Fast Beam Training Architecture for Hybrid mmWave Transceivers

Fast Beam Training Architecture for Hybrid mmWave Transceivers

Jie Yang, *Student Member, IEEE*, Shi Jin, *Senoir Member, IEEE*, Chao-Kai Wen, *Member, IEEE*, Xi Yang, and Michail Matthaiou, *Senoir Member, IEEE*

***Abstract*—Millimeter-wave (mmWave) communications attract considerable interest due to the massive available spectrum. However, to establish communication links, a beam training procedure is indispensable. How to accelerate the beam training process is one of the key challenges towards realizing mmWave communications in practice. In this study, we first propose a novel low-cost digital beamforming (DBF) module assisted hybrid (DA-hybrid) architecture, by exploiting both the capabilities of analog and digital modules. To make this topology practical, we deploy coarse radio frequency (RF) chains and low-resolution analog-to-digital converters in the low-cost DBF module to reduce cost and power consumption. Second, we design a fast beam training method (named DAH-BT) by utilizing the proposed DA-hybrid architecture and leveraging the sparse nature of mmWave channels, in which an internal calibration method is adapted to obtain the parameters of the RF impairments and the orthogonal matching pursuit algorithm is utilized to estimate beams. We also prove that the developed measurement matrices satisfy the restricted isometry property. Extensive simulation results show that the DA-hybrid architecture can not only provide close to 100% beam matching accuracy, but also dramatically reduce the system power consumption and cost. In addition, the proposed DAH-BT scheme consumes the shortest time for beam training over the state-of-art methods with comparable spectral efficiency.**

***Index Terms*—Beam training, digital beamforming architecture, hybrid architecture, low-resolution ADC, millimeter-wave system, RF impairments.**

I. Introduction

Millimeter-wave (mmWave) frequencies have been identified as a catalyst for the next generation wireless communication systems [1], since the mmWave band offers extremely large bandwidth and can therefore boost peak data rates. However, the path loss of mmWave signals is inherently large, and mmWave signals are sensitive to blockage, hence, large antenna arrays and highly directional transmission should be combined to compensate for the severe penetration path losses [2]. Unfortunately, directional communications complicate the link establishment, and the indispensable beam training is time-consuming [3]. To support high mobility in mmWave systems, which enables widespread applications, such as vehicular communications and wireless virtual reality [4], [5], there is an urgent need to overcome a critical challenge in mmWave communications: how to improve the reliability and reduce the latency of beam training. In this study, we design a novel system architecture by the cross-design of analog and digital modules and propose a compatible fast beam training scheme via compressed sensing (CS) tools to precisely address the above mentioned challenge.

We now recall that the practical implementation of mmWave systems faces several hardware challenges compared to sub-6 GHz communication systems [6]. The fully digital beamforming (DBF) architecture [7], [8], compared with other architectures, has the highest precoding freedom, flexible multi-beam ability, fast beam steering speed, and high beamforming precision. However, the comparison result in [8] showed that the total cost and power consumption of the fully DBF architecture is more than twice of that of hybrid architectures. The hybrid architecture utilizes a lens topology [9] or is connected to a bank of phase shifters to reduce the number of RF chains [10], so as to maintain the hardware cost and power consumption at reasonable levels. However, the signal processing becomes complicated, since it is split between the analog and digital domains in the hybrid architecture. In [11], different hybrid architectures were compared in terms of power consumption and performance.

Recently, to further reduce the power consumption and computational complexity, quantized multiple-input multiple-output (MIMO) systems were proposed in [12] that use low-cost low-resolution analog-to-digital converters (ADCs), e.g., 1-4 bits. Several aspects of quantized MIMO systems have been studied, such as capacity analysis [13], energy efficiency analysis [14], data detection [15], and channel estimation [16]. However, only very few works in the literature have investigated hardware imperfections in the DBF or hybrid architectures [17], [18], including phase noise, power amplifier nonlinearities, and In-phase and Quadrature-phase (IQ) imbalance [19], [20], which can seriously undermine the system performance especially in mmWave frequencies. The work in [17] considered the aggregate impact of hardware impairments and modeled it as additive Gaussian noise. The study in [18] characterized the superposition of different RF impairments by a more accurate extended error vector magnitude (EEVM) model. Capitalizing on [18], we will take low-resolution phase shifters, low-bit ADCs, and radio frequency (RF) impairments into consideration.

mmWave beam training has become a popular research topic in recent years [21]–[30]. The most straightforward method is the exhaustive search [21], which is prohibitively time-consuming, where the base station (BS) and the mobile station (MS) scan all the angle-of-arrival (AoA) and angle-of-departure (AoD) beam pairs until they find the strongest one. IEEE 802.11ad standardized a beam sector based scheme for beam

Manuscript received August 06, 2019; revised November 14, 2019; accepted December 19, 2019. This work was supported in part by the xxx. The associate editor coordinating the review of this paper and approving it for publication was C. Yuen. *(Corresponding author: Shi Jin)*

Jie Yang, Shi Jin, and Xi Yang are with the National Mobile Communications Research Laboratory, Southeast University, Nanjing, 210096, P. R. China (e-mail: yangjie, jinshi, ouyangxi@seu.edu.cn). Chao-Kai Wen is with the Institute of Communications Engineering, National Sun Yat-sen University, Kaohsiung, 804, Taiwan (e-mail: chaokai.wen@mail.nsysu.edu.tw). Michail Matthaiou is with the Institute of Electronics, Communications and Information Technology (ECIT), Queen's University Belfast, Belfast, BT3 9DT, U.K. (email: m.matthaiou@qub.ac.uk).

training [22], whose main idea is to start with sectors of wide beams to do a coarse beam estimation and then shrink the beamwidth adaptively and successively to obtain a more refined beam. Most works considered sequential and single-directional scanning during the beam training [23]–[26], while, work in [27] leveraged the multi-directional scanning abilities of the hybrid transceivers to accelerate the beam training process. The drawback of such methods, however, is the need of successive feedback between the BS and the MS, which is difficult to achieve at the initial channel acquisition stage [3]. The sparse nature of mmWave channels in the angular domain motivates the application of CS-based techniques to expedite the beam training process [28]–[30]. However, these solutions mainly focus on the channel estimation performance, which induce high computational complexity and rely on unrealistic hardware assumptions. In our study, to accelerate the beam training, eliminate the demand of feedback, and reduce the algorithmic complexity, we develop a more flexible beam training method, by embedding the idea of CS into the sector-based beam training scheme to change the fixed hierarchical sectors into adaptive refined beams.

In this study, we investigate the problem of beam training acceleration by proposing a novel low-cost DBF module assisted hybrid (DA-hybrid) architecture and a compatible beam training method (DAH-BT). According to the standards 3GPP TS 38.211 [31] and 3GPP TS 38.213 [32], the beam training is conducted within the periodic half frames, which can occupy high proportion of the total transmission process in 5G NR, especially for fast-moving communication scenarios. Hence, it is fundamentally important to greatly reduce the time of beam training by adding a low-cost auxiliary module. In contrast to almost all of the approaches in the literature [23]–[27], our design exploits both the abilities of the hybrid module to generate desired multiple beams and the low-cost DBF module to accurately capture the angular information. To make the proposed architecture more practically viable, we take into account the hardware imperfections, including RF impairments, low-bit ADCs, and low-resolution phase shifters, and develop robust algorithms and reduce the hardware cost and power consumption. The main contributions of our study are as follows:

- **Hardware Architecture (DA-hybrid)**: We propose a novel low-cost DBF module assisted hybrid architecture, named DA-hybrid, where we deploy coarse RF chains and low-resolution ADCs in the low-cost DBF module. We also model the RF chain impairments and low-bit quantization by the EEVM model and the additive quantization noise model, respectively. We provide a comprehensive cost and power consumption comparison of different architectures with respect to the price and power consumption information of commercial key devices. Through this numerical comparison, the DA-hybrid architecture composed of sub-connected hybrid module and low-cost DBF module is proven to be a promising solution for a good tradeoff between beam training performance and hardware cost.
- **Accelerate Beam Training (DAH-BT)**: We design a fast beam training method based on the proposed DA-hybrid architecture, named DAH-BT. We avail of the sparsity of mmWave channels in the angular domain to quickly estimate the AoAs and AoDs of beams via the adapted orthogonal matching pursuit (OMP) algorithm. Then, we develop an internal calibration method to obtain the parameters of RF impairments in the EEVM model before beam training, which simplifies the CS-based beam training algorithms by avoiding the relatively complex bi-linear techniques. To evaluate the adapted OMP algorithm, we also prove that the designed measurement matrices satisfy the restricted isometry property (RIP).

Extensive simulation results show that the DA-hybrid architecture composed of the sub-connected hybrid module and the low-cost DBF module can not only achieve close to 100% beam matching accuracy, but also reduce the system power consumption and cost. In addition, the proposed DAH-BT shows great potential in saving the time resources over traditional methods [25], [27] with comparable spectral efficiency.

The rest of this paper is organized as follows: The system model is presented in Section II. In Section III, we propose the DA-hybrid architecture, and analyze the cost and power consumption. In Section IV, we design a fast beam training method DAH-BT by utilizing the proposed system architecture. Our simulation results are presented in Section V.We conclude the study in Section VI.

Notations—Throughout this paper, uppercase boldface $\mathbf{A}$ and lowercase boldface $\mathbf{a}$ denote matrices and vectors, respectively. For any matrix $\mathbf{A}$, the superscripts $\mathbf{A}^*$, $\mathbf{A}^T$ and $\mathbf{A}^H$ stand for the conjugate, transpose and conjugate-transpose, respectively. A diagonal matrix is denoted by $\mathrm{diag}\{\cdot\}$ with diagonal entries given in $\{\cdot\}$, and $\mathrm{blkdiag}\{\mathbf{A}_1, \mathbf{A}_2, \cdots, \mathbf{A}_k\}$ denotes a block-diagonal matrix constructed by $\mathbf{A}_1$, $\mathbf{A}_2$, and $\mathbf{A}_k$. For any vector $\mathbf{a}$, $\mathbf{a}^*$ represents the conjugate, and the 2-norm is denoted by $\|\mathbf{a}\|_2$. The quantitative function is denoted by $Q(\cdot)$ and the vector operator is denoted by $\mathrm{vec}(\cdot)$. In addition, the Kronecker product is represented by $\otimes$, while $z \sim \mathcal{CN}(0, \sigma^2)$ denotes a complex-valued Gaussian random variable z with zero mean and variance σ^2. For any real numbers a and b, $\lfloor a \rfloor$ denotes the largest integer no greater than a, $\lceil a \rceil$ denotes the smallest integer no smaller than a, and $\mathrm{mod}\left\{\frac{a}{b}\right\}$ means the remainder of a being divided by b. For any complex number c, $|c|$ represents the modulus of c. For any set $\mathbb{S}$, $|\mathbb{S}|$ denotes the number of elements in set $\mathbb{S}$.

II. System Model

A. Signal Model

We consider a mmWave system in Fig. 1, where the BS and the MS are equipped with a uniform planar array (UPA) with $N_t = N_t^{az} \times N_t^{el}$ and $N_r = N_r^{az} \times N_r^{el}$ antennas,[1] respectively. The UPA is placed on the $y-z$ plane. We assume a narrowband point-to-point flat fading channel, and the received signal at the BS antenna array can be expressed as

$$\mathbf{y}^{\mathtt{UL}} = \mathbf{H}^{\mathtt{UL}}\mathbf{x}^{\mathtt{UL}} + \mathbf{n}^{\mathtt{UL}}, \tag{1}$$

where $\mathbf{H}^{\mathtt{UL}} \in \mathbb{C}^{N_t \times N_r}$ denotes the uplink channel matrix between the MS and the BS. The received signal at the MS antenna array is given by

$$\mathbf{y}^{\mathtt{DL}} = \mathbf{H}^{\mathtt{DL}}\mathbf{x}^{\mathtt{DL}} + \mathbf{n}^{\mathtt{DL}}, \tag{2}$$

[1]Throughout the paper, the superscripts "az" and "el" will be used to denote the azimuth and elevation domains, correspondingly. Likewise, the superscripts "$\mathtt{UL}$" and "$\mathtt{DL}$" will denote the uplink and downlink, respectively.

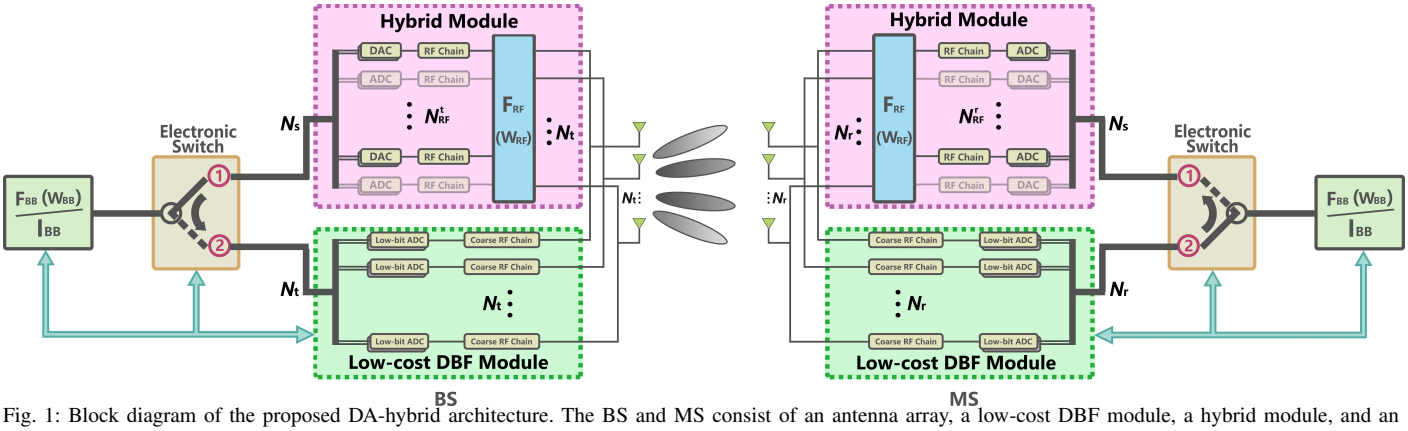


Fig. 1: Block diagram of the proposed DA-hybrid architecture. The BS and MS consist of an antenna array, a low-cost DBF module, a hybrid module, and an electronic switch.

where $\mathbf{H}^{\text{DL}} \in \mathbb{C}^{N_r \times N_t}$ represents the downlink channel matrix; $\mathbf{x}^{\text{UL}} \in \mathbb{C}^{N_r \times 1}$ and $\mathbf{x}^{\text{DL}} \in \mathbb{C}^{N_t \times 1}$ are the transmitted signal at the antenna array. Moreover, $\mathbf{n}^{\text{UL}}$ and $\mathbf{n}^{\text{DL}}$ are the additive complex Gaussian noise vectors. With the assumption of channel reciprocity, we have $\mathbf{H}^{\text{UL}} = (\mathbf{H}^{\text{DL}})^T$.

