Hybrid Precoding Based on Monopulse Ratio for Millimeter Wave Systems With Limited Feedback

HOIL KIM, (Member, IEEE), CHAN-BYOUNG CHAE, (Senior Member, IEEE), AND KWANG SOON KIM, (Senior Member, IEEE)

1School of Electrical and Electronic Engineering, Yonsei University, Seoul 03722, South Korea
2School of Integrated Technology, Yonsei University, Seoul 03722, South Korea

Corresponding author: Kwang Soon Kim (ks.kim@yonsei.ac.kr)

This work was supported in part by the National Research Foundation of Korea (NRF) funded by the Korea Government (MSIT) under Grant 2019R1A2C2007982 and in part by the Institute for Information & Communications Technology Promotion (IITP) funded by the Korea Government [Ministry of Science and ICT (MSIT)] (Virtual Presence in Moving Objects through 5G and High Accurate Positioning Enabled MIMO Transmission and Network Technologies for Next 5G-V2X Services) under Grant 2018-0-00170 and Grant 2016-0-00208.

ABSTRACT Multiuser hybrid precoding in 5G NR mmWave communication system faces significant challenges such as establishing accurate directional radio links and maintaining links for mobile stations (MSs) moving in outdoor environments. A conventional solution relies on finite codebook-based beam sweeping for initial first-stage beam acquisition and subsequent second-stage beam tracking by sweeping adjacent beam pairs. However, such a conventional solution has inevitable residual AoA/AoD errors even after the best beam pair is established in the first stage and incurs nonnegligible overheads to sound adjacent beam pairs for maintaining the best beam pair in the second stage. To overcome these problems, a novel codebook-based two-stage solution that combines a novel beam tracking protocol with a low sounding overhead, a unique receiver structure employing a beam scheduler and a beam tester, and a fine accuracy residual AoA/AoD error estimation algorithm based on the monopulse ratio concept is proposed. The solution is unique because of the receiver structure for the residual AoA/AoD error estimation that exploits the cyclic prefix in OFDM systems, inspired by the monopulse ratio in radar systems. Moreover, it can be applied to MSs with a single RF chain. This solution, using the proposed receiver structure and algorithm, can establish a more accurate directional beam pair right after the initial beam sweeping in the first stage. For beam tracking in the second stage, it estimates the residual AoA/AoD errors of the current best beam pair rather than sweeping adjacent beam pairs, thereby reducing beam tracking overheads. Numerical evaluation and computer simulations show that the proposed solution offers more accurate beam acquisition (i.e., average array gain improvement of several dB) and costs considerably reduced beam sounding overheads compared to the conventional solution. Lastly, a ray-tracing tool is used to demonstrate that our solution is effective in practical channel parameters for outdoor environments.

INDEX TERMS Monopulse ratio, mmWave, multiuser hybrid precoding, limited feedback, AoA/AoD, beam acquisition, beam tracking, cyclic prefix, OFDM, single RF chain.

I. INTRODUCTION

Millimeter wave (mmWave) band is highly anticipated for use in a mobile broadband radio access system. MmWave communication can solve the large bandwidth requirement issues caused by the ever-increasing data rate demands of user devices [1]–[6]. However, path loss in the mmWave band is more severe than that in lower frequencies [6]. To compensate for path loss, it is necessary to use directional transmission and reception using the array gain of large antenna arrays [2], [7]. Fortunately, large antenna arrays have feasible form factors because of the small wavelength in the mmWave band [8], [9]. Full digital baseband (BB) precoding is impractical because of the high power consumption and hardware (HW) costs of radio frequency (RF) chain components in the mmWave band [10]–[12]. Furthermore, analog-only beamforming exhibits a significantly degraded performance in downlink multiuser environments [13]. As a
tradeoff, researchers [6], [14]–[17] recently proposed hybrid precoding comprising analog RF precoding and digital BB precoding. In fact, this technology has low cost and good performance, and therefore, it is now considered a key enabling technology in mmWave communication systems [6], [7].

As a use case in the fifth generation (5G) new radio (NR) access specification, enhanced mobile broadband (eMBB) [6] aims at higher peak and average data rates than the fourth generation (4G) long-term evolution (LTE) technology. Therefore, researchers have studied how 5G NR supports large bandwidths in mmWave frequency bands for commercial deployment [18] and also (analog RF) beam-related operations, called beam management in 3GPP [19]; the objective of such research was to setup and maintain good-quality directional links. In addition, using multiple RF chains at the base station (BS), 5G NR supports digital BB precoding for spatial multiplexing in multiuser multiple-input multiple-output (MU-MIMO) with limited feedback, a feature already added to LTE to improve spectral efficiency. The number of simultaneous streams for multiusers (i.e., number of RF chains) is far less than the number of transmit antennas. Hence, a good analog RF precoding using a limited feedback is expected to be essential for hybrid precoding in mmWave MU-MIMO systems. In the initial access phase, finding the best RF beam pair between BS and each mobile station (MS), as shown in Fig. 1, is a crucial operation to achieve sufficient array gain and good link quality for subsequent data transmission. Further, fast movement or self-rotation of a MS may occur during the data transmission phase [20]. Thus, to maintain a stable link under mobility or time varying angles, the system needs sounding overheads in protocols for periodic candidate beam pair measurements (called beam tracking), as shown in Fig. 1.

**FIGURE 1.** A mmWave multiuser MIMO downlink system model.

There are several challenges to be considered when designing a hybrid precoding to meet the above mentioned requirements. First, as a tradeoff between accuracy and delay, most mmWave systems employ codebook-based hybrid precoding where RF precoders/combiners are selected from a finite size RF beam search codebook. Thus, there can be a considerably large sounding overhead for the periodic measurement of the candidate RF beam pairs to maintain the concurrent best beam pair, as in Fig. 1. Second, among many types of codebooks, discrete Fourier transform (DFT) based beam-steering codebooks have been widely used because they can achieve the maximum array gain at the designated beam directions and are well parameterized with beam directions. However, the performance of an average array gain can be degraded by the residual angle of arrival (AoA)/angle of departure (AoD) error because of the finite grid of angles in the codebook. Finally, MSs may have a single RF chain to lower HW cost or activate only a single RF chain to save power. As a result, when sweeping the adjacent candidate RF beam pairs to track the best beam pair, there might be nonnegligible sounding overheads in protocols, thereby causing loss of overall system throughput. In summary, multiuser hybrid precoding in the mmWave systems must overcome challenges so that beam acquisition is as accurate as possible and the beam tracking protocol is employed with sounding overheads as low as possible even for the MS with the single RF chain.

### A. RELATED WORK

We classified the literature on hybrid precoding in mmWave MIMO systems into three research trend categories. In the first category, researchers assume that downlink antenna-level channel state information is available at the transmitter (CSIT); they further assume that, to maximize spectral efficiency, joint/sequential optimization of analog/digital precoding matrix pair is performed toward the optimal, fully digital precoder $F_{opt}$ calculated from the provided perfect (or estimated) antenna-level CSIT [6], [21]–[29]. These research outputs are meaningful for revealing the performance upper bound of hybrid precoding with the perfect (or estimated) antenna level CSIT. Nevertheless, channel state information (CSI) is difficult to acquire in practice because of the considerable estimation overhead and low signal-to-noise (SNR) before analog beamforming [13], [30].

The second category is hybrid precoding that exploits statistical CSI [31] or spatial channel covariance [32]–[37]. In [31], by assuming small angular spreads, the transmit channel covariance is obtained from the statistics of the channel such as the mean of AoD, the angular spread of AoD, and the mean delay of channel clusters; the RF precoders are selected as the eigenvectors of the channel covariance matrix. In [32]–[37], the channel covariance matrix is utilized to design the analog precoding of a hybrid precoding without using instantaneous CSIT.

The third category is based on RF beam sweeping to find the best RF beam pair. Therefore, this category is better suited than the others for commercial systems such as 5G NR [13], [30], [38]–[46]. In the literature, RF beam sweeping has a long history of both research and practical deployment. In [38], codebook-based analog precoding was considered in wireless personal area networks for stationary device-to-device wireless connections. In [30], to attain fast beam alignment in mmWave backhaul links, the authors used adaptive subspace sampling and hierarchical codebook designed by
subarray-wise beam broadening. To estimate the AoA error, the authors in [41] used phase difference across spatially separated receive subarrays along with multiple receive RF chains. Their aim was to reduce RF beam codebook search complexity in single user and single carrier systems. Their method, however, does not compensate for the residual AoA error and cannot be applied to the single RF chain condition. In [40], the authors designed a sequential search on RF and BB codebook for a single user; this approach provides an initial version of two-stage hybrid precoding for a single user system. In [45], a novel analog codebook using a nonuniform quantization and a low-complexity hybrid precoding was proposed to reduce the feedback overhead. In [13], [42], [43], [46], two-stage hybrid precoding methods were considered for single RF chain users. Reference [13] uses a full sweeping over DFT-based RF beam search codebooks and a multiuser zero-forcing (ZF) BB precoding based on the reported effective BB channels for a multiuser narrowband system. Reference [42] and [43] are enhancements of [13] using Kalman or minimum mean squared error (MMSE) precoding. These, however, do not further investigate compensating for the residual AoA/AoD error. Reference [46] presents a tone-based AoA estimation for a narrowband multipath channel. Although it can improve AoA estimation accuracy, the AoA error compensation capability is not fully exploited. In [44], an auxiliary beam pair-enabled AoA/AoD error estimation is suggested by measuring/compensating the residual AoA/AoD error using the measurements of the left and right adjacent beams near the current best beam pair. It can be applied to multiuser wideband OFDM system with a single RF chain receiver. Although it requires additional auxiliary beams, it shows that there is still room for performance improvement in RF beam steering with a finite size DFT-based codebook. Afterwards, the conventional solution is defined as the method that uses the DFT-based codebook for beam acquisition/tracking and does not estimate/compensate the residual AoA/AoD errors.

**B. CONTRIBUTIONS**

In this paper, we present a novel solution for two-stage multiuser hybrid precoding based on monopulse ratio for mmWave systems. The contributions of the paper are as follows.

- We propose an enhanced solution for two-stage multiuser hybrid precoding that can achieve more accurate beam acquisition/tracking and considerably lower sounding overheads in a beam tracking protocol than that using the conventional solution. The proposed solution consists of a novel beam tracking protocol, a unique receiver structure, and an algorithm to estimate the residual AoA/AoD error. It works well for a single RF chain MS in both wideband multipath channels and time-varying angle conditions.
- Even with inter-block interference (IBI) and a multipath interference (MPI), cyclic prefix (CP) combined with monopulse ratio can be successfully exploited for residual AoA error estimation with a considerably reduced root of mean squared error (RMSE); thus more accurate beam acquisition and beam tracking are possible both in the first and second stages.
- The proposed solution requires sounding only at the current best beam pair in the second stage. This is more efficient than the conventional solution that sweeps all adjacent beam pairs around the current best beam pair.
- We demonstrate that it can reduce overheads significantly under realistic channel and mobility scenarios.

We use the following notation: $\mathbf{A}$ is a matrix, $a$ is a vector, $a$ is a scalar, and $\mathcal{A}$ is a set. $\| \mathbf{A} \|_F$ is the Frobenius norm of $\mathbf{A}$, whereas $\mathbf{A}^\dagger$, $\mathbf{A}^*$, $\mathbf{A}^{-1}$, and $\mathbf{A}^+$ are its transpose, Hermitian, inverse, and pseudoinverse, respectively. $\mathbf{I}_N$ is the $N \times N$ identity matrix. $\mathcal{U}(a, b)$ is a uniform distribution between $a$ and $b$. $\mathbf{A} \otimes \mathbf{B}$ is the Kronecker product of $\mathbf{A}$ and $\mathbf{B}$. $\mathbf{A}^a$ is a $n$-ary Cartesian power of a set $\mathcal{A}$.

**II. SYSTEM MODEL**

A single BS multiuser mmWave system is shown in Fig. 2. We assume that there are $U$ selected MSs by an appropriate user selection algorithm such as [47]–[51], and each MS has $N_f$ antennas and a single RF chain. To perform multiuser hybrid precoding, the BS has $N_t$ antennas and $N_{RF}$ RF chains. It is assumed that $U \leq N_{RF}$. Subsequently, we explain the transmitter, RF codebook constraints, channel model, receiver operation, and optimization problem as a fully connected architecture, and then, we mention its extensibility to sub-connected architectures.

We describe the transmitter operation for downlink hybrid precoding in a multiuser MIMO OFDM system with $N$ subcarriers. The transmit random data symbol vector $\mathbf{s}[k] = [s_1[k], s_2[k], \ldots, s_U[k]]^T$ for each subcarrier $k \in \{0, 1, \ldots, N - 1\}$ satisfies power allocation constraint $\mathbb{E}[\|\mathbf{s}[k]\|^2] = \frac{1}{N} P_T$ with the average transmit power $P_T$ of the BS. The BS applies a $U \times U$ BB digital precoder matrix $\mathbf{F}_{BB}[k] = [\mathbf{f}_{b1}^B[k], \mathbf{f}_{b2}^B[k], \ldots, \mathbf{f}_{bB}^B[k]]$ to $\mathbf{s}[k]$ for all $k$. The $N$-point inverse fast Fourier transform (IFFT) is performed at each $U$ RF chain to obtain the time domain signal $\mathbf{x}(n)$ for $n = 0, 1, \ldots, N - 1$ as

$$\mathbf{x}(n) = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} \mathbf{F}_{BB}[k] \mathbf{s}[k] e^{j \frac{2 \pi k n}{N}}. \quad (1)$$

The CP of length $N_c (\leq N)$ is added to the IFFT outputs to make one OFDM block of length $N_b = N_c + N$. Finally, an $N_t \times U$ frequency flat RF precoder $\mathbf{F}_{RF} = [\mathbf{f}_{1}^{RF}, \mathbf{f}_{2}^{RF}, \ldots, \mathbf{f}_{U}^{RF}]$ is applied so that an OFDM block at the output of BS transmitter antennas becomes $\mathbf{X}_b = \mathbf{F}_{RF} \mathbf{x}(0), \ldots, \mathbf{x}((N_b - 1)) = \mathbf{F}_{RF}[\mathbf{x}(N - N_c), \ldots, \mathbf{x}(N - 1), \mathbf{x}(0), \ldots, \mathbf{x}(N - 1)]$. Therefore, the effective transmit signal at subcarrier $k$ is $\mathbf{F}_{RF} \mathbf{F}_{BB}[k] \mathbf{s}[k]$, which is confirmed via IFFT/IFFT relation.

Considering analog phase shifters as RF precoding implementation, there are several constraints on RF precoders.
The entries of $F_{RF}$ satisfy $|F_{RF}|^2 = 1/N_t$. In addition, the phases of the analog phase shifters are discrete and have predefined finite values [6]. For a $B_{BS}$ bits phase shifter, $|F_{RF}|_{nm} = \frac{1}{\sqrt{N_c}} e^{j\phi_{nm}}$ with equally spaced phases $\phi_{nm} \in S_{BS} = \{2\pi v/2B_{BS} \mid v = 0, 1, \ldots, 2B_{BS} - 1\}$. Here, we assume $N_t \leq 2B_{BS}$ for uniform linear array (ULA), and max$(N_{t,x}, N_{t,y}) \leq 2B_{BS}$ for uniform planar array (UPA) of $N_t = N_{t,x} \times N_{t,y}$. The normalization constraint on $F_{BB}[k]$ is $|F_{RF}F_{BB}[k]|^2 = U$ for all $k$ to satisfy the total average power $P$. In this paper, our RF precoder codebook with notation $F_{HW}$ is an oversampled DFT-based beamsteered codebook with size $|F_{HW}| = N_{HW}N_t$ (ULA) or $|F_{HW}| = N_{HW}^2N_t$ (UPA), where $N_{HW} > 1$ denotes the oversampling factor per dimension (azimuth/elevation). In particular, for the BS phase shifters that have $B_{BS}$ bits, $F_{HW}$ can be defined for ULA or UPA as (for integers $a, b$)

$$F_{HW} = \left\{ a \left( \frac{2\pi a}{2B_{BS}} \right) \mid 0 \leq a < 2B_{BS} \right\}, \quad \text{or} \quad (2)$$

$$F_{HW} = \left\{ a \left( \frac{2\pi a}{2B_{BS}} - \frac{2\pi b}{2B_{BS}} \right) \mid 0 \leq a, b < 2B_{BS} \right\}, \quad \text{or} \quad (3)$$

where the meaning of the array response vector $a_t(\cdot)$ can be found in [13], [52], [53]. Here, each column vector $F_{RF}[u]$, $u$ becomes a member of $F_{HW}$. The RF combiner codebook $N_{HW}$ is defined similarly with $B_{BS}$.

A wideband geometric channel with multiple paths can be modeled as in [27], [54] for mmWave OFDM systems. The delay-$d$ MIMO $N_x \times N_t$ channel matrix $H_{u,d}$ of the $u$-th MS for ULA or UPA is written as

$$H_{u,d} = \frac{N_y}{d} \sum_{\ell=1}^{N_y} \alpha_{u,\ell} e^{j(\psi_{u,\ell} - \pi/2)} a_t(\psi_{u,\ell}) a_t^*(\psi_{u,\ell}'), \quad \text{or} \quad (4)$$

$$H_{u,d} = \frac{N_y}{\sqrt{d}} \sum_{\ell=1}^{N_y} \alpha_{u,\ell} e^{j(\psi_{u,\ell} - \pi/2)} a_t(\psi_{u,\ell} - \psi_{u,\ell}') a_t^*(\psi_{u,\ell} - \psi_{u,\ell}'), \quad \text{or} \quad (5)$$

where $N_p \equiv N_{cl}N_{ray}$ is the number of all rays (paths) in $N_{cl}$ channel clusters, $N_{ray}$ is the number of rays per cluster, $\ell$ is a ray index, $a_{u,\ell}$ is a complex gain with $E[|a_{u,\ell}|^2] = N_t/N_{p}$, and $T_s$ is a system sampling time, i.e., $T_s = \frac{1}{B} = \frac{1}{N_{HW}}$ for a system bandwidth $B$ and a subcarrier spacing $\Delta f$. Here, $p(dT_s - \tau_{u,\ell})$ denotes a pulse shaping filter, obtained from an evaluation at $t = dT_s$ with time delay $\tau_{u,\ell}$. The BS array response vector for ULA or UPA is written as

$$a_t(\psi_{u,\ell}) = N_{t} \frac{1}{2} \begin{bmatrix} 1, e^{j\psi_{u,\ell}}, \ldots, e^{j(N_{t}-1)\psi_{u,\ell}} \end{bmatrix}^T, \quad \text{or} \quad (6)$$

$$a_t(\psi_{u,\ell}^x, \psi_{u,\ell}^y) = N_{t} \frac{1}{2} \begin{bmatrix} 1, e^{j\psi_{u,\ell}^x}, \ldots, e^{j(N_{t}-1)\psi_{u,\ell}^x} \end{bmatrix}^T \otimes N_{t} \frac{1}{2} \begin{bmatrix} 1, e^{j\psi_{u,\ell}^y}, \ldots, e^{j(N_{t}-1)\psi_{u,\ell}^y} \end{bmatrix}^T. \quad (7)$$

The relation from the $\ell$-th ray’s spatial frequency $\psi_{u,\ell}$ of ULA or $(\psi_{u,\ell}^x, \psi_{u,\ell}^y)$ of UPA to the physical azimuth (elevation) angle of departure $\phi_{u,\ell}^x(\theta_{u,\ell}^x)$ is given by [52]

$$\psi_{u,\ell} = \pi \sin(\phi_{u,\ell}^x), \quad \text{or} \quad (8)$$

$$\psi_{u,\ell}^x = \pi \sin(\phi_{u,\ell}^x) \cos(\phi_{u,\ell}^y), \quad \text{or} \quad (9)$$

where the antenna element distance equals $0.5\lambda$, $\lambda = 3 \times 10^8/f_c$ is the carrier wavelength, and $\phi_{u,\ell}^x, \phi_{u,\ell}^y \in [-\pi, \pi], \theta_{u,\ell}^x \in [-\pi/2, \pi/2]$. The array response vector of the MS is defined similarly.

As a DFT of the sum of $D$-taps time domain channel matrix $H_p(t) = \sum_{d=0}^{D-1} H_{u,d} \delta(t - dT_s)$ with $D \leq N_x$ and $(D - 1)T_s \geq \tau_{u,\ell}$ for all paths $\ell$, the corresponding frequency domain channel matrix $H_u[k]$ for subcarrier $k$ is given as [53]

$$H_u[k] = \sum_{d=0}^{D-1} H_{u,d} e^{-j2\pi kd/N}, \quad (10)$$

FIGURE 2. A single BS and multisuser mmWave system where the RF precoder has a fully connected architecture as a representative case.
and it is further decomposed into rays’ gain, delay, and angle parameters for ULA or UPabhängig

\[ \mathbf{H}_u[k] = \sum_{\ell=1}^{N_p} \mathbf{a}_u(\psi'_{u,\ell}) \mathbf{a}_u^*(\psi'_{u,\ell}) \]  

(11)

\[ \mathbf{H}_u[k] = \sum_{\ell=1}^{N_p} \mathbf{a}_u(\psi'_{u,\ell}) \mathbf{a}_u^*(\psi'_{u,\ell}) \]  

(12)

Here, \( \beta_{u,\ell}[k] \) is the frequency selective gain caused by the rays’ delay \( \tau_{u,\ell} \) and the pulse shaping filter \( p(t) \) defined as [34]

\[ \beta_{u,\ell}[k] = \sum_{d=0}^{D-1} p(dT_s - \tau_{u,\ell}) e^{-j2\pi k d}. \]  

(13)

Note that \( \beta_{u,\ell}[k] \) becomes frequency flat if the system bandwidth is sufficiently small, i.e., \( \max \tau_{u,\ell} \ll T_s = 1/B \) [55].

Then, the received signal at the antenna of the \( u \)-th MS is

\[ \mathbf{r}_u(n) = \sum_{d=0}^{D-1} \mathbf{H}_{u,d} \mathbf{F}_\text{RF} \mathbf{x}_b(n-d) + \mathbf{n}_u(n). \]  

(14)

where \( \mathbf{n}_u(n) \sim \mathcal{CN}(0, \sigma^2 I_{N_u}) \) is an AWGN vector, and the signal after the RF combiner of the \( u \)-th MS is expressed as

\[ y_u(n) = \sum_{d=0}^{D-1} \mathbf{w}_u^* \mathbf{H}_{u,d} \mathbf{F}_\text{RF} \mathbf{x}_b(n-d) + \mathbf{w}_u^* \mathbf{n}_u(n). \]  

(15)

The frequency domain received signal in subcarrier \( k \) after FFT (discarding the cyclic prefix) is given as

\[ y_u[k] = \mathbf{w}_u^* \mathbf{H}_u[k] \mathbf{F}_\text{RF} \mathbf{x}_b[k] + \mathbf{w}_u^* \mathbf{n}_u[k] \]

(16)

\[ = \mathbf{w}_u^* \mathbf{H}_u[k] \sum_{i=1}^U \mathbf{F}_\text{RF}_i \mathbf{BB}[k] \mathbf{s}_i[k] + v_u[k], \]  

(17)

where \( \mathbf{w}_u \) has the same HW constraints as the BS RF precoders (i.e., equal gain magnitude and uniformly quantized phase set \( \mathcal{S}_\text{MS} \) with \( B_{\text{MS}} \) bits). \( \mathbf{n}_u[k] \) is a frequency domain representation of \( \{\mathbf{w}_u^* \mathbf{n}_u(n)\}_{n=0}^{N_y-1} \) and \( v_u[k] \sim \mathcal{CN}(0, \sigma^2) \) is an AWGN scalar in \( y_u[k] \).

The optimization problem of RF precoder/combiner selection from finite sets and BB precoder design is to find \( \mathbf{F}_\text{RF} \), \( [\mathbf{r}_u^{\text{BB}}[k]]_{k=0}^{N_y-1} \) and \( [\mathbf{w}_u]_{u=1}^U \) that satisfy

\[
\begin{aligned}
&\left\{ \mathbf{F}_\text{RF}, \left[ \left[ \mathbf{r}_u^{\text{BB}}[k] \right]_{k=0}^{N_y-1} \right]_{u=1}^U, \left[ \mathbf{w}_u \right]_{u=1}^U \right\} = \arg \max R_{\text{sum}} \\
&\text{s.t. } \left[ \mathbf{F}_\text{RF} \right]_{u} \in \mathcal{F}_\text{HW}, \mathbf{w}_u \in \mathcal{W}_\text{HW}, \forall u, \\\n&\left\| \mathbf{F}_\text{RF} \left[ \mathbf{r}_1^{\text{BB}}[k], \ldots, \mathbf{r}_U^{\text{BB}}[k] \right] \right\|_F^2 = U, \forall k.
\end{aligned}
\]  

(18)

Our chosen objective function is a system achievable sum-rate \( R_{\text{sum}} = \sum_{u=1}^U R_u \), where \( R_u \) is the achievable data rate of the \( u \)-th user per subcarrier. The optimal solution for the problem needs an exhaustive search over the discrete set \( \mathcal{F}_\text{HW} \times \mathcal{W}_\text{HW} \) space along with each \( \mathbf{r}_u^{\text{BB}}[k] \) in the continuous set \( \mathcal{C}_U^U \) provided that \( \mathbf{H}_u[k] \) is available for all \( k \) and \( u \). Currently, we identify \( \mathcal{W}_\text{HW} \) and \( \mathcal{F}_\text{HW} \) as RF beam HW codebooks, and distinguish \( \mathcal{W}_\text{HW} \) and \( \mathcal{F}_\text{HW} \) from the RF beam search codebooks, \( \mathcal{W} \) and \( \mathcal{F} \), which are used for beam sweeping in the first stage of two-stage multiuser hybrid precoding.

Even though we described the system model with a fully connected architecture as a representative case, it is straightforward to incorporate a sub-connected architecture into the system model [17]. The corresponding internal elements in the RF precoder \( \mathbf{F}_\text{RF} \), the codebooks \( \mathcal{F} \), \( \mathcal{F}_\text{HW} \) and the vectors \( \mathbf{a}_u(\cdot), \mathbf{a}_u(\cdot, \cdot) \) set to zeros based on fixed or even dynamic subarray configurations between RF chains and antenna elements in the sub-connected structure [34]–[37]. Similar reasoning can be applied to the MS side.

III. PROPOSED SOLUTION FOR TWO-STAGE MULTIUSER HYBRID PRECODING

We first explain the frame structure and concept of the proposed solution for two-stage multiuser hybrid precoding, and then, we describe its protocol, receiver structure, and algorithm.

A. FRAME STRUCTURE AND CONCEPT OF PROPOSED SOLUTION

Fig. 3 illustrates the concept of the proposed solution in comparison with the conventional solution. The frame comprises the first and second stages. The first stage can be divided into DL Beam Sweep and UL Beam Sweep. The second stage consists of multiple \( M_{\text{CSI}} \) repeats of a CSI period. The CSI period consists of CSI Acquisition and DL MU-MIMO Data Transmission. This frame structure is well suited for the two-stage multiuser hybrid precoding schemes in [13], [40], [42], [43].

The conventional solution in Fig. 3 can be viewed as a codebook-based method for beam acquisition and beam tracking procedures in 5G NR mmWave systems [56]. As shown in Fig. 3(a), the DL Beam Sweep in the first stage implies joint BS/MS beam sweeping using known reference signals (e.g., 5G NR synchronization signal (SS) blocks in an SS burst for initial directional cell search in the idle mode [57], [58]). Through beam pair tests (e.g., SNR measurement), each MS determines the best beam pair and best MS transmit beam for the following uplink beam sweeping. As shown in Fig. 3(b), the UL Beam Sweep is for the BS to perform beam sweeping and to find out the best BS beam for each MS. This is an abstraction of the 5G NR preamble based random access procedure in the standalone mode [58]. As shown in Fig. 3(c), the CSI Acquisition in the second stage has two functions. The first function is the measurement of candidate beam pairs adjacent to the current best beam pair for the MS under mobility. Based on the measurement, the BS and MS can switch to always-the-best beam pair. The \( N_{\text{CSI}} \) in the figure denotes the number of CSI reference signal (CSI-RS) symbols per CSI period. For the ULA type, \( N_{\text{CSI}} \) can be 9 (= 3 × 3) for the conventional solution. The second function is the BB CSI estimation using the CSI-RS symbols. Each MS estimates the BB CSI and reports the quantized BB.
CSI estimates to the BS. As shown in Fig. 3(d), during the DL MU-MIMO Data Transmission, the BS transmits simultaneous data streams using multiuser BB precoding. The conventional solution uses the RF precoder/combiner only in the RF beam search codebook and does not estimate/compensate the residual AoA/AoD errors.

The proposed solution shown in Fig. 3 has more accurate beam acquisition and reduced sounding overheads in beam tracking protocols owing to the use of monopulse ratio based AoA/AoD error estimators. As shown in Fig. 3(a), after DL beam sweeping, the residual AoA error of the best beam pair is identified using the monopulse ratio and then compensated. As indicated in Fig. 3(b), after UL beam sweeping, the residual AoD error of the best beam pair is estimated and compensated in the same manner. As illustrated in Fig. 3(c), the proposed solution needs $N_{CSI} = 2$ CSI-RS symbols (one for AoD tracking and the other for AoA tracking) for the ULA type (similar sounding overhead reduction can be achieved for the UPA type). Therefore, in this exemplary case, our solution can significantly reduce the number of CSI-RS symbols, $N_{CSI}$, for beam tracking, e.g., about 78% reduced CSI-RS overhead compared to the conventional solution in the ULA type. The difference is that the reporting of the candidate beam pairs measurement is replaced with AoD error reporting. In Fig. 3(d), similar to the conventional solution, multiuser MIMO data transmission using multiuser BB precoding, such as [48], [49], [59], follows. Note that the proposed solution selects the RF precoder/combiner from the RF beam HW codebook.

The advantage of the proposed solution over the conventional solution in Fig. 3 can be summarized as follows. First, the conventional solution has a limitation on array gain maximization. The BS and MSs cannot achieve maximum array gain because of the quantized grid of the RF beam search codebook even when the finer RF beam codebooks are available to the BS and MSs. However, the proposed solution estimates and compensates for the finer resolution AoA/AoD errors, and thus, it has better array gain and more accurate beam acquisition. Second, our proposed solution can detect and track changes in the AoA/AoD error of the current best beam pair even with a single RF chain. Therefore, it does not need to measure adjacent candidate beam pairs, and thus, it reduces the sounding overhead in the beam tracking protocol. In contrast, the conventional solution sweeps adjacent candidate beam pairs.

### B. DESCRIPTION OF THE PROPOSED SOLUTION

Our proposed solution consists of the protocol, receiver structure, and algorithm to enhance the two-stage multiuser hybrid precoding. First, we present the motivation of the proposed solution as follows.

The conventional solution is limited in the array gain maximization and incurs non-negligible sounding overheads in protocols for measuring adjacent candidate beam pairs under mobility. Increasing the codebook size (e.g., using oversampled DFT-based RF beam search codebooks) is not a good remedy, as it worsens beam-sweeping overheads in the protocols. To remove the limitations, we devise a receiver structure that can perform finer resolution AoA/AoD error estimation even though the nonoversampled DFT-based RF beam search codebook is used for beam sweeping.
We notice a monopulse ratio that is a well-established AoA error estimator in radar systems to track moving targets. However, simply adopting the monopulse ratio causes another obstacle that requires two separate RF chains for the respective sum and difference beam (nulling at the boresight) measurements [60]. To avoid the obstacle, one can measure the sum and difference beam outputs in a time division multiplexing manner with a single RF chain. However, if an entire OFDM block is allocated to obtain the difference beam output, then the entire OFDM block would be wasted in terms of data rate.

Currently, we are motivated by classic methods in which the CP in OFDM systems is exploited for time/frequency synchronization purposes [61]. In this paper, the CP is exploited for angle synchronization. Therefore, we design an algorithm for the unique receiver structure equipped with a monopulse ratio-based AoA/AoD error estimator; this is integrated within one OFDM block. Thus, our solution can be used to detect and track well, even with a single RF chain condition, the direction of change of the current best beam pair. Therefore, our solution does not need to measure adjacent candidate beam pairs, thereby considerably reducing the sounding overhead of beam tracking in protocols. The overhead reduction ratio is discussed in subsection V-B.

1) PROTOCOL

In Fig. 4, the protocol of the proposed solution is designed to have more accurate beam acquisition and lower sounding overhead beam tracking than the conventional solution. In the first stage, the proposed solution estimates and compensates the initial AoA/AoD error of the best beam pair found by DL/UL beam sweeping. This feature cannot be provided by the conventional solution. In the second stage, the sounding overhead $N_{\text{CSI}}$ for beam tracking has a significantly reduced value than that in the proposed solution because the proposed solution does not need to monitor adjacent candidate beam pairs.

As shown in the left half of Fig. 4, in the first stage of the conventional solution, the BS broadcasts directional reference signals and each MS finds the best beam pair by using codebook-based beam sweeping. Each MS sets the RF combiner corresponding to the identified best MS beam, and sends random access preambles. Then, the BS finds the best BS beam using codebook-based beam sweeping, sets the RF precoder as the determined best BS beam. Therefore, the BS and MSs establish directional radio links. In the second stage, at each CSI period, the MSs measure adjacent candidate beam pairs using $N_{\text{CSI}}$ CSI-RS symbols for beam tracking and BB CSI estimation. Based on the measurement results, each MS reports the quality of the candidate beam pairs and quantized BB CSI estimates. The BS sends the MU-MIMO data streams to MSs using the calculated BB precoders. At the end of the CSI period, the BS and MSs update the RF precoders/combiners based on the measurement results of the candidate beam pairs for the next CSI period.

As shown in the right half of Fig. 4, in the first stage of the proposed solution, the BS broadcasts directional reference signals and each MS finds the best beam pair by codebook-based beam sweeping. As a discriminating feature, each MS estimates/compensates the initial AoA error of the found best beam pair, sets the RF combiner corresponding to the estimated AoA, and sends random access preambles. Then, the BS finds the best BS beam using codebook-based beam sweeping, sets the RF precoder as the determined best BS beam. Therefore, the BS and MSs establish directional radio links. In the second stage, at each CSI period, for both AoD/AoA error estimation...
of the current best beam pair and BB CSI estimation, the MSs requires much smaller numbers of the CSI-RS symbols for sounding (smaller $N_{CSI}$) than those for the conventional solution. Based on the measurement results, each MS reports the AoD error and quantized BB CSI estimates. The BS sends MU-MIMO data streams to MSs using calculated BB precoders. At the end of a CSI period, the BS and MSs update RF precoders/combiners based on the estimated AoD/AoA error for the next CSI period.

2) RECEIVER STRUCTURE

a: MONOPULSE RATIO

We first explain the monopulse ratio [52], [60] for the ULA type. For a target AoA $\hat{\psi}_r$, the sum beam, which always belongs to the DFT-based HW codebook, is denoted as $w_{sum} = a_r(\hat{\psi}_r) \in \mathcal{W}_{HW}$, and the difference beam, which is used only for estimating the AoA error, is denoted as $w_{diff} = \text{DiffBeam}(w_{sum})$ with

$$\text{DiffBeam}(\mathbf{b}) = \left[ -b_0, \ldots, -b_{\frac{M}{2} - 1}, b_{\frac{M}{2}}, \ldots, b_{M - 1} \right]^T \quad (19)$$

for any array steering vector $\mathbf{b} = [b_0, \ldots, b_{\frac{M}{2} - 1}, b_M, \ldots, b_{M - 1}]^T$ with a length $M$ of an even number. For case $\psi_r = 0$ and $N_r = 16$, the beam patterns [53] of the difference beam and the sum beam illustrated in Fig. 5 are defined as

$$g_s = w^*_{diff} a_r(\psi_r), \quad g_d = w^*_{sum} a_r(\psi_r). \quad (20)$$

The monopulse ratio plotted in Fig. 6 together with the absolute values of $g_s$ and $g_d$ is defined by

$$r = \frac{\text{Im}(g_d g^*_s)}{|g_s|^2} = \tan\left(\frac{N_r e_{\psi_r}}{4}\right), \quad (21)$$

where the AoA error is $e_{\psi_r} = \psi_r - \hat{\psi}_r$ for a true AoA $\psi_r$. Note that monopulse ratio can be considered as a near-linear approximation of $e_{\psi_r}$ when $|e_{\psi_r}| \leq \pi/N_r$.

For the UPA, the sum beam for target AoA $(\hat{\psi}_r,x, \hat{\psi}_r,y)$ is $w_{sum} = w^x_{sum} \otimes w^y_{sum} = a_r(\psi_r,x) \otimes a_r(\psi_r,y)$. The difference beam for the spatial frequency $x$ or $y$ is $w^*_{diff} = \text{DiffBeam}(w^x_{sum} \otimes w_{sum}^y, w^y_{sum} \otimes w_{sum}^x)$.

The AoA errors are defined as $(e_{\psi_r,x,y}) = (\psi_r,x - \hat{\psi}_r,x, \psi_r,y - \hat{\psi}_r,y)$ for true AoAs $(\psi_r,x, \psi_r,y)$.

b: RECEIVER OPERATION

In Fig. 7, the MS receiver structure is designed to make the protocol of the proposed solution feasible assuming the ULA type. The beam scheduling module schedules the target MS beam to test in an OFDM block, and then, the beam testing module configures the target AoA $\hat{\psi}_r$ to the RF beam setter and gathers an AoA error report $\tilde{e}_r$ from the monopulse ratio based AoA error estimator, and the reference signal received power (RSRP) from the OFDM demodulator. The RF beam setter configures the difference beam $w_{diff}$ to the phase shifters in the CP duration, and subsequently, it switches to the sum beam $w_{sum}$ in the OFDM symbol duration. Using the difference beam and sum beam outputs, the monopulse ratio based AoA error estimator computes the estimate $\tilde{e}_r$.

The OFDM demodulator demodulates the OFDM symbol and reports the RSRP to the beam testing module. The RSRP can be used for best beam pair selection. When the OFDM block is not used for AoA error estimation, the RF beam setter holds the sum beam in the entire OFDM block.
the full bandwidth. After analog combining, the input signal assumed that BS uses only one RF chain and its RF precoder in Fig. 7, the received signal at the antennas of the

FIGURE 8. Monopulse ratio based AoA error estimator.

c: MONOPULSE RATIO BASED AOA ERROR ESTIMATOR

In Fig. 7, the received signal at the antennas of the nth MS is modeled with a sample index n for n = 0, . . . , N − 1 as

\[ r_u(n) = \sum_{d=0}^{D-1} H_{u,d} v_u(n - d) + n_u(n), \]

where it is assumed that BS uses only one RF chain and its RF precoder \( v_u \in \mathcal{F} \) for broadcasting the reference signals \( x_u(n) \) with the full bandwidth. After analog combining, the input signal \( y_u(n) \) of the AoA error estimator in Fig. 8 is respectively denoted as \( y_u^{\text{diff}}(n) = w_u^{\text{diff}} r_u(n) \) during the CP interval and \( y_u^{\text{sum}}(n) = w_u^{\text{sum}} r_u(n + N) \) during the last \( N_a \) samples in the current OFDM symbol; it can be represented as

\[
\begin{align*}
y_u^{\text{diff}}(n) &= \sum_{d=0}^{D-1} w_u^{\text{diff}} H_{u,d} v_u(n - d) + w_u^{\text{diff}} n_u(n), \quad (22) \\
y_u^{\text{sum}}(n) &= \sum_{d=0}^{D-1} w_u^{\text{sum}} H_{u,d} v_u(n + N - d) + w_u^{\text{sum}} n_u^{N_a}(n), \\
\end{align*}
\]

where \( n_u^{N_a}(n) = n_u(n + N) \) and \( \{x_u(n)\}_{n=-D+1}^{D-1} \) in (22) are the last \( D - 1 \) samples of the previous OFDM symbol.

First, by taking a cross-correlation with the pseudo-random reference signal sequences \( x_u(n) \), the least square estimates on the \( d \)-th tap of the time domain BB channel (i.e., \( \hat{h}_u^{\text{diff}}(d) = w_u^{\text{diff}} H_{u,d} v_u \) and \( \hat{h}_u^{\text{sum}}(d) = w_u^{\text{sum}} H_{u,d} v_u \) for \( 0 \leq d < W \), where \( W \) denotes the channel estimator tap length and \( D \) is a receiver design parameter) can be obtained as in [62] as

\[
\begin{align*}
\hat{h}_u^{\text{diff}}(d) &= \frac{1}{P} \sum_{p=0}^{P-1} y_u^{\text{diff}}(d + p)x_u^*(p), \quad (24) \\
\hat{h}_u^{\text{sum}}(d) &= \frac{1}{P} \sum_{p=0}^{P-1} y_u^{\text{sum}}(d + p)x_u^*(N + p). \quad (25)
\end{align*}
\]

Here, the cross-correlation window length \( P \) can be appropriately selected to suppress the IBI and the MPI from the previous and its own OFDM symbols as well as the AWGN\(^1\).

The effect of the selected \( P \) on the MSE of MS AoA estimation is discussed in Section V.

Then, the monopulse ratio estimate \( R \) is computed as

\[
R = \frac{3 \text{Im}[\hat{h}_u^{\text{diff}}(d^*)\hat{h}_u^{\text{sum}}(d^*)^*]}{|\hat{h}_u^{\text{sum}}(d^*)|^2}, \quad (26)
\]

where the dominant tap index \( d^* \) is found as

\[
d^* = \arg \max_{0 \leq d < W} |\hat{h}_u^{\text{sum}}(d)|^2. \quad (27)
\]

A detailed explanation on how (26) can estimate the monopulse ratio to estimate the residual AoA error w.r.t. the best receive beam steering direction is provided in Appendix A. For UPA, the monopulse ratio estimate \( R^t \) or \( R^b \) is computed with the difference beam \( w_{\text{diff}} \) or \( w_{\text{diff}} \), respectively.

Finally, based on \( R \) or \( R^t, R^b \), we can estimate the AoA error of the target MS beam for ULA or UPA by

\[
\hat{\psi}_r = \frac{4}{N_r} \tan^{-1}(R), \quad \text{or} \quad \hat{\psi}_r = \frac{4}{N_r} \tan^{-1}(R^t), \quad \hat{\psi}_r = \frac{4}{N_r} \tan^{-1}(R^b). \quad (28)
\]

Note that the proposed AoA error estimator at the MS can be easily applied to a sub-connected architecture with conventional fixed subarrays. The only difference is a reduced array gain compared to the fully connected architecture. However, for the dynamic subarray case, the situation is more complicated. The synthesis of the sum or difference beam under the dynamic subarray cannot utilize a simple closed-form analytic solution and requires a complex numerical computation [63]–[65]. However, assuming such a numerical method is available for providing sum and difference beams satisfying both (20) and (21), the proposed AoA error estimation can be applied to the dynamic subarray case in a straightforward manner.

Remark 1: The monopulse ratio based AoD estimator at the MS can be designed using the same principle. For the AoD \( \hat{\psi} \) of the target BS beam assuming the ULA type, the BS applies \( \hat{f}_{\text{diff}} = \text{DiffBeam}(f_{\text{sum}}) \) during the CP interval and then switches to \( f_{\text{sum}} = a_r(\hat{\psi}) \) during the OFDM symbol interval. The MS keeps the RF combiner as the current best MS beam (i.e., \( a_r(\hat{\psi}) \)) in the entire OFDM block.

Remark 2: The monopulse ratio based AoD estimator at BS is designed using TDD channel reciprocity.

Remark 3: Although not explicitly shown in this paper because of space limitations, we can see from the simulation results that we can improve the performance of AoA error estimation in the low SNR region by adopting a moving average filter employing several OFDM symbols in the case where the AoA does not vary much during the averaging window.

Remark 4: In this paper, the phase settling time of the analog phase shifter is assumed to be negligible compared to the CP duration, as reported in [66].

3) ALGORITHM

We present the algorithm of the proposed solution for the two-stage multiuser hybrid precoding as in Algorithm 1 for
Algorithm 1 Algorithm of the Proposed Solution

**Input:** $\mathcal{F}$ BS RF search codebook, $|\mathcal{F}| = N_f$, $\mathcal{W}$ MS RF search codebook, $|\mathcal{W}| = N_t$, $\mathcal{F} \subseteq \mathcal{F}_{HW}$, $\mathcal{W} \subseteq \mathcal{W}_{HW}$ $\mathcal{H}$ BB codebook, $|\mathcal{H}| = 2^b_{MB}$

**Output** $\mathbf{f}_{RF}$, $\mathbf{f}_{BB}$, and $\mathbf{w}_u$, $u = 1, 2, \ldots, U$

**First stage:** Beam acquisition

For each MS $u$, $u = 1, 2, \ldots, U$ (DL Beam Sweep) MS $u$ selects $\mathbf{g}_u^*$ and $\mathbf{g}_u^*$ using reference signals

$$\{\mathbf{g}_u^*, \mathbf{v}_u^*\} = \arg\max_{\mathbf{g}_u, \mathbf{v}_u \in \mathcal{H}} \|\mathbf{g}_u^T \mathbf{H}_u[k] \mathbf{v}_u^T\|^2$$

MS $u$ sets $\mathbf{w}_u$ as (31) (AoA error compensation)

For each MS $u$, $u = 1, 2, \ldots, U$ (UL Beam Sweep) BS selects $\mathbf{v}_u^*$ using UL access preamble

$$\mathbf{v}_u^* = \arg\max_{\mathbf{v}_u \in \mathcal{F}} \|\mathbf{v}_u^T \mathbf{H}_u[k] (\mathbf{w}_u)^T\|^2$$

BS sets $\mathbf{f}_{RF}^*$ as (33) (AoD error compensation)

**Second stage:** CSI acquisition and MU-MIMO data transmission

For each MS $u$, $u = 1, 2, \ldots, U$

MS $u$ estimates AoA/AoD error using $N_{CSI}$ CSI-RS

MS $u$ estimates BB channel using CSI-RS

$$\tilde{\mathbf{h}}_u[k] = \mathbf{w}_u^* \mathbf{H}_u[k] \mathbf{f}_{RF}, \forall k$$

MS $u$ quantizes $\tilde{\mathbf{h}}_u[k]$ using $\mathcal{H}$

$$\hat{\mathbf{h}}_u[k] = \arg\max_{\mathbf{h}_u[k] \in \mathcal{H}} \|\mathbf{h}_u[k] \hat{\mathbf{h}}_u[k]\|, \forall k$$

MS $u$ feedbacks $\hat{\mathbf{h}}_u[k], \forall k$ and AoD error

BS designs BB precoders $\mathbf{f}_{BB}[k]$ for all $k$

BS normalizes BB precoders for all $k$

$$\mathbf{f}_{BB}[k] = \frac{\mathbf{f}_{BB}[k]}{\|\mathbf{f}_{BB}[k]\|}, \forall k, u = 1, 2, \ldots, U$$

BS transmits MU-MIMO data by $\mathbf{f}_{RF}$, $\mathbf{f}_{BB}[k], \forall k, u$

BS updates RF combiners using reported AoD errors, and MSs update RF combiners using AOA errors

mmWave wideband OFDM systems. In the input, we state the relationship between the RF beam search codebook $\mathcal{F}$, $\mathcal{W}$ and the finer steering RF beam HW codebook $\mathcal{F}_{HW}$, $\mathcal{W}_{HW}$. In the first stage, the BS and MSs estimate and compensate the initial AoA/AoD error of the best beam pair found by beam sweeping to set up accurate directional links. In the second stage, the BS and MSs estimate and compensate other enhanced precodings in [42]. For illustrative purposes, the ZF precoding is used in Section V.

**Remark 5:** If $N_{CSI} = 1$ and UL beam sweeping plus AoA/AoD error compensation are removed in Algorithm 1, we get the conventional solution as in [13], [40].

**Remark 6:** The computational cost of the proposed solution for the downlink beam sweeping is $D \equiv 4 \times N \times |\mathcal{W}| \times |\mathcal{F}|$ real multiplications; this is common with the conventional solution. In addition, the proposed AOA error estimation in the first stage needs $8 \times P \times W + 2 \times W + 6$ real multiplications by treating one real division as 4 real multiplications as in [67]. Note that the operation in (28) is negated because it can be implemented using a look-up table. The increased cost is bounded above by $C \equiv 9 \times P \times W \leq 9 \times N_s^2$ and $C$ is far less (e.g., 2%) than cost $D$ when considering typical scenarios (e.g., $N_s = 144$, $N = 2048$, $|\mathcal{F}| = 64$, and $|\mathcal{W}| = 16$ [13], [57]). Thus, we can see that the computational cost of the proposed solution is comparable to that of the conventional solution.

The AOA error compensation routine in Algorithm 1 can be explained as follows. First, we assume that for the target beam pair $(\mathbf{g}_u^*, \mathbf{v}_u^*) = (\mathbf{a}_t(\hat{\psi}_t^u), \mathbf{a}_r(\hat{\psi}_r^u))$ or $(\mathbf{g}_u^*, \mathbf{v}_u^*) = (\mathbf{a}_t(\hat{\psi}_r^{t,x}, \hat{\psi}_r^{t,y}), \mathbf{a}_r(\hat{\psi}_t^{r,x}, \hat{\psi}_t^{r,y}))$, the AOA error estimate $\hat{\epsilon}_{\psi_t}$ or $(\hat{\epsilon}_{\psi_r^{t,x}}, \hat{\epsilon}_{\psi_r^{t,y}})$ is available. Subsequently, the AoA error is compensated for by setting the RF combiner for ULA or UPA as

$$\mathbf{w}_u = a_t(Q_{MS}(\hat{\psi}_t^u + \hat{\epsilon}_{\psi_t})), \text{ or } (30)$$

$$\mathbf{w}_u = a_r(Q_{MS}(\hat{\psi}_r^{t,x} + \hat{\epsilon}_{\psi_r^{t,x}}), Q_{MS}(\hat{\psi}_r^{t,y} + \hat{\epsilon}_{\psi_r^{t,y}})). \text{ (31)}$$

where the return value of the phase quantization function $Q_{MS}(\cdot)$ belongs to the set $\mathcal{S}_{MS}$ to meet $\mathbf{w}_u \in \mathcal{W}_{HW}$.

The AoD error compensation routine in Algorithm 1 is performed using the AoD error (measured at the BS or reported from the MSs) for ULA or UPA as

$$\mathbf{f}_{RF}^* = a_t(Q_{BS}(\hat{\psi}_t^u + \hat{\epsilon}_{\psi_t})), \text{ or } (32)$$

$$\mathbf{f}_{RF}^* = a_r(Q_{BS}(\hat{\psi}_r^{t,x} + \hat{\epsilon}_{\psi_r^{t,x}}), Q_{BS}(\hat{\psi}_r^{t,y} + \hat{\epsilon}_{\psi_r^{t,y}})). \text{ (33)}$$

**IV. PERFORMANCE ANALYSIS**

To verify that the proposed solution has more accurate beam acquisition, we evaluate the amount of gain achieved in the RF beam steering array gain and in the rate performance. The reduced sounding overhead in the beam tracking protocol under mobility is demonstrated in the next section.

**A. IMPROVEMENT IN ARRAY GAIN**

We present a numerical evaluation that shows that loss of the array gain in the conventional solution with the DFT-based RF beam search codebook can be compensated for using properly designed AoA estimators. For simplicity, a single path channel is assumed. For the evaluation, the RF beam search codebook size is set to $|\mathcal{W}| = N_{osf} \times N_t$ and $|\mathcal{F}| = N_{osf} \times N_f$ with an oversampling factor (osf) $N_{osf} \geq 1$.
1) CONVENTIONAL SOLUTION WITH $N_{osf} = 1$

The normalized array gain of a nonoversampled RF beam search codebook for ULA or UPA is given as [13], [52], [53]

$$
|G_{MS}(e_{\varphi_0})|^2 = |a_u^*(\hat{\varphi}_r) a_u(\varphi_0)|^2, \quad \text{or (34)}
$$

$$
|G_{MS}(e_{\varphi_0}, \varphi_0)|^2 = |a_u^*(\hat{\varphi}_r, \hat{\varphi}_0) a_u(\varphi_0)|^2.
$$

(35)

For ULA, the average normalized array gain can be obtained by assuming $e_{\varphi_0} \sim \mathcal{U}(-\pi/N_{r}, \pi/N_{r})$ and equations in [52] as

$$
E[|G_{MS}(e_{\varphi_0})|^2] = \int_{-\pi/N_r}^{\pi/N_r} |G_{MS}(e_{\varphi_0})|^2 \frac{N_r}{2\pi} de_{\varphi_0}
$$

$$
\approx \frac{N_r}{\pi} \int_{0}^{\pi/N_r} \frac{\sin^2(N_r e_{\varphi_0} - \varphi_0)}{N_r^2 \sin^2(\varphi_0)} \frac{de_{\varphi_0}}{de_{\varphi_0}}
$$

$$
\approx \frac{2}{\pi} \int_{0}^{\pi} \frac{\sin^2(t)}{t^2} dt, \quad \text{(36)}
$$

where $\sin x \approx x$ for small $x$ and a change of variable $N_r e_{\varphi_0} = t$ are used when deriving (a). By numerical integration, (36) amounts to $0.7737 \approx -1.1$ dB. Note that approximation (36) has no dependency on the number of receive antennas. For independent AoD and AoA errors, the total average normalized array gain is approximately $-2.2$ dB. For UPA, with independent azimuth and elevation AoA errors, the total average normalized array gain is approximately $-4.4$ dB.

### TABLE 1. Average normalized array gain in MS for oversampled codebook.

| $N_{osf}$ | 1 | 2 | 4 | 8 |
|-----------|---|---|---|---|
| $E[|G_{MS}(e_{\varphi_0})|^2]$ in dB | -1.1 | -0.29 | -0.07 | -0.02 |

2) CONVENTIONAL SOLUTION WITH $N_{osf} > 1$

In Table 1, we evaluate the average normalized array gain in MS for the ULA type for some values of $N_{osf}$ using a numerical integration assuming $e_{\varphi_r} \sim \mathcal{U}(-\pi/N_{rad}, \pi/N_{rad})$.

$$
E[|G_{MS}(e_{\varphi_0})|^2] = \int_{-\pi/N_{rad}}^{\pi/N_{rad}} |G_{MS}(e_{\varphi_0})|^2 \frac{N_{osf}N_r}{2\pi} de_{\varphi_0}
$$

$$
\approx \frac{2N_{osf}}{\pi} \int_{0}^{\pi/N_{rad}} \frac{\sin^2(t)}{t^2} dt. \quad \text{(37)}
$$

Note that, for a negligible loss (e.g., less than $0.1$ dB), $N_{osf} \geq 4$ is required. However, this causes $N_{osf}^2 \geq 16$ times increase in the overhead at the first stage, which may not be a practical choice.

3) PROPOSED SOLUTION

If we assume that the residual AoA error estimate $\hat{\varphi}_r$, when $e_{\varphi_0} = 0$, has a normal distribution with mean 0 and variance $\sigma_{\hat{\varphi}^2}$, i.e., $\hat{\varphi}_r \sim \mathcal{N}(0, \sigma_{\hat{\varphi}^2})$, then the average normalized array gain with the residual AoA error is obtained as

$$
E[|G_{MS}(e_{\hat{\varphi}_r})|^2]
$$

$$
= \int_{-\pi/N_r}^{\pi/N_r} |G_{MS}(e_{\hat{\varphi}_r})|^2 \exp \left( -\frac{e_{\hat{\varphi}_r}^2}{2\sigma_{\hat{\varphi}^2}} \right) \frac{de_{\hat{\varphi}_r}}{2\pi}\sigma_{\hat{\varphi}^2}
$$

$$
\approx 2 \int_{0}^{\pi/N_r} \frac{\sin^2(t)}{t^2} \exp \left( -\frac{t^2}{2\pi}\sigma_{\hat{\varphi}^2} \right) dt, \quad \text{(38)}
$$

which is numerically evaluated and summarized in Table 2.

### TABLE 2. Average normalized array gain after residual AoA compensation.

| $\sigma_{\hat{\varphi}^2}$ | $2^{-1} N_{r}$ | $2^{-2} N_{r}$ | $2^{-3} N_{r}$ | $2^{-4} N_{r}$ |
|--------------------------|----------------|----------------|----------------|----------------|
| $E[|G_{MS}(e_{\hat{\varphi}_r})|^2]$ in dB | -0.8 | -0.2 | -0.06 | -0.01 |

### B. RATE GAP ANALYSIS

The rate gap of the proposed solution for the two-stage hybrid precoding is defined as the difference between the rate $R_u$ achieved by the infinite sized RF/BB codebook and the rate $R_u^Q$ with finite sized codebooks as in [13]. We follow the rate gap analysis for single-path ($N_r = 1$) mmWave channels as in [13] to confirm the performance enhancement of the proposed solution.

The achievable rate of user $u$, assuming exact AoD/AoA matched RF beam steering with an infinitely large RF/BB codebook, infinite large number of bits in phase shifters, and perfect inter-user interference cancellation by ZF precoding, is given as

$$
R_u = \frac{1}{N} \sum_{k=0}^{N-1} \log_2 \left( 1 + \frac{SNR}{|h_u(k)|^2|\hat{h}_{BB}^Q[k]|^2} \right), \quad \text{(39)}
$$

where $SNR = \frac{P}{\sigma^2}$ and the $u$-th MS effective BB channel $h_u[k]$, assuming a single-path mmWave channel ($N_r = 1$) and ULA structure, can be written for each subcarrier $k$ as

$$
\hat{h}_u[k] = a_u^*(\hat{\varphi}_{r,u}) h_u[k] \left[ a_r(\hat{\varphi}_{r,1}), \ldots, a_r(\hat{\varphi}_{r,1}) \right]. \quad \text{(40)}
$$

The achievable rate of user $u$ assuming imperfect inter-user interference cancellation because of the finite size RF/BB codebooks is defined as

$$
R_u^Q = \frac{1}{N} \sum_{k=0}^{N-1} \log_2 \left( 1 + \frac{\text{SNR}}{U} \sum_{\pi \neq u} |h_u^Q[k]|^2 |\hat{h}_{BB}^Q[k]|^2 \right)^{-1} \left[ \sum_{\pi \neq u} \frac{|h_u^Q[k]|^2 |\hat{h}_{BB}^Q[k]|^2}{1 + \text{SNR}} \right]. \quad \text{(41)}
$$

Here, the $u$-th MS effective BB channel $h_u^Q[k]$ can be represented for the ULA type for subcarrier $k$ as

$$
\hat{h}^Q_u[k] = a_u^*(\hat{\varphi}_{r,u}) h_u[k] \left[ a_r(\hat{\varphi}_{r,1}), \ldots, a_r(\hat{\varphi}_{r,1}) \right]. \quad \text{(42)}
$$
where $\hat{\psi}_{u,1}$ and $\hat{\psi}_{u,1}^*$ are the estimated AoD and AoA, respectively.

Then, the average rate loss upper bound in terms of the normalized array gains after fixing the RF precoder/combiner for a single path channel can be obtained as (see [13])

$$\overline{\Delta R_u} = \mathbb{E}\left[ R_u - R_u^Q \right]$$

$$\leq \mathbb{E}\left[ \log_2 \left( \frac{1}{G_{BS}(\hat{\psi}_{u,1}, \hat{\psi}_{u,1}^* ) } \right) \right]$$

$$+ \mathbb{E}\left[ \frac{1}{N} \sum_{k=0}^{N-1} \log_2 \left( 1 + \frac{\text{SNR}}{U} \sum_{n \neq u} |\Gamma_n^Q[k] \Gamma_n^{BB}[k]|^2 \right) \right],$$

(43)

where the first term of the rate loss upper bound has an inverse power of the normalized array gains at the BS and MS. This shows that the array gain loss reduction in the first stage is transferred to the rate enhancement of the two-stage multiuser hybrid precoding for wideband OFDM systems as proved for narrowband systems in [13].

V. SIMULATION RESULTS

In this section, we perform extensive numerical simulations to compare the proposed solution with the conventional solution that has nonoversampled ($N_{osf} = 1$) or oversampled ($N_{osf} = 8$) RF beam search codebook.

A. MORE ACCURATE BEAM ACQUISITION

We randomly generate the channel matrix for each user, and we perform DL/UL beam sweeping during the first stage. The channel is not changed at transition from the first stage to the second stage. The parameter $N_{CSI}$ for beam tracking is set to 1 as beam tracking is not required in the second stage. Therefore, only BB CSI estimation is performed in CSI Acquisition. After the RF precoders/combiners and BB precoders are determined, the average achievable rates per user $\mathbb{E}\left[ \frac{1}{T} \sum_{t=1}^{T} R_u \right]$ is used as a metric for the performance comparison. Here, $R_u = \frac{1}{N} \sum_{k=0}^{N-1} \log_2 (1 + y_u[k])$, where

$$y_u[k] = \frac{\frac{P}{T} \left| \hat{w}_u^H[k] \Gamma_{RF} \Gamma_{BB}[k] \right|^2 }{\frac{P}{T} \sum_{n \neq u} \left| \hat{w}_u^H[k] \Gamma_{RF} \Gamma_{BB}[k] \right|^2 + \sigma^2}$$

(44)

denotes the signal-to-interference-noise ratio (SNIR) for subcarrier $k$. Note that the overhead for beam acquisition and beam tracking is not considered in the metric of performance comparison.

The simulation parameters are

- A BS with $N_r = 64$ (8 x 8 UPA) and 2 RF chains
- All MSs in the system have $N_r = 16$ (4 x 4 UPA) and $U$ MSs are assumed to be randomly selected out of all MSs in the first stage.

- The carrier frequency is 28 GHz and the system RF bandwidth is $B = 491.52$ MHz for the subcarrier spacing $\Delta f = 120$ kHz. The system sampling time is $T_s = 1/(N_{osf} \Delta f) \approx 2.0345$ ns. We adopt a raised cosine filter with a roll-off factor of 0.1 for the pulse shaping filter $p(t)$ [69] and the OFDM parameters are set to $N = 4096$ (8.33 $\mu$s) and $N_{osf} = 288$ (0.59 $\mu$s) [57].
- The used channel models are single path, single cluster multirays, and 4 clusters multirays models [54]. Each cluster has the same delay (i.e., the rays have the same delay in a cluster) [25], [70], [71]. The maximum allowable excess channel delay $D - 1$ is set to $N_{osf} = 1$, and the cluster delays are set to be uniformly distributed in $[0, (D - 1)T_s]$. The azimuth/elevation mean AoA/AoD of the cluster are assumed to be uniformly distributed in $[\pm \pi, \pi] / [-\pi/2, \pi/2]$, and the rays per cluster are set to have Laplacian angular distribution around the mean angle of the cluster with an angular spread of 10 degrees.
- The channel parameters are generated such that the strongest cluster resides in the first half of the CP duration and the AoA/AoD estimates of the first-stage beam sweeping with $N_{osf} = 1$ are in the neighborhood (i.e., half of beam width) of the finer ($N_{osf} = 8$) first-stage beam sweeping results. Note that the IBI caused by the previous OFDM symbol is considered in the DL/UL angle error estimation; MUI is also considered in the UL angle error estimation. The number of random channel realizations is $10^4$.
- The channel estimation correlation length $P$ can be 32, 64, and 128. The channel estimator length $W$ is set to $N_r$, because the MS receiver does not know the number of channel paths and associated delays
- The SNR in figures is user-wise SNR $\text{SNR}_{user} = \frac{P}{\sigma^2}$. The number of bits for analog phase shifters is set as $B_{BS} = \log_2(\max(N_{r,x}, N_{r,y})) + 3$, and $B_{MS} = \log_2(\max(N_{r,x}, N_{r,y})) + 3$. Thus $N_{BW}$ becomes $2^3$.
- The ZF precoding [13] is used so that the multiuser BB precoding for each subcarrier $k$ is set to

$$\Gamma_{BB}[k] = \hat{H}[k]^+, \quad (45)$$

where $\hat{H}[k] = [\hat{h}_1[k], \ldots, \hat{h}_U[k]]^T$.

As a fundamental performance for multicarrier channels with $N_{osf} = 1$, Fig. 9 shows an improved normalized RMSE of the proposed AoA estimation compared to the conventional solution according to the number of channel clusters $N_{ci}$, presence of IBI, and channel estimation correlation length $P$. Here, the normalized RMSE is obtained from the difference between the azimuth AoA of the strongest cluster and the AoA estimate, and the conventional solution with $N_{osf} = 1$ shows relatively larger normalized RMSE, i.e.,

$$\left( \int_{-\pi}^{\pi} (\epsilon(\phi))^2 N_r d\phi \right)^{1/2} / (\pi / N_r) = 1/\sqrt{3}.\quad \text{MPI caused by the increased number of the channel cluster and IBI resulting from the previous symbol worsen the RMSE of the proposed solution. However, we observe that as larger values of } P \text{ are used, the RMSE can approach the performance where}$$
FIGURE 9. MS AoA estimation performance in RMSE. Effect of $P$. The ideal (=no noise) sum beam and difference beam BB channel estimates are used for the monopulse ratio algorithm. In a noise limited region as shown in Fig. 10, the gain of using a large $P$ can be seen more clearly; therefore, we set $P = 128$ as a default value in the following tests. From the above observations, it is confirmed that CP can be successfully utilized for AoA error estimation considering the effects caused from typical IBI and MPI scenarios.

As a next step, we evaluate the SE for the proposed solution. In Fig. 11, for a single user analog-only beamforming case in the single path channel, the performance gain of the proposed solution over the conventional solution using a nonoversampled RF beam search codebook is almost the same as that obtained from the analysis in subsection IV-A. As another example, we show the performance of the system employing ULA ($N_t = 8$ and $N_r = 4$) instead of UPA. As already explained in IV-A, the performance gain is about 2.2dB lower than the gain obtained by using UPA. In Fig. 12, even for a wideband multicluster channel (4 clusters, 20 rays/cluster), the proposed solution has near comparable performance to the conventional solution using a large oversampled ($N_{osf} = 8$) RF beam search codebook. As an upper bound of all other solutions, the achievable rate of the full digital precoding obtained from a genie-aided antenna-level channel is plotted. Note that RF beamsteering to the strongest ray may not be better than RF beamsteering with the nonoversampled RF beam search codebook, especially when the channel has nonuniformly (e.g., Laplacian) distributed rays per cluster.

In Fig. 13, for multiuser hybrid precoding in the wideband multicluster channel (4 clusters, 20 rays/cluster), we compare the achievable rates of several solutions. It is observed that the proposed solution has comparable performance to the conventional solution using a large oversampled RF beam search codebook, and it shows about 4 dB performance gain over the conventional solution using a nonoversampled codebook. In the figure, we plot the performance of the analog-only beam steering to show that the multiuser interference in a high SNR regime causes considerable performance degradation if not properly cancelled. For comparison, the performance of the full digital block diagonalization performed per subcarrier in [13], [59] and the single user performance with perfect inter-user interference cancellation are plotted.
In Fig. 14, with emphasis on the number of limited feedback bits for the BB precoders, we evaluate the proposed solution for a wide range of sizes for the BB precoder random vector quantization codebook. The simulation results show that more than a certain number of BB CSI quantization bits are required to maintain the hybrid precoding gain over the analog-only beam steering.

**B. LOWER SOUNDING OVERHEAD BEAM TRACKING**

Although it is assumed that the channel does not vary during the time of the first stage beam sweeping or one CSI period, we may check the performance in a practical time-varying channel. Here, the channel parameters and matrix $H_{u,d}$ for a single cluster with 20 rays are assumed to be updated as a linear Gauss-Markov process in every $(N_g + N)T_s$ (i.e., at every OFDM symbol boundary) as

$$
\alpha_{u,\ell} = \rho \alpha_{u,\ell} + \sqrt{1 - \rho^2} \eta_{\ell},
$$

$$
\phi_{u,\ell} = \phi_{u,\ell} + \delta \pi / 180, \theta_{u,\ell} = \theta_{u,\ell} + \delta \pi / 180,
$$

where $\eta_{\ell} \sim \mathcal{N}(0, 1)$ denotes a per-ray innovation random variable and $\delta = 0.1$ degree. Here, the correlation coefficient $\rho$ is set to

$$
\rho = J_0(2\pi f_D(N_g + N)T_s) \approx 0.9947,
$$

where $J_0(\cdot)$ is the zero-th order Bessel function of the first kind, $f_D$ is the maximum Doppler frequency caused by a MS moving speed $v = 27.78$ m/s at $f_c = 28$ GHz, and the sample time $T_s = 1/(\Delta f \cdot N) = 2.0345$ ns. Then, $H_{u,d}$ for samples during an OFDM symbol are obtained by linearly interpolating the channels in the corresponding OFDM symbol boundaries. Fig. 15 shows the received signal power after analog beamforming at the MS when the SNR is 30 dB during 2000 OFDM symbols. The ‘‘Conven. ($N_{\text{osf}} = 8, N_{\text{CSI}} = 6561$)’’ or ‘‘Conven. ($N_{\text{osf}} = 1, N_{\text{CSI}} = 81$)’’ denotes the conventional method with the $N_{\text{osf}}$-oversampled DFT codebook in which $N_{\text{CSI}}$ candidate searches are performed as shown in Fig. 16 for every OFDM symbol. Note that these cases are actually impractical and used only for comparison. From Fig. 15, it is confirmed that our proposed solution tracks near the best beam pair quite well even in a practical fast time-varying channel with a far-less sounding overhead.

As a demonstration, rather than relying on simulations under a geometrical statistical channel model, we adopt a
The proposed solution is from the ray tracing tool. The sounding overhead for the channel with Doppler effects using the channel parameters as those used in the previous subsection. We construct the bandwidth is 491.52 MHz. The other parameters are all same distance of the MS trajectory points is set 0.2 m (0.2 m for MS is 2 m. The velocity of the MS is set as 20 m/s, and the set to 45 dBm. The height of the BS is 5 m, and that of the antenna element gain isotropic. The BS transmit power is half-power beamwidth and 180 degree first null and MS directional with 120 degree.

Commercial ray-tracing tool called Wireless InSite [74]. We set In Fig. 17, we show a map of Rosslyn as a test site in the scenario in a specific site, such as trends in [11], [12], [72], [73]. In Fig. 17, we show a map of Rosslyn as a test site in the commercial ray-tracing tool called Wireless InSite [74]. We set each BS antenna element gain directional with 120 degree half-power beamwidth and 180 degree first null and MS antenna element gain isotropic. The BS transmit power is set to 45 dBm. The height of the BS is 5 m, and that of the MS is 2 m. The velocity of the MS is set as 20 m/s, and the distance of the MS trajectory points is set 0.2 m (0.2 m for every 10 ms). The carrier frequency is 28 GHz, and the used bandwidth is 491.52 MHz. The other parameters are all same as those used in the previous subsection. We construct the channel with Doppler effects using the channel parameters (power and azimuth/elevation AoD and AoA of each paths) from the ray tracing tool. The sounding overhead for the proposed solution is \( N_{\text{CSI}} = 4 \) symbols per CSI period (=10 ms, 1121 symbols), whereas the conventional solution with the nonoversampled RF beam search codebook has an overhead of \( N_{\text{CSI}} = 81 \times 9 \times 10 = 81 \times 9 \times 10 \) symbols as shown in Fig. 16. We simulate the conventional solution with the oversampled RF beam search codebook, which has a sounding overhead of \( N_{\text{CSI}} = 6561 \times 81 \times 9 \times 10 \) symbols as shown in Fig. 16. This is an infeasible number in one CSI period; however, it is implemented as an offline search for performance comparison. The received signal power after the RF combiner is plotted for each solution in Fig. 18, which shows the beam tracking performance of the proposed solution is comparable to the conventional solution using an oversampled RF beam search codebook with a significantly lower sounding overhead in the beam tracking protocol.

**VI. CONCLUSION**

We proposed a novel solution for the two-stage multiuser hybrid precoding in mmWave OFDM systems such as 5G NR, which can achieve more accurate beam acquisition and lower sounding overheads in protocols than the conventional solution even for a single RF chain MS. To support the proposed two-stage protocol with lower overheads, a novel receiver design is proposed to employ AoA/AoD error estimators based on the monopulse ratio. We numerically evaluated the expected improvement in both average normalized array gain and rate gap, and we confirmed it via computer simulations under various channel conditions. In addition, as a practical demonstration, we used a ray-tracing tool in a realistic environment to confirm that the proposed solution works well in a realistic scenario. In the future work, the proposed solution will be extended to the case of using multiple RF chains at MSs in multicell environments.

**APPENDIX A ESTIMATION OF THE AOA ERROR**

We present our reasoning on how (26) can be used for the residual AoA error estimation in the following cases. For simplicity, we assume that the BS has a fully connected architecture with the ULA type and the delay of each cluster is obviously equal to \( u_0 \) [75]. With \( v_t = a_t(\psi^*_{u_t,0}) \), \( w_{\text{sum}} = a_t(\psi^*_{u_{t,\text{ini}}}) \), \( w_{\text{diff}} = \text{DiffBeam}(w_{\text{sum}}) \), and \( \Delta t = 0, 24 \) and \( 25 \) are

\[
\psi^*_{u_t,0} = \arg \max_{\psi^*_{u_{t,\text{ini}}}} |a_t(\psi^*_{u_{t,\text{ini}}})| H_{u_t,0} v_t |
\]
Then, we may assume that there is a dominant channel cluster with an additional MSE compared to the first case even in high SNR. When the channel has multiclusters with multirays, we may assume that there is a dominant channel cluster \( \ell^* \) with its delay tap \( d^* \) so that similar reasoning can be applied. Although some bias may occur because of the approximation as in the second case and further degradation due to MPI is expected in a multicluster case, (26) can successfully estimate the monopulse ratio in typical channel environments as shown in the simulation results.

**REFERENCES**

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

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

[3] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. K. Soong, and J. C. Zhang, “What will 5G be?” IEEE J. Sel. Areas Commun., vol. 32, no. 6, pp. 1065–1082, Jun. 2014.

[4] W. Chin, Z. Fan, and R. Haines, “Emerging technologies and research challenges for 5G wireless networks,” IEEE Wireless Commun., vol. 21, no. 2, pp. 106–112, Apr. 2014.

[5] F. Boccardi, R. W. Heath, Jr., A. Lozano, T. L. Marzetta, and P. Popovski, “Five disruptive technology directions for 5G,” IEEE Commun. Mag., vol. 52, no. 2, pp. 74–80, Feb. 2014.

[6] V. W. Wong. *Key Technologies for 5G Wireless Systems*. Cambridge, U.K.: Cambridge Univ. Press, 2017.

[7] W. Roh, J.-Y. Seol, J. Park, B. Lee, J. Lee, Y. Kim, J. Cho, K. Cheun, and F. Aryanfar, “Millimeter-wave beamforming as an enabling technology for 5G cellular communications: Theoretical feasibility and prototype results,” IEEE Commun. Mag., vol. 52, no. 2, pp. 106–113, Feb. 2014.

[8] B. Biglarbegian, M. Fakhrazadeh, D. Busiucu, M.-R. Nezhad-Ahmad, and S. Safavi-Naeini, “Optimized microstrip antenna arrays for emerging millimeter-wave wireless applications,” IEEE Trans. Antennas Propag., vol. 59, no. 5, pp. 1742–1747, May 2011.

[9] A. Alkhateeb, J. Mo, N. Gonzalez-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.

[10] T. S. Rappaport, R. W. Heath, Jr., R. C. Daniels, and J. N. Murdock, *Millimeter Wave Wireless Communications*. London, U.K.: Pearson, 2014.

[11] T. Kwon, Y.-G. Lim, B.-W. Min, and C.-B. Chae, “RF lens-embedded massive MIMO systems: Fabrication issues and codebook design,” IEEE Trans. Microw. Theory Techn., vol. 64, no. 7, pp. 2256–2271, Jul. 2016.

[12] Y.-J. Cho, G.-Y. Suk, B. Kim, D.-K. Kim, and C.-B. Chae, “RF lens embedded antenna array for mmWave MIMO: Design and performance,” IEEE Commun. Mag., vol. 56, no. 6, pp. 42–48, Jul. 2018.

[13] A. Alkhateeb, G. Leus, and R. W. Heath, Jr., “Limited feedback hybrid precoding for multi-user millimeter wave systems,” IEEE Trans. Wireless Commun., vol. 14, no. 11, pp. 6481–6494, Nov. 2015.

[14] X. Zhang, A. F. Molisch, and S.-Y. Kung, “Variable-phase-shifted RF-baseband code design for MIMO antenna selection,” IEEE Trans. Signal Process., vol. 53, no. 11, pp. 4091–4103, Nov. 2005.

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

[16] J. Jang, M. Chung, S. C. Hwang, Y.-G. Lim, H.-J. Yoon, T. Oh, B.-W. Min, Y. Lee, K. S. Kim, C.-B. Chae, and D. K. Kim, “Smart small cell with hybrid beamforming for 5G: Theoretical feasibility and prototype results,” IEEE Wireless Commun., vol. 23, no. 6, pp. 124–131, Dec. 2016.

[17] I. Ahmed, H. Khammari, A. Shahid, A. Musa, K. S. Kim, E. De Poorter, and I. Moerman, “A survey on hybrid beamforming techniques in 5G: Architecture and system model perspectives,” IEEE Commun. Surveys Tuts., vol. 20, no. 4, pp. 3060–3097, 4th Quart., 2018.

[18] Study On New Radio Access Technology Physical Layer Aspects, document (TR) 38.802, Version 14.2.0, 3rd Generation Partnership Project (3GPP), Sep. 2017.

[19] Study On New Radio Access Technology Radio Interface Protocol Aspects, document (TR) 38.804, Version 14.0.0, 3rd Generation Partnership Project (3GPP), Mar. 2017.

[20] M. S. Sim, Y.-G. Lim, S.-H. Park, L. Dai, and C.-B. Chae, “Deep learning-based mmWave beam selection for 5G NR/6G with sub-6 GHz channel information: Algorithms and prototype validation,” IEEE Access, vol. 8, no. 1, pp. 51634–51640, 2020.
Chan-Byoung Chae (Senior Member, IEEE) received the Ph.D. degree in electrical and computer engineering from The University of Texas at Austin, in 2008. From 2001 to 2005, he was a Research Engineer with the Telecommunications Research and Development Center, Samsung Electronics, Suwon, South Korea. From 2008 to 2009, he was a Post-doctoral Research Fellow with Harvard University, Cambridge, MA, USA. From 2009 to 2011, he was a member of Technical Staff with Bell Labs, Alcatel-Lucent, Murray Hill, NJ, USA. He was a member of the Wireless Networking and Communications Group (WNCG), The University of Texas at Austin. He is currently an Underwood Distinguished Professor with the School of Integrated Technology, Yonsei University, South Korea.

Dr. Chae was a recipient/co-recipient of the IEEE WCNC Best Demo Award, in 2020, the Young Engineer Award from the National Academy of Engineering of Korea (NAEK), in 2019, the IEEE DySPAN Best Demo Award, in 2018, the IEEE/KICS Journal of Communications and Networks Best Paper Award, in 2018, the Award of Excellence in Leadership of 100 Leading Core Technologies for Korea 2025 from the NAEK, in 2017, the Yonam Research Award from LG Yonam Foundation, in 2016, the IEEE INFOCOM Best Demo Award, in 2015, the IEIE/IEEE Joint Award for Young IT Engineer of the Year, in 2014, the KICS Hae Dong Young Scholar Award, in 2013, the IEEE Signal Processing Magazine Best Paper Award, in 2013, the IEEE ComSoc AP Outstanding Young Researcher Award, in 2012, and the IEEE VTS Dan. E. Noble Fellowship Award, in 2008. He is currently an Editor-in-Chief of the IEEE TRANSACTIONS ON MOLECULAR, BIOLOGICAL, AND MULTI-SCALE COMMUNICATIONS, since 2016, he has been serving as an Editor for the IEEE Communications Magazine. He served as an Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, from 2012 to 2017, the IEEE TRANSACTIONS ON MOLECULAR, BIOLOGICAL, AND MULTI-SCALE COMMUNICATIONS, from 2015 to 2018, the IEEE WIRELESS COMMUNICATIONS LETTERS, since 2016, and the IEEE/KICS JOURNAL OF COMMUNICATIONS AND NETWORKS, since 2012.

Hoil Kim (Member, IEEE) received the B.S. (summa cum laude), M.S.E., and Ph.D. degrees in electrical engineering from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea, in February 1994, February 1996, and February 1999, respectively.

From March 1999 to March 2000, he was a Postdoctoral Researcher with the Department of Electrical and Computer Engineering, University of California at San Diego, La Jolla, CA, USA. From April 2000 to February 2004, he was a Senior Member of Research Staff with the Mobile Telecommunication Research Laboratory, Electronics and Telecommunication Research Institute, Daejeon. Since March 2004, he has been with the Department of Electrical and Electronic Engineering, Yonsei University, Seoul, South Korea, where he is currently a Professor. His research interests include signal processing, communication theory, information theory, and stochastic geometry applied to wireless heterogeneous cellular networks, wireless local area networks, wireless D2D networks, wireless ad hoc networks, and new radio access technologies for 5G. He was a recipient of the Postdoctoral Fellowship from Korea Science and Engineering Foundation (KOSEF), in 1999. He received the Outstanding Researcher Award from Electronics and Telecommunication Research Institute (ETRI), in 2002, the Jack Neubauer Memorial Award (Best System Paper Award, IEEE Transactions on Vehicular Technology) from IEEE Vehicular Technology Society, in 2008, and the LG R&D Award: Industry-Academic Cooperation Prize, LG Electronics, in 2013. From 2006 to 2012, he served as an Editor for the Journal of the Korean Institute of Communications and Information Sciences (KICS). From 2013 to 2016, he served as the Editor-in-Chief for the Journal of KICS. Since 2008, he has been serving as an Editor of the Journal of Communications and Networks (JCN). From 2009 to 2014, he served as an Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS.

From 2004 to 2008, he was a System Engineer with Ericsson Technologies. Since 2008, he has been a Junior/Senior/Principal System Engineer with Samsung Electronics. His research interests include wireless communication and signal processing, including mmWave communications and vehicular communication.

Kwang Soon Kim (Senior Member, IEEE) received the B.S. degree in electrical engineering from POSTECH, Pohang, South Korea, in 2002, and the M.S. degree in mathematics from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea, in 2004. He is currently pursuing the Ph.D. degree with the School of Electrical and Electronic Engineering, Yonsei University, Seoul, South Korea. His research interests include signal processing, communication theory, information theory, and stochastic geometry applied to wireless heterogeneous cellular networks, wireless local area networks, wireless D2D networks, wireless ad hoc networks, and new radio access technologies for 5G. He was a recipient of the Postdoctoral Fellowship from Korea Science and Engineering Foundation (KOSEF), in 1999. He received the Outstanding Researcher Award from Electronics and Telecommunication Research Institute (ETRI), in 2002, the Jack Neubauer Memorial Award (Best System Paper Award, IEEE Transactions on Vehicular Technology) from IEEE Vehicular Technology Society, in 2008, and the LG R&D Award: Industry-Academic Cooperation Prize, LG Electronics, in 2013. From 2006 to 2012, he served as an Editor for the Journal of the Korean Institute of Communications and Information Sciences (KICS). From 2013 to 2016, he served as the Editor-in-Chief for the Journal of KICS. Since 2008, he has been serving as an Editor of the Journal of Communications and Networks (JCN). From 2009 to 2014, he served as an Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS.