A Scalable and Statistically Robust Beam Alignment Technique for mm-Wave Systems

Xiaoshen Song, Saeid Haghighatshoar, *Member, IEEE,* Giuseppe Caire, *Fellow, IEEE*

***Abstract*—Millimeter-Wave (mm-Wave) band provides orders of magnitude higher bandwidth compared with the traditional sub-6 GHz band. Communication at mm-Waves is, however, quite challenging due to the severe path loss. To cope with this problem, a directional beamforming both at the Base Station (BS) side and the user side is necessary to find a strong path connecting the BS and the user. Finding such beamforming directions is referred to as the *Beam Alignment* (BA) and is known to be a challenging problem. In this paper, we propose a new BA scheme that finds a strong path connecting the BS and the user by estimating the second order statistics of the channel. As a result, our proposed algorithm is highly robust to variations of the channel statistics compared with the other works in the literature. In our proposed scheme, the BS probes the channel in the *Downlink* (DL) letting each user to estimate its own channel (i.e., a strong path connecting the user to the BS), where all the users within the BS coverage are trained simultaneously. Thus, the complexity of our proposed BA (channel estimation) is independent of the number of users in the system. We pose the channel estimation at the user side as a *Compressed Sensing* (CS) of a non-negative signal and use the recently developed *Non-Negative Least Squares* (NNLS) technique to solve it efficiently. We use numerical simulation to assess the performance of our proposed algorithm and compare it with the performance of other competitive methods in the literature. The results illustrate that our approach incurs less training overhead, exhibits higher efficiency in multi-user scenarios, and is highly robust to variations in channel statistics.**

***Index Terms*—Millimeter-Wave, Beam Alignment, Compressed Sensing, Non-Negative Least Squares (NNLS).**

I. Introduction

Millimeter-wave (mm-Wave) provides an opportunity to fulfill the demand for high data rates in the next generation communication networks because of the huge bandwidth available at high frequency bands [1]. A critical challenge to signaling at mm-Waves compared with sub-6 GHz spectrum is the severe propagation loss and poor penetration into buildings [2]. Therefore, large antenna arrays yielding a high-gain directional beamforming at both the *Base Station* (BS) side and the user side are necessary to boost the *Signal-to-Noise Ratio* (SNR) to sufficiently high levels, such that small-cell outdoor communication is possible. Moreover, due to the large propagation loss at mm-Waves the communication between the BS and the user occurs only via a very sparse collection of scatterers in the angle domain [3–6]. This implies that to establish a reliable communication, the BS and the user need to focus their beam pattern in the direction of a strong path. For example, in the case of *Line-of-Sight* (LoS) propagation, the beams typically point at each other since the LoS path is typically the strongest one. More in general, we refer to the problem of finding a narrow beam direction at both the BS and the user side yielding an SNR after beamforming above a desired threshold as the *Beam Alignment* (BA) problem. This problem is quite well studied in the literature [3, 4, 6–15]. In particular, it is known to be a challenging problem since in mm-Waves the SNR before beamforming is typically very low, especially in outdoor non-LoS conditions. Moreover, systems operating at mm-Waves have large arrays with many antennas but typically only a limited number of *Radio Frequency* (RF) chains [7]. This implies that, at each time slot, only a limited span of *Angle-of-Departures* (AoDs) and *Angle-of-Arrivals* (AoAs) can be simultaneously probed to search for the location of strong scatterers. In particular, a naive exhaustive search over the whole AoA-AoDs is very time consuming and unfeasible in practice.

This has motivated the development of efficient BA algorithms based on hierarchical adaptive search, interactive search, and Compressed Sensing (CS) techniques [8–15]. The fundamental idea of hierarchical methods is to use wider beam patterns at the start and to refine them in several consecutive stages. In [11], for example, the authors develop a bisection algorithm in which the range of AoDs and AoAs are divided by a factor of 2 at each step and is refined by probing the resulting 2×2 sections and identifying the section with the maximum received power. A similar idea via using overlapping beam patterns is used in [12]. Such hierarchical techniques, however, require the interaction of a BS with a single user at each stage. Therefore, it is not obvious how to extend/adapt these approaches to a multiuser scenario, where a BS has to train the beams of many users. In [13], a novel method for BA is proposed where a BS and a user iteratively and collaboratively identify the dominant eigen-vector of their channel matrix via the well-known Power method, which gives an estimate of the location of strong scatterers. This requires, however, having many RF chains at the BS and the user and is difficult to implement in practical scenarios. In recent works, considering the natural channel sparsity in the AoA-AoD domain [3–6], CS-based algorithms have been widely used for BA in mm-Waves [14–18]. These algorithms are efficient and particularly attractive for multi-user scenarios, but they are typically based on the assumption that the instantaneous channel remains invariant during the whole probing/measurement stage (the same assumption is also adopted in [11, 12]), which is difficult to fulfill due to the large carrier frequency and Doppler spread, and fast channel

The authors are with the Communications and Information Theory Group, Technische Universität Berlin (xiaoshensong@yahoo.com, {saeid.haghighatshoar, caire}@tu-berlin.de).

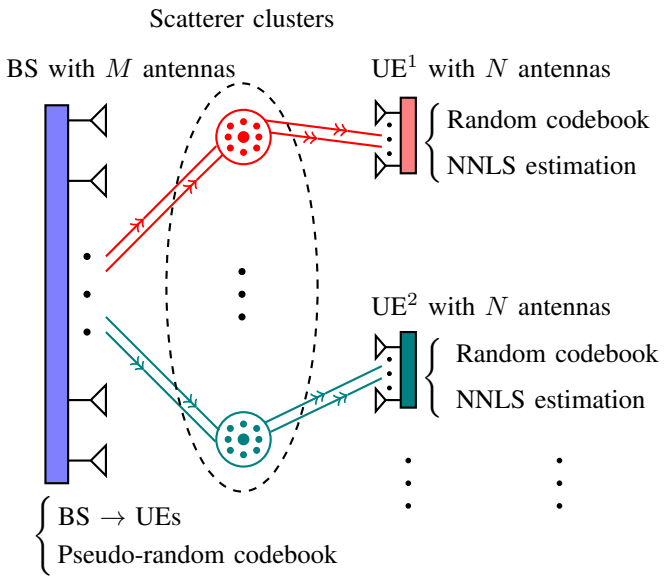

Fig. 1: Illustration of the physical channel model and our proposed *Beam Alignment* (BA) scheme.

variation in mm-Waves [19, 20]. Moreover, a naive application of the conventional CS techniques typically results in a wide spread of the power of the transmitted signal during the probing stage in the angle domain and diminishes the SNR of the resulting measurements. This is not problematic in sub-6 GHz but seems to be a big issue in mm-Waves due to the very low-SNR nature of the channel, and is widely overlooked in the literature.

A. Contributions

In this paper, we propose a novel BA scheme that has the following advantages compared with the existing works in the literature:

1) *System-level Scalability:* In our proposed approach, during the channel estimation, only the BS probes (via a beamforming codebook) the channel and all the users are in the listening mode. In particular, the BS transmits simultaneously to all the user a sufficient number of measurements to let them estimate their channel (i.e., AoA-AoD of a strong scatterer connecting them to the BS) reliably. Since there is no need for interaction between the BS and each user, the channel estimation is highly scalable and its complexity does not grow with the number of active users in the system.

2) *User-specific codebook:* In our scheme, although the BS probing codebook at the transmitter side is common to all the users, each user can still use its own probing codebook at the receiver side. We propose a technique in which each user selects its codebook based on its number of RF chains and its channel quality. In brief, when a user is close to the BS and has a sufficient SNR before beamforming, it uses wide beams as motivated by the conventional CS to speed up the channel estimation by taking less measurements (*exploration*) whereas a user far from the BS with a very low SNR before beamforming applies narrower probing beams to obtain measurements with a sufficiently large SNR (*exploitation*). In particular, we shall see that for a specific SNR level before beamforming, there is an optimal beam spreading factor that results in the fastest channel acquisition.

3) *Robustness to Variations in Channel Statistics:* As explained before, most of the existing works in the literature use the assumption that the instantaneous channel coefficients remain invariant during the whole BA phase. This is difficult to meet in mm-Waves due to the large carrier frequency, large Doppler spread, and fast channel variations [19, 20]. In this paper, we develop a novel technique that uses the second order statistics of the channel and is highly robust to variations in channel statistics. We also illustrate via numerical simulations that CS-based algorithms fail to estimate the channel when the channel is not invariant, whereas our scheme performs quite well in all ranges of channel variations.

4) *Low-complexity Channel Estimation:* In our scheme, each user needs to estimate the channel from its received measurements, thus, all the computation necessary for channel estimation is done at the user side. We show that the resulting channel estimation boils down to a *Non-Negative Least Squares* (NNLS) problem, which can be solved with an affordable complexity via standard techniques in the literature.

B. Notation

We denote vectors by boldface small (e.g., $\mathbf{a}$) and matrices by boldface capital (e.g., $\mathbf{A}$) letters. Scalars are denoted by non-boldface letters (e.g., a, A). We represent sets by calligraphic letter $\mathcal{A}$ and their cardinality with $|\mathcal{A}|$. We denote the empty set by $\emptyset$. We use $\mathbb{E}$ for the expectation, $\otimes$ for the Kronecker product of two matrices, and $\star$ for the convolution operator. We use $\mathbf{A}^{\mathsf{T}}$ for transpose and $\mathbf{A}^{\mathsf{H}}$ for conjugate transpose of a matrix $\mathbf{A}$. The output of an optimization problem such as $\arg\min_{x\in\mathcal{X}} f(x)$ is denoted by x^*. The complex circularly symmetric Gaussian distribution with a mean μ and a variance γ is denoted by $\mathcal{CN}(\mu,\gamma)$. For an integer $k \in \mathbb{Z}$, we use the shorthand notation $[k]$ for the set of non-negative integers $\{1,...,k\}$.

