Hybrid Beamforming with Random Analog Sampling for Wideband Channel Estimation in Millimeter Wave Massive MIMO Systems

Evangelos Vlachos[1], George C. Alexandropoulos[2], and John Thompson[1]
[1]Institute for Digital Communications, University of Edinburgh, Edinburgh, EH9 3JL, UK
[2]Department of Informatics and Telecommunications, National and Kapodistrian University of Athens, Greece
emails: e.vlachos@ed.ac.uk, alexandg@di.uoa.gr, j.s.thompson@ed.ac.uk

***Abstract*—Hybrid analog and digital BeamForming (HBF) transceiver architectures realizing directive communication over large bandwidths in millimeter Wave (mmWave) massive Multiple Input Multiple Output (MIMO) systems require the availability of accurate channel estimation. This is, however, a challenging task mainly due to the short channel coherence time and the hardware limitations imposed by HBF architectures. In this paper, we capitalize on recent matrix completion tools and develop a novel wideband channel estimation technique for mmWave massive MIMO systems with HBF reception. The proposed iterative algorithm exploits jointly the low rank and beamspace sparsity properties to provide more accurate recovery, especially for short beam training intervals. We introduce a novel analog combining architecture that includes a random subsampling step before the input of the analog received signals to the digital component of the HBF receiver. This step supports the proposed estimation technique in providing the sampling set of measurements for recovering the unknown channel matrix. The impact of various system parameters on the effectiveness of the designed algorithm and the performance improvement of our technique over representative state-of-the-art ones is demonstrated through indicative simulation results.**

***Index Terms*—Hybrid beamforming, millimeter wave, matrix completion, massive MIMO, wideband channel estimation.**

I. Introduction

A promising enabler for the highly demanding data rate requirements of fifth Generation (5G), and beyond, broadband wireless networks is the millimeter Wave (mmWave) frequency band that offers large bandwidths for wireless communication [1]. Although, communication signals at this band are expected to experience severe pathloss, their mm-level wavelengths enable many antenna elements to be packed in small-sized terminals, which allows for highly directional beamforming. The 5G New Radio (NR) technology [2] will support Hybrid BeamForming (HBF) [3] which capitalizes on both analog and digital signal processing. It enables a large number of antenna elements to be connected to a much smaller number of Radio Frequency (RF) chains.

Near-optimal HBF performance in mmWave massive Multiple Input Multiple Output (MIMO) systems necessitates reliable Channel State Information (CSI) knowledge. This knowledge is, however, very challenging to acquire in practice due to the very large numbers of transceiver antenna elements and the high channel variability [4]. Very recently in [5]–[7], the sparsity and low rank properties of certain narrowband wireless channels were jointly exploited for efficient CSI estimation. However, to the authors' best of knowledge, in the open technical literature of wideband channel estimation for mmWave massive MIMO systems, the sparsity and low rank channel properties have been exploited either independently or in a sequential manner, but not jointly. Targeting narrowband mmWave MIMO channel estimation with HBF transceivers, a two-stage procedure (one stage per channel property) was introduced in [6]. In [8], wideband mmWave MIMO channels with frequency selectivity were studied and a CSI estimation technique exploiting channel's sparsity in both the time and frequency domains was presented.

In this paper, we focus on the uplink of mmWave massive MIMO systems with HBF reception. We introduce a novel analog combining architecture that includes 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 the measurement collection during channel sounding. For the estimation of the massive MIMO channel matrix, we formulate a multi-objective optimization problem which capitalizes on the sparsity of the channel matrix in the beamspace domain, as well as the low rank property of the matrix with the received measurements. The designed iterative channel estimation algorithm is based on the Alternating Direction Method of Multipliers (ADMM) [9].

Notation: Fonts α, $\mathbf{a}$, and $\mathbf{A}$ denote a scalar, a vector, and a matrix, respectively. $\mathbf{A}^T$, $\mathbf{A}^*$, $\mathbf{A}^H$, and $\|\mathbf{A}\|_F$ represent $\mathbf{A}$'s transpose, conjugate transpose, Hermitian transpose, and Frobenius norm. $\mathbf{I}_N$ denotes the $N \times N$ identity matrix. Operands $\circ$ and $\otimes$ denote the matrix Hadamard and Kronecker products, respectively, $\mathrm{vec}(\cdot)$ concatenates the columns of a matrix into a vector, and $\mathrm{unvec}(\cdot)$ is the inverse operation. $\|\mathbf{A}\|_* \triangleq \sum_{i=1}^{r} \sigma_i$ is the nuclear norm with σ_i's being the r singular values of $\mathbf{A}$; $\|\mathbf{A}\|_1 \triangleq \max_{1\leq j\leq N} \sum_{i=1}^{M} |[\mathbf{A}]_{ij}|$ ($\mathbf{A} \in \mathbb{C}^{M\times N}$) with $[\mathbf{A}]_{ij}$ denoting $\mathbf{A}$'s (i,j)-th element.

E. Vlachos and J. Thompson gratefully acknowledge funding of this work by EPSRC grant EP/P000703/1.

978-1-5386-6528-2/19/$31.00 ©2019 IEEE

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 [10], 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 $L_{\mathrm{R}} \ll N_{\mathrm{R}}$ RF chains. Due to this hardware configuration, the TX is capable of digitally precoding up to N_{T} independent signals, each from one RF chain. On the other hand, we assume that RX is equipped with any of the available HBF architectures [3] supporting both analog and digital combining. As will be described in the sequel, the proposed framework applies also to HBF transmission, but we leave the full details of this extension for future work. The considered mmWave system can realize a wireless communication link comprising of $N_s \leq \min(N_{\mathrm{T}}, L_{\mathrm{R}})$ independent information data streams.

We adopt 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 the 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 is the gain of the k-th channel path drawn from the complex Gaussian distribution $\mathcal{CN}(0, 1/2)$. In addition, $p(\tau_k)$ is the band-limited pulse shaping filter response at τ_k. The vectors $\mathbf{a}_{\mathrm{T}}(\phi_k) \in \mathbb{C}^{N_{\mathrm{T}} \times 1}$ and $\mathbf{a}_{\mathrm{R}}(\theta_k) \in \mathbb{C}^{N_{\mathrm{R}} \times 1}$ represent the TX and RX normalized array response vectors, respectively, which are expressed as described in [4, Sec. II.C] for uniform linear arrays. Scalars ϕ_k and θ_k are the physical angles of arrival and departure, respectively, which are generated according to the Laplace distribution [11].

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 [12], which is defined as

$$\mathbf{H}(\ell) = \mathbf{D}_{\mathrm{R}} \mathbf{Z}(\ell) \mathbf{D}_{\mathrm{T}}^H, \quad (2)$$

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. We 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 (2).

III. Proposed Channel Estimation

In this section, we first introduce our novel analog combining architecture for measurement collection during the channel sounding process. Then, we present our new mathematical formulation for wideband mmWave MIMO channel estimation.

A. Proposed Analog Combining Architecture

To estimate the wideband mmWave MIMO channel, the TX utilizes the $N_{\mathrm{T}} \times T$ training matrix $\mathbf{S}$ that contains the symbols transmitted from the N_{T} TX antennas over T training periods. We assume that the channel remains static during each T-symbol training period. The baseband equivalent of the received signal $\mathbf{R} \in \mathbb{C}^{N_{\mathrm{R}} \times T}$ at the RX's antenna elements using $\mathbf{H}(\ell)$'s representation in (2) $\forall \ell$ can be expressed as

$$\mathbf{R} \triangleq \sum_{k=1}^{N_{\mathrm{T}}} [\mathbf{h}_k(0)\, \mathbf{h}_k(1) \ldots \mathbf{h}_k(L-1)] \mathbf{\Psi}_k + \mathbf{N} \quad (3)$$

$$= \sum_{\ell=0}^{L-1} \mathbf{H}(\ell) \bar{\mathbf{\Psi}}(\ell) + \mathbf{N} \quad (4)$$

$$= \sum_{\ell=0}^{L-1} \mathbf{D}_{\mathrm{R}} \mathbf{Z}(\ell) \mathbf{D}_{\mathrm{T}}^H \bar{\mathbf{\Psi}}(\ell) + \mathbf{N} = \underbrace{\mathbf{D}_{\mathrm{R}} \bar{\mathbf{Z}} \mathbf{B}}_{\triangleq \tilde{\mathbf{R}}} + \mathbf{N}, \quad (5)$$

where $\mathbf{N} \in \mathbb{C}^{N_{\mathrm{R}} \times T}$ is the complex-valued Additive White Gaussian Noise (AWGN) matrix with independent and identically distributed entries having zero mean and variance σ_n^2. In (4) and (5), we have used the following matrix definitions:

$$\bar{\mathbf{Z}} \triangleq [\mathbf{Z}(0)\, \mathbf{Z}(1)\, \cdots\, \mathbf{Z}(L-1)] \in \mathbb{C}^{N_{\mathrm{R}} \times L N_{\mathrm{T}}}, \quad (6)$$

as well as

$$\bar{\mathbf{\Psi}}(\ell) \triangleq [\boldsymbol{\psi}_1(\ell)\, \boldsymbol{\psi}_2(\ell)\, \cdots\, \boldsymbol{\psi}_{N_{\mathrm{T}}}(\ell)]^T \in \mathbb{C}^{N_{\mathrm{T}} \times T}, \quad (7)$$

$$\mathbf{B} \triangleq [\mathbf{D}_{\mathrm{T}}^H \bar{\mathbf{\Psi}}(0)\, \mathbf{D}_{\mathrm{T}}^H \bar{\mathbf{\Psi}}(1)\ \cdots\ \mathbf{D}_{\mathrm{T}}^H \bar{\mathbf{\Psi}}(L-1)]^T \in \mathbb{C}^{L N_{\mathrm{T}} \times T}, \quad (8)$$

where $\boldsymbol{\psi}_k(\ell)$ with $k = 1, 2, \ldots, N_{\mathrm{T}}$ represents the ℓ-th row of the Toeplitz matrix $\mathbf{\Psi}_k \triangleq \mathrm{toeplz}(\mathbf{s}_k) \in \mathbb{C}^{L \times T}$, which is obtained from the k-th row of the training symbol matrix $\mathbf{S}$. In the sequel, we focus on the estimation of the matrix $\bar{\mathbf{Z}}$ including the unknown mmWave MIMO channel gains.

To receive the training symbols intended for channel estimation with any of the standard HBF architectures [3], the RX utilizes its $N_{\mathrm{R}} \times L_{\mathrm{R}}$ analog combiners $\mathbf{W}_{\mathrm{RF}}$ available to its predefined beam codebook $\mathcal{W}^{N_{\mathrm{R}} \times L_{\mathrm{R}}}$. A network of phase shifters is usually considered for analog combining with all available $\mathbf{W}_{\mathrm{RF}}$ combiners having unit magnitude elements. After applying an analog combiner from the available codebook to the matrix $\mathbf{R}$ in (5), the baseband received signal at the L_{R} inputs of the RX's RF chains is given by $\mathbf{W}_{\mathrm{RF}}^H \mathbf{R} \in \mathbb{C}^{L_{\mathrm{R}} \times T}$.

Usually the number of receive RF chains is smaller than the number of receiving antennas, i.e., $L_{\mathrm{R}} < N_{\mathrm{R}}$, to provide low implementation complexity and reasonable power consumption for mmWave HBF reception. However, this reduction in RF chain hardware introduces a loss of information compared to the fully digital RX case, since the received signals lie in a smaller vector subspace. This is particularly true for the achievable spectral efficiency, which is related to the number of the eigenvalues of the effective communication channel.

To overcome this loss of information, in Fig. 1, we show the block diagram of our novel analog combining architecture. It comprises of an extended analog combiner with L_{R}^e outputs

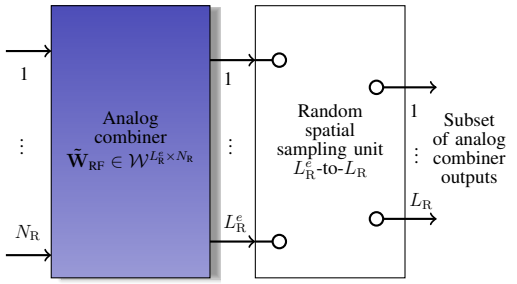

Fig. 1. Block diagram of the proposed analog combining architecture for wideband mmWave MIMO channel estimation with HBF reception. It includes an extended analog combiner of L^e_{R} outputs and a random sub-sampling unit.

and a random spatial sub-sampling unit with $L_{\text{R}} < L^e_{\text{R}}$ outputs. The inputs to this architecture are the analog received signals at the N_{R} RX antenna elements. From the L^e_{R} outputs of the analog combiner, a switching block will select *randomly* only L_{R} of them to forward to the RF chains. The proposed HBF architecture retains the low complexity and low power consumption characteristics of standard HBF reception, but also improves the estimation performance via the *random spatial sub-sampling strategy*. This is accomplished by only increasing the analog components of the RX,, i.e., by adding $L^e_{\text{R}} - L_{\text{R}}$ more phase shifters. Also, mmWave switches capable of switching at nanosecond or sub-nanosecond speeds can have very low power consumption [13].

In mathematical terms, the standard analog combiner $\mathbf{W}_{\text{RF}} \in \mathcal{W}^{N_{\text{R}} \times L_{\text{R}}}$ is replaced by the extended analog combiner $\tilde{\mathbf{W}}_{\text{RF}} \in \mathcal{W}^{N_{\text{R}} \times N_{\text{R}}}$ (having also unit magnitude elements) and a random sub-sampling step. The proposed extended analog combiner requires a larger number of phase shifters, compared to the conventional $\mathbf{W}_{\text{RF}}$, so as its outputs to be N_{R}. We argue, however, that those additional hardware components do not contribute significantly to RX's implementation complexity or power consumption. Let us define the output of the extended analog combiner as:

$$\mathbf{Y} \triangleq \tilde{\mathbf{W}}^H_{\text{RF}}\mathbf{R} = \tilde{\mathbf{W}}^H_{\text{RF}}(\mathbf{D}_{\text{R}}\bar{\mathbf{Z}}\mathbf{B} + \mathbf{N}) = \mathbf{A}\bar{\mathbf{Z}}\mathbf{B} + \tilde{\mathbf{W}}^H_{\text{RF}}\mathbf{N}, \quad (9)$$

