Channel Estimation and Hybrid Precoding for Millimeter Wave Cellular Systems

Ahmed Alkhateeb, *Student Member, IEEE*, Omar El Ayach, *Member, IEEE*, Geert Leus, and Robert W. Heath, Jr., *Fellow, IEEE*

***Abstract*—Millimeter wave (mmWave) cellular systems will enable gigabit-per-second data rates thanks to the large bandwidth available at mmWave frequencies. To realize sufficient link margin, mmWave systems will employ directional beamforming with large antenna arrays at both the transmitter and receiver. Due to the high cost and power consumption of gigasample mixed-signal devices, mmWave precoding will likely be divided among the analog and digital domains. The large number of antennas and the presence of analog beamforming requires the development of mmWave-specific channel estimation and precoding algorithms. This paper develops an adaptive algorithm to estimate the mmWave channel parameters that exploits the poor scattering nature of the channel. To enable the efficient operation of this algorithm, a novel hierarchical multi-resolution codebook is designed to construct training beamforming vectors with different beamwidths. For single-path channels, an upper bound on the estimation error probability using the proposed algorithm is derived, and some insights into the efficient allocation of the training power among the adaptive stages of the algorithm are obtained. The adaptive channel estimation algorithm is then extended to the multi-path case relying on the sparse nature of the channel. Using the estimated channel, this paper proposes a new hybrid analog/digital precoding algorithm that overcomes the hardware constraints on the analog-only beamforming, and approaches the performance of digital solutions. Simulation results show that the proposed low-complexity channel estimation algorithm achieves comparable precoding gains compared to exhaustive channel training algorithms. The results illustrate that the proposed channel estimation and precoding algorithms can approach the coverage probability achieved by perfect channel knowledge even in the presence of interference.**

***Index Terms*—Millimeter wave cellular systems, sparse channel estimation, adaptive compressed sensing, hybrid precoding.**

Manuscript received September 29, 2013; revised January 17, 2014; accepted June 10, 2014. Date of publication July 01, 2014; date of current version September 11, 2014. This work was supported by the National Science Foundation under Grant 1319556, and by a gift from Huawei Technologies. The guest editor coordinating the review of this manuscript and approving it for publication was Dr. Alexei Ashikhmin.

A. Alkhateeb is with the Department of Electrical and Computer Engineering, University of Texas at Austin, Austin, TX 78712-1687 USA (e-mail: aalkhateeb@utexas.edu).

O. El Ayach is with Corporate R&D, Qualcomm Technologies, Inc., San Diego, CA 92122-1714 USA (e-mail: oelayach@qti.qualcomm.com; omarayache@gmail.com).

G. Leus is with the Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, 2628 CD Delft, The Netherlands (e-mail: g.j.t.leus@tudelft.nl).

R. W. Heath, Jr. is with the Department of Electrical and Computer Engineering, University of Texas at Austin, Austin, TX 78712-1687 USA (e-mail: rheath@utexas.edu).

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

Digital Object Identifier 10.1109/JSTSP.2014.2334278

I. Introduction

MILLIMETER wave (mmWave) communication is a promising technology for future outdoor cellular systems [1]–[4]. Directional precoding with large antenna arrays appears to be inevitable to support longer outdoor links and to provide sufficient received signal power. Fortunately, large antenna arrays can be packed into small form factors at mmWave frequencies [5], [6], making it feasible to realize the large arrays needed for high precoding gains. The high power consumption of mixed signal components, however, makes digital baseband precoding impossible [1]. Moreover, the design of the precoding matrices is usually based on complete channel state information, which is difficult to achieve in mmWave due to the large number of antennas and the small signal-to-noise ratio (SNR) before beamforming. Because of the additional hardware constraints when compared with conventional microwave frequency multiple-input multiple-output (MIMO) systems, new channel estimation and precoding algorithms that are tailored to mmWave cellular systems must be developed.

To overcome the radio frequency (RF) hardware limitations, analog beamforming solutions were proposed in [3], [7]–[10]. The main idea is to control the phase of the signal transmitted by each antenna via a network of analog phase shifters. Several solutions, known as beam training algorithms, were proposed to iteratively design the analog beamforming coefficients in systems without channel knowledge at the transmitter. In [3], [7]–[9], adaptive beamwidth beamforming algorithms and multi-stage codebooks were developed by which the transmitter and receiver jointly design their beamforming vectors. In [10], multiple beams with unique signatures were simultaneously used to minimize the required beam training time. Despite the reduced complexity of [3], [7]–[11], they generally share the disadvantage of converging towards only one communication beam. Hence, these techniques are not capable of achieving multiplexing gains by sending multiple parallel streams. Moreover, the performance of analog strategies such as those in [3], [7]–[9] is sub-optimal compared with digital precoding solutions due to (i) the constant amplitude constraint on the analog phase shifters, and (ii) the potentially low-resolution signal phase control.

To achieve larger precoding gains, and to enable precoding multiple data streams, [11]–[14] propose to divide the precoding operations between the analog and digital domains. In [11], the joint analog-digital precoder design problem was considered for both spatial diversity and multiplexing systems. First, optimal unconstrained RF pre-processing signal transformations

followed by baseband precoding matrices were proposed, and then closed-form sub-optimal approximations when RF processing is constrained by variable phase-shifters were provided. In [12], hybrid analog/digital precoding algorithms were developed to minimize the received signal's mean-squared error in the presence of interference when phase shifters with only quantized phases are available. The work in [11], [12], however, was not specialized for mmWave systems, and did not account for mmWave channel characteristics. In [13], the mmWave channel's sparse multi-path structure [15]–[19], and the algorithmic concept of basis pursuit, were leveraged in the design of low-complexity hybrid precoders that attempt to approach capacity assuming perfect channel knowledge is available to the receiver. In [13], [14], the hybrid precoding design problem was considered in systems where the channel is partially known at the transmitter. While the developed hybrid precoding algorithms in [11], [13], [14] overcome the RF hardware limitations and can support the transmission of multiple streams, the realization of these gains require some knowledge about the channel at the transmitter prior to designing the precoding matrices. This motivates developing multi-path mmWave channel estimation algorithms, which enable hybrid precoding to approach the performance of the digital precoding algorithms.

In this paper, we develop low-complexity channel estimation and precoding algorithms for a mmWave system with large antenna arrays at both the base station (BS) and mobile station (MS). These algorithms account for practical assumptions on the mmWave hardware in which (i) the analog phase shifters have constant modulus and quantized phases, and (ii) the number of RF chains is limited, i.e., less than the number of antennas. The main contributions of the paper can be summarized as follows:

- We propose a new formulation for the mmWave channel estimation problem. This formulation captures the sparse nature of the channel, and enables leveraging tools developed in the adaptive compressed sensing (CS) field to design efficient estimation algorithms for mmWave channels.
- We design a novel multi-resolution codebook for the training precoders. The new codebook relies on joint analog/digital processing to generate beamforming vectors with different beamwidths, which is critical for proper operation of the adaptive channel estimation algorithms presented in the paper.
- We design an adaptive CS based algorithm that efficiently estimates the parameters of mmWave channels with a small number of iterations, and with high success probability. The advantage of the proposed algorithm over prior beam training work appears in multi-path channels where our algorithm is able to estimate channel parameters. Hence, it enables multi-stream multiplexing in mmWave systems, while prior work [7]–[9], [20], [21] was limited to the single-beam training and transmission.
- We analyze the performance of the proposed algorithm in single-path channels. We derive an upper bound on the error probability in estimating channel parameters, and find sufficient conditions on the total training power and its allocation over the adaptive stages of the algorithm to estimate the channel parameters with a certain bound on the maximum error probability.
- We propose a new hybrid analog/digital precoding algorithm for mmWave channels. In the proposed algorithm, instead of designing the precoding vectors as linear combinations of the steering vectors of the known angles of arrival/departure as assumed in [13], our design depends only on the quantized beamsteering directions to directly approximate the channel's dominant singular vectors. Hence, it implicitly considers the hardware limitations, and more easily generalizes to arbitrary antenna arrays.
- We evaluate the performance of the proposed estimation algorithm by simulations in a mmWave cellular system setting, assuming that both the BS and MS adopt hybrid precoding algorithms.

Simulation results indicate that the precoding gains given by the proposed channel estimation algorithm are close to that obtained when exhaustive search is used to design the precoding vectors. Multi-cell simulations show that the spectral efficiency and coverage probability achieved when hybrid precoding is used in conjunction with the proposed channel estimation strategy are comparable to that achieved when perfect channel knowledge and digital unconstrained solutions are assumed.

The rest of the paper is organized as follows. In Section II, we present the system model and main assumptions used in the paper. In Section III, we formulate the sparse channel estimation problem and present the idea of the proposed adaptive training/estimation algorithm. A hierarchical multi-resolution codebook for the training precoders and combiner is then designed in Section IV. Adaptive channel estimation algorithms are presented and discussed in Section V. The precoding design problem is formulated and a proposed hybrid RF/baseband precoding solution is presented in Section VI. In Section VII, simulation results demonstrating the performance of the proposed algorithms are given, before concluding the paper in Section VIII.

We use the following notation throughout this paper: $\mathbf{A}$ is a matrix, $\mathbf{a}$ is a vector, a is a scalar, and $\mathcal{A}$ is a set. $|\mathbf{A}|$ is the determinant of $\mathbf{A}$, $\|\mathbf{A}\|_F$ is its Frobenius norm, whereas $\mathbf{A}^T$, $\mathbf{A}^H$, $\mathbf{A}^*$, $\mathbf{A}^{-1}$, $\mathbf{A}^\dagger$ are its transpose, Hermitian (conjugate transpose), conjugate, inverse, and pseudo-inverse respectively. $[\mathbf{A}]_{\mathcal{R},:}$ ($[\mathbf{A}]_{:,\mathcal{R}}$) are the rows (columns) of the matrix $\mathbf{A}$ with indices in the set $\mathcal{R}$, and $\mathrm{diag}(\mathbf{a})$ is a diagonal matrix with the entries of $\mathbf{a}$ on its diagonal. $\mathbf{I}$ is the identity matrix and $\mathbf{1}_N$ is the N-dimensional all-ones vector. $\mathbf{A} \circ \mathbf{B}$ is the Khatri-Rao product of $\mathbf{A}$, and $\mathbf{B}$, $\mathbf{A} \otimes \mathbf{B}$ is the Kronecker product of $\mathbf{A}$, and $\mathbf{B}$, and $\mathbf{A} \odot \mathbf{B}$ denotes the Hadamard product of $\mathbf{A}$, and $\mathbf{B}$. $\mathcal{N}(\mathbf{m}, \mathbf{R})$ is a complex Gaussian random vector with mean $\mathbf{m}$ and covariance $\mathbf{R}$. $\mathbb{E}[\cdot]$ is used to denote expectation.

II. System Model

Consider the mmWave cellular system shown in Fig. 1. A BS with N_{BS} antennas and N_{RF} RF chains is assumed to communicate with a single MS with N_{MS} antennas and N_{RF} RF chains as shown in Fig. 2. The number of RF chains at the MSs is usually less than that of the BSs in practice, but we do not exploit this fact in our model. The BS and MS communicate via N_{S} data

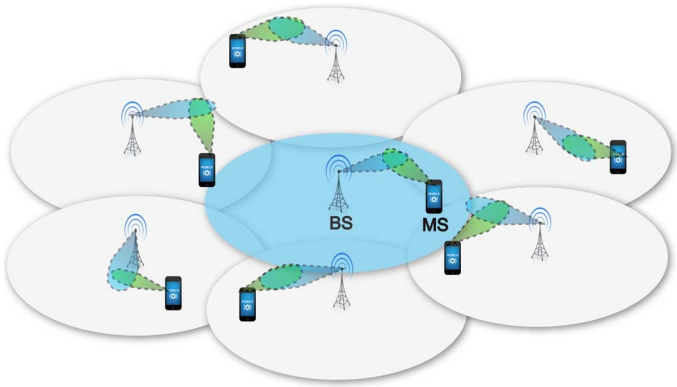

Fig. 1. A mmWave cellular system model, in which BSs and MSs communicate via directive beamforming using large antenna arrays.

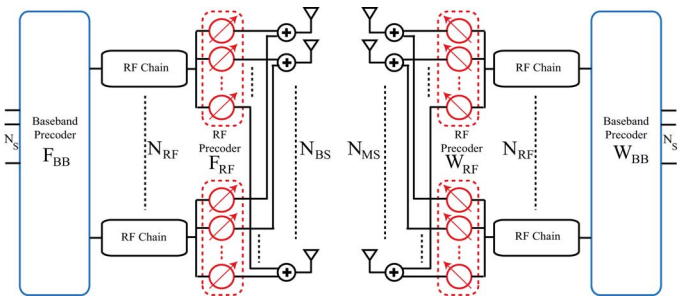

Fig. 2. Block diagram of BS-MS transceiver that uses RF and baseband beamformers at both ends.

streams, such that $N_{\mathrm{S}} \leq N_{\mathrm{RF}} \leq N_{\mathrm{BS}}$ and $N_{\mathrm{S}} \leq N_{\mathrm{RF}} \leq N_{\mathrm{MS}}$ [13], [22], [23].

In this paper, we will focus on the downlink transmission. The BS is assumed to apply an $N_{\mathrm{RF}} \times N_{\mathrm{S}}$ baseband precoder $\mathbf{F}_{\mathrm{BB}}$ followed by an $N_{\mathrm{BS}} \times N_{\mathrm{RF}}$ RF precoder, $\mathbf{F}_{\mathrm{RF}}$. If $\mathbf{F}_{\mathrm{T}} = \mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}$ is the $N_{\mathrm{BS}} \times N_{\mathrm{S}}$ combined BS precoding matrix, the discrete-time transmitted signal is then

$$\mathbf{x} = \mathbf{F}_{\mathrm{T}}\mathbf{s}, \tag{1}$$

where $\mathbf{s}$ is the $N_{\mathrm{S}} \times 1$ vector of transmitted symbols, such that $\mathbb{E}\left[\mathbf{s}\mathbf{s}^{H}\right] = (P_{\mathrm{S}}/N_{\mathrm{S}})\mathbf{I}_{N_{\mathrm{S}}}$, and P_{S} is the average total transmit power. Since $\mathbf{F}_{\mathrm{RF}}$ is implemented using analog phase shifters, its entries are of constant modulus. We normalize these entries to satisfy $\left|[\mathbf{F}_{\mathrm{RF}}]_{m,\mathrm{n}}\right|^2 = N_{\mathrm{BS}}^{-1}$, where $\left|[\mathbf{F}_{\mathrm{RF}}]_{m,\mathrm{n}}\right|$ denotes the magnitude of the (m,n)th element of $\mathbf{F}_{\mathrm{RF}}$. The total power constraint is enforced by normalizing $\mathbf{F}_{\mathrm{BB}}$ such that $\|\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}\|_F^2 = N_{\mathrm{S}}$.

We adopt a narrowband block-fading channel model in which an MS observes the received signal

$$\mathbf{r} = \mathbf{H}\mathbf{F}_{\mathrm{T}}\mathbf{s} + \mathbf{n}, \tag{2}$$

where $\mathbf{H}$ is the $N_{\mathrm{MS}} \times N_{\mathrm{BS}}$ matrix that represents the mmWave channel between the BS and MS, and $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$ is the Gaussian noise corrupting the received signal.

