CHALMERS
UNIVERSITY OF TECHNOLOGY

Low-Complexity Downlink Channel Estimation in mmWave Multiple-Input Single-Output Systems

Citation for the original published paper (version of record):

Fascista, A., De Monte, A., Coluccia, A. et al (2022). Low-Complexity Downlink Channel Estimation in mmWave Multiple-Input Single-Output Systems. IEEE Wireless Communications Letters, 11(3): 518-522. http://dx.doi.org/10.1109/LWC.2021.3134826

Low-Complexity Downlink Channel Estimation in mmWave Multiple-Input Single-Output Systems

Alessio Fascista, *Member, IEEE*, Alessandro De Monte, Angelo Coluccia, *Senior Member, IEEE*, Henk Wymeersch, *Senior Member, IEEE*, and Gonzalo Seco-Granados, *Senior Member, IEEE*

***Abstract*—This paper tackles the problem of channel estimation in mmWave multiple-input single-output systems, where users are equipped with single-antenna receivers. By leveraging broadcast transmissions in the downlink channel, two novel low-complexity estimation approaches are devised, able to operate even in presence of a reduced number of transmit antennas or limited bandwidth. Numerical results show that the proposed algorithms provide accurate estimates of the channel parameters, achieving at the same time about 50% complexity reduction compared to existing approaches.**

***Index Terms*—channel estimation, mmWave, MIMO, MISO, downlink, 5G, cellular networks**

I. Introduction

The use of multiple antennas in mobile cellular communications is one of the key enablers for the fifth generation (5G) and beyond [1], [2]. mmWaves significantly shrink the antenna array size and enable very narrow-beam spatial multiplexing [3], [4]. Current systems are however asymmetric, with many antennas at the base stations (BSs) and one or very few antennas at the user equipments (UEs). This is evidenced by the multiple-input multiple-output (MIMO) literature, where also the single-antenna receiver setup, namely multiple-input single-output (MISO), is very relevant [5]–[7].

The most common channel estimation approaches exploit the uplink (UL) channel, which has the advantage of many antennas and large computational capabilities at the BS side. On the other hand, UL channel estimation requires dedicated pilots per each UE, possibly leading to pilot contamination [8]. mmWave downlink (DL) channel estimation is more scalable with the number of UEs, since the channel parameters of multiple (potentially unlimited) UEs can be estimated through broadcast signals once precoding matrices are set, with no additional system overhead. This also provides timely estimates of delays and angles, which is crucial for positioning and mapping applications, among others [9], [10].

Different algorithms have been proposed based on the Compressive Sensing (CS) [11], to exploit the sparsity of mmWave channels. Bayesian learning has been also considered [12]. In [13], the use of the space-alternating generalized expectation-maximization (SAGE) algorithm is proposed to reduce the cost involved in the estimation of the whole multipath MIMO channel. However, the estimation task is more challenging in the MISO case: in fact, the angle of arrival (AOA) cannot be estimated, and the angle of departure (AOD) need to be determined from scalar data due to the lack spatial diversity. Furthermore, the limited processing capabilities available at the UE becomes a bottleneck, which calls for the need of low-complexity estimators.

In [14] a cascade approach for the MISO scenario is developed, in which the DL channel parameters, namely AOD and delay, are estimated in two subsequent stages. In particular, two separate 1D optimizations are performed: a first one on a non-linear unstructured maximum likelihood (UML) cost function for delay estimation, followed by a second one on a 1D-ML cost function for AOD estimation. This approach avoids the 2D optimization of the ML estimator, which searches for the AOD-delay pair that maximizes the compressed ML cost function after plugging the estimates of complex channel gain and noise power. The method can attain the Cramér-Rao lower bound (CRLB) for sufficient bandwidth; otherwise, the first stage may provide an insufficiently accurate delay estimate to the second stage, ultimately undermining the performance of the whole procedure. Moreover, although the full optimization of the optimal 2D-ML estimator is avoided, still two separate 1D optimizations of highly non-linear cost functions are involved, resulting in high complexity at the UE.

In this work, we provide two novel cascade estimation approaches to overcome the limitations of [14]. In the first one, a suitable reinterpretation of the received signal model in the fast Fourier transform (FFT) domain is exploited to directly obtain a delay estimate; the latter is subsequently plugged into the optimal 2D-ML estimator (which reduces to 1D) to produce an AOD estimate. The second approach is a cascade estimator that conversely starts with AOD estimation, by elaborating on the unstructured estimation idea as in [14], but following a completely different path. Then, instead of plugging the obtained estimate into the 2D-ML estimator (to finally estimate the delay), we propose for such a second stage a low-complexity FFT-based estimator, thus avoiding one of the two 1D non-linear optimizations.

Remarkably, the two proposed approaches exhibit good estimation performance, even in presence of multipath propagation and wideband channel effects such as the beam-squint, and provide about 50% reduction in the computational burden thanks to the suitably-designed FFT-based estimation steps.

A. Fascista, A. De Monte, and A. Coluccia are with the Department of Innovation Engineering, Università del Salento, Via Monteroni, 73100 Lecce, Italy (e-mail: alessio.fascista@unisalento.it; alessandro.demonte@studenti.unisalento.it; angelo.coluccia@unisalento.it).