II. Basic Setup

A. Channel Model

We consider a mm-Wave system with a BS with *Uniform Linear Array* (ULA) with M antennas and m RF chains where typically $m \ll M$. We consider a generic user denoted by UE and assume that it has also a ULA with N antennas and $n \ll N$ RF chains. We assume that both BS and UE arrays have the antenna spacing $d = \frac{\lambda}{2}$, where λ is the wavelength given by $\lambda = \frac{c_0}{f_0}$, where c_0 is the speed of the light and where f_0 is the carrier frequency. We denote by $\theta, \phi \in [-\frac{\pi}{2}, \frac{\pi}{2}]$ the steering angles with respect to the BS and UE arrays. We represent the array response of the BS and UE arrays to a planar wave coming from the angles θ and ϕ with respect to the BS and UE with the M-dim and N-dim array vectors $\mathbf{a}(\theta) \in \mathbb{C}^M$ and $\mathbf{b}(\phi) \in \mathbb{C}^N$ respectively, where

$$[\mathbf{a}(\theta)]_k = e^{j(k-1)\pi\sin(\theta)}, k \in [M], \quad (1)$$

$$[\mathbf{b}(\phi)]_l = e^{j(l-1)\pi\sin(\phi)}, \ l \in [N]. \quad (2)$$

We assume that the communication between the BS and the UE occurs via a collection of sparse *multi-path components*

(MPCs) in the AoA-AoD and delay domain [1], where the $M \times N$ low-pass equivalent impulse response of the channel at a time slot s is given by

$$\mathsf{H}_s(\tau) = \sum_{l=1}^{L} \rho_{s,l} \mathbf{a}(\theta_l)\mathbf{b}(\phi_l)^{\mathsf{H}} \delta(\tau - \tau_l), \tag{3}$$

where $\rho_{s,l}$ is the random channel gain of a MPC at AoA-AoD-delay $(\theta_l, \phi_l, \tau_l)$, $l \in [L]$, where typically $L \ll \max\{M, N\}$ [21]. Note that in (3), and also in (1) and (2), we made the implicit assumption that the communication bandwidth denoted by B is much smaller than the carrier frequency f_0 (i.e., $B \ll f_0$) such that the array response is approximately frequency-invariant. We adopt a block fading model, where the channel gains $\rho_{s,l}$, $l \in [L]$, remain invariant over the *coherence time* of the channel of duration Δt_c but change i.i.d. randomly across different coherence times. We also assume that each MPC is a superposition of several smaller components such that the channel gains $\rho_{s,l} \sim \mathcal{CN}(0, \gamma_l)$ have a zero-mean complex Gaussian distribution. The channel model (3) can be extended to the case where there is a continuum of MPCs connecting the BS and the UE, where the channel model is given by

$$\mathsf{H}_s(d\tau) = \rho_s(d\theta, d\phi, d\tau)\mathbf{a}(\theta)\mathbf{b}(\phi)^{\mathsf{H}}, \tag{4}$$

where $\rho_s(d\theta, d\phi, d\tau)$ denotes the angle-delay random impulse response of the channel. Our result in this paper holds for the general case (4) but for simplicity, we will use the discrete channel model in the sequel. We also assume that the AoA-AoDs of the scatterers $\{(\theta_l, \phi_l)\}_{l=1}^{L}$ change in a time-scale much longer that the channel coherence time Δt_c such that they can be treated as constant during the whole communication across many coherence blocks.

B. Signaling Model

Assume that the BS is going to communicate with a generic UE. Since the BS has m RF chains, it can transmit up to m different data streams as follows. Let $x_{s,i}(t)$, $t \in [0, t_0]$, be the continuous-time low-pass equivalent signal corresponding to the i-th data stream, where t_0 denotes the duration of a slot. We assume that the channel gains remain invariant over the channel *coherence time*, but can change i.i.d. randomly across different coherence times. Moreover, each coherence time is divided into several slots of duration t_0, where $t_0 \ll \Delta t_c$, thus, the channel gains remain invariant over each slot. To transmit the i-th data stream, the BS applies a beamforming vector $\mathbf{u}_{s,i} \in \mathbb{C}^M$, where the corresponding vector-valued signal is given by $\mathbf{x}_{s,i}(t) := x_{s,i}(t)\mathbf{u}_{s,i}$ for $t \in [0, t_0]$. We always assume that the beamforming vectors are normalized such that $\|\mathbf{u}_{s,i}\| = 1$[1]. The transmitted signal at time slot s is given by

$$\mathbf{x}_s(t) = \sum_{i=1}^{m} \mathbf{x}_{s,i}(t) = \sum_{i=1}^{m} x_{s,i}(t)\mathbf{u}_{s,i}. \tag{5}$$

[1]Also, note that here we are assuming that each beamforming vector $\mathbf{u}_{s,i}$, $i \in [m]$, is implemented in the RF domain via an analog beamforming network and it is frequency flat, i.e., it is the same across the whole bandwidth (e.g., along all subcarriers in an OFDM symbol).

The received low-pass equivalent signal at the UE array is

$$\mathbf{r}_s(t) = \mathsf{H}(t)^{\mathsf{H}} \star \mathbf{x}_s(t) = \int \mathsf{H}_s(d\tau)^{\mathsf{H}} \mathbf{x}_s(t - \tau) \tag{6}$$

$$= \sum_{l=1}^{L} \rho_{s,l} \mathbf{b}(\phi_l)\mathbf{a}(\theta_l)^{\mathsf{H}} \mathbf{x}_s(t - \tau_l) \tag{7}$$

$$= \sum_{l=1}^{L} \sum_{i=1}^{m} \rho_{s,l} x_{s,i}(t - \tau_l)\mathbf{b}(\phi_l)\mathbf{a}(\theta_l)^{\mathsf{H}} \mathbf{u}_{s,i} \tag{8}$$

$$= \sum_{l=1}^{L} \sum_{i=1}^{m} g_{s,l,i}^{\text{BS}} \rho_{s,l} x_{s,i}(t - \tau_l)\mathbf{b}(\phi_l) \tag{9}$$

where $g_{s,l,i}^{\text{BS}} := \mathbf{a}(\theta_l)^{\mathsf{H}} \mathbf{u}_{s,i}$ denotes the beamforming gain along the l-th MPC at the BS side for the i-th RF chain. As stated before, we assume that the UE is also equipped with n RF chains. The noisy received signal at the output of the j-th RF chain at the UE side is given by

$$\begin{aligned} y_{s,j}(t) &= \mathbf{v}_{s,j}^{\mathsf{H}} \mathbf{r}_s(t) + z_{s,j}(t) \\ &= \sum_{l=1}^{L} \sum_{i=1}^{m} g_{s,l,i}^{\text{BS}} \rho_{s,l} x_{s,i}(t - \tau_l) \mathbf{v}_{s,j}^{\mathsf{H}} \mathbf{b}(\phi_l) + z_{s,j}(t) \\ &= \sum_{l=1}^{L} \sum_{i=1}^{m} g_{s,l,i}^{\text{BS}} g_{s,l,j}^{\text{UE}} \rho_{s,l} x_{s,i}(t - \tau_l) + z_{s,j}(t) \qquad (10) \\ &=: \sum_{i=1}^{m} r_{s,i,j}(t) + z_{s,j}(t) \qquad (11) \end{aligned}$$

where $\mathbf{v}_{s,j} \in \mathbb{C}^N$ denotes the normalized beamforming vector of the j-th RF chain at the UE side, where $g_{s,l,j}^{\text{UE}} = \mathbf{v}_{s,j}^{\mathsf{H}} \mathbf{b}(\phi_l)$ denotes the array gain of the j-th RF chain along the l-th MPC, where $r_{s,i,j}(t) := \sum_{l=1}^{L} g_{s,l,i}^{\text{BS}} g_{s,l,j}^{\text{UE}} \rho_{s,l} x_{s,i}(t - \tau_l)$ denotes a fraction of the i-the data stream received at the j-th RF chain at the UE side, and where $z_{s,j}(t)$ is the continuous-time complex *Additive White Gaussian Noise* (AWGN) at the output of the j-th RF chain at the UE side with a one-sided *Power Spectral Density* (PSD) of N_0 Watt/Hz.

C. Beam Alignment

We assume that the signal corresponding to different data streams $x_{s,i}(t)$ are orthogonal, i.e.,

$$\langle x_{s,i}, x_{s,i'} \rangle := \frac{1}{t_0} \int_0^{t_0} x_{s,i}(t)^* x_{s,i'}(t) dt = 0, \ i \neq i', \tag{12}$$

thus, separable at the UE side. This holds, for example, when different data streams occupy disjoint frequency bands. We define the *Signal-to-Noise Ratio* (SNR) after beamforming (ABF) for the i-th data stream received at the j-th RF chain at the UE by

$$\begin{aligned} \mathsf{SNR}_{\text{ABF}}^{\text{CO } i,j} &:= \frac{\frac{1}{t_0}\mathbb{E}[\int |r_{s,i,j}(t)|^2 dt]}{N_0 B_i} \\ &= \frac{\frac{E_{s,i}}{t_0} \sum_{l=1}^{L} \gamma_l |g_{s,l,j}^{\text{UE}}|^2 |g_{s,l,i}^{\text{BS}}|^2}{N_0 B_i} \\ &= \frac{P_{s,i} \sum_{l=1}^{L} \gamma_l |g_{s,l,j}^{\text{UE}}|^2 |g_{s,l,i}^{\text{BS}}|^2}{N_0 B_i}, \end{aligned} \tag{13}$$

where B_i denotes the bandwidth of $x_{s,i}(t)$, where $E_{s,i} := \int_0^{t_0} |x_{s,i}(t)|^2 dt$ and $P_{s,i} = \frac{E_{s,i}}{t_0}$ denote the energy and the

average power of $x_{s,i}(t)$. We have $\sum_{i\in[m]} P_{s,i} = P_{\mathtt{tot}}$, where $P_{\mathtt{tot}}$ is the overall transmit power at the BS, which is distributed over m data streams. It is worthwhile to mention that the appropriate definition of $\mathsf{SNR}^{\mathrm{CO}}_{\mathrm{ABF}}$ depends on the application and on the number of independent data streams communicated between the BS and the UE. In most relevant cases, the UE and the BS communicate over a single data stream, where

$$\mathsf{SNR}^{\mathrm{CO}}_{\mathrm{ABF}} = \frac{P_{\mathtt{tot}} \sum_{l=1}^{L} \gamma_l |g^{\mathrm{UE}}_{s,l}|^2 |g^{\mathrm{BS}}_{s,l}|^2}{N_0 B'}, \tag{14}$$

where $g^{\mathrm{UE}}_{s,l}$ and $g^{\mathrm{BS}}_{s,l}$ are the array gains at the UE and BS via beamforming vectors $\mathbf{v}_s$ and $\mathbf{u}_s$ along the l-th MPC, and where $B' < B$ is a part of bandwidth B devoted to the data stream. We define the SNR before beamforming (BBF) by

$$\mathsf{SNR}^{\mathrm{CO}}_{\mathrm{BBF}} := \frac{P_{\mathtt{tot}} \sum_{l=1}^{L} \gamma_l}{N_0 B'}. \tag{15}$$

This is the SNR obtained when the signal $x_s(t)$ is transmitted through a single BS antenna and is received in a single UE antenna and can be simply obtained by setting $\mathbf{u}_s = (1, 0, \ldots, 0) \in \mathbb{C}^M$ and $\mathbf{v}_s = (1, 0, \ldots, 0) \in \mathbb{C}^N$ in (14). One of the challenges of communicating over mm-Waves is that the SNR before beamforming $\mathsf{SNR}^{\mathrm{CO}}_{\mathrm{BBF}}$ in (15) is very low. In theory, one can always improve the SNR by increasing the transmit power $P_{\mathtt{tot}}$ (loss in power) but this is not always feasible due to physical device constraints, or standard regulations on the maximum allowed transmit power. Another option is to communicate over a smaller bandwidth $B' \ll B$ to reduce the noise power at the cost of reducing the rate (loss in multiplexing gain and degree-of-freedom). A third alternative which seems viable is to use arrays with large number of antennas at the BS and UE. In such a setting, via suitable beam alignment one can hope to find *good* beamforming vectors $\mathbf{u}_s$ and $\mathbf{v}_s$ at the BS and the UE respectively to boost $\mathsf{SNR}^{\mathrm{CO}}_{\mathrm{BBF}}$ by a factor M at the BS and a factor N at the UE by beamforming along the AoA-AoD of strong MPCs in the channel.

D. Channel Model under OFDM Signaling

In this paper, we adopt an OFDM signaling between the BS and the UEs. For simplicity, we again consider a fixed UE and denote by $x_{s,i}(t)$ the continuous-time waveform corresponding to the i-th data stream transmitted from the BS to this UE at time slot s. In OFDM signaling, each $x_{s,i}(t)$ corresponds to an OFDM symbol and has a duration of t_0, to which we append a *Cyclic Prefix* (CP) of duration $\Delta\tau_c$, where $\Delta\tau_c$ denotes the effective delay spread of the channel. Thus, the frequency separation between the subcarriers is given by $\Delta f = \frac{1}{t_0}$ and the total number of subcarriers is given by $F = Bt_0$, where B denotes the whole communication bandwidth as before. For simplicity, and due to the linearity of the channel, we focus on a communication between the i-th RF chain at the BS and j-th RF chain at the UE, where the received signal is given by $r_{s,i,j}(t) = \sum_{l=1}^{L} g^{\mathrm{BS}}_{s,l,i} g^{\mathrm{UE}}_{s,l,j} \rho_{s,l} x_{s,i}(t-\tau_l)$ as in (10), which corresponds to the output of a channel with a time-variant impulse response $h_s(\tau) = \sum_{l=1}^{L} g^{\mathrm{BS}}_{s,l,i} g^{\mathrm{UE}}_{s,l,j} \rho_{s,l} \delta(\tau - \tau_l)$ for the input $x_{s,i}(t)$. After removing the CP, this is represented as

$$\check{h}_s(f) = \sum_{l=1}^{L} g^{\mathrm{BS}}_{s,l,i} g^{\mathrm{UE}}_{s,l,j} \rho_{s,l} e^{-j2\pi f \tau_l}. \tag{16}$$

We define the matrix-valued frequency response by

$$\check{\mathsf{H}}_s(f) = \sum_{l=1}^{L} \rho_{s,l} \mathbf{b}(\phi_l) \mathbf{a}(\theta_l)^{\mathsf{H}} e^{-j2\pi f \tau_l}, \tag{17}$$

which corresponds to the Fourier transform of the impulse response of the channel $\mathsf{H}(\tau)$ in (3). Note that $\check{h}_s(f) = \mathbf{u}_s^{\mathsf{H}} \check{\mathsf{H}}_s(f) \mathbf{v}_s$ is obtained by projecting the channel response $\check{\mathsf{H}}_s(f)$ along the beamforming vectors $\mathbf{u}_s$ and $\mathbf{v}_s$ at the BS and the UE respectively. In OFDM, the frequency response $\check{h}_s(f)$ or more generally $\check{\mathsf{H}}_s(f)$ are sampled at F discrete frequencies $f_\omega = \frac{\omega}{t_0}$, $\omega \in [F]$, corresponding to F subcarriers, where we denote the resulting channel matrix by

$$\mathbf{H}_s[\omega] := \check{\mathsf{H}}_s(f_\omega) = \sum_{l=1}^{L} \rho_{s,l} \mathbf{b}(\phi_l) \mathbf{a}(\theta_l)^{\mathsf{H}} e^{-j2\pi \frac{\omega}{t_0} \tau_l} \tag{18}$$

E. Quantized Channel Model

The AoA-AoD (θ_l, ϕ_l) and the delay parameters τ_l in (17) take continuous values. For later applications in the paper, we need a finite-dim representation of the channel. We obtain such a representation by quantizing the matrix-valued channel response (17) in a discrete quantized dictionary in the AoA-AoD domain. We consider the discrete set of AoA-AoDs

$$\Theta := \{\check{\theta} : (1 + \sin(\check{\theta}))/2 = \frac{k-1}{M}, k \in [M]\}, \tag{19}$$

$$\Phi := \{\check{\phi} : (1 + \sin(\check{\phi}))/2 = \frac{k'-1}{N}, k' \in [N]\}, \tag{20}$$

and use the corresponding array responses $\mathcal{A} := \{\mathbf{a}(\check{\theta}) : \check{\theta} \in \Theta\}$ and $\mathcal{B} := \{\mathbf{b}(\check{\phi}) : \check{\phi} \in \Phi\}$ as a discrete dictionary to quantize the channel response. For the ULAs considered in this paper, the dictionary $\mathcal{A}$ and $\mathcal{B}$ after suitable normalization correspond to the $M \times M$ and $N \times N$ DFT matrices $\mathbf{F}_M$ and $\mathbf{F}_N$ [5], where

$$[\mathbf{F}_M]_{k,k'} = \frac{1}{\sqrt{M}} e^{j2\pi(k-1)(\frac{k'-1}{M} - \frac{1}{2})}, k, k' \in [M] \tag{21}$$

$$[\mathbf{F}_N]_{k,k'} = \frac{1}{\sqrt{N}} e^{j2\pi(k-1)(\frac{k'-1}{N} - \frac{1}{2})}, k, k' \in [N]. \tag{22}$$

Using the discrete dictionaries $\mathbf{F}_M$ and $\mathbf{F}_N$, we quantize the channel matrix in (18) as $\mathbf{H}_s[\omega] = \mathbf{F}_M \check{\mathbf{H}}_s[\omega] \mathbf{F}_N^{\mathsf{H}}$, where

$$\check{\mathbf{H}}_s[\omega] = \mathbf{F}_M^{\mathsf{H}} \mathbf{H}_s[\omega] \mathbf{F}_N \tag{23}$$

$$= \sum_{l=1}^{L} \rho_{s,l} e^{-j2\pi \frac{\omega}{t_0} \tau_l} \left(\mathbf{F}_M^{\mathsf{H}} \mathbf{a}(\theta_l)\right) \left(\mathbf{F}_N^{\mathsf{H}} \mathbf{b}(\phi_l)\right)^{\mathsf{H}} \tag{24}$$

$$= \sum_{l=1}^{L} \rho_{s,l} e^{-j2\pi \frac{\omega}{t_0} \tau_l} \check{\mathbf{a}}(\theta_l) \check{\mathbf{b}}(\phi_l)^{\mathsf{H}}, \tag{25}$$

where $\check{\mathbf{a}}(\theta) = \mathbf{F}_M^{\mathsf{H}} \mathbf{a}(\theta)$, and $\check{\mathbf{b}}(\phi) = \mathbf{F}_N^{\mathsf{H}} \mathbf{b}(\phi)$ denote the projection of the array responses at the discrete dictionaries

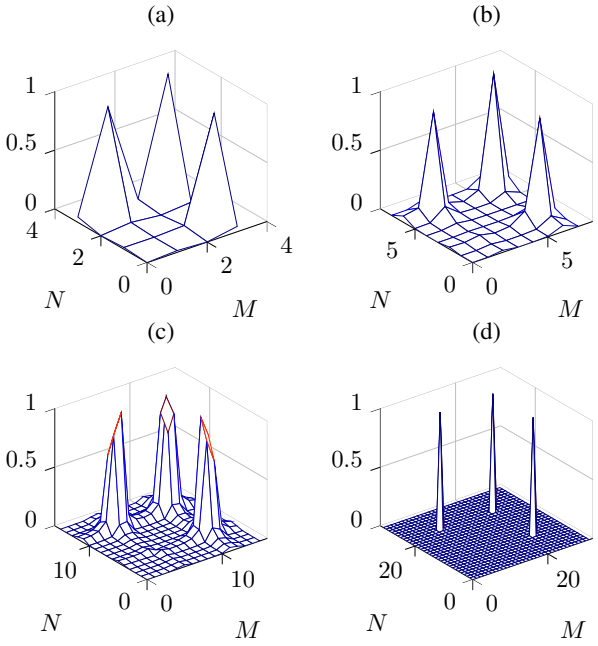

Fig. 2: Illustration of the sparsity of the channel matrix $\check{\mathbf{H}}_s[\omega]$ at an arbitrary subcarrier ω consisting of 3 off-grid AoA-AoDs with increasing number of antennas for $M = N = 4$ (a), $M = N = 8$ (b), $M = N = 16$ (c), $M = N = 32$ (d).

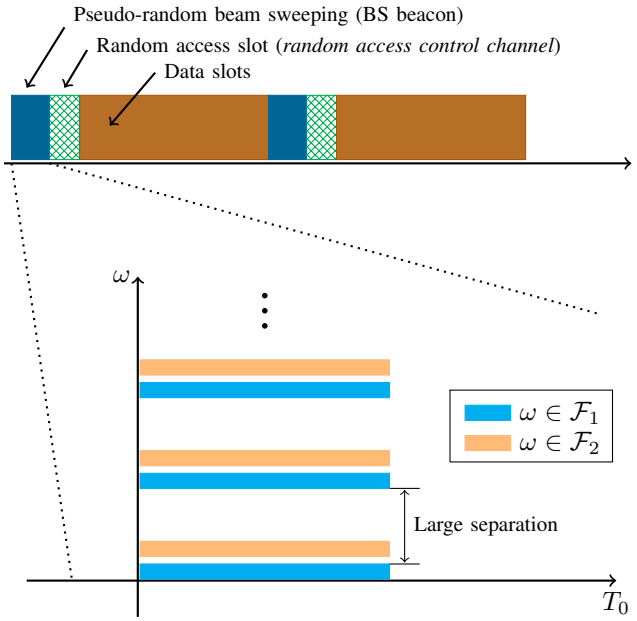

Fig. 3: Frame structure of the proposed *Beam Alignment* (BA) scheme at the BS side, where the beam training is done during the initial sweeping slots. The channel is always probed in the *Downlink* (DL), after which, the user feeds back its estimated AoA-AoD information to the BS during the *random access slot*. Having received the feedback message, the BS sends an *Acknowledgment* (ACK) packet to the user followed by a data transmission stage. Different data streams are assigned to disjoint sets of subcarriers $\omega \in \mathcal{F}_i$, $i \in [m]$. The beam sweeping phase consists of T_0 time slots.

