n / \sum_{n=1}^{N} \lambda_n \geq 95\%$. The metric is then $e(\mathbf{R}, \hat{\mathbf{R}}) := \sqrt{\sum_{n=1}^{p} \gamma_n^2}$, where $\cos(\gamma_n)$ are the eigenvalues of $\mathbf{U}_p^{\mathsf{H}} \hat{\mathbf{U}}_p$. This metric is particularly meaningful for the massive MIMO channel estimation problem, where a reliable signal subspace knowledge plays a crucial role.

The statistical mean is then obtained by Monte-Carlo simulations. For every Monte-Carlo run, a new APS and SNR_{est} level $\in [10, 30]$ (dB) are drawn.

Figure 1 compares the algorithms for different numbers of BS antennas N. The performance of both proposed algorithms approach that of the DL sample covariance estimator as N grows. Moreover, the performance of both algorithms are comparable or better (depending on the metric and on the number of antennas) than the *dictionary*-based method, which in principle can achieve extremely high accuracy given that the dictionary is sufficiently large (here we used only 1000 training samples). However, the proposed algorithms assume no dictionary, thus they do not require any overhead for dictionary acquisition. PLV has the same very low complexity as the *Fourier*-based method, but it achieves a much better accuracy. Compared to PLV, EAPM shows better performance, especially in the low N region, where the prior information about the positivity of the APS becomes important. However, the performance gains are achieved at a cost of a higher complexity.

B. Simulation with realistic channel model

In this section we evaluate the proposed algorithms by simulating a realistic communication scenario between a BS equipped with an 8×4 cross-polarized antenna array and a single antenna UE in a typical macro-cell environment, with system parameters given by Table III. We adopt the multipath channel model described in Section V-C which follows closely the simulation guidelines given by [29, Section 7.3], thus considering propagation in 3D environments and the effects of polarized antennas. We recall that the results shown here are valid for both narrow-band systems and for wide-band OFDM systems. Channel parameters are randomly drawn as follows:

- Cluster powers α_c are drawn uniformly from $[0, 1]$ and further normalized such that $\sum_{c=1}^{N_c} \alpha_c = 1$.
- The XPRs values K_{ic} are drawn from a log-Normal distribution with parameters $(\mu_{\text{XPR}}, \sigma_{\text{XPR}}) = (7, 3)$[dB]. This is identical to the 3GPP model [29, Sect. 7.3, Step 9], with parameters for 3D-UMa, NLOS propagation.
- The angles $\boldsymbol{\theta}_{ic}$, $\boldsymbol{\phi}_{ic}$ are generated from the jointly Gaussian distribution $f_c(\boldsymbol{\theta}, \boldsymbol{\phi}) = f_{\text{BS},c}(\boldsymbol{\theta}) f_{\text{UE},c}(\boldsymbol{\phi})$, where $f_{\text{BS},c} \sim \mathcal{N}(\boldsymbol{\mu}_{\text{BS}}, \boldsymbol{\sigma}_{\text{BS}}^2 \mathbf{I})$ and $f_{\text{UE},c} \sim \mathcal{N}(\boldsymbol{\mu}_{\text{UE}}, \boldsymbol{\sigma}_{\text{UE}}^2 \mathbf{I})$, and where the clusters means and angular spreads

$$\boldsymbol{\mu}_{\text{BS}} := [\mu_{\text{BS},a} \quad \mu_{\text{BS},z}], \quad \boldsymbol{\sigma}_{\text{BS}}^2 := [\sigma_{\text{BS},a}^2 \quad \sigma_{\text{BS},z}^2],$$
$$\boldsymbol{\mu}_{\text{UE}} := [\mu_{\text{UE},a} \quad \mu_{\text{UE},z}], \quad \boldsymbol{\sigma}_{\text{UE}}^2 := [\sigma_{\text{UE},a}^2 \quad \sigma_{\text{UE},z}^2],$$