H. Wymeersch is with the Department of Electrical Engineering, Chalmers University of Technology, 412 96 Gothenburg, Sweden (e-mail: henkw@chalmers.se).

G. Seco-Granados is with the Department of Telecommunications and Systems Engineering, Universitat Autònoma de Barcelona, 08193 Barcelona, Spain (e-mail: gonzalo.seco@uab.cat).

II. Signal and Channel Models

A single BS at known position $\boldsymbol{p}_{\text{BS}} = [0\ 0]^{\mathsf{T}}$, equipped with N_{BS} antennas, transmits G orthogonal frequency division multiplexing (OFDM) signals towards a single-antenna UE at unknown position $\boldsymbol{p} = [p_x\ p_y]^{\mathsf{T}}$. Each OFDM signal is transmitted at time g over each subcarrier $n = 0, \dots, N-1$, as $\boldsymbol{z}_g[n] = \boldsymbol{F}_g[n]\boldsymbol{x}_g[n]$, where $\boldsymbol{x}_g[n] \in \mathbb{C}^{M\times 1}$ is the vector of the transmitted symbols (with power $P_t = \mathbb{E}\left[\|\boldsymbol{x}_g[n]\|^2\right]$) and $\boldsymbol{F}_g[n] \in \mathbb{C}^{N_{\text{BS}}\times M}$ denotes the precoding matrix (either analog or hybrid) applied at the transmit side [15], [16], whose entries satisfy the total power constraint $\|\boldsymbol{F}_g[n]\|_{\text{F}} = 1$ [17].

The general channel model consists of a line-of-sight (LOS) path and P non-line-of-sight (NLOS) paths generated by local scatterers or reflectors. The channel parameters for each path include the AOD θ_i and the delay τ_i (accounting for the time-of-flight and the clock synchronization offset). We denote by $i = 0$ the LOS link, while $i > 0$ are NLOS paths. Assuming DL communications at a frequency f_c with bandwidth B, the complex channel vector over each subcarrier n is given by

$$\boldsymbol{h}^{\mathsf{T}}[n] = \boldsymbol{\Gamma}^{\mathsf{T}}[n]\boldsymbol{A}_{\text{BS}}^{\mathsf{T}}[n]. \tag{1}$$

$\boldsymbol{A}_{\text{BS}}[n] = [\boldsymbol{a}_{\text{BS}}(\theta_0)[n] \cdots \boldsymbol{a}_{\text{BS}}(\theta_P)[n]]$ is the frequency-dependent array response matrix, $\boldsymbol{\Gamma}[n] = [\alpha_0 \mathrm{e}^{\frac{-j2\pi n\tau_0}{NT_s}} \ \cdots\ \alpha_P \mathrm{e}^{\frac{-j2\pi n\tau_P}{NT_s}}]^{\mathsf{T}}$ where $\alpha_i = h_i/\sqrt{\rho_i}$, with ρ_i denoting the path loss and h_i the complex channel gain of the i-th path, respectively. To formulate the channel estimation problem, we exploit the fact that, under typical mmWave conditions, the channel is sparse and all the paths are resolvable either in the time or space domains, with negligible overlap among them. Therefore, the NLOS paths can be identified and treated separately with respect to the LOS [10], [18], and multipath parameter estimation can be then reduced to single-path estimation, that is, the channel vector given in (1) can be simplified to

$$\boldsymbol{h}^{\mathsf{T}}[n] = \alpha \mathrm{e}^{\frac{-j2\pi n\tau}{NT_s}} \boldsymbol{a}_{\text{BS}}^{\mathsf{T}}(\theta)[n] \tag{2}$$

with $\alpha \overset{\text{def}}{=} \alpha_0$, $\tau \overset{\text{def}}{=} \tau_0$, and $\theta \overset{\text{def}}{=} \theta_0$ the sole LOS parameters. This work aims at tackling the problem of DL channel estimation (namely, estimation of θ and τ) based on all the OFDM signals received by the UE at each time instant $g = 1, \dots, G$ over each subcarrier $n = 0, \dots, N-1$, i.e.,

$$y_g[n] = \boldsymbol{h}^{\mathsf{T}}[n]\boldsymbol{z}_g[n] + \nu_g[n] \tag{3}$$

with $\nu_g[n]$ the additive circularly complex Gaussian noise with zero mean and variance σ^2.

III. Low-Complexity Channel Estimation

In the following, we neglect the frequency dependency in the array steering vector for the sake of mathematical tractability[1] ($\boldsymbol{A}_{\text{BS}}[n] \approx \boldsymbol{A}_{\text{BS}}$). We will assess later in Sec. IV-B2 the sensitivity of the designed algorithms to actual frequency-selective channels. We also consider a uniform linear array (ULA) with steering vector $\boldsymbol{a}_{\text{BS}}(\theta) = [1\ e^{j\frac{2\pi}{\lambda_c} d\sin\theta} \cdots\ e^{j(N_{\text{BS}}-1)\frac{2\pi}{\lambda_c} d\sin\theta}]^{\mathsf{T}}$.

[1]This is a typical assumption in practice since the associated beam-squint effect is non-negligible only for very large bandwidths and array dimensions.

A. Delay-before-Angle (DbA) Estimator

In this section, we derive a two-stage low-complexity estimator that exploits a reinterpretation of (3) in the FFT domain to directly provide an estimate of the delay in the first stage, followed by a single 1D search in the optimal 2D-ML cost function to finally obtain an AOD estimate. In the following we denote this approach as delay-before-angle (DbA) estimator.

1) FFT-based Estimation of Delay: By considering the practical case in which the precoding matrices $\boldsymbol{F}_g[n]$ and the data symbols $\boldsymbol{x}_g[n]$ do not change across the different subcarriers n, the observation vectors $\boldsymbol{y}_g = [y_g[0]\ \cdots\ y_g[N-1]]^{\mathsf{T}}$, $g = 1, \dots, G$, can be expressed in the more convenient form[2]

$$\boldsymbol{y}_g = \boldsymbol{l}_g(\tau, \theta) + \boldsymbol{\nu}_g \tag{4}$$

where $\boldsymbol{\nu}_g = [\nu_g[0]\ \cdots\ \nu_g[N-1]]^{\mathsf{T}}$, and

$$\boldsymbol{l}_g(\tau, \theta) = p_g(\theta)\boldsymbol{s}(\tau) \tag{5}$$

with $p_g(\theta) = \boldsymbol{z}_g^{\mathsf{T}}\boldsymbol{a}_{\text{BS}}(\theta)$ a complex scalar that do not change over the different subcarriers n (since $\boldsymbol{z}_g[n] = \boldsymbol{z}_g\ \forall n$) and $\boldsymbol{s}(\tau) = [1\ \mathrm{e}^{\frac{-j2\pi\tau}{NT_s}} \cdots \mathrm{e}^{\frac{-j2\pi(N-1)\tau}{NT_s}}]^{\mathsf{T}}$ is the sole vector depending on the unknown delay τ. To obtain a low-complexity estimate of τ, we exploit the fact that the elements of $\boldsymbol{l}_g(\tau, \theta)$ in (5) can be interpreted as discrete samples of a complex exponential with normalized frequency $\nu_o = -\frac{\tau}{NT_s}$. This allows to estimate τ by searching for the dominant peak in the FFT of $\boldsymbol{y}_g, g = 1, \dots, G$. By defining $\boldsymbol{f}_g = \text{FFT}(\boldsymbol{y}_g)$ as the FFT-transformed observations on N_{F} points, we first seek for the index of the maximum element in the cost function

$$\hat{k} = \arg\max_k \left[\sum_{g=1}^{G} |f_g[k]| : 0 \le k \le N_F - 1\right] \tag{6}$$

with $|f_g[k]|$ denoting the absolute value of the k-th element of $\boldsymbol{f}_g$. Since the first $N_F/2+1$ elements correspond to positive values of the normalized frequency $\nu_o \in [0, 1/2]$, while the remaining $N_F/2-1$ are associated to the negative part of the spectrum, i.e., $\nu_o \in (-1/2, 0)$, the sought estimate is

$$\hat{\tau}_{\text{FFT}} = \begin{cases} -\frac{\hat{k}}{N_F} NT_S & \text{if } 0 \le \hat{k} \le N_F/2 \\ \left(1/2 - \frac{\hat{k}}{N_F}\right) NT_S & \text{if } N_F/2+1 \le \hat{k} \le N_F - 1 \end{cases}. \tag{7}$$

This approach directly estimates the delay τ without requiring knowledge of the AOD θ, as the corresponding terms $p_g(\theta)$ are constant over the different subcarriers, hence do not affect the FFT-based step. Furthermore, the complexity of the FFT peak search is far lower than the 1D optimization of the highly non-linear cost function in the UML approach in [14].

2) 1D-ML Estimation of AOD: In the second stage, we plug the above estimate $\hat{\tau}_{\text{FFT}}$ into the optimal 2D-ML estimator derived in [14] (which then reduces to a 1D-ML) and solve it for the unknown AOD by performing a single 1D optimization. The value estimated through this procedure will be denoted as $\hat{\theta}_{\text{1D-ML}}(\hat{\tau}_{\text{FFT}})$, to better highlight its dependency on $\hat{\tau}_{\text{FFT}}$.

[2]The same approach can be applied when $\boldsymbol{x}_g[n] = q[n]\boldsymbol{x}_g$, with $q[n] \in \mathbb{C}$; to ease the notation, hereafter we consider the case of $q[n] = 1\ \forall n$.

B. Angle-before-Delay (AbD) Estimator

In this section, we derive a two-stage low-complexity channel estimator that leverages a proper unstructured transformation of (2) to estimate the AOD, followed by an FFT-based estimation of the delay. Due to the order of the estimated parameters, we call it angle-before-delay (AbD) estimator.