where $\mathbf{A} \triangleq \tilde{\mathbf{W}}^H_{\text{RF}}\mathbf{D}_{\text{R}}$ and $\tilde{\mathbf{N}} \triangleq \tilde{\mathbf{W}}^H_{\text{RF}}\mathbf{N}$. As shown in Fig. 1, the N_{R} rows of $\mathbf{Y} \in \mathbb{C}^{N_{\text{R}} \times T}$ are fed to the random sub-sampler represented by $\mathbf{\Omega} \in \{0,1\}^{N_{\text{R}} \times T}$. This matrix is composed of TL_{R} ones and $T(N_{\text{R}} - L_{\text{R}})$ zeros, while the positions of its unity elements are randomly chosen in a uniform fashion over the set $\{1, 2, \ldots, N_{\text{R}}\}$ for each row of $\mathbf{\Omega}$. Putting all above together, the output of the random sub-sampler can be mathematically expressed as:

$$\mathbf{Y}_\Omega \triangleq \mathbf{\Omega} \circ \mathbf{Y} = \mathbf{\Omega} \circ (\tilde{\mathbf{W}}^H_{\text{RF}}\mathbf{R}) \in \mathbb{C}^{N_{\text{R}} \times T}. \quad (10)$$

Evidently, each column of $\mathbf{Y}_\Omega$ will have only L_{R} non-zero rows out of its N_{R} in total, which will be fed to the L_{R} receive RF chains.

B. Proposed Channel Estimation Algorithm

The intended channel estimation problem has two important properties. First, the time-domain beamspace channel matrices $\mathbf{Z}(\ell)$ $\forall \ell$ are sparse [10]; particularly, there exist $\|\mathbf{Z}_\ell\|_0$ non-zero values. The same holds for the concatenated matrix $\bar{\mathbf{Z}}$, which has exactly $\|\bar{\mathbf{Z}}\|_0 = \sum_{\ell=1}^{L} \|\mathbf{Z}_\ell\|_0$ non-zero values. Second, it can be shown, using the properties of the wideband mmWave MIMO channel model in (1), that $\mathbf{H}(\ell)$ $\forall \ell$ and $\mathbf{Y}$ are low rank matrices [14], [15].

Capitalizing on both the sparsity and low rank properties of the wideband MIMO channel, we formulate the following joint Optimization Problem (OP) for the estimation of $\bar{\mathbf{Z}}$:

$$\begin{aligned} &\min_{\mathbf{Y},\bar{\mathbf{Z}}} \tau_Y \|\mathbf{Y}\|_* + \tau_S \|\bar{\mathbf{Z}}\|_1 \\ &\text{subject to } \mathbf{\Omega} \circ \mathbf{Y} = \mathbf{Y}_\Omega \text{ and } \mathbf{Y} = \mathbf{A}\bar{\mathbf{Z}}\mathbf{B}, \end{aligned} \quad (11)$$

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 $\tau_Y > 0$ and $\tau_S > 0$ depend on the number of known entries $\mathbf{Y}_\Omega$.

In the following, we present an iterative algorithm which provides the optimal solution for (11). Note that a similar approach has been adopted in [7], however there, the sub-sampling is performed on the channel matrix, and not on the received matrix $\mathbf{Y}$. We first introduce the two auxiliary matrix variables $\mathbf{X} \in \mathbb{C}^{N_{\text{R}} \times T}$ and $\mathbf{C} \triangleq \mathbf{Y} - \mathbf{A}\bar{\mathbf{Z}}\mathbf{B}$ to reformulate the targeted OP in the following equivalent form:

$$\begin{aligned} &\min \tau_Y \|\mathbf{Y}\|_* + \tau_S \|\bar{\mathbf{Z}}\|_1 + \frac{1}{2}\|\mathbf{C}\|^2_F + \frac{1}{2}\|\mathbf{\Omega} \circ \mathbf{X} - \mathbf{Y}_\Omega\|^2_F \\ &\text{subject to } \mathbf{Y} = \mathbf{X} \text{ and } \mathbf{C} = \mathbf{Y} - \mathbf{A}\bar{\mathbf{Z}}\mathbf{B}, \end{aligned} \quad (12)$$

which is equivalent to (11), but the cost function has been decomposed into the sum of four variables: $\mathbf{Y}, \bar{\mathbf{Z}}, \mathbf{X}$, and $\mathbf{C}$. The Lagrangian function of the OP (12) can be easily expressed as:

