Exploiting Spatial Sparsity for Estimating Channels of Hybrid MIMO Systems in Millimeter Wave Communications

Junho Lee[†], Gye-Tae Gil[‡], and Yong H. Lee[†]
[†]Dept. of Electrical Engineering, Korea Advanced Institute of Science and Technology (KAIST), Korea
[‡]IT Convergence Lab., Korea Advanced Institute of Science and Technology (KAIST), Korea
Email: jhlee@stein.kaist.ac.kr, {gategil, yohlee}@kaist.ac.kr

***Abstract*—Hybrid multiple input multiple output (MIMO) systems consist of an analog beamformer with large antenna arrays followed by a digital MIMO processor. Channel estimation for hybrid MIMO systems in millimeter wave (mm-wave) communications is challenging because of the large antenna array and the low signal-to-noise ratio (SNR) before beamforming. In this paper, we propose an open-loop channel estimator for mm-wave hybrid MIMO systems exploiting the sparse nature of mm-wave channels. A sparse signal recovery problem is formulated for channel estimation and solved by the orthogonal matching pursuit (OMP) based methods. A modification of the OMP algorithm, called the multi-grid (MG) OMP, is proposed. It is shown that the MG-OMP can significantly reduce the computational load of the OMP method. A process for designing the training beams is also developed. Specifically, given the analog training beams the baseband processor for beam training is designed. Simulation results demonstrate the advantage of the OMP based methods over the conventional least squares (LS) method and the efficiency of the MG-OMP over the original OMP.**

I. Introduction

Recently, hybrid MIMO systems consisting of analog beamformers in RF domain and digital MIMO processors in baseband have been recognized as a useful technique for reducing the cost for implementing MIMO systems. In these systems more antennas can be employed without increasing the number of costly RF chains, consisting of amplifiers, mixers, and analog-to-digital (AD)/digital-to-analog (DA) converters. Hybrid MIMO processors have been proposed for both current microwave communications [1], [2] and millimeter wave (mm-wave) communications [3], [4].

Hybrid MIMO systems in mm-wave communications employ a large number of antennas, which is considerably greater than the number of RF chains, to improve the signal-to-noise ratio (SNR) by analog beamforming. In mm-wave systems, analog beamformers have been designed through a *closed-loop* beam training process consisting of beam pattern generation using a codebook at the transmitter, best beam selection at the receiver, and feedback of the selected beam indices [5]-[8]. The beam training is performed iteratively starting with a wide beam, and it reduces the beamwidth until it reaches the desired resolution. This closed-loop process can efficiently design the analog beamformers without channel knowledge at the transmitter. However, use of this technique for hybrid MIMO systems will result in a suboptimal scheme whose analog beamformer is designed without taking account of the digital MIMO processor at the baseband. Techniques for jointly designing the analog and digital parts of hybrid MIMO processors have been developed by exploiting the sparse nature of mm-wave channels [3], [4], [9]. Assuming channel knowledge at the transmitter, these schemes adopt the orthogonal matching pursuit (OMP) algorithm in the compressed sensing field [10], to perform iteratively the joint beam selection and baseband design. The OMP-based techniques can outperform the beam training-based design at the cost of channel knowledge at the transmitter. These days per antenna channel estimation is becoming an important issue in mm-wave systems, as in the case of conventional microwave systems.

Channel estimation for mm-wave hybrid MIMO systems is challenging because of the large antenna array and the low SNR before beamforming. Recently, an adaptive compressed sensing based algorithm in conjunction with closed-loop beam training has been proposed to estimate such mm-wave channels [11]. This technique designs the codebook for beam training in terms of analog/digital precoders, and it can exhibit excellent performance characteristics. However, application of this estimator to outdoor mm-wave channels would be difficult, because outdoor environmental factors require a much larger beamforming gain [8], which limits the use of wide beams for beam training and the feedback channel.

In this paper, we develop an *open-loop* channel estimator for mm-wave hybrid MIMO systems exploiting the sparse nature of mm-wave channels. The proposed scheme uses training beams with fixed beamwidth and estimates the channel without any feedback from the receiver. We formulate a sparse signal recovery problem and solve the problem by the OMP based algorithms [10], [12], [13]. A modification of the OMP algorithm, called the multi-grid (MG) OMP, is proposed to improve the efficiency the OMP algorithm. It is shown that the MG-OMP can significantly reduce the computational complexity of the OMP. We also develop a process for designing the baseband processor for beam training given the analog training

This work was supported by ICT R&D Program of MSIP/IITP. [1482304001, Development of small basestation supporting multiple streams based on LTE-A systems]

978-1-4799-3512-3/14/$31.00 ©2014 IEEE

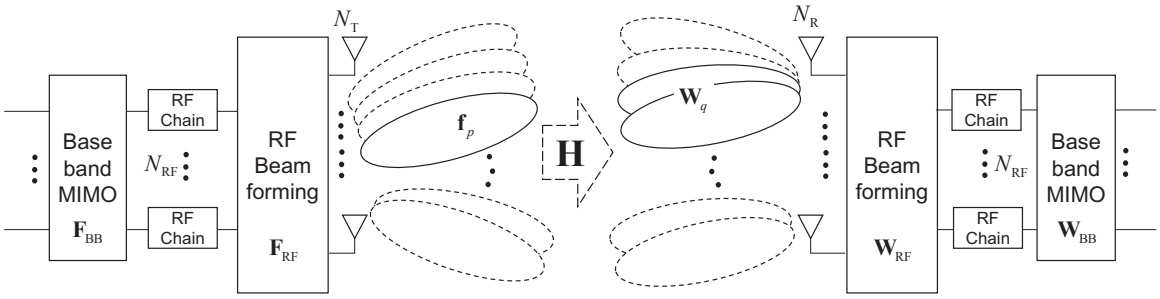

Fig. 1. A mm-wave system employing hybrid MIMO processors.

beams. Computer simulation results demonstrate the following: i) the proposed OMP based estimators can outperform the conventional least squares (LS) method and can efficiently achieve high resolution channel estimates without employing fine beams, ii) the computational saving achieved by the MG-OMP is significant.

The organization of the paper is as follows. Section II presents the system model. The compressed sensing based channel estimation problem is formulated and solved in Section III. The training beam patterns for the hybrid MIMO channel estimation are designed in Section IV. Simulation results showing the advantage of the proposed schemes over the conventional LS method are presented in Section V. Finally, the conclusion is presented in Section VI.

Notations: Bold uppercase $\mathbf{A}$ denotes a matrix and bold lowercase $\mathbf{a}$ denotes a vector. Superscripts $\mathbf{A}^*, \mathbf{A}^T, \mathbf{A}^H, \mathbf{A}^{-1}$ denote the conjugate, the transpose, the conjugate transpose, and the inverse of a matrix $\mathbf{A}$, respectively. $\mathbb{E}[\cdot]$ denotes the expectation and $\mathrm{diag}(\mathbf{A}_1, \ldots, \mathbf{A}_N)$ 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 $\mathbf{a}$. $\|\mathbf{A}\|_F$ is the Frobenius norm, and $\mathbf{A}(n)$ denotes the *n*-th column of a matrix $\mathbf{A}$. $\mathbf{I}_N$ denotes the $N \times N$ identity 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.

II. System Model

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

A. Signal model for open-loop beam training

We consider the single user 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 where $N_{\mathrm{RF}} \leq \min(N_{\mathrm{T}}, N_{\mathrm{R}})$. It is assumed that the RF beamformers are phased array beamformers having coefficients with unit magnitude. For channel estimation, the transmitter has $N_{\mathrm{T}}^{\mathrm{Beam}} \leq N_{\mathrm{T}}$ training beam patterns denoted as $\left\{\mathbf{f}_p \in \mathbb{C}^{N_{\mathrm{T}} \times 1} : p = 1, \ldots, N_{\mathrm{T}}^{\mathrm{Beam}}\right\}$, and the receiver has $N_{\mathrm{R}}^{\mathrm{Beam}} \leq N_{\mathrm{R}}$ beam patterns denoted as $\left\{\mathbf{w}_q \in \mathbb{C}^{N_{\mathrm{R}} \times 1} : q = 1, \ldots, N_{\mathrm{R}}^{\mathrm{Beam}}\right\}$. Here, we assume that $N_{\mathrm{T}}^{\mathrm{Beam}}$ and $N_{\mathrm{R}}^{\mathrm{Beam}}$ are multiples of N_{RF}, and denote $\frac{N_{\mathrm{T}}^{\mathrm{Beam}}}{N_{\mathrm{RF}}}$ and $\frac{N_{\mathrm{R}}^{\mathrm{Beam}}}{N_{\mathrm{RF}}}$ by $N_{\mathrm{T}}^{\mathrm{Block}}$ and $N_{\mathrm{R}}^{\mathrm{Block}}$, respectively. During the training period, the transmitter successively sends its training beams $\{\mathbf{f}_p\}$, and at the receiver each training beam is received through its $N_{\mathrm{R}}^{\mathrm{Beam}}$ beam patterns $\{\mathbf{w}_q\}$. Since the receiver has N_{RF} RF chains, it can generate N_{RF} beams simultaneously and receives the vector $\mathbf{y}_q \in \mathbb{C}^{N_{\mathrm{RF}} \times 1}$ for $q \in \left\{1, \ldots, N_{\mathrm{R}}^{\mathrm{Block}}\right\}$. Here q denotes the received block index, and $N_{\mathrm{R}}^{\mathrm{Block}}$ is the number of blocks. The received vector for the q-th block and the p-th transmit beam is given by

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

where P is the transmit power, $\mathbf{W}_q = \left[\mathbf{w}_{(q-1)N_{\mathrm{RF}}+1} \cdots \mathbf{w}_{qN_{\mathrm{RF}}}\right] \in \mathbb{C}^{N_{\mathrm{R}} \times N_{\mathrm{RF}}}$, $\mathbf{H} \in \mathbb{C}^{N_{\mathrm{R}} \times N_{\mathrm{T}}}$ represents the channel matrix, and $\mathbf{n} \in \mathbb{C}^{N_{\mathrm{R}} \times 1}$ is the noise vector with $\mathcal{CN}\left(0, \sigma_n^2 \mathbf{I}_{N_{\mathrm{R}}}\right)$ (here, the transmit signal is assumed to be 1). Collecting $\mathbf{y}_{q,p}$ for $q \in \left\{1, \ldots, N_{\mathrm{R}}^{\mathrm{Block}}\right\}$, we get $\mathbf{y}_p \in \mathbb{C}^{N_{\mathrm{R}}^{\mathrm{Beam}} \times 1}$ given by

$$\mathbf{y}_p = \sqrt{P}\mathbf{W}^H \mathbf{H}\mathbf{f}_p + \mathrm{diag}\left(\mathbf{W}_1^H, \ldots, \mathbf{W}_{N_{\mathrm{R}}^{\mathrm{Block}}}^H\right)\left[\mathbf{n}_{1,p}^T, \ldots, \mathbf{n}_{N_{\mathrm{R}}^{\mathrm{Block}},p}^T\right]^T, \tag{2}$$