1) Unstructured ML-based Estimation of AOD: We start by briefly recalling the unstructured transformation adopted in [14], which will be useful for the subsequent derivations. By stacking all the received signals from (3), we obtain

$$\begin{bmatrix} \boldsymbol{y}_1 \\ \vdots \\ \boldsymbol{y}_G \end{bmatrix} = \alpha \underbrace{\begin{bmatrix} \boldsymbol{S}(\tau)\boldsymbol{Z}_1^{\mathsf{T}} \\ \vdots \\ \boldsymbol{S}(\tau)\boldsymbol{Z}_G^{\mathsf{T}} \end{bmatrix}}_{\boldsymbol{A}(\tau)\in\mathbb{C}^{NG\times N_{BS}}} \boldsymbol{a}_{\text{BS}}(\theta) + \begin{bmatrix} \boldsymbol{\nu}_1 \\ \vdots \\ \boldsymbol{\nu}_G \end{bmatrix} \tag{8}$$

where $\boldsymbol{Z}_g = [\boldsymbol{z}_g[0] \ \cdots \ \boldsymbol{z}_g[N-1]] \in \mathbb{C}^{N_{\text{BS}}\times N}$, and $\boldsymbol{S}(\tau) = \text{diag}\left(\boldsymbol{s}(\tau)\right)$ is a diagonal matrix with $\boldsymbol{s}(\tau)$ as its diagonal elements. By relaxing the dependency of (8) on the AOD θ and using an unstructured model for $\boldsymbol{b} = \alpha\boldsymbol{a}_{\text{BS}}(\theta) \in \mathbb{C}^{N_{\text{BS}}\times 1}$ we can rewrite (8) in a more compact form as

$$\boldsymbol{y} = \boldsymbol{A}(\tau)\boldsymbol{b} + \boldsymbol{\nu} \tag{9}$$

where $\boldsymbol{y} = [\boldsymbol{y}_1^{\mathsf{T}} \ \cdots \ \boldsymbol{y}_G^{\mathsf{T}}]^{\mathsf{T}}$ and $\boldsymbol{\nu} = [\boldsymbol{\nu}_1^{\mathsf{T}} \ \cdots \ \boldsymbol{\nu}_G^{\mathsf{T}}]^{\mathsf{T}}$. The model in (9) is used in [14] to obtain an UML estimator of τ.

In this paper, we take a different path and propose an alternative decoupling of the dependencies of (2) on τ and θ, which allows us to rewrite the received signal model as

$$\begin{bmatrix} \boldsymbol{y}_1 \\ \vdots \\ \boldsymbol{y}_G \end{bmatrix} = \alpha \underbrace{\begin{bmatrix} \boldsymbol{\Psi}_1(\theta) \\ \vdots \\ \boldsymbol{\Psi}_G(\theta) \end{bmatrix}}_{\boldsymbol{\Psi}(\theta)\in\mathbb{C}^{NG\times N}} \boldsymbol{s}(\tau) + \begin{bmatrix} \boldsymbol{\nu}_1 \\ \vdots \\ \boldsymbol{\nu}_G \end{bmatrix} \tag{10}$$

where $\boldsymbol{\Psi}_g(\theta) = \text{diag}(\boldsymbol{a}_{\text{BS}}^{\mathsf{T}}(\theta)\boldsymbol{Z}_g)$. By relaxing the dependency of (10) on τ and using an unstructured model for the vector

$$\boldsymbol{t} = \alpha\boldsymbol{s}(\tau) \in \mathbb{C}^{N\times 1} \tag{11}$$

we can express (10) in the more compact form

$$\boldsymbol{y} = \boldsymbol{\Psi}(\theta)\boldsymbol{t} + \boldsymbol{\nu}. \tag{12}$$

Starting from this model, the UML estimator of θ is given by

$$\hat{\theta}_{\text{UML}} = \arg\min_{\theta} \left[\min_{\boldsymbol{t}} \|\boldsymbol{y} - \boldsymbol{\Psi}(\theta)\boldsymbol{t}\|^2\right] \tag{13}$$

whose inner minimization is solved in closed-form as

$$\hat{\boldsymbol{t}}(\theta) = (\boldsymbol{\Psi}^{\mathsf{H}}(\theta)\boldsymbol{\Psi}(\theta))^{-1}\boldsymbol{\Psi}^{\mathsf{H}}(\theta)\boldsymbol{y}. \tag{14}$$

Interestingly, the proposed unstructured transformation can avoid the inversion of the NG-dimensional matrix in (14), due to the peculiar structure of $\boldsymbol{\Psi}(\theta)$ which is a stack of diagonal matrices; in fact, through some basic manipulations, (13) can be more conveniently expressed as

$$\begin{aligned} \hat{\theta}_{\text{UML}} &= \arg\max_{\theta} \ \boldsymbol{y}^{\mathsf{H}}\boldsymbol{\Psi}(\theta)(\boldsymbol{\Psi}^{\mathsf{H}}(\theta)\boldsymbol{\Psi}(\theta))^{-1}\boldsymbol{\Psi}^{\mathsf{H}}(\theta)\boldsymbol{y} \\ &= \arg\max_{\theta} \ \left\|(\boldsymbol{\Psi}^{\mathsf{H}}(\theta)\boldsymbol{\Psi}(\theta))^{-1/2}\sum_{g=1}^{G}\boldsymbol{\Psi}_g^{\mathsf{H}}(\theta)\boldsymbol{y}_g\right\|^2 \\ &= \arg\max_{\theta} \ \|\boldsymbol{r}(\theta)\|^2 \end{aligned} \tag{15}$$

