This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSAC.2020.3000889, IEEE Journal on Selected Areas in Communications

Estimation of Wideband Dynamic mmWave and THz Channels for 5G Systems and Beyond

Alessandro Brighente, *Student Member, IEEE*, Mattia Cerutti, Monica Nicoli, *Member, IEEE*, Stefano Tomasin, *Senior Member, IEEE*, and Umberto Spagnolini, *Senior Member, IEEE*

***Abstract*—Millimeter wave (mmWave) wideband channels in a multiple-input multiple-output (MIMO) transmission are described by a sparse set of impulse responses in the angle-delay, or space-time (ST), domain. These characteristics will be even more prominent in the THz band used in future systems. We consider two approaches for channel estimation: compressed-sensing (CS), exploiting the sparsity in the angular/delay domain, and low-rank (LR), exploiting the algebraic structure of channel matrix. Both approaches share several commonalities, and this paper provides for the first time i) a comparison of the two approaches, and ii) new versions of CS and LR methods that significantly improve performance in terms of mean squared error (MSE), computational complexity, and latency. We derive the asymptotic MSE bound for *any estimator* of the ST-MIMO multipath channels with invariant angles/delays and time-varying fading, with unknown angle/delay diversity order: the bound also accounts for the degradation introduced by sub-optimal separable channel models. We will show that in the considered scenarios both CS and LR approaches attain the bound. Our performance assessment over ideal and 3^{rd} generation partnership project (3GPP) channel models, suitable for the fifth-generation (5G) and beyond of cellular networks, shows the trade-off obtained by the methods over various metrics: i) CS methods are converging faster than the LR methods, both attaining the asymptotic MSE bound; ii) the CS methods depend on the array manifold, while LR methods are independent of the array calibration; iii) CS solutions are more complex than LR solutions.**

***Index Terms*—Channel estimation, Compressed sensing, Low-rank approximation, mmWave, Multiple-input-multiple-output (MIMO).**

I. Introduction

MILLIMETER waves above 10 GHz communication are currently considered for the fifth generation (5G) of cellular networks [1]–[3], and the higher-frequency spectrum in the Terahertz (THz) band is raising the attention of the scientific community for future networks (see [4] and [5]). The high attenuation incurred at these frequencies makes mandatory the use of a large number of antennas at least on one end of the link, resulting in a multiple-input multiple-output (MIMO) system. Multipath MIMO channels are characterized by a sparse set of responses in the angles (of departure and/or arrival)-delays, or space-time (ST), domain. While this has been widely known for mmWave, it is also expected to occur in THz transmissions, as confirmed by preliminary measurement campaigns [6] that show a further reduced number of effective paths available for communication.

Training-based channel estimation is the preferred method in radio access systems to track the temporal channel variations. Channel estimation in millimeter wave (mmWave) MIMO involves several aspects that depend on the system configurations (see [7] for an overview), although all methods exploit the channel sparsity, either by constraining the optimization problem or by recognizing the algebraic properties of the channel matrix. More specifically, compressed-sensing (CS) techniques estimate the MIMO channel responses under the sparsity constraints of the ST domain, with the main drawback of a high sensitivity to antenna calibrations, e.g., irregular sampling of ST and phase errors/drift of radio frequency (RF) converters [8]. Off-grid methods have also been considered, using the implicit Dirichlet kernel structure in the angular domain [9] and continuous angle-delay parameters [10]. Moreover, in [9] the spatial-wideband effect is considered, i.e., the fact that the delay by which a transmitted signal reaches different receive antennas may span multiple data symbols. In low-rank (LR) methods sparsity is converted into a LR structure of the MIMO channel matrices, where powerful algebraic methods can be applied, regardless of the array configurations [11].

Although the complexity of transceivers is reduced by hybrid analog/digital structures, specific solutions for mmWave channel estimation have been proposed for a static environment in [9] and [12]–[15]. However, mmWave and THz MIMO channel estimation in a dynamic environment has not been considered in the literature, and this key aspect is addressed herein. Considering a radio access context with moving terminals connected to a fixed array unit, one can take for granted that the ST features of the mmWave MIMO channel remain constant for several (say L) temporal intervals organized in time-slots. This invariance depends on the temporal and spatial resolution of the MIMO system. To gain insight, for a speed of 50 km/h in an urban scenario, it can be easily proved that the angles are invariant for approximately 240 ms when the movement is tangential from the array at a distance of 200 m (for a beamwidth of 6 deg, or approximately 120 antennas in the array), and delays are invariant for approximately 108 ms for a radial movement (bandwidth is 200 MHz). Overall, the ST domain is invariant on, at least, 108 ms, that means $L \leq 108$ training slots, for a frame length of 1 ms (see [16] and [17]), but it scales accordingly for a smaller pilot repetition frequency, as reasonable for mmWave/THz transmissions in

M. Cerutti, M. Nicoli, and U. Spagnolini are with the Politecnico di Milano, Milano, Italy, email: first.lastname@polimi.it. A. Brighente and S. Tomasin are with the Department of Information Engineering, University of Padova, Padova, Italy, email: first.lastname@dei.unipd.it. Stefano Tomain is also with the Consorzio Nazionale Interuniversitario per le Telecomunicazioni (CNIT), Padova Research Unit. Part of this work was supported by MIUR (Italian Minister for Education) under the initiative *Departments of Excellence* (Law 232/2016).

a dynamic environment. This paper proposes to exploit this invariance in mmWave/THz wideband systems specialized for MIMO-OFDM communications. The CS and LR methods are tailored for dynamic environments constraining the invariance of the ST domain of the MIMO channel. More specifically, in the proposed CS method we find the most recurrent positions of the channel paths in the angular domain over the L training blocks, thus exploiting the slowly time-varying features. Similarly, in the proposed LR methods the spatial (or angular) and temporal signatures of the channel span the spatial and temporal subspaces that are invariant over the L slots [18]–[20]. The proposed solutions can be applied to existing mmWave systems; however, they will become even more relevant in the THz context, where more antennas are needed and more challenging channels are experienced [6].

The contributions of this paper are:

- For dynamic mmWave MIMO channels, based on time-invariant features of the ST domain (angles of arrivals and departures, and/or delays), we derive the lower bound to the mean squared error (MSE) of the channel estimate for asymptotic conditions ($L \to \infty$), regardless of the fading fluctuations, by extending the approach in [18] and [21]. With respect to these former works, the bound here is derived for ST-MIMO multipath channels with unknown angle/delay diversity orders, accounting for the MSE increase introduced by diversity order estimate errors. The fundamental limits are valid for any channel estimator and are specialized also for suboptimal methods assuming separable transmit-receive MIMO correlation and ST channel structures.
- With respect to the preliminary works [11], [18], and [20], here the LR channel estimator is tailored for mmWave/THz MIMO systems with realistic channel modeling. A major issue of the optimal LR approach [20] in MIMO scenarios is its slow convergence (typically $L > 1000$ slots), which is hardly compatible with the constraint of about 100 slots imposed by dynamic mmWave/THz systems. To accelerate the convergence here we propose a set of novel joint ST LR methods that are sub-optimal but fast-converging, and with far lower complexity.
- Starting from [22], where a narrowband communication was assumed, we propose three compressed-sensing (CS) methods targeting broadband communications. The first solution is based on the accelerated gradient descent with adaptive restart (AGDAR) algorithm, which has been shown to be fast and effective. Moreover, two further improved CS algorithms are introduced, namely the selective AGDAR (S-AGDAR) and the adaptive AGDAR (A-AGDAR): S-AGDAR compensates for the use of norm-1 (instead of norm-0) minimization by removing small paths and redistributing their power to the largest paths; A-AGDAR refines the estimation of the channel path positions in the angular domain, by processing multiple channel estimates, a solution particularly useful when operating at a low signal to noise ratio (SNR). Note that the broadband version of AGDAR and S-AGDAR can also be applied to mmWave transmissions subject to the spatial-wideband effect.

The rest of the paper is organized as follows. Section II introduces the mmWave/THz channel model. The novel LR and CS channel estimation techniques are introduced in Sections III and IV, respectively. In Section III we also derive MSE bounds on the channel estimation. Numerical results on both simplified and 3^{rd} generation partnership project (3GPP)-compliant channel models are presented in Section V, before conclusions are outlined in Section VI.

II. System Model

We consider[1] a mmWave wireless MIMO-orthogonal frequency division multiplexing (OFDM) system with $N_{\rm T}$ transmitting and $N_{\rm R}$ receiving antennas. Let $K_{\rm tot}$ be the total number of OFDM subcarriers, K the number of pilot subcarriers allocated for channel estimation and T the sampling time of the OFDM transmitter output. $\mathbf{X}_\ell^{(n_{\rm T})} \in \mathbb{C}^{K\times 1}$ is the vector containing the pilot symbols transmitted over the K subcarriers from antenna $n_{\rm T} = 1,\ldots,N_{\rm T}$, within the OFDM block $\ell = 1,\ldots,L$. Here, index ℓ runs over the OFDM blocks containing pilots, i.e., the training blocks. The cyclic prefix (CP) comprises $W-1$ samples, where W denotes the maximum temporal support of the channel response, with $K \geq N_{\rm T}W$.

At the receiver side, after CP removal and the computation of the $K_{\rm tot}$-point discrete Fourier transform (DFT), the vector signal $\mathbf{Y}_\ell^{(n_{\rm R})} \in \mathbb{C}^{K\times 1}$ received by the $n_{\rm R}$-th antenna during the ℓ-th block can be written as

$$\mathbf{Y}_\ell^{(n_{\rm R})} = \sum_{n_{\rm T}=1}^{N_{\rm T}} \mathrm{diag}\left(\mathbf{X}_\ell^{(n_{\rm T})}\right)\mathbf{H}_\ell^{(n_{\rm R},n_{\rm T})} + \mathbf{N}_\ell^{(n_{\rm R})}, \quad (1)$$

where $\mathbf{H}_\ell^{(n_{\rm R},n_{\rm T})} = \mathbf{F}\mathbf{h}_\ell^{(n_{\rm R},n_{\rm T})} \in \mathbb{C}^{K\times 1}$ is the channel for the link $(n_{\rm R},n_{\rm T})$ in the frequency domain, obtained as the DFT of the corresponding $W\times 1$ channel impulse response $\mathbf{h}_\ell^{(n_{\rm R},n_{\rm T})} = [h_\ell^{(n_{\rm R},n_{\rm T})}(1)\cdots h_\ell^{(n_{\rm R},n_{\rm T})}(W)]$. The DFT matrix has entries $[\mathbf{F}]_{k,w} = \frac{1}{\sqrt{K_{\rm tot}}}\exp(-j2\pi f_k w/K_{\rm tot})$, with f_k denoting the frequency index of the k-th pilot subcarrier, $k = 0,\ldots,K-1$, and $w = 0,\ldots,W-1$. The w-th tap of the ST MIMO channel for the ℓ-th training block is denoted as $\mathbf{h}_\ell(w) \in \mathbb{C}^{N_{\rm R}\times N_{\rm T}}$; note that $[\mathbf{h}_\ell(w)]_{n_{\rm R},n_{\rm T}} = h_\ell^{(n_{\rm R},n_{\rm T})}(w)$. The additive Gaussian noise-plus-interference signal $\mathbf{N}_\ell^{(n_{\rm R})} \in \mathbb{C}^{K\times 1}$ is assumed to be uncorrelated over the frequencies (i.e., $\mathbb{E}\left[\mathbf{N}_\ell^{(n_{\rm R})}\mathbf{N}_\ell^{(n_{\rm R})\rm H}\right] = K\sigma_n^2\mathbf{I}_K$) and spatially correlated, due to the geometrical distribution of the interferers. By stacking column-wise the $N_{\rm R}$ noise vectors into matrix $\mathbf{N}_\ell = \left[\mathbf{N}_\ell^{(1)}\cdots\mathbf{N}_\ell^{(N_{\rm R})}\right]$, the noise-plus-interference spatial

[1]Notation: scalar variables are in italic, while matrices and vectors are in boldface. $\mathbf{X}^{\rm T}$ and $\mathbf{X}^{\rm H}$ denote the transpose and Hermitian operators on matrix $\mathbf{X}$, respectively. $\mathrm{diag}(\mathbf{x})$ denotes a diagonal matrix with vector $\mathbf{x}$ on its diagonal. $\mathrm{vec}(\mathbf{X})$ is a column vector stacking the columns of $\mathbf{X}$. $\mathrm{vec}^{-1}(\mathbf{X})$ denotes the inverse operation of vec. $\mathbb{E}[x]$ denotes the expectation of random variable x. $\odot$ denotes the Hadamard product, and $\otimes$ denotes the Kronecker product. $\mathrm{rank}(\mathbf{x})$ denotes the rank of matrix $\mathbf{x}$. $\mathrm{tr}(\mathbf{X})$ denotes the trace of matrix $\mathbf{X}$, while $||\mathbf{X}||^2 = \mathrm{tr}[\mathbf{X}\mathbf{X}^{\rm H}]$ is the Frobenius norm of matrix $\mathbf{X}$. $\mathrm{cov}(\mathbf{x})$ is the covariance matrix of random vector $\mathbf{x}$.

covariance is $\mathbf{Q} = \frac{1}{K}\mathbb{E}\left[\mathbf{N}_\ell^{\mathrm{H}}\mathbf{N}_\ell\right]$, which is to be estimated as the arrangement of the interferers is unknown in general.

A. MmWave/THz Space-Time Channel

The mmWave/THz channel is modeled as in [23] and [24], by the sum of P paths,

$$\mathbf{h}_\ell(w) = \sum_{p=1}^{P} \alpha_{p,\ell}\mathbf{A}(\boldsymbol{\theta}_p)\, g((w-1)T - \tau_p), \qquad (2)$$

each path p having fading amplitude $\alpha_{p,\ell}$ and delay τ_p, and $g(\cdot)$ being the impulse response of the cascade between the transmitter and receiver filters. The matrix

$$\mathbf{A}(\boldsymbol{\theta}_p) = \mathbf{a}^{\mathrm{RX}}(\boldsymbol{\theta}_p^{\mathrm{RX}})\mathbf{a}^{\mathrm{TX}}(\boldsymbol{\theta}_p^{\mathrm{TX}})^{\mathrm{T}} \qquad (3)$$

accounts for the antenna array responses at the receiver, $\mathbf{a}^{\mathrm{RX}}(\boldsymbol{\theta}_p^{\mathrm{RX}})$, and at the transmitter, $\mathbf{a}^{\mathrm{TX}}(\boldsymbol{\theta}_p^{\mathrm{TX}})$. These responses depend on the azimuth (ϕ) and elevation (ψ) angles, at arrival $\boldsymbol{\theta}_p^{\mathrm{RX}} = \left[\phi_p^{\mathrm{RX}}, \psi_p^{\mathrm{RX}}\right]$ and at departure, $\boldsymbol{\theta}_p^{\mathrm{TX}} = \left[\phi_p^{\mathrm{TX}}, \psi_p^{\mathrm{TX}}\right]$. Fast varying fading amplitudes $\alpha_{p,\ell}$ are block-dependent, while delays τ_p and angles $\boldsymbol{\theta}_p = \left[\boldsymbol{\theta}_p^{\mathrm{TX}}, \boldsymbol{\theta}_p^{\mathrm{RX}}\right]$ are typically invariant over a large number of blocks and are thus assumed to be constant within L blocks. When comparing mmWave and THz channels, the THz channels are characterized by even fewer, more attenuated paths than mmWave channels [6].

The overall ST MIMO channel at block ℓ is arranged for analytical convenience into the $N_{\mathrm{T}}N_{\mathrm{R}} \times W$ matrix

$$\mathcal{H}_\ell = [\mathrm{vec}(\mathbf{h}_\ell(1)) \cdots \mathrm{vec}(\mathbf{h}_\ell(W))] = \mathcal{A}(\boldsymbol{\theta})\mathbf{D}_\ell\mathbf{G}(\tau)^{\mathrm{T}}, \quad (4)$$

where the $N_{\mathrm{R}}N_{\mathrm{T}} \times P$ complex matrix $\mathcal{A}(\boldsymbol{\theta}) = [\mathbf{a}(\boldsymbol{\theta}_1)\cdots\mathbf{a}(\boldsymbol{\theta}_P)]$ and the $W \times P$ real matrix $\mathbf{G}(\tau) = [\mathbf{g}(\tau_1)\cdots\mathbf{g}(\tau_P)]$ collect all the static ST channel components associated to the P paths, with $\mathbf{a}(\boldsymbol{\theta}_p) = \mathrm{vec}(\mathbf{A}(\boldsymbol{\theta}_p))$ and $\mathbf{g}(\tau_p) = [g(-\tau_p)\cdots g((W-1)T - \tau_p)]^{\mathrm{T}}$. The fading amplitudes $\mathbf{D}_\ell = \mathrm{diag}([\alpha_{1,\ell}\cdots\alpha_{P,\ell}]) \in \mathbb{C}^{P\times P}$ are assumed to follow the wide-sense stationary uncorrelated scattering (WSSUS) model, and to be uncorrelated over blocks:

$$\mathbb{E}\left[\mathbf{D}_{\ell+m}\mathbf{D}_\ell^{\mathrm{H}}\right] = \boldsymbol{\Omega}\delta(m), \qquad (5)$$

where $\boldsymbol{\Omega} = \mathrm{diag}([\Omega_1\cdots\Omega_P])$ and $\Omega_p > 0$ is the average power of the p-th path, normalized to have $\sum_{p=1}^{P}\Omega_p = 1$.

B. Pre-processing for ST Channel Estimation

Model (1) is conveniently rearranged to isolate the temporal channel responses as:

$$\mathbf{Y}_\ell^{(n_{\mathrm{R}})} = \mathbf{B}_\ell\mathbf{h}_\ell^{(n_{\mathrm{R}})} + \mathbf{N}_\ell^{(n_{\mathrm{R}})}, \qquad (6)$$

where the $K \times N_{\mathrm{T}}W$ matrix $\mathbf{B}_\ell = \left[\mathrm{diag}\left(\mathbf{X}_\ell^{(1)}\right)\mathbf{F}\cdots\mathrm{diag}\left(\mathbf{X}_\ell^{(N_{\mathrm{T}})}\right)\mathbf{F}\right]$ collects the N_{T} training matrices, and the $N_{\mathrm{T}}W \times 1$ vector $\mathbf{h}_\ell^{(n_{\mathrm{R}})} = \left[\mathbf{h}_\ell^{(n_{\mathrm{R}},1)\mathrm{T}}\cdots\mathbf{h}_\ell^{(n_{\mathrm{R}},N_{\mathrm{T}})\mathrm{T}}\right]^{\mathrm{T}}$ the N_{T} channel impulse responses. The unconstrained least squares (LS) estimate of $\mathbf{h}_\ell^{(n_{\mathrm{R}})}$ is obtained from (6) as:

$$\hat{\mathbf{h}}_\ell^{(n_{\mathrm{R}})} = \mathbf{B}_\ell^{\dagger}\mathbf{Y}_\ell^{(n_{\mathrm{R}})} = \mathbf{h}_\ell^{(n_{\mathrm{R}})} + \mathbf{B}_\ell^{\dagger}\mathbf{N}_\ell^{(n_{\mathrm{R}})}, \qquad (7)$$

TABLE I
LS CHANNEL ARRANGEMENTS

Notation	Meaning	Dimensions
$\hat{\mathbf{h}}_\ell$	Space Time Space	$N_{\mathrm{R}}N_{\mathrm{T}}W \times 1$
$\hat{\mathbf{h}}_\ell(w)$	Separate-Space Time	$N_{\mathrm{R}} \times N_{\mathrm{T}}$
$\hat{\mathcal{H}}_\ell$	Doubly-Space Time	$N_{\mathrm{T}}N_{\mathrm{R}} \times W$
$\hat{\boldsymbol{h}}_\ell$	Joint Space Time	$WN_{\mathrm{T}}N_{\mathrm{R}} \times 1$

with $\mathbf{B}_\ell^{\dagger} = \left(\mathbf{B}_\ell^{\mathrm{H}}\mathbf{B}_\ell\right)^{-1}\mathbf{B}_\ell^{\mathrm{H}}$. The LS estimate is optimal in a Gaussian setting, it is unbiased, and its variance is minimized by adopting optimally designed pilot sequences $\mathbf{X}_\ell^{(n_{\mathrm{T}})}$ with correlation matrix $\mathbf{R}_{\mathbf{BB}} = \mathbf{B}_\ell^{\mathrm{H}}\mathbf{B}_\ell = K\sigma_x^2\mathbf{I}_{N_TW}$ (see [25] and [26]). The noise spatial covariance is estimated as

$$\hat{\mathbf{Q}} = \frac{1}{KL}\sum_{\ell=1}^{L}\hat{\mathbf{N}}_\ell^{\mathrm{H}}\hat{\mathbf{N}}_\ell, \qquad (8)$$

from the noise-plus-interference estimates aggregated over the antennas, $\hat{\mathbf{N}}_\ell = \left[\hat{\mathbf{N}}_\ell^{(1)}\cdots\hat{\mathbf{N}}_\ell^{(N_{\mathrm{R}})}\right]$, where $\hat{\mathbf{N}}_\ell^{(n_{\mathrm{R}})} = \mathbf{Y}_\ell^{(n_{\mathrm{R}})} - \mathbf{B}_\ell\hat{\mathbf{h}}_\ell^{(n_{\mathrm{R}})}$.

Throughout the paper, we will use different rearrangements of the LS estimate as summarized in Table I. Denoting by $\hat{h}_\ell^{(n_{\mathrm{R}},n_{\mathrm{T}})}(w) = \left[\hat{\mathbf{h}}_\ell^{(n_{\mathrm{R}})}\right]_{(n_{\mathrm{T}}-1)W+w}$ the w-th tap for link $(n_{\mathrm{R}},n_{\mathrm{T}})$, we further define the $N_{\mathrm{R}} \times N_{\mathrm{T}}$ space matrix $\hat{\mathbf{h}}_\ell(w)$ still at w-th tap, and the $N_{\mathrm{T}}N_{\mathrm{R}} \times W$ ST matrix $\hat{\mathcal{H}}_\ell$, by re-ordering the LS estimates according to the channel arrangements in (2) and (4). The resulting equivalences are

$$\hat{h}_\ell^{(n_{\mathrm{R}},n_{\mathrm{T}})}(w) = \left[\hat{\mathbf{h}}_\ell(w)\right]_{n_{\mathrm{R}},n_{\mathrm{T}}} = \left[\hat{\mathcal{H}}_\ell\right]_{(n_{\mathrm{T}}-1)N_{\mathrm{R}}+n_{\mathrm{R}},w}. \quad (9)$$

To conclude, we define the LS channel vector $\hat{\mathbf{h}}_\ell = \left[\hat{\mathbf{h}}_\ell^{(1)\mathrm{T}}\cdots\hat{\mathbf{h}}_\ell^{(N_{\mathrm{R}})\mathrm{T}}\right]^{\mathrm{T}} \in \mathbb{C}^{N_{\mathrm{R}}N_{\mathrm{T}}W\times 1}$, with covariance $\mathbf{C} = \mathrm{cov}\left[\hat{\mathbf{h}}_\ell\right] = \mathbf{Q}\otimes\mathbf{R}_{\mathrm{BB}}^{-1}$. This is conveniently rearranged into the vectorization $\hat{\boldsymbol{h}}_\ell = \mathrm{vec}\left(\hat{\mathcal{H}}_\ell\right) \in \mathbb{C}^{WN_{\mathrm{T}}N_{\mathrm{R}}\times 1}$, whose covariance $\boldsymbol{C} = \mathrm{cov}\left[\hat{\boldsymbol{h}}_\ell\right]$ can be obtained by reordering the samples in $\mathbf{C}$, as detailed in the Appendix.

III. LOW-RANK CHANNEL ESTIMATION

The mmWave/THz MIMO wireless channel is sparse and the conventional LS estimation is extremely noisy in massive ST settings. A common strategy to reduce errors is to decouple the spatial and temporal domains by first aligning the transmit and receive beamformers, and then performing an LS estimation of the temporal channel components. Here we propose a different approach that exploits the invariance of the directions of arrival/departure (DOA/DOD), as well as the delays and the average powers of the propagation paths. This enables the identification of the LR algebraic structure of the channel: rather than estimating DOD/DOA and delays based on a parametric model, we describe the multipath components in terms of ST-invariant subspaces and estimate the channel by filtering the LS estimate through a set of ST projections. In this way, we avoid a joint angle and delay estimation that

TABLE II
LR CHANNEL ALGORITHMS

Method	Covariance $\mathbf{R}$	Projector $\mathbf{\Pi}_r(\hat{\mathbf{R}})$
JST	$\tilde{\mathbf{R}}_{\mathrm{ST}}$	$\hat{\mathbf{\Pi}}_{\mathrm{JST}}$
DST	$\tilde{\mathbf{R}}_{\mathrm{T}} \otimes \tilde{\mathbf{R}}_{\mathrm{S}}$	$\hat{\mathbf{\Pi}}^{*}_{\mathrm{T}} \otimes \hat{\mathbf{\Pi}}_{\mathrm{S}}$
SST	$\tilde{\mathbf{R}}_{\mathrm{T}} \otimes \tilde{\mathbf{R}}^{\mathrm{TX}}_{\mathrm{S}} \otimes \tilde{\mathbf{R}}^{\mathrm{RX}}_{\mathrm{S}}$	$\hat{\mathbf{\Pi}}^{*}_{\mathrm{T}} \otimes \hat{\mathbf{\Pi}}^{*}_{\mathrm{S,TX}} \otimes \hat{\mathbf{\Pi}}_{\mathrm{S,RX}}$
LL-JST	$\tilde{\mathbf{R}}_{\mathrm{T}} \otimes \tilde{\mathbf{R}}^{\mathrm{TX}}_{\mathrm{S}} \otimes \tilde{\mathbf{R}}^{\mathrm{RX}}_{\mathrm{S}}$	$\tilde{\mathbf{U}}_{\mathrm{SST}} \hat{\mathbf{\Pi}}_{\mathrm{LLJST}} \tilde{\mathbf{U}}^{\mathrm{H}}_{\mathrm{SST}}$

is computationally expensive and highly sensitive to antenna calibration. The LR channel models and the corresponding estimators are detailed in the following.

A. LR Algebraic Channel Structure

1) JST model: To highlight the relevant parameters and isolate the slowly-varying ones from the fast-varying terms, we consider the vectorized ST channel $\boldsymbol{h}_\ell = \mathrm{vec}\left(\mathcal{H}_\ell\right) \in \mathbb{C}^{WN_{\mathrm{T}}N_{\mathrm{R}}\times 1}$ re-parametrized as:

$$\boldsymbol{h}_\ell = \sum_{p=1}^{P} \underbrace{\left(\mathbf{g}\left(\tau_p\right) \otimes \mathbf{a}\left(\boldsymbol{\theta}_p\right)\right)}_{\mathbf{o}_p} \alpha_{p,\ell} = \mathbf{O}\left(\boldsymbol{\tau},\boldsymbol{\theta}\right)\boldsymbol{\alpha}_\ell, \tag{10}$$

where $\mathbf{O}\left(\boldsymbol{\tau},\boldsymbol{\theta}\right) = \left[\mathbf{o}_1 \cdots \mathbf{o}_P\right] \in \mathbb{C}^{N_{\mathrm{R}}N_{\mathrm{T}}W\times P}$ collects the block-invariant ST signatures of the P paths (depending on angles and delays), while $\boldsymbol{\alpha}_\ell = \left[\alpha_{1,\ell},\cdots,\alpha_{P,\ell}\right]^{\mathrm{T}} \in \mathbb{C}^{P\times 1}$ embeds the block-varying amplitudes. According to the WSSUS assumption, the ST channel correlation matrix is

$$\begin{aligned}\mathbf{R}_{\mathrm{ST}} =& \mathbb{E}\left[\boldsymbol{h}_\ell \boldsymbol{h}_\ell^{\mathrm{H}}\right] = \mathbf{O}\left(\boldsymbol{\tau},\boldsymbol{\theta}\right)\mathbf{\Omega}\mathbf{O}\left(\boldsymbol{\tau},\boldsymbol{\theta}\right)^{\mathrm{T}} \\ =& \sum_{p=1}^{P} \Omega_p \left(\mathbf{R}_{\mathrm{T},p} \otimes \mathbf{R}^{\mathrm{TX}}_{\mathrm{S},p} \otimes \mathbf{R}^{\mathrm{RX}}_{S,p}\right), \end{aligned} \tag{11}$$

(12)

where $\mathbf{R}_{\mathrm{T},p} = \mathbf{g}\left(\tau_p\right)\mathbf{g}\left(\tau_p\right)^{\mathrm{T}} \in \mathbb{C}^{W\times W}$ is the p-th path temporal term, while $\mathbf{R}^{\mathrm{TX}}_{\mathrm{S},p} = \mathbf{a}^{\mathrm{TX}}\left(\boldsymbol{\theta}^{\mathrm{TX}}_p\right)\mathbf{a}^{\mathrm{TX}}\left(\boldsymbol{\theta}^{\mathrm{TX}}_p\right)^{\mathrm{T}} \in \mathbb{C}^{N_{\mathrm{T}}\times N_{\mathrm{T}}}$ and $\mathbf{R}^{\mathrm{RX}}_{\mathrm{S},p} = \mathbf{a}^{\mathrm{RX}}\left(\boldsymbol{\theta}^{\mathrm{RX}}_p\right)\mathbf{a}^{\mathrm{RX}}\left(\boldsymbol{\theta}^{\mathrm{RX}}_p\right)^{\mathrm{T}} \in \mathbb{C}^{N_{\mathrm{R}}\times N_{\mathrm{R}}}$ are the spatial terms at the transmitter and receiver side, respectively. The number of resolvable paths given the system resolution (i.e., the antenna array aperture and the system bandwidth) is

$$r = \mathrm{rank}\left(\mathbf{O}\left(\boldsymbol{\tau},\boldsymbol{\theta}\right)\right) \leq \min\left(N_{\mathrm{R}}N_{\mathrm{T}}W, P\right). \tag{13}$$

Based on the LR constraint (13), we can rewrite the channel (10) according to the joint space-time (JST) LR model for MIMO channels [20]

$$\boldsymbol{h}_\ell = \mathbf{U}_{\mathrm{ST}}\boldsymbol{\gamma}_\ell, \tag{14}$$

where $\mathbf{U}_{\mathrm{ST}} \in \mathbb{C}^{WN_{\mathrm{T}}N_{\mathrm{R}}\times r}$ is a block-independent full-rank matrix that collects the r eigenvectors of $\mathbf{R}_{\mathrm{ST}}$ and spans the invariant ST subspace $\mathcal{R}\left(\mathbf{U}_{\mathrm{ST}}\right) = \mathcal{R}\left(\mathbf{O}\left(\boldsymbol{\tau},\boldsymbol{\theta}\right)\right)$, while $\boldsymbol{\gamma}_\ell \in \mathbb{C}^{r\times 1}$ contains the related block-dependent weights.

2) DST model: To reduce the complexity of channel estimation, we also introduce a simplified LR model that assumes a separable ST structure for the correlation matrix $\mathbf{R}_{\mathrm{ST}}$ in (12), according to the Kronecker model of [27] and [28], reported in Table II (second row). Note that this is an approximation as the Kronecker structure holds only for each single path in (12), not for the multipath combination. Following the approach in [18] for single-input-multiple-output (SIMO) channels, here we consider a separate ST LR channel model, referred to as doubly-space time (DST) model, with $\mathbf{U}_{\mathrm{ST}} = \mathbf{U}^{*}_{\mathrm{T}} \otimes \mathbf{U}_{\mathrm{S}}$, or equivalently:

$$\boldsymbol{h}_\ell = \left(\mathbf{U}^{*}_{\mathrm{T}} \otimes \mathbf{U}_{\mathrm{S}}\right)\boldsymbol{\gamma}_\ell, \tag{15}$$

where $\mathbf{U}_{\mathrm{S}} \in \mathbb{C}^{N_{\mathrm{T}}N_{\mathrm{R}}\times r_{\mathrm{S}}}$ and $\mathbf{U}_{\mathrm{T}} \in \mathbb{C}^{W\times r_{\mathrm{T}}}$ contain the leading eigenvectors of the channel correlation matrices

$$\mathbf{R}_{\mathrm{S}} = \mathbb{E}[\mathcal{H}_\ell \mathcal{H}_\ell^{\mathrm{H}}] = \mathcal{A}\left(\boldsymbol{\theta}\right)\mathbf{\Lambda}_{\mathrm{S}}\mathcal{A}\left(\boldsymbol{\theta}\right)^{\mathrm{H}}, \tag{16a}$$

$$\mathbf{R}_{\mathrm{T}} = \mathbb{E}[\mathcal{H}_\ell^{\mathrm{H}} \mathcal{H}_\ell] = \mathbf{G}(\boldsymbol{\tau})\mathbf{\Lambda}_{\mathrm{T}}\mathbf{G}(\boldsymbol{\tau})^{\mathrm{T}}, \tag{16b}$$

respectively, with $\mathbf{\Lambda}_{\mathrm{S}} = \mathbf{\Omega} \odot \left(\mathbf{G}(\boldsymbol{\tau})^{\mathrm{T}}\mathbf{G}(\boldsymbol{\tau})\right)$ and $\mathbf{\Lambda}_{\mathrm{T}} = \mathbf{\Omega} \odot \left(\mathcal{A}\left(\boldsymbol{\theta}\right)^{\mathrm{H}}\mathcal{A}\left(\boldsymbol{\theta}\right)\right)$. Matrices $\mathbf{U}_{\mathrm{S}}$ and $\mathbf{U}_{\mathrm{T}}$ span, respectively, the spatial and temporal channel subspaces $\mathcal{R}\left(\mathbf{U}_{\mathrm{S}}\right) = \mathcal{R}\left(\mathcal{A}\left(\boldsymbol{\theta}\right)\right)$ and $\mathcal{R}\left(\mathbf{U}_{\mathrm{T}}\right) = \mathcal{R}\left(\mathbf{G}\left(\boldsymbol{\tau}\right)\right)$. The corresponding spatial and temporal rank orders

$$r_{\mathrm{S}} = \mathrm{rank}(\mathcal{A}) \leq \min(P, N_{\mathrm{R}}N_{\mathrm{T}}), \tag{17a}$$

$$r_{\mathrm{T}} = \mathrm{rank}(\mathbf{G}) \leq \min(P, W), \tag{17b}$$

represent the number of resolvable angles and delays.