are drawn as follows: $\mu_{\text{BTS},a}, \mu_{\text{UE},a}$ are uniformly drawn from$[-2\pi/3, 2\pi/3]$, $\mu_{\text{BTS},z}, \mu_{\text{UE},z}$ from $[\pi/4, 3\pi/4]$, $\sigma_{\text{BTS},a}$ from $[3°, 5°]$, $\sigma_{\text{UE},a}$ from $[5°, 10°]$, $\sigma_{\text{BTS},z}$ from $[1°, 3°]$, and $\sigma_{\text{UE},z}$ from $[3°, 5°]$. This choice of parameters is inspired by experimental properties of ρ_V and ρ_H given by [29], e.g. the elevation angular spread is usually narrower than the azimuth one. The adopted procedure has some differences with that suggested by the 3GPP document [29, Sect. 7.3, Step 8]. In particular, here all angles are independently

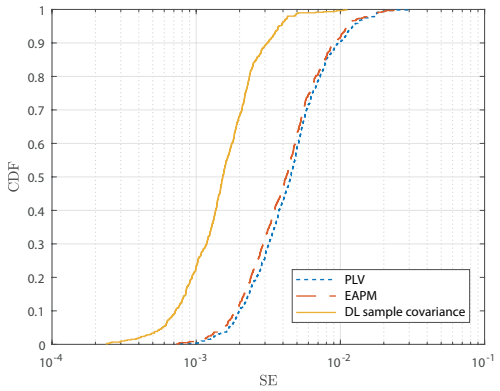
TABLE III: General simulation parameters

Carrier frequency (f_c)	1.8 GHz for UL, 1.9 GHz for DL
System type	Narrow-band or wide-band OFDM
BS	8x4 cross-polarized UPA
	$d = \lambda_u/2$
UE	Single antenna, vertically polarized
Antennas radiation pattern	3GPP [29, Section 7.1], 3D-UMa

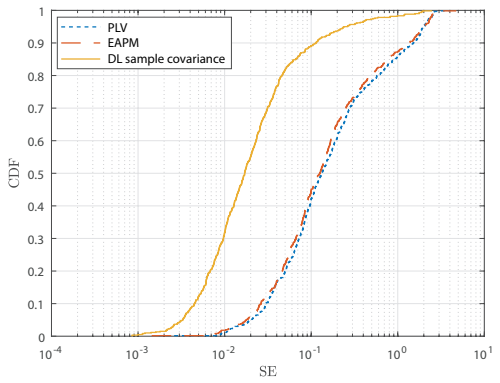
drawn, so we do not use the 3GPP simplification in which only the cluster mean is random, while the subpaths angles are defined from a predefined table.
- To simulate different UE antenna orientation, the UE antenna array response is given by applying a 3D rotation to the antenna radiation pattern as described in 3GPP [29, Sect. 5.1.3], with parameters $\alpha, \beta, \gamma \sim \mathcal{U}\left[0, \frac{\pi}{6}\right]$.

The BS is assumed to have access to the estimated UL covariance matrix $\hat{\mathbf{R}}^{\mathrm{u}}$ obtained from $N_s = 1000$ noisy channel estimates as described in Section III-C with $\mathbf{\Sigma}_z = \sigma_z^2 \mathbf{I}$ and noise power computed from $\mathrm{SNR}_{est} := \mathrm{tr}\{\mathbf{R}^{\mathrm{u}}\}/(N\sigma_z^2) = 10$ [dB], where $N = 2N_V N_H$ denotes the number of BS antennas. The estimate $\hat{\mathbf{R}}^{\mathrm{u}}$ is computed by projecting the sample covariance matrix as described in Section IV-B2. Furthermore, the proposed algorithms are implemented by exploiting the efficient vectorization for UPA described in Section IV-B1. The accuracy of an estimate $\hat{\mathbf{R}}^{\mathrm{d}}$ of $\mathbf{R}^{\mathrm{d}}$ is evaluated in terms of the square error $\mathrm{SE} := e^2(\mathbf{R}^{\mathrm{d}}, \hat{\mathbf{R}}^{\mathrm{d}})$, where $e(\cdot,\cdot)$ is a given error metric defined in Section VI-A. Specifically, we consider the normalized Frobenius norm and the 90% Grassmanian principal subspace distance.