where $\boldsymbol{r}(\theta) = \boldsymbol{D}(\theta)\sum_{g=1}^{G}\boldsymbol{\Psi}_g^{\mathsf{H}}(\theta)\boldsymbol{y}_g$, $\boldsymbol{D}(\theta) = \text{diag}(\xi_0^{-\frac{1}{2}}(\theta),\ldots,\xi_{N-1}^{-\frac{1}{2}}(\theta))$, $\xi_n(\theta) = \sum_{g=1}^{G}\left|\boldsymbol{a}_{\text{BS}}^{\mathsf{T}}(\theta)\boldsymbol{z}_g[n]\right|^2$. Eq. (15) yields the UML estimator of θ, which can be implemented with a 1D search over a quantized interval[3].

2) FFT-based Estimation of Delay: By recognizing the interesting analogy between (11) and the peculiar structure of the vectors in (5), the entries of the unstructured vector $\boldsymbol{t}$ can be reinterpreted as discrete samples of a complex exponential function having normalized frequency $\nu_o = -\frac{\tau}{NT_S}$. Accordingly, an estimate of τ can be obtained by searching for the normalized frequency $\hat{\nu}_o$ corresponding to the dominant peak in the FFT of the vector $\hat{\boldsymbol{t}}(\hat{\theta}_{\text{UML}})$ built by plugging the AOD estimate provided by the UML in (15) back into (14). Specifically, by defining $\boldsymbol{f}(\hat{\theta}_{\text{UML}}) = \text{FFT}(\hat{\boldsymbol{t}}(\hat{\theta}_{\text{UML}}))$, we first seek for the index corresponding to the maximum in $\boldsymbol{f}(\hat{\theta}_{\text{UML}})$

$$\hat{k}(\hat{\theta}_{\text{UML}}) = \arg\max_{k} \left[|f(\hat{\theta}_{\text{UML}})[k]| : 0 \le k \le N_F - 1\right]. \tag{16}$$

Then, the index $\hat{k}(\hat{\theta}_{\text{UML}})$ can be mapped to the corresponding estimate $\hat{\tau}_{\text{FFT}}(\hat{\theta}_{\text{UML}})$ according to (7).

IV. Simulation results

A. Simulation Setup

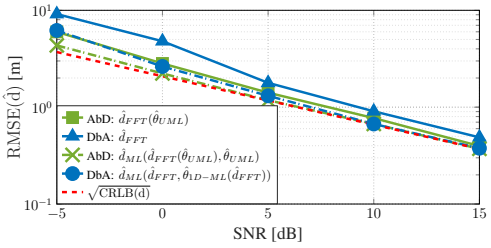
We consider a UE at $[9\ 5]^{\mathsf{T}}$ [m], a carrier frequency $f_c = 60$ GHz typical of mmWave communications, $G = 10$ OFDM symbols, and $N = 20$ subcarriers [13]. The elements of the precoding matrices $\boldsymbol{F}_g$ are generated as complex exponential terms $\text{e}^{j\phi}$ with random phases uniformly distributed in $[0, 2\pi)$, assuming $M = N_{\text{BS}}$. The same applies to the transmitted symbols, which for simplicity are also kept the same over all the subcarriers. The Root Mean Squared Error (RMSE) is adopted as metric to assess the estimation performance, computed based on 2000 independent Monte Carlo trials. For comparison, the performance provided by each individual estimator (of either θ or τ parameters) in the two proposed cascade approaches are reported. Moreover, it is shown the performance of the optimal 2D-ML estimator from [14] implemented using a gradient-based optimization, when initialized with the pair $(\hat{\tau}, \hat{\theta})$ obtained from the two-stage estimators.[4]

B. Results and Discussion

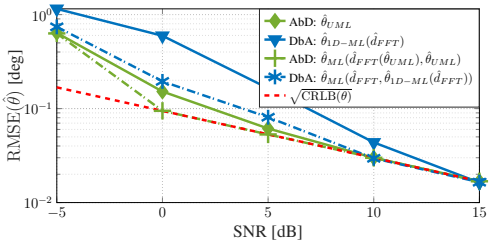
1) Analysis under Matched Conditions: We start by considering $N_{\text{BS}} = 64$ antennas and bandwidth limited to $B = 5$ MHz. In Fig. 1, we show the evolution of the RMSEs on the estimation of τ (reported in terms of distance as $d = c\tau$) and θ as a function of the SNR, also in comparison to the theoretical lower bounds derived in [14]. As Fig. 1a shows, the FFT-based estimator $\hat{d}_{\text{FFT}} = c\hat{\tau}_{\text{FFT}}$ ($\triangle$, ref. Sec. III-A1) of the DbA approach suffers from the lower time resolution due to the limited signal bandwidth, and indeed exhibits larger errors than the $\hat{d}_{\text{FFT}}(\hat{\theta}_{\text{UML}}) = c\hat{\tau}_{\text{FFT}}(\hat{\theta}_{\text{UML}})$ ($\square$, ref. Sec. III-B2)

[3]Notice that the computation of $\boldsymbol{r}(\theta)$ involves only diagonal matrices, hence can be very efficiently implemented via element-wise vector multiplications.