3) SST model: A further approximation, that we propose to lower the complexity of channel estimation, is based on the extension of the Kronecker model to the spatial MIMO correlation $\mathbf{R}_{\mathrm{S}}$ as detailed in Table II (third row). This reduces (15) to the separate space-time (SST) channel model

$$h_\ell = \left(\mathbf{U}^{*}_{\mathrm{T}} \otimes \mathbf{U}^{\mathrm{TX}*}_{\mathrm{S}} \otimes \mathbf{U}^{\mathrm{RX}}_{\mathrm{S}}\right)\boldsymbol{\gamma}_\ell, \tag{18}$$

where $\mathbf{U}^{\mathrm{TX}}_{\mathrm{S}} \in \mathbb{C}^{N_{\mathrm{T}}\times r_{\mathrm{S,TX}}}$ and $\mathbf{U}^{\mathrm{RX}}_{\mathrm{S}} \in \mathbb{C}^{N_{\mathrm{R}}\times r_{\mathrm{S,RX}}}$ are the transmit and receive spatial bases collecting the leading eigenvectors of the channel correlation matrices

$$\mathbf{R}^{\mathrm{TX}}_{\mathrm{S}} = \sum_{w=1}^{W} \mathbb{E}\left[\mathbf{h}\left(w\right)^{\mathrm{H}}\mathbf{h}\left(w\right)\right] = \mathbf{A}^{\mathrm{TX}}\mathbf{\Lambda}_{\mathrm{S}}\mathbf{A}^{\mathrm{TX,H}}, \tag{19a}$$

$$\mathbf{R}^{\mathrm{RX}}_{\mathrm{S}} = \sum_{w=1}^{W} \mathbb{E}\left[\mathbf{h}\left(w\right)\mathbf{h}\left(w\right)^{\mathrm{H}}\right] = \mathbf{A}^{\mathrm{RX}}\mathbf{\Lambda}_{\mathrm{S}}\mathbf{A}^{\mathrm{RX,H}}, \tag{19b}$$

respectively, with $\mathbf{A}^{\mathrm{TX}} = \left[\mathbf{a}^{\mathrm{TX}}(\boldsymbol{\theta}^{\mathrm{TX}}_1)\cdots\mathbf{a}^{\mathrm{TX}}(\boldsymbol{\theta}^{\mathrm{TX}}_P)\right] \in \mathbb{C}^{N_{\mathrm{T}}\times P}$ and $\mathbf{A}^{\mathrm{RX}} = \left[\mathbf{a}^{\mathrm{RX}}(\boldsymbol{\theta}^{\mathrm{RX}}_1)\cdots\mathbf{a}^{\mathrm{RX}}(\boldsymbol{\theta}^{\mathrm{RX}}_P)\right] \in \mathbb{C}^{N_{\mathrm{R}}\times P}$. The spatial diversity orders, which represent the number of angles that can be resolved by the transmitting and receiving array, respectively, are

$$r_{\mathrm{S,TX}} = \mathrm{rank}\left(\mathbf{A}^{\mathrm{TX}}\right) \leq \min\left(N_{\mathrm{T}}, P\right), \tag{20a}$$

$$r_{\mathrm{S,RX}} = \mathrm{rank}\left(\mathbf{A}^{\mathrm{RX}}\right) \leq \min\left(N_{\mathrm{R}}, P\right). \tag{20b}$$

This new parametrization yields $\mathcal{R}\left(\mathbf{U}^{\mathrm{TX}}_{\mathrm{S}}\right) = \mathcal{R}\left(\mathbf{A}^{\mathrm{TX}}\right)$ and $\mathcal{R}\left(\mathbf{U}^{\mathrm{RX}}_{\mathrm{S}}\right) = \mathcal{R}\left(\mathbf{A}^{\mathrm{RX}}\right)$.

B. LR Estimation Methods

The LR algorithms perform a maximum likelihood (ML) estimation of the ST MIMO channel from the signals (6), under the LR constraints (14), (15), or (18). Following the

approach in [18], the LR estimate can be computed, for a known rank order r, as

$$\hat{\boldsymbol{h}}_{\mathrm{LR},\ell} = \hat{\boldsymbol{C}}^{\frac{\mathrm{H}}{2}} \hat{\boldsymbol{\Pi}} \hat{\boldsymbol{C}}^{-\frac{\mathrm{H}}{2}} \boldsymbol{h}_\ell = \hat{\boldsymbol{C}}^{\frac{\mathrm{H}}{2}} \hat{\boldsymbol{\Pi}} \tilde{\boldsymbol{h}}_\ell, \tag{21}$$

i.e., by projecting the pre-whitened LS estimate

$$\tilde{\boldsymbol{h}}_\ell = \hat{\boldsymbol{C}}^{-\frac{\mathrm{H}}{2}} \hat{\boldsymbol{h}}_\ell, \tag{22}$$

onto the subspace spanned by the r leading vectors $\hat{\mathbf{U}} = \mathrm{eig}_{\mathrm{r}}[\hat{\mathbf{R}}]$ of a channel correlation estimate $\hat{\mathbf{R}}$, through the projector $\hat{\boldsymbol{\Pi}} = \boldsymbol{\Pi}_r(\hat{\mathbf{R}}) = \hat{\mathbf{U}}\hat{\mathbf{U}}^{\mathrm{H}}$. As detailed in the following for the specific methods, matrix $\hat{\mathbf{R}}$ is a sample estimate over L blocks of the ST channel correlation, obtained according to the JST model (12), or the related DST and SST Kronecker approximations in Table II. A low-latency low-complexity (LL) suboptimal algorithm is finally proposed to approach the optimal LR performance with lower complexity and faster convergence.

1) JST Estimator: The JST method [20] follows from the LR model (14) and is based on the projector

$$\hat{\boldsymbol{\Pi}}_{\mathrm{JST}} = \boldsymbol{\Pi}_r(\tilde{\mathbf{R}}_{\mathrm{ST}}), \tag{23}$$

computed from the ST sample correlation matrix

$$\tilde{\mathbf{R}}_{\mathrm{ST}} = \frac{1}{L}\sum_{\ell=1}^{L} \tilde{\boldsymbol{h}}_\ell \tilde{\boldsymbol{h}}_\ell^{\mathrm{H}}, \tag{24}$$

for a known rank order r.

2) DST Estimator: The DST method is a suboptimal approach that simplifies the projector computation by assuming the separable ST model (15), with known diversity orders $(r_{\mathrm{S}}, r_{\mathrm{T}})$. This provides the projector [18]

$$\hat{\boldsymbol{\Pi}}_{\mathrm{DST}} = \boldsymbol{\Pi}^*_{r_{\mathrm{T}}}\left(\tilde{\mathbf{R}}_{\mathrm{T}}\right) \otimes \boldsymbol{\Pi}_{r_{\mathrm{S}}}\left(\tilde{\mathbf{R}}_{\mathrm{S}}\right), \tag{25}$$

with spatial and temporal sample correlations

$$\tilde{\mathbf{R}}_{\mathrm{S}} = \frac{1}{L}\sum_{\ell=1}^{L} \tilde{\mathcal{H}}_\ell \tilde{\mathcal{H}}_\ell^{\mathrm{H}}, \tag{26a}$$

$$\tilde{\mathbf{R}}_{\mathrm{T}} = \frac{1}{L}\sum_{\ell=1}^{L} \tilde{\mathcal{H}}_\ell^{\mathrm{H}} \tilde{\mathcal{H}}_\ell, \tag{26b}$$

where $\tilde{\mathcal{H}}_\ell = \mathrm{vec}^{-1}(\tilde{\boldsymbol{h}}_\ell)$ is the whitened LS channel estimate rearranged into a matrix of dimensions $N_{\mathrm{T}}N_{\mathrm{R}} \times W$ (see Table I).

3) SST Estimator: The SST LR method is a new method that further extends the separable structure assumption to the spatial domain (18), with diversity orders $(r_{\mathrm{T}}, r_{\mathrm{S,RX}}, r_{\mathrm{S,TX}})$. It uses the projector

$$\hat{\boldsymbol{\Pi}}_{\mathrm{SST}} = \boldsymbol{\Pi}^*_{r_{\mathrm{T}}}\left(\tilde{\mathbf{R}}_{\mathrm{T}}\right) \otimes \boldsymbol{\Pi}^*_{r_{\mathrm{S,TX}}}\left(\tilde{\mathbf{R}}^{\mathrm{TX}}_{\mathrm{S}}\right) \otimes \boldsymbol{\Pi}_{r_{\mathrm{S,RX}}}\left(\tilde{\mathbf{R}}^{\mathrm{RX}}_{\mathrm{S}}\right), \tag{27}$$

with spatial correlations at the two sides of the MIMO link

$$\tilde{\mathbf{R}}^{\mathrm{RX}}_{\mathrm{S}} = \frac{1}{L}\sum_{\ell=1}^{L}\sum_{w=1}^{W} \tilde{\mathbf{h}}_\ell(w)\, \tilde{\mathbf{h}}_\ell^{\mathrm{H}}(w), \tag{28a}$$

$$\tilde{\mathbf{R}}^{\mathrm{TX}}_{\mathrm{S}} = \frac{1}{L}\sum_{\ell=1}^{L}\sum_{w=1}^{W} \tilde{\mathbf{h}}_\ell^{\mathrm{H}}(w)\, \tilde{\mathbf{h}}_\ell(w), \tag{28b}$$

and $\tilde{\mathbf{h}}_\ell(w)$ being the rearrangement of $\tilde{\mathcal{H}}_\ell$ according to (9).

4) LL-JST Estimator: The computational cost of the above LR algorithms mainly depends on the eigen-decomposition, as detailed in Table III. For the JST method the complexity becomes unbearably high and as $L \gg N_{\mathrm{R}}N_{\mathrm{T}}W$, the number of blocks required for the correlation estimation is unfeasible for practical (latency-constrained) mmWave/THz systems. This expensive and unfeasible computation is avoided by the low-latency low-complexity joint space-time (LL-JST) algorithm, herein proposed, which reduces the channel dimensions by exploiting the ST bases estimated by the SST algorithm (with $L \ll N_{\mathrm{R}}N_{\mathrm{T}}W$):

$$\tilde{\mathbf{U}}_{\mathrm{SST}} = \tilde{\mathbf{U}}^*_{\mathrm{T}} \otimes \left(\tilde{\mathbf{U}}^*_{\mathrm{S,TX}} \otimes \tilde{\mathbf{U}}_{\mathrm{S,RX}}\right), \tag{29}$$

where $\tilde{\mathbf{U}}_{\mathrm{T}} = \mathrm{eig}_{r_{\mathrm{T}}}\left[\tilde{\mathbf{R}}_{\mathrm{T}}\right]$, $\tilde{\mathbf{U}}_{\mathrm{S,TX}} = \mathrm{eig}_{r_{\mathrm{S,RX}}}\left[\tilde{\mathbf{R}}^{\mathrm{TX}}_{\mathrm{S}}\right]$, and $\tilde{\mathbf{U}}_{\mathrm{S,TX}} = \mathrm{eig}_{r_{\mathrm{S,TX}}}\left[\tilde{\mathbf{R}}^{\mathrm{TX}}_{\mathrm{S}}\right]$. The $r_{\mathrm{S,TX}}r_{\mathrm{S,RX}}r_{\mathrm{T}} \times 1$ projection of the LS channel estimate onto basis

$$\hat{\boldsymbol{s}}_{\mathrm{SST},\ell} = \tilde{\mathbf{U}}^{\mathrm{H}}_{\mathrm{SST}} \tilde{\boldsymbol{h}}_\ell = \boldsymbol{s}_\ell + \boldsymbol{n}_{\mathrm{SST},\ell}, \tag{30}$$

is the sum of the channel-related component $\boldsymbol{s}_\ell = \tilde{\mathbf{U}}^{\mathrm{H}}_{\mathrm{SST}} \boldsymbol{C}^{-\frac{\mathrm{H}}{2}} \boldsymbol{h}_\ell$ and the projected LS estimate error $\boldsymbol{n}_{\mathrm{SST},\ell} = \tilde{\mathbf{U}}^{\mathrm{H}}_{\mathrm{SST}} \boldsymbol{C}^{-\frac{\mathrm{H}}{2}} \triangle\boldsymbol{h}_{\mathrm{LS},\ell}$, with $\triangle\boldsymbol{h}_{\mathrm{LS},\ell} = \hat{\boldsymbol{h}}_\ell - \boldsymbol{h}_\ell$. Note that $\mathcal{R}(\mathbf{U}_{\mathrm{SST}}) \supseteq \mathcal{R}(\mathbf{U}_{\mathrm{JST}})$ and for $L \to \infty$, it is $\hat{\boldsymbol{C}} \to \boldsymbol{C}$ and $\tilde{\mathbf{U}}_{\mathrm{SST}} \to \boldsymbol{C}^{-\frac{\mathrm{H}}{2}}\mathbf{U}_{\mathrm{SST}}$, thereby the rank-$r$ channel is entirely embedded in $\boldsymbol{s}_\ell$. The residual $\boldsymbol{n}_{\mathrm{SST},\ell}$ is white and includes both the component laying in the channel subspace $\mathcal{R}(\tilde{\mathbf{U}}_{\mathrm{JST}})$ (which can no longer be removed) and the artifacts laying in the orthogonal subspace $\mathcal{R}(\tilde{\mathbf{U}}_{\mathrm{SST}}) \setminus \mathcal{R}(\tilde{\mathbf{U}}_{\mathrm{JST}})$, captured by the intersections of the three separate domains due to the SST Kronecker approximation. In order to remove these artifacts, we propose to apply the optimal JST approach to the rank-r signal $\hat{\boldsymbol{s}}_{\mathrm{SST},\ell}$ and therein identify the long-term channel subspace. The sample correlation:

$$\hat{\mathbf{R}}_v = \frac{1}{L}\sum_{\ell=1}^{L} \hat{\boldsymbol{s}}_{\mathrm{SST},\ell} \hat{\boldsymbol{s}}^{\mathrm{H}}_{\mathrm{SST},\ell} \tag{31}$$

isolates the truly invariant channel structure, thus the subspace projector is

$$\hat{\boldsymbol{\Pi}}_{\mathrm{LLJST}} = \boldsymbol{\Pi}_r(\hat{\mathbf{R}}_v) \tag{32}$$

and the LL-JST estimate is computed as:

$$\hat{\boldsymbol{h}}_{\mathrm{LR},\ell} = \hat{\boldsymbol{C}}^{\frac{\mathrm{H}}{2}} \tilde{\mathbf{U}}_{\mathrm{SST}} \hat{\boldsymbol{\Pi}}_{\mathrm{LLJST}} \hat{\boldsymbol{s}}_{\mathrm{SST},\ell}. \tag{33}$$

C. Asymptotic MSE Bounds

Following the approach in [21], a closed form expression for the MSE of the LR channel estimate is derived here as $\mathrm{MSE}_{\mathrm{LR}} = \mathbb{E}[||\triangle\boldsymbol{h}_{\mathrm{LR},\ell}||^2]$, with $\triangle\boldsymbol{h}_{\mathrm{LR},\ell} = \hat{\boldsymbol{h}}_{\mathrm{LR},\ell} - \boldsymbol{h}_\ell$. The impact of a possible mismatch between the estimated rank orders $\hat{\mathbf{r}} = [\hat{r}, \hat{r}_{\mathrm{T}}, \hat{r}_{\mathrm{S}}, \hat{r}_{\mathrm{S,TX}}, \hat{r}_{\mathrm{S,RX}}]$ and the true values $\mathbf{r} = [r, r_{\mathrm{T}}, r_{\mathrm{S}}, r_{\mathrm{S,TX}}, r_{\mathrm{S,RX}}]$ is taken into account. The choice $\hat{\mathbf{r}} = \mathbf{r}$ is the minimum order that gives an unbiased channel estimate, but in general it does not provide the lowest MSE.