At the MS, the combiner $\mathbf{W}_{\mathrm{T}}$ composed of the RF and baseband combiners $\mathbf{W}_{\mathrm{RF}}$ and $\mathbf{W}_{\mathrm{BB}}$ is used to process the received signal $\mathbf{r}$ which results in

$$\mathbf{y} = {\mathbf{W}_{\mathrm{T}}}^{H}\mathbf{H}\mathbf{F}_{\mathrm{T}}\mathbf{s} + {\mathbf{W}_{\mathrm{T}}}^{H}\mathbf{n}. \tag{3}$$

We will explain the proposed algorithms for the downlink model. The same algorithms, however, can be directly applied to the uplink system whose input-output relationship is identical to (3) with $\mathbf{H}$ replaced by the uplink channel, and the roles of the precoders ($\mathbf{F}_{\mathrm{RF}}$, $\mathbf{F}_{\mathrm{BB}}$) and combiners ($\mathbf{W}_{\mathrm{RF}}$, $\mathbf{W}_{\mathrm{BB}}$) switched.

While the mmWave channel estimation and precoding algorithms developed in the following sections consider only a BS-MS link with no interfering BSs, these algorithms will also be numerically evaluated by simulations in the case of mmWave cellular systems where out-of-cell interference exists in Section VII-B.

Since mmWave channels are expected to have limited scattering [15]–[19], we adopt a geometric channel model with L scatterers. Each scatterer is further assumed to contribute a single propagation path between the BS and MS [13], [24]. Under this model, the channel $\mathbf{H}$ can be expressed as

$$\mathbf{H} = \sqrt{\frac{N_{\mathrm{BS}}N_{\mathrm{MS}}}{\rho}} \sum_{\ell=1}^{L} \alpha_\ell \mathbf{a}_{\mathrm{MS}}\left(\theta_\ell\right) \mathbf{a}_{\mathrm{BS}}^{H}\left(\phi_\ell\right), \tag{4}$$

where ρ denotes the average path-loss between the BS and MS, and α_ℓ is the complex gain of the ℓ^{th} path. The path amplitudes are assumed to be Rayleigh distributed, i.e., $\alpha_\ell \sim \mathcal{N}\left(0, \bar{P}_R\right), \ell = 1, 2, \ldots, L$ with $\bar{P}_R$ the average power gain. The variables $\phi_\ell \in [0, 2\pi]$ and $\theta_\ell \in [0, 2\pi]$ are the ℓ^{th} path's azimuth angles of departure or arrival (AoDs/AoAs) of the BS and MS, respectively. Considering only the azimuth, and neglecting elevation, implies that all scattering happens in azimuth and that the BS and MS implement horizontal (2-D) beamforming only. Extensions to 3-D beamforming are possible [13]. Finally, $\mathbf{a}_{\mathrm{BS}}\left(\phi_\ell\right)$ and $\mathbf{a}_{\mathrm{MS}}\left(\theta_\ell\right)$ are the antenna array response vectors at the BS and MS, respectively. While the algorithms and results developed in the paper can be applied to arbitrary antenna arrays, we use uniform linear arrays (ULAs), in the simulations of Section VII. If a ULA is assumed, $\mathbf{a}_{\mathrm{BS}}\left(\phi_\ell\right)$ can be written as

$$\mathbf{a}_{\mathrm{BS}}\left(\phi_\ell\right) = \frac{1}{\sqrt{N_{\mathrm{BS}}}}\Big[1, e^{j(2\pi/\lambda)d\sin(\phi_\ell)}, \ldots \times \ldots, e^{j(N_{\mathrm{BS}}-1)(2\pi/\lambda)d\sin(\phi_\ell)}\Big]^{T}, \tag{5}$$

where λ is the signal wavelength, and d is the distance between antenna elements. The array response vectors at the MS, $\mathbf{a}_{\mathrm{MS}}\left(\theta_\ell\right)$, can be written in a similar fashion.

The channel in (4) is written in a more compact form as

$$\mathbf{H} = \mathbf{A}_{\mathrm{MS}}\mathrm{diag}\left(\boldsymbol{\alpha}\right)\mathbf{A}_{\mathrm{BS}}^{H}, \tag{6}$$

where $\boldsymbol{\alpha} = \sqrt{N_{\mathrm{BS}}N_{\mathrm{MS}}/\rho}\left[\alpha_1, \alpha_2, \ldots, \alpha_L\right]^{T}$. The matrices

$$\mathbf{A}_{\mathrm{BS}} = \left[\mathbf{a}_{\mathrm{BS}}\left(\phi_1\right), \mathbf{a}_{\mathrm{BS}}\left(\phi_2\right), \ldots, \mathbf{a}_{\mathrm{BS}}\left(\phi_L\right)\right], \tag{7}$$

and

$$\mathbf{A}_{\mathrm{MS}} = \left[\mathbf{a}_{\mathrm{MS}}\left(\theta_1\right), \mathbf{a}_{\mathrm{MS}}\left(\theta_2\right), \ldots, \mathbf{a}_{\mathrm{MS}}\left(\theta_L\right)\right], \tag{8}$$

contain the BS and MS array response vectors.

In this paper, we assume that both the BS and MS have no a priori knowledge of the channel. Hence, in the first part of the paper, namely, Sections III–V, the mmWave channel estimation

problem is formulated, and an adaptive CS based algorithm is developed and employed at the BS and MS to solve it. In the second part, i.e., Section VI, the estimated channel is used to construct the hybrid precoding and decoding matrices.

III. Formulation of the mmWave Channel Estimation Problem

Given the geometric mmWave channel model in (4), estimating the mmWave channel is equivalent to estimating the different parameters of the L channel paths; namely the AoA, the AoD, and the gain of each path. To do that accurately and with low training overhead, the BS and MS need to carefully design their training precoders and combiners. In this section, we exploit the poor scattering nature of the mmWave channel, and formulate the mmWave channel estimation problem as a sparse problem. We will also briefly show how adaptive CS work invokes some ideas for the design of the training precoders and combiners. Inspired by these ideas, and using the hybrid analog/digital system architecture, we will develop a novel hierarchical multi-resolution codebook for the training beamforming vectors in Section IV. We will then propose algorithms that adaptively use the developed codebook to estimate the mmWave channel along with evaluating their performance in Section V.

A. A Sparse Formulation of the mmWave Channel Estimation Problem

Consider the system and mmWave channel models described in Section II. If the BS uses a beamforming vector $\mathbf{f}_p$, and the MS employs a measurement vector $\mathbf{w}_q$ to combine the received signal, the resulting signal can be written as

$$y_{q,p} = \mathbf{w}_q^H \mathbf{H} \mathbf{f}_p s_p + \mathbf{w}_q^H \mathbf{n}_{q,p}, \tag{9}$$

where s_p is the transmitted symbol on the beamforming vector $\mathbf{f}_p$, such that $\mathbb{E}\left[s_p s_p^H\right] = P$, with P the average power used per transmission in the training phase. In Section IV, we will develop a hybrid analog/digital design for the beamforming/measurement vectors, $\mathbf{f}_p$ and $\mathbf{w}_q$. If M_{MS} such measurements are performed by the MS vectors $\mathbf{w}_q, q = 1, 2, \ldots, M_{\mathrm{MS}}$ at M_{MS} successive instants to detect the signal transmitted over the beamforming vector $\mathbf{f}_p$, the resulting vector will be

$$\mathbf{y}_p = \mathbf{W}^H \mathbf{H} \mathbf{f}_p s_p + \mathrm{diag}\big(\mathbf{W}^H \left[\mathbf{n}_{1,p}, \ldots, \mathbf{n}_{M_{\mathrm{MS}},p}\right]\big), \tag{10}$$

where $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_{M_{\mathrm{MS}}}]$ is the $N_{\mathrm{MS}} \times M_{\mathrm{MS}}$ measurement matrix. If the BS employs M_{BS} such beamforming vectors $\mathbf{f}_p, p = 1, \ldots, M_{\mathrm{BS}}$, at M_{BS} successive time slots, and the MS uses the same measurement matrix $\mathbf{W}$ to combine the received signal, the resultant matrix can then be written by concatenating the M_{BS} processed vectors $\mathbf{y}_p, p = 1, 2, \ldots, M_{\mathrm{BS}}$

$$\mathbf{Y} = \mathbf{W}^H \mathbf{H} \mathbf{F} \mathbf{S} + \mathbf{Q}, \tag{11}$$

where $\mathbf{F} = [\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_{M_{\mathrm{BS}}}]$ is the $N_{\mathrm{BS}} \times M_{\mathrm{BS}}$ beamforming matrix used by the BS, and $\mathbf{Q}$ is an $M_{\mathrm{MS}} \times M_{\mathrm{BS}}$ noise matrix given by concatenating the M_{BS} noise vectors. The matrix $\mathbf{S}$ is a diagonal matrix carrying the M_{BS} transmitted symbols $s_p, p = 1, \ldots, M_{\mathrm{BS}}$ on its diagonal. For the training phase, we assume that all transmitted symbols are equal, namely, $\mathbf{S} = \sqrt{P}\mathbf{I}_{M_{\mathrm{BS}}}$ and therefore

$$\mathbf{Y} = \sqrt{P}\mathbf{W}^H \mathbf{H} \mathbf{F} + \mathbf{Q}. \tag{12}$$

To exploit the sparse nature of the channel, we first vectorize the resultant matrix $\mathbf{Y}$

$$\mathbf{y}_v = \sqrt{P}\mathrm{vec}\left(\mathbf{W}^H \mathbf{H} \mathbf{F}\right) + \mathrm{vec}\left(\mathbf{Q}\right) \tag{13}$$

$$\overset{(a)}{=} \sqrt{P}\left(\mathbf{F}^T \otimes \mathbf{W}^H\right)\mathrm{vec}\left(\mathbf{H}\right) + \mathbf{n}_{\mathrm{Q}} \tag{14}$$

$$\overset{(b)}{=} \sqrt{P}\left(\mathbf{F}^T \otimes \mathbf{W}^H\right)\left(\mathbf{A}_{\mathrm{BS}}^* \circ \mathbf{A}_{\mathrm{MS}}\right)\boldsymbol{\alpha} + \mathbf{n}_{\mathrm{Q}}, \tag{15}$$

where (a) follows from [25, Theorem 13.26], (b) follows from the channel model in (6), and the properties of the Khatri-Rao product, [21]. The matrix $\left(\mathbf{A}_{\mathrm{BS}}^* \circ \mathbf{A}_{\mathrm{MS}}\right)$ is an $N_{\mathrm{BS}}N_{\mathrm{MS}} \times L$ matrix in which each column has the form $\left(\mathbf{a}_{\mathrm{BS}}^*\left(\phi_\ell\right) \otimes \mathbf{a}_{\mathrm{MS}}\left(\theta_\ell\right)\right), \ell = 1, 2, \ldots, L$, i.e., each column ℓ represents the Kronecker product of the BS and MS array response vectors associated with the AoA/AoD of the ℓth path of the channel.

To complete the problem formulation, we assume that the AoAs, and AoDs are taken from a uniform grid of N points, with $N \gg L$, i.e., we assume that $\phi_\ell, \theta_\ell \in \{0, 2\pi/N, \ldots, 2\pi(N-1)/N\}, \ell = 1, 2, \ldots, L$. As the values of the AoAs/AoDs are actually continuous, other off-grid based algorithms like sparse regularized total least squared [26], continuous basis pursuit [27], or Newton refinement ideas [28] can be incorporated to reduce the quantization error. In this paper, we consider only the case of quantized AoAs/AoDs, leaving possible improvements for future work. We evaluate the impact of this quantization error on the performance of the proposed algorithms in this paper by numerical simulations in Section VII.

By neglecting the grid quantization error, we can approximate $\mathbf{y}_v$ in (15) as

$$\mathbf{y}_v = \sqrt{P}\left(\mathbf{F}^T \otimes \mathbf{W}^H\right)\mathbf{A}_{\mathrm{D}}\mathbf{z} + \mathbf{n}_{\mathrm{Q}}, \tag{16}$$

where $\mathbf{A}_{\mathrm{D}}$ is an $N_{\mathrm{BS}}N_{\mathrm{MS}} \times N^2$ dictionary matrix that consists of the N^2 column vectors of the form $\left(\mathbf{a}_{\mathrm{BS}}^*\left(\bar{\phi}_u\right) \otimes \mathbf{a}_{\mathrm{MS}}\left(\bar{\theta}_v\right)\right)$, with $\bar{\phi}_u$, and $\bar{\theta}_v$ the uth, and vth points, respectively, of the angles uniform grid, i.e., $\bar{\phi}_u = 2\pi u/N, u = 0, 2, \ldots, N-1$, and $\bar{\theta}_v = 2\pi v/N, v = 0, 2, \ldots, N-1$. $\mathbf{z}$ is an $N^2 \times 1$ vector which carries the path gains of the corresponding quantized directions. Note that detecting the columns of $\mathbf{A}_{\mathrm{D}}$ that correspond to non-zero elements of $\mathbf{z}$, directly implies the detection of the AoAs and AoDs of the dominant paths of the channel. The path gains can be also determined by calculating the values of the corresponding elements in $\mathbf{z}$.

The formulation of the vectorized received signal $\mathbf{y}_v$ in (16) represents a sparse formulation of the channel estimation problem as $\mathbf{z}$ has only L non-zero elements and $L \ll N^2$. This implies that the number of required measurements, $M_{\mathrm{BS}}M_{\mathrm{MS}}$, to detect the non-zero elements of $\mathbf{z}$ is much less than N^2. Given this formulation in (16), CS tools can be leveraged to design estimation algorithms to determine the quantized AoAs/AoDs. If we define the sensing matrix $\boldsymbol{\Psi}$ as $\boldsymbol{\Psi} = \left(\mathbf{F}^T \otimes \mathbf{W}^H\right)\mathbf{A}_{\mathrm{D}}$, the objective of the CS algorithms will be to efficiently design

this sensing matrix to guarantee the recovery of the non-zero elements of the vector $\mathbf{z}$ with high probability, and with a small number of measurements. One common criterion for that is the restricted isometry property (RIP), which requires the matrix $\boldsymbol{\Psi}^H\boldsymbol{\Psi}$ to be close to diagonal on average [29].

To simplify the explanation of the BS-MS beamforming vectors' design problem in the later chapters, we prefer to use the Kronecker product properties and write (16) as [21]

$$\mathbf{y}_v = \sqrt{P}\left(\mathbf{F}^T\mathbf{A}^*_{\mathrm{BS,D}} \otimes \mathbf{W}^H\mathbf{A}_{\mathrm{MS,D}}\right)\mathbf{z} + \mathbf{n}_{\mathrm{Q}} \quad (17)$$

$$= \sqrt{P}\mathbf{F}^T\mathbf{A}^*_{\mathrm{BS,D}}\mathbf{z}_{\mathrm{BS}} \otimes \mathbf{W}^H\mathbf{A}_{\mathrm{MS,D}}\mathbf{z}_{\mathrm{MS}} + \mathbf{n}_{\mathrm{Q}}, \quad (18)$$

where $\mathbf{z}_{\mathrm{BS}}$, and $\mathbf{z}_{\mathrm{MS}}$ are two $N \times 1$ sparse vectors that have non-zero elements in the locations that correspond to the AoDs, and AoAs, respectively. $\mathbf{A}_{\mathrm{BS,D}}$, and $\mathbf{A}_{\mathrm{MS,D}}$ are $N_{\mathrm{BS}} \times N$, and $N_{\mathrm{MS}} \times N$ dictionary matrices that consist of column vectors of the forms $\mathbf{a}_{\mathrm{BS}}\left(\bar{\phi}_u\right)$, and $\mathbf{a}_{\mathrm{MS}}\left(\bar{\theta}_u\right)$, respectively.