[4]Notice that, differently from a direct resolution of the 2D-ML estimation problem which requires an exhaustive 2D grid search, the gradient-based optimization can be efficiently performed in a short time, as it involves a few iterations to obtain a solution starting from the initial point $(\hat{\tau}, \hat{\theta})$.

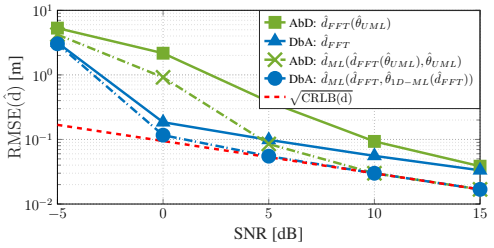


(a) RMSE on the estimation of $d = c\tau$.

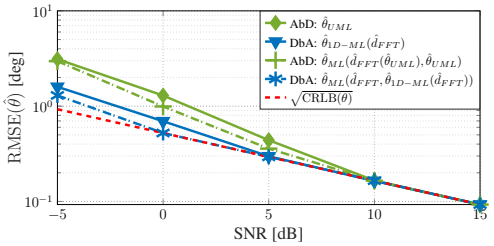


(b) RMSE on the estimation of θ.

Fig. 1: RMSEs on $d = c\tau$ and θ estimation compared to the CRLBs as a function of the SNR, for $N_{\text{BS}} = 64$ and $B = 5$ MHz.



(a) RMSE on the estimation of $d = c\tau$.



(b) RMSE on the estimation of θ.

Fig. 2: RMSEs on $d = c\tau$ and θ estimation compared to the CRLBs as a function of the SNR, for $N_{\text{BS}} = 8$ and $B = 50$ MHz.

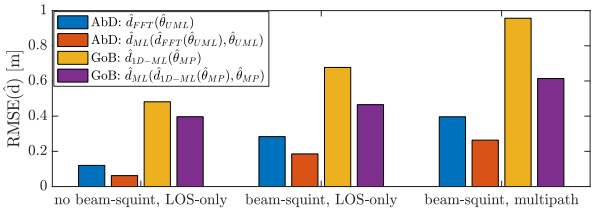
which instead benefits from a much more accurate estimate of the AOD provided by the UML-based estimator $\hat{\theta}_{\text{UML}}$ in the first stage of the AbD approach ($\diamond$, ref. Sec. III-B1). Indeed, as visible in Fig. 1b, the proposed UML estimator exhibits very good performance already at SNR of 0 dB and, despite its suboptimality, is able to attain the corresponding bound starting from an SNR of about 5 dB. Oppositely, the larger error of $\hat{d}_{\text{FFT}}$ negatively impacts onto the performance of the $\hat{\theta}_{\text{1D-ML}}(\hat{\tau}_{\text{FFT}})$ used in the second stage of the DbA approach ($\triangledown$, ref. Sec. III-A2), despite the sufficiently high number of available antennas for AOD estimation.

Fig. 1 also shows the improved performance that can be obtained when the proposed estimators are used to initialize the iterative (gradient-based) optimization of the optimal 2D-ML estimator. Remarkably, the RMSEs for the initialization pair $(\hat{d}_{\text{FFT}}(\hat{\theta}_{\text{UML}}), \hat{\theta}_{\text{UML}})$ (curves marked by $\times$ and $+$, respectively, in Figs. 1a and 1b) quickly attains the theoretical bounds as the SNR increases, so achieving the same accuracy of the approach in [14], which however requires two separate one-dimensional non-linear optimizations. Conversely, the RMSEs for the initialization pair $(\hat{d}_{\text{FFT}}, \hat{\theta}_{\text{1D-ML}}(\hat{d}_{\text{FFT}}))$ (curves marked by $\circ$ and $*$, respectively, in Figs. 1a and 1b), although lower than those of the plain estimators, are not able to attain the bounds unless the SNR is significantly higher.

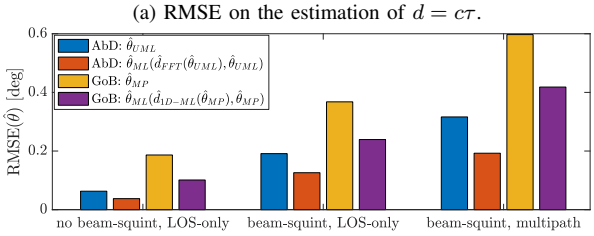
We now consider a different scenario with only $N_{\text{BS}} = 8$ antennas but $B = 50$ MHz bandwidth. Fig. 2 shows that considerations analogous to Fig. 1 can be made by exchanging the role of the two proposed approaches. Specifically, best performance are obtained by the $\hat{d}_{\text{FFT}}$ ($\triangle$) and subsequently by the $\hat{\theta}_{\text{1D-ML}}(\hat{d}_{\text{FFT}})$ ($\triangledown$), thanks to the larger available bandwidth: in fact, the better accuracy in the FFT-based estimation of the delay in the first stage can compensate for the reduced spatial resolution in the second stage of the DbA approach (due to the small number of antennas). For the same reason, instead, $\hat{\theta}_{\text{UML}}$ ($\diamond$) suffers in this scenario, as visible in Fig. 2b, and negatively impacts on the performance of $\hat{d}_{\text{FFT}}(\hat{\theta}_{\text{UML}})$ ($\square$, Fig. 2a) in the second stage of the AbD approach. Similarly, all estimators improve when refined through iterations of the optimal 2D-ML, but the relative ranking remains; indeed, the initialization pair $(\hat{d}_{\text{FFT}}, \hat{\theta}_{\text{1D-ML}}(\hat{d}_{\text{FFT}}))$ ($\circ$ and $*$) quickly leads to attaining the bounds already around 0 dB, while the pair $(\hat{d}_{\text{FFT}}(\hat{\theta}_{\text{UML}}), \hat{\theta}_{\text{UML}})$ ($\times$ and $+$) requires at least 10 dB of SNR.