B. Physical Channel Model

Before proceeding, we first specify the commonly used mmWave channel model which characterizes the geometrical structure and limited scattering nature of mmWave channels [2]. The channel model is expressed as

$$\mathbf{H}^{\text{DL}} = \sqrt{\frac{N_r N_t}{L}} \sum_{l=1}^{L} \alpha_l \mathbf{a}_{\text{Rx}}(\theta_l^{az}, \theta_l^{el}) \mathbf{a}_{\text{Tx}}^H(\phi_l^{az}, \phi_l^{el}), \tag{3}$$

where L is the number of paths, α_l is the complex path gain of the l-th path, θ_l^{az} represents the angle between the incident wave of the l-th path and the y axis; θ_l^{el} represents the angle between the incident wave of the l-th path and the z axis; ϕ_l^{az} represents the angle between the transmitted wave of the l-th path and the y axis; ϕ_l^{el} represents the angle between the transmitted wave of the l-th path and the z axis; θ_l^{az}, θ_l^{el}, ϕ_l^{az} and ϕ_l^{el} are modeled as uniformly distributed variables in $[0, \pi)$. Moreover, $\mathbf{a}_{\text{Rx}}(\theta_l^{az}, \theta_l^{el}) \in \mathbb{C}^{N_r \times 1}$ and $\mathbf{a}_{\text{Tx}}(\phi_l^{az}, \phi_l^{el}) \in \mathbb{C}^{N_t \times 1}$ denote the receiving and transmitting UPA steering vectors, respectively. The definition of $\mathbf{a}_{\text{Rx}}(\theta_l^{az}, \theta_l^{el})$ is given by

$$\mathbf{a}_{\text{Rx}}(\theta_l^{az}, \theta_l^{el}) = \mathbf{a}_{\text{Rx}}(\theta_l^{az}) \otimes \mathbf{a}_{\text{Rx}}(\theta_l^{el}), \tag{4}$$

where

$$\mathbf{a}_{\text{Rx}}(\theta_l^{az}) = \frac{1}{\sqrt{N_r^{az}}} \left[1, e^{j\frac{2\pi d}{\lambda}\cos(\theta_l^{az})}, \ldots, e^{j(N_r^{az}-1)\frac{2\pi d}{\lambda}\cos(\theta_l^{az})}\right]^T, \tag{5}$$

and

$$\mathbf{a}_{\text{Rx}}(\theta_l^{el}) = \frac{1}{\sqrt{N_r^{el}}} \left[1, e^{j\frac{2\pi d}{\lambda}\cos(\theta_l^{el})}, \ldots, e^{j(N_r^{el}-1)\frac{2\pi d}{\lambda}\cos(\theta_l^{el})}\right]^T, \tag{6}$$

where λ is the wavelength of the carrier and d is the distance between two neighboring antennas [33]–[35]. Likewise, by replacing θ_l^{az}, θ_l^{el}, N_r^{az}, and N_r^{el} in (4), (5), and (6) with ϕ_l^{az}, ϕ_l^{el}, N_t^{az}, and N_t^{el}, we can obtain the expression of $\mathbf{a}_{\text{Tx}}(\phi_l^{az}, \phi_l^{el})$. Note that, generally, we consider half-wavelength spaced UPAs in the remainder of the study.

C. Virtual Channel Representation

Unfortunately, it is difficult to analyze and estimate the physical model $\mathbf{H}^{\text{DL}}$ in (3) due to the nonlinear relationship of parameters $\{\alpha_l, \theta_l^{az}, \theta_l^{el}, \phi_l^{az}, \phi_l^{el}\}$. However, the physical channel model can be well approximated by the virtual channel model, which is a linear representation in the angular domain [36]. To harness the tractable virtual channel representation, we assume that the AoAs and AoDs are taken from angular grids of G_r^{az}, G_r^{el}, G_t^{az} and G_t^{el} points in $[0, \pi)$, respectively. The discrete angles are given by

$$\bar{\theta}_i^{az} = \arccos(2(i-1)/G_r^{az} - 1), \quad i = 1, 2, \ldots, G_r^{az}, \tag{7}$$

$$\bar{\theta}_i^{el} = \arccos(2(i-1)/G_r^{el} - 1), \quad i = 1, 2, \ldots, G_r^{el}, \tag{8}$$

$$\bar{\phi}_j^{az} = \arccos(2(j-1)/G_t^{az} - 1), \quad j = 1, 2, \ldots, G_t^{az}, \tag{9}$$

$$\bar{\phi}_j^{el} = \arccos(2(j-1)/G_t^{el} - 1), \quad j = 1, 2, \ldots, G_t^{el}. \tag{10}$$

Let $G_r = G_r^{az} \times G_r^{el}$ and $G_t = G_t^{az} \times G_t^{el}$. Then, we define the receiving and transmitting dictionary matrices $\bar{\mathbf{A}}_r$ and $\bar{\mathbf{A}}_t$, with $\bar{\mathbf{A}}_r \in \mathbb{C}^{N_r \times G_r}$ being given by

$$\bar{\mathbf{A}}_r = \bar{\mathbf{A}}_r^{az} \otimes \bar{\mathbf{A}}_r^{el}, \tag{11}$$

where

$$\bar{\mathbf{A}}_r^{az} = \left[\mathbf{a}_{\text{Rx}}(\bar{\theta}_1^{az}), \mathbf{a}_{\text{Rx}}(\bar{\theta}_2^{az}), \ldots, \mathbf{a}_{\text{Rx}}(\bar{\theta}_{G_r^{az}}^{az})\right], \tag{12}$$

and

$$\bar{\mathbf{A}}_r^{el} = \left[\mathbf{a}_{\text{Rx}}(\bar{\theta}_1^{el}), \mathbf{a}_{\text{Rx}}(\bar{\theta}_2^{el}), \ldots, \mathbf{a}_{\text{Rx}}(\bar{\theta}_{G_r^{el}}^{el})\right]. \tag{13}$$

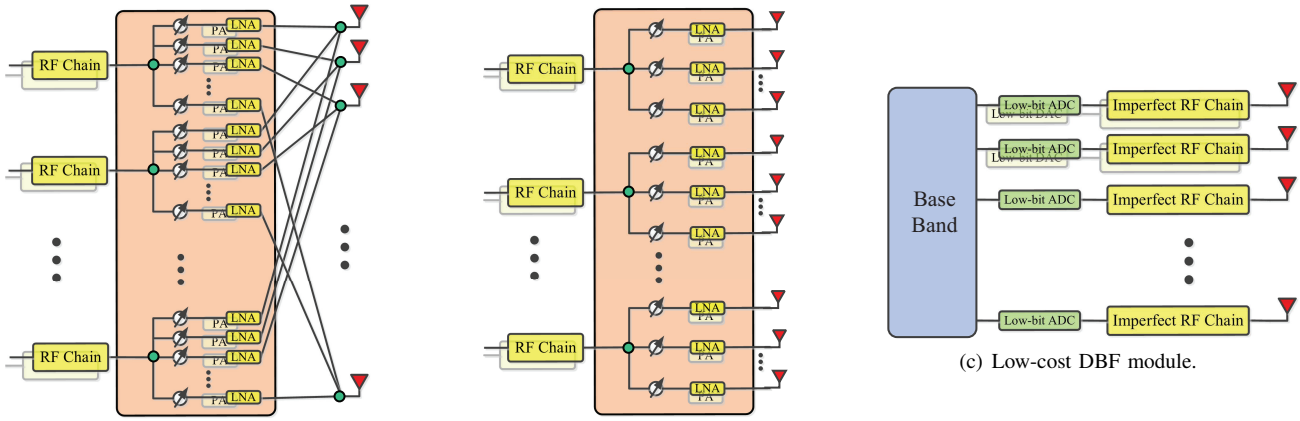
Similarly, by replacing $\mathbf{a}_{\text{Rx}}(\bar{\theta}_i^{az})$ and $\mathbf{a}_{\text{Rx}}(\bar{\theta}_i^{el})$ with $\mathbf{a}_{\text{Tx}}(\bar{\phi}_j^{az})$ and $\mathbf{a}_{\text{Tx}}(\bar{\phi}_j^{el})$ in (11), (12), and (13), we can obtain the expression of $\bar{\mathbf{A}}_t \in \mathbb{C}^{N_t \times G_t}$. Therefore, the virtual representation of $\mathbf{H}^{\text{DL}}$ is given by

$$\mathbf{H}_{\text{v}}^{\text{DL}} = \bar{\mathbf{A}}_r \mathbf{H}_a \bar{\mathbf{A}}_t^H, \tag{14}$$

where $\mathbf{H}_a$ is the virtual channel coefficients matrix (which will be defined in Section IV-B). The virtual channel $\mathbf{H}_{\text{v}}^{\text{DL}}$ approximates the physical channel $\mathbf{H}^{\text{DL}}$, and the approximation error will decrease as G_r and G_t increase. For a particular system architecture, there is a resolvable phase resolution. In this study, we assume that the resolution of the phase is π/N, hence, G_r^{az}, G_r^{el}, G_t^{az}, and G_t^{el} should be no greater than N. The virtual channel effectively captures the underlying sparse multipath environment comprising L physical paths through $G_r G_t$ resolvable paths.

III. Hardware Architecture of the Proposed Transceiver System

In this section, we propose a low-cost DBF module assisted hybrid architecture. In addition, we compare the cost and power consumption of the proposed DA-hybrid architecture with the traditional DBF and hybrid architectures.



(a) Fully-connected hybrid module: each RF chain is connected to all the antennas.

(b) Sub-connected hybrid module: each RF chain is connected to a subset of antennas.

(c) Low-cost DBF module.

Fig. 2: Illustration of different architectures.

A. System Topology and Frame Structure

As was previously mentioned, the fully DBF architectures have a number of performance advantages, however, the hardware cost and power consumption are viewed as their major constrains [4]. The hybrid architectures keep the cost and complexity to affordable levels by using fewer RF chains compared to the number of antennas and allow multi-terminal multi-stream precoding [2], while the signal processing becomes complicated. To leverage the advantages of both the DBF and hybrid architectures, we propose the DA-hybrid architecture, as illustrated in Fig. 1. The DA-hybrid architecture is composed of the following modules: an antenna array, a low-cost DBF module, a hybrid module, and an electronic switch.

•**The antenna array** is a UPA with N_t and N_r antennas in the BS and MS, respectively. Each antenna array is connected to both the low-cost DBF module and the hybrid module.

•**The hybrid module** consists of a digital precoding (combining) matrix $\mathbf{F}_{\mathrm{BB}} \in \mathbb{C}^{N_{RF}^t \times N_s}$ ($\mathbf{W}_{\mathrm{BB}} \in \mathbb{C}^{N_{RF}^t \times N_s}$) and an analog precoding (combining) matrix $\mathbf{F}_{\mathrm{RF}} \in \mathbb{C}^{N_t \times N_{RF}^t}$ ($\mathbf{W}_{\mathrm{RF}} \in \mathbb{C}^{N_t \times N_{RF}^t}$) at the BS, or a digital precoding (combining) matrix $\mathbf{F}_{\mathrm{BB}} \in \mathbb{C}^{N_{RF}^r \times N_s}$ ($\mathbf{W}_{\mathrm{BB}} \in \mathbb{C}^{N_{RF}^r \times N_s}$) and an analog precoding (combining) matrix $\mathbf{F}_{\mathrm{RF}} \in \mathbb{C}^{N_r \times N_{RF}^r}$ ($\mathbf{W}_{\mathrm{RF}} \in \mathbb{C}^{N_r \times N_{RF}^r}$) at the MS, where N_{RF}^t and N_{RF}^r represents the number of RF chains at the BS and MS, respectively, and N_s denotes the number of baseband data streams, with $N_s \leq N_{RF}^t \leq N_t$ and $N_s \leq N_{RF}^r \leq N_r$. Each RF chain carries a pair of amplitude-domain IQ high-resolution ADCs. Fig. 2(a) and Fig. 2(b) show the fully-connected and the sub-connected hybrid architectures, respectively. The signal transmitted by the hybrid architecture in a time slot is given by

$$\mathbf{x} = \mathbf{F}_{\mathrm{RF}}\mathbf{F}_{\mathrm{BB}}\mathbf{s}, \tag{15}$$

where $\mathbf{s} \in \mathbb{C}^{N_s \times 1}$ is the baseband data sequence. Through ingenious design of $\mathbf{F}_{\mathrm{RF}}$, $\mathbf{F}_{\mathrm{BB}}$ and $\mathbf{s}$, the beam can point at a particular direction, or cover the whole space [27]. Note that $\mathbf{F}_{\mathrm{RF}}$ is implemented by analog phase shifters, such that its entries are of constant modulus. Commercially used RF phase shifters generally have 6-bit resolution, hence, in this study, we assume that RF phase shifters in the hybrid modules have 6-bit phase resolution [37], without amplitude adjustment. Therefore, the phase resolution parameter N is $2^6 = 64$. According to (7)-(10) in Section II-C, the angular space of beams is meshed, with G_r^{az}, G_r^{el}, G_t^{az} and G_t^{el} no greater than 64, such that G_r (G_t) beam directions are stored in the beam dictionary matrix $\bar{\mathbf{A}}_r$ ($\bar{\mathbf{A}}_t$). We can design the analog precoding matrix $\mathbf{F}_{\mathrm{RF}}$ based on beam dictionary matrix (by combining different column vectors of the beam dictionary matrix). The hybrid architecture can reduce the number of RF chains, maintain the hardware cost and power consumption at reasonable levels, and guarantee satisfactory transmission performance. However, it splits the signal processing into the analog domains ($\mathbf{F}_{\mathrm{RF}}$ and $\mathbf{W}_{\mathrm{RF}}$) and digital domains ($\mathbf{F}_{\mathrm{BB}}$ and $\mathbf{W}_{\mathrm{BB}}$), which results in high computational complexity. Due to the complicated signal processing and the lack of RF chains, beam training has to rely on multi-level scanning and feedback, which are time consuming exercises.

•**The low-cost DBF module** consists of N_t and N_r low-cost coarse RF chains in the BS and MS, respectively. Each RF chain carries a pair of amplitude-domain IQ low-resolution ADCs. The block diagram of the low-cost DBF module is depicted in Fig. 2(c). Note that only the first two antennas have transmission capability. There is no analog combining matrix in the low-cost DBF module, therefore, data from all the antennas is retained directly but with a certain degree of resolution loss. The low-cost DBF module and the hybrid module will not work at the same time slot, hence, the power consumption will not be doubled. We adopt the EEVM model to capture the combined impact of all kinds of RF impairments [18], [19]. For illustration simplicity, let k denote the k-th RF chain of the receiver, $\chi(k) = \kappa(k)e^{j\varphi(k)}$ denotes the attenuation and phase shift effects on the k-th RF chain with $|\chi(k)| \leqslant 1$, and $n_{\mathtt{RF}}(k)$ is the additive Gaussian noise. Take downlink as an example, the downlink received data impaired by imperfect RF chains can be expressed as

$$\mathbf{y}_{\mathtt{RF}}^{\mathtt{DL}} = \boldsymbol{\chi}(\mathbf{H}^{\mathtt{DL}}\mathbf{x}^{\mathtt{DL}} + \mathbf{n}^{\mathtt{DL}}) + \mathbf{n}_{\mathtt{RF}}, \tag{16}$$

where $\boldsymbol{\chi} = \mathrm{diag}\{\chi(1), \ldots, \chi(N_r)\}$, and $\mathbf{n}_{\mathtt{RF}} = [n_{\mathtt{RF}}(1), \ldots, n_{\mathtt{RF}}(N_r)]^T$. Then, we use the additive quantization noise model to formulate the coarsely quantized outputs of low-resolution ADCs with ideal automatic gain control [13]. The quantized signal is expressed as

$$\mathbf{r}_{\mathtt{ADC}}^{\mathtt{DL}} = Q(\mathbf{y}_{\mathtt{RF}}^{\mathtt{DL}}) = \eta\mathbf{y}_{\mathtt{RF}}^{\mathtt{DL}} + \mathbf{n}_{\mathtt{q}}, \tag{17}$$

where $Q(\cdot)$ represents the non-linear quantitative function, which can be modeled as a linear function [38]. Let b denote the quantization bits of the ADC, then the value of η for $b \leqslant 5$

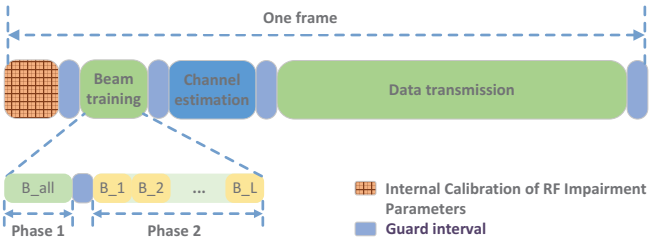


Fig. 3: Illustration of the frame structure, which consists of internal calibration phase, beam training phase, channel estimation phase, data transmission phase, and guard intervals between different phases. The beam training phase consists of two sub-phases.

can be found in [13]: when $b = 1$, $\eta = 0.6366$; when $b = 2$, $\eta = 0.8825$; when $b = 4$, $\eta = 0.9905$. Note that $\mathbf{n_q}$ denotes the additive Gaussian quantization noise vector. Although we apply coarse RF chains and low-bit ADCs to reduce the cost and power consumption, the low-cost DBF module still manages to retain some advantages of the ideal DBF module, e.g., a high enough number of RF chains can make the received measurements sparser, hence, by using compressed sensing algorithms, beam directions can be estimated even at low operating SNRs.