$\mathcal{A}$ and $\mathcal{B}$ respectively. From the expressions of $\mathbf{F}_M$ and $\mathbf{a}(\theta_l)$, the m'-th entry of $\check{\mathbf{a}}(\theta_l)$ is given by

$$[\,\check{\mathbf{a}}(\theta_l)\,]_{m'} = \frac{1}{\sqrt{M}} \sum_{i=0}^{M-1} e^{-j2\pi i(\frac{m'-1}{M}-\frac{1}{2})} e^{j\pi i \sin(\theta_l)} \tag{26}$$

$$= \frac{1}{\sqrt{M}} \frac{1-e^{-j\pi(\frac{2(m'-1)}{M}-\sin(\theta_l)-1)M}}{1-e^{-j\pi(\frac{2(m'-1)}{M}-\sin(\theta_l)-1)}} \tag{27}$$

$$= \frac{1}{\sqrt{M}} \frac{e^{j\pi\psi_l M}-e^{-j\pi\psi_l M}}{e^{j\pi\psi_l}-e^{-j\pi\psi_l}} e^{-j\pi\psi_l(M-1)} \tag{28}$$

$$= \frac{1}{\sqrt{M}} \frac{\sin(\pi\psi_l M)}{\sin(\pi\psi_l)} e^{-j\pi\psi_l(M-1)}, \tag{29}$$

where $\psi_l = \frac{m'-1}{M} - \frac{1}{2}\sin(\theta_l) - \frac{1}{2}$. A similar expression holds for $\check{\mathbf{b}}(\phi_l)$. It is seen from (29) that

$$|[\,\check{\mathbf{a}}(\theta_l)\,]_{m'}| = \frac{1}{\sqrt{M}} \frac{|\sin(\pi\psi_l M)|}{|\sin(\pi\psi_l)|}, \tag{30}$$

is a localized kernel around $\theta_l = \sin^{-1}[\frac{2(m'-1)}{M} - 1]$ with a resolution of $\frac{1}{M}$. In general, the AoA-AoDs of the MPCs are not aligned with the discrete grid $\mathcal{G} = \Theta \times \Phi$. However, as the number of antennas M at the BS and N at the UE increases, the discrete dictionaries $\mathbf{F}_M$ and $\mathbf{F}_N$ provide a *good* quantizer of the underlying channel giving a sparse approximation of the channel in the AoA-AoD domain. This is illustrated for a sparse channel with $L = 3$ off-grid AoA-AoD components in Fig. 2. It is seen that for a small M and N, the resulting representation $\check{\mathbf{H}}_s[\omega]$ is far from sparse, but becomes highly sparse as M and N increase.

III. Proposed Beam-Alignment Algorithm

A. Basic Setup

In this section, we describe our proposed scheme briefly. In our scheme, the channel is always probed in the *Downlink* (DL) and all the UEs stay in the listening mode. The BS broadcasts its pseudo-random beam patterns periodically over periods of T time slots, where a standard BS beacon (pilot signal) [11–18] is transmitted per subcarrier per time slot. This is illustrated in Fig. 3. The pseudo-random *transmit* codebook of the BS is shared among all the users in the system. However, each user can adopt its own *receive* codebook. The users gather channel measurements during the training until they obtain enough measurements to be able to estimate their channel. When a user finally estimates its channel, i.e., a strong AoA-AoD direction connecting it to the BS (details are given in the following sections), it feeds back this information in a *random access slot*, where the BS is in the listening mode. During a random access slot, each user transmits its signal by beamforming along its AoA-AoD estimate (user-side beamforming), where we assume that the SNR with one-sided beamforming gain at the user side and coarse beams at the BS side is sufficient for the feedback signal to be correctly received/decoded at the BS when there are no collisions between the users. At this stage, the BS knows which beam pattern to apply to each user, and will respond to the user on

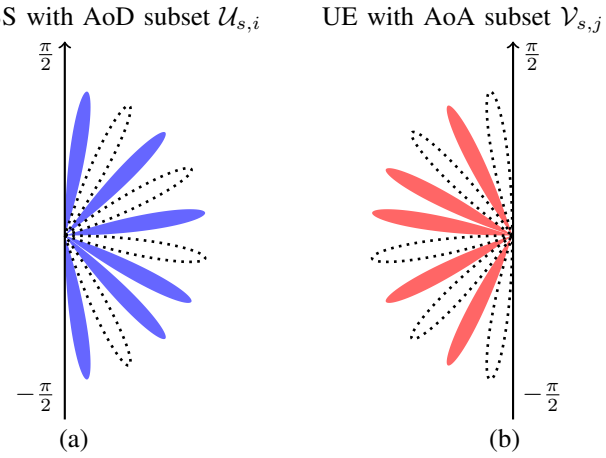

Fig. 4: An example of the probing beam patterns (counter-clockwise) at the BS side (s-th time slot, i-th RF chain) and the UE side (s-th time slot, j-th RF chain) side, where (a) corresponds to the quantized AoD subset $\mathcal{U}_{s,i} = [1,3,4,6,8,10]^{\mathsf{T}}$ and the probing vector $\check{\mathbf{u}}_{s,i} = \frac{1}{\sqrt{6}}[1,0,1,1,0,1,0,1,0,1]^{\mathsf{T}}$, and where (b) corresponds to the quantized AoA subset $\mathcal{V}_{s,j} = [2,4,5,7,9]^{\mathsf{T}}$ and the probing vector $\check{\mathbf{v}}_{s,j} = \frac{1}{\sqrt{5}}[0,1,0,1,1,0,1,0,1,0]^{\mathsf{T}}$.

a data slot with an *Acknowledgment* (ACK) packet and the reliable communication with beamforming gain at both sides continues afterwards. In the following sections, we provide a specific description of the proposed BA scheme.

B. BS Downlink (DL) Channel Probing

In this section, we consider the Gaussian channel given by $\mathbf{H}_s[\omega]$ as in Section II-E and its quantization $\check{\mathbf{H}}_s[\omega]$ at time slot $s \in [T]$ and subcarrier $\omega \in [F]$, where $T \leq T_0$ is the effective period of training. We assume that the geometry of the channel given by the locations of the scatterers and their second order strengths (covariance) remains invariant across many coherence blocks [1, 13, 19].

As mentioned before, in our proposed approach, the channel is always probed in the DL via the BS and all the UEs stay in the listening mode. At each training time slot s, the BS uses its m RF chains to probe the channel along m beamforming vectors $\mathbf{u}_{s,i}$, $i \in [m]$, by transmitting an OFDM symbol $x_{s,i}(t)$ along each $\mathbf{u}_{s,i}$. We assume that each OFDM symbol $x_{s,i}(t)$ probes only a subset $\mathcal{F}_i \subset [F]$ of subcarriers of size $F_i = |\mathcal{F}_i|$ and that different waveforms $x_{s,i}(t)$ have disjoint sets of subcarriers, i.e., $\mathcal{F}_i \cap \mathcal{F}_{i'} = \emptyset$ for $i \neq i'$. This is illustrated Fig. 3, We make the additional assumption that the comb of non-adjacent subcarriers in each subset $\mathcal{F}_i$ are sufficiently separated from each other (beyond channel coherence bandwidth) such that their corresponding channel matrices $\mathbf{H}_s[\omega]$, $\omega \in \mathcal{F}_i$, are uncorrelated, thus, independent due to Gaussian assumption.

Consider a specific UE as before and suppose that the UE has n RF chains. Assume that the UE applies beamforming vectors $\mathbf{v}_{s,j}$ via its j-th RF chain, where it receives

$$r_{s,i,j}[\omega] = \mathbf{v}_{s,j}^{\mathsf{H}} \mathbf{H}_s[\omega] \mathbf{u}_{s,i} \check{x}_{s,i}[\omega] + \check{z}_s[\omega], \omega \in \mathcal{F}_i, \qquad (31)$$

where $\check{x}_{s,i}[\omega] \in \mathbb{C}$ denotes the pilot symbol transmitted at ω-th subcarrier of the OFDM symbol $x_{s,i}(t)$ and where $\check{z}_s[\omega]$ denotes the noise at subcarrier $\omega \in \mathcal{F}_i$. Note that in (31), we used the fact that the signals corresponding to different OFDM symbols have disjoint set of subcarriers and are separable at the UE side. We can also write (31) directly in terms of the quantized channel model as follows

$$r_{s,i,j}[\omega] = \check{\mathbf{v}}_{s,j}^{\mathsf{H}} \check{\mathbf{H}}_s[\omega] \check{\mathbf{u}}_{s,i} \check{x}_{s,i}[\omega] + \check{z}_s[\omega], \omega \in \mathcal{F}_i, \qquad (32)$$

where $\check{\mathbf{v}}_{s,j} = \mathbf{F}_N^{\mathsf{H}} \mathbf{v}_{s,j}$ and $\check{\mathbf{u}}_{s,i} = \mathbf{F}_M^{\mathsf{H}} \mathbf{u}_{s,i}$ are representations of the channel vectors $\mathbf{v}_{s,j}$ and $\mathbf{u}_{s,i}$ in the DFT basis $\mathbf{F}_N$ and $\mathbf{F}_M$ respectively. A main ingredient of our proposed BA scheme is a pseudo-random probing (transmit) codebook at the BS given by the parameters

$$\mathcal{C}_{\text{BS}} := (P_{\mathtt{tot}}, \mathcal{F}_i, F_i, \mathbf{u}_{s,i}, \kappa_u, \{\check{x}_{s,i}[\omega]\}_{\omega \in \mathcal{F}_i}, s \in [T]), \qquad (33)$$

where $P_{\mathtt{tot}}$ denotes the total probing power, and where T is the effective probing length, i.e., the period (in slots) over which the probing is repeated. For simplicity, we will assume that $F_i = F'$ is the same for all transmitted OFDM symbols and that the pilot sequence $\{\check{x}_{s,i}[\omega]\}_{\omega \in \mathcal{F}_i}$ at each time slot $s \in [T]$ have a uniform power distribution with $\mathbb{E}[|\check{x}_{s,i}[\omega]|^2] = \frac{P_{\mathtt{tot}}}{mF'}$. We will also assume that each beamforming vector $\mathbf{u}_{s,i}$ probes a random subset of quantized AoDs of size κ_u as in (33). In particular, denoting by $\mathcal{U}_{s,i}$ such a subset of AoDs of size $|\mathcal{U}_{s,i}| = \kappa_u$, we will set $\check{\mathbf{u}}_{s,i} = \frac{\mathbf{1}_{\mathcal{U}_{s,i}}}{\sqrt{\kappa_u}}$, where $\mathbf{1}_{\mathcal{U}_{s,i}}$ denotes a vector with 1 at components belonging to $\mathcal{U}_{s,i}$ and 0 elsewhere. A simple example is illustrated in Fig. 4. We assume that the *spreading factor* κ_u and the probing subsets $\mathcal{U}_{s,i}$ are appropriately designed to guarantee a *good* channel estimation performance for all the UEs in the system. In particular, by varying κ_u the BS has the option to spread/focus its transmit power over AoDs of different sizes[2]. In this paper, we consider a random design for $\mathcal{U}_{s,i}$ leaving the suitable design of the codebook $\mathcal{U}_{s,i}$ for a future work. We assume that the codebook $\mathcal{C}_{\text{BS}}$ is shared across the whole system and is known to all the UEs. This can be done, for example, by sharing a random seed sequence such that each UE can reproduce locally the pseudo-random codebook of the BS from its own local clock, which is assumed to be globally distributed across the whole system.

Remark *1:* Note that in our scheme the pseudo-random beam patterns can be regarded as a generalization of the classical "coarse beam sweeping", where in the classical approach, the probing vector packs all directions in adjacent blocks [11, 12]. We show in later section that we can achieve a much better and more accurate estimation of the channel by using random beam patterns instead of classical coarse beam sweeping.

Remark *2:* In our proposed scheme only the BS probes the channel and all the UEs are in the receiving mode,

[2]This corresponds to the well-known *exploration-exploitation* trade-off in statistics. For larger κ_u, the BS is able to explore a larger set of AoDs but the measurements received at the UEs have a low SNR. For smaller κ_u, in contrast, the BS focuses its transmit power on a small subset of AoDs and the resulting measurements have a high SNR when the beamforming vector of the BS hits a strong scatterer but the drawback is that due to very small κ_u and the very sparse nature of the channel in mm-Waves the probability of hitting such a strong scatterer is quite low.

thus, the channel estimation is quite different than the interactive channel estimation schemes in which both the UE and the BS can probe the channel. All the interactive schemes require some coordination among the UEs, which is difficult to establish in the acquisition mode. In the absence of such a coordination, the simultaneous transmission of those UEs interested in joining the system might create huge non-coherent interference to the UEs already in the system.

C. UE Probing Scheme

The second ingredient of our proposed BA algorithm is a local receive codebook at each UE. In contrast with the BS codebook, which is shared among all the UEs, each UE can have its own local codebook $\mathcal{C}_{\text{UE}} := (\mathbf{v}_{s,j}, s \in [T])$. This has the advantage that each UE can change its codebook according to its channel state, e.g., its receive SNR, to speed up its channel estimation. In this paper, we assume that $\check{\mathbf{v}}_{s,j}$ is given by $\check{\mathbf{v}}_{s,j} = \frac{\mathbf{1}_{\mathcal{V}_{s,j}}}{\sqrt{\kappa_v}}$, where $\mathcal{V}_{s,j}$ denotes a subset of AoAs of size κ_v probed at the UE side, where $\mathcal{V}_{s,j}$ is selected completely randomly. Similar to the κ_u parameter which controls the spread of power along the AoDs at the BS side, the parameter κ_v controls the spread of probing window at the UE side. This is illustrated in Fig. 4.

D. Probing SNR

During the probing stage, the BS spreads its transmit power $P_{\mathtt{tot}}$ across m OFDM symbols transmitted via m RF chains, where only a subset of subcarriers $\mathcal{F}_i$ of size $F' = |\mathcal{F}_i|$ is used to transmit the pilot signal in each OFDM symbol. We define the probing SNR per subcarrier by

$$\begin{aligned}\mathsf{SNR}^{\text{CE}}_{\text{ABF}} &:= \frac{MNP_{\mathtt{tot}}\sum_{l=1}^{L}\gamma_l}{\kappa_u\kappa_v B' N_0} = \frac{MNP_{\mathtt{tot}}\sum_{l=1}^{L}\gamma_l}{\kappa_u\kappa_v mF'\Delta f N_0} \\ &= \frac{MNP_{\mathtt{dim}}\sum_{l=1}^{L}\gamma_l}{\kappa_u\kappa_v \Delta f N_0} = \frac{P_{\mathtt{dim}}}{\sigma^2}\frac{MN\sum_{l=1}^{L}\gamma_l}{\kappa_u\kappa_v} \end{aligned} \tag{34}$$

where $B' = mF'\Delta f = \frac{mF'}{t_0}$ is the fraction of bandwidth occupied by the transmitted probing signal, where $P_{\mathtt{dim}} = \frac{P_{\mathtt{tot}}}{mF'}$ denotes the power per orthogonal dimension (subcarrier), where N_0 is the one-sided noise PSD, and where $\sigma^2 = N_0\Delta f$ denotes the noise power per subcarrier with $\Delta f = \frac{1}{t_0}$ denoting the frequency separation between the subcarriers. Note that (34) gives the maximum SNR that one can have in each subcarrier in $\mathcal{F}_i$ and corresponds to the case where the probing window of AoA-AoDs $\mathcal{U}_{s,i} \times \mathcal{V}_{s,j}$ contains all the scatterers in $\mathcal{F}_i$. In the spacial case where $\kappa_u = M$ and $\kappa_v = N$, thus, the transmit power is uniformly spread over all AoDs at the BS side and there is effectively no beamforming gain at the UE side, this reduces to SNR before beamforming in (15) for a signaling bandwidth $B' = mF'\Delta f$. We also define the SNR per subcarrier before beamforming by

$$\mathsf{SNR}^{\text{CE}}_{\text{BBF}} := \frac{\mathsf{SNR}^{\text{CE}}_{\text{ABF}}}{MN} = \frac{P_{\mathtt{dim}}}{\sigma^2}\frac{\sum_{l=1}^{L}\gamma_l}{\kappa_u\kappa_v} \tag{35}$$

E. Channel Statistics in Time

In this paper, our goal is to design a BA scheme that is robust to variations in channel statistics. So, we assume that the channel $\check{\mathbf{H}}_s[\omega]$ might vary quite drastically in time $s \in [T]$, which is highly motivated due to the very short channel coherence time at mm-Waves due to high carrier frequency and high Doppler spread. For the rest of the paper, we consider an extreme case where the channel $\check{\mathbf{H}}_s[\omega]$ varies i.i.d. in time slot $s \in [T]$, where for a fixed s, $\check{\mathbf{H}}_s[\omega]$ is also independent across the subcarriers $\omega \in \mathcal{F}_i$ due to the large frequency separation (beyond the channel coherence bandwidth) as illustrated in Fig. 3. Under such channel statistics, we can assume without any loss of generality that the pilot sequences $\check{x}_{s,i}[\omega]$, $\omega \in \mathcal{F}_i$, satisfy $\check{x}_{s,i}[\omega] = \sqrt{\frac{P_{\mathtt{tot}}}{mF'}}$, where $\frac{P_{\mathtt{tot}}}{mF'}$ denotes the transmit power uniformly distributed among the set of all mF' subcarriers.

F. Proposed Channel Estimation at the UE side

In this section, we explain our proposed channel estimation algorithm. We consider the signal model as in (32)

$$\begin{aligned} r_{s,i,j}[\omega] &= \check{\mathbf{v}}_{s,j}^{\mathsf{H}}\check{\mathbf{H}}_s[\omega]^{\mathsf{H}}\check{\mathbf{u}}_{s,i}\check{x}_{s,i}[\omega] + \check{z}_s[\omega] && (36)\\ &= \sqrt{\frac{P_{\mathtt{tot}}}{mF'}}\check{\mathbf{v}}_{s,j}^{\mathsf{H}}\check{\mathbf{H}}_s[\omega]^{\mathsf{H}}\check{\mathbf{u}}_{s,i} + \check{z}_s[\omega] && (37)\\ &= \sqrt{P_{\mathtt{dim}}}(\check{\mathbf{u}}_{s,i}\otimes\check{\mathbf{v}}_{s,j})^{\mathsf{H}}\check{\mathbb{h}}_s[\omega] + \check{z}_s[\omega] && (38)\\ &= \sqrt{P_{\mathtt{dim}}}\,\mathbf{g}_{s,i,j}^{\mathsf{H}}\check{\mathbb{h}}_s[\omega] + \check{z}_s[\omega], && (39)\end{aligned}$$

where $\check{x}_{s,i}[\omega]$ denotes the pilot symbol transmitted along subcarrier $\omega \in \mathcal{F}_i$, where $\mathbb{E}[|\check{x}_{s,i}[\omega]|^2] = \frac{P_{\mathtt{tot}}}{mF'} = P_{\mathtt{dim}}$, where for simplicity, we assume that $\check{x}_{s,i}[\omega] = \sqrt{\frac{P_{\mathtt{tot}}}{mF'}} = \sqrt{P_{\mathtt{dim}}}$ as explained in Section III-E where $P_{\mathtt{dim}} = \frac{P_{\mathtt{tot}}}{mF'}$ denotes the average power per orthogonal dimension (subcarrier), where $\check{z}_s[\omega] \sim \mathcal{CN}(0, N_0\Delta f)$ is the AWGN noise at subcarrier ω, where $\check{\mathbb{h}}_s[\omega] = \text{vec}(\check{\mathbf{H}}_s[\omega]^{\mathsf{H}})$ denotes the vectorized channel at subcarrier $\omega \in \mathcal{F}_i$ with vec denoting the vec operator, where we used the well-known identity $\text{vec}(\mathbf{ABC}) = (\mathbf{C}^{\mathsf{T}} \otimes \mathbf{A})\text{vec}(\mathbf{B})$ for the vec operator, and where for simplicity of the notation we define the combined probing vector consisting of the beamforming vector $\check{\mathbf{u}}_{s,i}$ at the BS and beamforming vector $\check{\mathbf{v}}_{s,j}$ at the UE by $\mathbf{g}_{s,i,j} = \check{\mathbf{u}}_{s,i}\otimes\check{\mathbf{v}}_{s,j} \in \mathbb{C}^{MN}$, which is common across all the subcarriers $\omega \in \mathcal{F}_i$ but differs for different pairs of RF chains (i,j) at the BS and the UE.

We define the average instantaneous received power at the j-th RF chain at the UE from the signal transmitted from the i-th RF chain at the BS by

$$\begin{aligned}\check{q}_{s,i,j} &= \frac{1}{F'}\sum_{\omega\in\mathcal{F}_i}|r_{s,i,j}[\omega]|^2 \\ &= P_{\mathtt{dim}}\mathbf{g}_{s,i,j}^{\mathsf{H}}\Big(\frac{1}{F'}\sum_{\omega\in\mathcal{F}_i}\check{\mathbb{h}}_s[\omega]\check{\mathbb{h}}_s[\omega]^{\mathsf{H}}\Big)\mathbf{g}_{s,i,j} \\ &\quad + \frac{1}{F'}\sum_{\omega\in\mathcal{F}_i}|z_s[\omega]|^2 + \frac{1}{F'}\sum_{\omega\in\mathcal{F}_i}\xi_s[\omega], \end{aligned} \tag{40}$$

where the first and the second term correspond to the signal part and the noise part, and where

$$\xi_s[\omega] = \sqrt{P_{\mathtt{dim}}}\mathbf{g}_{s,i,j}^{\mathsf{H}}\check{\mathbb{h}}_s[\omega]z_s[\omega]^{\mathsf{H}} + \sqrt{P_{\mathtt{dim}}}z_s[\omega]\check{\mathbb{h}}_s[\omega]^{\mathsf{H}}\mathbf{g}_{s,i,j}$$

denotes the signal-noise cross term. Note that when the number of subcarriers F' in $\mathcal{F}_i$ is very large such that the cross term is negligible and the empirical covariance

$$\frac{1}{F'}\sum_{\omega\in\mathcal{F}_i}\check{\mathbb{h}}_s[\omega]\check{\mathbb{h}}_s[\omega]^{\mathsf{H}} \to \mathbb{E}[\mathbb{h}_s[\omega]\mathbb{h}_s[\omega]^{\mathsf{H}}] =: \mathbf{\Sigma}_{\mathbb{h}}, \tag{41}$$

(40) gives a 1-dim noisy projection of the covariance matrix $\mathbf{\Sigma}_{\mathbb{h}}$. It is important to note that $\mathbf{\Sigma}_{\mathbb{h}} := \mathbb{E}[\mathbb{h}_s[\omega]\mathbb{h}_s[\omega]^{\mathsf{H}}]$ is independent of the subcarrier index $\omega \in \mathcal{F}_i$ due to the channel stationarity in the frequency domain. We can write (40) as

$$\check{q}_{s,i,j} \approx P_{\mathtt{dim}}\mathbf{g}_{s,i,j}^{\mathsf{H}}\mathbf{\Sigma}_{\mathbb{h}}\mathbf{g}_{s,i,j} + \sigma^2, \tag{42}$$

where $\sigma^2 = N_0\Delta f$ denotes the noise variance in each subcarrier. When all the AoA-AoDs lie on a discrete grid, $\check{\mathbb{h}}_s[\omega]$ is a sparse vector with i.i.d. components with only a few nonzero coefficients corresponding to the scatterers. Due to the independence of the channel gain of the scatterers, the covariance matrix of $\check{\mathbb{h}}_s[\omega]$ would be a diagonal $MN \times MN$ matrix with only a few nonzero diagonal elements corresponding to the scatterers. $\mathbf{\Sigma}_{\mathbb{h}}$ is still sparse and approximately diagonal for sufficiently large M and N (as illustrated in Fig. 2) even if the AoA-AoDs of the scatterers do not lie on the discrete grid.

In this paper, we use equally weighted beamforming vectors $\check{\mathbf{u}}_{s,i} = \frac{\mathbf{1}_{\mathcal{U}_{s,i}}}{\sqrt{\kappa_u}}$ at the BS and $\check{\mathbf{v}}_{s,j} = \frac{\mathbf{1}_{\mathcal{V}_{s,j}}}{\sqrt{\kappa_v}}$ at the UE, for which (42) can be simplified to

$$\check{q}_{s,i,j} \approx \frac{P_{\mathtt{dim}}MN}{\kappa_u\kappa_v}\sum_{(r,c)\in\mathcal{N}_{s,i,j}}\check{\mathbf{\Gamma}}_{r,c} + \sigma^2, \tag{43}$$

where $\mathcal{N}_{s,i,j} = \mathcal{U}_{s,i}\times\mathcal{V}_{s,j}$ denotes a subset of AoA-AoDs probed by the beamforming vectors $\check{\mathbf{u}}_{s,i}$ and $\check{\mathbf{v}}_{s,j}$ and where $\check{\mathbf{\Gamma}}$ is an $M\times N$ with $\check{\mathbf{\Gamma}}_{r,c}$ denoting the channel gain (second order) of the scatterer located at the AoD r at the BS and the AoA c at the UE (if there is any) after the discrete quantization and is zero elsewhere. This is illustrated in Fig. 5. We define the normalized matrix $\mathbf{\Gamma} = \frac{P_{\mathtt{dim}}MN\check{\mathbf{\Gamma}}}{\kappa_u\kappa_v}$ and apply the vectorization to write (43) as

$$\check{q}_{s,i,j} \approx \mathbf{b}_{s,i,j}^{\mathsf{T}}\boldsymbol{\gamma} + \sigma^2, \tag{44}$$

where $\mathbf{b}_{s,i,j} = \mathbf{1}_{\mathcal{V}_{s,j}}\otimes\mathbf{1}_{\mathcal{U}_{s,i}}$ is a normalized binary vector containing 1 at the AoA-AoDs probed by $\mathcal{V}_{s,j}\times\mathcal{U}_{s,i}$ and is 0 elsewhere, and where $\boldsymbol{\gamma} = \mathrm{vec}(\mathbf{\Gamma}) \in \mathbb{R}_+^{MN}$. In general, for a finite number of subcarriers, (42) and (44) hold only approximately since the statistical fluctuations are not negligible. We consider this as a residual noise $\check{w}_{s,i,j}$ and write (44) as

$$\check{q}_{s,i,j} = \mathbf{b}_{s,i,j}^{\mathsf{T}}\boldsymbol{\gamma} + \sigma^2 + \check{w}_{s,i,j}. \tag{45}$$

Overall, when the BS transmits along m RF chain in each acquisition slot (over an OFDM symbol) and UE has n RF chains to probe the channel, the UE obtains mn equations for the unknown vector $\boldsymbol{\gamma}$ as in (45). Thus, over T acquisition slots the UE obtains mnT equations given by

$$\check{\mathbf{q}} = \mathbf{B}\boldsymbol{\gamma} + \sigma^2\mathbf{1} + \check{\mathbf{w}}, \tag{46}$$

where the vector $\check{\mathbf{q}} = [\check{q}_{1,1,1},\dots\check{q}_{1,m,n},\dots,\check{q}_{T,1,1},\dots,\check{q}_{T,m,n}]^{\mathsf{T}}$ consists of all mnT measurements, where the $mnT\times MN$ matrix $\mathbf{B} = [\mathbf{b}_{1,1,1},\dots,\mathbf{b}_{1,m,n},\dots,\mathbf{b}_{T,1,1},\dots,\mathbf{b}_{T,m,n}]^{\mathsf{T}}$ has the set of vectors $\mathbf{b}_{s,i,j}$, and where $\check{\mathbf{w}}\in\mathbb{R}^{mnT}$ is the residual noise in the measurements. Some remarks are in order here.

Remark *3:* In our proposed scheme, at each acquisition slot, each UE extracts its own set of measurements from its received signal. An implicit assumption, however, is that each UE is synchronized with the BS and knows the BS codebook $\mathcal{C}_{\mathsf{BS}}$ such that it can construct the equations (46) and estimate the channel parameter $\boldsymbol{\gamma}$. In contrast with the BS codebook $\mathcal{C}_{\mathsf{BS}}$, which is common to all the UEs, the probing codebook $\mathcal{C}_{\mathsf{UE}} := (\mathbf{v}_{s,j}, s\in[T])$ or equivalently $\mathcal{C}_{\mathsf{UE}} := (\mathcal{V}_{s,j}, s\in[T])$ can be different for the UEs. In particular, by changing the probing width κ_v, each UE is able to control its training SNR according to its channel state: *strong* users with a very good channel can select larger κ_v to explore the channel better and estimate $\boldsymbol{\gamma}$ faster, whereas *weak* users with a very bad channel should select a smaller κ_v to attain a reasonable training SNR. Although the *weak* users might need to wait longer to take more measurements before they are able to estimate their channels, (46) remains still valid since the channel gains (second order channel statistics) $\boldsymbol{\gamma}$ are stable across many coherence blocks. This is in contrast with the conventional CS-based techniques used for BA via estimating the instantaneous complex channel gains, where the underlying channel might change drastically while taking the measurements, especially when only very few number of RF chains m, n are available, thus, T needs to be large to guarantee a sufficient number of measurements mnT.

G. Non-Negative Least Squares

To estimate the channel geometry and, in particular, the AoA-AoDs of the dominant scatterers, we need to solve the equations (46) repeated here for the convenience

$$\check{\mathbf{q}} = \mathbf{B}\boldsymbol{\gamma} + \sigma^2\mathbf{1} + \check{\mathbf{w}}. \tag{47}$$

As an additional constraint, we know that $\boldsymbol{\gamma}$ is sparse, with only a few nonzero elements corresponding to the dominant scatterers which are also non-negative. Moreover, although the exact statistics of the noise $\check{\mathbf{w}}$ is not known, it is known that it is symmetric, has zero mean, and can be approximated with a complex Gaussian random variable when the number of subcarriers per data stream F' is sufficiently large (see (40)). To solve an equation such as (47), it is traditional to apply a Regularized Least Squares of the form

$$\boldsymbol{\gamma}^* = \arg\min_{\boldsymbol{\gamma}}\frac{1}{2}\|\mathbf{B}\boldsymbol{\gamma} + \sigma^2\mathbf{1} - \check{\mathbf{q}}\|^2 + \zeta R(\boldsymbol{\gamma}) \tag{48}$$

where $R(\boldsymbol{\gamma})$ is a regularization function and where $\zeta > 0$ is a regularization parameter. $R(\boldsymbol{\gamma})$ promotes a specific structure of the solution $\boldsymbol{\gamma}$ in (48) such as its sparsity or non-negativity. A regularizer traditionally used is the l_1-norm of $\boldsymbol{\gamma}$ given by $R(\boldsymbol{\gamma}) = \sum_{i=1}^{MN}|\gamma_i|$ which simplifies to $R(\boldsymbol{\gamma}) = \sum_{i=1}^{MN}\gamma_i$ when $\gamma_i > 0$. Recent progress in CS shows that when the underlying parameter $\boldsymbol{\gamma}$ is non-negative, a simple Least Squares as in

$$\boldsymbol{\gamma}^* = \arg\min_{\boldsymbol{\gamma}\in\mathbb{R}_+^{MN}}\frac{1}{2}\|\mathbf{B}\boldsymbol{\gamma} + \sigma^2\mathbf{1} - \check{\mathbf{q}}\|^2 \tag{49}$$

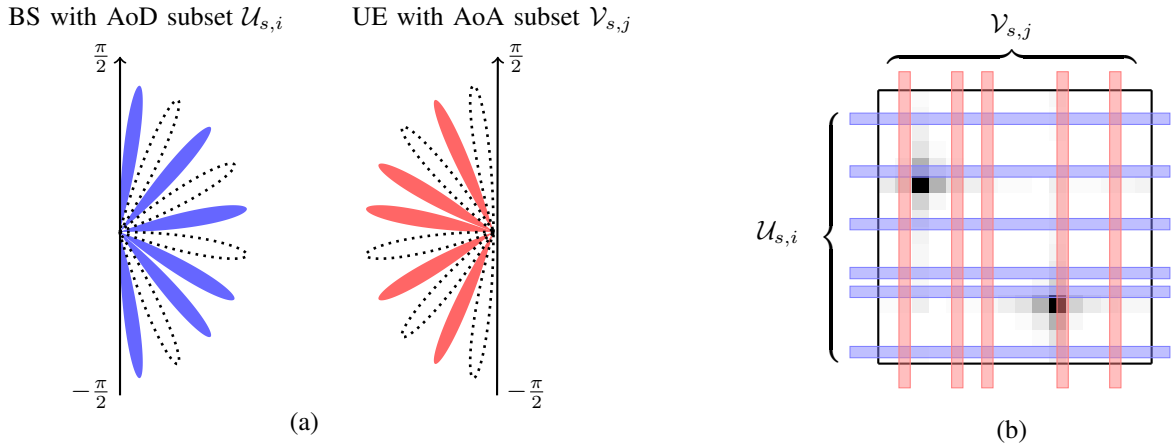

Fig. 5: Illustration of (a) the subset of AoA-AoDs at time slot s probed by i-th RF chain at the BS side and j-th RF chain at the user side respectively, (b) the channel gain matrix $\check{\mathbf{\Gamma}}$ (with two off-grid scatterers) measure along $\mathcal{U}_{s,i} \times \mathcal{V}_{s,j}$.

is still enough to impose the sparse structure of $\boldsymbol{\gamma}$ without any extra regularization [22, 23] . The Least Squares problem with additional non-negativity constraint in (49) is called the *Non-Negative Least Squares* (NNLS) and is quite well-known in the statistics. A quite early reference is [24] where it shows that depending on the structure of the measurement matrix (e.g., $\mathbf{B}$ in our case) NNLS might illustrate a "Super-Resolution" property, thus, allowing the underlying parameter $\boldsymbol{\gamma}$ to be estimated very well. More recently, by the advent of CS [25, 26], the NNLS has reemerged in the context of sparse signal recovery, where it has been shown that the non-negativity constraint alone might suffice to recover the underlying signal $\boldsymbol{\gamma}$ in the noiseless [27–30] as well as the noisy case [22, 23]. Moreover, [22] illustrates that the performance of NNLS is comparable to that of the LASSO [31] even though it does not apply any regularization per se. [22] also illustrates that NNLS along with an appropriate thresholding provides state-of-the-art performance in terms of support estimation. This is beneficial for identifying the underlying sparse geometry of a mm-Wave channel, encoded by the vectors $\boldsymbol{\gamma}$, that we address in this paper. Recently, NNLS has also been of huge interest in other related areas such as Machine Learning.

In terms of numerical implementations, the NNLS can be posed as an unconstrained Least Squares problem over the positive orthant and can be solved by many possible techniques such as Gradient Projection, Primal-Dual techniques, etc. [32] with an affordable computational complexity. We refer to [33, 34] for the recent progress on the numerical solution of NNLS and a discussion on other related work in the literature.

IV. Performance Evaluation

In this section, we evaluate the the performance of our proposed algorithm via numerical simulations. To run the NNLS optimization in (49), we use the implementation of NNLS in MATLAB© called `lsqnonneg.m`.

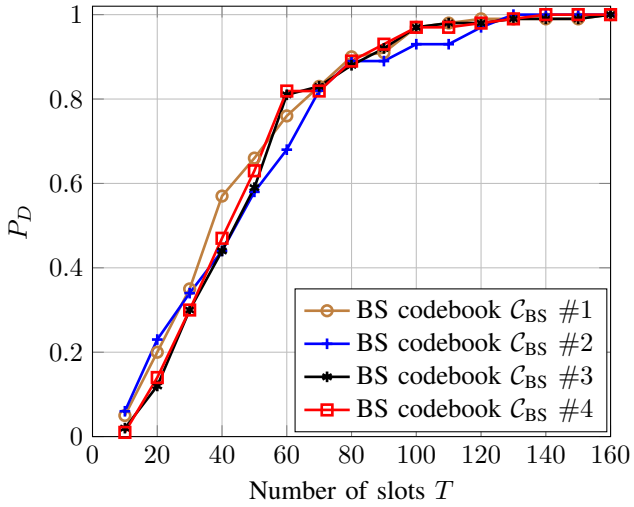

Fig. 6: Comparison of the detection probability P_D of different pseudo-random codebooks (denoted by $\mathcal{C}_{\text{BS}}$) achieved by the proposed NNLS scheme, for $M = N = 32$, $F' = 3$, $m = 3$, $n = 2$, $\mathsf{SNR}^{\text{CE}}_{\text{BBF}} = -10$ dB, $\kappa_u = \kappa_v = 16$.

A. Channel and Signal Model

For simplicity, we consider a very sparse channel model with only one path (one scatterer). The system is assumed to operate at 28 GHz carrier frequency and has a bandwidth of $B = 100$ MHz [11]. We adopt OFDM signaling, where each OFDM symbol has an effective duration of 71.3 μs with 66.7 μs and a CP of 4.7 μs. The subcarrier spacing is 15 KHz, namely around 6.7×10^3 orthogonal subcarriers are available [35]. We assume that the BS has $M = 32$ antennas and $m = 3$ RF chains, and the UE has $N = 32$ antennas and $n = 2$ RF chains. For simulations, we assume that each OFDM signal used for channel estimation probes $F' = 3$ subcarriers, i.e., the total probing power $P_{\mathtt{tot}}$ is uniformly distributed across $m = 3$ RF chains and across $F' = 3$ subcarriers in each OFDM symbol ($mF' = 9$ subcarriers in total). The SNR before beamforming at each subcarrier is $\mathsf{SNR}^{\text{CE}}_{\text{BBF}} = -10$ dB. We announce a individual experiment to be successful if the

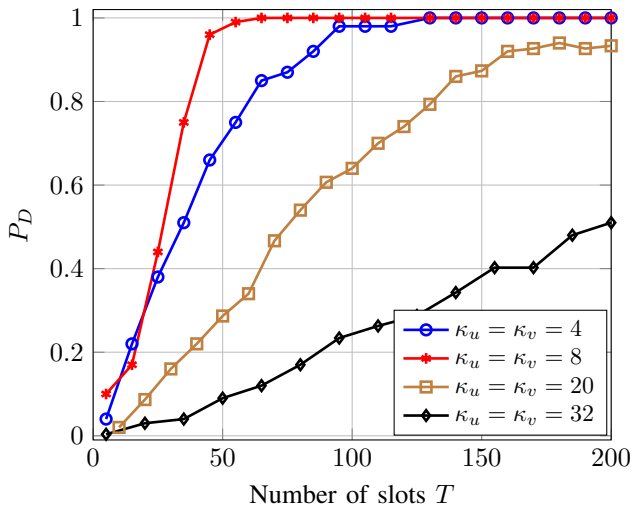

Fig. 7: Detection probability of our proposed scheme for different power spreading factors (κ_u, κ_v), where $M = N = 32$, $F' = 3$, $m = 3$, $n = 2$, $\mathsf{SNR}_{\text{BBF}}^{\text{CE}} = -10$ dB per subcarrier.

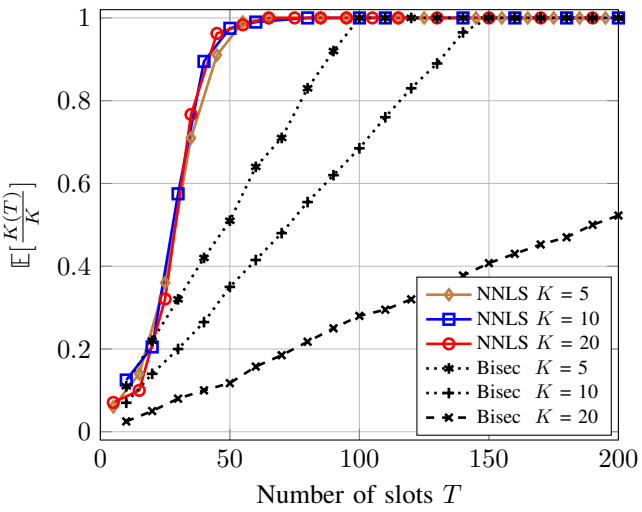

Fig. 8: Comparison of the performance of our proposed scheme with that of interactive bisection method in [11] in terms of the fraction of users whose channel is estimated until a given time slot T given by $\frac{K(T)}{K}$. We take $M = N = 32$, $F' = 3$, $m = 3$, $n = 2$, $\mathsf{SNR}_{\text{BBF}}^{\text{CE}} = -10$ dB per subcarrier, $\kappa_u = \kappa_v = 8$.

index of the strongest component in $\boldsymbol{\gamma}$ is correctly estimated.

B. Concentration Result for the Random BS Codebook

We randomly generate 4 probing codebooks at the BS side, which are known to all the users. Fig. 6 illustrates the detection probability for different pseudo-random codebooks, where the power spreading factors at the BS and the user sides are set to $\kappa_u = \kappa_v = 16$ respectively. We repeat the experiment 200 times and plot the resulting detection probability versus training period length T. It is seen that, as expected, increasing T improves the detection probability significantly. More importantly, different codebooks have quite similar performances. This illustrates that the performance is statistically concentrated for a random BS codebook, i.e., it is not sensitive to the specific transmit codebook design, as far as it is random.

C. Dependence on the Beam Spreading factors κ_u and κ_v

Note that in the proposed NNLS method, the spreading factors κ_u and κ_v impose a trade-off between the coverage of the probing window $\mathbf{B}$ (exploration) and its receive SNR at the user side (exploitation). This is illustrated in Fig. 7. It is seen that increasing the spreading factor from $\kappa_u = \kappa_v = 4$ to $\kappa_u = \kappa_v = 8$ improves the performance. However, increasing κ_u, κ_v to $\kappa_u = \kappa_v = 20, 32$ degrades the performance considerably.

D. System-level Scalability

We consider a multi-user scenario. We denote by K the total number of active users in the system, and by $K(T)$ the number of users that are able to estimate their channel within T time slots. Fig. 8 compares the fraction $\frac{K(T)}{K}$ of those users in our scheme with that in the interactive bisection method proposed in [11]. As we can see, the training overhead of hierarchical interactive methods scales proportionally with the number of active users, whereas in our scheme all the users are trained more or less simultaneously, thus, the overhead does not grow with the number of users. This is due to the fact in our proposed algorithm, the BS probing patterns are not adapted to any specific user (in contrast with the interactive methods) and cover the channels of all the users within the BS coverage simultaneously.

E. Robustness w.r.t. Variations in Channel Statistics

As we explained in Section I, the existing CS-based methods in the literature assume that the instantaneous channel remains constant during the whole training stage, which is quite difficult to meet in mm-Waves. To investigate the sensitivity of the proposed CS techniques to channel variations, we consider a very simple model for temporal channel variations given by

$$\rho_{s,l} = \alpha \rho_{s-1,l} + \sqrt{1 - |\alpha|^2}\, \nu_{s,l}, s \in \mathbb{Z}_+, \qquad (50)$$

where we adopt the initialization $\rho_{0,l} \sim \mathcal{CN}(0, \gamma_l)$, and where $\nu_{s,l} \sim \mathcal{CN}(0, \gamma_l)$ is the white noise innovation term for the instantaneous channel gain $\rho_{s,l}$, which is i.i.d. across time (temporally white) as well as across l (spatially white), and where $|\alpha| \in [0, 1]$ is an update factor controlling the temporal variations of the channel gain $\rho_{s,l}$, which is assumed, for simplicity, to be the same for all the scatterers. Note that for $|\alpha| = 0$ the path gains $\rho_{s,l}$ change i.i.d. whereas for $|\alpha| = 1$ they are sampled once (initialized with $\rho_{0,l}$ at time slot $s = 0$) and remain invariant afterwards during the whole channel estimation. Fig. 9 illustrates the comparison of the performance of our proposed scheme with that of CS-based technique (OMP method) in [15]. It is seen that our proposed method exhibits much robust performance across a wide range of variations in channel statistics whereas the algorithm in [15] is quite fragile and fails to estimate the channel properly when the channel varies during the channel estimation.

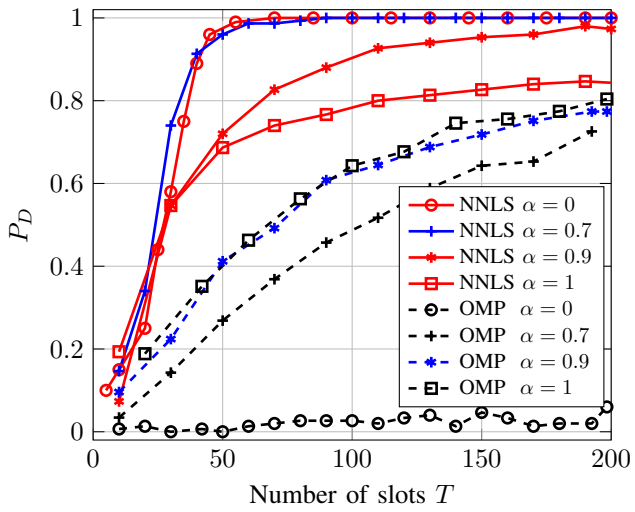

Fig. 9: Comparison of detection probability between proposed NNLS and OMP in [15] for different statistic path gains, where $M = N = 32$, $F' = 3$, $m = 3$, $n = 2$, $\mathsf{SNR}_{\mathrm{BBF}}^{\mathrm{CE}} = -10$ dB per subcarrier, $\kappa_u = \kappa_v = 8$, the path gains change from i.i.d. ($\alpha = 0$) to constant ($\alpha = 1$).

V. Conclusion

In this paper, we proposed an efficient beam alignment scheme for mm-Wave multi-user MIMO systems. In our proposed scheme, the channel is always probed by the BS in the DL, providing all the users within the BS coverage sufficiently many measurements to estimate the AoA-AoD of strong scatterers connecting them to the BS. In contrast with the conventional interactive BA algorithms, where at each time slot the BS interacts with only a single user, in our proposed scheme all the users are trained simultaneously, thus, the BA scales very well with the number of active users in the system. We posed the BA as the estimation of the second order statistics of the channel and proposed a novel NNLS technique to guarantee a reliable recovery. We illustrated, via numerical simulations and comparison with other competitive techniques in the literature, that our algorithm is highly robust to variations in the channel statistics.

Acknowledgments

X.S. is sponsored by the China Scholarship Council (201604910530).

References

[1] R. W. Heath, N. Gonzalez-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, "An overview of signal processing techniques for millimeter wave mimo systems," *IEEE journal of selected topics in signal processing*, vol. 10, no. 3, pp. 436–453, 2016.

[2] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez, "Millimeter wave mobile communications for 5g cellular: It will work!" *Access, IEEE*, vol. 1, pp. 335–349, 2013.

[3] A. M. Sayeed, "Deconstructing multiantenna fading channels," *IEEE Transactions on Signal Processing*, vol. 50, no. 10, pp. 2563–2579, 2002.

[4] T. Nitsche, C. Cordeiro, A. B. Flores, E. W. Knightly, E. Perahia, and J. C. Widmer, "Ieee 802.11 ad: directional 60 ghz communication for multi-gigabit-per-second wi-fi [invited paper]," *IEEE Communications Magazine*, vol. 52, no. 12, pp. 132–141, 2014.

[5] Z. Chen and C. Yang, "Pilot decontamination in wideband massive mimo systems by exploiting channel sparsity," *IEEE Transactions on Wireless Communications*, vol. 15, no. 7, pp. 5087–5100, 2016.

[6] S. Haghighatshoar and G. Caire, "The beam alignment problem in mmwave wireless networks," in *Signals, Systems and Computers, 2016 50th Asilomar Conference on*. IEEE, 2016, Conference Proceedings, pp. 741–745.

[7] V. Desai, L. Krzymien, P. Sartori, W. Xiao, A. Soong, and A. Alkhateeb, "Initial beamforming for mmwave communications," in *2014 48th Asilomar Conference on Signals, Systems and Computers*, 2014, Conference Proceedings, pp. 1926–1930.

[8] J. Wang, Z. Lan, C.-W. Pyo, T. Baykas, C.-S. Sum, M. A. Rahman, J. Gao, R. Funada, F. Kojima, H. Harada *et al.*, "Beam codebook based beamforming protocol for multi-gbps millimeter-wave wpan systems," *Selected Areas in Communications, IEEE Journal on*, vol. 27, no. 8, pp. 1390–1399, 2009.

[9] L. Chen, Y. Yang, X. Chen, and W. Wang, "Multi-stage beamforming codebook for 60ghz wpan," in *Communications and Networking in China (CHINACOM), 2011 6th International ICST Conference on*. IEEE, 2011, pp. 361–365.

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

[11] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, "Channel estimation and hybrid precoding for millimeter wave cellular systems," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 8, no. 5, pp. 831–846, 2014.

[12] M. Kokshoorn, H. Chen, P. Wang, Y. Li, and B. Vucetic, "Millimeter wave mimo channel estimation using overlapped beam patterns and rate adaptation," *IEEE Transactions on Signal Processing*, vol. 65, no. 3, pp. 601–616, 2017.

[13] P. Xia, R. W. Heath, and N. Gonzalez-Prelcic, "Robust analog precoding designs for millimeter wave mimo transceivers with frequency and time division duplexing," *IEEE Transactions on Communications*, vol. 64, no. 11, pp. 4622–4634, Nov 2016.

[14] D. E. Berraki, S. M. D. Armour, and A. R. Nix, "Application of compressive sensing in sparse spatial channel recovery for beamforming in mmwave outdoor systems," in *2014 IEEE Wireless Communications and Networking Conference (WCNC)*, 2014, Conference Proceedings, pp. 887–892.

[15] A. Alkhateeby, G. Leusz, and R. W. Heath, "Compressed sensing based multi-user millimeter wave systems: How many measurements are needed?" in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2015, Conference Proceedings, pp. 2909–2913.

[16] J. Choi, "Beam selection in mm-wave multiuser mimo systems using compressive sensing," *IEEE Transactions on Communications*, vol. 63, no. 8, pp. 2936–2947, 2015.

[17] R. Mndez-Rial, C. Rusu, N. Gonzlez-Prelcic, A. Alkhateeb, and R. W. Heath, "Hybrid mimo architectures for millimeter wave communications: Phase shifters or switches?" *IEEE Access*, vol. 4, pp. 247–267, 2016.

[18] K. Venugopal, A. Alkhateeb, N. G. Prelcic, and R. W. Heath, "Channel estimation for hybrid architecture based wideband millimeter wave systems," *IEEE Journal on Selected Areas in Communications*, 2017.

[19] R. J. Weiler, M. Peter, W. Keusgen, and M. Wisotzki, "Measuring the busy urban 60 ghz outdoor access radio channel," in *2014 IEEE International Conference on Ultra-WideBand (ICUWB)*, 2014, Conference Proceedings, pp. 166–170.

[20] V. Va, J. Choi, and R. W. Heath, "The impact of beamwidth on temporal channel variation in vehicular channels and its implications," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 6, pp. 5014–5029, 2017.

[21] A. Adhikary, E. Al Safadi, M. K. Samimi, R. Wang, G. Caire, T. S. Rappaport, and A. F. Molisch, "Joint spatial division and multiplexing for mm-wave channels," *IEEE J. on Sel. Areas on Commun. (JSAC)*, vol. 32, no. 6, pp. 1239–1255, 2014.

[22] M. Slawski, M. Hein *et al.*, "Non-negative least squares for high-dimensional linear models: Consistency and sparse recovery without regularization," *Electronic Journal of Statistics*, vol. 7, pp. 3004–3056, 2013.

[23] R. Kueng and P. Jung, "Robust nonnegative sparse recovery and the nullspace property of 0/1 measurements," *arXiv preprint arXiv:1603.07997*, 2016.

[24] D. L. Donoho, I. M. Johnstone, J. C. Hoch, and A. S. Stern, "Maximum entropy and the nearly black object," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 41–81, 1992.

[25] D. L. Donoho, "Compressed sensing," *Information Theory, IEEE Transactions on*, vol. 52, no. 4, pp. 1289–1306, 2006.

[26] E. J. Candes and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?" *Information Theory, IEEE Transactions on*, vol. 52, no. 12, pp. 5406–5425, 2006.

[27] A. M. Bruckstein, M. Elad, and M. Zibulevsky, "On the uniqueness of non-negative sparse & redundant representations," in *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*. IEEE, 2008, pp. 5145–5148.

[28] D. L. Donoho and J. Tanner, "Counting the faces of randomly-projected hypercubes and orthants, with applications," *Discrete & computational geometry*, vol. 43, no. 3, pp. 522–541, 2010.

[29] M. Wang and A. Tang, "Conditions for a unique non-negative solution to an underdetermined system," in *Communication, Control, and Computing, 2009. Allerton 2009. 47th Annual Allerton Conference on*. IEEE, 2009, pp. 301–307.

[30] M. Wang, W. Xu, and A. Tang, "A unique nonnegative solution to an underdetermined system: From vectors to matrices," *IEEE Transactions on Signal Processing*, vol. 59, no. 3, pp. 1007–1016, 2011.

[31] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288, 1996.

[32] D. P. Bertsekas and A. Scientific, *Convex optimization algorithms*. Athena Scientific Belmont, 2015.

[33] D. Kim, S. Sra, and I. S. Dhillon, "Tackling box-constrained optimization via a new projected quasi-newton approach," *SIAM Journal on Scientific Computing*, vol. 32, no. 6, pp. 3548–3563, 2010.

[34] D. K. Nguyen and T. B. Ho, "Anti-lopsided algorithm for large-scale nonnegative least square problems," *arXiv preprint arXiv:1502.01645*, 2015.

[35] J. Zyren and W. McCoy, "Overview of the 3gpp long term evolution physical layer," *Freescale Semiconductor, Inc., white paper*, 2007.