Wideband MIMO Channel Estimation for Hybrid Beamforming Millimeter Wave Systems via Random Spatial Sampling

Evangelos Vlachos, *Member, IEEE*, George C. Alexandropoulos, *Senior Member, IEEE*, and John Thompson, *Fellow, IEEE*

***Abstract*—Millimeter Wave (mmWave) massive Multiple Input Multiple Output (MIMO) systems realizing directive communication over large bandwidths via Hybrid analog and digital BeamForming (HBF) require reliable estimation of the wideband wireless channel. However, the hardware limitations with HBF architectures in conjunction with the short coherence time inherit in mmWave communication render this estimation a challenging task. In this paper, we develop a novel wideband channel estimation framework for mmWave massive MIMO systems with HBF reception. The proposed framework jointly exploits the low rank property and the available angular information to provide more accurate channel recovery, especially for short beam training intervals. We introduce a novel analog combining architecture that includes a random spatial sampling structure placed before the input of the analog received signals to the digital component of the HBF receiver. This architecture supports the proposed matrix-completion-based estimation approach in providing the sampling set of measurements for recovering the unknown channel matrix. The performance improvement of the proposed approach over representative state-of-the-art techniques is demonstrated through numerous computer simulation results.**

***Index Terms*—Wideband channel estimation, angular information, random spatial sampling, matrix completion, massive MIMO, hybrid beamforming, millimeter wave communications.**

I. Introduction

THE millimeter Wave (mmWave) frequency band offers large bandwidths for wireless communication, and thus, has been considered as a promising enabler for the highly demanding data rate requirements of fifth Generation (5G), and beyond, broadband wireless networks [1]. Although communication signals at mmWave systems are expected to experience severe pathloss and atmospheric attenuation, their mm-level wavelengths enable many antenna elements to be packed in small-sized terminals, which allows for highly directional beamforming. The IEEE 802.11ad standard adopts analog beamforming [2] with single-stream communication, where many antenna elements are connected to one Radio Frequency (RF) chain via an analog network that is usually comprised of phase shifters. To provide higher data rates, the 5G New Radio (NR) technology [3] will support Hybrid BeamForming (HBF) [4] with up to four spatial streams for communication. This technology capitalizes on both analog and digital signal processing to enable a large number of antenna elements to be connected to a much smaller number of RF chains.

Near-optimal HBF performance in mmWave massive Multiple Input Multiple Output (MIMO) systems necessitates reliable Channel State Information (CSI) knowledge. This is however very challenging to acquire in practice due to the very large numbers of transceiver antenna elements and the high channel variability [5]. Several approaches requiring receiver feedback have been proposed for designing BF vectors suitable for CSI estimation [6], [7]. On the other side, static dictionaries or beam training techniques without receiver feedback have also been adopted for beam codebook designs [8]–[12]. In most of these studies, CSI estimation has been treated as a compressive sensing problem [13], where the Orthogonal Matching Pursuit (OMP) algorithm [14] has been usually adopted to recover the sparse channel gain vector. However, the performance of these channel estimation techniques is usually limited by the codebook design, since beam dictionaries suffer from power leakage due to the discretization of the Angle of Arrival (AoA) and Angle of Departure (AoD). Very recently in [15]–[17], the sparsity and low rank properties of certain wireless channels were jointly exploited for efficient CSI estimation. Targeting narrowband mmWave MIMO channel estimation with HBF transceivers, a two-stage procedure (one stage per channel property) was introduced in [16]. For the same narrowband channels and considering transceivers equipped with simple switches, [17] presented a matrix completion algorithm leveraging jointly from both properties. This algorithm was shown to outperform the techniques of [8], [16] in terms of estimation performance, requiring relatively short channel sounding intervals. Wideband mmWave MIMO channels with frequency selectivity were recently considered in [11], where a CSI estimation technique exploiting channel's sparsity in both time and frequency domains was presented.

Manuscript received February 15, 2019; revised June 26, 2019; accepted July 17, 2019. Date of publication August 26, 2019; date of current version September 20, 2019. This work was supported by Engineering and Physical Sciences Research Council under Grant EP/P000703/1. The guest editor coordinating the review of this paper and approving it for publication was Prof. Feifei Gao. *(Corresponding author: Evangelos Vlachos.)*

E. Vlachos and J. Thompson are with the Institute for Digital Communications, University of Edinburgh, Edinburgh EH9 3JL, U.K. (e-mail: e.vlachos@ed.ac.uk; j.s.thompson@ed.ac.uk).

G. C. Alexandropoulos is with the Department of Informatics and Telecommunications, National and Kapodistrian University of Athens, Athens 15784, Greece (e-mail: alexandg@di.uoa.gr).

Digital Object Identifier 10.1109/JSTSP.2019.2937633

A. Novelty and Contributions

In this paper, we consider the uplink of wideband mmWave massive MIMO systems with HBF reception and focus on accurate pilot-assisted channel estimation with short training lengths and low computational complexity. For the estimation of the unknown channel, we formulate a multi-objective optimization problem that capitalizes on the sparsity of the channel matrix in the beamspace domain, as well as on the low-rank property of the received signal matrix during channel sounding. The proposed channel estimation algorithm is based on the Alternating Direction Method of Multipliers (ADMM) [18]. In particular, the main contributions of this work are summarized as follows:

- We introduce a novel analog combining architecture including a *random spatial sampling* structure, which comprises of a random selection step before the input of the analog received signals to the digital component of the HBF receiver. This architecture is used for measurement collection during channel sounding, and is radically different from the one proposed in [17], where an analog combiner with switches was considered to estimate the signals at the received antennas one-by-one. The key characteristic of the proposed architecture is the random selection of a subset of the available analog receive beams per channel sounding interval, which in conjunction with the proposed CSI estimation technique yields improved performance with relatively short training lengths.
- We show that when the wideband mmWave massive MIMO channel matrix is of a low rank, the same holds for the matrix obtained at the HBF receiver after the analog processing of training signals with multiple receive beams. Capitalizing on this finding, we formulate an optimization problem for CSI estimation that jointly exploits the low-rank property of the received training signal matrix and the sparsity of the channel matrix in the beamspace domain. The proposed optimization problem formulation, which aims at finding the directions and gains of the wideband channel paths, is different from that in [17]. The differences refer to the mathematical formulations, the measurement requirements needed for them and the applicability of the approach in wideband mmWave massive MIMO channels; [17] is based on the low-rank property of narrowband mmWave MIMO channels.
- We extend the proposed CSI estimation algorithm to incorporate prior knowledge of the angles of the propagation channel paths. This is achieved by imposing a specific structure on the beamspace representation of the wideband MIMO channel matrix using a hard thresholding operation that does not add significant complexity to the designed technique. It is shown via representative simulation results that the proposed angle-aware CSI estimation is robust in regimes of high noise exhibiting improved performance in cases of reduced number of spatial training measurements, i.e., when the number of RF chains in the HBF receiver is small.
- Apart from the proposed iterative ADMM-based optimization algorithms for solving the targeted wideband mmWave massive MIMO channel estimation problem, we design their low-complexity approximations that exhibit very similar performance, but with significantly reduced computational complexity requirements.

Our extensive simulation results showcase that the proposed algorithms exhibit improved performance in terms of Mean Squared Error (MSE) for channel estimation with a short beam training length and under high noise conditions, when compared with other state-of-the-art wideband mmWave massive MIMO channel estimation techniques.

Notations: A summary of the notation used throughout this manuscript can be found in Table I.

TABLE I
THE NOTATIONS OF THIS PAPER

$a, \mathbf{a}$ and $\mathbf{A}$	Scalar, vector, and matrix
$j \triangleq \sqrt{-1}$	The imaginary unit
$\mathbf{A}^T$ and $\mathbf{A}^H$	Matrix transpose and Hermitian transpose
$\mathbf{A}^{-1}$ and $\mathbf{A}^\dagger$	Matrix inverse and pseudo-inverse
$[\mathbf{A}]_{i,j}$	Matrix element at the i-th row and j-th column
$[\mathbf{a}]_i$	The i-th vector element
$\mathbf{A}$ and $\hat{\mathbf{A}}$	Actual and estimated matrix
$\mathbf{I}_N$	$N \times N$ identity matrix
$\mathbf{0}_{N\times K}$	$N \times K$ matrix with zeros
$\mathbf{I}_{N\times K}$	Column concatenated matrix $[\mathbf{I}_N\ \mathbf{0}_{N\times K}]$
$\mathbf{\Omega}$	Matrix containing 0's and 1's
$\|\cdot\|_*$, $\|\cdot\|_F$	Matrix nuclear and Frobenius norms
$\times$	Scalar multiplication
$\circ$	Element-wise (Hadamard) matrix product
$\otimes$	Kronecker product
$\mathrm{vec}(\mathbf{A})$	Vectorization of $\mathbf{A}$
$\mathrm{unvec}(\mathbf{A})$	Inverse operation of $\mathrm{vec}(\cdot)$
$\mathrm{diag}(\mathbf{a})$	Diagonal matrix with $\mathbf{a}$ on the main diagonal
$\mathrm{toeplitz}(\mathbf{a})$	Toeplitz matrix with $\mathbf{a}$ on the first row
SVT_t	Singular Value Thresholding [19] with threshold t
$\mathcal{L}(\mathbf{X},\mathbf{Y},\mathbf{V})$	Lagrangian with primal $(\mathbf{X},\mathbf{Y})$ and dual $(\mathbf{V})$ variables
$\mathcal{S}_{\mathbf{\Omega}}(\mathbf{A})$	Hadamard operator ($\mathcal{S}_{\mathbf{\Omega}}(\mathbf{A}) \triangleq \mathbf{\Omega} \circ \mathbf{A}$) which samples the entries of $\mathbf{A}$ based on $\mathbf{\Omega}$

II. Channel Model

We consider a point-to-point $N_{\mathrm{R}} \times N_{\mathrm{T}}$ massive MIMO communication system operating over wideband mmWave channels. Similar to [20], we assume that each of the N_{T} antenna elements of the Transmitter (TX) is attached to a dedicated RF chain, while the N_{R} Receiver (RX) antenna elements are connected (all or in small groups) to $M_{\mathrm{R}} < N_{\mathrm{R}}$ RF chains. Due to this hardware configuration, the TX is capable of digitally precoding up to N_{T} independent signals, each provided from one dedicated RF chain. On the other hand, we assume that RX is equipped with any of the available HBF architectures [4] supporting both analog and digital combining. As described next, the proposed framework applies also to HBF transmission, but we leave the full details of this extension for future work. From an information theoretic point of view, the considered mmWave MIMO system can realize a wireless communication link comprising of $d_{\mathrm{s}} \leq \min(N_{\mathrm{T}}, M_{\mathrm{R}})$ independent information data streams.

A. Frequency-Selective MIMO Channel Model

Let us utilize the geometric representation for the frequency-selective mmWave MIMO channel, according to which the channel has L delay taps with $\mathbf{H}(\ell) \in \mathbb{C}^{N_{\mathrm{R}}\times N_{\mathrm{T}}}$ ($\ell = 0, 1, \ldots,$

$L-1$) denoting the MIMO channel gain matrix for each ℓ-th channel tap. For the ℓ-th delay tap, $\mathbf{H}(\ell)$ can be mathematically expressed as

$$\mathbf{H}(\ell) \triangleq \sqrt{\frac{N_{\mathrm{T}}N_{\mathrm{R}}}{N_p}} \sum_{k=1}^{N_p} \alpha_k p\left(\ell T_s - \tau_k\right) \mathbf{a}_{\mathrm{R}}(\phi_k)\mathbf{a}_{\mathrm{T}}^H(\theta_k), \quad (1)$$

where N_p denotes the number of propagation paths per channel delay tap ℓ, and α_k with $k = 1, 2, \ldots, N_p$ is the gain of the k-th channel path drawn from the complex Gaussian distribution $\mathcal{CN}(0, 1/2)$. In addition, $p(\tau_k)$ denotes the band-limited pulse shaping filter response at the discrete time τ_k. The vectors $\mathbf{a}_{\mathrm{T}}(\phi_k) \in \mathbb{C}^{N_{\mathrm{T}}}$ and $\mathbf{a}_{\mathrm{R}}(\theta_k) \in \mathbb{C}^{N_{\mathrm{R}}}$ represent the TX and RX normalized array response vectors, respectively, which are expressed as described in [5, Sec. II.C] for uniform linear arrays. Scalars ϕ_k and θ_k are the physical AoA and AoD, respectively, which are commonly generated according to the Laplace distribution [21].

In a similar way to the case of narrowband mmWave MIMO channel estimation, techniques that capitalize on the low-rank property of the channel (e.g., [16], [17]) can be employed. Before proceeding with the proposed CSI estimation framework, we first investigate the rank of the considered wideband mmWave MIMO, which will witness the relevance of the latter techniques to the targeted estimation problem. Specifically, the rank of each ℓ-th delay tap channel $\mathbf{H}(\ell)$ given by (1) can be obtained as

$$\mathrm{rank}(\mathbf{H}(\ell)) = \mathrm{rank}\left(\sum_{k=1}^{N_p} \alpha_k p(\ell T_s - \tau_k)\mathbf{a}_{\mathrm{R}}(\phi_k)\mathbf{a}_{\mathrm{T}}^H(\theta_k)\right) \quad (2)$$

$$\leq \sum_{k=1}^{N_p} \mathrm{rank}\left(\mathbf{a}_{\mathrm{R}}(\phi_k)\mathbf{a}_{\mathrm{T}}^H(\theta_k)\right) = N_p, \quad (3)$$

where the equality between the expressions in (2) and (3) holds when the following steering matrices

$$\mathbf{A}_{\mathrm{T}} \triangleq \left[\mathbf{a}_{\mathrm{T}}(\theta_1)\, \mathbf{a}_{\mathrm{T}}(\theta_2)\, \cdots\, \mathbf{a}_{\mathrm{T}}(\theta_{N_p})\right], \quad (4)$$

and

$$\mathbf{A}_{\mathrm{R}} \triangleq \left[\mathbf{a}_{\mathrm{R}}(\phi_1)\, \mathbf{a}_{\mathrm{R}}(\phi_2)\, \cdots\, \mathbf{a}_{\mathrm{R}}(\phi_{N_p})\right] \quad (5)$$

are orthogonal. Hence, all channel matrices $\mathbf{H}(\ell)$ for $\ell = 0, 1, \ldots, L-1$ have the same rank, which proves that the rank of the concatenated channel matrix

$$\bar{\mathbf{H}} \triangleq [\mathbf{H}(0)\, \mathbf{H}(1)\, \cdots\, \mathbf{H}(L-1)] \in \mathbb{C}^{N_{\mathrm{R}} \times LN_{\mathrm{T}}}$$

will also be upper bounded by:

$$\mathrm{rank}(\bar{\mathbf{H}}) \leq N_p. \quad (6)$$

It is already well known (see, e.g., [22]–[24]) that $N_p \ll N_{\mathrm{R}}$, which reveals that $\bar{\mathbf{H}}$ is of a low rank depending directly on the geometry of the mmWave channel propagation paths.