In the standard CS theory, the number of measurement vectors required to guarantee the recovery of the L-sparse vector with high probability is of order $\mathcal{O}(L\log(N/L))$ [30]. While these results are theoretically proved, their implementations to specific applications, and the development of efficient algorithms, require further work. We therefore resort to adaptive CS tools which invoke some ideas for the design of the training beamforming vectors.

B. Adaptive Compressed Sensing Solution

In adaptive CS [31]–[33], the training process is divided into a number of stages. The training precoding, and measurement matrices used at each stage are not determined a priori, but rather depend on the output of the earlier stages. More specifically, if the training process is divided into S stages, then the vectorized received signals of these stages are

$$\begin{aligned}
\mathbf{y}_{(1)} &= \sqrt{P_{(1)}}\left(\mathbf{F}^T_{(1)}\mathbf{A}^*_{\mathrm{BS,D}} \otimes \mathbf{W}^H_{(1)}\mathbf{A}_{\mathrm{MS,D}}\right)\mathbf{z} + \mathbf{n}_1 \\
\mathbf{y}_{(2)} &= \sqrt{P_{(2)}}\left(\mathbf{F}^T_{(2)}\mathbf{A}^*_{\mathrm{BS,D}} \otimes \mathbf{W}^H_{(2)}\mathbf{A}_{\mathrm{MS,D}}\right)\mathbf{z} + \mathbf{n}_2 \\
&\vdots \\
\mathbf{y}_{(\mathrm{S})} &= \sqrt{P_{(\mathrm{S})}}\left(\mathbf{F}^T_{(\mathrm{S})}\mathbf{A}^*_{\mathrm{BS,D}} \otimes \mathbf{W}^H_{(\mathrm{S})}\mathbf{A}_{\mathrm{MS,D}}\right)\mathbf{z} + \mathbf{n}_{\mathrm{S}} \quad (19)
\end{aligned}$$

The design of the sth stage training precoders and combiners, $\mathbf{F}_{(s)}, \mathbf{W}_{(s)}$, depends on $\mathbf{y}_{(1)}, \mathbf{y}_{(2)}, \ldots, \mathbf{y}_{(s-1)}$. Recent research in [31]–[33] shows that adaptive CS algorithms yield better performance than standard CS tools at low SNR, which is the typical case at mmWave systems before beamforming. Moreover, these adaptive CS ideas that rely on successive bisections provide important insights that can be used in the design of the training beamforming vectors.

In our proposed channel estimation algorithm described in Section V, the training beamforming vectors are adaptively designed based on the bisection concept. In particular, the algorithm starts initially by dividing the vector $\mathbf{z}$ in (19) into a number of partitions, which equivalently divides the AoAs/AoDs range into a number of intervals, and design the training precoding and combining matrices of the first stage, $\mathbf{F}_{(1)}, \mathbf{W}_{(1)}$, to sense those partitions. The received signal $\mathbf{y}_{(1)}$ is then used to determine the partition(s) that are highly likely to have non-zero element(s) which are further divided into smaller partitions in the later stages until detecting the non-zero elements, the AoAs/AoDs, with the required resolution. If the number of BS precoding vectors used in each stage of the adaptive algorithm equals K, where K is a design parameter, then the number of adaptive stages needed to detect the AoAs/AoDs with a resolution $2\pi/N$ is $\mathrm{S} = \log_K N$, which we assume to be integer for ease of exposition. Before delving into the details of the algorithm, we will focus in the following section on the design of a multi-resolution beamforming codebook which is essential for the proper operation of the adaptive channel estimation algorithm.

IV. Hybrid Precoding Based Multi-Resolution Hierarchical Codebook

In this section, we present a novel hybrid analog/digital based approach for the design of a multi-resolution beamforming codebook. Besides considering the RF limitations, namely, the constant amplitude phase shifters with quantized phases, the proposed approach for constructing the beamforming vectors is general for ULAs/non-ULAs, has a very-low complexity, and outperforms the analog-only beamforming codebooks thanks to its additional digital processing layer.

The design of a multi-resolution or variant beamwidth beamforming vector codebook has been studied before in [7]–[9], [20], [21]. This prior work focused on analog-only beamforming vectors, and on the physical design of the beam patterns. Unfortunately, the design of analog-only multi-resolution codebooks is subject to practical limitations in mmWave. (1) The existence of quantized phase shifters makes the design of non-overlapping beam patterns difficult, and may require an exhaustive search over a large space given the large number of antennas. (2) The design of analog-only beamforming vectors with certain beamwidths relies mostly on the beamsteering beam patterns of ULAs, and is hard to apply for non-ULAs due to the lack of intuition about their beam patterns.

To simplify explaining the codebook structure and design, we focus on the design of the BS training precoding codebook $\mathcal{F}$; a similar approach can be followed to construct the MS training codebook $\mathcal{W}$.

A. Codebook Structure

The proposed hierarchical codebook consists of S levels, $\mathcal{F}_s, s = 1, 2, \ldots, \mathrm{S}$. Each level contains beamforming vectors with a certain beamwidth to be used in the corresponding training stage of the adaptive mmWave channel estimation algorithm. Fig. 3 shows the first three levels of an example codebook with $N = 256$, and $K = 2$, and Fig. 4 illustrates the beam patterns of the beamforming vectors of each codebook level.

In each codebook level s, the beamforming vectors are divided into K^{s-1} subsets, with K beamforming vectors in each of them. Each subset k, of the codebook level s is associated with a unique range of the AoDs equal to $\{2\pi u/N\}_{u\in\mathcal{I}_{(s,k)}}$, where $\mathcal{I}_{(s,k)} = \left\{(k-1)N/K^{s-1}, \ldots, kN/K^{s-1}\right\}$. This AoD range is further divided into K sub-ranges, and each of the K beamforming vectors in this subset is designed so as to have an almost equal projection on the vectors $\mathbf{a}_{\mathrm{BS}}\left(\bar{\phi}_u\right)$, with u in this

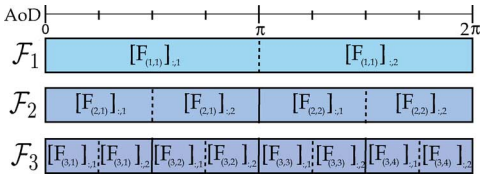

Fig. 3. An example of the structure of a multi-resolution codebook with a resolution parameter $N = 8$, and with $K = 2$ beamforming vectors in each subset.

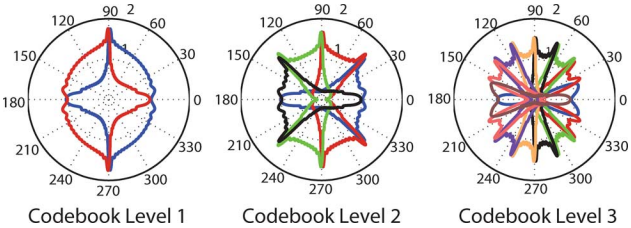

Fig. 4. The resulting beam patterns of the beamforming vectors in the first three codebook levels of an example hierarchical codebook.

sub-range, and zero projection on the other vectors. Physically, this implies the implementation of a beamforming vector with a certain beamwidth determined by these sub-ranges, and steered in pre-defined directions.

While the proposed codebook structure is similar to the codebooks in [7], [9], which also have multiple levels, each with beamforming vectors with a certain beamwidth, we adopt a different way for defining each beamforming vector in terms of the set of *quantized* angles that it covers. This is different from the previous work in [7], [9] which defined each vector by the central beamforming angle and the beamwidth. This difference leads to a novel formulation of the arbitrary beamwidth beamforming design problem in addition to a completely new way for realizing these vectors using analog/digital architecture as will be explained shortly. To the best of our knowledge, this is the first work that considers implementing beamforming vectors of different beamwidth using joint analog/digital processing; as the previous work relied on analog-only designs. This additional digital processing layer adds more degrees of freedom to the beamforming design problem which can be leveraged to obtain better characteristics in the beamforming patterns.

B. Design of the Codebook Beamforming Vectors

In each codebook level s, and subset k, the beamforming vectors $\left[\mathbf{F}_{(s,k)}\right]_{:,m}, m = 1, 2, \ldots, K$ are designed such that

$$\left[\mathbf{F}_{(s,k)}\right]_{:,m}^{H} \mathbf{a}_{\mathrm{BS}}\left(\bar{\phi}_u\right) = \begin{cases} C_s & \text{if } u \in \mathcal{I}_{(s,k,m)} \\ 0 & \text{if } u \notin \mathcal{I}_{(s,k,m)} \end{cases}, \qquad (20)$$

where

$$\mathcal{I}_{(k,s,m)} = \left\{ \frac{N}{K^s}\left(K(k-1)+m-1\right)+1, \ldots, \frac{N}{K^s}\left(K(k-1)+m\right) \right\}$$

defines the sub-range of AoDs associated with the beamforming vector $\left[\mathbf{F}_{(s,k)}\right]_{:,m}$, and C_s is a normalization constant that satisfies $\|\mathbf{F}_{(s,k)}\|_F = K$. For example, the beamforming vector $\left[\mathbf{F}_{(2,1)}\right]_{:,1}$ in Fig. 3 is designed such that it has a constant projection on the array response vectors $\mathbf{a}_{\mathrm{BS}}\left(\bar{\phi}_u\right)$, u is in $\{0, 1, \ldots, 63\}$, i.e., $\bar{\phi}_u$ is in $\{0, \ldots, 2\pi 63/256\}$, and zero projection on the other directions.

In a more compact form, we can write the design objective of the beamforming vectors $\mathbf{F}_{(s,k)}$ in (20) as the solution of

$$\mathbf{A}_{\mathrm{BS,D}}^{H}\mathbf{F}_{(s,k)} = C_s\mathbf{G}_{(s,k)}, \qquad (21)$$

where $\mathbf{G}_{(s,k)}$ is an $N \times K$ matrix where each column m containing $1's$ in the locations $u, u \in \mathcal{I}_{(s,k,m)}$, and zeros in the locations $u, u \notin \mathcal{I}_{(s,k,m)}$. Now, we note that the BS AoDs matrix $\mathbf{A}_{\mathrm{BS,D}}$ is an over-complete dictionary with $N \geq N_{\mathrm{BS}}$, i.e., (21) represents an inconsistent system of which the approximate solution is given by $\mathbf{F}_{(s,k)} = C_s(\mathbf{A}_{\mathrm{BS,D}}\mathbf{A}_{\mathrm{BS,D}}^{H})^{-1}\mathbf{A}_{\mathrm{BS,D}}\mathbf{G}_{(s,k)}$. Further, given the available system model in Section II, the precoding matrix $\mathbf{F}_{(s,k)}$ is defined as $\mathbf{F}_{(s,k)} = \mathbf{F}_{\mathrm{RF},(s,k)}\mathbf{F}_{\mathrm{BS},(s,k)}$. As each beamforming vector will be individually used in a certain time instant, we will design each of them independently in terms of the hybrid analog/digital precoders. Consequently, the design of the hybrid analog and digital training precoding matrices is accomplished by solving

$$\begin{aligned} \left\{\mathbf{F}_{\mathrm{RF},(s,k)}^{\star}, \left[\mathbf{F}_{\mathrm{BB},(s,k)}^{\star}\right]_{:,m}\right\} = \\ \arg\min \ \ & \| \left[\mathbf{F}_{(s,k)}\right]_{:,m} - \mathbf{F}_{\mathrm{RF},(s,k)} \left[\mathbf{F}_{\mathrm{BB},(s,k)}\right]_{:,m} \|_F, \\ \text{s.t.} \ \ & \left[\mathbf{F}_{\mathrm{RF},(s,k)}\right]_{:,i} \in \left\{ [\mathbf{A}_{\mathrm{can}}]_{:,\ell} \,|\, 1 \leq \ell \leq N_{\mathrm{can}} \right\}, \\ & i = 1, 2, \ldots, N_{\mathrm{RF}} \\ & \|\mathbf{F}_{\mathrm{RF},(s,k)} \left[\mathbf{F}_{\mathrm{BB},(s,k)}\right]_{:,m} \|_F^2 = 1, \end{aligned} \qquad (22)$$

where $\left[\mathbf{F}_{(s,k)}\right]_{:,m} = C_s(\mathbf{A}_{\mathrm{BS,D}}\mathbf{A}_{\mathrm{BS,D}}^{H})^{-1}\mathbf{A}_{\mathrm{BS,D}} \left[\mathbf{G}_{(s,k)}\right]_{:,m}$, and $\mathbf{A}_{\mathrm{can}}$ is an $N_{\mathrm{BS}} \times N_{\mathrm{can}}$ matrix which carries the finite set of possible analog beamforming vectors. The columns of the candidate matrix $\mathbf{A}_{\mathrm{can}}$ can be chosen to satisfy arbitrary analog beamforming constraints. Two example candidate beamformer designs we consider in the simulations of Section VII are summarized as follows.

1) Equally spaced ULA beam steering vectors [13], i.e., a set of N_{can} vectors of the form $\mathbf{a}_{\mathrm{BS}}(t_{\mathrm{can}}\pi/N)$ for $t_{\mathrm{can}} = 0, 1, 2, \ldots, N_{\mathrm{can}} - 1$.
2) Beamforming vectors whose elements can be represented as quantized phase shifts. In the case of quantized phase shifts, if each phase shifter is controlled by an N_{Q}-bit input, the entries of the candidate precoding matrix $\mathbf{A}_{\mathrm{can}}$ can all be written as $e^{jk_{\mathrm{Q}}2\pi/2^{N_{\mathrm{Q}}}}$ for some $k_{\mathrm{Q}} = 0, 1, 2, \ldots, 2^{N_{\mathrm{Q}}} - 1$.

Now, given the matrix of possible analog beamforming vectors $\mathbf{A}_{\mathrm{can}}$, the optimization problem in (22) can be reformulated as a sparse approximation problem similar to the optimization problem in [13, equation (17)], with the matrices $\tilde{\mathbf{F}}_{\mathrm{BB}}^{\mathrm{opt}}$, $\mathbf{F}_{\mathrm{opt}}$, $\mathbf{F}_{\mathrm{BB}}$ and $\mathbf{A}_{\mathrm{t}}$ in [13, equation (17)] taking the values $\left[\mathbf{F}_{\mathrm{BB},(s,k)}^{\star}\right]_{:,m}$, $\left[\mathbf{F}_{(s,k)}\right]_{:,m}$, $\left[\mathbf{F}_{\mathrm{BB},(s,k)}\right]_{:,m}$, and $\mathbf{A}_{\mathrm{can}}$, respectively, and with setting $N_s = 1$. This sparse problem can be then solved using Algorithm 1 in [13] which is a variant of orthogonal matching pursuit algorithms.

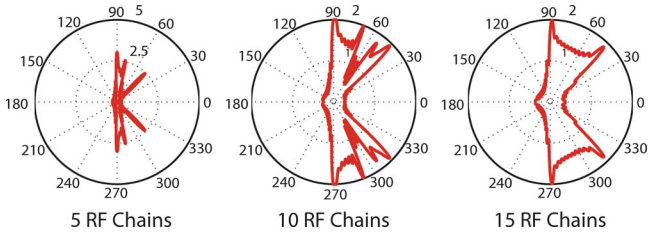

Fig. 5. Beam patterns approximation of one of the beamforming vectors in the second codebook level with different numbers of RF chains.

An example of the beam patterns resulting from applying the proposed algorithm is shown in Fig. 5. These patterns are generated by a BS has 32 antennas, and a number of RF chains $N_{\mathrm{RF}} = 5, 10, 15$ to approximate one of the unconstrained beamforming vectors in the second codebook level shown in Fig. 4.

After the design of the BS training beamforming vectors for the kth subset of the sth codebook, the following quantities are calculated, as they will be used after that in the channel estimation algorithm in Section V:

- **Beamforming Gain**: Given the channel model in (4), and the codebook beamforming design criteria in (20), we define the beamforming gain of the BS training vectors at the sth stage as $G^{\mathrm{BS}}_{(s)} = N_{\mathrm{BS}} C_s^2$. A similar definition can be used for the MS beamforming vectors, yielding a total training beamforming gain at the sth stage equal to $G^{(s)} = G^{\mathrm{BS}}_{(s)} G^{\mathrm{MS}}_{(s)}$.
- **Error Matrix**: As the system in (21) is inconsistent, the solution given by the pseudo-inverse means that $\mathbf{A}^H_{\mathrm{BS,D}} \mathbf{F}_{(s,k)}$ may not be exactly equal to $C_s \mathbf{G}_{(s,k)}$. Moreover, the limitations of the RF beamforming vectors and the approximate solution of the optimization problem in (22) result in an additional error in satisfying (21). This error physically means (i) the existence of a spectral leakage of the beamforming vectors outside their supposed AoD sub-ranges, and (ii) the beamforming gain is not exactly uniform over the desired AoD ranges. To take the effect of this error into the performance analysis of the proposed channel estimation algorithm in Section V, we define the error matrix of each subset k of the sth BS beamforming codebook level as

$$\mathbf{E}^{\mathrm{BS}}_{(s,k)} = \mathbf{A}^H_{\mathrm{BS,D}} \mathbf{F}_{(s,k)} - C_s \mathbf{G}_{(s,k)}. \tag{23}$$

As a similar error exists in the MS combining codebook, we can define the overall beamforming error matrix experienced by the received vector $\mathbf{y}_s$ in (18) after applying the Kronecker product as

$$\begin{aligned}\mathbf{E}^{(s)}_{k_{\mathrm{BS}},k_{\mathrm{MS}}} &= {\mathbf{E}^{\mathrm{BS}}_{(s,k_{\mathrm{BS}})}}^T \otimes {\mathbf{E}^{\mathrm{MS}}_{(s,k_{\mathrm{MS}})}}^T + {\mathbf{E}^{\mathrm{BS}}_{(s,k_{\mathrm{BS}})}}^T \\ &\otimes C^{\mathrm{MS}}_s \mathbf{G}^T_{(s,k_{\mathrm{MS}})} + C^{\mathrm{BS}}_s \mathbf{G}^T_{(s,k_{\mathrm{BS}})} \otimes {\mathbf{E}^{\mathrm{MS}}_{(s,k_{\mathrm{MS}})}}^T. \end{aligned} \tag{24}$$

Now, if we also define the overall gain matrix as $\mathbf{G}^{(s)}_{k_{\mathrm{BS}},k_{\mathrm{MS}}} = \mathbf{G}_{(s,k_{\mathrm{BS}})} \otimes \mathbf{G}_{(s,k_{\mathrm{MS}})}$, then we can rewrite $\mathbf{y}_s$ in (19) as follows assuming the subsets k_{BS} of $\mathcal{F}_s$, and k_{MS} of $\mathcal{W}_s$ are used at the BS and MS

$$\mathbf{y}_{(s)} = \sqrt{P_{(s)}} \left(\mathbf{F}^T_{(s)} \otimes \mathbf{W}^H_{(s)} \right) \mathbf{A}_{\mathrm{D}} \mathbf{z} + \mathbf{n}_{\mathrm{Q}} \tag{25}$$

$$= \sqrt{P_{(s)}} \left(\mathbf{F}^T_{(s)} \mathbf{A}_{\mathrm{BS,D}} \otimes \mathbf{W}^H_{(s)} \mathbf{A}_{\mathrm{MS,D}} \right) \mathbf{z} + \mathbf{n}_{\mathrm{Q}} \tag{26}$$

$$= \sqrt{P_{(s)}} \left(\sqrt{\frac{G_{(s)}}{N_{\mathrm{BS}} N_{\mathrm{MS}}}} \mathbf{G}^{(s)}_{k_{\mathrm{BS}},k_{\mathrm{MS}}} + \mathbf{E}^{(s)}_{k_{\mathrm{BS}},k_{\mathrm{MS}}} \right) \mathbf{z} + \mathbf{n}_{\mathrm{Q}}. \tag{27}$$

- **Actual Beamforming Gain**: To include the effect of the previously defined error matrix on the beamforming gain, we will define the actual forward and backward gains of the designed beamforming vectors. First, note that $\left(\mathbf{G}^{(s)}_{k_{\mathrm{BS}},k_{\mathrm{MS}}} + \mathbf{E}^{(s)}_{k_{\mathrm{BS}},k_{\mathrm{MS}}} \right)$ is a $K^2 \times N^2$ matrix in which each row corresponds to a certain pair of the precoding/measurement vectors, and each column corresponds to a certain quantized AoA/AoD pair. We denote each AoA/AoD pair as a direction d, $d = 1, 2, \ldots, N^2$, and the set of directions covered by each pair m of the precoder/measurement vectors as $\mathcal{D}(m), m = 1, 2, \ldots, K^2$. Now, if d is one of the directions covered by the joint beamforming vector m, i.e., $d \in \mathcal{D}(m)$, then the actual beamforming gain of the vector m in the direction d can be defined as

$$\hat{G}^{(s)}_d = \left| \left[\sqrt{G^{(s)}} \mathbf{G}^{(s)}_{k_{\mathrm{BS}},k_{\mathrm{MS}}} + \sqrt{N_{\mathrm{BS}} N_{\mathrm{MS}}} \mathbf{E}^{(s)}_{k_{\mathrm{BS}},k_{\mathrm{MS}}} \right]_{m,d} \right|^2, \tag{28}$$

$$\overset{(a)}{=} \left| \sqrt{G_{(s)}} + \sqrt{N_{\mathrm{BS}} N_{\mathrm{MS}}} \left[\mathbf{E}^{(s)}_{k_{\mathrm{BS}},k_{\mathrm{MS}}} \right]_{m(d),d} \right|^2, \tag{29}$$

where (a) is by noting that $\left[\mathbf{G}^{(s)}_{k_{\mathrm{BS}},k_{\mathrm{MS}}} \right]_{m,d} = 1$ if $d \in \mathcal{D}(m)$. We can also define the side lobe gain introduced by each one of the other precoding/measurement vectors, $\bar{m}$, in the direction d as $\Delta^{(s)}_{\bar{m},d} = N_{\mathrm{BS}} N_{\mathrm{MS}} \left| \left[\mathbf{E}^{(s)}_{k_{\mathrm{BS}},k_{\mathrm{MS}}} \right]_{\bar{m},d} \right|^2$, $\bar{m} = 1, 2, \ldots, K^2, \bar{m} \neq m$. To simplify the notation, we omitted the beamforming subset symbols $k_{\mathrm{BS}}, k_{\mathrm{MS}}$ from the previous definition; as each direction d is covered by only one beamforming/combining vector on each codebook level.

In the analysis of the proposed channel estimation algorithm in Section V, we will be interested in the ratio between the previously defined gains. Hence, we define the ratio between the actual forward gain in a certain direction d and the side lobe gain, in the same direction, induced by beamforming vector $\bar{m}$ as

$$\beta^{(s)}_{d,\bar{m}} = \frac{\hat{G}^{(s)}_d}{\Delta^{(s)}_{d,\bar{m}}}. \tag{30}$$

V. Adaptive Estimation Algorithms for mmWave Channels

In this section, we consider the sparse channel estimation problem formulated in (19) of Section III, and propose algorithms that adaptively use the hierarchical codebook developed in Section IV to estimate the mmWave channel. We firstly address this problem for the rank-one channel model, i.e., when the channel has only one-path, in Section V-A. We then extend the proposed algorithm for the multi-path case in Section V-B.

A. Adaptive Channel Estimation Algorithm for Single-Path mmWave Channels

Given the problem formulation in (19), the single-path channel implies that the vector $\mathbf{z}$ has only one non-zero element. Hence, estimating the single-path channel is accomplished by determining the location of this non-zero element, which in turn defines the AoA/AoD, and the value of this element, which decides the channel path gain. To efficiently do that with low training overhead, we propose Algorithm 1 which adaptively searches for the non-zero element of $\mathbf{z}$ by using the multi-resolution beamforming vectors designed in Section IV.

Algorithm 1 Adaptive Estimation Algorithm for Single-Path mmWave Channels

Input: BS and MS know N, K, and have $\mathcal{F}, \mathcal{W}$.

Initialization: $k_1^{\mathrm{BS}} = 1, k_1^{\mathrm{MS}} = 1$

// Initialize the subsets to be used of codebooks $\mathcal{F}, \mathcal{W}$

$\mathrm{S} = \log_K N$ // The number of adaptive stages

for $s \leq \mathrm{S}$ **do**

  for $m_{\mathrm{BS}} \leq K$ **do**

    BS uses $\left[\mathbf{F}_{(s,k_s^{\mathrm{BS}})}\right]_{:,m_{\mathrm{BS}}}$

    for $m_{\mathrm{MS}} \leq K$ **do**

      MS uses $\left[\mathbf{W}_{(s,k_s^{\mathrm{MS}})}\right]_{:,m_{\mathrm{MS}}}$

    After MS measurements:

    $\mathbf{y}_{m_{\mathrm{BS}}} = \sqrt{P_s}\left[\mathbf{W}_{(s,k_s^{\mathrm{MS}})}\right]\mathbf{H}\left[\mathbf{F}_{(s,k_s^{\mathrm{BS}})}\right]_{:,m_{\mathrm{BS}}} + \mathbf{n}_{m_{\mathrm{BS}}}$

  $\mathbf{Y}_{(s)} = [\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_K]$

  $(m_{\mathrm{BS}}^\star, m_{\mathrm{MS}}^\star) = \underset{\forall m_{\mathrm{BS}}, m_{\mathrm{MS}}=1,2,\ldots,K}{\arg\max} \left[\mathbf{Y}_{(s)} \odot \mathbf{Y}_{(s)}^*\right]_{m_{\mathrm{MS}}, m_{\mathrm{BS}}}$

  $k_{s+1}^{\mathrm{BS}} = K(m_{\mathrm{BS}}^\star - 1) + 1, k_{s+1}^{\mathrm{MS}} = K(m_{\mathrm{MS}}^\star - 1) + 1$

$\hat{\phi} = \bar{\phi}_{k_{\mathrm{S}+1}^{\mathrm{BS}}}, \hat{\theta} = \bar{\theta}_{k_{\mathrm{S}+1}^{\mathrm{MS}}}$

$\hat{\alpha} = \sqrt{\rho / P_{(\mathrm{S})} G_{(\mathrm{S})}}\left[\mathbf{Y}_{(\mathrm{S})}\right]_{m_{\mathrm{MS}}^\star, m_{\mathrm{BS}}^\star}$

Algorithm 1 operates as follows. In the initial stage, the BS uses the K training precoding vectors of the first level of the codebook $\mathcal{F}$ in Section IV. For each of those vectors, the MS uses the K measurement vectors of the first level of $\mathcal{W}$ to combine the received signal. Note that the first level of the hierarchical codebook in Section IV has only one subset of beamforming vectors. After the K^2 precoding-measurement steps of this stage, the MS compares the power of the K^2 received signals to determine the one with the maximum received power. As each one of the precoding/measurement vectors is associated with a certain range of the quantized AoA/AoD, the operation of the first stage divides the vector $\mathbf{z}$ in (19) into K^2 partitions, and compares between the power of the sum of each of them. Hence, the selection of the maximum power received signal implies the selection of the partition of $\mathbf{z}$, and consequently the range of the quantized AoA/AoD, that is highly likely to contain the single path of the channel. The output of the maximum power problem is then used to determine the subsets of the beamforming vectors of level $s+1$ of $\mathcal{F}$, and $\mathcal{W}$ to be used in the next stage. The MS then feeds back the selected subset of the BS precoders to the BS to use it in the next stage, which needs only $\log_2 K$ bits. As the beamforming vectors of the next levels have higher and higher resolution, the AoA/AoD ranges are further refined adaptively as we proceed in the algorithm stages until the desired resolution, $2\pi/N$, is achieved. Note that the training powers in the S stages are generally different as will be discussed shortly.

Based on the proposed algorithm, the total number of stages required to estimate the AoA/AoD with a resolution $2\pi/N$ is $\log_K N$. Also, since we need K beamforming vectors, and K measurement vectors for each beamforming vector in each stage, the total number of steps needed to estimate the mmWave channel using the proposed algorithm becomes $K^2 \log_K N$ steps. Moreover, since N_{RF} RF chains can be simultaneously used at the MS to combine the measurements, the required number of steps can be further reduced to be $K\lceil K/N_{\mathrm{RF}}\rceil \log_K N$.

In the following theorem, we characterize the performance of the proposed algorithm for the case of single dominant path channels, i.e., assuming that the channel model in (4) has $L = 1$. We find an upper bound of the probability of error in estimating the AoA/AoD with a certain resolution using Algorithm 1. We will then use Theorem 1 to derive sufficient conditions on the total training power and its distribution over the adaptive stages of Algorithm 1 to guarantee estimating the AoA/AoD of the mmWave channel with a desired resolution, and a certain bound on the maximum error probability.

Theorem 1: Algorithm 1 succeeds in estimating the correct AoA and AoD of the single-path channel model in (4), for a desired resolution $2\pi/N$, with an average probability of error $\bar{p}$ which is upper bounded by

$$\bar{p} \leq \frac{K^2-1}{2N^2}\sum_{s=1}^{S}\sum_{d=1}^{N^2}\left(1 - \frac{1}{4} \times \frac{\left(1-\frac{1}{\overline{\beta}_d^{(s)}}\right)P_{(s)}\hat{G}_d^{(s)}\bar{\gamma}}{\sqrt{1+\left(1+\frac{1}{\overline{\beta}_d^{(s)}}\right)\frac{P_{(s)}\hat{G}_d^{(s)}\bar{\gamma}}{2} + \frac{P_{(s)}^2\hat{G}_d^{(s)^2}\bar{\gamma}^2}{16}\left(1-\frac{1}{\overline{\beta}_d^{(s)}}\right)^2}}\right), \tag{31}$$

where $\overline{\beta}_d^{(s)} = \min_{\substack{\forall \bar{m}=1,2,\ldots,K^2,\\ d \notin \mathcal{D}(\bar{m})}} \beta_{d,\bar{m}}^{(s)}$ is the minimum forward to backward ratio in the direction d, $\hat{G}_d^{(s)}$ is the actual beamforming gain in this direction, and $\bar{\gamma}$ is the average channel SNR defined as $\bar{\gamma} = \bar{P}_R/\rho\sigma^2$.