$$\begin{aligned} &\mathcal{L}(\mathbf{Y}, \bar{\mathbf{Z}}, \mathbf{X}, \mathbf{C}, \mathbf{V}_1, \mathbf{V}_2) = \tau_Y \|\mathbf{Y}\|_* + \tau_S \|\bar{\mathbf{Z}}\|_1 \\ &+ \frac{1}{2}\|\mathbf{C}\|^2_F + \frac{1}{2}\|\mathbf{\Omega} \circ \mathbf{X} - \mathbf{Y}_\Omega\|^2_F + \text{tr}(\mathbf{V}_1^H(\mathbf{Y} - \mathbf{X})) \\ &+ \frac{\rho}{2}\|\mathbf{Y} - \mathbf{X}\|^2_F + \text{tr}(\mathbf{V}_2^H(\mathbf{C} - \mathbf{X} + \mathbf{A}\bar{\mathbf{Z}}\mathbf{B})) \\ &+ \frac{\rho}{2}\|\mathbf{C} - \mathbf{X} + \mathbf{A}\bar{\mathbf{Z}}\mathbf{B})\|^2_F, \end{aligned} \quad (13)$$

where $\mathbf{V}_1 \in \mathbb{C}^{N_{\text{R}} \times T}$ and $\mathbf{V}_2 \in \mathbb{C}^{N_{\text{R}} T \times 1}$ are dual variables (the Lagrange multipliers) adding the constraints of (12) to the cost function, and ρ denotes ADMM's stepsize.

To optimality conditions are obtained by differentiating sequentially with respect to the primal variables. Therefore, differentiating with respect to $\mathbf{Y}$, we get:

$$\frac{\partial \mathcal{L}}{\partial \mathbf{Y}} = \frac{\partial}{\partial \mathbf{Y}} \left(\tau_Y \|\mathbf{Y}\|_* + \frac{\rho}{2}\|\mathbf{Y} - (\mathbf{X} - \frac{1}{\rho}\mathbf{V}_1)\|^2_F \right). \quad (14)$$

Setting (14) equal to zero, the solution of the resulting equation can be obtained from the Singular Value Thresholding (SVT) operator [16] as expressed in line 2 of Algorithm 1, where $\mathbf{U} \in \mathbb{C}^{N_{\text{R}} \times r}$ and $\mathbf{V} \in \mathbb{C}^{N_{\text{R}} \times r}$ are the left and right singular vector matrices, respectively, of the matrix $\mathbf{X} - \frac{1}{\rho}\mathbf{V}_1$, and $\zeta_i \triangleq \sigma_i - \tau/\rho$ with σ_i's denoting its r singular values.

Next, differentiating with respect to $\mathbf{X}$, we have $\frac{\partial \mathcal{L}}{\partial \mathbf{X}} = \mathbf{\Omega} \circ \mathbf{X} - \mathbf{Y}_\Omega - \mathbf{V}_1 - \rho(\mathbf{Y} - \mathbf{X})$, and if we set it equal to zero,

Algorithm 1 Proposed ADMM-based Channel Estimation

Input: $\mathbf{Y}_\Omega$, $\mathbf{\Omega}$, $\mathbf{A}$, $\mathbf{B}$, ρ, τ_Y, τ_S, and $I_{\max}$.
Output: $\bar{\mathbf{Z}}^{(I_{\max})}$
Initialization: $\mathbf{Y}^{(0)} = \bar{\mathbf{Z}}^{(0)} = \mathbf{C}^{(0)} = \mathbf{V}_1^{(0)} = \mathbf{V}_2^{(0)} = \mathbf{0}$
1: **for** $\ell = 0, 1, \ldots, I_{\max} - 1$ **do**
2: $\quad \mathbf{Y}^{(\ell+1)} = \mathbf{U}\text{diag}\big(\{\text{sign}(\zeta_i) \times \max(\zeta_i, 0)\}_{1 \le i \le r}\big)\mathbf{V}^H$
3: $\quad \mathbf{k}_1 = \mathbf{v}_1 + \rho\mathbf{y} + \mathbf{K}_1^H\mathbf{y}_\Omega + \mathbf{v}_2 + \rho\mathbf{c} + \rho\mathbf{K}_2\bar{\mathbf{z}}$
4: $\quad \mathbf{X}^{(\ell+1)} = \text{unvec}((\mathbf{K}_1^H\mathbf{K}_1 + 2\rho\mathbf{I})^{-1}\mathbf{k}_1)$
5: $\quad$ Obtain $\bar{\mathbf{z}}^{(\ell+1)}$ via (16).
6: $\quad \bar{\mathbf{Z}}^{(\ell+1)} = \text{unvec}(\bar{\mathbf{z}}^{(\ell+1)})$
7: $\quad \mathbf{C}^{(\ell+1)} = \frac{\rho}{1+\rho}\left(\mathbf{X} - \mathbf{A}\bar{\mathbf{Z}}\mathbf{B} - \frac{1}{\rho}\mathbf{V}_2\right)$
8: $\quad \mathbf{V}_1^{(i+1)} = \mathbf{V}_1^{(i)} + \rho\big(\mathbf{X}^{(i+1)} - \mathbf{Y}^{(i+1)}\big)$
9: $\quad \mathbf{V}_2^{(i+1)} = \mathbf{V}_2^{(i)} + \rho\big(\mathbf{C} - \mathbf{X}^{(i+1)} + \mathbf{A}\bar{\mathbf{Z}}^{(i+1)}\mathbf{B}\big)\big)$
10: **end for**

the resulting problem is equivalent to solving the following system of equations:

$$\mathbf{x} = (\mathbf{K}_1^H\mathbf{K}_1 + 2\rho\mathbf{I}_{TN_{\rm R}})^{-1}\mathbf{k}_1, \tag{15}$$

where $\mathbf{k}_1 \triangleq \mathbf{v}_1 + \rho\mathbf{y} + \mathbf{K}_1^H\mathbf{y}_\Omega + \mathbf{v}_2 + \rho\mathbf{c} + \rho\mathbf{K}_2\bar{\mathbf{z}}$, $\mathbf{K}_2 \triangleq \mathbf{B}^T \otimes \mathbf{A} \in \mathbb{C}^{TN_{\rm R} \times N_{\rm T}N_{\rm R}}$ and $\mathbf{K}_1 \triangleq \sum_{i=1}^{N_{\rm R}} \text{diag}([\mathbf{\Omega}]_i)^T \otimes \mathbf{E}_{ii} \in \mathbb{C}^{TN_{\rm R} \times TN_{\rm R}}$ with $\mathbf{E}_{ii}$ obtained from the $N_{\rm R} \times N_{\rm R}$ all-zero matrix after inserting a unity value at its (i,i)-th position. Also, the small boldfaced letters $\mathbf{v}_1, \mathbf{y}, \mathbf{y}_\Omega, \mathbf{c}$ and $\bar{\mathbf{z}}$, are the $\text{vec}(\cdot)$ results of their capital letter equivalents.

Differentiating with respect to $\bar{\mathbf{Z}}$, one gets $\frac{\partial \mathcal{L}}{\partial \bar{\mathbf{Z}}} = \frac{\partial}{\partial \bar{\mathbf{Z}}}\left(\tau_S\|\bar{\mathbf{Z}}\|_1 + \frac{\rho}{2}\|\frac{1}{\rho}\mathbf{V}_2 + \mathbf{C} - \mathbf{X} + \mathbf{A}\bar{\mathbf{Z}}\mathbf{B}\|_F^2\right)$, the minimization of which is equivalent to the following sparse optimization problem:

$$\min_{\bar{\mathbf{z}}} \|\bar{\mathbf{z}}\|_1 + \|\mathbf{K}_2\bar{\mathbf{z}} - \mathbf{k}\|_2^2, \tag{16}$$

where $\mathbf{k} \triangleq \mathbf{x} - \mathbf{c} - \frac{1}{\rho}\mathbf{v}_2 \in \mathbb{C}^{TN_{\rm R} \times 1}$. Finally, differentiating with respect to $\mathbf{C}$ and equating to zero leads to the solution, which is expressed in line 7 of Algorithm 1. After obtaining the updates for the primal variables, we update the dual variables $\mathbf{V}_1$ and $\mathbf{V}_2$ according to the lines 8 and 9 of Algorithm 1, respectively. The termination of the algorithm is done by a maximum number of iterations $I_{\max}$.

Computational Complexity: The overall computational order of Algorithm 1 is dictated by the most demanding step, i.e., line 5 which requires the solution of (16). For instance, using a standard soft-thresholding operator for solving (16), the computational order is $\mathcal{O}(N_{\rm T}N_{\rm R}T)$ for each iteration of Algorithm 1. Line 3 involves the inversion of the matrix $(\mathbf{K}_1^H\mathbf{K}_1 + 2\rho\mathbf{I}) \in \mathbb{C}^{TN_{\rm R} \times TN_{\rm R}}$. However, this matrix is diagonal, hence, the order of the complexity is $\mathcal{O}(TN_{\rm R})$. In this work, we have solved (16) using the CVX and Mosek packages. Investigation of more computational efficient algorithms for (15) and (16) is left as future work.

IV. Performance Evaluation Results

In this section, we evaluate the performance of the proposed wideband mmWave massive MIMO channel estimation technique via computer simulations obtained using MATLAB$^{\rm TM}$ with CVX and Mosek packages. All presented results have been averaged over 100 Monte Carlo realizations. The number of the mmWave channel paths was set to $N_p = 6$, while the number of the delay taps to $L = 2$. For Algorithm 1 we have used $\rho = \sqrt{\lambda_{\min}\tau_Y/2}$ and $I_{\max} = 50$, where $\lambda_{\min}$ is the minimum eigenvalue of the matrix $\mathbf{Y}$. We have considered a point-to-point $N_{\rm T} \times N_{\rm R}$ MIMO system, where TX broadcasts a random Gaussian symbol vector $\mathbf{s}(t) \in \mathbb{C}^{N_{\rm T} \times 1}$ at each frame t with $t = 1, 2, \ldots, T$. We have assumed block transmissions of the training frames with a zero prefix appended to each frame [8]. The RX adopts HBF with $N_{\rm R}$ antenna elements attached to $L_{\rm R} \le N_{\rm R}$ RF chains with $L_{\rm R}^e = N_{\rm R}$. We have assumed perfect time and frequency synchronization between the communicating TX and RX nodes.

As benchmark CSI estimation techniques we have considered recent techniques for wideband mmWave channel estimation, which are based on compressive sampling tools: *i*) the technique in [8] for the Time Domain (TD) case based on the Orthogonal Matching Pursuit (OMP) algorithm; and *ii*) the Vector Approximate Message Passing (VAMP) algorithm of [17] applied for our problem formulation. We have also considered *iii*) the Two-Stage estimation technique [6] exploiting both Sparsity and low Rank (TSSR), which we have applied to the targeted in this work wideband channel case. Note that OMP and VAMP exploit only the sparsity of the channel matrix, and VAMP is a statistical estimator. TSSR exploits the sparsity and low rank properties by first employing the SVT operator to recover the channel matrix $\mathbf{Y}$, then transforms it to the beamspace domain, and finally uses it as input to the VAMP sparse estimator. Concerning the computational order, note that the benchmark techniques involve the solution of a sparse problem similar to (16), hence, they have comparable complexity with the proposed technique.

As for the estimation performance metric, we used the Normalized MSE (NMSE) performance, which is defined as NMSE $\triangleq \|\mathbf{Z}^* - \bar{\mathbf{Z}}\|/\|\mathbf{Z}^*\|$, where $\bar{\mathbf{Z}}$ is the estimation matrix obtained by each of the compared techniques, and $\mathbf{Z}^* = (\tilde{\mathbf{W}}_{\rm RF}^H\mathbf{D}_{\rm R})^\dagger\bar{\mathbf{H}}\mathbf{B}^\dagger$ where $\bar{\mathbf{H}} \triangleq [\mathbf{H}(0)\,\mathbf{H}(1) \cdots \mathbf{H}(L-1)] \in \mathbb{C}^{N_{\rm R} \times LN_{\rm T}}$. In the figures that follow, we investigate NMSE with respect to (w.r.t.) the transmit Signal-to-Noise-Ratio (SNR), the number of the training blocks T, the number of the channel spatial paths N_p, and the number of the transmitted data streams N_s. Note that for the proposed design, we have that $L_{\rm R} = \|\mathbf{\Omega}\|_0 T^{-1}$, i.e., based on the random subsampling block, we randomly select $L_{\rm R}$ outputs of the analog combiner $\tilde{\mathbf{W}}_{\rm RF}$ at each frame instance $t = 1, 2, \ldots, T$.

The NMSE performance as a function of the transmit SNR in dB for the four considered CSI estimation algorithms is illustrated in Fig. 2(a). It can be observed that the OMP technique is the least affected by the SNR increase, exhibiting similar performance over all plotted SNR range. This behavior can be explained from the angular discretization noise that this technique is suffering. On the other side, the estimation accuracy of VAMP improves significantly with increasing SNR, since this estimator obtains the noise statistics from measured data. Nevertheless, it is shown that the proposed

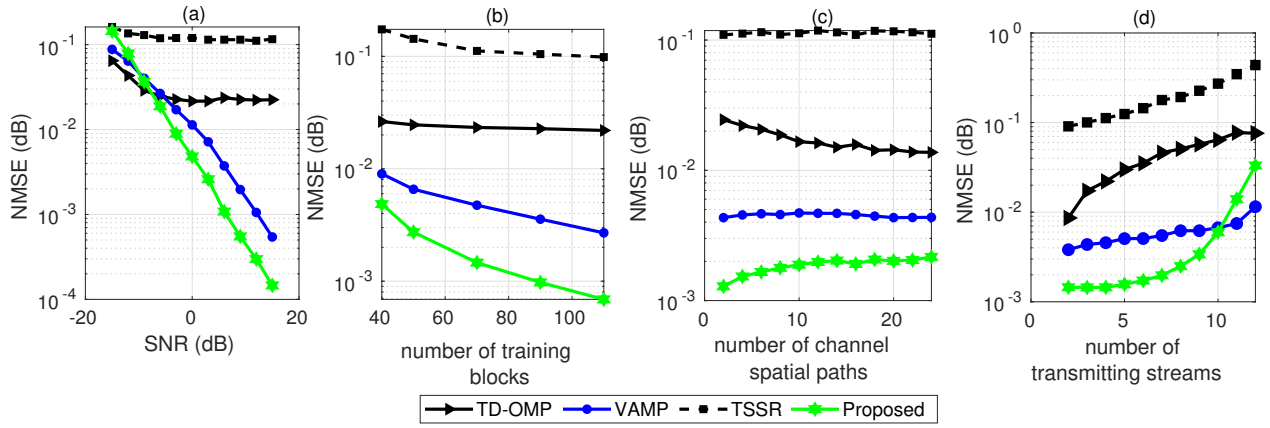

Fig. 2. NMSE of the channel estimation for a $32 \times N_{\mathrm{T}}$ wideband mmWave MIMO channel. (a) NMSE versus SNR in dB. The number of transmit antennas is $N_{\mathrm{T}} = 4$, the number of receive RF chains is set as $L_{\mathrm{R}} = 24$ and the number of training blocks is $T = 70$. (b) NMSE versus the number of training blocks T. The SNR is set at 5dB, the number of transmit antennas is $N_{\mathrm{T}} = 4$ and the number of RF chains is $L_{\mathrm{R}} = 24$. (c) NMSE versus the number of the channel spatial paths N_p for transmit SNR equal to 5dB and $T = 70$ training blocks and the number of transmit antennas is $N_{\mathrm{T}} = 4$. (d) NMSE versus the number of transmitting antennas N_{T} for transmit SNR equal to 5dB and $T = 70$ training blocks.

technique provides superior estimation performance compared to all others over -5dB SNR values. This trend can be justified by the fact that it exploits jointly the sparsity and low rank properties of the wideband mmWave MIMO channel. In this work, the values of $\tau_Y = \frac{1}{\|\mathbf{Y}_\Omega\|_F^2}, \tau_S = \frac{\tau_Y}{2}$ have been chosen so as to optimize its performance at higher SNR values. Selecting a higher value for τ_S could result into better performance at the low SNR regime. In Figs. 2(b)-(c), we depict the NMSE results versus the number of the training blocks T and the channel spatial paths N_p. It is shown that, for all T values, the proposed estimator seems to outperform all others considered. Finally, Fig. 2(d) includes NMSE curves versus the number of the transmit antennas N_{T} for transmit SNR equal to 5dB and $T = 70$ training blocks. These results indicate that for $N_{\mathrm{T}} > 8$ the NMSE of the proposed technique increases due to the increase of the rank of the measurement matrix $\mathbf{Y}$. This is due to the fact that Algorithm 1 is not able to converge to a small steady-state error due to the increased rank of $\mathbf{Y}$. To overcome this, a higher number of receive antennas or training symbols is required. Specifically, the minimum number of measurements is related to the the rank of the under-sampled matrix according to the relation [16] $\|\mathbf{Y}_\Omega\|_0 \geq \mathrm{rank}(\mathbf{Y}) \times T^{1.2} \times \log(T)$ for $N_{\mathrm{R}} = T$.

V. Conclusion

We have presented an iterative algorithm for wideband mmWave MIMO channel estimation that exploits jointly the beamspace channel's sparsity and the low rank of the measurement matrix. It provides more accurate recovery, especially for short beam training intervals. The presented algorithm capitalizes on our proposed analog combining architecture that includes an extended combiner and a random sub-sampler before the input of the analog received signals to the digital component of the HBF receiver. It has been shown through representative simulation results that the proposed algorithm exhibits improved performance in terms of MSE for channel estimation with short beam training length.

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] 3GPP, "Study on New Radio (NR) Access Technology Physical Layer Aspects- Release 14," *ETSI*, TR 38.802, 2017.
[3] A. F. Molisch *et al.*, "Hybrid beamforming for massive MIMO: A survey," *IEEE Commun. Mag.*, vol. 55, no. 9, pp. 134–141, Sep. 2017.
[4] R. W. Heath, Jr. *et al.*, "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.
[5] 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.
[6] 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.
[7] 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.
[8] K. Venugopal *et al.*, "Channel estimation for hybrid architecture-based wideband millimeter wave systems," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 1996–2009, Sep. 2017.
[9] S. Boyd *et al.*, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.
[10] 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.
[11] A. Forenza *et al.*, "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.
[12] A. M. Sayeed, "Deconstructing multiantenna fading channels," *IEEE Trans. Signal Process.*, vol. 50, no. 10, pp. 2563–2579, Oct. 2002.
[13] R. Méndez-Rial et al., "Hybrid MIMO architectures for millimeter wave communications: Phase shifters or switches?" *IEEE Access*, vol. 4, pp. 247–267, Jan. 2016.
[14] 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.
[15] H. Ghauch *et al.*, "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.
[16] J. F. Cai et al., "A singular value thresholding algorithm for matrix completion," *SIAM J. Opt.*, vol. 20, no. 4, pp. 1956–1982, 2010.
[17] P. Schniter, S. Rangan, and A. K. Fletcher, "Vector approximate message passing for the generalized linear model," in *Proc. Asilomar CSSC*, Pacific Grove, USA, Nov. 2016, pp. 1525–1529.