Frequency-Domain Compressive Channel Estimation for Frequency-Selective Hybrid Millimeter Wave MIMO Systems

Javier Rodríguez-Fernández, Nuria González-Prelcic, Kiran Venugopal, *Student Member, IEEE*, and Robert W. Heath, Jr., *Fellow, IEEE*

***Abstract*— Channel estimation is useful in millimeter wave (mm-wave) MIMO communication systems. Channel state information allows optimized designs of precoders and combiners under different metrics, such as mutual information or signal-to-interference noise ratio. At mm-wave, MIMO precoders and combiners are usually hybrid, since this architecture provides a means to trade-off power consumption and achievable rate. Channel estimation is challenging when using these architectures, however, since there is no direct access to the outputs of the different antenna elements in the array. The MIMO channel can only be observed through the analog combining network, which acts as a compression stage of the received signal. Most of the prior work on channel estimation for hybrid architectures assumes a frequency-flat mm-wave channel model. In this paper, we consider a frequency-selective mm-wave channel and propose compressed sensing-based strategies to estimate the channel in the frequency domain. We evaluate different algorithms and compute their complexity to expose tradeoffs in complexity overhead performance as compared with those of previous approaches.**

***Index Terms*— Wideband channel estimation, millimeter wave MIMO, hybrid architecture.**

I. Introduction

MIMO architectures with large arrays are a key ingredient of mm-wave communication systems, providing gigabit-per-second data rates [1]. Hybrid MIMO structures have been proposed to operate at mm-wave because the cost and power consumption of an all-digital achitecture is prohibitive at these frequencies [2]. Optimally configuring the digital and analog precoders and combiners requires channel knowledge when the design goal is maximizing metrics such as the achievable rate or the SINR. Acquiring the mm-wave channel is challenging with a hybrid architecture, however, because the channel is seen through the analog combining network, the SNR is low before beamforming, and the size of the channel matrices is large [3].

A. Prior Work

Two different approaches to increase SNR after spatial processing are beam training and channel estimation [2]. On the one hand, beam training, adopted in IEEE 802.11ad [4], is a technique that relies on searching for transmit and receive beam pairs that maximize the received SNR, thereby enabling reliable decoding of transmitted data [5]–[7]. Though they can increase link quality, beam training strategies are restricted to single-stream communication, which disables spatial multiplexing capabilities to obtain high data rate communications [8].

Channel estimation, on the other hand, allows transmission of several data streams, overcoming the limitation of beam training strategies. Most of these strategies exploit the spatially sparse structure in the mm-wave MIMO channel, formulating its estimation as a sparse recovery problem. The support of the estimated sparse vector identifies the pairs of direction-of-arrival/direction-of-departure (DoA/DoD) for each path in the mm-wave channel, while the amplitudes of the non-zero coefficients provide the channel gains for each path. Compressive estimation leads to a reduction in the channel training length when compared to conventional approaches such as those based on least squares (LS) estimation [9]. The main limitation of most of this prior work with hybrid MIMO architectures comes from considering a frequency-flat channel model [9]–[20], since the mm-wave channel is frequency-selective.

Recently, some approaches for channel estimation in frequency-selective mm-wave channels have been proposed. In [21], we designed a time-domain approach to estimate the wideband mm-wave channel assuming a hybrid MIMO architecture. This algorithm exploits the sparsity of the wideband millimeter wave channel in both the angular and delay domains. The sparse formulation of the problem in [21]

Manuscript received April 26, 2017; revised August 26, 2017 and December 1, 2017; accepted January 15, 2018. Date of publication March 2, 2018; date of current version May 8, 2018. This work was supported in part by the Agencia Estatal de Investigacin (Spain) and the European Regional Development Fund through the MYRADA Project under Grant TEC2016-75103-C2-2-R, in part by the U.S. Department of Transportation through the Data-Supported Transportation Operations and Planning Tier 1 University Transportation Center, in part by the Texas Department of Transportation through the Communications and Radar-Supported Transportation Operations and Planning Project under Grant 0-6877, and in part by the National Science Foundation under Grant NSF-CCF-1319556 and NSF-CCF-1527079. The associate editor coordinating the review of this paper and approving it for publication was W. H. Mow. *(Corresponding author: Robert W. Heath, Jr.)*

J. Rodríguez-Fernández and N. González-Prelcic are with the University of Vigo, 36310 Vigo, Spain (e-mail: jrodriguez@gts.uvigo.es; nuria@gts.uvigo.es).

K. Venugopal was with The University of Texas at Austin, Austin, TX 78701 USA. He is now with the Systems R&D, Qualcomm Flarion Technologies, Inc., Bridgewater, NJ 08807 USA (e-mail: kiranv@utexas.edu).

R. W. Heath, Jr., is with the University of Texas at Austin, Austin, TX 78701 USA (e-mail: rheath@utexas.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TWC.2018.2804943

includes the effect of non-integer sampling of the transmit pulse shaping filter, with the subsequent leakage effect and increase of sparsity level in the channel matrix. The main limitation of [21] is the high computational complexity of the algorithm. A frequency-domain strategy to estimate frequency-selective mm-wave channels was also proposed in [22]. A sparse reconstruction problem was formulated there to estimate the channel independently for every subcarrier, without exploiting spatial congruence between subbands. Another approach in the frequency domain was designed in [23], but only exploting the information from a reduced number of subcarriers. Based on the structured SAMP (SSAMP) algorithm, first proposed in [24], another approach to estimate the mm-wave channel was proposed in [25]. Exploiting the fact that spatial propagation characteristics do not change significantly within the system bandwidth, [25] assumed spatially common sparsity between the channels corresponding to the different subcarriers. The SSAMP algorithm from [24] was then considered to reconstruct the channels in the frequency domain. Thus, [25] is an interesting initial solution to the problem, but has several limitations when applied to a mm-wave communications system:

1) The effect of sampling the pulse shaping filter delayed by a non integer factor was not considered in the channel model for a given delay tap. As shown in [26], not accounting for this effect leads to virtual MIMO matrices with an artificially enhanced sparsity.
2) The algorithm was evaluated only for medium and high SNR regimes (larger than 10 dBs), not realistic at mm-wave, where the expected SNR is below 0 dB.
3) The reconstruction algorithm provides accurate results when Gaussian measurement matrices are employed; to generate the Gaussian matrices, unquantized phases were considered in the training precoders, which is unrealistic in a practical implementation of a mm-wave system based on a hybrid architecture.

Another algorithm exploiting common sparsity in the frequency domain at mm-wave was proposed in [27]. Unlike the SSAMP algorithm proposed in [24], the algorithm in [27] is proposed to estimate mm-wave wideband MU-MIMO channels. Besides the limitations 1)-3) described above, which also hold in this case, this algorithm exhibits another problem that makes it less feasible to be applied in a real mm-wave communication system. A Line-of-Sight (LoS) Rician channel model with $K_{\text{factor}} = 20$ dB was considered, which is only applicable when there is a strong LoS path. Owing to this artifact of the channel model, the algorithm in [27] estimates only a single path for each user, such that the task of channel estimation in a general mm-wave system can not be successfully accomplished.

B. Contributions

This paper proposes two novel frequency-domain approaches for the estimation of frequency-selective mm-wave MIMO channels. These approaches overcome the limitations of prior work and provide different trade-offs between complexity and achievable rate for a fixed training length. As in recent work on hybrid architectures for frequency-selective mm-wave channels [25], [28], we also consider a MIMO-OFDM communications system. Similar to [22], we use zero-padding (ZP) as a cyclic prefix to avoid loss and/or distortion of training data during reconfiguration of RF circuitry. A geometric channel is considered to model the different scattering clusters as in [21], [27], and [26], including the band-limiting property in the overall channel response.

Our two proposed approaches exploit the spatially common sparsity within the system bandwidth. The first algorithm aims at exploiting the information on the support coming from every subcarrier in the MIMO-OFDM system and provides the best performance. In contrast, the second algorithm uses less information to estimate the different frequency-domain subchannels, thereby managing to significantly reduce the computational complexity. We show that both strategies are asymptotically efficient when the channel DoAs/DoDs are correctly recovered, since they both attain the Cramér-Rao Lower Bound (CRLB). Further, we show that asymptotic efficiency can be achieved without using frequency-selective baseband precoders and combiners during the training stage, thereby reducing computational complexity.

Simulation results in the low SNR regime show that the two proposed algorithms significantly outperform the approach in the frequency domain developed in [22]. Comparisons with the algorithms proposed in [25] and [27] are also provided to show their performance in terms of estimation error in the SNR regime where mm-wave systems are expected to work. To the best of our knowledge, there are no prior work that evaluates channel estimation algorithms with real frequency-selective channel models. We evaluate the proposed algorithms with channel realizations obtained from the NYUSIM channel simulator [29], which was developped based on the statistical model in [30]. We show that the proposed algorithms are suitable for estimation of more general mm-wave frequency-selective channels than the model in [26], thereby demonstrating their practical application. The two proposed channel estimation algorithms provide a good trade-off between performance and overhead. Results show that using a reasonably small training length, approximately in the range of $60 - 100$ frames, leads to low estimation errors. The computational complexity of the proposed algorithms and previous strategies is also analyzed to compare the trade-offs between performance-complexity provided by the different algorithms. Finally, we also show that it is not necessary to exploit the information on the support coming from every OFDM subcarrier to estimate the different mm-wave subchannels. Yet, a reduced number of subcarriers is enough to asymptotically attain the CRLB.

The paper is organized as follows: Section II introduces the system model that is used throughout this work. Section III proposes two frequency-domain compressive channel estimation approaches, including the computation of the CRLB and suitable estimators derived from it to obtain the different frequency-domain channel matrices for each subcarrier. Thereafter, Section IV provides the main simulation results for the two proposed algorithms, the OMP-based compressive

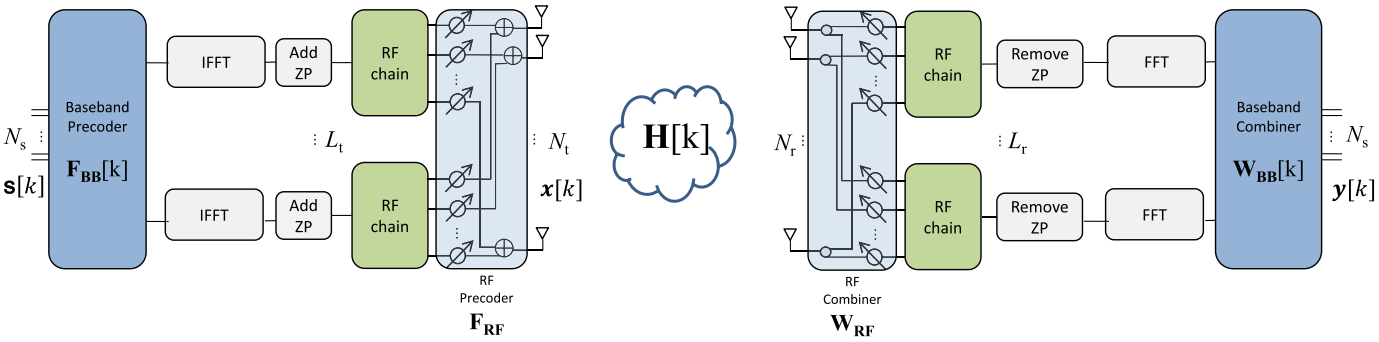


Fig. 1. Illustration of the structure of a hybrid MIMO architecture, which include analog and digital precoders and combiners.

approach proposed in [21] and the SSAMP and DGMP algorithms proposed in [25] and [27], respectively. Finally, Section V collects the main conclusions derived from the simulation results and also describes future work to be conducted.

Notation: We use the following notation throughout this paper: bold uppercase $\mathbf{A}$ is used to denote matrices, bold lowercase $\mathbf{a}$ denotes a column vector and non-bold lowercase a denotes a scalar value. We use $\mathcal{A}$ to denote a set. Further, $\|\mathbf{A}\|_F$ is the Frobenius norm and $\mathbf{A}^*$, $\overline{\mathbf{A}}$, $\mathbf{A}^T$ and $\mathbf{A}^\dagger$ denote the conjugate transpose, conjugate, transpose and Moore-Penrose pseudoinverse of a matrix $\mathbf{A}$, respectively. The (i,j)-th entry of a matrix $\mathbf{A}$ is denoted using $\{\mathbf{A}\}_{i,j}$. Similarly, the i-th entry of a column vector $\mathbf{a}$ is denoted as $\{\mathbf{a}\}_i$. The identity matrix of order N is denoted as $\mathbf{I}_N$. If $\mathbf{A}$ and $\mathbf{B}$ are two matrices, $\mathbf{A} \circ \mathbf{B}$ is the Khatri-Rao product of $\mathbf{A}$ and $\mathbf{B}$ and $\mathbf{A} \otimes \mathbf{B}$ is their Kronecker product. We use $\mathcal{N}(\mu, \mathbf{C})$ to denote a circularly symmetric complex Gaussian random vector with mean μ and covariance matrix $\mathbf{C}$. We use $\mathbb{E}$ to denote expectation. Discrete-time signals are represented as $\mathbf{x}[n]$, while frequency-domain signals are denoted using $\mathbf{x}[k]$.

II. System Model

We consider an OFDM based mm-wave MIMO link employing K subcarriers to send N_s data streams using a transmitter with N_t antennas and a receiver with N_r antennas. The system is based on a hybrid MIMO architecture as shown in Fig. 1, with L_t and L_r RF chains at the transmitter and receiver sides. For a general exposition, a frequency-selective hybrid precoder is used, with $\mathsf{F}[k] = \mathbf{F}_\mathrm{RF}\mathsf{F}_\mathrm{BB}[k] \in \mathbb{C}^{N_\mathrm{t} \times N_\mathrm{s}}$, $k = 0, \ldots, K-1$, where $\mathbf{F}_\mathrm{RF}$ is the analog precoder and $\mathsf{F}_\mathrm{BB}[k]$ the digital one. Note that the analog precoder is frequency-flat, while the digital precoder is different for every subcarrier. The RF precoder and combiner are implemented using a fully connected network of phase shifters, as described in [9]. The symbol blocks are transformed into the time domain using L_t parallel K-point IFFTs. As in [22] and [31], we consider ZP to both suppress Inter Block Interference (IBI) and account for the RF circuitry reconfiguration time. The discrete-time complex baseband signal at subcarrier k can be written as

$$\mathbf{x}[k] = \mathbf{F}_\mathrm{RF}\mathsf{F}_\mathrm{BB}[k]\mathbf{s}[k], \tag{1}$$

where the transmitted symbol sequence at subcarrier k of size $N_\mathrm{s} \times 1$ is denoted as $\mathbf{s}[k]$.

