RESEARCH
Open Access

mm-Wave channel estimation with accelerated gradient descent algorithms

Hossein Soleimani[1], Danilo De Donno[2] and Stefano Tomasin[1,3]* (ID)

Abstract

The availability of millimeter wave (mm-Wave) band in conjunction with massive multiple-input-multiple-output (MIMO) technology is expected to boost the data rates of the fifth-generation (5G) cellular systems. However, in order to achieve high spectral efficiencies, an accurate channel estimate is required, which is a challenging task in massive MIMO. By exploiting the small number of paths that characterize the mm-Wave channel, the estimation problem can be solved by compressed-sensing (CS) techniques. In this paper, we propose a novel CS channel estimation method based on the accelerated gradient descent with adaptive restart (AGDAR) algorithm exploiting a ℓ_1-norm approximation of the sparsity constraint. Moreover, a modified re-weighted compressed-sensing (RCS) technique is considered that iterates AGDAR using a weighted version of the ℓ_1-norm term, where weights are adapted at each iteration. We also discuss the impact of cell sectorization and tracking on the channel estimation algorithm. We compare the proposed solutions with existing channel estimations with an extensive simulation campaign on downlink third-generation partnership project (3GPP) channel models.

Keywords: Compressed sensing, Estimation, mm-Wave

1 Introduction

Due to its huge spectrum availability, the millimeter wave (mm-Wave) band is currently considered for the fifth generation (5G) of cellular networks [1–3]. The high attenuation incurred at those frequencies imposes the use of multiple antennas at each device, typically resulting in massive multiple-input-multiple-output (MIMO) systems, giving rise to various challenges. We focus here on channel estimation that is needed for proper transmit beamforming. In fact, the least square (LS) estimate using short training sequences and limited transmit power (to reduce overhead in massive MIMO systems) is not accurate enough for capacity achieving beamforming. However, the mm-Wave MIMO channel comprises a small number of dominant clusters of paths and even with many antennas a small set of parameters characterizes the entire channel. This induces a sparsity of the mm-Wave channel matrix when transformed by a Fourier transform into the so-called *virtual channel*, and compressed-sensing (CS) techniques can be used for channel estimation.

Various solutions have been proposed for channel estimation in mm-Wave communication systems, and the reader may refer to [4] for their survey. Part of the literature has considered transceivers with hybrid beamformers (cascade of beamformers before and after the digital to analog converters): the joint optimization of both training and estimation has been pursued in [5] for these structures using a feedback channel. Orthogonal matching pursuit (OMP) solutions have been considered with both single path [6, 7] and multiple-path cancelation [8, 9]. In [10], an enhanced approach for generating the beamforming codebook has been proposed, using the continuous basis pursuit (CBP) method, while [11] considers a fast iterative shrinkage-thresholding algorithm (FISTA) approach.

For fully digital beamformers, [12] considers the sparse channel estimation as a least absolute shrinkage and selection operator (LASSO) problem. In [13], OMP is used to estimate the channel by iteratively detecting and canceling paths from the virtual channel estimate. In [14], a basis pursuit denoise (BPDN) approach is suggested where a weighted version of ℓ_1-norm term is considered in the

*Correspondence: tomasin@dei.unipd.it
[1]Department of Information Engineering, University of Padova, via Gradenigo 6/A, Padua, Italy
[3]Consorzio Nazionale Interuniversitario per le Telecomunicazioni, Padua, Italy
Full list of author information is available at the end of the article

LASSO problem and weights are iteratively adapted. A sparsity adaptive matching pursuit (SSAMP) approach is instead used in [15], while in [16] the LASSO problem is solved by applying a generalized approximate message passing (GAMP) algorithm exploiting the Bernoulli-Gaussian distribution of paths in the virtual channel.

In this paper, we propose a novel sparse channel estimation method based on the accelerated gradient descent with adaptive restart (AGDAR) algorithm [17]. Focusing on a scenario where the receiver obtains first the LS estimate of the narrowband mm-Wave MIMO channel, we relax the sparse optimization problem using LASSO, wherein the ℓ_0-norm is replaced by the ℓ_1-norm. We apply then the AGDAR algorithm [17] to solve the sparse channel estimation problem. In order to further enhance the channel estimation procedure, a re-weighted ℓ_1-norm problem is considered leading to the re-weighted compressed-sensing (RCS) algorithm [18], which iterates AGDAR with different weights of the ℓ_1-norm term. We also discuss the impact of cell sectorization and channel tracking on the channel estimation algorithm. We compare the proposed solutions with OMP solutions [6, 8, 13]. With respect to the rest of the literature we reduce the complexity (with respect to the random search of A-LASSO in [12]), we swap the objective functions and the constraints with respect to [14], effectively minimizing the mean square error (MSE) and providing details on the implementation of the optimization algorithm. Compared to [15], we use different algorithms (AGDAR and RCS instead of SSAMP) that trade-off between sparsity and noise reduction. Lastly, we consider a single user and a static pilot transmission for the initial estimate, while [16] considers the adaptation of the transmit and receive beamformers to allow channel estimation simultaneously for more users. An extensive simulation campaign on third-generation partnership project (3GPP) channel models [3] for a downlink scenario has been conducted to show the merits of the proposed approach in terms of both estimate MSE and computational complexity.

The rest of the paper is organized as follows. We introduce the system model in Section 2, providing the description of both the mm-Wave channel model and the existing OMP solutions. The sparse channel estimation problem is introduced in Section 3, together with a discussion on sectorization and channel tracking. The proposed AGDAR technique is described in Section 4, together with the refined RCS approach. Numerical results are presented in Section 5 to assess the performance of the considered techniques in a 5G scenario, before conclusions are driven in Section 6.

2 System model

We consider a massive MIMO narrowband communication system with N_t antennas at the transmitter and N_r antennas at the receiver. This models indifferently either the uplink or the downlink of a cellular communication system. Let $\mathbf{H} \in \mathbb{C}^{N_r \times N_t}$ be the channel matrix with complex entries. Antennas are organized into either uniform linear arrays (ULAs) [19] or uniform planar arrays (UPAs) [20] at both the transmitter and receiver: ULA antennas are uniformly spaced along the z axis while UPA antennas are uniformly tiled over the yz-plane[1]. For the sake of a clearer explanation in the main body of the paper, we only provide derivations for ULA, while we report in Appendix A the results for an UPA with $D_2 \times D_3 = N_t$ transmit antennas and $D_0 \times D_1 = N_r$ receive antennas.

We indicate with L the number of paths for the signal from the transmitter to the receiver, so that the channel matrix entries can written as

$$[\boldsymbol{H}]_{i_1,i_2} = \sum_{l=1}^{L} \alpha_l e^{2\pi \eta_1^l i_1 j} e^{2\pi \eta_2^l i_2 j}, \qquad (1)$$

where $i_1 = 0, \ldots, N_r - 1$, $i_2 = 0, \ldots, N_t - 1$, $|\eta_d^l| \leq \frac{\delta}{\lambda}$, $d = 1, 2$, λ is the carrier wave length, δ is the antenna spacing, and α_l is the l-path amplitude including path loss, shadowing, and fading. Note that parameters η_d^l are related to the angles of departure and arrival of the l-th path. By assuming $\delta \leq \lambda/2$, we have $\eta_d^l \in \left[-\frac{1}{2}, \frac{1}{2}\right]$. The statistics of each parameter depend on the considered propagation scenario, and various relevant cases can be found for example in the 3GPP mm-Wave channel model [3] including channel models with clustered sub-paths [3], where L becomes the total number of (sub-)paths from all clusters. Typically, in mm-Wave systems, the number of paths (or sub-paths) L is small [21].

Figure 1 shows an example of receiver with $N_r = 3$ receive antennas and a single path arriving at the antennas with an angle ϑ from a distance D: in this case $\eta_1^1 = \frac{\delta}{\lambda} \cos \vartheta$ and $\alpha_1 = \frac{1}{D^2} e^{-j2\pi \frac{D}{\lambda}}$.
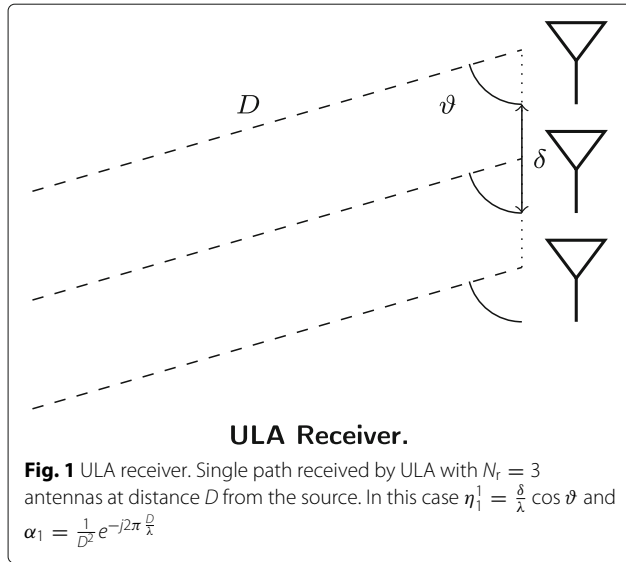
2.1 LS estimate

The considered channel estimation techniques in this paper are all based on the LS channel estimate, briefly summarized here.

The set of N_t training symbols[2] transmitted with the N_t antennas are collected into the $N_t \times N_t$ matrix $\boldsymbol{S}$, assumed here to be unitary. The corresponding received $N_t \times N_r$ matrix signal is

$$\boldsymbol{R} = \boldsymbol{H}\boldsymbol{S} + \boldsymbol{N}, \qquad (2)$$

where $\boldsymbol{N} \in \mathbb{C}^{N_r \times N_t}$ is the noise matrix with independent and identically distributed (iid) zero-mean complex Gaussian entries, each with power σ^2. The LS channel estimate at the receiver is obtained as [22]

$$\boldsymbol{H}' = \boldsymbol{R}\boldsymbol{S}^{-1} = \boldsymbol{H} + \boldsymbol{N}', \qquad (3)$$

Fig. 1 ULA receiver. Single path received by ULA with $N_r = 3$ antennas at distance D from the source. In this case $\eta_1^1 = \frac{\delta}{\lambda} \cos \vartheta$ and $\alpha_1 = \frac{1}{D^2} e^{-j2\pi \frac{D}{\lambda}}$

where N' is a matrix with iid zero-mean complex Gaussian entries having power σ^2 (thanks to the unitary property of S). Note that this estimation procedure may yield a significant overhead for the transmission of training sequences only when the number of transmit antennas grows large [23], since the number of transmitted symbols (the columns of S) is N_t. We will further address this problem in Section 3.

2.2 OMP methods

We will compare our channel estimation algorithm with two OMP techniques: single peak cancelation (SPC) [6] and joint peak cancelation (JPC) [8]. Both methods use the Fourier transform of the channel matrix, in what is usually denoted as *virtual channel* or *angular domain representation* [24, Sec. 7.3.3]. With reference to ULA, let $I_n(x) = \frac{\sin(\pi x)}{n \sin(\pi \frac{x}{n})}$ be the 1D-periodic sinc function and let

$$[\, W\left(\mathbf{\Omega}^l\right)\,]_f = \left[\frac{N_r}{M_1} I_{N_r}\left(\frac{N_r\left(f_1 + \Omega_1^l\right)}{M_1}\right) e^{j\pi\left(f_1 + \Omega_1^l\right)\frac{N_r-1}{M_1}}\right] \times \left[\frac{N_t}{M_2} I_{N_t}\left(\frac{N_t(f_2 + \Omega_2^l)}{M_2}\right) e^{j\pi\left(f_2 + \Omega_2^l\right)\frac{N_t-1}{M_2}}\right],$$

$$(4)$$

be the two-dimensional (2D) sampled periodic sinc function, where M_1 and M_2 are the number of samples per period in the 2D virtual channel domain, $\mathbf{f} = (f_1, f_2)$, $f_d = 0, \ldots, M_d - 1, d = 1, 2$, are the indices of the samples, and $\mathbf{\Omega}^l = \left(\Omega_1^l, \Omega_2^l\right) = \left(M_1 \eta_1^l, M_2 \eta_2^l\right)$.