B. Beamspace Decomposition

An alternative representation for $\mathbf{H}(\ell)$, that will be exploited later in the proposed channel estimation algorithm, is based on the beamspace channel of [25], which is defined as

$$\mathbf{H}(\ell) = \mathbf{D}_{\mathrm{R}}\mathbf{Z}(\ell)\mathbf{D}_{\mathrm{T}}^H, \quad (7)$$

where $\mathbf{D}_{\mathrm{R}} \in \mathbb{C}^{N_{\mathrm{R}} \times N_{\mathrm{R}}}$ and $\mathbf{D}_{\mathrm{T}} \in \mathbb{C}^{N_{\mathrm{T}} \times N_{\mathrm{T}}}$ are unitary matrices based on the Discrete Fourier Transform (DFT), and $\mathbf{Z}(\ell) \in \mathbb{C}^{N_{\mathrm{R}} \times N_{\mathrm{T}}}$ includes the virtual channel gains of $\mathbf{H}_\ell$. Motivated by the low rank property of this matrix, we further assume that $\mathbf{Z}(\ell)$ $\forall \ell$ contains only few virtual channel gains with high amplitude, i.e., it is a sparse matrix. The sparsity level of $\mathbf{Z}(\ell)$ depends on the angular discretization in the beamspace representation given by (7).

Low-rank-based mmWave MIMO channel estimation leads to better approximation of the channel matrix, at the expense of higher number of training blocks [26], [27]. On the opposite side, CSI estimation approaches capitalizing on channel's spatial sparsity may approximate the channel matrix faster, but with lower accuracy due to the angular discretization errors [24]. In [17], an estimation framework for narrowband mmWave MIMO channels was introduced leveraging jointly the latter two properties. Although the extension of [17]'s approach is feasible for wideband channels it may introduce increased design complexity, given that the number of the receiving antennas N_{R} becomes very large. To overcome this limitation, in this work we introduce a novel architecture for the analog part of HBF reception including a *random spatial sampling* structure. Owing to this new structure, the number of RF chains is reduced to M_{R} which is usually much lower than the number of the receiving antennas N_{R}, thus, the hardware design complexity is significantly reduced.

III. Proposed Analog Combining Architecture

In this section, we first describe the sounding procedure for the considered wideband MIMO channel, and then present the proposed analog combining architecture for HBF RXs with large number of antenna elements.

A. Channel Sounding Procedure

We assume frame-by-frame communication, where the wireless channel remains constant during each frame, but might change independently from one frame to another. Every frame consists of T blocks dedicated for channel estimation, whereas the rest of the frame is used for data communication. Naturally, a large T yields improved channel estimation, but leaves less frame length for actual data communication. To estimate the intended wideband mmWave MIMO channel, the N_{T}-antenna TX utilizes the $N_{\mathrm{T}} \times 1$ training symbols' vector $\mathbf{s}[t]$ for each block t with $t = 1, 2, \ldots, T$. Ignoring for the moment for the clarity of the exposition the impact of the Additive White Gaussian Noise (AWGN), the N_{R}-dimensional received training signal can be expressed as:

$$\tilde{\mathbf{y}}[t] = \sum_{\ell=0}^{L-1} \mathbf{H}(\ell)\mathbf{s}[t-\ell], \quad (8)$$

which represents the convolution of the L tap delay channel matrices and the L training vectors $\mathbf{s}[t-\ell] \in \mathbb{C}^{N_{\mathrm{T}} \times 1}$. Equivalently, (8) can be re-expressed as:

$$\tilde{\mathbf{y}}[t] = \sum_{\ell=0}^{L-1} \sum_{k=1}^{N_{\mathrm{T}}} \mathbf{h}_k(\ell) s_k[t-\ell], \quad (9)$$

where $\mathbf{h}_k(\ell)$ is the k-th column of $\mathbf{H}(\ell)$ and $s_k[t]$ denotes the k-th element of $\mathbf{s}[t]$. Interchanging the summation order in (9), we can express the inner convolution sum in terms of $L \times T$ Toeplitz matrix. Particularly, by introducing the $L \times T$ Toeplitz matrix $\tilde{\mathbf{\Psi}}_k$ with its (ℓ, t)-th element given by:

$$[\tilde{\mathbf{\Psi}}_k]_{\ell,t} = s_k[t - L - \ell + 2], \tag{10}$$

where $\ell = 0, 1, \ldots, L-1$ and $k = 1, 2, \ldots, N_{\mathrm{T}}$, (9) can be re-written as:

$$\tilde{\mathbf{Y}} \triangleq \sum_{k=1}^{N_{\mathrm{T}}} \tilde{\mathbf{H}}_k \tilde{\mathbf{\Psi}}_k, \tag{11}$$

where $\tilde{\mathbf{Y}} \in \mathbb{C}^{N_{\mathrm{R}} \times T}$ and $\tilde{\mathbf{H}}_k \triangleq [\mathbf{h}_k(0) \cdots \mathbf{h}_k(L-1)] \in \mathbb{C}^{N_{\mathrm{R}} \times L}$. Reorganizing the structure of $\tilde{\mathbf{\Psi}}$ and $\tilde{\mathbf{H}}$ so as to be grouped over the transmitting antennas (see Appendix B), (11) can be equivalently expressed as:

$$\tilde{\mathbf{Y}} = \bar{\mathbf{H}} \bar{\mathbf{\Psi}}, \tag{12}$$

where $\bar{\mathbf{H}} \triangleq [\mathbf{H}(0) \cdots \mathbf{H}(L-1)] \in \mathbb{C}^{N_{\mathrm{R}} \times L N_{\mathrm{T}}}$ and $\bar{\mathbf{\Psi}} \triangleq [\mathbf{\Psi}^T(0) \cdots \mathbf{\Psi}^T(L-1)]^T \in \mathbb{C}^{L N_{\mathrm{T}} \times T}$. Moreover, using the beamspace decomposition of each delay's path matrix we have that

$$\bar{\mathbf{H}} = \mathbf{D}_{\mathrm{R}} \bar{\mathbf{Z}} (\mathbf{I}_L \otimes \mathbf{D}_{\mathrm{T}}^H), \tag{13}$$

where $\bar{\mathbf{Z}} \triangleq [\mathbf{Z}(0) \cdots \mathbf{Z}(L-1)] \in \mathbb{C}^{N_{\mathrm{R}} \times L N_{\mathrm{T}}}$. The decomposition of (13) can be seen as an *equivalent* to the beamspace decomposition of (7) for the case of the wideband channel matrix $\bar{\mathbf{H}}$. Putting all above together, the matrix including the received training symbols is given by

$$\tilde{\mathbf{Y}} = \mathbf{D}_{\mathrm{R}} \bar{\mathbf{Z}} (\mathbf{I}_L \otimes \mathbf{D}_{\mathrm{T}}^H) \bar{\mathbf{\Psi}}. \tag{14}$$

The difference of (14) with respect to (13) is the right hand matrix $\bar{\mathbf{\Psi}}$, which contains the training symbols. The latter expression will be used in our approach in order to express the received training data with respect to the concatenated virtual channel gains $\bar{\mathbf{Z}}$.

B. Analog Combining With Random Spatial Sampling

1) Motivation: Before introducing the proposed analog combining architecture, we provide some motivation arguments. To receive the training symbols intended for channel estimation, the HBF RX utilizes its $N_{\mathrm{R}} \times M_{\mathrm{R}}$ analog combiner denoted by $\mathbf{W}_{\mathrm{RF}}$ and belonging to a predefined beam codebook $\mathcal{W}^{N_{\mathrm{R}} \times M_{\mathrm{R}}}$. Its block diagram for standard HBF RX architectures (see, e.g., [4]) is sketched in Fig. 1. A network of Phase Shifters (PSs) is usually considered for analog combining, hence all available $\mathbf{W}_{\mathrm{RF}}$ combiners have unit magnitude elements. For the combiner's phases, the following generic quantized set is usually considered:

$$\mathcal{W}^{N_{\mathrm{R}} \times M_{\mathrm{R}}} = \left\{ 0, \frac{2\pi}{2^{N_{\mathrm{Q}}}}, \ldots, \frac{(2^{N_{\mathrm{Q}}} - 1)2\pi}{2^{N_{\mathrm{Q}}}} \right\}, \tag{15}$$

with N_Q being the quantization resolution. After applying an analog combiner from the available codebook to the received training symbols' matrix $\tilde{\mathbf{Y}}$ in (13), the baseband received signal at the M_{R} inputs of the RX's RF chains is given by $\mathbf{W}_{\mathrm{RF}}^H \tilde{\mathbf{Y}} \in \mathbb{C}^{M_{\mathrm{R}} \times T}$.

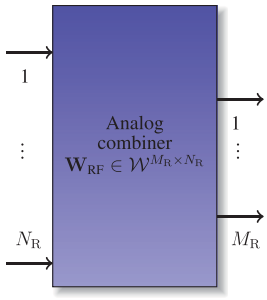

Fig. 1. Block diagram of the conventional analog combiner [4] used for wideband mmWave MIMO channel estimation with HBF reception. The combiner connects via phase shifters its N_{R} inputs from the respective RX antennas to the M_{R} receive RF chains.

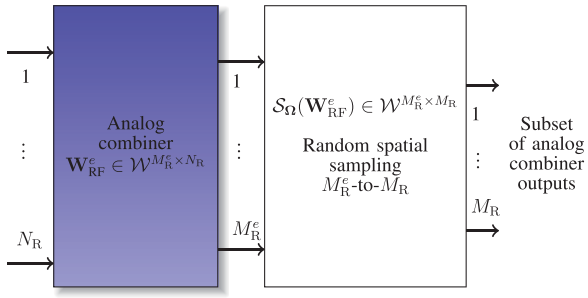

Fig. 2. Block diagram of the proposed analog combining architecture for wideband mmWave MIMO channel estimation with HBF reception. The architecture includes an extended analog combiner having as inputs the N_{R} analog signals from the respective RX antennas. Those signals are fed via a network of phase shifters to the M_{R}^e outputs of the combiner, M_{R} out of which are selected from the subsequent random spatial sampling unit.

Usually the number of receive RF chains is much smaller than the number of receiving antennas, i.e., $M_{\mathrm{R}} \ll N_{\mathrm{R}}$, to provide low implementation complexity and reasonable power consumption for mmWave HBF reception [4]. However, this reduction in the RF chain hardware introduces a loss of information compared to the fully digital RX case, since the received signals lie in a smaller vector space. This is particularly true for the achievable Average Spectral Efficiency (ASE), which is related to the orthogonality of the analog combiner $\mathbf{W}_{\mathrm{RF}}$. Specifically, assuming equal allocation of the unit transmit power to the N_{T} transmit antennas and AWGN with variance σ_n^2, the achievable ASE with conventional HBF reception can be upper bounded as (see Appendix C):

$$C_{\mathrm{HBF}} = \log_2 \det \left(\mathbf{I}_{M_{\mathrm{R}}} + \frac{1}{\sigma_n^2 N_{\mathrm{T}}} \mathbf{W}_{\mathrm{RF}}^H \mathcal{Y} \mathcal{Y}^H \mathbf{W}_{\mathrm{RF}} \right) \tag{16}$$

$$\leq \log_2 \det \left(\mathbf{I}_{N_{\mathrm{R}}} + \frac{1}{\sigma_n^2 N_{\mathrm{T}}} \mathcal{Y} \mathcal{Y}^H \right), \tag{17}$$

where $\mathcal{Y} \in \mathbb{C}^{N_{\mathrm{R}} \times d_s}$ represents the received information bearing data matrix having a similar expression to (12). It is noted that the inequality (17) is satisfied with equality when $\mathbf{W}_{\mathrm{RF}} \mathbf{W}_{\mathrm{RF}}^H = \mathbf{I}_{N_{\mathrm{R}}}$.

2) Architectural Components: The block diagram of our novel analog combining architecture comprising of an extended analog combiner and a random spatial sampling unit is illustrated in Fig. 2. As shown, the inputs of this architecture are the M_{R}^e analog signals received at the respective RX antenna elements,

TABLE II
COMPARISON OF THE DIFFERENT COMPONENTS BETWEEN THE CONVENTIONAL AND THE PROPOSED ANALOG COMBINER ARCHITECTURES

	Conventional	Proposed
Num. of PSs	$M_{\mathrm{R}}N_{\mathrm{R}}$	$M_{\mathrm{R}}^{e}N_{\mathrm{R}}$
Num. of LNAs	$(M_{\mathrm{R}}+1)N_{\mathrm{R}}$	$(M_{\mathrm{R}}^{e}+1)N_{\mathrm{R}}$
Total Power consumption (P)	$M_{\mathrm{R}}N_{\mathrm{R}}P_{\mathrm{PS}}+$ $N_{\mathrm{R}}(M_{\mathrm{R}}+1)P_{\mathrm{LNA}}$	$M_{\mathrm{R}}^{e}N_{\mathrm{R}}P_{\mathrm{PS}}+M_{\mathrm{R}}P_{\mathrm{SW}}$ $+N_{\mathrm{R}}(M_{\mathrm{R}}^{e}+1)P_{\mathrm{LNA}}$

TABLE III
SIMULATION PARAMETERS FOR THE PERFORMANCE EVALUATION

Carrier frequency	30 GHz
Antenna array characteristics	Uniform Linear Array (ULA)
Length of zero prefix	$L-1$ symbols
Parameters for Algorithms 1, 2	$\rho=\sqrt{\lambda_{\min}\frac{\tau_R+\tau_Z}{2}}$ [37] $\tau_R=1/\|\mathbf{R}_\Omega\|_F^2$ $\tau_Z=1/2\|\bar{\mathbf{Z}}\|_F^2$ $I_{\max}=100$
Maximum Iterations for OMP [14]	$I_{\max}=100$
Maximum Iterations for VAMP [38]	$I_{\max}=100$
Step size for SVT [16]	0.1

and its outputs are fed to the M_{R} receive RF chains. In mathematical terms, the conventional analog combiner $\mathbf{W}_{\mathrm{RF}}\in\mathcal{W}^{N_{\mathrm{R}}\times M_{\mathrm{R}}}$ is replaced in the proposed architecture by the extended analog combiner $\mathbf{W}_{\mathrm{RF}}^{e}\in\mathcal{W}^{N_{\mathrm{R}}\times M_{\mathrm{R}}^{e}}$ (having also unit magnitude elements) which is followed by a random spatial sampling structure. The latter structure performs sampling of the M_{R}^{e} outputs of the extended combiner $\mathbf{W}_{\mathrm{RF}}^{e}$ forwarding only M_{R} of these signals to the receive RF chains, as shown in Fig. 2. In mathematical terms, the functionality of the random spatial sampler can be represented by the operator $\mathcal{S}_\Omega(\mathbf{W}_{\mathrm{RF}}^{e})\in\mathcal{W}^{M_{\mathrm{R}}^{e}\times M_{\mathrm{R}}}$, which randomly selects M_{R} out of M_{R}^{e} columns of the extended analog combiner $\mathbf{W}_{\mathrm{RF}}^{e}$. Following the mode of operation of the proposed analog combiner, the achievable spectral efficiency is given by

$$C_{\mathrm{p}}=\log_2\det\left(\mathbf{I}_{M_{\mathrm{R}}}+\frac{1}{\sigma_n^2N_{\mathrm{T}}}\mathcal{S}_\Omega(\mathbf{W}_{\mathrm{RF}}^{e})^H\mathcal{Y}\mathcal{Y}^H\mathcal{S}_\Omega(\mathbf{W}_{\mathrm{RF}}^{e})\right). \tag{18}$$

Since the sampling via the operator $\mathcal{S}_\Omega$ is performed to a larger vector space, generated by the columns of $\mathbf{W}_{\mathrm{RF}}^{e}$, it can be shown that $\mathcal{S}_\Omega(\mathbf{W}_{\mathrm{RF}}^{e})(\mathcal{S}_\Omega(\mathbf{W}_{\mathrm{RF}}^{e}))^H$ has better orthogonality properties than $\mathbf{W}_{\mathrm{RF}}\mathbf{W}_{\mathrm{RF}}^{H}$. This means that the achievable rate of the proposed design is expected to be higher than the conventional HBF architecture (c.f., Section VI.A).

The proposed extended analog combiner requires larger number of PSs compared to the conventional combiner in order to generate $M_{\mathrm{R}}^{e}>M_{\mathrm{R}}$ outputs. We argue, however, that those additional hardware components do not contribute significantly to RX's hardware complexity or power consumption. In Table II, we compare the conventional and the proposed analog parts of the HBF RX in terms of the total power consumption P. We consider the power consumption required by three main components: (i) P_{PS} for each PS, (ii) P_{LNA} for each Low Noise Amplifier (LNA), and (iii) P_{SW} for each SWitch (SW). Specifically, the increase in power consumption required by the proposed architecture is $(M_{\mathrm{R}}^{e}-M_{\mathrm{R}})(P_{\mathrm{LNA}}+P_{\mathrm{PS}})+M_{\mathrm{R}}P_{\mathrm{SW}}$. The impact of the increased hardware components to the power consumption and the energy efficiency are thoroughly investigated in Section VI.

3) Random Spatial Sampling: Capitalizing on expression (13) for the received training signal and the functionality of the extended analog combiner, the output of the latter structure in the proposed HBF RX architecture can be expressed by the following $M_{\mathrm{R}}^{e}\times T$ complex-valued matrix:

$$\mathbf{R}\triangleq(\mathbf{W}_{\mathrm{RF}}^{e})^H(\tilde{\mathbf{Y}}+\tilde{\mathbf{N}}) \tag{19}$$

$$=(\mathbf{W}_{\mathrm{RF}}^{e})^H(\mathbf{D}_{\mathrm{R}}\bar{\mathbf{Z}}\mathbf{B}+\tilde{\mathbf{N}}) \tag{20}$$

$$=\mathbf{A}\bar{\mathbf{Z}}\mathbf{B}+(\mathbf{W}_{\mathrm{RF}}^{e})^H\tilde{\mathbf{N}} \tag{21}$$

with $\mathbf{A}\triangleq(\mathbf{W}_{\mathrm{RF}}^{e})^H\mathbf{D}_{\mathrm{R}}\in\mathbb{C}^{M_{\mathrm{R}}^{e}\times N_{\mathrm{R}}}$ and $\mathbf{B}\triangleq(\mathbf{I}_L\otimes\mathbf{D}_{\mathrm{T}}^{H})^H\bar{\mathbf{\Psi}}\in\mathbb{C}^{LN_{\mathrm{T}}\times T}$ where $\tilde{\mathbf{N}}\in\mathbb{C}^{N_{\mathrm{R}}\times T}$ is the AWGN matrix with independent and identically distributed (i.i.d.) entries each having zero mean and variance σ_n^2.

An equivalent way to represent the sampling operator $\mathcal{S}_\Omega(\cdot)$ in the proposed architecture is via the Hadamard product. Let us consider the matrix $\mathbf{\Omega}\in\{0,1\}^{N_{\mathrm{R}}\times T}$ composed of TM_{R} ones and $T(N_{\mathrm{R}}-M_{\mathrm{R}})$ zeros, where the positions of its unity elements in each of its rows are randomly chosen in a uniform fashion over the set $\{1,2,\ldots,N_{\mathrm{R}}\}$. The role of the random spatial sampling unit is to process the output $\mathbf{R}$ of the extended analog combiner as follows:

$$\mathbf{R}_\Omega\triangleq\mathbf{\Omega}\circ\mathbf{R} \tag{22}$$

$$=\mathbf{\Omega}\circ\big(\mathbf{Y}+\mathbf{N}\big) \tag{23}$$

with $\mathbf{R}_\Omega\in\mathbb{C}^{M_{\mathrm{R}}\times T}$, $\mathbf{Y}\triangleq\mathbf{A}\bar{\mathbf{Z}}\mathbf{B}$, and $\mathbf{N}\triangleq(\mathbf{W}_{\mathrm{RF}}^{e})^H\tilde{\mathbf{N}}$. Evidently, each column of $\mathbf{R}_\Omega$ will have only M_{R} non-zero rows out of its N_{R} in total, which will be fed to the M_{R} receive RF chains.

The proposed analog combiner architecture for HBF reception retains the low complexity and low power consumption characteristics of standard HBF reception. It also increases the degrees of freedom for channel estimation since the resulting effective channel has larger number of dominant directions. Indeed, by reconstructing the training signal matrix $\tilde{\mathbf{Y}}\in\mathbb{C}^{N_{\mathrm{R}}\times T}$ we recover the $N_{\mathrm{R}}-M_{\mathrm{R}}$ lost directions due to the reduced number of RF chains. In the ideal case where $\mathbf{W}_{\mathrm{RF}}^{e}=\mathbf{D}_{\mathrm{R}}$, it can be easily concluded that:

$$(\mathbf{W}_{\mathrm{RF}}^{e})^H\tilde{\mathbf{Y}}=\bar{\mathbf{Z}}(\mathbf{I}_L\otimes\mathbf{D}_{\mathrm{T}}^{H})\bar{\mathbf{\Psi}}\in\mathbb{C}^{N_{\mathrm{R}}\times T}. \tag{24}$$

The latter equality is true since $(\mathbf{W}_{\mathrm{RF}}^{e})^H\mathbf{D}_{\mathrm{R}}=\mathbf{I}_{N_{\mathrm{R}}}$ holds.

IV. PROPOSED WIDEBAND MMWAVE MIMO CHANNEL ESTIMATION

In this section, we present the proposed channel estimation problem formulation together with its detailed algorithmic solution.

A. Problem Formulation

It follows from (21) that the received training signal matrix at the output of the extended combiner is given by the sum of the *low-rank matrix* $(\mathbf{W}_{\mathrm{RF}}^{e})^H\tilde{\mathbf{Y}}$ that includes the training symbols after passing through the wideband mmWave MIMO

channel, and the AWGN matrix $\mathbf{N}$. In this work, we propose the exploitation of the low-rank property of $\tilde{\mathbf{Y}}$ for the estimation of the wideband channel matrix $\bar{\mathbf{Z}}$. To do so, let us first investigate the rank properties of the latter matrix. For simplicity, let us consider the ideal case where $(\mathbf{W}_{\mathrm{RF}}^{e})^{H}\mathbf{D}_{\mathrm{R}} = \mathbf{I}_{N_{\mathrm{R}}}$. In this case, the next proposition provides a *worst-case upper bound* for the rank of $\bar{\mathbf{Z}}(\mathbf{I}_L \otimes \mathbf{D}_{\mathrm{T}}^{H})\bar{\mathbf{\Psi}}$, which represents the received training signal after applying the extended analog combiner denoted by $\mathbf{W}_{\mathrm{RF}}^{e}$, i.e, $(\mathbf{W}_{\mathrm{RF}}^{e})^{H}\tilde{\mathbf{Y}}$.

Proposition 1: The rank of $\mathbf{Q} \triangleq \bar{\mathbf{Z}}(\mathbf{I}_L \otimes \mathbf{D}_{\mathrm{T}}^{H})\bar{\mathbf{\Psi}}$ is upper bounded by

$$\mathrm{rank}(\mathbf{Q}) \leq \min(N_p, LN_{\mathrm{T}}). \tag{25}$$

with L denoting the number of delay taps of the wideband mmWave MIMO channel and N_p is the number of propagation channel paths.

Proof: See Appendix A. ■

Capitalizing on both the sparse structure of $\bar{\mathbf{Z}}$ and the low rank property of the training symbols' matrix $\tilde{\mathbf{Y}}$, we formulate the following dual objective Optimization Problem (OP) for the estimation of $\bar{\mathbf{Z}}$:

$$\begin{aligned} &\min_{\mathbf{Y},\bar{\mathbf{Z}}} \ \tau_R\|\mathbf{Y}\|_* + \tau_Z\|\bar{\mathbf{Z}}\|_1 \\ &\text{subject to } \ \mathbf{R}_{\Omega} = \mathbf{\Omega} \circ \big(\mathbf{Y} + \mathbf{N}\big), \\ &\qquad\qquad \mathbf{Y} = \mathbf{A}\bar{\mathbf{Z}}\mathbf{B}, \end{aligned} \tag{26}$$

where $\mathbf{Y}$'s nuclear norm in the objective function imposes its low rank property, whereas the ℓ_1-norm of $\bar{\mathbf{Z}}$ enforces its sparse structure. The weighting factors $\tau_R, \tau_Z > 0$ depend in general on the number L of the distinct mmWave MIMO channel propagation paths. Of course the noise matrix $\mathbf{N}$ is unknown, so in the following, we replace the first constraint in (26) with its least-squares estimate $\|\mathbf{R}_{\Omega} - \mathbf{\Omega} \circ \mathbf{Y}\|_F^2$.

B. Solution via Alternating Minimization

The OP (26) can be solved optimally using the ADMM technique. A similar approach has been adopted in [17], however there, the sampling is performed on the channel matrix, and not on the received training signal, as shown from (26). To solve (26), we proceed as follows. We first introduce the two auxiliary matrix variables $\mathbf{X} \in \mathbb{C}^{N_{\mathrm{R}} \times T}$ and $\mathbf{C} \triangleq \mathbf{Y} - \mathbf{A}\bar{\mathbf{Z}}\mathbf{B}$ to reformulate the targeted OP in the following equivalent form:

$$\begin{aligned} &\min_{\mathbf{Y},\bar{\mathbf{Z}},\mathbf{X},\mathbf{C}} \tau_R\|\mathbf{Y}\|_* + \tau_Z\|\bar{\mathbf{Z}}\|_1 \\ &\qquad + \frac{1}{2}\|\mathbf{C}\|_F^2 + \frac{1}{2}\|\mathbf{\Omega} \circ \mathbf{X} - \mathbf{R}_{\Omega}\|_F^2 \\ &\text{subject to } \mathbf{Y} = \mathbf{X} \text{ and } \mathbf{C} = \mathbf{Y} - \mathbf{A}\bar{\mathbf{Z}}\mathbf{B}, \end{aligned} \tag{27}$$

which is equivalent to (26), but the cost function has been decomposed into the sum of four variables: $\mathbf{Y}, \bar{\mathbf{Z}}, \mathbf{X}$, and $\mathbf{C}$. Note that now the third term in the objective function takes into account the discretization error, while the fourth term is the AWGN noise. The Lagrangian function of the OP in (27) is easily expressed as:

$$\begin{aligned} \mathcal{L}(\mathbf{Y}, \bar{\mathbf{Z}}, \mathbf{X}, \mathbf{C}, \mathbf{V}_1, \mathbf{V}_2) &= \tau_R\|\mathbf{Y}\|_* + \tau_Z\|\bar{\mathbf{Z}}\|_1 \\ &+ \frac{1}{2}\|\mathbf{C}\|_F^2 + \frac{1}{2}\|\mathbf{\Omega} \circ \mathbf{X} - \mathbf{R}_{\Omega}\|_F^2 + \mathrm{tr}(\mathbf{V}_1^{H}(\mathbf{Y} - \mathbf{X})) \\ &+ \frac{\rho}{2}\|\mathbf{Y} - \mathbf{X}\|_F^2 + \mathrm{tr}(\mathbf{V}_2^{H}(\mathbf{C} - \mathbf{X} + \mathbf{A}\bar{\mathbf{Z}}\mathbf{B})) \\ &+ \frac{\rho}{2}\|\mathbf{C} - \mathbf{X} + \mathbf{A}\bar{\mathbf{Z}}\mathbf{B})\|_F^2, \end{aligned} \tag{28}$$

where $\mathbf{V}_1 \in \mathbb{C}^{N_{\mathrm{R}} \times T}$ and $\mathbf{V}_2 \in \mathbb{C}^{N_{\mathrm{R}}T \times 1}$ are dual variables (the Lagrange multipliers) adding the constraints of (27) to the cost function, and ρ denotes the stepsize for ADMM.

According to the standard ADMM approach, at the i-th algorithmic iteration, with $i = 0, 1, \ldots, I_{\max}$, the following separate sub-problems need to be solved:

$$\mathbf{Y}^{(i+1)} = \arg\min_{\mathbf{Y}} \mathcal{L}_1\big(\mathbf{Y}, \mathbf{X}^{(i)}, \bar{\mathbf{Z}}^{(i)}, \mathbf{C}^{(i)}, \mathbf{V}_1^{(i)}, \mathbf{V}_2^{(i)}\big), \tag{29}$$

$$\mathbf{X}^{(i+1)} = \arg\min_{\mathbf{X}} \mathcal{L}_1\big(\mathbf{Y}^{(i+1)}, \mathbf{X}, \bar{\mathbf{Z}}^{(i)}, \mathbf{C}^{(i)}, \mathbf{V}_1^{(i)}, \mathbf{V}_2^{(i)}\big), \tag{30}$$

$$\bar{\mathbf{Z}}^{(i+1)} = \arg\min_{\bar{\mathbf{Z}}} \mathcal{L}_1\big(\mathbf{Y}^{(i+1)}, \mathbf{X}^{(i+1)}, \bar{\mathbf{Z}}, \mathbf{C}^{(i)}, \mathbf{V}_1^{(i)}, \mathbf{V}_2^{(i)}\big), \tag{31}$$

$$\mathbf{C}^{(i+1)} = \arg\min_{\mathbf{C}} \mathcal{L}_1\big(\mathbf{Y}^{(i+1)}, \mathbf{X}^{(i+1)}, \bar{\mathbf{Z}}^{(i+1)}, \mathbf{C}, \mathbf{V}_1^{(i)}, \mathbf{V}_2^{(i)}\big), \tag{32}$$

$$\mathbf{V}_1^{(i+1)} = \mathbf{V}_1^{(i)} + \rho\big(\mathbf{X}^{(i+1)} - \mathbf{Y}^{(i+1)}\big), \tag{33}$$

$$\mathbf{V}_2^{(i+1)} = \mathbf{V}_2^{(i)} + \rho\big(\mathbf{C}^{(i+1)} - \mathbf{X}^{(i+1)} + \mathbf{A}\bar{\mathbf{Z}}^{(i+1)}\mathbf{B}\big)\big). \tag{34}$$

Note that for the initialization $i = 0$: $\mathbf{Y}^{(0)} = \bar{\mathbf{Z}}^{(0)} = \mathbf{C}^{(0)} = \mathbf{V}_1^{(0)} = \mathbf{V}_2^{(0)} = \mathbf{0}$, $\mathbf{X}^{(0)} = \mathbf{R}_{\Omega}$.

The partial derivatives of each of the latter sub-problems can be obtained as follows.