Proof: If the BS and MS use Algorithm 1 to estimate their AoA/AoD with a resolution $2\pi/N$, and employ K precoding and measurement vectors of codebooks $\mathcal{F}$ and $\mathcal{W}$ at each stage, then the output of the first stage can be written as (27)

$$\mathbf{y}_{(1)} = \sqrt{P_{(1)}}\left(\sqrt{\frac{G_{(1)}}{N_{\mathrm{BS}}N_{\mathrm{MS}}}}\mathbf{G}^{(1)}_{k_{\mathrm{BS}},k_{\mathrm{MS}}} + \mathbf{E}^{(1)}_{k_{\mathrm{BS}},k_{\mathrm{MS}}}\right)\mathbf{z} + \mathbf{n}_1 \tag{32}$$

$$= \sqrt{P_{(1)}} \times \begin{bmatrix} \sqrt{\frac{G_{(1)}}{N_{\mathrm{BS}}N_{\mathrm{MS}}}}\sum_{i=1}^{N^2/K^2}[x]_i + \sum_{i=1}^{N^2}\left[\mathbf{E}^{(1)}_{k_{\mathrm{BS}},k_{\mathrm{MS}}}\right]_{1,i}[x]_i \\ \vdots \\ \sqrt{\frac{G_{(1)}}{N_{\mathrm{BS}}N_{\mathrm{MS}}}}\sum_{i=(K^2-1)N^2/K^2+1}^{N^2}[x]_i + \sum_{i=1}^{N^2}\left[\mathbf{E}^{(1)}_{k_{\mathrm{BS}},k_{\mathrm{MS}}}\right]_{K^2,i}[x]_i \end{bmatrix} + \mathbf{n}_1. \tag{33}$$

Without loss of generality, if we assume that the single non-zero element of $\mathbf{z}$ is the dth element, and it is covered by the first precoding/measurement vector, then by the definition of $\boldsymbol{\alpha}$ in (6), we get

$$\mathbf{y}_{(1)} = \begin{bmatrix} \sqrt{\frac{P_{(1)}N_{\mathrm{BS}}N_{\mathrm{MS}}}{\rho}}\left(\sqrt{\frac{G_{(1)}}{N_{\mathrm{BS}}N_{\mathrm{MS}}}}+\left[\mathbf{E}^{(1)}_{k_{\mathrm{BS}},k_{\mathrm{MS}}}\right]_{1,d}\right)\alpha+n_1 \\ \sqrt{\frac{P_{(1)}N_{\mathrm{BS}}N_{\mathrm{MS}}}{\rho}}\left[\mathbf{E}^{(1)}_{k_{\mathrm{BS}},k_{\mathrm{MS}}}\right]_{2,d}\alpha+n_2 \\ \vdots \\ \sqrt{\frac{P_{(1)}N_{\mathrm{BS}}N_{\mathrm{MS}}}{\rho}}\left[\mathbf{E}^{(1)}_{k_{\mathrm{BS}},k_{\mathrm{MS}}}\right]_{K^2,d}\alpha+n_{K^2} \end{bmatrix}. \tag{34}$$

To select the partition of $\mathbf{z}$ with the highest probability to carry the non-zero element, Algorithm 1 chooses the partition with the maximum received power. Hence, the probability of successfully estimating the correct AoA/AoD range at this stage is the probability of the event $\bigcap_{\bar{m}=2}^{K^2}\left\{\left[\mathbf{y}_{(1)}\right]_1^2 > \left[\mathbf{y}_{(1)}\right]_{\bar{m}}^2\right\}$. Taking the complement of this event, and using the union bound, we write the probability of error at stage (1) conditioned on the channel gain $p_{(1)}(\alpha)$ as

$$p_{(1)}(\alpha) = \mathrm{P}\left(\bigcup_{\bar{m}=2}^{K^2}\left\{\left[\mathbf{y}_{(1)}\right]_1^2 < \left[\mathbf{y}_{(1)}\right]_{\bar{m}}^2\right\}\middle|\,\alpha\right) \tag{35}$$

$$\leq \sum_{\bar{m}=2}^{K^2}\mathrm{P}\left(\left[\mathbf{y}_{(1)}\right]_1^2 < \left[\mathbf{y}_{(1)}\right]_{\bar{m}}^2\middle|\,\alpha\right). \tag{36}$$

Now, note that $\left[\mathbf{y}_{(1)}\right]_1 \sim \mathcal{N}\left(\mu_1,\sigma^2\right)$ with

$$\mu_1 = \sqrt{\frac{P_{(1)}N_{\mathrm{BS}}N_{\mathrm{MS}}}{\rho}}\left(\sqrt{\frac{G^{(1)}}{N_{\mathrm{BS}}N_{\mathrm{MS}}}}+\left[\mathbf{E}^{(1)}_{k_{\mathrm{BS}},k_{\mathrm{MS}}}\right]_{1,d}\right)\alpha,$$

and $\left[\mathbf{y}_{(1)}\right]_{\bar{m}} \sim \mathcal{N}\left(\mu_{\bar{m}},\sigma^2\right)$ with the mean value $\mu_{\bar{m}} = \sqrt{P_{(1)}N_{\mathrm{BS}}N_{\mathrm{MS}}/\rho}\left[\mathbf{E}^{(s)}_{k_{\mathrm{BS}},k_{\mathrm{MS}}}\right]_{\bar{m},d}\alpha$, $\bar{m}$ takes the values $2,3,\ldots,K^2$. Using the result of [34] for the probability that the difference between the magnitude squares of two Gaussian random variables is less than zero, we reach

$$\mathrm{P}\left(\left[\mathbf{y}_{(1)}\right]_1^2 < \left[\mathbf{y}_{(1)}\right]_{\bar{m}}^2\middle|\,\alpha\right) = \mathrm{Q}_1(a,b) - \frac{1}{2}\mathrm{I}_0(ab)\exp\left(-\frac{1}{2}\left(a^2+b^2\right)\right), \tag{37}$$

with $a = |\mu_1|/\sqrt{2\sigma^2} = \sqrt{P_{(1)}\hat{G}^{(1)}_d/2\rho\sigma^2}|\alpha|$, and $b = |\mu_{\bar{m}}|/\sqrt{2\sigma^2} = \sqrt{P_{(1)}\Delta^{(1)}_{d,\bar{m}}/2\rho\sigma^2}|\alpha|$. Q_1 is the first-order Marcum Q-function, and I_0 is the 0th order modified Bessel function of the first kind. We recall here that $\hat{G}^{(1)}_d$ represents the actual beamforming gain in the direction (the joint quantized AoA/AoD) d, $\Delta^{(s)}_{d,\bar{m}}$ denotes the sidelobe gain of the beamforming vector $\bar{m}$ in the direction d. We also note that the ratio a^2/b^2 equals to the ratio between those two gains, hence, a^2/b^2 equals $\beta^{(s)}_{d,\bar{m}}$. In [35], [36], a new integral form of the Q-function was derived, by which we get

$$\mathrm{P}\left(\left[\mathbf{y}_{(1)}\right]_1^2 < \left[\mathbf{y}_{(1)}\right]_{\bar{m}}^2\middle|\,\alpha\right) = \frac{1}{4\pi}\int_{-\pi}^{\pi}\left(1-\frac{1}{\beta^{(1)}_{d,\bar{m}}}\right) \times\frac{\exp\left(-\frac{P_{(1)}\hat{G}^{(1)}_d|\alpha|}{4\rho\sigma^2}\left(1+\frac{2\sin(\phi)}{\sqrt{\beta^{(1)}_{d,\bar{m}}}}+\frac{1}{\beta^{(1)}_{d,\bar{m}}}\right)\right)}{1+\frac{2\sin(\phi)}{\sqrt{\beta^{(1)}_{d,\bar{m}}}}+\frac{1}{\beta^{(1)}_{d,\bar{m}}}}d\phi. \tag{38}$$

We can now substitute by (38) in (36) to obtain

$$p_{(1)}(\alpha) \leq \frac{1}{4\pi}\int_{-\pi}^{\pi}\sum_{r=1}^{K^2}\frac{1-\frac{1}{\beta^{(1)}_{d,\bar{m}}}}{1+2\sin(\phi)\sqrt{\frac{1}{\beta^{(1)}_{d,\bar{m}}}}+\frac{1}{\beta^{(1)}_{d,\bar{m}}}} \times\exp\left(\frac{-P_{(1)}\hat{G}^{(s)}_d|\alpha|}{4\rho\sigma^2}\left(1+\frac{2\sin(\phi)}{\sqrt{\beta^{(1)}_{d,\bar{m}}}}+\frac{1}{\beta^{(1)}_{d,\bar{m}}}\right)\right)d\phi. \tag{39}$$

From (39), we can show that $\partial p_{(1)}(\alpha)/\partial\beta^{(1)}_{d,\bar{m}} < 0$ for $0 \leq \beta^{(1)}_{d,\bar{m}} < 1$. This is expected as $\beta^{(1)}_{d,\bar{m}}$ represents the forward to backward gain which is intuitively negatively proportional with the probability of error. Hence, we can bound $p_{(1)}(\alpha)$ using the worst forward to backward ratio as

$$p_{(1)}(\alpha) \leq \frac{K^2-1}{4\pi}\int_{-\pi}^{\pi}\frac{1-\frac{1}{\beta^{(1)}_{d}}}{1+2\sin(\phi)\sqrt{\frac{1}{\beta^{(1)}_{d}}}+\frac{1}{\beta^{(1)}_{d}}} \times\exp\left(-\frac{P_{(1)}\hat{G}^{(1)}_d|\alpha|}{4\rho\sigma^2}\left(1+2\sin(\phi)\sqrt{\frac{1}{\beta^{(1)}_{d}}}+\frac{1}{\beta^{(1)}_{d}}\right)\right)d\phi, \tag{40}$$

where $\beta^{(1)}_{d} = \min_{\substack{\forall d=1,2,\ldots,N^2 \\ d\notin\bar{m}}}\beta^{(1)}_{d,\bar{m}}$, i.e., the worst forward to backward gain in the direction d.

Using a similar analysis for each stage s, the total probability of error conditioned on the path gain can be now defined, and bounded again using the union bound as

$$\begin{aligned} &p(\alpha) \\ &= \mathrm{P}\left(\bigcup_{s=1}^{S}\left(\bigcup_{\bar{m}=2}^{K^2}\left[\mathbf{y}_{(s)}\right]_1^2 < \left[\mathbf{y}_{(s)}\right]_{\bar{m}}^2\middle|\,\alpha\right)\right) \\ &\leq \frac{K^2-1}{4\pi}\sum_{s=1}^{S}\int_{-\pi}^{\pi}\frac{1-\frac{1}{\beta^{(s)}_{d}}}{1+2\sin(\phi)\sqrt{\frac{1}{\beta^{(s)}_{d}}}+\frac{1}{\beta^{(s)}_{d}}} \times\exp\left(-\frac{P_{(s)}\hat{G}^{s}_d|\alpha|}{4\rho\sigma^2}\left(1+2\sin(\phi)\sqrt{\frac{1}{\beta^{(s)}_{d}}}+\frac{1}{\beta^{(s)}_{d}}\right)\right)d\phi, \end{aligned} \tag{41}$$

Finally, to obtain the average probability of error $\bar{p}$, we need to integrate over the exponential distribution of $|\alpha|^2$. However, by swapping the summation with the integration sign, we will get again an integral similar to that given and solved in [36, equations (27)-(35)]. Hence, using this result and averaging

over all the N^2 directions, we can directly obtain the bound in (31). ■

While the central idea of the proof of Theorem 1 is similar to the beam misalignment analysis in [9], this theorem has a number of important contributions over the prior work. One key difference is that Theorem 1 considers the case when the channel path gains are taken from a fading distribution; the analysis in [9] assumes constant line-of-sight (LOS) channels. Both the LOS and NLOS cases are important for mmWave systems, and using blockage models, e.g., the model in [37], we can combine the probability of estimation error in the two cases to get more accurate performance evaluation in mmWave systems. Theorem 1 also characterizes the AoA/AoD estimation error as a function of the different parameters of the designed hierarchical codebook in Section IV: the forward to backward gain ratios and the actual beamforming gains with imperfect realization errors. Hence, it provides a realistic evaluation of the proposed adaptive channel estimation algorithm with more practical beamforming patterns. Further, while the beam alignment analysis in [9] focused on the exhaustive search case, i.e., when only the highest resolution beams are used, Theorem 1 considers the case when the AoAs/AoDs are estimated using adaptive algorithms. This is of particular interest in this channel estimation paper, as the bound in Theorem 1 hints that the training power can be distributed among the different adaptive stages to reduce the probability of estimation error. These insights into the power allocation are explained in the remainder of this section.

Now, for the case when $\beta_d^{(s)} \to \infty$, i.e., when the backward gain in negligible and $\mathbf{E}^{(s)}_{k_{\mathrm{BS}},k_{\mathrm{MS}}} \to \mathbf{0}$, we can proceed further, and obtain a sufficient condition on the training power distribution to guarantee estimating the AoA/AoD of the channel with a certain bound on the maximum probability of error.

Corollary 2: Consider using Algorithm 1 to estimate the AoA and AoD of the single-path mmWave channel of model (4), with a resolution $2\pi/N$, with K precoding and measurement vectors of $\mathcal{F}, \mathcal{W}$ used at each stage, and with $\beta_d^{(s)} \to \infty$, and $\mathbf{E}^{(s)}_{k_{\mathrm{BS}},\mathrm{MS}} \to 0$. If the power at each stage $P_{(s)}, s = 1, 2, \ldots, \mathrm{S}$ satisfies:

$$P_{(s)} \geq \frac{\Gamma}{G_{(s)}} \tag{42}$$

with

$$\Gamma = \frac{2}{\bar{\gamma}} \left(\frac{(K^2-1)\mathrm{S}}{\delta} - 2 \right), \tag{43}$$

then, the AoA and AoD are guaranteed to be estimated with an average probability of error $\bar{p} \leq \delta$.

To prove Corollary 2, it is sufficient to substitute with the given $P_{(s)}$, and Γ in (31) to get $\bar{p} \leq \delta$.

We can also note that the power allocation strategy described in Corollary 2 makes the probability of AoA/AoD estimation error equal for the different stages. The intuition behind this result is that the stages with narrower, i.e., higher resolution, beamforming vectors have higher beamforming gains, and hence need less training power to achieve the same estimation success probability compared with the staged with wider beamforming vectors. Another advantage of Corollary 2 is that it characterizes an upper bound on the required training power to achieve a certain success probability. From Corollary 2, it is easy to show that a total training power P_{T}, with $P_{\mathrm{T}} \geq K^2 \Gamma \sum_{s=1}^{S} 1/G_{(s)}$ is sufficient to estimate the AoA/AoD of the single-path mmWave channel with $\bar{p} \leq \delta$ if it is distributed according to the way described in Corollary 2.