The MIMO channel between the transmitter and the receiver is assumed to be frequency-selective, having a delay tap length N_c in the time domain. The d-th delay tap of the channel is represented by a $N_\mathrm{r} \times N_\mathrm{t}$ matrix denoted as $\mathbf{H}_d$, $d = 0, 1, \ldots, N_\mathrm{c} - 1$, which, assuming a geometric channel model [26], can be written as

$$\mathbf{H}_d = \sqrt{\frac{N_\mathrm{t} N_\mathrm{r}}{L \rho_\mathrm{L}}} \sum_{\ell=1}^{L} \alpha_\ell p_\mathrm{rc}(dT_\mathrm{s} - \tau_\ell) \mathbf{a}_\mathrm{R}(\phi_\ell) \mathbf{a}_\mathrm{T}^*(\theta_\ell), \tag{2}$$

where ρ_L denotes the path loss between the transmitter and the receiver, L denotes the number of paths, T_s denotes the sampling period, $p_\mathrm{rc}(\tau)$ is a filter that includes the effects of pulse-shaping and other lowpass filtering evaluated at τ, $\alpha_\ell \in \mathbb{C}$ is the complex gain of the ℓth path, $\tau_\ell \in \mathbb{R}$ is the delay of the ℓth path, $\phi_\ell \in [0, 2\pi)$ and $\theta_\ell \in [0, 2\pi)$ are the angles-of-arrival and departure (AoA/AoD), of the ℓth path, and $\mathbf{a}_\mathrm{R}(\phi_\ell) \in \mathbb{C}^{N_\mathrm{r} \times 1}$ and $\mathbf{a}_\mathrm{T}(\theta_\ell) \in \mathbb{C}^{N_\mathrm{t} \times 1}$ are the array steering vectors for the receive and transmit antennas. Each one of these matrices can be written in a more compact way as

$$\mathbf{H}_d = \mathbf{A}_\mathrm{R} \mathbf{\Delta}_d \mathbf{A}_\mathrm{T}^*, \tag{3}$$

where $\mathbf{\Delta}_d \in \mathbb{C}^{L \times L}$ is diagonal with non-zero complex entries, and $\mathbf{A}_\mathrm{R} \in \mathbb{C}^{N_\mathrm{r} \times L}$ and $\mathbf{A}_\mathrm{T} \in \mathbb{C}^{N_\mathrm{t} \times L}$ contain the receive and transmit array steering vectors $\mathbf{a}_\mathrm{R}(\phi_\ell)$ and $\mathbf{a}_\mathrm{T}(\theta_\ell)$, respectively. The channel $\mathbf{H}_d$ can be approximated using the extended virtual channel model defined in [2] as

$$\mathbf{H}_d \approx \tilde{\mathbf{A}}_\mathrm{R} \mathbf{\Delta}_d^v \tilde{\mathbf{A}}_\mathrm{T}^*, \tag{4}$$

where $\mathbf{\Delta}_d^v \in \mathbb{C}^{G_\mathrm{r} \times G_\mathrm{t}}$ is a sparse matrix which contains the path gains of the quantized spatial frequencies in the non-zero elements. The dictionary matrices $\tilde{\mathbf{A}}_\mathrm{T}$ and $\tilde{\mathbf{A}}_\mathrm{R}$ contain the transmitter and receiver array response vectors evaluated on a grid of size G_r for the AoA and a grid of size G_t for the AoD. Due to the few scattering clusters in mm-wave channels, the sparse assumption for $\mathbf{\Delta}_d^v \in \mathbb{C}^{G_\mathrm{r} \times G_\mathrm{t}}$ is commonly accepted.

Finally, the channel at subcarrier k can be written in terms of the different delay taps as

$$\mathbf{H}[k] = \sum_{d=0}^{N_\mathrm{c}-1} \mathbf{H}_d e^{-\mathrm{j}\frac{2\pi k}{K} d} = \mathbf{A}_\mathrm{R} \mathbf{\Delta}[k] \mathbf{A}_\mathrm{T}^*. \tag{5}$$

It is also useful to write this matrix in terms of the sparse matrices $\mathbf{\Delta}_d^v$ and the dictionaries

$$\mathbf{H}[k] \approx \tilde{\mathbf{A}}_{\mathrm{R}} \left(\sum_{d=0}^{N_{\mathrm{c}}-1} \mathbf{\Delta}_d^v e^{-\mathrm{j}\frac{2\pi k}{K} d} \right) \tilde{\mathbf{A}}_{\mathrm{T}}^* \approx \tilde{\mathbf{A}}_{\mathrm{R}} \mathbf{\Delta}^{\mathrm{v}}[k] \tilde{\mathbf{A}}_{\mathrm{T}}^* \quad (6)$$

to help expose the sparse structure later.

Assuming that the receiver applies a hybrid combiner $\mathbf{W}[k] = \mathbf{W}_{\mathrm{RF}}\mathbf{W}_{\mathrm{BB}}[k] \in \mathbb{C}^{N_{\mathrm{r}} \times N_{\mathrm{s}}}$, the received signal at subcarrier k can be written as

$$\mathbf{y}[k] = \mathbf{W}_{\mathrm{BB}}^*[k]\mathbf{W}_{\mathrm{RF}}^*\mathbf{H}[k]\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}[k]\mathbf{s}[k] + \mathbf{W}_{\mathrm{BB}}^*[k]\mathbf{W}_{\mathrm{RF}}^*\mathbf{n}[k], \quad (7)$$

where $\mathbf{n}[k] \sim \mathcal{N}\left(0, \sigma^2\mathbf{I}\right)$ is the circularly symmetric complex Gaussian distributed additive noise vector. The receive signal model in (7) corresponds to the data transmission phase. As we will see in Section III, during the channel acquisition phase, we will consider frequency-flat training precoders and combiners to reduce complexity.

III. Compressive Channel Estimation in the Frequency Domain

In this section, we formulate a compressed sensing problem to estimate the vectorized sparse channel vector in the frequency domain. We also propose two algorithms to solve this problem that leverage the common support between the channel matrices for every subcarrier, providing different tradeoffs complexity-performance. The first algorithm leverages the common support between the K different subchannels providing very good performance, while the second one only exploits information from a reduced number of subcarriers, thereby keeping computational complexity at a lower level.

A. Problem Formulation

We assume that L_{t} and L_{r} RF chains are used at the transmitter and receiver. During the training phase, for the m-th frame transmitter and receiver use a training precoder $\mathbf{F}_{\mathrm{tr}}^{(m)} \in \mathbb{C}^{N_{\mathrm{t}} \times L_{\mathrm{t}}}$ and a training combiner $\mathbf{W}_{\mathrm{tr}}^{(m)} \in \mathbb{C}^{N_{\mathrm{r}} \times L_{\mathrm{r}}}$. This means that during the training phase, frequency-flat precoders and combiners are considered to keep the complexity of the sparse recovery algorithms low. We assume that the transmitted symbols satisfy $\mathbb{E}\{\mathbf{s}^{(m)}[k]\mathbf{s}^{(m)*}[k]\} = \frac{P}{N_{\mathrm{s}}}\mathbf{I}_{N_{\mathrm{s}}}$, with P the total transmitted power and $N_{\mathrm{s}} = L_{\mathrm{t}}$. To reduce computational complexity, we decompose the transmitted symbol $\mathbf{s}^{(m)}[k]$ as $\mathbf{s}^{(m)}[k] = \mathbf{q}^{(m)}\mathrm{t}^{(m)}[k]$, with $\mathbf{q}^{(m)} \in \mathbb{C}^{L_{\mathrm{t}} \times 1}$ a frequency-flat vector and $\mathrm{t}^{(m)}[k]$ a pilot symbol known at the receiver. This choice is motivated to simultaneously enable exploitation of the L_{t} spatial degrees of freedom coming from L_{t} RF chains and to allow channel estimation with a single subcarrier-independent measurement matrix. This enables both online and offline complexity reduction when the noise statistics are used to estimate the MIMO channel at the different subcarriers. Furthermore, each entry in $\mathbf{F}_{\mathrm{tr}}^{(m)}$, $\mathbf{W}_{\mathrm{tr}}^{(m)}$ is normalized to have squared-modulus N_{t}^{-1} and N_{r}^{-1}, respectively. Then, the received samples in the frequency domain for the m-th training frame can be written as

$$\mathbf{y}^{(m)}[k] = \mathbf{W}_{\mathrm{tr}}^{(m)^*}\mathbf{H}[k]\mathbf{F}_{\mathrm{tr}}^{(m)}\mathbf{q}^{(m)}\mathrm{t}^{(m)}[k] + \mathbf{n}_{\mathrm{c}}^{(m)}[k], \quad (8)$$

where $\mathbf{H}[k] \in \mathbb{C}^{N_r \times N_t}$ is the frequency-domain MIMO channel response at the k-th subcarrier and $\mathbf{n}_{\mathrm{c}}^{(m)}[k] \in \mathbb{C}^{L_r \times 1}$, $\mathbf{n}_{\mathrm{c}}^{(m)}[k] = \mathbf{W}_{\mathrm{tr}}^{(m)^*}\mathbf{n}^{(m)}[k]$, is the frequency-domain combined noise vector received at the k-th subcarrier. The average received SNR is given by $\mathrm{SNR} = \frac{P}{\rho_{\mathrm{L}}\sigma^2}$. We assume that the channel coherence time is larger than the frame duration and that the same channel can be considered for several consecutive frames. To enable sparse reconstruction with a single, subcarrier-independent measurement matrix, we will invert the effect of the scalar $\mathrm{t}^{(m)}[k]$ by means of multiplying the received signal by $(\mathrm{t}^{(m)}[k])^{-1}$. Using the result $\mathrm{vec}\{\mathbf{AXC}\} = (\mathbf{C}^T \otimes \mathbf{A})\,\mathrm{vec}\{\mathbf{X}\}$, the vectorized received signal after compensating for $\mathrm{s}^{(m)}[k]$ is

$$\mathrm{vec}\{\mathbf{y}^{(m)}[k]\} = (\mathbf{q}^{(m)T}\mathbf{F}_{\mathrm{tr}}^{(m)T} \otimes \mathbf{W}_{\mathrm{tr}}^{(m)^*})\,\mathrm{vec}\{\mathbf{H}[k]\} + \mathbf{n}_{\mathrm{c}}^{(m)}[k]. \quad (9)$$

Taking into account the expression in (6), the vectorized channel matrix can be written as $\mathrm{vec}\{\mathbf{H}[k]\} = (\tilde{\bar{\mathbf{A}}}_{\mathrm{T}} \otimes \tilde{\mathbf{A}}_{\mathrm{R}})\,\mathrm{vec}\{\mathbf{\Delta}[k]\}$. Therefore, if we define the measurement matrix $\mathbf{\Phi}^{(m)} \in \mathbb{C}^{L_{\mathrm{r}} \times N_{\mathrm{t}} N_{\mathrm{r}}}$ as

$$\mathbf{\Phi}^{(m)} = (\mathbf{q}^{(m)T}\mathbf{F}_{\mathrm{tr}}^{(m)T} \otimes \mathbf{W}_{\mathrm{tr}}^{(m)*}), \quad (10)$$

and the dictionary $\mathbf{\Psi} \in \mathbb{C}^{N_{\mathrm{t}}N_{\mathrm{r}} \times G_{\mathrm{t}}G_{\mathrm{r}}}$ as

$$\mathbf{\Psi} = (\tilde{\bar{\mathbf{A}}}_{\mathrm{T}} \otimes \tilde{\mathbf{A}}_{\mathrm{R}}), \quad (11)$$

(9) can be rewritten as

$$\mathrm{vec}\{\mathbf{y}^{(m)}[k]\} = \mathbf{\Phi}^{(m)}\mathbf{\Psi}\mathbf{h}^{\mathrm{v}}[k] + \mathbf{n}_{\mathrm{c}}^{(m)}[k], \quad (12)$$

where $\mathbf{h}^{\mathrm{v}}[k] = \mathrm{vec}\{\mathbf{\Delta}[k]\} \in \mathbb{C}^{G_{\mathrm{r}}G_{\mathrm{t}} \times 1}$ is the sparse vector containing the complex channel gains. To have enough measurements and accurately reconstruct the sparse vector $\mathbf{h}^{\mathrm{v}}[k]$, it is necessary to use several training frames, especially in the very-low SNR regime. If the transmitter and receiver communicate during M training steps using different pseudorandomly built precoders and combiners, (12) can be extended to

$$\underbrace{\begin{bmatrix} \mathbf{y}^{(1)}[k] \\ \vdots \\ \mathbf{y}^{(M)}[k] \end{bmatrix}}_{\mathbf{y}[k]} = \underbrace{\begin{bmatrix} \mathbf{\Phi}^{(1)} \\ \vdots \\ \mathbf{\Phi}^{(M)} \end{bmatrix}^T}_{\mathbf{\Phi}} \mathbf{\Psi}\mathbf{h}^{\mathrm{v}}[k] + \underbrace{\begin{bmatrix} \mathbf{n}_{\mathrm{c}}^{(1)}[k] \\ \vdots \\ \mathbf{n}_{\mathrm{c}}^{(M)}[k] \end{bmatrix}}_{\mathbf{n}_{\mathrm{c}}[k]}. \quad (13)$$

Finally, the vector $\mathbf{h}^{\mathrm{v}}[k]$ can be found by solving the sparse reconstruction problem

$$\min \|\mathbf{h}^{\mathrm{v}}[k]\|_1 \quad \text{subject to } \|\mathbf{y}[k] - \mathbf{\Phi}\mathbf{\Psi}\mathbf{h}^{\mathrm{v}}[k]\|_2^2 < \epsilon, \quad (14)$$

where ϵ is a tunable parameter defining the maximum error between the measurement and the received signal assuming the reconstructed channel between the transmitter and the receiver. Since the sparsity (number of channel paths) is usually unknown, the choice of ϵ is critical to solve (14). The choice of this parameter will be explained in Section III-C.

There are a great variety of algorithms to solve (14). For example, Orthogonal Matching Pursuit (OMP) was considered in [22]. However, this requires running the algorithm as many times as the number of OFDM subcarriers. In the next subsections we consider an additional assumption to solve this

problem, which avoids the need to run K OMP algorithms in parallel as proposed in [22].