Solution of (29): The first subproblem considers the optimization over the variable $\mathbf{Y}$, so let us express (29) by keeping only the terms that are related to it and completing the square, i.e.,

$$\mathbf{Y}^{(i+1)} = \arg\min_{\mathbf{Y}} \tau_Y\|\mathbf{Y}\|_* + \frac{\rho}{2}\|\mathbf{Y} - \left(\mathbf{X}^{(i)} - \frac{1}{\rho}\mathbf{V}_1^{(i)}\right)\|_F^2. \tag{35}$$

It is known that the solution of (35) can be obtained from the Singular Value Thresholding (SVT) operator as follows [19]:

$$\begin{aligned} \mathbf{Y}^{(i+1)} = \mathbf{U}_{\mathrm{L}}^{(i)}\mathrm{diag}\big(\{\mathrm{sign}(\zeta_j^{(i)}) \\ \times \max(\zeta_j^{(i)}, 0)\}_{1 \leq j \leq r}\big)(\mathbf{U}_{\mathrm{R}}^{(i)})^{H}, \end{aligned} \tag{36}$$

where $\mathbf{U}_{\mathrm{L}}^{(i)} \in \mathbb{C}^{N_{\mathrm{R}} \times r}$ and $\mathbf{U}_{\mathrm{R}}^{(i)} \in \mathbb{C}^{N_{\mathrm{R}} \times r}$ contain the left and right singular vectors, respectively, of the matrix $(\mathbf{X}^{(i)} - \frac{1}{\rho}\mathbf{V}_1^{(i)})$, and $\zeta_j^{(i)} \triangleq \sigma_j - \tau/\rho$ with σ_j denoting its r singular values.

Solution of (30): Differentiating with respect to $\mathbf{X}$, we have:

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \mathbf{X}} = \frac{\partial}{\partial \mathbf{X}} \Bigg(&\frac{1}{2}\|\mathbf{\Omega} \circ \mathbf{X} - \mathbf{R}_{\Omega}\|_F^2 + \mathrm{tr}((\mathbf{V}_1^{(i)})^{H}(\mathbf{Y}^{(i+1)} - \mathbf{X})) \\ &+ \frac{\rho}{2}\|\mathbf{Y}^{(i)} - \mathbf{X}\|_F^2 + \mathrm{tr}((\mathbf{V}_2^{(i)})^{H}(\mathbf{C}^{(i)} - \mathbf{X} + \mathbf{A}\bar{\mathbf{Z}}^{(i)}\mathbf{B})) \\ &+ \frac{\rho}{2}\left\|\mathbf{C}^{(i)} - \mathbf{X} + \mathbf{A}\bar{\mathbf{Z}}^{(i)}\mathbf{B}\right\|_F^2\Bigg) \end{aligned} \tag{37}$$

$$\begin{aligned} &= \mathbf{\Omega} \circ \mathbf{X} - \mathbf{R}_{\Omega} - \mathbf{V}_1^{(i)} - \rho(\mathbf{Y}^{(i+1)} - \mathbf{X}) \\ &\quad - \mathbf{V}_2^{(i)} - \rho(\mathbf{C}^{(i)} - \mathbf{X} + \mathbf{A}\bar{\mathbf{Z}}^{(i)}\mathbf{B}). \end{aligned} \tag{38}$$

If we set (38) equal to zero, the resulting problem is equivalent to solving the following system of equations (for details see Appendix D):

$$\mathbf{x} = (\mathbf{K}_1 + 2\rho\mathbf{I})^{-1}(\mathbf{v}_1^{(i)} + \rho\mathbf{y}^{(i+1)} + \mathbf{R}_\Omega + \mathbf{v}_2^{(i)} + \rho\mathbf{c}^{(i)} + \rho\mathbf{K}_2\bar{\mathbf{z}}^{(i)}), \quad (39)$$

where

$$\mathbf{K}_1 \triangleq \sum_{j=1}^{N_\mathrm{R}} \mathrm{diag}([\mathbf{\Omega}]_j)^T \otimes \mathbf{E}_{jj} \in \mathbb{C}^{TN_\mathrm{R} \times TN_\mathrm{R}}, \quad (40)$$

with $\mathbf{E}_{jj}$ obtained from the $N_\mathrm{R} \times N_\mathrm{R}$ all-zero matrix after inserting a unity value at its (j, j)-th position, and

$$\mathbf{K}_2 \triangleq \mathbf{B}^T \otimes \mathbf{A} \in \mathbb{C}^{TN_\mathrm{R} \times LN_\mathrm{T}N_\mathrm{R}}. \quad (41)$$

Also, small boldfaced letters in (39) are the $\mathrm{vec}(\cdot)$ results of their capital equivalents, e.g., $\mathbf{c} = \mathrm{vec}(\mathbf{C})$.

Solution of (31): The subproblem (31) concerns the unknown variable $\bar{\mathbf{Z}}$, which can be equivalently expressed as:

$$\bar{\mathbf{Z}}^{(i+1)} = \arg\min \tau_Z \|\bar{\mathbf{Z}}\|_1 + \frac{\rho}{2} \left\| \frac{1}{\rho}\mathbf{V}_2^{(i)} + \mathbf{C}^{(i)} - \mathbf{X}^{(i+1)} + \mathbf{A}\bar{\mathbf{Z}}\mathbf{B} \right\|_F^2. \quad (42)$$

By performing vectorization, the minimization of (42) is equivalent to the following sparse optimization problem:

$$\min_{\bar{\mathbf{z}}} \tau_Z \|\bar{\mathbf{z}}\|_1 + \|\mathbf{K}_2\bar{\mathbf{z}} - \mathbf{k}^{(i)}\|_2^2, \quad (43)$$

where $\mathbf{k}^{(i)} \triangleq \mathbf{x}^{(i+1)} - \mathbf{c}^{(i)} - \frac{1}{\rho}\mathbf{v}_2^{(i)} \in \mathbb{C}^{TN_\mathrm{R} \times 1}$ and $\bar{\mathbf{z}} \in \mathbb{C}^{LN_\mathrm{T}N_\mathrm{R} \times 1}$.

To solve the problem of (43) we express it in the standard LASSO form [28], i.e.,

$$\min_{\bar{\mathbf{z}}} \|\bar{\mathbf{z}}\|_1 + \|\bar{\mathbf{z}} - \boldsymbol{\beta}^{(i)}\|_2^2, \quad (44)$$

where

$$\boldsymbol{\beta}^{(i)} \triangleq \mathbf{K}_2^\dagger \mathbf{k}^{(i)} \in \mathbb{C}^{LN_\mathrm{T}N_\mathrm{R} \times 1}.$$

Hence, a soft-thresholding operator can be applied to provide an estimate of $\bar{\mathbf{z}}$ in (44), which is expressed as follows:

$$\bar{\mathbf{z}}^{(i+1)} = \mathrm{sign}(\mathtt{Re}(\boldsymbol{\beta}^{(i)})) \circ \max\left(|\mathtt{Re}(\boldsymbol{\beta}^{(i)})| - \tau_Z', 0\right) + \mathrm{sign}(\mathtt{Im}(\boldsymbol{\beta}^{(i)})) \circ \max\left(|\mathtt{Im}(\boldsymbol{\beta}^{(i)})| - \tau_Z', 0\right), \quad (45)$$

where $\tau_Z' \triangleq \tau_Z/\rho$, and the $\max(\cdot)$ and the sign operator $\mathrm{sign}(\cdot)$ are applied component wise. Note that the superscript $(i+1)$ will be used in the proposed iterative algorithm. The resulting vector in (45) is then transformed into matrix form as $\bar{\mathbf{Z}}^{(i+1)} = \mathrm{unvec}(\bar{\mathbf{z}}^{(i+1)}) \in \mathbb{C}^{LN_\mathrm{R} \times N_\mathrm{T}}$.

Solution of (32): Finally, differentiating with respect to $\mathbf{C}$ yields:

$$\frac{\partial \mathcal{L}}{\partial \mathbf{C}} = \frac{\partial}{\partial \mathbf{C}} \left(\frac{1}{2}\|\mathbf{C}\|_F^2 + \mathrm{tr}((\mathbf{V}_2^{(i)})^H(\mathbf{C} - \mathbf{X}^{(i+1)} + \mathbf{A}\bar{\mathbf{Z}}^{(i+1)}\mathbf{B} + \frac{\rho}{2}\|\mathbf{C} - \mathbf{X}^{(i+1)} + \mathbf{A}\bar{\mathbf{Z}}^{(i+1)}\mathbf{B}\|_F^2 \right) = (1+\rho)\mathbf{C} - \rho\left(\mathbf{X}^{(i+1)} - \mathbf{A}\bar{\mathbf{Z}}^{(i+1)}\mathbf{B} - \frac{1}{\rho}\mathbf{V}_2^{(i)}\right), \quad (46)$$

Algorithm 1: Proposed ADMM-based Channel Estimation.

Input: $\mathbf{R}_\Omega$, $\mathbf{\Omega}$, $\mathbf{A}$, $\mathbf{B}$, ρ, τ_R, τ_Z, and $I_{\max}$.
Output: $\bar{\mathbf{Z}}^{(I_{\max})}$
Initialization: $\mathbf{Y}^{(0)} = \bar{\mathbf{Z}}^{(0)} = \mathbf{C}^{(0)} = \mathbf{V}_1^{(0)} = \mathbf{V}_2^{(0)} = \mathbf{0}, \mathbf{X}^{(0)} = \mathbf{R}_\Omega$
1: **for** $i = 0, 1, \ldots, I_{\max} - 1$ **do**
2: Update $\mathbf{Y}^{(i+1)}$ using (36).
3: Solve the system of equations in (39).
4: Update $\mathbf{X}^{(i+1)} = \mathrm{unvec}(\mathbf{x}^{(i+1)})$.
5: Solve the sparse optimization (44) using (45).
6: Update $\bar{\mathbf{Z}}^{(i+1)} = \mathrm{unvec}(\bar{\mathbf{z}}^{(i+1)})$.
7: Update $\mathbf{C}^{(i+1)}$ using (47).
8: Update $\mathbf{V}_1^{(i+1)}$ and $\mathbf{V}_2^{(i+1)}$ using (33) and (34).
9: **end for**

and setting (46) equal to zero leads to the solution:

$$\mathbf{C}^{(i+1)} = \frac{\rho}{1+\rho}\left(\mathbf{X}^{(i+1)} - \mathbf{A}\bar{\mathbf{Z}}^{(i+1)}\mathbf{B} - \frac{1}{\rho}\mathbf{V}_2^{(i)}\right). \quad (47)$$

We have summarized the algorithmic steps of the proposed ADMM-based channel estimation method in Algorithm 1.

C. Computational Complexity

Algorithm 1 involves some algorithmic steps which are computationally demanding in terms of processing and memory operations. To have a better understanding of the requirements, we next describe the computational complexity of Algorithm 1, considering the basic steps separately.

- In Line 2 of the Algorithm 1, the solution of (36) is obtained by employing the SVT operator on the dense non-square matrix $\mathbf{Y} \in \mathbb{C}^{N_\mathrm{R} \times T}$. Recall that SVT is based on the Singular Value Decomposition (SVD). Thus, in general, this step requires complexity which is proportional to $M_\mathrm{R}^2 T$ [29, Chapter 8.6]. However, for large values of T and M_R, the dominant singular values and vectors can be efficiently computed via incomplete SVD methods (e.g., Lanczos bidiagonalization algorithm [30]) or via subspace tracking (e.g., [31]), where the complexity can be reduced to $\mathcal{O}(M_\mathrm{R}T)$.
- In the 3rd Line of the algorithm, the solution of the system of equations in (39) is required, which involves the inversion of the matrix $\mathbf{K}_1 + 2\rho\mathbf{I} \in \mathbb{C}^{TN_\mathrm{R} \times TN_\mathrm{R}}$. However, this matrix is diagonal, hence, the order of the complexity is $\mathcal{O}(TN_\mathrm{R})$.
- In Line 5 of the Algorithm 1, the pseudo-inverse of $\mathbf{K}_2 \in \mathbb{C}^{TN_\mathrm{R} \times LN_\mathrm{T}N_\mathrm{R}}$ needs to be computed. This requires the computation and the inversion of the Gram matrix $\mathbf{K}_2^H\mathbf{K}_2 \in \mathbb{C}^{LN_\mathrm{T}N_\mathrm{R} \times LN_\mathrm{T}N_\mathrm{R}}$, which is the most costly step of the proposed algorithm. However, it can be seen that $\mathbf{K}_2^H\mathbf{K}_2$ is a dominant diagonal matrix, hence gradient-based iterative algorithms can be used to reduce the complexity for the inversion of the Gram matrix to $\mathcal{O}(LN_\mathrm{T}N_\mathrm{R})$ [32].

The other lines of Algorithm 1 involve the computation of matrix-matrix and matrix-vector products, which have smaller computational order than the aforementioned steps.

V. Extensions

In this section, we detail a couple of extensions for the proposed wideband mmWave MIMO channel estimation framework. We first present an algorithm for the case where information for the angles of the propagation paths is available. For this case, we also present a low complexity solution. Finally, we discuss the consideration of the beam squint effect in the proposed estimation framework.

A. Exploitation of Angle Information

The previously described wideband mmWave MIMO channel estimation algorithm can be extended to incorporate prior knowledge of the angles of the propagation channel paths, e.g., obtained via dedicated direction estimation techniques [33], [34]. As is will be shown in the section with the simulation results that follow, this knowledge can boost the CSI estimation performance in certain cases.

To exploit angle information in the considered beamspace representation of the wideband mmWave MIMO channel matrix $\bar{\mathbf{Z}}$, we introduce the selection matrix $\mathbf{\Omega}_S$ which contains 1's at its elements corresponding to the receive or transmit angles, and 0's elsewhere. For those cases, instead of minimizing the ℓ_1 norm of the sparse matrix $\bar{\mathbf{Z}}$, we first impose the *a priori* known structure (offered by the availability of the path angles) to this matrix, and then insert the term $\|\mathbf{\Omega}_S \circ \bar{\mathbf{Z}}\|_1$ into the problem formulation (26) in the place of $\|\mathbf{Z}\|_1$, where $\mathbf{\Omega}_S \in \{0,1\}^{N_{\mathrm{R}} \times T}$. This matrix is composed of S ones and $T(N_{\mathrm{R}} - S)$ zeros, while the positions of its unity elements are chosen based on prior information of the angles. Using the latter considerations, we formulate the following OP for wideband mmWave MIMO channel estimation for the case where prior angle information is available

$$\begin{aligned} \min_{\mathbf{Y},\bar{\mathbf{Z}}} \ & \tau_Y \|\mathbf{Y}\|_* + \tau_Z \|\mathbf{\Omega}_S \circ \bar{\mathbf{Z}}\|_1 \\ \text{subject to} \ & \mathbf{R}_\Omega = \mathbf{\Omega} \circ \big(\mathbf{Y} + \mathbf{N}\big), \\ & \mathbf{Y} = \mathbf{A}\bar{\mathbf{Z}}\mathbf{B}, \end{aligned} \tag{48}$$

where the objective is to obtain the estimation of $\bar{\mathbf{Z}}$. This OP can be solved following the previously developed ADMM-based methodology, while replacing the optimization of (31) as follows. More specifically, the subproblem (42) is expressed for the considered prior angle knowledge case as:

$$\bar{\mathbf{Z}} = \arg\min \tau_Z \|\mathbf{\Omega}_S \circ \bar{\mathbf{Z}}\|_1 + \frac{\rho}{2}\|\frac{1}{\rho}\mathbf{V}_2 + \mathbf{C} - \mathbf{X} + \mathbf{A}\bar{\mathbf{Z}}\mathbf{B}\|_F^2, \tag{49}$$

which can be shown to be equivalent to the following optimization problem (using similar steps as in Appendix D):

$$\min_{\bar{\mathbf{z}}} \|\mathbf{K}_3 \bar{\mathbf{z}}\|_1 + \|\bar{\mathbf{z}} - \boldsymbol{\beta}'^{(i)}\|_2^2, \tag{50}$$

with $\bar{\mathbf{z}} \in \mathbb{C}^{N_{\mathrm{T}} N_{\mathrm{R}} \times 1}$, $\boldsymbol{\beta}'^{(i)}$ given by

$$\boldsymbol{\beta}'^{(i)} = \mathbf{K}_2^{\dagger}\left(\mathbf{x}^{(i+1)} - \mathbf{c}^{(i)} - \frac{1}{\rho}\mathbf{v}_2^{(i)}\right), \tag{51}$$

and

$$\mathbf{K}_3 \triangleq \sum_{j=1}^{N_{\mathrm{R}}} \mathrm{diag}([\mathbf{\Omega}]_j)^T \otimes \mathbf{E}_{jj}\mathrm{vec}(\mathbf{X}). \tag{52}$$

Note that $\mathbf{K}_3 \in \mathbb{C}^{N_{\mathrm{T}} N_{\mathrm{R}} \times N_{\mathrm{T}} N_{\mathrm{R}}}$ applies a hard threshold to the known structure of $\bar{\mathbf{Z}}$, while for the remaining entries we apply the soft-thresholding operator as follows:

$$\begin{aligned} \bar{\mathbf{z}}^{(i+1)} = & \ \mathrm{sign}(\mathtt{Re}(\boldsymbol{\gamma}^{(i)})) \circ \max\left(|\mathtt{Re}(\boldsymbol{\gamma}^{(i)})| - \tau_Z', 0\right) \\ & + \mathrm{sign}(\mathtt{Im}(\boldsymbol{\gamma}^{(i)})) \circ \max\left(|\mathtt{Im}(\boldsymbol{\gamma}^{(i)})| - \tau_Z', 0\right). \end{aligned} \tag{53}$$

In the latter expression, we have defined for each algorithmic iteration the generic vector

$$\boldsymbol{\gamma}^{(i+1)} \triangleq \mathbf{K}_3\mathbf{K}_2^{\dagger}\left(\mathbf{x}^{(i+1)} - \mathbf{c}^{(i)} - \frac{1}{\rho}\mathbf{v}_2^{(i)}\right) \in \mathbb{C}^{LN_{\mathrm{T}}N_{\mathrm{R}} \times 1}. \tag{54}$$

Putting all above together, by replacing (45) in line 5 of Algorithm 1 with (53) yields our proposed ADMM-based channel estimation algorithm for the special case where angle information is *a priori* known. The computational complexity of this new algorithm follows the properties of Algorithm 1 requiring also the construction and the application of the diagonal matrix $\mathbf{K}_3 \in \mathbb{C}^{N_{\mathrm{T}} N_{\mathrm{R}} \times N_{\mathrm{T}} N_{\mathrm{R}}}$, i.e., $\mathcal{O}(N_{\mathrm{T}} N_{\mathrm{R}})$ operations. Concerning the computational complexity of the technique that exploits the prior angle information, it only requires the construction and the application of the diagonal matrix $\mathbf{K}_3 \in \mathbb{C}^{N_{\mathrm{T}} N_{\mathrm{R}} \times N_{\mathrm{T}} N_{\mathrm{R}}}$, i.e., $\mathcal{O}(N_{\mathrm{T}} N_{\mathrm{R}})$ operations.

B. Low-Complexity Implementation

To deal with the increased computational complexity overhead of Algorithm 1, we hereinafter modify its part with the highest computational burden, i.e., line 5 which calls for the solution of $\boldsymbol{\beta}^{(i)} \triangleq \mathbf{K}_2^{\dagger}\mathbf{k}^{(i)} \in \mathbb{C}^{N_{\mathrm{R}} N_{\mathrm{T}} \times 1}$. Specifically, we consider the following system of normal equations:

$$\mathbf{\Phi}\boldsymbol{\beta}^{(i)} = \mathbf{b}^{(i)}, \tag{55}$$

where

$$\mathbf{\Phi} \triangleq \mathbf{K}_2^H\mathbf{K}_2 \in \mathbb{C}^{LN_{\mathrm{T}}N_{\mathrm{R}} \times LN_{\mathrm{T}}N_{\mathrm{R}}}, \tag{56}$$

and

$$\mathbf{b}^{(i)} \triangleq \mathbf{K}_2^H\mathbf{k}^{(i)} \in \mathbb{C}^{LN_{\mathrm{T}}N_{\mathrm{R}} \times 1}. \tag{57}$$

Instead of employing the exact least-squares solution, at each ADMM iteration step i, we approximate $\boldsymbol{\beta}^{(i)}$ with one step from the iterative Gradient Descent (GD) algorithm. This GD step for the i-th iteration is expressed as:

$$\tilde{\boldsymbol{\beta}}^{(i)} = \tilde{\boldsymbol{\beta}}^{(i-1)} - \alpha^{(i)}\mathbf{r}^{(i)}, \tag{58}$$

where $\mathbf{r}^{(i)}$ is the residual of the i-th GD step, defined as:

$$\mathbf{r}^{(i)} = \mathbf{b}^{(i)} - \mathbf{\Phi}\tilde{\boldsymbol{\beta}}^{(i-1)}. \tag{59}$$

In other words, we employ an inexact gradient method to approximate $\boldsymbol{\beta}^{(i)}$, since at each step the right hand side is also varying over i. The step size $\alpha^{(i)}$ can be obtained using a pre-fixed value,

Algorithm 2: Low-complexity Channel Estimation Exploiting Angle Information.

Input: $\mathbf{R}_\Omega$, $\mathbf{\Omega}$, $\mathbf{\Omega}_S$, $\mathbf{A}$, $\mathbf{B}$, ρ, τ_R, τ_Z, and $I_{\max}$.
Output: $\bar{\mathbf{Z}}^{(I_{\max})}$

Initialization: $\mathbf{Y}^{(0)} = \bar{\mathbf{Z}}^{(0)} = \mathbf{C}^{(0)} = \mathbf{V}_1^{(0)} = \mathbf{V}_2^{(0)} = \mathbf{0}_{N_\mathrm{R}\times T}$, $\mathbf{v}^{(0)} = \mathbf{0}_{N_\mathrm{R}N_\mathrm{T}\times 1}$

1: **for** $i = 0, 1, \ldots, I_{\max} - 1$ **do**
2: Update $\mathbf{Y}^{(i+1)}$ using (36).
3: Solve the system of equations in (39).
4: Update $\mathbf{X}^{(i+1)} = \mathrm{unvec}(\mathbf{x}^{(i+1)})$.
5: Compute the GD step size and residual using (60) and (59).
6: Compute one GD step to obtain $\boldsymbol{\beta}^{(i)}$ using (58).
7: Solve the sparse optimization (44) using (45) and $\boldsymbol{\beta}^{(i)}$.
8: Update $\bar{\mathbf{Z}}^{(i+1)} = \mathrm{unvec}(\bar{\mathbf{z}}^{(i+1)})$.
9: Update $\mathbf{C}^{(i+1)}$ using (47).
10: Update $\mathbf{V}_1^{(i+1)}$ and $\mathbf{V}_2^{(i+1)}$ using (33) and (34).
11: **end for**

i.e., $\alpha^{(i)} = \alpha > 0$, or the exact line search, i.e.,

$$\alpha^{(i)} = \frac{(\mathbf{r}^{(i)})^H \mathbf{r}^{(i)}}{(\mathbf{r}^{(i)})^H \mathbf{A}\mathbf{r}^{(i)}}. \tag{60}$$

Remark: The convergence rate of the GD algorithm depends on the *spectral condition number* of the matrix $\mathbf{\Phi}$, which is the ratio of the largest to the smallest eigenvalue, $\kappa \triangleq \frac{\lambda_{\max}}{\lambda_{\min}}$. Specifically, the error of the i-th step is upper bounded by:

$$\|\mathbf{e}^{(i)}\|_\mathbf{A} \leq \left(\frac{\kappa - 1}{\kappa + 1}\right) \|\mathbf{e}^{(0)}\|_\mathbf{A}, \tag{61}$$

where $\mathbf{e}^{(i)} \triangleq \boldsymbol{\beta}^{\mathrm{opt}} - \boldsymbol{\beta}^{(i)}$ is the error vector. In our case, $\mathbf{\Phi}$ is a diagonal dominant matrix, which constraints the eigenvalues spread and guarantees fast convergence rate.

The aforementioned algorithmic steps can be incorporated into the proposed CSI methods of the current and previous sections by replacing line 5 with eqs. (58)–(60). Algorithm 2 describes the proposed low-complexity channel estimation method exploiting angle information. In terms of computational efficiency, instead of the computation of the pseudo-inverse of $\mathbf{K}_2$, Algorithm 2 requires only the computation of two matrix-vector products for the computation of the GD step size and residual. While the presented GD-based modification yields an inexact solution, it turns out that it behaves remarkably well. This is due to the interesting property of ADMM according to which, under certain conditions, it is observed to converge even in cases where the alternating minimization steps are not carried out exactly. In the next section with the performance evaluation results that follows, we will verify this behavior and we will compare the MSE estimation performance of both Algorithms 1 and 2.

C. Consideration of the Beam Squint Effect

In this section, we briefly discuss the impact of the beam squint effect on our problem formulation. Large size antenna arrays realizing directive beams in wideband communication systems are susceptible to the beam squint phenomenon that imposes selectivity both in the frequency and the spatial domains [35], [36]. This spatio-frequency selectivity needs to carefully accounted for when designing wideband mmWave massive MIMO systems. By considering the beam squint effect in the time domain, the received training signal at the t-th channel block at each m-th receiving antenna element with $m = 1, 2, \ldots, N_\mathrm{R}$ can be mathematically expressed as

$$[\mathbf{y}[t]]_m = \sum_{\ell=0}^{L-1} \sum_{k=1}^{N_\mathrm{T}} [\mathbf{h}_k(\ell)]_m s_k(t - \ell - m\psi_\ell), \tag{62}$$

where ψ_ℓ is a function of the direction-of-arrival for the ℓ-th path [35]. Following (62), the N_R-dimensional vector with the received training signal can be written as

$$\mathbf{y}[t] = \sum_{\ell=0}^{L-1} \sum_{k=1}^{N_\mathrm{T}} \mathbf{S}_{\ell,k}[t] \mathbf{h}_k(\ell), \tag{63}$$

where $\mathbf{S}_{\ell,k}[t]$ is a $N_\mathrm{R} \times N_\mathrm{R}$ diagonal matrix with the diagonal entries $s_k(t - \ell - m\psi_\ell)\ \forall m = 1, 2, \ldots, N_\mathrm{R}$. Alternatively, (63) can be written in the following vector form:

$$\mathbf{y}[t] = \mathbf{\Xi}[t] \begin{bmatrix} \mathrm{vec}(\mathbf{H}(0)) \\ \vdots \\ \mathrm{vec}(\mathbf{H}(L-1)) \end{bmatrix}, \tag{64}$$

where $\mathbf{\Xi}[t]$ is a $(N_\mathrm{T}N_\mathrm{R}(L-1)) \times N_\mathrm{R}$ matrix created by the horizontal concatenation of $N_\mathrm{T}(L-1)$ matrices $\mathbf{S}_{\ell,k}[t]\ \forall \ell = 0, 1, \ldots, L-1$ and $\forall k = 1, 2, \ldots, N_\mathrm{T}$. Equivalently, using the beamspace decomposition of the channel matrix for each delay tap, $\mathbf{y}[t]$ can be re-written as

$$\mathbf{y}[t] = \mathbf{\Xi}[t] \begin{bmatrix} (\mathbf{D}_\mathrm{T}^T \otimes \mathbf{D}_\mathrm{R})\mathrm{vec}(\mathbf{Z}(0)) \\ \vdots \\ (\mathbf{D}_\mathrm{T}^T \otimes \mathbf{D}_\mathrm{R})\mathrm{vec}(\mathbf{Z}(L-1)) \end{bmatrix} \tag{65}$$

$$= \mathbf{\Xi}[t]\mathbf{D}_\mathrm{bs} \begin{bmatrix} \mathrm{vec}(\mathbf{Z}(0)) \\ \vdots \\ \mathrm{vec}(\mathbf{Z}(L-1)) \end{bmatrix}, \tag{66}$$

where $\mathbf{D}_\mathrm{bs} \triangleq \mathbf{I}_{L-1} \otimes \mathbf{D}_\mathrm{T}^T \otimes \mathbf{D}_\mathrm{R}$. Finally, concatenating the columns of T received training blocks, one gets the following compact expression for the received training signal matrix $\mathbf{Y}_\mathrm{bs} \in \mathbb{C}^{N_\mathrm{T}\times T}$:

$$\mathbf{Y}_\mathrm{bs} \triangleq \mathbf{\Xi}_\mathrm{R} \left(\mathbf{I}_T \otimes \mathbf{D}_\mathrm{bs}\mathrm{vec}(\bar{\mathbf{Z}})\right), \tag{67}$$

where $\mathbf{\Xi}_\mathrm{R} \triangleq [\mathbf{\Xi}[1] \cdots \mathbf{\Xi}[T]]$. It is noted that the latter expression extends (12) to incorporate the beam squint effect.

Using the latter discussion for the case of beam squint, the optimization problem (26) can be extended to include the unknown beam squint matrix $\mathbf{\Xi}_\mathrm{R}$ as follows:

$$\begin{aligned} \min_{\mathbf{Y},\bar{\mathbf{Z}},\mathbf{\Xi}_\mathrm{R}} \ & \tau_R\|\mathbf{Y}_\mathrm{bs}\|_* + \tau_Z\|\bar{\mathbf{Z}}\|_1 \\ \text{subject to } & \mathbf{R}_\Omega = \mathbf{\Omega} \circ \left(\mathbf{Y}_\mathrm{bs} + \mathbf{N}\right), \\ & \mathbf{Y}_\mathrm{bs} = \mathbf{\Xi}_\mathrm{R} \left(\mathbf{I}_T \otimes \mathbf{D}_\mathrm{bs}\mathrm{vec}(\bar{\mathbf{Z}})\right). \end{aligned} \tag{68}$$

However, the investigation of (68)'s solution is out of the scope of this paper, and is left for future work.

VI. Performance Evaluation Results

In this section, we consider a mmWave point-to-point $N_{\mathrm{R}} \times N_{\mathrm{T}}$ MIMO system for various large values of N_{R} and N_{T} and investigate the performance of the proposed wideband channel estimation technique using the proposed analog combining architecture for HBF reception. All computer simulation results have been obtained using MATLAB. We have specifically simulated the Average Spectral Efficiency (ASE), Energy Efficiency (EE), and MSE performances, which were averaged over 100 Monte-Carlo realizations. We have also investigated the convergence speed and the computationally complexity of the proposed approach in comparison with relevant state-of-the-art techniques. Block transmissions with zero padding appended to each transmitted frame have been assumed, and the channel estimation has been performed in the time domain at RX over T frames. We have assumed that the channels coherence time is larger than the duration of the T frames required for channel estimation.

During each t-th channel estimation frame with $t = 1, 2, \ldots, T$, N_{T} complex-valued training symbols $\mathbf{s}(t)$ have being used such that $\mathcal{E}\{\mathbf{s}(t)\mathbf{s}(t)^H\} = \mathbf{I}_{N_{\mathrm{T}}}$. The training signal is received at the N_{R} antenna elements of the HBF RX, which are connected to $M_{\mathrm{R}} < N_{\mathrm{R}}$ RF chains. Finally, we have assumed perfect time and frequency synchronization between the communicating TX and RX sides and fully connected PS networks at all considered HBF RXs.

A. ASE and EE Performances

For the design of the PS network at the RX analog combiner, we have considered quantized angles, i.e., $[\mathbf{W}_{\mathrm{RF}}]_{i,j} = N_{\mathrm{R}}^{-1} e^{j\omega_{i,j}}$ with $\omega_{i,j} \in \mathcal{W}$ as defined in (15), for the two cases of angle quantization $N_Q = 4$ and 6. We have also included performance results for the extreme case of very high angle quantization resolution, where the analog combiner was designed based on Q-th root Zadoff-Chu (ZC) sequence with $Q = 11$ [32]. Moreover, we have considered the fully digital combining case, where the number of the RX RF chains is equal to the number of receiving antenna elements, i.e., $M_{\mathrm{R}} = N_{\mathrm{R}}$.

The ASE performance is illustrated in Fig. 3 as a function of the number of RX RF chains M_{R} for three different cases of system parameters and $\mathrm{SNR} = 15$ dB (SNR stands for Signal to Noise Ratio and is mathematically defined as $\mathrm{SNR} \triangleq \sigma_n^{-2}$). We have particularly numerically evaluated expressions (16) and (18) for the proposed analog BeamForming (BF) approach with 6-bit PSs and compared it with the ASE for the fully digital RX case, conventional analog BF with ZC combining, and conventional analog BF with 6-bit PSs. The obtained performance curves demonstrate that the proposed design is able to recover part the lost performance that happened due to the non-orthogonality of the conventional analog combining matrix. The ASE performance from conventional analog BF reception is more pronounced for low to mid number of RF chains, cases which are particularly relevant with the design specifications of the current cellular base stations. It is also evident in this figure that the proposed analog BF design exhibits smoother ASE curves over M_{R} due to the randomness of the selection operator happening in the random sampling unit.

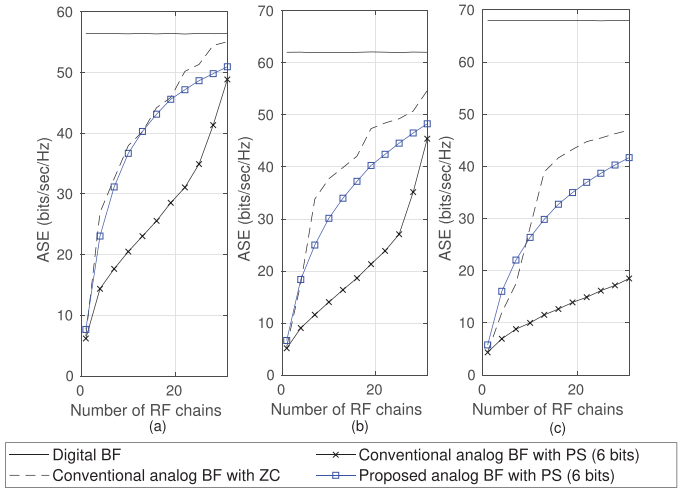

Fig. 3. ASE as a function of the number of RF chains M_{R} for $T = 40$, $N_p = 6$, $N_{\mathrm{T}} = 16$ and $\mathrm{SNR} = 15$ dB. (a) $N_{\mathrm{R}} = 32$, $M_{\mathrm{R}}^e = 32$, (b) $N_{\mathrm{R}} = 64$, $M_{\mathrm{R}}^e = 32$, and (c) $N_{\mathrm{T}} = 128$, $M_{\mathrm{R}}^e = 64$,.

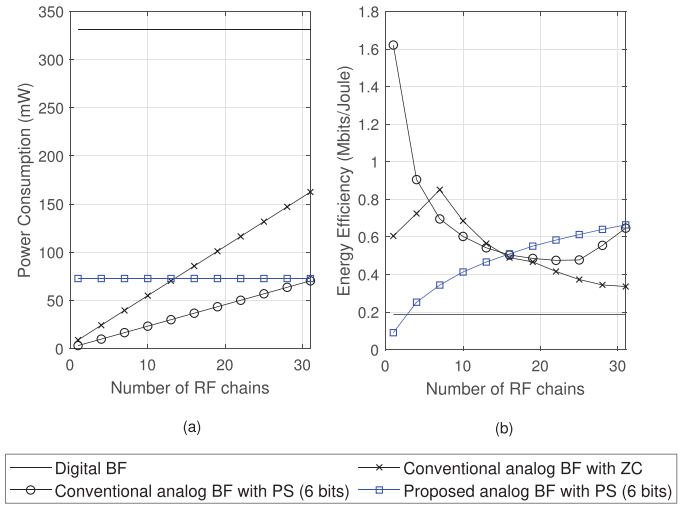

Fig. 4. Power Consumption (a) and EE (b) with respect to the number of RF chains M_{R} for $T = 40$, $N_p = 6$, $N_{\mathrm{T}} = 16$, $N_{\mathrm{R}} = 64$, $M_{\mathrm{R}}^e = 32$, and $\mathrm{SNR} = 15$ dB.

We now investigate the power consumption of the extended analog combining design, which has been detailed componentwise in Table II. In Fig. 4, we compare the required power in mW of the conventional and the proposed analog BF designs as functions of the number of RX RF chains M_{R}. We used the realistic values $P_{\mathrm{SW}} = 0.005$ mW and $P_{\mathrm{LNA}} = 0.02$ mW. For the conventional analog BF with PS we used $P_{\mathrm{PS}} = 0.015$ mW, while for the conventional analog BF with ZC we assume an equivalent to PS power consumption equal to $P_{\mathrm{PS}} = 0.06$ mW [8]. As shown in Fig. 4, the power consumption of the proposed design does not depend on M_{R}, but on the number of outputs of the extended analog BF $M_{\mathrm{R}}^e \geq M_{\mathrm{R}}$. This means that has always higher power requirements than the conventional analog BF. However, when $M_{\mathrm{R}} > N_{\mathrm{R}}/4$, the power consumption of the proposed analog BF is lower than the conventional ZC case. In particular, for the case of $N_{\mathrm{R}} = 64$ and $M_{\mathrm{R}}^e = 22$, the increase

in the power consumption with the proposed analog BF for 6-bit PSs compared to the conventional case the same PS resolution is close to 44%, while the increase in ASE is almost 77%.

The power consumption improvement offered by proposed analog combining architecture over the conventional one, as showcased in Fig. 4(a), can also be witnessed from the EE performance inspection. EE is mathematical expressed as:

$$\text{EE} \triangleq \frac{\text{ASE}}{P}(\text{Mbit/Joule}), \tag{69}$$

where P denotes the power consumption given by Table II. Figure 4(b) plots EE versus M_{R} and depicts that the proposed analog BF outperforms all compared designs when $M_{\text{R}} > N_{\text{R}}/4$. In Fig. 4(b) we show the EE results, where the proposed analog BF design exhibits higher efficiency when the number of RF chains is more than the $1/4$ of the receiving antennas.

B. Channel Estimation Performance

In this subsection, we evaluate the performance of the proposed algorithms for the estimation of beamspace channel matrix $\bar{\mathbf{Z}}$ for different system parameters. For the training data symbols we have considered Quadrature Amplitude Modulation (QAM) with the four constellation points $\frac{1}{\sqrt{2}}\{1+j, 1-j, -1+j, -1-j\}$. In assessing the performance of the algorithms and comparing with relevant state-of-the-art wideband mmWave MIMO channel estimation techniques, we have simulated the Normalized Mean-Square-Error (NMSE) criterion, which is defined as:

$$\text{NMSE} \triangleq \frac{\|\bar{\mathbf{Z}} - \hat{\bar{\mathbf{Z}}}\|}{\|\bar{\mathbf{Z}}\|}, \tag{70}$$

where $\hat{\bar{\mathbf{Z}}}$ represents the estimated channel matrix in the beamspace domain. The considered benchmark channel estimation techniques are based on compressive sensing and matrix completion tools. We have particularly simulated the performance of iterative thresholding and message passing for solving the ℓ_1-minimization problem, and the SVT [19] technique for solving the rank minimization problem. The iterative thresholding algorithms, e.g., Orthogonal Matching Pursuit (OMP) [11] and CoSaMP [39], can efficiency solve the ℓ_1-minimization problem with low computational complexity but they yield good estimation performance only for the case of highly sparse vectors, i.e., when the unknown vector has only few non-zero values. On the contrary, message passing techniques, e.g., Approximate Message Passing (AMP) [40] and Vector AMP (VAMP) [38], provide more robust estimation performance in the cases of signals with lower sparsity as well as for measurements with increased noise level.

The NMSE performance of the proposed channel estimation algorithms versus the number of training frames T is sketched in Fig. 5, where we also include NMSE curves for the Least Squares (LS) estimation given by:

$$\hat{\mathbf{Z}}_{\text{LS}} = (\tilde{\mathbf{Y}} + \tilde{\mathbf{N}})\big((\mathbf{I}_L \otimes \mathbf{D}_{\text{T}}^H)\bar{\mathbf{\Psi}}\big)^{\dagger} \in \mathbb{C}^{N_{\text{R}} \times L N_{\text{T}}}. \tag{71}$$

Additionally, we have included performance results for the compressed sensing: OMP [14], CoSaMP [39], and VAMP [38],

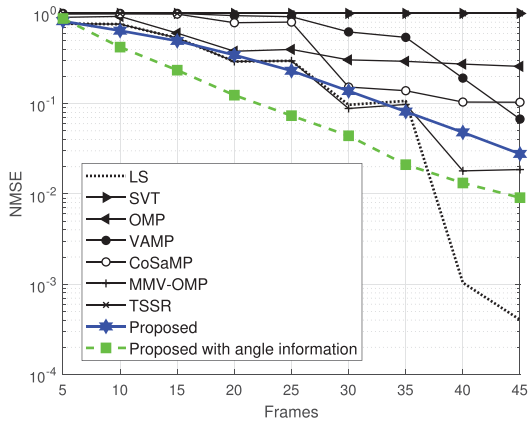

Fig. 5. NMSE performance as a function of the number of training frames T for $N_{\text{T}} = 8, N_{\text{R}} = 32, M_{\text{R}} = 4, M_{\text{R}}^e = 32, N_p = 6, L = 4$ and $\text{SNR} = 15$ dB.

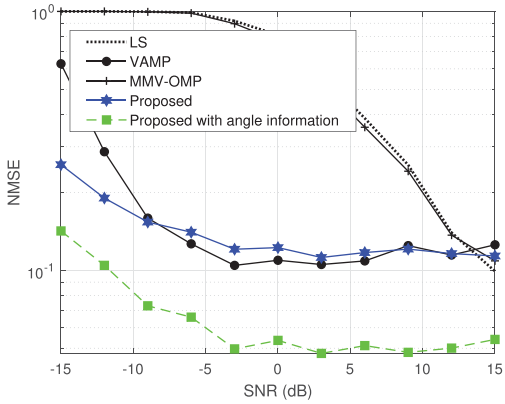

Fig. 6. NMSE performance as a function of the SNR for $N_{\text{T}} = 8, N_{\text{R}} = 32, M_{\text{R}} = 4, M_{\text{R}}^e = 32, N_p = 6, L = 4$, and $T = 20$.

which solve the following problem:

$$\min_{\bar{\mathbf{z}}} \|\bar{\mathbf{z}}\|_1 \text{ s.t. } \|\mathbf{y} - \big((\mathbf{I}_L \otimes \mathbf{D}_{\text{T}}^H)\bar{\mathbf{\Psi}} \otimes \mathbf{I}_{N_{\text{R}}}\big)\bar{\mathbf{z}}\|_F^2 \leq \epsilon, \tag{72}$$

where $\mathbf{y} \triangleq \text{vec}(\tilde{\mathbf{Y}} + \tilde{\mathbf{N}})$, $\bar{\mathbf{z}} \triangleq \text{vec}(\bar{\mathbf{Z}})$, and $\epsilon \in \mathbb{R}^+$ represents a very low positive real number. Note that VAMP represents a statistical learning estimator that is based on the training data, thus, larger training period is in principal required. The OMP-MMV algorithm is based on OMP but exploits the common sparsity pattern between the L beamspace matrices. This algorithm actually solves the following OP:

$$\min_{\bar{\mathbf{Z}}} \|\bar{\mathbf{Z}}\|_1 \text{ s.t. } \|\tilde{\mathbf{Y}} - \bar{\mathbf{Z}}\big((\mathbf{I}_L \otimes \mathbf{D}_{\text{T}}^H)\bar{\mathbf{\Psi}}\big)\|_F^2 \leq \epsilon. \tag{73}$$

In Fig. 6, we have also simulated the NMSE performance of the subspace thresholding techniques SVT and TSSR (define the abbreviation, it's not defined), which exploit the low-rank property of the received training signal matrix. The SVT algorithm solves the following OP:

$$\min_{\mathbf{Y}} \|\mathbf{Y}\|_* \text{ subject to } \mathbf{R}_{\Omega} = \mathbf{\Omega} \circ (\mathbf{Y} + \tilde{\mathbf{N}}), \tag{74}$$

whereas, TSSR offers a two-stage procedure which exploits both the sparsity and low-rank properties. This is accomplished by first employing the SVT operator to recover the received signal $\mathbf{Y}$ from (74), then by transforming it to the beamspace domain,

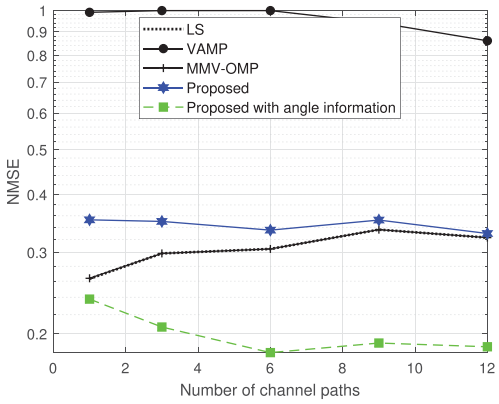

Fig. 7. NMSE performance as a function of the number of propagation channels N_p for $N_{\mathrm{T}} = 8$, $N_{\mathrm{R}} = 32$, $M_{\mathrm{R}} = 4$, $M_{\mathrm{R}}^e = 32$, $L = 4$, $T = 20$, and $\mathrm{SNR} = 15$ dB.

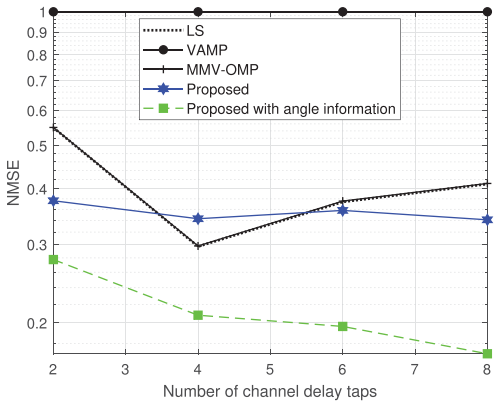

Fig. 8. NMSE performance as a function of the number of the channel delay taps L for $N_{\mathrm{T}} = 8$, $N_{\mathrm{R}} = 32$, $M_{\mathrm{R}} = 4$, $M_{\mathrm{R}}^e = 32$, $N_p = 6$, $T = 20$, and $\mathrm{SNR} = 15$ dB.

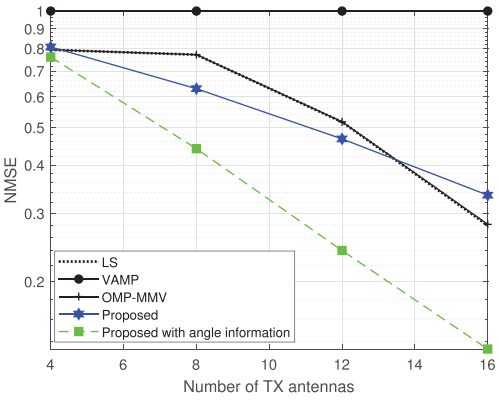

Fig. 9. NMSE performance as a function of the number of TX antennas N_{T} for $N_{\mathrm{R}} = 32$, $M_{\mathrm{R}} = 4$, $M_{\mathrm{R}}^e = 32$, $L = 4$, $N_p = 6$, $T = 20$, and $\mathrm{SNR} = 15$ dB.

and finally using it as input to the OMP-MMV algorithm to solve (73).

It is evident that the LS solution provides the lowest NMSE after $T = 35$ training frames with the cost of high computational complexity, i.e., $\mathcal{O}(L^3 N_{\mathrm{T}}^3 T^3)$. However, to be able to consider large number of training frames, we restrict the wideband mmWave channel to be static for longer time, i.e., longer coherence time is required. The techniques SVT and TSSR are

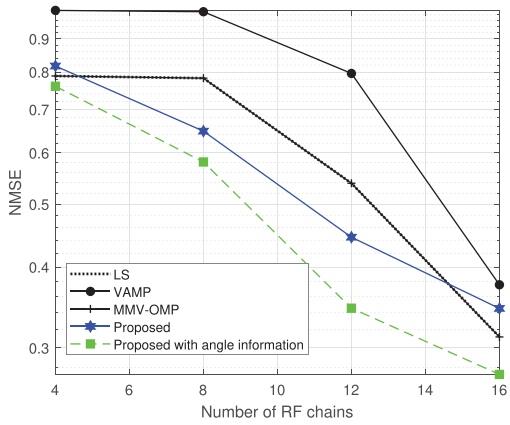

Fig. 10. NMSE performance as a function of the number of RX RF chains M_{R} for $N_{\mathrm{R}} = 8$, $N_{\mathrm{R}} = 32$, $M_{\mathrm{R}}^e = 32$, $L = 4$, $N_p = 6$, $T = 20$, and $\mathrm{SNR} = 15$ dB.

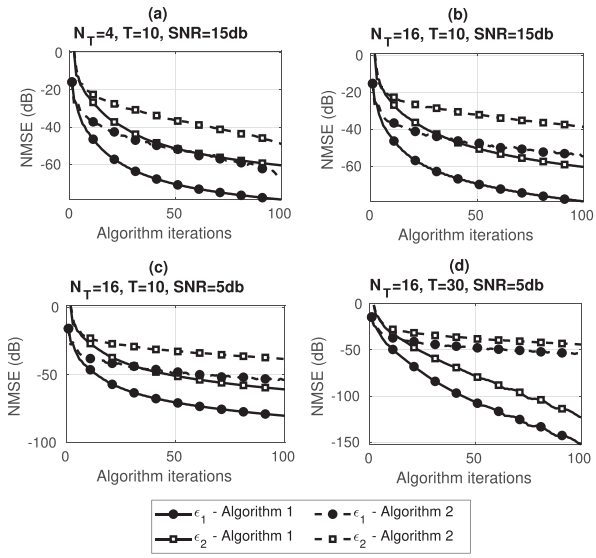

Fig. 11. NMSE performance as a function of the number of RX RF antennas M_{R} for $N_{\mathrm{R}} = 32$, $M_{\mathrm{R}} = 16$, $M_{\mathrm{R}}^e = 32$, $L = 4$ and $N_p = 6$.

not able to converge under this scenario, while OMP, VAMP and CoSaMP converge slowly compared to LS. The OMP-MMV and the proposed technique (Algorithm 1) achieve similar NMSE with the LS, up to 35 frames. For $T > 35$, OMP-MMV outperforms Algorithm 1.

In Figs. 6, 7, 8, 9, and 10 we plot the NMSE performance as functions of the SNR, the number of the channel propagation paths N_p, the number of the wideband mmWave channel delay taps L, the number of the transmitting antenna elements N_{T}, and the number of the RX RF chains, respectively. We actually focus on comparing the performances among the LS, VAMP, OMP-MMV, and the proposed Algorithms 1 and 2, since the OMP, SVT, CoSaMP, and TSSR approaches results in higher NMSE, as showcased in

Finally, in Fig. 11, we plot NMSE versus the number of RX RF chains M_{R} for $N_{\mathrm{T}} = 8$, $N_{\mathrm{R}} = 31$, $M_{\mathrm{R}}^e = 32$, and $\mathrm{SNR} = 5$ dB. It is shown that both proposed algorithms yield the best performance with Algorithm 2 being the best algorithm overall. For all considered techniques, NMSE improves as M_{R} increases.

TABLE IV
RUN TIME IN SECONDS FOR THE CONSIDERED WIDEBAND MMWAVE MIMO CHANNEL ESTIMATION ALGORITHMS

	OMP	VAMP	SVT	Algorithm 1	GD-based Algorithm 1
$N_{\mathrm{T}}=4,\ N_{\mathrm{R}}=32,\ T=70$	0.5085	6.0917	0.0419	15.1963	**1.1699**
$N_{\mathrm{T}}=4,\ N_{\mathrm{R}}=64,\ T=70$	1.0993	29.9625	0.0744	147.8875	**5.9474**
$N_{\mathrm{T}}=8,\ N_{\mathrm{R}}=32,\ T=70$	0.8079	17.5677	0.0407	67.2175	**1.7646**
$N_{\mathrm{T}}=8,\ N_{\mathrm{R}}=64,\ T=70$	1.3836	79.4238	0.0750	535.5001	**8.3759**
$N_{\mathrm{T}}=4,\ N_{\mathrm{R}}=32,\ T=120$	0.9974	10.5556	0.0663	39.3253	**2.8179**

C. Algorithmic Convergence and Computational Complexity

In Fig. 11, the convergence curves of the proposed iterative Algorithms 1 and 2 are illustrated. Particularly, we plot the ADMM error concerning the estimation of the received training signal matrix $\mathbf{Y}$, which is defined as:

$$\epsilon_1 \triangleq \frac{\|\mathbf{X}^{(I_{\max})}-\mathbf{Y}^{(I_{\max})}\|^2}{\|\mathbf{Y}^{(I_{\max})}\|^2}, \tag{75}$$

where $\mathbf{Y}^{(I_{\max})}$ and $\mathbf{X}^{(I_{\max})}$ are given by (29) and (30) after $I_{\max}$ iterations, respectively. The GD residual error is expressed as:

$$\epsilon_2 \triangleq \frac{\|\boldsymbol{\beta}^{(i+1)}-\boldsymbol{\beta}^{(i)}\|^2}{\|\boldsymbol{\beta}^{(i)}\|^2}, \tag{76}$$

expecting that both they will go to zero after $I_{\max}$ iterations. We have considered different scenarios for investigating the convergence rates of ϵ_1 and ϵ_2. The first one, shown in Fig. 11(a), assumes only $N_{\mathrm{T}}=4$ transmitting antenna elements and high SNR. As shown in this subfigure, the convergence rates are fast for both proposed algorithms, particularly, e.g. $\epsilon_1 < 1e-6$ is achieved after 14 algorithmic iterations. The same trend happens when N_T increases, as shown in Fig. 11(b), and when lower SNR values are used, as considered in Fig. 11(c). In Fig. 11(d), the impact of the number of propagation channel paths N_p is depicted. It is evident form the latter subfigures that for Algorithm 1 both errors converge very fast to 0. Algorithm 2 exhibits faster convergence, however its error floor is higher than that of Algorithm 1. This error floor depends on the reliability of the available angle information. In the formulation of Algorithm 2, this information was expressed based on the number of the non-zero values of $\mathbf{K}_3$. In Fig. 12, we have used a $\mathbf{K}_3$ such that $\|\mathbf{K}_3\|=10$. Increasing the latter value will result in lower error floors for Algorithm 2.

In Fig. 12, we take a closer investigation of the NMSE performance gap between Algorithm 1 and its low-complexity approximate Algorithm 2 for $N_{\mathrm{R}}=32$. It is shown in the figure that the performance loss of GD-based Algorithm 1 is negligible for all considered values of the maximum number of ADMM iterations $I_{\max}$. It can be also observed that both proposed algorithms converge to the steady state after 30 ADMM iterations, i.e., for $I_{\max}>30$ the NMSE is constant. It is noted that Algorithm 2, which is based on a GD technique, converges very fast when the involved matrix is strongly diagonal dominant.

Finally, in Table IV, we compare the algorithmic run time in seconds of the two proposed algorithms and representative ones from the state-of-the-art. It is noted that the run time values depend on the system specifications where the algorithms are executed, however, Table IV provides their comparison for the sketched parameters when executed in in a desktop computer with Intel(R) Core(TM) i3-8350K CPU at 4.00 GHz and with a 32 GB DDR4 2666 MHz random access memory. It is shown that OMP and SVT techniques yield the smallest elapsed time, while Algorithm 1 is the slowest for all system configurations. However, a fair comparison should focus on techniques having equivalent estimation performance. Hence, comparing VAMP and Algorithm 2, reveals that the proposed low-complexity channel estimation technique exhibits the faster execution time, rendering it suitable for latency demanding real-time applications.

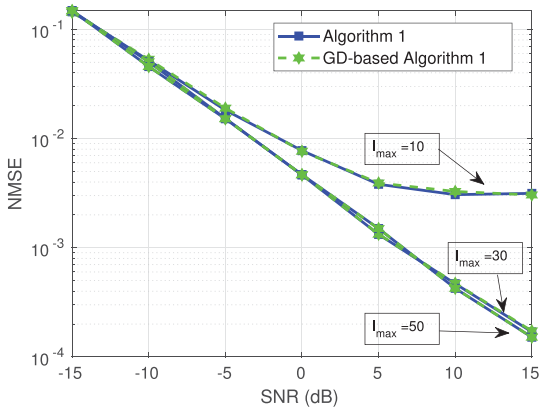

Fig. 12. NMSE performance as a function of SNR for the proposed Algorithm 1 and the GD-based low-complexity approximation for three different values of the maximum number of ADMM iterations Imax. The parameters values $T=35$, $N_{\mathrm{T}}=4$, $N_{\mathrm{R}}=32$, $M_{\mathrm{R}}^{e}=32$, $M_{\mathrm{R}}=24$ have been used.

VII. CONCLUSIONS

In this paper, we have presented iterative ADMM-based algorithms for wideband mmWave MIMO channel estimation that exploit jointly the channel's low rank and beamspace sparsity in order to provide more accurate channel recovery, especially for short beam training intervals. The presented algorithms capitalize on our proposed analog combining architecture that includes an extended analog combiner incorporating a random spatial sampling structure placed before the input of the analog received signals to the digital component of the HBF receiver. It has been shown through extensive simulation results that the proposed algorithms exhibit improved performance in terms of MSE for channel estimation requiring only short beam training lengths and when operating under high noise conditions. Interestingly, in scenarios with small numbers of receiver RF chains and severe noise, angle information can be exploited to boost channel

estimation performance. Extension to the case of soft angular information and HBF transmission, as well as the solution of the optimization framework for the case where beam squint is present are left for future works.

APPENDIX

A. Proof of Proposition 1

Based on (12) and the properties of the rank operator, we have that:

$$\text{rank}(\mathbf{Q}) \leq \min\left(\text{rank}(\bar{\mathbf{Z}}), \text{rank}(\mathbf{I}_L \otimes \mathbf{D}_{\text{T}}^H)), \text{rank}(\bar{\mathbf{\Psi}})\right), \quad (77)$$

where $\mathbf{Q} \triangleq \bar{\mathbf{Z}}(\mathbf{I}_L \otimes \mathbf{D}_{\text{T}}^H)\bar{\mathbf{\Psi}}$. The rank of the concatenated wideband channel matrix $\bar{\mathbf{Z}}$ is upper bounded by $\text{rank}(\bar{\mathbf{Z}}) \leq N_p$ based on (2). Additionally, based on the properties of the Kronecker product, $\text{rank}(\mathbf{I}_L \otimes \mathbf{D}_{\text{T}}^H) = \text{rank}(I_L)\text{rank}(\mathbf{D}_{\text{T}}) = LN_{\text{T}}$. The rank of L concatenated Toeplitz matrices $\mathbf{\Psi} \in \mathbb{C}^{N_{\text{T}} \times T}$ is $\text{rank}(\bar{\mathbf{\Psi}}) \leq LN_{\text{T}}$. Putting all together, we have that $\text{rank}(\mathbf{Q}) \leq \min(N_p, LN_{\text{T}})$.

B. Derivation of (12)

The matrix product given by (12) can be obtained by reorganizing the columns and rows of the involved matrices. In particular, the finite summation of (11) can be expressed as the product of two concatenated matrices, as follows:

$$\tilde{\mathbf{Y}} = \sum_{k=1}^{N_{\text{T}}} \tilde{\mathbf{H}}_k \tilde{\mathbf{\Psi}}_k = \left[\tilde{\mathbf{H}}_1 \cdots \tilde{\mathbf{H}}_{N_{\text{T}}}\right] \begin{bmatrix} \tilde{\mathbf{\Psi}}_1 \\ \vdots \\ \tilde{\mathbf{\Psi}}_{N_{\text{T}}} \end{bmatrix} = \bar{\mathbf{H}}\bar{\mathbf{\Psi}} \quad (78)$$

where $\bar{\mathbf{H}} \in \mathbb{C}^{N_{\text{R}} \times LN_{\text{T}}}$ is given by

$$\begin{aligned} \bar{\mathbf{H}} &\triangleq \left[\tilde{\mathbf{H}}_1 \cdots \tilde{\mathbf{H}}_{N_{\text{T}}}\right] \\ &= \left[\left[\mathbf{h}_1(0) \cdots \mathbf{h}_1(L-1)\right] \right. \\ &\quad \left. \cdots \left[\mathbf{h}_{N_{\text{T}}}(0) \cdots \mathbf{h}_{N_{\text{T}}}(L-1)\right]\right] \\ &= \left[\left[\mathbf{h}_1(0) \cdots \mathbf{h}_{N_{\text{T}}}(0)\right] \right. \\ &\quad \left. \cdots \left[\mathbf{h}_{N_{\text{T}}}(L-1) \cdots \mathbf{h}_{N_{\text{T}}}(L-1)\right]\right] \\ &= \left[\mathbf{H}(0) \cdots \mathbf{H}(L-1)\right], \end{aligned}$$

while $\bar{\mathbf{\Psi}} \in \mathbb{C}^{LN_{\text{T}} \times T}$ is defined as

$$\bar{\mathbf{\Psi}} = \begin{bmatrix} \tilde{\mathbf{\Psi}}_1 \\ \vdots \\ \tilde{\mathbf{\Psi}}_{N_{\text{T}}} \end{bmatrix} = \begin{bmatrix} \text{toeplitz}(\mathbf{s}_1) \\ \vdots \\ \text{toeplitz}(\mathbf{s}_{N_{\text{T}}}) \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} \text{toeplitz}_1(\mathbf{s}_1) \\ \vdots \\ \text{toeplitz}(\mathbf{s}_{N_{\text{T}}}) \end{bmatrix} \\ \vdots \\ \begin{bmatrix} \text{toeplitz}_L(\mathbf{s}_1) \\ \vdots \\ \text{toeplitz}_L(\mathbf{s}_{N_{\text{T}}}) \end{bmatrix} \end{bmatrix},$$

where $\text{toeplitz}_\ell(\mathbf{s}_k)$ represents the ℓ-th row of an $L \times T$ Toeplitz matrix.

C. Derivation of (16)

We first consider the low AWGN regime, i.e., $\sigma_n^2 \ll 1$. In this case, the achievable ASE can be approximated by [41]

$$\begin{aligned} C_{\text{HBF}} &= \log_2 \det\left(\mathbf{I}_{M_{\text{R}}} + \frac{1}{\sigma_n^2 N_{\text{T}}} \mathbf{W}_{\text{RF}}^H \mathcal{Y}\mathcal{Y}^H \mathbf{W}_{\text{RF}}\right) && (79) \\ &= \log_2 \det\left(\mathbf{I}_{N_{\text{R}}} + \frac{1}{\sigma_n^2 N_{\text{T}}} \mathcal{Y}\mathcal{Y}^H \mathbf{W}_{\text{RF}} \mathbf{W}_{\text{RF}}^H\right) && (80) \\ &\approx \log_2 \det\left(\frac{1}{\sigma_n^2 N_{\text{T}}} \mathcal{Y}\mathcal{Y}^H \mathbf{W}_{\text{RF}} \mathbf{W}_{\text{RF}}^H\right) && (81) \\ &= \log_2 \frac{1}{\sigma_n^2 N_{\text{T}}} + \log_2 \det\left(\mathcal{Y}\mathcal{Y}^H\right) \\ &\quad + \log_2 \det\left(\mathbf{W}_{\text{RF}} \mathbf{W}_{\text{RF}}^H\right) && (82) \\ &\leq \log_2 \frac{1}{\sigma_n^2 N_{\text{T}}} + \log_2 \det\left(\mathcal{Y}\mathcal{Y}^H\right), && (83) \end{aligned}$$

where the last inequality holds since the analog combining matrix $\mathbf{W}_{\text{RF}}$ is composed by unit norm entries, which yields $\det\left(\mathbf{W}_{\text{RF}} \mathbf{W}_{\text{RF}}^H\right) \geq 1$.

Similarly, for the high AWGN regime, the achievable ASE can be upper bounded as follows:

$$\begin{aligned} C_{\text{HBF}} &\approx \det\left(\frac{1}{\sigma_n^2 N_{\text{T}}} \mathbf{W}_{\text{RF}}^H \mathcal{Y}\mathcal{Y}^H \mathbf{W}_{\text{RF}}\right) && (84) \\ &= \frac{1}{\sigma_n^2 N_{\text{T}}} + \det\left(\mathcal{Y}\mathcal{Y}^H\right) + \det\left(\mathbf{W}_{\text{RF}} \mathbf{W}_{\text{RF}}^H\right) && (85) \\ &\leq \frac{1}{\sigma_n^2 N_{\text{T}}} + \det\left(\mathcal{Y}\mathcal{Y}^H\right). && (86) \end{aligned}$$

It is noted that the expressions (83) and (86) can be seen as the low and high AWGN approximations, respectively, of the achievable ASE of a fully Digital BeamForming (DBF) system (i.e., $M_{\text{R}} = N_{\text{R}}$). In this case, ASE can be computed as

$$\mathcal{C}_{\text{DBF}} = \log_2 \det\left(\mathbf{I}_{N_{\text{R}}} + \frac{1}{\sigma_n^2 N_{\text{T}}} \mathcal{Y}\mathcal{Y}^H\right). \quad (87)$$

D. Derivation of (39)

Given (38), it can be straightforwardly obtained that:

$$\begin{aligned} &\mathbf{\Omega} \circ \mathbf{X} - \mathbf{R}_\Omega - \mathbf{V}_1 - \rho(\mathbf{Y} - \mathbf{X}) \\ &\quad - \mathbf{V}_2 - \rho(\mathbf{C} - \mathbf{X} + \mathbf{A}\bar{\mathbf{Z}}\mathbf{B}) = \mathbf{0} && (88) \\ &\Rightarrow \mathbf{\Omega} \circ \mathbf{X} + 2\rho\mathbf{X} = \mathbf{R}_\Omega + \mathbf{V}_1 + \rho\mathbf{Y} + \mathbf{V}_2 + \rho\mathbf{C} + \rho\mathbf{A}\bar{\mathbf{Z}}\mathbf{B} && (89) \\ &\Rightarrow \sum_{j=1}^{N_{\text{R}}} \mathbf{E}_{jj} \mathbf{X} \,\text{diag}([\mathbf{\Omega}]_j) + 2\rho\mathbf{X} \\ &\quad = \mathbf{R}_\Omega + \mathbf{V}_1 - \rho\mathbf{Y} - \mathbf{V}_2 + \rho\mathbf{C} - \rho\mathbf{A}\bar{\mathbf{Z}}\mathbf{B} && (90) \\ &\Rightarrow \text{vec}\left(\sum_{j=1}^{N_{\text{R}}} \mathbf{E}_{jj} \mathbf{X} \,\text{diag}([\mathbf{\Omega}]_j) + 2\rho\mathbf{X}\right) \\ &\quad = \text{vec}\left(\mathbf{R}_\Omega + \mathbf{V}_1 - \rho\mathbf{Y} - \mathbf{V}_2 + \rho\mathbf{C} - \rho\mathbf{A}\bar{\mathbf{Z}}\mathbf{B}\right) && (91) \end{aligned}$$

$$\Rightarrow \sum_{j=1}^{N_{\mathrm{R}}} \mathrm{diag}([\mathbf{\Omega}]_j)^T \otimes \mathbf{E}_{jj}\mathrm{vec}(\mathbf{X}) + 2\rho\mathrm{vec}(\mathbf{X})$$

$$= \mathrm{vec}\left(\mathbf{R}_\Omega + \mathbf{V}_1 - \rho\mathbf{Y} - \mathbf{V}_2 + \rho\mathbf{C} - \rho\mathbf{A}\bar{\mathbf{Z}}\mathbf{B}\right), \qquad (92)$$

where from the last equality one easily obtains (39).

REFERENCES

[1] T. S. Rappaport *et al.*, "Millimeter wave mobile communications for 5G cellular: It will work!," *IEEE Access*, vol. 1, pp. 335–349, May 2013.

[2] V. Venkateswaran and A.-J. van der Veen, "Analog beamforming in MIMO communications with phase shift networks and online channel estimation," *IEEE Trans. Signal Process.*, vol. 58, no. 8, pp. 4131–4143, Aug. 2010.

[3] 3GPP, "Study on New Radio (NR) Access Technology- Physical Layer Aspects- Release 14," 3GPP, Sophia Antipolis Cedex, France, TR 38.802, 2017.

[4] A. F. Molisch *et al.*, "Hybrid beamforming for massive MIMO: A survey," *IEEE Commun. Mag.*, vol. 55, no. 9, pp. 134–141, Sep. 2017.

[5] 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. Topics Signal Process.*, vol. 10, no. 3, pp. 436–453, Apr. 2016.

[6] 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.*, New Orleans, LA, USA, 2017, pp. 6493–6497.

[7] A. Alkhateeb, O. E. 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.

[8] 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, Jan. 2016.

[9] J. Lee, G. T. Gil, and Y. H. Lee, "Channel estimation via orthogonal matching pursuit for hybrid MIMO systems in millimeter wave communications," *IEEE Trans. Commun.*, vol. 64, no. 6, pp. 2370–2386, Jun. 2016.

[10] G. C. Alexandropoulos and S. Chouvardas, "Low complexity channel estimation for millimeter wave systems with hybrid A/D antenna processing," in *Proc. IEEE GLOBECOM*, Washington, DC, USA, Dec. 2016, pp. 1–6.

[11] 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.

[12] G. C. Alexandropoulos, "Position aided beam alignment for millimeter wave backhaul systems with large phased arrays," in *Proc. IEEE 7th Int. Workshop Computational Adv. Multi-Sensor Adaptive Process.*, Dec. 2017, pp. 1–5.

[13] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[14] T. T. Cai and L. Wang, "Orthogonal matching pursuit for sparse signal recovery with noise," *IEEE Trans. Inf. Theory*, vol. 57, no. 7, pp. 4680–4688, Jul. 2011.

[15] D. Lee, S.-J. Kim, and G. B. Giannakis, "Channel gain cartography for cognitive radios leveraging low rank and sparsity," *IEEE Trans. Wireless Commun.*, vol. 16, no. 9, pp. 5953–5966, Sep. 2017.

[16] X. Li, J. Fang, H. Li, and P. Wang, "Millimeter wave channel estimation via exploiting joint sparse and low-rank structures," *IEEE Trans. Wireless Commun.*, vol. 17, no. 2, pp. 1123–1133, Feb. 2018.

[17] E. Vlachos, G. C. Alexandropoulos, and J. Thompson, "Massive MIMO channel estimation for millimeter wave systems via matrix completion," *IEEE Signal Process. Lett.*, vol. 25, no. 11, pp. 1675–1679, Nov. 2018.

[18] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, pp. 1–122, Jan. 2011.

[19] J. F. Cai, E. J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM J. Opt.*, vol. 20, no. 4, pp. 1956–1982, 2010.

[20] J. Mo, P. Schniter, and R. W. Heath, Jr., "Channel estimation in broadband millimeter wave MIMO systems with few-bit ADCs," *IEEE Trans. Signal Process.*, vol. 66, no. 5, pp. 1141–1154, Mar. 2018.

[21] A. Forenza, D. J. Love, and R. W. Heath Jr., "Simplified spatial correlation models for clustered MIMO channels with different array configurations," *IEEE Trans. Veh. Technol.*, vol. 56, no. 4, pp. 1924–1934, Jul. 2007.

[22] Z. Gao, L. Dai, Z. Wang, and S. Chen, "Spatially common sparsity based adaptive channel estimation and feedback for FDD massive MIMO," *IEEE Trans. Signal Process.*, vol. 63, no. 23, pp. 6169–6183, Dec. 2015.

[23] K. Venugopal, N. González-Prelcic, and R. W. Heath, "Optimality of frequency flat precoding in frequency selective millimeter wave channels," *IEEE Wireless Commun. Lett.*, vol. 6, no. 3, pp. 330–333, Jun. 2017.

[24] J. Rodríguez-Fernández, N. González-Prelcic, K. Venugopal, and R. W. Heath, "Frequency-domain compressive channel estimation for frequency-selective hybrid millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 5, pp. 2946–2960, May 2018.

[25] A. M. Sayeed, "Deconstructing multiantenna fading channels," *IEEE Trans. Signal Process.*, vol. 50, no. 10, pp. 2563–2579, Oct. 2002.

[26] H. Ghauch, T. Kim, M. Bengtsson, and M. Skoglund, "Subspace estimation and decomposition for large millimeter-wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 528–542, Apr. 2016.

[27] P. A. Eliasi, S. Rangan, and T. S. Rappaport, "Low-rank spatial channel estimation for millimeter wave cellular systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 5, pp. 2748–2759, May 2017.

[28] R. Tibshirani, "Regression shrinkage and selection via the Lasso," *J. Roy. Statist. Soc., Ser. B*, vol. 58, no. 1, pp. 267–288, 1996.

[29] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 4th ed. Baltimore, MD, USA: Johns Hopkins Univ. Press, 2013.

[30] H. Simon, "The Lanczos algorithm with partial reorthogonalization," *Math. Comput.—Math. Comput.*, vol. 42, pp. 115–115, Jan. 1984.

[31] E. Vlachos and K. Berberidis, "Adaptive completion of the correlation matrix in wireless sensor networks," in *Proc. 24th Eur. Signal Process. Conf.*, Budapest, Hungary, Sep. 2016, pp. 1403–1407.

[32] H. A. V. D. Vorst, *Iterative Krylov Methods for Large Linear Systems*. Cambridge, U.K.: Cambridge Univ. Press, 2003.

[33] Z. Guo, X. Wang, and W. Heng, "Millimeter-wave channel estimation based on 2-D beamspace MUSIC method," *IEEE Trans. Wireless Commun.*, vol. 16, no. 8, pp. 5384–5394, Aug. 2017.

[34] J. Zhang and M. Haardt, "Channel estimation and training design for hybrid multi-carrier mmWave massive MIMO systems: The beamspace ESPRIT approach," in *Proc. 25th Eur. Signal Process. Conf.*, Kos, Greece, Sep. 2017, pp. 385–389.

[35] B. Wang, F. Gao, S. Jin, H. Lin, and G. Y. Li, "Spatial- and frequency-wideband effects in millimeter-wave massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 66, no. 13, pp. 3393–3406, Jul. 2018.

[36] B. Wang *et al.*, "Beam squint and channel estimation for wideband mmWave massive MIMO-OFDM systems," *CoRR*, vol. abs/1903.01340, 2019. [Online]. Available: http://arxiv.org/abs/1903.01340

[37] E. Ghadimi, A. Teixeira, I. Shames, and M. Johansson, "On the optimal step-size selection for the alternating direction method of multipliers," in *Proc. NecSys*, Santa Barbara, CA, USA, Sep. 2012, pp. 139–144.

[38] P. Schniter, S. Rangan, and A. K. Fletcher, "Vector approximate message passing for the generalized linear model," in *Proc. Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, USA, Nov. 2016, pp. 1525–1529.

[39] D. Needell and J. A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Commun. ACM*, vol. 53, pp. 93–100, Dec. 2010.

[40] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," *Proc. Nat. Acad. Sci. USA*, vol. 106, no. 45, pp. 18 914–18 919, 2009. [Online]. Available: https://www.pnas.org/content/106/45/18914

[41] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, MA, USA: Cambridge Univ. Press, 2005.