The virtual channel matrix $V \in \mathbb{C}^{M_1 \times M_2}$ is the 2D-discrete Fourier transform (DFT) of H with entries [6]

$$[V]_f = \frac{1}{M_1 M_2} \sum_{i_1=0}^{N_r-1} \sum_{i_2=0}^{N_t-1} [H]_{(i_1, i_2)} e^{-\frac{2\pi f_1 i_1 j}{M_1}} e^{-\frac{2\pi f_2 i_2 j}{M_2}}$$

$$= \sum_{l=1}^{L} \alpha_l \left[W(\mathbf{\Omega}^l) \right]_f, \qquad (5)$$

where $\mathbf{f} = (f_1, f_2)$ are the indices of the channel sample in the virtual domain.

The SPC method [6] reported in Algorithm 1 (for *algo* = SPC), iteratively estimates the amplitude α_l and the *discrete* positions $\mathbf{\Omega}^l$ of I paths in the virtual channel and cancels their corresponding periodic sinc functions in the virtual channel. After I iterations, the channel estimate $\widehat{H}$ is obtained by taking the 2D-inverse discrete Fourier transform (IDFT) of the estimated virtual channel $\widehat{V}$ reconstructed by summing the contributions of all the detected paths.

Algorithm 1 SPC and JPC channel estimation methods

Input: $H', I, algo$
Output: $\widehat{H}$.

1: $V' \leftarrow$ DFT of H'
2: **for** $l = 1$ to I **do**
3: $\quad \widehat{\mathbf{\Omega}}^l = \underset{f}{\arg\max} |[\, V']_f|^2$
4: $\quad$ **if** $algo =$ JPC **then**
5: $\quad\quad V' \leftarrow$ DFT of H'
6: $\quad\quad v' = \text{vec}(V')$
7: $\quad\quad w_{\cdot,l} = \text{vec}\left(W\left(\widehat{\mathbf{\Omega}}^l\right)\right)$
8: $\quad\quad [\widehat{\alpha}_1, \ldots, \widehat{\alpha}_l]^T = \left(w^H w\right)^{-1} w^H v'$
9: $\quad\quad$ **for** $q = 1$ to l **do**
10: $\quad\quad\quad V' = V' - \widehat{\alpha}_q W\left(\widehat{\mathbf{\Omega}}^q\right)$
11: $\quad\quad$ **end for**
12: $\quad$ **else if** $algo =$ SPC **then**
13: $\quad\quad \widehat{\alpha}_l = [\, V']_{\widehat{\mathbf{\Omega}}^l}$
14: $\quad\quad V' = V' - \widehat{\alpha}_l W\left(\widehat{\mathbf{\Omega}}^l\right)$
15: $\quad$ **end if**
16: **end for**
17: $\widehat{V} = \sum_{l=1}^{I} \widehat{\alpha}_l W\left(\widehat{\mathbf{\Omega}}^l\right)$
18: $\widehat{H} \leftarrow$ IDFT of $\widehat{V}$

The number of iterations (i.e., the number of detected paths) is a trade-off between L (the number of paths) and the noise level. On the one hand, it is advisable to estimate all L paths, and on the other hand, noise can make small-power paths not detectable; therefore, it is better not to estimate all of them by using $I < L$. In Section 5, we determine by simulations the optimal I that minimizes the

MSE of the channel estimate. Note that SPC provides an intrinsically approximated solution even in the absence of noise, since the peak positions $\boldsymbol{\Omega}^l$ are estimated on a fixed discrete grid.

The JPC algorithm of [8] reported in Algorithm 1 (for *algo* = JPC) is a modification of SPC that at each iteration jointly estimates the amplitudes of all previously detected peaks by the LS approach and cancels the corresponding periodic sinc's from the virtual channel. In particular, $\boldsymbol{x} = \text{vec}(\boldsymbol{X})$ stacks the columns of matrix $\boldsymbol{X}$ into the column vector $\boldsymbol{x}$; at iteration l, one peak is detected (line 3) and then the amplitudes of all previously detected peaks are jointly estimated (line 8), and the new virtual channel with removed peaks is obtained (lines 9–11). This is achieved by building matrix $\boldsymbol{w}$ that contains in column l the vector version of $\boldsymbol{W}\left(\widehat{\boldsymbol{\Omega}}^l\right)$ (line 6).

This algorithm has the advantage over the SPC that each amplitude estimate is refined at each iteration thus taking advantage also of the peaks detected in further iterations.

3 Sparse dual channel estimation

In order to obtain an efficient and simple channel estimator, we exploit the specific channel structure described in the previous section. In particular, we use the fact that the channel is composed of a small number of paths with respect to the typically large number of transmit and receive antennas.

In this paper, we directly refer to the representation (1) and interpret it as 2D-IDFT of a sparse matrix having only L non-zero entries. First, the channel $\boldsymbol{H}$ is rearranged into the channel column vector $\boldsymbol{h} = \text{vec}(\boldsymbol{H}) \in \mathbb{C}^{N_r N_t \times 1}$ with entries

$$[\boldsymbol{h}]_{i_1 + N_r i_2} = [\boldsymbol{H}]_{i_1, i_2}, \tag{6}$$

where $i_1 = 0, \ldots, N_r - 1$, $i_2 = 0, \ldots, N_t - 1$, while the 2D-IDFT matrix is $\boldsymbol{F} \in \mathbb{C}^{N_t N_r \times M_2 M_1}$ with entries

$$[\boldsymbol{F}]_{(i_1 + N_r i_2, f_1 + M_1 f_2)} = \prod_{d \in \{1,2\}} e^{\frac{2\pi i f_d i_d}{M_d}}, \tag{7}$$

where $f_d = 0, \ldots, M_d - 1$, $d = 1, 2$. Lastly, we define the column vector $\boldsymbol{v}$ of length $M_1 M_2$ with L non-zero entries at position $\bar{\Omega}_1^l + M_1 \bar{\Omega}_2^l$, for $l = 1, \ldots, L$, i.e.,

$$[\boldsymbol{v}]_{\bar{\Omega}_1^l + M_1 \bar{\Omega}_2^l} = \alpha_l, \tag{8}$$

with

$$\bar{\Omega}_1^l = \langle \eta_1^l M_1 \rangle, \, \bar{\Omega}_2^l = \langle \eta_2^l M_2 \rangle, \tag{9}$$

and $\langle x \rangle$ denotes the integer part of x. From (1), we can approximate the channel vector as

$$\boldsymbol{h} \approx \boldsymbol{F} \boldsymbol{v}. \tag{10}$$

We will denote with $\boldsymbol{v}$ as the dual channel in the $M_1 \times M_2$ domain. Note that the dual channel is sparse as it contains only L non-zero entries.

Remark 1 *The approximation (10) stems from the rounding of (9), i.e., from approximating Ω_d^l with $\bar{\Omega}_d^l$. As $M_d \to \infty$, the approximation becomes more accurate. Moreover, we have used DFTs with M_d points along dimension d, as for the dual channel representation, in order to make a simpler comparison among various channel estimation schemes. Lastly, note that $\boldsymbol{v}$ is not the vectorial representation of the virtual channel, since the DFT used to obtain the virtual channel does not invert the IDFT of (10): in fact, the DFT is taken on the reduced set of $N_r \times N_t$ samples, thus yielding the periodic* sinc*'s of (4).*

Similarly to the vectorial representation of the channel, we define

$$\boldsymbol{h}' = \text{vec}(\boldsymbol{H}') \approx \boldsymbol{F} \boldsymbol{v} + \boldsymbol{n}', \tag{11}$$

where indices of $\boldsymbol{H}'$ to obtain $\boldsymbol{h}'$ are selected similarly to (6). From (11), we observe that the LS estimate is a noisy version of a linear transformation of the dual channel $\boldsymbol{v}$.

We propose an algorithm that improves the LS channel estimate by exploiting the sparsity of $\boldsymbol{v}$. In particular, we define by $\hat{\boldsymbol{v}}$ the new estimate of the dual channel $\boldsymbol{v}$ and write the sparse channel estimation problem as

$$\widehat{\boldsymbol{v}} = \underset{\boldsymbol{v}'}{\text{argmin}} \left(\|\boldsymbol{F} \boldsymbol{v}' - \boldsymbol{h}'\|_2^2 + \rho \|\boldsymbol{v}'\|_0 \right), \tag{12}$$

where $\|\boldsymbol{v}'\|_0$ is the ℓ_0-norm that counts the non-zero elements in $\boldsymbol{v}'$ and ρ is a parameter that controls the sparsity of the solution. This problem formulation aims at minimizing the MSE between the estimated channel and the LS channel estimate, under a constraint on the sparsity of vector $\hat{\boldsymbol{v}}$, imposed by the norm-zero term.

Unfortunately, problem (12) is non-convex and NP-hard [25]. Thus, we relax the problem by replacing the ℓ_0-norm with the ℓ_1-norm obtaining the LASSO problem

$$\widehat{\boldsymbol{v}} = \underset{\boldsymbol{v}}{\text{argmin}} \left(\|\boldsymbol{F} \boldsymbol{v} - \boldsymbol{h}'\|_2^2 + \rho \|\boldsymbol{v}\|_1 \right), \tag{13}$$

with $\|\boldsymbol{v}\|_1 = \sum_{i=0}^{M_1 M_2 - 1} |v_i|$, which is now convex.

Remark 2 *Note that estimating the dual channel opens the possibility of reducing the training overhead. We observe that systems with different numbers of antennas (placed at the same position) share the same dual channel. Thus, once we have an estimate of $\boldsymbol{v}$, we can change $\boldsymbol{F}$ to obtain the channel estimate for a different antenna setting. Indeed, we can use a fewer transmit antennas to transmit the training sequence, then obtain an estimate of the dual channel and finally project the estimate into a larger number of antennas by modifying the size of the IDFT matrix in (10). Typically paths are concentrated in clusters on part of the dual channel, thus by an iterative channel estimation procedure, we can beamform training signals in the part of the dual channel covered by the clusters.*

4 Solution of the sparse channel estimation problem

A vast literature is available for the solution of the sparse channel estimation problem (13), see for example the survey [26]. We propose here to use two recent and efficient methods based on the gradient descent algorithm with improved convergence speed, namely the AGDAR algorithm (also named FISTA with adaptive restart [27]) and the RCS algorithm [18].

4.1 Accelerated gradient descent with adaptive restart

The AGDAR algorithm [27] has been developed to solve problems where the objective function is the sum of a differentiable function and a general but simple closed convex function.

Here, we briefly summarize the motivation of the AGDAR algorithm. We first observe that the minimization problem $\min_{\boldsymbol{x} \in \mathbb{R}^N} f(\boldsymbol{x})$ when $f(\cdot)$ is convex and *smooth* can be solved by the gradient descent algorithm that iteratively updates the solution, computing at iteration p

$$\boldsymbol{x}_p = \boldsymbol{x}_{p-1} - t\nabla f(\boldsymbol{x}_{p-1})\,, \tag{14}$$

where t is the step size and $\nabla f(\boldsymbol{x})$ is the gradient of $f(\cdot)$ computed in $\boldsymbol{x}$. An alternative formulation of (14) is provided by the *proximal form* [28]

$$\boldsymbol{x}_p = \operatorname{argmin}_{\boldsymbol{x}} \nabla f(\boldsymbol{x}_{p-1})^T(\boldsymbol{x} - \boldsymbol{x}_{p-1}) + \frac{1}{2t}||\boldsymbol{x} - \boldsymbol{x}_{p-1}||_2^2\,. \tag{15}$$

Now, in order to minimize $f(\boldsymbol{x}) + g(\boldsymbol{x})$ with $f(\cdot)$ still convex and smooth but $g(\cdot)$ convex, nondifferentiable, and lower semicontinuous, the proximal form must be modified as follows [28]

$$\begin{aligned}\boldsymbol{x}_p =&\operatorname{argmin}_{\boldsymbol{x}} \nabla f(\boldsymbol{x}_{p-1})^T(\boldsymbol{x} - \boldsymbol{x}_{p-1}) + \frac{1}{2t}||\boldsymbol{x} - \boldsymbol{x}_{p-1}||_2^2 \\ &+ g(\boldsymbol{x}) = \operatorname{argmin}_{\boldsymbol{x}} g(\boldsymbol{x}) + \frac{1}{2t}||\boldsymbol{x} - (\boldsymbol{x}_{p-1} - t\nabla f(\boldsymbol{x}_{p-1}))||_2^2\,.\end{aligned} \tag{16}$$

In general, this optimization problem may be hard to solve; however, when $g(\boldsymbol{x}) = \rho||\boldsymbol{x}||_1$, problem (16) is efficiently solved by splitting it into N separate one-dimensional problems for each entry of $\boldsymbol{x} \in \mathbb{R}^N$, i.e.,

$$[\boldsymbol{x}_p]_n = \operatorname{argmin}_x \rho|x| + \frac{1}{2t}|x - z_n|^2\,, \tag{17}$$

with $z_n = [\boldsymbol{x}_{p-1} - t\nabla f(\boldsymbol{x}_{p-1})]_n$. Problem (17) can be solved in closed form [29] as $[\boldsymbol{x}_p]_n = \mathcal{T}_{\rho t}(z_n)$, with the shrinkage operator defined as

$$\mathcal{T}_a(x) = \operatorname{sgn}(x)\max(|x| - a, 0)\,, \tag{18}$$

obtaining the iterative shrinkage-thresholding algorithm (ISTA) algorithm. This solution can be made faster by applying the Nesterov acceleration principle [30]: instead of using the gradient descent (14), $\boldsymbol{x}_p$ is updated as a linear combination of the gradient descent terms (14) at the current and previous iterations, i.e.,

$$\boldsymbol{y}_p = \boldsymbol{x}_{p-1} - t\nabla f(\boldsymbol{x}_{p-1})\,, \tag{19a}$$

$$\boldsymbol{x}_p = (1 - \gamma_{p-1})\boldsymbol{y}_p + \gamma_{p-1}\boldsymbol{y}_{p-1}\,. \tag{19b}$$

Combining this approach with ISTA, we obtain the FISTA algorithm where $[\boldsymbol{y}_p]_n = \mathcal{T}_{\rho t}(z_n)$ and $\boldsymbol{x}_p$ is updated using (19b) It turns out that this approach fastens the convergence of the algorithm for example by choosing as linear combination coefficients [30]

$$\gamma_p = \begin{cases} 0 & p = 0 \\ \frac{1 - \theta_{p-1}}{\theta_p} & p > 0\,, \end{cases} \tag{20}$$

$$\theta_p = \begin{cases} 1 & p = 0 \\ \left(1 + \sqrt{1 + 4\theta_{p-1}^2}\right)/2 & p > 0\,. \end{cases} \tag{21}$$

The explanation of the Nesterov iteration is not very intuitive, and the interested reader can find more details in [30].

The parameter choice (20) is not in general optimal while its optimization is a difficult task. An alternative approach is the *adaptive restart* technique [27], in which γ_p is set according to (20) (thus in a suboptimal way) but the FISTA algorithm is restarted whenever the objective function is locally increasing (thus the iterative solution is moving in the wrong direction), i.e., when

$$\nabla f(\boldsymbol{x}_{p-1})^T(\boldsymbol{y}_p - \boldsymbol{y}_{p-1}) > 0\,, \tag{22}$$

From (19a) we obtain that the restarting condition (22) can be written as

$$(\boldsymbol{x}_{p-1} - \boldsymbol{y}_p)(\boldsymbol{y}_p - \boldsymbol{y}_{p-1}) > 0\,. \tag{23}$$

The restart consists in resetting $\theta_p = 1$ and using as initial point the last point produced by the algorithm.

4.2 Re-weighted compressed sensing

The RCS method is proposed [18] to improve the sparsity of the gradient descent solution of (13). The algorithm weights the entries of $\boldsymbol{x}_p$ in the ℓ_1-norm in (13) in order to better approximate the ℓ_0-norm term in (12). Therefore, instead of solving (12), the RCS method aims at solving problem

$$\widehat{\boldsymbol{v}} = \operatorname{argmin}_{\boldsymbol{v}'} \left(||\boldsymbol{F}\boldsymbol{v}' - \boldsymbol{h}'||_2 + \rho||\boldsymbol{D}\boldsymbol{v}'||_1\right)\,, \tag{24}$$

where the diagonal matrix $\boldsymbol{D}$ contains the weights. This problem can be seen as a ℓ_1-norm relaxation of a weighted version of the ℓ_0-norm problem (12), i.e.,

$$\widehat{\boldsymbol{v}} = \operatorname{argmin}_{\boldsymbol{v}'} \left(||\boldsymbol{F}\boldsymbol{v}' - \boldsymbol{h}'||_2^2 + \rho||\boldsymbol{D}\boldsymbol{v}'||_0\right)\,. \tag{25}$$

Algorithm 2 AGDAR algorithm

Input: $\boldsymbol{h}', \rho, \epsilon, q_1, t$
Output: $\widehat{\boldsymbol{h}}$

1: $\boldsymbol{x}_0 = \boldsymbol{y}_0 = \boldsymbol{0}, \theta_0 = 1, p = 1, e = 2\epsilon$
2: **while** $(p < q_1$ **and** $e > \epsilon)$ **do**
3:     $\boldsymbol{z} = \boldsymbol{x}_{p-1} - t\boldsymbol{F}^H \left(\boldsymbol{F}\boldsymbol{x}_{p-1} - \boldsymbol{h}'\right)$
4:     $[\boldsymbol{y}_p]_n = \mathcal{T}_{\rho t}(z_n), \quad n = 1, \ldots, M_1 M_2$
5:     $\theta_p = \left(1 + \sqrt{1 + 4\theta_{p-1}^2}\right)/2$
6:     **if** $\left(\boldsymbol{x}_{p-1} - \boldsymbol{y}_p\right)\left(\boldsymbol{y}_p - \boldsymbol{y}_{p-1}\right) > 0$ **then**
7:       $\boldsymbol{x}_p = \boldsymbol{y}_p; \theta_p = 1$
8:     **else**
9:       $\boldsymbol{x}_p = \boldsymbol{y}_p + \frac{\theta_{p-1}-1}{\theta_p}\left(\boldsymbol{y}_p - \boldsymbol{y}_{p-1}\right)$
10:    **end if**
11:    $e = \|\boldsymbol{x}_{p-1} - \boldsymbol{x}_p\|_2^2$
12:    $p = p + 1$
13: **end while**
14: $\widehat{\boldsymbol{h}} = \boldsymbol{F}\boldsymbol{x}_p$

Algorithm 3 RCS algorithm

Input: $\boldsymbol{h}', \rho, \epsilon, q_1, q_2, t$ and ζ
Output: $\widehat{\boldsymbol{h}}$

1: $\boldsymbol{D}_0 = \text{diag}\{[1, \ldots, 1]\}$
2: **for** $i = 0$ to $q_2 - 1$ **do**
3:    $\boldsymbol{x}_0 = \boldsymbol{y}_0 = \boldsymbol{0}, \theta_0 = 1, p = 1, e = 2\epsilon$
4:    **while** $(p < q_1$ **and** $e > \epsilon)$ **do**
5:      $\boldsymbol{z} = \boldsymbol{x}_{p-1} - t\boldsymbol{F}^H \left(\boldsymbol{F}\boldsymbol{x}_{p-1} - \boldsymbol{h}'\right)$
6:      $[\boldsymbol{y}_p]_n = \mathcal{T}_{D_{n,n}\rho t}(z_n), n = 1, \ldots, M_1 M_2$
7:      $\theta_p = \left(1 + \sqrt{1 + 4\theta_{p-1}^2}\right)/2$
8:      **if** $\left(\boldsymbol{x}_{p-1} - \boldsymbol{y}_p\right)\left(\boldsymbol{y}_p - \boldsymbol{y}_{p-1}\right) > 0$ **then**
9:        $\boldsymbol{x}_p = \boldsymbol{y}_p; \theta_p = 1$
10:     **else**
11:       $\boldsymbol{x}_p = \boldsymbol{y}_p + \frac{\theta_{p-1}-1}{\theta_p}\left(\boldsymbol{y}_p - \boldsymbol{y}_{p-1}\right)$
12:     **end if**
13:     $e = \|\boldsymbol{x}_{p-1} - \boldsymbol{x}_p\|_2^2$
14:     $p = p + 1$
15:    **end while**
16:    $\boldsymbol{D}_{i+1} = \text{diag}\left\{\frac{1}{|\boldsymbol{x}_p| + \zeta}\right\}$
17: **end for**
18: $\widehat{\boldsymbol{h}} = \boldsymbol{F}\boldsymbol{x}_p$

As the ℓ_0-norm counts the non-zero entries, regardless of their amplitude, for non-zero weights, the two problems (12) and (25) have the same solutions.

About the choice of the weights, they are meant to provide a good approximation of the ℓ_0-norm using the (weighted) ℓ_1-norm. Therefore, imposing that at solution

$$\|\boldsymbol{D}\hat{\boldsymbol{v}}\|_1 = \|\hat{\boldsymbol{v}}\|_0, \tag{26}$$

we obtain the optimum weights (diagonal entries of matrix $\boldsymbol{D}$)

$$[\boldsymbol{D}]_{k,k} = \begin{cases} \frac{1}{|\hat{v}_k|}, & \hat{v}_k \neq 0, \\ *, & \hat{v}_k = 0, \end{cases} \tag{27}$$

where $*$ denotes any non-zero value. However, this choice requires the knowledge of the problem solution $\hat{\boldsymbol{v}}$, which is not available while solving the problem.

In [18], an iterative approach has been proposed, where the weights are adapted to converge to (27) without knowing the optimal solution. In particular, RCS runs q_2 times the AGDAR algorithm, using at each iteration a different set of weights chosen according to (28). It has been shown by extensive simulations over a variety of examples that the following weight adaptation strategy is performing well: starting from $\boldsymbol{D}_0 = \text{diag}\{[1, \ldots, 1]\}$, for which (24) corresponds to (13), and then at iteration $i + 1$ update the weights as

$$[\boldsymbol{D}_{i+1}]_{j,j} = \frac{1}{|[\boldsymbol{x}_p]_j| + \zeta}, \quad j = 1, \ldots, M_1 M_2, \tag{28}$$

where $\boldsymbol{x}_p$ is the solution of (24) for weights $\boldsymbol{D}_i$ and ζ is a small number. Note that at convergence (for $\zeta \approx 0$) we obtain (27).

The implementation of RCS is obtained by running q_2 times the AGDAR algorithm and computing the shrinkage function with a weighted parameter, i.e., $\mathcal{T}_{w_n \rho t}(z_n)$. The resulting procedure is reported in Algorithm 3.

It has been shown [18] that the RCS algorithm is a majorization-minimization algorithm that iteratively minimizes a simple surrogate function majoring the objective function, and indeed provides in general a better approximation to the original ℓ_0-norm problem.

4.3 Sectorization and channel tracking
When the antennas at either or both the transmitter and the receiver are transmitting/receiving in a focused direction (in what is known as cell *sectorization*), the departure and arrival angles are within sub-intervals of $[0, 2\pi)$; therefore, also η_d^l will take value in sub-intervals of $[-1/2, 1/2)$ and the rounding of $\eta_d^l M_d$ will be in a sub-interval of $\left(-\frac{M_d}{2}, \frac{M_d}{2}\right)$. Therefore, vector $\boldsymbol{v}$ can be reduced by eliminating the entries corresponding to values of η_d^l that are never taken by the channel realization. Correspondingly, the columns of $\boldsymbol{F}$ are removed and the AGDAR algorithm is run over a reduced space, thus increasing its accuracy.

About channel tracking, once the channel has been estimated, it may slowly change due to the variations of the propagation environment. In this case, we can reduce the complexity of the channel estimation and make it

more effective by simply tracking its changes rather than starting from scratch its estimation. We propose to focus the search of the paths in the dual channel within intervals around the initial estimates of arrival and departure angles. Therefore, for both AGDAR and RCS approaches, we have a reduction of the dual channel vector $\boldsymbol{v}$. Indeed, this is similar to sectorization; however, for channel tracking we must consider multiple angle intervals, one for each initially estimated path.

Sectorization or tracking are also possible for SPC and JPC, wherein the search of the peak positions $\boldsymbol{\Omega}^l$ will be done on a sub-grid of the $M_1 \times M_2$ grid, according to the intervals of η_1^l and η_2^l. Also in this case, the benefits for the channel estimation process will be a lower probability of periodic sinc misplacement, as the search space is smaller.

4.4 Computational complexity
The computational complexity of the considered algorithms is evaluated in terms of complex multiplications (CMUX), complex additions (CADD), and comparisons (COM). Let $M_{tot} = \prod_d M_d$, where the product is along all dimensions, depending on the use of either UPA or ULA. Hence, using the fast Fourier transform algorithm, a (I)DFT of M_{tot} samples requires $M_{tot} \log_2 M_{tot}$ CMUX, $M_{tot} \log_2 M_{tot}$ CADD and no comparison. A summary of the computational complexity of the various algorithms is reported in Table 1, where q_3 denotes the effective number iterations over the variable p of the AGDAR and RCS algorithms. In the following section, we will present numerical results also for a complexity comparison among the various techniques.

5 Numerical results
In this section, we compare the proposed channel estimation techniques by evaluating the MSE in decibels (dB)

$$\text{MSE} = 10 \log_{10} \mathbb{E}\left[||\widehat{\boldsymbol{H}} - \boldsymbol{H}||_2^2\right], \tag{29}$$

where $\mathbb{E}[\cdot]$ denotes the expectation operator and $\hat{\boldsymbol{H}}$ is the estimated channel matrix.

We consider the urban macro cell (UMa), urban micro cell (UMi), rural macro cell (RMa), and Indoor Hotspot (InH) 3GPP channel models [3], with both line-of-sight (LOS) and non-line-of-sight (NLOS). In these scenarios, the number of clusters (typically from 4 to 20) depends on the channel model and the number of sub-paths is 20 per cluster, thus totaling L in the range of tens to hundreds. Note that although the number of sub-paths is large, only three or four sub-paths have a notable power. Therefore de facto, we find the sparse channel model described in this paper and in many literature papers and measurement campaign results. Channels are obtained for a downlink, where the base station (BS) and user equipment (UE) are on the same plane, with parameters defined in Table 2. The average channel gain is unitary, so we assume that transmit power has been adapted to compensate for the path loss; therefore, the average SNR is the reciprocal of the noise power. This also provides that the MSE for the LS estimate is simply the reciprocal of the SNR, which is then not reported in the figures.

For the proposed AGDAR and RCS algorithms, we use parameters as in Table 2. In the following, we will always consider the same antenna geometry (either ULA or UPA) at both BS and UE, with a different number of antennas at the two ends.

5.1 Parameters setting
We first evaluate the impact of the parameters on the channel estimate MSE. The performance of both AGDAR and RCS is determined by the parameter ρ that weights the ℓ_1-norm term in the objective function and should be chosen according to the channel sparsity and the operating signal to noise ratio (SNR). For each value of SNR, we have assessed the optimum value of ρ that minimizes the average MSE of the channel estimate. The results are reported in Fig. 2 for both (at both devices) with $N_r = 16 = 4 \times 4$ ($D_0 = D_1 = 4$) and $N_t = 4 = 2 \times 2$

Table 1 Computational complexity of the channel estimation methods

Method	Complex multiplications	Additions	Comparisons
SPC	$2M_{tot} \log_2 M_{tot} + 3IM_{tot}$	$2M_{tot} \log_2 M_{tot} + 2IM_{tot}$	IM_{tot}
JPC	$2M_{tot} \log_2 M_{tot} + \frac{I(I+1)}{2}M_{tot}$	$2M_{tot} \log_2 M_{tot} + \frac{I(I+1)}{2}M_{tot} + 2IM_{tot}$	IM_{tot}
AGDAR	$3M_{tot}q_3 + 2q_3 M_{tot} \log_2 M_{tot}$	$2q_3 M_{tot} \log_2 M_{tot} + 5M_{tot}q_3$	$M_{tot}q_3$
RCS	$3M_{tot}q_2q_3 + 2M_{tot}q_3 + 2q_2q_3 M_{tot} \log_2 M_{tot}$	$M_{tot}q_3 + 2q_2q_3 M_{tot} \log_2 M_{tot} + 5M_{tot}q_2q_3$	$M_{tot}q_2q_3$

Table 2 Simulation parameters

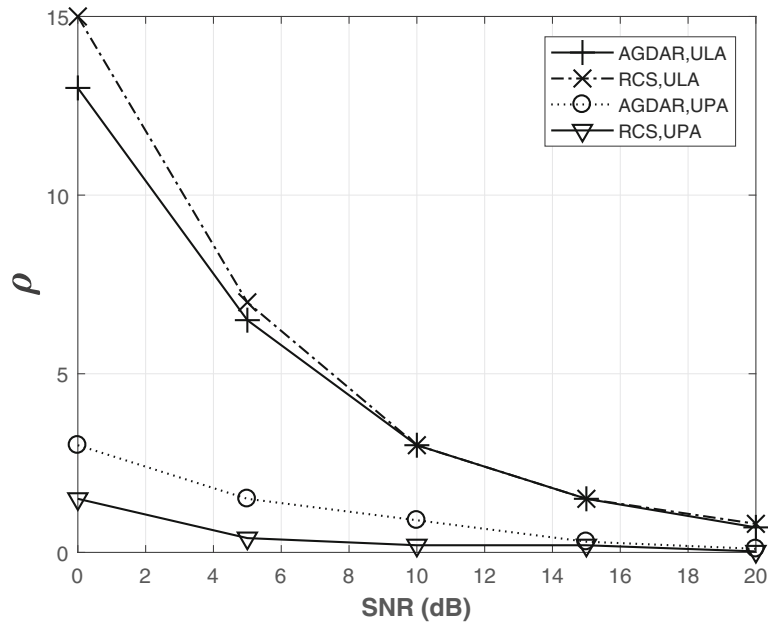| Parameter	Value
number of sub-paths per cluster	20
BS position	$(0, 0)$ m
BS antenna altitude	25 m
UE position	$(90, 15)$ m
UE antenna altitude	1.5 m
Carrier frequency	$f = 28$ GHz
δ	$\lambda/2$
Average channel gain	0 dB
ϵ	10^{-6}
q_2	3
ζ for ULA, UPA	0.1, 0.99

Fig. 2 Choice of ρ. Optimized value of ρ vs *SNR*, for both UPA, with $N_r = 16 = 4 \times 4$ ($D_0 = D_1 = 4$) and $N_t = 4 = 2 \times 2$ ($D_2 = D_3 = 2$) and $M_0 = M_1 = M_2 = M_3 = 8$, and ULAs with $N_r = 16$, $N_t = 4$ and $M_1 = M_2 = 32$. UMi LOS model

($D_2 = D_3 = 2$) and $M_0 = M_1 = M_2 = M_3 = 8$, and ULAs (at both devices) with $N_r = 16$, $N_t = 4$ and $M_1 = M_2 = 32$. In this case the channel model is UMi LOS for which the number of clusters is random, between 3 and 12, while the number of sub-paths per cluster is 20, for a maximum total of $L = 240$ paths. We can see a smooth behavior of ρ with respect to the average SNR, that can be described with simple functions for its adaptation to operating conditions. Moreover, as the average SNR increases, the optimal ρ decreases, since the LS estimate is less noisy and a limited sparsification of the channel is required. We have optimized the value of this parameter also for other conditions (e.g., different number of antennas) and forthcoming results are obtained with the optimized ρ.

A second relevant parameter for both AGDAR and RCS is the maximum number of iterations q_1. Figure 3 shows the average MSE as a function of the number of maximum allowed iterations q_1 for ULAs with $N_r = 16$, $N_t = 4$, and various DFT sizes. We note that for all DFT sizes the MSE is flooring as q_1 increases: in particular, with $\log_{10} q_1 = 2.5$, all algorithms converge to the minimum MSE. We also observe that the required number of iterations grows with (M_1, M_2), and RCS achieves a lower asymptotic MSE and converges faster.

For a comparison with SPC and JPC, we have also optimized parameter I, i.e., the number of detected paths. Note that the existing literature typically assumes the knowledge of L and sets $I = L$. We instead observe that I does not necessarily correspond to the number of paths,

since small paths can be neglected as may be easily confused with noise artifacts. This is particularly true in the 3GPP channel model, where many clusters and sub-paths are present, most of which however have a very limited power. Figure 4 shows the value of I that minimizes the average MSE versus the average SNR for various values of $M = M_1 = M_2$, using ULAs with $N_r = 16$, $N_t = 4$ and SPC. The channel model is UMi LOS. We observe that we need a large value of I when the SNR is high, as the considered channel model has many (sub-)paths, and at high SNRs, they can be distinguished from the noise. Also, note that the optimal value of I is decreasing as M increases: indeed, for higher M the approximation between positions Ω_l and $\hat{\Omega}_l$ becomes more accurate, thus fewer sinc functions (closer to the actual number of paths) is enough (and better) for channel estimation. The results reported in the rest of this section are obtained with the optimized value of I.