The matrices $\mathbf{\Delta}[k]$ exhibit an interesting property that can be exploited when solving the compressed channel estimation problems defined in (14). Let us define the $G_{\mathrm{t}}G_{\mathrm{r}}\times 1$ vectorized virtual channel matrix for a given delay tap as

$$\mathbf{h}_d^{\mathrm{v}} \triangleq \mathrm{vec}\{\mathbf{\Delta}_d^v\}. \tag{15}$$

Let us denote the supports of the virtual channel matrices $\mathbf{\Delta}_d^v$ as $\mathcal{T}_0, \mathcal{T}_1, \ldots, \mathcal{T}_{N_{\mathrm{c}}-1}$, $d = 0, \ldots, N_{\mathrm{c}} - 1$. Then, since $\mathbf{h}^{\mathrm{v}}[k] = \mathrm{vec}\{\mathbf{\Delta}[k]\}$, with $\mathbf{\Delta}[k] = \sum_{d=0}^{N_{\mathrm{c}}-1} \mathbf{\Delta}_d^v e^{-j\frac{2\pi k}{N}d}$, $k = 0, \ldots, K-1$, it is clear that, in general

$$\mathrm{supp}\{\mathbf{h}^{\mathrm{v}}[k]\} = \bigcup_{d=0}^{N_{\mathrm{c}}-1} \mathrm{supp}\{\mathbf{h}_d^{\mathrm{v}}\} \quad k = 0, \ldots, K-1, \tag{16}$$

where the union of the supports of the time-domain virtual channel matrices comes from the additive nature of the Fourier transform. Therefore, the channel model in (6) exhibits the same sparse structure for all k, since the AoA and AoD do not change with frequency in the transmission bandwidth. Notice, however, that the AoA/AoD in (2) do not depend on the delay tap d, such that (16) reduces to $\mathrm{supp}\{\mathbf{h}^{\mathrm{v}}[k]\} = \mathrm{supp}\{\mathbf{h}_d^{\mathrm{v}}\}$, $k = 0, \ldots, K-1$. The sparse assumption on the vectorized channel matrix for a given delay tap $\mathbf{h}_d^{\mathrm{v}}$ is commonly accepted, since in mm-wave channels $L \ll G_{\mathrm{r}}G_{\mathrm{t}}$. In general, the vectorized channel matrix $\mathbf{h}^{\mathrm{v}}[k]$ will have, in the worst case, $N_{\mathrm{c}}L$ non-zero coefficients. Typical values for N_{c} in mm-wave channels are usually lower than 64 symbols (for example IEEE 802.11ad has been designed to work robustly for a maximum of 64 delay taps in the channel), while the number of measured paths usually satisfies $L < 30$ for outdoor and indoor scenarios [32]. From these values, using dictionaries of size $G_{\mathrm{r}} \geq 64$ $G_{\mathrm{t}} \geq 64$, allows us to assume a spatially sparse structure for $\mathbf{h}^{\mathrm{v}}[k]$ as well.

B. Simultaneous Weighted Estimation Exploiting Common Support Between Subcarriers (SW-OMP)

To develop a channel estimation algorithm that leverages the sparse nature and the common support property for all $\mathbf{h}^{\mathrm{v}}[k]$, we propose to modify the S-OMP algorithm proposed in [33]. For a given iteration, this algorithm aims at finding a new index of the support exploiting information coming from several signals, thus increasing the reliability of the final support estimate. If the algorithm processes K received signals simultaneously, the amount of information it obtain is K times larger. Consequently, the likelihood of the estimated AoA/AoD, or equivalently, the support of the channel, increases accordingly. The S-OMP algorithm in [33] computes the non-zero values of the sparse vector using a LS approach once the support is obtained, assuming that the noise covariance matrix is the identity matrix $\mathbf{I}_{ML_r}$. In this paper, however, we generalize the S-OMP algorithm to account for the correlated noise after combining, and show that the proposed algorithm attains the CRLB.

1) Support Computation With Correlated Noise: Before explicit estimation of the channel gains, it is necessary to compute the atom, i.e., vector in the measurement matrix, which yields the largest sum-correlation with the received signals, since the support of the different sparse vectors is the same. The S-OMP algorithm is based on the assumption that the perturbation (noise) covariance matrix is diagonal, such that no correlation between the different noise components is present. The correlation vector $\mathbf{c}[k] \in \mathbb{C}^{ML_{\mathrm{r}}}$ is defined as

$$\mathbf{c}[k] = \mathbf{\Upsilon}^* \mathbf{y}[k], \tag{17}$$

in which $\mathbf{\Upsilon} \in \mathbb{C}^{ML_{\mathrm{r}}\times G_{\mathrm{t}}G_{\mathrm{r}}}$, $\mathbf{\Upsilon} = \mathbf{\Phi\Psi}$ is the equivalent measurement matrix and $\mathbf{y}[k] \in \mathbb{C}^{ML_{\mathrm{r}}\times 1}$ is the received signal for a given k, $k = 0, \ldots, K-1$. If there is correlation between noise components, the atom estimated from the projection in (17) is not likely to be the actual one. To introduce the appropriate correction in the correlation, the specific form the noise covariance matrix takes needs to be used. Let us consider two arbitrary (hybrid) combiners $\mathbf{W}_{\mathrm{tr}}^{(m)(i)}$, $\mathbf{W}_{\mathrm{tr}}^{(m)(j)} \in \mathbb{C}^{N_{\mathrm{r}}\times L_{\mathrm{r}}}$ for two arbitrary training steps i,j and a given subcarrier k. Therefore, if the combined noise at a given training step i and subcarrier k is denoted as $\mathbf{n}_{\mathrm{c}}^{(i)}[k] = \mathbf{W}_{\mathrm{tr}}^{(i)*}\mathbf{n}^{(i)}[k]$, with $\mathbf{n}^{(i)}[k] \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I}_{L_{\mathrm{r}}})$, the noise cross-covariance matrix is given by $\mathbb{E}\{\mathbf{n}_{\mathrm{c}}^{(i)}[k]\mathbf{n}_{\mathrm{c}}^{(j)*}[k]\} = \mathbf{W}_{\mathrm{tr}}^{(i)*}\sigma^2\delta[i-j]\mathbf{W}_{\mathrm{tr}}^{(j)}$. The noise covariance matrix of $\mathbf{y}[k]$ can be written as a block diagonal matrix $\mathbf{C} \in \mathbb{C}^{ML_{\mathrm{r}}\times ML_{\mathrm{r}}}$, $\mathbf{C}_{\mathrm{w}} = \mathrm{blkdiag}\{\mathbf{W}_{\mathrm{tr}}{}^{(1)*}\mathbf{W}_{\mathrm{tr}}{}^{(1)}, \mathbf{W}_{\mathrm{tr}}{}^{(2)*}\mathbf{W}_{\mathrm{tr}}{}^{(2)}, \ldots, \mathbf{W}_{\mathrm{tr}}{}^{(M)*}\mathbf{W}_{\mathrm{tr}}{}^{(M)}\}$. Now, we use the Cholesky factorization to write $\mathbf{C}_{\mathrm{w}} = \mathbf{D}_{\mathrm{w}}^*\mathbf{D}_{\mathrm{w}}$, where $\mathbf{D}_{\mathrm{w}} \in \mathbb{C}^{ML_{\mathrm{r}}\times ML_{\mathrm{r}}}$ is an upper triangular matrix. The subscript in $\mathbf{C}_{\mathrm{w}}$ and $\mathbf{D}_{\mathrm{w}}$ indicates that these matrices only depend on the combiners $\mathbf{W}_{\mathrm{tr}}^{(m)}$. Then, the correlation step is performed as

$$\mathbf{c}[k] = \mathbf{\Upsilon}_{\mathrm{w}}^*\mathbf{y}_{\mathrm{w}}[k], \tag{18}$$

where $\mathbf{\Upsilon}_{\mathrm{w}} \in \mathbb{C}^{ML_{\mathrm{r}}\times G_{\mathrm{t}}G_{\mathrm{r}}}$ is the whitened measurement matrix given by $\mathbf{\Upsilon}_{\mathrm{w}} = \mathbf{D}_{\mathrm{w}}^{-*}\mathbf{\Upsilon}$. Likewise, the $ML_{\mathrm{r}} \times 1$ whitened received signal $\mathbf{y}_{\mathrm{w}}[k]$ is given by $\mathbf{y}_{\mathrm{w}}[k] = \mathbf{D}_{\mathrm{w}}^{-*}\mathbf{y}[k]$. The matrix $\mathbf{D}_{\mathrm{w}}^{-1} \in \mathbb{C}^{ML_{\mathrm{r}}\times ML_{\mathrm{r}}}$ is given by $\mathbf{D}_{\mathrm{w}}^{-1} = \mathrm{blkdiag}\left\{\left(\mathbf{D}_{\mathrm{w}}^{(1)}\right)^{-1}, \ldots, \left(\mathbf{D}_{\mathrm{w}}^{(M)}\right)^{-1}\right\}$, where $\left(\mathbf{D}_{\mathrm{w}}^{(m)}\right)^{-1}$ can be seen as a frequency-flat baseband combiner $\mathbf{W}_{\mathrm{BB},tr}^{(m)}$ used in the m-th training step. In this way, the resulting correlation simultaneously whitens the spatial noise components and estimates the most likely support index in the sparse vectors $\mathbf{h}^{\mathrm{v}}[k]$.

2) Computation of the Channel Gains: If the vectorized channel matrix for a given subcarrier $k = 0, 1, \ldots, K-1$, is given by $\mathrm{vec}\{\mathbf{H}[k]\} = (\bar{\mathbf{A}}_{\mathrm{T}} \circ \mathbf{A}_{\mathrm{R}})\boldsymbol{\xi}[k]$, with $\boldsymbol{\xi}[k]$ the $L \times 1$ vector of channel gains, then the received signal $\mathbf{y}[k]$ is distributed according to $\mathcal{N}(\mathbf{\Phi\Psi}\boldsymbol{\xi}[k], \mathbf{C})$. Once an estimate $\hat{\mathcal{T}}$ of the support of the sparse channel vectors is found, with $\hat{L} = |\hat{\mathcal{T}}|$ the estimated sparsity level, we can define the matrix $[\mathbf{\Upsilon}]_{:,\hat{\mathcal{T}}} \in \mathbb{C}^{ML_{\mathrm{r}}\times \hat{L}}$ as $[\mathbf{\Upsilon}]_{:,\hat{\mathcal{T}}} = [\mathbf{\Phi\Psi}]_{:,\hat{\mathcal{T}}}$. Accordingly, the signal model for the k-th subcarrier can be written as

$$\mathbf{y}[k] = [\mathbf{\Upsilon}]_{:,\hat{\mathcal{T}}}\,\tilde{\boldsymbol{\xi}}[k] + \tilde{\mathbf{n}}_{\mathrm{c}}[k], \tag{19}$$

where $\tilde{\mathbf{n}}_{\mathrm{c}}[k] \in \mathbb{C}^{ML_{\mathrm{r}}\times 1}$ is the residual noise in our linear model after estimating the channel support. If the estimation of the support is accurate, $\tilde{\mathbf{n}}_{\mathrm{c}}[k]$ will be close to the post-combining noise vector $\mathbf{n}_{\mathrm{c}}[k]$. The $\hat{L}\times 1$ vector $\tilde{\boldsymbol{\xi}}[k]$ is the vector of channel gains to be estimated after sparse recovery. It is important to remark that the support estimated by the proposed algorithm may be different from the actual channel support. In general, $\hat{\mathcal{T}}$ can be different from the actual support. Therefore, the vector $\tilde{\boldsymbol{\xi}}[k] \in \mathbb{C}^{\hat{L}\times 1}$ can be also different from $\boldsymbol{\xi}[k] \in \mathbb{C}^{L\times 1}$, $\boldsymbol{\xi}[k] = \mathrm{vec}\{\mathrm{diag}\{\boldsymbol{\Delta}[k]\}\}$. Since the model in (19) is linear on the parameter vector $\boldsymbol{\xi}[k]$, there is a Minimum Variance Unbiased (MVU) Estimator that happens to be the Best Linear Unbiased Estimator (BLUE) as well [34].

The mathematical equation in (19) is usually referred to as the *General Linear Model* (GLM), for which the solution of $\tilde{\boldsymbol{\xi}}[k]$ for real parameters is provided in [34]. The extension for a complex vector of parameters is straightforward and given by

$$\hat{\tilde{\boldsymbol{\xi}}}[k] = \left([\boldsymbol{\Upsilon}]^{*}_{:,\hat{\mathcal{T}}}\,\mathbf{C}_{\mathrm{w}}^{-1}\,[\boldsymbol{\Upsilon}]_{:,\hat{\mathcal{T}}}\right)^{-1}[\boldsymbol{\Upsilon}]^{*}_{:,\hat{\mathcal{T}}}\,\mathbf{C}_{\mathrm{w}}^{-1}\mathbf{y}[k], \tag{20}$$

which can be further reduced to

$$\hat{\tilde{\boldsymbol{\xi}}}[k] = \left([\boldsymbol{\Upsilon}_{\mathrm{w}}]_{:,\hat{\mathcal{T}}}\right)^{\dagger}\mathbf{y}_{\mathrm{w}}[k]. \tag{21}$$

Therefore, $\hat{\tilde{\boldsymbol{\xi}}}[k]$ is the MVU estimator for our parameter vector $\tilde{\boldsymbol{\xi}}[k]$, $k = 0,\ldots,K-1$. Hence, it is unbiased and attains the CRLB if the support is estimated correctly. It is interesting to note that this corresponds to a Weighted Least-Squares (WLS) estimator, with the corresponding weights given by the inverse noise covariance matrix. An important feature of this estimator is that the difference in performance given by the LS and the WLS estimators is more accentuated as the number of RF chains grows (if and only if the hybrid combiner is not built from orthonormal vectors).

We are interested in finding the CRLB for the estimation of the channel matrices at each subcarrier assuming perfect sparse reconstruction. To that end, taking into account only the non-zero entries in $\mathbf{h}^{\mathrm{v}}[k]$, the Fisher Information Matrix (FIM) is derived from the GLM in (19) as

$$\mathbf{I}(\boldsymbol{\xi}[k]) = [\boldsymbol{\Upsilon}]^{*}_{:,\mathcal{T}}\,\mathbf{C}^{-1}\,[\boldsymbol{\Upsilon}]_{:,\mathcal{T}}\,. \tag{22}$$

Note that (22) gives the FIM for the vector $\boldsymbol{\xi}[k]$, which contains the actual channel gains. To compute the CRLB for the estimation of the frequency-domain channel matrix $\mathbf{H}[k]$, we can vectorize (5) as

$$\mathrm{vec}\{\mathbf{H}[k]\} = (\overline{\mathbf{A}_{\mathrm{T}}} \circ \mathbf{A}_{\mathrm{R}})\boldsymbol{\xi}[k]. \tag{23}$$

The decomposition in (23) is expressed with equality, since our focus is on finding the CRLB when the estimation of the support is perfect. Owing to lack of space, the solution to the problem of finding the CRLB in a more general case accounting for continuous AoA and AoD is left for future work.

The overall minimum variance for an unbiased estimator of the $N_{\mathrm{t}}N_{\mathrm{r}}$ entries in $\mathbf{H}[k]$ is given by the sum of the variances for each element $\{\mathbf{H}[k]\}_{\{i,j\}}$, $i = 1,\ldots,N_{\mathrm{r}}$, $j = 1,\ldots,N_{\mathrm{t}}$. We will denote the overall variance of the estimator for $\mathbf{H}[k]$ as $\gamma(\mathrm{vec}\{\mathbf{H}[k]\})$, for a given subcarrier k. Then, $\gamma(\mathrm{vec}\{\mathbf{H}[k]\})$ is derived as

$$\begin{aligned}\gamma(\mathrm{vec}\{\mathbf{H}[k]\}) &= \mathrm{trace}\left\{\frac{\partial\,\mathrm{vec}\{\mathbf{H}[k]\}}{\partial\boldsymbol{\xi}[k]}\mathbf{I}^{-1}(\boldsymbol{\xi}[k])\frac{\partial\,\mathrm{vec}\{\mathbf{H}^{*}[k]\}}{\partial\boldsymbol{\xi}[k]}\right\}\\ &= \mathrm{trace}\left\{(\overline{\mathbf{A}_{\mathrm{T}}}\circ\mathbf{A}_{\mathrm{R}})\mathbf{I}^{-1}(\boldsymbol{\xi}[k])(\overline{\mathbf{A}_{\mathrm{T}}}\circ\mathbf{A}_{\mathrm{R}})^{*}\right\}.\end{aligned} \tag{24}$$

C. Computation of the Residual and Noise Variance Estimation

After estimating the channel gains, it is necessary to determine whether a sufficient number of paths has already been estimated or not. To solve this detection problem, we need some prior information to compare the received signals $\mathbf{y}[k]$ to the reconstructed signals $\hat{\mathbf{x}}_{\mathrm{rec}}[k] = [\boldsymbol{\Upsilon}]_{:,\hat{\mathcal{T}}}\,\hat{\tilde{\boldsymbol{\xi}}}[k]$. For this reason, we assume that the noise variance is known at the receiver. This is a practical assumption since the receiver can accurately estimate the noise variance before the training stage takes place. Therefore, the received signal $\mathbf{y}[k]$ can be approximately modeled as $\mathbf{y}[k] \approx \hat{\mathbf{x}}_{\mathrm{rec}}[k] + \tilde{\mathbf{n}}_{\mathrm{c}}[k]$, which is the same model as in (7), since $\hat{\mathbf{x}}_{\mathrm{rec}}[k]$ is an estimate of the mean of $\mathbf{y}[k]$. Let us define the complete received signal as $\mathbf{y} \triangleq \mathrm{vec}\{\mathbf{y}[0],\ldots,\mathbf{y}[K-1]\}$, and the complete reconstructed signal as $\hat{\mathbf{x}}_{\mathrm{rec}} \triangleq \mathrm{vec}\{\hat{\mathbf{x}}_{\mathrm{rec}}[0],\ldots,\hat{\mathbf{x}}_{\mathrm{rec}}[K-1]\}$. Then, the estimation of the noise variance can be formulated as a Maximum-Likelihood estimation problem [34],

$$\hat{\sigma^2}_{\mathrm{ML}} = \arg\max_{\sigma^2} L(\mathbf{y},\hat{\mathbf{x}}_{\mathrm{rec}},\sigma^2), \tag{25}$$

where $L(\mathbf{y},\hat{\mathbf{x}}_{\mathrm{rec}},\sigma^2)$ denotes the Log-Likelihood Function (LLF) of $\mathbf{y}$. This function is given by [34]

$$\begin{aligned}&L(\mathbf{y},\hat{\mathbf{x}}_{\mathrm{rec}},\sigma^2)\\ &\quad= -ML_{\mathrm{r}}\ln\pi\sigma^2 - \ln\det\{\mathbf{C}_{\mathrm{w}}\}\\ &\qquad-\frac{1}{\sigma^2}\sum_{k=0}^{K-1}\big(\mathbf{y}[k]-\hat{\mathbf{x}}_{\mathrm{rec}}[k]\big)^{*}\mathbf{C}_{\mathrm{w}}^{-1}\big(\mathbf{y}[k]-\hat{\mathbf{x}}_{\mathrm{rec}}[k]\big).\end{aligned} \tag{26}$$

The ML estimator of the noise variance is then obtained by taking $\partial L(\mathbf{y},\hat{\mathbf{x}}_{\mathrm{rec}},\sigma^2)/\partial\sigma^2 = 0$, such that $\widehat{\sigma^2}_{\mathrm{ML}}$ is given by

$$\widehat{\sigma^2}_{\mathrm{ML}} = \frac{1}{KML_{\mathrm{r}}}\sum_{k=0}^{K-1}\underbrace{\big(\mathbf{y}[k]-\hat{\mathbf{x}}_{\mathrm{rec}}[k]\big)^{*}\,\mathbf{C}_{\mathrm{w}}^{-1}\,\big(\mathbf{y}[k]-\hat{\mathbf{x}}_{\mathrm{rec}}[k]\big)}_{\mathbf{r}^{*}[k]\mathbf{r}[k]} \tag{27}$$

where the $ML_{\mathrm{r}}\times 1$ vector $\mathbf{r}[k] \triangleq \mathbf{y}_{\mathrm{w}}[k] - \mathbf{D}_{\mathrm{w}}^{-*}\hat{\mathbf{x}}_{\mathrm{rec}}$ is the residual. Observe that $\mathbf{r}[k]$ can also be written as $\mathbf{r}[k] = (\mathbf{I}_{ML_{\mathrm{r}}} - \mathbf{P})\,\mathbf{y}_{\mathrm{w}}[k]$, where $\mathbf{P} \in \mathbb{C}^{ML_{\mathrm{r}}\times ML_{\mathrm{r}}}$ is the projection matrix given by $\mathbf{P} = [\boldsymbol{\Upsilon}_{\mathrm{w}}]^{\dagger}_{:,\hat{\mathcal{T}}}\,[\boldsymbol{\Upsilon}_{\mathrm{w}}]_{:,\hat{\mathcal{T}}}$. Thereby, after a sufficient number of iterations, $\hat{L}$ different paths are expected to be estimated. Consequently, the estimated noise variance will be comparable to the true noise variance of the received signal, such that the detection process is accomplished by setting ϵ in (14) to σ^2.

Of particular importance: the larger the number of subcarriers, the smaller the estimation variance the ML estimator can

Algorithm 1 Simultaneous Weighted Orthogonal Matching Pursuit (SW-OMP)

procedure SW-OMP($\mathbf{y}[k]$,$\mathbf{\Phi}$,$\mathbf{\Psi}$,ϵ)
  Compute the whitened equivalent observation matrix
    $\mathbf{\Upsilon}_{\mathrm{w}} = \mathbf{D}_{\mathrm{w}}^{-*}\mathbf{\Phi}\mathbf{\Psi}$
  Initialize the residual vectors to the input signal vectors and support estimate
    $\mathbf{y}_{\mathrm{w}}[k] = \mathbf{D}_{\mathrm{w}}^{-*}\mathbf{y}[k]$, $\mathbf{r}[k] = \mathbf{y}_{\mathrm{w}}[k]$, $k = 0, \ldots, K-1$, $\hat{\mathcal{T}} = \{\emptyset\}$
  while MSE $> \epsilon$ **do**
    Distributed Correlation
      $\mathbf{c}[k] = \mathbf{\Upsilon}_{\mathrm{w}}^{*}\mathbf{r}[k], \quad k = 0, \ldots, K-1$
    Find the maximum projection along the different spaces
      $p^{\star} = \arg\max_{p} \sum_{k=0}^{K-1} |\{\mathbf{c}[k]\}_p|$
    Update the current guess of the common support
      $\hat{\mathcal{T}} = \hat{\mathcal{T}} \cup p^{\star}$
    Project the input signal onto the subspace given by the support using WLS
      $\mathbf{x}_{\hat{\mathcal{T}}}[k] = \left([\mathbf{\Upsilon}_{\mathrm{w}}]_{:,\hat{\mathcal{T}}}\right)^{\dagger}\mathbf{y}_{\mathrm{w}}[k]$
      $k = 0, \ldots, K-1$
    Update residual
      $\mathbf{r}[k] = \mathbf{y}_{\mathrm{w}}[k] - [\mathbf{\Upsilon}_{\mathrm{w}}]_{:,\hat{\mathcal{T}}}\hat{\hat{\boldsymbol{\xi}}}[k]$
      where $\hat{\hat{\boldsymbol{\xi}}}[k] = \mathbf{x}_{\hat{\mathcal{T}}}[k], \quad k = 0, \ldots, K-1$
    Compute the current MSE
      $\mathrm{MSE} = \frac{1}{KML_{\mathrm{r}}}\sum_{k=0}^{K-1}\mathbf{r}^{*}[k]\mathbf{r}[k]$
  end while
end procedure

achieve. Thereby, if the number of averaging subcarriers K is large enough, the lack of knowledge of the sparsity level is not so critical because of two reasons: 1) the computation of the support is more precise due to noise averaging during the projection process, and 2) if the support is estimated correctly, a particular estimate of σ^2 will be very close to the true noise variance, such that the chosen halting criterion is optimal from the Maximum Likelihood perspective. It should be clear that the higher the covariance between adjacent noise components, the larger the performance gap between the S-OMP and the SW-OMP algorithm, which depends on the ratio between N_{r} and L_{r}. The modification of the S-OMP algorithm to include the MVU estimator for the channel gains, as well as the whitening matrix to estimate the support and the residual is provided in Algorithm 1.

D. Subcarrier Selection Simultaneous Weighted-OMP + Thresholding

Despite the use of a single, subcarrier-independent measurement matrix $\mathbf{\Upsilon}$ to estimate the frequency-domain channels, the algorithm presented in the previous section exhibits high computational complexity. The SW-OMP algorithm considers the distributed projection coming from every subcarrier; however, a trade-off between performance and computational complexity can be achieved if a small number of subcarriers $K_{\mathrm{p}} \ll K$ is used, instead. The problem amounts as to how to choose those subcarriers, since no quality measure is available beforehand. The ideal situation would require knowledge of the Signal-to-Noise Ratio (SNR), which is unknown so far. Nonetheless, we have access to different frequency-domain received vectors $\mathbf{y}[k]$, $k = 0, 1, \ldots, K-1$. Therefore, we propose to use the l_2-norm of the different vectors as quality metric. Owing to the triangle inequality, $||\mathbf{y}[k]||_2^2 \leq ||\mathbf{\Phi}\mathbf{h}^{\mathrm{v}}[k]||_2^2 + ||\mathbf{n}_{\mathrm{c}}[k]||_2^2$, such that the K_{p} selected signals are expected to exhibit the strongest channel response. Thereby, the K_{p} subcarriers having largest l_2-norm will be used to derive an estimate of the support of the already defined sparse channel vectors $\mathbf{h}^{\mathrm{v}}[k]$, $k = 0, \ldots, K-1$.

The main problem concerning Matching Pursuit (MP) algorithms comes from the lack of knowledge of the channel sparsity L. For this reason, there is usually an iteration at which L paths have been detected but the estimate of the average residual energy is a little larger than the noise variance itself. This makes the algorithm find additional paths which are not actually contained in the MIMO channel. These paths usually have low power, and a pruning procedure is needed to filter out these undesired components. An approach is to remove those components whose power falls below a given threshold, which can be related to the average power of the component in the estimated sparse vectors having maximum average power. Let us denote this power by $P^{\star}$. Then, the threshold can be defined as $\eta = \beta P^{\star}$, $\beta \in (0, 1)$. The value $P^{\star}$ is taken as $P^{\star} = \max_{\ell} \frac{1}{K}\sum_{k=0}^{K-1} |\{\hat{\boldsymbol{\xi}}[k]\}_{\ell}|^2$. To keep the common sparsity property we must ensure that the channel support after thresholding remains invariant. For this purpose, we define a signal $\hat{\mathbf{p}}_{\mathrm{av}} \in \mathbb{C}^{\hat{L}\times 1}$ whose i-th component is given by $\hat{p}_{\mathrm{av},i} = \frac{1}{K}\sum_{k=0}^{K-1} |\{\hat{\boldsymbol{\xi}}[k]\}_i|^2$, $i = 1, \ldots, \hat{L}$, such that $\hat{\mathbf{p}}_{\mathrm{av}}$ measures the average power along the different subbands of each spatial component in the quantized angle grid. The final support after thresholding $\hat{\mathcal{T}}_{\mathrm{Th}}$ is defined as $\hat{\mathcal{T}}_{\mathrm{Th}} = \bigcup_{i=1}^{\hat{L}} i \,/\, \hat{p}_{\mathrm{av},i} \geq \beta P^{\star}$. Therefore, the components in $\hat{\boldsymbol{\xi}}[k]$ indexed by $\hat{\mathcal{T}}_{\mathrm{Th}}$ are the final channel gains estimates for each subcarrier. The modification of the proposed SW-OMP algorithm to reduce computational complexity and implement this pruning procedure is provided in Algorithm 2.

E. Convergence Analysis

In this section, the proposed algorithms SW-OMP and SS-SW-OMP+Th are theoretically analyzed to show its convergence to a local optimum. We assume that the dictionary sizes G_{t}, G_{r} are large enough such that the coarsely quantized AoA/AoD are accurately estimated, which holds for large enough values of M [9] and K. Since the residual computation for both algorithms is identical, they are analyzed following the same procedure. A sufficient condition for convergence to a local optimum is that the energy of the residual computed at the $(n+1)$-th iteration is strictly smaller than that of the n-th iteration, i.e.,

$$||\mathbf{r}^{(n+1)}[k]||_2^2 < ||\mathbf{r}^{(n)}[k]||_2^2, \quad k = 0, \ldots, K-1. \tag{28}$$

Algorithm 2 Subcarrier Selection Simultaneous Weighted Orthogonal Matching Pursuit + Thresholding (SS-SW-OMP+Th)

procedure SS-SW-OMP($\mathbf{y}[k]$,$\mathbf{\Phi}$,$\mathbf{\Psi}$,K_P,β,ϵ)
  Initialize counter, set of subcarriers and residual vectors
  $i=0,\ \mathcal{K}=\{\emptyset\},\ \mathbf{y}_\mathrm{w}[k]=\mathbf{D}_\mathrm{w}^{-*}\mathbf{y}[k],\ \mathbf{r}[k]=\mathbf{y}_\mathrm{w}[k],\ k=0,\ldots,K-1$
  Find the K_p strongest subcarriers
  while $i \leq K_\mathrm{p}$ **do**
    $\mathcal{K}=\mathcal{K}\cup \arg\max_{k\notin\mathcal{K}} \|\mathbf{y}[k]\|_2^2$
    $i=i+1$
  end while
  Compute the whitened equivalent measurement matrix
    $\mathbf{\Upsilon}_\mathrm{w}=\mathbf{D}_\mathrm{w}^{-*}\mathbf{\Phi}\mathbf{\Psi}$
  while MSE $>\epsilon$ **do**
    Distributed Correlation
      $\mathbf{c}[k]=\mathbf{\Upsilon}_\mathrm{w}^{*}\mathbf{r}[k],\quad k\in\mathcal{K}$
    Find the maximum projection along the different spaces
      $p^\star=\arg\max_p \sum_{k\in\mathcal{K}}|\{\mathbf{c}[k]\}_p|$
    Update the current guess of the common support
      $\hat{\mathcal{T}}=\hat{\mathcal{T}}\cup p^\star$
    Project the input signal onto the subspace given by the support using WLS
      $\mathbf{x}_{\hat{\mathcal{T}}}[k]=\left([\mathbf{\Upsilon}_\mathrm{w}]_{:,\hat{\mathcal{T}}}\right)^\dagger \mathbf{y}_\mathrm{w}[k],\ k=0,\ldots,K-1$
    Update residual
      $\mathbf{r}[k]=\mathbf{y}_\mathrm{w}[k]-[\mathbf{\Upsilon}_\mathrm{w}]_{:,\hat{\mathcal{T}}}\hat{\hat{\boldsymbol{\xi}}}[k]$, where $\hat{\hat{\boldsymbol{\xi}}}[k]=\mathbf{x}_{\hat{\mathcal{T}}}[k],\quad k=0,\ldots,K-1$
    Compute the current MSE
      $\mathrm{MSE}=\frac{1}{KML_\mathrm{r}}\sum_{k=0}^{K-1}\mathbf{r}^*[k]\mathbf{r}[k]$
  end while
  Thresholding based on maximum average power
    $P^\star=\max_\ell \frac{1}{K}\sum_{k=0}^{K-1}|\{\hat{\boldsymbol{\xi}}[k]\}_\ell|^2,$
    $\hat{p}_{\mathrm{av},i}=\frac{1}{K}\sum_{k=0}^{K-1}|\{\hat{\boldsymbol{\xi}}[k]\}_i|^2,\quad i=1,\ldots,\hat{L}$
    $\hat{\mathcal{T}}_\mathrm{Th}=\bigcup i\,/\,\hat{p}_{\mathrm{av},i}\geq\beta P^\star,\ i\in\hat{\mathcal{T}}$
    $\hat{\boldsymbol{\xi}}[k]=\{\hat{\boldsymbol{\xi}}[k]\}_{\hat{\mathcal{T}}_\mathrm{Th}},\ k=0,\ldots,K-1$
end procedure

The residual for a given iteration n can be written as

$$\mathbf{r}^{(n)}[k]=\left(\mathbf{I}_{ML_\mathrm{r}}-\mathbf{P}^{(n)}\right)\mathbf{y}_\mathrm{w}[k], \tag{29}$$

with $\mathbf{P}^{(n)}\in\mathbb{C}^{ML_\mathrm{r}\times ML_\mathrm{r}}$ a projection matrix given by $\mathbf{P}^{(n)}\triangleq[\mathbf{\Upsilon}_\mathrm{w}]_{:,\hat{\mathcal{T}}^{(n)}}\,[\mathbf{\Upsilon}_\mathrm{w}]^\dagger_{:,\hat{\mathcal{T}}^{(n)}}$. Accordingly, $\mathbf{r}^{(n)}[k]$ is the vector resulting from projecting $\mathbf{y}_\mathrm{w}[k]$ onto the subspace orthogonal to the column space of $[\mathbf{\Upsilon}_\mathrm{w}]_{:,\hat{\mathcal{T}}^{(n)}}$. Therefore, the condition in (28) can be rewritten as a function of the projection onto the column space of $[\mathbf{\Upsilon}_\mathrm{w}]_{:,\hat{\mathcal{T}}^{(n)}}$ as

$$||\mathbf{P}^{(n+1)}\mathbf{y}_\mathrm{w}[k]||_2^2>||\mathbf{P}^{(n)}\mathbf{y}_\mathrm{w}[k]||_2^2. \tag{30}$$

TABLE I
ONLINE COMPUTATIONAL COMPLEXITY OF PROPOSED ALGORITHMS

SW-OMP	
Operation	Complexity
$(K)\times\mathbf{c}[k]=\mathbf{\Upsilon}_\mathrm{w}^{*}\mathbf{r}[k]$	$\mathcal{O}(K(G_\mathrm{r}G_\mathrm{t}-(j-1))L_\mathrm{r}M)$
Maximum of $\sum_{k=0}^{N-1}\|\{\mathbf{c}[k]\}_p\|$	$\mathcal{O}(K(G_\mathrm{r}G_\mathrm{t}-(j-1)))$
$(K)\times\mathbf{x}_{\hat{\mathcal{T}}}[k]=\left([\mathbf{\Upsilon}_\mathrm{w}]_{:,\hat{\mathcal{T}}}\right)^\dagger\mathbf{y}_\mathrm{w}[k]$	$\mathcal{O}(2j^2L_\mathrm{r}M+j^3)$
$(K)\times\mathbf{r}[k]=\mathbf{y}_\mathrm{w}[k]-[\mathbf{\Upsilon}_\mathrm{w}]_{:,\hat{\mathcal{T}}}\hat{\hat{\boldsymbol{\xi}}}[k]$	$\mathcal{O}(KL_\mathrm{r}M)$
$\mathrm{MSE}=\frac{1}{KML_\mathrm{r}}\sum_{k=0}^{K-1}\mathbf{r}^*[k]\mathbf{r}[k]$	$\mathcal{O}(KL_\mathrm{r}M)$
Overall	$\mathcal{O}(L_\mathrm{r}M(K(G_\mathrm{r}G_\mathrm{t}-(j-1))))$
SS-SW-OMP+Thresholding	
Operation	Complexity
Find the K_p strongest subcarriers	$\mathcal{O}(KL_\mathrm{r}M)$
$(K_\mathrm{p})\times\mathbf{c}[k]=\mathbf{\Upsilon}_\mathrm{w}^{*}\mathbf{r}[k]$	$\mathcal{O}(K_\mathrm{p}(G_\mathrm{r}G_\mathrm{t}-(j-1))L_\mathrm{r}M)$
Maximum of $\sum_{k\in\mathcal{K}}\|\{\mathbf{c}[k]\}_p\|$	$\mathcal{O}(K_\mathrm{p}(G_\mathrm{r}G_\mathrm{t}-(j-1)))$
$(K)\times\mathbf{x}_{\hat{\mathcal{T}}}[k]=\left([\mathbf{\Upsilon}_\mathrm{w}]_{:,\hat{\mathcal{T}}}\right)^\dagger\mathbf{y}_\mathrm{w}[k]$	$\mathcal{O}(2j^2L_\mathrm{r}M+j^3)$
$(K)\times\mathbf{r}[k]=\mathbf{y}_\mathrm{w}[k]-[\mathbf{\Upsilon}_\mathrm{w}]_{:,\hat{\mathcal{T}}}\hat{\hat{\boldsymbol{\xi}}}[k]$	$\mathcal{O}(KL_\mathrm{r}M)$
$\mathrm{MSE}=\frac{1}{KML_\mathrm{r}}\sum_{k=0}^{K-1}\mathbf{r}^*[k]\mathbf{r}[k]$	$\mathcal{O}(KL_\mathrm{r}M)$
Thresholding	$\mathcal{O}(K\hat{L})$
Overall	$\mathcal{O}(L_\mathrm{r}M(K_\mathrm{p}(G_\mathrm{r}G_\mathrm{t}-(j-1))))$

Observe that the term inside the ℓ_2-norm on the left side of (30) can be expressed as

$$\mathbf{P}^{(n+1)}\mathbf{y}_\mathrm{w}[k]=\left[\,[\mathbf{\Upsilon}_\mathrm{w}]_{:,\hat{\mathcal{T}}^{(n)}}\quad[\mathbf{\Upsilon}_\mathrm{w}]_{:,\hat{p}^{(n+1)*}}\,\right]\times\left[\,[\mathbf{\Upsilon}_\mathrm{w}]_{:,\hat{\mathcal{T}}^{(n)}}\quad[\mathbf{\Upsilon}_\mathrm{w}]_{:,\hat{p}^{(n+1)*}}\,\right]^\dagger\mathbf{y}_\mathrm{w}[k], \tag{31}$$

with $\hat{p}^{(n+1)*}$ the estimate for the support index found during the $(n+1)$-th iteration, satisfying $\hat{p}^{(n+1)*}\notin\hat{\mathcal{T}}^{(n)}$. Thereby, we can recursively write the projection matrix $\mathbf{P}^{(n+1)}$ as a function of $\mathbf{P}^{(n)}$ using the formula for the inverse of a 2×2 block matrix [34] as

$$\mathbf{P}^{(n+1)}=\mathbf{P}^{(n)}+\underbrace{\frac{\left(\mathbf{I}_{ML_\mathrm{r}}-\mathbf{P}^{(n)}\right)[\mathbf{\Upsilon}_\mathrm{w}]_{:,\hat{p}^{(n)*}}[\mathbf{\Upsilon}_\mathrm{w}]^*_{:,\hat{p}^{(n)*}}\left(\mathbf{I}_{ML_\mathrm{r}}-\mathbf{P}^{(n)}\right)}{[\mathbf{\Upsilon}_\mathrm{w}]^*_{:,\hat{p}^{(n)*}}\left(\mathbf{I}_{ML_\mathrm{r}}-\mathbf{P}^{(n)}\right)[\mathbf{\Upsilon}_\mathrm{w}]_{:,\hat{p}^{(n)*}}}}_{\mathbf{\Delta P}^{(n+1)}}, \tag{32}$$

with $\mathbf{\Delta P}^{(n+1)}\in\mathbb{C}^{ML_\mathrm{r}\times ML_\mathrm{r}}$ another projection matrix accounting for the relation between the projections at the n-th and $(n+1)$-th iterations. The equation in (32) can be noticed to fulfill the orthogonality principle, $\mathbf{P}^{(n+1)}\mathbf{\Delta P}^{(n+1)}=\mathbf{0}$. The left-handed term in (30) then can be expressed as

$$\begin{aligned}||\mathbf{P}^{(n+1)}\mathbf{y}_\mathrm{w}[k]||_2^2&=||\mathbf{P}^{(n)}\mathbf{y}_\mathrm{w}[k]+\mathbf{\Delta P}^{(n+1)}\mathbf{y}_\mathrm{w}[k]||_2^2\\&=||\mathbf{P}^{(n)}\mathbf{y}_\mathrm{w}[k]||_2^2+||\mathbf{\Delta P}^{(n+1)}\mathbf{y}_\mathrm{w}[k]||_2^2,\end{aligned} \tag{33}$$

satisfying thereby the triangle equality. Finally, from (33), it is inmediate that $||\mathbf{P}^{(n+1)}\mathbf{y}_\mathrm{w}[k]||_2^2>||\mathbf{P}^{(n)}\mathbf{y}_\mathrm{w}[k]||_2^2$ since $\mathbf{\Delta P}^{(n+1)}$ has a non-zero eigenvalue equal to 1. Thereby, the condition in (30) is satisfied and convergence of the proposed algorithms to a local optimum is guaranteed.

TABLE II
ONLINE COMPUTATIONAL COMPLEXITY OF PREVIOUSLY PROPOSED ALGORITHMS

OMP from [22]	
Operation	Complexity
$(K) \times \mathbf{c}[k] = \mathbf{\Upsilon}^*[k]\mathbf{r}[k]$	$\mathcal{O}(K(G_\mathrm{r}G_\mathrm{t} - (j-1))L_\mathrm{r}M)$
$(K)\times$ Maximum of $\lvert\{\mathbf{c}[k]\}_p\rvert$	$\mathcal{O}(K(G_\mathrm{r}G_\mathrm{t} - (j-1)))$
$(K) \times \mathbf{x}_{\hat{\mathcal{T}}}[k] = [\mathbf{\Upsilon}]^\dagger_{:,\hat{\mathcal{T}}}[k]\mathbf{z}[k]$	$\mathcal{O}(K(2j^2L_\mathrm{r}M + j^3))$
$(K) \times \mathbf{r}[k] = \mathbf{y}[k] - [\mathbf{\Upsilon}]_{:,\hat{\mathcal{T}}}[k]\hat{\boldsymbol{\xi}}[k]$	$\mathcal{O}(KL_\mathrm{r}M)$
$(K) \times \mathrm{MSE} = \frac{1}{ML_\mathrm{r}}\mathbf{r}^*[k]\mathbf{r}[k]$	$\mathcal{O}(KL_\mathrm{r}M)$
Overall	$\mathcal{O}(KL_\mathrm{r}M(G_\mathrm{r}G_\mathrm{t} - (j-1) + 2j^2))$
SSAMP from [24], [25]	
Operation	Complexity
$(K) \times \mathbf{a}_p = \mathbf{\Phi}_p^*\mathbf{b}_p^{i-1}$	$\mathcal{O}(KG_\mathrm{r}G_\mathrm{t}L_\mathrm{r}M)$
Maximum of $\sum_{p=0}^{N-1}\lVert\mathbf{a}_p\rVert_2^2$	$\mathcal{O}(NG_\mathrm{t}G_\mathrm{r})$
$(K) \times \{\mathbf{t}_p\}_{\Omega^{i-1}\cup\Gamma} = ((\mathbf{\Phi}_p)^\dagger_{\Omega^{i-1}\cup\Gamma}\mathbf{r}_p$	$\mathcal{O}(K(4j^2ML_\mathrm{r} + (2j)^3))$
Prune support $\Omega = \arg\max_{\tilde{\Omega}}\{\sum_{p=1}^{P}\lVert\{\mathbf{t}_p\}_{\tilde{\Omega}}\rVert_2^2\}, \lvert\tilde{\Omega}\rvert = \mathcal{T}$	$\mathcal{O}(2Kj)$
$(K) \times \{\mathbf{c}_p\}_\Omega = (\{\mathbf{\Phi}_p\}_\Omega)^\dagger\mathbf{r}_p$	$\mathcal{O}(K(2j^2ML_\mathrm{r} + j^3))$
$(K) \times \mathbf{b}_p = \mathbf{r}_p - \mathbf{\Phi}_p\mathbf{c}_p$	$\mathcal{O}(KL_\mathrm{r}MG_\mathrm{t}G_\mathrm{r})$
Computation of total error $\sum_{p=0}^{K-1}\lVert\mathbf{b}_p\rVert_2^2$	$\mathcal{O}(KML_\mathrm{r})$
Overall	$\mathcal{O}(2KL_\mathrm{r}M(G_\mathrm{r}G_\mathrm{t} + 6j^2))$
DGMP from [27]	
Operation	Complexity
$(K) \times \mathbf{a}_p = \mathbf{\Upsilon}_p^*\mathbf{r}_p$	$\mathcal{O}(KG_\mathrm{r}G_\mathrm{t}L_\mathrm{r}M)$
Maximum of $\rho = \arg\max_{\tilde{\rho}}\sum_{p=0}^{N-1}\lVert\mathbf{a}_p\rVert_2^2$	$\mathcal{O}(KG_\mathrm{t}G_\mathrm{r})$
$(K) \times \{\alpha_p\}_\rho = (\{\mathbf{\Phi}_p\}^\dagger_\rho\mathbf{r}_p$	$\mathcal{O}(K(2j^2ML_\mathrm{r} + j^3))$
Overall	$\mathcal{O}(KL_\mathrm{r}M(G_\mathrm{r}G_\mathrm{t} + 2j^2))$

IV. RESULTS

This section includes the main empirical results obtained with the two proposed algorithms, SW-OMP and SS-SW-OMP + Thresholding, and comparisons with other frequency-domain channel estimation algorithms including SSAMP [25] and DGMP [27]. To obtain these results, we perform Monte Carlo simulations averaged over many trials to evaluate the normalized mean squared error (NMSE) and the ergodic rate as a function of SNR and number of training frames M. We also provide calculations of the computational complexity for the proposed algorithms in Table I and prior work in Table II.

The typical parameters for our system configuration are summarized as follows. Both the transmitter and the receiver are assumed to use Uniform Linear Arrays (ULAs) with half-wavelength separation. Such a ULA has steering vectors obeying the expressions $\{\mathbf{a}_\mathrm{T}(\theta_\ell)\}_n = \sqrt{\frac{1}{N_\mathrm{t}}}e^{\mathrm{j}n\pi\cos(\theta_\ell)}$, $n = 0,\ldots,N_\mathrm{t}-1$ and $\{\mathbf{a}_\mathrm{R}(\phi_\ell)\}_m = \sqrt{\frac{1}{N_\mathrm{r}}}e^{\mathrm{j}m\pi\cos(\phi_\ell)}$, $m = 0,\ldots,N_\mathrm{r}-1$. We take $N_\mathrm{t} = N_\mathrm{r} = 32$ and $G_\mathrm{t} = G_\mathrm{r} = 64$. The phase-shifters used in both the transmitter and the receiver are assumed to have N_Q quantization bits, so that the entries of the training vectors $\mathbf{f}_\mathrm{tr}^{(m)}$, $\mathbf{w}_\mathrm{tr}^{(m)}$, $m = 1,2,\ldots,M$ are drawn from a set $\mathcal{A} = \left\{0, \frac{2\pi}{2^{N_\mathrm{Q}}},\ldots,\frac{2\pi(2^{N_\mathrm{Q}}-1)}{2^{N_\mathrm{Q}}}\right\}$. The number of quantization bits is set to $N_\mathrm{Q} = 2$. The number of RF chains is set to $L_\mathrm{t} = 1$ at the transmitter and $L_\mathrm{r} = 4$ at the receiver. The number of OFDM subcarriers is set to $K = 16$.

We generate channels according to (2) with the following parameters:

- The $L = 4$ channel paths are assumed to be independent and identically distributed, with delay τ_ℓ chosen uniformly from $[0,(N_\mathrm{c}-1)T_\mathrm{s}]$, with $T_\mathrm{s} = \frac{1}{1760}$ μs, as in the IEEE 802.11ad wireless standard.
- The AoAs/AoDs are assumed to be uniformly distributed in $[0,\pi)$.
- The gains of each path are zero-mean complex Gaussian distributed such that $\mathbb{E}_k\{\lVert\mathbf{H}[k]\rVert_F^2\} = \frac{N_\mathrm{r}N_\mathrm{t}}{\rho_\mathrm{L}}$.
- The band-limiting filter $p_\mathrm{rc}(t)$ is assumed to be a raised-cosine filter with roll-off factor of 0.8.
- The number of delay taps of the channel is set to $N_\mathrm{c} = 4$ symbols.

The simulations we perform consider channel realizations in which the AoA/AoD are off-grid, i.e. do not correspond to the angles used to build the dictionary, and also the on-grid case, to analyze the loss due to the model mismatch.

A. NMSE Comparison

One performance metric is the Normalized Mean Squared Error (NMSE) of a channel estimate $\hat{\mathbf{H}}[k]$ for a given realization, defined as

$$\mathrm{NMSE} = \frac{\sum_{k=0}^{K-1}\lVert\hat{\mathbf{H}}[k] - \mathbf{H}[k]\rVert_F^2}{\sum_{k=0}^{K-1}\lVert\mathbf{H}[k]\rVert_F^2}. \tag{34}$$

The NMSE will be our baseline metric to compute the performance of the different estimators, and will be averaged over many channel realizations. The normalized CRLB (NCRLB) is also provided to compare the average performance of each

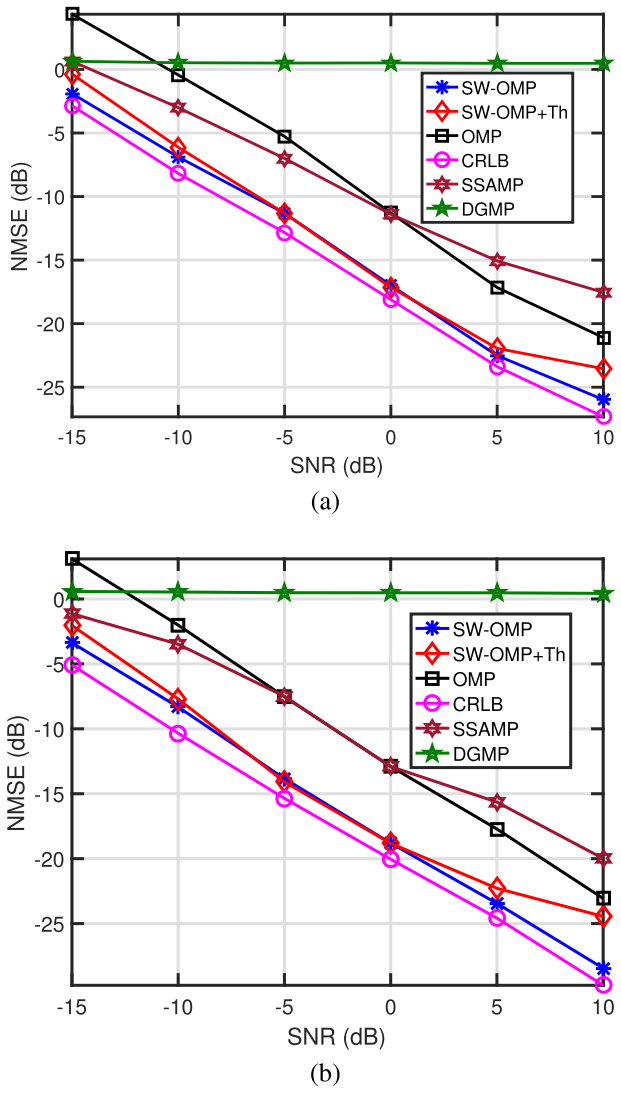


Fig. 2. Evolution of the NMSE versus SNR for the different frequency-domain algorithms when the AoA/AoDs are assumed to lie on the dictionary grid. The number of training frames is set to $M = 80$ (a) and $M = 120$ (b).

algorithm with the lowest achievable NMSE, and will also be averaged over many channel realizations

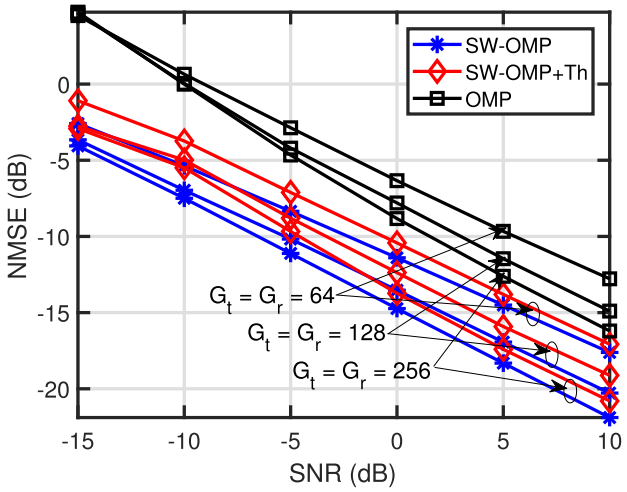
$$\gamma(\{\text{vec}\{\mathbf{H}[k]\}\}_{k=0}^{K-1}) = \frac{\sum_{k=0}^{K-1} \mathbf{I}^{-1}(\text{vec}\{\mathbf{H}[k]\})}{\sum_{k=0}^{K-1} \|\mathbf{H}[k]\|_F^2}. \quad (35)$$

We compare the average NMSE versus SNR obtained for the different channel estimation algorithms in Fig. 2 for a practical SNR range of -15 dB to 10 dB, on-grid AoA/AoDs, and two different values for the number of training frames. SW-OMP performs the best, achieving NMSE values very close to the NCRLB. SS-SW-OMP+Th performs similarly to SW-OMP, although there is some performance loss due to the fact that SS-SW-OMP+Th does not employ every subcarrier to estimate the common support of the sparse channel vectors. In SS-SW-OMP+Th, the number of selected subcarriers for the estimation of the support was set to $K_\text{p} = 4$ for illustration and the parameter β was chosen as $\beta = 0.025$, which is a reasonably small value to filter out undesired components in the sparse channel estimate. From the curves shown in Fig. 2, OMP performs poorly for all the SNR range since it is not designed to process several vectors which are sparse in a common vector basis. Exploiting common sparsity provides an NMSE reduction of approximately 7 dB, although there are slight variations depending on the SNR. This improvement comes at the cost of a higher offline computational complexity in the proposed algorithms in comparison with OMP, as we show in Section IV-C.

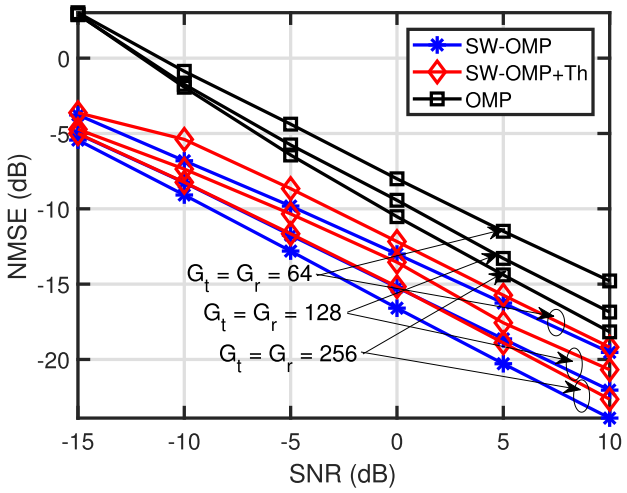
We also observe that the DGMP algorithm from [27] performs the worst, as it has been designed to estimate near-LoS mm-wave MIMO channels. Since it only estimates a single path, the estimation error for NLOS channels is large. The algorithm SSAMP is also shown for comparison. We can see that, at low SNR regime, the information on the common support is enough to outperform the OMP algorithm, but not at the high SNR regime. This comes from the fact that the algorithm is estimating more than a single path per iteration. Since the dictionary matrices are not square in this setting, the redundancy between columns in the transmit and receive array matrices makes it difficult to properly estimate more than a support index per iteration.

Using a larger number of training frames M increases performance, but at the cost of both higher overhead and computational complexity, since the complexity of estimating the support, channel gains and noise variance grows linearly with $L_\text{r}M$. Nonetheless, there is an SNR-increasing performance gap between SS-SW-OMP+Th and SW-OMP. Even further, this gap increases with the number of training frames, which may seem counter-intuitive. This is because the variance of the MVU estimator for the channel gains depends on the SNR. Clearly, in high SNR regime, the estimates of the weakest paths have smaller estimation variance than in low SNR regime. Therefore, since the threshold (dependent on β) is not adapted to the noise variance, these paths are more likely to be removed at higher SNR. When M increases, the estimation variance also decreases, which further increases this gap, as observed in Fig. 2(a) and 2(b).

The previous simulations showed the performance of the algorithms when the channel fits the on-grid model, but it is also important to analyze the performance in a practical scenario, when the AoA/AoD do not fall within the quantized spatial grid. Fig. 3 shows the performance of the different algorithms under a more challenging scenario with channel realizations extracted from the NYUSIM channel simulator [29]. The simulation parameters for this scenario are chosen as $K = 256$, $K_\text{p} = K/4 = 64$, $\beta = 0.01\sigma^2$ and $L_\text{t} = L_\text{r} = 4$. A ZP of length $\lceil K/3 \rceil = 86$ samples is chosen to guarantee that most of channel energy, i.e., 99% is contained in the ZP. The remaining parameters are the same as in Fig. 2. Results are shown for $M = 60$ frames (a), and $M = 100$ frames (b). The estimation error is observed to be below -10 dB for values of SNR in the order of 0 and beyond. On the other hand, since the SNR expected in mm-wave communication systems is in the order of -20 dB up to 0 dB, the gap between the attained NMSE and the NCRLB should be reduced. Increasing the size of the dictionary is one of the possible solutions, as shown by the curves corresponding to $G_\text{t} = G_\text{r} = 128$ and $G_\text{t} = G_\text{r} = 256$.



(a)



(b)

Fig. 3. Evolution of the NMSE versus SNR for the different frequency-domain algorithms. The number of training frames is set to $M = 60$ (a) and $M = 100$ (b). The channel realizations are taken from NYUSIM channel simulator.

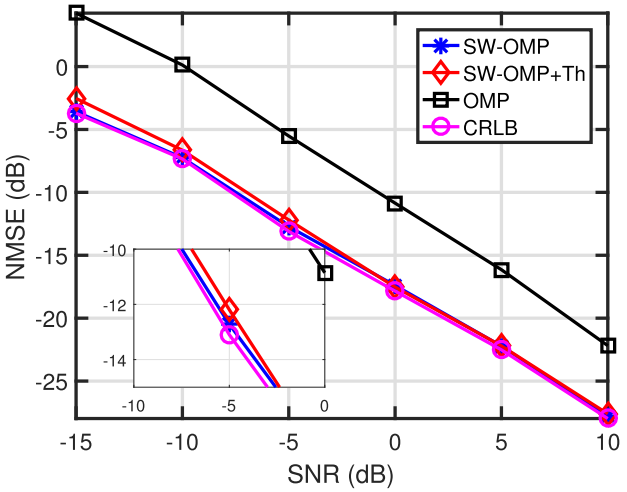


Fig. 4. Comparison of evolution of the NMSE versus SNR for the different frequency-domain algorithms. The number of training frames is set to $M = 80$. The number of subcarriers is set to $K = 64$.

We show in Fig. 4 the performance of the different frequency-domain algorithms when increasing the number of subcarriers. The parameters for the simulation scenario are the same as in Fig. 2, however, the number of subcarriers is set to $K = 64$ in this case. K_{p} is set to 32 subcarriers and $\beta = 0.025\sigma^2$. Interestingly, both SW-OMP and SS-SW-OMP+Th are asymptotically efficient since they are both unbiased and attain the NCRLB. A magnified plot for SNR = -5 dB is also shown to clearly see the performance gap between the different algorithms and also the NCRLB.

Fig. 5 shows the average NMSE vs number of training frames M. The number of training frames M is increased from 20 to 100. The remaining parameters in the simulation scenario are the same as in Fig. 3, with $G_{\mathrm{t}} = G_{\mathrm{r}} = 128$. Results are shown for channel realizations in which the angular parameters fall within the quantized angle grid and when they do not. The average performance of OMP algorithm is poor for all the considered cases, which comes from its inability to exploit the common support property shared by the different subchannels. SW-OMP can be observed to provide the best performance for all the values of M and SNR. The larger the number of training frames and the higher the SNR, the estimation of the support is more robust and gets closer to the actual one. In the on-grid case, if the number of training frames is large enough and the SNR is not low, the performance gap between SW-OMP and the NCRLB is lower than 1 dB. The difference in performance between SS-SW-OMP+Th and SW-OMP reduces when either M or SNR is increased. As in Figs. 2 and 3, there is a big difference in performance between OMP and SW-OMP, depending on both the SNR and the number of frames.

B. Spectral Efficiency Comparison

Another performance metric is the spectral efficiency, computed assuming fully-digital precoding and combining using estimates for the N_{s} dominant left and right singular vectors of the channel estimate. This gives K parallel effective channels $\mathbf{H}_{\mathrm{eff}}[k] = \left[\hat{\mathbf{U}}[k]\right]^{*}_{:,1:N_{\mathrm{s}}} \mathbf{H}[k] \left[\hat{\mathbf{V}}[k]\right]_{:,1:N_{\mathrm{s}}}$. Accordingly, the average spectral efficiency can be expressed as [35]

$$R = \frac{1}{K} \sum_{k=0}^{K-1} \sum_{n=1}^{N_{\mathrm{s}}} \log \left(1 + \frac{\mathrm{SNR}}{N_{\mathrm{s}}} \lambda_n (\mathbf{H}_{\mathrm{eff}}[k])^2 \right), \quad (36)$$

with $\lambda_n(\mathbf{H}_{\mathrm{eff}}[k])$, $n = 1, \ldots, N_{\mathrm{s}}$ the eigenvalues of each effective channel $\mathbf{H}_{\mathrm{eff}}[k]$.

In Fig. 6(a), we show the achievable spectral efficiency as a function of the SNR for the different channel estimation algorithms when realistic channel realizations are considered. The simulation parameters are the same as in Fig. 3. The difference in performance between OMP and either of our two proposed algorithms is noticeable, which comes from the fact that OMP does not force the channel estimates to share the same support. The two proposed algorithms perform similarly for all the range of SNR, which is an indicator that $K_{\mathrm{p}} < K$ subcarriers are enough to obtain a reliable channel estimate.

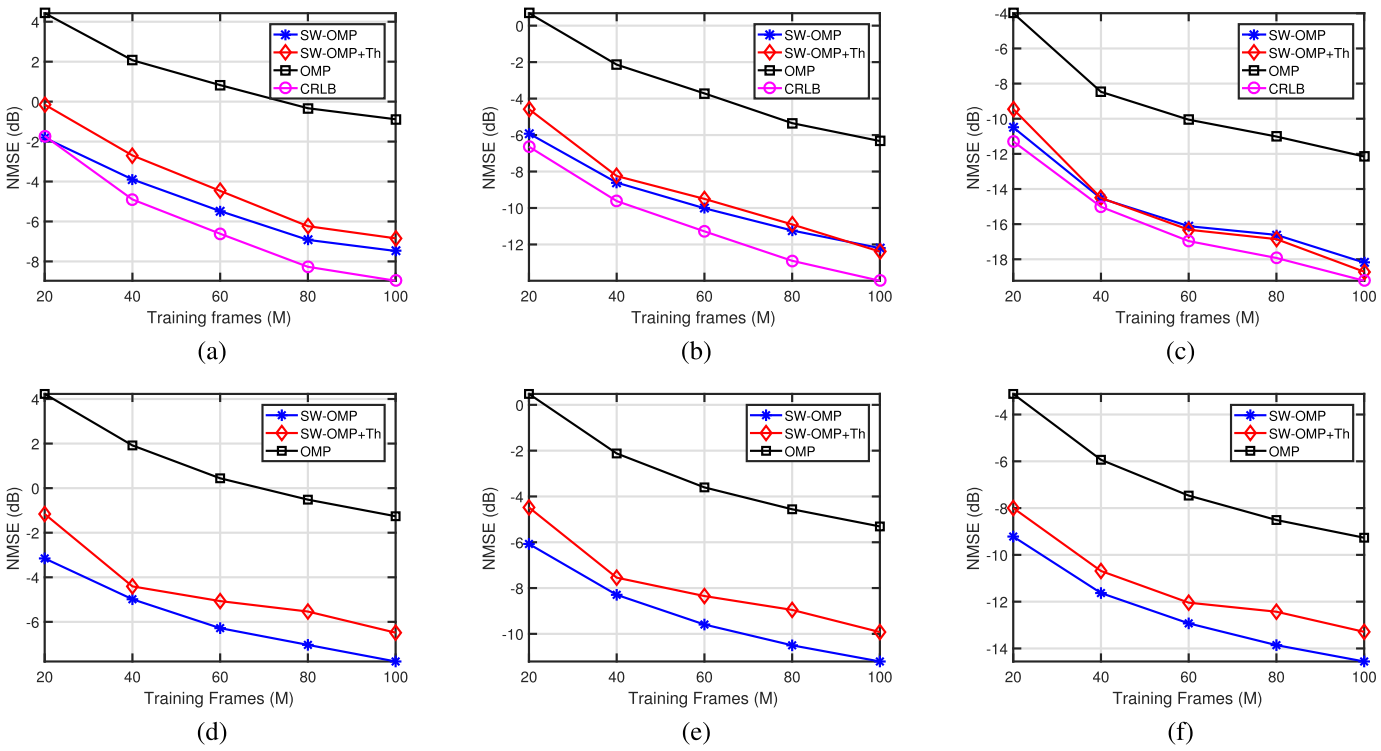


Fig. 5. Comparison of evolution of the NMSE versus number of training frames M at different SNR for the different channel estimation algorithms. The SNR is set to -10 dB (a) and (d), -5 dB (b) and (e) and 0 dB (c) and (f). The curves in the first row consider on-grid angular parameters on the channel realizations, while the ones in the second row consider the off-grid case with channel realizations extracted from the NYUSIM channel simulator [29].

Therefore, SS-SW-OMP+Th can be claimed to be a good trade-off between performance and computational complexity.

Finally, Fig. 6(b) shows the spectral efficiency as a function of the number of training frames M for the different channel estimation algorithms. Comparisons are provided for SNR $= \{-10, -5, 0\}$ dB. We observe that SW-OMP is the algorithm providing the best performance, followed closely by SS-SW-OMP+Th, whilst OMP performs the worst. For low values of SNR, there is a noticeable gap between SW-OMP and the perfect CSI case, which becomes smaller as M increases. We observe that using $M \geq 40$ frames does not bring about significant improvement in performance, which leverages the robustness of the two proposed approaches. Simulations also show that near-optimal achievable rates can be achieved by using a reasonable number of frames, i.e., $40 \leq M \leq 100$.

C. Computational Complexity

The computational complexity for each step in the different algorithms is also provided for the j-th iteration in Table I. Since some steps can be performed before running the channel estimation algorithms, we will distinguish between on-line and off-line operations. Values are provided for a single iteration. In the case of the SSAMP algorithm [24], [25] and the DGMP algorithm in [27], we take the notation used in the corresponding papers for frequency-domain vectors and measurement matrices.

The computational complexity of SS-SW-OMP+Th is lower than in SW-OMP, since only a smaller number of projections are computed to estimate the channel support. It must be noticed that the complexity of SS-SW-OMP+Th is lower than that of OMP, since the matrix product $\boldsymbol{\Upsilon}_{\mathrm{w}} = \mathbf{D}_{\mathrm{w}}^{-*}\boldsymbol{\Upsilon}$ can be computed before explicit channel estimation. The online computational complexity of SW-OMP is lower than that of OMP, since OMP computes K matrix pseudoinverse while SW-OMP only computes one. Conversely, the offline computational complexity of both SW-OMP and SS-SW-OMP+Th are higher than those for the other algorithms, since the matrix $\boldsymbol{\Upsilon}_{\mathrm{w}}$ must be computed before explicit channel estimation.

We observe that the computational complexity of SW-OMP is lower than its SSAMP counterpart. SSAMP exhibits an increase in complexity of at most $\mathcal{O}(4j^2)$ owing to the estimation of j paths at the j-th iteration of SSAMP. More especifically, this algorithm uses an iteration index i to estimate the sparsity level L, and a stage index j to estimate the j channel paths found at the current iteration. Afterwards, the support of the channel estimate is pruned to select the j most likely channel paths. Therefore, at a given iteration i and stage j, at most $2j = |\tilde{\Omega}^{i-1} \cup \Gamma|$ paths are estimated and then pruned, such that only j paths are selected among the $2j$ candidates. The union of the sets Ω and Γ comes from the possibility of finding new potential paths at the i-th iteration and the j-th stage. This is done by jointly considering the paths found at $(i-1)$-th iteration and the ones found in the j-th stage within the i-th iteration. While both SW-OMP, SS-SW-OMP+Th and OMP estimate a single path at a given iteration j, SSAMP estimates at most $2j$ different paths by using LS. When computing the pseudoinverse during LS estimation, this results in an additional increase in complexity of $\mathcal{O}(4j^2)$, as shown in Table II. By contrast, as shown

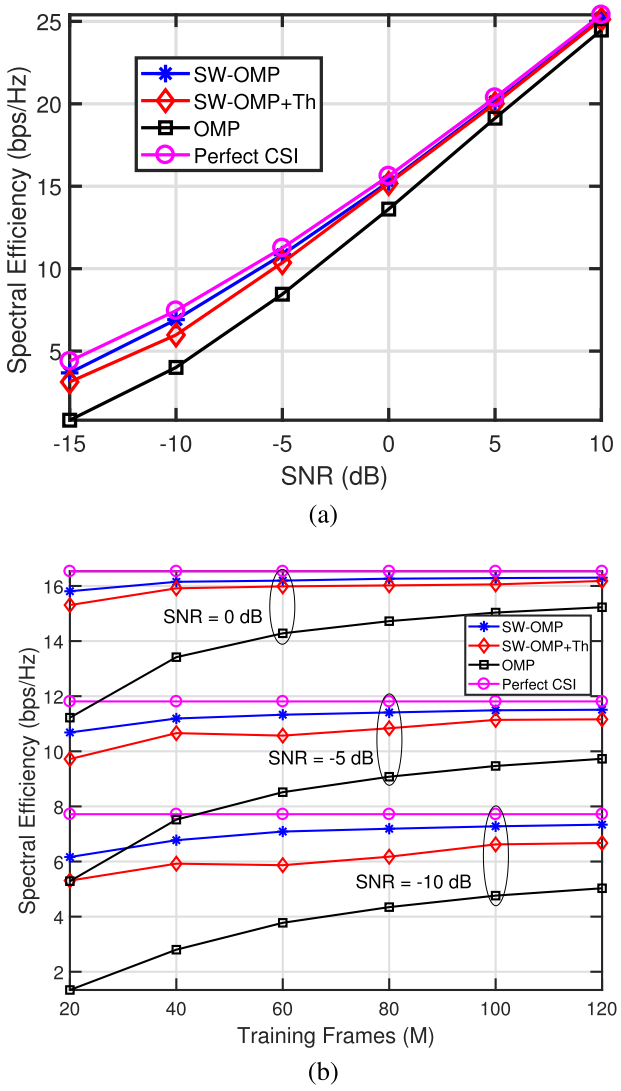


Fig. 6. (a) Evolution of the spectral efficiency versus SNR for the different frequency-domain algorithms. The number of training frames is set to $M = 60$. (b) Evolution of the spectral efficiency versus number of training frames M at different SNR for the different frequency-domain algorithms.

in Table I, our proposed ML estimator for the channel gains exhibits computational complexity in the order of $\mathcal{O}(2j^2)$, thereby slightly reducing the number of operations.

While OMP does not require any offline operation, both SW-OMP and SS-SW-OMP+Th need to compute the matrix $\boldsymbol{\Upsilon}_{\mathrm{w}}$. The offline computation of $\mathbf{D}_{\mathrm{w}}^{-1}$ has complexity of $\mathcal{O}(\frac{M}{3}L_{\mathrm{r}}^3)$, since $\mathbf{C}_{\mathrm{w}}^{-1}$ is a block diagonal matrix containing M hermitian matrices. It is important to remark that this cost comes from the use of frequency-flat precoders/combiners. This entails a reduction in computational complexity with respect to the case in which frequency-selective baseband combiners were used during the channel estimation stage. In summary, the proposed algorithms reduce the computational complexity by approximately a factor of K.

D. Bit Error Rate (BER)

The last performance metric we consider in this paper is the Bit Error Rate (BER). Simulation results on the BER achieved by our proposed algorithms are provided, as well as comparisons with the baseline approach in [22]. We assume that channel estimation has already been performed, and a data transmission phase takes places. We provide both uncoded and coded BER results. For the latter, we use Low Density Parity Check (LDPC) codes, inspired by the IEEE 802.11ad mm-wave communications standard [4], to encode the data bits to be transmitted. We consider the OFDM-PHY transmission frame specified in the aforementioned standard, and use the Modulation and Coding Scheme (MCS) 18 for illustration. Such MCS considers a 16-QAM constellation using dual carrier modulation [4], in which two constellation symbols are shared between two subcarriers, and LDPC coding with rate $R_{\mathrm{LDPC}} = \frac{1}{2}$. To estimate the equivalent (beamformed) channel, we use known symbols from the pilot subcarriers contained in this frame, and then use spline interpolation to estimate the beamformed channel at the data subcarriers. We denote the set of positions of pilot subcarriers as $\mathcal{P}$. Likewise, the set of positions of data subcarriers is denoted as $\mathcal{D}$. Let us express the SVD of the MIMO channel estimate $\hat{\mathbf{H}}[k]$ as $\hat{\mathbf{H}}[k] = \hat{\mathbf{U}}[k]\hat{\boldsymbol{\Sigma}}[k]\hat{\mathbf{V}}^*[k]$. For the data phase, we transmit information using a precoder $\mathbf{f}[k] \in \mathbb{C}^{N_{\mathrm{t}} \times 1}$, $\mathbf{f}[k] = \left[\hat{\mathbf{V}}[k]\right]_{:,1}$, and combine the received MIMO signal using a combiner $\mathbf{w}[k] \in \mathbb{C}^{N_{\mathrm{r}} \times 1}$, $\mathbf{w}[k] = \left[\hat{\mathbf{U}}[k]\right]_{:,1}$. Transmitted data in IEEE 802.11ad is partitioned into B_{data} OFDM symbols with $K = 512$ subcarriers, with $N_{\mathrm{SD}} = 336$ data subcarriers, and $N_{\mathrm{SP}} = 16$ pilot subcarriers [4].

Let $\mathsf{s}^{(l)}[p] \in \mathbb{C}$ be the transmitted pilot in the l-th data block at subcarrier $p \in \mathcal{P}$, $\mathsf{s}^{(l)}[d] \in \mathbb{C}$ the data constellation symbol at subcarrier $d \in \mathcal{D}$, $\mathsf{g}[k] \in \mathbb{C}$ be the equivalent channel $\mathsf{g}[k] = \mathbf{w}^*[k]\mathbf{H}[k]\mathbf{f}[k]$, $\mathsf{g}[k] = \alpha[k]e^{j\phi[k]}$, and $\mathsf{n}^{(i)}[k] \in \mathbb{C}$ the combined Gaussian noise sample, at subcarrier k. Then, the l-th received block at subcarrier k is given by

$$\mathsf{r}^{(l)}[k] = \mathsf{g}[k]\mathsf{s}^{(l)}[k] + \mathsf{n}^{(l)}[k], \quad k = 0, \ldots, K-1. \tag{37}$$

The unknown parameters $\alpha[k]$, $\phi[k]$ are estimated using the ML criterion, for which the ML estimators are [34]

$$\hat{\phi}_{\mathrm{ML}}[p] = \tan^{-1}\left\{\frac{\sum_{l=1}^{B_{\mathrm{data}}} \mathrm{Im}\{\mathsf{r}^{(l)}[p]\bar{\mathsf{s}}^{(l)}[k]\}}{\sum_{l=1}^{B_{\mathrm{data}}} \mathrm{Re}\{\mathsf{r}^{(l)}[p]\bar{\mathsf{s}}^{(l)}[k]\}}\right\}$$
$$\hat{\alpha}_{\mathrm{ML}}[p] = \frac{1}{B_{\mathrm{data}}}\sum_{l=1}^{B_{\mathrm{data}}}\left|\mathsf{r}^{(l)}[p]\bar{\mathsf{s}}^{(l)}[p]e^{-j\hat{\phi}_{\mathrm{ML}}[p]}\right| \tag{38}$$

After interpolating the estimates $\mathsf{g}[k]$ at $k \in \mathcal{D}$ using $\mathsf{g}[p]$, $p \in \mathcal{P}$, the receiver estimates Log-Likelihood Ratios (LLRs) for the coded bits, and the trasnmitted data bits are estimated using the belief propagation algorithm.

We show in Fig. 7 the empirical BER obtained with the proposed algorithms and the baseline approach, as a function of the SNR, for the same simulation parameters as with Fig. 3, and with $K = 512$, $M = 60$, and $G_{\mathrm{t}} = G_{\mathrm{r}} = 128$. We observe that the proposed approaches outperform the baseline channel estimation technique for both uncoded and coded BER, and for all values of SNR. At the very low SNR regime, our proposed algorithms show a noticeable performance gain, which reduces as the SNR increases because estimating the channel support becomes easier. There is, however, still a performance gain at

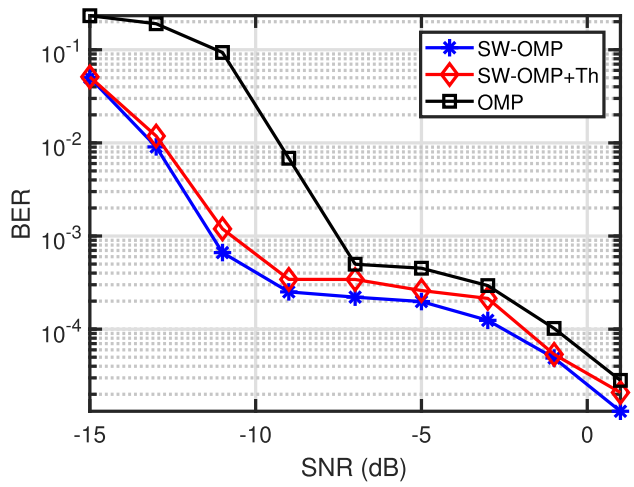


Fig. 7. (a) Evolution of the coded BER versus SNR for the different frequency-domain algorithms. The number of training frames is set to $M = 60$. (b) Evolution of the uncoded BER versus SNR for the different frequency-domain algorithms. The number of training frames is set to $M = 60$.

the medium-high SNR regime, which comes from the use of the noise covariance matrix in our proposed algorithms. This results in more accurate estimates of the channel gains, and also better estimates of the sparsity level.

V. Conclusions

In this paper, we proposed two compressive channel estimation approaches suitable for OFDM-based communication systems. These two strategies are based on jointly-sparse recovery to exploit information on the common basis that is shared for every subcarrier. Our compressive approaches enable MIMO operation in mm-wave systems since the different subchannels are simultaneously estimated during the training phase. Further, if there is no grid quantization error and the estimation of the support is correct, we showed that our algorithms are asymptotically efficient since they asymptotically attain the CRLB. In simulations, we found that only a small number of subcarriers provide a high probability of correct support detection, thus our estimators approach the CRLB with reduced computational complexity. The approaches were also found to work well even with off-grid parameters, and to outperform competitive frequency-domain channel estimation approaches. For future work, it would be interesting to analytically calculate the minimum required number of subcarriers to guarantee a high probability of correctly recovering the support of the sparse channel vectors. It would also be interesting to study the effects of other impairments including array miscalibration, beam squint, and synchronization errors.

References

[1] T. S. Rappaport, R. W. Heath, Jr., R. C. Daniels, and J. N. Murdock, *Millimeter Wave Wireless Communications*. London, U.K.: Pearson Education, 2014.

[2] R. W. Heath, Jr., N. González-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE J. Sel. Areas Commun.*, vol. 10, no. 3, pp. 436–453, Apr. 2016.

[3] A. Alkhateeb, J. Mo, N. González-Prelcic, and R. W. Heath, Jr., "MIMO precoding and combining solutions for millimeter-wave systems," *IEEE Commun. Mag.*, vol. 52, no. 12, pp. 122–131, Dec. 2014.

[4] E. Perahia, C. Cordeiro, M. Park, and L. L. Yang, "IEEE 802.11ad: Defining the next generation multi-Gbps Wi-Fi," in *Proc. IEEE Consum. Commun. Netw. Conf. (CCNC)*, Jan. 2010, pp. 1–5.

[5] J. Wang *et al.*, "Beam codebook based beamforming protocol for multi-Gbps millimeter-wave WPAN systems," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 8, pp. 1390–1399, Oct. 2009.

[6] L. Chen, Y. Yang, X. Chen, and W. Wang, "Multi-stage beamforming codebook for 60 GHz WPAN," in *Proc. 6th Int. ICST Conf. Commun. Netw. China (CHINACOM)*, Aug. 2011, pp. 361–365.

[7] S. Hur, T. Kim, D. J. Love, J. V. Krogmeier, T. A. Thomas, and A. Ghosh, "Millimeter wave beamforming for wireless backhaul and access in small cell networks," *IEEE Trans. Commun.*, vol. 61, no. 10, pp. 4391–4403, Oct. 2013.

[8] J. G. Andrews *et al.*, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, Jun. 2014.

[9] R. Méndez-Rial, C. Rusu, N. González-Prelcic, A. Alkhateeb, and R. W. Heath, Jr., "Hybrid MIMO Architectures for millimeter wave communications: Phase shifters or switches?" *IEEE Access*, vol. 4, pp. 247–267, 2016.

[10] M. L. Malloy and R. D. Nowak, "Near-optimal adaptive compressed sensing," in *Proc. 46th Asilomar Conf. Signals, Syst. Comput. (ASILOMAR)*, Pacific Grove, CA, USA, Nov. 2012, pp. 1935–1939.

[11] M. L. Malloy and R. D. Nowak. (2012). "Near-optimal compressive binary search." [Online]. Available: https://arxiv.org/abs/1203.1804

[12] M. A. Iwen and A. H. Tewfik, "Adaptive strategies for target detection and localization in noisy environments," *IEEE Trans. Signal Process.*, vol. 60, no. 5, pp. 2344–2353, May 2012.

[13] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, Jr., "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 831–846, Oct. 2014.

[14] D. Ramasamy, S. Venkateswaran, and U. Madhow, "Compressive adaptation of large steerable arrays," in *Proc. Inf. Theory Appl. Workshop (ITA)*, Feb. 2012, pp. 234–239.

[15] D. Ramasamy, S. Venkateswaran, and U. Madhow, "Compressive tracking with 1000-element arrays: A framework for multi-Gbps mm wave cellular downlinks," in *Proc. Annu. Allerton Conf. Commun., Control, Comput. (Allerton)*, Oct. 2012, pp. 690–697.

[16] D. E. Berraki, S. M. D. Armour, and A. R. Nix, "Application of compressive sensing in sparse spatial channel recovery for beamforming in mmWave outdoor systems," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Apr. 2014, pp. 887–892.

[17] J. Lee, G.-T. Gil, and Y. H. Lee, "Exploiting spatial sparsity for estimating channels of hybrid MIMO systems in millimeter wave communications," in *Proc. IEEE GLOBECOM*, Dec. 2014, pp. 3326–3331.

[18] Z. Marzi, D. Ramasamy, and U. Madhow, "Compressive channel estimation and tracking for large arrays in mm-wave picocells," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 514–527, Apr. 2016.

[19] I. Chafaa and M. Djeddou, "Improved channel estimation in mmWave communication system," in *Proc. Seminar Detection Syst. Archit. Technol. (DAT)*, Feb. 2017, pp. 1–5.
[20] Z. Xiao, P. Xia, and X.-G. Xia, "Channel estimation and hybrid precoding for millimeter-wave MIMO systems: A low-complexity overall solution," *IEEE Access*, vol. 5, pp. 16100–16110, 2017.
[21] K. Venugopal, A. Alkhateeb, R. W. Heath, Jr., and N. González-Prelcic, "Time-domain channel estimation for wideband millimeter wave systems with hybrid architecture," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2017, pp. 6493–6497.
[22] K. Venugopal, A. Alkhateeb, N. González-Prelcic, and R. W. Heath, Jr., "Channel estimation for hybrid architecture-based wideband millimeter wave systems," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 1996–2009, Sep. 2017.
[23] J. Rodríguez-Fernández, K. Venugopal, N. González-Prelcic, and R. W. Heath, Jr., "A frequency-domain approach to wideband channel estimation in millimeter wave systems," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2017, pp. 1–7.
[24] S. Liu, F. Yang, W. Ding, X. Wang, and J. Song, "Two-dimensional structured-compressed-sensing-based NBI cancelation exploiting spatial and temporal correlations in MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 65, no. 11, pp. 9020–9028, Nov. 2016.
[25] Z. Gao, L. Dai, and Z. Wang, "Channel estimation for mmWave massive MIMO based access and backhaul in ultra-dense network," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2016, pp. 1–6.
[26] P. Schniter and A. Sayeed, "Channel estimation and precoder design for millimeter-wave communications: The sparse way," in *Proc. 48th Asilomar Conf. Signals, Syst., Comput.*, Nov. 2014, pp. 273–277.
[27] Z. Gao, C. Hu, L. Dai, and Z. Wang, "Channel estimation for millimeter-wave massive MIMO with hybrid precoding over frequency-selective fading channels," *IEEE Commun. Lett.*, vol. 20, no. 6, pp. 1259–1262, Jun. 2016.
[28] A. Alkhateeb and R. W. Heath, Jr., "Frequency selective hybrid precoding for limited feedback millimeter wave systems," *IEEE Trans. Commun.*, vol. 64, no. 5, pp. 1801–1818, May 2016.
[29] S. Sun, G. R. MacCartney, Jr., and T. S. Rappaport, "A novel millimeter-wave channel simulator and applications for 5G wireless communications," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2017, pp. 1–7.
[30] M. K. Samimi and T. S. Rappaport, "3-D millimeter-wave statistical channel model for 5G wireless system design," *IEEE Trans. Microw. Theory Techn.*, vol. 64, no. 7, pp. 2207–2225, Jul. 2016.
[31] S. G. Larew, T. A. Thomas, M. Cudak, and A. Ghosh, "Air interface design and ray tracing study for 5G millimeter wave communications," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Dec. 2013, pp. 117–122.
[32] T. S. Rappaport, G. R. Maccartney, M. K. Samimi, and S. Sun, "Wideband millimeter-wave propagation measurements and channel models for future wireless communication system design," *IEEE Trans. Commun.*, vol. 63, no. 9, pp. 3029–3056, Sep. 2015.
[33] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, "Simultaneous sparse approximation via greedy pursuit," in *Proc. IEEE Conf. Acous., Speech, Signal Process. (ICASSP)*, Mar. 2005, pp. v/721–v/724.
[34] S. M. Kay, *Fundamentals of Statistical Processing: Estimation Theory*, vol. 1. Upper Saddle River, NJ, USA: Prentice-Hall, 1993.
[35] K. Venugopal, N. González-Prelcic, and R. W. Heath, Jr., "Optimality of frequency flat precoding in frequency selective millimeter wave channels," *IEEE Wireless Commun. Lett.*, vol. 6, no. 3, pp. 330–333, Jun. 2017.

Javier Rodríguez-Fernández received the B.Sc.Eng. degree in telecommunication technologies and the M.Sc. degree in telecommunication engineering from The University of Vigo, Spain, in 2014 and 2016, respectively, where he is currently pursuing the Ph.D. degree in information and communication technologies. His research interests lie in signal processing for communications.

Nuria González-Prelcic received the Ph.D. in electrical engineering in 1998. She has held visiting positions with the University of New Mexico in 2012 and The University of Texas at Austin from 2014 to 2017. She has been the Founding Director of the Atlantic Research Center for Information and Communication Technologies, University of Vigo, Spain, where she is currently an Associate Professor with the Signal Theory and Communications Department. She has authored around 80 papers in the topic of signal processing for communications, including a tutorial on millimeter wave communication published in the IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING. She has also served as a Guest Editor for the special issue of this journal on signal processing for mm-wave wireless communications. Her main research interests include signal processing theory and signal processing for wireless communications, such as filter banks, compressive sampling and estimation, multicarrier modulation, channel estimation, and MIMO processing for millimeter wave communications, including V2X at millimeter wave. She is an Editor of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS. She has co-organized several special sessions on mm-wave communications at SPAWC 2015, Asilomar 2016, and CAMSAP 2017.

Kiran Venugopal (S'15) received the B.Tech. degree in electronics and communication engineering from the National Institute of Technology Calicut, India, in 2011, the M.E. degree in telecommunication engineering from the Indian Institute of Science, Bengaluru, India, in 2013, and the Ph.D. degree in electrical and computer engineering from the University of Texas at Austin, TX, USA, in 2017. From 2013 to 2014, he was a Systems Design Staff Engineer with Broadcom Communication Technologies Pvt. Ltd., Bengaluru. Since 2017, he has been a Senior Engineer at Qualcomm R&D, Qualcomm Flarion Technologies, Inc., NJ, USA, where he has been involved in 5G design and standardization. His research interests include wireless communication, signal processing, stochastic geometry, and the analysis and physical layer development of mmWave systems. He received the Prof. S. V. C. Aiya Medal for the best M.E. (Telecommunications) student in 2013, from the Indian Institute of Science, and the Er. M. L. Bapna Gold Medal in 2011, from the National Institute of Technology, Calicut, for topping the under graduate programs in the institute.

Robert W. Heath, Jr., (S'96–M'01–SM'06–F'11) received the B.S. and M.S. degrees from the University of Virginia, Charlottesville, VA, USA, in 1996 and 1997, respectively, and the Ph.D. degree from Stanford University, Stanford, CA, USA, in 2002, all in electrical engineering. From 1998 to 2001, he was a Senior Member of the Technical Staff and then a Senior Consultant at Iospan Wireless Inc., San Jose, CA, USA, where he was involved in the design and implementation of the physical and link layers of the first commercial MIMO-OFDM communication system. Since 2002, he has been with the Department of Electrical and Computer Engineering, The University of Texas at Austin, where he is currently a Cullen Trust for Higher Education Endowed Professor, and a Member of the Wireless Networking and Communications Group. He is also the President and the CEO of MIMO Wireless Inc. He has authored *Introduction to Wireless Digital Communication* (Prentice Hall, 2017) and *Digital Wireless Communication: Physical Layer Exploration Lab Using the NI USRP* (National Technology and Science Press, 2012), and co-authored *Millimeter Wave Wireless Communications* (Prentice Hall, 2014).

Dr. Heath has been a co-author of sixteen award winning conference and journal papers, including the 2010 and 2013 *EURASIP Journal on Wireless Communications and Networking* Best Paper Award, the 2012 *Signal Processing Magazine* Best Paper Award, the 2013 Signal Processing Society Best Paper Award, the 2014 *EURASIP Journal on Advances in Signal Processing* Best Paper Award, the 2014 *Journal of Communications and Networks* Best Paper Award, the 2016 IEEE Communications Society Fred W. Ellersick Prize, the 2016 IEEE Communications and Information Theory Societies Joint Paper Award, and the 2017 Marconi Prize Paper Award. He received the 2017 EURASIP Technical Achievement Award. He was a Distinguished Lecturer of the IEEE Signal Processing Society and is an ISI Highly Cited Researcher. He is also an Elected Member of the Board of Governors for the IEEE Signal Processing Society, a licensed amateur radio operator, a private pilot, and a registered professional engineer in Texas.