To evaluate the proposed algorithms we use as a baseline an estimate of the DL covariance matrix obtained with the same technique for the estimation of $\hat{\mathbf{R}}^{\mathrm{u}}$, including the correction step that takes into account the structure of the covariance matrix for UPA, but using DL pilots. The results are shown in Figure 2, which shows the empirical cumulative distribution function (CDF) of the SE for the two chosen metrics, obtained by drawing independent realizations of the quantities that are assumed to stay fixed for T_{WSS} (i.e. by drawing a new V-APS and H-APS). The simulation confirms that the proposed algorithms are able to provide an accurate DL estimate by using only UL training, thus it can be used as an effective solution to the DL channel covariance acquisition problem.



(a) Normalized Euclidean distance



(b) Principal subspaces distance

Fig. 2: Empirical CDF of the squared error (SE)

VII. Concluding Remarks and Future Directions

A. Relation with alternative approaches

The proposed algorithms for covariance conversion present several advantages with respect to existing approaches available in the literature, which can be roughly divided into two main categories, i.e. *model-driven* methods [11]–[13] and *data-driven* methods [14]–[16].

Firstly, the PLV method achieves better or comparable performance than existing model-driven approaches. Moreover, it can be implemented via simple matrix-vector multiplication, hence it is at now the best known candidate for low complexity implementations. Furthermore, in contrast to existing model-based approaches, both algorithms do not assume any particular array design, and, in particular, they consider realistic antenna radiation patterns and dual-polarized array models.

Secondly, in contrast to data-driven methods, the proposed algorithms do not require any dataset acquisition and training phase. Although in principle the considered data-driven approaches [14]–[16] can achieve extremely high accuracies while being robust to model mismatches, they are in practice limited by small dataset sizes and sensitive to changes in distribution between the estimands and the dataset. These issues are particularly relevant for all practical applications where scarce resources and the variability of the propagation environment (hence, of the distribution of its parameters) do not allow for timely acquisition of sufficiently large datasets. As a consequence, we envision that the proposed model-based approach may be a more robust solution for such settings. As a final remark, results merging the benefits of the proposed algorithms and data-driven approaches are reported in [41].

B. Exploiting side information of the APS

In many applications, additional side information about the APS is available. Clearly, such information can be useful to obtain better APS estimates. Although this interesting topic is subject of ongoing work (see e.g. [41]), in this section we briefly discuss some related aspects that can be directly applied to the proposed algorithms.

First, note that by separating the real and imaginary part of $\mathbf{R}^f$, and by working in the space of real functions, the proposed algorithms already implicitly take into account the knowledge that the APS is real valued. More interestingly, if such side information can be expressed in terms of closed convex sets, the set of candidate solutions of (6) and (12) can be modified to obtain better APS estimates. For example, [42] shows that PLV and EAPM can be readily extended so that (not necessarily perfect) support information of the APS is taken into account, since support information can be again expressed as a closed subspace.

Another interesting feature of the APS is that it is typically slowly-variant with time. This can be modelled for example by assuming that the APSs in adjacent T_{WSS} windows are strongly correlated. Hence, intuitively, side information coming from the APS estimate $\tilde{\rho}$ obtained in the previous T_{WSS} window could be exploited. A trivial tracking variant of *Algorithm 2* can be readily implemented by replacing the initial condition $\rho^{(0)} = P_V(0)$ by $\rho^{(0)} = P_V(\tilde{\rho})$. Unfortunately, such approach is not beneficial for PLV, since $P_V(P_{\tilde{V}}(0)) = P_V(0)$, where $\tilde{V}$ denotes the linear variety obtained from the previous UL covariance matrix.

C. Potentials of the proposed covariance modelling framework

As already discussed, because of its generality, the proposed covariance modelling framework is particularly suitable for developing solutions which abstract away the particular array design. For example, the extension of the analytical performance analysis for covariance conversion given by [13] to arbitrary array design is left by the authors as an open problem [13, Sect. VIII]. In contrast, as shown in [42], related performance bounds can be directly derived by using the proposed modelling framework, hence readily carrying over to arbitrary array designs.

More generally, we believe that the interest of the proposed framework goes beyond the covariance conversion problem studied in here. Owing to its capability of formally pinpointing the effects of the propagation environment and of the antenna array, the proposed framework can be also applied to many other problems involving statistical description of propagation in MIMO channels. Examples of further successful applications include multi-user covariance estimation and *favourable-propagation* analysis [43].

REFERENCES

[1] T. L. Marzetta, E. G. Larsson, H. Yang, and H. Q. Ngo, *Fundamentals of massive MIMO*, Cambridge University Press, 2016.

[2] E. Björnson, J. Hoydis, and L. Sanguinetti, "Massive MIMO networks: Spectral, energy, and hardware efficiency," *Foundations and Trends® in Signal Processing*, vol. 11, no. 3-4, pp. 154–655, 2017.

[3] E. Björnson, J. Hoydis, and L. Sanguinetti, "Massive MIMO has unlimited capacity," *IEEE Trans. Wireless Commun.*, vol. 17, no. 1, pp. 574–590, Jan. 2018.

[4] E. Björnson, E. G. Larsson, and T. L. Marzetta, "Massive MIMO: Ten myths and one critical question," *IEEE Commun. Mag.*, vol. 54, no. 2, pp. 114–123, Feb. 2016.

[5] H. Yin, D. Gesbert, and L. Cottatellucci, "Dealing with interference in distributed large-scale MIMO systems: A statistical approach," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 942–953, May 2014.

[6] A. Adhikary, J. Nam, J.-Y. Ahn, and G. Caire, "Joint spatial division and multiplexing–The large-scale array regime," *IEEE Trans. Info. Theory*, vol. 59, no. 10, pp. 6441–6463, Oct. 2013.

[7] M. B. Khalilsarai, S. Haghighatshoar, X. Yi, and G. Caire, "FDD massive MIMO via UL/DL channel covariance extrapolation and active channel sparsification," *IEEE Trans. Wireless Commun.*, vol. 18, no. 1, pp. 121–135, Jan. 2019.

[8] A. Decurninge, L. G. Ordóñez, P. Ferrand, H. Gaoning, L. Bojie, Z. Wei, and M. Guillaud, "CSI-based outdoor localization for massive MIMO: Experiments with a learning approach," *15th Int. Symp. Wireless Commun. Systems*, pp. 1–6, 2018.

[9] X. Gao, O. Edfors, F. Rusek, and F. Tufvesson, "Massive MIMO performance evaluation based on measured propagation data," *IEEE Trans. Wireless Commun.*, vol. 14, no. 7, pp. 3899–3911, July 2015.

[10] S. Haghighatshoar and G. Caire, "Massive MIMO channel subspace estimation from low-dimensional projections," *IEEE Trans. Signal Process.*, vol. 65, no. 2, pp. 303–318, Feb. 2017.

[11] Y.-C. Liang and F. P. S. Chin, "Downlink channel covariance matrix (DCCM) estimation and its applications in wireless DS-CDMA systems," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 2, pp. 222–232, Feb. 2001.

[12] M. Jordan, A. Dimofte, X. Gong, and G. Ascheid, "Conversion from uplink to downlink spatio-temporal correlation with cubic splines," *Proc. IEEE 69th Veh. Technol. Conf.*, pp. 1–5, 2009.

[13] S. Haghighatshoar, M. B. Khalilsarai, and G. Caire, "Multi-band covariance interpolation with applications in massive MIMO," 2018, [Online]. Available: https://arxiv.org/abs/1801.03714.

[14] A. Decurninge, M. Guillaud, and D. T. M. Slock, "Channel covariance estimation in massive MIMO frequency division duplex systems," in *Proc. IEEE Global Conf. Communications*, 2015, pp. 1–6.

[15] A. Decurninge, M. Guillaud, and D. T. M. Slock, "Riemannian coding for covariance interpolation in massive MIMO frequency division duplex systems," *Proc. IEEE Sens. Array and Multichannel Sig. Process. Workshop*, pp. 1–5, 2016.

[16] Y. Song, M. B. Khalilsarai, S. Haghighatshoar, and G. Caire, "Machine learning for geometrically-consistent angular spread function estimation in massive MIMO," 2019, [Online]. Available: https://arXiv.org/abs/1910.13795.

[17] L. Miretti, R. L. G. Cavalcante, and S. Stańczak, "FDD Massive MIMO channel spatial covariance conversion using projection methods," *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, pp. 3609–3613, 2018.

[18] L. Miretti, R. L. G. Cavalcante, and S. Stańczak, "Downlink channel spatial covariance estimation in realistic FDD massive MIMO systems," *Proc. IEEE Global Conf. Signal Inf. Processing*, pp. 161–165, 2018.

[19] H. Stark and Y. Yang, *Vector space projections: a numerical approach to signal and image processing, neural nets, and optics*, John Wiley & Sons, Inc., 1998.

[20] D. G. Luenberger, *Optimization by vector space methods*, John Wiley & Sons, Inc., 1998.

[21] H. H. Bauschke and P. L. Combettes, *Convex analysis and monotone operator theory in Hilbert spaces*, vol. 408, Springer, 2011.

[22] R. A. Kennedy and P. Sadeghi, *Hilbert space methods in signal processing*, Cambridge University Press, 2013.

[23] D. Tse and P. Viswanath, *Fundamentals of wireless communication*, Cambridge University Press, 2005.

[24] A. F. Molish, *Wireless communications*, John Wiley & Sons, Inc., 2010.

[25] T. Trump and B. Ottersten, "Estimation of nominal direction of arrival and angular spread using an array of sensors," *Signal Processing*, vol. 50, no. 1-2, pp. 57–69, 1996.

[26] O. Besson, L. L. Scharf, and F. Vincent, "Matched direction detectors and estimators for array processing with subspace steering vector uncertainties," *IEEE Transactions on Signal Processing*, vol. 53, no. 12, pp. 4453–4463, Dec. 2005.

[27] A. Pezeshki, B. D. Van Veen, L. L. Scharf, H. Cox, and M. Lundberg Nordenvaad, "Eigenvalue beamforming using a multirank MVDR beamformer and subspace selection," *IEEE Transactions on Signal Processing*, vol. 56, no. 5, pp. 1954–1967, May 2008.

[28] K. Hugl, K. Kalliola, and J. Laurila, "Spatial reciprocity of uplink and downlink radio channels in FDD systems," *Proc. COST 273 Tech. Document TD(02)*, vol. 66, 2002.

[29] 3GPP, "Study on 3D channel model for LTE (release 12)," Tech. Rep. TR 36.873 V12.6.0, 3rd Generation Partnership Project (3GPP), 2017.

[30] P. Stoica and R. L. Moses, *Spectral analysis of signals*, Prentice Hall, 2005.

[31] P. L. Combettes, "The foundations of set theoretic estimation," *Proc. IEEE*, vol. 81, no. 2, pp. 182–208, Feb. 1993.

[32] I. Yamada and N. Ogura, "Adaptive projected subgradient method for asymptotic minimization of sequence of nonnegative convex functions," *Numerical functional analysis and optimization*, vol. 25, pp. 593–617, 2005.

[33] S. Theodoridis, K. Slavakis, and I. Yamada, "Adaptive learning in a world of projections," *IEEE Trans. Signal Process.*, vol. 28, no. 1, pp. 97–123, Jan. 2011.

[34] H. H. Bauschke, P. L. Combettes, and S. G. Kruk, "Extrapolation algorithm for affine-convex feasibility problems," *Numerical Algorithms*, vol. 41, no. 3, pp. 239–274, 2006.

[35] H. H. Bauschke, M. N. Dao, D. Noll, and H. M. Phan, "On Slater's condition and finite convergence of the Douglas-Rachford algorithm for solving convex feasibility problems in euclidean spaces," *Journal of Global Optimization*, vol. 65, no. 2, pp. 329–349, 2016.

[36] D. Neumann, M. Joham, L. Weiland, and W. Utschick, "Low-complexity computation of LMMSE channel estimates in massive MIMO," *Proc. 19th Int. ITG Workshop Smart Antennas*, pp. 1–6, 2015.

[37] K. M. Grigoriadis, A. E. Frazho, and R. E. Skelton, "Application of alternating convex projection methods for computation of positive toeplitz matrices," *IEEE Trans. Signal Process.*, vol. 42, no. 7, pp. 1873–1875, July 1994.
[38] M. B. Khalilsarai, T. Yang, S. Haghighatshoar, X. Yi, and G. Caire, "Active channel sparsification and precoding for dual-polarized FDD massive MIMO," *24th International ITG Workshop on Smart Antennas*, pp. 1–6, 2020.
[39] 3GPP, "Spatial channel model for multiple input multiple output MIMO simulations (release 14)," Tech. Rep. TR 25.996 V14.0.0, 3rd Generation Partnership Project (3GPP), 2017.
[40] S. T. Smith, "Covariance, subspace, and intrinsic Cramér-Rao bounds," *IEEE Trans. Signal Process.*, vol. 53, no. 5, pp. 1610–1630, May 2005.
[41] R. L. G. Cavalcante and Slawomir Stańczak, "Hybrid data and model driven algorithms for angular power spectrum estimation," *Proc. IEEE Global Conf. Communications*, 2020.
[42] R. L. G. Cavalcante, L. Miretti, and S. Stańczak, "Error bounds for FDD massive MIMO channel covariance conversion with set-theoretic methods," *Proc. IEEE Global Conf. Communications*, 2018.
[43] R. L. G. Cavalcante and S. Stańczak, "Channel covariance estimation in multiuser massive MIMO systems with an approach based on infinite dimensional Hilbert spaces," *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, pp. 5180–5184, 2020.

Lorenzo Miretti (S'18) received the B.Sc. degree in Telecommunication Engineering from Politecnico di Torino, Turin, Italy, in 2015 and the M.Sc. degree in Communications and Computer Networks Engineering from Politecnico di Torino and Télécom ParisTech, in 2018, both cum laude. He is currently working toward the Ph.D. degree with the Department of Communicaiton Systems, EURECOM, Sophia Antipolis, France. He studies the physical layer of wireless networks, signal processing, and multiuser information theory.

Renato Luís Garrido Cavalcante (M'09) received the Electronics Engineering degree from the Instituto Tecnológico de Aeronáutica, São José dos Campos, Brazil, in 2002, and the M.E. and Ph.D. degrees in communications and integrated systems from the Tokyo Institute of Technology, Tokyo, Japan, in 2006 and 2008, respectively. He is currently a Research Fellow with the Fraunhofer Institute for Telecommunications, Heinrich Hertz Institute, Berlin, Germany, and a lecturer with the Technical University of Berlin. Previously, he held appointments as a Research Fellow with the University of Southampton, Southampton, U.K., and as a Research Associate with the University of Edinburgh, Edinburgh, U.K. His current interests include signal processing for distributed systems, multiagent systems, convex analysis, machine learning, and wireless communications. Dr. Cavalcante received the Excellent Paper Award from the Institute of Electronics, Information and Communication Engineers in 2006 and the IEEE Signal Processing Society (Japan Chapter) Student Paper Award in 2008. He also co-authored a study that received a Best Student Paper Award at the 13th IEEE International Workshop on Signal Processing Advances in Wireless Communications in 2012. From April 2003 to April 2008, he was a recipient of the Japanese Government (MEXT) Scholarship.

Slawomir Stańczak (M'04-SM'11) studied electrical engineering with specialization in control theory at the Wroclaw University of Technology and at the Technical University of Berlin (TU Berlin). He received the Dipl.-Ing. degree in 1998 and the Dr.-Ing. degree (summa cum laude) in electrical engineering in 2003, both from TU Berlin; the Habilitation degree (venialegendi) followed in 2006. Since 2015, he has been a Full Professor for network information theory with TU Berlin and the head of the Wireless Communications and Networks department at Fraunhofer Institute for Telecommunications, Heinrich Hertz Institute. Prof. Stańczak is a co-author of two books and more than 200 peer-reviewed journal articles and conference papers in the area of information theory, wireless communications, signal processing and machine learning. He was an Associate Editor of the *IEEE Transactions on Signal Processing* between 2012 and 2015. Since February 2018 Prof. Stańczak has been the chairman of the ITU-T focus group on machine learning for future networks including 5G.