2) Analysis under Mismatched Conditions: We assess the performance under different, gradually introduced mismatches on the assumed model, representative of the main phenomena found in mmWave communications. We compare our approaches against the state-of-the-art method adopted in the 3GPP standard, known as grid-of-beams (GoB) [19]. The GoB estimates the AOD by measuring the power of each received beam. The beam associated to the maximum power is retained as AOD estimate (denoted as $\hat{\theta}_{\text{MP}}$) and is subsequently plugged into the 2D-ML estimator to obtain an estimate of the delay. The analysis is carried out by considering $N_{\text{BS}} = 64$ and $B = 200$ MHz, and an increased number of transmissions $G = N_{\text{BS}}$ so that the GoB method can achieve its maximum spatial resolution (while we recall that our approaches already achieved a good accuracy with a much smaller G). The performance are evaluated by considering another realistic mmWave setup operating at $f_c = 38$ GHz, which represents the upper limit of the FR2 band used by 5G New Radio (NR).

We start by evaluating the beam-squint effect of frequency-selective OFDM channels, i.e., the actual array response is frequency-dependent while the estimators neglect such a dependency at the design stage, a scenario labeled "beam-squint, LOS-only". The obtained results, represented by the first two groups of bars in Fig. 3, reveal that the AbD exhibits good robustness against this effect. To further challenge the estimators, in a second scenario "beam-squint, multipath" we consider the simultaneous presence of beam-squint and multipath propagation. Specifically, we introduce an additional NLOS path generated by a scatterer located at (unknown) position $[15\ 3]^T$ [m], with $|\alpha_i|^2 = (\lambda_c/4\pi)^2\sigma^2_{\text{RCS}}/(d_{1,i}d_{2,i})^2$



(a) RMSE on the estimation of $d = c\tau$.



(b) RMSE on the estimation of θ.

Fig. 3: RMSEs on $d = c\tau$ and θ estimation compared to the GoB approach in different mismatched scenarios, for SNR = 5 dB.

and $\sigma^2_{\text{RCS}} = 10$ m^2 radar cross section (to emulate a moderate multipath propagation) and $d_i = d_{1,i} + d_{2,i} = c\tau_i$ [18]. Results reported in the third group of bars in Fig. 3 demonstrate that the proposed AbD algorithm is effective also in this scenario, with only a slight performance degradation. Remarkably, the AbD approach significantly outperforms the GoB method in all the three considered scenarios, with a gap that becomes more pronounced as the mismatches are progressively included.

3) Complexity Comparison: Asymptotically, the complexity in performing the two 1D optimizations required by [14] is $O(QG(NN_{\text{BS}}M + N^2))$, with Q the number of evaluation points per dimension (delay or AOD), which is the same of the two proposed approaches. However, the big-O notation hides constants that significantly impact onto the actual computational cost. We thus compared the average runtimes normalized by the overall average runtime of the cascade approach in [14]. It turns out that the complexity of the proposed algorithms is dominated by either the first (UML estimation) stage in the AbD or the second (1D-ML estimation) stage in the DbA, since they involve a 1D search to optimize their respective nonlinear cost functions. On the other hand, the FFT-based estimation stages can be efficiently performed in less than 5% out of the total execution time. The approach in [14] has the highest complexity, being the two involved estimation stages as complex as the sum between the first UML estimation stage in the AbD and the second 1D-ML estimation stage in the DbA. Overall, the proposed two-stage channel estimators saves about 50% computational cost compared to [14].

V. Conclusions

The paper addressed the problem of delay and AOD estimation in a mmWave MISO system. A first proposed approach exploits a reinterpretation of the observations in the FFT domain to directly estimate the delay, then plugged into the optimal 2D-ML estimator to finally obtain an AOD estimate, by means of a single 1D search. The second approach leverages a proper unstructured transformation of the channel model to estimate the AOD, followed by an FFT-based estimation of the delay. The proposed algorithms turn out to provide accurate channel estimates and, remarkably, allow to save about 50% of computational cost compared to state-of-the-art algorithms. As future work, it would be worth investigating scenarios where the beam-squint effect is dominant, namely massive number of antennas combined with very broad bandwidths.

References

[1] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Comm. Magazine*, vol. 52, no. 2, pp. 186–195, 2014.

[2] J. Zhang, E. Björnson, M. Matthaiou, D. W. K Ng, H. Yang, and D. J. Love, "Prospective multiple antenna technologies for beyond 5G," *IEEE Journ. on Selec. Areas in Comm.*, vol. 38, no. 8, pp. 1637–1660, 2020.

[3] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez, "Millimeter wave mobile communications for 5G cellular: It will work!," *IEEE Access*, vol. 1, pp. 335–349, 2013.

[4] T. Bai and R. W. Heath, "Coverage and rate analysis for millimeter-wave cellular networks," *IEEE Trans. on Wireless Comm.*, vol. 14, no. 2, pp. 1100–1114, 2015.

[5] Y. R. Ramadan, H. Minn, and A. S. Ibrahim, "Hybrid analog–digital precoding design for secrecy mmWave MISO-OFDM systems," *IEEE Trans. on Comm.*, vol. 65, no. 11, pp. 5009–5026, 2017.

[6] H. Li, M. Li, and Q. Liu, "Hybrid beamforming with dynamic subarrays and low-resolution PSs for mmWave MU-MISO systems," *IEEE Trans. on Comm.*, vol. 68, no. 1, pp. 602–614, 2020.

[7] G. Lee, Y. Sung, and J. Seo, "Randomly-directional beamforming in Millimeter-Wave multiuser MISO downlink," *IEEE Trans. on Wireless Comm.*, vol. 15, no. 2, pp. 1086–1100, 2016.

[8] A. Zaib, M. Masood, A. Ali, W. Xu, and T. Y. Al-Naffouri, "Distributed channel estimation and pilot contamination analysis for massive MIMO-OFDM systems," *IEEE Trans. on Comm.*, vol. 64, no. 11, pp. 4607–4621, 2016.

[9] A. Fascista, A. Coluccia, H. Wymeersch, and G. Seco-Granados, "Low-complexity accurate mmwave positioning for single-antenna users based on angle-of-departure and adaptive beamforming," in *IEEE Intern. Conf. on Acoustics, Speech and Signal Proc. (ICASSP)*, 2020, pp. 4866–4870.

[10] A. Fascista, A. Coluccia, H. Wymeersch, and G. Seco-Granados, "Downlink single-snapshot localization and mapping with a single-antenna receiver," *IEEE Trans. on Wireless Comm.*, vol. 20, no. 7, pp. 4672–4684, 2021.

[11] Z. Wan, Z. Gao, B. Shim, K. Yangand G. Mao, and M. S. Alouini, "Compressive Sensing Based Channel Estimation for Millimeter-Wave Full-Dimensional MIMO With Lens-Array," *IEEE Trans. on Vehic. Techn.*, vol. 69, no. 2, pp. 2337–2342, 2020.

[12] S. Srivastava, A. Mishra, A. Rajoriya, A. Jagannatham, and G. Ascheid, "Quasi-static and time-selective channel estimation for block-sparse millimeter wave hybrid MIMO systems: Sparse bayesian learning (SBL) based approaches," *IEEE Trans. on Sign. Proc.*, vol. 67, no. 5, pp. 1251–1266, 2018.

[13] A. Shahmansoori, G. Garcia, G. Destino, G. Seco-Granados, and H. Wymeersch, "Position and orientation estimation through millimeter-wave MIMO in 5G systems," *IEEE Trans. on Wireless Comm.*, vol. 17, no. 3, pp. 1822–1835, 2018 .

[14] A. Fascista, A. Coluccia, H. Wymeersch, and G. Seco-Granados, "Millimeter-wave downlink positioning with a single-antenna receiver," *IEEE Trans. on Wireless Comm.*, vol. 18, no. 9, pp. 4479–4490, 2019.

[15] A. Alkhateeb, J. Mo, N. Gonzalez-Prelcic, and R. W. Heath, "MIMO precoding and combining solutions for millimeter-wave systems," *IEEE Comm. Magazine*, vol. 52, no. 12, pp. 122–131, 2014.

[16] C. Huang, L. Liu, C. Yuen, and S. Sun, "Iterative channel estimation using LSE and sparse message passing for mmwave MIMO systems," *IEEE Trans. on Signal Proc.*, vol. 67, no. 1, pp. 245–259, 2019 .

[17] A. Alkhateeb and R. W. Heath, "Frequency selective hybrid precoding for limited feedback millimeter wave systems," *IEEE Trans. on Comm.*, vol. 64, no. 5, pp. 1801–1818, 2016.

[18] Z. Abu-Shaban, X. Zhou, T. Abhayapala, G. Seco-Granados, and H. Wymeersch, "Error bounds for uplink and downlink 3D localization in 5G millimeter wave systems," *IEEE Trans. on Wireless Comm.*, vol. 17, no. 8, pp. 4939–4954, 2018.

[19] J. Singh and S. Ramakrishna, "On the feasibility of codebook-based beamforming in millimeter wave systems with multiple antenna arrays," *IEEE Trans. on Wireless Comm.*, vol. 14, no. 5, pp. 2670–2683, 2015.