•**The electronic switch** is strictly controlled by the frame structure to make the system switch between the hybrid module and the low-cost DBF module.

The proposed frame structure is illustrated in Fig. 3: The first phase is the internal calibration phase (introduced in detail in Section IV-A), during which both of the BS and the MS compute the values of the RF impairments' parameters of their own low-cost DBF module. The next is the beam training phase (introduced in detail in Section IV-B), which is divided into two phases: in phase 1, the BS switches to the hybrid module to omni-directionally transmit the beam training sequence, and the MS switches to the low-cost DBF module to obtain AoAs by the standard CS algorithm[2]. In phase 2, by taking advantage of channel reciprocity, the MS switches to the hybrid module to single-directionally transmit beam training sequences along the estimated beam directions successively, and the BS switches to the low-cost DBF module to obtain AoDs by the standard CS algorithm. Finally, during the channel estimation phase and the data transmission phase, both of the BS and MS switch to the hybrid module.

B. Cost and Power Consumption Analysis

In this section, we analyze the cost and power consumption of the following architectures: the fully-connected hybrid architecture (A1), the sub-connected hybrid architecture (A2), the fully DBF architecture (A3) [2], and the low-cost DBF architecture (A4). The previous work in [11] summarizes the power consumption of the ADC and the phase shifter, however, these values are relatively low since the considered devices' designs are not commercial products. Here, we assume that the transceiver operates at the 28-GHz band with a 500-MHz signal bandwidth, and we refer to commercial products in Applied Dynamics International (ADI) company website that meet the above conditions. With the development of chip technology, the price and power consumption of products will change, but the trend we revealed in this study is valuable for reference. We pursue the analysis on the receiver side as an example, and the required devices are summarized in Table I. The choice of the reference value is difficult, since the cost and power consumption for ADI devices present high variability, e.g., the price of one ADC device with around 1Gs/s sampling rate ranges from \$292.19 to \$1326.00, and the power consumption ranges from 0.71W to 5.1W per ADC device. For A1, A2, and A3, which we assume that good device performance is required, we choose the average price and power consumption as the reference value, e.g., \$809 and 3W per ADC device. For A4, with coarse RF chains and low-bit ADCs, we choose a relatively cheaper device with lower power consumption as reference, e.g., the No. HMCAD1511 8-bit ADC device costs \$300 with 1W power consumption.[3] Similarly, we can obtain the reference cost and power consumption of the low noise amplifier (LNA), mixer, local oscillator (LO), low pass filter (LPF),[4] and base band amplifier (BBA). Since there is no phase shifter in A4, we only choose the average cost and power consumption as the reference values for the phase shifter, by also noting that the previous work [8] considers similar reference values. According to [8], [11], the electronic switch has negligible power consumption and cost.

To compare the cost and power consumption of the considered architectures, we calculate the total cost and power consumption of each architecture based on Table I with $N_r = 64$ and N_{RF}^r as a variable. Since we assume that the baseband processor consumes almost the same power in different modules [11], we can ignore its impact in our analysis without severely sacrificing the modeling accuracy. The simulation result is shown in Fig. 4: (1) A1 and A3 are much more expensive than A2 and A4, and with N_{RF}^r increasing, the cost of A4 becomes lower than A2; (2) the power consumption of A3 is the highest and that of A2 is the lowest among four architectures, nevertheless, the power consumption of A4 will be lower than A2 with N_{RF}^r increasing. (3) the low-cost DBF module assisted sub-connected hybrid architecture (A2+A4) has lower cost and power consumption than A1 and A3. Given the above analysis, as an additional module, A4 has relative low cost and power consumption. Moreover, compared to the low-cost DBF module assisted fully-connected hybrid module, the low-cost DBF module assisted sub-connected hybrid architecture can further reduce the total cost and power consumption with comparable beam training performance shown in Section V-A.

IV. Accelerating Beam Training by the Proposed Transceiver System

The proposed DA-hybrid transceiver architecture can in principle accelerate the beam training process. In this section, we propose a fast beam training method, which only requires $L_{est}+1$ time slots, where L_{est} is the number of the estimated paths. As briefly introduced in Section III-A, there are internal calibration phase and beam training phase in the frame structure. In the following, we will elaborate on these phases.

[2]The low-cost DBF module and the hybrid module share the same antenna array and beam codebook, hence, the angular information obtained by the low-cost DBF module can be directly used for the hybrid module.

[3]Actually, 4 or less bits ADC devices will have lower cost and power consumption [2].

[4]The power consumption of the LPF is relatively low, which is around 0.02W, we directly set the reference value to 0.02W.

TABLE I: Comparison of cost and power consumption for different devices on the receiver side

Device	ADI device cost ($)	Reference cost low/avg ($)	ADI device power (W)	Reference power low/avg (W)	Number in A1	Number in A2	Number in A3	Number in A4
LNA	25.85-223.13	40/125	0.09-0.64	0.09/0.4	N_r	N_r	N_r	N_r
Mixer	25.09-102.91	26/64	0.595-1.442	0.6/2	N^r_{RF}	N^r_{RF}	N_r	N_r
LO	7.63-200.01	8/104	0.03-1.75	0.1/0.9	N^r_{RF}	N^r_{RF}	N_r	N_r
LPF	2.10-17.50	2.5/10	0.02	0.02	N^r_{RF}	N^r_{RF}	N_r	N_r
BBA	1.31-5.65	1.5/3.5	0.02-0.16	0.02/0.09	N^r_{RF}	N^r_{RF}	N_r	N_r
ADC	292.19-1326.00	300/809	0.71-5.1	1/3	N^r_{RF}	N^r_{RF}	N_r	N_r
Phase shifter	170	170	0.4	0.4	$N^r_{RF}N_r$	N_r	-	-

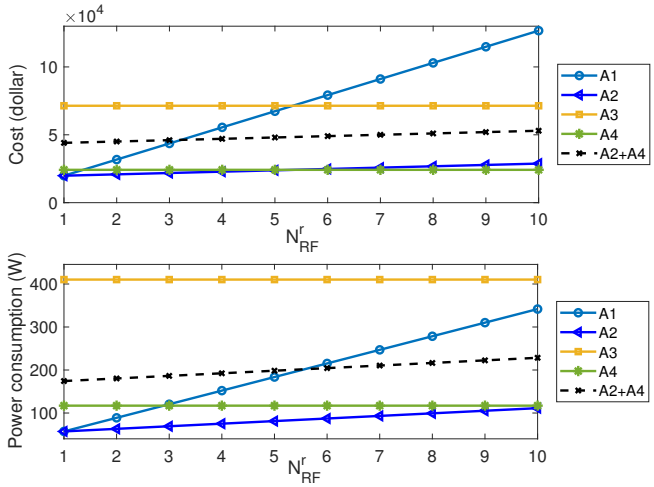


Fig. 4: Cost and power consumption of different architectures versus the number of RF chains N^r_{RF}.

A. Internal Calibration of RF Impairments' Parameters

The RF impairments include phase noise, IQ imbalance and nonlinearities. The EEVM model in (16) generally describes the aggregation of multiple impairments [39]–[41]. In the subsequent beam training phase, we are going to use the OMP algorithm, however, $\boldsymbol{\chi}$ as a part of measurement matrices is unknown. To fulfill the requirements of the OMP algorithm [42], we need to know the values of $\boldsymbol{\chi}$. Since all RF chains in the same low-cost DBF module share one clock, the values of $\boldsymbol{\chi}$ are in fact stable over a long coherence time. Thus, it is possible to internally calibrate the RF impairments' parameters by a similar calibration method in [43]. For calibration, the low-cost DBF module should have transmission capability in at least the first two antennas. Therefore, as shown in Fig. 2(c), the first two antennas carry digital-to-analog converters (DAC). The adapted calibration is summarized in Algorithm 1.

Algorithm 1 Internal Calibration

Step 1: The antenna connected to RF chain 1 sends a training signal $s = 1$.

Step 2: The antennas connected to RF chains numbered $(2, \ldots, N_r)$ receive the signal; after the low-resolution ADC, the quantized received signal is $b_{1\to i}$, where $2 \leqslant i \leqslant N_r$.

Step 3: The antenna connected to RF chain 2 sends training signal $s = 1$.

Step 4: The antennas connected to RF chains numbered $(1, 3, \ldots, N_r)$ receive the signal; after the low-resolution ADC, the quantized received signal is $b_{2\to i}$, where $1 \leqslant i \leqslant N_r$ and $i \neq 2$.

Output:

$$\boldsymbol{\chi}_{cal} = Q(\chi(2))\text{diag}\left\{\frac{b_{2\to1}b_{1\to3}}{b_{1\to2}b_{2\to3}}, 1, \frac{b_{1\to3}}{b_{1\to2}}, \frac{b_{1\to4}}{b_{1\to2}}, \ldots, \frac{b_{1\to N_r}}{b_{1\to2}}\right\}.$$

To be specific, firstly, the antenna connected to RF chain 1 sends a training signal $s = 1$. Then, the antennas connected to RF chains numbered $(2, \ldots, N_r)$ receive the signal; after the low-resolution ADC, the quantized received signal $b_{1\to i}$ is given by

$$b_{1\to i} = Q\left(ht_1\chi(i)\right) \approx ht_1 Q(\chi(i)) \quad 2 \leqslant i \leqslant N_r, \tag{18}$$

where t_1 and $\chi(i)$ denote the unknown RF chain impairments' coefficient of the transmit antenna 1 and the receive antenna i, respectively; h represents the channel between the transmission and receiver antennas. Since the distance between the transmission and receiver antennas is quite close, the effective signal-to-noise ratio (SNR) can be high enough. Therefore, it is reasonable to assume that the Gaussian thermal noise is negligible and the channel between different antennas is the same, denoted as h in (18).

Next, the antenna connected to RF chain 2 sends a training signal $s = 1$. Then, the antennas connected to RF chains numbered $(1, 3, \ldots, N_r)$ receive the signal, similarly, after the low-resolution ADC, the quantized received signal $b_{2\to i}$ is given by

$$b_{2\to i} \approx ht_2 Q(\chi(i)) \quad 1 \leq i \leq N_r, i \neq 2, \tag{19}$$

where t_2 is the unknown RF chain impairments' coefficient of the transmit antenna 2. According to (18) and (19), we can obtain that

$$\frac{Q(\chi(i))}{Q(\chi(2))} \approx \begin{cases} \frac{b_{2\to1}b_{1\to3}}{b_{1\to2}b_{2\to3}} & i = 1, \\ \frac{b_{1\to i}}{b_{1\to2}} & 2 < i \leq N_r. \end{cases} \tag{20}$$

Then, we obtain the relative quantized estimation of RF chain impairments' parameters as

$$\boldsymbol{\chi}_{cal} = Q(\chi(2))\text{diag}\left\{\frac{b_{2\to1}b_{1\to3}}{b_{1\to2}b_{2\to3}}, 1, \frac{b_{1\to3}}{b_{1\to2}}, \frac{b_{1\to4}}{b_{1\to2}}, \ldots, \frac{b_{1\to N_r}}{b_{1\to2}}\right\}. \tag{21}$$

We can assume that $Q(\chi(2)) = 1$ in this study, since the performance of AoA/AoD estimation based on CS algorithms will not deteriorate when each RF chain deviates from the real RF chain impairments by the same multiplicative factor. If more accurate results are needed, it is not difficult to measure the value of $Q(\chi(2))$ by external measuring facilities.

B. Fast Beam Training Method

In this section, we propose a fast beam training method by utilizing the proposed DA-hybrid architecture, therefore, we name the proposed beam training method DAH-BT. Since the MS can be devices which are equipped with large antenna arrays, such as vehicles, in this study we assume that $N_t \gg L$ and $N_r \gg L$.[5] Our design leverages a simple observation: the angular equivalent channel matrix $\mathbf{H}_a$ is sparse and there is a one-to-one relationship between AoAs/AoDs and the indexes of entries in $\mathbf{H}_a$.

Definition 1. *Let $\mathbb{S}_d$ denote the set of dominant entries in $\mathbf{H}_a$, where $\mathbb{S}_d = \{(i,k) : |\mathbf{H}_a(i,k)| > \epsilon\}$, and ϵ is an appropriately*

[5]To be more practical, in simulations, we consider less RF chains in the MS than BS.

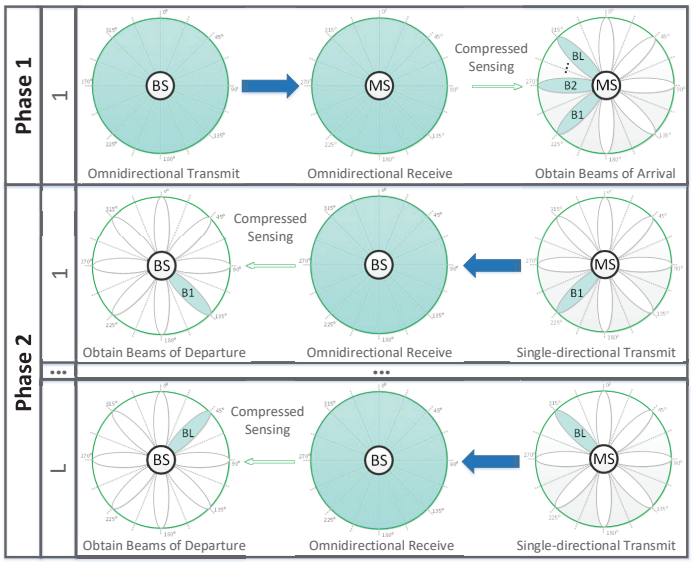


Fig. 5: Illustration of the proposed DAH-BT method, where L_{est} is the number of the estimated paths.

chosen threshold. If $d = |\mathbb{S}_d| \ll G_r G_t$, we say that $\mathbf{H}_a$ is effectively d-sparse.