Lastly, we consider the optimization of the DFT size, i.e., the number of points used to approximate η_d^l in all considered channel estimation methods. In Appendix B, we provide the analytical derivation of the MSE as a function of M, for ULA and a Rayleigh fading channel with a small number of taps. Figure 5 shows the MSE as a function of $M = M_1 = M_2$ for ULA with $N_t = 8$, $N_r = 2$ and the channel model described in Appendix B. As expected the MSE decreases for a higher value of M, as the quantization error is reduced, flooring for high values of M. We then have considered a ULA system with $N_r = 32$
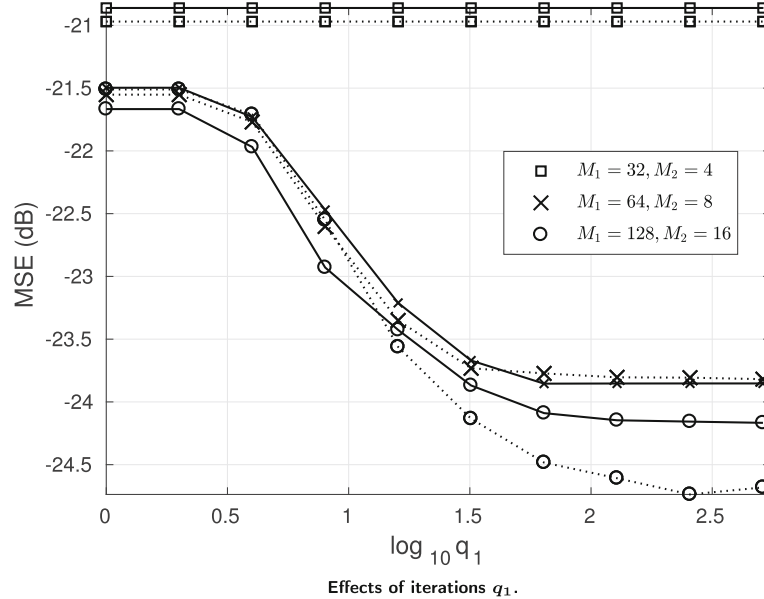
Fig. 3 Effects of iterations q_1. MSE vs the maximum number of iterations q_1 for ULA with $N_r = 32, N_t = 4$ using AGDAR (solid lines) and RCS (dotted lines). UMi LOS model

and $N_t = 4$ and a DFT size multiple μ of the number of antennas, i.e., $M_1 = \mu N_r$ and $M_2 = \mu N_t$. For a UMi LOS model, Fig. 6 shows the average MSE for an SNR of 0, 10, and 20 dB, and different channel estimation techniques. We observe that in all cases by increasing μ we obtain a better channel estimate, thanks to a better quantization precision of either the virtual channel domain (for

SPC and JPC) or the values of η_d^l (for both AGDAR and RCS). Moreover, both AGDAR and RCS methods achieve a lower MSE than SPC and JPC techniques at both low and high SNRs, thanks to their better exploitation of compact channel representation in the dual domain. The RCS has almost negligible performance improvement with respect to AGDAR, as both have a gain from 3 to 5 dB with respect
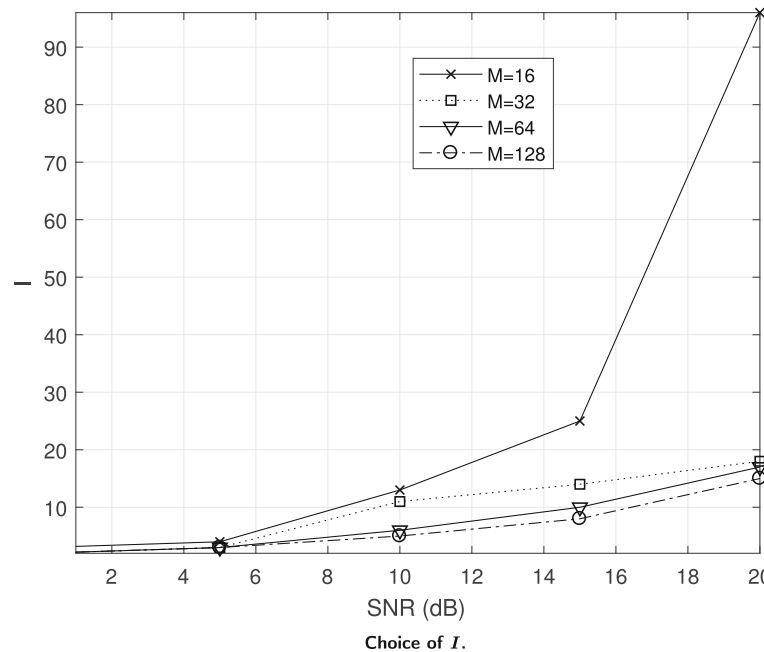
Fig. 4 Choice of I. Optimal I for both SPC and JPC methods and various values of $M = M_1 = M_2$ using ULA with $N_r = 16, N_t = 4$. UMi LOS model
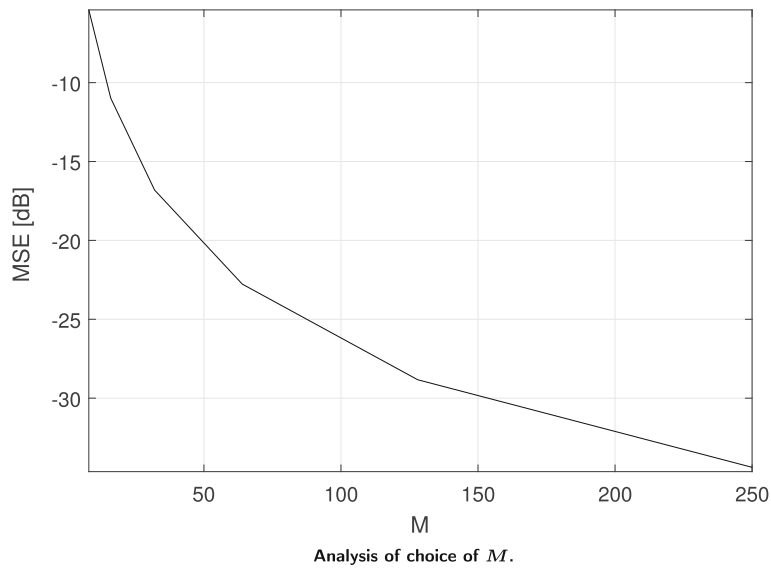
Fig. 5 Analysis of choice of M. MSE vs $M = M_1 = M_2$ for the scenario of Appendix B, with $N_t = 16$ and $N_r = 4$

to interference cancelation techniques. Note that the gain is more remarkable at a lower SNR, showing that the compressed-sensing techniques are able to better reject the noise. Lastly, note that JPC has an almost negligible improvement over SPC; thus, we can conclude that the detection of the peaks is already accurate when performed sequentially rather than in parallel. Overall, we conclude that $\mu = 2$ already provides close-to-optimal results for all methods. Similar observations can be drawn from Fig. 7, where we report the MSE for a UPA configuration with 2×2 antennas at the UE and 6×6 antennas at BS in a UMa LOS channel model.
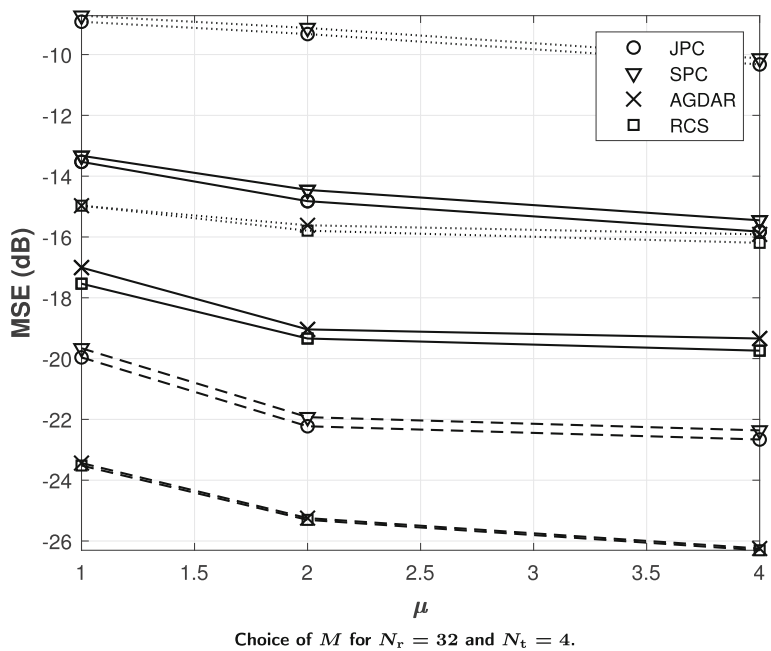
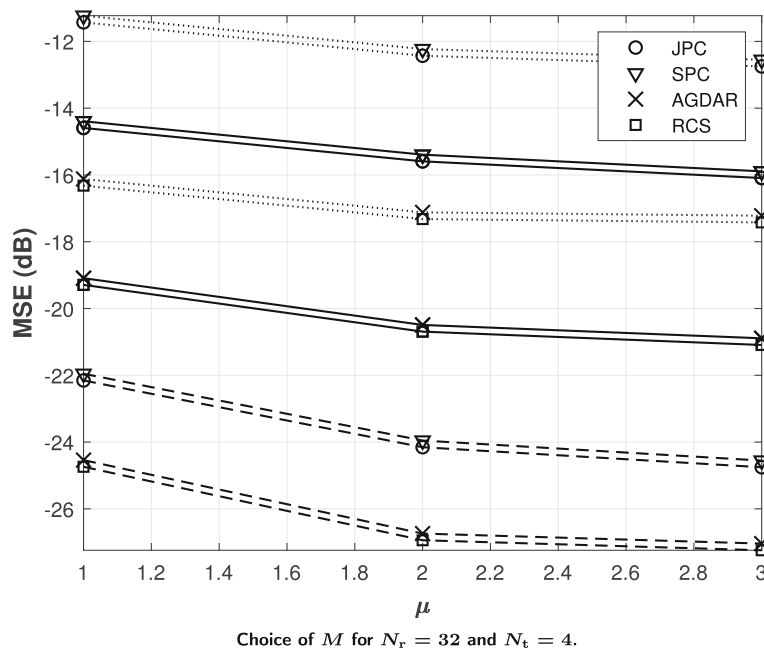**Fig. 6** Choice of M for $N_r = 32$ and $N_t = 4$. MSE vs μ for ULA with $N_r = 32$ and $N_t = 4$ at an SNR of 0 (dotted lines), 10 (solid lines) and 20 dB (dashed lines) and $M_1 = \mu N_r$ and $M_2 = \mu N_t$. UMi LOS model
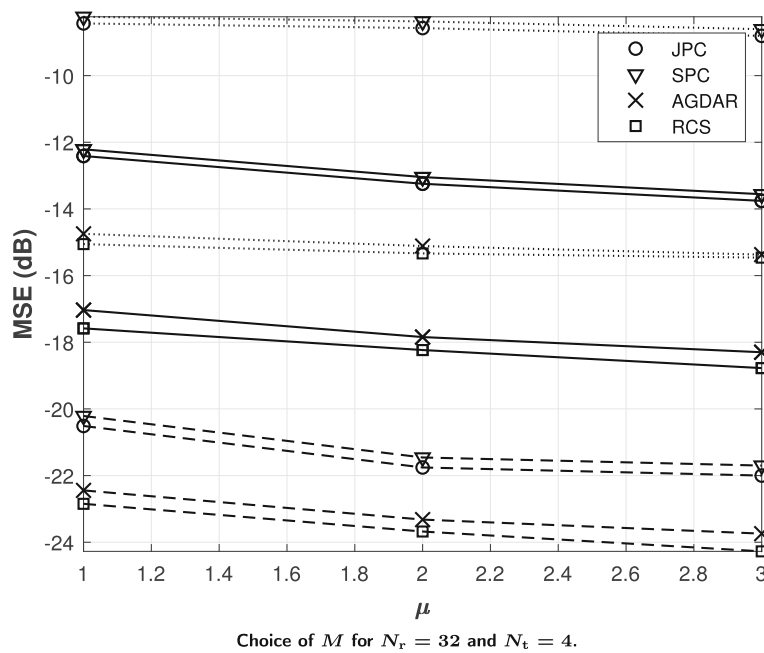
Fig. 7 Choice of M for $N_{\mathrm{r}} = 32$ and $N_{\mathrm{t}} = 4$. MSE vs μ for UPA with $N_{\mathrm{r}} = 36$ and $N_{\mathrm{t}} = 4$ at an SNR of 0 (dotted lines), 10 (solid lines) and 20 dB (dashed lines) and $M_1 = \mu N_{\mathrm{r}}$ and $M_2 = \mu N_{\mathrm{t}}$. UMa LOS model

In order to show the performance of the proposed solution in a scenario with a large number of antennas, Fig. 8 shows the MSE as a function of μ for ULA with $N_{\mathrm{r}} = 128$ and $N_{\mathrm{t}} = 16$ and the UMi LOS model. Also in this case, we can appreciate the advantage of all techniques with respect to the LS method, as we recall that for LS the MSE is the reciprocal of the SNR, thus 0 dB in correspondence of the dotted lines, −10 dB for the solid lines and −20 dB for the dashed lines. Indeed, a higher number of antennas with respect to Fig. 6 increases

Fig. 8 Choice of M for $N_{\mathrm{r}} = 32$ and $N_{\mathrm{t}} = 4$. MSE vs μ for ULA with $N_{\mathrm{r}} = 128$ and $N_{\mathrm{t}} = 16$ at an SNR of 0 (dotted lines), 10 (solid lines), and 20 dB (dashed lines) and $M_1 = \mu N_{\mathrm{r}}$ and $M_2 = \mu N_{\mathrm{t}}$. UMi LOS model

the gain of the other channel estimation techniques with respect to LS. About the comparison among the various methods, we can derive similar conclusions as those of Fig. 6, confirming also that other results are representative of a massive MIMO scenario.

5.2 Sectorization and channel tracking

As we already discussed, sectorization provides a faster and more accurate search of the channel paths. Here, we consider a system where channel angles are uniformly distributed in intervals of 6, 60, and 180 degrees. We have $L = 14$ paths, with independent Gaussian-distributed amplitudes α_l: by this simple channel model, we better capture the effects of sectorization and channel tracking.

Figure 9 shows the resulting MSE as a function of $M = M_1 = M_2$ for the various systems, when the average SNR is 10 dB. We observe that sectorization indeed reduces the MSE of all channel estimates, and sectors of 6° provide a MSE of 10 to 16 dB smaller than that of 6° sectors. Comparing the various techniques we observe that with large (180°) and small (6°) sectors all techniques take advantage of the sectorization in a similar way, while for intermediate values (60°) the compressed-sensing methods have a higher gains than interference cancelation methods.

We also consider channel tracking where, after an initial channel estimation performed according to the various considered techniques and with an angle span of

360°, when channels are time-invariant. Estimators are run using an angle span of 6° around the angles of each path. Figure 10 shows the MSE of the channel estimates at an average SNR of 10 dB, and for various values of M_1 and M_2. We observe that, thanks to the search over a smaller angle span both SPC and JPC achieve similar performance to the proposed approaches, a phenomenon that we already observed with sectorization. Also, the difference between RCS and AGDAR is further reduced, again because of the easier task of channel estimation in this case. We still note instead a high sensitivity to the DFT-size, which corresponds to an accuracy in the estimate of the angles of arrival and departure. Lastly, both sectorization and channel tracking reduce the complexity of the proposed solutions, as path search operations can be performed on a reduced space.

5.3 3GPP channel scenarios comparison

Until now, we considered only the UMi LOS channel model: in this section instead, we consider also the other 3GPP channel models. Figure 11 shows the average MSE for various algorithms and UPAs with $N_t = 8 \times 8$ at the BS and $N_r = 2 \times 2$ at the UE, and $M_0 = M_1 = 10$ and $M_2 = M_3 = 4$. We compare various channel estimation techniques for an average SNR of 10 dB. At this intermediate SNR value, we observe that the proposed AGDAR and RCS significantly outperform both SPC and JPC for all the considered channel models, by 6 to 7 dB.
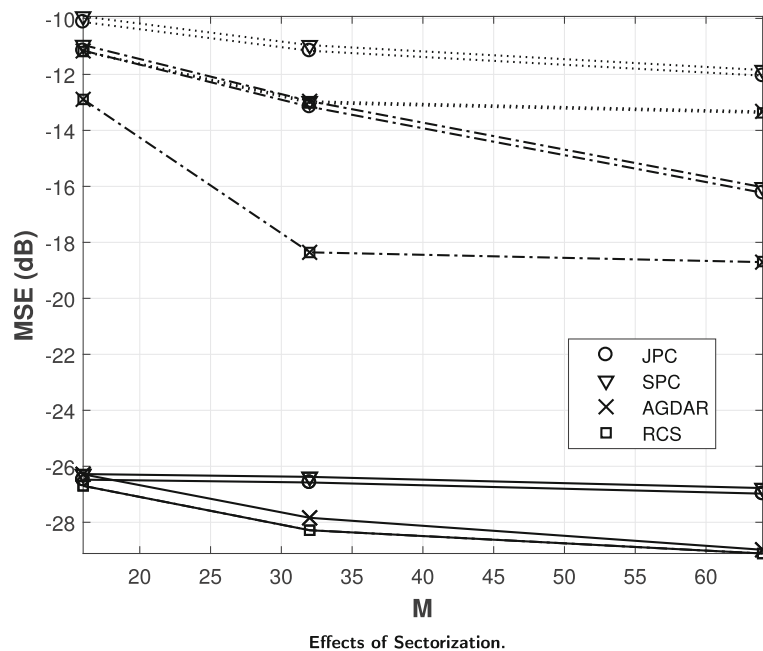

Fig. 9 Effects of sectorization. MSE vs $M = M_1 = M_2$ for ULAs and various channel estimation techniques, when the average SNR is 10 dB when sectorization is used, with 6 (solid lines), 60 (dashed lines), and 180° (dotted lines). $N_r = N_t = 16$
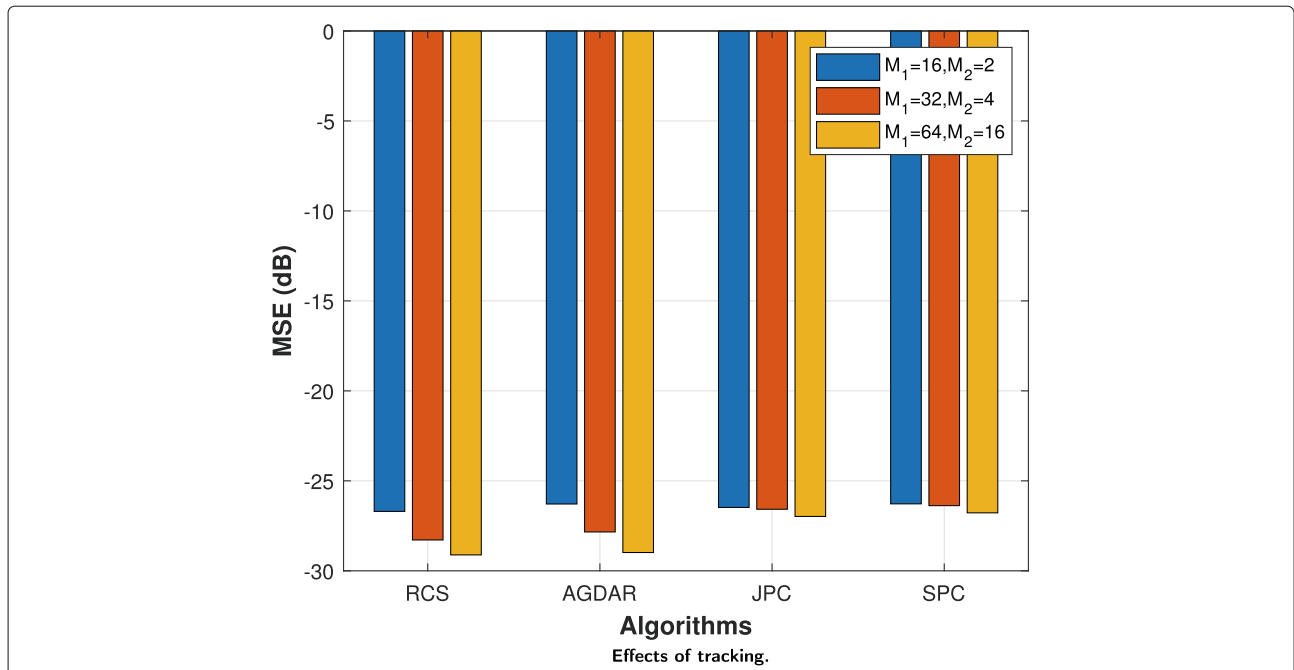
Fig. 10 Effects of tracking. MSE for ULAs of various channel estimation techniques, when the average SNR is 10 dB with channel tracking at 6°, and various values of M. $N_r = 16$ and $N_t = 4$

The indoor-office model (InH Mixed) has a few significant taps with reduced dispersion, a favorable condition for SPC and JPC, which exhibit a reduced gap with respect to the proposed (still better performing) techniques. On the other hand, dispersive channels with many low-power taps (UMi Street Canyon NLOS) make the channel estimation more problematic for SPC and JPC, while can be handled very efficiently by the compressed-sensing techniques, thanks to their ability to better distinguish between noise and channel components. This provides a
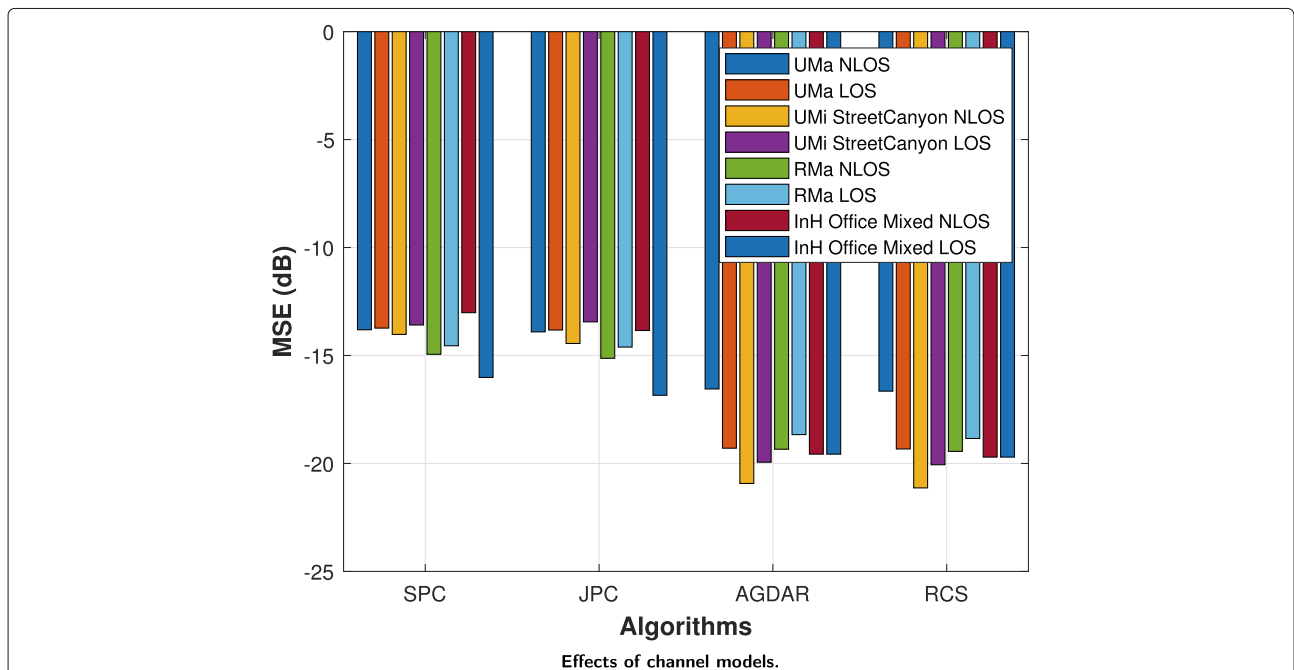
Fig. 11 Effects of channel models. MSE under different channel models with SNR of 10 dB for UPA with $N_t = 8 \times 8$ at the BS and $N_r = 2 \times 2$ at the UE, and $M_0 = M_1 = 10$ and $M_2 = M_3 = 4$

gain of 7 dB between SPC and AGDAR. We also note that both AGDAR and RCS have comparable performance across all channel models. Similar results are obtained at low and high SNR (results are not reported here for the sake of conciseness).

5.4 Complexity comparison

In order to assess the complexity of the various channel estimation methods, we first report in Fig. 12 the effective number of iterations q_3 of AGDAR as a function of the maximum allowed number of iterations q_1, for ULAs with $N_r = 32$, $N_t = 4$, and average SNR of 10 dB on a UMi LOS channel model. As expected, when the number of allowed iterations increases, also the number of effective iterations increases, until reaching a floor. Moreover, a higher value of M_1 and M_3 requires a higher number of iterations. We also observe that RCS requires fewer iterations (while achieving a better performance in terms of MSE) as the reweighting fastens the convergence.

