MmWave Multiuser MIMO Precoding with Fixed Subarrays and Quantized Phase Shifters

Junquan Deng, Olav Tirkkonen, and Christoph Studer

***Abstract*—Millimeter-wave (mmWave) wireless communication promises high data-rates when combined with multiuser multiple-input multiple-output (MU-MIMO) technology. A practical deployment of these technologies, however, faces numerous challenges, including the design of energy-efficient analog hardware. To address these challenges, we consider a hybrid base station architecture that consists of a set of fixed subarrays with quantized phase shifters (FS-QPS). For this system, we investigate the multiuser beamforming gains with different phase-quantization levels and subarray geometries. We show that for zero forcing baseband precoding, analog precoder optimization becomes an eigenvalue maximization problem, which can be approximated efficiently by received power maximization. We develop an efficient, optimal analog precoder for well-established mmWave channel models, and provide performance bounds characterizing the required phase-shifting accuracy for a beam-steering codebook as a function of the geometric size of the subarray. To demonstrate the efficacy of the proposed FS-QPS architecture, we show simulation results using the latest 3GPP mmWave channel model for multiuser spectral efficiency, and compare our solution to existing architectures.**

I. Introduction

In fifth-generation (5G) millimeter-wave (mmWave) cellular networks, large antenna arrays at the base station (BS) are indispensable in serving large numbers of user equipments (UEs) while enabling beamforming and multiplexing gains [1]. Furthermore, mmWave signals are vulnerable to blockage. To achieve both high capacity and consistent user experience, mmWave infrastructures need to be densely deployed to increase the line-of-sight (LoS) probability, and to tackle the pathloss and blockage problems. It is estimated that an inter-site distance (ISD) of 75-100 m is required in standalone mmWave deployments [2]. As mmWave BSs need to be densely deployed to provide seamless coverage, it is critical to keep the BS cost and power consumption at a minimum.

Due to the extremely short wavelengths, and severe path loss, mmWave channels are sparse in the angular domain [3], and dominated by LoS and low-order-reflection paths, with reduced diffractions. Such channel characteristics provide an opportunity to design low-complexity architectures that achieve beamforming and multiplexing gains comparable to a fully-digital architecture, but at reduced hardware costs. To this end, hybrid beamforming with a small number of radio frequency (RF) chains has been considered in, e.g., [3]–[10]. In hybrid architectures, a phase-shifting network is generally deployed for RF beamforming before the digital processing. The RF phase-shifting network may be fully or partially connected [4]. In a fully-connected architecture, each RF chain is connected to all antennas. For partially connected architectures [9], [11], [12], each RF chain is connected to a subarray only. As an alternative method to reduce the number of RF chains in mmWave MIMO, lens arrays have been considered in [13], [14].

A range of hybrid precoding algorithms [3], [7], [8], [10], [11], [15]–[18] have been discussed in the literature, many of which are designed for single-user MIMO communication with a fully-connected architecture. Some of these can be directly adapted for MU-MIMO and subarray architectures. For example, sparse precoding [3] via orthogonal matching pursuit (OMP) has been considered for MU-MIMO in [19].

In [7], an alternative minimization technique using manifold optimization (MO) and semidefinite relaxation (SDR) for hybrid precoder design was investigated, assuming perfect channel state information (CSI) and infinite-resolution phase shifters are available. Both fully and partially connected architectures were investigated. Gradient descent [10] and coordinate descent methods [16] were proposed to find an optimized RF precoder, where the phase of a single phase shifter is updated at a time by minimizing the objective of a single-variable subproblem. If each user channel comprises one dominant path, a fully-connected hybrid architecture performs similarly to a fully-digital architecture [8]. For generic mmWave channels, it was shown in [10] that if the number of RF chains is larger than twice the number of total data streams, hybrid precoding with a fully-connected architecture can achieve the same performance as achieved via fully digital precoding.

In the literature, infinite resolution phase shifters are generally assumed for hybrid precoding [3], [6], [7], [18], [20]. A high-resolution phase-shifting network is costly, however, especially when the number of antennas is large [21]. The RF phase-shifting network in a hybrid architecture can be implemented in the analog RF domain before the frequency mixers, or in the local oscillator path [22]. Real-world mmWave RF phase-shifter networks are subject to a finite resolution with a few controllable bits, and phase-shifting errors. To further reduce the cost and energy consumption, phase shifts

Junquan Deng (e-mail: junquan.deng@aalto.fi) was with the Department of Communications and Networking, Aalto University, Finland. He is now with the sixty-third research institute, National University of Defence Technology, Nanjing, China. Olav Tirkkonen (e-mail: olav.tirkkonen@aalto.fi) is with the Department of Communications and Networking, Aalto University, Finland. Christoph Studer (e-mail: studer@cornell.edu) is with the School of Electrical and Computer Engineering, Cornell University, NY, USA. This work was supperted in part by the China Scholarship Council, grant 201403170444, and the Academy of Finland, grant 319484. The work of Christoph Studer was supported in part by Xilinx Inc. and the US NSF under grants ECCS-1408006, CCF-1535897, CCF-1652065, CNS-1717559, and ECCS-1824379.

for signals from all antennas can be jointly controlled, e.g., using a Butler matrix [23], [24] to perform beam-steering. Considering such hardware constraints, some of the proposed hybrid precoding solutions, e.g., those based on OMP [3], manifold optimization [7], successive interference cancelation (SIC) [11], which require high-resolution and independent phase shifters, become difficult to implement in practice.

The effects of phase shifter quantization to hybrid precoding performance have been investigated, e.g., in [16], [25], [26] for the single user scenario, and in [8], [10], [15], [17] for multiple users. Generally, the quantization constraint for each phase shifter is taken into account by quantizing the RF precoder given by a solution with infinite precision, either during [10] or after [15] the optimization process. It was shown that the fully-connected hybrid architecture with coarse (e.g. 1-bit and 2-bit) and independently controllable phase shifters incurs tolerable performance loss compared to architectures with high resolution phase shifters. Accordingly, using a fully-connected hybrid architecture with low-resolution phase shifters provides a method to lower the hardware cost, without jeopardizing performance. In [15], the *RF-Quantized Maximum Ratio Transmission (MRT)* precoder was applied to quantize the phase infinite resolution precoders in a fully connected architecture, to maximize received power. Such direct quantization has linear complexity in the number of antennas. It has been extensively discussed in the literature on CSI feedback under the name of co-phasing feedback [27]—phases of antennas are adjusted with a finite granularity to maximize the received signal amplitude. There is an integer programming component in finding the optimal precoder, however, and simple algorithms to find the optimal co-phasing precoder are not known.

A fully connected low-resolution architecture is still difficult to implement, as the number of required phase shifters is large, and the RF routing circuit is complex. In [8], complexity in a fully connected architecture was reduced by applying a beam-steering codebook, while in [28], subarray hybrid beamforming via low-cost Butler phase-shifting was considered, providing initial simulation analysis. A multi-subarray mmWave architecture with a switch network is considered in [29], where a joint subarray selection and baseband precoding problem is formulated and solved using a group sparse approach, considering fixed subarray beams aligned to LoS paths are applied. In [17], a Discrete-Fourier-Transform (DFT) based finite resolution beam-steering fully connected hybrid precoding scheme is analyzed in multi-user scenario. Spectral efficiency bounds are provided as functions of the number of antennas, users, and RF-chains. Subarrays were however not considered in this work. In this regard, the performance of subarray-based multiuser hybrid precoding with low-resolution phase shifting is an important problem, which, however, has not been thoroughly investigated. Furthermore, simple algorithms for finding the optimal RF-precoders are not known, and the loss from using beam-steering as opposed to independently controlling the phases of the antennas is not known. The effect of subarray geometry has not been investigated.

This paper fills these gaps. We consider a low-hardware-complexity Fixed-Subarray Quantized-Phase-Shifter (FS-QPS) hybrid architecture with only a few RF chains and a moderate number of low-resolution phase shifters. Two hardware-constrained RF codebooks are applied for the RF precoding. Concretely, our key contributions are as follows:

- We show that when finding a baseband precoder, the constraint related to additional power reduction caused by the non-ideality of hybrid precoding can be left out from the optimization, and treated afterwards. This simplifies baseband precoder design.
- We show that if a zero forcing baseband precoder is used, the RF-precoding problem becomes an eigenvalue maximization problem, which can be approximated by a received power maximization problem.
- We provide an $N \log N$ time algorithm to find the optimal precoder in a finite resolution independent phase shifting codebook. We provide a bound on the loss from using per-antenna quantized MRT, and find that the loss is negligible already for moderate phase-shifter resolutions.
- We show that when beam-steering codebooks with low-resolution phase quantization are used, contiguous subarrays are optimal, and provide an analytical bound on the required phase-shifter resolution as a function of the geometric size of subarrays.
- We evaluate downlink (DL) multiuser spectral efficiency performance by simulations in a geometry-based stochastic channel model based on the most recent mmWave channel modeling efforts. The effects of hardware constraints, including the subarray geometries and the quantized phase-shifting network, are investigated, and analytic insights are confirmed.

As a reference fully digital precoder we shall consider zero-forcing (ZF). Optimal fully digital precoders would be based on dirty-paper coding (DPC) [30], which is known to achieve the MU-MIMO capacity if perfect CSI is available at the BS. Other non-linear precoding methods to approach the MU-MIMO capacity have also been devised (see e.g. [31], [32]). However, most of these precoding schemes are computationally demanding when the number of BS antennas is large. Linear precoding such as ZF is an attractive MU-MIMO downlink precoding approach, which offers comparable performance to DPC for large antenna arrays [33]–[35], and where linearity in the operation leads to moderate complexity. Also, it is known that when the number of tranmsit antennas increases, the difference between ZF and optimal linear precoding based on linear minimum mean-square error (LMMSE) transmission, becomes small [36].

The following notation is used. $\mathbf{A}$ is a matrix, $\mathbf{a}$ is a vector, and $\mathcal{A}$ is a set. The cardinality of $\mathcal{A}$ is $|\mathcal{A}|$. For a matrix, $(\mathbf{A})_{\mathcal{I}}$ denotes a sub-matrix of $\mathbf{A}$ with columns indexed by a set $\mathcal{I}$, while $(\mathbf{a})_{\mathcal{I}}$ denotes the vector with entries of $\mathbf{a}$ indexed by $\mathcal{I}$. The Frobenius norm, transpose, Hermitian transpose, conjugate, inverse, and pseudo-inverse of $\mathbf{A}$ are denoted by $\|\mathbf{A}\|_{\mathrm{F}}$, $\mathbf{A}^{\mathrm{T}}$, $\mathbf{A}^{\mathrm{H}}$, $\mathbf{A}^{*}$, $\mathbf{A}^{-1}$, and $\mathbf{A}^{\dagger}$ respectively. $\mathbf{I}_N$ is the $N \times N$ identity matrix, $\mathrm{diag}(\mathbf{a})$ is a diagonal matrix with diagonal entries from the vector $\mathbf{a}$. $\mathbf{A} \otimes \mathbf{B}$ is the Kronecker product of $\mathbf{A}$ and $\mathbf{B}$; $\mathbf{A} \odot \mathbf{B}$ denotes the Hadamard product.

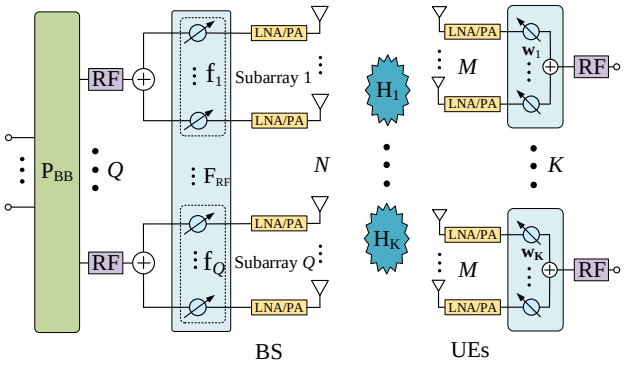


Fig. 1. Fixed-Subarray Quantized-Phase-Shifters (FS-QPS) system architecture for the BS. LNA denotes low noise amplifier and PA denotes the power amplifier. BS antennas are grouped into subarrays, each subarray is associated with one specific RF chain. Each UE has a single RF chain with M phase shifters.

II. System Model

A. System Architectures of BS and UE

The considered BS and UE architectures are depicted in Fig. 1. The BS has N antennas and Q RF chains to serve $K = Q$ UEs in the DL. The BS antennas are indexed by N consecutive integers in $\mathcal{N} = \{1, 2, \ldots, N\}$, and the K UEs are indexed by $\mathcal{K} = \{1, 2, \ldots, K\}$. The indexes of antennas associated to the qth RF chain are denoted by $\mathcal{S}_q$. We assume that the number of RF chains Q is a factor of N, and subarrays have the same size. For fixed subarrays, the N BS antennas are grouped into Q disjoint subsets $\mathcal{S}_1, \mathcal{S}_2, \ldots, \mathcal{S}_Q \subset \mathcal{N}$, and we have $\bigcup_{q=1}^{Q} \mathcal{S}_q = \mathcal{N}$. At the UE side, each UE is assumed to have a single RF chain with M antennas and M phase shifters.

The system is assumed to work in time division duplex (TDD) mode, and the DL and uplink (UL) channels are assumed to be reciprocal. In a wideband channel with orthogonal frequency-division multiplexing (OFDM), the same analog precoders and combiners have to be used across all subcarriers [9], [37], [38]. Given the analog precoders and combiners, digital precoder design and performance analysis happens on a per-subcarrier basis, with a narrowband channel model. The narrowband DL MIMO channel matrix for UE $k \in \mathcal{K}$ on a subcarrier is a $M \times N$ matrix with complex entries, $\mathbf{H}_k \in \mathbb{C}^{M \times N}$, the BS baseband precoder is denoted by $\mathbf{P}_{\mathrm{BB}} = [\mathbf{p}_1, \ldots, \mathbf{p}_K] \in \mathbb{C}^{Q \times K}$, the BS RF precoder by $\mathbf{F}_{\mathrm{RF}} = [\mathbf{f}_1, \ldots, \mathbf{f}_Q] \in \mathbb{C}^{N \times Q}$, and the UE RF combiners by $\mathbf{W} = [\mathbf{w}_1, \ldots, \mathbf{w}_K] \in \mathbb{C}^{M \times K}$. In numerical evaluations, we evaluate performance in a multicarrier OFDM system with frequency selective channels.

On one OFDM subcarrier, the DL received signal plus interference and noise for UE k is then

$$y_k = \mathbf{w}_k^{\mathrm{H}} \mathbf{H}_k \mathbf{F}_{\mathrm{RF}} \left(\mathbf{p}_k x_k + \sum\nolimits_{j \in \mathcal{K} \setminus k} \mathbf{p}_j x_j \right) + \mathbf{w}_k^{\mathrm{H}} \mathbf{n}_k, \qquad (1)$$

where x_k is the DL signal for UE k satisfying $\mathbb{E}\{x_k x_k^*\} = K^{-1}\rho_{\mathrm{BS}}$, and UE noise $\mathbf{n}_k \in \mathbb{C}^{M \times 1}$ is modeled as $\mathcal{CN}(0, \sigma^2 \mathbf{I}_M)$ with noise power σ^2. The multiuser precoder $\mathbf{F} = \mathbf{F}_{\mathrm{RF}} \mathbf{P}_{\mathrm{BB}}$ satisfies the per-subcarrier power constraint $\mathrm{tr}(\mathbf{F}\mathbf{F}^{\mathrm{H}}) \leq 1$. For frequency-domain power allocation, and we assume that the transmit power is equally distributed over all OFDM subcarriers. Such a power allocation constraint is also reasonable as the number of antennas is relatively large, and the channel hardening phenomenon [39] starts to take effect.

Note that if similar hybrid precoding would be applied in an frequency division duplexing (FDD) system, the RF-precoder $\mathbf{F}_{\mathrm{RF}}$ might still be based on reciprocity, while feedback from the UEs to the BS would be needed for choosing the digital baseband precoders $\mathbf{P}_{\mathrm{BB}}$.

B. RF Codebooks with Hardware Constraints

The BS is assumed to adopt quantized phase shifters. If each phase shifter can be independently adjusted, the qth RF chain applies an RF codeword $\mathbf{f}_q$ in the *independent phase-shifting codebook*

$$\mathcal{P}_q \triangleq \left\{ \mathbf{1}_{\mathcal{S}_q} \odot \mathbf{f} : \mathbf{f} \in \mathbb{P}^{N \times 1} \right\}, \qquad (2)$$

where $\mathbf{1}_{\mathcal{S}_q}$ is a $N \times 1$ binary vector with entries indexed by $\mathcal{S}_q$ being one and others being zero, $\mathbb{P} = \{\omega^0, \omega^1, \ldots, \omega^{2^B - 1}\}$ is the available phase set for the quantized phase shifters, with $\omega = \mathrm{e}^{\frac{\mathrm{j}2\pi}{2^B}}$, and B is the quantization level of each independent phase shifter. For example, if $B = 2$, the entries in $\mathbf{f}_q$ indexed by $\mathcal{S}_q$ would be selected from $\mathbb{P} = \{1, \mathrm{j}, -1, -\mathrm{j}\}$ independently, while the other entries would vanish. Such an RF codebook has been considered, for example, in [10]. However, independent phase control for each phase shifter is difficult to realize in practice. As the angular resolution of an array is related to the array size, given by the number of antennas, and antenna separation, a larger B is needed for larger arrays, if losses due to hybrid precoding are to be minimized. We shall see a relation between array size and B in Proposition 2 below.

Practical large-scale phase-shifter networks generally have a limited number of fixed beams generated by low-complexity circuits such as the Butler matrix [23], [24]. To steer fixed beams, the phased vector $\mathbf{f}_q$ should take values in a *beam-steering codebook*

$$\mathcal{F}_q \triangleq \left\{ \mathbf{1}_{\mathcal{S}_q} \odot \boldsymbol{f}_\omega(c, N) : c \in \{0, 1, \ldots, 2^B - 1\} \right\}, \qquad (3)$$

where B is the quantization level for the beam-steering codebook, such that there are 2^B beams. We assume that the individual entries in the beam-steering codebook come from the alphabet consisting of integer powers of $\omega = \mathrm{e}^{\frac{j2\pi}{2^B}}$, so that $\boldsymbol{f}_\omega(c, X) \triangleq [1, \omega^c, \omega^{2c}, \ldots, \omega^{(X-1)c}]^T$. A consequence of this is that if a beam-steering and independent phase-shifting codebook have the same resolution B, the beam-steering codebook is a subset, $\mathcal{F}_q \subset \mathcal{P}_q$, while the cardinality $|\mathcal{F}_q| = 2^B$ is significantly smaller than the cardinality $|\mathcal{P}_q| = 2^{B|\mathcal{S}_q|}$.

For UEs, low complexity is more important. We assume that the UE combiner $\mathbf{w}_k$ takes values in the UE beam-steering codebook

$$\mathcal{U} \triangleq \left\{ \boldsymbol{f}_\omega(c, M) : c \in \{0, 1, \ldots, 2^B - 1\} \right\}. \qquad (4)$$

As an example, for $B = 4$ and $M = 8$, there are 16 UE beam-steering codewords. The codeword associated with the beam pointing at direction 0° is $\mathbf{w} = [1, 1, \ldots, 1]^{\mathrm{T}}$, and the codeword associated with the beam pointing at direction 7.18° is $\mathbf{w} = [1, \mathrm{e}^{\mathrm{j}\frac{\pi}{8}}, \mathrm{e}^{\mathrm{j}\frac{2\pi}{8}}, \ldots, \mathrm{e}^{\mathrm{j}\frac{7\pi}{8}}]^{\mathrm{T}}$.

C. Channel Model

According to the results obtained from ray tracing [40] and measurement campaigns [41]–[43], a mmWave channel is typically comprised of a small number of dominant multi-path components in the angular domain. Following [9], for rectangular OFDM sample pulses, the DL narrowband channel matrix for UE k on subcarrier n is given by

$$\mathbf{H}_{k,n} = \sum\nolimits_{l=1}^{L} \alpha_{l,n}\ \mathbf{a}_{\mathrm{UE}}(\theta_l)\mathbf{a}_{\mathrm{BS}}^{\mathrm{H}}(\phi_l)\ . \tag{5}$$

Here, L represents the number of propagation paths, and $\alpha_{l,n} = \tilde{\alpha}_l/\sqrt{N_c}\ \sum_{d=0}^{D-1} p(dT_c - \tau_l)\exp\left(\frac{j2\pi nd}{N_c}\right)$ denotes the complex gain of the lth path depending on several factors including the complex propagation gain, and antenna element patterns [9], [41], encoded in $\tilde{\alpha}_l$, as well as the subcarrier-specific contribution of the delay τ_l in the time-domain filter $p(t)$ with a order of D. The OFDM-symbol duration is T_c, and the number of subcarriers is N_c. Furthermore, $\mathbf{a}_{\mathrm{BS}}(\phi_l)$ and $\mathbf{a}_{\mathrm{UE}}(\theta_l)$ represent the BS and UE array response vectors for the lth path, where ϕ_l is the direction of departure (DoD) at the BS, and θ_l is the direction of arrival (DoA) at the UE. When designing and analyzing digital precoders, we consider subcarrier-specific processing, and drop the index n.

We assume that uniform linear arrays (ULAs) are adopted at the BS and UEs with array response vectors

$$\begin{aligned}\mathbf{a}_{\mathrm{BS}}(\phi) &= \left[1, \mathrm{e}^{j\frac{2\pi}{\lambda}d\sin(\phi)}, \ldots, \mathrm{e}^{j(N-1)\frac{2\pi}{\lambda}d\sin(\phi)}\right]^{\mathrm{T}},\\ \mathbf{a}_{\mathrm{UE}}(\theta) &= \left[1, \mathrm{e}^{j\frac{2\pi}{\lambda}d\sin(\theta)}, \ldots, \mathrm{e}^{j(M-1)\frac{2\pi}{\lambda}d\sin(\theta)}\right]^{\mathrm{T}}.\end{aligned} \tag{6}$$

Here $j^2 = -1$, λ is the carrier wavelength, and d is the antenna spacing, assumed to be $\lambda/2$ in this paper. The beam-steering vectors in codebooks $\mathcal{F}_q$ and $\mathcal{U}$ can be written as $\mathbf{f}_q = \mathbf{1}_{\mathcal{S}_q} \odot \mathbf{a}_{\mathrm{BS}}(\phi_c)$ and $\mathbf{w}_k = \mathbf{a}_{\mathrm{UE}}(\theta_c)$, where $\phi_c = \sin^{-1}(\frac{2c}{2^B})$ is the pointing direction of a BS beam, $\theta_c = \sin^{-1}(\frac{2c'}{2^B})$ is the direction of the UE beam, and $c, c' \in \{0, \ldots, 2^B - 1\}$.

D. Effective Downlink MISO channel

When a UE beam $\mathbf{w}_k = \mathbf{a}_{\mathrm{UE}}(\theta_c) \in \mathcal{U}$ is used for DL reception, the effective DL multiple-input single-output (MISO) channel for UE k on a subcarrier is

$$\mathbf{h}_k^{\mathrm{H}} = \mathbf{w}_k^{\mathrm{H}}\mathbf{H}_k = \sum\nolimits_{l=1}^{L} \underbrace{\alpha_l \mathbf{a}_{\mathrm{UE}}^{\mathrm{H}}(\theta_c)\mathbf{a}_{\mathrm{UE}}(\theta_l)}_{\beta_l} \mathbf{a}_{\mathrm{BS}}^{\mathrm{H}}(\phi_l). \tag{7}$$

The effective MISO channel with UE phased combining contains a smaller number of significant paths than the original MIMO channel $\mathbf{H}_k$, as only those paths with θ_l close to θ_c would have significant effective path gain, here denoted by $\beta_l = \alpha_l \mathbf{a}_{\mathrm{UE}}^{\mathrm{H}}(\theta_c)\mathbf{a}_{\mathrm{UE}}(\theta_l)$. We assume that a UE uses a beam $\mathbf{w}_k$ that maximizes $\|\mathbf{w}_k^{\mathrm{H}}\mathbf{H}_k\|_2$.

For reliable performance, it is important that user channels in (7) are linearly independent, such that the joint channel covariance across all users would have rank K. If this is the case, reliable communication can be guaranteed to all users with shared analog beamforming. The rank of the covariance can be ensured by applying a user selection or grouping algorithm, e.g., using the joint spatial division and multiplexing algorithm [44].

To characterize the dominance of the strongest path in the effective channel (7), we shall be interested in the path dominance ratio

$$D = \frac{\rho_{\max}}{\rho_{\mathrm{tot}} - \rho_{\max}} \tag{8}$$

where ρ_{tot} is the total channel power, and $\rho_{\max} = |\beta_1|^2$ is the power of the strongest path, assumed to be $l = 1$.

To get intuition on path-dominance, consider a $d = \lambda/2$ ULA at the UE. The inner product between a UE precoder selected according to the strongest beam and the UE array steering vector for path i is

$$\mathbf{a}_{\mathrm{UE}}^{\mathrm{H}}(\theta_1)\ \mathbf{a}_{\mathrm{UE}}(\theta_i) = \sum_{k=0}^{M-1} e^{\pi jk(\sin\theta_i - \sin\theta_1)}\ . \tag{9}$$

If

$$|\sin\theta_i - \sin\theta_1| < \frac{1}{2(M-1)}\ , \tag{10}$$

all of the elements in the sum (9) lie within an angular cone of width $\pi/4$, and combine in a constructive manner, while for larger differences of steering angles, some elements in the sum combine in a destructive manner. In a $\lambda/2$ ULA, the half space between directions $\theta \in [-\pi/2, \pi/2]$ can be divided into M orthogonal beam directions, and the probability for two beams to fulfill (10) is proportional to the probability of the beams to have considerably overlapping main lobes, which is $\sim 1/M$. For two beams with randomly selected directions, with a large enough beam separation such that (10) is not fulfilled, the inner product (9) is well approximated by a random variate from a 2D random walk with M steps and step size 1. The expected absolute value of the inner product would thus be $\sqrt{M}$. Take a channel with L paths having the same path gain $|\alpha_i| = \alpha$ before Rx-combining, and where none of the paths $i = 2, \ldots, L$ fulfill (10). The typical path dominance ratio (8) for such a channel would be $D = M/L$, and this would increase when one of the path is stronger than the others.

III. Low-complexity Multiuser Precoding with Subarrays and Quantized Phase-shifters

We consider a low-complexity multiuser precoding scheme which applies low-resolution phase-shifters at the BS. This RF precoder is subject to both constant modulus and quantization constraints. The effective multiuser DL multiple-input multiple-output MU-MIMO channel when UEs use their selected beam codewords for data reception is $\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_K]^{\mathrm{H}} \in \mathbb{C}^{K\times N}$. The DL received signals (1), including inter-user interference and noise for all K UEs on a subcarrier can thus be written as

$$\mathbf{y} = \begin{bmatrix} \mathbf{h}_1^{\mathrm{H}}\mathbf{f}_1 & \cdots & \mathbf{h}_1^{\mathrm{H}}\mathbf{f}_Q \\ \vdots & \ddots & \vdots \\ \mathbf{h}_K^{\mathrm{H}}\mathbf{f}_1 & \cdots & \mathbf{h}_K^{\mathrm{H}}\mathbf{f}_Q \end{bmatrix} \mathbf{P}_{\mathrm{BB}}\mathbf{x} + \begin{bmatrix} \mathbf{w}_1^{\mathrm{H}}\mathbf{n}_1 \\ \vdots \\ \mathbf{w}_K^{\mathrm{H}}\mathbf{n}_K \end{bmatrix}, \tag{11}$$

where $\mathbf{x} = [x_1, \ldots, x_K]^{\mathrm{T}}$ is the transmit signal vector. Here, $\mathbf{H}\mathbf{F}_{\mathrm{RF}} \in \mathbb{C}^{K\times Q}$ is the effective channel for the digital baseband precoder.

Before digital precoding, one would expect $\mathbf{H}\mathbf{F}_{\mathrm{RF}}$ to be a diagonally dominant matrix to control the inter-user interference, i.e., $|\mathbf{h}_k^{\mathrm{H}}\mathbf{f}_q|$ is large for $k = q$ while it is small for

$k \neq q$. To achieve this property, the UE effective channels $\{\mathbf{h}_k\}_{k\in\{1,\ldots,K\}}$ should be separable in the angular domain. As a consequence, the BS would use different RF beamformers which direct energy to different directions for different UEs. We shall see in Section III-B how this emerges from SINR optimization when zero forcing basedband precoding is assumed.

Analog precoding alone is unable to mitigate inter-user interference, especially if the UEs are close to each other. In a narrowband system, the best hybrid precoding approximation of a fully digital precoder $\mathbf{F}_{\mathrm{D}}$ would be given by

$$\begin{aligned} &\underset{\mathbf{F}_{\mathrm{RF}},\mathbf{P}_{\mathrm{BB}}}{\text{minimize}} \quad \left\|\mathbf{H}\mathbf{F}_{\mathrm{D}} - \mathbf{H}\mathbf{F}_{\mathrm{RF}}\mathbf{P}_{\mathrm{BB}}\right\|_{\mathrm{F}}, \\ &\text{subject to} \quad \begin{cases} \mathbf{F}_{\mathrm{RF}} = [\mathbf{f}_1,\ldots,\mathbf{f}_Q],\ \mathbf{f}_q \in \mathcal{P}_q \text{ or } \mathcal{F}_q, \\ \left\|\mathbf{F}_{\mathrm{RF}}\mathbf{P}_{\mathrm{BB}}\right\|_{\mathrm{F}}^2 \leq 1, \end{cases} \end{aligned} \tag{12}$$

while in a multicarrier system, the selection of $\mathbf{F}_{\mathrm{RF}}$ would be coupled across subcarriers.

To attain low complexity in baseband processing, we consider the ZF precoder [33] as the reference precoder. The key feature of ZF is to eliminate all multiuser interference based on CSI. If $\mathbf{H}\mathbf{H}^{\mathrm{H}}$ is invertible, a ZF precoder is given by

$$\mathbf{F}_{\mathrm{D}} = \mathbf{H}^{\mathrm{H}}\left(\mathbf{H}\mathbf{H}^{\mathrm{H}}\right)^{-1}\mathbf{\Lambda}_{\rho}, \tag{13}$$

where $\mathbf{\Lambda}_{\rho} = \mathrm{diag}(\rho_1,\ldots,\rho_K)$ is a multiuser power allocation matrix which guarantees that $\mathbf{F}_{\mathrm{D}}$ satisfies the power constraint $\|\mathbf{F}_{\mathrm{D}}\|_{\mathrm{F}}^2 \leq 1$. For equal power allocation across the users, $\mathbf{\Lambda}_{\rho} = \rho\mathbf{I}_K$ where

$$\rho = 1/\|\mathbf{H}^{\mathrm{H}}\left(\mathbf{H}\mathbf{H}^{\mathrm{H}}\right)^{-1}\|_{\mathrm{F}}\ . \tag{14}$$

Note that in a multicarrier system, joint power allocation across users and subcarriers would be optimal. Power allocation across subcarriers may, however, have a limited effect on performance, see e.g. [45].

Finding the best hybrid precoding approximation for a fully digital ZF precoder then changes (12) to

$$\begin{aligned} &\underset{\mathbf{F}_{\mathrm{RF}},\mathbf{P}_{\mathrm{BB}}}{\text{minimize}} \quad \left\|\mathbf{\Lambda}_{\rho} - \mathbf{H}\mathbf{F}_{\mathrm{RF}}\mathbf{P}_{\mathrm{BB}}\right\|_{\mathrm{F}}, \\ &\text{subject to} \quad \begin{cases} \mathbf{F}_{\mathrm{RF}} = [\mathbf{f}_1,\ldots,\mathbf{f}_Q],\ \mathbf{f}_q \in \mathcal{P}_q \text{ or } \mathcal{F}_q, \\ \left\|\mathbf{F}_{\mathrm{RF}}\mathbf{P}_{\mathrm{BB}}\right\|_{\mathrm{F}}^2 \leq 1. \end{cases} \end{aligned} \tag{15}$$

The optimization problem in (12) can be treated as a constrained matrix factorization problem for the targeted fully-digital precoder. This problem is difficult to solve exactly due to the special requirements of $\mathbf{F}_{\mathrm{RF}}$. In addition, due to the power constraint, subarray geometric constraints and the constant amplitude constraint on phase shifters, the optimal solution given by (15) cannot ensure $\|\mathbf{\Lambda}_{\rho} - \mathbf{H}\mathbf{F}_{\mathrm{RF}}\mathbf{P}_{\mathrm{BB}}\|_{\mathrm{F}} = 0$ for generic channel conditions, even in a narrowband system. The situation in a multicarrier system differs in that one $\mathbf{F}_{\mathrm{RF}}$ should be chosen for all subcarriers. Conceptually, the difficulty in approximating subcarrier-specific digital precoders, however, are the same in a narrowband and frequency selective multicarrier system, as we shall see below.

A. Baseband Precoder Design

Given a $\mathbf{F}_{\mathrm{RF}}$, which may be a joint analog precoder for a multicarrier system, one can solve $\mathbf{P}_{\mathrm{BB}}$ for a subcarrier from (15). If $\mathbf{H}\mathbf{F}_{\mathrm{RF}}$ is invertible, a non-normalized ZF solution is given by $\tilde{\mathbf{P}}_{\mathrm{BB}} = \left(\mathbf{H}\mathbf{F}_{\mathrm{RF}}\right)^{-1}\mathbf{\Lambda}_{\rho}$. To satisfy the power constraint, $\tilde{\mathbf{P}}_{\mathrm{BB}}$ is then normalized to be

$$\mathbf{P}_{\mathrm{BB}} = \frac{1}{\beta}\tilde{\mathbf{P}}_{\mathrm{BB}}\ , \tag{16}$$

where $\beta = \|\mathbf{F}_{\mathrm{RF}}\tilde{\mathbf{P}}_{\mathrm{BB}}\|_{\mathrm{F}}$. With ZF baseband precoding, and equal power allocation for the approximated fully digital precoding as in (14), the power constraint for $\mathbf{F}_{\mathrm{RF}}\mathbf{P}_{\mathrm{BB}}$ is satisfied with

$$\beta = \|\mathbf{F}_{\mathrm{RF}}\left(\mathbf{H}\mathbf{F}_{\mathrm{RF}}\right)^{-1}\mathbf{\Lambda}_{\rho}\|_{\mathrm{F}} = \rho\|\mathbf{F}_{\mathrm{RF}}\left(\mathbf{H}\mathbf{F}_{\mathrm{RF}}\right)^{-1}\|_{\mathrm{F}}\ . \tag{17}$$

It is important to understand the possible suboptimality of the sequential approach, where Problem (15) is first solved without the normalization constraint, followed by a normalization step. As compared to normalization of the fully digital precoder (13) handled by choosing $\mathbf{\Lambda}_{\rho}$, the need for additional normalization in (16) arises from the non-ideal factorization of the precoder in the hybrid precoding architecture. If the non-normalized hybrid precoder $\mathbf{F}_{\mathrm{RF}}\tilde{\mathbf{P}}_{\mathrm{BB}}$ is a good approximation of the fully digital precoder $\mathbf{F}_{\mathrm{D}}$, the additional errors caused by the sequential normalization (16) are under control. We have

Proposition 1. *If a non-normalized hybrid precoder $\tilde{\mathbf{P}}_{\mathrm{BB}}$ approximates a fully digital precoder so that $\|\mathbf{F}_{\mathrm{D}} - \mathbf{F}_{\mathrm{RF}}\tilde{\mathbf{P}}_{\mathrm{BB}}\|_{\mathrm{F}} \leq \delta$, the approximation error from using the normalized hybrid precoder $\mathbf{P}_{\mathrm{BB}}$ in Problem (12) is bounded as $\|\mathbf{H}\mathbf{F}_{\mathrm{D}} - \mathbf{H}\mathbf{F}_{\mathrm{RF}}\mathbf{P}_{\mathrm{BB}}\|_{\mathrm{F}} \leq 2\delta\|\mathbf{H}\|_{\mathrm{F}}$.*

Proof. Assuming $\|\mathbf{F}_{\mathrm{D}}\|_{\mathrm{F}} = 1$, denote the normalization factor $\beta = \|\mathbf{F}_{\mathrm{RF}}\tilde{\mathbf{P}}_{\mathrm{BB}}\|_{\mathrm{F}}$ with $\beta \neq 1$. Using reverse triangle inequality, we have

$$\|\mathbf{F}_{\mathrm{D}} - \mathbf{F}_{\mathrm{RF}}\tilde{\mathbf{P}}_{\mathrm{BB}}\|_{\mathrm{F}} \geq |1-\beta|\,\|\mathbf{F}_{\mathrm{D}}\|_{\mathrm{F}}\,,$$

which implies $\|\mathbf{F}_{\mathrm{D}}\|_{\mathrm{F}} \leq \frac{1}{|\beta-1|}\delta$. As the Frobenius norm is submultiplicative, we have

$$\begin{aligned} &\|\mathbf{H}\mathbf{F}_{\mathrm{D}} - \mathbf{H}\mathbf{F}_{\mathrm{RF}}\mathbf{P}_{\mathrm{BB}}\|_{\mathrm{F}} \leq \|\mathbf{H}\|_{\mathrm{F}}\,\|\mathbf{F}_{\mathrm{D}} - \mathbf{F}_{\mathrm{RF}}\mathbf{P}_{\mathrm{BB}}\|_{\mathrm{F}} \\ &= \|\mathbf{H}\|_{\mathrm{F}}\left\|\mathbf{F}_{\mathrm{D}} - \mathbf{F}_{\mathrm{RF}}\tilde{\mathbf{P}}_{\mathrm{BB}} + (1-\beta^{-1})\mathbf{F}_{\mathrm{RF}}\tilde{\mathbf{P}}_{\mathrm{BB}}\right\|_{\mathrm{F}} \\ &\leq \|\mathbf{H}\|_{\mathrm{F}}\left(\delta + |\beta-1|\cdot\|\mathbf{F}_{\mathrm{D}}\|_{\mathrm{F}}\right) \leq 2\delta\|\mathbf{H}\|_{\mathrm{F}}. \end{aligned}$$

■

Note that Proposition 1 is not limited to ZF baseband precoding. Using any method to find an unconstrained hybrid precoder $\mathbf{F}_{\mathrm{RF}}\tilde{\mathbf{P}}_{\mathrm{BB}}$ that is close to the digital precoder $\mathbf{F}_{\mathrm{D}}$, and then normalizing the found solution, provides predictable good performance. For example, LMMSE baseband precoding may be considered.

Given a $\mathbf{F}_{\mathrm{RF}}$, and using the ZF precoder $\mathbf{P}_{\mathrm{BB}}$ from (16) with normalization (17) and assuming perfect CSI, the DL received signals for all UEs on a subcarrier can be written as

$$\mathbf{y} = \frac{1}{\beta}\mathbf{\Lambda}_{\rho}\mathbf{x} + \mathrm{diag}(\mathbf{W}^{\mathrm{H}}\mathbf{N})\ . \tag{18}$$

Note that the useful signal power in (18) is carried by the diagonal matrix $\beta^{-1}\mathbf{\Lambda}_{\rho}$. Compared to the fully digital precoder $\mathbf{F}_{\mathrm{D}}$ in (13), the hybrid precoder $\mathbf{F}_{\mathrm{RF}}\mathbf{P}_{\mathrm{BB}}$ suffers from a loss in the beamforming gain, which is characterized

by $1/\beta$. One should notice that $\mathrm{rank}(\mathbf{HF}_{\mathrm{RF}}) = K$ is required to arrive at (18). The SINR for user k is then

$$\gamma_k = \frac{\rho^2}{\beta^2}\frac{\rho_{\mathrm{BS}}}{MK\sigma^2} = \frac{\rho_{\mathrm{BS}}}{MK\sigma^2\|\mathbf{F}_{\mathrm{RF}}(\mathbf{HF}_{\mathrm{RF}})^{-1}\|_{\mathrm{F}}^2}\ , \qquad (19)$$

where ρ_{BS} is the BS transmit power per subcarrier.

B. RF Precoding for ZF Baseband Precoding

A ZF digital baseband precoder aims to mitigate inter-user interference, and the resulting received SINRs (19) are inversely proportional to $\|\mathbf{F}_{\mathrm{RF}}(\mathbf{HF}_{\mathrm{RF}})^{-1}\|_{\mathrm{F}}^2$. In a narrowband system, $\mathbf{F}_{\mathrm{RF}}$ should be designed to minimize this. The objective of RF precoding is thus to steer beams towards the users to increase the received power. The RF precoder design problem becomes

$$\begin{aligned} \underset{\mathbf{F}_{\mathrm{RF}}}{\text{minimize}} \quad & f(\mathbf{F}_{\mathrm{RF}}) = \left\|\mathbf{F}_{\mathrm{RF}}(\mathbf{HF}_{\mathrm{RF}})^{-1}\right\|_{\mathrm{F}}, \\ \text{subject to} \quad & \mathbf{F}_{\mathrm{RF}} = [\mathbf{f}_1,\ldots,\mathbf{f}_Q],\ \mathbf{f}_q \in \mathcal{P}_q \text{ or } \mathcal{F}_q. \end{aligned} \qquad (20)$$

The objective function $f(\mathbf{F}_{\mathrm{RF}})$ is non-convex and difficult to evaluate due to the need of matrix inversion. Furthermore, the size of search space grows exponentially as the number of antennas and the phase shifter resolution increases. As a result, directly searching over the codebooks to find the optimal solution for $\mathbf{F}_{\mathrm{RF}}$ is infeasible.

The objective function $f(\mathbf{F}_{\mathrm{RF}})$ is bounded by

$$\frac{\|\mathbf{F}_{\mathrm{RF}}\|_{\mathrm{F}}}{\|\mathbf{HF}_{\mathrm{RF}}\|_{\mathrm{F}}} \le f(\mathbf{F}_{\mathrm{RF}}) \le \|\mathbf{F}_{\mathrm{RF}}\|_{\mathrm{F}} \cdot \|(\mathbf{HF}_{\mathrm{RF}})^{-1}\|_{\mathrm{F}},$$

where $\|\mathbf{F}_{\mathrm{RF}}\|_{\mathrm{F}} = \sqrt{\sum_q |\mathcal{S}_q|}$ is fixed and

$$\|(\mathbf{HF}_{\mathrm{RF}})^{-1}\|_{\mathrm{F}} = \sqrt{\sum\nolimits_{i=1}^{K} \frac{1}{\lambda_i}},$$

with λ_i denoting the ith eigenvalue of the Hermitian matrix

$$\mathbf{A} = (\mathbf{HF}_{\mathrm{RF}})^{\mathrm{H}}(\mathbf{HF}_{\mathrm{RF}}). \qquad (21)$$

This implies that to minimize the objective function $f(\mathbf{F}_{\mathrm{RF}})$, one can maximize the eigenvalues of $\mathbf{A}$. According to the Gershgorin circle theorem [46, Chapter 6], all eigenvalues of $\mathbf{A}$ lie within at least one of the Gershgorin discs, which are

$$\{z \in \mathbb{C} : |z - a_{ii}| \le \sum\nolimits_{j\neq i} |a_{ij}|\}, \quad i = 1, 2, \ldots, K.$$

Thus, to attain large λ_i, one should maximize the diagonal entries a_{ii} of $\mathbf{A}$ while minimizing its off-diagonal elements, i.e., $\mathbf{A}$ should be diagonally dominant. The diagonal entries of $\mathbf{A}$ are $a_{qq} = |\mathbf{h}_q^{\mathrm{H}}\mathbf{f}_q|^2$. A tractable approach for creating a diagonally dominant $\mathbf{A}$ is thus to select the $\mathbf{F}_{\mathrm{RF}}$ so that the diagonal elements of $\mathbf{A}$ are maximized. The qth element of $\mathbf{F}_{\mathrm{RF}}$ would thus be a quantized spatial matched filter where the channel $\mathbf{h}_q$ of UE q is quantized, to *maximize the received power* at UE q on beam q.

An *Rx-power maximizing RF-precoder* can be found in a frequency selective channel as well. Denote the effective channel of UE k on subcarrier n as $\mathbf{h}_{k,n} = \mathbf{w}_k\mathbf{H}_{k,n}$, with the channel matrix given in (5). With RF-precoder $\mathbf{f}_k$, the sum received power for this UE across all subcarriers is $\sum_n |\mathbf{h}_{k,n}^{\mathrm{H}}\mathbf{f}_k|^2 = \mathbf{f}_k^{\mathrm{H}}\mathbf{R}_k\mathbf{f}_k$, where the wideband channel covariance matrix is

$$\mathbf{R}_k = \sum_n \mathbf{h}_{k,n}\mathbf{h}_{k,n}^{\mathrm{H}}\ . \qquad (22)$$

According to Rayleigh-Ritz theorem, the unquantized $\mathbf{f}_k$ that maximizes the sum received power is the maximum eigenvector $\tilde{\mathbf{h}}_k$ of $\mathbf{R}_k$.

A quantized precoder can then be found that maximally aligns with $\tilde{\mathbf{h}}_k$. In mmWave channels with strong LoS components, such a wideband RF-precoder would be close to what would be an optimal per subcarrier RF-precoder.

Note that the analysis in this section is based on using a ZF baseband precoder. For other precoders, such as LMMSE, the SINR expression (19) would involve interference terms, and the RF-precoding problem for maximizing SINR would become less tractable.

C. Rx-power Maximizing Finite Resolution RF precoding

In this section we address the problem of finding a Rx-power maximizing quantized precoder, if the UE channel $\mathbf{h}_k$, or the maximum eigenvector $\tilde{\mathbf{h}}_k$, of the wideband covariance matrix (22) is known.

The optimum RF-precoder depends on the codebook, and the subarray configuration. When a beam-steering codebook $\mathcal{F}_q$ is used for precoding, the phase shifters are jointly adjusted and the optimal RF precoder should be chosen as

$$\hat{\mathbf{f}}_q = \arg\max_{\mathbf{f}_q\in\mathcal{F}_q} |\mathbf{h}_q^{\mathrm{H}}\mathbf{f}_q|, \qquad (23)$$

while for independent phase-shifting with $\mathcal{P}_q$, the precoder maximizing the Rx-power for UE q is

$$\mathbf{f}_q^{\star} = \arg\max_{\mathbf{f}_q\in\mathcal{P}_q} |\mathbf{h}_q^{\mathrm{H}}\mathbf{f}_q|\ . \qquad (24)$$

Note that for each subarray $q = 1, \ldots, Q$, the set of subarray antennas $|\mathcal{S}_q|$ is fixed by the subarray partition. We shall discuss possible subarray partitions in sections III-E and IV-A.

The search space for beam-steering in (23) is of size 2^B, while for $\mathcal{P}_q$ in (24), the search space is of size $2^{B|\mathcal{S}_q|}$. Both search spaces grow exponentially with the phase quantization level B, while for independent phase shifting, there is an additional exponential growth in the number of subarray antennas $|\mathcal{S}_q|$. An exhaustive search over $\mathcal{P}_q$ is infeasible even when B and $|\mathcal{S}_q|$ are small. For example, if $|\mathcal{S}_q|=8, B=4$, a search over more than $4\cdot 10^9$ alternatives is needed.

Due to the geometry of the problem, it is not necessary to perform exhaustive search over $\mathcal{P}_q$ to find the optimal precoder (24), however. To formulate a reduced complexity algorithm, denote the phase quantization granularity by

$$\nu = \pi/2^{B-1}, \qquad (25)$$

so that the entries in precoders are of the form $(\mathbf{f}_q)_i = e^{j\nu c_i}$ with integer c_i and $i \in \mathcal{S}_q$. Furthermore, decompose the effective channel from antenna i as

$$(\mathbf{h}_k)_i = a_i e^{j\theta_i},\ i \in \mathcal{S}_q. \qquad (26)$$

When searching for the optimum independent phase-shifting codeword in (24), the objective is to align the phases of the

rotated per-antenna channels such that the amplitude of

$$h_c = \sum_{i \in \mathcal{S}_q} a_i e^{j(\theta_i - \nu c_i)} \tag{27}$$

is maximized, by finding a suitable set of integers $\{c_i\}$.

If a target direction ζ in the complex plane is selected, it is straight forward to find the per antenna phase-shift that maximally aligns h_i with the target direction. The resulting RF-precoder $\tilde{\mathbf{f}}_q$, found by directly quantizing the conjugate phase vector of the channel w.r.t. the pre-assigned phase ζ, is the RF-Quantized Maximum Ratio Transmission (MRT) precoder [15], [27]. In [15], when applied for RF-quantization in a fully connected hybrid architecture, $\zeta = 0$ was used. The entries of the Quantized MRT precoder are $(\tilde{\mathbf{f}}_q)_i = e^{j\nu\tilde{c}_i}$, where the integer quantized phases are

$$\tilde{c}_i = \lfloor (\theta_i - \zeta)/\nu \rfloor, \tag{28}$$

and $\lfloor \cdot \rfloor$ is the rounding function. The target direction ζ thus acts as a quantization bias.

As argued in [27], the overall phase of a precoder $\mathbf{f}_q$ is irrelevant, so that the precoder on one antenna, e.g. the first, can be fixed. This amounts to choosing the quantization bias as $\zeta = \theta_1$.

When Quantized MRT related to direction ζ is performed, all the per-antenna complex numbers $a_i e^{j(\theta_i - \nu c_i)}$ contributing to h_c in (27) lie in the cone between phase angles $\zeta \pm \nu/2$.

The codeword found by using a preassigned ζ may be suboptimal. The non-triviality of (24) arises from searching over the alignment direction ζ. If the optimal precoder $\mathbf{f}_q^\star$ were known, the phase of the resulting combined channel in (27) would be $\zeta^\star$. It is easy to see that the optimal precoder is then described by the per-antenna quantization w.r.t. $\zeta^\star$:

$$c_i^\star = \lfloor (\theta_i - \zeta^\star)/\nu \rfloor. \tag{29}$$

Now consider methods to refine an RF-Quantized MRT codeword with an a priori target direction ζ to an optimum precoder. As any shift $c_i \to c_i + m$ for a fixed m for all i lead to the same amplitude in (27), for any a priori ζ, there exist an optimum $\zeta^\star$ in the interval $\zeta \pm \nu/2$. An infeasible way to find the optimum precoder would be to search over the continuum of candidates ζ' in this interval, perform (28) for each ζ', and select the one which maximize the amplitude of h_c. Simpler methods can be devised by closely analyzing the possible phase values.

For simplicity assume that the a priori quantization bias is $\zeta = 0$. Applying (28), we find for each antenna an integer $\tilde{c}_i$, and a remainder angle

$$\tilde{\theta}_i = \theta_i - \nu\tilde{c}_i \ \in [-\nu/2, \nu/2]. \tag{30}$$

The corresponding contribution to (27) thus lies in the cone with angular width ν around the positive real axis in the complex plane. Now consider all possible values ζ_i' in this cone. If $\tilde{\theta}_1 > 0$, there exists a switching value $\zeta_i = \tilde{\theta}_i - \nu/2$ such that

$$c_i(\zeta') = \begin{cases} \tilde{c}_i & \text{if } \zeta_i \le \zeta' \le \nu/2 \\ \tilde{c}_i - 1 & \text{if } \zeta_i \ge \zeta' \ge -\nu/2 \end{cases} . \tag{31}$$

Here, $c_i(\zeta') = \lfloor (\theta_i - \zeta')/\nu \rfloor$ is the per-antenna quantization of θ_i with the bias ζ'. Thus if $\zeta' < \zeta_i$, the angle $\tilde{\theta}_i$ is not in the cone of angular width ν around ζ', but $\tilde{\theta}_i - \nu$ is. Similarly, if $\tilde{\theta}_i < 0$, there exists a positive $\zeta_i = \tilde{\theta}_i + \nu/2$, at which $c_i(\zeta')$ changes from $\tilde{c}_i$ to $\tilde{c}_i + 1$. Thus to find the optimal precoder it is sufficient to search over two values for each antenna, $\tilde{c}_i$ and $\tilde{c}_i - \mathrm{sign}(\tilde{\theta}_i)$, where $\tilde{c}_i$ are the outcome of Quantized MRT. The optimum precoder can thus be found with complexity $2^{|\mathcal{S}_q|-1}$, recalling that the precoder in one antenna can be fixed.

Further simplification can be achieved by sorting the ζ_i, and constructing a Quantized MRT precoder for each of the $|\mathcal{S}_q| + 1$ intervals that these values divide $[-\nu/2, \nu/2]$ to. For values of ζ' within each interval, all $c_i(\zeta')$ are fixed, and at $\zeta' = \zeta_i$, precisely one integer, $c_i(\zeta')$, changes its value, while all other remain constant. If multiple ζ' coincide, multiple c_i change value. This leads to a discrete line search for finding the optimum RF-precoder in $\mathcal{P}_q$, summarized as Algorithm 1. The dominant additive complexity of this algorithm is in sorting, with order $|\mathcal{S}_q| \log |\mathcal{S}_q|$. The multiplicative complexity is linear in $|\mathcal{S}_q|$; the channel gain has to be computed for $|\mathcal{S}_q| + 1$ alternatives. Algorithm 1 needs to be performed for each RF chain, and the overall computation complexity is $\mathcal{O}(Q|\mathcal{S}_q| \log |\mathcal{S}_q|)$.

Algorithm 1 Discrete Line Search for Optimal $\mathcal{P}_q$ Precoder

Input: Antenna-specific phases $\{\theta_i\}_{i \in \mathcal{S}_q}$.

1: Compute c_i from (28) with $\zeta = 0$.
2: $\tilde{\theta}_i = \theta_i - \nu c_i$ for $i \in \mathcal{S}_q$
3: $c_i = c_i + 1$ for all i with $\tilde{\theta}_i < 0$
4: Find permutation $\sigma(i)$ which orders the $\tilde{\theta}_i$ so that first come the negative ones in decreasing order, then the positive ones in decreasing order.
5: $g_0 = \left|\sum_i a_i e^{j(\theta_i - \nu c_i)}\right|^2$
6: **for** $n = 1$ to $|\mathcal{S}_q|$ **do**
7: $\quad c_{\sigma^{-1}(n)} = c_{\sigma^{-1}(n)} - 1$ ▷ inverse permutation σ^{-1}
8: $\quad g_n = \left|\sum_i a_i e^{j(\theta_i - \nu c_i)}\right|^2$
9: **end for**
10: $n^* = \arg\max_n g_n$ ▷ find largest channel gain
11: **for** $n = n^* + 1$ to $|\mathcal{S}_q|$ **do**
12: $\quad c_{\sigma^{-1}(n)} = c_{\sigma^{-1}(n)} + 1$
13: **end for**
14: $c_i^\star = c_i$

Output: Optimal precoder $\mathbf{f}_q^\star$ with entries $e^{j\nu c_i^\star}$

The difference between a Quantized MRT precoder $\tilde{\mathbf{f}}_q$ from (28) and the optimum precoder $\mathbf{f}_q^\star$ is that in $\mathbf{f}_q^\star$, the per-antenna contributions in (27) are better aligned with the overall phase of the combined channel h_c than in $\tilde{\mathbf{f}}_q$. The maximal misalignment of a term in (27) from the phase is ν. The relative RF-beamforming gain of the quantized MRT precoder, as compared to the optimum precoder is thus bounded as

$$\frac{|\mathbf{h}_k^{\mathrm{H}} \tilde{\mathbf{f}}_q|^2}{|\mathbf{h}_k^{\mathrm{H}} \mathbf{f}_q^\star|^2} \ge \cos^2\left(\frac{\pi}{2^{B-1}}\right) \ge 1 - 4\pi^2 2^{-2B} \ . \tag{32}$$

The difference in power gain thus vanishes exponentially in B. Moreover, this bound is loose. Especially when $|\mathcal{S}_q|$ is large, the mean loss becomes negligible, as statistically the $\tilde{\theta}_i$ are expected to be uniformly distributed in the quantization error cone.

D. Power Loss from Beam-steering Precoding

For a given resolution B, the beam-steering codebook $\mathcal{F}_q$ is a subset of the independently phase-shifting codebook $\mathcal{P}_q$. This means that one often suffers from a power loss when using $\mathcal{F}_q$ instead of $\mathcal{P}_q$, while one never gains from this shift. Conversely, the hardware and computational complexity is reduced when using $\mathcal{F}_q$. To quantify the complexity—performance tradeoff, the power loss incurred from using $\mathcal{F}_q$ should be understood. It turns out that this power loss depends on the *geometric size of the subarrays*.

Intuitively, the difference between $\mathcal{F}_q$ and $\mathcal{P}_q$ is related to multipath propagation. If there is only one DoD for a BS to transmit to in an effective channel (7), one can find a beam-steering codeword with a departure angle rather close to the optimal D2D. For the performance loss from beamsteering in single path channels we have

Proposition 2. *Consider a single path channel $\mathbf{h}_k$ with amplitude β_{l_0}, and ULA RF-precoding with a given BS beam-steering resolution B. The difference in the power gain of the beam-steering RF precoder $\hat{\mathbf{f}}_q$ in* (23) *and the Quantized MRT precoder $\tilde{\mathbf{f}}_q$ of Equation (28) is bounded by*

$$|\mathbf{h}_k^{\mathrm{H}}\tilde{\mathbf{f}}_q|^2 - |\mathbf{h}_k^{\mathrm{H}}\hat{\mathbf{f}}_q|^2 < \varepsilon^2|\beta_{l_0}|^2|\mathcal{S}_q|^2 \ ,$$

where $|\beta_{l_0}|^2|\mathcal{S}_q|^2$ is the maximum power gain, and

$$\varepsilon = \frac{\pi\sqrt{S_{\mathrm{m}}^2+1}}{2^B} \ , \tag{33}$$

with S_{m} the geometric size of the subarray in units of wavelength.

Proof. For ease of notation, we assume that the gain of the path $|\beta_1|^2 = 1$ within this proof, while the steering angle is $\sin(\phi_1)$, and the antenna separation in the ULA is $d = \lambda/2$. The absloute power scale can be simply recovered if needed. Denote $\kappa = 2^{B-1}$, $(\hat{\mathbf{f}}_q)_i = \omega^{\hat{c}_i}$ and $\hat{c} = \lfloor \kappa \sin(\phi_1) + \zeta \rfloor$, where ζ is a bias to be optimized over. Let $\hat{c}_i = (i-1)\hat{c} = (i-1)\ \kappa\sin(\phi_1) + \hat{\epsilon}_i$ be the quantized phase in the beam-steering codebook, with quantization error $\hat{\epsilon}_i$, while the Quantized MRT yields the integers $c_i = \lfloor (i-1)\kappa\sin(\phi_1) \rfloor$, with quantization errors $\epsilon_i = c_i - (i-1)\kappa\sin(\phi_1)$. Then $|\epsilon_i| \leq |\hat{\epsilon}_i|$, reflecting the fact that for the same B, we have $\mathcal{F}_q \subset \mathcal{P}_q$. With $\zeta = 0$, we have $|\hat{\epsilon}_i| \leq \frac{i-1}{2}$. Choosing the bias we can tune the error so that the maximum error at the edges of the array are halved. Recalling that in this analysis we have antenna separation $\lambda/2$, we thus get $|\hat{\epsilon}_i| \leq \frac{i-1}{4} \leq S_{\mathrm{m}}/2$. Now denote $x_i = \frac{\pi\epsilon_i}{\kappa}$, $y_i = \frac{\pi\hat{\epsilon}_i}{\kappa}$. We are interested in the difference

$$\Delta = |\mathbf{h}_k^{\mathrm{H}}\tilde{\mathbf{f}}_q|^2 - |\mathbf{h}_k^{\mathrm{H}}\hat{\mathbf{f}}_q|^2 = \left|\sum\nolimits_i \mathrm{e}^{jx_i}\right|^2 - \left|\sum\nolimits_i \mathrm{e}^{jy_i}\right|^2 .$$

Using Euler's equation, Δ can be written as a sum Δ_c of cosine-terms, and a sum Δ_s of sine-terms. For the cosine terms we have

$$\begin{aligned}\Delta_c &\leq \left(\sum\nolimits_i \cos(x_i)+\cos(y_i)\right)\left(\sum\nolimits_i \cos(x_i)-\cos(y_i)\right) \\ &< |\mathcal{S}_q| \sum\nolimits_i (|y_i|^2 - |x_i|^2).\end{aligned}$$

As $|y_i| = |\frac{\pi\hat{\epsilon}_i}{\kappa}| \leq \frac{\pi S_{\mathrm{m}}}{2^B}$, we then have $\Delta_c < |\mathcal{S}_q|^2\left(\frac{\pi S_{\mathrm{m}}}{2^B}\right)^2$. For the sine terms we have

$$\Delta_s = \left|\sum\nolimits_i \sin(x_i)\right|^2 - \left|\sum\nolimits_i \sin(y_i)\right|^2 \leq \left(\sum\nolimits_i |x_i|\right)^2 .$$

As $|x_i| = |\frac{\pi\epsilon_i}{\kappa}| \leq \frac{\pi}{2^B}$ we then have $\Delta_s \leq |\mathcal{S}_q|^2\left(\frac{\pi}{2^B}\right)^2$. The statement of the proposition follows. ∎

Proposition 2 considers beamforming in single path channels. In multipath channels, an independently phase shifting codeword can be optimized to transmit energy to multiple directions, while a beam-steering codeword always transmits to a single direction. This leads to additional power loss for beam-steering. Proposition 2 can be refined to multipath effective channels by using the path dominance ratio D. Recall that mmWave channels often have strong LoS components, which would lead to large values of D. We have

Proposition 3. *Consider using precoding codebooks with phase resolution B for the effective multipath channel $\mathbf{h}_k$ of (7) with path dominance ratio D from (8). The relative loss in power gain from using the optimal finite resolution beam-steering precoder $\hat{\mathbf{f}}_q \in \mathcal{F}_q$ fulfilling (23), as compared to using the optimal independent phase-shifting precoder $\mathbf{f}_q^\star \in \mathcal{P}_q$ found by Algorithm 1, is bounded as*

$$1 - \frac{\left|\mathbf{h}_k^{\mathrm{H}}\hat{\mathbf{f}}_q\right|^2}{\left|\mathbf{h}_k^{\mathrm{H}}\mathbf{f}_q^\star\right|^2} \quad \leq \quad 2\sqrt{\frac{1}{D}} + \frac{1}{D} + \sin^2(\nu) + \frac{\varepsilon^2}{\cos^2(\nu)} \ ,$$

where ε is given in (33) and $\nu = \pi/2^{B-1}$.

The proof can be found in the Appendix. The $\sin^2(\nu)$ term is related to analyzing loss in the proof w.r.t. Quantized MRT as opposed to the optimal precoder from Algorithm 1. This part of the bound is loose. The first terms of the bound arise from the difficulty to address beam-steering in a multipath channel. The effect of the geometric size of the subarray is captured by ε.

Propositions 2 and 3 are useful as the effective channel $\mathbf{h}_k$ typically has single dominant path (e.g., a LoS path) which is associated with the best UE beam. Correspondingly D would be large.

Optimally, the RF precoders should be chosen based on (24) for independent phase shifting codebooks $\{\mathcal{P}_q\}$, and on (23), for beam-steering codebooks $\{\mathcal{F}_q\}$. According to Proposition 3, an optimal beam-steering codeword approximates an optimal independent phase-shifting codeword when there is a dominant path, and the codebook resolution B is large as compared to the geometric size of subarrays. Similarly, when B and/or $|\mathcal{S}_q|$ is large, a simple RF-precoder selection based on quantizing the effective channel is a good selection in mmWave channels with independent phase-shifting.

For simulations we thus consider a simple low-complexity hybrid beamforming architecture, in which for beam-steering RF-codebooks, an exhaustive search is used, while for independent phase-shifting RF-codebooks, Algorithm 1, or a per antenna quantization is applied. In the digital domain, zero forcing is then used.

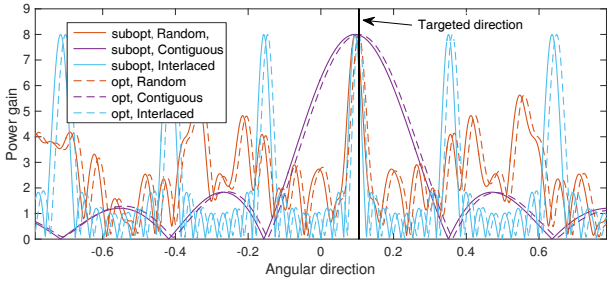


Fig. 2. Example of analog beam patterns for random, contiguous and interlaced subarrays with 6-bit codebook $\mathcal{F}_q$ (sub-optimal) and 6-bit codebook $\mathcal{P}_q$ (optimal), all subarrays have 8 antennas.

E. Effects of Subarray Geometries

Proposition 2 shows that simple beam-steering can achieve satisfying SNR performance when the BS has a moderate phase-shifter resolution compared to the geometric size $S_{\rm m}$ of the subarrays. For beam-steering with a specific phase-shifter resolution, contiguous subarrays where adjacent antennas are used achieve the best SNR performance as they have the smallest geometric size.

To understand the role of ε (33), and its relation to subarray geometry, a numeric example is in place. Consider a $\lambda/2$ ULA with a subarray consisting of $|\mathcal{S}_q| = 8$ contiguous antennas, such that $S_m = 7/2$. For $B = 3$, we have $\varepsilon = 1.43$, while for $B = 4$, we have $\varepsilon = 0.71$. This indicates that when $B \gtrsim 3$, beam-steering starts to work in a reliable manner. Conversely, if we have an interlaced subarray of size $|\mathcal{S}_q| = 8$, taken from an $N = 64$ ULA, the geometric array size is $S_m = 56/2$. In this case, ε becomes smaller than one between $B = 6$ and $B = 7$. Accoringly, we would predict that beam-steering starts to work in a reliable manner if $B \gtrsim 6$.

Fig. 2 shows the subarray power gain achieved by the optimal RF precoder in (24) and the sub-optimal one in (23) both with the quantization constraint, for three types of subarrays. In addition to contiguous subarrays, random and interlaced ones are considered. In the latter, the antennas are chosen from the main array in a regular grid [9]. The contiguous subarray has the widest main-lobe, and requires the lowest phase-shifter resolution to perform beam-steering to cover the entire angular domain. The main-lobes of random and interlaced subarrays have approximately $1/Q$ of the beam width of the contiguous subarray. As the beam width decreases, one needs a higher phase-shifter resolution to steer these narrow beams to cover the entire angular domain if a beam-steering codebook $\mathcal{F}_q$ is applied. If the phase-shifter resolution is low, using the beam-steering with narrow beams would also suffer from beam misalignment. Furthermore, when UEs are uniformly distributed in the angular domain, interlaced and random subarrays would create side-lobe interference.

IV. Performance Evaluation

In this section, we evaluate the performance of the proposed channel estimation method and MU-MIMO precoding scheme via numerical simulations. The UEs have $M = 8$ antennas and 4-bit phase-shifter resolution. The BS has a ULA with $N = 64$ antennas and $\lambda/2$ antenna separation. In hybrid architectures, there are $Q = 8$ RF chains in the base station. For the multiuser simulation, a sectorized cell with a azimuth width of 120° is considered [41].

In each simulation instance, 80 UEs are dropped randomly in the cell at a horizontal distance between 10 m and 80 m from the BS. The UEs are divided into ten groups, so that $K = 8$ UEs in each group are served at a time.

To select the parameters of the propagation paths in Channel Model (5), we use a geometry-based stochastic channel model, following [41], [42]. These models provide a controlled method to select attenuated multipath components with angles of arrival and departure, pertinent for a given mobile communication scenario. We consider a typical outdoor microcell network in the Urban Microcellular (UMi) street canyon scenario discussed in [42] and detailed in 3GPP [41]. There, the multipath components in (5) are grouped into $N_{\rm cl}$ clusters of nearby components with similar path delays and directions. and there are $L_{\rm cl}$ subpaths in each cluster. If the UE is in LoS condition, a LoS path is added. Thus, if an UE is in LoS condition, there are in total $L = N_{\rm cl}L_{\rm cl} + 1$ paths in (5), otherwise there are $L = N_{\rm cl}L_{\rm cl}$ paths with different DoA and DoD. Note that not all of these paths are necessarily separable in Rx/Tx signal processing—they are used to generate a realistic channel model. The path parameters are generated stochastically to reflect the geometry of the environment. The LoS/NLoS condition for an UE is also generated according to the LoS probability model proposed in [41]. The details of the channel model parameters are given in Table I.

Following [9] and [41], the baseband complex gain $\alpha_{l,n}$ within the nth subcarrier in (5) for the l-th path is given by

$$\alpha_{l,n} = e^{j\psi}\sqrt{\frac{P_l}{\ell(x)\kappa_{\rm SF}}g_1(\theta_l)g_2(\phi_l)} \times \frac{1}{\sqrt{N_c}}\sum_{d=0}^{D-1} p(dT_c - \tau_l)\exp\left(\frac{j2\pi nd}{N_c}\right), \tag{34}$$

where P_l represents the path power, ψ is a random phase, $g_1(\theta)$ and $g_2(\phi)$ are the UE and BS antenna patterns, $\ell(x)$ is the pathloss for a UE-to-BS distance x, T_c is the sampling interval selected as $\frac{1}{2B_{\rm w}} \approx 1.95$ ns and N_c is the number of OFDM subcarriers. In addition, $p(t)$ is the baseband pulse-shaping filter, and τ_l represents the path delay. The $p(t)$ is chosen as a root-raised-cosine filter [9] with a roll-off factor of 0.22 and an order of $D = 64$.

The multiuser precoding performance for the architecture with fixed subarray and quantized phase shifters (FS-QPS) is investigated. Its performance is compared to a fully-connected hybrid architecture with quantized phase shifters (FC-QPS) and the fully-digital (FD) architecture. The effects of channel estimation inaccuracy and phase-shifter resolution are studied.

Note that the channel model in the simulations is a wideband one—there is delay spread both within and between the path clusters, and a multicarrier OFDM-system is simulated. The RF-precoders are selected for the full band, based on covariance (22), and used for all subcarriers. The ZF baseband precoder is chosen separately for each subcarrier.

The cumulative distribution function (CDF) for the mul-

TABLE I
CHANNEL MODEL PARAMETERS

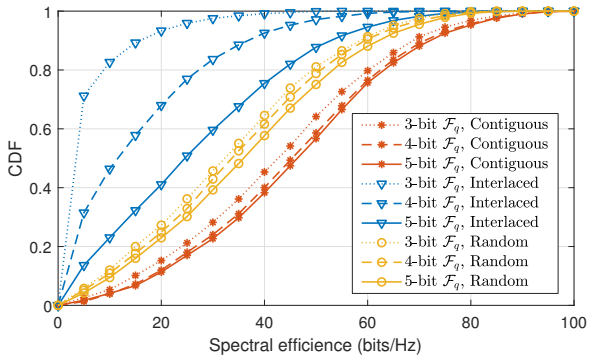
Parameter	Symbol	Value
BS hight	h_{BS}	5 m
UE hight	h_{UE}	1.5 m
Carrier frequency	f_c	28 GHz
System bandwidth	B_{w}	256 MHz
Subcarrier number	N_c	256
BS total Tx power	$N_c \times \rho_{\mathrm{BS}}$	35 dBm
UE noise power	$N_c \times \sigma^2$	-83 dBm
Pathloss	$10 \lg \ell(x)$	UMi, Table 7.4.1 [41]
LoS probability	$p_{\mathrm{LoS}}(x)$	UMi, Table 7.4.2 [41]
UE antenna pattern	$g_1(\theta)$	omni-directional
BS antenna pattern	$g_2(\phi)$	given in ITU-R M.2135
Shadowing factor (SF)	$10 \lg \kappa_{\mathrm{SF}}$	$\mathcal{N}(0, \sigma^2_{\mathrm{SF}})$, $\sigma_{\mathrm{SF}} = 2$
Number of clusters	N_{cl}	Poisson(8)
Number of subpaths	L_{cl}	20
Per cluster SF std	ζ	3 dB
Delay spread (DS)	σ_τ	given in Table 7.5-6 [41]
BS angular spread (AS)	σ_ϕ	ASD in Table 7.5-6 [41]
UE angular spread (AS)	σ_θ	ASA in Table 7.5-6 [41]
BS per cluster AS	c_ϕ	c_{ASD} in Table 7.5-6 [41]
UE per cluster AS	c_θ	c_{ASA} in Table 7.5-6 [41]
LoS K-factor	K_{r}	given in Table 7.5-6 [41]
Subpath power	P_l	defined in [41]

tiuser spectral efficiency is collected for 100 iterations. The multiuser spectral efficiency is estimated as $\sum_{k=1}^{K} \log_2(1+\gamma_k)$ from the Shannon formula with γ_k the SINR for UE k. Note that due to the power allocation principle, all simultaneously scheduled users have similar spectral efficiency, with variations only created by the inefficiency of RF-precoding.

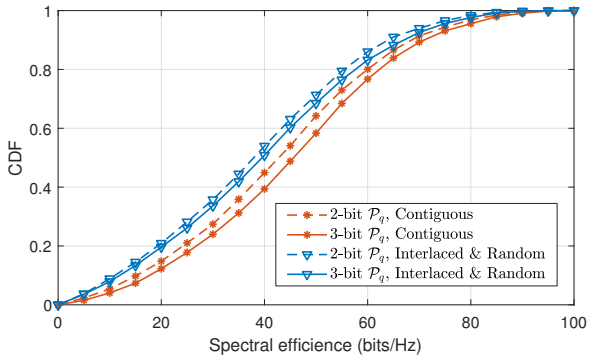
A. Performance of Different Subarray Geometries

First, we investigate the performance of FS-QPS architecture with the precoders of Algorithm 1 and (23), using different subarray geometries and different RF codebooks. The subarray size is $|\mathcal{S}_q| = N/Q = 8$. Interlaced subarrays are based on regular partitions of the ULA with $N = 64$ antennas to $Q = 8$ subarrays where there are $Q-1 = 7$ antennas between neighboring antennas in a subarray. Spectral efficiency results are given in Fig. 3. The spectral efficiency increases as the BS phase-shifter resolution B increases, especially when the beam-steering codebooks are applied. Contiguous subarrays with 3-bit, 4-bit and 5-bit $\mathcal{F}_q$ achieve similar performance, while the other two subarray geometries are more sensitive to the phase shifting resolution when using the beam-steering codebooks. This is explained by Proposition 2; as the interlaced and random subarrays have larger geometry sizes, more control bits are required for fine-grained beamforming. These results confirm the estimates in Section III-E, where the B needed for reliable operation of beam-steering was predicted based on ε.

Compared to $\mathcal{F}_q$, independently controllable phase-shifting codebooks $\mathcal{P}_q$ suffer smaller losses due to the coarse phase-shifting quantization. In fact, $\mathcal{P}_q$ with 3 control bits can provide almost the same performance as with an infinite resolution. We can also see that, for all RF codebooks, contiguous subarrays achieve the best performance.



(a) Beam-steering codebook $\mathcal{F}_q$



(b) Independently phase-shifting codebook $\mathcal{P}_q$

Fig. 3. Multiuser spectral efficiency using random, contiguous and interlaced subarrays, with different RF codebooks. Here, perfect channel estimation is assumed.

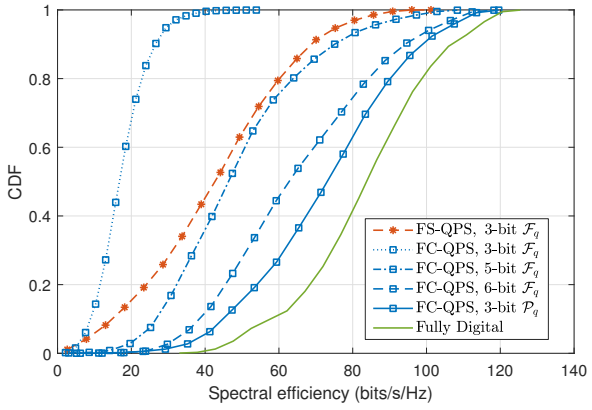


Fig. 4. Multiuser spectral efficiency performance achieved by different BS architectures with perfect CSI.

B. Performance Comparison with Fully-connected Hybrid and Fully Digital Architectures

We now compare the performance of the FS-QPS architecture with contiguous subarrays to fully-connected hybrid (FC-QPS) with different RF codebooks, and the fully-digital architectures. Perfect CSI is assumed. Experimental distributions of spectral efficiency performance are reported in Fig. 4. In Table. II, mean spectral efficiency and total codebook complexity are compared. For $\mathcal{F}_q$ codebooks, the total complexity, i.e., the total number of bits required to specify the RF-codewords for one transmission instance is QB, while for $\mathcal{P}_q$ codebooks it

TABLE II
MEAN MULTIUSER SPECTRAL EFFICIENCY AND TOTAL CODEBOOK SIZE FOR DIFFERENT ARCHITECTURES.

Architecture	RF codebook Type	B	Tot RF CB size	Spec. eff.
FS-QPS 64 phase shifters	$\mathcal{F}_q$	3	24	41
	$\mathcal{F}_q$	4	32	43
	$\mathcal{F}_q$	5	40	44
	$\mathcal{P}_q$	1	64	34
	$\mathcal{P}_q$	2	128	40
	$\mathcal{P}_q$	3	192	43
	$\mathcal{P}_q$	4	256	44
FC-QPS 512 phase shifters	$\mathcal{F}_q$	3	24	17
	$\mathcal{F}_q$	4	32	29
	$\mathcal{F}_q$	5	40	46
	$\mathcal{F}_q$	6	48	62
	$\mathcal{P}_q$	1	512	49
	$\mathcal{P}_q$	2	1024	64
	$\mathcal{P}_q$	3	1536	68
	$\mathcal{P}_q$	4	2048	69
Fully Digital			—	82

is NQB for fully connected and $|\mathcal{S}|_q QB$ for fixed subarrays. The total number of phase shifters required in the two hybrid precoding architectures is also reported.

The FS-QPS architecture with contiguous subarrays and 3-bit $\mathcal{F}_q$ can achieve 49% of the mean spectral efficiency of the fully-digital architecture, while a fully-connected hybrid architecture with 6-bit $\mathcal{F}_q$ achieves 76%. Interestingly, a fully-connected hybrid architecture with 3- and 4-bit $\mathcal{F}_q$ performs worse than a fixed subarray architecture with the same codebooks. This is explained by Proposition 2. To have sufficient beam granularity, at least a 6-bit phase-shifter resolution is required for a fully connected array with a beam-steering codebook.

With the same phase-shifter resolution, precoding with $\mathcal{P}_q$ always outperforms $\mathcal{F}_q$. However, the RF hardware complexity to realize $\mathcal{P}_q$ is much higher than $\mathcal{F}_q$, as the phase shifters in $\mathcal{P}_q$ require independent control, and accordingly the total number of states in the RF-codebooks is significantly larger.

C. Comparison with Other Algorithms

Here, we compare the algorithms discussed in this paper with multiuser hybrid precoding algorithms from the literature. We consider five typical algorithms with various complexities, which are a) spatially sparse precoding algorithm via OMP [19]; b) two-stage hybrid precoding [8] with the analog RF beam-steering codebook and a random vector quantization (RVQ) codebook in its second stage; c) hybrid precoding with quantized MRT for RF precoding and a baseband MMSE precoder adopted in [15]; d) hybrid precoding with iterative coordinate descent RF precoding and MMSE baseband precoding [47] and e) SDR based hybrid precoding for the partially-connected structure via alternating minimization (SDR-AltMin) [7]. For meaningful comparisons, a common contiguous-subarray architecture is considered for all algorithms.

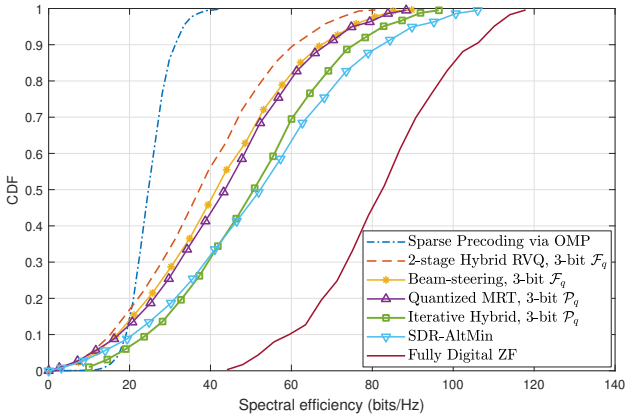


Fig. 5. Multiuser spectral efficiency achieved by different multiuser hybrid precoding algorithms. For reasonable comparison, the contiguous-subarray architecture with $N = 64$, $Q = 8$ is considered for all algorithms.

The Quantized MRT RF precoding in [15] can be directly applied in the FS-QPS architecture by setting part of the entries in the phased vector be zero. It leads essentially to the same performance as the independently controllable phase shifting studied here, subject to the slight non-optimality of quantized MRT discussed in (32). For the two-stage hybrid precoding [8], we use beam steering codebooks $\mathcal{F}_q$ for BS subarrays and $\mathcal{U}$ for UEs at its first stage, and a RVQ codebook with size of 2^{10} at its second stage. For the iterative hybrid precoding method [47], it can also be implemented in the subarray architecture. Furthermore, a quantized RF precoder can be obtained by quantizing the solution of the analog precoder elements in each iteration [10], [47].

The multiuser spectral efficiencies for the considered algorithms are given in Fig. 5. The OMP-based sparse precoding algorithm, which works well in the fully-connected multiuser hybrid architecture considered in [19], does not yield satisfying performance in the simulated mmWave system with subarrays. The two-stage hybrid precoding in [8] performs worse than the proposed beam-steering method, while they both have a computation complexity of $\mathcal{O}(Q|\mathcal{S}_q|)$ at the RF precoding design stage. The spectral efficiency and computation complexity of RF Quantized MRT with 3-bit independent phase shifters is similar to (23) with 3-bit beam-steering codebooks. The required hardware cost, however, is higher. In quantized MRT, one needs 64^8 hardware states, while in the beam-steering architecture, 8^8 states are needed. The best average performance is given by SDR-AltMin at the cost of solving a SDR problem via semidefinite programming at a high computation complexity of $\mathcal{O}((QK)^6)$ [48]. The iterative hybrid precoding scheme exhibits similar performance as SDR-AltMin, with a computation complexity of $\mathcal{O}((Q|\mathcal{S}_q|)^3)$ [47]. The hardware cost in the iterative hybrid precoding scheme is the same as in quantized MRT, while in SDR-AltMin, infinite-resolution phase shifters are required.

V. CONCLUSION

We have considered a low-complexity hybrid architecture with fixed subarrays and quantized RF phase-shifting networks

for mmWave multiuser MIMO systems. Assuming that linear zero-forcing is applied at the digital baseband, we have simplified the complicated hybrid precoder optimization problem to an eigenvalue maximization problem. An efficient method has been developed to address this problem for maximizing the multiuser SINRs.

We found that for independent phase-shifting precoding, a direct quantization of an MRT precoder has a gap to an optimal Rx-power maximizing precoder that vanishes exponentially in phase-shifting resolution B. Related to subarray geometry, we found that contiguous subarrays outperform other types of subarrays. Moreover, for finite resolution beam-steering, performance is inversely proportional to the geometric size of the subarray.

For the same resolution, beam-steering codebooks are subsets of independent phase shifting codebooks, and accordingly, independent phase shifting always outperforms beam-steering. However, for moderate size subarray architectures, which provide low RF-complexity, the gain from independent beam steering is limited.

The effectiveness of the discussed hybrid precoding design was verified by extensive numerical simulations, confirming the analytic results. The hybrid architecture with contiguous subarrays and beam-steering codebooks has a low RF and computational complexity and provides mean multiuser spectral efficiency comparable to the fully-connected hybrid architectures, making it a viable solution for mmWave MU-MIMO systems.

APPENDIX

To prove Proposition 3, we first prove a couple of lemmas. First consider the difference of the power gain of a precoder from the dominant path as compared to the full power gain. We have

Lemma 1. *Assume that the effective channel $\mathbf{h}_k$ of (7) has a dominant path with path dominance ratio D from (8), and that a precoder $\mathbf{f}_q$ is used. The difference of the power gain achieved when choosing $\mathbf{f}_q$ based on the whole channel as compared to the power gain achieved by choosing $\mathbf{f}_q$ based on the dominant path only is upper bounded by*

$$\xi < \left(2\sqrt{\frac{1}{D}} + \frac{1}{D}\right) |\beta_{l_0}|^2 \left|\mathbf{a}_{\mathrm{BS}}^{\mathrm{H}}(\phi_{l_0})\mathbf{f}_q\right|^2.$$

Proof. Denote the steering vector of a generic path by $\mathbf{a}_l = \mathbf{a}_{\mathrm{BS}}(\phi_l)$ and the dominant path by $\mathbf{a}_0 = \mathbf{a}_{\mathrm{BS}}(\phi_{l_0})$. The power gain of $\mathbf{f}_q$ in the channel $\mathbf{h}_k$ is

$$|\mathbf{h}_k^{\mathrm{H}}\mathbf{f}_q|^2 = \left|\sum_{l=1}^{L}\beta_l \mathbf{a}_l^{\mathrm{H}}\mathbf{f}_q\right|^2 = |\beta_{l_0}|^2 |\mathbf{a}_0^{\mathrm{H}}\mathbf{f}_q|^2 + \xi \qquad (35)$$

where ξ is given by

$$\xi = 2\mathrm{Re}\left\{\sum_{l\neq l_0}\beta_l \mathbf{a}_l^{\mathrm{H}}\mathbf{f}_q\mathbf{f}_q^{\mathrm{H}}\mathbf{a}_0\beta_{l_0}^*\right\} + \left|\sum_{l\neq l_0}\beta_l \mathbf{a}_l^{\mathrm{H}}\mathbf{f}_q\right|^2. \qquad (36)$$

According to the Cauchy-Schwarz inequality, the term ξ achieves its maximum when all other paths $l \neq l_0$ align with the dominant path l_0. As

$$\begin{aligned}\mathrm{Re}\sum_{l\neq l_0}|\beta_l|^2\mathbf{a}_l^{\mathrm{H}}\mathbf{f}_q\mathbf{f}_q^{\mathrm{H}}\mathbf{a}_0 &\leq \mathrm{Re}\left\{\sqrt{\frac{1}{D}}|\beta_{l_0}|^2\mathbf{a}_0^{\mathrm{H}}\mathbf{f}_q\mathbf{f}_q^{\mathrm{H}}\mathbf{a}_0\right\}\\ &\leq \sqrt{\frac{1}{D}}|\beta_{l_0}|^2\left|\mathbf{a}_0^{\mathrm{H}}\mathbf{f}_q\right|^2,\end{aligned}$$

and

$$\left|\sum_{l\neq l_0}\beta_l\mathbf{a}_l^{\mathrm{H}}\mathbf{f}_q\right|^2 \leq \sum_{l\neq l_0}|\beta_l|^2\left|\mathbf{a}_0^{\mathrm{H}}\mathbf{f}_q\right|^2 = \frac{1}{D}|\beta_{l_0}|^2\left|\mathbf{a}_0^{\mathrm{H}}\mathbf{f}_q\right|^2,$$

the statement follows. ∎

For a given D, the first term in (36) is negligible when $\{\phi_l\}_{l\neq l_0}$ are outside the beam width of $\mathbf{a}_{\mathrm{BS}}(\phi_{l_0})$. On the other hand, if $\{\phi_l\}_{l\neq l_0}$ are close to ϕ_{l_0}, maximizing $|\mathbf{a}_{\mathrm{BS}}^{\mathrm{H}}(\phi_{l_0})\mathbf{f}_q|^2$ also leads to a larger ξ. In a word, $|\mathbf{h}_k^{\mathrm{H}}\mathbf{f}_q|^2$ is increased when $|\mathbf{a}_{\mathrm{BS}}^{\mathrm{H}}(\phi_{l_0})\mathbf{f}_q|^2$ is maximized.

Lemma 1 indicates that if one uses an RF precoder constructed based on the dominant path, the relative power loss is bounded. We have

Lemma 2. *Assume that the effective channel $\mathbf{h}_k$ of (7) has a dominant path with path dominance ratio D, and that a precoder $\mathbf{f}_q^0$ selected from a codebook to maximize the power gain w.r.t. the dominant path. The relative power loss from using $\mathbf{f}_q^0$ instead of an optimum precoder $\mathbf{f}_q^\star$ from the same codebook is upper bounded by $2\sqrt{\frac{1}{D}} + \frac{1}{D}$.*

Proof. Denote the steering vector of the dominant path by $\mathbf{a}_0 = \mathbf{a}_{\mathrm{BS}}(\phi_{l_0})$. Applying (35) for $\mathbf{f}_q^0$ and $\mathbf{f}_q^\star$ separately, defining the respective quantities ξ^0 and $\xi^\star$, the relative power loss is

$$\begin{aligned}R &= \frac{|\beta_{l_0}|^2\left|\mathbf{a}_0^{\mathrm{H}}\mathbf{f}_q^\star\right|^2 + \xi^\star - |\beta_{l_0}|^2\left|\mathbf{a}_0^{\mathrm{H}}\mathbf{f}_q^0\right|^2 - \xi^0}{|\beta_{l_0}|^2\left|\mathbf{a}_0^{\mathrm{H}}\mathbf{f}_q^\star\right|^2 + \xi^\star}\\ &\leq \frac{\left|\mathbf{a}_0^{\mathrm{H}}\mathbf{f}_q^\star\right|^2 + \xi^\star/|\beta_{l_0}|^2 - \left|\mathbf{a}_0^{\mathrm{H}}\mathbf{f}_q^0\right|^2}{\left|\mathbf{a}_0^{\mathrm{H}}\mathbf{f}_q^\star\right|^2},\end{aligned}$$

where $\xi^0 \geq 0$ was used. Now $\mathbf{f}_q^0$ is the optimal precoder for the steering vector $\mathbf{a}_0$, so that $\left|\mathbf{a}_0^{\mathrm{H}}\mathbf{f}_q^\star\right|^2 \leq \left|\mathbf{a}_0^{\mathrm{H}}\mathbf{f}_q^0\right|^2$. It then follows that

$$R \leq \frac{\xi^\star}{|\beta_{l_0}|^2\left|\mathbf{a}_0^{\mathrm{H}}\mathbf{f}_q^\star\right|^2} \leq 2\sqrt{\frac{1}{D}} + \frac{1}{D}$$

where we used Lemma 1 in the second step. ∎

We then can prove Proposition 3.

Proof. The difference between the optimal beamformer $\mathbf{f}_q^\star$ and a dominant path beamformer can be found in Lemma 2. The difference between an optimal single-path precoder, and a Quantized MRT single-path precoder is given in (32), which can be loosened by comparing a single-path Quantized MRT precoder to the full channel precoder $\mathbf{f}_q^\star$. The difference between a Quantized MRT precoder and an optimal beam-steering precoder in single-path channels is bounded in Proposition 2. This is given in terms of the maximum precoding gain, which in the relative gain is normalized as $|\beta_{l_0}|^2|\mathcal{S}_q|^2/\left|\mathbf{h}_k^{\mathrm{H}}\mathbf{f}_q^\star\right|^2 \leq |\mathbf{h}_k|^2/\left|\mathbf{h}_k^{\mathrm{H}}\mathbf{f}_q^\star\right|^2$. With the same line

of argument that lead to (32), one can show that $\left|\mathbf{h}_k^{\mathrm{H}}\mathbf{f}_q^{\star}\right|^2 \geq |\mathbf{h}_k|^2\cos^2(\pi/2^{B-1})$. Observing that the beam-steering codeword optimized for the whole channel $\mathbf{h}_k$ provides larger power gain than optimizing it for the dominant beam, the statement follows. ■

REFERENCES

[1] S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter-wave cellular wireless networks: Potentials and challenges," *Proceedings of the IEEE*, vol. 102, no. 3, pp. 366–385, March 2014.

[2] T. Bai and R. W. Heath, "Coverage and rate analysis for millimeter-wave cellular networks," *IEEE Transactions on Wireless Communications*, vol. 14, no. 2, pp. 1100–1114, Feb 2015.

[3] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Transactions on Wireless Communications*, vol. 13, no. 3, pp. 1499–1513, March 2014.

[4] A. F. Molisch, V. V. Ratnam, S. Han, Z. Li, S. L. H. Nguyen, L. Li, and K. Haneda, "Hybrid beamforming for massive MIMO: A survey," *IEEE Communications Magazine*, vol. 55, no. 9, pp. 134–141, 2017.

[5] A. Alkhateeb, J. Mo, N. Gonzalez-Prelcic, and R. W. Heath, "MIMO precoding and combining solutions for millimeter-wave systems," *IEEE Communications Magazine*, vol. 52, no. 12, pp. 122–131, December 2014.

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

[7] X. Yu, J. C. Shen, J. Zhang, and K. B. Letaief, "Alternating minimization algorithms for hybrid precoding in millimeter wave MIMO systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 3, pp. 485–500, April 2016.

[8] A. Alkhateeb, G. Leus, and R. W. Heath, "Limited feedback hybrid precoding for multi-user millimeter wave systems," *IEEE Transactions on Wireless Communications*, vol. 14, no. 11, pp. 6481–6494, Nov 2015.

[9] S. Park, A. Alkhateeb, and R. W. Heath, "Dynamic subarrays for hybrid precoding in wideband mmWave MIMO systems," *IEEE Transactions on Wireless Communications*, vol. 16, no. 5, pp. 2907–2920, May 2017.

[10] F. Sohrabi and W. Yu, "Hybrid digital and analog beamforming design for large-scale antenna arrays," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 3, pp. 501–513, April 2016.

[11] X. Gao, L. Dai, S. Han, C. I, and R. W. Heath, "Energy-efficient hybrid analog and digital precoding for mmWave MIMO systems with large antenna arrays," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 4, pp. 998–1009, April 2016.

[12] J. Zhang, Y. Huang, T. Yu, J. Wang, and M. Xiao, "Hybrid precoding for multi-subarray millimeter-wave communication systems," *IEEE Wireless Communications Letters*, vol. 7, no. 3, pp. 440–443, June 2018.

[13] Y. Zeng and R. Zhang, "Millimeter wave MIMO with lens antenna array: A new path division multiplexing paradigm," *IEEE Transactions on Communications*, vol. 64, no. 4, pp. 1557–1571, April 2016.

[14] W. Huang, Y. Huang, Y. Zeng, and L. Yang, "Wideband millimeter wave communication with lens antenna array: Joint beamforming and antenna selection with group sparse optimization," *IEEE Transactions on Wireless Communications*, vol. 17, no. 10, pp. 6575–6589, Oct 2018.

[15] L. Liang, W. Xu, and X. Dong, "Low-complexity hybrid precoding in massive multiuser MIMO systems," *IEEE Wireless Communications Letters*, vol. 3, no. 6, pp. 653–656, Dec 2014.

[16] J. Chen, "Hybrid beamforming with discrete phase shifters for millimeter-wave massive MIMO systems," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 8, pp. 7604–7608, Aug 2017.

[17] W. Tan, M. Matthaiou, S. Jin, and X. Li, "Spectral efficiency of DFT-based processing hybrid architectures in massive MIMO," *IEEE Wireless Communications Letters*, vol. 6, no. 5, pp. 586–589, Oct 2017.

[18] W. Tan, D. Xie, J. Xia, W. Tan, L. Fan, and S. Jin, "Spectral and energy efficiency of massive mimo for hybrid architectures based on phase shifters," *IEEE Access*, vol. 6, pp. 11 751–11 759, 2018.

[19] D. H. N. Nguyen, L. B. Le, and T. Le-Ngoc, "Hybrid MMSE precoding for mmWave multiuser MIMO systems," in *2016 IEEE International Conference on Communications (ICC)*, May 2016, pp. 1–6.

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

[21] J. D. Krieger, C. P. Yeang, and G. W. Wornell, "Dense Delta-Sigma phased arrays," *IEEE Transactions on Antennas and Propagation*, vol. 61, no. 4, pp. 1825–1837, April 2013.

[22] D. Liu, U. Pfeiffer, J. Grzyb, and B. Gaucher, *Advanced millimeter-wave technologies: antennas, packaging and circuits*. John Wiley & Sons, 2009.

[23] J. S. Park, T. Chi, and H. Wang, "An ultra-broadband compact mm-wave Butler matrix in CMOS for array-based MIMO systems," in *Proceedings of the IEEE 2013 Custom Integrated Circuits Conference*, Sept 2013, pp. 1–4.

[24] C. Chang, R. Lee, and T. Shih, "Design of a beam switching/steering Butler matrix for phased array system," *IEEE Transactions on Antennas and Propagation*, vol. 58, no. 2, pp. 367–374, Feb 2010.

[25] S. He, J. Wang, Y. Huang, B. Ottersten, and W. Hong, "Codebook-based hybrid precoding for millimeter wave multiuser systems," *IEEE Transactions on Signal Processing*, vol. 65, no. 20, pp. 5289–5304, Oct 2017.

[26] Y. Lin, "On the quantization of phase shifters for hybrid precoding systems," *IEEE Transactions on Signal Processing*, vol. 65, no. 9, pp. 2237–2246, May 2017.

[27] J. Hämäläinen and R. Wichman, "Closed-loop transmit diversity for FDD WCDMA systems," in *Asilomar Conference on Signals, Systems and Computers*, Oct. 2000, pp. 111–115.

[28] J. Li, L. Xiao, X. Xu, X. Su, and S. Zhou, "Energy-efficient Butler-matrix-based hybrid beamforming for multiuser mmWave MIMO system," *Science China Information Sciences*, vol. 60, no. 8, p. 080304, May 2017.

[29] W. Huang, Z. Lu, Y. Huang, and L. Yang, "Hybrid precoding for single carrier wideband multi-subarray millimeter wave systems," *IEEE Wireless Communications Letters*, vol. 8, no. 2, pp. 484–487, April 2019.

[30] G. Caire, N. Jindal, M. Kobayashi, and N. Ravindran, "Multiuser MIMO achievable rates with downlink training and channel state feedback," *IEEE Transactions on Information Theory*, vol. 56, no. 6, pp. 2845–2866, June 2010.

[31] R. D. Wesel and J. M. Cioffi, "Achievable rates for Tomlinson-Harashima precoding," *IEEE Transactions on Information Theory*, vol. 44, no. 2, pp. 824–831, March 1998.

[32] K. Zu, R. C. de Lamare, and M. Haardt, "Multi-branch Tomlinson-Harashima precoding design for MU-MIMO systems: Theory and algorithms," *IEEE Transactions on Communications*, vol. 62, no. 3, pp. 939–951, March 2014.

[33] Z. Shen, R. Chen, J. G. Andrews, R. W. Heath, and B. L. Evans, "Low complexity user selection algorithms for multiuser MIMO systems with block diagonalization," *IEEE Transactions on Signal Processing*, vol. 54, no. 9, pp. 3658–3663, Sept 2006.

[34] H. Yang and T. L. Marzetta, "Performance of conjugate and zero-forcing beamforming in large-scale antenna systems," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 2, pp. 172–179, February 2013.

[35] J. Lee and N. Jindal, "Dirty paper coding vs. linear precoding for MIMO broadcast channels," in *2006 Fortieth Asilomar Conference on Signals, Systems and Computers*, Oct 2006, pp. 779–783.

[36] E. Björnson, M. Bengtsson, and B. Ottersten, "Optimal multiuser transmit beamforming: A difficult problem with a simple solution structure," *IEEE Signal Processing Magazine*, vol. 31, no. 4, pp. 142–148, July 2014.

[37] A. Alkhateeb and R. W. Heath, "Frequency selective hybrid precoding for limited feedback millimeter wave systems," *IEEE Transactions on Communications*, vol. 64, no. 5, pp. 1801–1818, May 2016.

[38] M. Iwanow, N. Vucic, M. H. Castaneda, J. Luo, W. Xu, and W. Utschick, "Some aspects on hybrid wideband transceiver design for mmWave communication systems," in *20th International ITG Workshop on Smart Antennas (WSA)*, March 2016, pp. 1–8.

[39] E. Bjőrnson, E. G. Larsson, and T. L. Marzetta, "Massive MIMO: ten myths and one critical question," *IEEE Communications Magazine*, vol. 54, no. 2, pp. 114–123, February 2016.

[40] S. Hur, S. Baek, B. Kim, J. Park, A. F. Molisch, K. Haneda, and M. Peter, "28 GHz channel modeling using 3D ray-tracing in urban environments," in *2015 9th European Conference on Antennas and Propagation (EuCAP)*, May 2015, pp. 1–5.

[41] 3GPP, "Study on channel model for frequency spectrum above 6 GHz," 3rd Generation Partnership Project (3GPP), Technical Report (TR) 38.900, June 2017, version 14.3.1.

[42] M. Peter, "Measurement Results and Final mmMAGIC Channel Models," mmMAGIC D2.2, Technical Report, May 2017.

[43] M. K. Samimi and T. S. Rappaport, "3-D millimeter-wave statistical channel model for 5G wireless system design," *IEEE Transactions on*

Microwave Theory and Techniques, vol. 64, no. 7, pp. 2207–2225, July 2016.

[44] A. Adhikary, E. A. Safadi, M. K. Samimi, R. Wang, G. Caire, T. S. Rappaport, and A. F. Molisch, "Joint spatial division and multiplexing for mm-Wave channels," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1239–1255, June 2014.

[45] H. Karaa, R. S. Adve, and A. J. Tenenbaum, "Linear precoding for multiuser MIMO-OFDM systems," in *IEEE International Conference on Communications*, June 2007, pp. 2797–2802.

[46] R. A. Horn, R. A. Horn, and C. R. Johnson, *Matrix analysis*. Cambridge university press, 1990.

[47] F. Sohrabi and W. Yu, "Hybrid analog and digital beamforming for mmwave ofdm large-scale antenna arrays," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 7, pp. 1432–1443, July 2017.

[48] K.-C. Toh, M. J. Todd, and R. H. Tütüncü, "Sdpt3 - a matlab software package for semidefinite programming, version 1.3," *Optimization methods and software*, vol. 11, no. 1-4, pp. 545–581, 1999.

Junquan Deng received his Ph.D. degree in information theory from Aalto university, Finland, in 2018, B.Eng. degree in automation engineering from Tsinghua university, Beijing, China, in 2011 and M.Sc. degree in computer science from National University of Defense Technology (NUDT), Changsha, China, in 2013. Since 2019, he is an assistant research fellow at the sixty-third Research Institute, National University of Defence Technology, Nanjing, China. His research interests include device-to-device communication, millimeter-wave communication, mobile relaying in 5G cellular networks, and machine learning with wireless network data.

Olav Tirkkonen is an Associate Professor in communication theory at the Department of Communications and Networking in Aalto University, Finland, where he has held a faculty position since August 2006. He received his M.Sc. and Ph.D. degrees in theoretical physics from Helsinki University of Technology in 1990 and 1994, respectively. Between 1994 and 1999 he held post-doctoral positions at the University of British Columbia, Vancouver, Canada, and the Nordic Institute for Theoretical Physics, Copenhagen, Denmark. From 1999 to 2010 he was with Nokia Research Center (NRC), Helsinki, Finland. In 2016-2017 he was Visiting Associate Professor at Cornell University, Ithaca, NY, USA. He has published some 200 papers, and is coauthor of the book *Multiantenna transceiver techniques for 3G and beyond*. His current research interests are in coding theory, multiantenna techniques, and cognitive management of 5G cellular networks.

Christoph Studer received his Ph.D. degree in Information Technology and Electrical Engineering from ETH Zurich in 2009. In 2005, he was a Visiting Researcher with the Smart Antennas Research Group at Stanford University. From 2006 to 2009, he was a Research Assistant in both the Integrated Systems Laboratory and the Communication Technology Laboratory (CTL) at ETH Zurich. From 2009 to 2012, Dr. Studer was a Postdoctoral Researcher at CTL, ETH Zurich, and the Digital Signal Processing Group at Rice University. In 2013, he has held the position of Research Scientist at Rice University. Since 2014, Dr. Studer is an Assistant Professor at Cornell University and an Adjunct Assistant Professor at Rice University. Dr. Studer's research interests include the design of very large-scale integration (VLSI) circuits, as well as wireless communications, signal and image processing, and convex optimization.

Dr. Studer received ETH Medals for his M.S. and Ph.D. theses in 2006 and 2009, respectively. He received a Swiss National Science Foundation fellowship for Advanced Researchers in 2011 and a US National Science Foundation CAREER Award in 2017. Dr. Studer won a Michael Tien Excellence in Teaching Award from the College of Engineering, Cornell University, in 2016. He shared the Swisscom/ICTnet Innovations Award in both 2010 and 2013. Dr. Studer was the winner of the Student Paper Contest of the 2007 Asilomar Conf. on Signals, Systems, and Computers, received a Best Student Paper Award of the 2008 IEEE Int. Symp. on Circuits and Systems (ISCAS), and shared the best Live Demonstration Award at the IEEE ISCAS in 2013.