The work in [28] deduced the expression of $\mathbf{H}_a$ with uniform linear array (ULA) deployed at both BS and MS. We can infer that the (i,k)-th entry of $\mathbf{H}_a$ for UPA equipped at both BS and MS is given by

$$\begin{aligned}\mathbf{H}_a(i,k) \approx \sum_{l=1}^{L} \alpha_l f_{N_r^{az}}\left(\frac{m(i)}{N_r^{az}} - \theta_l^{az}\right) f_{N_r^{el}}\left(\frac{n(i)}{N_r^{el}} - \theta_l^{el}\right) \\ \times f^*_{N_r^{az}}\left(\frac{p(k)}{N_t^{az}} - \phi_l^{az}\right) f^*_{N_r^{az}}\left(\frac{q(k)}{N_t^{el}} - \phi_l^{el}\right),\end{aligned} \tag{22}$$

where $m(i) \triangleq \left\lfloor \frac{i}{G_r^{el}} \right\rfloor + 1$, $n(i) \triangleq \mathrm{mod}\left\{\frac{i}{G_r^{el}}\right\}$, $p(k) \triangleq \left\lfloor \frac{k}{G_t^{el}} \right\rfloor + 1$, $q(k) \triangleq \mathrm{mod}\left\{\frac{k}{G_t^{el}}\right\}$; $f_{N_r^{az}}(\cdot)$, $f_{N_r^{el}}(\cdot)$, $f_{N_t^{az}}(\cdot)$ and $f_{N_t^{el}}(\cdot)$ are the Dirichlet kernels, where $f_N(\theta) = \frac{1}{N}\sum_{i=1}^{N} e^{-j2\pi i\theta}$. According to the summation formula of geometric sequences, $f_N(\theta) = \frac{1}{N} e^{-j\pi(N+1)\theta} \frac{\sin(N\pi\theta)}{\sin(\pi\theta)}$. The property of the sinc function infers that the indexes of dominant entries in $\mathbf{H}_a$ are corresponding to AoA/AoD pairs, which can be calculated according to (7)-(10), and (22). Then, it is obvious that $\mathbf{H}_a$ is effectively L-sparse due to L paths. Therefore, we can find the matching beams by utilizing standard CS techniques to coarsely reconstruct $\mathbf{H}_a$.

The schematic diagram of the DAH-BT procedure is depicted in Fig. 5. Simply put, the DAH-BT is divided into two phases[6]: during phase 1, the BS omni-directionally transmits the training signal, and the MS omni-directionally receives the signal, then, via a standard compressed sensing algorithm, the MS estimates the AoAs of the beams; during phase 2, by exploiting channel reciprocity, the MS single-directionally transmits training signals along the estimated beam directions successively, and the BS omni-directionally receives the signal and estimates the AoDs of the beams by a method similar to phase 1. Finally, the beams are matched. A more detailed introduction of the beam training is given as follows.

[6]Considering that the maximum uplink transmit power is less than that of downlink, it is reasonable to design omnidirectional transmission at the BS and directional transmission at the MS.

1) DAH-BT Phase 1: The BS omni-directionally sends the training signal, as shown in Fig. 5. The transmitted data is given by

$$\mathbf{x}^{\mathtt{DL}} = P_t[1, 0, \ldots, 0]^T, \tag{23}$$

where $\mathbf{x}^{\mathtt{DL}} \in \mathbb{C}^{N_t \times 1}$ and P_t is the transmit power, which is set to 1 in this study. The signal received at the MS by the low-cost DBF module is given by

$$\mathbf{y}_{\mathtt{RF}}^{\mathtt{DL}} = \boldsymbol{\chi}_{\mathtt{ms}}(\mathbf{H}^{\mathtt{DL}}\mathbf{x}^{\mathtt{DL}} + \mathbf{n}^{\mathtt{DL}}) + \mathbf{n}_{\mathtt{RF}} = \boldsymbol{\chi}_{\mathtt{ms}}\mathbf{H}^{\mathtt{DL}}\mathbf{x}^{\mathtt{DL}} + \mathbf{n}_{\mathtt{eff}}^{\mathtt{DL}}, \tag{24}$$

where $\boldsymbol{\chi}_{\mathtt{ms}}$ contains the RF chain impairments' parameters at the MS, and $\mathbf{n}_{\mathtt{eff}}^{\mathtt{DL}} \triangleq \boldsymbol{\chi}_{\mathtt{ms}}\mathbf{n}^{\mathtt{DL}} + \mathbf{n}_{\mathtt{RF}}$ is still the additive Gaussian noise. Then, the received signal quantized by the low-bit ADC can be written as

$$\mathbf{r}_{\mathtt{ADC}}^{\mathtt{DL}} = Q\left(\boldsymbol{\chi}_{\mathtt{ms}}\mathbf{H}^{\mathtt{DL}}\mathbf{x}^{\mathtt{DL}} + \mathbf{n}_{\mathtt{eff}}^{\mathtt{DL}}\right) = \eta\boldsymbol{\chi}_{\mathtt{ms}}\mathbf{H}^{\mathtt{DL}}\mathbf{x}^{\mathtt{DL}} + \mathbf{n}_{\mathtt{e}}^{\mathtt{DL}}, \tag{25}$$

where $\mathbf{n}_{\mathtt{e}}^{\mathtt{DL}} \triangleq \eta\mathbf{n}_{\mathtt{eff}}^{\mathtt{DL}} + \mathbf{n}_{\mathtt{q}}$, and we assume that $\mathbf{n}_{\mathtt{e}}^{\mathtt{DL}} \sim \mathcal{CN}(0, \sigma_{\mathtt{dl,e}}^2\mathbf{I})$. Note that a Gaussian distribution is the worst-case assumption. Here, we define the downlink receive SNR as

$$\mathrm{SNR}^{\mathtt{DL}} = 10\log_{10}\left(\|\eta\boldsymbol{\chi}_{\mathtt{ms}}\mathbf{H}^{\mathtt{DL}}\mathbf{x}^{\mathtt{DL}}\|_2^2/(N_r\sigma_{\mathtt{dl,e}}^2)\right). \tag{26}$$

According to (14), by approximately replacing $\mathbf{H}^{\mathtt{DL}}$ with $\mathbf{H}_{\mathtt{v}}^{\mathtt{DL}} = \bar{\mathbf{A}}_r\mathbf{H}_a\bar{\mathbf{A}}_t^H$, we obtain

$$\mathbf{r}_{\mathtt{ADC}}^{\mathtt{DL}} = \eta\boldsymbol{\chi}_{\mathtt{ms}}\bar{\mathbf{A}}_r\mathbf{H}_a\bar{\mathbf{A}}_t^H\mathbf{x}^{\mathtt{DL}} + \mathbf{n}_{\mathtt{e}}^{\mathtt{DL}}. \tag{27}$$

Since $\frac{\sqrt{N_t}}{G_t}\bar{\mathbf{A}}_t\mathbf{1} = [1, 0, 0, \ldots, 0]^T$, where $\mathbf{1}$ represents a vector in which all elements are one, we make $\mathbf{x}^{\mathtt{DL}} = \frac{\sqrt{N_t}}{G_t}\bar{\mathbf{A}}_t\mathbf{1}$. Then, we have

$$\mathbf{r}_{\mathtt{ADC}}^{\mathtt{DL}} = \eta\frac{\sqrt{N_t}}{G_t}\boldsymbol{\chi}_{\mathtt{ms}}\bar{\mathbf{A}}_r\mathbf{h}_v^{\mathtt{DL}} + \mathbf{n}_{\mathtt{e}}^{\mathtt{DL}}, \tag{28}$$

where $\mathbf{h}_v^{\mathtt{DL}} \triangleq \mathbf{H}_a\bar{\mathbf{A}}_t^H\bar{\mathbf{A}}_t\mathbf{1}$. Since $\bar{\mathbf{A}}_t^H\bar{\mathbf{A}}_t \approx \mathbf{I}$, we have $\mathbf{h}_v^{\mathtt{DL}} \approx \mathbf{H}_a\mathbf{1}$. Hence, the row index of $\mathbf{h}_v$ corresponds to the AoAs in grid. Denote $\boldsymbol{\Phi} \triangleq \eta\frac{\sqrt{N_t}}{G_t}\boldsymbol{\chi}_{\mathtt{ms}}\bar{\mathbf{A}}_r$, then, (28) can be simplified into

$$\mathbf{r}_{\mathtt{ADC}}^{\mathtt{DL}} = \boldsymbol{\Phi}\mathbf{h}_v^{\mathtt{DL}} + \mathbf{n}_e^{\mathtt{DL}}. \tag{29}$$

According to the Definition 1 and the property of $\mathbf{H}_a$, $\mathbf{h}_v^{\mathtt{DL}}$ also has a sparse nature. Hence, the equation (29) fits with the CS model. Nevertheless, we also need to verify that the derived matrix $\boldsymbol{\Phi}$ in (29) meets the requirements of the measurement matrix in CS. To analyze the property of $\boldsymbol{\Phi}$, we should review some concepts of CS. In particular, one fundamental property of the measurement matrix that has been very useful in proving the optimality of CS reconstruction procedures is the RIP [44], [45].

Definition 2. *Let $\mathbf{A}$ be an $p \times q$ matrix with $p < q$. If there exists $\delta_k \in (0,1)$, for all $\mathbf{x}$ with $\|\mathbf{x}\|_0 \leqslant k$, the following inequality holds, we say that the matrix $\mathbf{A}$ satisfies the k-order RIP:*

$$(1-\delta_k)\|\mathbf{x}\|_2^2 \leqslant \|\mathbf{A}\mathbf{x}\|_2^2 \leqslant (1+\delta_k)\|\mathbf{x}\|_2^2, \tag{30}$$

where $\|\cdot\|_0$ counts the number of nonzero entries of a vector.

Note that RIP can be explained as approximately preserving the distance of the k-sparse vector pairs. If a matrix meets the RIP condition, many algorithms will have great chance to recover a sparse signal from measurement with noise successfully [28]. A. C. Gilbert *et al.* have proven that partial Fourier matrices satisfy the RIP [46]. The partial Fourier matrices are formed by randomly choosing P rows from $N \times N$ Fourier matrices, and normalizing each column. It is obvious that $\bar{\mathbf{A}}_t$ and $\bar{\mathbf{A}}_r$ are partial Fourier matrices. Then, we will prove that the measurement matrix $\boldsymbol{\Phi}$ satisfies the RIP.

Lemma 1. *Let $\boldsymbol{\Phi} = \eta\frac{\sqrt{N_t}}{G_t}\boldsymbol{\chi}_{\mathtt{ms}}\bar{\mathbf{A}}_r$, where the coefficient*

$\eta\frac{\sqrt{N_t}}{G_t} \leqslant 1$, *the entries in* $\boldsymbol{\chi}_{\mathtt{ms}}$ *satisfy* $0 < |\chi_{\mathtt{ms}}(i)| < 1$, *and* $\bar{\mathbf{A}}_r$ *is the matrix defined in* (11). *Then, for an effectively* d*-sparse vector* $\mathbf{h}_v$, *there exists* $\delta_d \in (0,1)$ *such that*

$$(1-\delta_d)\|\mathbf{h}_v^{\mathtt{DL}}\|_2^2 \leqslant \|\boldsymbol{\Phi}\mathbf{h}_v^{\mathtt{DL}}\|_2^2 \leqslant (1+\delta_d)\|\mathbf{h}_v^{\mathtt{DL}}\|_2^2. \tag{31}$$

Namely, $\boldsymbol{\Phi}$ *satisfies the* d*-order RIP.*

Proof. Please refer to Appendix A. □

Hence, with the knowledge of $\boldsymbol{\Phi}$ and $\mathbf{r}_{\mathtt{ADC}}^{\mathtt{DL}}$, we can estimate $\mathbf{h}_v^{\mathtt{DL}}$ by the standard OMP algorithm. We denote the estimation of $\mathbf{h}_v^{\mathtt{DL}}$ as $\hat{\mathbf{h}}_v^{\mathtt{DL}}$. There are L_{est} entries in $\hat{\mathbf{h}}_v^{\mathtt{DL}}$ with non-vanishing power, where L_{est} is the number of estimated paths. To prevent some paths from being missed, we take L_{est} slightly larger than L in simulations. Based on the indexes of these L_{est} entries, we find the azimuth and elevation AoAs of the corresponding paths from the discrete angles, according to (7), (8), and (22). The set of estimated AoAs is denoted as

$$\mathbb{S}_{AoA}=\Big[(\bar{\theta}_{AoA_1}^{az},\bar{\theta}_{AoA_1}^{el}),(\bar{\theta}_{AoA_2}^{az},\bar{\theta}_{AoA_2}^{el}),\ldots,(\bar{\theta}_{AoA_{L_{est}}}^{az},\bar{\theta}_{AoA_{L_{est}}}^{el})\Big].$$

The process of beam training phase 1 is summarized in Algorithm 2.

Algorithm 2 OMP-based DAH-BT Phase 1

Step 1: The BS omni-directionally transmits $\mathbf{x}^{\mathtt{DL}}$ by the hybrid module.

Step 2: The MS omni-directionally receives signal $\mathbf{r}_{\mathtt{ADC}}^{\mathtt{DL}} = \eta\frac{\sqrt{N_t}}{G_t}\boldsymbol{\chi}_{\mathtt{ms}}\bar{\mathbf{A}}_r\mathbf{h}_v^{\mathtt{DL}} + \mathbf{n}_{\mathtt{e}}^{\mathtt{DL}}$ by the low-cost DBF module.

Step 3: OMP-based AoAs estimation.

Input: $\mathbf{r}_{\mathtt{ADC}}^{\mathtt{DL}}$, $\boldsymbol{\Phi} = \eta\frac{\sqrt{N_t}}{G_t}\boldsymbol{\chi}_{\mathtt{ms}}\bar{\mathbf{A}}_r$ and stopping criterion

Initialization: $\mathbf{r}^0 = \mathbf{r}_{\mathtt{ADC}}^{\mathtt{DL}}$, $\mathbf{x}^0 = \mathbf{0}$, $\Lambda^0 = \emptyset$, $t = 0$

While not converged **do**

Match: $\mathbf{v}^t = \boldsymbol{\Phi}^T\mathbf{r}^t$

Identify: $\Lambda^{t+1} = \Lambda^t \cup \{\underset{j}{\operatorname{argmax}}\,|v^t(j)|\}$

(where $v^t(j)$ is the j-th entry of $\mathbf{v}^t$)

Update: $\mathbf{x}^{t+1} = \underset{\mathbf{z}:\mathrm{supp}(\mathbf{z})\subseteq\Lambda^t}{\operatorname{argmin}} \|\mathbf{r}-\boldsymbol{\Phi}\mathbf{z}\|_2$

$\mathbf{r}^{t+1} = \mathbf{r} - \boldsymbol{\Phi}\mathbf{x}^{t+1}$

$t = t+1$

end While

Output: $\hat{\mathbf{h}}_v^{\mathtt{DL}} = \mathbf{x}^t$

Step 4: Find the largest L_{est} entries in $\hat{\mathbf{h}}_v^{\mathtt{DL}}$, look for the corresponding angles according to (7), (8), and (22), and record them in set $\mathbb{S}_{AoA}$.

2) DAH-BT Phase 2: After obtaining the set of receiving beams $\mathbb{S}_{AoA}$, we avail of channel reciprocity, and the MS transmits the training signal across the directions in set $\mathbb{S}_{AoA}$ successively in L_{est} time slots by the hybrid module, as shown in Fig. 5. The reasons for this successive transmission are (1) match the beams one by one without feedback; (2) better focus the transmission energy of the MS. The transmitted data is given by

$$\mathbf{X}^{\mathtt{UL}}=P_t\Big[\mathbf{a}_{\mathtt{Rx}}^*(\bar{\theta}_{AoA_1}^{az},\bar{\theta}_{AoA_1}^{el}),\ldots,\mathbf{a}_{\mathtt{Rx}}^*(\bar{\theta}_{AoA_{L_{est}}}^{az},\bar{\theta}_{AoA_{L_{est}}}^{el})\Big], \tag{32}$$

where $\mathbf{X}^{\mathtt{UL}} \in \mathbb{C}^{N_r\times L_{est}}$, and P_t is the transmit power, which is set to 1 in this study. Then, the signal received at the BS by the low-cost DBF module is given by

$$\mathbf{Y}_{\mathtt{RF}}^{\mathtt{UL}}=\boldsymbol{\chi}_{\mathtt{bs}}(\mathbf{H}^{\mathtt{UL}}\mathbf{X}^{\mathtt{UL}}+\mathbf{N}^{\mathtt{UL}})+\mathbf{N}_{\mathtt{RF}}^{\mathtt{UL}}=\boldsymbol{\chi}_{\mathtt{bs}}\mathbf{H}^{\mathtt{UL}}\mathbf{X}^{\mathtt{UL}}+\mathbf{N}_{\mathtt{eff}}^{\mathtt{UL}}, \tag{33}$$

where $\boldsymbol{\chi}_{\mathtt{bs}}$ contains the RF chain impairments' parameters at the BS, and $\mathbf{N}_{\mathtt{eff}}^{\mathtt{UL}} \triangleq \boldsymbol{\chi}_{\mathtt{bs}}\mathbf{N}^{\mathtt{UL}}+\mathbf{N}_{\mathtt{RF}}^{\mathtt{UL}}$ is still the additive Gaussian noise. Then, the received signal quantized by the low-bit ADC can be written as

$$\mathbf{R}_{\mathtt{ADC}}^{\mathtt{UL}}=Q\left(\boldsymbol{\chi}_{\mathtt{bs}}\mathbf{H}^{\mathtt{UL}}\mathbf{X}^{\mathtt{UL}}+\mathbf{N}_{\mathtt{eff}}^{\mathtt{UL}}\right)=\eta\boldsymbol{\chi}_{\mathtt{bs}}\mathbf{H}^{\mathtt{UL}}\mathbf{X}^{\mathtt{UL}}+\mathbf{N}_{\mathtt{e}}^{\mathtt{UL}}, \tag{34}$$

where $\mathbf{N}_{\mathtt{e}}^{\mathtt{UL}} \triangleq \eta\mathbf{N}_{\mathtt{eff}}^{\mathtt{UL}} + \mathbf{N}_{\mathtt{q}}$, and we assume that the elements in $\mathbf{N}_{\mathtt{e}}^{\mathtt{UL}}$ follow Gaussian distribution with zero mean and $\sigma_{\mathtt{ul,e}}^2$ variance. Here, we define the uplink receive SNR as

$$\mathrm{SNR}_l^{\mathtt{UL}} = 10\log_{10}\left(\|\eta\boldsymbol{\chi}_{\mathtt{bs}}\mathbf{H}^T\mathbf{x}_l^{\mathtt{UL}}\|_2^2/(N_t\sigma_{\mathtt{ul,e}}^2)\right), \tag{35}$$

where $\mathbf{x}_l^{\mathtt{UL}}$ is the l-th column of $\mathbf{X}^{\mathtt{UL}}$ with $1 \leq l \leq L_{est}$. According to (14), by replacing $\mathbf{H}^{\mathtt{UL}}$ approximately to $\mathbf{H}_{\mathtt{v}}^{\mathtt{UL}} = \bar{\mathbf{A}}_t^*\mathbf{H}_a^T\bar{\mathbf{A}}_r^T$, we obtain

$$\mathbf{R}_{\mathtt{ADC}}^{\mathtt{UL}} = \eta\boldsymbol{\chi}_{\mathtt{bs}}\bar{\mathbf{A}}_t^*\mathbf{H}_a^T\bar{\mathbf{A}}_r^T\mathbf{X}^{\mathtt{UL}} + \mathbf{N}_{\mathtt{e}}^{\mathtt{UL}}. \tag{36}$$

According to (32), $\bar{\mathbf{A}}_r^T\mathbf{X}^{\mathtt{UL}}$ is a $G_r \times L_{est}$ matrix, and each column of $\bar{\mathbf{A}}_r^T\mathbf{X}^{\mathtt{UL}}$ has approximately a single 1 at the index of the corresponding AoA. Denote $\tilde{\mathbf{H}}_a \triangleq \mathbf{H}_a^T\bar{\mathbf{A}}_r^T\mathbf{X}^{\mathtt{UL}}$, and $\tilde{\mathbf{H}}_a$ is a $G_t \times L_{est}$ matrix. Each column of $\tilde{\mathbf{H}}_a$ corresponds to AoAs in set $\mathbb{S}_{AoA}$, and the row indexes of $\tilde{\mathbf{H}}_a$ correspond to AoDs in grid. Then, we have

$$\mathbf{R}_{\mathtt{ADC}}^{\mathtt{UL}} = \eta\boldsymbol{\chi}_{\mathtt{bs}}\bar{\mathbf{A}}_t^*\tilde{\mathbf{H}}_a + \mathbf{N}_{\mathtt{e}}^{\mathtt{UL}}. \tag{37}$$

Vectorizing $\mathbf{R}_{\mathtt{ADC}}^{\mathtt{UL}}$, we have

$$\mathbf{r}_{\mathtt{ADC}}^{\mathtt{UL}} = \mathrm{vec}(\mathbf{R}_{\mathtt{ADC}}^{\mathtt{UL}}) = \eta(\mathbf{I}\otimes\boldsymbol{\chi}_{\mathtt{bs}}\bar{\mathbf{A}}_t^*)\mathbf{h}_v^{\mathtt{UL}} + \mathbf{n}_{\mathtt{e}}^{\mathtt{UL}}, \tag{38}$$

where $\mathbf{h}_v^{\mathtt{UL}} \triangleq \mathrm{vec}(\tilde{\mathbf{H}}_a)$ and $\mathbf{n}_{\mathtt{e}}^{\mathtt{UL}} \triangleq \mathrm{vec}(\mathbf{N}_{\mathtt{e}}^{\mathtt{UL}})$, Then, denote $\tilde{\boldsymbol{\Phi}} \triangleq \eta(\mathbf{I}\otimes\boldsymbol{\chi}_{\mathtt{bs}}\bar{\mathbf{A}}_t^*)$, and (38) can be simplified to

$$\mathbf{r}_{\mathtt{ADC}}^{\mathtt{UL}} = \tilde{\boldsymbol{\Phi}}\mathbf{h}_v^{\mathtt{UL}} + \mathbf{n}_{\mathtt{e}}^{\mathtt{UL}}. \tag{39}$$

Similarly, the following lemma shows that the measurement matrix $\tilde{\boldsymbol{\Phi}}$ satisfies the RIP.

Lemma 2. *Let* $\tilde{\boldsymbol{\Phi}} = \eta(\mathbf{I}\otimes\boldsymbol{\chi}_{\mathtt{bs}}\bar{\mathbf{A}}_t^*)$, *where the coefficient* $\eta \leqslant 1$, *and the entries in* $\boldsymbol{\chi}_{\mathtt{bs}}$ *satisfy* $0 < |\chi_{\mathtt{bs}}(i)| < 1$, *and* $\bar{\mathbf{A}}_t$ *is the matrix defined in Section* II-C. *Then for an effectively* $\tilde{d}$*-sparse vector* $\mathbf{h}_v^{\mathtt{UL}}$, *there exists* $\delta_{\tilde{d}} \in (0,1)$ *such that*

$$(1-\delta_{\tilde{d}})\|\mathbf{h}_v^{\mathtt{UL}}\|_2^2 \leqslant \|\tilde{\boldsymbol{\Phi}}\mathbf{h}_v^{\mathtt{UL}}\|_2^2 \leqslant (1+\delta_{\tilde{d}})\|\mathbf{h}_v^{\mathtt{UL}}\|_2^2. \tag{40}$$

Namely, $\tilde{\boldsymbol{\Phi}}$ *satisfies the* $\tilde{d}$*-order RIP.*

Proof. Please refer to Appendix B. □

Therefore, $\mathbf{h}_v^{\mathtt{UL}}$ can be estimated by the OMP algorithm, and the estimate of $\mathbf{h}_v^{\mathtt{UL}}$ is denoted as $\hat{\mathbf{h}}_v^{\mathtt{UL}}$. Then, we reshape $\hat{\mathbf{h}}_v^{\mathtt{UL}} \in \mathbb{C}^{G_tL_{est}\times 1}$ into $\hat{\mathbf{H}}_a \in \mathbb{C}^{G_t\times L_{est}}$. The index of the largest entry in the l-th column of $\hat{\mathbf{H}}_a$ represents the AoD of the l-th path. According to (9), (10), and (22), we can obtain the set of estimated AoDs denoted as

$$\mathbb{S}_{AoD}=\Big[(\bar{\phi}_{AoD_1}^{az},\bar{\phi}_{AoD_1}^{el}),(\bar{\phi}_{AoD_2}^{az},\bar{\phi}_{AoD_2}^{el}),\ldots,(\bar{\phi}_{AoD_{L_{est}}}^{az},\bar{\phi}_{AoD_{L_{est}}}^{el})\Big].$$

To recapitulate, the process of beam training phase 2 is summarized in Algorithm 3. After phase 1 and phase 2 of the DAH-BT, we finally identify the pairs of transmitting and receiving beams recorded in $\mathbb{S}_{AoD}$ and $\mathbb{S}_{AoA}$, respectively.

V. Numerical Results

In this section, we carry out simulations to investigate the feasibility of the proposed DA-hybrid architecture and the performance of the proposed DAH-BT method. In all of the simulations, N_t, N_r, G_t^{az}, G_t^{el}, G_r^{az} and G_r^{el} are set to 64.[7] Other parameters will be defined with each simulation. Finally,

[7]The phase resolution parameter N is $2^6 = 64$, due to the 6-bit resolution of the RF phase shifters. Hence, G_r^{az}, G_r^{el}, G_t^{az}, and G_t^{el} should be no greater than 64.

Algorithm 3 OMP-based DAH-BT Phase 2

Step 1: The MS transmits $\mathbf{x}_l^{\text{UL}} = \mathbf{a}_{\text{Rx}}^*(\bar{\theta}_{AoA_l}^{az}, \bar{\theta}_{AoA_l}^{el})$, where $l = 1, \ldots, L_{est}$, successively by the hybrid module.
Step 2: The BS omni-directionally receives signal $\mathbf{r}_{\text{ADC}}^{\text{UL}} = \eta(\mathbf{I} \otimes \boldsymbol{\chi}_{\text{bs}} \bar{\mathbf{A}}_t^*)\mathbf{h}_v^{\text{UL}} + \mathbf{n}_{\text{e}}^{\text{UL}}$ by the low-cost DBF module.
Step 3: OMP-based AoDs estimation is similar to **Step 3** in Algorithm 2, where **Input** is $\mathbf{r}_{\text{ADC}}^{\text{UL}}$ and $\tilde{\boldsymbol{\Phi}} = \eta(\mathbf{I} \otimes \boldsymbol{\chi}_{\text{bs}} \bar{\mathbf{A}}_t^*)$, and **Output** is $\mathbf{h}_v^{\text{UL}}$.
Step 4: Reshape $\mathbf{h}_v^{\text{UL}} \in \mathbb{C}^{G_t L_{est} \times 1}$ into $\hat{\mathbf{H}}_a \in \mathbb{C}^{G_t \times L_{est}}$, the index of the largest entry in the l-th column of $\hat{\mathbf{H}}_a$ represents the AoD of the l-th path, then, record AoDs in set $\mathbb{S}_{AoD}$.

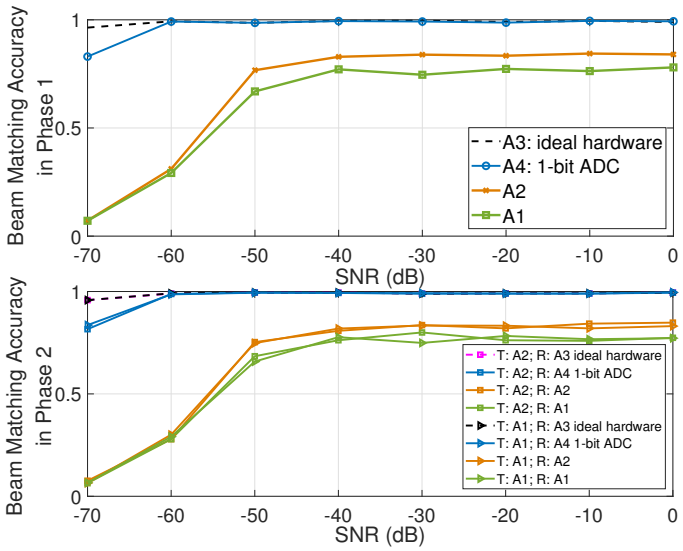


Fig. 6: Comparison of the beam matching accuracy of the proposed beam training method in different architectures with $L = 1$. The receive SNR of the upper figure is defined in (26), and the receive SNR of the lower figure is defined in (35).

all the numerical results provided in this section are obtained from Monte Carlo simulations with 1000 independent channel realizations for each BS-MS configuration.

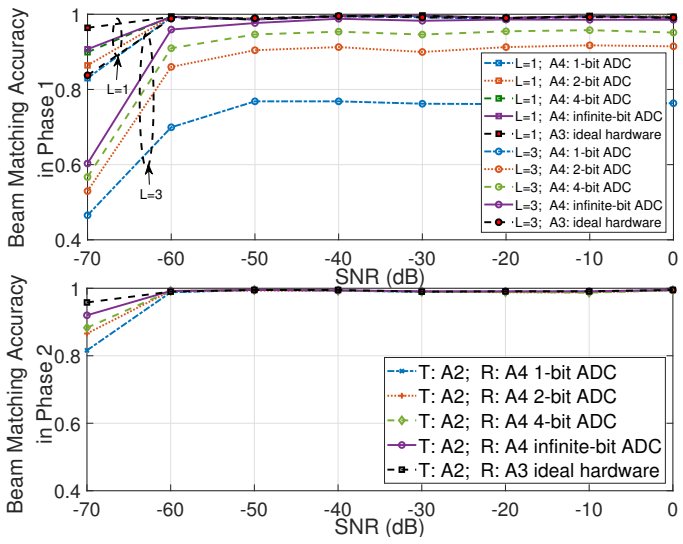


Fig. 7: Beam matching accuracy of the proposed beam training method with different ADC bits and L. The receive SNR of the upper figure is defined in (26), and the receive SNR of the lower figure is defined in (35).

A. Beam Matching Accuracy

In the first set of simulations, we aim at comparing the beam training performance of the proposed DA-hybrid architecture with the traditional DBF and hybrid architectures [2], and analyzing the impact of replacing the fully-connected hybrid module with sub-connected hybrid module in the proposed DA-hybrid architecture. In addition, we further study the effect of increasing the number of dominant paths L on the proposed DAH-BT method. Since the angles are meshed, according to (7)-(10), the index of the grid point corresponds to the angle one by one. In our algorithm, we estimate the index of the grid point that is closest to the real angle. The indexes of the real angle can also be obtained, which may not be integers, by the inverse operation of (7)-(10). When the difference between the estimated indexes and the true indexes is smaller than a given threshold, we determine that the beam is correctly matched. In this section, the threshold is set to 5. We define the beam matching accuracy as T_a/T, which signifies that the beams are correctly matched T_a times out of T Monte Carlo simulations.

Fig. 6 shows the beam matching accuracy for different architectures. During the simulation, the path number L is set to 1. For beam training phase 1, since the signal is omni-directionally transmitted, for different transmitter architectures, the transmitted signals have the same expression $\mathbf{x}^{\text{DL}} = [1, 0, \ldots, 0]^T$. Hence, we do not have to consider different transmitter architectures during phase 1. We compare the performance of four receiver architectures by setting N_{RF}^r for A1 and A2 to 8 and L_r for A2 to 8. For beam training phase 2, beamforming is considered and different transmitter architectures have different beamforming performance. Obviously, the fully DBF architecture A3 has the highest precoding freedom and beamforming precision among four architectures mentioned in this study, while the low-cost DBF architecture A4 with imperfect RF chains and low-bit ADCs cannot be used as transmitter, since the RF impairments' parameters of the transmitter are unknown to the receiver. Therefore, we only compare two transmitter architectures, the fully-connected hybrid architecture A1 and the sub-connected hybrid architecture A2. Considering that the number of RF chains at the MS is smaller than at the BS, we make the following settings: at the transmitters of the MS, the number of RF chains for A1 and A2 is set to 4, hence each subarray in A2 has 16 antennas; while at the receivers of the BS, the number of RF chains for A1 and A2 is set to 8, and therefore each subarray in A2 has 8 antennas. At the receivers, to fit in the CS framework, the $\mathbf{W}_{\text{RF}}$ in A1 is a random matrix with elements consisting of $1/\sqrt{N_{RF}^t}$ and $-1/\sqrt{N_{RF}^t}$, while $\mathbf{W}_{\text{RF}}$ in A2 is a block diagonal matrix, where $\mathbf{W}_{\text{RF}} = \text{blkdiag}\{\mathbf{w}_1, \ldots, \mathbf{w}_i, \ldots, \mathbf{w}_{N_{RF}^r}\}$ and $\mathbf{w}_i$ is a random vector with elements consisting of $1/\sqrt{L_t}$ and $-1/\sqrt{L_t}$. Actually, more suitable measurement matrices and more efficient CS algorithms can be selected to further improve the performance of A1 and A2. However, this goes outside the scope of this study. In this section, we aim at qualitatively comparing the performance of different architectures.

TABLE II: Beam matching accuracy comparison of different receiver architectures in phase 1.

Receiver	BMA in Phase 1
A1	0.771
A2	0.829
A3	0.995
A4: 1-bit ADC	0.994

TABLE III: Beam matching accuracy comparison of different architectures in phase 2.

Transmitter	Receiver	BMA in Phase 2
A1	A1	0.777
A1	A2	0.819
A1	A3	0.995
A1	A4: 1-bit ADC	0.993
A2	A1	0.762
A2	A2	0.808
A2	A3	0.995
A2	A4: 1-bit ADC	0.993

There are several interesting findings shown in Fig. 6: both in phase 1 and 2, as the SNR increases, A4 as a receiver gradually reaches the performance of A3, which achieves close to 100% beam matching accuracy, while A1 and A2 are far from such precision, because A4 can get more information despite of the RF impairments and low ADC bits, making the measurements more sparse than A1 and A2, in which the measurements are compressed by the combining matrix; in phase 2, although A2 is less flexible than A1, A2 as a transmitter can achieve performance close to A1, because in phase 2, only single beam transmission is needed. More specifically, we list the beam matching accuracy for different architectures in Table II and III, the values of which are obtained in Fig. 6 with receive $\mathrm{SNR}^{\mathrm{DL}} = -40$ dB for phase 1 and $\mathrm{SNR}^{\mathrm{UL}} = -40$ dB for phase 2, respectively. In terms of beam matching accuracy, A4 as a receiver has performance close to A3 from Table II, and A2 as a transmitter has performance close to A1 from Table III. Hence, we can conclude that the DA-hybrid architecture composed of the sub-connected hybrid (A2) module and the low-cost DBF (A4) module can not only achieve high beam matching accuracy, but also reduce the system power consumption and cost, which confirms the feasibility of the proposed DA-hybrid architecture.

Fig. 7 further displays the beam matching accuracy for A4 with different ADC bits and L. For beam training phase 1, as the number of L increases, the beam matching accuracy decreases accordingly, due to the reduced sparsity. By comparing the red dotted line "A3", which represents the fully DBF architecture with ideal hardware, with solid lines, which represent imperfect RF chains and low-bit ADCs, we find that the performance loss caused by the aggregate impact of residual hardware impairments increases when $L = 3$. However, with the module "A4: 4-bit ADC", our proposed algorithm can gradually get close to the performance of ideal hardware module A3, within a certain angle error of beams set by the threshold, as the receive SNR increases. During beam training phase 2, since beams are sent and received one by one, the beam matching accuracy for beam training phase 2 will not be affected by the increase of L, but the beam training time grows linearly with L.

As mentioned before, the beam matching accuracy is calculated by setting a particular threshold. For a more intuitive understanding of the beam matching accuracy, we compare the estimated azimuth and elevation AoA indexes with the true indexes in Fig. 8, where we take DAH-BT phase 1 for example by setting L to 3 and $\mathrm{SNR}^{\mathrm{DL}}$ in (26) to -30 dB. As shown in Fig. 8, for A4 with 1-bit ADC, the average deviation of the estimated index in both azimuth and elevation is around 2; for A4 with 4-bit ADC, the average deviation of the estimated index is further reduced to less than 1; when the ADC in A4 is infinite-bit, the accuracy reaches the level of A3. Particularly, from Fig. 8 for A4 with 4-bit ADC, we can see that $L_{est} \geq L$, and the estimated paths cluster around the true values. Actually,

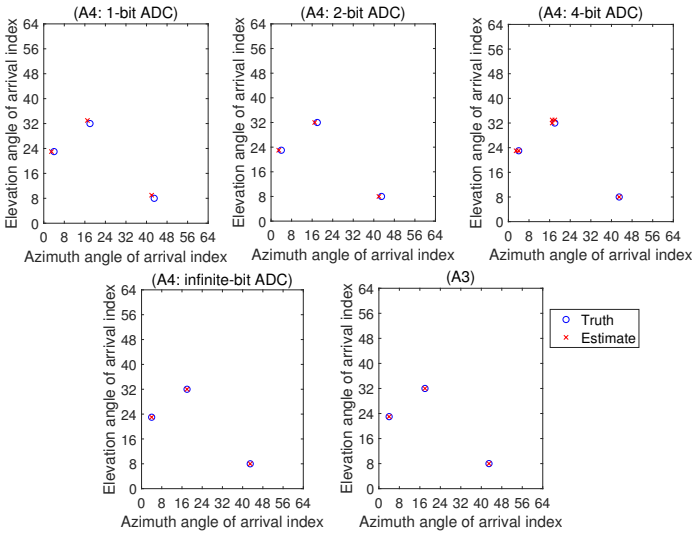


Fig. 8: Comparison of the estimated azimuth and elevation AoA indexes with true values in phase 1 with $L = 3$ and $\mathrm{SNR}^{\mathrm{DL}} = -30$ dB.

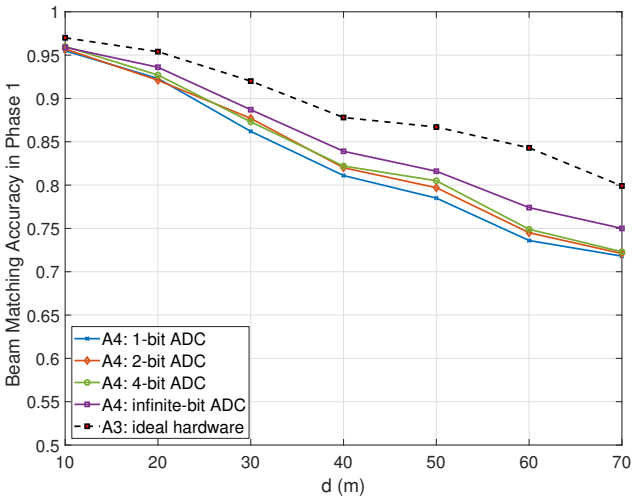


Fig. 9: Beam matching accuracy of the proposed beam training method with different BS-MS distance with $L = 1$.

in the case where the beam matching accuracy requirement is slightly lower, A4 with 1-bit ADC can be used.

Finally, we analyze the impact of BS-MS distance on beam matching accuracy. In the simulations, we model the path gain of the l-th path as $\alpha_l \sim \mathcal{CN}(0, \gamma 10^{-0.1PL})$ [27]. According to [47], for omnidirectional path loss, $PL = \alpha + 10\beta \log_{10}(d) + \xi$ dB, where $\xi \sim \mathcal{N}(0, \sigma^2)$ denotes lognormal shadowing, $\gamma = U^{r_\tau - 1} 10^{0.1Z}$ is the cluster power fraction with $Z \sim \mathcal{N}(0, \varsigma^2)$ and $U \sim \mathcal{U}[0, 1]$, and d represents the BS-MS distance in meters. The system is assumed to operate at 28 GHz carrier frequency with 500 MHz bandwidth, hence, the reference values of α, β, σ, r_τ, and ς can be found in [47, Table I], where $r_\tau = 2.8$ and $\varsigma = 4.0$; for the LoS path, $\alpha = 61.4$, $\beta = 2$, and $\sigma = 5.8$ dB; for the NLoS path, $\alpha = 72.0$, $\beta = 2.92$, and $\sigma = 8.7$ dB. We assume that the transmit power at the BS is set to 33 dBm, the noise power at the MS is set to -83 dBm, the threshold is set to 5. The simulation results are shown in Fig. 9 and Fig. 10, where the beam matching accuracy decreases when the BS-MS distance d increases. Since the BS transmit power is fixed, the received energy of the signal decreases with d increases, which causes this performance degradation. However,

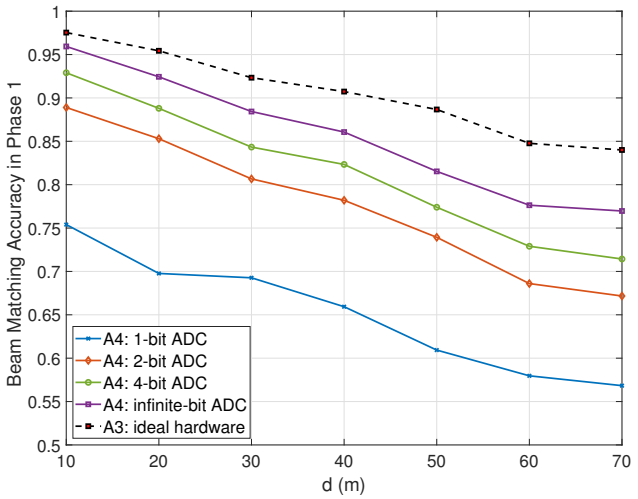


Fig. 10: Beam matching accuracy of the proposed beam training method with different BS-MS distance with $L = 3$.

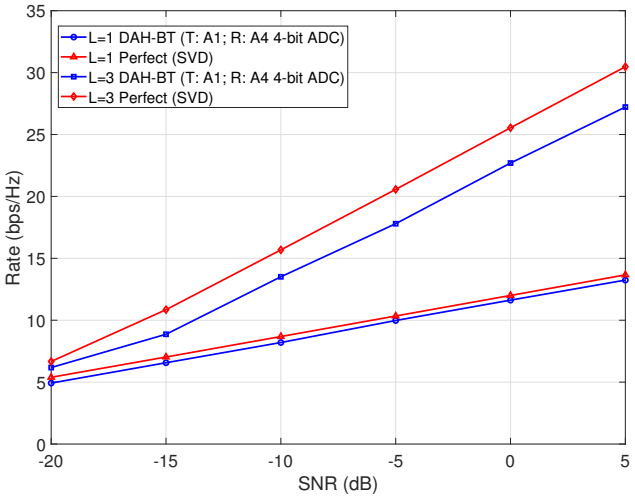


Fig. 11: Spectral efficiency comparison between SVD-based precoding/combining with perfect channel knowledge and the proposed method when $L = 1$ and $L = 3$.

under the given simulation parameters, with the module "A4: 4-bit ADC", the beam matching accuracy at beam training phase 1 (omnidirectional transmission) can reach 80% or even more within 50 (m) when $L = 1$ and within 45 (m) when $L = 3$. Since mmWave frequency band is used for short-distance transmission, the simulation results have confirmed the feasibility of the proposed beam training algorithm.

B. Outage Probability

After beam training, the BS and MS use the estimated beam directions to build their precoder $\mathbf{P}_{\mathtt{DATA}} \in \mathbb{C}^{N_t \times L_{est}}$ (at the BS) and combiner $\mathbf{C}_{\mathtt{DATA}} \in \mathbb{C}^{N_r \times L_{est}}$ (at the MS). To further analyze the performance of the proposed DA-hybrid architecture and DAH-BT method, we quantify the quality of the precoder and combiner by considering the spectral efficiency [23]:

$$R = \log_2 \left| \mathbf{I}_{L_{est}} + \frac{P}{L_{est}} \mathbf{R}_n^{-1} \mathbf{C}_{\mathtt{DATA}}^H \mathbf{H}^{\mathtt{DL}} \mathbf{P}_{\mathtt{DATA}} \mathbf{P}_{\mathtt{DATA}}^H \mathbf{H}^{\mathtt{DL}\,H} \mathbf{C}_{\mathtt{DATA}} \right|, \tag{41}$$

where L_{est} is the number of estimated main paths via beam training, P is the transmit power, $\mathbf{R}_n$ is the post-processing noise covariance matrix, i.e., $\mathbf{R}_n = \sigma_{\mathtt{dl}}^2 \mathbf{C}_{\mathtt{DATA}}^H \mathbf{C}_{\mathtt{DATA}}$, and $\sigma_{\mathtt{dl}}^2$ is the average downlink noise power. Theoretically, R in (41) reaches the maximum value when $\mathbf{P}_{\mathtt{DATA}}$ and $\mathbf{C}_{\mathtt{DATA}}$ equal to the right and left sigular vectors of the channel matrix $\mathbf{H}^{\mathtt{DL}}$, respectively. Here, we define the downlink transmit $\mathrm{SNR} = 10\log_{10}(P/\sigma_{\mathtt{dl}}^2)$. Eq. (41) can be used as a valid indicator of how good the estimated beam steering directions are compared to the optimal ones.

To analyze the spectral efficiency of the proposed method, we deploy A1 as transceivers in both BS and MS during the data transmission phase. Note that when $L = 1$ (only one beam is sent at a time), the beamforming and combining vector are obtained by $\mathbf{P}_{\mathtt{DATA}} = \mathbf{a}_{\mathtt{Tx}}(\bar{\phi}_{AoD}^{az}, \bar{\phi}_{AoD}^{el})$ and $\mathbf{C}_{\mathtt{DATA}} = \mathbf{a}_{\mathtt{Rx}}(\bar{\theta}_{AoA}^{az}, \bar{\theta}_{AoA}^{el})$, respectively, where $(\bar{\theta}_{AoA}^{az}, \bar{\theta}_{AoA}^{el})$ and $(\bar{\phi}_{AoD}^{az}, \bar{\phi}_{AoD}^{el})$ are defined in sets $\mathbb{S}_{AoA}$ and $\mathbb{S}_{AoD}$ with $L_{est} = 1$, respectively. When $L > 1$, multiple beams will be transmitted at the same time, the precoder and combiner are given by

$$\mathbf{P}_{\mathtt{DATA}} = \mathrm{blkdiag}\left\{ \mathbf{a}_{\mathtt{Tx}}(\bar{\phi}_{AoD_1}^{az}, \bar{\phi}_{AoD_1}^{el}), \ldots, \mathbf{a}_{\mathtt{Tx}}(\bar{\phi}_{AoD_{L_{est}}}^{az}, \bar{\phi}_{AoD_{L_{est}}}^{el}) \right\}, \tag{42}$$

and

$$\mathbf{C}_{\mathtt{DATA}} = \mathrm{blkdiag}\left\{ \mathbf{a}_{\mathtt{Rx}}(\bar{\theta}_{AoA_1}^{az}, \bar{\theta}_{AoA_1}^{el}), \ldots, \mathbf{a}_{\mathtt{Rx}}(\bar{\theta}_{AoA_{L_{est}}}^{az}, \bar{\theta}_{AoA_{L_{est}}}^{el}) \right\}, \tag{43}$$

respectively. Fig. 11 shows that the spectral efficiency of proposed method is slightly worse than the optimal one when $L = 1$ and $L = 3$., since the estimated beam directions are taken from limited angular grids.

For further analysis, we define the outage as the event that the spectral efficiency delivered to the MS is below a target value R_{th}. In Fig. 12, we plot the measured outage probability for $R_{th} = 0.1$ bps/Hz as a function of the downlink transmit SNR. Obviously, when $L = 1$ (only one beam is sent at a time) A2 has the same beamforming and combining performance as A1. Therefore, the outage probability of the proposed method is 0, which can be inferred from Fig. 11. As expected, when $L = 3$, the outage probability increases, since the beamforming performance of the sub-connected hybrid module deteriorates when multiple beams are transmitted. The sequential adaptive beam training (SA-BT) was proposed in [23], [25] and parallel adaptive beam training (PA-BT) was proposed in [27]. When the downlink transmit SNR exceeds -10 dB, the performance of the proposed DAH-BT method exceeds that of SA-BT and PA-BT.