To highlight this, in the Appendix we derive the LR performance for *any* rank-order selection, comprised within the correlations' dimensions. The performance, here referred to as asymptotic MSE bound, is computed analytically for $L \to \infty$. The asymptotic MSE bound will be used in the following sections to analyze the impact of rank-order selection on the performance of mmWave/THz systems and provide a baseline for the algorithms' comparison. Note that for $L \to \infty$ the noise covariance $\mathbf{Q}$ can be considered as known (as $\hat{\mathbf{Q}}_\ell \to \mathbf{Q}$) and the estimated ST projectors equal to the projectors onto the subspaces spanned by $\boldsymbol{C}^{-\frac{\mathrm{H}}{2}}\boldsymbol{h}_\ell$. Combining (21) and (22) with (7), it can be shown that the error based on the rank orders $\hat{\mathbf{r}}$ is the sum of two terms, one accounting for the distortion due to rank mismatch (i.e., for $[\hat{\mathbf{r}}]_i < [\mathbf{r}]_i, 1 \leq i \leq 5$) and the other for the noise, i.e.,

$$\triangle \boldsymbol{h}_{\mathrm{LR},\ell} = \boldsymbol{C}^{\frac{\mathrm{H}}{2}}\boldsymbol{\Pi}_{\hat{\mathbf{r}}}\boldsymbol{C}^{-\frac{\mathrm{H}}{2}}\triangle \boldsymbol{h}_{\mathrm{LS},\ell} + \boldsymbol{C}^{\frac{\mathrm{H}}{2}}\triangle\boldsymbol{\Pi}_{\hat{\mathbf{r}}}\boldsymbol{C}^{-\frac{\mathrm{H}}{2}}\boldsymbol{h}_\ell, \quad (34)$$

where $\boldsymbol{C}^{-\frac{\mathrm{H}}{2}}\triangle\boldsymbol{h}_{\mathrm{LS},\ell}$ is the white error on the LS channel estimate and $\triangle\boldsymbol{\Pi}_{\hat{\mathbf{r}}} = \boldsymbol{\Pi}_{\mathbf{r}} - \boldsymbol{\Pi}_{\hat{\mathbf{r}}}$. This leads to the additive MSE model

$$\mathrm{MSE}_{\mathrm{LR}}(\hat{\mathbf{r}}) = \mathrm{MSE}_n(\hat{\mathbf{r}}) + \mathrm{MSE}_b(\hat{\mathbf{r}}), \quad (35)$$

where

$$\mathrm{MSE}_n(\hat{\mathbf{r}}) = \mathrm{tr}\left(\boldsymbol{C}^{\frac{\mathrm{H}}{2}}\boldsymbol{\Pi}_{\hat{\mathbf{r}}}\boldsymbol{C}^{\frac{1}{2}}\right) \quad (36)$$

represents the noise filtered by the base of the projector, hence increasing with the rank orders in $\hat{\mathbf{r}}$ as a larger base is selected, while

$$\mathrm{MSE}_b(\hat{\mathbf{r}}) = \mathrm{tr}\left(\boldsymbol{C}^{\frac{\mathrm{H}}{2}}\triangle\boldsymbol{\Pi}_{\hat{\mathbf{r}}}\boldsymbol{C}^{-\frac{\mathrm{H}}{2}}\mathbf{R}_{\mathrm{ST}}\right) \quad (37)$$

represents the bias due to the under-parametrization in the rank order selection, decreasing with the values in $\hat{\mathbf{r}}$ as the mismatch between the projectors becomes smaller. The closed forms of MSE for the LR algorithms are obtained by specializing the generic projector $\boldsymbol{\Pi}_{\hat{\mathbf{r}}}$ with the expressions derived in Section III-B and summarized in Table II.

IV. Time-dependent Accelerated Gradient Descent

We now introduce a second method for channel estimation, based on the sparsity of the channel in a dual domain. First we introduce the angular domain channel representation and then we exploit it for the estimate.

The problem of efficiently estimating the entries of the angular-domain channel for the proposed model has already been tackled in [22], for a narrowband transmission scenario. In this work we exploit the AGDAR algorithm of [22], for a *broadband* transmission scenario. Furthermore, by noticing that the AGDAR algorithm is an upper-bound of the optimal solution (more on this later), we propose two new estimators, namely S-AGDAR and A-AGDAR, which improve the identification of channel paths and the estimation of their gains. Therefore, the AGDAR algorithm provides a baseline for all the proposed CS-based algorithms.

A. Angular Domain Channel Representation

In this section we introduce the space-space angular-domain representation of the channel in the angular domain [29] and the novel ST angular domain representation. We assume that uniform linear arrays (ULAs) are deployed at both the transmitter and the receiver.

The w-th tap of the channel impulse response between all transmitter and receiver antennas can be represented by the angular domain matrix $\boldsymbol{V}_\ell(w)$ as

$$\boldsymbol{h}_\ell(w) = \boldsymbol{F}_R^H \boldsymbol{V}_\ell(w)\boldsymbol{F}_T, \quad (38)$$

where $\boldsymbol{F}_R$ and $\boldsymbol{F}_T$ are the $N_{\mathrm{R}} \times M_{\mathrm{R}}$ and $N_{\mathrm{T}} \times M_{\mathrm{T}}$ DFT matrices respectively. The vector form of $\boldsymbol{h}_\ell(w)$ is obtained as

$$\mathrm{vec}(\boldsymbol{h}_\ell(w)) = \boldsymbol{\mathcal{F}}\boldsymbol{v}_\ell(w), \quad (39)$$

where $\boldsymbol{v}_\ell(w)$ is the vectorized form of $\boldsymbol{V}_\ell(w)$, and $\boldsymbol{\mathcal{F}} = \left(\boldsymbol{F}_T^T \otimes \boldsymbol{F}_R^H\right)$. The overall channel matrix is obtained as (see (4))

$$\mathcal{H}_\ell = [\boldsymbol{\mathcal{F}}\boldsymbol{v}_\ell(1), \cdots, \boldsymbol{\mathcal{F}}\boldsymbol{v}_\ell(W)]. \quad (40)$$

B. Channel Estimation In The Angular Domain

The AGDAR estimation of the w-th tap is obtained by solving the following minimization problem

$$\boldsymbol{v}_{\mathrm{LS},\ell}(w) = \underset{\boldsymbol{\varrho}}{\mathrm{argmin}} \left(\|\boldsymbol{\mathcal{F}}\boldsymbol{\varrho} - \mathrm{vec}\left(\boldsymbol{h}_\ell(w)\right)\|_2^2 + \rho \|\boldsymbol{v}_\ell\|_0 \right), \quad (41)$$

as described in [22], where ρ is the weight assigned to the sparsity imposing constraint. As the norm 0 is not convex, the objective function (41) is approximated via the norm 1 as

$$\boldsymbol{v}_{\mathrm{LS},\ell}(w) = \underset{\boldsymbol{\varrho}}{\mathrm{argmin}} \left(\|\boldsymbol{\mathcal{F}}\boldsymbol{\varrho} - \mathrm{vec}\left(\boldsymbol{h}_\ell(w)\right)\|_2^2 + \rho \|\boldsymbol{v}_\ell\|_1 \right), \quad (42)$$

which is convex. However, as the wideband channel has most of the paths concentrated in the first taps and considering that the sparse solution is obtained by approximating the norm 0 with a norm 1, solving a CS problem for each tap is widely suboptimal, as the same power is allocated to taps that contain different channel paths. In order to obtain a sparser solution we jointly estimate the wideband channel taps.

As outlined in (40) the angular domain representation of the wideband channel is obtained by a DFT of each of the W taps. Therefore, starting from (40), the ST LS estimate of the channel $\mathcal{H}_{\mathrm{LS},\ell}$ can be approximated by

$$\mathcal{H}_{\mathrm{LS},\ell} \approx \boldsymbol{\mathcal{F}}[\boldsymbol{v}_\ell(1) \cdots \boldsymbol{v}_\ell(W)] + \boldsymbol{N}_\ell, \quad (43)$$

whose vector form is obtained as

$$\mathbf{h}_{\mathrm{LS},\ell} \approx \boldsymbol{\mathcal{F}}_K \boldsymbol{v}_\ell + \boldsymbol{n}_\ell, \quad (44)$$

where $\boldsymbol{v}_\ell = \mathrm{vec}\left([\boldsymbol{v}_\ell(1) \cdots \boldsymbol{v}_\ell(W)]\right)$, $\boldsymbol{\mathcal{F}}_K = \boldsymbol{I} \otimes \boldsymbol{\mathcal{F}}$, and $\boldsymbol{I}$ is the $W \times W$ identity matrix.

We now define the sparse channel estimation problem as

$$\bar{\boldsymbol{v}}_{\mathrm{LS},\ell} = \underset{\boldsymbol{\varpi}}{\mathrm{argmin}} \left(\|\boldsymbol{\mathcal{F}}_K\boldsymbol{\varpi} - \boldsymbol{h}_{\mathrm{LS},\ell}\|_2^2 + \rho \|\boldsymbol{\varpi}\|_0 \right), \quad (45)$$

where the sparsity constraint imposed via the norm 0 is approximated via the norm 1, in order to have a convex objective function. The sparsity of vector $\boldsymbol{v}_\ell$ depends on M_R and M_T: if the antenna array is not aligned toward the direction of the DFT, a DFT of the same size of the number of antennas may not be accurate enough to isolate the P paths.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSAC.2020.3000889, IEEE Journal on Selected Areas in Communications

Algorithm 1: AGDAR algorithm

Data: $\boldsymbol{\mathcal{F}}_K$, $\boldsymbol{h}_{\mathrm{LS},\ell}$, ρ, ϵ, α, β, I_t
Result: $\bar{\boldsymbol{v}}_{\mathrm{LS},\ell}$

$\boldsymbol{x}_0 = \boldsymbol{\mathcal{F}}_K^H \boldsymbol{h}_{\mathrm{LS},\ell}$;
$\boldsymbol{y}_0 = \boldsymbol{0}$;
$\omega_0 = j = 1, \quad \Xi = 2\epsilon$;
$t = 1/||\nabla \boldsymbol{\mathcal{F}}_K||_2^2$;
$\boldsymbol{\nabla}_0 = \boldsymbol{\mathcal{F}}_K^H (\boldsymbol{\mathcal{F}}_K \boldsymbol{x}_0 - \boldsymbol{h}_{\mathrm{LS},\ell})$;
while $j \leq I_t$ **And** $\Xi > \epsilon$ **do**
  $\boldsymbol{z}_j = \boldsymbol{x}_{j-1} - t\boldsymbol{\nabla}_{j-1}$;
  $[\boldsymbol{z}_j]_n = sgn([\boldsymbol{z}_j]_n) \max(|[\boldsymbol{z}_j]_n| - \rho t, 0)$, $\forall n = 1, \cdots, M_R M_T$;
  $\omega_j = 2/(1 + \sqrt{1 + 4\omega_{j-1}^2})$;
  if $(\boldsymbol{x}_{j-1} - \boldsymbol{z}_j)(\boldsymbol{z}_j - \boldsymbol{z}_{j-1}) > 0$ **then**
    $\boldsymbol{x}_j = \boldsymbol{z}_j$;
  else
    $\boldsymbol{x}_j = \boldsymbol{z}_j + (1 - \omega_j)(\boldsymbol{z}_j - \boldsymbol{z}_{j-1})$;
  end
  $\Xi = ||\boldsymbol{x}_{j-1} - \boldsymbol{x}_j||_2^2$;
  $\boldsymbol{\nabla}_{j-1} = \boldsymbol{\nabla}_j$;
  $\boldsymbol{\nabla}_j = \boldsymbol{\mathcal{F}}_K^H (\boldsymbol{\mathcal{F}}_K \boldsymbol{x}_j - \boldsymbol{h}_{\mathrm{LS},\ell})$;
  $\hat{t} = \frac{1\left(||\boldsymbol{z}_j - \boldsymbol{z}_{-j-1}||_2^2\right)}{2\left((\boldsymbol{z}_j - \boldsymbol{z}_{-j-1})^H (\boldsymbol{\nabla}_{j-1} - \boldsymbol{\nabla}_j)\right)}$;
  $t = \min(\alpha t, \max(\beta t, \hat{t}))$;
  $j = j + 1$;
end

By increasing the precision of the DFT matrix, i.e., with larger values of M_R and M_T, we obtain that different paths are aligned with the DFT grid and are hence associated to a single entry in the angular-domain channel matrix. [2]

Problem (43) is similar to that analyzed in [22] and we can hence resort to the proposed AGDAR algorithm. The steps of the AGDAR algorithm are reported in Algorithm 1: at the input it requires the extended DFT matrix (used to compute $\boldsymbol{\mathcal{F}}_K$), the LS estimate, and the sparsity regulation parameter ρ. Additional parameters are: α and β (to update the step size), ϵ (representing the precision required as stopping criterion), and the maximum number of iterations I_t. The algorithm is described in more detail in [22].

C. Selective AGDAR

In the AGDAR algorithm, due to the non-convexity of the norm-0 constraint, sparsity is imposed via a norm-1 constraint. As a consequence, spurious paths are present in the estimate, having small gains. Therefore, we propose to improve our estimate by removing entries with gain below a threshold value θ_{th}, and redistributing their power among the paths with gains above θ_{th}. In particular, given the angular domain channel estimate $\bar{\boldsymbol{v}}_{\mathrm{LS},\ell}$ obtained with the AGDAR algorithm, let $\mathcal{I}_\ell$ be the set of indices of the paths with power above θ_{th}, i.e.,

$$\mathcal{I}_\ell = \{i : |\bar{\boldsymbol{v}}_{\mathrm{LS},\ell}^{(i)}|^2 \geq \theta_{\mathrm{th}}\}, \tag{46}$$

where $\bar{\boldsymbol{v}}_{\mathrm{LS},\ell}^{(i)}$ denotes the i-th entry of vector $\bar{\boldsymbol{v}}_{\mathrm{LS},\ell}$. Then we re-estimate the gains of paths in $\mathcal{I}_\ell$ as

$$\hat{\boldsymbol{v}}_{\mathrm{LS},\ell}^{(\mathcal{I}_\ell)} = \underset{\boldsymbol{\varpi}}{\operatorname{argmin}} \left\| \boldsymbol{\mathcal{F}}_K(\mathcal{I}_\ell)\boldsymbol{\varpi} - \boldsymbol{h}_{\mathrm{LS},\ell} \right\|_2^2, \tag{47}$$

[2]An alternative approach would consider directly off-grid path estimation.

Algorithm 2: A-AGDAR algorithm

Data: L, $\boldsymbol{h}_{\mathrm{LS},\ell}\, \forall \ell = 1, \cdots, L$
Result: $\hat{\boldsymbol{v}}_{\mathrm{LS},\ell}^{(\mathcal{I})}\, \forall \ell = 1, \cdots, L$

for $\ell = 1, \cdots, L$ **do**
  compute $\bar{\boldsymbol{v}}_{\mathrm{LS},\ell}$ via AGDAR;
  $\mathcal{I}_\ell = \{i : |\bar{\boldsymbol{v}}_{\mathrm{LS},\ell}^{(i)}|^2 \geq \theta_{\mathrm{th}}\}$;
end
$\hat{P} = \frac{1}{L}\sum_{\ell=1}^{L} |\mathcal{I}_\ell|$;
$\mathcal{I}$ = {$\hat{P}$ most recurrent entries in $\cup_{\ell=1}^{L} \mathcal{I}_\ell$};
for $\ell = 1, \cdots, L$ **do**
  $\hat{\boldsymbol{v}}_{\mathrm{LS},\ell}^{(\mathcal{I})} = \underset{\boldsymbol{\varpi}}{\operatorname{argmin}} \left(\|\boldsymbol{\mathcal{F}}_K(\mathcal{I})\boldsymbol{\varpi} - \boldsymbol{h}_{\mathrm{LS},\ell}\|_2^2 \right)$;
end

where columns of $\boldsymbol{\mathcal{F}}_K(\mathcal{I}_\ell)$ are those of $\boldsymbol{\mathcal{F}}_K$ with indices in $\mathcal{I}_\ell$. The overall estimation process is defined as the S-AGDAR algorithm, and $\hat{\boldsymbol{v}}_{\mathrm{LS},\ell}^{(\mathcal{I}_\ell)}$ represents the new angular domain channel estimate for entries in $\mathcal{I}_\ell$, while other entries are assumed to be zero.

