

COMPRESSED SENSING BASED CHANNEL ESTIMATION AND OPEN-LOOP TRAINING DESIGN FOR HYBRID ANALOG-DIGITAL MASSIVE MIMO SYSTEMS

Khaled Ardah[‡], Bruno Sokal[‡], André L. F. de Almeida[†], and Martin Haardt[‡]

[‡]Communications Research Laboratory (CRL), TU Ilmenau, Ilmenau, Germany

[†] Wireless Telecom Research Group (GTEL), Federal University of Ceará, Fortaleza, Brazil
{khaled.ardah, bruno.sokal, martin.haardt}@tu-ilmenau.de, andre@gtel.ufc.br.

ABSTRACT

Channel estimation in hybrid analog-digital massive MIMO systems is a challenging problem due to the high channel dimension, low signal-to-noise ratio before beamforming, and reduced number of radio-frequency chains. Compressed sensing based algorithms have been adopted to address these challenges by leveraging the sparse nature of millimeter-wave MIMO channels. In compressed sensing-based methods, the training vectors should be designed carefully to guarantee recoverability. Although using random vectors has an overwhelming recoverability guarantee, it has been recently shown that an optimized update, which could be obtained so that the mutual coherence of the resulting sensing matrix is minimized, can improve the recoverability guarantee. In this paper, we propose an open-loop hybrid analog-digital beam-training framework, where a given sensing matrix is decomposed into analog and digital beamformers. The given sensing matrix can be designed efficiently offline to reduce computational complexity. Simulation results show that the proposed training method achieves a lower mutual coherence and an improved channel estimation performance than the other benchmark methods.

Index Terms— Hybrid analog-digital, massive MIMO, compressed sensing, open-loop training design, mutual coherence.

1. INTRODUCTION

The combination of millimeter-wave (mmWave) and massive MIMO wireless technologies is seen as a key enabler in future wireless networks [1]. To facilitate its practical implementation, hybrid analog-digital (HAD) beamforming architectures have been proposed [2–7] to reduce cost and energy consumption compared to fully digital (FD) architectures. To realize its full potential, channel state information (CSI) is required at the transceivers to perform efficient adaptive beamforming techniques. However, the CSI estimation problem in massive MIMO systems is hard to solve because of the large channel dimension and it becomes harder with HAD systems due to the reduced number of radio frequency (RF) chains and low the signal-to-noise ratio (SNR) before beamforming and beam alignment.

In the practice, however, it is observed that the massive MIMO channel in mmWave communication has a sparse structure in the angular domain due to the limited number of scatters compared to the large number of antennas [2, 8]. Exploiting this sparse structure, compressed sensing (CS) tools [9–12] can be used to estimate the MIMO channel [4, 13–15], where the problem can be turned into estimating the dominant channel path parameters. In CS-based methods,

channel training vectors should be designed carefully to guarantee recoverability. In general, channel training can be categorized into closed-loop and open-loop methods. In the former, the channel training is implemented as a multistage process, where at each stage the transceivers engage in a beam-training process to select the best beam [8]. However, the mmWave channel has a very short time-coherence, which makes such time-consuming adaptive beam-training method limited in practice. Further, in multi-user MIMO systems, the training overhead tends to increase linearly with the number of users. Therefore, the open-loop based training methods are more practical, where the transmitter emits pilot vectors. After that, the receiver estimates the parameters of the dominant channel paths from the received signals [4, 15]. In contrast to closed-loop methods, open-loop methods only have a single-stage and the training overhead remains the same irrespective of the number of users [12, 15].

Although using random vectors as training sequences has an overwhelming recoverability guarantee [9–12], it has been recently shown that an optimized update, which could be obtained so that the mutual coherence¹ of the resulting sensing matrix is minimized, can improve the signal recoverability guarantee [4, 15]. Considering hybrid analog-digital MIMO systems, the authors in [15] proposed an open-loop training design, where the columns of the fully-connected analog beamformers are selected from a given DFT-matrix, while the digital beamformers are updated afterward so that they have a unitary structure. The upper-bound and lower-bound of sum-of-squared errors are derived analytically, which are then confirmed using computer simulations. In [15], it is shown that carefully updated training vectors can lead to a lower mutual coherence and a better CSI estimation accuracy than the random training approach.

In this paper, we consider a downlink hybrid analog-digital massive MIMO system operating over mmWave bands. We formulate the CSI estimation problem as a sparse recovery problem of the dominant channel path parameters, which is solved using orthogonal matching pursuit (OMP). To improve the recoverability guarantee, we propose an open-loop hybrid analog-digital beam-training framework and evaluate its performance against the DFT-based training method [15] and the random-based approach [4]. In the proposed method, we assume that a sensing matrix with an appropriate size is efficiently designed offline using one of the proposed methods in the literature [16–19]. After that, the given sensing matrix is decomposed into analog and digital beamformers, which are then used for the channel training and estimation. Simulations results show that the proposed training method achieves a lower mutual coherence and an improved channel estimation accuracy than the state-of-the-art.

The authors gratefully acknowledge the support of the German Research Foundation (DFG) under contract no. HA 2239/6-2 (EXPRESS II) and the support of CAPES/PROBRAL project (Grant no. 88887.144009/2017-00). The research of André L. F. de Almeida is partially supported by the CNPq.

¹The mutual-coherence of a given sensing matrix is defined by (9).

2. SYSTEM MODEL

We consider a mmWave MIMO system where a single transmitter equipped with M_T antennas and $N_T \leq M_T$ RF chains is communicating with a single receiver equipped with M_R antennas and $N_R \leq M_R$ RF chains. The total transmission time is divided into $T = T_R \cdot T_T$ time slots: for each of the T_T time slots at the transmitter, there are T_R time slots at the receiver. The received signal $\mathbf{y}_{r,t}$ at the (r,t) -th transmission time, $r \in \{1, \dots, T_R\}$, $t \in \{1, \dots, T_T\}$, is given as

$$\mathbf{y}_{r,t} = \mathbf{U}_r^H \mathbf{H} \mathbf{V}_t \mathbf{s}_t + \mathbf{n}_{r,t} \in \mathbb{C}^{N_R}, \quad (1)$$

where $\mathbf{U}_r \in \mathbb{C}^{M_R \times N_R}$ is the receive hybrid analog-digital combining matrix with $\|\mathbf{U}_r\|_F^2 = 1$, $\mathbf{V}_t \in \mathbb{C}^{M_T \times N_T}$ is the transmit hybrid analog-digital precoding matrix with $\|\mathbf{V}_t\|_F^2 = 1$, $\mathbf{H} \in \mathbb{C}^{M_R \times M_T}$ is the mmWave channel matrix with $\mathbb{E}\{\|\mathbf{H}\|_F^2\} = M_R M_T$, $\mathbf{s}_t \in \mathbb{C}^{N_T}$ is the pilot vector with $\mathbb{E}\{\mathbf{s}_t^H \mathbf{s}_t\} = \mathbf{I}$, and $\mathbf{n}_{r,t} \in \mathbb{C}^{N_R}$ is the additive white Gaussian noise with variance σ_n^2 . Here, the hybrid analog-digital precoding and combining matrices $\mathbf{V}_t, \forall t, \mathbf{U}_r, \forall r$, are constructed as

$$\mathbf{V}_t = \mathbf{V}_t^{\text{RF}} \mathbf{V}_t^{\text{BB}} \text{ and } \mathbf{U}_r = \mathbf{U}_r^{\text{RF}} \mathbf{U}_r^{\text{BB}}, \quad (2)$$

where $\mathbf{V}_t^{\text{RF}} \in \mathbb{C}^{M_T \times N_T}$ and $\mathbf{U}_r^{\text{RF}} \in \mathbb{C}^{M_R \times N_R}$ represent the analog domain beamformers, while $\mathbf{V}_t^{\text{BB}} \in \mathbb{C}^{N_T \times N_T}$ and $\mathbf{U}_r^{\text{BB}} \in \mathbb{C}^{N_R \times N_R}$ are the digital baseband domain beamformers. We assume that the analog domain beamformers are realized using a network of fully-connected phase-shifters [5, 6]. Therefore, the entries of \mathbf{V}_t^{RF} and \mathbf{U}_r^{RF} should satisfy the constant modulus constraints given as $|\mathbf{V}_t^{\text{RF}}[i,j]| = \frac{1}{\sqrt{M_T}}, \forall t, i, j$, and $|\mathbf{U}_r^{\text{RF}}[i,j]| = \frac{1}{\sqrt{M_R}}, \forall r, i, j$.

Similar to [4, 8, 15], we assume a geometric channel model where the mmWave channel matrix \mathbf{H} is modeled as

$$\mathbf{H} = \sqrt{M_T M_R / P} \cdot \sum_{p=1}^P \alpha_p \mathbf{a}_{M_R}(\theta_{R,p}) \mathbf{a}_{M_T}^T(\theta_{T,p}), \quad (3)$$

where P is the total number of channel paths, $\theta_{T,p}, \theta_{R,p}$ and ϕ_p are, respectively, the complex path gain, angle-of-arrival (AoA), and angle-of-departure (AoD) of the p -th path. Further, $\mathbf{a}_{M_R}(\theta_{R,p})$ and $\mathbf{a}_{M_T}(\theta_{T,p})$ denote the array response/steering vectors at the receiver and transmitter, respectively. Assuming a uniform linear array with half-wavelength spacing between the array elements, the steering vector $\mathbf{a}_M(\xi) \in \mathbb{C}^M$ is given as $\mathbf{a}_M(\xi) = 1/\sqrt{M} [1, e^{j\xi}, \dots, e^{j\xi(M-1)}]^T$, where $\xi \in [0, 2\pi]$ is the spatial frequency [4].

Let $\mathbf{U} = [\mathbf{U}_1, \dots, \mathbf{U}_{T_R}] \in \mathbb{C}^{M_R \times T_R N_R}$. Then, after stacking the received signals $\{\mathbf{y}_{r,t}, \forall r\}$ over each other we have a signal vector \mathbf{y}_t expressed as

$$\mathbf{y}_t = [\mathbf{y}_{1,t}^T, \dots, \mathbf{y}_{T_R,t}^T]^T = \mathbf{U}^H \mathbf{H} \mathbf{V}_t \mathbf{s}_t + \mathbf{n}_t \in \mathbb{C}^{T_R N_R}, \quad (4)$$

where $\mathbf{n}_t = [\mathbf{n}_{1,t}^T, \dots, \mathbf{n}_{T_R,t}^T]^T \in \mathbb{C}^{T_R N_R}$. Next, we stack the received signals $\{\mathbf{y}_t, \forall t\}$ over each other to have a signal vector \mathbf{y} expressed as

$$\mathbf{y} = (\mathbf{V}^T \otimes \mathbf{U}^H) \mathbf{h} + \mathbf{n} \in \mathbb{C}^{T_R N_R T_T}, \quad (5)$$

where \otimes denotes the Kronecker product, $\mathbf{V} = [\mathbf{V}_1 \mathbf{s}_1, \dots, \mathbf{V}_{T_T} \mathbf{s}_{T_T}] \in \mathbb{C}^{M_T \times T_T}$, $\mathbf{h} = \text{vec}(\mathbf{H}) \in \mathbb{C}^{M_T M_R}$ is the vectorized channel, and $\mathbf{n} = [\mathbf{n}_1^T, \dots, \mathbf{n}_{T_T}^T]^T \in \mathbb{C}^{T_R N_R T_T}$. Conventionally, the least squares estimate of the channel vector \mathbf{h} from (5) is computed by $\hat{\mathbf{h}} = (\mathbf{V}^T \otimes \mathbf{U}^H)^+ \mathbf{y}$. This requires that $(\mathbf{V}^T \otimes \mathbf{U}^H) \in \mathbb{C}^{T_R N_R T_T \times M_R M_T}$ has full column rank, i.e., $\text{rank}(\mathbf{V}^T \otimes \mathbf{U}^H) = M_R M_T \leq T_R N_R T_T$. To have an accurate channel estimate, the minimum number of time slots should

satisfy $T_T T_R \geq \frac{M_R M_T}{N_R}$, which can be seen as a limiting factor in practice, especially with massive MIMO configurations. Therefore, a more efficient channel estimation approach is required.

In the following, we formulate the channel estimation problem as a sparse recovery problem by exploiting the sparse (i.e., low-rank) structure of the mmWave channel. Let $\bar{\boldsymbol{\theta}}_R = [\bar{\theta}_{R,1}, \dots, \bar{\theta}_{R,G_R}]^T$ and $\bar{\boldsymbol{\theta}}_T = [\bar{\theta}_{T,1}, \dots, \bar{\theta}_{T,G_T}]^T$ represent the grid points of a $G_R \times G_T$ grid, where $\bar{\theta}_{R,n} = \frac{2\pi n}{G_R}, n \in \{0, \dots, G_R - 1\}$, $\bar{\theta}_{T,m} = \frac{2\pi m}{G_T}, m \in \{0, \dots, G_T - 1\}$, $G_R = \beta_R M_R$, $G_T = \beta_T M_T$, and $\beta_R, \beta_T \in \{1, 2, \dots\}$. Using $\bar{\boldsymbol{\theta}}_R$ and $\bar{\boldsymbol{\theta}}_T$, we define the dictionaries $\bar{\mathbf{A}}_R$ and $\bar{\mathbf{A}}_T$ as

$$\bar{\mathbf{A}}_R = [\mathbf{a}_{M_R}(\bar{\theta}_{R,1}), \dots, \mathbf{a}_{M_R}(\bar{\theta}_{R,G_R})] \in \mathbb{C}^{M_R \times G_R}, \quad (6)$$

$$\bar{\mathbf{A}}_T = [\mathbf{a}_{M_T}(\bar{\theta}_{T,1}), \dots, \mathbf{a}_{M_T}(\bar{\theta}_{T,G_T})] \in \mathbb{C}^{M_T \times G_T}. \quad (7)$$

Let us assume that $\theta_{R,p} \in \bar{\boldsymbol{\theta}}_R, \forall p, \theta_{T,p} \in \bar{\boldsymbol{\theta}}_T, \forall p$ (i.e., the AoAs and AoDs fall perfectly on the grid), and the number of channel paths $P \ll \min\{M_R, M_T\}$, which is usually the case in the mmWave bands [2, 8]. Then, the channel matrix \mathbf{H} can be represented by an P -sparse matrix $\bar{\mathbf{D}}$ as $\mathbf{H} = \bar{\mathbf{A}}_R \bar{\mathbf{D}} \bar{\mathbf{A}}_T^T$, where $\bar{\mathbf{D}} \in \mathbb{R}^{G_R \times G_T}$ contains P nonzero entries: their positions indicate the active AoA and AoD $\{\theta_{R,p}, \theta_{T,p}, \forall p\}$, where their values represent the complex numbers $\{\alpha_p, \forall p\}$. Using the channel sparse representation, we can write $\mathbf{h} = \text{vec}\{\mathbf{H}\} = (\bar{\mathbf{A}}_T \otimes \bar{\mathbf{A}}_R) \text{vec}\{\bar{\mathbf{D}}\} \in \mathbb{C}^{M_R M_T}$. Substituting the latter expression into (5), we have

$$\mathbf{y} = (\mathbf{V}^T \otimes \mathbf{U}^H) (\bar{\mathbf{A}}_T \otimes \bar{\mathbf{A}}_R) \mathbf{d} + \mathbf{n} = \boldsymbol{\Phi} \mathbf{d} + \mathbf{n} \in \mathbb{C}^{T_R N_R T_T}, \quad (8)$$

where $\boldsymbol{\Phi} = (\mathbf{V}^T \bar{\mathbf{A}}_T \otimes \mathbf{U}^H \bar{\mathbf{A}}_R) \in \mathbb{C}^{T_T T_R N_R \times G_T G_R}$. Note that the matrix $\boldsymbol{\Phi}$ is formed using pre-known information and \mathbf{d} is a sparse vector, i.e., it contains, on the ideal case, P nonzero entries. Therefore, the signal model in (8) fulfills a sparse formulation of the channel estimation problem [8, 9], where the P -sparse vector \mathbf{d} can be recovered using any known sparse-recovery technique, e.g., the OMP [20]. This implies that the number of required measurements T to detect the non-zero entries in \mathbf{d} is much less than $G_T G_R$, which can be approximated by $\mathcal{O}(P \ln M_T)$ [15]. However, to guarantee the recoverability of \mathbf{d} with high probability, $\boldsymbol{\Phi}$ should be efficiently designed, e.g., by minimizing its mutual coherence $\mu(\boldsymbol{\Phi})$ defined as [11, 12]

$$\mu(\boldsymbol{\Phi}) = \max_{j \neq k} \frac{|\phi_j^H \phi_k|}{\|\phi_j\| \|\phi_k\|}, \quad (9)$$

with columns $\phi_i \in \mathbb{C}^{T_R N_R T_T}$. Here, a large coherence $\mu(\boldsymbol{\Phi})$ means that there exist, at least, two highly correlated columns in $\boldsymbol{\Phi}$, which may confuse any pursuit technique, such as OMP. However, if $\mu(\boldsymbol{\Phi})$ is sufficiently small so that $\|\mathbf{d}\|_0 < \frac{1}{2}(1 + \mu^{-1}(\boldsymbol{\Phi}))$, the OMP technique is guaranteed to recover \mathbf{d} with an overwhelming probability [9–12]. Therefore, the goal of the next section is to design the sensing matrix $\boldsymbol{\Phi}$ to have a small mutual coherence $\mu(\boldsymbol{\Phi})$.

3. PROPOSED SENSING MATRIX DESIGN

First, we note that due to the Kronecker structure of $\boldsymbol{\Phi} = (\mathbf{V}^T \bar{\mathbf{A}}_T \otimes \mathbf{U}^H \bar{\mathbf{A}}_R)$, the coherence $\mu(\boldsymbol{\Phi})$ can be expressed as

$$\mu(\boldsymbol{\Phi}) = \max\{\mu(\boldsymbol{\Phi}_T), \mu(\boldsymbol{\Phi}_R)\}, \quad (10)$$

where $\boldsymbol{\Phi}_T = \mathbf{V}^T \bar{\mathbf{A}}_T \in \mathbb{C}^{T_T \times G_T}$ and $\boldsymbol{\Phi}_R = \mathbf{U}^H \bar{\mathbf{A}}_R \in \mathbb{C}^{T_R N_R \times G_R}$. The above result suggests that the sensing matrix design can be carried out using two independent steps, one for $\mu(\boldsymbol{\Phi}_T)$ and another

for $\mu(\Phi_R)$. Note that, since the dictionary matrices $\bar{\mathbf{A}}_T$ and $\bar{\mathbf{A}}_R$ are assumed fixed, the sensing matrix design of Φ_T and Φ_R boils down to efficiently choosing suitable precoding and decoding matrices $\mathbf{V} = [\mathbf{V}_T \mathbf{s}_T, \dots, \mathbf{V}_{T_T} \mathbf{s}_{T_T}] \in \mathbb{C}^{M_T \times T_T}$ and $\mathbf{U} = [\mathbf{U}_1, \dots, \mathbf{U}_{T_R}] \in \mathbb{C}^{M_R \times T_R N_R}$, respectively. In the literature, there exists a number of sensing matrix design methods with different objectives and solution approaches [16–19]. In general, the objective is to design a sensing matrix \mathbf{F} of size $[I \times J]$ such that it has a low mutual coherence, while assuming $I \leq J$. In the following, we utilize such results to design the sensing matrices Φ_T and Φ_R .

3.1. Sensing matrix design for Φ_T

We start with the design of the transmit precoding matrices $\{\mathbf{V}_t\}_{t=1}^{T_T}$ to minimize the mutual coherence of $\Phi_T = \mathbf{V}^T \bar{\mathbf{A}}_T \in \mathbb{C}^{T_T \times G_T}$. Let $\mathbf{F}_T \in \mathbb{C}^{T_T \times G_T}$ denote a sensing matrix designed efficiently using, e.g., the methods in [16–19]. For a given \mathbf{F}_T , the objective is to decompose it into \mathbf{V}^T and $\bar{\mathbf{A}}_T$ such that the error $\eta_v = \|\mathbf{F}_T - \mathbf{V}^T \bar{\mathbf{A}}_T\|_F^2$ is minimized. Since the dictionary matrix $\bar{\mathbf{A}}_T \in \mathbb{C}^{M_T \times G_T}$ is fixed, a solution for \mathbf{V} that minimizes η_v can be obtained using the least-squared method as

$$\mathbf{V}^{\text{LS}} = [\mathbf{v}_1^{\text{LS}}, \dots, \mathbf{v}_{T_T}^{\text{LS}}] = (\mathbf{F}_T \bar{\mathbf{A}}_T^+)^T \in \mathbb{C}^{M_T \times T_T}. \quad (11)$$

which is optimal only if $G_T \leq M_T$, where $\bar{\mathbf{A}}_T^+ = (\bar{\mathbf{A}}_T^H \bar{\mathbf{A}}_T)^{-1} \bar{\mathbf{A}}_T^H$, if $M_T \geq G_T$, or $\bar{\mathbf{A}}_T^+ = \bar{\mathbf{A}}_T^H (\bar{\mathbf{A}}_T \bar{\mathbf{A}}_T^H)^{-1}$, if $M_T < G_T$ [21]. From the above, we know that each t -th column-vector \mathbf{v}_t^{LS} should be expressed as $\mathbf{v}_t^{\text{LS}} = \mathbf{V}_t \mathbf{s}_t$. Therefore, for a given symbol vector $\mathbf{s}_t \in \mathbb{C}^{N_T}$, we seek for a precoding matrix $\mathbf{V}_t \in \mathbb{C}^{M_T \times N_T}$ that minimizes the error $\eta_s = \|\mathbf{v}_t^{\text{LS}} - \mathbf{V}_t \mathbf{s}_t\|_F^2$. The optimal solution for \mathbf{V}_t that minimizes η_s can be calculated using the least-squared method as

$$\mathbf{V}_t^{\text{LS}} = \mathbf{v}_t^{\text{LS}} \mathbf{s}_t^+ = \mathbf{v}_t^{\text{LS}} (\mathbf{s}_t^H \mathbf{s}_t)^{-1} \mathbf{s}_t^H \stackrel{(a)}{=} \mathbf{v}_t^{\text{LS}} \mathbf{s}_t^H \in \mathbb{C}^{M_T \times N_T}, \quad (12)$$

which always has a rank equal to one, i.e., $\text{rank}(\mathbf{V}_t^{\text{LS}}) = 1$, where (a) follows from $\mathbf{s}_t^H \mathbf{s}_t \stackrel{\text{def}}{=} 1$. From (12), we have $\eta_s = 0$, since $\mathbf{v}_t^{\text{LS}} = \mathbf{V}_t^{\text{LS}} \mathbf{s}_t = \mathbf{v}_t^{\text{LS}} \mathbf{s}_t^H \mathbf{s}_t = \mathbf{v}_t^{\text{LS}}$. Thus, if $G_T \leq M_T$, the mutual coherence of the resulting sensing matrix $\hat{\Phi}_T = [\mathbf{V}_1^{\text{LS}} \mathbf{s}_1, \dots, \mathbf{V}_{T_T}^{\text{LS}} \mathbf{s}_{T_T}]^T \bar{\mathbf{A}}_T$ has an equal mutual coherence to that of the given matrix \mathbf{F}_T , i.e., $\mu(\hat{\Phi}_T) = \mu(\mathbf{F}_T)$.

Note that the precoding matrix \mathbf{V}_t^{LS} , which can be seen as a FD precoding matrix, has to be decomposed into an analog \mathbf{V}_t^{RF} and a digital \mathbf{V}_t^{BB} matrix. One approach to decompose \mathbf{V}_t^{LS} into \mathbf{V}_t^{RF} and \mathbf{V}_t^{BB} , $\forall t = 1, \dots, T_T$, is given as [7]

$$\begin{aligned} \min_{\mathbf{V}_t^{\text{RF}}, \mathbf{V}_t^{\text{BB}}} \quad & \|\mathbf{V}_t^{\text{LS}} - \mathbf{V}_t^{\text{RF}} \mathbf{V}_t^{\text{BB}}\|_F^2 \\ \text{s.t.} \quad & \mathbf{V}_t^{\text{RF}} \in \mathcal{F} \text{ and } (\mathbf{V}_t^{\text{BB}})^H \mathbf{V}_t^{\text{BB}} = \mathbf{V}_t^{\text{BB}} (\mathbf{V}_t^{\text{BB}})^H = \mathbf{I}_{N_T}, \end{aligned} \quad (13)$$

where \mathcal{F} denotes set of matrices satisfying the constant modulus constraints. In [7], an iterative algorithm based on alternating minimization (AM) is proposed to solve the above problem, as summarized in Algorithm 1, which is guaranteed to convergence monotonically to, at least, a local minimum. Using Algorithm 1, if $N_T = 1$, we always have that $\mathbf{V}_t^{\text{BB}} = 1$. Therefore, the precoding vector is updated simply as $\mathbf{V}_t^{\text{RF}} = \mathbf{V}_t^{\text{LS}} \oslash |\mathbf{V}_t^{\text{LS}}|$, i.e., the analog precoding is obtained in a closed-form.

Assuming that \mathbf{V}_t^{LS} is decomposed into \mathbf{V}_t^{RF} and \mathbf{V}_t^{BB} using Algorithm 1, we note that the mutual coherence of the estimated sensing matrix $\hat{\Phi}_T = [\mathbf{V}_1^{\text{RF}} \mathbf{V}_1^{\text{BB}} \mathbf{s}_1, \dots, \mathbf{V}_{T_T}^{\text{RF}} \mathbf{V}_{T_T}^{\text{BB}} \mathbf{s}_{T_T}]^T \bar{\mathbf{A}}_T$ has,

Algorithm 1 HAD design by alternating minimization (AM) [7].

- 1: **Input:** \mathbf{V}_t^{LS}
 - 2: **Initial:** $\mathbf{V}_t^{\text{RF}} = \mathbf{V}_t^{\text{LS}} \oslash |\mathbf{V}_t^{\text{LS}}|$ (\oslash is the element-wise normalization function.)
 - 3: **while** not converged **do**
 - 4: Update $\mathbf{V}_t^{\text{BB}} = \mathbf{P} \mathbf{Q}^H$, where $(\mathbf{V}_t^{\text{LS}})^H \mathbf{V}_t^{\text{RF}} = \mathbf{Q} \mathbf{S} \mathbf{P}^H$
 - 5: Update $\mathbf{V}_t^{\text{RF}} = \mathbf{V}_t^{\text{LS}} (\mathbf{V}_t^{\text{BB}})^H \oslash |\mathbf{V}_t^{\text{LS}} (\mathbf{V}_t^{\text{BB}})^H|$
 - 6: **end while**
 - 7: **Output:** \mathbf{V}_t^{RF} and $\mathbf{V}_t^{\text{BB}} = \mathbf{V}_t^{\text{LS}} \oslash |\mathbf{V}_t^{\text{LS}}| \mathbf{V}_t^{\text{RF}} \mathbf{V}_t^{\text{BB}}$
-

in general, a larger mutual coherence compared to that of \mathbf{F}_T , i.e., $\mu(\hat{\Phi}_T) \geq \mu(\mathbf{F}_T)$.

3.2. Sensing matrix design for Φ_R

Next, we show the design of the receive decoding matrices to minimize the mutual coherence of the sensing matrix $\Phi_R = \mathbf{U}^H \bar{\mathbf{A}}_R \in \mathbb{C}^{T_R N_R \times G_R}$. Using a similar approach as above, let $\mathbf{F}_R \in \mathbb{C}^{T_R N_R \times G_R}$ denote a sensing matrix designed efficiently using, e.g., the methods in [16–19]. The objective is to decompose \mathbf{F}_R into \mathbf{U} and $\bar{\mathbf{A}}_R$ such that the error $\eta_u = \|\mathbf{F}_R - \mathbf{U}^H \bar{\mathbf{A}}_R\|_F^2$ is minimized. Since the dictionary matrix $\bar{\mathbf{A}}_R \in \mathbb{C}^{M_R \times G_R}$ is fixed, the optimal solution for \mathbf{U} that minimizes η_u can be obtained using the least-squared method as

$$\mathbf{U}^{\text{LS}} = [\mathbf{U}_1^{\text{LS}}, \dots, \mathbf{U}_{T_R}^{\text{LS}}] = (\mathbf{F}_R \bar{\mathbf{A}}_R^+)^H \in \mathbb{C}^{M_R \times T_R N_R}, \quad (14)$$

assuming that $G_R \leq M_R$. Using a similar approach as in (13), we update \mathbf{U}_r^{RF} and \mathbf{U}_r^{BB} , $r = 1, \dots, T_R$, as [7]

$$\begin{aligned} \min_{\mathbf{U}_r^{\text{RF}}, \mathbf{U}_r^{\text{BB}}} \quad & \|\mathbf{U}_r^{\text{LS}} - \mathbf{U}_r^{\text{RF}} \mathbf{U}_r^{\text{BB}}\|_F^2 \\ \text{s.t.} \quad & \mathbf{U}_r^{\text{RF}} \in \mathcal{U} \text{ and } (\mathbf{U}_r^{\text{BB}})^H \mathbf{U}_r^{\text{BB}} = \mathbf{U}_r^{\text{BB}} (\mathbf{U}_r^{\text{BB}})^H = \mathbf{I}_{N_R}, \end{aligned} \quad (15)$$

where \mathcal{U} denotes set of matrices satisfying the constant modulus constraints. Problem (15) can be solved using a similar approach as in Algorithm 1. At the end, we have $\hat{\Phi}_R = [\mathbf{U}_1^{\text{RF}} \mathbf{U}_1^{\text{BB}}, \dots, \mathbf{U}_{T_R}^{\text{RF}} \mathbf{U}_{T_R}^{\text{BB}}]^H \bar{\mathbf{A}}_R$.

4. NUMERICAL RESULTS

In this section, we show simulation results evaluating the proposed channel training method. As a benchmark, the precoding and combining matrices are updated using 1) the random training approach [4], termed Random and 2) the hybrid analog-digital training method proposed in [15], termed DFT-based since it updates the analog matrices using sub-matrices of a DFT-matrix. We assume that the given sensing matrices \mathbf{F}_R and \mathbf{F}_T are designed offline using the method proposed in [19]. The performance metrics are 1) the mutual coherence $\mu(\hat{\Phi})$ of the sensing matrix $\hat{\Phi} = (\hat{\Phi}_T \otimes \hat{\Phi}_R)$ and 2) the normalized mean-squared-errors (NMSE) that is defined as $\gamma_{\text{NMSE}} = \mathbb{E}[\|\mathbf{H} - \hat{\mathbf{H}}\|_F^2 / \|\mathbf{H}\|_F^2]$, where \mathbf{H} is the true channel and $\hat{\mathbf{H}}$ is the estimated channel. We define the SNR as $\rho_{\text{SNR}} = \frac{1}{\sigma_n^2}$ (dB) and assume that $\alpha_p \sim \mathcal{CN}(0, 1)$, $\forall p$ and the pilot symbols \mathbf{s}_t are generated randomly following a zero-mean circularly symmetric complex Gaussian distribution such that $\mathbb{E}\{\mathbf{s}_t^H \mathbf{s}_t\} = 1, \forall t$.

In Fig. 1 (a), we show the mutual coherence $\mu(\hat{\Phi})$ of the resulting sensing matrix $\hat{\Phi}$, while assuming $M_T = 128$, $N_R = 8$, $T_R = 2$, $\beta_R = \beta_T = 1$. To see the impact of the hybrid analog-digital decomposition performed by Algorithm 1, in Fig. 1 (a), we show $\mu(\hat{\Phi})$ when the obtained precoding and decoding matrices are implemented in the fully-digital (FD) form, i.e., we set $\mathbf{V}_t = \mathbf{V}_t^{\text{LS}} / \|\mathbf{V}_t^{\text{LS}}\|_F, \forall t$, and $\mathbf{U}_r = \mathbf{U}_r^{\text{LS}} / \|\mathbf{U}_r^{\text{LS}}\|_F, \forall r$, where \mathbf{V}_t^{LS} and \mathbf{U}_r are given by (12) and

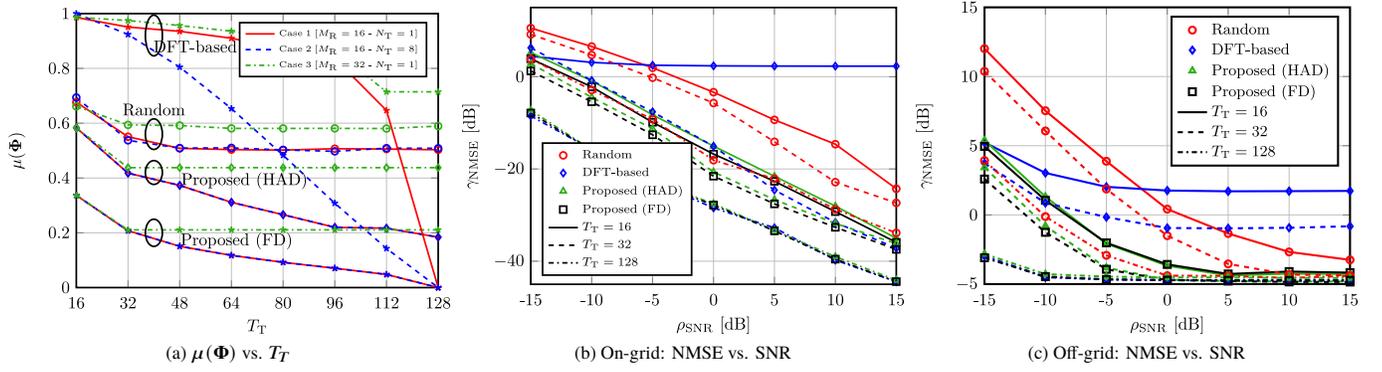


Fig. 1: Performance analysis: (a) $\mu(\Phi)$ vs. T_T , (b) NMSE vs. SNR ($\theta_{R,p} \in \bar{\theta}_R$, $\theta_{T,p} \in \bar{\theta}_T$, $\forall p$), and (c) NMSE vs. SNR ($\theta_{R,p}, \theta_{T,p} \in [0, 2\pi]$, $\forall p$).

(14), respectively. Note that the normalization is done to satisfy the total power constraint of $\|\mathbf{V}_T\|_F = 1$ and $\|\mathbf{U}_R\|_F = 1$.

From Fig. 1, we can see that the proposed training method achieves a mutual coherence that is significantly lower than that of the benchmark methods. As expected above, the achieved mutual coherence $\mu(\Phi)$ is equal irrespective of the number transmit RF chains N_T . Moreover, we note that the hybrid analog-digital (HAD) implementation of our proposed approach always leads to a higher mutual coherence than the FD counterpart. This can be explained by the fact that the objective functions of problems (13) and (15) are always larger than zero, i.e., $\|\mathbf{V}_T^{\text{LS}} - \mathbf{V}_T^{\text{RF}} \mathbf{V}_T^{\text{BB}}\|_F^2 > 0$ and $\|\mathbf{U}_R^{\text{LS}} - \mathbf{U}_R^{\text{RF}} \mathbf{U}_R^{\text{BB}}\|_F^2 > 0$, even when $N_T = 1$. Nonetheless, we can see that the proposed training method still achieves a significantly lower mutual coherence than the benchmark methods. It is important to note that in Case 1 and Case 2, both proposed FD and DFT-based methods achieve the optimal coherence of $\mu(\Phi) = 0$ when $T_T = 128$. To explain this, note that when $\beta_R = \beta_T = 1$, $N_R = 8$, and $T_R = 2$, it implies that $\Phi_R = \mathbf{U}^H \bar{\mathbf{A}}_R \in \mathbb{C}^{T_R N_R \times G_R}$ is a 16×16 matrix and $\Phi_T = \mathbf{V}^T \bar{\mathbf{A}}_T \in \mathbb{C}^{T_T \times G_T}$ is a 128×128 matrix. Here, for square matrices, we can always find a matrix with an optimal mutual coherence, i.e., $\mu(\Phi_T) = 0$ and $\mu(\Phi_R) = 0$ [19], thus $\mu(\Phi) = \max\{\mu(\Phi_T), \mu(\Phi_R)\} = 0$. This shows that whenever $\beta_R = \beta_T = 1$, $T_T = G_T$ and $T_R N_R = G_R$, the optimal hybrid analog-digital training method is achieved by the DFT-based method [15]. It is also important to note that when M_R increases to 32, i.e., Case 3, the mutual coherence of the proposed and random method (resp. DFT-based method) saturates as T_T increases above 32 (resp. 112). The main reason is that when $T_T < 32$ (resp. $T_T < 112$), the mutual coherence $\mu(\Phi)$ seems to be dominated by $\mu(\Phi_T)$. Thus, increasing T_T from 16 to 32 (resp. from 16 to 112), reduces $\mu(\Phi)$. However, when $T_T > 32$ (resp. $T_T > 112$), the mutual coherence $\mu(\Phi)$ seems to be dominated by $\mu(\Phi_R)$. Thus, the mutual coherence $\mu(\Phi)$ does not reduce when we increase T_T .

In Fig. 1 (b), we show simulation results of the averaged NMSE versus SNR, while assuming $M_T = 128$, $N_T = 8$, $M_R = 32$, $N_R = 8$, $T_R = 4$, $P = 4$, $\beta_R = \beta_T = 1$. For this figure, we assume that $\theta_{R,p} \in \bar{\theta}_R, \forall p$, and $\theta_{T,p} \in \bar{\theta}_T, \forall p$, i.e., the AoA and AoD fall perfectly on a $G_R \times G_T$ grid, where $\beta_R = \beta_T = 1$. With this assumption we avoid the impact of the quantization errors on the channel estimation, thus, we focus only on the effectiveness of the training design methods. From Fig. 1 (b), we can see that when $T_T < 128$, the proposed method significantly outperforms the random and the DFT-based training

methods. This can be explained by noting that when $T_T < 128$, the proposed method achieves a mutual coherence $\mu(\Phi)$ that is much lower than that achieved with the competing benchmark methods, as can be seen from Fig. 1 (a). Meanwhile, when the value of T_T increases to 128, i.e., when $T_T = G_T$, the DFT-based method has a similar performance as our proposed training method. It is interesting to note that the HAD implementation of our proposed method does not incur a significant degradation on the estimation performance, as compared to its FD implementation. Even with small T_T , i.e., $T_T = 16$, we can see that this performance degradation is negligible, while it vanishes as T_T increases to 128.

In Fig. 1 (c) we show simulation results of the averaged NMSE versus SNR, while assuming $M_T = 128$, $N_T = 8$, $M_R = 32$, $N_R = 8$, $T_R = 4$, $P = 4$, $\beta_R = \beta_T = 1$. For this figure, we assume that $\theta_{R,p}, \theta_{T,p} \in [0, 2\pi], \forall p$, i.e., the AoA and AoD do not have to lie on the grid. With this assumption, different from Fig. 1 (b), we include the impact of the quantization errors on the channel estimation. From Fig. 1 (c), we can see that the above observations from Fig. 1 (b) hold true as well. The main difference is that the NMSE performance in Fig. 1 (c) is worse than that in Fig. 1 (b). Moreover, the NMSE performance, with all the training design methods, seems to saturate as the SNR value increases. This is rather an expected behavior, since the NMSE performance in this case is lower-bounded by the AoA and AoD quantization errors. Dealing with the quantization errors is out of the scope of this paper, which can be reduced or even eliminated by using off-grid and/or gridless sparse recovery techniques, see [14, 22] for more details.

5. CONCLUSIONS

In CS-based channel estimation methods, channel training vectors should be designed carefully to guarantee recoverability. In this paper, we consider a downlink hybrid analog-digital MIMO system operating over millimeter-wave bands and formulate the CSI estimation problem as a sparse recovery of the dominant channel path parameters. To improve the recoverability guarantee, an open-loop hybrid analog-digital beam-training framework is proposed, which decomposes a given sensing matrix that is designed offline into an analog and digital beamformers, which are then used for the channel training and estimation. Simulations results show that the proposed training method achieves a lower mutual coherence and an improved channel estimation accuracy compared to benchmark methods.

6. REFERENCES

- [1] Z. Pi and F. Khan, "An introduction to millimeter-wave mobile broadband systems," *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 101–107, Jun. 2011.
- [2] R. W. Heath, 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.
- [3] J. Zhang, A. Wiesel, and M. Haardt, "Low rank approximation based hybrid precoding schemes for multi-carrier single-user massive MIMO systems," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Mar. 2016, pp. 3281–3285.
- [4] J. Zhang, I. Podkurkov, M. Haardt, and A. Nadeev, "Channel estimation and training design for hybrid analog-digital multi-carrier single-user massive MIMO systems," in *Proc. 20th International ITG Workshop on Smart Antennas (WSA)*, Mar. 2016, pp. 1–8.
- [5] R. Mendez-Rial, N. González-Prelcic, A. Alkhateeb, and R. W. Heath, "Hybrid MIMO architectures for millimeter wave communications: Phase shifters or switches?," *IEEE Access*, vol. 4, pp. 247–267, Jan. 2016.
- [6] K. Ardah, G. Fodor, Y. C. B. Silva, W. Cruz, and F. R. Cavalcanti, "A unifying design of hybrid beamforming architectures employing phase-shifters or switches," *IEEE Trans. Veh. Technol.*, pp. 1–1, Nov. 2018.
- [7] C. Rusu, R. Mendez-Rial, N. González-Prelcic, and R. W. Heath, "Low complexity hybrid precoding strategies for millimeter wave communication systems," *IEEE Trans. Wireless Commun.*, vol. 15, no. 12, pp. 8380–8393, Dec. 2016.
- [8] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, "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.
- [9] J. W. Choi, B. Shim, Y. Ding, B. Rao, and D. I. Kim, "Compressed sensing for wireless communications: Useful tips and tricks," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 3, pp. 1527–1550, 2017.
- [10] I. Orović, V. Papić, C. Ioana, Xi. Li, and S. Stanković, "Compressive sensing in signal processing: Algorithms and transform domain formulations," *Mathematical Problems in Engineering*, vol. 16, 2016.
- [11] M. Rani, S. B. Dhok, and R. B. Deshmukh, "A systematic review of compressive sensing: Concepts, implementations and applications," *IEEE Access*, vol. 6, pp. 4875–4894, 2018.
- [12] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
- [13] Z. Gao, C. Hu, L. Dai, and Z. Wang, "Channel estimation for millimeter-wave massive MIMO with hybrid precoding over frequency-selective fading channels," *IEEE Commun. Lett.*, vol. 20, no. 6, pp. 1259–1262, Jun. 2016.
- [14] K. Ardah, A. L. F. de Almeida, and M. Haardt, "A gridless CS approach for channel estimation in hybrid massive MIMO systems," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May. 2019.
- [15] J. Lee, G. 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.
- [16] C. Lu, H. Li, and Z. Lin, "Optimized projections for compressed sensing via direct mutual coherence minimization," *Signal Processing*, vol. 151, pp. 45 – 55, 2018.
- [17] V. Abolghasemi, S. Ferdowsi, B. Makkiabadi, and S. Sanei, "On optimization of the measurement matrix for compressive sensing," in *Proc. 18th European Signal Processing Conference*, Aug. 2010, pp. 427–431.
- [18] M. Elad, "Optimized projections for compressed sensing," *IEEE Trans. Signal Process.*, vol. 55, no. 12, pp. 5695–5702, Dec. 2007.
- [19] K. Ardah, M. Pesavento, and M. Haardt, "A novel sensing matrix design for compressed sensing via mutual coherence minimization," in *Proc. IEEE Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, Dec. 2019, (Accepted. Available online at: <http://www5.tu-ilmenau.de/nt/en/publications/index.html>).
- [20] B. L. Sturm and M. G. Christensen, "Comparison of orthogonal matching pursuit implementations," in *Proc. of the 20th European Signal Processing Conference (EUSIPCO)*, Aug. 2012, pp. 220–224.
- [21] M. A. Murray-Lasso, "Alternative methods of calculation of the pseudo inverse of a non full-rank matrix," *Journal of applied research and technology*, vol. 6, pp. 170 – 183, 12 2008.
- [22] Z. Tian, Z. Zhang, and Y. Wang, "Low-complexity optimization for two-dimensional direction-of-arrival estimation via decoupled atomic norm minimization," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Mar. 2017, pp. 3071–3075.