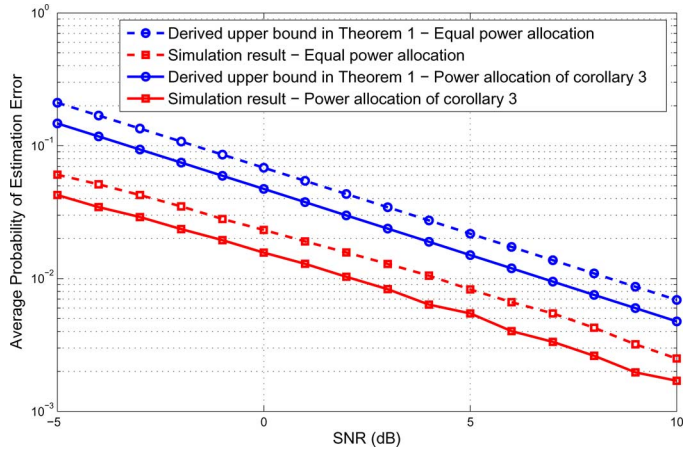

Fig. 6. Average probability of error in estimating the AoA/AoD of single-path channels using Algorithm 1.

Finally, if we have a bound on the total training power, we can use Theorem 1 to get an upper bound on the error probability.

Corollary 3: Consider using Algorithm 1 to estimate the AoA and AoD of the single-path mmWave channel of model (4), with a resolution $2\pi/N$, with K precoding and measurement vectors of $\mathcal{F}, \mathcal{W}$ used at each stage, and with $\beta_d^{(s)} \to \infty$, and $\mathbf{E}^{(s)}_{k_{\mathrm{BS}},\mathrm{MS}} \to 0$. If the total training power is P_{T}, and if this power is distributed over the adaptive stages of Algorithm 1 such that:

$$P_{(s)} = \frac{P_{\mathrm{T}}}{K^2 \sum_{\mathrm{n}=1}^{S} \frac{G_{(s)}}{G_{(\mathrm{n})}}}, \quad s = 1, 2, \ldots, \mathrm{S} \tag{44}$$

Then, the AoA and AoD are guaranteed to be estimated with an average probability of error $\bar{p}$ where

$$\bar{p} \leq \frac{(K^2-1)S}{\frac{P_{\mathrm{T}}\bar{\gamma}}{2K^2 \sum_{s=1}^{S} \frac{1}{G_{(s)}}} + 2}. \tag{45}$$

To prove Corollary 3, it is sufficient to substitute with the given $P_{(s)}$ in (31) to get the bound on $\bar{p}$. Corollary 3 is important as it gives an indication of the reliability of the AoA/AoD estimation when a certain training power is used.

To study the accurateness of the derived bound in Theorem 1 and the power allocation strategy in Corollary 3, Fig. 6 compares the average probability of the AoA/AoD estimation error when Algorithm 1 is used with the bound in Theorem 1. In this figure, we assume a BS with 64 antennas and 10 RF chains is communicating with a MS that has 32 antennas and 6 RF chains. The actual average estimation error and the bound in Theorem 1 are plotted for the case when an equal power allocation over the adaptive stages is used and when the power allocation strategy in Corollary 3 is adopted. The results in Fig. 6 indicate that although the derived bound is not very tight, it gives very useful insights on how the training power can be distributed among the adaptive stages of the proposed channel estimation algorithm to reduce the average probability of estimation error.

B. Adaptive Channel Estimation Algorithm for Multi-Path mmWave Channels

In this section, we consider the case when multiple paths exist between the BS and MS. Thanks to the poor scattering nature of the mmWave channels, the channel estimation problem can be formulated as a sparse compressed sensing problem as discussed in Section III. Consequently, a modified matching pursuit algorithm can be used to estimate the AoAs and AoDs along with the corresponding path gains of L_d paths of the channel, where L_d is the number of dominant paths need to be resolved. Given the problem formulation in (19), the objective now is to determine the L_d non-zero elements of $\mathbf{z}$ with the maximum power. Based on the single-path case, we propose Algorithm 2 to adaptively estimate the different channel parameters.

1) Modified Hierarchical Codebook: For the multi-path case, we need to make a small modification to the structure of the hierarchical codebook described in Section IV. As will be explained shortly, the adaptive algorithm in the multi-path case starts by using KL_d precoding and measurement vectors at the BS and MS instead of K. In each stage, L_d of those KL_d partitions are selected for further refinement by dividing each one into K smaller partitions in the next stage. Hence, to take this into account, the first level of the codebook $\mathcal{F}$ in Section IV consists of one subset with KL_d beamforming vectors that divide the initial AoD range into KL_d ranges. Similarly, in each level $s, s > 1$, the codebook $\mathcal{F}_s$ has $K^{s-1}L_d$ levels, and the ranges $\mathcal{I}_{(s,k)}$, and $\mathcal{I}_{(s,k,m)}$ are consequently defined as $\mathcal{I}_{(s,k)} = \left\{(k-1)N/L_dK^{s-1}, \ldots, kN/L_dK^{s-1}\right\}$, and

$$\mathcal{I}_{(k,s,m)} = \left\{ \frac{N}{L_dK^s}\left(K(k-1)+m-1\right)+1, \ldots, \frac{N}{L_dK^s}\left(K(k-1)+m\right) \right\}.$$

Given these definitions of the quantized AoD ranges associated with each beamforming vector m, of the subset k, of level s, the design of the beamforming vectors proceeds identical to that described in Section IV-B.

To estimate the L_d dominant paths of the mmWave channel, Algorithm 2 makes L_d outer iterations. In each one, an algorithm similar to Algorithm 1 is executed to detect one more path after subtracting the contributions of the previously estimated paths. More specifically, Algorithm 2 operates as follows: In the initial stage, both the BS and MS use KL_d beamforming vectors defined by the codebooks in Section IV to divide the AoA, and AoD range into KL_d sub-ranges each. Similar to the single-path case, the algorithm proceeds by selecting the maximum received signal power to determine the L_d most promising sections to carry the dominant paths of the channel. This process is repeated until we reach the required AoD resolution, and only one path is estimated at this iteration. The trajectories used by the BS to detect the first path is stored in the matrix $\mathbf{T}^{\mathrm{BS}}$ to be used in the later iterations. In the next iteration, a similar BS-MS precoding/measurement step is repeated. However, at each stage s, the contribution of the first path that has been already estimated in the previous iteration, which is stored in $\mathbf{T}^{\mathrm{BS}}$, is projected out before determining the new promising AoD ranges. In the next stage $s+1$, two AoD ranges are selected for further refinement, namely, the one selected at stage s of this iteration, and the one selected by the first path at stage $s+1$ of the first iteration which is stored in $\mathbf{T}^{\mathrm{BS}}$. The selection of those two AoD ranges enables the algorithm to detect different path with AoDs separated by a resolution up to $2\pi/N$. The algorithm proceeds in the same way until the L_d paths are solved. After estimating the AoAs/AoDs with the desired resolution, the algorithm finally calculates the estimated path gains using a linear least squares estimator (LLSE).

Note that one disadvantage of the adaptive beamwidth algorithm in the multi-path case is the possible destructive interference between the path gains when they are summed up in the earlier stages of the algorithm. This disadvantage does not appear in the exhaustive search training algorithms; as only high resolution beams are used in estimating the dominant paths of the channel. The impact of this advantage on the operation of the proposed algorithm, however, is smaller in the case of mmWave channels thanks to the sparse nature of the channel.

The total number of adaptive stages required by Algorithm 2 to estimate the AoAs/AoDs of the L_d paths of the channel with a resolution $2\pi/N$ is $\log_K(N/L_d)$. Since we need KL_d precoding vectors, and KL_d measurement vectors for each precoding direction in each stage, and since these adaptive stages are repeated for each path, the total number of steps required to estimate L_d paths of the mmWave channel using the proposed algorithm is $K^2L_d^3\log_K(N/L_d)$. If multiple RF chains are used in the MS to combine the measurements, the required number of training time slots is then reduced to be $KL_d^2\lceil KL_d/N_{\mathrm{RF}}\rceil\log_K(N/L_d)$.

VI. Hybrid Precoding Design

We seek now to design the hybrid precoders/combiners, $(\mathbf{F}_{\mathrm{RF}}, \mathbf{F}_{\mathrm{BB}}, \mathbf{W}_{\mathrm{RF}}, \mathbf{W}_{\mathrm{BB}})$, at both the BS and MS to maximize the mutual information achieved with Gaussian signaling over the mmWave link in (3) [38] while taking the different RF precoding constraints into consideration. Regardless of whether uplink or downlink transmission is considered, the hybrid precoding problem can be summarized as directly maximizing the rate expression

$$R = \log_2\left|I_{N_{\mathrm{S}}} + \frac{P}{N_{\mathrm{S}}}\mathbf{R}_{\mathrm{n}}^{-1}\mathbf{W}_{\mathrm{BB}}{}^H\mathbf{W}_{\mathrm{RF}}{}^H\mathbf{H}\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}\mathbf{F}_{\mathrm{BB}}{}^H\mathbf{F}_{\mathrm{RF}}{}^H\mathbf{H}^H\mathbf{W}_{\mathrm{RF}}\mathbf{W}_{\mathrm{BB}}\right|, \quad (46)$$

over the choice of feasible analog and digital processing matrices $(\mathbf{F}_{\mathrm{RF}}, \mathbf{F}_{\mathrm{BB}}, \mathbf{W}_{\mathrm{RF}}, \mathbf{W}_{\mathrm{BB}})$. Note that in (46), $\mathbf{R}_{\mathrm{n}}$ is the post-processing noise covariance matrix, i.e., $\mathbf{R}_{\mathrm{n}} = \mathbf{W}_{\mathrm{BB}}{}^H\mathbf{W}_{\mathrm{RF}}{}^H\mathbf{W}_{\mathrm{RF}}\mathbf{W}_{\mathrm{BB}}$ in the downlink, and $\mathbf{R}_{\mathrm{n}} = \mathbf{F}_{\mathrm{BB}}{}^H\mathbf{F}_{\mathrm{RF}}{}^H\mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}$ in the uplink.

For simplicity of exposition, we summarize the process with which the BS calculates the hybrid precoding matrices, $(\mathbf{F}_{\mathrm{RF}}, \mathbf{F}_{\mathrm{BB}})$, to be used on the downlink. Calculation of the uplink precoders used by the MS follows in an identical manner.

We propose to split the precoding problem into two phases. In the first phase, the BS and MS apply the adaptive channel estimation algorithm of Section III to estimate the mmWave channel parameters. At the end of the channel training/estimation phase, the BS constructs the downlink channel's matrix leveraging the geometric structure of the channel. If the channel

Algorithm 2 Adaptive Estimation Algorithm for Multi-Path mmWave Channels

Input: BS and MS know N, K, L_d, and have $\mathcal{F}, \mathcal{W}$

Initialization: $\mathbf{T}^{\mathrm{BS}}_{(1,1)} = \{1, \ldots, 1\}, \mathbf{T}^{\mathrm{MS}}_{(1,1)} = \{1, \ldots, 1\}$

$\mathrm{S} = \log_K (N/L_d)$

for $\ell \leq L_d$ **do**

  for $s \leq \mathrm{S}$ **do**

    for $m_{\mathrm{BS}} \leq KL_d$ **do**

      BS uses $\left[\mathbf{F}_{(s,\mathbf{T}^{\mathrm{BS}}_{(\ell,s)})}\right]_{:,m_{\mathrm{BS}}}$

      for $m_{\mathrm{MS}} \leq KL_d$ **do**

        MS uses $\left[\mathbf{W}_{(s,\mathbf{T}^{\mathrm{BS}}_{(\ell,s)})}\right]_{:,m_{\mathrm{MS}}}$

      After MS measurements:

      $\mathbf{y}_{m_{\mathrm{BS}}} = \sqrt{P_s}\left[\mathbf{W}_{(s,\mathbf{T}^{\mathrm{MS}}_{(\ell,s)})}\right]\mathbf{H}\left[\mathbf{F}_{(s,\mathbf{T}^{\mathrm{BS}}_{(\ell,s)})}\right]_{:,m_{\mathrm{BS}}} + \mathbf{n}_{m_{\mathrm{BS}}}$

    $\mathbf{y}_{(s)} = [\mathbf{y}_1^T, \mathbf{y}_2^T, \ldots, \mathbf{y}_K^T]^T$

    for $p = 1 \leq \ell - 1$ **do**

      Project out previous path contributions

      $\mathbf{g} = \mathbf{F}^T_{(s,\mathbf{T}^{\mathrm{BS}}_{(p,s)})}\left[\mathbf{A}_{\mathrm{BS,D}}\right]^*_{:,\mathbf{T}^{\mathrm{BS}}_{(p,s)}(1)} \otimes \mathbf{W}^H_{(s,\mathbf{T}^{\mathrm{MS}}_{(p,s)})}\left[\mathbf{A}_{\mathrm{MS,D}}\right]_{:,\mathbf{T}^{\mathrm{MS}}_{(p,s)}(1)}$

      $\mathbf{y}_{(s)} = \mathbf{y}_{(s)} - \mathbf{y}^H_{(s)}\mathbf{g}\left(\mathbf{g}^H\mathbf{g}\right)\mathbf{g}$

    $\mathbf{Y} = \mathrm{matrix}(\mathbf{y}_{(s)})$ Return $\mathbf{y}_{(s)}$ to the matrix form

    $(m^\star_{\mathrm{BS}}, m^\star_{\mathrm{MS}}) = \underset{\forall m_{\mathrm{BS}}, m_{\mathrm{MS}}=1,2,\ldots,K}{\arg\max} [\mathbf{Y} \odot \mathbf{Y}^*]_{m_{\mathrm{MS}}, m_{\mathrm{BS}}}$

    $\mathbf{T}^{\mathrm{BS}}_{(\ell,s+1)}(1) = K(m^\star_{\mathrm{BS}} - 1) + 1$

    $\mathbf{T}^{\mathrm{MS}}_{(\ell,s+1)}(1) = K(m^\star_{\mathrm{MS}} - 1) + 1$

    for $p = 1 \leq \ell - 1$ **do**

      $\mathbf{T}^{\mathrm{BS}}_{(\ell,s+1)}(p) = \mathbf{T}^{\mathrm{BS}}_{(p,s+1)}(1)$

      $\mathbf{T}^{\mathrm{MS}}_{(\ell,s+1)}(p) = \mathbf{T}^{\mathrm{MS}}_{(p,s+1)}(1)$

$\hat{\phi}_\ell = \bar{\phi}_{\mathbf{T}^{\mathrm{BS}}_{(\ell,\mathrm{S}+1)}(1)}, \hat{\theta}_\ell = \bar{\theta}_{\mathbf{T}^{\mathrm{MS}}_{(\ell,\mathrm{S}+1)}(1)}$

$\mathbf{g} = \mathbf{F}^T_{(\mathrm{S},\mathbf{T}^{\mathrm{BS}}_{(\ell,\mathrm{S})})}\left[\mathbf{A}^*_{\mathrm{BS,D}}\right]_{:,\mathbf{T}^{\mathrm{BS}}_{(\ell,\mathrm{S}+1)}(1)} \otimes \mathbf{W}^H_{(\mathrm{S},\mathbf{T}^{\mathrm{MS}}_{(\ell,\mathrm{S})})}\left[\mathbf{A}_{\mathrm{MS,D}}\right]_{:,\mathbf{T}^{\mathrm{MS}}_{(\ell,\mathrm{S}+1)}(1)}$

$\hat{\alpha}_\ell = \sqrt{\rho/P_{(\mathrm{S})}G_{(\mathrm{S})}}\mathbf{y}^H_{(\mathrm{S})}\mathbf{g}/\mathbf{g}^H\mathbf{g}$

is not reciprocal, the estimation algorithm of Section V can be used to construct the uplink channel matrix at the MS.