D. Adaptive AGDAR

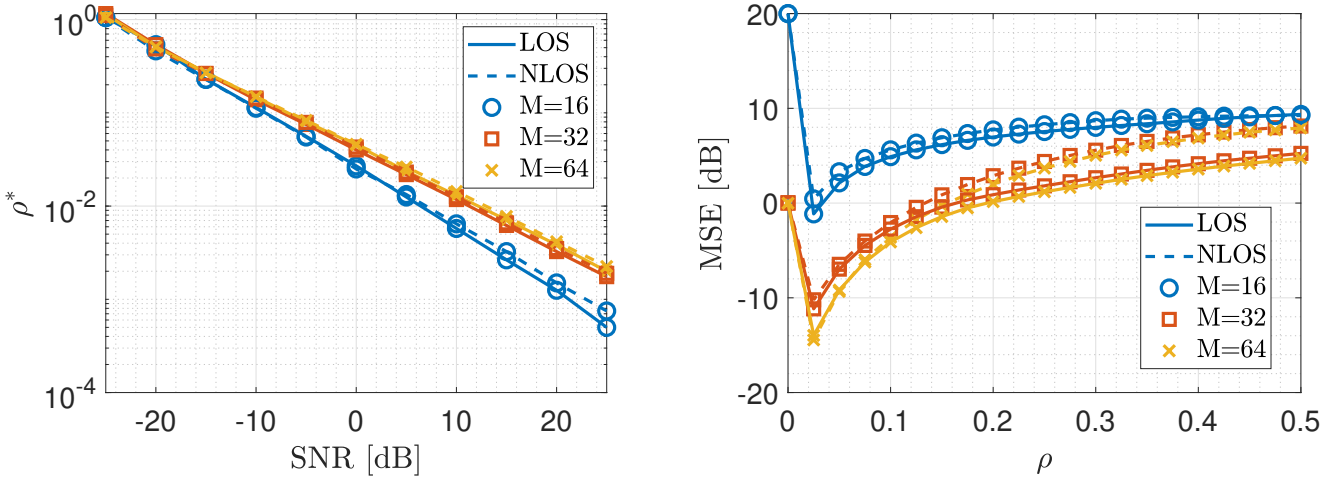
The S-AGDAR algorithm operates on a slot-by-slot basis. However some features of fading channels remain unchanged for more slots. In particular, typically DOAs and DODs are slowly time-varying and we can exploit this in the path selection process (46). In particular, we propose to refine the construction of set $\mathcal{I}_\ell$ as follows. First we compute the average number of channel paths over L LS estimates as

$$\hat{P} = \frac{1}{L}\sum_{\ell=1}^{L} |\mathcal{I}_\ell|, \tag{48}$$

where $|\mathcal{I}_\ell|$ is the cardinality of $\mathcal{I}_\ell$. Then, the set of indexes over which the path gain estimation (47) is solved is given by the $\hat{P}$ most recurrent entries in the union of the sets of the selected indexes, $\cup_{\ell=1}^{L} \mathcal{I}_\ell$. The overall estimation process is denoted as adaptive AGDAR (A-AGDAR). The steps for the A-AGDAR algorithm are shown in Algorithm 2. We note that initially L angular domain channel estimates must be obtained via the AGDAR algorithm, and for each one we create the set of indexes $\mathcal{I}_\ell$ of the paths having gain larger than θ_{th}. Then we compute the average number of paths $\hat{P}$ and create the set $\mathcal{I}$ of the $\hat{P}$ most recurrent paths. For each of the L estimates obtained via AGDAR, the gains of the paths are then computed via (47) on set $\mathcal{I}$.

Comparing A-AGDAR and S-AGDAR we observe that in general A-AGDAR improves over S-AGDAR, by better detecting paths of the virtual channel. The improvement is more relevant at low SNRs (and high L), where A-AGDAR is able to average out the effects of noise, while for a lower power S-AGDAR detects the correct path positions over the discrete DFT grid, and both S-AGDAR and A-AGDAR have the same performance. Lastly, the benefits of A-AGDAR come at the price of a longer channel estimate, as a larger number (L) of LS estimates must be collected. Indeed, when L is small the averaging process (48) may lead to the wrong estimate of $\hat{P}$ which then affects L channel estimates, possibly increasing the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSAC.2020.3000889, IEEE Journal on Selected Areas in Communications



(a) Optimal value ρ^*.

(b) MSE vs parameter ρ.

Fig. 1. a) Optimal value ρ^* for AGDAR methods for different SNR values and different DFT sizes M for the LOS and NLOS synthetic channels. b) Sensitivity of MSE to parameter ρ. Results are obtained for SNR $= 0$ dB and different DFT sizes M.

MSE with respect to S-AGDAR, which performs independent estimations at each slot.

E. Computational Complexity

The computational complexity of the AGDAR algorithms is assessed through the number of complex multiplications following the derivation of [22]. By letting $M_{\text{tot}} = M_R M_T W$ being the dimension of the overall DFT computed on the vectorized ST channel, the overall computational complexity is given by [22]

$$\mathcal{C}_{\text{AGDAR}} = O(3M_{\text{tot}}N_{\text{it}} + 2N_{\text{it}}M_{\text{tot}}\log_2 M_{\text{tot}}), \quad (49)$$

where N_{it} is the number of iterations. Note that we assume $M_{\text{tot}} > N_{\text{TOT}}$, as the angular domain has typically more entries than the number of antennas.

The S-AGDAR algorithm requires an additional comparison step with respect to the AGDAR algorithm and the solution of problem (47). In particular, each entry of the obtained angular-domain channel estimate is compared to a threshold value to select relevant paths. Then, the LS estimation of the selected paths' gain is obtained by solving an overdetermined system of $\hat{P}$ linear equations, where $\hat{P}$ represents the average number of channel entries above θ_{th}, with complexity at most $O(\hat{P}^3)$. Therefore we can express the computational complexity of S-AGDAR as

$$\mathcal{C}_{\text{S AGDAR}} = O(3M_{\text{tot}}N_{\text{it}} + 2N_{\text{it}}M_{\text{tot}}\log_2 M_{\text{tot}} + \hat{P}^3). \quad (50)$$

The A-AGDAR algorithm, which obtains an estimate of the locations of the channel paths in the angular domain by exploiting L LS estimates, requires the computation of L angular-domain channel estimates and the selection of the main paths by comparison with a threshold value for all of them. For each of the L LS estimates A-AGDAR requires the computation of the location of the paths in the angular domain via AGDAR and a successive comparison with a threshold value. Then, as the mean locations have been computed from the L LS estimates, A-AGDAR requires the solution of a linear system similar to the one in S-AGDAR, where the average number of channel entries above θ_{th} is denoted as $\hat{P}$. Therefore, the complexity of A-AGDAR once the channel paths have been computed is

$$\mathcal{C}_{A-AGDAR} = O(L(3M_{\text{tot}}N_{\text{it}} + 2N_{\text{it}}M_{\text{tot}}\log_2 M_{\text{tot}}) + \hat{P}^3). \quad (51)$$

V. Numerical Results

In this section we assess the (normalized) MSE of the proposed estimation methods, i.e.,

$$\text{MSE} = \frac{\mathbb{E}\left[|\Delta\mathbf{h}_\ell|^2\right]}{\mathbb{E}\left[|\mathbf{h}_\ell|^2\right]}, \quad (52)$$

where $\mathbf{h}_\ell = \left[\mathbf{h}_\ell^{(1)} \cdots \mathbf{h}_\ell^{(N_{\text{R}})}\right]$ and $\Delta\mathbf{h}_\ell = \mathbf{h}_\ell - \mathbf{h}_{\text{EST},\ell}$, with $\mathbf{h}_{\text{EST},\ell}$ denoting the generic $N_{\text{T}}W \times N$ estimated channel. Performance is assessed in terms of MSE, obtained as a function of the signal-to-interference-and-noise ratio (SINR) $\text{SINR} = \frac{\mathbb{E}\left[|\mathbf{h}_\ell|^2\right]}{\mathbb{E}\left[|\hat{\mathbf{N}}_\ell|^2\right]}$. The noise covariance can be expressed as:

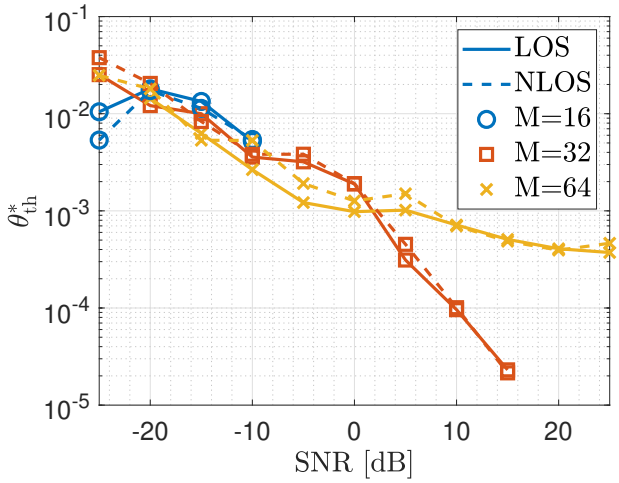
$$\mathbf{Q} = \sigma_n^2\mathbf{I}_{\text{N}_\text{R}} + \mathbf{Q}_{\text{I}}, \quad (53)$$

where the two terms represent the receiver noise and the interferers' contribution, respectively. We model the interferers as single rays impinging the receiving array, thus

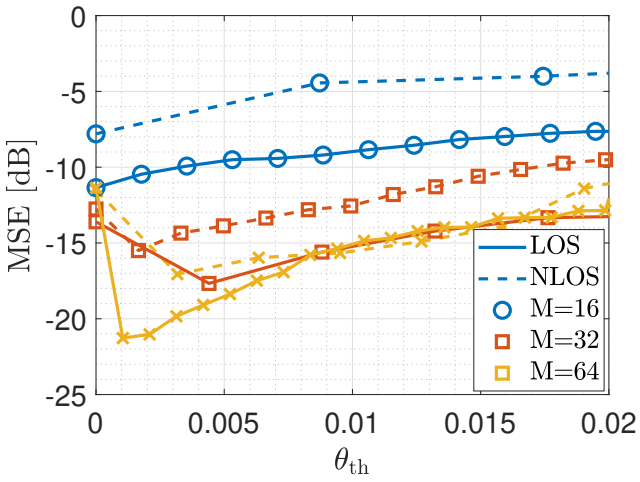
$$\mathbf{Q}_{\text{I}} = \sum_{i=1}^{N_{\text{I}}} \Omega_{\text{I},i}\sigma_{\text{I},i}^2 \mathbf{a}^{\text{RX}}(\boldsymbol{\theta}_{\text{I},i}^{\text{RX}})\mathbf{a}^{\text{RX}}(\boldsymbol{\theta}_{\text{I},i}^{\text{RX}})^{\text{H}}, \quad (54)$$

where N_{I} is the total number of interferers, while $\sigma_{\text{I},i}^2$, $\Omega_{\text{I},i}$, and $\boldsymbol{\theta}_{\text{I},i}^{\text{RX}}$ represent the single interferer transmitted power, the path gain, and the DOA, respectively. The SNR is the SINR in the absence of interference. In order to evaluate the convergence rate to the asymptotic bounds, we compute the MSE versus the number of slots L used for the estimation of the channel subspaces and path gains for the LR and

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSAC.2020.3000889, IEEE Journal on Selected Areas in Communications



(a) Optimal value $\theta^*_{\rm th}$.



(b) MSE vs parameter $\theta_{\rm th}$.

Fig. 2. a) Optimal value $\theta^*_{\rm th}$ to be used for different SNR values and different DFT sizes M for the LOS and NLOS synthetic channels. b) Sensitivity of MSE to parameter $\theta_{\rm th}$. Results are obtained for 0 dB SNR for different DFT size values M.

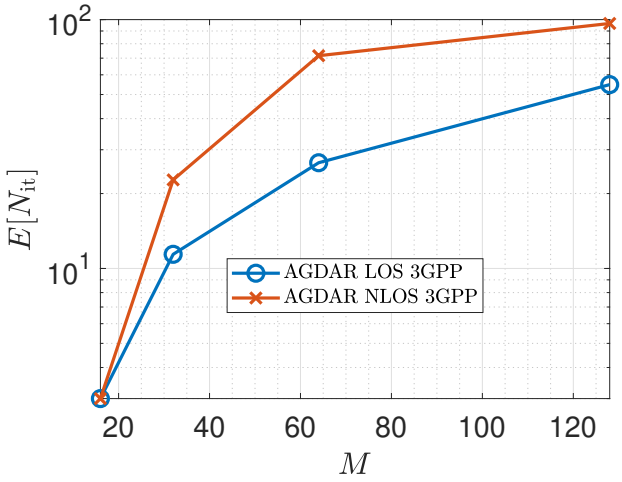


Fig. 3. Average number of iterations vs M for the AGDAR algorithm.

CS algorithms, respectively. The number of pilot subcarriers allocated for the channel estimation is $K \geq N_{\rm T}W$, and the optimal sequences introduced in Section II are used for pilots.

[b]

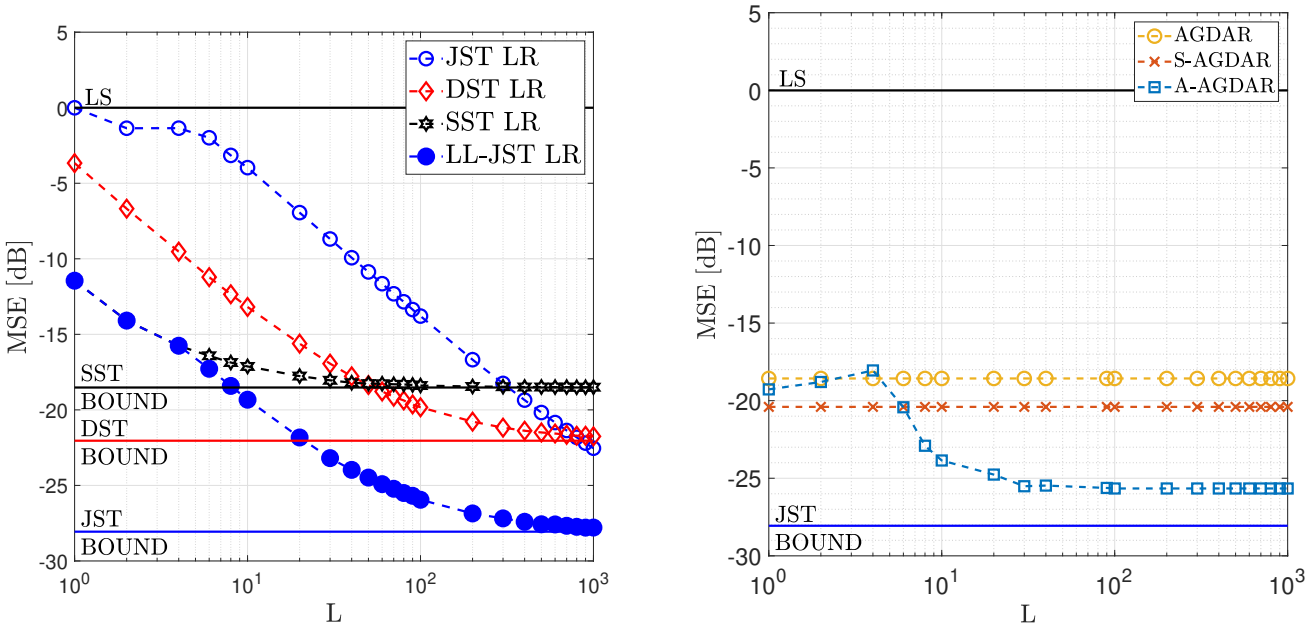
The MIMO-OFDM system consists of two vertical ULAs, with $N_{\rm R} = 16$ and $N_{\rm T} = 16$ elements[3] and inter-element spacing $d = \frac{\lambda}{2}$. Both the transmit and receive devices use polarized antenna elements with polarization angles of $-45°$ and $45°$ for each array, respectively, and antenna gains as specified in Section 7.3 of [30]. The signal bandwidth is set to 100 MHz and the cascade of the transmit and receive filters is a raised cosine with roll-off factor 0.2. The time-slot duration is 1 ms. The performance is evaluated over two different sets of channels: simplified deterministic channels and 3GPP stochastic channels, generated according to the specifications in [30], assuming a urban macro cell (UMa) propagation environment.

[3]Note that mmWave and THz links may require a larger number of antennas, but the relative performance between CS and LR method is found also with other array sizes. Moreover, the considered scenario may also represent the equivalent digital channel when both the transmitter and the receiver are equipped with hybrid analog/digital transceivers.

A. CS Parameters Selection

The effectiveness of the channel estimation based on the CS approaches partially depends on the selection of its parameters. In particular, we must properly select ρ, the weight of the norm-1 term, M, the DFT size, and $\theta_{\rm th}$, the threshold of S-AGDAR. The choice of the optimal value of ρ (common to AGDAR, S-AGDAR, and A-AGDAR) is a known problem in literature [31]–[33]. Among the different approaches we consider the exhaustive search of the optimal value, computing the MSE over a set of testing channels. Fig. 1(a) shows the optimal value ρ^* for different SNR values, different DFT sizes M, and both the line-of-sight (LOS) and non-line-of-sight (NLOS) simple channel models described in Section V-B, assuming $M_T = M_R = M$. We notice that the optimal ρ is linearly decreasing with the SNR. Therefore, given the knowledge of the SNR, the optimal ρ can be obtained by interpolation from a look-up table built off-line. We notice that a high weight of the norm-1 path leads to worse performance in terms of MSE, as we are imposing an amount of sparsity, which is not present in the original channel. Recalling that, as outlined in Section IV-B, paths may not fall on the DFT grid, we explore different values of M. Figs. 1 and 2 show the MSE obtained with different DFT sizes M, and we notice that increasing M leads to a lower MSE. Furthermore, from Fig. 1(b), we notice that, by setting $M = 64$, all the channel paths are properly estimated and we do not benefit from a higher DFT precision. From Fig. 1(b) we note that errors up to 20% on the value of ρ lead to a performance degradation of only 2 dB (from -20 to -18 dB) in MSE, thus showing that AGDAR methods are reasonably robust to a wrong choice of ρ. Note however that the optimal ρ is also related to the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSAC.2020.3000889, IEEE Journal on Selected Areas in Communications



(a) LR (b) CS

Fig. 4. MSE vs L for a) LR and b) CS channel estimation algorithms over the simplified NLOS channel with white noise. $M = 128$.

specific channel model. Still, from Fig. 1(a) we note that LOS and NLOS models have approximately the same optimal ρ, thus the choice of ρ is (in this case) robust also with respect to the channel model.

Parameter θ_{th} controls instead whether an entry in the estimated virtual channel matrix is to be considered as noise or as useful channel. Its value is chosen in order to minimize the MSE, through simulations. Fig. 2(b) shows the optimal value θ^*_{th} vs SNR for different DFT sizes for both the LOS and NLOS synthetic channels. Note that θ^*_{th} is not reported for all SNR values: its absence in the figure means that AGDAR is outperforming S-AGDAR. This occurs when the DFT size is small and we do not obtain a precise location of channel paths by sparsification. As for ρ^*, the optimal θ^*_{th} may be obtained by interpolation from a pre-built look up table. Fig. 2(b) shows the MSE vs the threshold parameter θ_{th} for different DFT sizes $M_T = M_R = M$, using the optimal parameter ρ^*. By comparing results in Figs. 1 and 2 for the same value of M, we note that a proper selection of the channel paths and a successive estimation of their gains via LS effectively decreases the MSE. This confirms that the AGDAR algorithm estimates spurious channels due to the use of the norm 1.

As outlined in Section IV-E, the computational complexity of the AGDAR solutions depends on the number of iterations N_{it}, which is common to all AGDAR algorithms. Fig. 3 shows the average number of AGDAR iterations vs the DFT size M, assuming $M_T = M_R = M$, for the different channel models with white noise. We notice that, as the DFT size grows, the number of iterations needed to reach convergence increases. Indeed, for a larger DFT size, the channel has more entries in the angular domain, therefore more variables must be optimized by the gradient-descent algorithm, regardless of the considered channel model.

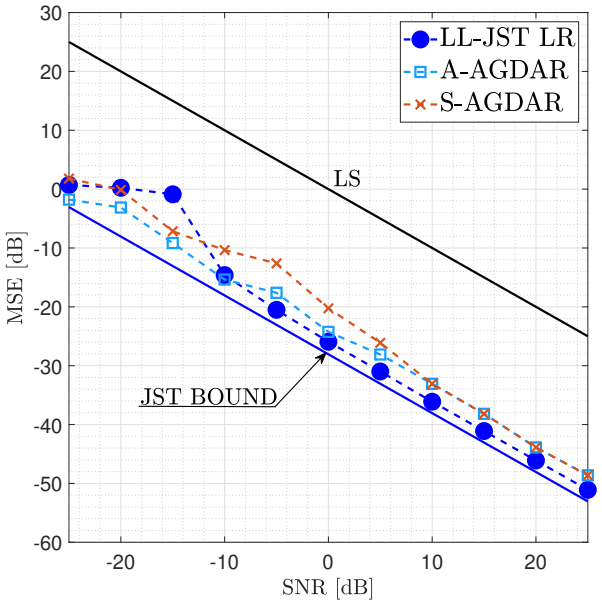
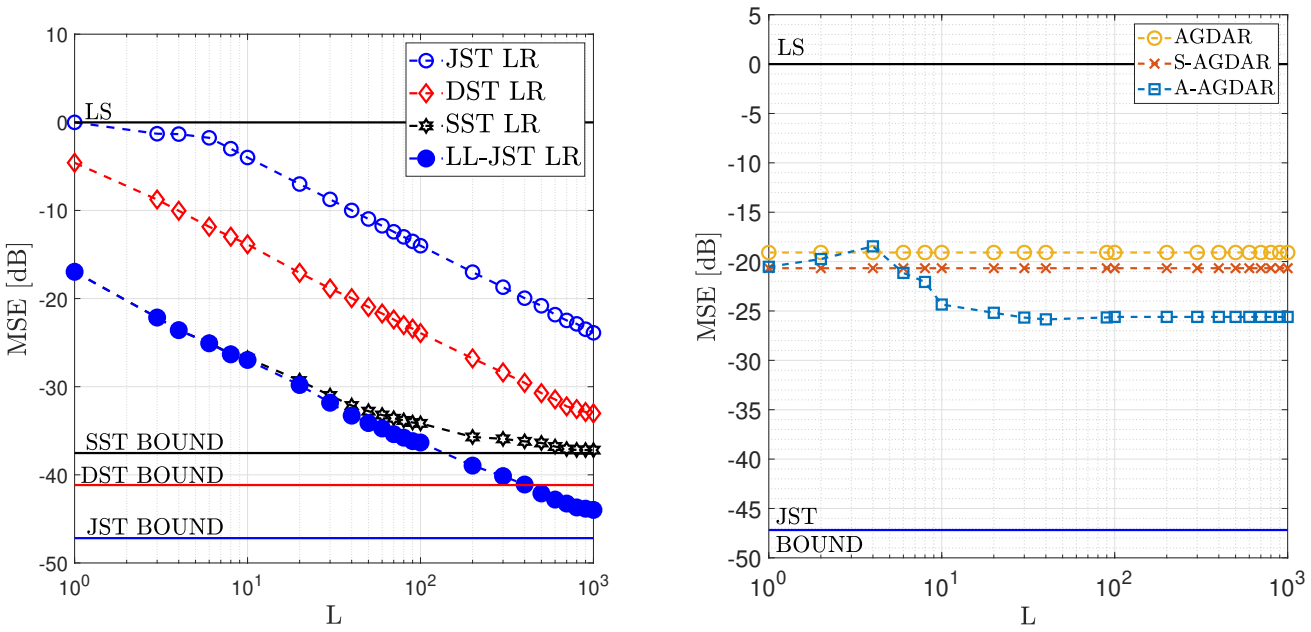


Fig. 5. MSE vs SNR for the best LR, S-AGDAR and A-AGDAR channel estimation algorithms over the simplified NLOS channel, for $M = 128$ and $L_{A-AGDAR} = 30$.

B. Simplified Channel Scenario

We first analyze the algorithms' performance over a simplified channel model characterized by four paths, both in LOS and NLOS propagation conditions.The four paths have

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSAC.2020.3000889, IEEE Journal on Selected Areas in Communications



(a) LR

(b) CS

Fig. 6. MSE vs L for a) LR and b) CS channel estimation algorithms over the simplified LOS channel with colored noise for $M = 128$.

elevation angles $\left(\psi_p^{\mathrm{RX}}, \psi_p^{\mathrm{TX}}\right)$ (115.17,64.83), (145.17,94.83), (85.17,94.83), and (85.17,34.83) [deg], for $p = 1, \ldots, 4$, respectively, compatible with a base station placed at $(0, 0, 25)$ and a user at $(0, 50, 1.5)$ in 3D Cartesian coordinates. With this choice of angles, the spatial diversity orders become $r_{\mathrm{S,RX}} = 3, r_{\mathrm{S,TX}} = 3$, and $r_{\mathrm{S}} = 4$, leading to a low-rank channel structure. To account for the temporal sparsity, path delays are set as $\tau_p = (p-1)T$, $p = 1, \ldots, 4$. Being T the temporal resolution of the system, the temporal diversity order is $r_{\mathrm{T}} = 4$, while $W = 10$. Path powers are compatible with the distributions for the UMa channels in [30], with a delay spread of $DS = 30$ ns and delay scaling parameter $r_\tau = 2.3$ and 2.5, for NLOS and LOS, respectively. Furthermore, our model includes antenna gains, per-cluster shadowing and, for the LOS case, the Rice component with factor 9 dB. Powers are then set as $\mathbf{\Omega}_{\mathrm{NLOS}} = [0.319, 0.272, 0.227, 0.182]$, for the simplified NLOS scenario, and $\mathbf{\Omega}_{\mathrm{LOS}} = [0.924, 0.030, 0.025, 0.021]$, for the simplified LOS scenario. Notice that the provided results are consistent as long as the rank values are preserved, while variations of the channel powers affect the MSE for $SNR < 0$, as the asymptotic bounds for high SNR values depend only on the ratio between the rank values and the channel entries (see the Appendix for a proof).

Fig. 4(a) shows the MSE of the LR algorithms against L, for SNR $= 0$ dB over the simplified NLOS channel and without interference. The performance is compared to the asymptotic bound derived in Section III-C, assuming optimal rank selection. The JST algorithm has been included in the analysis for comparison purposes with the existing literature. We notice that all algorithms attain the corresponding lower bound, outperforming the conventional unconstrained LS estimator. However, the four estimators converge at different rates: a larger number of slots is required by the JST method with respect to the other methods, due to the higher number of channel parameters to be estimated in the joint ST domain. On the other hand, the LL-JST method reduces the dimensions of the estimation problem as the ST bases are identified in the projected domain selected by the first stage of decoupled filtering. It follows that LL-JST outperforms all the other LR algorithms for $L > 4$, while at lower values it is interestingly upperbounded by the SST algorithm. This holds as long as the channel rank r is properly estimated: selecting a wrong number of channel bases leads to a higher MSE than that of the SST algorithm. Within $L = 100$ slots, the MSE of LL-JST is 2 dB above the asymptotic bound. Moreover, we observe that, despite of the better asymptotic performance, DST outperforms SST only for $L > 50$ slots, while JST needs more than $L = 100$ slots to outperform both SST and DST, which can be critical for mmWave/THz channels.

Fig. 4(b) shows the performance of the CS algorithms, for the same scenario of Fig. 4(a). The figure includes the JST lower bound, which holds for any estimator operating on the considered channel model. We note a very fast convergence within $L = 40$ for A-AGDAR and even faster for the other solutions, that however achieve a higher MSE. We here recall that the performance of the S-AGDAR algorithm is independent of L, as the set of non-zero indexes in the virtual channel domain is obtained independently in each slot. Therefore, the MSE is constant with respect to L for the S-AGDAR algorithm. For the A-AGDAR algorithm instead we observe that for $L < 6$ the MSE can be higher than S-AGDAR, since the estimate of $\hat{P}$ is inaccurate and affects L channel estimates, as discussed in Section IV-D. Comparing Figs. 4(a) and 4(b) we note that for $L = 100$ both the LL-JST and the A-

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSAC.2020.3000889, IEEE Journal on Selected Areas in Communications

AGDAR techniques achieve the same MSE, while A-AGDAR shows a faster convergence, thus we do not show the MSE for $L > 100$ in Fig. 4(a).

We now investigate the behavior of the estimation methods as a function of the SNR. Fig. 5 reports the MSE vs SNR for the best LR and the S-AGDAR and A-AGDAR channel estimation algorithms. Also in this case the two methods have a similar performance and provide an MSE reduction of more than 20 dB with respect to the LS estimate. Moreover, when comparing S-AGDAR and A-AGDAR we observe that for $SNR < 10$ dB A-AGDAR has a significantly reduced MSE, as it averages out the noise before detecting paths in the angular domain; at higher SNR the gap between S-AGDAR and A-AGDAR is reduced, as they both detect the correct path positions on the discrete DFT grid.

Fig. 6(a) shows the MSE of LR algorithms over the simplified LOS channel, assuming the presence of 3 interferers paths impinging the receiving array, so that one interferer path is superimposed to one channel path, i.e., has the same DOA as the channel's path, while the other interferers and channel's path have different DOAs. We assumed $\mathbf{Q}$ known, in order to perform the whitening and de-whitening operations. The MSE vs L is shown for SINR $= 0$ dB. With respect to Fig. 4(a), we observe lower asymptotic bounds, as most of the interference is concentrated within directions that do not correspond to channel's DOAs. Since the convergence rate is similar, we observe that it takes more slots for all the algorithms to converge and only SST is at 3 dB from its lower bound after $L = 100$ slots, while only the LL-JST algorithm reaches the JST bound, with a longer latency of $L = 1000$ slots.

Fig. 6(b) shows the performance of the CS algorithms, for the same scenario of Fig. 6(a). Note that in this case the CS algorithms do not take into account that the noise is colored, i.e., has different powers at the various entries in the angular domain. Also in this case the convergence is fast, and both S-AGDAR and A-AGDAR outperform AGDAR. We see that the A-AGDAR algorithm has a lower MSE than the S-AGDAR algorithm, due to fact that also in this case angular domain channel entries are better located when considering an increasing number of slots. Overall, when comparing Figs. 6(a) and 6(b) we observe that both the SST and LL-JST methods outperform the best CS solution for any $L > 10$, reducing the MSE by about 10 dB at $L = 100$. This is due to the fact that AGDAR does not consider the presence of the interferers, while they are accounted for by the LR methods through $\hat{\mathbf{C}}$.

C. 3GPP Channel Model

We now show the MSE of the proposed estimation algorithms over the 3GPP stochastic channel, generated according to the specifications in [30] assuming a UMa propagation environment, white noise, and the worst-case synchronization error, i.e., the first arrival placed between two sampling instants (here we set $\tau_1 = 5.5T$).

Fig. 7 shows the MSE for both LR and CS algorithms vs L. As for the simplified scenarios, all LR algorithms have a significant gain with respect the LS estimator, but

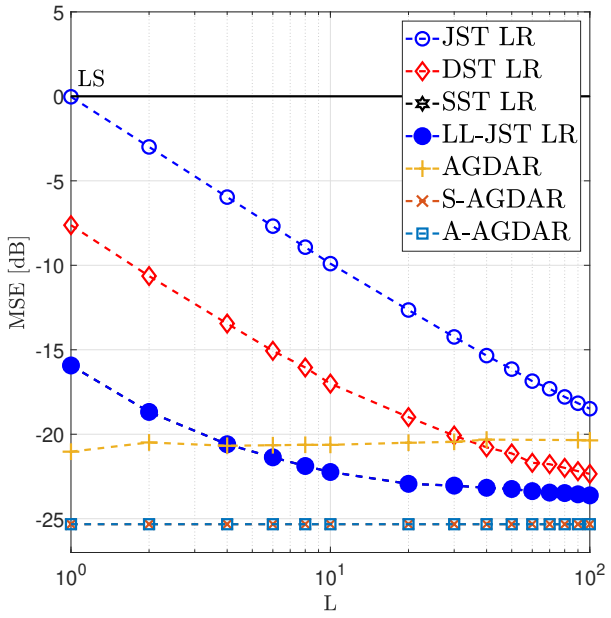


Fig. 7. MSE vs L for the LR and AGDAR channel estimation algorithms over the 3GPP LOS channel, for $M = 128$.

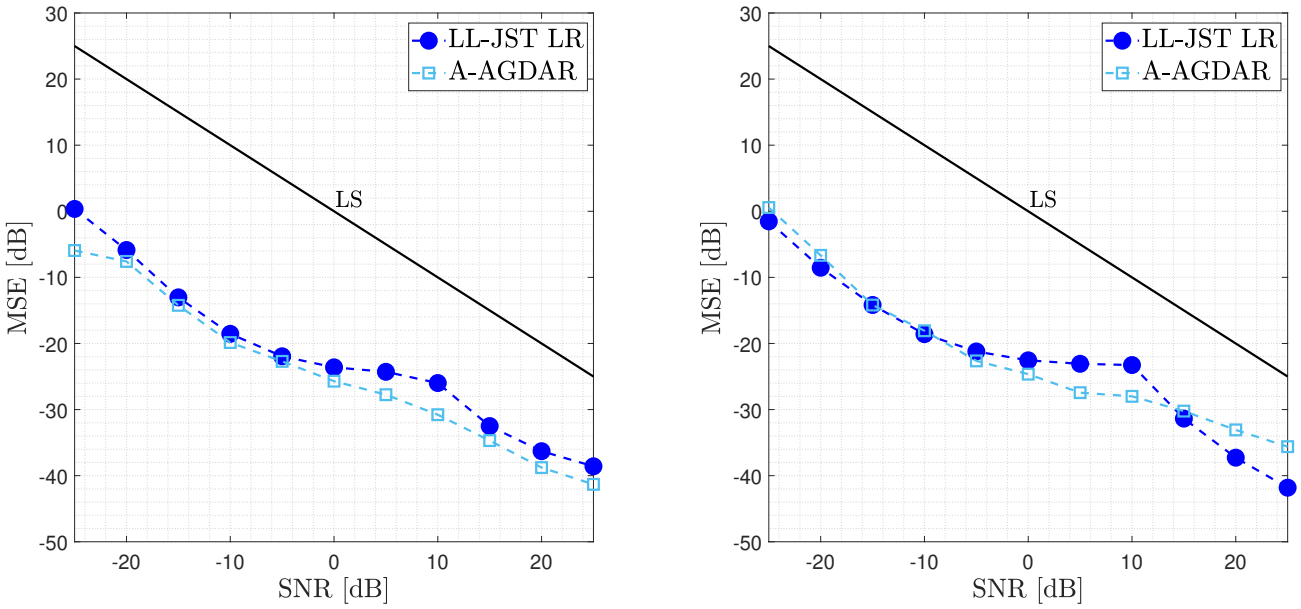
faster rate of convergence due to the low-rank nature of the channel, despite of its numerous clusters and rays. In this scenario, the LL-JST does not gain significantly with respect to the SST algorithm, but yields a lower MSE than both DST and JST, as in all of the other considered scenarios. The figure shows that synchronization errors have a negligible impact on performance. About the CS methods, we observe an even faster convergence (within $L = 2$) with respect to the simplified channel model.

D. Algorithm Comparison

The comparison among the proposed algorithms should take into account various metrics. Here we focus on the estimation MSE, the computational complexity, and the latency required for convergence, i.e., L.

In terms of MSE, from Fig. 7 we conclude that the two approaches have a similar performance in the considered scenario, with MSE reductions at $SNR = 0$ dB that range between 20 and 25 dB, with respect to the common baseline LS. This is further confirmed by Fig. 8, showing the estimate MSE as a function of the channel SNR, for the best performing LR and CS algorithms. Also in this case we note that both proposed algorithms considerably reduce the MSE with respect to the LS estimate for any value of SNR, by a factor between 15 dB (at $SNR = 10$ dB) and 30 dB (at $SNR = -10$ dB). Also, below $L = 100$ slots, A-AGDAR outperforms LL-JST. By Fig. 8(b) we observe that the improvement of the proposed methods holds also for asymmetric MIMO configurations (here $N_T = 8$ and $N_R = 32$). We notice that A-AGDAR exhibits a smoother behavior than LL-JST in both plots of Fig. 8. However, we also notice that the A-AGDAR MSE decreases more slowly

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSAC.2020.3000889, IEEE Journal on Selected Areas in Communications



(a) $N_{\mathrm{T}} = N_{\mathrm{R}} = 16$ (b) $N_{\mathrm{T}} = 8$ and $N_{\mathrm{R}} = 32$

Fig. 8. MSE vs SNR comparison of the best LR and AGDAR channel estimation algorithms over 3GPP LOS Channel with different array sizes, for $M = 128$ and $L = 100$ (10) slots for LL-JST (A-AGDAR).

at low SNRs in Fig. 8(a), due to the effects of noise and the non-linearity of CS.

The computational complexity of the algorithms for the results of Fig. 8 with $SNR = 0$ dB, $L = 100$ for LR algorithms and $L = 2$ for A-AGDAR (providing the same MSE) is reported in Table III, in terms of multiplications per slot, normalized to the multiplications required by one LS estimate, i.e., $\mathcal{C}_{\mathrm{LS}} = \left(N_{\mathrm{R}} \left(N_{\mathrm{T}} W\right)^2\right)$. The computational complexity of the LR methods is obtained using of a subspace tracking algorithm such as [34], in order to reduce both the latency due to the batch processing of L slots and the costs of the eigenvalue decomposition. When spatially-correlated noise is considered, (22) requires an additional cost of $O_{LR} = O(N_{\mathrm{R}}^3) + O(W N_{\mathrm{T}} N_{\mathrm{R}}^2)$ for the computation of both the whitening factors for $\hat{\boldsymbol{C}}_{\mathrm{LS}}$ and the overall filtering operation; this complexity can be significantly reduced when the interference has low-rank spatial correlation (e.g., in case of few directional interferers), by exploiting efficient approaches as subspace tracking. Note that the JST complexity is ruled by the overestimation of $\hat{r}$ (here $\hat{r} = 100$, as $L << W N_{\mathrm{T}} N_{\mathrm{R}}$), while for all other cases the selected ranks are unitary, leading to similar computational costs for the DST, SST, and LL-JST methods, with a slightly heavier load for SST and LL-JST, when compared to DST. Indeed, the computational advantage of the former algorithms becomes evident only when $\hat{r}_{\mathrm{S}} >> \hat{r}_{\mathrm{S,TX}} + \hat{r}_{\mathrm{S,RX}}$. For the CS method, formulas have been obtained in Section IV-E. It can be observed that for low-ranks channels, as in mmWave/THz systems, the complexity grows only linearly with the ST channel dimensions, when using an efficient subspace-tracking implementation. We notice that, in general, the CS-based methods require a significantly larger number of operations than the LR-based ones. In terms of convergence latency, from Fig. 8 we note that CS methods have a very fast convergence ($L < 10$) with respect to LR methods that require 10 times more LS estimates before convergence. Still, we observe that LR methods can capture also the spatial correlation of noise (or interference). Both LR and CS methods require the tuning of channel-model-dependent parameters, namely the rank and the norm-1 coefficient (and the threshold for S-AGDAR), respectively. Notice that in dense multipath scenarios, even if still grouped into few clusters as for 3GPP model, the rank estimation via MDL may over/under-estimate the ranks. This can be seen in 8(b) where some fluctuations are introduced in the LR method, peaking around the SNR=10 dB. In summary, i) CS methods are converging faster than LR; ii) the CS methods depends on the array manifold, while LR methods are independent on the array calibration; iii) the complexity of the LR methods is reduced with respect to the CS methods; iv) the LR methods require accurate estimate of the rank, while for CS methods processing parameters (ρ and θ_{th}) must be optimized.

VI. Conclusions

We have proposed new estimation methods for the mmWave/THz channel, based on LR and CS approaches. Their new design takes into account many peculiarities of the MIMO channel at those frequencies. Estimated channel statistics over a few transmission periods are exploited to improve both approaches over previously existing techniques. Extensive numerical results on both simple and well-established 3GPP models show the merit of the proposed solutions, providing gains up to 30 dB over the LS channel estimate, with gains from 5 to 25 dB over existing LR and CS solutions. The solutions have also been compared in terms of achieved MSE,

computational complexity, and convergence latency, concluding that a different trade-off among these metrics is achieved by each solution.

APPENDIX

The MSE bound for the LR estimators is here derived assuming that $\mathbf{R}_{\mathrm{BB}}$ is diagonal. The channel covariance reduces to:

$$\boldsymbol{C}_{\mathrm{LS}} = \gamma \underbrace{\left(\mathbf{I}_W \otimes \mathbf{I}_{N_{\mathrm{T}}} \otimes \mathbf{Q}^{\mathrm{T}}\right)}_{\bar{\boldsymbol{C}}_{\mathrm{LS}}}, \tag{55}$$

where $\gamma = 1/(\sigma_x^2 K)$. The closed form expressions of the noise and bias MSE contributions can be obtained by plugging (55) into (36) and (37), respectively.

Let us first consider the noise contribution and simplify the notation with $\Phi\left(\mathbf{A}, \mathbf{B}\right) = \mathrm{tr}\left(\mathbf{B}^{\frac{\mathrm{H}}{2}}\mathbf{A}\mathbf{B}^{\frac{1}{2}}\right)$, we get:

$$\mathrm{MSE}_n(\hat{\mathbf{r}}) = \gamma\Phi\left(\boldsymbol{\Pi}_{\mathrm{LR},\hat{\mathbf{r}}}, \bar{\boldsymbol{C}}_{\mathrm{LS}}\right). \tag{56}$$

By substituting the projectors of Table II we obtain, after some rearrangements,

$$\mathrm{MSE}_n^{\mathrm{JST}}(\hat{\mathbf{r}}) = \gamma\Phi\left(\boldsymbol{\Pi}_{\mathrm{LR},\hat{r}}, \bar{\boldsymbol{C}}_{\mathrm{LS}}\right), \tag{57a}$$

$$\mathrm{MSE}_n^{\mathrm{DST}}(\hat{\mathbf{r}}) = \gamma\mathrm{tr}\left(\boldsymbol{\Pi}_{\mathrm{T},\hat{r}_{\mathrm{T}}}\right)\Phi\left(\boldsymbol{\Pi}_{\mathrm{S},\hat{r}_{\mathrm{S}}}, \mathbf{I}_{N_{\mathrm{T}}} \otimes \mathbf{Q}^{\mathrm{T}}\right), \tag{57b}$$

$$\mathrm{MSE}_n^{\mathrm{SST}}(\hat{\mathbf{r}}) = \gamma\mathrm{tr}\left(\boldsymbol{\Pi}_{\mathrm{T},\hat{r_{\mathrm{T}}}}\right)\mathrm{tr}\left(\boldsymbol{\Pi}_{\mathrm{S},\hat{r}_{\mathrm{S,TX}}}\right)\Phi\left(\boldsymbol{\Pi}_{\mathrm{S},\hat{r}_{\mathrm{S,RX}}}, \mathbf{Q}^{\mathrm{T}}\right), \tag{57c}$$

which simplify in presence of white noise, i.e., when $\mathbf{Q} = \sigma_n^2\mathbf{I}_{N_{\mathrm{R}}}$, into

$$\mathrm{MSE}_n^{\mathrm{JST}}(\hat{\mathbf{r}}) = \gamma\sigma_n^2\hat{r}, \tag{58a}$$

$$\mathrm{MSE}_n^{\mathrm{DST}}(\hat{\mathbf{r}}) = \gamma\sigma_n^2\hat{r}_{\mathrm{T}}\hat{r}_{\mathrm{S}}, \tag{58b}$$

$$\mathrm{MSE}_n^{\mathrm{SST}}(\hat{\mathbf{r}}) = \gamma\sigma_n^2\hat{r}_{\mathrm{T}}\hat{r}_{\mathrm{S,RX}}\hat{r}_{\mathrm{S,TX}}. \tag{58c}$$

For the bias contribution, we can apply the cyclic-shift property of the trace operator, obtaining $\mathrm{MSE}_b(\hat{\mathbf{r}}) = \gamma\mathrm{tr}\left(\triangle\boldsymbol{\Pi}_{\mathrm{LR},\hat{\mathbf{r}}}\tilde{\mathbf{R}}_{\mathrm{ST}}\bar{\mathbf{C}}_{\mathrm{LS}}^{\frac{1}{2}}\bar{\mathbf{C}}_{\mathrm{LS}}^{\frac{\mathrm{H}}{2}}\right)$, which can be rearranged as:

$$\mathrm{MSE}_b(\hat{\mathbf{r}}) = \sum_{p=1}^{P}\Omega_p\mathrm{tr}\left(\triangle\boldsymbol{\Pi}_{\mathrm{LR},\hat{\mathbf{r}}}\left(\mathbf{R}_{\mathrm{T},p} \otimes \mathbf{R}_{\mathrm{S},p}^{\mathrm{TX}} \otimes \bar{\mathbf{R}}_{\mathbf{S},p}^{\mathrm{RX}}\right)\right), \tag{59}$$

where $\bar{\mathbf{R}}_{\mathbf{S},p}^{\mathrm{RX}} = \mathbf{Q}^{-\frac{*}{2}}\mathbf{R}_{\mathbf{S},p}^{\mathrm{RX}}\mathbf{Q}^{\frac{*}{2}}$. It can be shown that for the DST channel model, the projector can be computed as $\triangle\boldsymbol{\Pi}_{\mathrm{LR},\hat{r}} = \triangle\boldsymbol{\Pi}_{\mathrm{T},\hat{r}_{\mathrm{T}}}^{\mathrm{T}} \otimes \triangle\boldsymbol{\Pi}_{\mathrm{S},\hat{r}_{\mathrm{S}}} - \mathbf{I}_{\mathrm{W}} \otimes \triangle\boldsymbol{\Pi}_{\mathrm{S},\hat{r}_{\mathrm{S}}} - \triangle\boldsymbol{\Pi}_{\mathrm{T},\hat{r}_{\mathrm{T}}}^{\mathrm{T}} \otimes \mathbf{I}_{N_{\mathrm{R}}N_{\mathrm{T}}}$, where $\triangle\boldsymbol{\Pi}_{\mathrm{T},\hat{r}_{\mathrm{T}}} = \boldsymbol{\Pi}_{\mathrm{T},r_{\mathrm{T}}} - \boldsymbol{\Pi}_{\mathrm{T},\hat{r}_{\mathrm{T}}}$ and $\triangle\boldsymbol{\Pi}_{\mathrm{S},\hat{r}_{\mathrm{S}}} = \boldsymbol{\Pi}_{\mathrm{S},r_{\mathrm{S}}} - \boldsymbol{\Pi}_{\mathrm{S},\hat{r}_{\mathrm{S}}}$. This holds also for the SST channel model, where $\triangle\boldsymbol{\Pi}_{\mathrm{S},\hat{r}_{\mathrm{S}}}$ is replaced with $\triangle\boldsymbol{\Pi}_{\mathrm{S},\hat{r}_{\mathrm{S,TX}}}^{\mathrm{T}} \otimes \triangle\boldsymbol{\Pi}_{\mathrm{S},\hat{r}_{\mathrm{S,RX}}} - \mathbf{I}_{N_{\mathrm{T}}} \otimes \triangle\boldsymbol{\Pi}_{\mathrm{S},\hat{r}_{\mathrm{S,RX}}} - \triangle\boldsymbol{\Pi}_{\mathrm{S},\hat{r}_{\mathrm{S,TX}}}^{\mathrm{T}} \otimes \mathbf{I}_{N_{\mathrm{R}}}$, with $\triangle\boldsymbol{\Pi}_{\mathrm{S},\hat{r}_{\mathrm{S,TX}}} = \boldsymbol{\Pi}_{\mathrm{S},r_{\mathrm{S,TX}}} - \boldsymbol{\Pi}_{\mathrm{S},\hat{r}_{\mathrm{S,TX}}}$ and $\triangle\boldsymbol{\Pi}_{\mathrm{S},\hat{r}_{\mathrm{S,RX}}} = \boldsymbol{\Pi}_{\mathrm{S},r_{\mathrm{S,RX}}} - \boldsymbol{\Pi}_{\mathrm{S},\hat{r}_{\mathrm{S,RX}}}$. The final expressions for the bias are:

$$\mathrm{MSE}_b^{\mathrm{JST}}(\hat{\mathbf{r}}) = \sum_{p=1}^{P}\Omega_p\mathrm{tr}\left(\triangle\boldsymbol{\Pi}_{\mathrm{LR},\hat{r}}\left(\mathbf{R}_{\mathrm{T},p} \otimes \underbrace{\mathbf{R}_{\mathrm{S},p}^{\mathrm{TX}} \otimes \bar{\mathbf{R}}_{S,p}^{\mathrm{RX}}}_{\bar{\mathbf{R}}_{\mathrm{S},p}}\right)\right), \tag{60a}$$

$$\begin{aligned}\mathrm{MSE}_b^{\mathrm{DST}}(\hat{\mathbf{r}}) = \sum_{p=1}^{P}\Omega_p\Big(&-\mathrm{tr}\left(\triangle\boldsymbol{\Pi}_{\mathrm{T},\hat{r}_{\mathrm{T}}}^{\mathrm{T}}\mathbf{R}_{\mathrm{T},p} \otimes \bar{\mathbf{R}}_{\mathrm{S},p}\right)\\ &- \mathrm{tr}\left(\mathbf{R}_{\mathrm{T},p} \otimes \triangle\boldsymbol{\Pi}_{\hat{r}_{\mathrm{S}}}\bar{\mathbf{R}}_{\mathrm{S},p}\right)\\ &+\mathrm{tr}\left(\triangle\boldsymbol{\Pi}_{\mathrm{T},\hat{r}_{\mathrm{T}}}^{\mathrm{T}}\mathbf{R}_{\mathrm{T},p} \otimes \triangle\boldsymbol{\Pi}_{\hat{r}_{\mathrm{S}}}\bar{\mathbf{R}}_{\mathrm{S},p}\right)\Big),\end{aligned} \tag{60b}$$

$$\begin{aligned}\mathrm{MSE}_b^{\mathrm{SST}}(\hat{\mathbf{r}}) = \sum_{p=1}^{P}\Omega_p\Big(&-\mathrm{tr}\left(\triangle\boldsymbol{\Pi}_{\mathrm{T},\hat{r}_{\mathrm{T}}}^{\mathrm{T}}\mathbf{R}_{\mathrm{T},p} \otimes \bar{\mathbf{R}}_{\mathrm{S},p}\right)\\ &-\mathrm{tr}\left(\mathbf{R}_{\mathrm{T},p} \otimes \mathbf{P}\right) + \mathrm{tr}\left(\triangle\boldsymbol{\Pi}_{\mathrm{T},\hat{r}_{\mathrm{T}}}^{\mathrm{T}}\mathbf{R}_{\mathrm{T},p} \otimes \mathbf{P}\right)\Big),\end{aligned} \tag{60c}$$

where $\mathbf{P} = \triangle\boldsymbol{\Pi}_{\mathrm{S},\hat{r}_{\mathrm{S,TX}}}^{\mathrm{T}}\mathbf{R}_{\mathrm{S},p}^{\mathrm{TX}} \otimes \triangle\boldsymbol{\Pi}_{\mathrm{S},\hat{r}_{\mathrm{S,RX}}}\bar{\mathbf{R}}_{\mathrm{S},p}^{\mathrm{RX}} - \mathbf{R}_{\mathrm{S},p}^{\mathrm{TX}} \otimes \triangle\boldsymbol{\Pi}_{\mathrm{S},\hat{r}_{\mathrm{S,RX}}}\bar{\mathbf{R}}_{\mathrm{S},p}^{\mathrm{RX}} - \triangle\boldsymbol{\Pi}_{\mathrm{S},\hat{r}_{\mathrm{S,TX}}}^{\mathrm{T}}\mathbf{R}_{\mathrm{S},p}^{\mathrm{TX}} \otimes \bar{\mathbf{R}}_{\mathrm{S},p}^{\mathrm{RX}}$.

REFERENCES

[1] T. S. Rappaport, J. N. Murdock, and F. Gutierrez, "State of the art in 60-GHz integrated circuits and systems for wireless communications," *Proceedings of the IEEE*, vol. 99, no. 8, pp. 1390–1436, Aug. 2011.

[2] S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter-wave cellular wireless networks: Potentials and challenges," *Proceedings of the IEEE*, vol. 102, no. 3, pp. 366–385, Mar. 2014.

[3] 3GPP TR 38.901 V14.3.0, "5G; study on channel model for frequencies from 0.5 to 100 GHz," 3GPP, Tech. Rep., 2018.

[4] Z. Zhang *et al.*, "6G wireless networks: Vision, requirements, architecture, and key technologies," *IEEE Vehicular Technology Mag.*, vol. 14, no. 3, pp. 28–41, Sep. 2019.

[5] I. Akyildiz, J. Jornet, and C. Han, "Terahertz band: Next frontier for wireless communications," *Physical Communication*, vol. 12, pp. 16–32, 2014.

[6] C. Han, A. O. Bicen, and I. F. Akyildiz, "Multi-ray channel modeling and wideband characterization for wireless communications in the Terahertz band," *IEEE Trans. on Wireless Communications*, vol. 14, no. 5, pp. 2402–2412, May 2015.

[7] R. W. Heath *et al.*, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 3, pp. 436–453, April 2016.

[8] P. Roblin, M. Rawat, and V. Ratnasamy, "RF front-end flexibility, self-calibration, and self-linearization: Characterizing and mitigating nonlinearities in SDR MIMO systems for concurrent multiband operation," *IEEE Microwave Magazine*, vol. 19, no. 2, pp. 49–61, March 2018.

[9] C. K. Anjinappa *et al.*, "Channel estimation in mmWave hybrid MIMO system via off-grid Dirichlet kernels," 2019, 1907.04427.

[10] M. Jian, F. Gao, Z. Tian, S. Jin, and S. Ma, "Angle-domain aided UL/DL channel estimation for wideband mmwave massive MIMO systems with beam squint," *IEEE Trans. on Wireless Commun.*, vol. 18, no. 7, pp. 3515–3527, July 2019.

[11] G. Soatti *et al.*, "Low-rank channel and interference estimation in mm-Wave massive antenna arrays," in *Proc. European Signal Processing Conf. (EUSIPCO)*, Sep. 2018, pp. 922–926.

[12] J. Rodriguez-Fernandez and N. Gonzalez-Prelcic, "Channel estimation for hybrid mmwave MIMO systems with CFO uncertainties," *IEEE Trans. on Wireless Communications*, pp. 1–1, 2019.

[13] E. Vlachos, G. C. Alexandropoulos, and J. Thompson, "Wideband MIMO channel estimation for hybrid beamforming millimeter wave systems via random spatial sampling," *IEEE Journal of Selected Topics in Signal Processing*, vol. 13, no. 5, pp. 1136–1150, Sep. 2019.

[14] J. Zhang and M. Haardt, "Channel estimation and training design for hybrid multi-carrier MmWave massive MIMO systems: The beamspace ESPRIT approach," in *Proc. 25th European Signal Processing Conf. (EUSIPCO)*, Aug 2017, pp. 385–389.

[15] H. Ghauch *et al.*, "Subspace estimation and decomposition for large millimeter-wave MIMO systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 3, pp. 528–542, April 2016.

[16] Y. K. andothers, "Feasibility of mobile cellular communications at millimeter wave frequency," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 3, pp. 589–599, April 2016.

[17] M. Mezzavilla *et al.*, "End-to-end simulation of 5G mmWave networks," *IEEE Commun. Surveys Tutorials*, vol. 20, no. 3, pp. 2237–2263, thirdquarter 2018.

TABLE III
COMPUTATIONAL COMPLEXITY OF THE CHANNEL ESTIMATION METHODS

Method	Computational Complexity	3GPP LOS
JST	$O(\hat{r}^2 N_{\mathrm{TOT}})$	$1 \cdot 10^2$
DST	$O(\hat{r}_{\mathrm{S}}^2 N_{\mathrm{T}} N_{\mathrm{R}}) + O(\hat{r}_{\mathrm{T}}^2 W) + O((\hat{r}_{\mathrm{T}} + \hat{r}_{\mathrm{S}}) N_{\mathrm{TOT}})$	$2 \cdot 10^{-2}$
SST	$O_{\mathrm{SST}} = O(\hat{r}_{\mathrm{S,TX}}^2 N_{\mathrm{T}}) + O(\hat{r}_{\mathrm{S,RX}}^2 N_{\mathrm{R}})$ + $O(\hat{r}_{\mathrm{T}}^2 W) + O((\hat{r}_{\mathrm{T}} + \hat{r}_{\mathrm{S,RX}} + \hat{r}_{\mathrm{S,TX}}) N_{\mathrm{TOT}})$	$3 \cdot 10^{-2}$
LLJST	$O_{\mathrm{SST}} + O(\hat{r}^2 \hat{r}_{\mathrm{S,TX}} \hat{r}_{\mathrm{S,RX}} \hat{r}_{\mathrm{T}})$	$3 \cdot 10^{-2}$
AGDAR	$O(3M_{\mathrm{tot}} N_{\mathrm{it}} + 2N_{\mathrm{it}} M_{\mathrm{tot}} \log_2 M_{\mathrm{tot}})$	$1.3 \cdot 10^3$
S-AGDAR	$O(3M_{\mathrm{tot}} N_{\mathrm{it}} + 2N_{\mathrm{it}} M_{\mathrm{tot}} \log_2 M_{\mathrm{tot}} + \hat{P}^3)$	$1.3 \cdot 10^3$
A-AGDAR	$O(L(3M_{\mathrm{tot}} N_{\mathrm{it}} + 2N_{\mathrm{it}} M_{\mathrm{tot}} \log_2 M_{\mathrm{tot}}) + \hat{P}^3)$	$2.6 \cdot 10^3$

[18] M. Nicoli, O. Simeone, and U. Spagnolini, "Multislot estimation of fast-varying space-time communication channels," *IEEE Trans. on Signal Proc.*, vol. 51, no. 5, pp. 1184–1195, May 2003.

[19] M. Nicoli and U. Spagnolini, "Subspace-methods for space-time processing," in *Proc. Smart Antennas: State of the Art (Eurasip Book Series on Signal Proc. and Communications)*. Hindawi, 2005, ch. 3, pp. 27–52.

[20] M. Cicerone, O. Simeone, and U. Spagnolini, "Channel estimation for MIMO-OFDM systems by modal analysis/filtering," *IEEE Trans. on Commun.*, vol. 54, no. 11, pp. 2062–2074, Nov 2006.

[21] O. Simeone and U. Spagnolini, "Lower bound on training-based channel estimation error for frequency-selective block-fading Rayleigh mimo channels," *IEEE Trans. on Signal Processing*, vol. 52, no. 11, pp. 3265–3277, Nov 2004.

[22] H. Soleimani, D. De Donno, and S. Tomasin, "Mm-Wave channel estimation with accelerated gradient descent algorithms," *EURASIP Jour. on Wireless Commun. and Net.*, vol. 2018, no. 1, p. 272, 2018.

[23] O. E. Ayach, R. W. Heath, S. Abu-Surra, S. Rajagopal, and Z. Pi, "Low complexity precoding for large millimeter wave MIMO systems," in *Proc. IEEE Int. Conf. on Communications (ICC)*, June 2012, pp. 3724–3729.

[24] C. Lin and G. Y. Li, "Indoor Terahertz communications: How many antenna arrays are needed?" *IEEE Trans. on Wireless Commun.*, vol. 14, no. 6, pp. 3097–3107, June 2015.

[25] I. Barhumi, G. Leus, and M. Moonen, "Optimal training design for MIMO OFDM systems in mobile wireless channels," *IEEE Trans. on Signal Proc.*, vol. 51, no. 6, pp. 1615–1624, 2003.

[26] Y. Li, "Optimum training sequences for OFDM systems with multiple transmit antennas," in *Proc. IEEE Global Telecommunications Conf. (GLOBECOM)*, vol. 3, 2000, pp. 1478–1482.

[27] C.-N. Chuah, J. M. Kahn, and D. Tse, "Capacity of multi-antenna array systems in indoor wireless environment," in *Proc. IEEE Global Telecommunications Conf. (GLOBECOM)*, vol. 4, Nov 1998, pp. 1894–1899 vol.4.

[28] D.-S. Shiu, G. J. Foschini, M. J. Gans, and J. M. Kahn, "Fading correlation and its effect on the capacity of multielement antenna systems," *IEEE Trans. on Commun.*, vol. 48, no. 3, pp. 502–513, March 2000.

[29] D. Tse and P. Viswanath, *Fundamentals of wireless communication*. Cambridge university press, 2005.

[30] ETSI TR 138 900 V14.2.0, "Study on channel model for frequency spectrum above 6 GHz," 3GPP, Tech. Rep., Jun. 2017.

[31] H. Zou, T. Hastie, R. Tibshirani *et al.*, "On the "degrees of freedom" of the LASSO," *The Annals of Statistics*, vol. 35, no. 5, pp. 2173–2192, 2007.

[32] S. Kaufman and S. Rosset, "When does more regularization imply fewer degrees of freedom? sufficient conditions and counterexamples," *Biometrika*, vol. 101, no. 4, pp. 771–784, 2014.

[33] A. Lykou and I. Ntzoufras, "On bayesian LASSO variable selection and the specification of the shrinkage parameter," *Statistics and Computing*, vol. 23, no. 3, pp. 361–390, 2013.

[34] P. Strobach, "Low-rank adaptive filters," *IEEE Trans. on Signal Proc.*, vol. 44, no. 12, pp. 2932–2947, Dec 1996.

Alessandro Brighente received the master degree in Telecommunications Engineering from the University of Padova in 2016. From January to September 2017 he was a research fellow at the University of Padova. From June to November 2019 he was visitor researcher at Nokia Bell Labs, Stuttgart. He is now pursuing its Ph.D. degree in Information Engineering at the University of Padova. His current research interests include signal processing for millimeter-wave systems and machine learning for wireless communications.

Mattia Cerutti received the B.Sc. degree and the M.Sc. degree in telecommunication engineering from Politecnico di Milano, Milan, Italy, in 2016 and 2018, respectively. Currently, he is a researcher with the Dipartimento di Ingegneria Gestionale, Politecnico di Milano. His research interests include algorithms and protocols for the Internet of Things and signal processing for wireless systems, with emphasis on channel estimation for mmWave MIMO systems.

Monica Nicoli received the M.Sc. in Telecommunication Engineering (Hons.) and the Ph.D. in Electronic and Communication Engineering from Politecnico di Milano, in 1998 and 2002, respectively. She was a Visiting Researcher with Uppsala University in 2001. She joined Politecnico di Milano as a Faculty Member in 2002 and she is currently an Associate Professor. Her research interests are in the area of statistical signal processing, with a focus on wireless communications, localization and navigation, smart mobility and cooperative ITS. She has co-authored over 100 scientific publications. She is a co-recipient of the Best Paper Awards of the IEEE Statistical Signal Processing Workshop 2018 and the IET ITS Journal 2014, a recipient of the 1999 Marisa Bellisario Award and the 1999 SPE-EAGE Best Thesis Award. She served as Associate Editor of the EURASIP Journal on Wireless Communications and Networking from 2010 to 2017, and as Lead Guest Editor for the Special Issue on Localization in Mobile Wireless and Sensor Networks in 2011.

Stefano Tomasin received the Ph.D. degree in Telecommunications Engineering from the University of Padova, Italy, in 2003, where he is currently Associate Professor. He has been on leave at Philips Research (Eindhoven, Netherlands) in 2002, Qualcomm Research Laboratories (San Diego, California) in 2004, Polytechnic University (Brooklyn, New York) in 2007 and Huawei Mathematical and Algorithmic Sciences Laboratory (Boulogne-Billancourt, France) in 2014. His current research interests include physical layer security and signal processing for wireless communications, with application to 5th generation cellular systems. In 2011-2017 he was Editor of the IEEE Transactions of Vehicular Technologies and since 2016 he has been Editor of the IEEE Transactions on Signal Processing. Since 2011 he has been also Editor of the EURASIP Journal of Wireless Communications and Networking.

Umberto Spagnolini is Faculty member at Politecnico di Milano since 1990, where he is Professor of Statistical Signal Processing. His research in statistical signal processing covers remote sensing and communication systems with more than 300 papers on peer-reviewed journals/conferences and patents. He is author of the book Statistical Signal Processing in Engineering (J. Wiley, 2017). The specific areas of interest include channel estimation and space-time processing for single/multi-user wireless communication systems, cooperative and distributed inference methods including V2X systems, mmWave communication systems, parameter estimation/tracking and wavefield interpolation for remote sensing (UWB radar and oil exploration). He was recipient/co-recipient of Best Paper Awards from EAGE on geophysical signal processing methods (1991, 1998), and IEEE on array processing (ICASSP 2006) and distributed synchronization for wireless sensor networks (SPAWC 2007, WRECOM 2007). He served as part of IEEE Editorial boards as well as member in technical program committees of several conferences for all the areas of interests.