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

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

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

In the hybrid MIMO framework, the transmit and receive training matrices are decomposed as $\mathbf{F} = \mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}$ and $\mathbf{W} = \mathbf{W}_{\mathrm{RF}}\mathbf{W}_{\mathrm{BB}}$, where $\mathbf{F}_{\mathrm{RF}} \in \mathbb{C}^{N_{\mathrm{T}} \times N_{\mathrm{T}}^{\mathrm{Beam}}}$ and $\mathbf{W}_{\mathrm{RF}} \in \mathbb{C}^{N_{\mathrm{R}} \times N_{\mathrm{R}}^{\mathrm{Beam}}}$ represent the RF beamforming matrices, and $\mathbf{F}_{\mathrm{BB}} \in \mathbb{C}^{N_{\mathrm{T}}^{\mathrm{Beam}} \times N_{\mathrm{T}}^{\mathrm{Beam}}}$ and $\mathbf{W}_{\mathrm{BB}} \in \mathbb{C}^{N_{\mathrm{R}}^{\mathrm{Beam}} \times N_{\mathrm{R}}^{\mathrm{Beam}}}$ represent

the baseband processing matrices. These matrices are designed under the following assumptions: i) $\mathbf{F}_{\mathrm{RF}}$ and $\mathbf{W}_{\mathrm{RF}}$ are unitary matrices, ii) $N_{\mathrm{T}}^{\mathrm{Beam}}$ beams generated by the columns of $\mathbf{F}_{\mathrm{RF}}$ cover all angles of departures (AoDs), iii) $N_{\mathrm{R}}^{\mathrm{Beam}}$ beams generated by the columns of $\mathbf{W}_{\mathrm{RF}}$ cover all angles of arrivals (AoAs), iv) $\mathbf{F}_{\mathrm{BB}}$ and $\mathbf{W}_{\mathrm{BB}}$ are block diagonal matrices given by $\mathbf{F}_{\mathrm{BB}} = \mathrm{diag}\left(\mathbf{F}_{\mathrm{BB},1},\ \ldots,\ \mathbf{F}_{\mathrm{BB},i},\ \ldots,\ \mathbf{F}_{\mathrm{BB},N_{\mathrm{T}}^{\mathrm{Block}}}\right)$ and $\mathbf{W}_{\mathrm{BB}} = \mathrm{diag}\left(\mathbf{W}_{\mathrm{BB},1},\ldots,\mathbf{W}_{\mathrm{BB},i},\ldots,\mathbf{W}_{\mathrm{BB},N_{\mathrm{R}}^{\mathrm{Block}}}\right)$ whose diagonal entries, $\mathbf{F}_{\mathrm{BB},i}$ and $\mathbf{W}_{\mathrm{BB},i}$, consist of $N_{\mathrm{RF}} \times N_{\mathrm{RF}}$ complex valued matrices.

B. Channel Model

We use the parametric channel model [3] given by

$$\mathbf{H} = \sqrt{\frac{N_{\mathrm{T}}N_{\mathrm{R}}}{L}} \sum_{l=1}^{L} \alpha_l \mathbf{a}_{\mathrm{r}}\left(\theta_l^{\mathrm{r}}\right) \mathbf{a}_{\mathrm{t}}^{H}\left(\theta_l^{\mathrm{t}}\right), \tag{4}$$

where L is the number of scatterers, 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. We assume the uniform linear arrays 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. 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 $[-\pi,\pi)$, and angular standard deviation is σ_{AS}. The channel model in (4) can be rewritten in matrix form as

$$\mathbf{H} = \mathbf{A}_{\mathrm{R}}\mathbf{H}_{\mathrm{a}}\mathbf{A}_{\mathrm{T}}^{H}, \tag{5}$$

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

III. Proposed Sparse Channel Estimation

To formulate the sparse estimation problem, it is necessary to vectorize the received signal matrix $\mathbf{Y}$ in (3). Denoting $\mathrm{vec}\left(\mathbf{Y}\right)$ by $\bar{\mathbf{y}}$, (3) is rewritten as

$$\begin{aligned}\bar{\mathbf{y}} &= \sqrt{P}\left(\left(\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}\right)^{T} \otimes \mathbf{W}_{\mathrm{BB}}^{H}\mathbf{W}_{\mathrm{RF}}^{H}\right)\cdot \mathrm{vec}\left(\mathbf{H}\right) + \bar{\mathbf{n}}, &(6)\\ &= \mathbf{Q}\cdot\mathrm{vec}\left(\mathbf{H}\right) + \bar{\mathbf{n}}, &(7)\end{aligned}$$

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 $\mathbf{Q} = \sqrt{P}\left(\left(\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}\right)^{T}\otimes\mathbf{W}_{\mathrm{BB}}^{H}\mathbf{W}_{\mathrm{RF}}^{H}\right) \in \mathbb{C}^{N_{\mathrm{T}}^{\mathrm{Beam}}N_{\mathrm{R}}^{\mathrm{Beam}}\times N_{\mathrm{T}}N_{\mathrm{R}}}$. Given (7), a natural approach to estimating $\mathrm{vec}\left(\mathbf{H}\right)$ is the LS approach, which results in a closed-form solution given by $\left(\mathbf{Q}^{H}\mathbf{Q}\right)^{-1}\mathbf{Q}^{H}\bar{\mathbf{y}}$ when $N_{\mathrm{T}}^{\mathrm{Beam}}N_{\mathrm{R}}^{\mathrm{Beam}} \geq N_{\mathrm{T}}N_{\mathrm{R}}$. However, use of this solution for mm-wave communication is difficult because $(N_{\mathrm{T}}, N_{\mathrm{R}})$ are large integers and evaluating the inverse of $\mathbf{Q}^{H}\mathbf{Q} \in \mathbb{C}^{N_{\mathrm{T}}N_{\mathrm{R}}\times N_{\mathrm{T}}N_{\mathrm{R}}}$ needs heavy computations. The compressed sensing based channel estimation reduces the computational load by exploiting the sparse nature of the channel.

To apply compressed sensing techniques to the channel estimation, we first select the set of discrete angles, called the grid, defined as $\mathbf{\Psi}_G = \{\varphi_g \in [0,\pi) : g = 1,\ldots,G\}$. Here $\mathbf{\Psi}_G$ includes all candidate angles of departures and arrivals. $\{\varphi_g\}$ are uniformly distributed in the candidate angle space, and $G \gg L$ to achieve the desired resolution. Then, we define the array response matrices $\bar{\mathbf{A}}_{\mathrm{T}} \in \mathbb{C}^{N_{\mathrm{T}}\times G}$ and $\bar{\mathbf{A}}_{\mathrm{R}} \in \mathbb{C}^{N_{\mathrm{R}}\times G}$ whose columns are the array response vectors corresponding to the candidate angles in $\mathbf{\Psi}_G$. Specifically, $\bar{\mathbf{A}}_{\mathrm{T}} = \left[\mathbf{a}_{\mathrm{t}}\left(\varphi_1\right),\ldots,\mathbf{a}_{\mathrm{t}}\left(\varphi_g\right),\ldots,\mathbf{a}_{\mathrm{t}}\left(\varphi_G\right)\right]$, and $\bar{\mathbf{A}}_{\mathrm{R}} = \left[\mathbf{a}_{\mathrm{r}}\left(\varphi_1\right),\ldots,\mathbf{a}_{\mathrm{r}}\left(\varphi_g\right),\ldots,\mathbf{a}_{\mathrm{r}}\left(\varphi_G\right)\right]$. Using these matrices, the channel matrix $\mathbf{H}$ in (5) can be approximated as $\mathbf{H} \cong \bar{\mathbf{A}}_{\mathrm{R}}\bar{\mathbf{H}}_{\mathrm{a}}\bar{\mathbf{A}}_{\mathrm{T}}^{H}$ where $\bar{\mathbf{H}}_{\mathrm{a}} \in \mathbb{C}^{G\times G}$ is an L-sparse matrix having L non-zero elements corresponding to AoDs and AoAs and zeros, otherwise. Unlike $\mathbf{H}_{\mathrm{a}} \in \mathbb{C}^{L\times L}$ in (5) which is diagonal, $\bar{\mathbf{H}}_{\mathrm{a}} \in \mathbb{C}^{G\times G}$ is not a diagonal matrix but a sparse matrix. To simplify notations, we ignore the error caused by the discretization of angles and rewrite (5) as

$$\mathbf{H} = \bar{\mathbf{A}}_{\mathrm{R}}\bar{\mathbf{H}}_{\mathrm{a}}\bar{\mathbf{A}}_{\mathrm{T}}^{H}. \tag{8}$$

Using (8) in (6), we have

$$\begin{aligned}\bar{\mathbf{y}} &\overset{(a)}{=} \sqrt{P}\left(\left(\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}\right)^{T}\otimes\mathbf{W}_{\mathrm{BB}}^{H}\mathbf{W}_{\mathrm{RF}}^{H}\right)\left(\bar{\mathbf{A}}_{\mathrm{T}}^{*}\otimes\bar{\mathbf{A}}_{\mathrm{R}}\right)\\ &\quad\times\mathrm{vec}\left(\bar{\mathbf{H}}_{\mathrm{a}}\right) + \bar{\mathbf{n}}\\ &\overset{(b)}{=} \sqrt{P}\left(\left(\bar{\mathbf{A}}_{\mathrm{T}}^{H}\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}\right)^{T}\otimes\mathbf{W}_{\mathrm{BB}}^{H}\mathbf{W}_{\mathrm{RF}}^{H}\bar{\mathbf{A}}_{\mathrm{R}}\right)\\ &\quad\times\mathrm{vec}\left(\bar{\mathbf{H}}_{\mathrm{a}}\right) + \bar{\mathbf{n}}\\ &= \bar{\mathbf{Q}}\cdot\mathrm{vec}\left(\bar{\mathbf{H}}_{\mathrm{a}}\right) + \bar{\mathbf{n}}, &(9)\end{aligned}$$

where the equalities (a) and (b) hold because $\mathrm{vec}\left(\mathbf{ABC}\right) = \left(\mathbf{C}^{T}\otimes\mathbf{A}\right)\cdot\mathrm{vec}\left(\mathbf{B}\right)$ and $\left(\mathbf{A}\otimes\mathbf{B}\right)\left(\mathbf{C}\otimes\mathbf{D}\right) = \mathbf{AC}\otimes\mathbf{BD}$, and $\bar{\mathbf{Q}} = \sqrt{P}\left(\left(\bar{\mathbf{A}}_{\mathrm{T}}^{H}\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}\right)^{T}\otimes\mathbf{W}_{\mathrm{BB}}^{H}\mathbf{W}_{\mathrm{RF}}^{H}\bar{\mathbf{A}}_{\mathrm{R}}\right) \in \mathbb{C}^{N_{\mathrm{T}}^{\mathrm{Beam}}N_{\mathrm{R}}^{\mathrm{Beam}}\times G^2}$. Since $\mathrm{vec}\left(\bar{\mathbf{H}}_{\mathrm{a}}\right) \in \mathbb{C}^{G^2\times 1}$ is an L-sparse vector, (9) is seen to be a sparse reconstruction problem with the sensing matrix $\bar{\mathbf{Q}}$ and can be solved by a sparse signal recovery technique, such as the OMP algorithm, which has been used for channel estimation [12], [13]. The optimization problem for compressed sensing based channel estimation can be written as

$$\begin{aligned}\mathrm{vec}\left(\bar{\mathbf{H}}_{\mathrm{a}}^{\mathrm{CS}}\right) = \arg\min_{\bar{\mathbf{H}}_{\mathrm{a}}} &\left\|\bar{\mathbf{y}} - \bar{\mathbf{Q}}\cdot\mathrm{vec}\left(\bar{\mathbf{H}}_{\mathrm{a}}\right)\right\|_2\\ &\text{subject to } \left\|\mathrm{vec}\left(\bar{\mathbf{H}}_{\mathrm{a}}\right)\right\|_0 = L, &(10)\end{aligned}$$

where $\bar{\mathbf{H}}_{\mathrm{a}}^{\mathrm{CS}}$ denotes the estimate of the sparse matrix $\bar{\mathbf{H}}_{\mathrm{a}}$ through compressed sensing, and the estimate of the desired channel, denoted as $\mathbf{H}^{\mathrm{CS}}$, is given by

$$\mathbf{H}^{\mathrm{CS}} = \bar{\mathbf{A}}_{\mathrm{R}}\bar{\mathbf{H}}_{\mathrm{a}}^{\mathrm{CS}}\bar{\mathbf{A}}_{\mathrm{T}}^{H}. \tag{11}$$

The OMP algorithm solving (10) 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 an AoD/AoA pair of the grid and is called the AoD/AoA pair index. Then, the channel gains associated with the chosen grid points are

obtained by evaluating the LS solution of $\bar{\mathbf{y}} = \bar{\mathbf{Q}}_{\Omega_t}\mathbf{h}$ in step 5, where $\bar{\mathbf{Q}}_{\Omega_t} \in \mathbb{C}^{N_{\mathrm{T}}^{\mathrm{Beam}}N_{\mathrm{T}}^{\mathrm{Beam}} \times t}$ is the sub-matrix of $\bar{\mathbf{Q}}$ that only contains the columns whose indices are included in Ω_t. 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 $\|\mathbf{r}_{t-1} - \mathbf{r}_{t-2}\|_2^2$ falls bellow the predetermined threshold δ. In step 9, the algorithm constructs the vector $\hat{\mathbf{h}}_{\mathrm{a}} \in \mathbb{C}^{G^2 \times 1}$ so that $\hat{\mathbf{h}}_{\mathrm{a}}(i) = \mathbf{h}_{t-1}(i)$ for $i \in \Omega_{t-1}$ and $\hat{\mathbf{h}}_{\mathrm{a}}(i) = 0$, otherwise. The desired estimate is given by $\bar{\mathbf{H}}_{\mathrm{a}}^{\mathrm{CS}} = \mathrm{vec}^{-1}\left(\hat{\mathbf{h}}_{\mathrm{a}}\right)$.

Algorithm 1 OMP based mmWave channel estimator

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

1: Ω_0=empty set, residual $\mathbf{r}_{-1} = \mathbf{0}$, $\mathbf{r}_0 = \bar{\mathbf{y}}$, set the iteration counter $t = 1$
2: **while** $\|\mathbf{r}_{t-1} - \mathbf{r}_{t-2}\|_2^2 > \delta$ **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: $\Omega_t = \Omega_{t-1} \cup \{j\}$ ▷ Update AoD/AoA pair set
5: $\mathbf{h}_t = \arg\min_{\mathbf{h}} \left\|\bar{\mathbf{y}} - \bar{\mathbf{Q}}_{\Omega_t}\mathbf{h}\right\|_2$ ▷ Estimate channel gains
6: $\mathbf{r}_t = \bar{\mathbf{y}} - \bar{\mathbf{Q}}_{\Omega_t}\mathbf{h}_t$ ▷ Update residual
7: $t = t + 1$
8: **end while**
9: $\hat{\mathbf{h}}_{\mathrm{a}}(i) = \mathbf{h}_{t-1}(i)$ for $i \in \Omega_{t-1}$ and $\hat{\mathbf{h}}_{\mathrm{a}}(i) = 0$ otherwise
10: **return** $\bar{\mathbf{H}}_{\mathrm{a}}^{\mathrm{CS}} = \mathrm{vec}^{-1}\left(\hat{\mathbf{h}}_{\mathrm{a}}\right)$

Since the AoDs and AoAs are generated from the continuous Laplacian distribution, increasing the number of grid points G can improve the estimation performance. However, using large G in Algorithm 1 leads to heavy computational load. To avoid this difficulty, we adopt the adaptive multi-grid (MG) approach [13], [14], which adaptively refines the grid to achieve better precision. In the MG based OMP, called the MG-OMP, the algorithm starts with a coarse grid and makes the grid fine only around the regions where the AoDs and AoAs are present. To describe the *r*-th stage of the MG-OMP algorithm we rewrite (9) as

$$\bar{\mathbf{y}} = \bar{\mathbf{Q}}_r \cdot \mathrm{vec}\left(\bar{\mathbf{H}}_{\mathrm{a},r}\right) + \bar{\mathbf{n}}, \tag{12}$$

where $\bar{\mathbf{Q}}_r$ is the sensing matrix at the *r*-th stage given by $\bar{\mathbf{Q}}_r = \sqrt{P}\left(\left(\bar{\mathbf{A}}_{\mathrm{T},r}^H \mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}\right)^T \otimes \mathbf{W}_{\mathrm{BB}}^H \mathbf{W}_{\mathrm{RF}}^H \bar{\mathbf{A}}_{\mathrm{R},r}\right) \in \mathbb{C}^{N_{\mathrm{T}}^{\mathrm{Beam}}N_{\mathrm{R}}^{\mathrm{Beam}} \times G_r^2}$, $\bar{\mathbf{A}}_{\mathrm{T},r} \in \mathbb{C}^{N_{\mathrm{T}} \times G_r}$ and $\bar{\mathbf{A}}_{\mathrm{R},r} \in \mathbb{C}^{N_{\mathrm{R}} \times G_r}$ are the array response matrices corresponding to the candidate angles in the AoD grid $\mathbf{\Psi}_{G_r}^{\mathrm{AoD}}$ and the AoA grid $\mathbf{\Psi}_{G_r}^{\mathrm{AoA}}$, respectively, of the *r*-th stage whose number of points is G_r.

The MG-OMP algorithm is the same as Algorithm 1 with exception of step 3. In the MG-OMP, this step is replaced with the following:

Initialization: Set the initial sensing matrix $\bar{\mathbf{Q}}_0 = \bar{\mathbf{Q}}$, the initial grid size $G_0 = G$, and $r = 1$. Find the coarse AoD/AoA pair index, $j = \arg\max_{i=1,\ldots,G_0^2} \left|\bar{\mathbf{Q}}_0(i)^H \mathbf{r}_{t-1}\right|$.

3-1) Obtain the indices of AoD and AoA corresponding to j by $g_{r-1} = \left\lceil \frac{j}{G_{r-1}} \right\rceil$ and $g'_{r-1} = \mathrm{mod}\,(j-1, G_{r-1}) + 1$, respectively, where $\lceil\cdot\rceil$ denotes the ceiling operator and $\mathrm{mod}\,(a,b)$ is the remainder of a when divided by b.
3-2) Set G_r to achieve the desired resolution at the *r*-th stage and get the refined grids $\mathbf{\Psi}_{G_r}^{\mathrm{AoD}} = \left\{\varphi_{g_r} \in \left[\varphi_{g_{r-1}-1}, \varphi_{g_{r-1}+1}\right] : g_r = 1, \ldots, G_r\right\}$ and $\mathbf{\Psi}_{G_r}^{\mathrm{AoA}} = \left\{\varphi_{g'_r} \in \left[\varphi_{g'_{r-1}-1}, \varphi_{g'_{r-1}+1}\right] : g'_r = 1, \ldots, G_r\right\}$.
3-3) Define the array response matrices, $\bar{\mathbf{A}}_{\mathrm{T},r} = \left[\mathbf{a}_t(\varphi_1), \ldots, \mathbf{a}_t(\varphi_{g_r}), \ldots, \mathbf{a}_t(\varphi_{G_r})\right]$ and $\bar{\mathbf{A}}_{\mathrm{R},r} = \left[\mathbf{a}_r(\varphi_1), \ldots, \mathbf{a}_r(\varphi_{g'_r}) \ldots, \mathbf{a}_r(\varphi_{G_r})\right]$.
3-4) Form the sensing matrix, $\bar{\mathbf{Q}}_r = \sqrt{P}\left(\left(\bar{\mathbf{A}}_{\mathrm{T},r}^H \mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}\right)^T \otimes \mathbf{W}_{\mathrm{BB}}^H \mathbf{W}_{\mathrm{RF}}^H \bar{\mathbf{A}}_{\mathrm{R},r}\right)$.
3-5) Find the fine AoD/AoA pair index, $j = \arg\max_{i=1,\ldots,G_r^2} \left|\bar{\mathbf{Q}}_r(i)^H \mathbf{r}_{t-1}\right|$.
3-6) Set $r = r + 1$ and return to step 3-1 until the grids reach the desired resolution.

IV. Training Beam Pattern Design

In this section, we first briefly discuss about the RF beams and then design the baseband processors following the procedure in [15].

We suggest the use of the DFT beams for RF beamforming whose transmit and receive weight vectors are given by the columns of $N_{\mathrm{T}}^{\mathrm{Beam}} \times N_{\mathrm{T}}^{\mathrm{Beam}}$ and $N_{\mathrm{R}}^{\mathrm{Beam}} \times N_{\mathrm{R}}^{\mathrm{Beam}}$ DFT matrices, respectively. In this case, it is convenient to assume that $\left\{N_{\mathrm{T}}, N_{\mathrm{R}}, N_{\mathrm{T}}^{\mathrm{Beam}}, N_{\mathrm{R}}^{\mathrm{Beam}}\right\}$ are powers of two and that the antennas are properly decimated when implementing the beams with $N_{\mathrm{T}}^{\mathrm{Beam}} < N_{\mathrm{T}}$ and $N_{\mathrm{R}}^{\mathrm{Beam}} < N_{\mathrm{R}}$. The transmit and receive beamwidths defined as $\frac{180^{\mathrm{o}}}{N_{\mathrm{T}}^{\mathrm{Beam}}}$ and $\frac{180^{\mathrm{o}}}{N_{\mathrm{R}}^{\mathrm{Beam}}}$, respectively, should be determined depending on the required beamforming gain (or SNR).

To design the baseband processors for given RF beamformers, we consider the coherence $\mu\left(\bar{\mathbf{Q}}\right)$ defined as

$$\mu\left(\bar{\mathbf{Q}}\right) \triangleq \max_{1 \le m,n \le G^2, m \ne n} \frac{\left|\bar{\mathbf{Q}}(m)^H \bar{\mathbf{Q}}(n)\right|}{\left\|\bar{\mathbf{Q}}(m)\right\|_2 \cdot \left\|\bar{\mathbf{Q}}(n)\right\|_2}. \tag{13}$$

In compressed sensing, it is known that a small $\mu\left(\bar{\mathbf{Q}}\right)$ improves the estimation performance. Thus it is desirable to design the sensing matrix $\bar{\mathbf{Q}}$ so that $\mu\left(\bar{\mathbf{Q}}\right)$ is minimized. Now due to the identity, $\mu(\mathbf{A} \otimes \mathbf{B}) = \max\{\mu(\mathbf{A}), \mu(\mathbf{B})\}$, we have $\mu\left(\bar{\mathbf{Q}}\right) = \max\left\{\mu\left(\left(\bar{\mathbf{A}}_{\mathrm{T}}^H \mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}\right)^T\right), \mu\left(\mathbf{W}_{\mathrm{BB}}^H \mathbf{W}_{\mathrm{RF}}^H \bar{\mathbf{A}}_{\mathrm{R}}\right)\right\}$, indicating that the design problem for $\bar{\mathbf{Q}}$ can be decomposed into the design of $\mathbf{F}_{\mathrm{BB}}$ and $\mathbf{W}_{\mathrm{BB}}$ minimizing $\mu\left(\left(\bar{\mathbf{A}}_{\mathrm{T}}^H \mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}\right)^T\right)$ and $\mu\left(\mathbf{W}_{\mathrm{BB}}^H \mathbf{W}_{\mathrm{RF}}^H \bar{\mathbf{A}}_{\mathrm{R}}\right)$, respectively. Next we describe the design of $\mathbf{W}_{\mathrm{BB}}$ (the process for designing $\mathbf{F}_{\mathrm{BB}}$ is similar to that for $\mathbf{W}_{\mathrm{BB}}$ and will be omitted).

Following the approach in [15], we first modify the objective function $\mu\left(\mathbf{W}_{\mathrm{BB}}^H \mathbf{W}_{\mathrm{RF}}^H \bar{\mathbf{A}}_{\mathrm{R}}\right)$ so that the objective becomes the sum of the squared inner products of all column pairs

of $\mathbf{W}_{\mathrm{BB}}^H \mathbf{W}_{\mathrm{RF}}^H \bar{\mathbf{A}}_{\mathrm{R}}$. For the block diagonal matrix $\mathbf{W}_{\mathrm{BB}} = \mathrm{diag}\left(\mathbf{W}_{\mathrm{BB},1}, \ldots, \mathbf{W}_{\mathrm{BB},i}, \ldots, \mathbf{W}_{\mathrm{BB},N_{\mathrm{R}}^{\mathrm{Block}}}\right)$ the new objective is written as

$$\sum_{m}^{G} \sum_{n,m\neq n}^{G} \left|\bar{\mathbf{W}}(m)^H \bar{\mathbf{W}}(n)\right|^2 = \sum_{i}^{N_{\mathrm{R}}^{\mathrm{Block}}} \left\|\left(\mathbf{W}_{\mathrm{BB},i}^H \mathbf{W}_{\mathrm{RF},i}^H \bar{\mathbf{A}}_{\mathrm{R}}\right)^H \mathbf{W}_{\mathrm{BB},i}^H \mathbf{W}_{\mathrm{RF},i}^H \bar{\mathbf{A}}_{\mathrm{R}} - \mathbf{I}_G\right\|_F^2, \quad (14)$$

where $\bar{\mathbf{W}} \triangleq \mathbf{W}_{\mathrm{BB},i}^H \mathbf{W}_{\mathrm{RF},i}^H \bar{\mathbf{A}}_{\mathrm{R}}$ and $\mathbf{W}_{\mathrm{RF},i} \in \mathbb{C}^{N_{\mathrm{R}} \times N_{\mathrm{RF}}}$ is the i-th sub-matrix of $\mathbf{W}_{\mathrm{RF}} = \left[\mathbf{W}_{\mathrm{RF},1}, \ldots, \mathbf{W}_{\mathrm{RF},i}, \ldots, \mathbf{W}_{\mathrm{RF},N_{\mathrm{R}}^{\mathrm{Block}}}\right]$. Thus designing $\mathbf{W}_{\mathrm{BB}}$ is decomposed into designing $\left\{\mathbf{W}_{\mathrm{BB},i} : 1, \ldots, N_{\mathrm{R}}^{\mathrm{Block}}\right\}$ by solving

$$\mathbf{W}_{\mathrm{BB},i} = \arg \min_{\mathbf{W}_{\mathrm{BB},i}} \left\|\left(\mathbf{W}_{\mathrm{BB},i}^H \mathbf{W}_{\mathrm{RF},i}^H \bar{\mathbf{A}}_{\mathrm{R}}\right)^H \times \mathbf{W}_{\mathrm{BB},i}^H \mathbf{W}_{\mathrm{RF},i}^H \bar{\mathbf{A}}_{\mathrm{R}} - \mathbf{I}_G\right\|_F^2, 1 \le i \le N_{\mathrm{R}}^{\mathrm{Blcok}}. \quad (15)$$

It was shown in [15] that the optimal solution to (15) is given by

$$\mathbf{W}_{\mathrm{BB},i} = \mathbf{U}\left(\mathbf{\Lambda}^{-1/2}\right)^H, \quad 1 \le i \le N_{\mathrm{R}}^{\mathrm{Blcok}}, \quad (16)$$

where $\mathbf{U}$ and $\mathbf{\Lambda}$ are the matrices of the eigenvectors and eigenvalues, respectively, satisfying $\mathbf{W}_{\mathrm{RF},i}^H \bar{\mathbf{A}}_{\mathrm{R}} \bar{\mathbf{A}}_{\mathrm{R}}^H \mathbf{W}_{\mathrm{RF},i} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H$.

V. Simulation Results

The performance of the proposed channel estimators is examined through computer simulation with the following parameters. The transmitter and the receiver are equipped with the uniform linear arrays with $N_{\mathrm{T}} = N_{\mathrm{R}} = 32$ and $N_{\mathrm{RF}} = 4$. They have DFT training beams with $N_{\mathrm{T}}^{\mathrm{Beam}} = N_{\mathrm{R}}^{\mathrm{Beam}} = 32$. The results in this simulation are obtained through 500 channel realizations with $\sigma_\alpha^2 = 1$ and $\sigma_{\mathrm{AS}} = 20$. At each channel realization, the number of scatterers L is determined by $L = \max\{\mathrm{P}_{10}, 1\}$ where P_{10} is the outcome of Poisson random variable with mean 10. We consider two OMP algorithms with $G \in \{60, 180\}$, called OMP1 for $G = 60$ and OMP2 for $G = 180$, and two MG-OMP algorithms having two grids (two stages) with $(G_0, G_1) = (60, 7)$ and $(60, 13)$, called the MG-OMP1 and MG-OMP2, respectively. The grid points of the OMP algorithms are uniformly distributed over $[0, \pi)$, and thus the OMP1 and OMP2 have angular resolutions of 3^{o} and 1^{o}, respectively. On the other hand, the second grid points of the MG-OMPs are distributed over an angle of 7^{o}, and thus the MG-OMP1 and MG-OMP2 have angular resolutions of 1^{o} and 0.5^{o}, respectively. From these resolutions, we expect the following: OMP1 performs the worst, MG-OMP2 performs the best, and OMP2 and MG-OMP1 exhibit comparable behaviors. The simulation results comparing the normalized mean square errors (NMSEs) will confirm these expectations. For comparison, we also consider the conventional LS algorithm based on (7).

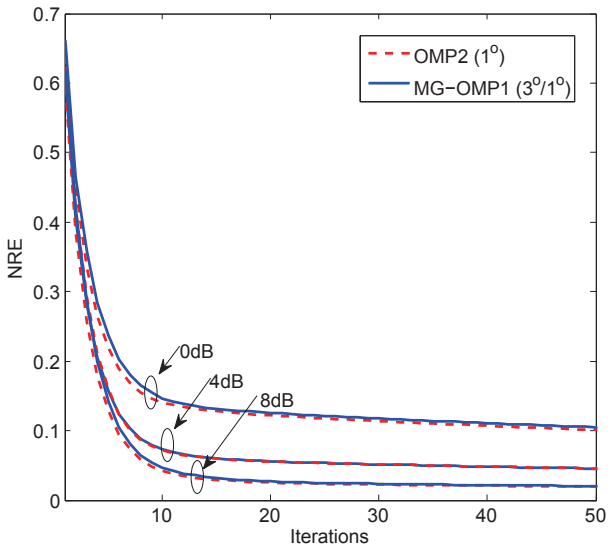

Fig. 2. The convergence characteristics of the OMP2 and MG-OMP1 algorithms in terms of NRE.

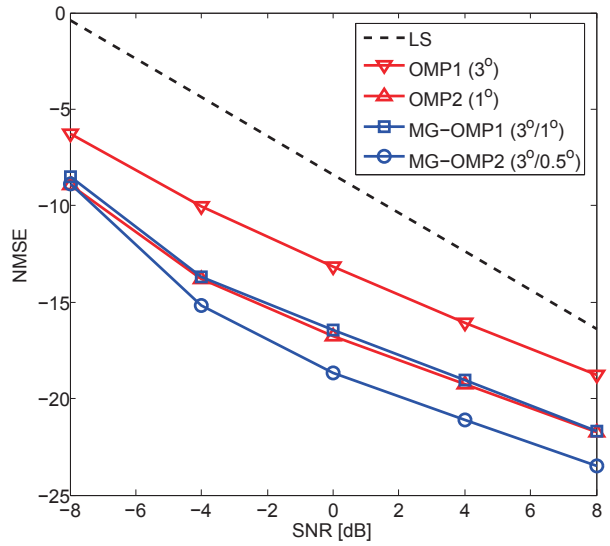

Fig. 3. NMSEs at different SNR levels (dB).

Before comparing the NMSEs, we examine the convergence characteristics of the OMP and MG-OMP algorithms by evaluating the normalized residual error (NRE) at each iteration of the OMP, defined as $\|\mathbf{r}_t\|_2^2 / \|\bar{\mathbf{y}}\|_2^2$. Fig. 2 compares the NREs of OMP2 and MG-OMP1 when the SNRs defined as P/σ_n^2 is 0dB, 4dB, and 8dB. As expected, the two algorithms exhibit almost identical behaviors and converge after about 20 iterations. In fact, after 20 iterations $\|\mathbf{r}_{t-1} - \mathbf{r}_{t-2}\|_2^2$, which is evaluated in step 2 of Algorithm 1, is less than $0.1\,\sigma_n^2$. Therefore, we set $\delta = 0.1\,\sigma_n^2$ in step 2 of Algorithm 1.

Fig. 3 compares the NMSE defined as $10\log_{10}\left(\mathbb{E}\left[\|\mathbf{H} - \mathbf{H}^{\mathrm{LS}}\|_F^2 / \|\mathbf{H}\|_F^2\right]\right)$. The LS method, whose complexity is $O\left((N_{\mathrm{T}} N_{\mathrm{R}})^2 N_{\mathrm{T}}^{\mathrm{Beam}} N_{\mathrm{R}}^{\mathrm{Beam}}\right)$, exhibits the worst performance. Among the OMP based algorithms,

as expected, the OMP1 performs the worst, the MG-OMP2 performs the best, and the OMP2 and the MG-OMP1 exhibit almost identical performance. The complexities of the OMP and MG-OMP are given by $O\left(LN_{\mathrm{T}}^{\mathrm{Beam}}N_{\mathrm{R}}^{\mathrm{Beam}}G^2\right)$ and $O\left(LN_{\mathrm{T}}^{\mathrm{Beam}}N_{\mathrm{R}}^{\mathrm{Beam}}\left({G_0}^2+{G_1}^2\right)\right)$, respectively, and the OMP based methods need less computation than the LS method. Comparing the complexities of the OMP2 and MG-OMP1 showing similar performance characteristics, the latter requires much less computation than the former. Computational savings achieved by the MG approach can be significant.

VI. Conclusion

An open-loop channel estimator for hybrid MIMO systems in mm-wave communication was proposed. By exploiting the sparse nature of mm-wave channels, a sparse signal recovery problem was formulated for channel estimation and solved by the OMP based methods. To reduce the computational load of the OMP algorithm employing a dense grid, the MG-OMP that adaptively uses dense grids only in the neighborhood of AoDs/AoAs is proposed. Given the analog training beams the baseband processor for training is designed to minimize the coherence of the sensing matrix of the sparse signal recovery problem. The simulation results demonstrate that the OMP based methods can outperform the LS method, while requiring less computation, and that the computational saving achieved by the MG-OMP can be significant. Further work in this area includes the extension of the proposed method to orthogonal frequency division multiplexing systems.

References

[1] X. Zhang, A. F. Molisch, and S. Y. Kung, "Variable-phase-shift-based RF baseband codesign for MIMO antenna selection," *IEEE Trans. Signal Process.*, vol. 53, no. 11, pp. 4091-4103, Nov. 2005.

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

[3] O. El Ayach, R. W. Heath, Jr., S. Abu-Surra, S. Rajagopal, and Z. Pi, "Low complexity precoding for large millimeter wave MIMO systems," *in Proc. IEEE Int. Conf. Commun.*, June 2012.

[4] 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.

[5] IEEE 802.15 WPAN Millimeter Wave Alternative PHY Task Group 3c. [Online]. Available: www.ieee802.org/15/pub/TG3c.html, Sep. 2011.

[6] IEEE P80211ad, Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications - Amendment 3: Enhancements for Very High Throughput in the 60GHz band, Dec. 2012.

[7] J. Wang, Z. Lan, C. Pyo, T. Baykas, C. Sum, M. Rahman, J. Gao, R. Funada, F. Kojima, H. Harada 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.

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

[9] 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.*, June 2014.

[10] 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.

[11] 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.*, 2014.

[12] G. Taubock, F. Hlawatsch, D. Eiwen, and H. Rauhut, "Compressive estimation of doubly selective channels in multicarrier systems: Leakage effects and sparsity-enhancing processing," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 2, pp. 255-271, Apr. 2010.

[13] 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.

[14] D. Malioutov, M. Cetin, and A. S. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 3010-3022, Aug. 2005.

[15] L. Zelnik-Manor, K. Rosenblum, and Y. Eldar, "Sensing matrix optimization for block-sparse decoding," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4300-4312, Sept. 2011.