As a result of the downlink channel training/estimation phase in Section V, the BS now has estimated knowledge of its own steering matrix $\hat{\mathbf{A}}_{\mathrm{BS}}$, the MS steering matrix $\hat{\mathbf{A}}_{\mathrm{MS}}$, and the estimated path gain vector $\hat{\boldsymbol{\alpha}}$. Thus, the BS may construct the estimated downlink channel matrix as

$$\hat{\mathbf{H}} = \hat{\mathbf{A}}_{\mathrm{MS}}\mathrm{diag}\left(\hat{\boldsymbol{\alpha}}\right)\hat{\mathbf{A}}^H_{\mathrm{BS}}. \tag{47}$$

The BS can now build its hybrid data precoders $\mathbf{F}_{\mathrm{RF}}$ and $\mathbf{F}_{\mathrm{BB}}$ to approximate the dominant singular vectors of the channel, $\hat{\mathbf{H}}$, denoted by the unconstrained precoder $\mathbf{F}_{\mathrm{opt}}$ using a similar procedure to [13, equation (16)-(18)] and using Algorithm 1 in [13].

VII. Simulation Results

In this section, we present numerical results to evaluate the performance of the proposed training codebook, adaptive channel estimation algorithm, and hybrid precoding algorithm. We firstly consider a single BS-MS link, and then show some results for the mmWave cellular channel model.

A. Performance Evaluation With Point-to-Point Channels

In these simulations, we consider the case when there is only one BS and one MS, i.e., without any interference. The system model and the simulation scenario are as follows:

1) System Model: We adopt the hybrid analog/digital system architecture presented in Fig. 2. The BS has $N_{\mathrm{BS}} = 64$ antennas, and 10 RF chains, the MS has $N_{\mathrm{MS}} = 32$ antennas and 6 RF chains. The antenna arrays are ULAs, with spacing between antennas equal to $\lambda/2$, and the RF phase shifters are assumed to have only quantized phases. Hence, only a finite set of the RF beamforming vectors is allowed, and assumed to be beam-steering vectors, as discussed in Section IV-B, with 7 quantization bits.

2) Channel Model: We consider the channel model described in (4), with $\bar{P}_R = 1$, and a number of paths $L = 3$. The AoAs/AoDs are assumed to take continuous values, i.e., not quantized, and are uniformly distributed in $[0, 2\pi]$. The system is assumed to operate at 28 GHz carrier frequency, has a bandwidth of 100 MHz, and with path-loss exponent $n_{\mathrm{pl}} = 3$.

3) Simulation Scenario: All the simulations in this section will present spectral efficiency results with different system, and algorithms parameters. To generate these results, the channel parameters are estimated using the algorithms presented in Section V. After estimating its parameters, the geometrical channel is reconstructed according to (47), and is used in the design of the hybrid precoders and decoders according to Section VI. Unless otherwise mentioned, these are the parameters used for both of the two steps:

1) Channel estimation parameters: For the single-path channels, Algorithm 1 is used to estimate the channel parameters with AoA/AoD resolution parameter $N = 64$, and with $K = 2$ beamforming vectors at each stage. For the multi-path case, the parameters N, K, L_d will be defined with each simulation. The training power are determined according to Corollary 2, with a desired maximum probability of error $\delta = 0.05$. Hence, the training power changes based on the parameter K, and N. Also, the total training power is distributed over the adaptive estimation stages according to Corollary 2.
2) Hybrid precoding parameters: The precoding matrices are constructed with the same system architecture, and assuming a number of multiplexed streams $N_{\mathrm{S}} = L_d$.

In Fig. 7, the precoding gains given by the proposed mmWave channel estimation algorithms are simulated for the cases when the desired number of estimated paths L_d equals 1, 2, and 3. Algorithm 1, and Algorithm 2 are simulated for different values

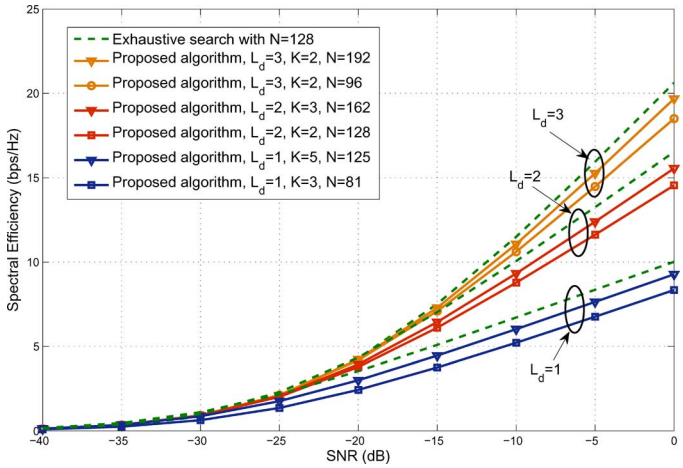

Fig. 7. Spectral efficiency achieved when the precoding matrices are built using the mmWave channel estimated by the proposed algorithms in a channel with $L = 3$, and $L_d = 1, 2, 3$. The figure compares the performance of the algorithm when different values of the parameter K are chosen. The results indicate that a very close performance to the exhaustive search case can be achieved with $K \ll N$, which maps to much smaller numbers of iterations.

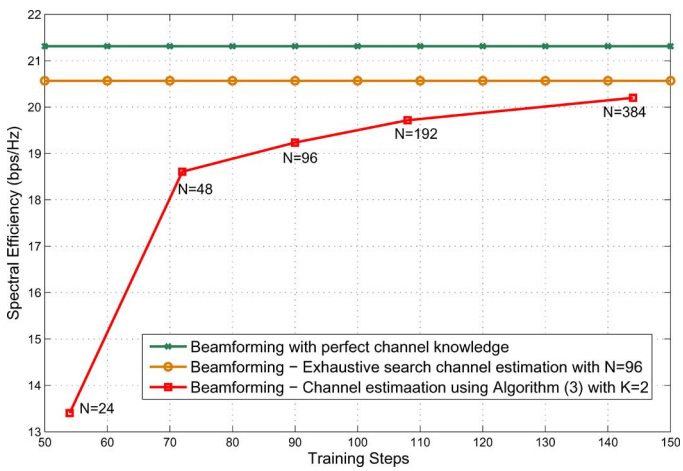

Fig. 8. The improvement of the spectral efficiency with the development of the adaptive channel estimation algorithm is shown and compared with the exhaustive search and perfect channel knowledge cases.

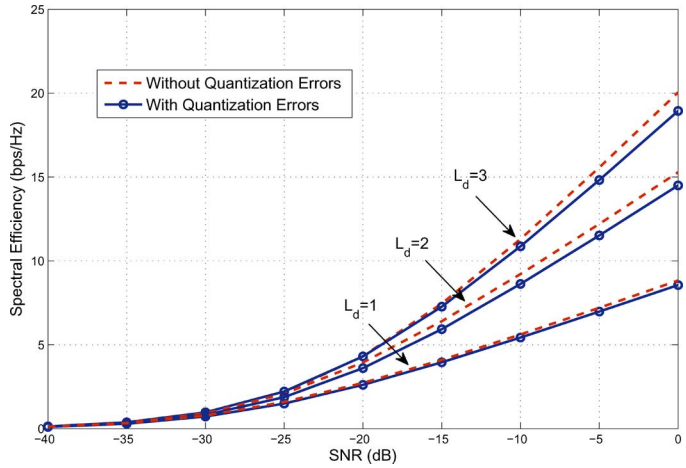

Fig. 9. The performance error due to the AoAs/AoDs quantization assumption in (16) is evaluated. The performance error is the difference between the curve with continuous angles, and the one with quantization, as this continuity of angles' values is not taken into consideration while designing the algorithm.

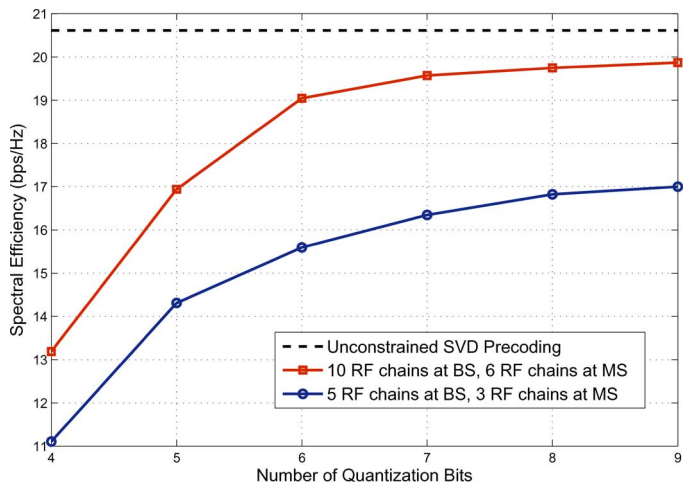

Fig. 10. Spectral efficiency as a function of phase quantization bits in a hybrid system with only quantized analog phase control. Results compare the performance of the hybrid analog digital channel estimation and precoding algorithms with the unconstrained digital system.

of K, and compared with the precoding gain of the exhaustive search solution. The results indicate that comparable gains can be achieved using the proposed algorithms despite their low-complexity, and the requirement of a much smaller number of iterations. For example, for $L_d = 3$, and $K = 2$, although only $96 \ll N_{\mathrm{BS}}N_{\mathrm{MS}} = 2048$ training steps are required, the spectral efficiency performance degradation is less than 1 bps/Hz compared with the exhaustive search solution that requires much more iterations.

In Fig. 8, the improvement of the precoding gains achieved by the proposed algorithm for $L_d = 3$ with the training iterations is simulated. The results show that more than 90% of the exhaustive search gain can be achieved with only 70 iterations with $K = 2$. These results also indicate that a wise choice of the desired resolution parameter N is needed in order to have a good compromise between performance and training overhead. For example, the figure shows that doubling the number of training steps, i.e., from 70 to 140, achieves an improvement of only 1 bps/Hz in the spectral efficiency.

In Fig. 9, we evaluate the error in the performance of the proposed channel estimation algorithm caused by the AoAs/AoDs quantization assumption made in (16), the proposed algorithms are simulated for the cases when the channel AoAs/AoDs are quantized, i.e., when the used quantization assumption is exact, and when the AoAs/AoDs are continuous, i.e., with quantization error induced in our formulation. The figure plots the performance of the proposed algorithms for the cases $L_d = 1, K = 2, N = 81$, $L_d = 2$, $K = 2, N = 128$, and $L_d = 3, K = 3, N = 96$, and show that the performance loss in our algorithms due to the quantization assumption is very small for large enough resolution parameters N.

In Fig. 10, the impact of the RF system limitations on the performance of the proposed channel estimation, and precoding algorithms, is evaluated, and compared with the case of constraints-free system. Two system models are considered in Fig. 10, one with 10 RF chains at the BS, 6 RF chains at the MS, and the other with 5 RF chains at the BS, and 3 RF chains at the MS. The other parameters are the same as the

previous simulations with $L_d = 3$. The performance achieved by those two systems is further simulated with different number of quantization bits of the phase shifters. Simulation results show that the proposed hybrid precoding algorithm can achieve near-optimal data-rates compared with the unconstrained solutions if a sufficient number of RF chains, and quantization bits exist. Also, the results show that 5 quantization bits may be sufficient to accomplish more than 90% of the maximum gain.

B. Performance Evaluation With mmWave Cellular System

Now, we consider evaluating the proposed algorithm in a mmWave cellular system setting with the following stochastic geometry model.

1) Network and System Models: The desired BS, in a cell of radius $R_c = 100m$, is assumed to communicate with a MS using the channel estimation, and hybrid precoding algorithms derived. Each MS is assumed to receive its desired signal s_d in addition to cellular interference. The interfering BSs follow a Poisson point process (PPP) $\Phi(\lambda)$ with $\lambda = 1/\pi R_c^2$ to model the downlink out-of-cell interference [37], [39], [40]. To simulate a cellular setting, the nearest BS to the MS is always considered as the desired BS. The received signal at the MS can be then written as

$$\mathbf{y} = \mathbf{W}\mathbf{H}_d\mathbf{F}_d\mathbf{s}_d + \sum_{\substack{r_i \in \Phi(\lambda) \\ r_i \geq r_d}} \mathbf{W}\mathbf{H}_i\mathbf{F}_i\mathbf{s}_i + \mathbf{n} \qquad (48)$$

where r_d, r_i are the distances from the MS to the desired and the ith interfering BSs, respectively. Each interfering BS is assumed to have the same number of ULA antennas $N_{\mathrm{BS}} = 64$, and to have the same horizontal orientation of the antenna arrays, i.e., all the beamforming is in the azimuth domain. Further, each BS generates a beamsteering beamforming vectors that steers its signal in a uniform random direction, i.e., $\mathbf{F}_i = \mathbf{a}_{\mathrm{BS}}(\phi_i)$, ϕ_i is uniformly chosen in $[0, 2\pi]$. $\mathbf{H}_i$ has the same definition in (4) with the path loss calculated for each BS. For fairness, all BSs are assumed to transmit with the same average power P. All the other system parameters are similar to the previous section.

In each stage s of the estimation phase, the received signal at the MS is given by (48) with $\mathbf{W}$ and $\mathbf{F}$ equal to the BS and MS training precoders and combiners descried in Section V. Hence, the cellular interference affects the maximum power detection problem at every stage of the channel estimation algorithm. After the channel is estimated, the precoders $\mathbf{W}$ and $\mathbf{F}$ are designed as shown in Section VI.

To evaluate the performance of the proposed hybrid precoding algorithm, we adopt the coverage probability as a performance metric. As we are interested in multiplexing many streams per user, we define the coverage probability relative to the rate instead of the signal to interference and noise ratio (SINR). Consequently, we define the coverage probability as

$$P_{(c)}(\eta) = \mathrm{P}(R \geq \eta). \qquad (49)$$

An outage occurs if the rate falls below a threshold η.

2) Scenario and Results: In Fig. 11, the coverage probability is evaluated as described before. The curves with 'Estimated Channel' label represents the case when Algorithm (3) is used to estimate the channel parameters in the presence of interference. After estimating the channel, this interference is taken into consideration again in calculating the coverage probability in the curve labeled 'With Interference', and omitted for the curve with the label 'No Interference'. Hence, those two curves represent the cases when cellular interference affects both the channel estimation and data transmission phases, or the channel estimation phase only. The presented results compare the performance of the mentioned scenario using the proposed algorithms, with the case when the hybrid precoding algorithm in Section VI is designed based on perfect channel state information (CSI). They are also compared with the case when only analog beamforming is used to steer the signal towards the dominant channel paths. The results show that a reasonable gain can be achieved with the proposed hybrid precoding algorithm due to its higher capability of managing the inter-stream interference, in addition to overcoming the RF hardware constraints. The simulations also indicate that the effect of the cellular interference of the performance of the channel estimation and precoding algorithms is not critical despite of the low-complexity of the proposed algorithms.

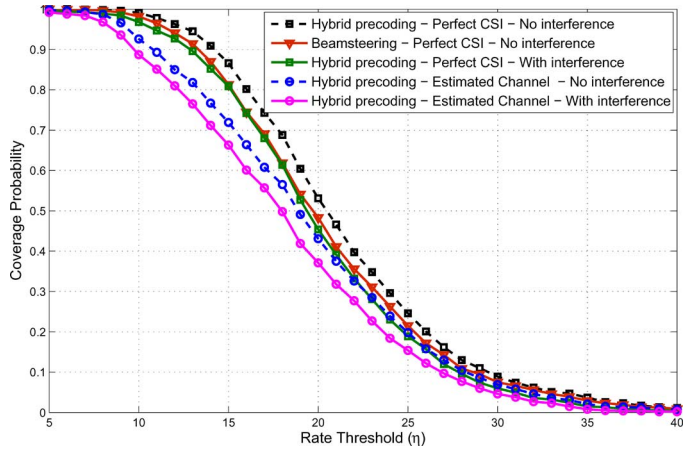

Fig. 11. Coverage probabilities of the proposed channel estimation and precoding algorithms in a mmWave cellular system setting with PPP interference. The figure compares the different cases when the estimation and/or interference error exist.

VIII. Conclusions

In this paper, we considered a single-user mmWave system setting, and investigated the design of suitable mmWave channel estimation and precoding algorithms. First, we formulated, and developed a hierarchical multi-resolution codebook based on hybrid analog/digital precoding. We then proposed mmWave channel estimation algorithms that efficiently detect the different parameters of the mmWave channel with a low training overhead. The performance of the proposed algorithm is analytically evaluated for the single-path channel case, and some insights into efficient training power distributions are obtained. Despite the low-complexity, simulation results showed that the proposed channel estimation algorithm realizes spectral efficiency and precoding gain that are comparable to that obtained by exhaustive search. The attained precoding gains can be also stated in terms of the coverage probability of mmWave cellular systems. For future work, it would be interesting to consider mmWave channels with random blockage between

the BS and MS [41], and seek the design of robust adaptive channel estimation algorithms. Besides the channel estimation algorithms developed in this paper assuming fixed and known array structures, it would be also important for mmWave systems to develop efficient algorithms that adaptively estimate the channel with random or time-varying array manifolds.

REFERENCES

[1] Z. Pi and F. Khan, "An introduction to millimeter-wave mobile broadband systems," *IEEE Commun.. Mag.*, vol. 49, no. 6, pp. 101–107, Jun. 2011.

[2] *IEEE 802.11ad Standard Draft D0.1*, [Online]. Available: www.ieee802.org/11/Reports/tgad update.htm

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

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

[5] A. Abbaspour-Tamijani and K. Sarabandi, "An affordable millimeter-wave beam-steerable antenna using interleaved planar subarrays," *IEEE Trans. Antennas Propag.*, vol. 51, no. 9, pp. 2193–2202, Sep. 2003.

[6] B. Biglarbegian, M. Fakharzadeh, D. Busuioc, M. Nezhad-Ahmadi, and S. Safavi-Naeini, "Optimized microstrip antenna arrays for emerging millimeter-wave wireless applications," *IEEE Trans. Antennas Propag.*, vol. 59, no. 5, pp. 1742–1747, May 2011.

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

[8] L. Chen, Y. Yang, X. Chen, and W. Wang, "Multi-stage beamforming codebook for 60GHz WPAN," in *Proc. 6th Int. ICST Conf. Commun. Network. China*, China, 2011, pp. 361–365.

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

[10] Y. Tsang, A. Poon, and S. Addepalli, "Coding the beams: Improving beamforming training in mmwave communication system," in *Proc. IEEE Global Telecomm. Conf. (GLOBECOM)*, Houston, TX, USA, 2011, pp. 1–6.

[11] X. Zhang, A. Molisch, and S. 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.

[12] 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.

[13] O. E. 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. 2013.

[14] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, Jr, "Hybrid precoding for millimeter wave cellular systems with partial channel knowledge," in *Proc. Inf. Theory Applicat. Workshop (ITA)*, San Diego, CA, USA, Feb. 2013, pp. 1–5.

[15] T. Rappaport, Y. Qiao, J. Tamir, J. Murdock, and E. Ben-Dor, "Cellular broadband millimeter wave propagation and angle of arrival for adaptive beam steering systems," in *Proc. Radio and Wireless Symp. (RWS)*, Santa Clara, CA, USA, Jan. 2012, pp. 151–154.

[16] J. Murdock, E. Ben-Dor, Y. Qiao, J. Tamir, and T. Rappaport, "A 38 ghz cellular outage study for an urban outdoor campus environment," in *Proc. Wireless Commun. Netw. Conf. (WCNC)*, Shanghai, China, Apr. 2012, pp. 3085–3090.

[17] H. Zhang, S. Venkateswaran, and U. Madhow, "Channel modeling and MIMO capacity for outdoor millimeter wave links," in *Proc. Wireless Commun. Netw. Conf. (WCNC)*, Sydney, NSW, Australia, Apr. 2010, pp. 1–6.

[18] T. Rappaport, F. Gutierrez, E. Ben-Dor, J. Murdock, Y. Qiao, and J. Tamir, "Broadband millimeter-wave propagation measurements and models using adaptive-beam antennas for outdoor urban cellular communications," *IEEE Trans. Antennas Propag.*, vol. 61, no. 4, pp. 1850–1859, Apr. 2013.

[19] A. Sayeed and V. Raghavan, "Maximizing MIMO capacity in sparse multipath with reconfigurable antenna arrays," *IEEE J. Sel. Topics Signal Process.*, vol. 1, no. 1, pp. 156–166, Jun. 2007.

[20] Y. M. Tsang and A. S. Poon, "Successive AoA estimation: Revealing the second path for 60 GHz communication system," in *Proc. 49th Annu. Allerton Conf. Commun., Control, Comput. (Allerton)*, 2011, pp. 508–515.

[21] H. L. Van Trees, "Optimum array processing (detection, estimation, and modulation theory, Part IV)," *Wiley-Interscience, Mar*, no. 50, p. 100, 2002.

[22] P. Xia, S.-K. Yong, J. Oh, and C. Ngo, "A practical SDMA protocol for 60 GHz millimeter wave communications," in *Proc. 42nd Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, USA, Oct. 2008, pp. 2019–2023.

[23] P. Xia, S.-K. Yong, J. Oh, and C. Ngo, "Multi-stage iterative antenna training for millimeter wave communications," in *Proc. Global Telecommun. Conf. (GLOBECOM)*, New Orleans, LA, USA, 2008, pp. 1–6.

[24] V. Raghavan and A. Sayeed, "Sublinear capacity scaling laws for sparse MIMO channels," *IEEE Trans. Inf. Theory*, vol. 57, no. 1, pp. 345–364, Jan 2011.

[25] A. J. Laub, *Matrix Analysis for Scientists and Engineers*. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 2004.

[26] H. Zhu, G. Leus, and G. Giannakis, "Sparse regularized total least squares for sensing applications," in *Proc. IEEE 11th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Marrakech, Morocco, Jul. 2010, pp. 1–5.

[27] C. Ekanadham, D. Tranchina, and E. P. Simoncelli, "Recovery of sparse translation-invariant signals with continuous basis pursuit," *IEEE Trans. Signal Process.*, vol. 59, no. 10, pp. 4735–4744, Oct. 2011.

[28] D. Ramasamy, S. Venkateswaran, and U. Madhow, "Compressive parameter estimation in AWGN," ArXiv, 2013 [Online]. Available: arXiv:1304.7539, to be published

[29] M. Rossi, A. Haimovich, and Y. Eldar, "Spatial compressive sensing for MIMO radar," *IEEE Trans. Signal Process.*, vol. 62, no. 2, pp. 419–430, Feb. 2014.

[30] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[31] M. L. Malloy and R. D. Nowak, "Near-optimal compressive binary search," ArXiv, 2012 [Online]. Available: arXiv:1306.6239, to be published

[32] M. L. Malloy and R. D. Nowak, "Near-optimal adaptive compressed sensing," in *Proc. 46th Asilomar Conf. Signals, Syst. Comput. (ASILOMAR)*, Pacific Grove, CA, USA, 2012, pp. 1935–1939.

[33] M. Iwen and A. Tewfik, "Adaptive strategies for target detection and localization in noisy environments," *IEEE Trans. Signal Process.*, vol. 60, no. 5, pp. 2344–2353, May 2012.

[34] J. G. Proakis, *Digital Communications*. New York, NY, USA: McGraw-Hill, 1995.

[35] M. K. Simon, "A new twist on the Marcum Q-function and its application," *IEEE Commun. Lett.*, vol. 2, no. 2, pp. 39–41, Feb. 1998.

[36] M. K. Simon and M.-S. Alouini, "A unified approach to the probability of error for noncoherent and differentially coherent modulations over generalized fading channels," *IEEE Trans. Commun.*, vol. 46, no. 12, pp. 1625–1638, Dec. 1998.

[37] T. Bai and R. W. Heath Jr., "Coverage analysis in dense millimeter wave cellular networks," in *Proc. 47th Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, USA, Nov. 2013.

[38] A. Goldsmith, S. Jafar, N. Jindal, and S. Vishwanath, "Capacity limits of MIMO channels," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 5, pp. 684–702, Jun. 2003.

[39] J. G. Andrews, F. Baccelli, and R. K. Ganti, "A tractable approach to coverage and rate in cellular networks," *IEEE Trans. Commun.*, vol. 59, no. 11, pp. 3122–3134, Nov. 2011.

[40] S. Akoum, O. El Ayach, and R. W. Heath, "Coverage and capacity in mmwave cellular systems," in *Proc. 46th Asilomar Conf. Signals, Syst. Comput. (ASILOMAR)*, Pacific Grove, CA, USA, 2012, pp. 688–692.

[41] T. Bai, R. Vaze, and R. W. J. Heath, "analysis of blockage effects on urban cellular networks," ArXiv, 2013 [Online]. Available: arXiv:1309.4141, to be published

Ahmed Alkhateeb (S'08) received the B.S. (with highest honors) and M.S. degrees from Cairo University, Egypt, in 2008 and 2012, respectively. He is currently a graduate student at the Wireless Networking and Communication Group (WNCG) in the Department of Electrical and Computer Engineering, at the University of Texas at Austin, USA. His research interests are in the broad area of network information theory, communication theory, and signal processing. In the context of wireless communication, his interests include cooperative communications, MIMO systems, and mmWave communication.

Omar El Ayach (S'08–M'13) is currently a Senior Systems Engineer at Qualcomm Research in San Diego, CA. He received his M.S. and Ph.D. in electrical and computer engineering from The University of Texas at Austin in 2010 and 2013, respectively. Before joining the University of Texas, he received his B.E. degree in computer and communications engineering from the American University of Beirut, Lebanon, in 2008. He was an intern at the University of California, Berkeley in the Summer of 2007 and an intern at Samsung Research America—Dallas in the Summers of 2011 and 2012. His research interests are in the broad area of network sciences, signal processing and information theory. In the context of wireless communication, his interests are in MIMO systems, interference management, and mmWave communication.

Geert Leus received the electrical engineering degree and the Ph.D. degree in applied sciences from the Katholieke Universiteit Leuven, Belgium, in June 1996 and May 2000, respectively. Currently, he is an "Antoni van Leeuwenhoek" Full Professor at the Faculty of Electrical Engineering, Mathematics and Computer Science of the Delft University of Technology, The Netherlands. His research interests are in the area of signal processing for communications. He received a 2002 IEEE Signal Processing Society Young Author Best Paper Award and a 2005 IEEE Signal Processing Society Best Paper Award. He is a Fellow of the IEEE. He was the Chair of the IEEE Signal Processing for Communications and Networking Technical Committee, and an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING, the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, the IEEE SIGNAL PROCESSING LETTERS, and the *EURASIP Journal on Advances in Signal Processing*. Currently, he is a Member-at-Large to the Board of Governors of the IEEE Signal Processing Society and a member of the IEEE Sensor Array and Multichannel Technical Committee. He finally serves as the Editor in Chief of the *EURASIP Journal on Advances in Signal Processing*.

Robert W. Heath, Jr. (S'96–M'01–SM'06–F'11) received the B.S. and M.S. degrees from the University of Virginia, Charlottesville, VA, in 1996 and 1997 respectively, and the Ph.D. from Stanford University, Stanford, CA, in 2002, all in electrical engineering. From 1998 to 2001, he was a Senior Member of the Technical Staff then a Senior Consultant at Iospan Wireless Inc, San Jose, CA, where he worked on the design and implementation of the physical and link layers of the first commercial MIMO-OFDM communication system. Since January 2002, he has been with the Department of Electrical and Computer Engineering at The University of Texas at Austin where he is a Cullen Trust for Higher Education Endowed Professor and is Director of the Wireless Networking and Communications Group. He is also President and CEO of MIMO Wireless Inc. and Chief Innovation Officer at Kuma Signals LLC. His research interests include several aspects of wireless communication and signal processing: limited feedback techniques, multihop networking, multiuser and multicell MIMO, interference alignment, adaptive video transmission, manifold signal processing, and millimeter wave communication techniques.

Dr. Heath has been an Editor for the IEEE TRANSACTIONS ON COMMUNICATION, an Associate Editor for the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, and lead guest editor for an IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS special issue on limited feedback communication, and lead guest editor for an IEEE JOURNAL ON SELECTED TOPICS IN SIGNAL PROCESSING special issue on Heterogenous Networks. He currently serves on the steering committee for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS. He was a member of the Signal Processing for Communications Technical Committee in the IEEE Signal Processing Society and is a former Chair of the IEEE COMSOC Communications Technical Theory Committee. He was a technical co-chair for the 2007 Fall Vehicular Technology Conference, general chair of the 2008 Communication Theory Workshop, general co-chair, technical co-chair and co-organizer of the 2009 IEEE Signal Processing for Wireless Communications Workshop, local co-organizer for the 2009 IEEE CAMSAP Conference, technical co-chair for the 2010 IEEE International Symposium on Information Theory, the technical chair for the 2011 Asilomar Conference on Signals, Systems, and Computers, general chair for the 2013 Asilomar Conference on Signals, Systems, and Computers, founding general co-chair for the 2013 IEEE GlobalSIP conference, and is technical co-chair for the 2014 IEEE GLOBECOM conference.

Dr. Heath was a co-author of best student paper awards at several conferences as well as co-recipient of the Grand Prize in the 2008 WinTech WinCool Demo Contest. He was co-recipient of the 2010 and 2013 EURASIP Journal on Wireless Communications and Networking best paper awards, the 2012 *Signal Processing Magazine* best paper award, a 2013 Signal Processing Society best paper award, the 2014 *EURASIP Journal on Advances in Signal Processing* best paper award, and the 2014 *Journal of Communications and Networks* best paper award. He was a 2003 Frontiers in Education New Faculty Fellow. He is also a licensed Amateur Radio Operator and is a registered Professional Engineer in Texas.