C. Time Resources

To provide further insights into the time resources required by the proposed DAH-BT, let us briefly review the procedure of the SA-BT and PA-BT shown in Fig. 13 and Fig. 14, respectively. In SA-BT, the BS and MS can respectively transmit and receive training signals through only one sector at each time slot. At each stage, the BS (MS) divides its angular domain into $K_{BS}(K_{MS})$ sectors, hence $K_{BS}K_{MS}$ time slots are needed to train all the possible transmit/receive sector pairs, besides one additional slot is required for feedback. Then, the number of

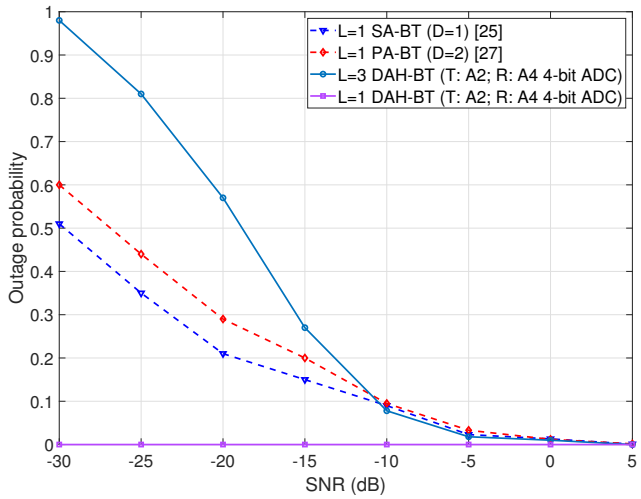


Fig. 12: Outage probability referred to a target value $R_{th} = 0.1$ bps/Hz versus transmit SNR for different beam training methods.

beam training stages required to estimate one AoA/AoD pair with angular resolution $2\pi/N_a$ is

$$S = \max\left[\log_{K_{BS}} N_a, \log_{K_{MS}} N_a\right]. \tag{44}$$

Therefore, the total beam training time for SA-BT is $L_{est}S(K_{BS}K_{MS}+1)T_{\mathtt{slot}}$, where $T_{\mathtt{slot}}$ denotes one time slot duration. In PA-BT, the beam training time is further shortened, since the BS (MS) can transmit the training signal over $D_{BS}(D_{MS})$ sectors simultaneously, then the total beam training time for PA-BT is $L_{est}S\left(\left\lceil\frac{K_{BS}}{D_{BS}}\right\rceil\left\lceil\frac{K_{MS}}{D_{MS}}\right\rceil+1\right)T_{\mathtt{slot}}$. However, when N_a is large, the time overhead of the SA-BT and PA-BT will become prohibitive. The time overhead of our proposed method DAH-BT is $(L_{est}+1)T_{\mathtt{slot}}$, which is obviously small in mmWave communication since L_{est} is small. It is easy to observe from Fig. 15 that the proposed DAH-BT has great potential in saving time resources.

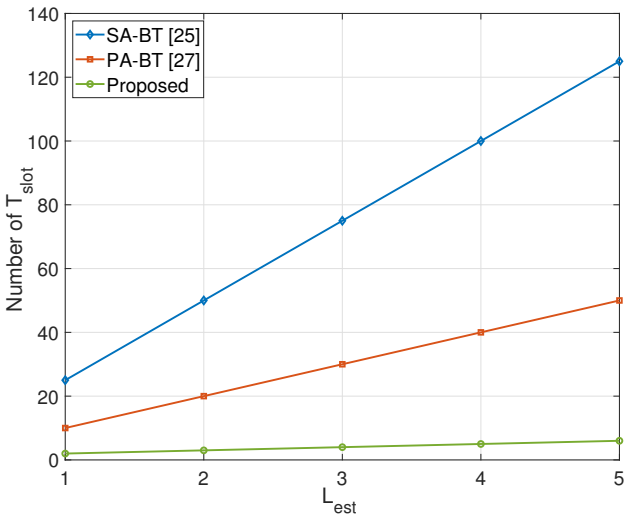


Fig. 15: Beam training time versus L_{est} for different methods with $N_a = 32$ and $K_{BS} = K_{MS} = D_{BS} = D_{MS} = 2$.

VI. Conclusion

In this study, we researched the problem of beam training acceleration by proposing a novel DA-hybrid architecture and a compatible DAH-BT method. Specifically, we first designed a low-cost DBF module assisted hybrid architecture, where coarse RF chains and low-resolution ADCs were considered in the low-cost DBF module. We also modeled the RF impairments and low-bit quantization by the EEVM model and the additive quantization noise model, respectively. Through numerical comparison of the cost and power consumption, the DA-hybrid architecture composed of sub-connected hybrid module and low-cost DBF module was proven to be an alternative solution for a good tradeoff between beam training performance and hardware cost. Additionally, by leveraging the sparsity of mmWave channels in the angular domain, we developed a fast DAH-BT method based on the proposed DA-hybrid architecture, which included the internal calibration phase to acquire the parameters of RF impairments and the beam training phase to obtain AoAs and AoDs of beams via the OMP algorithm. We also evaluated the algorithm by proving that the designed measurement matrices satisfy the RIP. In contrast to almost all of the approaches in the literature, our design exploited both the abilities of the hybrid module to generate desired multiple beams and the low-cost DBF module to accurately capture the angular information. Finally, simulation results revealed that the DA-hybrid architecture composed of the sub-connected hybrid module and the low-cost DBF module can not only approach close to 100% beam matching accuracy, but also reduce the system power consumption and cost. Moreover, the proposed DAH-BT (which requires only $L_{est}+1$ time slots) showed great advantage in saving time resources over traditional methods with comparable spectral efficiency.

Appendix A
Proof of Lemma 1

Denote $a = \eta\frac{\sqrt{N_t}}{G_t}$ and $|a| < 1$, then $\mathbf{\Phi} = a\boldsymbol{\chi}_{\mathtt{ms}}\bar{\mathbf{A}}_r$, and $\mathbf{\Phi}\mathbf{h}_v^{\mathtt{DL}} = a\boldsymbol{\chi}_{\mathtt{ms}}\bar{\mathbf{A}}_r\mathbf{h}_v^{\mathtt{DL}}$. By denoting $\bar{\mathbf{A}}_r\mathbf{h}_v^{\mathtt{DL}} = [g_1, g_2, \ldots, g_{N_r}]^T$, we have

$$\|\mathbf{\Phi}\mathbf{h}_v^{\mathtt{DL}}\|_2^2 = \sum_{i=1}^{N_r}|a\chi_{\mathtt{ms}}(i)g_i|^2 = a^2\sum_{i=1}^{N_r}|\chi_{\mathtt{ms}}(i)|^2|g_i|^2, \tag{45}$$

since $|\chi_{\mathtt{ms}}(i)|^2 < 1$, then

$$a^2\sum_{i=1}^{N_r}|\chi_{\mathtt{ms}}(i)|^2|g_i|^2 \leqslant a^2\sum_{i=1}^{N_r}|g_i|^2 = a^2\|\bar{\mathbf{A}}_r\mathbf{h}_v^{\mathtt{DL}}\|_2^2. \tag{46}$$

As known that $\bar{\mathbf{A}}_r$ satisfies the d-th order RIP, there exists $\delta_+ \in (0,1)$ such that

$$a^2\|\bar{\mathbf{A}}_r\mathbf{h}_v^{\mathtt{DL}}\|_2^2 \leqslant a^2(1+\delta_+)\|\mathbf{h}_v^{\mathtt{DL}}\|_2^2. \tag{47}$$

Since $a^2 < 1$, we have

$$a^2(1+\delta_+)\|\mathbf{h}_v^{\mathtt{DL}}\|_2^2 \leqslant (1+\delta_+)\|\mathbf{h}_v^{\mathtt{DL}}\|_2^2. \tag{48}$$

Hence, we obtain $\|\mathbf{\Phi}\mathbf{h}_v^{\mathtt{DL}}\|_2^2 \leqslant (1+\delta_+)\|\mathbf{h}_v^{\mathtt{DL}}\|_2^2$. Denote $\chi_{min} = \min\limits_{1\leqslant i\leqslant N_r}|\chi_{\mathtt{ms}}(i)|$, then

$$a^2\sum_{i=1}^{N_r}|\chi_{\mathtt{ms}}(i)|^2|g_i|^2 \geqslant a^2\chi_{min}^2\sum_{i=1}^{N_r}|g_i|^2 = a^2\chi_{min}^2\|\bar{\mathbf{A}}_r\mathbf{h}_v^{\mathtt{DL}}\|_2^2. \tag{49}$$

Since $\bar{\mathbf{A}}_r$ satisfies d-th RIP, there exists $\delta \in (0,1)$ such that

$$a^2\chi_{min}^2\|\bar{\mathbf{A}}_r\mathbf{h}_v^{\mathtt{DL}}\|_2^2 \geqslant a^2\chi_{min}^2(1-\delta)\|\mathbf{h}_v^{\mathtt{DL}}\|_2^2 = \left[1-\left(1-a^2\chi_{min}^2(1-\delta)\right)\right]\|\mathbf{h}_v^{\mathtt{DL}}\|_2^2. \tag{50}$$

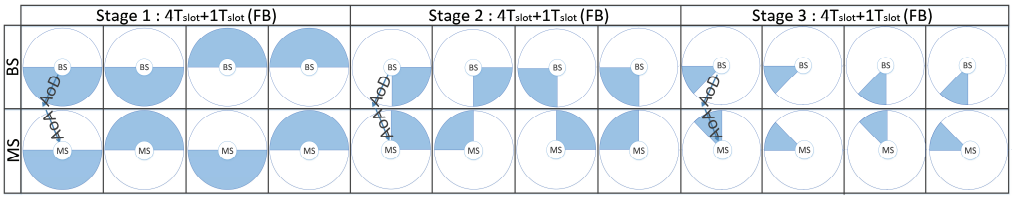


Fig. 13: Illustration of the SA-BT protocol with $K_{BS} = K_{MS} = 2$.

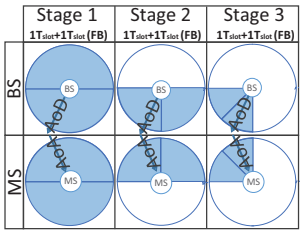


Fig. 14: Illustration of the PA-BT protocol with $K_{BS} = K_{MS} = D_{BS} = D_{MS} = 2$.

Let $\delta_- = 1 - a^2\chi_{min}^2(1-\delta)$, we have

$$a^2\chi_{min}^2\|\bar{\mathbf{A}}_r\mathbf{h}_v^{\mathrm{DL}}\|_2^2 \geqslant (1-\delta_-)\|\mathbf{h}_v^{\mathrm{DL}}\|_2^2. \tag{51}$$

Hence, we obtain $\|\mathbf{\Phi}\mathbf{h}_v^{\mathrm{DL}}\|_2^2 \geqslant (1-\delta_-)\|\mathbf{h}_v^{\mathrm{DL}}\|_2^2$. Let $\delta_d = \max(\delta_-, \delta_+)$, according to (48) and (51), it is obviously that $(1-\delta_d)\|\mathbf{h}_v^{\mathrm{DL}}\|_2^2 \leqslant \|\mathbf{\Phi}\mathbf{h}_v^{\mathrm{DL}}\|_2^2 \leqslant (1+\delta_d)\|\mathbf{h}_v^{\mathrm{DL}}\|_2^2$.

APPENDIX B
PROOF OF LEMMA 2

Let $\tilde{\mathbf{H}}_a(i)$ represent the i-th column of matrix $\tilde{\mathbf{H}}_a$, where $1 \leqslant i \leqslant L$. Then

$$\|\tilde{\mathbf{\Phi}}\mathbf{h}_v^{\mathrm{UL}}\|_2^2 = \|\eta(\mathbf{I}\otimes\boldsymbol{\chi}_{\mathtt{bs}}\bar{\mathbf{A}}_t^*)\mathbf{h}_v^{\mathrm{UL}}\|_2^2 \overset{(a)}{=} \eta^2\sum_{i=1}^{L}\|\boldsymbol{\chi}_{\mathtt{bs}}\bar{\mathbf{A}}_t^*\tilde{\mathbf{H}}_a(i)\|_2^2, \tag{52}$$

where (a) holds because $\mathbf{h}_v^{\mathrm{UL}}$ is the vectorized version of $\tilde{\mathbf{H}}_a$. As we know that $\bar{\mathbf{A}}_t^*$ satisfies the $\tilde{d}$-th order RIP. Similar to (45), (46) and (47), there exists $\tilde{\delta}_+ \in (0,1)$ such that

$$\begin{aligned}&\eta^2\sum_{i=1}^{L}\|\boldsymbol{\chi}_{\mathtt{bs}}\bar{\mathbf{A}}_t^*\tilde{\mathbf{H}}_a(i)\|_2^2 \overset{(a)}{\leqslant} \sum_{i=1}^{L}\|\bar{\mathbf{A}}_t^*\tilde{\mathbf{H}}_a(i)\|_2^2\\ &\leqslant (1+\tilde{\delta}_+)\sum_{i=1}^{L}\|\tilde{\mathbf{H}}_a(i)\|_2^2 = (1+\tilde{\delta}_+)\|\mathbf{h}_v^{\mathrm{UL}}\|_2^2,\end{aligned} \tag{53}$$

where (a) holds because $\eta < 1$ and $|\chi_{\mathtt{bs}}(i)|^2 < 1$. Hence, we get $\|\tilde{\mathbf{\Phi}}\mathbf{h}_v^{\mathrm{UL}}\|_2^2 \leqslant (1+\tilde{\delta}_+)\|\mathbf{h}_v^{\mathrm{UL}}\|_2^2$. For the same reason as in (49) and (50) , we have

$$\begin{aligned}&\eta^2\sum_{i=1}^{L}\|\boldsymbol{\chi}_{\mathtt{bs}}\bar{\mathbf{A}}_t^*\tilde{\mathbf{H}}_a(i)\|_2^2 \geqslant \eta^2\chi_{min}^2\sum_{i=1}^{L}\|\bar{\mathbf{A}}_t^*\tilde{\mathbf{H}}_a(i)\|_2^2\\ &\geqslant \eta^2\chi_{min}^2(1-\tilde{\delta})\sum_{i=1}^{L}\|\tilde{\mathbf{H}}_a(i)\|_2^2,\end{aligned} \tag{54}$$

and we have

$$\begin{aligned}&\eta^2\chi_{min}^2(1-\tilde{\delta})\sum_{i=1}^{L}\|\tilde{\mathbf{H}}_a(i)\|_2^2 = \eta^2\chi_{min}^2(1-\tilde{\delta})\|\mathbf{h}_v^{\mathrm{UL}}\|_2^2\\ &= [1-(1-\eta^2\chi_{min}^2(1-\tilde{\delta}))]\|\mathbf{h}_v^{\mathrm{UL}}\|_2^2.\end{aligned} \tag{55}$$

Let $\tilde{\delta}_- = 1-\eta^2\chi_{min}^2(1-\tilde{\delta})$, we obtain

$$\eta^2\sum_{i=1}^{L}\|\boldsymbol{\chi}_{\mathtt{bs}}\bar{\mathbf{A}}_t^*\tilde{\mathbf{H}}_a(i)\|_2^2 \geqslant (1-\tilde{\delta}_-)\|\mathbf{h}_v^{\mathrm{UL}}\|_2^2, \tag{56}$$

hence, we have $\|\tilde{\mathbf{\Phi}}\mathbf{h}_v^{\mathrm{UL}}\|_2^2 \geqslant (1-\tilde{\delta}_-)\|\mathbf{h}_v^{\mathrm{UL}}\|_2^2$. Finally we obtain

$$(1-\delta_{\tilde{d}})\|\mathbf{h}_v^{\mathrm{UL}}\|_2^2 \leqslant \|\tilde{\mathbf{\Phi}}\mathbf{h}_v^{\mathrm{UL}}\|_2^2 \leqslant (1+\delta_{\tilde{d}})\|\mathbf{h}_v^{\mathrm{UL}}\|_2^2, \tag{57}$$

where $\delta_{\tilde{d}} = \max(\tilde{\delta}_-, \tilde{\delta}_+)$.

