Channel Estimation via Orthogonal Matching Pursuit for Hybrid MIMO Systems in Millimeter Wave Communications

Junho Lee, Gye-Tae Gil, *Member, IEEE*, and Yong H. Lee, *Senior Member, IEEE*

***Abstract*— We propose an efficient open-loop channel estimator for a millimeter-wave (mm-wave) hybrid multiple-input multiple-output (MIMO) system consisting of radio-frequency (RF) beamformers with large antenna arrays followed by a baseband MIMO processor. A sparse signal recovery problem exploiting the sparse nature of mm-wave channels is formulated for channel estimation based on the parametric channel model with quantized angles of departures/arrivals (AoDs/AoAs), called the angle grids. The problem is solved by the orthogonal matching pursuit (OMP) algorithm employing a redundant dictionary consisting of array response vectors with finely quantized angle grids. We suggest the use of non-uniformly quantized angle grids and show that such grids reduce the coherence of the redundant dictionary. The lower and upper bounds of the sum-of-squared errors of the proposed OMP-based estimator are derived analytically: the lower bound is derived by considering the oracle estimator that assumes the knowledge of AoDs/AoAs, and the upper bound is derived based on the results of the OMP performance guarantees. The design of training vectors (or sensing matrix) is particularly important in hybrid MIMO systems, because the RF beamformer prevents the use of independent and identically distributed random training vectors, which are popular in compressed sensing. We design training vectors so that the total coherence of the equivalent sensing matrix is minimized for a given RF beamforming matrix, which is assumed to be unitary. It is observed that the estimation accuracy can be improved significantly by randomly permuting the columns of the RF beamforming matrix. The simulation results demonstrate the advantage of the proposed OMP with a redundant dictionary over the existing methods such as the least squares method and the OMP based on the virtual channel model.**

***Index Terms*— Channel estimation, hybrid RF/baseband processing, millimeter wave communication, orthogonal matching pursuit, sparsity.**

I. Introduction

RECENTLY, millimeter-wave (mm-wave) communication systems are emerging as a promising technology for next-generation wireless communications [1], [2]. By employing large antenna arrays that can be packed into a very small area thanks to the short wave-length, mm-wave systems can achieve the beamforming gain needed to overcome the propagation loss which is higher than the existing micro-wave communication systems. Furthermore, mm-wave systems can utilize the vast bandwidth available in the mm-wave spectrum. These characteristics allow the design of multi-Gbps mm-wave systems, as demonstrated in indoor wireless systems such as wireless local area network (LAN) [3] and personal area network (PAN) [4].

The channels of mm-wave communications are sparse in the sense that impulse responses are dominated by a small number of clusters of significant paths. This sparsity causes rank-deficient channel matrices having a few dominant singular values, and the effective rank of an mm-wave channel, which is the number of dominant singular values, tends to be considerably smaller than the number of transmit/receive antennas [5]. Based on this fact, use of a hybrid MIMO processor consisting of an analog beamformer in radio frequency (RF) domain cascaded with a digital MIMO processor in baseband has been proposed for mm-wave communications [6]–[8]. In a hybrid MIMO processor, the number of RF chains is determined by the effective rank of the channel, and its implementation can be much simpler than conventional MIMO systems that need one RF chain per antenna.

As in conventional multiple-input multiple-output (MIMO) systems, channel state information (CSI) is useful for designing efficient hybrid MIMO processors of mm-wave systems. For example, if CSI is given, the RF beamformer and the baseband MIMO processor can be jointly designed [6], [7], resulting in better performance than the beam training-based hybrid beamformers which first steer the RF beams to angle of departure/arrival (AoD/AoA) directions and then optimize the baseband MIMO processor. However, estimating CSI in mm-wave systems is challenging because the signal-to-noise ratio (SNR) before beamforming is low and the number of antennas is large. To increase the SNR it would be necessary to employ proper beamforming during channel estimation; and to relieve the effect of large antenna size the sparsity of mm-wave channels would be exploited: instead of estimating all the entries of the channel matrix, one would estimate only the AoDs/AoAs of dominant paths and the corresponding path gains.

In mm-wave communication, closed-loop beam training-based methods are proposed for CSI estimation [3], [4], [9]–[11]. These methods first estimate the AoDs/AoAs by closed-loop beam training and then estimate

Manuscript received November 2, 2015; revised March 14, 2016; accepted April 13, 2016. Date of publication April 20, 2016; date of current version June 14, 2016. This work was supported by ICT R&D program of MSIP/IITP [B0101-16-1369, Development of small basestation supporting multiple streams based on LTE-A systems]. This paper was presented at the IEEE Global Communications Conference, Dec. 2014. The associate editor coordinating the review of this paper and approving it for publication was X. Gao. *(Corresponding author: Yong H. Lee.)*

J. Lee and Y. H. Lee are with the Department of Electrical Engineering, Korea Advanced Institute of Science and Technology, Daejeon 305-701, South Korea (e-mail: junho515@kaist.ac.kr; junho515@kaist.ac.kr).

G.-T. Gil is with the Institute for Information Technology Convergence, Korea Advanced Institute of Science and Technology, Daejeon 305-701, South Korea (e-mail: gategil@kaist.ac.kr).

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

Digital Object Identifier 10.1109/TCOMM.2016.2557791

the path gain associated with each pair of AoD and AOA. The closed-loop beam training is a multistage process that can avoid an exhaustive beam search. At each stage the transmitter emits the pilot beams, and the receiver selects the best beam and feeds back its decision. This process starts with wide beams that cover all of the angles of interest and makes the beams fine only around the angles where AoDs/AoAs are present. While closed-loop beam training-based methods have been adopted in practical systems [3], [4], their performance tends to be limited by the training beam patterns (or codebook); furthermore, in multiuser MIMO systems, their training overhead tends to increase linearly with the number of users.

An alternative approach to AoD/AoA estimation of mm-wave channels is to apply the compressed sensing (CS) techniques in [12]–[14], which have been proposed for sparse multipath estimation in angular domain of massive MIMO channels. These CS-based methods are open-loop techniques that perform explicit channel estimation: the transmitter emits pilot vectors for channel estimation, and the receiver estimates the channel from the received pilot signals.[1] The pilot overhead of the CS-based methods is given by $\mathcal{O}(L \ln N_{\mathrm{T}})$, where L is the number of non-zero spatial channel paths (or the sparsity level) and N_{T} is the number of transmit antennas [16], [17].[2] When $L \ll N_{\mathrm{T}}$, which is often true in massive MIMO systems, this pilot overhead of the CS-based methods represents substantial reduction of the overhead as compared with the conventional least squares (LS) method, whose overhead is given by $\mathcal{O}(N_{\mathrm{T}})$. In addition, the pilot overhead $\mathcal{O}(L \ln N_{\mathrm{T}})$ remains the same irrelevant to the number of users in multiuser MIMO systems, due to the open-loop characteristic of the CS techniques. These CS-based channel estimators have been derived based on the virtual angular domain representation of MIMO channels [18], [19], called the *virtual channel model*, which describes the channel with respect to fixed basis functions corresponding to the angles whose resolution is determined by the spatial resolution of arrays. As a result, their estimation accuracy is limited by the array resolution. Another difficulty encountered when applying the CS-based schemes to mm-wave channel estimation is caused by their pilot vectors: they assume conventional microwave communication where the SNR without beamforming is reasonably high and suggest the use of i.i.d. Rademacher vectors as pilots. In hybrid MIMO systems over a mm-wave channel use of such i.i.d. random pilots is difficult because the RF beamformers, which are employed to increase the SNR, yield directional training beams that are correlated with each other.

In this paper, we develop an alternative CS-based open-loop channel estimator for mm-wave hybrid MIMO systems. This scheme is based on the parametric channel model with quantized AoDs/AoAs, called the angle grids, and the pilot vectors of the proposed scheme is generated by the cascade of the RF beamformer and the baseband MIMO processor. For channel estimation, we adopt the grid-based orthogonal matching pursuit (OMP) algorithm which has been proposed for estimating channels of multicarrier underwater acoustic channels [20] and time-varying orthogonal frequency division multiplexing (OFDM) systems [21]. The contributions of this paper are described as follows:

- We formulate a sparse signal recovery problem and develop a grid-based OMP algorithm for estimating channels of hybrid MIMO systems. In the CS formulation, we suggest the use of a *redundant* dictionary consisting of array response vectors with finely quantized angle grids which are non-uniformly distributed in $[0, \pi]$. Specifically, we determine the angle grids, denoted as $\{\varphi_g\}$, so that $\{\cos(\varphi_g)\}$ are uniformly distributed in $[1, -1)$. The angular resolution of the proposed scheme can be much finer than that of the virtual channel model, and the *non-uniformly* distributed angle grids are shown to reduce the coherence of the redundant dictionary.
- The lower- and upper-bounds of the sum-of-squared errors (SSE) of the proposed OMP-based estimator are derived analytically: the lower-bound is derived by considering the oracle estimator that assumes the knowledge of AoDs/AoAs, and the upper-bound is derived based on the results of the OMP performance guarantees. The results indicate that the SSE increases with the number of channel paths (or the sparsity level).
- Efficient pilot vectors are designed for given RF beamformers: we design the baseband MIMO processor so that the total coherence of the equivalent sensing matrix is minimized for a given RF beamforming matrix which is assumed to be unitary. Here, the equivalent sensing matrix is the product of the sensing matrix and the dictionary, and the total coherence is the sum of the squared inner products of all pairs of columns in the equivalent sensing matrix [22]. It is observed that the estimation accuracy can be improved significantly by randomly permuting the columns of the RF beamforming matrix.

The organization of the paper is as follows. Section II presents the system model. In Section III, we formulate the CS based channel estimation problem with a redundant dictionary and present the grid-based OMP algorithm. The SSE is analyzed in Section IV. In Section V, the pilot beam patterns are designed under the beam power constraints for a given RF beamforming matrix. Simulation results showing the advantages of the proposed schemes over the existing methods in terms of both the estimation accuracy and the training overhead are presented in Section VI. Finally, the conclusion is presented in Section VII.

Notations: Bold uppercase $\mathbf{A}$ denotes a matrix and bold lowercase $\mathbf{a}$ denotes a vector. Superscripts $\mathbf{A}^*, \mathbf{A}^T, \mathbf{A}^H, \mathbf{A}^{-1}, \mathbf{A}^\dagger$ denote the conjugate, the transpose, the conjugate transpose, the inverse, and the pseudo-inverse of a matrix $\mathbf{A}$, respectively. $\mathrm{diag}(\mathbf{A}_1, \ldots, \mathbf{A}_N)$ represents a block diagonal matrix whose diagonal entries are given by $\{\mathbf{A}_1, \ldots, \mathbf{A}_N\}$. $\|\mathbf{a}\|_0$ and $\|\mathbf{a}\|_2$ are the $\mathcal{L}_0$ and $\mathcal{L}_2$ norms, respectively, and $\mathbf{a}(n)$ denotes the n-th entry of a vector $\mathbf{a}$. The support of a vector $\mathbf{a}$ is defined as

[1] A closed-loop technique that performs coarse channel estimation by beam training and fine channel estimation by CS is introduced in [15].

[2] This overhead, which is the number of measurements needed for recovering the sparse vector in the CS formulation, has been derived for different settings, which include recovery with high probability using the orthogonal matching pursuit (OMP) in [16] and the basis-pursuit (BP) in [17].

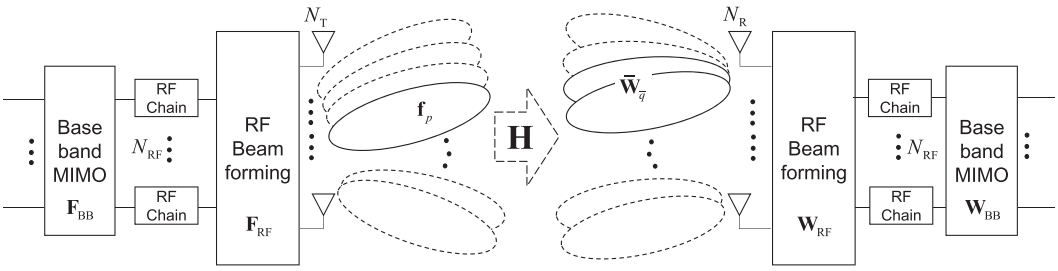

Fig. 1. A mm-wave system employing hybrid MIMO processors.

$\mathrm{supp}(\mathbf{a}) \triangleq \{n|\mathbf{a}(n) \neq 0\}$. $\|\mathbf{A}\|_F$ is the Frobenius norm; $\mathrm{Tr}(\mathbf{A})$ is the trace; and $\mathbf{A}(n)$ and $\mathbf{A}(m,n)$ denote the n-th column and the (m, n)-th entry of a matrix $\mathbf{A}$, respectively. $\mathbf{A}_{\mathcal{I}}$ is the sub-matrix of a matrix $\mathbf{A}$ that only contains the columns whose indices are included in a set of column indices $\mathcal{I}$. Similarly, $\mathbf{a}_{\mathcal{I}}$ is the sub-vector of a vector $\mathbf{a}$ that only contains the entries whose indices are included in a set of indices $\mathcal{I}$. $\mathbf{I}_N$ denotes the $N \times N$ identity matrix and $\mathbf{O}_{M,N}$ denotes the $M \times N$ all-zero matrix. $\mathrm{vec}(\mathbf{A})$ is a vector obtained through the vectorization of a matrix $\mathbf{A}$, and $\mathrm{vec}^{-1}(\mathbf{A})$ represents a matrix obtained by the inverse of vectorization. In addition, for an $N \times N$ square matrix $\mathbf{A}$, $\mathrm{vecd}(\mathbf{A})$ is an N-dimensional vector consisting of the diagonal entries of $\mathbf{A}$ (the n-th entry of $\mathrm{vecd}(\mathbf{A})$ is given by $\mathbf{A}(n,n)$). $\lambda_{\max}(\mathbf{A})$ and $\lambda_{\min}(\mathbf{A})$ represent the maximum and the minimum eigenvalues of $\mathbf{A}$, respectively. For $M \times N$ matrices $\mathbf{A}$ and $\mathbf{B}$, $\mathbf{A} \otimes \mathbf{B}$ denotes the $M^2 \times N^2$ matrix of Kronecker product between $\mathbf{A}$ and $\mathbf{B}$; and $\mathbf{A} \odot \mathbf{B}$ denotes the $M^2 \times N$ matrix of Khatri-Rao product [23] defined as $\mathbf{A} \odot \mathbf{B} = [\mathbf{A}(1) \otimes \mathbf{B}(1), \ldots, \mathbf{A}(N) \otimes \mathbf{B}(N)]$. $\mathbb{E}[\cdot]$ is the expectation operator, and $\Pr\{\mathcal{B}\}$ denotes the probability that an event $\mathcal{B}$ occurs.

II. System Model

In this section, we present the signal model for the open-loop beam training and the channel model in mm-wave communications.

A. Signal Model for Open-Loop Beam Training

We consider the hybrid MIMO system shown in Fig. 1, where the transmitter and the receiver are equipped with N_{T} and N_{R} antennas, respectively, and both of them have N_{RF} RF chains[3] where $N_{\mathrm{RF}} \leq \min(N_{\mathrm{T}}, N_{\mathrm{R}})$. Here, we assume that N_{T} and N_{R} are multiples of N_{RF}, and denote $\frac{N_{\mathrm{T}}}{N_{\mathrm{RF}}}$ and $\frac{N_{\mathrm{R}}}{N_{\mathrm{RF}}}$ by $N_{\mathrm{T}}^{\mathrm{Block}}$ and $N_{\mathrm{R}}^{\mathrm{Block}}$, respectively. The RF beamformers are assumed to be phased array beamformers, realized using analog phase shifters. For channel estimation, the transmitter uses $N_{\mathrm{T}}^{\mathrm{Beam}} \leq N_{\mathrm{T}}$ pilot beam patterns denoted as $\left\{\mathbf{f}_p \in {}^{N_{\mathrm{T}}\times 1} : \|\mathbf{f}_p\|_2^2 = 1, p = 1, \ldots, N_{\mathrm{T}}^{\mathrm{Beam}}\right\}$, and the receiver uses $N_{\mathrm{R}}^{\mathrm{Beam}} \leq N_{\mathrm{R}}$ beam patterns denoted as $\left\{\mathbf{w}_q \in {}^{N_{\mathrm{R}}\times 1} : \|\mathbf{w}_q\|_2^2 = 1, q = 1, \ldots, N_{\mathrm{R}}^{\mathrm{Beam}}\right\}$. Here, we also assume that $N_{\mathrm{T}}^{\mathrm{Beam}}$ and $N_{\mathrm{R}}^{\mathrm{Beam}}$ are multiples of N_{RF}.

[3]For simplicity, we assume the same number of RF chains at the transmitter and the receiver. The proposed framework can be extended to the case where different numbers of RF chains are employed.

During the training period, the transmitter successively emits its training beams $\{\mathbf{f}_p\}$, and at the receiver each training beam is received through its beam patterns $\{\mathbf{w}_q\}$. Since the receiver has N_{RF} RF chains, it can generate N_{RF} beams simultaneously and receives the vector $\mathbf{y}_{\bar{q},p} \in \mathbb{C}^{N_{\mathrm{RF}}\times 1}$, $\bar{q} \in \left\{1, \ldots, \frac{N_{\mathrm{R}}^{\mathrm{Beam}}}{N_{\mathrm{RF}}}\right\}$, for the p-th transmit beam. Specifically, the $\bar{q}$-th received vector for the p-th transmit beam is given by

$$\mathbf{y}_{\bar{q},p} = \bar{\mathbf{W}}_{\bar{q}}^H \mathbf{H}\mathbf{f}_p x_p + \bar{\mathbf{W}}_{\bar{q}}^H \mathbf{n}_{\bar{q},p}, \tag{1}$$

where x_p is the transmitted pilot symbol, $\bar{\mathbf{W}}_{\bar{q}} = \left[\mathbf{w}_{(\bar{q}-1)N_{\mathrm{RF}}+1} \ldots \mathbf{w}_{\bar{q}N_{\mathrm{RF}}}\right] \in \mathbb{C}^{N_{\mathrm{R}}\times N_{\mathrm{RF}}}$, $\mathbf{H} \in \mathbb{C}^{N_{\mathrm{R}}\times N_{\mathrm{T}}}$ represents the channel matrix, and $\mathbf{n}_{\bar{q},p} \in \mathbb{C}^{N_{\mathrm{R}}\times 1}$ is a noise vector with $\mathcal{CN}(0, \sigma_n^2 \mathbf{I}_{N_{\mathrm{R}}})$. Collecting $\mathbf{y}_{\bar{q},p}$ for $\bar{q} \in \left\{1, \ldots, \frac{N_{\mathrm{R}}^{\mathrm{Beam}}}{N_{\mathrm{RF}}}\right\}$, we get $\mathbf{y}_p \in \mathbb{C}^{N_{\mathrm{R}}^{\mathrm{Beam}}\times 1}$ given by

$$\mathbf{y}_p = \mathbf{W}^H \mathbf{H}\mathbf{f}_p x_p + \mathrm{diag}\left(\bar{\mathbf{W}}_1^H, \ldots, \bar{\mathbf{W}}_{N_{\mathrm{R}}^{\mathrm{Beam}}/N_{\mathrm{RF}}}^H\right) \times \left[\mathbf{n}_{1,p}^T, \ldots, \mathbf{n}_{N_{\mathrm{R}}^{\mathrm{Beam}}/N_{\mathrm{RF}},p}^T\right]^T, \tag{2}$$

where $\mathbf{W} = \left[\bar{\mathbf{W}}_1, \ldots, \bar{\mathbf{W}}_{N_{\mathrm{R}}^{\mathrm{Beam}}/N_{\mathrm{RF}}}\right] \in \mathbb{C}^{N_{\mathrm{R}}\times N_{\mathrm{R}}^{\mathrm{Beam}}}$. To represent the received signals for all transmit beams, we collect $\mathbf{y}_p$ for $p \in \{1, \ldots, N_{\mathrm{T}}^{\mathrm{Beam}}\}$ yielding

$$\mathbf{Y} = \mathbf{W}^H \mathbf{H}\mathbf{F}\mathbf{X} + \mathbf{N}, \tag{3}$$

where $\mathbf{Y} = \left[\mathbf{y}_1, \ldots, \mathbf{y}_{N_{\mathrm{T}}^{\mathrm{Beam}}}\right] \in \mathbb{C}^{N_{\mathrm{R}}^{\mathrm{Beam}}\times N_{\mathrm{T}}^{\mathrm{Beam}}}$, $\mathbf{F} = \left[\mathbf{f}_1, \ldots, \mathbf{f}_{N_{\mathrm{T}}^{\mathrm{Beam}}}\right] \in \mathbb{C}^{N_{\mathrm{T}}\times N_{\mathrm{T}}^{\mathrm{Beam}}}$, and $\mathbf{N} \in \mathbb{C}^{N_{\mathrm{R}}^{\mathrm{Beam}}\times N_{\mathrm{T}}^{\mathrm{Beam}}}$ is the noise matrix given by

$$\mathrm{diag}\left(\bar{\mathbf{W}}_1^H, \ldots, \bar{\mathbf{W}}_{N_{\mathrm{R}}^{\mathrm{Beam}}/N_{\mathrm{RF}}}^H\right) \left[\left[\mathbf{n}_{1,1}^T, \ldots, \mathbf{n}_{N_{\mathrm{R}}^{\mathrm{Beam}}/N_{\mathrm{RF}},1}^T\right]^T, \ldots, \left[\mathbf{n}_{1,N_{\mathrm{T}}^{\mathrm{Beam}}}^T, \ldots, \mathbf{n}_{N_{\mathrm{R}}^{\mathrm{Beam}}/N_{\mathrm{RF}},N_{\mathrm{T}}^{\mathrm{Beam}}}^T\right]^T\right].$$

The matrix $\mathbf{X} \in \mathbb{C}^{N_{\mathrm{T}}^{\mathrm{Beam}}\times N_{\mathrm{T}}^{\mathrm{Beam}}}$ is a diagonal matrix with $\{x_p\}$ on its diagonal. Throughout the paper, we assume identical pilot symbols so that $\mathbf{X} = \sqrt{P}\mathbf{I}_{N_{\mathrm{T}}^{\mathrm{Beam}}}$ where P is the pilot power.

In the hybrid MIMO framework, the transmit and receive processing matrices are decomposed as $\mathbf{F} = \mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}$ and $\mathbf{W} = \mathbf{W}_{\mathrm{RF}}\mathbf{W}_{\mathrm{BB}}$, and the received signal (3) is rewritten as

$$\mathbf{Y} = \sqrt{P}\mathbf{W}_{\mathrm{BB}}^H \mathbf{W}_{\mathrm{RF}}^H \mathbf{H}\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}} + \mathbf{N}, \tag{4}$$

where $\mathbf{F}_{\mathrm{RF}} \in \mathbb{C}^{N_{\mathrm{T}}\times N_{\mathrm{T}}}$ and $\mathbf{W}_{\mathrm{RF}} \in \mathbb{C}^{N_{\mathrm{R}}\times N_{\mathrm{R}}}$ denote the transmit and receive RF beamforming matrices, respectively, and

$\mathbf{F}_{\mathrm{BB}} \in \mathbb{C}^{N_{\mathrm{T}} \times N_{\mathrm{T}}^{\mathrm{Beam}}}$ and $\mathbf{W}_{\mathrm{BB}} \in \mathbb{C}^{N_{\mathrm{R}} \times N_{\mathrm{R}}^{\mathrm{Beam}}}$ denote the transmit and receive baseband processing matrices, respectively. These matrices are designed under the following assumptions: i) $\mathbf{F}_{\mathrm{RF}}$ and $\mathbf{W}_{\mathrm{RF}}$ are unitary matrices whose entries have unit magnitude; ii) N_{T}-beam RF pilot beams generated by the columns of $\mathbf{F}_{\mathrm{RF}}$ cover the full range of AoDs, $[-\pi, \pi]$; and iii) N_{R}-beam RF pilot beams generated by the columns of $\mathbf{W}_{\mathrm{RF}}$ cover the full range of AoAs, $[-\pi, \pi]$; iv) since there are N_{RF} RF chains, $\mathbf{F}_{\mathrm{RF}}$ and $\mathbf{W}_{\mathrm{RF}}$ are partitioned into $N_{\mathrm{T}}^{\mathrm{Block}}$ and $N_{\mathrm{R}}^{\mathrm{Block}}$ sub-RF beams: $\mathbf{F}_{\mathrm{RF}} = \left[\mathbf{F}_{\mathrm{RF},1}, \ldots, \mathbf{F}_{\mathrm{RF},N_{\mathrm{T}}^{\mathrm{Block}}}\right]$ and $\mathbf{W}_{\mathrm{RF}} = \left[\mathbf{W}_{\mathrm{RF},1}, \ldots, \mathbf{W}_{\mathrm{RF},N_{\mathrm{T}}^{\mathrm{Block}}}\right]$ where $\mathbf{F}_{\mathrm{RF},\bar{p}} \in \mathbb{C}^{N_{\mathrm{T}} \times N_{\mathrm{RF}}}$ and $\mathbf{W}_{\mathrm{RF},\bar{q}} \in \mathbb{C}^{N_{\mathrm{R}} \times N_{\mathrm{RF}}}$. Similarly, $\mathbf{F}_{\mathrm{BB}}$ and $\mathbf{W}_{\mathrm{BB}}$ are block diagonal matrices given by $\mathbf{F}_{\mathrm{BB}} = \mathrm{diag}\left(\mathbf{F}_{\mathrm{BB},1}, \ldots, \mathbf{F}_{\mathrm{BB},N_{\mathrm{T}}^{\mathrm{Block}}}\right)$ and $\mathbf{W}_{\mathrm{BB}} = \mathrm{diag}\left(\mathbf{W}_{\mathrm{BB},1}, \ldots, \mathbf{W}_{\mathrm{BB},N_{\mathrm{R}}^{\mathrm{Block}}}\right)$, where $\mathbf{F}_{\mathrm{BB},\bar{p}} \in \mathbb{C}^{N_{\mathrm{RF}} \times \left(N_{\mathrm{T}}^{\mathrm{Beam}}/N_{\mathrm{T}}^{\mathrm{Block}}\right)}$ and $\mathbf{W}_{\mathrm{BB},\bar{q}} \in \mathbb{C}^{N_{\mathrm{RF}} \times \left(N_{\mathrm{R}}^{\mathrm{Beam}}/N_{\mathrm{R}}^{\mathrm{Block}}\right)}$. Therefore, $\mathbf{F} = \mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}$ can be represented as

$$\mathbf{F} = \left[\mathbf{F}_1, \ldots, \mathbf{F}_{N_{\mathrm{T}}^{\mathrm{Block}}}\right], \tag{5}$$

where $\mathbf{F}_{\bar{p}} = \mathbf{F}_{\mathrm{RF},\bar{p}}\mathbf{F}_{\mathrm{BB},\bar{p}}$ and $1 \le \bar{p} \le N_{\mathrm{T}}^{\mathrm{Block}}$. Similarly $\mathbf{W} = \mathbf{W}_{\mathrm{RF}}\mathbf{W}_{\mathrm{BB}}$ can be represented as

$$\mathbf{W} = \left[\mathbf{W}_1, \ldots, \mathbf{W}_{N_{\mathrm{R}}^{\mathrm{Block}}}\right], \tag{6}$$

where $\mathbf{W}_{\bar{q}} = \mathbf{W}_{\mathrm{RF},\bar{q}}\mathbf{W}_{\mathrm{BB},\bar{q}}$ and $1 \le \bar{q} \le N_{\mathrm{R}}^{\mathrm{Block}}$.[4]

B. Channel Model

We use the parametric channel model [6], [24] given by

$$\mathbf{H} = \sqrt{\frac{N_{\mathrm{T}}N_{\mathrm{R}}}{L}} \sum_{l=1}^{L} \alpha_l \mathbf{a}_{\mathrm{R}}\left(\theta_l^{\mathrm{r}}\right) \mathbf{a}_{\mathrm{T}}^H\left(\theta_l^{\mathrm{t}}\right), \tag{7}$$

where L is the number of scatterers that is assumed to be less than the antenna sizes, $L < \min\left(N_{\mathrm{T}}, N_{\mathrm{R}}\right)$, α_l is the complex gain, and θ_l^{r} and θ_l^{t} are the AoA and AoD of the l-th path, respectively. Here, we assume the use of a uniform linear array (ULA) whose array response vectors are denoted as $\mathbf{a}_{\mathrm{R}}\left(\theta_l^{\mathrm{r}}\right) \in \mathbb{C}^{N_{\mathrm{R}} \times 1}$ for the receiver and $\mathbf{a}_{\mathrm{T}}\left(\theta_l^{\mathrm{t}}\right) \in \mathbb{C}^{N_{\mathrm{T}} \times 1}$ for the transmitter. Dropping the subscript/superscript (r, t, l), the array response vector of an ULA is represented as

$$\mathbf{a}\left(\theta\right) = \frac{1}{\sqrt{N}}\left[1, e^{-j\frac{2\pi}{\lambda}d\cos(\theta)}, \ldots, e^{-j\frac{2\pi}{\lambda}(N-1)d\cos(\theta)}\right]^T, \tag{8}$$

where λ is the signal wavelength, and d is the antenna spacing which is assumed to be $d = \frac{\lambda}{2}$. Note that the array response vector is an even function of θ.

For simplicity, each scatterer is assumed to contribute a single propagation path. The channel gains $\{\alpha_l\}_{l=1}^{L}$ are modeled by i.i.d. random variables with distribution $\mathcal{CN}\left(0, \sigma_\alpha^2\right)$. The AoAs and AoDs are modeled by the Laplacian distribution whose mean is uniformly distributed over $[0, 2\pi)$, and angular standard deviation is σ_{AS} [6], [24]. The channel model in (7) can be written in matrix form as

$$\mathbf{H} = \mathbf{A}_{\mathrm{R}}\mathbf{H}_{\mathrm{a}}\mathbf{A}_{\mathrm{T}}^H, \tag{9}$$

where $\mathbf{H}_{\mathrm{a}} = \sqrt{\frac{N_{\mathrm{T}}N_{\mathrm{R}}}{L}}\mathrm{diag}\left(\alpha_1, \ldots, \alpha_L\right)$, $\mathbf{A}_{\mathrm{R}} = \left[\mathbf{a}_{\mathrm{R}}\left(\theta_1^{\mathrm{r}}\right), \ldots, \mathbf{a}_{\mathrm{R}}\left(\theta_L^{\mathrm{r}}\right)\right] \in \mathbb{C}^{N_{\mathrm{R}} \times L}$, and $\mathbf{A}_{\mathrm{T}} = \left[\mathbf{a}_{\mathrm{T}}\left(\theta_1^{\mathrm{t}}\right), \ldots, \mathbf{a}_{\mathrm{T}}\left(\theta_L^{\mathrm{t}}\right)\right] \in \mathbb{C}^{N_{\mathrm{T}} \times L}$. In vector form, (9) can be rewritten as

$$\mathrm{vec}\left(\mathbf{H}\right) = \left(\mathbf{A}_{\mathrm{T}}^* \otimes \mathbf{A}_{\mathrm{R}}\right) \cdot \mathrm{vec}\left(\mathbf{H}_{\mathrm{a}}\right) \tag{10}$$

$$= \left(\mathbf{A}_{\mathrm{T}}^* \odot \mathbf{A}_{\mathrm{R}}\right) \cdot \mathrm{vecd}\left(\mathbf{H}_{\mathrm{a}}\right), \tag{11}$$

where the first equality follows from the identity, $\mathrm{vec}\left(\mathbf{ABC}\right) = \left(\mathbf{C}^T \otimes \mathbf{A}\right) \cdot \mathrm{vec}\left(\mathbf{B}\right)$; the second equality holds because $\mathbf{H}_{\mathrm{a}}$ is a diagonal matrix and $\mathbf{A}_{\mathrm{T}}^* \odot \mathbf{A}_{\mathrm{R}} \in \mathbb{C}^{N_{\mathrm{T}}N_{\mathrm{R}} \times L}$.

C. LS and Oracle Estimators

To formulate the channel estimation problem, it is necessary to vectorize the received signal matrix $\mathbf{Y}$ in (4). Denoting $\mathrm{vec}\left(\mathbf{Y}\right)$ by $\bar{\mathbf{y}} \in \mathbb{C}^{N_{\mathrm{T}}^{\mathrm{Beam}}N_{\mathrm{R}}^{\mathrm{Beam}} \times 1}$, (4) is rewritten as

$$\bar{\mathbf{y}} = \sqrt{P}\left(\mathbf{F}_{\mathrm{BB}}^T\mathbf{F}_{\mathrm{RF}}^T \otimes \mathbf{W}_{\mathrm{BB}}^H\mathbf{W}_{\mathrm{RF}}^H\right) \cdot \mathrm{vec}\left(\mathbf{H}\right) + \bar{\mathbf{n}}$$

$$= \sqrt{P}\mathbf{Q} \cdot \mathrm{vec}\left(\mathbf{H}\right) + \bar{\mathbf{n}}, \tag{12}$$

where the first equality comes from the identity, $\mathrm{vec}\left(\mathbf{ABC}\right) = \left(\mathbf{C}^T \otimes \mathbf{A}\right) \cdot \mathrm{vec}\left(\mathbf{B}\right)$, $\bar{\mathbf{n}} = \mathrm{vec}\left(\mathbf{N}\right)$, and $\mathbf{Q} \triangleq \left(\mathbf{F}_{\mathrm{BB}}^T\mathbf{F}_{\mathrm{RF}}^T \otimes \mathbf{W}_{\mathrm{BB}}^H\mathbf{W}_{\mathrm{RF}}^H\right) \in \mathbb{C}^{N_{\mathrm{T}}^{\mathrm{Beam}}N_{\mathrm{R}}^{\mathrm{Beam}} \times N_{\mathrm{T}}N_{\mathrm{R}}}$. Given (12), a natural approach to estimating $\mathrm{vec}\left(\mathbf{H}\right)$ is to use the LS method which directly estimates the $N_{\mathrm{T}}N_{\mathrm{R}}$ entries of $\mathrm{vec}\left(\mathbf{H}\right)$: the LS estimate, denoted as $\mathrm{vec}\left(\mathbf{H}^{\mathrm{LS}}\right)$, is given by

$$\mathrm{vec}\left(\mathbf{H}^{\mathrm{LS}}\right) = \frac{1}{\sqrt{P}}\left(\mathbf{Q}^H\mathbf{Q}\right)^{-1}\mathbf{Q}^H\bar{\mathbf{y}}. \tag{13}$$

The LS solution requires $N_{\mathrm{T}}^{\mathrm{Beam}}N_{\mathrm{R}}^{\mathrm{Beam}} \ge N_{\mathrm{T}}N_{\mathrm{R}}$ so that $\mathbf{Q}^H\mathbf{Q}$ has full rank. Use of the LS solution for mm-wave communication is difficult because $(N_{\mathrm{T}}, N_{\mathrm{R}})$ are usually large integers and the dimension of the estimation problem $(N_{\mathrm{T}}N_{\mathrm{R}})$ can be excessively large. This difficulty can be overcome in the CS approach because the number of entries to be estimated in the CS formulation is proportional to the sparsity level L which is much less than $N_{\mathrm{T}}N_{\mathrm{R}}$.

The oracle estimator of $\mathbf{H}$ estimates the diagonal channel gain matrix $\mathbf{H}_{\mathrm{a}}$ in (9) under the assumption that AoDs and AoAs (or equivalently, $\mathbf{A}_{\mathrm{T}}$ and $\mathbf{A}_{\mathrm{R}}$) are exactly known [25]. To derive the oracle estimator, $\bar{\mathbf{y}}$ in (12) is rewritten as follows: using (11) in (12)

$$\bar{\mathbf{y}} = \sqrt{P}\mathbf{Q} \cdot \left(\mathbf{A}_{\mathrm{T}}^* \odot \mathbf{A}_{\mathrm{R}}\right) \cdot \mathrm{vecd}\left(\mathbf{H}_{\mathrm{a}}\right) + \bar{\mathbf{n}}$$

$$= \sqrt{P}\mathbf{Q}^{\mathrm{o}} \cdot \mathrm{vecd}\left(\mathbf{H}_{\mathrm{a}}\right) + \bar{\mathbf{n}}, \tag{14}$$

where $\mathbf{Q}^{\mathrm{o}} \triangleq \mathbf{Q} \cdot \left(\mathbf{A}_{\mathrm{T}}^* \odot \mathbf{A}_{\mathrm{R}}\right) \in \mathbb{C}^{N_{\mathrm{T}}^{\mathrm{Beam}}N_{\mathrm{R}}^{\mathrm{Beam}} \times L}$. The matrix $\mathbf{Q}^{\mathrm{o}}$ can be represented as

$$\mathbf{Q}^{\mathrm{o}} = \left(\mathbf{F}_{\mathrm{BB}}^T\mathbf{F}_{\mathrm{RF}}^T\mathbf{A}_{\mathrm{T}}^* \odot \mathbf{W}_{\mathrm{BB}}^H\mathbf{W}_{\mathrm{RF}}^H\mathbf{A}_{\mathrm{R}}\right), \tag{15}$$

because $\mathbf{Q} = \mathbf{F}_{\mathrm{BB}}^T\mathbf{F}_{\mathrm{RF}}^T \otimes \mathbf{W}_{\mathrm{BB}}^H\mathbf{W}_{\mathrm{RF}}^H$ and $\left(\mathbf{A} \otimes \mathbf{B}\right)\left(\mathbf{C} \odot \mathbf{D}\right) = \mathbf{AC} \odot \mathbf{BD}$. The oracle estimator estimates $\mathrm{vecd}\left(\mathbf{H}_{\mathrm{a}}\right)$ in

[4]The partition of $\mathbf{W}$ in (6) is different from the partition in (2): the former partitions $\mathbf{W}$ into $N_{\mathrm{R}}^{\mathrm{Block}}$ sub-matrices, while the latter partitions $\mathbf{W}$ into $N_{\mathrm{R}}^{\mathrm{Beam}}/N_{\mathrm{RF}}$ sub-matrices.

the LS sense. From (14), the LS estimate of vecd $(\mathbf{H}_{\mathrm{a}})$, denoted as vec $\left(\mathbf{H}_{\mathrm{a}}^{\mathrm{o}}\right)$, is given by

$$\operatorname{vecd}\left(\mathbf{H}_{\mathrm{a}}^{\mathrm{o}}\right)=\frac{1}{\sqrt{P}}\left(\left(\mathbf{Q}^{\mathrm{o}}\right)^{H}\mathbf{Q}^{\mathrm{o}}\right)^{-1}\left(\mathbf{Q}^{\mathrm{o}}\right)^{H}\bar{\mathbf{y}}. \tag{16}$$

The oracle estimator requires $N_{\mathrm{T}}^{\mathrm{Beam}}N_{\mathrm{R}}^{\mathrm{Beam}}\geq L$ so that $(\mathbf{Q}^{\mathrm{o}})^{H}\mathbf{Q}^{\mathrm{o}}$ has full rank. Then the oracle estimate of $\mathbf{H}$ is given by

$$\mathbf{H}^{\mathrm{o}}=\mathbf{A}_{\mathrm{R}}\mathbf{H}_{\mathrm{a}}^{\mathrm{o}}\mathbf{A}_{\mathrm{T}}^{H}, \tag{17}$$

where $\mathbf{H}_{\mathrm{a}}^{\mathrm{o}}$ is the diagonal matrix with diagonal entries given by vecd $\left(\mathbf{H}_{\mathrm{a}}^{\mathrm{o}}\right)$. The oracle estimator is useful for examining the performance lower-bound of a CS-based channel estimator.

III. Proposed Sparse Channel Estimation

A. Problem Formulation

To apply the OMP algorithm to channel estimation, we first select the set of quantized angle parameters, called the grids, defined as

$$\mathcal{G}=\left\{\varphi_g:\varphi_g\in[0,\pi],g=1,\ldots,G\right\}, \tag{18}$$

where $\varphi_1=0$ and $\varphi_G=\pi$. The grids $\left\{\varphi_g\right\}$ in (18) are determined so that $\left\{\cos\left(\varphi_g\right)\right\}$ that appear in the definition of the array response vector in (8) are uniformly distributed in $[1,-1)$: specifically, φ_g is determined to satisfy

$$\cos\left(\varphi_g\right)=\frac{2}{G}(g-1)-1, \tag{19}$$

for $g\in\{1,2,\ldots,G\}$. These grids are distributed non-uniformly in the angular interval $[0,\pi]$. Note that the angles in (18) are limited to $[0,\pi]$ because the array response vector in (8) is an even function; any $\varphi_g\in[-\pi,0]$ can be replaced with $-\varphi_g$, giving the same array response vector. It is assumed that the number of grids G is larger than the antenna size, i.e., $G\geq\max\{N_{\mathrm{T}},N_{\mathrm{R}}\}$. Collecting all the array response vectors with angles $\left\{\varphi_g\right\}$, we define the array response matrices $\bar{\mathbf{A}}_{\mathrm{T}}=[\mathbf{a}_{\mathrm{T}}(\varphi_1),\ldots,\mathbf{a}_{\mathrm{T}}(\varphi_G)]$ and $\bar{\mathbf{A}}_{\mathrm{R}}=[\mathbf{a}_{\mathrm{R}}(\varphi_1),\ldots,\mathbf{a}_{\mathrm{R}}(\varphi_G)]$, which constitute the dictionary of our CS formulation. The rows of these array response matrices are orthogonal, as shown below.

Lemma 1: When the angle grids $\left\{\varphi_g\right\}$ satisfy (19) and $d=\frac{\lambda}{2}$, the array response matrices $\bar{\mathbf{A}}_{\mathrm{T}}$ and $\bar{\mathbf{A}}_{\mathrm{R}}$ meet the following:

$$\bar{\mathbf{A}}_{\mathrm{T}}\bar{\mathbf{A}}_{\mathrm{T}}^{H}=\frac{G}{N_{\mathrm{T}}}\mathbf{I}_{N_{\mathrm{T}}}\ \textit{and}\ \bar{\mathbf{A}}_{\mathrm{R}}\bar{\mathbf{A}}_{\mathrm{R}}^{H}=\frac{G}{N_{\mathrm{R}}}\mathbf{I}_{N_{\mathrm{R}}}. \tag{20}$$

Proof: From (8) and (19), the n-th column of $\bar{\mathbf{A}}_{\mathrm{T}}^{T}\in\mathbb{C}^{G\times N_{\mathrm{T}}}$ can be written as

$$\bar{\mathbf{A}}_{\mathrm{T}}^{T}(n)=\frac{1}{\sqrt{N_{\mathrm{T}}}}\left[e^{j\pi(n-1)},e^{-j\pi\left(\frac{2}{G}-1\right)(n-1)},\ldots,e^{-j\pi\left(\frac{2}{G}(g-1)-1\right)(n-1)},\ldots,e^{-j\pi\left(\frac{2}{G}(G-1)-1\right)(n-1)}\right]^{T}, \tag{21}$$

and the (m,n)-th entry of $\bar{\mathbf{A}}_{\mathrm{T}}\bar{\mathbf{A}}_{\mathrm{T}}^{H}\in\mathbb{C}^{N_{\mathrm{T}}\times N_{\mathrm{T}}}$ is given by

$$\begin{aligned}\left[\bar{\mathbf{A}}_{\mathrm{T}}\bar{\mathbf{A}}_{\mathrm{T}}^{H}\right](m,n)&=\frac{1}{N_{\mathrm{T}}}\left(e^{j\pi(m-n)}+e^{-j\pi\left(\frac{2}{G}-1\right)(m-n)}\right.\\&\qquad\left.+\cdots+e^{-j\pi\left(\frac{2}{G}(G-1)-1\right)(m-n)}\right)\\&=\sum\nolimits_{g=1}^{G}e^{-j\pi\left(\frac{2}{G}(g-1)-1\right)(m-n)}.\end{aligned}$$

Thus, the main $(m=n)$ and off-diagonal $(m\neq n)$ entries of $\bar{\mathbf{A}}_{\mathrm{T}}\bar{\mathbf{A}}_{\mathrm{T}}^{H}$ are given by $\frac{G}{N_{\mathrm{T}}}$ and $\pm\frac{e^{-j2\pi(m-n)}-1}{e^{-j2\pi(m-n)/G}-1}=0$, respectively. This indicates $\bar{\mathbf{A}}_{\mathrm{T}}\bar{\mathbf{A}}_{\mathrm{T}}^{H}=\frac{G}{N_{\mathrm{T}}}\mathbf{I}_{N_{\mathrm{T}}}$. In a similar manner, $\bar{\mathbf{A}}_{\mathrm{R}}\bar{\mathbf{A}}_{\mathrm{R}}^{H}=\frac{G}{N_{\mathrm{R}}}\mathbf{I}_{N_{\mathrm{R}}}$ can be proved. □

This type of orthogonality does not hold for conventional array response matrices with uniformly distributed angle grids in $[0,\pi]$ [10]. It will be shown in Sections IV and V that the orthogonality property in Lemma 1 helps reduce the coherence of the proposed CS formulation and design the pilot beam patterns.

Using the array response matrices, the channel matrix $\mathbf{H}$ in (9) can be represented as

$$\begin{aligned}\mathbf{H}&=\bar{\mathbf{A}}_{\mathrm{R}}\bar{\mathbf{H}}_{\mathrm{a}}\bar{\mathbf{A}}_{\mathrm{T}}^{H}+\mathbf{E}\\&=\bar{\mathbf{H}}+\mathbf{E},\end{aligned} \tag{22}$$

where $\bar{\mathbf{H}}\triangleq\bar{\mathbf{A}}_{\mathrm{R}}\bar{\mathbf{H}}_{\mathrm{a}}\bar{\mathbf{A}}_{\mathrm{T}}^{H}$, which is called the *quantized*-channel, is the approximated channel matrix defined on the quantized angle space; $\bar{\mathbf{H}}_{\mathrm{a}}\in\mathbb{C}^{G\times G}$ is an L-sparse matrix having L non-zero entries corresponding to the AoDs/AoAs; and $\mathbf{E}\in\mathbb{C}^{N_{\mathrm{R}}\times N_{\mathrm{T}}}$ is the quantization error matrix caused by quantizing the AoDs/AoAs by the grid points in $\mathcal{G}$. Note that $\bar{\mathbf{H}}_{\mathrm{a}}\in\mathbb{C}^{G\times G}$ is not a diagonal matrix, unlike the diagonal matrix $\mathbf{H}_{\mathrm{a}}\in\mathbb{C}^{L\times L}$ in (9). The first term in the right-hand-side (RHS) of (22) reduces to the virtual channel model when $G=N_{\mathrm{T}}=N_{\mathrm{R}}$ in [18], [19]. Referring to (10), (22) can be rewritten in vector form as

$$\operatorname{vec}(\mathbf{H})=\left(\bar{\mathbf{A}}_{\mathrm{T}}^{*}\otimes\bar{\mathbf{A}}_{R}\right)\cdot\operatorname{vec}\left(\bar{\mathbf{H}}_{\mathrm{a}}\right)+\operatorname{vec}(\mathbf{E}) \tag{23}$$

Using (23) in (12), we have

$$\begin{aligned}\bar{\mathbf{y}}&=\sqrt{P}\mathbf{Q}\left(\bar{\mathbf{A}}_{\mathrm{T}}^{*}\otimes\bar{\mathbf{A}}_{\mathrm{R}}\right)\cdot\operatorname{vec}\left(\bar{\mathbf{H}}_{\mathrm{a}}\right)+\bar{\mathbf{e}}+\bar{\mathbf{n}}\\&=\sqrt{P}\bar{\mathbf{Q}}\cdot\operatorname{vec}\left(\bar{\mathbf{H}}_{\mathrm{a}}\right)+\bar{\mathbf{e}}+\bar{\mathbf{n}},\end{aligned} \tag{24}$$

where $\bar{\mathbf{Q}}\triangleq\mathbf{Q}\cdot\left(\bar{\mathbf{A}}_{\mathrm{T}}^{*}\otimes\bar{\mathbf{A}}_{\mathrm{R}}\right)\in\mathbb{C}^{N_{\mathrm{T}}^{\mathrm{Beam}}N_{\mathrm{R}}^{\mathrm{Beam}}\times G^{2}}$ and $\bar{\mathbf{e}}\triangleq\sqrt{P}\mathbf{Q}\cdot\operatorname{vec}(\mathbf{E})\in\mathbb{C}^{N_{\mathrm{T}}^{\mathrm{Beam}}N_{\mathrm{R}}^{\mathrm{Beam}}\times 1}$. The matrix $\bar{\mathbf{Q}}$ can be rewritten as

$$\bar{\mathbf{Q}}=\left(\mathbf{F}_{\mathrm{BB}}^{T}\mathbf{F}_{\mathrm{RF}}^{T}\bar{\mathbf{A}}_{\mathrm{T}}^{*}\otimes\mathbf{W}_{\mathrm{BB}}^{H}\mathbf{W}_{\mathrm{RF}}^{H}\bar{\mathbf{A}}_{\mathrm{R}}\right), \tag{25}$$

because $\mathbf{Q}=\mathbf{F}_{\mathrm{BB}}^{T}\mathbf{F}_{\mathrm{RF}}^{T}\otimes\mathbf{W}_{\mathrm{BB}}^{H}\mathbf{W}_{\mathrm{RF}}^{H}$ and $(\mathbf{A}\otimes\mathbf{B})(\mathbf{C}\otimes\mathbf{D})=\mathbf{AC}\otimes\mathbf{BD}$. Estimating the L-sparse vector $\operatorname{vec}\left(\bar{\mathbf{H}}_{\mathrm{a}}\right)\in\mathbb{C}^{G^{2}\times 1}$ in (24) is a CS problem with a redundant dictionary, because $\left(\bar{\mathbf{A}}_{\mathrm{T}}^{*}\otimes\bar{\mathbf{A}}_{\mathrm{R}}\right)\in\mathbb{C}^{N_{\mathrm{T}}N_{\mathrm{R}}\times G^{2}}$ and $L<\max\{N_{\mathrm{T}},N_{\mathrm{R}}\}\leq G$. The matrix $\mathbf{Q}$ is the sensing matrix consisting of the RF and baseband processors of the hybrid MIMO system. The product of the sensing matrix and the dictionary is referred to as the *equivalent* sensing matrix, denoted as $\bar{\mathbf{Q}}$.[5] When $G=N_{\mathrm{T}}=N_{\mathrm{R}}$, the redundant

[5]Comparing $\mathbf{Q}^{\mathrm{o}}\in\mathbb{C}^{N_{\mathrm{T}}^{\mathrm{Beam}}N_{\mathrm{R}}^{\mathrm{Beam}}\times L}$ in (15) and $\bar{\mathbf{Q}}\in\mathbb{C}^{N_{\mathrm{T}}^{\mathrm{Beam}}N_{\mathrm{R}}^{\mathrm{Beam}}\times G^{2}}$ in (25), the former is the system matrix for the LS problem and the latter is the equivalent sensing matrix for the CS problem.

Algorithm 1 OMP Based mmWave Channel Estimator

Require: sensing matrix $\bar{\mathbf{Q}}$, measurement vector $\bar{\mathbf{y}}$, and a threshold δ

1: $\mathcal{I}_0 =$ empty set, residual $\mathbf{r}_{-1} = \mathbf{0}$, $\mathbf{r}_0 = \bar{\mathbf{y}}$, set the iteration counter $t = 1$
2: **while** $\|\mathbf{r}_{t-1}\|_2^2 - \|\mathbf{r}_{t-2}\|_2^2 > \delta$ **do**
3: $j = \arg \max_{i=1,\ldots,G^2} \left|\bar{\mathbf{Q}}(i)^H \mathbf{r}_{t-1}\right|$ ▷ Find AoD/AoA pair
4: $\mathcal{I}_t = \mathcal{I}_{t-1} \cup \{j\}$ ▷ Update AoD/AoA pair set
5: $\mathbf{h}_t = \arg\min_{\mathbf{h}} \left\|\bar{\mathbf{y}} - \sqrt{P}\bar{\mathbf{Q}}_{\mathcal{I}_t}\mathbf{h}\right\|_2$ ▷ Estimate channel gains
6: $\mathbf{r}_t = \bar{\mathbf{y}} - \sqrt{P}\bar{\mathbf{Q}}_{\mathcal{I}_t}\mathbf{h}_t$ ▷ Update residual
7: $t = t + 1$
8: **end while**
9: $\hat{\mathbf{h}}_{\mathrm{a}}(i) = \mathbf{h}_{t-1}(i)$ for $i \in \mathcal{I}_{t-1}$ and $\hat{\mathbf{h}}_{\mathrm{a}}(i) = 0$ otherwise
10: **return** $\bar{\mathbf{H}}_{\mathrm{a}}^{\mathrm{CS}} = \mathrm{vec}^{-1}\left(\hat{\mathbf{h}}_{\mathrm{a}}\right)$

dictionary $\bar{\mathbf{A}}_{\mathrm{T}}^* \otimes \bar{\mathbf{A}}_{\mathrm{R}}$ reduces to the orthonormal dictionary that can be derived from the virtual channel model. The optimization problem for CS based channel estimation can be written as

$$\mathrm{vec}\left(\bar{\mathbf{H}}_{\mathrm{a}}^{\mathrm{CS}}\right) = \arg\min_{\bar{\mathbf{H}}_{\mathrm{a}}} \left\|\bar{\mathbf{y}} - \sqrt{P}\bar{\mathbf{Q}} \cdot \mathrm{vec}\left(\bar{\mathbf{H}}_{\mathrm{a}}\right)\right\|_2 \quad \text{subject to } \left\|\mathrm{vec}\left(\bar{\mathbf{H}}_{\mathrm{a}}\right)\right\|_0 = L, \tag{26}$$

where $\bar{\mathbf{H}}_{\mathrm{a}}^{\mathrm{CS}}$ denotes the CS estimate of $\bar{\mathbf{H}}_{\mathrm{a}}$. From (26) the CS estimate, denoted as $\bar{\mathbf{H}}^{\mathrm{CS}}$, of the desired quantized-channel $\bar{\mathbf{H}}$ is given by

$$\bar{\mathbf{H}}^{\mathrm{CS}} = \bar{\mathbf{A}}_{\mathrm{R}} \bar{\mathbf{H}}_{\mathrm{a}}^{\mathrm{CS}} \bar{\mathbf{A}}_{\mathrm{T}}^H, \tag{27}$$

and the SSE is defined as the Frobenius norm of the difference between $\mathbf{H}$ and $\bar{\mathbf{H}}^{\mathrm{CS}}$, i.e., $\left\|\mathbf{H} - \bar{\mathbf{H}}^{\mathrm{CS}}\right\|_F^2$. The optimization problem in (26) is a non-convex optimization with $\mathcal{L}_0$ norm and is difficult to solve. The OMP algorithm presented below is one of the most common and simplest greedy CS algorithms that can solve (26).

B. Channel Estimation via OMP

The OMP algorithm solving (26) is summarized in Algorithm 1. At the t-th iteration, this algorithm chooses the column of $\bar{\mathbf{Q}}$ that is most strongly correlated with the residual $\mathbf{r}_{t-1}$ (step 3), and updates the column index set (step 4). Each column index obtained in step 3 corresponds to a pair of grids representing the quantized AoD/AoA. Then, the channel gains associated with the chosen grid points are obtained by solving the LS problem in step 5. In step 6, the contributions of the chosen column vectors to $\bar{\mathbf{y}}$ are subtracted to update the residual $\mathbf{r}_{t-1}$. This procedure is repeated until $\|\mathbf{r}_{t-1}\|_2^2 - \|\mathbf{r}_{t-2}\|_2^2$ falls bellow the predetermined threshold δ. (Here we assume that the sparsity level L is unknown. In the case where L is known, a stopping criterion that depends on L would be preferred.) In step 9, the vector $\hat{\mathbf{h}}_{\mathrm{a}} \in \mathbb{C}^{G^2 \times 1}$ is constructed so that its i-th element, $1 \le i \le G^2$, $\hat{\mathbf{h}}_{\mathrm{a}}(i) = \mathbf{h}_{t-1}(i)$ if $i \in \mathcal{I}_{t-1}$ and $\hat{\mathbf{h}}_{\mathrm{a}}(i) = 0$, otherwise. The desired estimate is given by $\bar{\mathbf{H}}_{\mathrm{a}}^{\mathrm{CS}} = \mathrm{vec}^{-1}\left(\hat{\mathbf{h}}_{\mathrm{a}}\right)$.

The computational load of the OMP algorithm is proportional to G^2 and can be excessive in mm-wave channel estimation where $G > \max\{N_{\mathrm{T}}, N_{\mathrm{R}}\}$ and the antenna sizes are large. Specifically, in step 3, the multiplication of $\bar{\mathbf{Q}}(i)^H$ and $\mathbf{r}_{t-1}$ needs $\mathcal{O}\left(N_{\mathrm{T}}^{\mathrm{Beam}} N_{\mathrm{R}}^{\mathrm{Beam}} G^2\right)$ computations and the maximum selection needs $\mathcal{O}\left(G^2\right)$ computations; in step 5, assuming the use of the QR factorization of $\bar{\mathbf{Q}}_{\mathcal{I}_{t-1}}$ and the modified Gram-Schmidt algorithm, solving the LS problems requires $\mathcal{O}\left(|\mathcal{I}_t| N_{\mathrm{T}}^{\mathrm{Beam}} N_{\mathrm{R}}^{\mathrm{Beam}}\right)$, where $|\mathcal{I}_t|$ denotes the cardinality of $\mathcal{I}_t$. The computational complexity of the OMP algorithm can be reduced by employing efficient OMP algorithms such as the multi-stage OMP [21].

IV. Analysis on Sum of Squared Errors

In this section, we analyze the SSE of the proposed OMP estimator, defined as $\left\|\mathbf{H} - \bar{\mathbf{H}}^{\mathrm{CS}}\right\|_F^2$. Specifically, a performance lower-bound is derived by considering the mean SSE, $\mathbb{E}\left[\|\mathbf{H} - \mathbf{H}^{\mathrm{o}}\|_F^2\right]$, of the oracle estimator, and a performance upper-bound is derived based on the results of OMP performance guarantees [24]. The validity of the derived bounds will be examined through computer simulation in Section VI.

A. Lower-Bound Analysis

Consider the oracle estimator which estimates the channel gain $\mathbf{H}_{\mathrm{a}}$ in (9) under the assumption that $\mathbf{A}_{\mathrm{T}}$ and $\mathbf{A}_{\mathrm{R}}$ are known. The mean SSE of the oracle estimator can be written as

$$\begin{aligned}\mathbb{E}\left[\left\|\mathbf{H} - \mathbf{H}^{\mathrm{o}}\right\|_F^2\right] &= \mathbb{E}\left[\left\|\mathrm{vec}\left(\mathbf{H}\right) - \mathrm{vec}\left(\mathbf{H}^{\mathrm{o}}\right)\right\|_2^2\right] \\ &= \mathbb{E}\left[\left\|\left(\mathbf{A}_{\mathrm{T}}^* \odot \mathbf{A}_{\mathrm{R}}\right)\left(\mathrm{vecd}\left(\mathbf{H}_{\mathrm{a}}\right) - \mathrm{vecd}\left(\mathbf{H}_{\mathrm{a}}^{\mathrm{o}}\right)\right)\right\|_2^2\right], \end{aligned} \tag{28}$$

where the second equality follows from (11) and (17). A lower-bound of the mean SSE is derived as follows.

Lemma 2: Suppose that $\mathbb{E}\left[\bar{\mathbf{n}}\bar{\mathbf{n}}^H\right] = \sigma_n^2 \mathbf{I}_{N_{\mathrm{T}}^{\mathrm{Beam}} N_{\mathrm{R}}^{\mathrm{Beam}}}$ *for* $\bar{\mathbf{n}}$ *in (14). The mean SSE in (28) is lower-bounded as*

$$\mathbb{E}\left[\left\|\mathbf{H} - \mathbf{H}^{\mathrm{o}}\right\|_F^2\right] \ge \frac{\kappa L^2 \sigma_n^2}{P} \tag{29}$$

where $\kappa \triangleq \lambda_{\min}\left(\mathbf{B}^H\mathbf{B}\right)/\mathrm{Tr}\left((\mathbf{Q}^{\mathrm{o}})^H\mathbf{Q}^{\mathrm{o}}\right)$ *and* $\mathbf{B} \triangleq \mathbf{A}_{\mathrm{T}}^* \odot \mathbf{A}_{\mathrm{R}}$.

Proof: To simplify notations, let $\mathbf{B} \triangleq \mathbf{A}_{\mathrm{T}}^* \odot \mathbf{A}_{\mathrm{R}}$ and $\mathbf{e} \triangleq \mathrm{vecd}\left(\mathbf{H}_{\mathrm{a}}\right) - \mathrm{vecd}\left(\mathbf{H}_{\mathrm{a}}^{\mathrm{o}}\right)$. The mean SSE in (28) can be written as

$$\mathbb{E}\left[\|\mathbf{B}\mathbf{e}\|_2^2\right] \ge \lambda_{\min}\left(\mathbf{B}^H\mathbf{B}\right)\mathbb{E}\left[\|\mathbf{e}\|_2^2\right], \tag{30}$$

where the inequality holds due to the Rayleigh-Ritz ratio [27]. Using (14) in (16), $\mathrm{vecd}\left(\mathbf{H}_{\mathrm{a}}^{\mathrm{o}}\right)$ can be rewritten as

$$\mathrm{vecd}\left(\mathbf{H}_{\mathrm{a}}^{\mathrm{o}}\right) = \mathrm{vecd}\left(\mathbf{H}_{\mathrm{a}}\right) - \frac{1}{\sqrt{P}}\mathbf{C}\bar{\mathbf{n}},$$

where $\mathbf{C} \triangleq \left((\mathbf{Q}^{\mathrm{o}})^H\mathbf{Q}^{\mathrm{o}}\right)^{-1}(\mathbf{Q}^{\mathrm{o}})^H$. Thus $\mathbf{e} = \frac{1}{\sqrt{P}}\mathbf{C}\bar{\mathbf{n}}$, and

$$\begin{aligned}\mathbb{E}\left[\|\mathbf{e}\|_2^2\right] &= \frac{1}{P}\mathbb{E}\left[\|\mathbf{C}\bar{\mathbf{n}}\|_2^2\right] \\ &= \frac{\sigma_n^2}{P}\mathrm{Tr}\left(\mathbf{C}\mathbf{C}^H\right). \end{aligned} \tag{31}$$

Since $\mathbf{C}\mathbf{C}^H = \left(\left((\mathbf{Q}^{\mathrm{o}})^H\mathbf{Q}^{\mathrm{o}}\right)^{-1}\right)^H = \left((\mathbf{Q}^{\mathrm{o}})^H\mathbf{Q}^{\mathrm{o}}\right)^{-1} \in \mathbb{C}^{L\times L}$, then

$$\begin{aligned}\mathrm{Tr}\left(\mathbf{C}\mathbf{C}^H\right) &= \sum_{l=1}^{L}\lambda_l\left(\left((\mathbf{Q}^{\mathrm{o}})^H\mathbf{Q}^{\mathrm{o}}\right)^{-1}\right)\\ &= L\left(\frac{1}{L}\sum_{l=1}^{L}\lambda_l\left(\left((\mathbf{Q}^{\mathrm{o}})^H\mathbf{Q}^{\mathrm{o}}\right)^{-1}\right)\right)\\ &\geq \frac{L^2}{\sum_{l=1}^{L}1/\lambda_l\left(\left((\mathbf{Q}^{\mathrm{o}})^H\mathbf{Q}^{\mathrm{o}}\right)^{-1}\right)}\\ &= \frac{L^2}{\sum_{l=1}^{L}\lambda_l\left((\mathbf{Q}^{\mathrm{o}})^H\mathbf{Q}^{\mathrm{o}}\right)} = \frac{L^2}{\mathrm{Tr}\left((\mathbf{Q}^{\mathrm{o}})^H\mathbf{Q}^{\mathrm{o}}\right)}\end{aligned}$$

where the inequality comes from the inequality between the arithmetic and harmonic means [28], and the third equality holds because the l-th eigenvalue of $\left((\mathbf{Q}^{\mathrm{o}})^H\mathbf{Q}^{\mathrm{o}}\right)^{-1}$ can be written as $\lambda_l\left(\left((\mathbf{Q}^{\mathrm{o}})^H\mathbf{Q}^{\mathrm{o}}\right)^{-1}\right) = 1/\lambda_l\left((\mathbf{Q}^{\mathrm{o}})^H\mathbf{Q}^{\mathrm{o}}\right)$. Thus $\mathbb{E}\left[\|\mathbf{e}\|_2^2\right]$ in (31) satisfies

$$\mathbb{E}\left[\|\mathbf{e}\|_2^2\right] \geq \frac{\sigma_n^2}{P}\frac{L^2}{\mathrm{Tr}\left((\mathbf{Q}^{\mathrm{o}})^H\mathbf{Q}^{\mathrm{o}}\right)} \tag{32}$$

The lower-bound in (29) can be obtained by using (32) in (30). □

B. Upper-Bound Analysis

We derive an upper-bound of the SSE under the following assumptions: i) $N_{\mathrm{T}}^{\mathrm{Beam}} = N_{\mathrm{T}}$ and $N_{\mathrm{R}}^{\mathrm{Beam}} = N_{\mathrm{R}}$; ii) the transmit and receive matrices, $\mathbf{F} = \mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}$ and $\mathbf{W} = \mathbf{W}_{\mathrm{RF}}\mathbf{W}_{\mathrm{BB}}$ in (4), are unitary; iii) both $\bar{\mathbf{H}}_{\mathrm{a}}$ and $\mathbf{E}$ in (23) are unknown deterministic matrices; and iv) the sparsity level L of $\bar{\mathbf{H}}_{\mathrm{a}}$ is known.

The SSE of the sparse estimation problem in (26) can be examined based on the coherence of the equivalent sensing matrix $\bar{\mathbf{Q}}$ defined as

$$\mu\left(\bar{\mathbf{Q}}\right) \triangleq \max_{1\leq m,n\leq G^2, m\neq n}\left|\bar{\mathbf{Q}}(m)^H\bar{\mathbf{Q}}(n)\right|. \tag{33}$$

This coherence can be decomposed in terms of the coherences of the transmit and receive array response matrices (dictionaries), denoted as $\mu\left(\bar{\mathbf{A}}_{\mathrm{T}}\right)$ and $\mu\left(\bar{\mathbf{A}}_{\mathrm{R}}\right)$, respectively, as shown below.

Lemma 3: When the angle grids $\{\varphi_g\}$ satisfy (19) and $d = \frac{\lambda}{2}$, the coherence of the equivalent dictionary $\bar{\mathbf{Q}} = \left(\mathbf{F}_{\mathrm{BB}}^T\mathbf{F}_{\mathrm{RF}}^T\bar{\mathbf{A}}_{\mathrm{T}}^* \otimes \mathbf{W}_{\mathrm{BB}}^H\mathbf{W}_{\mathrm{RF}}^H\bar{\mathbf{A}}_{\mathrm{R}}\right)$ in (25) can be expressed as

$$\begin{aligned}\mu\left(\bar{\mathbf{Q}}\right) &= \max\left\{\mu\left(\mathbf{F}_{\mathrm{BB}}^T\mathbf{F}_{\mathrm{RF}}^T\bar{\mathbf{A}}_{\mathrm{T}}^*\right), \mu\left(\mathbf{W}_{\mathrm{BB}}^H\mathbf{W}_{\mathrm{RF}}^H\bar{\mathbf{A}}_{\mathrm{R}}\right)\right\}\\ &= \max\left\{\mu\left(\bar{\mathbf{A}}_{\mathrm{T}}\right), \mu\left(\bar{\mathbf{A}}_{\mathrm{R}}\right)\right\}\\ &= \max\left\{\left|D_{N_{\mathrm{T}}}\left(\frac{2\pi}{G}\right)\right|, \left|D_{N_{\mathrm{R}}}\left(\frac{2\pi}{G}\right)\right|\right\},\end{aligned} \tag{34}$$

and $D_N(\phi)$ is the Dirichlet kernel defined as

$$D_N(\phi) \triangleq \frac{1}{N}\sum_{n=0}^{N-1}e^{-jn\phi} = \frac{1}{N}e^{j\phi(N-1)/2}\frac{\sin(\phi N/2)}{\sin(\phi/2)}. \tag{35}$$

Proof: The first equality in (34) follows from the decomposition of the coherence of Kronecker product [29], and the second equality is due to the assumption that $\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}$ and $\mathbf{W}_{\mathrm{RF}}\mathbf{W}_{\mathrm{BB}}$ are unitary. The coherence of each array response matrix is

$$\begin{aligned}\mu\left(\bar{\mathbf{A}}_{\mathrm{T}}\right) &= \max_{1\leq i,j\leq G, i\neq j}\left|\mathbf{a}_{\mathrm{T}}^H(\varphi_i)\,\mathbf{a}_{\mathrm{T}}(\varphi_j)\right|\\ &\overset{(a)}{=} \max_{1\leq i,j\leq G, i\neq j}\left|D_{N_{\mathrm{T}}}\left(\pi\left(\cos(\varphi_j)-\cos(\varphi_i)\right)\right)\right|\\ &\overset{(b)}{=} \left|D_{N_{\mathrm{T}}}\left(\pi\left(\cos(\varphi_{i+1})-\cos(\varphi_i)\right)\right)\right| = \left|D_{N_{\mathrm{T}}}\left(\frac{2\pi}{G}\right)\right|,\end{aligned} \tag{36}$$

where the equality (a) follows from (8) when $d = \frac{\lambda}{2}$; the equality (b) is due to the following facts: i) $\cos(\varphi_i) - \cos(\varphi_j)$ depends only on $|i-j|$, because of (19); ii) $D_{N_{\mathrm{T}}}\left(\pi\left(\cos(\varphi_i)-\cos(\varphi_j)\right)\right)$ is maximized when $j = i+1$. In a similar manner, $\mu\left(\bar{\mathbf{A}}_{\mathrm{R}}\right) = \left|D_{N_{\mathrm{R}}}\left(\frac{2\pi}{G}\right)\right|$ can be derived. □

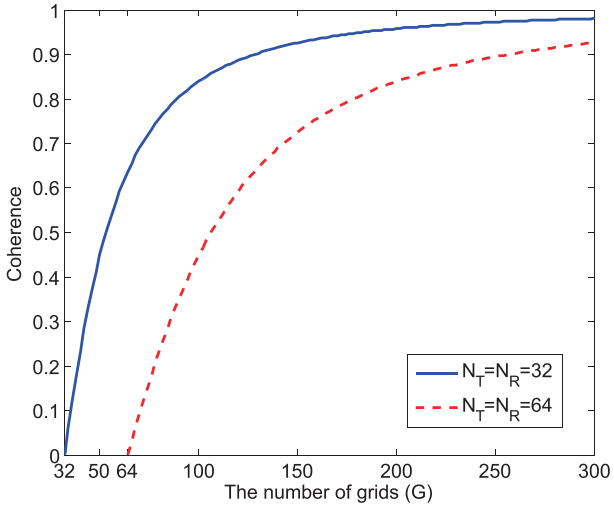

Fig. 2. The coherence $\mu\left(\bar{\mathbf{Q}}\right)$ against the number of grids G.

Since $D_N(\phi)$ monotonically increases and converges to one as ϕ decreases from $2\pi/N$ to zero, the coherences $\mu\left(\bar{\mathbf{A}}_{\mathrm{T}}\right)$, $\mu\left(\bar{\mathbf{A}}_{\mathrm{R}}\right)$ and $\mu\left(\bar{\mathbf{Q}}\right)$ monotonically increase and converge to one as G increases from $\max\{N_{\mathrm{T}}, N_{\mathrm{R}}\}$ to infinity. This is illustrated in Fig. 2 for $\mu\left(\bar{\mathbf{Q}}\right)$. If the grids were chosen such that $\{\varphi_g\}$ are uniformly distributed in $[0,\pi]$, the coherence $\mu\left(\bar{\mathbf{A}}_{\mathrm{T}}\right)$ would be $\left|D_{N_{\mathrm{T}}}\left(\pi\left(\cos(0)-\cos\left(\frac{\pi}{G}\right)\right)\right)\right|$ which is larger than $\left|D_{N_{\mathrm{T}}}\left(\frac{2\pi}{G}\right)\right|$ in (34), because $1-\cos\left(\frac{\pi}{G}\right) < \frac{2\pi}{G}$. This observation indicates that the proposed non-uniformly distributed angle grids reduce the coherence $\mu\left(\bar{\mathbf{Q}}\right)$.

Our derivation for the SSE upper-bound starts with the following inequality showing that the SSE is upper-bounded by the sum of the quantized-channel estimation error and the angle quantization error:

$$\begin{aligned}\left\|\mathbf{H}-\bar{\mathbf{H}}^{\mathrm{CS}}\right\|_F^2 &= \left\|\left(\bar{\mathbf{H}}+\mathbf{E}\right)-\bar{\mathbf{H}}^{\mathrm{CS}}\right\|_F^2\\ &= \left\|\mathrm{vec}\left(\bar{\mathbf{H}}+\mathbf{E}\right)-\mathrm{vec}\left(\bar{\mathbf{H}}^{\mathrm{CS}}\right)\right\|_2^2\\ &\leq \underbrace{\left\|\mathrm{vec}\left(\bar{\mathbf{H}}\right)-\mathrm{vec}\left(\bar{\mathbf{H}}^{\mathrm{CS}}\right)\right\|_2^2}_{\text{Quantized-channel estimation error}} + \underbrace{\|\mathrm{vec}(\mathbf{E})\|_2^2}_{\text{Quantization error}},\end{aligned} \tag{37}$$

where $\bar{\mathbf{H}}$ is the quantized-channel in (22), and $\bar{\mathbf{H}}^{\mathrm{CS}}$ is the CS estimate of $\bar{\mathbf{H}}$ in (27). Since $\mathrm{vec}\left(\bar{\mathbf{H}}\right) = \bar{\mathbf{A}}_{\mathrm{T}}^{*} \otimes \bar{\mathbf{A}}_{\mathrm{R}} \cdot \mathrm{vec}\left(\bar{\mathbf{H}}_{\mathrm{a}}\right)$ and $\mathrm{vec}\left(\bar{\mathbf{H}}^{\mathrm{CS}}\right) = \bar{\mathbf{A}}_{\mathrm{T}}^{*} \otimes \bar{\mathbf{A}}_{\mathrm{R}} \cdot \mathrm{vec}\left(\bar{\mathbf{H}}_{\mathrm{a}}^{\mathrm{CS}}\right)$, the quantized-channel estimation error in the RHS of (37) can be represented as

$$\left\|\mathrm{vec}\left(\bar{\mathbf{H}}\right) - \mathrm{vec}\left(\bar{\mathbf{H}}^{\mathrm{CS}}\right)\right\|_2^2 = \left\|\bar{\mathbf{A}}_{\mathrm{T}}^{*} \otimes \bar{\mathbf{A}}_{\mathrm{R}}\left(\mathrm{vec}\left(\bar{\mathbf{H}}_{\mathrm{a}}\right) - \mathrm{vec}\left(\bar{\mathbf{H}}_{\mathrm{a}}^{\mathrm{CS}}\right)\right)\right\|_2^2. \quad (38)$$

We derive an upper bound of the quantized-channel estimation error when $\mathrm{supp}\left(\mathrm{vec}\left(\bar{\mathbf{H}}_{\mathrm{a}}\right)\right) = \mathrm{supp}\left(\mathrm{vec}\left(\bar{\mathbf{H}}_{\mathrm{a}}^{\mathrm{CS}}\right)\right)$ (the condition under which this equality holds will be presented in Lemma 5 below).

Lemma 4: Suppose that the supports of $\mathrm{vec}\left(\bar{\mathbf{H}}_{\mathrm{a}}\right)$ *and* $\mathrm{vec}\left(\bar{\mathbf{H}}_{\mathrm{a}}^{\mathrm{CS}}\right)$ *are the same. Let* $\mathcal{I}$ *denote their support:* $\mathcal{I} = \mathrm{supp}\left(\mathrm{vec}\left(\bar{\mathbf{H}}_{\mathrm{a}}\right)\right) = \mathrm{supp}\left(\mathrm{vec}\left(\bar{\mathbf{H}}_{\mathrm{a}}^{\mathrm{CS}}\right)\right)$ *and* $|\mathcal{I}| = L$*. Then the quantized-channel estimation error in (38) satisfies*

$$\left\|\mathrm{vec}\left(\bar{\mathbf{H}}\right) - \mathrm{vec}\left(\bar{\mathbf{H}}^{\mathrm{CS}}\right)\right\|_2^2 \le \left(1 + (L-1)\,\mu\left(\bar{\mathbf{A}}_{\mathrm{T}} \otimes \bar{\mathbf{A}}_{\mathrm{R}}\right)\right) \times \left\|\mathrm{vec}\left(\bar{\mathbf{H}}_{\mathrm{a}}\right) - \mathrm{vec}\left(\bar{\mathbf{H}}_{\mathrm{a}}^{\mathrm{CS}}\right)\right\|_2^2. \quad (39)$$

Proof: The RHS of (38) satisfies the following:

$$\begin{aligned}
&\left\|\bar{\mathbf{A}}_{\mathrm{T}}^{*} \otimes \bar{\mathbf{A}}_{\mathrm{R}}\left(\mathrm{vec}\left(\bar{\mathbf{H}}_{\mathrm{a}}\right) - \mathrm{vec}\left(\bar{\mathbf{H}}_{\mathrm{a}}^{\mathrm{CS}}\right)\right)\right\|_2^2 \\
&= \left\|\left(\bar{\mathbf{A}}_{\mathrm{T}}^{*} \otimes \bar{\mathbf{A}}_{\mathrm{R}}\right)_{\mathcal{I}}\left(\mathrm{vec}\left(\bar{\mathbf{H}}_{\mathrm{a}}\right)_{\mathcal{I}} - \mathrm{vec}\left(\bar{\mathbf{H}}_{\mathrm{a}}^{\mathrm{CS}}\right)_{\mathcal{I}}\right)\right\|_2^2 \\
&\le \left\|\left(\bar{\mathbf{A}}_{\mathrm{T}}^{*} \otimes \bar{\mathbf{A}}_{\mathrm{R}}\right)_{\mathcal{I}}\right\|_2^2 \left\|\left(\mathrm{vec}\left(\bar{\mathbf{H}}_{\mathrm{a}}\right)_{\mathcal{I}} - \mathrm{vec}\left(\bar{\mathbf{H}}_{\mathrm{a}}^{\mathrm{CS}}\right)_{\mathcal{I}}\right)\right\|_2^2 \\
&= \left\|\left(\bar{\mathbf{A}}_{\mathrm{T}}^{*} \otimes \bar{\mathbf{A}}_{\mathrm{R}}\right)_{\mathcal{I}}\right\|_2^2 \left\|\left(\mathrm{vec}\left(\bar{\mathbf{H}}_{\mathrm{a}}\right) - \mathrm{vec}\left(\bar{\mathbf{H}}_{\mathrm{a}}^{\mathrm{CS}}\right)\right)\right\|_2^2,
\end{aligned}$$

where the inequality comes from the Cauchy-schwarz inequality. To show that $\left\|\left(\bar{\mathbf{A}}_{\mathrm{T}}^{*} \otimes \bar{\mathbf{A}}_{\mathrm{R}}\right)_{\mathcal{I}}\right\|_2^2 \le 1 + (L-1)\,\mu\left(\bar{\mathbf{A}}_{\mathrm{T}} \otimes \bar{\mathbf{A}}_{\mathrm{R}}\right)$, we define $\mathbf{G}_{\mathcal{I}} \triangleq \left(\bar{\mathbf{A}}_{\mathrm{T}}^{*} \otimes \bar{\mathbf{A}}_{\mathrm{R}}\right)_{\mathcal{I}}^{H}\left(\bar{\mathbf{A}}_{\mathrm{T}}^{*} \otimes \bar{\mathbf{A}}_{\mathrm{R}}\right)_{\mathcal{I}}$. Then the squared spectral norm satisfies $\left\|\left(\bar{\mathbf{A}}_{\mathrm{T}}^{*} \otimes \bar{\mathbf{A}}_{\mathrm{R}}\right)_{\mathcal{I}}\right\|_2^2 = \lambda_{\max}\left(\mathbf{G}_{\mathcal{I}}\right)$. Since all the diagonal entries of $\mathbf{G}_{\mathcal{I}}$ are equal to one, $\mathbf{G}_{\mathcal{I}}$ can be represented as $\mathbf{G}_{\mathcal{I}} = \mathbf{I}_L + \Gamma$, where $\Gamma \in \mathbb{C}^{L \times L}$ is a matrix having zero diagonal entries. From the definition of the coherence in (33), we observe that the off-diagonal entries of Γ (or equivalently $\mathbf{G}_{\mathcal{I}}$) are upper-bounded by $\mu\left(\bar{\mathbf{A}}_{\mathrm{T}}^{*} \otimes \bar{\mathbf{A}}_{\mathrm{R}}\right) = \mu\left(\bar{\mathbf{A}}_{\mathrm{T}} \otimes \bar{\mathbf{A}}_{\mathrm{R}}\right)$ and define $\Gamma' \in \mathbb{C}^{L \times L}$ by replacing all the off-diagonal entries of Γ by the upper-bound $\mu\left(\bar{\mathbf{A}}_{\mathrm{T}} \otimes \bar{\mathbf{A}}_{\mathrm{R}}\right)$. Now we apply the Gershgorin circle theorem [30] to $\mathbf{I}_L + \Gamma'$ to find an upper-bound of $\lambda_{\max}\left(\mathbf{I}_L + \Gamma'\right)$. To be specific, let $\mathbf{A} = \mathbf{I}_L + \Gamma'$. Then the Gershgorin circle theorem states that

$$\Lambda\left(\mathbf{A}\right) \subseteq \bigcup_{l=1}^{L} D_l,$$

where $\Lambda\left(\mathbf{A}\right)$ is the set of eigenvalues of $\mathbf{A}$ and $D_l = \left\{z \in \mathbb{C} : |z - 1| \le \sum_{m=1}^{L}\left|\Gamma'\left(l, m\right)\right|\right\}$. Thus $\lambda_{\max}\left(\mathbf{I}_L + \Gamma'\right) \le 1 + (L-1)\,\mu\left(\bar{\mathbf{A}}_{\mathrm{T}} \otimes \bar{\mathbf{A}}_{\mathrm{R}}\right)$. The Gershgorin circle theorem also indicates that the upper-bound of $\lambda_{\max}\left(\mathbf{G}_{\mathcal{I}}\right)$ is less than or equal to that of $\lambda_{\max}\left(\mathbf{I}_L + \Gamma'\right)$. Thus, the squared spectral norm satisfies $\left\|\left(\bar{\mathbf{A}}_{\mathrm{T}}^{*} \otimes \bar{\mathbf{A}}_{\mathrm{R}}\right)_{\mathcal{I}}\right\|_2^2 \le 1 + (L-1)\,\mu\left(\bar{\mathbf{A}}_{\mathrm{T}} \otimes \bar{\mathbf{A}}_{\mathrm{R}}\right)$ which is the desired result. □

In the RHS of (39), $\left\|\mathrm{vec}\left(\bar{\mathbf{H}}_{\mathrm{a}}\right) - \mathrm{vec}\left(\bar{\mathbf{H}}_{\mathrm{a}}^{\mathrm{CS}}\right)\right\|_2^2$ represents the sparse estimation error of the CS problem in (26). In what follows, an upper-bound of this estimation error is derived by extending the OMP performance guarantees in [26] to the case of (24) where both the quantization error $\bar{\mathbf{e}}$ and noise $\bar{\mathbf{n}}$ exist.

Lemma 5 (OMP Performance Guarantees): Consider the sparse estimation problem in (26) with the sparsity level L. Suppose that

$$\sqrt{P}\left|\bar{\mathbf{H}}_{\mathrm{a}}\right|_{\min}\left(1 - (2L-1)\,\mu\left(\bar{\mathbf{Q}}\right)\right) \ge 2\left(\sigma_{\bar{n}}\sqrt{(1+\beta)\log G^2} + \sqrt{P}\|\mathrm{vec}\left(\mathbf{E}\right)\|_2\right), \quad (40)$$

where $\left|\bar{\mathbf{H}}_{\mathrm{a}}\right|_{\min}$ *is the smallest magnitude of the non-zero entries of* $\bar{\mathbf{H}}_{\mathrm{a}}$ *in (23),* $\sigma_{\bar{n}}$ *is the standard deviation of* $\bar{\mathbf{n}}$*, and* β *is a positive constant. Then, with probability at least* $1 - \left(G^{2\beta}\pi\right)^{-1}$ *the OMP correctly identifies the* $\mathrm{supp}\left(\mathrm{vec}\left(\bar{\mathbf{H}}_{\mathrm{a}}\right)\right)$*, and* $\mathrm{vec}\left(\bar{\mathbf{H}}_{\mathrm{a}}^{\mathrm{CS}}\right)$ *satisfies*

$$\left\|\mathrm{vec}\left(\bar{\mathbf{H}}_{\mathrm{a}}\right) - \mathrm{vec}\left(\bar{\mathbf{H}}_{\mathrm{a}}^{\mathrm{CS}}\right)\right\|_2^2 \le \frac{L}{\left(1 - (L-1)\,\mu\left(\bar{\mathbf{Q}}\right)\right)^2} \times \left(\sqrt{(1+\beta)\,\frac{\sigma_{\bar{n}}^2}{P}\log G^2} + \|\mathrm{vec}\left(\mathbf{E}\right)\|_2\right)^2. \quad (41)$$

Proof: We define an event $\mathcal{B} = \left\{\max_{1 \le m \le G^2}\left|\bar{\mathbf{Q}}^{H}\left(m\right)\left(\bar{\mathbf{n}} + \bar{\mathbf{e}}\right)\right| < \tau\right\}$ for $\tau = \sigma_{\bar{n}}\sqrt{(1+\beta)\log G^2} + \sqrt{P}\|\mathrm{vec}\left(\mathbf{E}\right)\|_2$ and first show that $\Pr\left\{\mathcal{B}\right\} \ge 1 - \left(G^{2\beta}\pi\right)^{-1}$. Note that

$$\begin{aligned}
&\max_{1 \le m \le G^2}\left|\bar{\mathbf{Q}}^{H}\left(m\right)\left(\bar{\mathbf{n}} + \bar{\mathbf{e}}\right)\right| \\
&\le \max_{1 \le m \le G^2}\left|\bar{\mathbf{Q}}^{H}\left(m\right)\bar{\mathbf{n}}\right| + \max_{1 \le m \le G^2}\left|\bar{\mathbf{Q}}^{H}\left(m\right)\bar{\mathbf{e}}\right| \\
&\le \max_{1 \le m \le G^2}\left|\bar{\mathbf{Q}}^{H}\left(m\right)\bar{\mathbf{n}}\right| + \max_{1 \le m \le G^2}\left\|\bar{\mathbf{Q}}\left(m\right)\right\|_2\|\bar{\mathbf{e}}\|_2 \\
&= \max_{1 \le m \le G^2}\left|\bar{\mathbf{Q}}^{H}\left(m\right)\bar{\mathbf{n}}\right| + \sqrt{P}\|\mathrm{vec}\left(\mathbf{E}\right)\|_2,
\end{aligned}$$

where the equality holds because $\left\|\bar{\mathbf{Q}}\left(m\right)\right\|_2 = 1$ for all m and $\|\bar{\mathbf{e}}\|_2 = \sqrt{P}\|\mathrm{vec}\left(\mathbf{E}\right)\|_2$. Hence,

$$\begin{aligned}
\Pr\left\{\mathcal{B}\right\} &= \Pr\left\{\max_{1 \le m \le G^2}\left|\bar{\mathbf{Q}}^{H}\left(m\right)\left(\bar{\mathbf{n}} + \bar{\mathbf{e}}\right)\right| < \tau\right\} \\
&\ge \Pr\left\{\max_{1 \le m \le G^2}\left|\bar{\mathbf{Q}}^{H}\left(m\right)\bar{\mathbf{n}}\right| + \sqrt{P}\|\mathrm{vec}\left(\mathbf{E}\right)\|_2 < \sigma_{\bar{n}}\sqrt{(1+\beta)\log G^2} + \sqrt{P}\|\mathrm{vec}\left(\mathbf{E}\right)\|_2\right\} \\
&= \Pr\left\{\max_{1 \le m \le G^2}\left|\bar{\mathbf{Q}}^{H}\left(m\right)\bar{\mathbf{n}}\right| < \sigma_{\bar{n}}\sqrt{(1+\beta)\log G^2}\right\} \\
&\ge 1 - \left(G^{2\beta}\pi\right)^{-1},
\end{aligned}$$

where the last inequality follows from [31, Proposition 5]. Next, we derive a condition that the OMP correctly identifies the support $\mathcal{I} = \mathrm{supp}\left(\mathrm{vec}\left(\bar{\mathbf{H}}_{\mathrm{a}}\right)\right)$. Suppose that the event $\mathcal{B}$ occurs and the condition (40) is satisfied. Then, by modifying the proof of [26, Lemma 3] to take account of complex

variables, it can be shown that

$$\max_{m\in\mathcal{I}}\left|\bar{\mathbf{Q}}^H(m)\,\bar{\mathbf{y}}\right| > \max_{m\notin\mathcal{I}}\left|\bar{\mathbf{Q}}^H(m)\,\bar{\mathbf{y}}\right|. \tag{42}$$

The inequality (42) is the condition that the OMP correctly identifies $\mathcal{I} = \mathrm{supp}\left(\mathrm{vec}\left(\bar{\mathbf{H}}_{\mathrm{a}}\right)\right)$ [26]. When the support $\mathcal{I}$ is successfully identified, the CS estimator becomes the oracle estimator represented as $\mathrm{vec}\left(\bar{\mathbf{H}}_{\mathrm{a}}^{\mathrm{o}}\right)_{\mathcal{I}} = P^{-1/2}\bar{\mathbf{Q}}_{\mathcal{I}}^{\dagger}\bar{\mathbf{y}}$, and its squared-error is written as

$$\begin{aligned}
&\left\|\mathrm{vec}\left(\bar{\mathbf{H}}_{\mathrm{a}}\right)_{\mathcal{I}} - \mathrm{vec}\left(\bar{\mathbf{H}}_{\mathrm{a}}^{\mathrm{o}}\right)_{\mathcal{I}}\right\|_2^2 \\
&= \left\|\mathrm{vec}\left(\bar{\mathbf{H}}_{\mathrm{a}}\right)_{\mathcal{I}} - \bar{\mathbf{Q}}_{\mathcal{I}}^{\dagger}\bar{\mathbf{Q}}_{\mathcal{I}}\mathrm{vec}\left(\bar{\mathbf{H}}_{\mathrm{a}}\right)_{\mathcal{I}} - P^{-1/2}\bar{\mathbf{Q}}_{\mathcal{I}}^{\dagger}\left(\bar{\mathbf{n}}+\bar{\mathbf{e}}\right)\right\|_2^2 \\
&= \left\|P^{-1/2}\bar{\mathbf{Q}}_{\mathcal{I}}^{\dagger}\left(\bar{\mathbf{n}}+\bar{\mathbf{e}}\right)\right\|_2^2 \\
&= \left\|P^{-1/2}\left(\bar{\mathbf{Q}}_{\mathcal{I}}^H\bar{\mathbf{Q}}_{\mathcal{I}}\right)^{-1}\bar{\mathbf{Q}}_{\mathcal{I}}^H\left(\bar{\mathbf{n}}+\bar{\mathbf{e}}\right)\right\|_2^2 \\
&\le \left\|\left(\bar{\mathbf{Q}}_{\mathcal{I}}^H\bar{\mathbf{Q}}_{\mathcal{I}}\right)^{-1}\right\|_2^2\sum_{m\in\mathcal{I}}\left(P^{-1/2}\bar{\mathbf{Q}}^H(m)\left(\bar{\mathbf{n}}+\bar{\mathbf{e}}\right)\right)^2 \\
&\le \left(1-(L-1)\,\mu\left(\bar{\mathbf{Q}}\right)\right)^{-2} \\
&\quad\times L\left(\sqrt{(1+\beta)\,\frac{\sigma_{\bar{n}}^2}{P}\log G^2} + \|\mathrm{vec}\,(\mathbf{E})\|_2\right)^2,
\end{aligned}$$

where the last inequality is derived by using the definition of $\mathcal{B}$ and the fact that $\left\|\left(\bar{\mathbf{Q}}_{\mathcal{I}}^H\bar{\mathbf{Q}}_{\mathcal{I}}\right)^{-1}\right\|_2^2 = 1/\lambda_{\min}^2\left(\bar{\mathbf{Q}}_{\mathcal{I}}^H\bar{\mathbf{Q}}_{\mathcal{I}}\right) \le 1/\left(1-(L-1)\,\mu\left(\bar{\mathbf{Q}}\right)\right)^2$. Here, the inequality can be proved by applying the Gershgorin circle theorem to $\lambda_{\min}\left(\bar{\mathbf{Q}}_{\mathcal{I}}^H\bar{\mathbf{Q}}_{\mathcal{I}}\right)$ as in the proof of Lemma 4. □

The condition in (40) requires $\left|\bar{\mathbf{H}}_{\mathrm{a}}\right|_{\min}$ to be larger than the sum of a constant multiple of the noise standard deviation and the quantization noise. This is needed to ensure that the support of $\mathrm{vec}\left(\bar{\mathbf{H}}_{\mathrm{a}}\right)$ will be identified with high probability. Furthermore, we can make some interesting observations from this condition. Because the RHS of (40) is positive, the condition can be reduced to the well-known bound [32] given by

$$L < \frac{1}{2}\left(1+\frac{1}{\mu\left(\bar{\mathbf{Q}}\right)}\right) \tag{43}$$

or

$$\mu\left(\bar{\mathbf{Q}}\right) < \frac{1}{2L-1}. \tag{44}$$

The bounds in (43) and (44) lead to the following inequalities:

$$1 - L\mu\left(\bar{\mathbf{Q}}\right) + \mu\left(\bar{\mathbf{Q}}\right) > L\mu\left(\bar{\mathbf{Q}}\right) > 0, \tag{45}$$

and

$$L\mu\left(\bar{\mathbf{Q}}\right) < \frac{1+\mu\left(\bar{\mathbf{Q}}\right)}{2} \le 1. \tag{46}$$

Due to (45), the denominator in the RHS of (41) is always nonzero. In addition, the inequality in (44) limits the grid size G because the RHS of (44) is much less than one even for a small L and $\mu\left(\bar{\mathbf{Q}}\right)$ is a monotonically increasing function of G. Lemma 5 is valid only for those G values that make $\mu\left(\bar{\mathbf{Q}}\right)$ meet (45). The following example illustrates Lemma 5.

Example 1: Let $N_{\mathrm{T}} = N_{\mathrm{R}} = 32$, $|\mathbf{H}_{\mathrm{a}}|_{\min} = 32$, $L = 3$, $\beta = 0.5$, $\|\mathrm{vec}\,(\mathbf{E})\|_2 = 1$, *and* $10\log_{10}\left(P/\sigma_{\bar{n}}^2\right) = 20$dB. *Then* $\mu\left(\bar{\mathbf{Q}}\right) \le 0.1834$, *and the condition (40) is satisfied for* $G \le 38$. *When* $G = 38$, *the probability for successfully identifying the* $\mathrm{supp}\left(\mathrm{vec}\left(\bar{\mathbf{H}}_{\mathrm{a}}\right)\right)$ *is given by* $1-\left(G^{2\beta}\pi\right)^{-1} = 0.99$ *and the upper-bound in (41) is* 12.97.

Next, we derive an upper-bound of the quantization error $\|\mathrm{vec}\,(\mathbf{E})\|_2^2$ in (37). From (7) and (9),

$$\begin{aligned}
\mathrm{vec}\,(\mathbf{H}) &= \mathrm{vec}\left(\sum\nolimits_{l=1}^{L}\mathbf{H}_{\mathrm{a}}(l,l)\,\mathbf{a}_{\mathrm{R}}\left(\theta_l^{\mathrm{r}}\right)\mathbf{a}_{\mathrm{T}}^H\left(\theta_l^{\mathrm{t}}\right)\right) \\
&= \sum\nolimits_{l=1}^{L}\mathrm{vec}\left(\mathbf{H}_{\mathrm{a}}(l,l)\,\mathbf{a}_{\mathrm{R}}\left(\theta_l^{\mathrm{r}}\right)\mathbf{a}_{\mathrm{T}}^H\left(\theta_l^{\mathrm{t}}\right)\right) \\
&= \sum\nolimits_{l=1}^{L}\mathbf{H}_{\mathrm{a}}(l,l)\,\mathbf{a}_{\mathrm{T}}^*\left(\theta_l^{\mathrm{t}}\right)\otimes\mathbf{a}_{\mathrm{R}}\left(\theta_l^{\mathrm{r}}\right),
\end{aligned}$$

where the third equality holds due to the identity, $\mathrm{vec}\,(\mathbf{ABC}) = \mathbf{C}^T\otimes\mathbf{A}\cdot\mathrm{vec}\,(\mathbf{B})$ (we set $\mathbf{A} = \mathbf{a}_{\mathrm{R}}\left(\theta_l^{\mathrm{r}}\right)$, $\mathbf{B} = \mathbf{H}_{\mathrm{a}}(l,l)$, $\mathbf{C} = \mathbf{a}_{\mathrm{T}}^H\left(\theta_l^{\mathrm{t}}\right)$). To simplify notations, we define $\mathbf{g}_l \triangleq \mathbf{H}_{\mathrm{a}}(l,l)\,\mathbf{a}_{\mathrm{T}}^*\left(\theta_l^{\mathrm{t}}\right)\otimes\mathbf{a}_{\mathrm{R}}\left(\theta_l^{\mathrm{r}}\right)$. Then,

$$\mathrm{vec}\,(\mathbf{H}) = \sum_{l=1}^{L}\mathbf{g}_l. \tag{47}$$

Similarly, $\mathrm{vec}\left(\bar{\mathbf{H}}\right)$ can be represented as

$$\mathrm{vec}\left(\bar{\mathbf{H}}\right) = \sum_{l=1}^{L}\bar{\mathbf{g}}_l, \tag{48}$$

where $\bar{\mathbf{g}}_l \triangleq \bar{\mathbf{H}}_{\mathrm{a}}(l,l)\,\mathbf{a}_{\mathrm{T}}^*\left(\varphi_l^{\mathrm{t}}\right)\otimes\mathbf{a}_{\mathrm{R}}\left(\varphi_l^{\mathrm{r}}\right)$ and φ_l^{t} and φ_l^{r} are the quantized angles of θ_l^{t} and θ_l^{r}, respectively. Suppose that $\bar{\mathbf{g}}_l$ is the projection of $\mathbf{g}_l$ onto the line through the vector $\mathbf{a}_{\mathrm{T}}^*\left(\varphi_l^{\mathrm{t}}\right)\otimes\mathbf{a}_{\mathrm{R}}\left(\varphi_l^{\mathrm{r}}\right)$ and that φ_l^{t} and φ_l^{r} are selected from the set $\mathcal{G}$ in (18) according to the following:

$$\left\{\varphi_l^{\mathrm{t}},\varphi_l^{\mathrm{r}}\right\} = \underset{\varphi_l^{\mathrm{t}},\varphi_l^{\mathrm{r}}\in\mathcal{G}}{\arg\max}\left|\left(\mathbf{a}_{\mathrm{T}}^*\left(\theta_l^{\mathrm{t}}\right)\otimes\mathbf{a}_{\mathrm{R}}\left(\theta_l^{\mathrm{r}}\right)\right)^H\mathbf{a}_{\mathrm{T}}^*\left(\varphi_l^{\mathrm{t}}\right)\otimes\mathbf{a}_{\mathrm{R}}\left(\varphi_l^{\mathrm{r}}\right)\right|. \tag{49}$$

Then, since $\left\|\mathbf{a}_{\mathrm{T}}^*\left(\varphi_l^{\mathrm{t}}\right)\otimes\mathbf{a}_{\mathrm{R}}\left(\varphi_l^{\mathrm{r}}\right)\right\|_2^2 = 1$, we get

$$\bar{\mathbf{H}}_{\mathrm{a}}(l,l) = \left(\mathbf{a}_{\mathrm{T}}^*\left(\varphi_l^{\mathrm{t}}\right)\otimes\mathbf{a}_{\mathrm{R}}\left(\varphi_l^{\mathrm{r}}\right)\right)^H\mathbf{g}_l. \tag{50}$$

Using (47)-(50), an upper-bound of the quantization error is derived as follows.

Lemma 6: Suppose that $\bar{\mathbf{g}}_l$ *is obtained via the projection associated with (50). When the angle grids* $\{\varphi_g\}$ *satisfy (19) and* $d = \frac{\lambda}{2}$, *the quantization error in (37) is bounded by*

$$\begin{aligned}
\|\mathrm{vec}\,(\mathbf{E})\|_2^2 &\le \left(\sum\nolimits_{l=1}^{L}|\mathbf{H}_{\mathrm{a}}(l,l)|\right)^2 \\
&\quad\times\left(1-\left|D_{N_{\mathrm{T}}}^*\left(\frac{\pi}{G}\right)D_{N_{\mathrm{R}}}\left(\frac{\pi}{G}\right)\right|^2\right) \triangleq E_{\mathrm{UB}},
\end{aligned} \tag{51}$$

for $\mathbf{H}_{\mathrm{a}}$ *in (9).*

Proof: $\mathrm{vec}(\mathbf{E}) = \mathrm{vec}(\mathbf{H}) - \mathrm{vec}(\bar{\mathbf{H}}) = \sum_{l=1}^{L} \mathbf{g}_l - \sum_{l=1}^{L} \bar{\mathbf{g}}_l = \sum_{l=1}^{L} \mathbf{H}_{\mathrm{a}}(l,l)\, \mathbf{a}_{\mathrm{T}}^{*}(\theta_l^{\mathrm{t}}) \otimes \mathbf{a}_{\mathrm{R}}(\theta_l^{\mathrm{r}}) - \sum_{l=1}^{L} \bar{\mathbf{H}}_{\mathrm{a}}(l,l)\, \mathbf{a}_{\mathrm{T}}^{*}(\varphi_l^{\mathrm{t}}) \otimes \mathbf{a}_{\mathrm{R}}(\varphi_l^{\mathrm{r}})$. From (50),

$$\begin{aligned}
\bar{\mathbf{H}}_{\mathrm{a}}(l,l) &= \left(\mathbf{a}_{\mathrm{T}}^{*}(\varphi_l^{\mathrm{t}}) \otimes \mathbf{a}_{\mathrm{R}}(\varphi_l^{\mathrm{r}})\right)^{H} \mathbf{g}_l \\
&= \mathbf{H}_{\mathrm{a}}(l,l) \left(\mathbf{a}_{\mathrm{T}}^{T}(\varphi_l^{\mathrm{t}}) \otimes \mathbf{a}_{\mathrm{R}}^{H}(\varphi_l^{\mathrm{r}})\right) \left(\mathbf{a}_{\mathrm{T}}^{*}(\theta_l^{\mathrm{t}}) \otimes \mathbf{a}_{\mathrm{R}}(\theta_l^{\mathrm{r}})\right) \qquad (52) \\
&= \mathbf{H}_{\mathrm{a}}(l,l)\, \mathbf{a}_{\mathrm{T}}^{T}(\varphi_l^{\mathrm{t}})\, \mathbf{a}_{\mathrm{T}}^{*}(\theta_l^{\mathrm{t}})\, \mathbf{a}_{\mathrm{R}}^{H}(\varphi_l^{\mathrm{r}})\, \mathbf{a}_{\mathrm{R}}(\theta_l^{\mathrm{r}}) \\
&\overset{(a)}{=} \mathbf{H}_{\mathrm{a}}(l,l)\, D_{N_{\mathrm{T}}}^{*}\left(\pi\left(\cos(\varphi_l^{\mathrm{t}}) - \cos(\theta_l^{\mathrm{t}})\right)\right) \\
&\quad \times D_{N_{\mathrm{R}}}\left(\pi\left(\cos(\varphi_l^{\mathrm{r}}) - \cos(\theta_l^{\mathrm{r}})\right)\right) \\
&= \mathbf{H}_{\mathrm{a}}(l,l)\, D_{N_{\mathrm{T}}}^{*}\left(\pi \cdot \Delta_l^{\mathrm{t}}\right) D_{N_{\mathrm{R}}}\left(\pi \cdot \Delta_l^{\mathrm{r}}\right) \qquad (53)
\end{aligned}$$

where the equality (a) follows from (8) when $d = \frac{\lambda}{2}$; $\Delta_l^{\mathrm{t}} \triangleq \cos(\varphi_l^{\mathrm{t}}) - \cos(\theta_l^{\mathrm{t}})$ and $\Delta_l^{\mathrm{r}} \triangleq \cos(\varphi_l^{\mathrm{r}}) - \cos(\theta_l^{\mathrm{r}})$. Here, the third equality holds due to the identity $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = \mathbf{AC} \otimes \mathbf{BD}$. The quantization error is represented as

$$\begin{aligned}
&\|\mathrm{vec}(\mathbf{E})\|_2 \\
&= \Big\| \sum\nolimits_{l=1}^{L} \mathbf{H}_{\mathrm{a}}(l,l)\, \mathbf{a}_{\mathrm{T}}^{*}(\theta_l^{\mathrm{t}}) \otimes \mathbf{a}_{\mathrm{R}}(\theta_l^{\mathrm{r}}) \\
&\qquad - \sum\nolimits_{l=1}^{L} \bar{\mathbf{H}}_{\mathrm{a}}(l,l)\, \mathbf{a}_{\mathrm{T}}^{*}(\varphi_l^{\mathrm{t}}) \otimes \mathbf{a}_{\mathrm{R}}(\varphi_l^{\mathrm{r}}) \Big\|_2 \\
&\overset{(a)}{\le} \sum\nolimits_{l=1}^{L} |\mathbf{H}_{\mathrm{a}}(l,l)| \left\| \mathbf{a}_{\mathrm{T}}^{*}(\theta_l^{\mathrm{t}}) \otimes \mathbf{a}_{\mathrm{R}}(\theta_l^{\mathrm{r}}) \right. \\
&\qquad \left. - D_{N_{\mathrm{T}}}^{*}\left(\pi \cdot \Delta_l^{\mathrm{t}}\right) D_{N_{\mathrm{R}}}\left(\pi \cdot \Delta_l^{\mathrm{r}}\right) \cdot \mathbf{a}_{\mathrm{T}}^{*}(\varphi_l^{\mathrm{t}}) \otimes \mathbf{a}_{\mathrm{R}}(\varphi_l^{\mathrm{r}}) \right\|_2 \\
&\overset{(b)}{=} \sum\nolimits_{l=1}^{L} |\mathbf{H}_{\mathrm{a}}(l,l)| \sqrt{1 - \left| D_{N_{\mathrm{T}}}^{*}\left(\pi \cdot \Delta_l^{\mathrm{t}}\right) D_{N_{\mathrm{R}}}\left(\pi \cdot \Delta_l^{\mathrm{r}}\right) \right|^2} \\
&\overset{(c)}{\le} \sum\nolimits_{l=1}^{L} |\mathbf{H}_{\mathrm{a}}(l,l)| \sqrt{1 - \left| D_{N_{\mathrm{T}}}^{*}\left(\frac{\pi}{G}\right) D_{N_{\mathrm{R}}}\left(\frac{\pi}{G}\right) \right|^2}, \qquad (54)
\end{aligned}$$

where (a) follows from the Cauchy-Schwarz inequality. The equality (b) holds because $\left\|\mathbf{a}_{\mathrm{T}}^{*}(\theta_l^{\mathrm{t}}) \otimes \mathbf{a}_{\mathrm{R}}(\theta_l^{\mathrm{r}})\right\|_2^2 = 1$, $\left\|\mathbf{a}_{\mathrm{T}}^{*}(\varphi_l^{\mathrm{t}}) \otimes \mathbf{a}_{\mathrm{R}}(\varphi_l^{\mathrm{r}})\right\|_2^2 = 1$, and $\left(\mathbf{a}_{\mathrm{T}}^{T}(\theta_l^{\mathrm{t}}) \otimes \mathbf{a}_{\mathrm{R}}^{H}(\theta_l^{\mathrm{r}})\right)\left(\mathbf{a}_{\mathrm{T}}^{*}(\varphi_l^{\mathrm{t}}) \otimes \mathbf{a}_{\mathrm{R}}(\varphi_l^{\mathrm{r}})\right) = D_{N_{\mathrm{T}}}\left(\pi \cdot \Delta_l^{\mathrm{t}}\right) D_{N_{\mathrm{R}}}^{*}\left(\pi \cdot \Delta_l^{\mathrm{r}}\right)$ where the third equality follows from (52) and (53); the inequality (c) holds because both Δ_l^{t} and Δ_l^{r} are less than or equal to $1/G$, due to (19). □

The Dirichlet kernel $D_N(\pi/G)$ in (51) monotonically increases and converges to one as G increases from $N/2$ to infinity. Thus, as expected, the quantization error upper bound monotonically decreases to zero as G increases.

Combining Lemmas 3-6, we have the desired upper-bound for $\left\|\mathbf{H} - \bar{\mathbf{H}}^{\mathrm{CS}}\right\|_F^2$ in (37).

Theorem 1: Suppose that the condition in (40) is satisfied for given $\bar{\mathbf{H}}_{\mathrm{a}}$ and $\mathbf{E}$, and that the angles are quantized according to (19). The SSE in (37) is upper-bounded as

$$\begin{aligned}
&\left\|\mathbf{H} - \bar{\mathbf{H}}^{\mathrm{CS}}\right\|_F^2 \\
&\le \frac{L\left(1 + (L-1)\,\mu\left(\bar{\mathbf{A}}_{\mathrm{T}} \otimes \bar{\mathbf{A}}_{\mathrm{R}}\right)\right)}{\left(1 - (L-1)\,\mu\left(\bar{\mathbf{Q}}\right)\right)^2} \\
&\quad \times \left(\sqrt{(1+\beta)\,\frac{\sigma_{\tilde{n}}^2}{P}\log G^2} + \sqrt{E_{\mathrm{UB}}}\right)^2 + E_{\mathrm{UB}} \\
&\le \frac{2 - L\mu\left(\bar{\mathbf{Q}}\right)}{L\mu^2\left(\bar{\mathbf{Q}}\right)} \left(\sqrt{(1+\beta)\,\frac{\sigma_{\tilde{n}}^2}{P}\log G^2} + \sqrt{E_{\mathrm{UB}}}\right)^2 + E_{\mathrm{UB}}, \qquad (55)
\end{aligned}$$

with probability at least $1 - \left(G^{2\beta}\pi\right)^{-1}$, where $\mu\left(\bar{\mathbf{Q}}\right) = \mu\left(\bar{\mathbf{A}}_{\mathrm{T}} \otimes \bar{\mathbf{A}}_{\mathrm{R}}\right) = \max\left\{\left|D_{N_{\mathrm{T}}}(2\pi/G)\right|, \left|D_{N_{\mathrm{R}}}(2\pi/G)\right|\right\}$ from Lemma 3; $E_{\mathrm{UB}} = \|\mathrm{vec}(\mathbf{E})\|_2^2 \le \left(\sum_{l=1}^{L} |\mathbf{H}_{\mathrm{a}}(l,l)|\right)^2 \left(1 - \left|D_{N_{\mathrm{T}}}^{*}\left(\frac{\pi}{G}\right) D_{N_{\mathrm{R}}}\left(\frac{\pi}{G}\right)\right|^2\right)$ from Lemma 6; and the second inequality comes from (45).

Proof: This theorem is a consequence of Lemmas 3-6. □

Corollary 1: In the high SNR regime where $\frac{\sigma_{\tilde{n}}^2}{P}\log G^2$ is approximately zero, the upper-bound in (55) reduces to $(1+L)\,E_{\mathrm{UB}}$ in the limit as $L\mu\left(\bar{\mathbf{Q}}\right)$ approaches one.

Proof: In the high SNR regime, the limit of the bound in (55) as $L\mu\left(\bar{\mathbf{Q}}\right) \to 1$ can be written as $\lim_{L\mu(\bar{\mathbf{Q}})\to 1} \left(1 + \frac{2 - L\mu(\bar{\mathbf{Q}})}{L\mu^2(\bar{\mathbf{Q}})}\right) E_{\mathrm{UB}} = \lim_{\mu(\bar{\mathbf{Q}})\to 1/L} \left(1 + \mu^{-1}\left(\bar{\mathbf{Q}}\right)\right) E_{\mathrm{UB}} = (1+L)\,E_{\mathrm{UB}}$. □

Since $0 < L\mu\left(\bar{\mathbf{Q}}\right) < 1$ from (45) and (46), L^{-1} is the supremum of $\mu\left(\bar{\mathbf{Q}}\right)$ and $(1+L)\,E_{\mathrm{UB}}$ is an SSE upper-bound when $\mu\left(\bar{\mathbf{Q}}\right)$ approaches its maximum value. This bound indicates the following: i) the SSE tends to increase with the number of scatterers L and ii) the SSE tends to decreases as the number of grids G increases. These observations are intuitively reasonable and will be confirmed through computer simulation in Section VI.

V. Pilot Beam Pattern Design

In [22], and [33]–[35], it has been observed that a carefully designed sensing matrix, which minimizes the total coherence of the equivalent dictionary, can improve the CS performance, as compared with randomized sensing matrices which are popular in CS and have been used for the CS-based channel estimators in [12] and [13]. In this section, we consider the design of the sensing matrix $\mathbf{Q}$ in (24), for a given dictionary $\bar{\mathbf{A}}_{\mathrm{T}}^{*} \otimes \bar{\mathbf{A}}_{\mathrm{R}}$, by minimizing the total coherence under the beam constraints for $\mathbf{f}_p$ and $\mathbf{w}_q$ in (1). This design problem involves joint optimization of the RF and baseband processors, which is a difficult task. To avoid the joint design, we shall focus on the design of the baseband processors, $\mathbf{F}_{\mathrm{BB}}$ and $\mathbf{W}_{\mathrm{BB}}$, under the assumption that the RF processors, $\mathbf{F}_{\mathrm{RF}}$ and $\mathbf{W}_{\mathrm{RF}}$, are given (in (4) they are assumed to be unitary matrices).

The total coherence, denoted as $\mu^t\left(\bar{\mathbf{Q}}\right)$, is defined by

$$\mu^t\left(\bar{\mathbf{Q}}\right) \triangleq \sum\nolimits_{m}^{G} \sum\nolimits_{n,m\neq n}^{G} \left(\bar{\mathbf{Q}}(m)^H \bar{\mathbf{Q}}(n)\right)^2, \qquad (56)$$

where $\bar{\mathbf{Q}}$ is the equivalent sensing matrix in (25). The total coherence is the sum of the squared inner products of all

pairs of columns in $\bar{\mathbf{Q}}$ (or equivalently the sum of all squared off-diagonal entries of $\bar{\mathbf{Q}}^H\bar{\mathbf{Q}}$). As in the case of the coherence in Lemma 3, the total coherence can be represented in terms of the transmit and receive total coherences, which are denoted as $\mu^t\left(\mathbf{F}_{\mathrm{BB}}^T\mathbf{F}_{\mathrm{RF}}^T\bar{\mathbf{A}}_{\mathrm{T}}^*\right)$ and $\mu^t\left(\mathbf{W}_{\mathrm{BB}}^H\mathbf{W}_{\mathrm{RF}}^H\bar{\mathbf{A}}_{\mathrm{R}}\right)$, respectively, as shown below.

Lemma 7: The total coherence of equivalent sensing matrix $\bar{\mathbf{Q}} = \left(\mathbf{F}_{\mathrm{BB}}^T\mathbf{F}_{\mathrm{RF}}^T\bar{\mathbf{A}}_{\mathrm{T}}^* \otimes \mathbf{W}_{\mathrm{BB}}^H\mathbf{W}_{\mathrm{RF}}^H\bar{\mathbf{A}}_{\mathrm{R}}\right)$ *in (25) is upper-bounded as*

$$\mu^t\left(\bar{\mathbf{Q}}\right) \le \mu^t\left(\mathbf{F}_{\mathrm{BB}}^T\mathbf{F}_{\mathrm{RF}}^T\bar{\mathbf{A}}_{\mathrm{T}}^*\right)\cdot\mu^t\left(\mathbf{W}_{\mathrm{BB}}^H\mathbf{W}_{\mathrm{RF}}^H\bar{\mathbf{A}}_{\mathrm{R}}\right) \tag{57}$$

Proof: Using the notations $\tilde{\mathbf{F}} \triangleq \mathbf{F}_{\mathrm{BB}}^T\mathbf{F}_{\mathrm{RF}}^T\bar{\mathbf{A}}_{\mathrm{T}}^*$ and $\tilde{\mathbf{W}} \triangleq \mathbf{W}_{\mathrm{BB}}^H\mathbf{W}_{\mathrm{RF}}^H\bar{\mathbf{A}}_{\mathrm{R}}$ in the RHS of (57), we get

$$\begin{aligned}
&\sum\nolimits_m^G\sum\nolimits_{n,m\neq n}^G\left(\bar{\mathbf{Q}}(m)^H\bar{\mathbf{Q}}(n)\right)^2\\
&= \sum\nolimits_m^G\sum\nolimits_{n,m\neq n}^G\left(\left(\tilde{\mathbf{F}}(m)^H\otimes\tilde{\mathbf{W}}(m)^H\right)\left(\tilde{\mathbf{F}}(n)\otimes\tilde{\mathbf{W}}(n)\right)\right)^2\\
&\overset{(a)}{=} \sum\nolimits_m^G\sum\nolimits_{n,m\neq n}^G\left(\tilde{\mathbf{F}}(m)^H\tilde{\mathbf{F}}(n)\right)^2\left(\tilde{\mathbf{W}}(m)^H\tilde{\mathbf{W}}(n)\right)^2\\
&\overset{(b)}{\le} \sqrt{\sum\nolimits_m^G\sum\nolimits_{n,m\neq n}^G\left(\tilde{\mathbf{F}}(m)^H\tilde{\mathbf{F}}(n)\right)^4}\\
&\quad\times\sqrt{\sum\nolimits_m^G\sum\nolimits_{n,m\neq n}^G\left(\tilde{\mathbf{W}}(m)^H\tilde{\mathbf{W}}(n)\right)^4}\\
&\le \sum\nolimits_m^G\sum\nolimits_{n,m\neq n}^G\left(\tilde{\mathbf{F}}(m)^H\tilde{\mathbf{F}}(n)\right)^2\\
&\quad\times\sum\nolimits_m^G\sum\nolimits_{n,m\neq n}^G\left(\tilde{\mathbf{W}}(m)^H\tilde{\mathbf{W}}(n)\right)^2,
\end{aligned}\tag{58}$$

where the equality (a) follows from $(\mathbf{A}\otimes\mathbf{B})(\mathbf{C}\otimes\mathbf{D}) = \mathbf{AC}\otimes\mathbf{BD}$ and the inequality (b) holds due to the Cauchy-Schwarz inequality. □

Based on this lemma, we decompose the design problem for minimizing $\mu^t\left(\bar{\mathbf{Q}}\right)$ into two separate designs: the design of $\mathbf{F}_{\mathrm{BB}}$ minimizing $\mu^t\left(\mathbf{F}_{\mathrm{BB}}^T\mathbf{F}_{\mathrm{RF}}^T\bar{\mathbf{A}}_{\mathrm{T}}^*\right)$ and that of $\mathbf{W}_{\mathrm{BB}}$ for minimizing $\mu^t\left(\mathbf{W}_{\mathrm{BB}}^H\mathbf{W}_{\mathrm{RF}}^H\bar{\mathbf{A}}_{\mathrm{R}}\right)$. Next we describe the process for designing $\mathbf{F}_{\mathrm{BB}}$, starting with the following lemma.

Lemma 8: When the angle grids $\{\varphi_g\}$ *satisfy (19) and* $d = \frac{\lambda}{2}$, *minimizing the total coherence* $\mu^t\left(\mathbf{F}_{\mathrm{BB}}^T\mathbf{F}_{\mathrm{RF}}^T\bar{\mathbf{A}}_{\mathrm{T}}^*\right)$ *is equivalent to minimizing* $\left\|\frac{G}{N_{\mathrm{T}}}\mathbf{F}_{\mathrm{BB}}^T\mathbf{F}_{\mathrm{BB}}^* - \mathbf{I}_{N_{\mathrm{T}}^{\mathrm{Beam}}}\right\|_F^2$.

Proof: Minimizing $\mu^t\left(\mathbf{F}_{\mathrm{BB}}^T\mathbf{F}_{\mathrm{RF}}^T\bar{\mathbf{A}}_{\mathrm{T}}^*\right)$ with respect to $\mathbf{F}_{\mathrm{BB}}$ is equivalent to minimizing

$$\left\|\bar{\mathbf{A}}_{\mathrm{T}}^T\mathbf{F}_{\mathrm{RF}}^*\mathbf{F}_{\mathrm{BB}}^*\mathbf{F}_{\mathrm{BB}}^T\mathbf{F}_{\mathrm{RF}}^T\bar{\mathbf{A}}_{\mathrm{T}}^* - \mathbf{I}_G\right\|_F^2$$

because the off-diagonal terms in $\bar{\mathbf{A}}_{\mathrm{T}}^T\mathbf{F}_{\mathrm{RF}}^*\mathbf{F}_{\mathrm{BB}}^*\mathbf{F}_{\mathrm{BB}}^T\mathbf{F}_{\mathrm{RF}}^T\bar{\mathbf{A}}_{\mathrm{T}}^*$ are the inner products that appear in (56). Now,

$$\begin{aligned}
&\left\|\bar{\mathbf{A}}_{\mathrm{T}}^T\mathbf{F}_{\mathrm{RF}}^*\mathbf{F}_{\mathrm{BB}}^*\mathbf{F}_{\mathrm{BB}}^T\mathbf{F}_{\mathrm{RF}}^T\bar{\mathbf{A}}_{\mathrm{T}}^* - \mathbf{I}_G\right\|_F^2\\
&= \mathrm{Tr}\left(\tilde{\mathbf{F}}^H\tilde{\mathbf{F}}\tilde{\mathbf{F}}^H\tilde{\mathbf{F}} - 2\tilde{\mathbf{F}}^H\tilde{\mathbf{F}} + \mathbf{I}_G\right)\\
&= \mathrm{Tr}\left(\tilde{\mathbf{F}}\tilde{\mathbf{F}}^H\tilde{\mathbf{F}}\tilde{\mathbf{F}}^H - 2\tilde{\mathbf{F}}\tilde{\mathbf{F}}^H + \mathbf{I}_{N_{\mathrm{T}}^{\mathrm{Beam}}}\right) + \left(G - N_{\mathrm{T}}^{\mathrm{Beam}}\right)\\
&= \left\|\mathbf{F}_{\mathrm{BB}}^T\mathbf{F}_{\mathrm{RF}}^T\bar{\mathbf{A}}_{\mathrm{T}}^*\bar{\mathbf{A}}_{\mathrm{T}}^T\mathbf{F}_{\mathrm{RF}}^*\mathbf{F}_{\mathrm{BB}}^* - \mathbf{I}_{N_{\mathrm{T}}^{\mathrm{Beam}}}\right\|_F^2 + \left(G - N_{\mathrm{T}}^{\mathrm{Beam}}\right)\\
&= \left\|\frac{G}{N_{\mathrm{T}}}\mathbf{F}_{\mathrm{BB}}^T\mathbf{F}_{\mathrm{BB}}^* - \mathbf{I}_{N_{\mathrm{T}}^{\mathrm{Beam}}}\right\|_F^2 + \left(G - N_{\mathrm{T}}^{\mathrm{Beam}}\right),
\end{aligned}\tag{59}$$

where the first three equalities follow from the relation between the Frobenious norm and the trace, and the last equality holds because $\bar{\mathbf{A}}_{\mathrm{T}}^*\bar{\mathbf{A}}_{\mathrm{T}}^T = \frac{G}{N_{\mathrm{T}}}\mathbf{I}_{N_{\mathrm{T}}}$, as shown in Lemma 1, and $\mathbf{F}_{\mathrm{RF}}^T\mathbf{F}_{\mathrm{RF}}^* = \mathbf{I}_{N_{\mathrm{T}}}$. This completes the proof. □

The optimization problem under the beam normalization constraint in (1) is written as

$$\begin{aligned}
&\min_{\mathbf{F}_{\mathrm{BB}}}\ \left\|\frac{G}{N_{\mathrm{T}}}\mathbf{F}_{\mathrm{BB}}^T\mathbf{F}_{\mathrm{BB}}^* - \mathbf{I}_{N_{\mathrm{T}}^{\mathrm{Beam}}}\right\|_F^2\\
&\text{subject to}\ \left\|\mathbf{F}_{\mathrm{BB}}(m)\right\|_2^2 = 1,\ m = 1,\ldots,N_{\mathrm{T}}^{\mathrm{Beam}}.
\end{aligned}\tag{60}$$

For the block diagonal $\mathbf{F}_{\mathrm{BB}} = \mathrm{diag}\left(\mathbf{F}_{\mathrm{BB},1},\ldots,\mathbf{F}_{\mathrm{BB},N_{\mathrm{T}}^{\mathrm{Block}}}\right)$ assumed in our system model, the objective function can be decomposed into the following block optimization problems:

$$\begin{aligned}
&\min_{\mathbf{F}_{\mathrm{BB},\bar{p}}}\ \left\|\frac{G}{N_{\mathrm{T}}}\mathbf{F}_{\mathrm{BB},\bar{p}}^T\mathbf{F}_{\mathrm{BB},\bar{p}}^* - \mathbf{I}_{N_{\mathrm{T}}^{\mathrm{Beam}}/N_{\mathrm{T}}^{\mathrm{Block}}}\right\|_F^2\\
&\text{subject to}\ \left\|\mathbf{F}_{\mathrm{BB},\bar{p}}(m)\right\|_2^2 = 1,\quad m = 1,\ldots,\frac{N_{\mathrm{T}}^{\mathrm{Beam}}}{N_{\mathrm{T}}^{\mathrm{Block}}}.
\end{aligned}\tag{61}$$

This optimization problem (61) is non-convex due to the multiple spherical constraints, and may not be efficiently solvable [36]. Fortunately, we can relax the individual power constraints in (61) into a single constraint, as shown below.

Theorem 2: The solution of (61) can be obtained by solving the following optimization problem:

$$\begin{aligned}
&\min_{\mathbf{F}_{\mathrm{BB},\bar{p}}}\ \left\|\frac{G}{N_{\mathrm{T}}}\mathbf{F}_{\mathrm{BB},\bar{p}}^T\mathbf{F}_{\mathrm{BB},\bar{p}}^* - \mathbf{I}_{N_{\mathrm{T}}^{\mathrm{Beam}}/N_{\mathrm{T}}^{\mathrm{Block}}}\right\|_F^2\\
&\text{subject to}\ \sum\nolimits_{m=1}^{N_{\mathrm{T}}^{\mathrm{Beam}}/N_{\mathrm{T}}^{\mathrm{Block}}}\left\|\mathbf{F}_{\mathrm{BB},\bar{p}}(m)\right\|_2^2 = \frac{N_{\mathrm{T}}^{\mathrm{Beam}}}{N_{\mathrm{T}}^{\mathrm{Block}}}.
\end{aligned}\tag{62}$$

Its solution is given by

$$\mathbf{F}_{\mathrm{BB},\bar{p}}^{\mathrm{MTC}} = \bar{\mathbf{U}}\left[\mathbf{I}_{\frac{N_{\mathrm{T}}^{\mathrm{Beam}}}{N_{\mathrm{T}}^{\mathrm{Block}}}}\ \mathbf{O}_{\frac{N_{\mathrm{T}}^{\mathrm{Beam}}}{N_{\mathrm{T}}^{\mathrm{Block}}},N_{\mathrm{RF}}-\frac{N_{\mathrm{T}}^{\mathrm{Beam}}}{N_{\mathrm{T}}^{\mathrm{Block}}}}\right]^T\bar{\mathbf{V}}^H,\tag{63}$$

where $1 \le \bar{p} \le N_{\mathrm{T}}^{\mathrm{block}} = \frac{N_{\mathrm{T}}}{N_{\mathrm{RF}}}$, $\bar{\mathbf{U}} \in \mathbb{C}^{N_{\mathrm{RF}}\times N_{\mathrm{RF}}}$, *and* $\bar{\mathbf{V}} \in \mathbb{C}^{(N_{\mathrm{T}}^{\mathrm{Beam}}/N_{\mathrm{T}}^{\mathrm{Block}})\times(N_{\mathrm{T}}^{\mathrm{Beam}}/N_{\mathrm{T}}^{\mathrm{Block}})}$ *are arbitrary unitary matrices. The superscript 'MTC' in the left-hand-side (LHS) of (63) stands for minimal total coherence (MTC). The columns of* $\mathbf{F}_{\mathrm{BB},\bar{p}}^{\mathrm{MTC}}$ *satisfy the individual constraints of (61).*

Proof: Consider the singular value decomposition (SVD) of $\mathbf{F}_{\mathrm{BB},\bar{p}} = \mathbf{U}_{\mathrm{T}}\Lambda_{\mathrm{T}}\mathbf{V}_{\mathrm{T}}^H$ where $\mathbf{U}_{\mathrm{T}} \in \mathbb{C}^{N_{\mathrm{RF}}\times N_{\mathrm{RF}}}$ and $\mathbf{V}_{\mathrm{T}} \in \mathbb{C}^{(N_{\mathrm{T}}^{\mathrm{Beam}}/N_{\mathrm{T}}^{\mathrm{Block}})\times(N_{\mathrm{T}}^{\mathrm{Beam}}/N_{\mathrm{T}}^{\mathrm{Block}})}$ are unitary matrices and

$$\Lambda_{\mathrm{T}} = \left[\mathrm{diag}\left(\lambda_1,\ldots,\lambda_{\frac{N_{\mathrm{T}}^{\mathrm{Beam}}}{N_{\mathrm{T}}^{\mathrm{Block}}}}\right)\ \mathbf{O}_{\frac{N_{\mathrm{T}}^{\mathrm{Beam}}}{N_{\mathrm{T}}^{\mathrm{Block}}},N_{\mathrm{RF}}-\frac{N_{\mathrm{T}}^{\mathrm{Beam}}}{N_{\mathrm{T}}^{\mathrm{Block}}}}\right]^T$$

with singular values λ_j. The objective function in (62) can be rewritten as

$$\begin{aligned}
&\min_{\mathbf{F}_{\mathrm{BB},\bar{p}}} \left\| \frac{G}{N_{\mathrm{T}}} \mathbf{F}_{\mathrm{BB},\bar{p}}^{T} \mathbf{F}_{\mathrm{BB},\bar{p}}^{*} - \mathbf{I}_{N_{\mathrm{T}}^{\mathrm{Beam}}/N_{\mathrm{T}}^{\mathrm{Block}}} \right\|_F^2 \\
&\quad = \min \left\| \frac{G}{N_{\mathrm{T}}} \mathbf{V}_{\mathrm{T}}^{*} \Lambda_{\mathrm{T}}^{T} \mathbf{U}_{\mathrm{T}}^{T} \mathbf{U}_{\mathrm{T}}^{*} \Lambda_{\mathrm{T}}^{*} \mathbf{V}_{\mathrm{T}}^{T} - \mathbf{I}_{N_{\mathrm{T}}^{\mathrm{Beam}}/N_{\mathrm{T}}^{\mathrm{Block}}} \right\|_F^2 \\
&\quad = \min \left\| \frac{G}{N_{\mathrm{T}}} \mathbf{V}_{\mathrm{T}}^{*} \Lambda_{\mathrm{T}}^{T} \Lambda_{\mathrm{T}}^{*} \mathbf{V}_{\mathrm{T}}^{T} - \mathbf{I}_{N_{\mathrm{T}}^{\mathrm{Beam}}/N_{\mathrm{T}}^{\mathrm{Block}}} \right\|_F^2 \\
&\quad = \min \left\| \frac{G}{N_{\mathrm{T}}} \mathbf{V}_{\mathrm{T}}^{*} \left(\Lambda_{\mathrm{T}}^{T} \Lambda_{\mathrm{T}}^{*} - \frac{N_{\mathrm{T}}}{G} \mathbf{I}_{N_{\mathrm{T}}^{\mathrm{Beam}}/N_{\mathrm{T}}^{\mathrm{Block}}} \right) \mathbf{V}_{\mathrm{T}}^{T} \right\|_F^2 \\
&\quad = \min \left\| \Lambda_{\mathrm{T}}^{T} \Lambda_{\mathrm{T}}^{*} - \frac{N_{\mathrm{T}}}{G} \mathbf{I}_{N_{\mathrm{T}}^{\mathrm{Beam}}/N_{\mathrm{T}}^{\mathrm{Block}}} \right\|_F^2 .
\end{aligned}$$

In a similar manner, the constraint of (62) can be written as $\sum_{m=1}^{N_{\mathrm{T}}^{\mathrm{Beam}}/N_{\mathrm{T}}^{\mathrm{Block}}} \left\| \mathbf{F}_{\mathrm{BB},\bar{p}}(m) \right\|_2^2 = \sum_{j=1}^{N_{\mathrm{T}}^{\mathrm{Beam}}/N_{\mathrm{T}}^{\mathrm{Block}}} \lambda_j^2$. Hence, the problem (62) reduces to

$$\begin{aligned}
&\min_{\lambda_j^2} \ \sum\nolimits_{j=1}^{N_{\mathrm{T}}^{\mathrm{Beam}}/N_{\mathrm{T}}^{\mathrm{Block}}} \left(\lambda_j^2 - \frac{N_{\mathrm{T}}}{G} \right)^2 \\
&\text{subject to} \ \sum\nolimits_{j=1}^{N_{\mathrm{T}}^{\mathrm{Beam}}/N_{\mathrm{T}}^{\mathrm{Block}}} \lambda_j^2 = \frac{N_{\mathrm{T}}^{\mathrm{Beam}}}{N_{\mathrm{T}}^{\mathrm{Block}}}. \qquad (64)
\end{aligned}$$

This problem can be solved by the method of Lagrange multipliers. The Lagrangian is given by

$$\begin{aligned}
\mathcal{L}\left(\lambda_j^2, \nu\right) &= \sum\nolimits_{j=1}^{N_{\mathrm{T}}^{\mathrm{Beam}}/N_{\mathrm{T}}^{\mathrm{Block}}} \left(\lambda_j^2 - \frac{N_{\mathrm{T}}}{G} \right)^2 \\
&\quad + \nu \left(\sum\nolimits_{j=1}^{N_{\mathrm{T}}^{\mathrm{Beam}}/N_{\mathrm{T}}^{\mathrm{Block}}} \lambda_j^2 - \frac{N_{\mathrm{T}}^{\mathrm{Beam}}}{N_{\mathrm{T}}^{\mathrm{Block}}} \right), \qquad (65)
\end{aligned}$$

where ν is the Lagrange multiplier. Since $\mathcal{L}\left(\lambda_i^2, \nu\right)$ is a convex quadratic function of λ_j^2, we can find the optimal λ_j^2 by solving $\nabla_{\lambda_j^2} \mathcal{L}\left(\lambda_j^2, \nu\right) = 2\lambda_j^2 - 2\frac{N_{\mathrm{T}}}{G} + \nu = 0$. The optimal λ_j^2 is given by $\lambda_j^2 = 1$ for all $j = 1, \ldots, \frac{N_{\mathrm{T}}^{\mathrm{Beam}}}{N_{\mathrm{T}}^{\mathrm{Block}}}$. Therefore, the matrix of optimal singular values is $\left[\mathbf{I}_{\frac{N_{\mathrm{T}}^{\mathrm{Beam}}}{N_{\mathrm{T}}^{\mathrm{Block}}}} \ \mathbf{O}_{\frac{N_{\mathrm{T}}^{\mathrm{Beam}}}{N_{\mathrm{T}}^{\mathrm{Block}}}, N_{\mathrm{RF}} - \frac{N_{\mathrm{T}}^{\mathrm{Beam}}}{N_{\mathrm{T}}^{\mathrm{Block}}}} \right]^T$, and the optimal $\mathbf{F}_{\mathrm{BB},\bar{p}}^{\mathrm{MTC}}$ can be written as (63). For each m, the optimal $\mathbf{F}_{\mathrm{BB},\bar{p}}^{\mathrm{MTC}}(m)$ satisfies the constraint in (61) because $\left(\mathbf{F}_{\mathrm{BB},\bar{p}}^{\mathrm{MTC}}\right)^H \mathbf{F}_{\mathrm{BB},\bar{p}}^{\mathrm{MTC}} = \bar{\mathbf{V}} \mathbf{I}_{N_{\mathrm{T}}^{\mathrm{Beam}}/N_{\mathrm{T}}^{\mathrm{Block}}} \bar{\mathbf{V}}^H = \mathbf{I}_{N_{\mathrm{T}}^{\mathrm{Beam}}/N_{\mathrm{T}}^{\mathrm{Block}}}$. □

In a similar manner, we can derive the optimal $\mathbf{W}_{\mathrm{BB}}$ minimizing $\mu^t\left(\mathbf{W}_{\mathrm{BB}}^H \mathbf{W}_{\mathrm{RF}}^H \bar{\mathbf{A}}_{\mathrm{R}}\right)$. The $\bar{q}$-th block of the optimal $\mathbf{W}_{\mathrm{BB}}$ with MTC is given by

$$\mathbf{W}_{\mathrm{BB},\bar{q}}^{\mathrm{MTC}} = \bar{\mathbf{U}} \left[\mathbf{I}_{\frac{N_{\mathrm{R}}^{\mathrm{Beam}}}{N_{\mathrm{R}}^{\mathrm{Block}}}} \ \mathbf{O}_{\frac{N_{\mathrm{R}}^{\mathrm{Beam}}}{N_{\mathrm{R}}^{\mathrm{Block}}}, N_{\mathrm{RF}} - \frac{N_{\mathrm{R}}^{\mathrm{Beam}}}{N_{\mathrm{R}}^{\mathrm{Block}}}} \right]^T \bar{\mathbf{V}}^H, \qquad (66)$$

where $1 \le \bar{q} \le N_{\mathrm{R}}^{\mathrm{block}} = \frac{N_{\mathrm{R}}}{N_{\mathrm{RF}}}$, and both $\bar{\mathbf{U}} \in \mathbb{C}^{N_{\mathrm{RF}} \times N_{\mathrm{RF}}}$ and $\bar{\mathbf{V}} \in \mathbb{C}^{(N_{\mathrm{R}}^{\mathrm{Beam}}/N_{\mathrm{R}}^{\mathrm{Block}}) \times (N_{\mathrm{R}}^{\mathrm{Beam}}/N_{\mathrm{R}}^{\mathrm{Block}})}$ are arbitrary unitary matrices.

Now, we compare the total coherences of the proposed MTC beam patterns and the conventional beams generated by employing randomized $\mathbf{F}_{\mathrm{BB}}$ and $\mathbf{W}_{\mathrm{BB}}$ consisting of i.i.d. Rademacher random variables [12], [13]. Both the beam patterns use the DFT matrix for RF processing ($\mathbf{F}_{\mathrm{RF}}$ and $\mathbf{W}_{\mathrm{RF}}$). Fig. 3 shows histograms of the magnitudes of off-diagonal entries of $\bar{\mathbf{Q}}^H \bar{\mathbf{Q}}$ when $N_{\mathrm{T}} = N_{\mathrm{R}} = 32$, $N_{\mathrm{T}}^{\mathrm{Beam}} = N_{\mathrm{R}}^{\mathrm{Beam}} = 24$, $N_{\mathrm{RF}} = 8$, and $G = 100$. The proposed optimization tends to reduce the magnitudes of off-diagonal entries and reshapes the histogram such that there is a marked shift towards the origin. The reduction of the total coherence is significant: the normalized total coherences, $\mu^t\left(\bar{\mathbf{Q}}\right)/G^2$, are 15.836 for the random $\mathbf{F}_{\mathrm{BB}}/\mathbf{W}_{\mathrm{BB}}$, and 5.175 for the optimal $\mathbf{F}_{\mathrm{BB}}/\mathbf{W}_{\mathrm{BB}}$ with MTC.

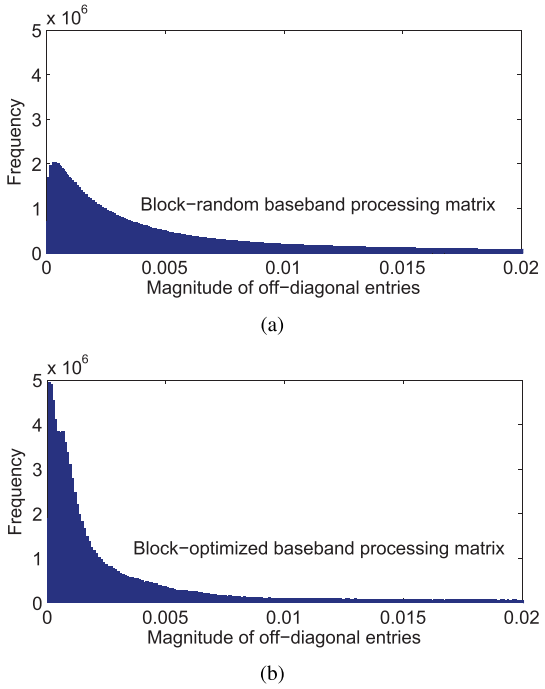

Fig. 3. Histograms of the magnitudes of off-diagonal entries of $\bar{\mathbf{Q}}^H \bar{\mathbf{Q}}$ when $N_{\mathrm{T}} = N_{\mathrm{R}} = 32$, $N_{\mathrm{T}}^{\mathrm{Beam}} = N_{\mathrm{R}}^{\mathrm{Beam}} = 24$, $N_{\mathrm{RF}} = 8$, $G = 100$, and (32×32) DFT matrix is used for RF processing. (a) Randomized $\mathbf{F}_{\mathrm{BB}}$ and $\mathbf{W}_{\mathrm{BB}}$ consisting of i.i.d. Rademacher random variables. (b) Proposed optimal $\mathbf{F}_{\mathrm{BB}}$ and $\mathbf{W}_{\mathrm{BB}}$ with MTC when $\bar{\mathbf{U}}$ and $\bar{\mathbf{V}}$ are DFT matrices.

Some additional remarks of interest illustrating the characteristics of the optimal training and receiving beams are presented below.

1) Although any unitary $\bar{\mathbf{U}}$ and $\bar{\mathbf{V}}$ in (63) and (66) minimizes the total coherence, different types of unitary $(\bar{\mathbf{U}}, \bar{\mathbf{V}})$ matrices result in different coherence values, defined in (33). For example, if we employ the DFT matrix as $(\bar{\mathbf{U}}, \bar{\mathbf{V}})$, the coherence is 0.575 for the parameters in Fig. 3. The coherence increases to 0.82 and 0.868 for the discrete cosine transform (DCT) and the identity matrix, respectively. Since the system performance can be improved further by reducing the coherence, we suggest the use of the DFT matrix as $(\bar{\mathbf{U}}, \bar{\mathbf{V}})$ in (63) and (66).
2) Due to the blocking nature of hybrid MIMO processor, the magnitudes of the entries of $\bar{\mathbf{Q}}^H \bar{\mathbf{Q}}$ exhibit a blocking pattern, as shown in Fig. 4(a), for the following parameters: $N_{\mathrm{T}} = N_{\mathrm{R}} = 32$, $N_{\mathrm{T}}^{\mathrm{Beam}} = N_{\mathrm{R}}^{\mathrm{Beam}} = 24$, $N_{\mathrm{RF}} = 8$, and $G = 32$. In this figure, larger values of the inner products (entries of $\bar{\mathbf{Q}}^H \bar{\mathbf{Q}}$) are clustered, resulting

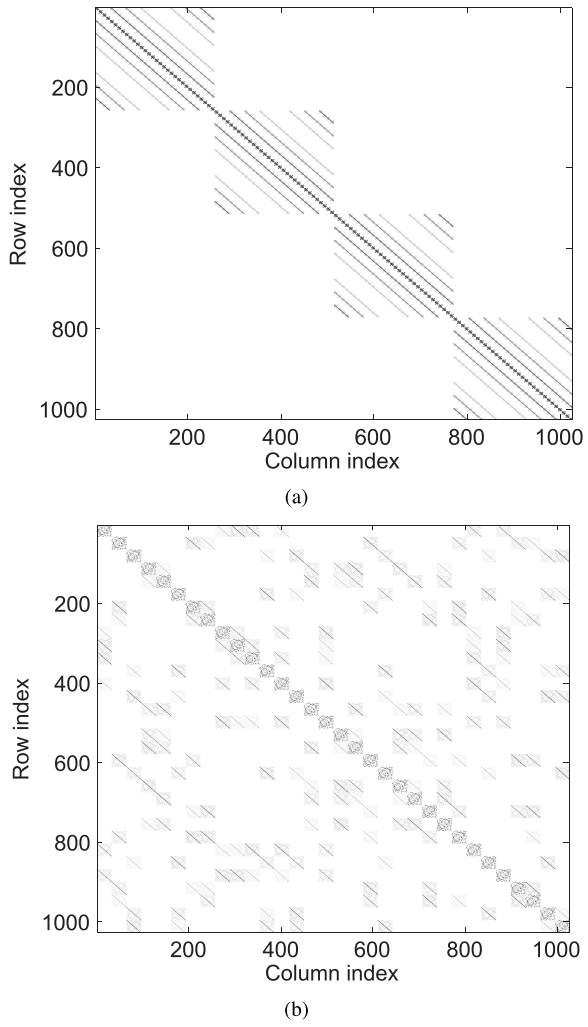

Fig. 4. Magnitudes of the entries of $\bar{\mathbf{Q}}^H\bar{\mathbf{Q}}$ when $N_{\mathrm{T}} = N_{\mathrm{R}} = 32$, $N_{\mathrm{T}}^{\mathrm{Beam}} = N_{\mathrm{R}}^{\mathrm{Beam}} = 24$, $N_{\mathrm{RF}} = 8$, and $G = 32$. (32×32) DFT matrix is used for RF processing. (a) No column permutation. (b) Randomly permuting the columns of $\mathbf{F}_{\mathrm{RF}}$ and $\mathbf{W}_{\mathrm{RF}}$.

in a blocking pattern. This pattern tends to cause performance degradation because the channel consists of clusters of significant paths and the performance for estimating the AoDs/AoAs of each cluster of paths tends to be limited by one or two clusters of larger inner product values. Therefore, it is desirable to remove the pattern by spreading out the larger values. Fortunately, this can be achieved by randomly permuting the columns of $\mathbf{F}_{\mathrm{RF}}$ and $\mathbf{W}_{\mathrm{RF}}$. Fig. 4(b) shows the magnitudes of the entries of $\bar{\mathbf{Q}}^H\bar{\mathbf{Q}}$ after the random permutation. Note that larger values of the inner products are spread out, and there is no blocking pattern. The total coherence of $\bar{\mathbf{Q}}$ after the permutation is equivalent to that of $\bar{\mathbf{Q}}$ without permutation. However, as demonstrated in the following section through simulation, considerable performance improvement can be achieved by randomly permuting the columns of $\mathbf{F}_{\mathrm{RF}}$ and $\mathbf{W}_{\mathrm{RF}}$.

3) The optimal baseband processor $\mathbf{W}_{\mathrm{BB},\bar{q}}^{\mathrm{MTC}}$ in conjunction with the corresponding RF processor $\mathbf{W}_{\mathrm{RF},\bar{q}}$ exhibits a desirable property that the joint RF and baseband processing preserves the whiteness of noise: specifically,

$$\begin{aligned}&\mathbb{E}\left[\left(\mathbf{W}_{\mathrm{BB},\bar{q}}^{\mathrm{MTC}}\right)^H\mathbf{W}_{\mathrm{RF},\bar{q}}^H\mathbf{n}\mathbf{n}^H\mathbf{W}_{\mathrm{RF},\bar{q}}\mathbf{W}_{\mathrm{BB},\bar{q}}^{\mathrm{MTC}}\right]\\&\quad=\sigma_n^2\left(\mathbf{W}_{\mathrm{BB},\bar{q}}^{\mathrm{MTC}}\right)^H\mathbf{W}_{\mathrm{RF},\bar{q}}^H\mathbf{W}_{\mathrm{RF},\bar{q}}\mathbf{W}_{\mathrm{BB},\bar{q}}^{\mathrm{MTC}}\\&\quad=\sigma_n^2\mathbf{I}_{N_{\mathrm{R}}^{\mathrm{Beam}}/N_{\mathrm{R}}^{\mathrm{Block}}},\end{aligned}\tag{67}$$

where $\mathbf{n} \in \mathbb{C}^{N_{\mathrm{R}}\times 1}$ is the white noise vector with $\mathcal{CN}\left(0, \sigma_n^2\mathbf{I}_{N_{\mathrm{R}}}\right)$. (67) indicates that the noise vector after the RF and baseband processing remains i.i.d. Gaussian with $\mathcal{CN}\left(0, \sigma_n^2\mathbf{I}_{N_{\mathrm{R}}^{\mathrm{Beam}}/N_{\mathrm{R}}^{\mathrm{Block}}}\right)$ so that no prewhitening is needed. In contrast, this property does not hold when randomized $\mathbf{W}_{\mathrm{BB},\bar{q}}$ is employed.

4) The $N_{\mathrm{T}}^{\mathrm{Beam}}$ transmitting beams and $N_{\mathrm{R}}^{\mathrm{Beam}}$ receiving beams are used for obtaining the $N_{\mathrm{T}}^{\mathrm{Beam}}N_{\mathrm{R}}^{\mathrm{Beam}}$ measurements which are the entries of $\bar{\mathbf{y}}$ in (12) and (24). In the case of LS estimation in (13), the number of measurements needed for a non-singular $\mathbf{Q}^H\mathbf{Q}$ is at least $N_{\mathrm{T}}^{\mathrm{Beam}}N_{\mathrm{R}}^{\mathrm{Beam}} = N_{\mathrm{T}}N_{\mathrm{R}}$. On the other hand, the OMP-based estimator needs $\mathcal{O}\left(L\ln G^2\right)$ measurements that can be much less than $N_{\mathrm{T}}N_{\mathrm{R}}$ [37], [38]. For example, if $N_{\mathrm{T}} = N_{\mathrm{R}} = 32$, $L = 5$, and $G = 100$, then $L\ln G^2 = 46.05 \ll 1024 = N_{\mathrm{T}}N_{\mathrm{R}}$. Hereafter, the cases with $N_{\mathrm{T}}^{\mathrm{Beam}}N_{\mathrm{R}}^{\mathrm{Beam}} = N_{\mathrm{T}}N_{\mathrm{R}}$ and $N_{\mathrm{T}}^{\mathrm{Beam}}N_{\mathrm{R}}^{\mathrm{Beam}} < N_{\mathrm{T}}N_{\mathrm{R}}$ are referred to as "full-training" and "partial-training", respectively.

5) The optimal patterns in (63) and (66) are designed by minimizing the total coherence and should be useful for any type of CS algorithms [25] that can be applied to solve (26). Furthermore, it can be seen from [39] that the optimal beam patterns are also optimal for the LS estimator in (13) when $N_{\mathrm{T}}^{\mathrm{Beam}}N_{\mathrm{R}}^{\mathrm{Beam}} = N_{\mathrm{T}}N_{\mathrm{R}}$. In our simulation, the derived beam patterns are used for both the OMP and LS estimators.

VI. Simulation Results

The performance of the proposed channel estimator is examined through computer simulation with the following parameters. The transmitter and the receiver are equipped with ULAs with $N_{\mathrm{T}} = N_{\mathrm{R}} \in \{32, 64\}$ and $N_{\mathrm{RF}} = 8$ $\left(N_{\mathrm{T}}^{\mathrm{Block}} = N_{\mathrm{R}}^{\mathrm{Block}} = 4\right)$. For RF processing ($\mathbf{F}_{\mathrm{RF}}$ and $\mathbf{W}_{\mathrm{RF}}$), the DFT matrices are employed. The pilot beam patterns ($\mathbf{F}_{\mathrm{BB},\bar{p}}^{\mathrm{MTC}}$ and $\mathbf{W}_{\mathrm{BB},\bar{q}}^{\mathrm{MTC}}$) are designed using (63) and (66), while setting $\bar{\mathbf{U}}$ and $\bar{\mathbf{V}}$ at DFT matrices. The results in this simulation are obtained through 500 channel realizations with $\sigma_\alpha^2 = 1$ and $\sigma_{\mathrm{AS}} = 15$. We set $\delta = 0.1\sigma_n^2$ in step 2 of Algorithm 1.

The proposed OMP estimator, termed the OMP-redundant, is compared with the LS estimator in (13), the oracle estimator in (16), and the OMP estimator based on the virtual channel model [13], termed the OMP-unitary. The metrics for performance comparison are the normalized mean square error (NMSE) and the achievable spectral efficiency (ASE), which are defined as $10\log_{10}\left(\mathbb{E}\left[\left\|\mathbf{H}-\hat{\mathbf{H}}\right\|_F^2 / \|\mathbf{H}\|_F^2\right]\right)$ and $\log_2\left|\mathbf{I}_{N_{\mathrm{RF}}} + \frac{\tilde{P}}{N_{\mathrm{RF}}}\mathbf{R}_{\mathbf{n}}^{-1}\tilde{\mathbf{W}}^H\mathbf{H}\tilde{\mathbf{F}}\tilde{\mathbf{F}}^H\mathbf{H}^H\tilde{\mathbf{W}}\right|$, respectively, for the channel estimate $\hat{\mathbf{H}}$. Here $\tilde{\mathbf{F}}$ and $\tilde{\mathbf{W}}$ are the optimal precoder and combiner, respectively, which are designed via the SVD

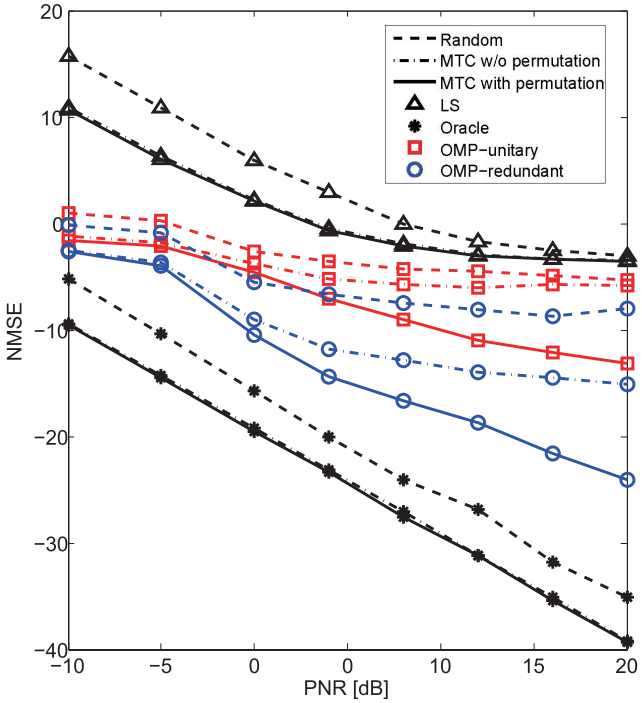

Fig. 5. NMSE performance against PNR for the LS, OMP-unitary, and OMP-redundant estimators for the three types of training beams when $N_{\mathrm{T}} = N_{\mathrm{R}} = 32$ and $N_{\mathrm{T}}^{\mathrm{Beam}} = N_{\mathrm{R}}^{\mathrm{Beam}} = 24$.

of $\hat{\mathbf{H}}$; $\mathbf{R_n} \triangleq \sigma_n^2 \tilde{\mathbf{W}}^H \tilde{\mathbf{W}}$ (see the Appendix for details of ASE derivation). Two types of signal-to-noise ratios (SNRs) are considered: one is the pilot-to-noise ratio (PNR), defined as $10\log_{10}\left(P/\sigma_n^2\right)$; and the other is the data-to-noise ratio (DNR), $10\log_{10}\left(\tilde{P}/\sigma_n^2\right)$, where P and $\tilde{P}$ denote the pilot power and the data power, respectively. We whill show the NMSE curves against PNR and the ASE curves against DNR.

This section consists of two parts. In the first part, the performance of the proposed scheme is compared with those of the other estimators, while employing various training beam patterns. In the second part, the proposed scheme with the optimal training beam is analyzed, and the NMSEs of the proposed scheme are compared with the SSE lower- and upper-bounds derived in Section IV.

A. Performance Comparison

We consider three types of training beam patterns. i) $\mathbf{F}_{\mathrm{BB}}/\mathbf{W}_{\mathrm{BB}}$ are obtained by generating i.i.d. Rademacher random variables and $\mathbf{F}_{\mathrm{RF}}/\mathbf{W}_{\mathrm{RF}}$ are the DFT matrices without any column permutation. This beam is called the random beam. ii) $\mathbf{F}_{\mathrm{BB}}/\mathbf{W}_{\mathrm{BB}}$ are the MTC beams in (63) and (66), and $\mathbf{F}_{\mathrm{RF}}/\mathbf{W}_{\mathrm{RF}}$ are the DFT matrices without any column permutation. This is called the MTC beam without permutation. iii) $\mathbf{F}_{\mathrm{BB}}/\mathbf{W}_{\mathrm{BB}}$ are the MTC beams, and $\mathbf{F}_{\mathrm{RF}}/\mathbf{W}_{\mathrm{RF}}$ are the permuted DFT matrix whose columns are randomly permuted. This is called the MTC beam with permutation. It is assumed that $N_{\mathrm{T}}^{\mathrm{Beam}} = N_{\mathrm{R}}^{\mathrm{Beam}} = N^{\mathrm{Beam}}$. The number of scatterers in the channel (sparsity level) is set at $L = 5$, and the number of angle grids for the OMP-redundant is set at $G = 180$.

Fig. 5 shows the NMSEs of the LS, oracle, OMP-unitary, and OMP-redundant against PNR when $N_{\mathrm{T}} = N_{\mathrm{R}} = 32$ and

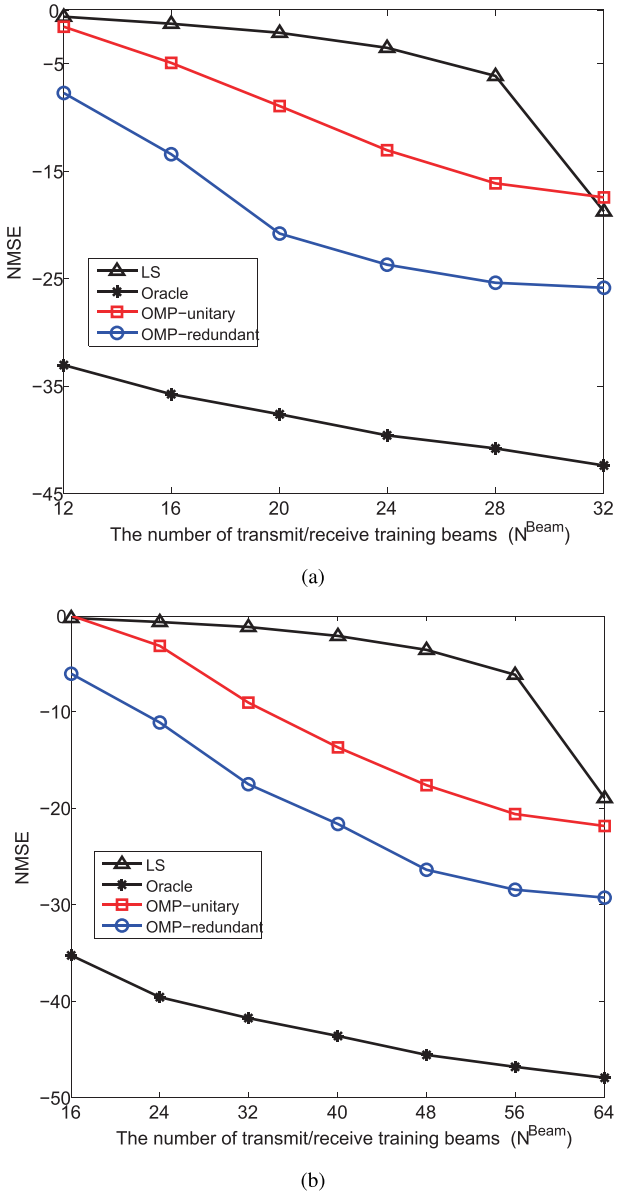

Fig. 6. NMSE performance against the number of training beams when $N_{\mathrm{T}}^{\mathrm{Beam}} = N_{\mathrm{R}}^{\mathrm{Beam}} = N^{\mathrm{Beam}}$, PNR = 20dB, and the MTC training beams with permutation are employed. (a) $N_{\mathrm{T}} = N_{\mathrm{R}} = 32$, (b) $N_{\mathrm{T}} = N_{\mathrm{R}} = 64$.

$N_{\mathrm{T}}^{\mathrm{Beam}} = N_{\mathrm{R}}^{\mathrm{Beam}} = 24$ $\left(N^{\mathrm{Beam}} = 24\right)$. This is a case of partial-training and, compared to the full-training $\left(N^{\mathrm{Beam}} = 32\right)$, the number of measurements $N_{\mathrm{T}}^{\mathrm{Beam}} N_{\mathrm{R}}^{\mathrm{Beam}}$ is reduced by 9/16. As expected, the NMSEs of the oracle estimator, which assumes the knowledge of AoDs/AoAs, are smaller than those of the others. Among the practical estimators, the proposed OMP-redundant always outperforms the others for each training beam pattern. The LS method performs the worst because of partial-training. Comparing the beam patterns, the MTC beams with permutation exhibit the best performance, while the random beams exhibit the worst. The performance gain achieved by permutation is significant: in the case of the OMP-redundant, the NMSE is reduced by about 10dB when PNR = 20dB.

Fig. 6 shows the NMSEs against the number of training beams when PNR = 20dB, $N_{\mathrm{T}} = N_{\mathrm{R}} \in \{32, 64\}$, and the

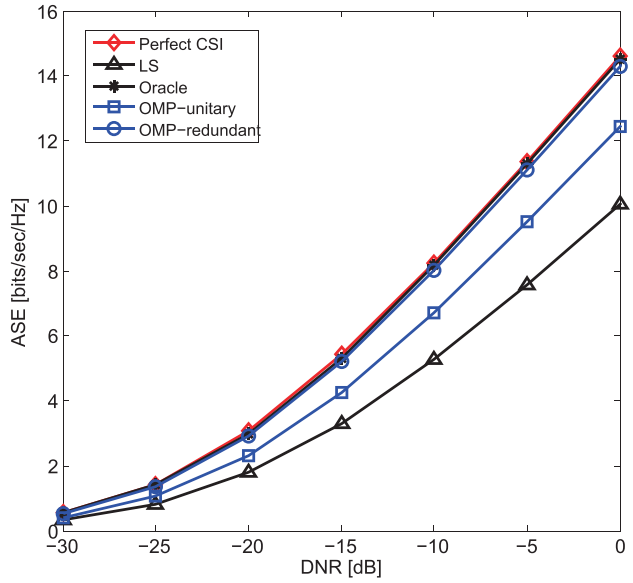

Fig. 7. ASE against DNR when Spectral efficiency against DNR when $N_{\mathrm{T}} = N_{\mathrm{R}} = 32$, $N_{\mathrm{T}}^{\mathrm{Beam}} = N_{\mathrm{R}}^{\mathrm{Beam}} = 24$, $N_{\mathrm{RF}} = 2$, PNR = 0dB and the MTC training beams with permutation are employed.

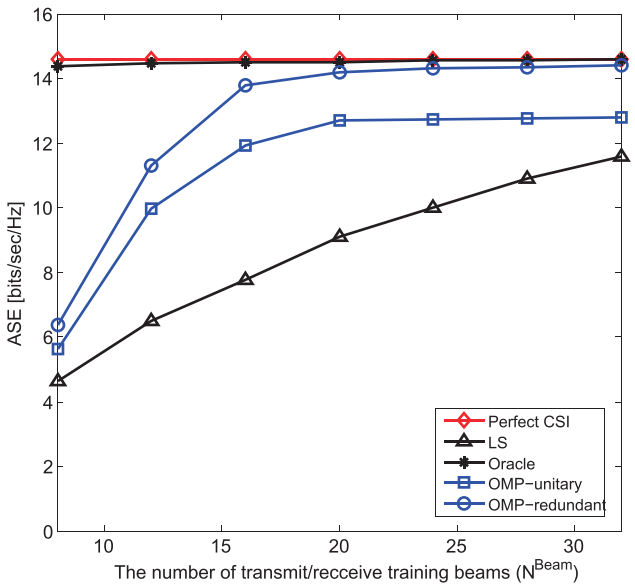

Fig. 8. ASE against the number of training beams when $N_{\mathrm{T}} = N_{\mathrm{R}} = 32$, $N_{\mathrm{T}}^{\mathrm{Beam}} = N_{\mathrm{R}}^{\mathrm{Beam}} = N^{\mathrm{Beam}}$, $N_{\mathrm{RF}} = 2$, PNR = DNR = 0dB, and the MTC training beams with permutation are employed.

MTC beams with permutation are employed. As the number of training beams N^{Beam} increases, the NMSEs of all three estimators decrease monotonically. In the case of the LS estimation, the NMSE drops rapidly when N^{Beam} is increased from 56 to 64. This happens because $\mathbf{Q}^H\mathbf{Q}$ in (13) becomes nonsingular when $N^{\mathrm{Beam}} = 64$. On the other hand, the OMP and the oracle curves tend to decrease slowly. Again, among the practical estimators, the OMP-redundant outperforms the others; its NMSE is considerably less than that of the LS method, which is true even for the cases of full-training. As a result, the OMP-redundant with partial-training can perform like the LS with full-training. For example, in Fig 6(b), the performance of the proposed with $N^{\mathrm{Beam}} = 32$ (partial training) is comparable to that of the LS with $N^{\mathrm{Beam}} = 64$ (full-training). This indicates that the training overhead can be reduced significantly by the OMP redundant.

Fig. 7 shows the ASE curves against DNR when $N_{\mathrm{T}} = N_{\mathrm{R}} = 32$, $N_{\mathrm{T}}^{\mathrm{Beam}} = N_{\mathrm{R}}^{\mathrm{Beam}} = 24$, $N_{\mathrm{RF}} = 2$, PNR = 0dB, and the MTC beams with permutation are employed. It is interesting to note that the ASE curve of the proposed OMP-redundant is close to that of the oracle estimator for all DNR values, while the latter almost overlaps with the ideal ASE curve assuming the perfect CSI. Comparing the ASEs of the practical estimators, the ASE gain achieved by the proposed scheme can be significant. For example, the gain over the LS method is about 5bits/sec/Hz when DNR = 0dB, and the gain tends to increase with the DNR.

Fig. 8 shows the ASE curves against the number of beams when $N_{\mathrm{T}} = N_{\mathrm{R}} = 32$, $N_{\mathrm{RF}} = 2$, PNR = DNR = 0dB, and the MTC beams with permutation are employed. The ASEs of the OMP-based estimators tend to remain constant for $N^{\mathrm{Beam}} \geq 20$, while that of the LS estimator monotonically increases with N^{Beam}. When $N^{\mathrm{Beam}} \geq 20$, the performance gap between the proposed and the oracle estimators is less than 0.4bits/sec/Hz.

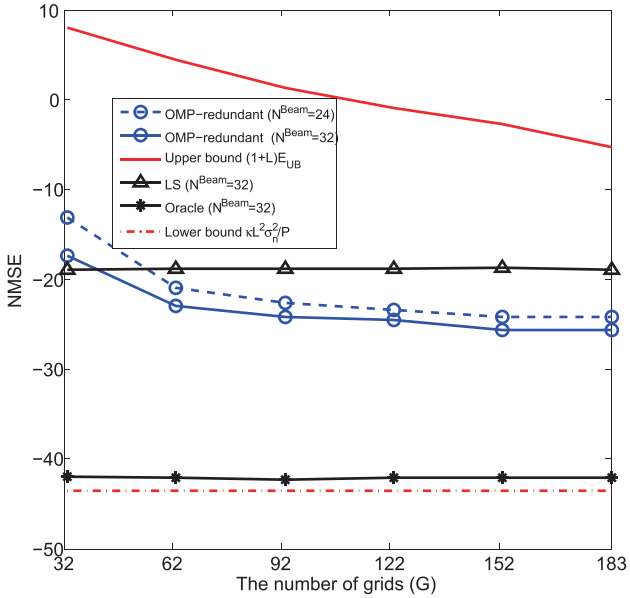

Fig. 9. NMSE performance against the number of grids G when $L = 5$, PNR = 20dB, and $N^{\mathrm{Beam}} \in \{24, 32\}$.

B. Analysis of Proposed OMP

The NMSEs of the OMP-redundant are obtained for various values of G and L, and the results are compared with the SSE lower-bound $\kappa L^2\sigma_n^2/P$ in (29) and the upper bound, $(1+L)\,E_{\mathrm{UB}}$, derived in Corollary 1. The PNR is set at PNR = 20dB. Both full-training ($N^{\mathrm{Beam}} = 32$) and partial-training with $N^{\mathrm{Beam}} = 24$ are considered.

Fig. 9 shows the NMSE curves of the OMP-redundant against the number of grids G. For comparison, the NMSEs of the oracle estimator and the fully trained LS estimator, which are constant for different values of G, and the SSE lower-/upper-bounds are also shown. Both the fully- and partially-trained OMPs outperform the LS when $G \geq 62$, and the

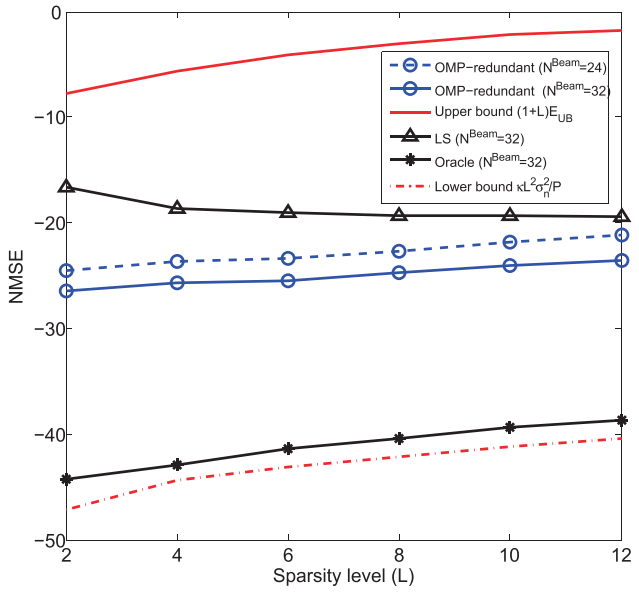

Fig. 10. NMSE performance against the sparsity level L when $G = 180$, PNR $= 20$dB, and $N^{\text{Beam}} \in \{24, 32\}$.

NMSEs of the OMPs monotonically decrease with G, as expected by the SSE upper-bound.

Fig. 10 shows the NMSEs when L varies from 2 to 12. Again the OMPs outperform the LS, and their NMSEs monotonically increase with L, as expected by the SSE lower- and upper-bounds.

Finally, it is noted that there are big gaps between the lower-/upper-bounds and simulation results in Figs. 9 and 10. In the case of the lower-bound, the knowledge of AoDs/AoAs, which is assumed in oracle estimation, has a dominant effect on the gap. For the upper-bound, the gap is caused by the inequalities in equations (37), (39), (41), and (51). (In this case it is difficult to find out which inequality has a dominant effect on the gap.)

VII. Conclusion

An open-loop channel estimator for hybrid MIMO systems in mm-wave communications was proposed. Mm-wave channels are sparse in the sense that they are dominated by a few clusters of significant paths. Exploiting this sparsity, we formulated a CS problem that estimates the AoD, AoA, and the corresponding gain of each significant path. This problem is based on the parametric channel model with quantized angle grids and solved by the OMP method that employs a redundant dictionary consisting of array response vectors with finely quantized angle grids. Given the RF beamforming matrix, which is assumed to be unitary, the baseband processor for training (or the sensing matrix) was designed by minimizing the total coherence of the equivalent sensing matrix. It has been observed that the estimation accuracy can be improved considerably by randomly permuting the RF beamforming matrix. The computer simulation results demonstrated that the proposed OMP method can outperform existing methods such as the LS method and the OMP based on the virtual channel model.

Further work in this area includes the application of more sophisticated CS algorithms than the OMP. Various greedy algorithms and $\mathcal{L}_1$ minimization algorithms [25] can be used for solving the channel estimation problem in (26). Furthermore, the "off the grid" techniques that can solve CS problems with interpolation-based algorithms [40] or without defining the grids [41] would be useful for reducing the angle quantization error, and their application to hybrid MIMO channel estimation needs investigation.

Appendix

We consider the single-user mm-wave system where the transmitter sends N_{RF} streams to the receiver. The transmitted signal can be written as $\mathbf{x} = \tilde{\mathbf{F}}\mathbf{s}$, where $\tilde{\mathbf{F}} \in \mathbb{C}^{N_{\text{T}} \times N_{\text{RF}}}$ is the precoder matrix, $\mathbf{s} \in \mathbb{C}^{N_{\text{RF}} \times 1}$ is the symbol vector with $\mathbb{E}\left[\mathbf{s}\mathbf{s}^H\right] = \frac{\tilde{P}}{N_{\text{RF}}}\mathbf{I}$, and $\tilde{P}$ is the average total transmit power. The received signal after combining can be written as

$$\tilde{\mathbf{y}} = \tilde{\mathbf{W}}^H \mathbf{H}\tilde{\mathbf{F}}\mathbf{s} + \tilde{\mathbf{W}}^H \tilde{\mathbf{n}},$$

where $\tilde{\mathbf{W}} \in \mathbb{C}^{N_{\text{R}} \times N_{\text{RF}}}$ is the combiner matrix, $\mathbf{H}$ is the channel matrix in (7), and $\tilde{\mathbf{n}} \in \mathbb{C}^{N_{\text{R}} \times 1}$ is the noise vector with $\mathcal{CN}\left(0, \sigma_n^2 \mathbf{I}_{N_{\text{R}}}\right)$. The ASE can be written as

$$R = \log_2 \left| \mathbf{I}_{N_{\text{RF}}} + \frac{\tilde{P}}{N_{\text{RF}}} \mathbf{R}_{\mathbf{n}}^{-1} \tilde{\mathbf{W}}^H \mathbf{H}\tilde{\mathbf{F}}\tilde{\mathbf{F}}^H \mathbf{H}^H \tilde{\mathbf{W}} \right|,$$

where $\mathbf{R}_{\mathbf{n}} = \sigma_n^2 \tilde{\mathbf{W}}^H \tilde{\mathbf{W}}$; and $\left\{\tilde{\mathbf{F}}, \tilde{\mathbf{W}}\right\}$ are designed via the SVD of $\hat{\mathbf{H}}$, denoted as $\hat{\mathbf{H}} = \hat{\mathbf{U}}\hat{\Sigma}\hat{\mathbf{V}}^H$. Here $\left\{\hat{\mathbf{U}}, \hat{\mathbf{V}}\right\}$ are the singular vector matrices and $\hat{\Sigma}$ is the singular value matrix whose diagonal entries are arranged in decreasing order. The precoder $\tilde{\mathbf{F}}$ and the combiner $\tilde{\mathbf{W}}$ consist of the first N_{RF} columns of $\hat{\mathbf{V}}$ and $\hat{\mathbf{U}}$, respectively.

References

[1] S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter-wave cellular wireless networks: Potentials and challenges," *Proc. IEEE*, vol. 102, no. 3, pp. 366–385, Mar. 2014.

[2] T. Bai, A. Alkhateeb, and R. W. Heath, Jr., "Coverage and capacity of millimeter-wave cellular networks," *IEEE Commun. Mag.*, vol. 52, no. 9, pp. 70–77, Sep. 2014.

[3] *Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications—Amendment 3: Enhancements for Very High Throughput in the 60 GHz Band*, IEEE Standard P802.11.ad, Jan. 2013.

[4] *Part 15.3: Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for High Rate Wireless Personal Area Networks (WPANs)—Amendment 2: Millimeter-Wave Based Alternative Physical Layer Extension*, IEEE Standard 802.15.3c, Oct. 2009.

[5] M. R. Akdeniz *et al.*, "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1164–1179, Jun. 2014.

[6] O. El Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, Jr., "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, Mar. 2014.

[7] J. Lee and Y. H. Lee, "AF relaying for millimeter wave communication systems with hybrid RF/baseband MIMO processing," in *Proc. IEEE Int. Conf. Commun.*, Jun. 2014, pp. 5838–5842.

[8] S. Han, I. Chih-Lin, Z. Xu, and C. Rowell, "Large-scale antenna systems with hybrid analog and digital beamforming for millimeter wave 5G," *IEEE Commun. Mag.*, vol. 53, no. 1, pp. 186–194, Jan. 2015.

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

[10] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, Jr., "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 831–846, Oct. 2014.
[11] 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.
[12] W. U. Bajwa, J. Haupt, A. M. Sayeed, and R. Nowak, "Compressed channel sensing: A new approach to estimating sparse multipath channels," *Proc. IEEE*, vol. 98, no. 6, pp. 1058–1076, Jun. 2010.
[13] X. Rao and V. K. N. Lau, "Distributed compressive CSIT estimation and feedback for FDD multi-user massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 62, no. 12, pp. 3261–3271, Jun. 2014.
[14] 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.
[15] Z. Gao, L. Dai, D. Mi, Z. Wang, M. A. Imran, and M. Z. Shakir, "MmWave massive-MIMO-based wireless backhaul for the 5G ultra-dense network," *IEEE Wireless Commun.*, vol. 22, no. 5, pp. 13–21, Oct. 2015.
[16] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.
[17] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
[18] A. M. Sayeed, "Deconstructing multiantenna fading channels," *IEEE Trans. Signal Process.*, vol. 50, no. 10, pp. 2563–2579, Oct. 2002.
[19] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005.
[20] C. R. Berger, S. Zhou, J. C. Preisig, and P. Willett, "Sparse channel estimation for multicarrier underwater acoustic communication: From subspace methods to compressed sensing," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1708–1721, Mar. 2010.
[21] D. Hu, X. Wang, and L. He, "A new sparse channel estimation and tracking method for time-varying OFDM systems," *IEEE Trans. Veh. Technol.*, vol. 62, no. 9, pp. 4648–4653, Nov. 2013.
[22] L. Zelnik-Manor, K. Rosenblum, and Y. C. Eldar, "Sensing matrix optimization for block-sparse decoding," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4300–4312, Sep. 2011.
[23] S. Li, "Matrix results on the Khatri-Rao and Tracy-Singh products," *Linear Algebra Appl.*, vol. 289, pp. 267–277, Mar. 1999.
[24] 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.
[25] Y. C. Eldar and G. Kutyniok, Eds., *Compressed Sensing: Theory and Applications*. New York, NY, USA: Cambridge Univ. Press, 2012.
[26] Z. Ben-Haim, Y. C. Eldar, and M. Elad, "Coherence-based performance guarantees for estimating a sparse vector under random noise," *IEEE Trans. Signal Process.*, vol. 58, no. 10, pp. 5030–5043, Oct. 2010.
[27] R. A. Horn and C. R. Johnson, *Matrix Analysis*. New York, NY, USA: Cambridge Univ. Press, 1988.
[28] P. S. Bullen, *Handbook of Means and Their Inequalities*. Boston, MA, USA: Kluwer, 2003.
[29] S. Jokar and V. Mehrmann, "Sparse solutions to underdetermined Kronecker product systems," *Linear Algebra Appl.*, vol. 431, no. 12, pp. 2437–2447, Dec. 2009.
[30] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed. Baltimore, MD, USA: The Johns Hopkins Univ. Press, 1996.
[31] Y. Chi and R. Calderbank, "Coherence-based performance guarantees of orthogonal matching pursuit," in *Proc. 50th Annu. Allerton Conf. Commun., Control, Comput.*, Sep. 2012, pp. 2003–2009.
[32] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2231–2242, Oct. 2004.
[33] M. Elad, "Optimized projections for compressed sensing," *IEEE Trans. Signal Process.*, vol. 55, no. 12, pp. 5695–5702, Dec. 2007.
[34] J. M. Duarte-Carvajalino and G. Sapiro, "Learning to sense sparse signals: Simultaneous sensing matrix and sparsifying dictionary optimization," *IEEE Trans. Image Process.*, vol. 18, no. 7, pp. 1395–1408, Jul. 2009.
[35] G. Li, Z. Zhu, D. Yang, L. Chang, and H. Bai, "On projection matrix optimization for compressive sensing systems," *IEEE Trans. Signal Process.*, vol. 61, no. 11, pp. 2887–2898, Jun. 2013.
[36] R. Lai and S. Osher, "A splitting method for orthogonality constrained problems," *J. Sci. Comput.*, vol. 58, no. 2, pp. 431–449, Feb. 2014.
[37] H. Rauhut, K. Schnass, and P. Vandergheynst, "Compressed sensing and redundant dictionaries," *IEEE Trans. Inf. Theory*, vol. 54, no. 5, pp. 2210–2219, May 2008.
[38] M. A. Davenport, D. Needell, and M. B. Wakin, "Signal space CoSaMP for sparse recovery with redundant dictionaries," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6820–6829, Oct. 2013.
[39] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.
[40] K. Fyhn, M. F. Duarte, and S. H. Jensen, "Compressive parameter estimation for sparse translation-invariant signals using polar interpolation," *IEEE Trans. Signal Process.*, vol. 63, no. 4, pp. 870–881, Feb. 2015.
[41] G. Tang, B. N. Bhaskar, P. Shah, and B. Recht, "Compressed sensing off the grid," *IEEE Trans. Inf. Theory*, vol. 59, no. 11, pp. 7465–7490, Nov. 2013.

Junho Lee received the B.S. degree in electrical engineering from Kwangwoon University, Seoul, South Korea, in 2010, and the M.S. degree from the University of Science and Technology–Electronics and Telecommunications Research Institute (ETRI), Daejeon, South Korea, in 2013. He is currently pursuing the Ph.D. degree with the Korea Advanced Institute of Science and Technology, Daejeon. From 2010 to 2011, he was an Intern with the Smart Radio Research Team, ETRI. His research interests include sparsity-aware statistical signal processing and its applications, especially in wireless communications.

Gye-Tae Gil (M'04) was born in Daejeon, South Korea, in 1966. He received the B.S. degree in electronic communication engineering from Hanyang University, Seoul, South Korea, in 1989, and the M.S. and Ph.D. degrees in electrical engineering from the Korea Advanced Institute of Science and Technology (KAIST), South Korea, in 1992 and 2004, respectively. Since 1991, he has been with the Research Center of Korea Telecom, and joined the KAIST Institute in 2013, where he is currently a Research Professor. He is also interested in orbital angular momentum transmission and massive antenna technologies for cellular mobile communication systems. His research interests are in the area of communication signal processing, which includes synchronization, interference cancellation, and adaptive filter design.

Yong H. Lee (S'80–M'84–SM'98) received the B.S. and M.S. degrees from Seoul National University, Seoul, South Korea, in 1978 and 1980, respectively, and the Ph.D. degree from the University of Pennsylvania, Philadelphia, USA, in 1984, all in electrical engineering.

He was an Assistant Professor with the Department of Electrical and Computer Engineering, State University of New York, Buffalo, from 1984 to 1988. Since 1989, he has been with the Department of Electrical Engineering, Korea Advanced Institute of Science and Technology (KAIST), where he is currently a Professor. He was the Provost from 2011 to 2013, the Dean of Engineering from 2005 to 2011, and the Head of Electrical Engineering from 2004 to 2005 with KAIST. His research activities are in the area of communication signal processing, which includes new waveform design, interference management, resource allocation, synchronization, detection and estimation for multiple-input multiple-output, wireless communication systems, including line-of-sight communications, and millimeter wave communication systems.