In Fig. 13, we compare the complexity of the various scenarios, by considering the number of complex multiplications, as derived in Section 4.4, as a function of the number of canceled paths I. Parameters are those of Fig. 3, so that we can compare the achieved MSE of the various schemes. We observe that the number of multiplications grows exponentially for the SPC and JPC techniques. We also report the number of complex multiplications for the AGDAR and RCS methods that do not depend on I. We

note that the AGDAR has a remarkably lower complexity than other methods. However, we also notice that for $M_1 = 64$ and $M_2 = 8$, RCS has a significantly higher complexity with respect to the other methods. When comparing the MSE performance (Fig. 3), we see that AGDAR achieves a much lower MSE than SPC and JPC methods for a lower complexity (in terms of CMUX and CADD).

6 Conclusions

In this paper, we have proposed channel estimation techniques for mm-Wave massive MIMO systems, based on a CS approach, where we have exploited the sparse nature of the channel, considering in particular the small number of channel paths at those frequencies. Efficient innovative solutions based on the adaptive restart of the Nesterov accelerated gradient algorithm have been explored. Numerical results have shown the superiority of the proposed approaches with respect to existing procedures, with similar or lower computational complexity. We have also considered the effects of sectorization and proposed a channel tracking technique that exploits slow channel variations.

Endnotes

[1] Note that other configurations (e.g., UPA on one side and ULA on the other side) can be obtained with similar derivations.
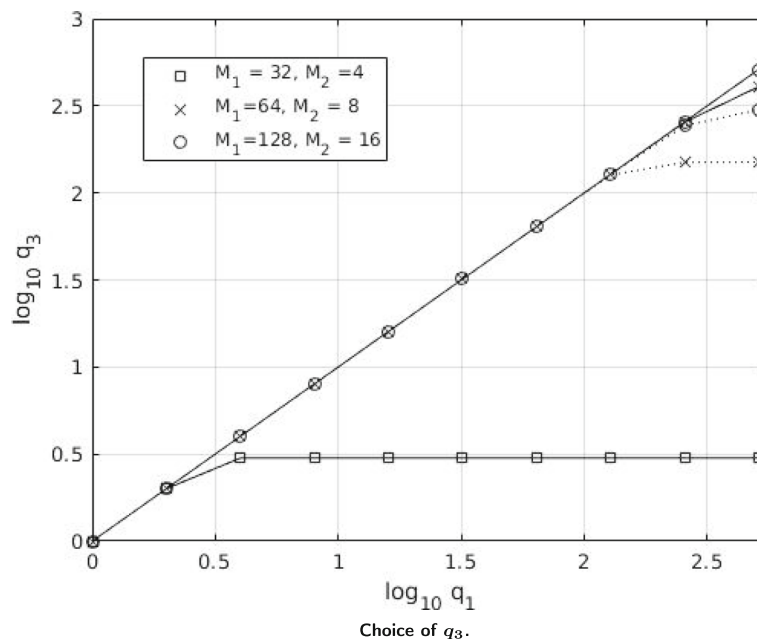

Fig. 12 Choice of q_3. Number of iterations q_3 (in log scale) vs the number iterations q_1 (in log scale) for a ULA system with $N_r = 32$ and $N_t = 4$ for AGDAR (solid lines) and RCS (dotted lines). UMi LOS model
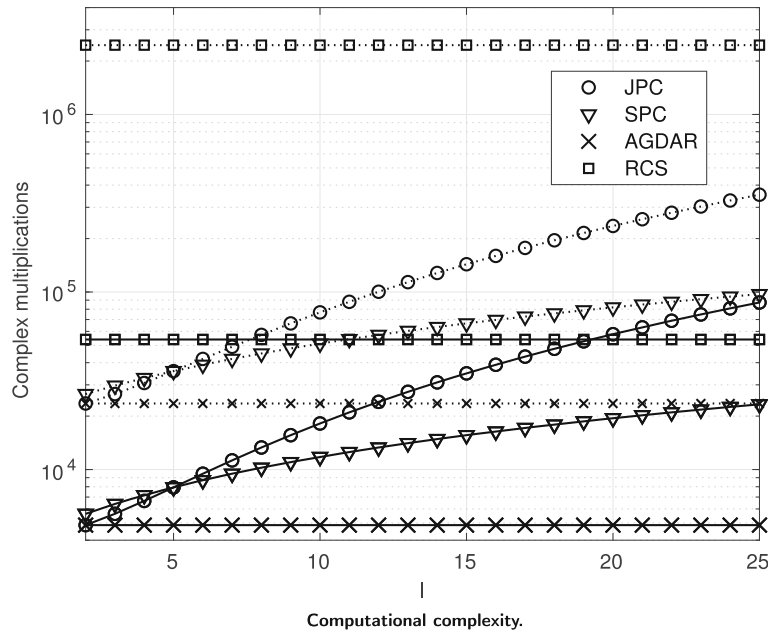
Fig. 13 Computational complexity. Average number of complex multiplications as function of *l* for various systems with ($M_1 = 32, M_2 = 4$) (solid lines) and ($M_1 = 64, M_2 = 8$) (dotted lines), ULA system with ($N_r = 32, N_t = 4$) and SNR of 10 dB

[2] In general, the number of symbols can be larges than N_t but we consider here this simpler case for the sake of conciseness.

Appendix A: CS channel estimation for UPA

For a UPA system, we denote as D_2 and D_3 the number of transmit antennas along the y- and z-axes, respectively; therefore, $N_t = D_2 \times D_3$. Similarly D_0 and D_1 are the numbers of antennas at the receiver along the two axes; therefore, $N_r = D_0 \times D_1$. The channel matrix has entries

$$\left[\boldsymbol{H}^{\text{UPA}}\right]_{i_1+D_1 i_0, i_3+D_3 i_2} = \sum_{l=1}^{L} \alpha_l e^{2\pi \eta_0^l i_0 j} e^{2\pi \eta_1^l i_1 j} e^{2\pi \eta_2^l i_2 j} e^{2\pi \eta_3^l i_3 j}. \tag{30}$$

The channel $\boldsymbol{H}^{\text{UPA}}$ is transformed into the channel column vector $\boldsymbol{h}^{\text{UPA}} = \text{vec}\left(\boldsymbol{H}^{\text{UPA}}\right) \in \mathbb{C}^{N_r N_t \times 1}$ with entries

$$\left[\boldsymbol{h}^{\text{UPA}}\right]_{i_0+D_0 i_1+D_0 D_1 i_2+D_2 D_1 D_0 i_3} = \left[\boldsymbol{H}^{\text{UPA}}\right]_{i_1+D_1 i_0, i_3+D_3 i_2}. \tag{31}$$

We also define the 4D-DFT matrix as $\boldsymbol{F}^{\text{4D}} \in \mathbb{C}^{N_r N_t \times (M_0 M_1 M_2 M_3)}$ with entries

$$\left[\boldsymbol{F}^{\text{4D}}\right]_{(i_0+D_0 i_1+D_1 D_0 i_2+D_2 D_1 D_0 i_3, f_0+M_0 f_1+M_1 M_0 f_2+M_2 M_1 M_0 f_3)}$$
$$= \prod_{d=0}^{3} e^{-\frac{2\pi i_d f_d j}{M_d}}. \tag{32}$$

Lastly, we define the column vector $\boldsymbol{v}^{\text{UPA}}$ of length $M_0 M_1 M_2 M_3$ with L non-zero entries, namely

$$\left[\boldsymbol{v}^{\text{UPA}}\right]_{\bar{\Omega}_0^l+M_0 \bar{\Omega}_1^l+M_1 M_0 \bar{\Omega}_2^l+M_2 M_1 M_0 \bar{\Omega}_3^l} = \alpha_l \tag{33}$$

where

$$\bar{\Omega}_0^l = \left\langle \eta_0^l M_0 \right\rangle, \ \bar{\Omega}_1^l = \left\langle \eta_1^l M_1 \right\rangle, \ \bar{\Omega}_2^l = \left\langle \eta_3^l M_3 \right\rangle, \ \bar{\Omega}_3^l = \left\langle \eta_3^l M_3 \right\rangle. \tag{34}$$

From (30), we can approximate the channel as

$$\boldsymbol{h}^{\text{UPA}} \approx \boldsymbol{F}^{\text{4D}} \boldsymbol{v}^{\text{UPA}}. \tag{35}$$

Similarly, we can define $\boldsymbol{h}'^{\text{UPA}}$ and $\boldsymbol{v}'^{\text{UPA}}$ for the LS estimate of the channel and its dual representation. The AGDAR and RCS algorithms for UPA can be obtained as described in Section 4, with $\boldsymbol{F}$, $\boldsymbol{h}'$, and $\boldsymbol{v}'$ replaced by $\boldsymbol{F}^{\text{4D}}$, $\boldsymbol{h}'^{\text{UPA}}$, and $\boldsymbol{v}'^{\text{UPA}}$, respectively.

Appendix B: On the choice of M_d

The choice of the number of DFT points per dimension M_d is important to determine the performance of the channel estimation algorithm. From (34), we have that the AGDAR solution approximates η_d^l with a quantized value taken over M_d possible points. Therefore, we can write the quantization error on the l-th path as

$$\epsilon_d^l = \eta_d^l - \frac{\left\langle \eta_d^l M_d \right\rangle}{M_d}. \tag{36}$$

Focusing now on a scenario wherein ULAs are used at both transmitter and receiver, assuming that all other estimates (i.e., the amplitude angle estimates of each path) are correct and the only imperfection is the quantization error, from (36), the estimated channel can be written as (compared with (1))

$$[\hat{\boldsymbol{H}}]_{i_1,i_2} = \sum_{l=1}^{L} \alpha_l e^{2\pi\left(\eta_1^l - \epsilon_1^l\right)i_1 j} e^{2\pi\left(\eta_2^l - \epsilon_2^l\right)i_2 j} \tag{37}$$

and using the definition of $\boldsymbol{H}$ of (1), we have that the MSE of the channel estimate is

$$\gamma^{(q)} = \frac{1}{N_r N_t} \sum_{i_1=0}^{N_r-1} \sum_{i_2=0}^{N_t-1} \mathbb{E}\left[|[\boldsymbol{H}]_{i_1,i_2} - \hat{\boldsymbol{H}}]_{i_1,i_2}|^2 \right]$$
$$= \frac{1}{N_r N_t} \sum_{i_1=0}^{N_r-1} \sum_{i_2=0}^{N_t-1} \mathbb{E}\left[\left| \sum_{l=1}^{L} \alpha_l e^{2\pi\eta_1^l i_1 j} e^{2\pi\eta_2^l i_2 j} \left(1 - e^{-2\pi\epsilon_1^l i_1 j} e^{-2\pi\epsilon_2^l i_2 j}\right) \right|^2 \right]. \tag{38}$$

Now, assuming that the amplitudes and angles are independent random variables, we have

$$\gamma^{(q)} = \frac{1}{N_r N_t} \sum_{i_1=0}^{N_r-1} \sum_{i_2=0}^{N_t-1} \sum_{l=1}^{L} \sigma_\alpha^2(l) \mathbb{E}\left[\left| 1 - e^{-2\pi\epsilon_1^l i_1 j} e^{-2\pi\epsilon_2^l i_2 j} \right|^2 \right], \tag{39}$$

where $\sigma_\alpha^2(l) = \mathbb{E}[|\alpha_l|^2]$. Expanding the expectation we have

$$\gamma^{(q)} = \frac{1}{N_r N_t} \sum_{i_1=0}^{N_r-1} \sum_{i_2=0}^{N_t-1} \sum_{l=1}^{L} \sigma_\alpha^2(l) \mathbb{E}\left\{ \left[1 - \cos\left(2\pi\left(\epsilon_1^l i_1 + \epsilon_2^l i_2\right)\right)\right]^2 \right.$$
$$\left. + \sin^2\left(2\pi\left(\epsilon_1^l i_1 + \epsilon_2^l i_2\right)\right)\right\}$$
$$= \frac{1}{N_r N_t} \sum_{i_1=0}^{N_r-1} \sum_{i_2=0}^{N_t-1} \sum_{l=1}^{L} \sigma_\alpha^2(l) \mathbb{E}\left\{ 2\left[1 - \cos\left(2\pi\left(\epsilon_1^l i_1 + \epsilon_2^l i_2\right)\right)\right]\right\}$$
$$= \frac{1}{N_r N_t} \sum_{i_1=0}^{N_r-1} \sum_{i_2=0}^{N_t-1} \sum_{l=1}^{L} \sigma_\alpha^2(l) 2\times$$
$$\mathbb{E}\left\{ 1 - \cos\left[2\pi\left(\eta_1^l - \frac{\langle\eta_1^l M_1\rangle}{M_1}\right)i_1 + \left(\eta_2^l - \frac{\langle\eta_2^l M_2\rangle}{M_2}\right)i_2\right]\right\}. \tag{40}$$

Let us assume that arrival and departure angles (ϑ_l^r and ϑ_l^t) are uniformly distributed in the interval $[0, 2\pi)$ and $\eta_1^l = \frac{\delta}{\lambda}\cos\vartheta_l^r, \eta_2^l = -\frac{\delta}{\lambda}\cos\vartheta_l^t$. Then, the probability density function of η_1^l and η_2^l for $\delta = \frac{\lambda}{2}$ becomes

$$p_{\eta_1^l}(a) = \begin{cases} \frac{\lambda}{\delta\pi\sqrt{1-\lambda^2 a^2/d^2}}, & a \in [-1/2, 1/2) \\ 0 & \text{otherwise,} \end{cases} \tag{41}$$

and (40) becomes

$$\gamma^{(q)} = \frac{1}{N_r N_t} \sum_{i_1=0}^{N_r-1} \sum_{i_2=0}^{N_t-1} \sum_{l=1}^{L} \sigma_\alpha^2(l) \sum_{m_1=0}^{M_1-1} \sum_{m_2=0}^{M_2-1}$$
$$\int_{m_1/M_1-1/2}^{(m_1+1)/M_1-1/2} \int_{m_2/M_2-1/2}^{(m_2+1)/M_2-1/2} p_{\eta_1^l}(a) p_{\eta_2^l}(b)\times$$
$$\left\{ 1 - \cos\left[2\pi\left(a - \frac{m_1}{M_1}\right)i_1 + \left(b - \frac{m_2}{M_2}\right)i_2\right]\right\} da\, db. \tag{42}$$

This MSE provides a guideline for the choice of M_d, as we must have at least $\gamma^{(q)} > \sigma^2$, so that quantization does not introduce more errors (in terms of its power) than noise already present.

Abbreviations

3GPP: Third-generation partnership project; AGDAR: Accelerated gradient descent with adaptive restart; AWGN: Additive white Gaussian noise; BPDN: Basis pursuit denoise; BS: Base station; CBP: Continuous basis pursuit; CS: Compressed sensing; DFT: Discrete Fourier transform; FISTA: Fast iterative shrinkage-thresholding algorithm; IDFT: Inverse discrete Fourier transform; InH: Indoor Hotspot; ISTA: Iterative shrinkage-thresholding algorithm; JPC: Joint peak cancelation; LASSO: Least absolute shrinkage and selection operator; LOS: Line-of-sight; LS: Least square; MIMO: Multiple-input-multiple-output; ML: Maximum likelihood; MSE: Mean square error; NLOS: Non-line-of-sight OMP: Orthogonal matching pursuit; RCS: Re-weighted compressed sensing; RF: Radio frequency; RMa: Rural macro cell; SNR: Signal to noise ratio; SPC: Single peak cancelation; SSAMP: Sparsity adaptive matching pursuit; UE: User equipment; ULA: Uniform linear array; UMa: Urban macro cell; UMi: Urban micro cell; UPA: Uniform planar array

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details
[1] Department of Information Engineering, University of Padova, via Gradenigo 6/A, Padua, Italy. [2] Huawei Technologies Italia, Milan, Italy. [3] Consorzio Nazionale Interuniversitario per le Telecomunicazioni, Padua, Italy.

References
1. T. S. Rappaport, J. N. Murdock, F. Gutierrez, State of the art in 60-ghz integrated circuits and systems for wireless communications. Proc. IEEE. **99**(8), 1390–1436 (2011)
2. S. Rangan, T. S. Rappaport, E. Erkip, Millimeter-wave cellular wireless networks: Potentials and challenges. Proc. IEEE. **102**(3), 366–385 (2014)

3.   Tecnical report, 5G; study on channel model for frequencies from 0.5 to 100 GHz (3GPP TR 38.901 version 14.3.0 release 14) (2018)

4.   R. W. Heath, N. González-Prelcic, S. Rangan, W. Roh, A. M. Sayeed, An overview of signal processing techniques for millimeter wave MIMO systems. IEEE J. Sel. Top. Sig. Proc. **10**(3), 436–453 (2016). https://doi.org/10.1109/JSTSP.2016.2523924

5.   A. Alkhateeb, O. E. Ayach, G. Leus, R. W. Heath, Channel estimation and hybrid precoding for millimeter wave cellular systems. IEEE J. Sel. Top. Sig. Proc. **8**(5), 831–846 (2014). https://doi.org/10.1109/JSTSP.2014.2334278

6.   S. Montagner, N. Benvenuto, S. Tomasin, in *Proc 2015 IEEE Int. Conf. on Communication Workshop (ICCW)*. Taming the complexity of mm-wave massive MIMO systems: Efficient channel estimation and beamforming, (2015), pp. 1251–1256. https://doi.org/10.1109/ICCW.2015.7247349

7.   D. De Donno, J. P. Beltrán, D. Giustiniano, J. Widmer, in *2016 IEEE International Conference on Communications Workshops (ICC), Kuala Lumpur*. Hybrid analog-digital beam training for mmWave systems with low-resolution RF phase shifters, (2016), pp. 700–705. https://doi.org/10.1109/ICCW.2016.7503869

8.   J. Lee, G. T. Gil, Y. H. Lee, Channel estimation via orthogonal matching pursuit for hybrid MIMO systems in millimeter wave communications. IEEE Trans. Commun. **64**(6), 2370–2386 (2016). https://doi.org/10.1109/TCOMM.2016.2557791

9.   J. Palacios, D. De Donno, D. Giustiniano, J. Widmer, in *2016 IEEE 27th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), Valencia*. Speeding up mmWave beam training through low-complexity hybrid transceivers, (2016), pp. 1–7. https://doi.org/10.1109/PIMRC.2016.7794709

10. S. Sun, T. S. Rappaport, in *2017 IEEE International Conference on Communications Workshops (ICC Workshops), Paris*. Millimeter Wave MIMO channel estimation based on adaptive compressed sensing, (2017), pp. 47–53. https://doi.org/10.1109/ICCW.2017.7962632

11. X. Li, J. Fang, H. Li, P. Wang, Millimeter wave channel estimation via exploiting joint sparse and low-rank structures. IEEE Trans. Wirel. Commun. **17**(2), 1123–1133 (2018). https://doi.org/10.1109/TWC.2017.2776108

12. G. Destino, M. Juntti, S. Nagaraj, in *Proc 2015 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*. Leveraging sparsity into massive MIMO channel estimation with the adaptive-LASSO, (2015), pp. 166–170. https://doi.org/10.1109/GlobalSIP.2015.7418178

13. Z. Marzi, D. Ramasamy, U. Madhow, Compressive channel estimation and tracking for large arrays in mm-wave picocells. IEEE J. Sel. Top. Sig. Proc. **10**(3), 514–527 (2016). https://doi.org/10.1109/JSTSP.2016.2520899

14. S. Malla, G. Abreu, in *Proc 2016 Int. Symposium on Wireless Comm. Systems (ISWCS)*. Channel estimation in millimeter wave MIMO systems: Sparsity enhancement via reweighting, (2016), pp. 230–234. https://doi.org/10.1109/ISWCS.2016.7600906

15. Z. Gao, L. Dai, Z. Wang, in *Proc 2016 IEEE Int. Conf. on Commun. (ICC)*. Channel estimation for mmwave massive MIMO based access and backhaul in ultra-dense network, (2016), pp. 1–6. https://doi.org/10.1109/ICC.2016.7511578

16. M. Kokshoorn, H. Chen, Y. Li, B. Vucetic, Beam-on-graph: Simultaneous channel estimation for mmwave MIMO systems with multiple users. IEEE Trans. Commun. **PP**(99), 1–1 (2018). https://doi.org/10.1109/TCOMM.2018.2791540

17. Y. Nesterov, Gradient methods for minimizing composite objective function. Math. Program. Ser. B. **140**, 125–161 (2007)

18. E. J. Candès, M. B. Wakin, S. P. Boyd, Enhancing sparsity by reweighted ℓ_1 minimization. J. Fourier Anal. Appl. **14**(5), 877–905 (2008). https://doi.org/10.1007/s00041-008-9045-x

19. A. M. Sayeed, Deconstructing multiantenna fading channels. IEEE Trans. Sig. Process. **50**(10), 2563–2579 (2002). https://doi.org/10.1109/TSP.2002.803324

20. J. Mo, P. Schniter, N. G. Prelcic, R. W. Heath, in *Proc 2014 48th Asilomar Conference on Signals, Systems and Computers*. Channel estimation in millimeter wave MIMO systems with one-bit quantization, (2014), pp. 957–961. https://doi.org/10.1109/ACSSC.2014.7094595

21. C. A. Balanis, *Antenna Theory: Analysis and Design*. (Wiley-Interscience, Hoboken, New Jersey, 2005)

22. Y. S. Cho, J. Kim, W. Y. Yang, C. G. Kang, *MIMO-OFDM Wireless Communications with MATLAB*. (Wiley, Hoboken, New Jersey, 2010)

23. E. Björnson, E. G. Larsson, T. L. Marzetta, Massive mimo: ten myths and one critical question. IEEE Commun. Mag. **54**(2), 114–123 (2016). https://doi.org/10.1109/MCOM.2016.7402270

24. D. Tse, P. Viswanath, *Fundamentals of Wireless Communication*. (Cambridge University Press, New York, 2005)

25. S. Boyd, L. Vandenberghe, *Convex Optimization*. (Cambridge university press, Cambridge, 2004)

26. M. A. T. Figueiredo, R. D. Nowak, S. J. Wright, Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems. IEEE J. Sel. Top. Sig. Proc. **1**(4), 586–597 (2007). https://doi.org/10.1109/JSTSP.2007.910281

27. B. O'Donoghue, E. Candès, Adaptive restart for accelerated gradient schemes. Found. Comput. Math. **15**(3), 715–732 (2015). https://doi.org/10.1007/s10208-013-9150-3

28. N. Parikh, S. Boyd, Proximal algorithms. Found. Trends Optim. **1**(3), 127–239 (2014). https://doi.org/10.1561/2400000003

29. A. Chambolle, R. A. D. Vore, N.-Y. Lee, B. J. Lucier, Nonlinear wavelet image processing: variational problems, compression, and noise removal through wavelet shrinkage. IEEE Trans. Image Process. **7**(3), 319–335 (1998). https://doi.org/10.1109/83.661182

30. Y. Nesterov, A method of solving a convex programming problem with convergence rate o(1/k 2). Sov. Math. Dokl. **27**, 372–376 (1983)