REFERENCES

[1] T. R. Rappaport, *et al.*, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE Access*, vol. 1, pp. 335-349, May 2013.

[2] R. W. Heath, Jr., N. G. Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE J. Sel. Top. Signal Process.*, vol. 10, no. 3, pp. 436-453, Apr. 2016.

[3] C. N. Barati, *et al.*, "Initial access in milllimeter wave cellular systems," *IEEE Trans. Wireless Commun.*, vol. 15, no. 12, pp. 7926-7940, Dec. 2016.

[4] M. Xiao, *et al.*, "Millimeter wave communications for future mobile networks," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 1909-1935, Sept. 2017.

[5] A. Alkhateeb, S. Alex, P. Varkey, Y. Li, Q. Qu, and D. Tujkovic, "Deep learning coordinated beamforming for highly-mobile millimeter wave systems," *IEEE Access*, vol. 6, pp. 37328-37348, Jun. 2018.

[6] X. Yang, M. Matthaiou, J. Yang, C. K. Wen, F. Gao, and S. Jin, "Hardware-constrained millimeter wave systems for 5G: Challenges, opportunities, and solutions," *IEEE Commun. Mag.*, vol. 57, no. 1, pp. 44-50, Jan. 2019.

[7] N. Tawa, T. Kuwabara, Y. Maruta, M. Tanio, and T. Kaneko, "28 GHz downlink multi-user MIMO experimental verification using 360 element digital AAS for 5G massive MIMO," in *Proc. IEEE European Microwave Conference (EuMC)*, Sept. 2018.

[8] B. Yang, Z. Yu, J. Lan, R. Zhang, J. Zhou, and W. Hong, "Digital beamforming-based massive MIMO transceiver for 5G millimeter-wave communications," *IEEE Trans. Microw. Theory Techn.*, vol. 66, no. 7, pp. 3403-3418, Jul. 2018.

[9] M. A. B. Abbasi, V. F. Fusco, H. Tataria, and M. Matthaiou, "Constant-ϵ_r lens beamformer for low-complexity millimeter-wave hybrid MIMO," *IEEE Trans. Microw. Theory Techn.*, vol. 67, no. 7, pp. 2894-2903, Jul. 2019.

[10] A. Alkhateeb, J. Mo, N. González-Prelcic, and R. W. Heath, Jr., "MIMO precoding and combining solutions for millimeter-wave systems," *IEEE Commun. Mag.*, vol. 52, no. 12, pp. 122-131, Dec. 2014.

[11] R. Méndez-Rial, C. Rusu, N. González-Prelcic, A. Alkhateeb, and R. W. Heath, Jr., "Hybrid MIMO architectures for millimeter wave communications: Phase shifters or switches?" *IEEE Access*, vol. 4, no. 1, pp. 247-267, Mar. 2016.

[12] C. Mollén, J. Choi, E. G. Larsson, and R. W. Heath, Jr., "Uplink performance of wideband massive MIMO with one-bit ADCs," *IEEE Trans. Wireless Commun.*, vol. 16, no. 1, pp. 87-100, Oct. 2016.

[13] L. Fan, S. Jin, C. K. Wen, and H. Zhang, "Uplink achievable rate for massive MIMO systems with low-resolution ADC," *IEEE Commun. Lett.*, vol. 19, no. 12, pp. 2186-2189, Dec. 2015.

[14] Q. Bai and J. Nossek, "Energy efficiency maximization for 5G multi-antenna receivers," *Trans. Emerg. Telecommun. Techn.*, vol. 26, no. 1, pp. 3-14, Jan. 2015.

[15] S. Wang, Y. Li, and J. Wang, "Multiuser detection in massive spatial modulation MIMO with low-resolution ADCs," *IEEE Trans. Wireless Commun.*, vol. 14, no. 4, pp. 2156-2168, Apr. 2015.

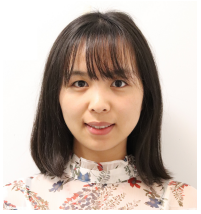
[16] C. K. Wen, C. J. Wang, S. Jin, K. -K. Wong, and P. Ting, "Bayes-optimal joint channel-and-data estimation for massive MIMO with low-precision ADCs," *IEEE Trans. Signal Process.*, vol. 64, no. 10, pp. 2541-2556, May 2016.

[17] E. Björnson, J. Hoydis, M. Kountouris, and M. Debbah, "Massive MIMO systems with non-ideal hardware: Energy efficiency, estimation, and capacity limits," *IEEE Trans. Inf. Theory*, vol. 60, no. 11, pp. 7112-7139, Nov. 2014.

[18] L. Xu, X. Lu, S. Jin, F. Gao, and Y. Zhu, "On the uplink achievable rate of massive MIMO system with low-resolution ADC and RF impairments," *IEEE Commun. Lett.*, vol. 23, no. 3, pp. 502-505, Jan. 2019.

[19] T. C. W. Schenk, *RF Imperfections in High-rate Wireless Systems: Impact and Digital Compensation*. Springer Netherlands, 2008.

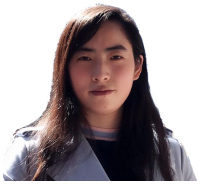
[20] N. Kolomvakis, M. Matthaiou, and M. Coldrey, "IQ imbalance in multiuser systems: Channel estimation and compensation," *IEEE Trans. Commun.*, vol. 64, no. 7, pp. 3039-3051, Jul. 2016.
[21] T. Nitsche, C. Cordeiro, A. B. Flores, E. W. Knightly, E. Perahia, and J. C. Widmer, "IEEE 802.11ad: Directional 60 GHz communication for multi-gigabit-per-second Wi-Fi," *IEEE Commun. Mag.*, vol. 52, no. 12, pp. 132-141, Dec. 2014.
[22] "IEEE 802.11ad specification. part 11: wireless LAN medium access control (MAC) and physical layer (PHY) specifications amendment 3: enhancements for very high throughput in the 60 GHz band," Dec. 2012.
[23] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, Jr., "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, pp. 831-846, Oct. 2014.
[24] O. El Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, Jr., "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499-1513, Mar. 2014.
[25] D. Donno, J. Palacios, D. Giustiniano, and J. Widmer, "Hybrid analog-digital beam training for mmWave systems with low-resolution RF phase shifters," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2016, pp. 700-705.
[26] S. Noh, M. D. Zoltowski, and D. J. Love, "Multi-resolution codebook and adaptive beamforming sequence design for millimeter wave beam alignment," *IEEE Trans. Wireless Commun.*, vol. 16, no. 9, pp. 5689-5701, Sept. 2017.
[27] D. Donno, J. Palacios, and J. Widmer, "Millimeter-wave beam training acceleration through low-complexity hybrid transceivers," *IEEE Trans. Wireless Commun.*, vol. 16, no. 6, pp. 3646-3660, Jun. 2017.
[28] W. U. Bajwa, J. Haupt, A. M. Sayeed, and R. Nowak, "Compressed channel sensing: A new approach to estimating sparse multipath channels," *Proc. IEEE*, vol. 98, no. 6, pp. 1058-1076, Apr. 2010.
[29] K. Venugopal, A. Alkhateeb, N. G. Prelcic, and R. W. Heath, Jr., "Channel estimation for hybrid architecture-based wideband millimeter wave systems," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 1996-2009, Sept. 2017.
[30] J. Yang, C. K. Wen, S. Jin, and F. Gao, "Beamspace channel estimation in mmWave systems via cosparse image reconstruction technique," *IEEE Trans. Commun.*, vol. 66, no. 10, pp. 4767-4782, Oct. 2018.
[31] Technical Specification Group Radio Access Network; Study on New Radio (NR) Physical channels and modulation (Release 15), document 3GPP TS 38.211 V15.7.0, Sep. 2019.
[32] Technical Specification Group Radio Access Network; Study on New Radio (NR) Physical layer procedures for control (Release 15), document 3GPP TS 38.213 V15.7.0, Sep. 2019.
[33] A. M. Sayeed, "Deconstructing multiantenna fading channels," *IEEE Trans. Signal Process.*, vol. 50, no. 10, pp. 2563-2579, Oct. 2002.
[34] J. Brady and A. M. Sayeed, "Beamspace MU-MIMO high density gigabit small-cell access at millimeter-wave frequencies," in *Proc. IEEE Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Jun. 2014, pp. 80-84.
[35] J. Song, J. Choi, and D. J. Love, "Common codebook millimeter wave beam design: Designing beams for both sounding and communication with uniform planar arrays," *IEEE Trans. Commun.*, vol. 65, no. 4, pp. 1859-1872, Apr. 2017.
[36] A. M. Sayeed, "A virtual representation for time- and frequency-selective correlated MIMO channels," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, May 2003, vol. 4, pp. 648-651.
[37] X. Quan, *et al.*, "A 52-57 GHz 6-bit phase shifter with hybrid of passive and active structures," *IEEE Microw. Wireless Compon. Lett.*, vol. 28, no. 3, pp. 236-238, Mar. 2018.
[38] J. J. Bussgang, "Crosscorrelation functions of amplitude-distorted Gaussian signals," *MIT Res. Lab. Electron.*, Cambridge, MA, USA, Tech. Rep. 216, Mar. 1952.
[39] X. Zhang, M. Matthaiou, E. Björnson, and M. Coldrey, "Impact of residual transmit RF impairments on training-based MIMO systems," *IEEE Trans. Commun.*, vol. 63, no. 8, pp. 2899-2911, Aug. 2015.
[40] E. Björnson, M. Matthaiou, and M. Debbah, "Massive MIMO with non-ideal arbitrary arrays: Hardware scaling laws and circuit-aware design," *IEEE Trans. Wireless Commun.*, vol. 14, no. 8, pp. 4353-4368, Aug. 2015.
[41] E. Björnson, M. Matthaiou, and M. Debbah, "A new look at dual-hop relaying: Performance limits with hardware impairments," *IEEE Trans. Commun.*, vol. 61, no. 11, pp. 4512-4525, Nov. 2013.
[42] T. T. Cai and L. Wang, "Orthogonal matching pursuit for sparse signal recovery with noise," *IEEE Trans. Inf. Theory*, vol. 57, no. 7, pp. 4680-4688, Jun. 2011.
[43] C. Shepard, *et al.*, "Argos: Practical many-antenna base stations," in *Proc. ACM Int'l. Conf. Mobile Computing and Networking*, Aug. 2012, pp. 53-64.
[44] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289-1306, Apr. 2006.
[45] E. J. Candès and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203-4215, Dec. 2005.
[46] A. C. Gilbert, S. Guha, P. Indyk, S. Muthukrishman, and M. J. Strauss, "Near-optimal sparse Fourier representations via sampling," in *Proc. 34th ACM Symp. Theory Comput.*, May 2002, pp. 152-161.
[47] M. R. Akdeniz, et al., "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1164-1179, Jun. 2014.



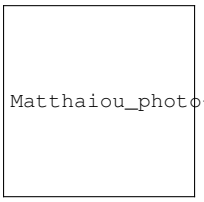
Jie Yang (S'18) received the B.S. degree in communication engineering from Nanjing University of Science and Technology, Nanjing, China, in 2015, the M.S. degree in information and communications engineering from Southeast University, Nanjing, China, in 2018. She is currently working towards the Ph.D. degree in information and communications engineering with Southeast University, Nanjing, China. Her current research interests include millimeter-wave wireless communication, massive multi-in multi-out, and compressed sensing.

Shi Jin (S'06-M'07-SM'17) received the B.S. degree in communications engineering from Guilin University of Electronic Technology, Guilin, China, in 1996, the M.S. degree from Nanjing University of Posts and Telecommunications, Nanjing, China, in 2003, and the Ph.D. degree in information and communications engineering from the Southeast University, Nanjing, in 2007. From June 2007 to October 2009, he was a Research Fellow with the Adastral Park Research Campus, University College London, London, U.K. He is currently with the faculty of the National Mobile Communications Research Laboratory, Southeast University. His research interests include space time wireless communications, random matrix theory, and information theory. He serves as an Associate Editor for the IEEE Transactions on Wireless Communications, and IEEE Communications Letters, and IET Communications. Dr. Jin and his co-authors have been awarded the 2011 IEEE Communications Society Stephen O. Rice Prize Paper Award in the field of communication theory and a 2010 Young Author Best Paper Award by the IEEE Signal Processing Society.

Chao-Kai Wen (S'00–M'04) received the Ph.D. degree from the Institute of Communications Engineering, National Tsing Hua University, Taiwan, in 2004. He was with Industrial Technology Research Institute, Hsinchu, Taiwan and MediaTek Inc., Hsinchu, Taiwan, from 2004 to 2009. Since 2009, he has been with National Sun Yat-sen University, Taiwan, where he is Professor of the Institute of Communications Engineering. His research interests center around the optimization in wireless multimedia networks.



Xi Yang received the the B.S. degree and the M.S. degree from Southeast University, Nanjing, China in 2013 and 2016 respectively. She is currently working toward the Ph.D. degree with the School of Information Science and Engineering, Southeast University. Her main research interests include wireless communication system prototyping, massive MIMO, and millimeter wave communications.



Michail Matthaiou (S'05–M'08–SM'13) was born in Thessaloniki, Greece in 1981. He obtained the Diploma degree (5 years) in Electrical and Computer Engineering from the Aristotle University of Thessaloniki, Greece in 2004. He then received the M.Sc. (with distinction) in Communication Systems and Signal Processing from the University of Bristol, U.K. and Ph.D. degrees from the University of Edinburgh, U.K. in 2005 and 2008, respectively. From September 2008 through May 2010, he was with the Institute for Circuit Theory and Signal Processing, Munich University of Technology (TUM), Germany working as a Postdoctoral Research Associate. He is currently a Reader (equivalent to Associate Professor) in Multiple-Antenna Systems at Queen's University Belfast, U.K. after holding an Assistant Professor position at Chalmers University of Technology, Sweden. His research interests span signal processing for wireless communications, massive MIMO systems, hardware-constrained communications, mm-wave systems and deep learning for communications.

Dr. Matthaiou and his coauthors received the IEEE Communications Society (ComSoc) Leonard G. Abraham Prize in 2017. He was awarded the prestigious 2018/2019 Royal Academy of Engineering/The Leverhulme Trust Senior Research Fellowship and recently received the 2019 EURASIP Early Career Award. His team was also the Grand Winner of the 2019 Mobile World Congress Challenge. He was the recipient of the 2011 IEEE ComSoc Best Young Researcher Award for the Europe, Middle East and Africa Region and a co-recipient of the 2006 IEEE Communications Chapter Project Prize for the best M.Sc. dissertation in the area of communications. He has co-authored papers that received best paper awards at the 2018 IEEE WCSP and 2014 IEEE ICC and was an Exemplary Reviewer for IEEE COMMUNICATIONS LETTERS for 2010. In 2014, he received the Research Fund for International Young Scientists from the National Natural Science Foundation of China. He is currently the Editor-in-Chief of Elsevier Physical Communication and a Senior Editor for IEEE WIRELESS COMMUNICATIONS LETTERS. In the past, he was an Associate Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS, Associate Editor/Senior Editor for IEEE COMMUNICATIONS LETTERS and was the Lead Guest Editor of the special issue on "Large-scale multiple antenna wireless systems" of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS.