Millimeter Wave MIMO-OFDM With Index Modulation: A Pareto Paradigm on Spectral-Energy Efficiency Trade-Off

Yan Yang, *Member, IEEE*, Shuping Dang, *Member, IEEE*, Miaowen Wen, *Senior Member, IEEE*, and Mohsen Guizani, *Fellow, IEEE*

***Abstract*— Multiple-input multiple-output orthogonal frequency division multiplexing with index modulation (MIMO-OFDM-IM) has recently received increased attention, due to the potential advantage to balance the trade-off between spectral efficiency (SE) and energy efficiency (EE). In this paper, we investigate the application of MIMO-OFDM-IM to millimeter wave (mmWave) communication systems, where a hybrid analogy-digital (HAD) beamforming architecture is employed. Taking advantage of the Pareto-optimal beam design, we propose a feasible solution to approximately achieve a globally Pareto-optimal trade-off between SE and EE, and the collision constraints of the multi-objective optimization problem (MOP) can be solved efficiently. Correspondingly, the MOP of SE-EE trade-off can be converted into a feasible solution for energy-efficient resource usage, by finding the Pareto-optimal set (POS) towards the Pareto front. This combinatorial-oriented resource allocation approach on the SE-EE relation considers the optimal beam design and power control strategies for downlink multi-user mmWave transmission. To ease the system performance evaluation, we adopt the Poisson point process (PPP) to model the mobile data traffic, and the evolutionary algorithm is applied to speed up the search efficiency of the Pareto front. Compared with benchmarks, the experimental results collected from extensive simulations demonstrate that the proposed optimization approach is vastly superior to existing algorithms.**

***Index Terms*— MIMO-OFDM, index modulation, spectral efficiency, energy efficiency, Pareto-optimal set, mmWave communication.**

Manuscript received October 21, 2020; revised February 2, 2021 and April 12, 2021; accepted April 12, 2021. Date of publication April 26, 2021; date of current version October 11, 2021. This work was supported in part by the State Key Laboratory of Rail Traffic Control and Safety under Contract RCS2020ZT012, in part by Beijing Jiaotong University, and in part by the Key Research Task of China Railway Corporation under Contract N2019G028. This article was presented in part at the IEEE GLOBECOM 2020, Taibei, Taiwan, December 2020. The associate editor coordinating the review of this article and approving it for publication was A. S. Cacciapuoti. *(Corresponding author: Yan Yang.)*

Yan Yang is with the State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing 100044, China (e-mail: yyang@bjtu.edu.cn).

Shuping Dang is with the Computer, Electrical and Mathematical Science and Engineering Division, King Abdullah University of Science and Technology (KAUST), Thuwal 23955, Saudi Arabia (e-mail: shuping.dang@kaust.edu.sa).

Miaowen Wen is with the School of Electronic and Information Engineering, South China University of Technology, Guangzhou 510640, China (e-mail: eemwwen@scut.edu.cn).

Mohsen Guizani is with the Department of Computer Science and Engineering, Qatar University, Doha 2713, Qatar (e-mail: mguizani@ieee.org).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TWC.2021.3073692.

Digital Object Identifier 10.1109/TWC.2021.3073692

I. Introduction

MULTIPLE-INPUT multiple-output (MIMO) millimeter wave (mmWave) communication has aroused great expectations on enabling unprecedented high-rate transmission up to multi-gigabit for future wireless communication networks, since a wide chunk of available bandwidth in mmWave frequency band could be authorized to use [2]–[4]. In practical implementation, the wireless data transmission in the mmWave band will incur high path loss and thereby severe signal attenuation. It inevitably leads to a shorter transmission range (roughly hundreds of meters), and necessitates light-of-sight (LoS) communication links. Until now, the key to get around this issue has been largely dependent on the use of massive-antenna architectures at base stations (BS) and the exploitation of highly directional beamforming at both transmitter and receiver [5]–[7]. Additionally, the extremely small wavelengths at mmWave frequencies also enable massive antenna arrays to integrate a larger number of antenna elements, resulting in a sufficient spatial gain to combat fading. To reduce energy consumption and system design complexity, a two-stage hybrid analogy-digital (HAD) beamforming architecture with reduced RF chains has emerged. As a scalable and economically efficient technique, HAD consists of analog beamformers in the RF and digital beamformers in the baseband [4], [8]. Mobile mmWave communication systems will likely operate over wideband channels with frequency selectivity. In this case, the orthogonal frequency division multiplexing (OFDM) modulation can be employed to effectively combat the channel's frequency selectivity, as well as to provide further improvement in spectral efficiency (SE) [9].

As a matter of fact, the above-mentioned sophisticated technologies that require a large number of phase shifters usually consume a large amount of power, which becomes an obvious drawback, especially for non-chargeable devices. In addition to a large amount of power consumption for transmission, RF chains also contain some of the most energy-hungry components in a transmission system, e.g., digital-to-analog converters (DACs), amplifiers, and frequency synthesizers. These components substantially increase the circuit power dissipation of the BS. In this context, many studies discussed circuits design challenges in implementing energy-efficient multi-antenna architectures [10]–[15]. Meanwhile, it has been demonstrated that the energy costs

represent a significant portion of the total energy consumption of a network. Seriously, the radio network itself could be the most energy-consuming part, occupying ca. 80% of an operator's entire energy consumption. This results in major economic and technical challenges [16], [17]. Due to these facts, wireless operators resort to green wireless networks, where energy efficiency (EE) and SE are the main performance metrics for reducing the prohibitive cost and energy consumption. Unfortunately, according to the Shannon-Hartley theorem, conflicts of objects are usually difficult to balance while optimizing both SE and EE simultaneously. For mmWave MIMO systems with HAD beamforming, spectral- and energy-efficient system-level design is still an imminent challenge. It is mainly because the power consumption is very high owing to a large number of radiating elements, ultra-dense BS sites, and heavy data traffic load, etc., [18].

It should be emphasized that a large number of the existing investigations on the general SE-EE relation have been comprehensively conducted. They provided good insights into the joint SE-EE trade-off for different scenarios, e.g., single/multiple cell deployment etc [18]–[22]. To jointly solve the multi-objective optimization problem (MOP) in the wideband regime, there have been some works focusing on energy-efficient resource allocation/scheduling with guaranteed quality of service (QoS) [19], [23], and optimal resource allocation policy [19], [24], [25]. For example, a complete analysis of the SE and EE of two hybrid structures was provided in [13], [26]. The relationship between SE and EE with partially-connected HAD architecture was examined for optimal trade-off in [26]–[28]. The work in [27] proposed a successive interference cancelation (SIC)-based HAD beamforming for mmWave MIMO systems. The authors of [29] formulated a decoupled two-stage HAD design to maximize the SE and EE in a mmWave massive MIMO system. In [30], Riberio *et al*. investigated the EE of quantized hybrid transmitters and proved that the topology of phase-shifting components can offer a better SE-EE trade-off. Similarly, the authors of [31] studied the trade-off between SE and EE in consideration of the impact of nonlinear power amplifiers. Furthermore, the extensive investigations in [32]–[34] showed that the configurable hybrid precoding and energy-efficient beam designs are capable of effectively improving the SE and EE, respectively.

In an effort to relax the paradox in the SE-EE trade-off, an alternative way is to decompose the MOP into a number of subproblems and optimize them simultaneously. The prospective study on multi-objective signal processing [35], [36], revealed some facts, such as the respective scalarized problems, the resource optimization and allocation, as well as algorithmic tools in related fields. The authors of [37] highlighted the fact that the multi-component Pareto-optimization will gradually become the norm. It differs from simply minimizing a single metric of the system, such as the bit error rate (BER), the power consumption or the complexity. In [38], Di Renzo *et al.* derived an explicit analytical formulation of the SE-EE Pareto front to solve a bi-objective optimization problem, and proved that the Pareto front is constituted by a subset of the SE-EE trade-off. In brief, the use of Pareto property has recently emerged as an attractive solution, showing a connection of an allocation state of resources with Pareto-optimal transmission design [39].

As a novel digital modulation scheme with high SE and EE, index modulation uses the indices of the building blocks of the communication system to implicitly convey additional information bits. These approaches thereby create completely new dimensions for data transmission. Inspired by the concept of subcarrier index modulation (SIM) in [40], [41], OFDM with index modulation (OFDM-IM) has been regarded as a possible candidate for next-generation wireless networks. More specifically, the extensions of OFDM-IM in various formats have been regarded as appealing modulation candidates for mmWave communications and MIMO-OFDM systems [42]–[44]. Among different IM schemes, MIMO-OFDM with index modulation (MIMO-OFDM-IM) provides a beneficial transmission paradigm [45]. The study demonstrated that MIMO-OFDM-IM can offer significantly improved transmission rates for practical systems, as well as a better error performance than conventional MIMO-OFDM. In the MIMO-OFDM-IM scheme, each parallel stream of information is modulated by both subcarrier indices and M-ary constellation symbols. Therefore, it has the potential to provide a flexible trade-off between SE and EE [45]–[47]. For a typical MIMO-OFDM mmWave system, it is worth noting that with the extremely increasing of bandwidth and frequency at mmWave frequencies, the escalating energy consumption necessitates a high EE as well as a desirable SE. In this context, MIMO-OFDM-IM has the potential to satisfy the above requirements.

Motivated by these facts, we propose an SE-EE maximization IM scheme for multi-user mmWave MIMO-OFDM systems. Pareto-optimal beam design is taken into account with respect to the energy-efficient resource allocation in beamspace. Because the total energy consumption of cellular system is dominated by the BS, we focus on the SE-EE trade-off in downlink. The main contributions of this paper can be summarized as follows:

- We propose a MIMO-OFDM-IM scheme for HAD beamforming mmWave systems, and a maximum likelihood (ML) detector is employed to decode the information bits from each subblock of MIMO-OFDM-IM. To the best of our knowledge, this is the first work that integrates the concept of IM into mmWave MIMO-OFDM communication systems. Meanwhile, we investigate the energy-efficient aspects on designing the HAD precoder and combiner. Interestingly, the proposed scheme integrating MIMO-OFDM-IM can improve the SE-EE and transmission reliability with low complexity. It has the potential to extend the coverage without capacity penalty. This, collaborating with the HAD beamforming architecture, allows more degrees of freedom to achieve realistic SE-EE maximization in mmWave cellular networks.
- From the perspective of Pareto principle, we propose a Pareto-optimal beam design scheme for energy-efficient resource usage for downlink mmWave transmissions. By the construction of the Pareto-optimal set (POS), we propose a feasible combinatorial-oriented

power control strategy, i.e., resource reallocation scheme, to approximately achieve a Pareto-optimal trade-off between SE and EE. We give a fundamental guideline to tackle the MOP, where Pareto front is constituted by a subset of the SE-EE trade-off.

- In this new paradigm, our approach for solving the SE-EE trade-off is to convert the MOP into an evolutionary search process of POS. We show that there exists a globally optimal solution that maximizes EE, while still maintaining an increased SE. Moreover, we show that the combinational-oriented transmit power control strategy is effective to balance the total transmit power, and the globally optimal solution to the SE-EE maximization can be achieved. To systematically evaluate the performance of multiuser networks, we introduce a Poisson point process PPP) to model the spatial distribution of users, and the evolutionary algorithm is applied to speed up the search process for approaching the Pareto front asymptotically. The solving process of POSs associated with all users is abstracted as an evolutionary population-based MOP, which is potentially capable of applying to other multiuser communication systems.

The rest of the paper is organized as follows. In Section II, an mmWave MIMO-OFDM-IM system with HAD beamforming architecture is presented, as well as a preliminary analysis of SE and EE. In Sections III, the technical details of the proposed Pareto paradigm on the SE-EE trade-off are described, and a Pareto-optimal beam design strategy is mathematically formulated. Section IV develops a system-level evaluation technique based on the quantifiable PPP. Numerical results are presented and discussed in Section V, and finally conclusions are drawn in Section VI.

We use the following notations throughout this paper: bold upper and lower case letters denote matrices and vectors, respectively; $(\cdot)^*$, $(\cdot)^T$, and $(\cdot)^H$ denote complex conjugation, transpose, and Hermitian transpose, respectively. The field of complex numbers is denoted by $\mathbb{C}$, and an m by n dimensional complex space is correspondingly represented by $\mathbb{C}^{m\times n}$.

II. System Model

A. Transmission Model of mmWave MIMO-OFDM-IM Systems

The transceiver block diagram of MIMO-OFDM-IM for the multi-user mmWave system is illustrated in Fig. 1, where the conventional configuration of HAD beamforming architecture is adopted. In this paper, we focus on the downlink multi-user transmission and consider a single cell MIMO-OFDM network. A BS with N_t transmit antennas and M_t RF chains serves K active users, each of them using $N_{r,k}$ receive antennas and $M_{r,k}$ RF chains, where $k \in \{1,\ldots,K\}$. For any user k, we assume that the BS transmits $J_k \le N_{r,k}$ data streams with $M_{t,k}$ RF chains and $N_{t,k}$ transmit antennas ($\sum_{k=1}^{K} J_k \le M_t \le N_t$). The HAD beamforming architecture of BS is constructed by the concatenation of a digital precoder associated with the nth subcarrier $\mathbf{U}_{n,k} \in \mathbb{C}^{M_{t,k}\times N_{t,k}}$ and an analog beamformer $\widetilde{\mathbf{U}}_k \in \mathbb{C}^{M_{t,k}\times J_k}$ with a fully-connected structure. For simplicity, we suppose that all users have the same structure, i.e., an RF combiner $\widetilde{\mathbf{V}}_k \in \mathbb{C}^{J_k\times M_{r,k}}$ followed by a mapped digital baseband combiner $\mathbf{V}_{n,k} \in \mathbb{C}^{N_{r,k}\times M_{r,k}}$.

For such a system incorporating the OFDM-IM transceiver, a total of $T_k J_k$ incoming bits from the input alphabet are first split into J_k parallel streams. Each T_k-bit stream to the digital precoder $\mathbf{U}_{n,k}$ is pre-processed in each branch of the transmitter by the OFDM-IM modulator. Afterwards, the BS applies the baseband digital precoder $\mathbf{U}_k$ to modify the obtained OFDM-IM data blocks. Typically, the inverse fast Fourier transform (IFFT) is applied to derive a time-domain signal, and the cyclic prefix (CP) is appended to prevent the OFDM symbol from inter-symbol interference. At the receiver, the mobile station (MS) performs an FFT of the time-domain received signal and removes the CP. At the end, MS applies the digital combiner $\mathbf{V}_{n,k}$, and the received signal can be separated and demodulated by the ML or minimum mean square error (MMSE) detector. In our work, the effect of CP on SE and EE could be regarded as a stable impact factor, since the length of CP is conservatively chosen and fixed in most current standards.

Within the OFDM-IM modulator, the incoming T_k-bit stream is equally divided into G groups, in which $p = T_k/G$ bits for each group are split into two subgroups, i.e., the index selection and M-ary modulation subgroups. For a feasible frequency bandwidth B_T with N_{total} OFDM subcarriers, we assume that N consecutive OFDM subcarriers are assigned for each given subblock g. For each subblock g, considering the number of active subcarriers $N_{A,k}$, the corresponding index selection subgroup contains $p_1 = \lfloor \log_2 C(N, N_{A,k}) \rfloor$ bits for determining the indices of the active subcarriers, where $g \in \{1,\ldots,G\}$; The M-ary modulation subgroup contains the remaining $p_2 = N_{A,k}\log_2 M$ bits, which are mapped onto a predefined M-ary quadrature amplitude modulation (M-QAM) signal constellation $\mathcal{S}$ to obtain the first-stage modulation subblock from the OFDM-IM subblock creator. For simplicity, we denote $\mathfrak{N}_k$ as the *index pattern* of user k. Thereafter, each modulated subblock is fed into a concatenated OFDM block creator to construct the main OFDM-IM blocks [45]. For the output of each OFDM-IM block, the kth user's data stream vector $\mathbf{x}_{j_k} = [x_{j_k}(1)\ \ldots\ x_{j_k}(N)]^T$ is independently modified by the digital baseband precoder $\mathbf{U}_k \in \mathbb{C}^{N_{t,k}\times N_{r,k}}$ for subsequent IFFT operation, which is the same as in a conventional OFDM transmitter [48].

In our work, a total of N_{total} OFDM subcarriers are split into G groups, each of which consists of N subcarriers to transmit M-ary PSK/QAM modulated symbols. Generally, an index selection procedure can be performed by the use of reference lookup tables for simple cases with a small number of subcarriers. Alternatively, the combinatorial number theory can be employed to design an efficient strategy for selecting a subset of active subcarriers from a large number of subcarriers. The unselected subcarriers remain inactive and silent throughout the entire transmission phase [45], [49]. At the input of the OFDM index modulators, the modulated symbol can be expressed as a $T_k \times J_k$ matrix

$$\widetilde{\mathbf{X}}_k = [\widetilde{\mathbf{X}}_1\ \widetilde{\mathbf{X}}_2\ \ldots\ \widetilde{\mathbf{X}}_{J_k}], \tag{1}$$

where $\widetilde{\mathbf{X}}_{j_k} = [\widetilde{x}_1, \widetilde{x}_2, \ldots, \widetilde{x}_{T_k}]^T$ is the vector of the incoming data bits at the j_kth input of the index select and the M-ary

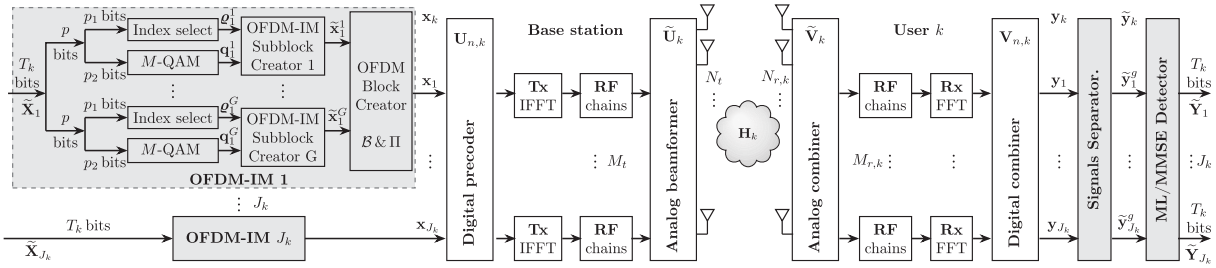


Fig. 1. Block diagram of the mmWave MIMO-OFDM-IM transceiver that employs the HAD beamforming architecture.

modulator. $N_{A,k}$ active subcarriers are selected by the index selector, and the indices of the active subcarriers can be denoted as

$$\boldsymbol{\varrho}_{j_k}^{g} = [\varrho_{j_k}^{g}(1)\ \varrho_{j_k}^{g}(2)\ \ldots\ \varrho_{j_k}^{g}(N_{A,k})]^T, \quad g = 1, 2, \ldots, G. \tag{2}$$

Meanwhile, M-QAM signal constellation with element $q_{j_k}^{g}(n_A) \in \mathcal{S}$ can be obtained:

$$\mathbf{q}_{j_k}^{g} = [q_{j_k}^{g}(1)\ q_{j_k}^{g}(2)\ \ldots\ q_{j_k}^{g}(N_{A,k})]^T, \quad g = 1, 2, \ldots, G. \tag{3}$$

For the j_kth transmit data steam, the gth OFDM-IM subblock element with $p = p_1 + p_2$ bits, is used to form OFDM-IM subblocks:

$$\widetilde{\mathbf{x}}_{j_k}^{g} = [\widetilde{x}_{j_k}^{g}(1)\ \widetilde{x}_{j_k}^{g}(2)\ \ldots\ \widetilde{x}_{j_k}^{g}(N)]^T, \quad g = 1, 2, \ldots, G. \tag{4}$$

With the OFDM block creator $\mathcal{B}$ followed by $J_k \times N$ block interleavers Π for traditional IFFT operation, the transmitted IFFT components for users are given by

$$\mathbf{x}_k = [\overline{\mathbf{x}}_1\, \overline{\mathbf{x}}_2\ \ldots\ \overline{\mathbf{x}}_N]^T,\ \overline{\mathbf{x}}_n \in \{0, \mathcal{S}\}, \tag{5}$$

where $\overline{\mathbf{x}}_n = [x_1, \ldots, x_{J_k}]$, $n = 1, 2, \ldots, N$ is the unit-power transmitted OFDM symbol at the j_k antenna for the kth user. Suppose $P_{t,k}$ is the transmitter's total power. Utilizing the power reallocation policy proposed in [39], the power allocated to each active subcarrier is $P_{t,k}/(N - N_{A,k})$.

With OFDM block creator, each OFDM-IM block is created by concatenating G OFDM-IM subblocks in each branch of the transmitter. To exemplify the mapping of the block creator $\mathcal{B}$, a simple case for $N = 4$ and $N_{A,k} = 2$ taking the following mapping is demonstrated [45]:

$$\mathcal{B} : (\mathbf{p}, \boldsymbol{\varrho}) \rightarrow \widetilde{\mathbf{x}}_{j_k}^{g}\ \big|_{N=4,\ N_{A,k}=2}\ ,\ \underbrace{\begin{bmatrix} 0 & 0 \\ 0 & 1 \\ 1 & 0 \\ 1 & 1 \end{bmatrix}}_{p_1 = 2 \text{ bits}}\ ,\ \underbrace{\begin{bmatrix} 1 & 3 \\ 2 & 4 \\ 1 & 4 \\ 2 & 3 \end{bmatrix}}_{\text{Indices } (\boldsymbol{\varrho}_{j_k}^{g})^T}$$
$$\rightarrow \underbrace{\begin{bmatrix} q_{j_k}^{g}(1) & 0 & q_{j_k}^{g}(2) & 0 \\ 0 & q_{j_k}^{g}(1) & 0 & q_{j_k}^{g}(2) \\ q_{j_k}^{g}(1) & 0 & 0 & q_{j_k}^{g}(2) \\ 0 & q_{j_k}^{g}(1) & q_{j_k}^{g}(2) & 0 \end{bmatrix}}_{\text{OFDM-IM subblocks } (\widetilde{\mathbf{x}}_{j_k}^{g})^T}, \tag{6}$$

where $p_1 = 2$ bits can be used to determine the indices of the two active subcarriers out of four available subcarriers; $N \in \mathfrak{N}_k$ is the size of FFT.

To proceed more explicitly, we temporarily put aside the discussions on the HAD issue in this subsection. At the receiver side, we also suppose that the receiver removes the CP of the received signal and performs FFT operations for the subsequent signal separation and detection. With conventional MIMO configurations, each transmit antenna transmits its own OFDM-IM frame to boost into a subframe.

Assume that the channel matrix for each subcarrier is available at the receiver. Generally, a near-ML joint detector of the MIMO-OFDM-IM scheme can be used to achieve near-optimal error performance and computationally efficient detection (based on the results shown in [50]). For an arbitrary OFDM-IM subblock from different transmit antennas, a straightforward solution to the ML detection of MIMO-OFDM-IM can be represented by

$$\hat{\widetilde{\mathbf{x}}}_{j_k}^{g} = \arg\max_{\widetilde{\mathbf{x}}_{j_k}^{g}} \left\| \widetilde{\mathbf{y}}_{j_k}^{g} - \sqrt{\frac{\gamma_k}{N_{t,k}}} \sqrt{\frac{N}{N_{A,k}}} \mathbf{A}_{j_k} \widetilde{\mathbf{x}}_{j_k}^{g} \right\|^2, \tag{7}$$

where $\mathbf{A}_{j_k}$ is an $N_{t,k} \times 1$ channel gain vector with zero-mean and identity covariance matrix, and γ_k denotes the signal-to-noise-ratio (SNR) per receive antenna.

Let $\widetilde{\mathbf{Y}}_{j_k} = \{\widetilde{y}_1, \widetilde{y}_2, \ldots, \widetilde{y}_{j_k}\}$ and $p(\widetilde{\mathbf{X}}_{j_k}, \widetilde{\mathbf{Y}}_{j_k})$ be the output and the transition probability, respectively. Considering a discrete memoryless channel, the mutual information between $\mathbf{X}_{j_k}$ and $\mathbf{Y}_{j_k}$ at each transmission epoch is given by

$$\mathcal{I}(\widetilde{\mathbf{X}}_{j_k}; \widetilde{\mathbf{Y}}_{j_k}) = \sum_{\widetilde{x}_{j_k} \in \widetilde{\mathbf{X}}_{j_k}} \sum_{\widetilde{y}_{j_k} \in \widetilde{Y}_{j_k}} p(\widetilde{x}_{j_k}, \widetilde{y}_{j_k}) \log_2 \left(\frac{p(\widetilde{x}_{j_k}, \widetilde{y}_{j_k})}{p(\widetilde{x}_{j_k}) p(\widetilde{y}_{j_k})} \right). \tag{8}$$

Accordingly, the channel capacity ζ_{j_k} is given by

$$\zeta_{j_k} = \sup_{x_{j_k}} \mathcal{I}(\mathbf{X}_{j_k}; \mathbf{Y}_{j_k}). \tag{9}$$

B. Spectrum Efficiency and Energy Efficiency

In practice, mmWave channels are expected to have a limited number of channel taps due to the small number of scatterers compared to the number of antennas [22], [23]. We thereby adopt a sparse geometric multipath channel model. In this paper, we assume all channels to be slow Rayleigh fading channels. The channel response does not change within one symbol block but could vary from block to block. Without loss of generality, the composite MIMO channel between BS

and user k can be represented by the widely used Saleh-Valenzuela channel model [51], which is given by

$$\mathbf{H}_k = \sqrt{\frac{N_{t,k}N_{r,k}}{L}} \sum\nolimits_{\ell=1}^{L} \kappa_{k,\ell} \mathbf{a}_{\mathrm{R}}(\theta_{k,\ell}) \mathbf{a}_{\mathrm{T}}^H(\phi_{k,\ell}), \quad (10)$$

where $\kappa_{k,\ell}$ denotes the complex channel coefficient of the ℓth propagation path. In our work, the mmWave channel estimation is simplified to estimate the channel gains, the angle-of-arrivals (AoAs), and angle-of-departures (AoDs) of the propagation paths [29], [52], [53]. The response of the array associated with user k can be modeled as

$$\begin{aligned} \mathbf{a}_{\mathrm{R}}(\theta_{k,\ell}) &= \frac{1}{\sqrt{N_{r,k}}} \left[1, e^{\cos(\theta_{k,\ell})}, \ldots, e^{(N_{r,k}-1)\cos(\theta_{k,\ell})}\right]^T, \\ \mathbf{a}_{\mathrm{T}}(\phi_{k,\ell}) &= \frac{1}{\sqrt{N_{t,k}}} \left[1, e^{\cos(\phi_{k,\ell})}, \ldots, e^{(N_{t,k}-1)\cos(\phi_{k,\ell})}\right]^T. \end{aligned} \quad (11)$$

For the convenience of derivation, we define the transmitted data symbols over the nth subcarrier as $\mathbf{x}_{n,k} \in \mathbb{C}^{J_k \times 1}, n = 1, \ldots, N$. The individual data stream is transmitted via $J_k N$-length data symbol blocks from the view of implementation in the frequency domain. Assuming perfect synchronization and FFT operation, the received MIMO-OFDM-IM signal of user k over the nth subcarrier in the frequency domain is given by [4], [9]

$$\mathbf{y}_{n,k} = \underbrace{\mathbf{W}_{n,k}^H \mathbf{H}_k \mathbf{F}_{n,k} \mathbf{x}_{n,k}}_{\text{primitive signal}} + \underbrace{\sum\nolimits_{j \neq k} \mathbf{W}_{n,j}^H \mathbf{H}_k \mathbf{F}_{n,j} \mathbf{x}_{n,j}}_{\text{interference component}} + \mathbf{W}_{n,k}^H \mathbf{z}_{n,k}, \quad n \in \mathfrak{N}_k, \quad (12)$$

where $\mathbf{W}_{n,k} = \mathbf{V}_k \widetilde{\mathbf{V}}_{n,k}$; $\mathbf{F}_{n,k} = \mathbf{U}_k \widetilde{\mathbf{U}}_{n,k}$; $\mathbf{I}_{n,j} = \mathbf{W}_{n,j}^H \mathbf{H}_k \mathbf{F}_{n,j} \mathbf{x}_{n,j}$ is the interference component from other users; $\mathbf{z}_{n,k} \sim \mathcal{CN}(0, \boldsymbol{\sigma}_k^2)$ is the circularly symmetric complex additive white Gaussian noise (AWGN) vector with mean zero and covariance matrix $\boldsymbol{\sigma}_k^2$.

The received signal-to-interference-plus-noise ratio (SINR) at the kth user can be represented by

$$\mathrm{SINR}_k = \frac{\sum_{n \in \mathfrak{N}_k} ||\mathbf{W}_{n,k}^H \mathbf{H}_k \mathbf{F}_{n,k} \mathbf{x}_{n,k}||^2}{\sum_{j \neq k} \sum_{n \in \mathfrak{N}_j} ||\mathbf{W}_{n,j}^H \mathbf{H}_k \mathbf{F}_{n,j} \mathbf{x}_{n,j}||^2 + \sum_{n \in \mathfrak{N}_k} ||\mathbf{W}_{n,k}^H \mathbf{z}_{n,k}||^2}. \quad (13)$$

Assuming Gaussian signaling and single-user detection, the interference can be treated as additional noise. We directly apply Theorem 1 proven in [49] to mmWave MIMO-OFDM-IM systems. Correspondingly, the SE of user k based on the Shannon formula can be written as (14), shown at the bottom of the page, (detailed derivation can be found in [49]), where $\mathbf{\Lambda}_{n,k}^{-1} = \mathbf{W}_{n,k}^H \mathbf{H}_k (\sum_{j \neq k} \mathbf{F}_{n,k} \mathbf{F}_{n,k}^H) \mathbf{H}_k^H \mathbf{W}_{n,k} + \sigma^2 \mathbf{W}_{n,k}^H \mathbf{W}_{n,k}$.

The total SE in bits/s/Hz of an arbitrary user can be found by maximizing the following objective function with the constraint of transmitted power:

$$\zeta_{\text{total}} = \sum\nolimits_{k=1}^{K} \omega_k \zeta_k, \quad (15)$$

where ω_k is the weight factor for user k; $\zeta_k = \sum_{j_k} \zeta_{j_k}$ is the SE of kth user. We can determine the weights according to user's traffic types, fairness, and other priority requirements. Note that in a multiuser network with concurrent transmissions, the PPP model can be conveniently applied to simulate the locations of the users as well as their mobile data traffic. Correspondingly, the SINR in (13) can be characterized, and the system performance can be experimentally evaluated [54].

In practice, the total energy consumption of a cellular system, consisting of both circuit power consumption and transmit power consumption, is dominated by the BS. Generally, the power consumption model at the BS consists of static and dynamic power consumption. The static power consumption model is constructed by the power consumption of all power amplifiers. The transmit power (dynamic) contains all the other circuit power. In this paper, we adopt a linear power consumption model proposed by [55]:

$$P_{\text{sys},k} = P_{t,k} + J_k P_{\text{RF},k} + P_{c,k} + N_{t,k} P_{\text{shift}}, \quad (16)$$

where $P_{\text{sys},k}$ is the system power dissipation of downlink transmission of user k; $P_{\text{RF},k}$ and P_{shift} denote the power consumption of a single RF chain and a phase shifter, respectively; $P_{c,k}$ is the other power consumption of circuitry components to support downlink transmission.[1] The transmitter's total power constraint $P_{t,k}$ is enforced by normalizing $\mathbf{F}_k$, i.e., $\sum_{n=1}^{N} ||\mathbf{F}_{n,k}||^2$, $n \in \mathfrak{N}_k$ [56]. The EE metric of the transmitter can be defined as the transmitted information bits per unit energy in the unit of bit per Joule (bit/J). It is worth to note that an exact computation of the dissipated power is a very difficult task [28]. Therefore, our work focuses on a generalized power consumption model, whose energy consumption can have a direct impact on energy-efficient optimal beam design.

For a set of SINRs $\{\mathrm{SINR}_k\}_{k=1}^K$, the total SINR scales in a multi-user system can be represented by

$$\mathrm{SINR}_{\text{sys}} = \sum\nolimits_{k=1}^{K} \omega_k \mathrm{SINR}_k. \quad (17)$$

To guarantee the fairness among users in the multiuser network, the max-min SINR problem is considered, which deals with the sum-rate maximization problem and guarantees the best performance of receivers. It is worth highlighting that the theoretical and algorithmic connection between maximizing the weighted sum rate and the max-min SINR problem

[1]Depending on the restriction on cost, HAD beamforming is currently an appropriate solution for mmWave systems. Alternatively, the future systems will likely use all-digital architectures, which can be interpreted as a digital counterpart. The proposed algorithms in this paper are applicable to the general case, and the analog component listed above can be directly substituted by a special type of digital fashion.

$$\zeta_k = J_k \left[\mathbb{E}\left(\frac{N_{A,k}}{N}\right) \log_2 \det(\mathbf{I}_{J_k} + \mathbf{W}_{n,k} \mathbf{\Lambda}_{n,k}^{-1} \mathbf{W}_{n,k}^H \mathbf{H}_k \mathbf{F}_{n,k}^H \mathbf{F}_{n,k} \mathbf{H}_k^H) + \frac{1}{N} \log_2 C(N, N_{A,k}) \right], \quad (14)$$

was revealed in [57], [58]. Leveraging on this fact, the joint optimization of SE-EE maximization is equivalent to the max-min SINR power control. This transformed problem has already been constructed as in [57], [58], and the interested readers can refer to a detailed proof in the aforementioned literature. Typically, the max-min SINR problem satisfying the overall transmit power constraint can be formulated as

$$\max_{\{\mathbf{F}_k,\mathbf{W}_k\}_{k=1}^K} \min \quad \mathrm{SINR}_k(P_{t,k}) \\ \text{s.t. } P_{t,k} \le P_{t,k}^{\max}, \quad \forall\, k,\ K \le \sum\nolimits_{k=1}^{K} J_k \le M_t, \quad (18)$$

where $P_{t,k}^{\max}$ is the maximum downlink transmission power of user k. It should be noted that (18) only considers a single-user case, which represents a locally optimal solution, only with respect to feasible solutions close to that point. To find the globally optimal solution, a global coordination is naturally required for every feasible solution of whole users [57], [58]. Equivalently, since each global maximum is also a local maximum, we can determine the overall optimization problem by maximizing the minimum weighted sum of SINR in (17).

The objective function in terms of EE is defined as the system capacity (bits/s) divided by the total power consumption. The EE (bits/J) of user k is then defined as

$$\max_{\{\mathbf{F}_k,\mathbf{W}_k\}_{k=1}^K} \eta_k = \frac{B_{T_k}\cdot\zeta_k}{\sum_{n=1}^{N}\|\mathbf{F}_{n,k}\|^2 + J_k P_{\mathrm{RF},k} + P_{c,k} + N_{t,k}P_{\mathrm{shift}}} \\ \text{s.t. } \sum\nolimits_{n=1}^{N} \|\mathbf{F}_{n,k}\|^2 \le P_{t,k}^{\max}. \quad (19)$$

In the following, our goal is to concurrently optimize the SE and EE under the individual SINR constraints given in (18). It inevitably brings about conflicts of interest among objective functions and needs a trade-off. According to the classical SE-EE trade-off paradigm, the objective function can be rewritten as

$$\eta_{\text{total}} = \frac{B_T\cdot\zeta_{\text{total}}}{\log_2(\sum_{k=1}^{K} P_{\text{sys},k})}. \quad (20)$$

III. Spectral-Energy Efficiency Trade-off With Pareto-Optimal Beam Design

A. Problem Statement

Using the aforementioned theoretical and algorithmic connection between the SE-EE problem and the SINR problem, in this section, we reformulate the problem of interest as an SE-EE maximization problem for the holistic system under the optimal beam design criteria, while guaranteeing a prescribed minimum SINR and maximizing the overall minimum SINR under a total power constraint for the transmit antennas. Specifically, it is equivalent to concurrently solve the following problems:

$$\begin{cases} \mathcal{Q}_{\mathrm{SE}}: \max\limits_{\{\mathbf{F}_k,\mathbf{W}_k\}_{k=1}^K} \sum\limits_{k=1}^{K} \zeta_k(\mathbf{F}_k,\mathbf{W}_k,\mathbf{P},\mathbf{J}) \\ \qquad \text{s.t. } \mathbf{1}^T\mathbf{P} \le P_t^{\max},\ \mathbf{1}^T\mathbf{J} \le M_t, \\ \mathcal{Q}_{\mathrm{EE}}: \max\limits_{\{\mathbf{F}_k,\mathbf{W}_k,\,\forall k\}} \sum\limits_{k=1}^{K} \eta_k(\mathbf{F}_k,\mathbf{W}_k,\mathbf{P},\mathbf{J}) \\ \qquad \text{s.t.} \quad \mathbf{1}^T\mathbf{P} \le P_t^{\max},\ \mathbf{1}^T\mathbf{J} \le M_t \end{cases}$$

and

$$\begin{cases} \mathcal{Q}_{\mathrm{SINR}}: \max\limits_{\mathbf{P},\,\mathbf{J}\in\mathbf{\Psi}} \quad \min \quad \mathrm{SINR}_{\mathrm{sys}}(\mathbf{P},\mathbf{J}) \\ \qquad \text{s.t.} \quad \mathbf{1}^T\mathbf{P} \le P_t^{\max}, \end{cases}$$

where $\mathbf{P} = \{P_1,\ldots,P_K\}$; $\mathbf{J} = \{J_1,\ldots,J_K\}$; $\mathbf{\Psi}$ is the set consisting of all feasible resource allocation strategies.

Mathematical optimization problems above can be abstracted as solving the maximum (or minimum) problem of one or more objective functions under certain constraints, which is identified by MOPs. Generally, a maximization MOP with $\mathfrak{m}$ objective functions $\{f_1(\boldsymbol{\mathcal{X}}), f_2(\boldsymbol{\mathcal{X}}),\ldots,f_{\mathfrak{m}}(\boldsymbol{\mathcal{X}})\}$ can be formulated as follows [59]:

$$\text{maximize } \boldsymbol{\mathcal{F}}(\boldsymbol{\mathcal{X}}) = (f_1(\boldsymbol{\mathcal{X}}),\ldots,f_{\mathfrak{m}}(\boldsymbol{\mathcal{X}}))^T \\ \text{s.t. } \boldsymbol{\mathcal{X}} \in E^{\mathfrak{n}}, \quad (21)$$

where $\boldsymbol{\mathcal{F}}(\boldsymbol{\mathcal{X}}) \in E^{\mathfrak{m}}$ is a vector function, that is, each element is an objective function; $\boldsymbol{\mathcal{X}} = \{\mathcal{X}_1,\ldots,\mathcal{X}_{\mathfrak{n}}\}$ is a feasible variable set selected from the decision variable space, aiming to maximize each objective function by searching to identify an optimal solution. $E^{\mathfrak{m}}$ and $E^{\mathfrak{n}}$ are the objective function space and decision variable space, respectively. The decision space represented by MOPs can be characterized by tuple $\mathcal{M} = (E^{\mathfrak{m}}, E^{\mathfrak{n}})$. If there are conflicts of interest among multiple objectives, it is generally difficult to find the globally optimal solution to the MOPs in (21). In particular, when a set of non-convex objective functions conflict with each other, solving the mixed MOP with high-dimensional parameters becomes extremely difficult, and the problem could even be NP-hard. From a methodological perspective, solving such an optimization problem may also incur an excessive amount of computational resource and a large search space.

In the following subsections, we focus on the relaxation of the conflicting objectives and the Pareto-optimal SE-EE trade-off to obtain a tractable solution. By applying the problem transformation, we recast the above MOPs as several suboptimal problems. We show that the paradox in the conflicting objectives can thus be alleviated, and the globally optimal SE-EE trade-off can be achieved with lower computational complexity.

B. Relaxation of SE-EE Maximization

In this subsection, we investigate the relaxation of SE-EE maximization, as well as the impact of SINR under multi-user interference constraints. To ease expositions, we use $f_{\mathrm{SE}}(\cdot)$, $f_{\mathrm{EE}}(\cdot)$, and $f_{\mathrm{SINR}}(\cdot)$ to represent the objective functions with respect to SE, EE and SINR, respectively. For the kth user, the objective function can be expressed as

$$\boldsymbol{\mathcal{F}}(\boldsymbol{\mathcal{X}}_k) = (f_{\mathrm{SE}}(\boldsymbol{\mu}_k), f_{\mathrm{EE}}(\boldsymbol{\nu}_k), f_{\mathrm{SINR}}(\boldsymbol{\mathcal{X}}_k))^T, \quad \boldsymbol{\mathcal{F}}(\boldsymbol{\mathcal{X}}_k) \in \mathcal{M}, \quad (22)$$

where $\boldsymbol{\mu}_k$, $\boldsymbol{\nu}_k$, $\boldsymbol{\xi}_k \in \boldsymbol{\mathcal{X}}_k$ is the set of the decision vectors with the same or different variable size(s), respectively. We accordingly use the sets $\{\boldsymbol{\mathcal{F}}(\boldsymbol{\mathcal{X}}_k)\}_{k=1}^K$ to denote the objective functions of all users.

1) Problem Transformation: For some practical applications, although MOPs with multi-dimensional variables (parameters) are common in various engineering domains, the joint

optimization over more than two objectives is extremely difficult to solve. Since the optimization of max-min SINR is usually subject to non-convex constraints, these trade-offs should be carefully considered. Otherwise, it would incur prohibitively high computational demands. To achieve the globally optimal SE-EE trade-off, our solution, which resorts to the decomposition of the optimization process, consists of two independent stages. First, SE and EE maximization of every user is performed in a best-effort fashion without setting other priority balance.[2] In this case, we assume that SINR is approximately constant and does not vary until the end of the first stage. Then, the SE-EE trade-off between the power and number of RF chains as well as throughput is optimized to approach the Pareto-optimal beam design. As a result, the vector objective function of the MOP can be reduced to a single objective function, which is monotonically increasing with SINR. Note that this assumption is only made at the starting stage for initialization purposes, which is no longer needed in the subsequent processing.

Then, a globally optimal solution for SINR can be solved in a straightforward manner by properly adjusting the reconfigurable power and the number of RF chains. In the case of mmWave HAD systems, it is important to mention the fact that finding the Pareto-optimal SE and EE can be equivalently transformed into a design problem of the optimal beam. Meanwhile, it is equivalent to the max-min weighted SINR power control problem, as well as the practical constraints on the number of active RF chains. Using this problem transformation, the paradox in (22) can be relaxed by sequentially solving the following optimization problem:

$$\begin{aligned} &\max \quad f_{\text{SE}}(\boldsymbol{\mu}_k) = f_{\text{SE}}(\mathbf{F}_k, \mathbf{W}_k, \mathbf{P}_k, J_k), \\ &\max \quad f_{\text{EE}}(\boldsymbol{\upsilon}_k) = f_{\text{EE}}(\mathbf{F}_k, \mathbf{W}_k, \mathbf{P}_k, J_k) \\ &\text{s.t.} \sum\nolimits_{k=1}^{K} J_k \le M_t, \ \sum\nolimits_{k=1}^{K} P_{t,k} \le \sum\nolimits_{k=1}^{K} P_{t,k}^{\max}, \end{aligned} \tag{23}$$

and

$$\begin{aligned} &\max_{\bar{\boldsymbol{\alpha}}} \min_{\{\mathbf{P}_k, J_k\}_{k=1}^{K}} \quad f_{\text{SINR}}(\boldsymbol{\xi}_k) = \sum\nolimits_{k=1}^{K} \bar{\alpha}_k f_{\text{SINR}}(\mathbf{P}_k, J_k), \\ &\text{s.t.} \sum\nolimits_{k=1}^{K} \bar{\alpha}_k = 1, \end{aligned} \tag{24}$$

where $\bar{\boldsymbol{\alpha}}^T \in \mathbb{R}^{1\times K}$ with rows $\bar{\alpha}_k \in [0,1]$; $\boldsymbol{\mu}$, $\boldsymbol{\nu}$ and $\boldsymbol{\xi}$ are constructed by those limited variable combinations that have a significant impact on the objective function, so as to reduce complexity; $\mathbf{P}_k = (P_1, \ldots, P_{J_k})$ represents the link power allocation outcome with respect to the kth user. For simplicity, the power allocated to the reference subcarrier is assumed to be equal in this paper. From the perspective of the Pareto principle, the key to solve (23) and (24) is to find the Pareto-optimal decision vector from the feasible decision vectors. Meanwhile, it may be unrealistic to consider all possible decision variables for solving MOPs. For this reason, this paper only considers a limited set of decision vectors corresponding to an approximately optimal solution so as to limit the computational complexity. To be precise, the feasible set $\boldsymbol{\mathcal{X}}$ can be constructed by finite options of decisions, i.e., a linear combination of all chosen decision vectors, which can be viewed as a closed and bounded subset of $\boldsymbol{\mathcal{G}}$ (all possible decision sets). For K active users, it is given by $\boldsymbol{\mathcal{X}}$: $\{\boldsymbol{\mathcal{X}}_k = \{\mathbf{F}_k, \mathbf{W}_k, \mathbf{P}_k, J_k\}\}_{k=1}^{K}$. In the case of mmWave HAD beamforming, we refer to this strategy as a reconfigurable combinatorial-oriented beam management approach.

2) Energy-Efficient Transmit Beam Design: It is important to stress that in practice, the prerequisites of mmWave communications with high SE and EE have necessitated the requirement: (1) highly directive beams are required to compensate for the severe path loss; (2) these beams need to be aligned for affording robust transmission links. In particular, high-directivity beam as well as the alignment of beam pairs at the transmit and receive antennas depend on the steering vector, which is dominated by the phased array, i.e., the analog beamformer/combiner. These facts in turn imply that the aligned beam pairs are always Pareto-optimal. The optimization of the analog and digital parts can be thereby decoupled and treated separately. To this end, the key consideration of this paper is that the design of Pareto-optimal beam can be cast as the optimization problem on digital baseband precoder and combiner. By decoupling the optimization variables, the analog part can be tackled by conventional optimization techniques and regarded as a special type of the Pareto-optimal beam design. The advantage of this strategy is that it not only can greatly reduce the dimension of MOP, but also prevent the impacts of nonlinearly associated with the analog components. In other words, we are able to independently optimize $\widetilde{\mathbf{U}}_k$ and $\widetilde{\mathbf{V}}_k$, i.e., the optimal $\widetilde{\mathbf{U}}_k^{\star}$ and $\widetilde{\mathbf{V}}_k^{\star}$ can be found by assuming that $\mathbf{U}_k$ and $\mathbf{V}_k$ are fixed. To obtain optimal $\widetilde{\mathbf{U}}_k^{*}$, the perfect channel state information (CSI) is assumed to be available at the BS. Based on (12), the $\mathbf{F}_k$ with locally optimal beam design at the BS can be thereby determined by jointly solving the following optimization problem:

$$\begin{aligned} \widetilde{\mathbf{U}}_{n,k}^{\star} = \underset{\widetilde{\mathbf{U}}_{n,k},\ \forall\, k}{\arg\max} \ & \mathbf{F}_{n,k}^H \mathbf{x}_{n,k}^H \mathbf{x}_{n,k} \mathbf{F}_{n,k} \\ \text{s.t.} \quad & \sum\nolimits_{n=1}^{N} \text{Tr}(\mathbf{F}^H{}_{n,k} \mathbf{F}_{n,k}) < P_{t,k}. \end{aligned} \tag{25}$$

At the receiver, the feasible $\mathbf{W}_k^{\star}$ associated with $\widetilde{\mathbf{V}}_k^{*}$ for each user can be computed in the same way as

$$\begin{aligned} \widetilde{\mathbf{V}}_{n,k}^{\star} = \underset{\widetilde{\mathbf{V}}_{n,k},\ \forall\, k}{\arg\max} \ & \frac{\left|\left|\mathbf{W}_{n,k}^H \mathbf{H}[k] \mathbf{F}_{n,k}\right|\right|^2 ||\mathbf{x}_{n,k}||^2}{\left|\left|\mathbf{W}_{n,k}^H \mathbf{z}_{n,k}\right|\right|^2} \\ \text{s.t.} \ & \sum\nolimits_{n=1}^{N} ||\mathbf{W}_{n,k}||^2 = 1. \end{aligned} \tag{26}$$

By explicitly decoupling the optimization operation, $\widetilde{\mathbf{U}}_k$ and $\widetilde{\mathbf{V}}_k$ can be determined first. After that, the desired digital precoder and combiner can be treated as a matrix factorization problem, resulting in low complexity. The optimal digital combiner can be independently tackled via the MMSE criterion [4]:

$$\mathbf{V}_{n,k} = \boldsymbol{\Lambda}^{-1} \widetilde{\mathbf{U}}_{n,k}^H \mathbf{H}_{n,k} \widetilde{\mathbf{U}}_{n,k} \mathbf{U}_{n,k}. \tag{27}$$

Although the CSI is necessary for mmWave transmissions, imperfect CSI is always present in practice due to fast-varying channels or channel estimation errors. To evaluate the impact

[2]In this context, a best-effort fashion refers to a mechanism in which all users make their best effort to separately achieve SE-EE maximization as much as possible. Intuitively, the behavior of best-effort optimization is supposed to capture only a locally optimal SE-EE trade-off.

of imperfect CSI, the estimated channel matrix $\hat{\mathbf{H}}_k$ with imperfect CSI can be modeled as [27]

$$\hat{\mathbf{H}}_k = \varepsilon \mathbf{H}_k + \sqrt{1-\varepsilon^2}\boldsymbol{\Delta}, \tag{28}$$

where ε denotes the CSI accuracy, and $\boldsymbol{\Delta} \sim \mathcal{CN}(0, \mathbf{I})$ is the error matrix that characterizes the channel estimation error.

By removing $\widetilde{\mathbf{U}}_k$ and $\widetilde{\mathbf{V}}_k$ from the problem formulated in (23) for the kth user, the problem of finding a set of decision vectors $\boldsymbol{\mathcal{X}}_k^\star$ that maximizes the SE and EE can be represented by

$$\max_{\{\boldsymbol{\mathcal{X}}_k^\star\}_{k=1}^K} (f_{\text{SE}}(\mathbf{U}_k, \mathbf{V}_k, \mathbf{P}_k, J_k), f_{\text{EE}}(\mathbf{U}_k, \mathbf{V}_k, \mathbf{P}_k, J_k)). \tag{29}$$

Then, the transferred problem given in (29) can temporarily achieve a single locally optimal solution associated with the Pareto-optimal beam design of two objective functions. Since the transmit beamforming mainly depends on the channel direction, the individual beam alignment should be a necessary and sufficient condition to accommodate the SE-EE maximization. From this standpoint, the analog part of beamformers can be fixed, while is to relax the constraint. We start from this point and focus on solving the globally Pareto-optimal beam design in the rest of this section.

C. Pareto-Optimal Solution to SE-EE Trade-Off

The objective of this subsection is to provide an efficient procedure for the SE-EE maximization of entire system. To generate universally Pareto-optimal SE-EE maximization, we show that there exists a globally optimal solution that maximizes EE while having an increasing SE. For satisfying various rate requirements of the users, our proposed strategies involve properly balancing SINR and choosing the beam management strategy with an adequate number of active RF chains. To do this, we apply a weighted-sum criterion to capture multiple objectives, which is a standard technique for finding POS for vector optimization problems. This linear weighted-sum method scalarizes multiple performance metrics into a single-objective function. The optimal solution, i.e., $\boldsymbol{\mathcal{X}}_k^{\text{POS}}$, is based on the solution to $\boldsymbol{\mathcal{X}}_k^\star$, resulting from a particular set of weights. This method, also called *scalarization*, can be expressed as

$$\begin{aligned} &\arg\max_{\boldsymbol{\mathcal{X}}, \bar{\boldsymbol{\alpha}}} \quad \sum\nolimits_{k=1}^K \bar{\alpha}_k \mathcal{F}(\boldsymbol{\mu}_k, \boldsymbol{\nu}_k, \boldsymbol{\xi}_k) \\ &\text{s.t.} \quad \boldsymbol{\mu}_k, \boldsymbol{\nu}_k, \boldsymbol{\xi}_k \in \boldsymbol{\mathcal{X}}^\star, \quad \sum\nolimits_{k=1}^K \bar{\alpha}_k = 1, \end{aligned} \tag{30}$$

where $\bar{\boldsymbol{\alpha}}^T \in \mathbf{R}^{K\times 1}$ with rows $\bar{\alpha}_k$.

1) Compromise a Solution to the Pareto-Optimal Beam Design: According to the criterion of the Pareto principle, once the objective values of $\boldsymbol{\mathcal{X}}^{\text{POS}}$ with respect to the Pareto-optimal beam design are achieved, it means that $\boldsymbol{\mathcal{X}}^{\text{POS}}$ is not dominated by any other feasible decision vectors and satisfies $\mathcal{F}(\boldsymbol{\mathcal{X}}^{\text{POS}}) \succeq \mathcal{F}(\boldsymbol{\mathcal{X}}^\star)$, with objective set $\{\mathbf{U}_k^{\text{POS}}, \mathbf{V}_k^{\text{POS}}, \mathbf{P}_k^{\text{POS}}, J_k^{\text{POS}}\}$. The objective value of a set of Pareto-optimal decision vectors, being chosen as optimal, constitutes the Pareto front. Note that the relaxation strategy effectively avoids the singularity problems caused by non-linearity and non-convexity. In this case, the feasible $\mathcal{X}^\star$ is spanned by linear space because the elements of objective vector set $\{\mathbf{U}_k^{\text{POS}}, \mathbf{V}_k^{\text{POS}}, \mathbf{P}_k^{\text{POS}}, J_k^{\text{POS}}\}$ are linear. To this end, a feasible solution to finding optimal $\boldsymbol{\mathcal{X}}^{\text{POS}}$ can be easily found by numerical methods.

In the following, we apply Cobb-Douglas production function to characterize the trade-off between SE and EE. It can be expressed by a utility function U, which has the form [60]

$$U(f_{\text{SE}}(\boldsymbol{\mathcal{X}}_k^\star), f_{\text{SE}}(\boldsymbol{\mathcal{X}}_k^\star)) = (\widehat{f_{\text{SE}}}(\boldsymbol{\mathcal{X}}_k^\star))^{\alpha_k} (\widehat{f_{\text{EE}}}(\boldsymbol{\mathcal{X}}_k^\star))^{1-\alpha_k}, \tag{31a}$$

$$\begin{aligned} \widehat{f_{\text{SE}}}(\boldsymbol{\mathcal{X}}_k^\star) &= \frac{f_{\text{SE}}(\boldsymbol{\mathcal{X}}_k^\star)}{f_{\text{SE}}^{\max}(\boldsymbol{\mathcal{X}}_k^\star)} \text{ and } \widehat{f_{\text{EE}}}(\boldsymbol{\mathcal{X}}_k^\star) \\ &= \frac{f_{\text{EE}}(\boldsymbol{\mathcal{X}}_k^\star)}{f_{\text{EE}}^{\max}(\boldsymbol{\mathcal{X}}_k^\star)}, \end{aligned} \tag{31b}$$

where $\alpha_k \in [0, 1]$ is a normalized weight metric on a particular trade-off between SE and EE for the conflicts of interest. With this empirical production function, different weights can be viewed as the benefit-cost ratio of the system. Accordingly, the objective value can be evaluated by taking the sum of the elasticities of output. For the purpose of evaluation, a logarithmic form of above function can be given as follows:

$$\begin{aligned} \mathcal{H}_k: \ \mathcal{H}(\boldsymbol{\mathcal{X}}_k^\star) = \alpha_k &\log(\widehat{f_{\text{SE}}}(\mathbf{U}_k^\star, \mathbf{V}_k^\star, \mathbf{P}_k^\star, J_k^\star)) \\ &+ (1-\alpha_k)\log\widehat{f_{\text{EE}}}(\mathbf{U}_k^\star, \mathbf{P}_k^\star, J_k^\star). \end{aligned} \tag{32}$$

As can be observed, the above MOP of finding an optimal decision vector set is transformed into a single globally optimal problem:

$$\begin{aligned} &\max_{\{\mathbf{U}_k^\star, \mathbf{V}_k^\star, \mathbf{P}_k^\star, J_k^\star\}} \mathcal{H}(\mathbf{U}_k^\star, \mathbf{V}_k^\star, \mathbf{P}_k^\star, J_k^\star) \\ &\quad \text{s.t. } \{\mathbf{U}_k^\star, \mathbf{V}_k^\star, \mathbf{P}_k^\star, J_k^\star\} \in \boldsymbol{\mathcal{X}}^\star, \quad \sum\nolimits_{k=1}^K J_k^\star \le \sum\nolimits_{k=1}^K J_k. \end{aligned} \tag{33}$$

Recall (30) and (32) that are for mmWave HAD beamforming systems with interference constraints. We have considered the optimal beam design that satisfies multiple objectives in a multi-user environment. It is worth pointing out that aforementioned $\bar{\boldsymbol{\alpha}}$ and the weight vector with respect to the utility function are equivalent. Note that, considering the optimal target set of the overall user in the whole beamspace, the globally optimal trade-off between SE and EE can be expressed as

$$\begin{aligned} \mathcal{H}_{\text{global}} \succeq \sum\nolimits_{k=1}^K &(\alpha_k f_{\text{SE}}(\mathbf{U}_k^\star, \mathbf{V}_k^\star, \mathbf{P}_k^\star, J_k^\star) \\ &+ (1-\alpha_k) f_{\text{EE}}(\mathbf{U}_k^\star, \mathbf{V}_k^\star, \mathbf{P}_k^\star, J_k^\star)). \end{aligned} \tag{34}$$

Even though the scalarized problem in (34) would naturally come up with a compromise, since it may be convex, it is feasible to explore its Pareto front by this linear function with a reasonable cost (that is, solvable in polynomial time) [35], [61].

In fact, due to the diversity of rate requirements, different users have different load priorities on the SE-EE trade-off. Typically, the weights $\boldsymbol{\alpha}_k$ can be used to balance aggregate utility and/or user fairness. In the user-centric scenario, EE maximizations are generally considered with respect to minimum-rate and/or SINR constraints, including the reconfigurable number of multiplexed data streams. For

cell-center users (strong users) and cell-edge users (weak users), the low and the high power allocation strategies are better to gain a higher network throughput. For strong users, the over-large transmit power does not necessarily help to improve the total SE, since the interference would dramatically increase. To achieve the optimal SINR in a multi-user network, the worst-case (weak users) and best-case (strong users) SINR optimization problems need to be balanced.

2) Combinatorial Resource Reallocation: From a systematic perspective, the optimization problem formulated in (34) under the SINR constraint would be still arduous if a great deal of adjustment of different parameters of the decision vector is imposed. To this end, we propose a combinational-oriented transmit power control strategy to adaptively handle the worst and the best cases. With the aid of MIMO-OFDM-IM, a power reallocation and/or power saving scheme can be directly applied to balance the total transmit power. This additional degree of freedom is capable of providing an appropriate trade-off between SE and EE as well as error performance. Let $\boldsymbol{\mathcal{P}} = (\beta_1, \ldots, \beta_K)^T$ be the instantaneous power reallocation set, which is an additional product of the power saved by those inactive subcarriers of OFDM-IM. Note that (33) is equivalent to the classical sum rate maximization [57]. After a beam design which is proven to be optimal, the power allocation solution can be yielded by applying the Lagrangian method, e.g., the so-called water-filling solution (proved in [61], p. 245).

To achieve the optimal SINR, we then turn to define a resource reallocation policy $\boldsymbol{\Psi}^{\mathcal{T}}$, which is a new combination of decision vector sets $\boldsymbol{\Psi}^{\mathcal{T}} : \{\boldsymbol{\Psi}_k = \{\mathbf{P}_k, J_k; \beta_k\}\}_{k=1}^K$ at decision epoch $\mathcal{T}$. Without loss of generality, this resource reallocation problem can be interpreted as searching for a feasible strategy and has been formulated as in [35]

$$\begin{aligned} \underset{\boldsymbol{\Psi}^{\mathcal{T}+1} \succ \boldsymbol{\Psi}^{\mathcal{T}}, \ldots, \boldsymbol{\Psi}^{1} \succ \boldsymbol{\Psi}^{0}}{\text{maximize}} \quad & \{f_{\text{SINR}}(\boldsymbol{\Psi}_1), \ldots, f_{\text{SINR}}(\boldsymbol{\Psi}_K)\} \\ \text{s.t.} \ \ & \sum\nolimits_{k=1}^{K} \text{Tr}(\mathbf{1}^T \mathbf{P}_k) \le P_t^{\max}, \quad \sum\nolimits_{k=1}^{K} J_k \le N_t. \end{aligned} \tag{35}$$

In this way, the effective source reallocation among the users will be gradually achieved over each iteration, and then be extracted as $\boldsymbol{\Psi}^{\mathcal{T}+1} | \boldsymbol{\Psi}^{\mathcal{T}}$ that satisfies the power constraints and maximizes the performance $f_{\text{SINR}}(\boldsymbol{\mathcal{X}}_k), \forall\, k$.

A key characteristic of such an approach is to allocate power, RF chains and bandwidth optimally among users, whilst the beamforming matrixes can be temporarily fixed (see the necessary and sufficient conditions discussed in Section III-B). More precisely, the combinatorial optimization strategy $\boldsymbol{\Psi}$ is equivalent to select a limited subset from all feasible combinations of decision variables, and the search-space for finding optimal decision vector sets is thereby shrunken. Accordingly, the optimization problem of (24) can be transferred to the following single-objective optimization problem:

$$\begin{aligned} \max_{\{\mathbf{U}_k^\star, \mathbf{V}_k^\star\}_{k=1}^K,\ \bar{\boldsymbol{\alpha}}} \ \min_{\{\mathbf{P}_k^{\text{POS}}, J_k^{\text{POS}}\}_{k=1}^K} & \sum\nolimits_{k=1}^{K} \bar{\alpha}_k f_{\text{SINR}}(\mathbf{P}_k^\star, J_k^\star; \beta_k) \\ \text{s.t.} \ \ & P_{t,k} \le P_{t,k}^{\max}, \quad \forall\, k. \end{aligned} \tag{36}$$

In particular, by combinatorially altering the values of the members of $\boldsymbol{\Psi}$, the proposed policy is able to find the best one from a set of globally optimal solutions, and (36) yields the complete Pareto-optimal set [35], [36].

3) Approaching Complete Pareto-Optimal Beam Design: By applying the method in [35], [36], one can explicitly see that the proposed policy can also lead to the sum rate maximization as well as to maximize the minimum rate of the users, especially in the low SNR region. In our work, the desired quantization level of SINR is indicated by $\boldsymbol{\alpha}$ and $\boldsymbol{\mathcal{P}}$, which can be used to improve SINR of weak users. If all of the constraints in (36) are satisfied, the reduced J_k^{POS} and $\mathbf{P}_k^{\text{POS}}$ produce the optimal power allocation scheme. In this case, when coefficients of $\boldsymbol{\alpha}$ are properly balanced, it is effective to perform optimal downlink power transmission for all active users.

Importantly, the expression in (36) also indicates that we can achieve an approximately complete Pareto-optimal beam design. The constrained weight vector yields a part of POS, i.e., $\mathbf{P}_k^{\text{POS}}$ and J_k^{POS}. Obviously, if all of the constraints in (36) are satisfied, we can obtain the partially suppressed J_k^{POS} and the feasible power reallocation scheme $\mathbf{P}_k^{\text{POS}}$ for all users. It is possible that the network can dynamically choose its most desired weight vector. When these weighting coefficients are properly adjusted, i.e., $\boldsymbol{\alpha}^\star \to \boldsymbol{\alpha}^{\text{POS}}$, the SINR of the entire downlink approaches the globally Pareto-optimal design.

Recalling (34), it can provide the complete Pareto-optimal set, despite the nonconvexity (if any) of the considered problems. With the combined decision variable set $\boldsymbol{\Psi}$ for achieving the Pareto-optimal beam design, mathematically, the globally optimal Pareto solution can be found by alternately solving the following problem:

$$\begin{aligned} \max_{\{\mathbf{U}_k^{\text{POS}}, \mathbf{V}_k^{\text{POS}}\}_{k=1}^K} \ & \mathcal{H}_{\text{global}} \succeq \sum\nolimits_{k=1}^{K} \mathcal{H}(\mathbf{U}_k^\star, \mathbf{V}_k^\star, \mathbf{P}_k^{\text{POS}}, J_k^{\text{POS}}, \beta_k) \\ \text{s.t.} \ & \{\mathbf{P}_k^{\text{POS}}, J_k^{\text{POS}}, \beta_k\}_{k=1}^K \in \boldsymbol{\Psi}. \end{aligned} \tag{37}$$

Note that the above approaches focus on one specific subject, i.e., finding the optimal points towards the true Pareto front in the space of objective functions in MOPs. Correspondingly, the obtained result is approximately Pareto optimal, i.e., optimized beam designs towards the Pareto front. To achieve this, a more efficient algorithm needs be devised to find the globally optimal Pareto solution to the SE-EE maximization. Most existing multi-objective evolutionary algorithms have been proven to be efficient to approximate POS in both of the decision and objective spaces [62]. Subsequently, we will provide a generic evaluation framework and give an alternative way of finding POS towards the true Pareto front. Based on the classical evaluation algorithm, a set of candidates $\boldsymbol{\mathcal{X}}^\star$ can gradually evolve to seek a set of promising points that are as close to the Pareto-optimal $\boldsymbol{\mathcal{X}}^{\text{POS}}$ as possible.

IV. Pareto-Optimal SE-EE Evaluation

A. Spatial Poisson Point Process

To systematically evaluate the proposed optimization strategy, we consider the case of a single-cell mmWave MIMO-OFDM-IM system. The location of the BS is fixed and set as the original point of the coordinate, and a single cell system

is considered to be a circular disc of radius R. The active K users are located on a Euclidean plane, and randomly distributed around the BS. The location of the kth user, represented by polar radius r_k and polar angle φ_k, can be modeled as an independent PPP in $\mathbb{R}^m$ [54]. Denote λ_k as the corresponding intensity of these K users. Each user can be modeled as an independent homogeneous PPP with intensity λ_k, i.e., $\Gamma = \{(r_k, \varphi_k; \lambda_k)\}_{k=1}^{K}$, where (r_k, φ_k) is a two-dimensional polar coordinate. Since the locations of K active users with $\{\Gamma_1, \ldots, \Gamma_K\}$ are random variables in a given area, they can be randomly generated according to the PPP. Fig. 2 illustrates a snapshot of PPP in the beamspace, where the size of the circle represents the corresponding scaled intensity of the PPP. Note that although we consider a single cell multiuser downlink system, the model also works for multi-cell systems, since the case of single-cell downlink transmission can be easily extended to multi-cell setups. Importantly, by applying a simple but effective optimization approach that was proposed in [35], [36], the multi-cell resource allocation can be generally transformed into a monotonic problem that can be solved with the global trade-off.

As shown in Fig. 2, the distributions of users are modeled as points of a homogeneous PPP on the plane, and the differentiated rate requirements of users are also approximated by using stochastic geometry tools for the performance evaluation of the proposed optimization strategy. In order to make the evaluation tractable, we assume that the proposed optimization strategy operates in a succession of time intervals, i.e., epochs, where each PPP realization maps to the current decision epoch $\mathcal{T}$. Then, we can evaluate the impact of superimposing performance from one epoch to the next. For any given decision epoch, the intensity of the underlying PPP can be obtained in a closed-form solution, by which we can signify the strong users and weak users of interest. Note that this systematic approach can actually be interpreted as a priori articulation of preferences. Most importantly, such a predefined configuration can be reproduced, and is thereby more suited to fairly evaluate the performance of optimization strategy in practice with the SE-EE trade-off.

According to the aforementioned Pareto-optimal solution, α_k is proportional to λ_k. If the full set of λ_k is given, then a specific vector $\boldsymbol{\alpha}$ can be estimated. As a result, the optimal compromise can be obtained for optimizing the SE-EE trade-off. In this paper, it is worth pointing out that by making use of each user's channel quality indicator (CQI) feedback, $\boldsymbol{\alpha}$ can be easily acquired at the BS, and the weights can be flexibly altered according to the dynamic of the network load as well as the CQI information.

Define $\mathcal{R}$ to be the state transition rate corresponding to the gradient of intensity between transient states. The dynamic of PPP over a specified decision epoch $\mathcal{T}$ can be modeled as a backward Markov chain:

$$p(\Gamma_k^i | \Gamma_k^{i-1}) = \prod_{j=0}^{j} \Pr\left(1 + \mathcal{R}(\Gamma_k^{i+j} | \Gamma_k^{i+j-1})\right), \quad \forall\, k, \quad (38)$$

where $p(\Gamma_k^i | \Gamma_k^{i-1})$ is an accumulated one-step transition probability from the $(i-1)$th decision epoch.

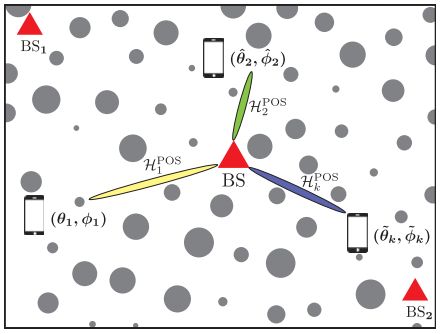


Fig. 2. A snapshot of the spatial PPP is directly applied in beamspace, where nomadic users are uniformly and independently distributed, and the randomly generating intensity of the mobile data traffic is proportional to the size of the circle. Each beam denoted as (θ, ϕ) is associated with a desirable $\mathcal{H}^{\mathrm{POS}}$.

B. Determining Pareto-Optimal Set

In this section, we show that the proposed MOP solution to finding POS towards the true Pareto front can be calculated by means of a multi-objective evolutionary algorithm, e.g., non-dominated sorting genetic algorithm-II (NSGA-II) [62]. Many studies reveal that NSGA-II has a better search ability than exhaustive search. Thus, it is often more efficient to find multiple Pareto-optimal solutions in one single simulation run, and many NP-hard MOPs can thereby be solved using this technique.

Our target is to determine the Pareto-optimal SE-EE trade-off through the feasible decision vector set $\boldsymbol{\mathcal{X}}^{\star}$ by identifying the globally POS. Once searched outcome converges to the optimal or near-optimal solution, the Pareto-optimal beams for K active users are available to achieve SE-EE maximization under SINR constraints. As described in Section III-B, the dimension of MOP is reduced by the decomposition-based method for the optimization of the analog and digital beamforming. By relaxing the design problem of the objective function in (22), this dimension reduction approach can considerably reduce the search space and still guarantees globally optimal solutions. Note that NSGA-II is an algorithmic tool for finding out the Pareto front, and other advanced evolutionary algorithms can be alternatively applied to solve such problems.

Algorithm 1 summarizes the steps to search through the feasible vector space for identifying the globally POS, please find it at the next page. It is initialized by the initial user discovery as well as the beam training procedure. In the first stage, SE and EE maximization of each user is performed in a best-afford fashion without setting any priority balance. The joint maximization of SE and EE for each user is performed concurrently, and we thereby achieve a single locally Pareto-optimal solution. In the second stage, the globally optimal trade-off between SE and EE is taken into account. It mainly aims to apply NSGA-II for finding the globally POS under the SINR constraints. We alternately solve the problem formulated in (33) and (37) based on the combinatorial-oriented power control strategy for the optimal transmission.

For given (30), we solve the two subproblems (33) and (36) separately to obtain the optimal decision vector sets $\{\mathbf{U}_k^{\star}, \mathbf{V}_k^{\star}, \mathbf{P}_k^{\star}, J_k^{\star}\}_{k=1}^{K}$ and $\boldsymbol{\Psi}$. In each iteration, NSGA-II repeats this process until the convergence is achieved, i.e., $\boldsymbol{\mathcal{X}}^{\mathrm{POS}}$ is identified. By NSGA-II, ϵ is a ratio of Euclidean

Algorithm 1 Evolutionary Algorithm Searching POS for the Globally Pareto-Optimal SE-EE Trade-off

Input: $\boldsymbol{\mathcal{F}} = \{f_{\text{SE}}, f_{\text{EE}}; f_{\text{SINR}}\}$, $\boldsymbol{\mathcal{X}} = \{\mathbf{F}_k, \mathbf{W}_k, \mathbf{P}_k, J_k\}_{k=1}^{K}$.
▷ K active users

Output: $\mathcal{H}_{\text{global}}^{\text{POS}} \leftarrow \{\mathbf{U}_k^{\text{POS}}, \mathbf{V}_k^{\text{POS}}, \mathbf{P}_k^{\text{POS}}, J_k^{\text{POS}}\}_{k=1}^{K}$.

1: **initialization:** Generate K active users with $\{\Gamma_1, \ldots, \Gamma_K\}$.
2: ▷ PPP
3: $\mathbf{U}_k,\ \mathbf{V}_k,\ \forall k \rightarrow \widetilde{\mathbf{U}}_k^{\star},\ \widetilde{\mathbf{V}}_k^{\star}$. ▷ user discovery
4: **for** k **from** 1 **to** K **do** ▷ beam alignment
5: Search $\boldsymbol{\mathcal{X}}_k^i$: $(\mathbf{U}_k, \mathbf{V}_k, \mathbf{P}_k, J_k)$ to maximize $\boldsymbol{\mathcal{F}}(f_{\text{SE}}, f_{\text{EE}})$.
6: Set $\boldsymbol{\mathcal{X}}_k^{i+1} \leftarrow \boldsymbol{\mathcal{X}}_k^i$.
7: return $\mathcal{H}_{\text{global}}^{(i)} : \boldsymbol{\mathcal{X}}_k^{(0)} \leftarrow \boldsymbol{\mathcal{X}}_k^{\star}$.
8: ▷ initial population: locally optimal trade-off
9: **end for**
10: **repeat**
11: SE-EE Trade-off: $\mathcal{H}_{\text{global}}^{(i)}$: $\mathcal{H}(\mathbf{U}_k^{\star}, \mathbf{V}_k^{\star}, \mathbf{P}_k^{\star}, J_k^{\star})$.
12: Measure $\boldsymbol{\alpha}$ weight metrics. ▷ Cobb-Douglas
13: **for all** k **do** ▷ complexity reduction
14: max-min: $f_{\text{SINR}}(\boldsymbol{\xi}_k) = \sum_{k=1}^{K} \alpha_k f_{\text{SINR}}(\mathbf{P}_k, J_k)$.
15: *Subprocedure*: an instantiation of NSGA-II
16: evoluatary subroutine for solving (37) ▷ search $\boldsymbol{\mathcal{X}}_k^{\text{POS}}$
17: **If** $\mathcal{H}_{\text{global}}^{(j)} \prec \mathcal{H}_{\text{global}}^{(j-1)}$ **then** $\boldsymbol{\mathcal{X}}_k^{(j+1)} \leftarrow \boldsymbol{\mathcal{X}}_k^{(j)}$.
18: **else if** $\mathcal{H}_{\text{global}}^{(j)} \succeq \mathcal{H}_{\text{global}}^{(j-1)}$ **then** $\boldsymbol{\mathcal{X}}_k^{\text{POS}} \leftarrow \boldsymbol{\mathcal{X}}_k^{(j)}$.
19: **end if**
20: **end for**
21: Find $\boldsymbol{\Psi}$: $\{\mathbf{P}_k^{\text{POS}}, J_k^{\text{POS}}, \beta_k\}_{k=1}^{K}$.
22: ▷ combinatorial resource reallocation mechanism
23: **Until** $|\mathcal{H}_{\text{global}}^{j} - \mathcal{H}_{\text{global}}^{j-1}| \leq \epsilon$. ▷ Euclidean distances
24: Determine $\mathcal{H}_{\text{global}}^{\text{POS}}$: $\{\mathcal{H}(\mathbf{U}_k^{\text{POS}}, \mathbf{V}_k^{\text{POS}}, \mathbf{P}_k^{\text{POS}}, J_k^{\text{POS}})\}_{k=1}^{K}$ from the last fronts $\{\boldsymbol{\mathcal{X}}_k^{\text{POS}}\}_{k=1}^{K}$.
25: Optimal beam redesign: $\{\mathbf{F}_k^{\text{POS}}, \mathbf{W}_k^{\text{POS}}\}_{k=1}^{K}$.
26: ▷ globally optimal trade-off

distances among different Pareto fronts. In our work, ϵ is used to compute the minimum Euclidean distance of the objective function from chosen solutions towards Pareto front. The smaller this metric is, the better the convergence toward the Pareto front will be.

As shown in Algorithm 1, NSGA-II starts with an initial population $\boldsymbol{\mathcal{X}}_P^{(0)} = \{\boldsymbol{\mathcal{X}}_k^{\star}\}_{k=1}^{K}$, corresponding to the locally optimal SE-EE trade-off $\mathcal{H}_{\text{global}}^{(0)}$ (c.f. line 6 of Algorithm 1). The population then evolves toward nearby globally optimal SE-EE trade-off, i.e., sorting the best non-dominated solution $\{\boldsymbol{\mathcal{X}}_k^{\text{POS}}\}_{k=1}^{K}$, through subsequent iterations, called generations. Finding the optimal solution to (37) toward the Pareto front involves three key steps for the decision making process, given as follows:

Subprocedure: an instantiation of NSGA-II evoluatary subroutine for solving (37):

1) Create an offspring population $\boldsymbol{\mathcal{X}}_Q^{(i)} = \{\nabla \boldsymbol{\mathcal{X}}_k^{(i)}\}_{k=1}^{K}$ by a mutation operator ∇ with different variable step sizes and form a combined population $\boldsymbol{\mathcal{X}}_R^{(i)} = \boldsymbol{\mathcal{X}}_P^{(i)} \cup \boldsymbol{\mathcal{X}}_Q^{(i)}$. Classify all non-dominated solutions $\boldsymbol{\mathcal{X}}_R^{(i)}$ into different dominance sets $\{\boldsymbol{\mathcal{Y}}_l\}_{l=1}^{l}$, where $\boldsymbol{\mathcal{Y}}_l$ corresponds to the best solutions, given $\boldsymbol{\mathcal{Y}}^{(i)} \succ \boldsymbol{\mathcal{Y}}^{(i+1)}$.
2) Sort the members of each set $\boldsymbol{\mathcal{Y}}_l^{(i)}$ within the same dominance level, according to the descending order of crowding distance, while those solutions located in a lesser crowded region are preferred.
3) Select a better subset of non-dominated solutions from $\{\boldsymbol{\mathcal{Y}}_l\}_{l=1}^{l}$ to form a new population $\boldsymbol{\mathcal{X}}_P^{(i+1)}$ while preserving an already found $\boldsymbol{\mathcal{X}}_k^{\text{POS}}$. Repeat until satisfying the stop criteria (c.f. line 19, Algorithm 1).

C. Complexity Analysis

The computational complexity of Algorithm 1 comes from the size of decision space $\mathcal{M}$, i.e., the number of objectives and the size of decision vectors. According to the description in Section IV, the objective space $\boldsymbol{\mathcal{F}}$ is transformed into a single globally optimal problem. The overall computational complexity of the algorithm is within an acceptable range. To search for the optimal $\boldsymbol{\mathcal{X}}^{\text{POS}}$, the time and space complexity of NSGA-II is related to the dimension of the feasible variable set. It mainly depends on two aspects: (1) the optimization of HAD beamforming, and (2) the number of simultaneously transmitting and receiving beams in the beamspace. It is important to mention the fact that, in the case of the proposed algorithm, we are able to independently optimize the digital parts of HAD beamforming, which is dominated by calculating all of the optimal $\mathbf{U}_{n,k}^{\star}$. By this decoupled optimization operation, the first-generation non-dominated $\{\boldsymbol{\mathcal{X}}_k^{(0)}\}_{k=1}^{K}$ can be obtained from initial user finding and beam training procedure. In mmWave communications, a fundamental limit is that the number of transmission beams cannot be larger than the number of RF chains. Thus, the number of simultaneous transmission users is limited. For this reason, the search space can be further reduced, and thereby leads to fast convergence. On the other hand, it has been proved that NSGA-II with $\mathcal{O}(MN^2)$ computational complexity is a fast and elitist multi-objective genetic algorithm. Most results indicate that it is superior to the genetic algorithm in terms of both its accuracy and convergence rate in the obtained non-dominated solution [62]. In our proposed approach, the size of search space is effectively shrunken. Therefore, it is feasible to search for the globally Pareto-optimal solutions at an affordable complexity.

V. Numerical Results

A. Simulation Settings

In the considered simulation scenario, a cellular mmWave network with the HAD architecture is considered. The azimuth angles are assumed to be uniformly distributed over $[0; \pi]$, and the AoA/AoD elevation angles are uniformly distributed over $[-\pi/2; \pi/2]$. As shown in Section IV-A, the users are independent and uniformly distributed abiding a spatial PPP. Correspondingly, the user own expected throughput is independently characterized by the PPP intensity λ. To simulate strong users and weak users, the desirable users are chosen from 100 user samples, which are all generated by an independent two-dimensional homogeneous PPP. The performance evaluation is carried out through Monte Carlo simulations, and each result is the average of 100 independent realizations. Finally,

TABLE I
SYSTEM CONFIGURATION AND SIMULATION SETTING PARAMETERS

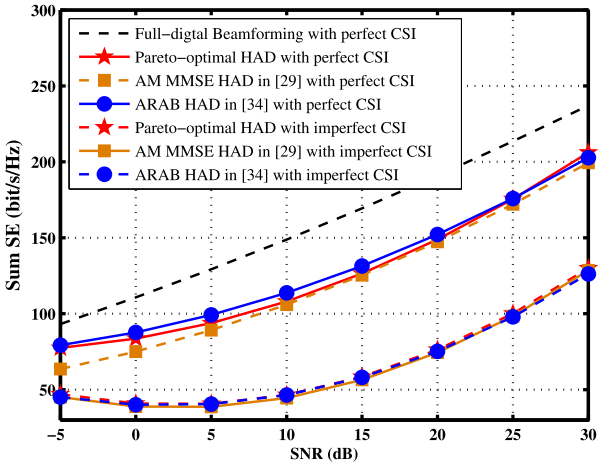
Parameters	Value	Parameters	Value
Carrier frequency	28 GHz	K	8
Bandwidth	200 MHz	Maximum J_k at BS/MS	16/4
Number of subcarriers	32	N_t	256
Symbol duration	3.7 μs	$N_{r,k}$ with per-user	16
Cell radius	200 m	$P_{k,t}^{\max}$	38 dBm
Modulation	QPSK	$P_{\mathrm{RF},k}$	250 mW [6]
Number of channel paths	4	$P_{c,k}$	200 mW
Receiver Noise	-174 dBm/Hz	P_{shift}	88 mW [28]

the convergence behavior of the proposed POS searching algorithm is investigated. Table I summarizes the simulation parameters and the experimental setup.

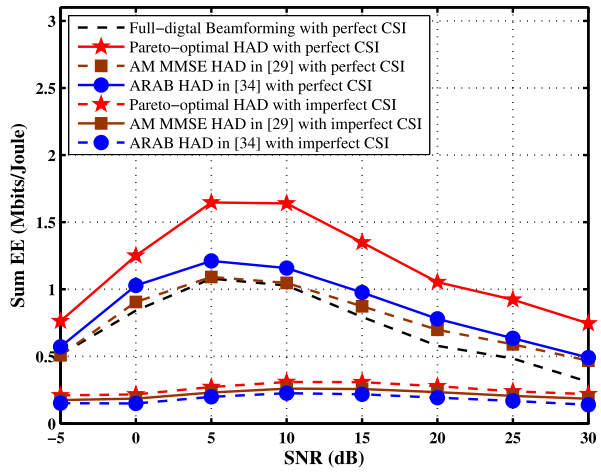
B. Discussion of Numerical Results

In this part, we present and discuss the detailed simulation results to evaluate the performance of the proposed optimization approach as compared to state-of-the-art association algorithms for both high and low SNR regimes. In Fig. 3(a) and 3(b), we evaluate the SE and EE performance in the high SNR regime by considering the impact of perfect and imperfect CSI, respectively. We compare the proposed Pareto-optimal HAD beamforming scheme with the approximation methods, including a decoupled two-stage HAD design proposed in [34] and an improved alternating minimization (AM) algorithm proposed in [29]. The work of [34] proposed an alternating optimization technique, called the ARAB algorithm, for multi-user massive MIMO system, where OFDM modulation is employed. In [29], the AM algorithm can be extended easily to advanced MIMO-OFDM configurations without changing the antenna setup at both the transmitter and the receiver. Therefore, the above algorithms considered for our experiments can be uniformly set to the MIMO-OFDM transmission mode for fair comparisons. For all cases, we can see that imperfect CSI greatly degrades the performance of SE. In Fig. 3(a), we notice that the SE performance of the algorithms is close when the perfect CSI is assumed. In Fig. 3(b), the first important note is that our proposed approach has a much higher EE than those of the algorithms proposed in [29], [34]. As expected, the full-digital beamforming achieves better SE performance, but has poorer EE performance.

Next, we consider the low SNR condition with the same system configurations. The experimental results for achievable SE and EE in the lower SNR regime are shown in Fig. 4(a) and 4(b), respectively. We can observe that except for the full-digital beamforming, the proposed approach with equal power allocation can significantly outperform its counterpart. That is, the proposed approach achieves higher SE and EE than the existing approaches designed in [29], [34]. For weak users, it is a competitive technology to provide a much higher data rate and a longer transmission distance. In addition, it can be observed that our proposed scheme has the ability to extend the coverage at the edge of the cell and make the cell-edge users obtain better performance. This is because for MIMO-OFDM-IM with power allocation, the subcarriers with high power can be scheduled for weak users. Even with a total transmit power constraint, the achievable sum rate in the low SNR regime is maximized by exploiting the systematic optimization of power allocation.



(a) Comparison of the maximum achievable SE at high SNR.

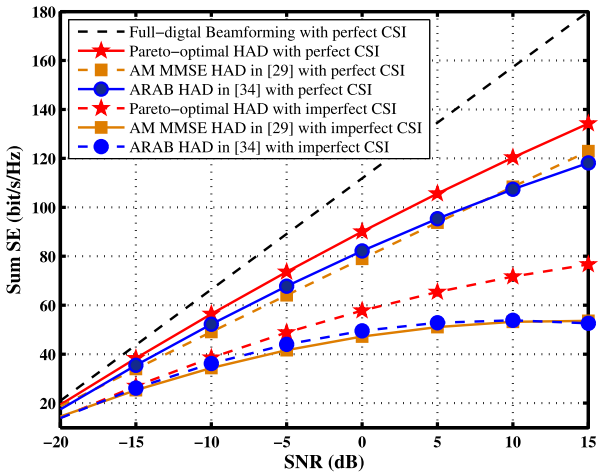


(b) Comparison of the EE performance at high SNR.

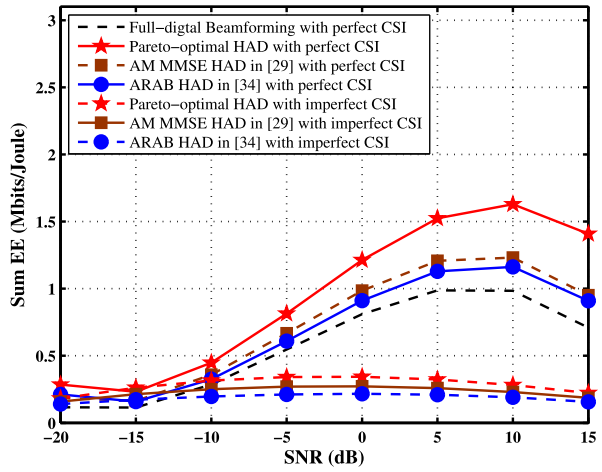
Fig. 3. SE vs. SNR and EE vs. SNR with perfect CSI and imperfect CSI, given $\varepsilon = 0.8$.

For any given individual user link, the relation of the SE and EE as a function of the ratio of α is depicted in Fig. 5. Here, the normalized weight factor ω represents the expected throughput of specific user, i.e., SE demand. Generally, we can determine the weight according to the actual traffic distribution by considering different throughputs. From this figure, as the weight ω grows from 0.6 to 0.9, the SE and EE achieved by the proposed SE-EE trade-off can adapt with the demand of throughput. After experimental comparison of these aforementioned algorithms, the results demonstrate that our proposed approach can explicitly control the SE and EE ratios among users. In particular, we find that the proposed Pareto-optimal beam design is efficient to improve EE performance while guaranteeing SE. It means that if we increase capacity demand of the user, Pareto-optimal trade-off can provide a balance between SE and EE. For the same reasons pointed out above, with the help of OFDM-IM, it allows a higher degree of freedom to achieve an adequate SE-EE trade-off. When the reduction in energy consumption is taken into account, an inverse relationship between SE and EE can be avoided. Our approach thereby achieves superiority over the traditional approaches.

In Fig. 6, we examine the sum transmission power yielded by different approaches with respect to assigning different



(a) Comparison of the maximum achievable SE at low SNR.



(b) Comparison of the EE performance at low SNR.

Fig. 4. SE vs. SNR and EE vs. SNR with perfect CSI and imperfect CSI, given $\varepsilon = 0.8$, where the proposed method is based on the equal power allocation scheme.

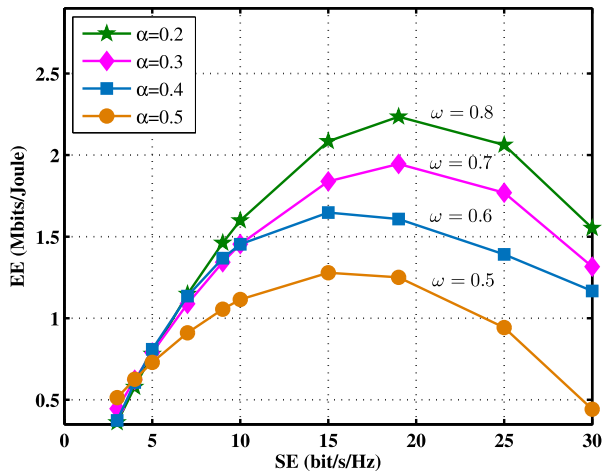


Fig. 5. SE vs. EE with locally optimal trade-off: A case of single-user SE-EE trade-off.

numbers of RF chains J_k at the BS. In order to characterize the proposed power allocation, the sum transmission power is investigated without considering SE-EE trade-off. The figure shows that the total transmit power increases with an increasing number of active RF chains based on the distributed number of users $K = (8, 4, 1)$. As described

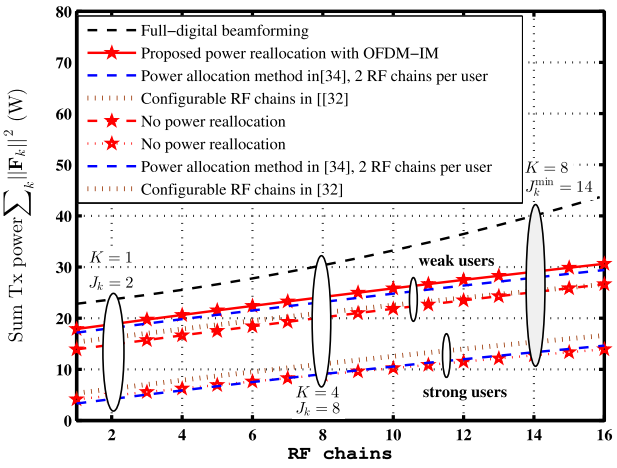


Fig. 6. Sum Tx power vs. the number of RF chains at the BS based on the optimal power adaptation scheme for strong users and weak users with different transmission power reallocation strategies.

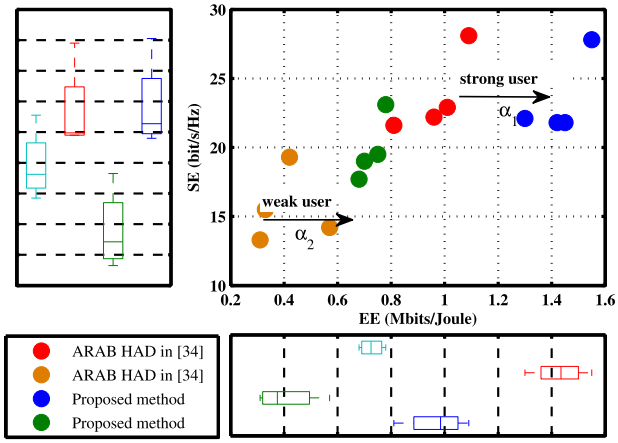


Fig. 7. SE vs. EE with globally optimal trade-off: A case of globally optimal SE-EE trade-off based on the globally POS, where $\alpha_1 = 0.4$ and $\alpha_2 = 0.6$ are assigned for strong users and weak users, respectively.

in (33), we perform the max-min power allocation strategy in the downlink of mmWave transmissions. In [32], the authors considered a flexible hybrid precoding scheme for mmWave MIMO communications, where the number of active RF chains can be adjusted to achieve the optimal EE. Similarly, with the proposed combinatorial-oriented beam design approach, we also prove that there exists a trade-off between SE and EE by varying the number of activated RF chains. For weak users, the proposed power reallocation scheme shows great superiority over those of the schemes proposed in [29], [32], [34] under a total power constraint. Besides, we can observe that in the case of $k = 8, J_k = 14$, the proposed power allocation algorithm and the approach proposed in [32] outperform their counterparts (requiring 16 RF chains) proposed in [29], [34]. From the perspective of degree of freedom introduced by MIMO-OFDM-IM, our approach is capable of efficiently utilizing downlink wireless resource as well as achieving a lower expected power consumption. This also implies that the proposed power control strategy can partially eliminate the limits of the SE-EE trade-off.

Fig. 7 shows the SE-EE Pareto front in terms of a particular MOP solution, in which this MOP with constraints is addressed by two possible solutions in the feasible region.

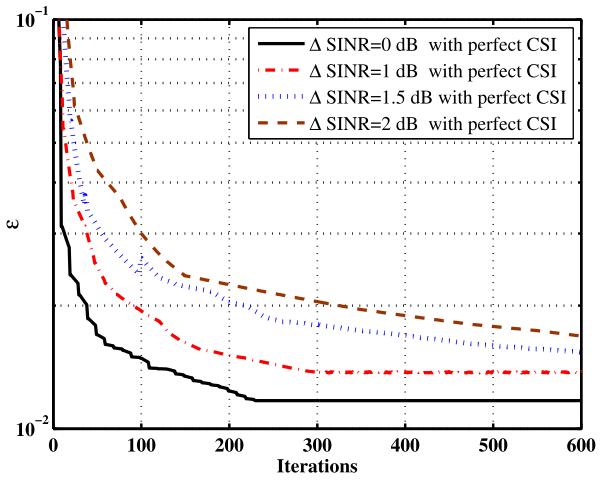


Fig. 8. Convergence property of using NSGA-II algorithm to enumerate all feasible POSs with the SINR threshold.

The green and blue points correspond to weak users and strong users that are identified on the Pareto front, respectively. As a case study, we consider a realistic scenario, which consists of four strong users and four weak users, and we aim to maximize the SE and EE, simultaneously. In particular, our goal is to find the globally POS by using the NSGA-II algorithm to enumerate all of the feasible POSs. In other words, deriving the globally optimal SE-EE trade-off is equivalent to finding the POS as close as possible to the Pareto front, i.e., $\boldsymbol{\mathcal{X}}^{\text{POS}}$. By comparing $\mathcal{H}_{\text{global}}$ for each solution $\boldsymbol{\mathcal{X}}^{\star}$ in the objective space, we can identify a Pareto-optimal solution $\boldsymbol{\mathcal{X}}^{\text{POS}}$ and thereby make a final decision on the SE-EE trade-off with $\boldsymbol{\mathcal{X}}^{\text{POS}}$. A key observation of Fig. 7 is that by properly adjusting weighting coefficients α, the performance of SE and EE can be significantly improved, which fully verifies the theoretical analysis. It is obvious that the SE-EE improvement achieved by the proposed approach is clearly superior to the one given in [34]. Here, we show a globally optimal SE-EE trade-off. As can be observed from the figure, the blue points can be regarded as one of possible POS to improve the EE of strong users, and the green points represent another POS to maximize both SE and EE for weak users. Once both SE and EE are maximized, these points can be approximately viewed as Pareto front.

Finally, the convergence behavior of NSGA-II based optimization strategy with respect to different SINRs is shown in Fig. 8. The metric ϵ is iteratively computed until a convergence threshold is reached, resulting in the globally POS. It is seen that as the iteration number increases, a sufficiently small ϵ can be achieved. After 100 iterations, the objective values are close to the final converged results. We examine the effect of the SINR threshold and find that the rate of convergence seems to be very sensitive to the SINR factor. Note that when the improvement of Δ SINR $= 0$, it actually means that no SINR constraint is imposed. In this case, the proposed algorithm can provide the highest convergence rate with the smallest ϵ, but SE cannot increase significantly in the absence of SINR improvement. From this figure, it is shown that in the best case, the algorithm can achieve a 2 dB SINR improvement, whilst the convergence attained at this optimal seems to be relatively slow, requiring about 500 iterations. Intuitively, this phenomenon coincides with the realistic situation. That is, the larger SINR improvement would introduce a higher degree of complexity or uncertainty and requires more steps to obtain the stationary point of the globally POS. By user discovery and initial access procedures, we can run our algorithm with different initial points independently and allows us to adaptively handle the SE-EE optimization problem with tractable solutions.

VI. Conclusion

In this paper, an energy-efficient mmWave MIMO-OFDM-IM system with the HAD beamforming architecture was proposed and investigated. We provided the optimal solution that allows a higher degree of freedom to achieve realistic SE-EE maximization in mmWave cellular networks. We have given a baseline design to solve the SE-EE trade-off for mmWave MIMO-OFDM-IM systems. The key finding of this study is that the use of Pareto-optimal beam design can achieve a globally optimal trade-off between SE and EE, and the collision constraints of MOP can be efficiently released. Also, the flexible power reallocation scheme can significantly extend the coverage for the cell-edge users. It is expected that the density of the BS may have a more significant impact on the network SE and/or spatial SE, as well as EE scales with the user density. In future work, we will consider a wide range of conditions as well as the equivalent form of the SE-EE trade-off, such as combining some fundamental results from random matrix theory.

Acknowledgment

The authors would like to thank the anonymous reviewers for many valuable comments that considerably improved this article.

References

[1] Y. Yang, S. Dang, M. Wen, and M. Guizani, "MmWave MIMO-OFDM with index modulation: A Pareto-optimal trade-off on spectral-energy efficiency," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Taipei, Taiwan, Dec. 2020, pp. 1–6.

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

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

[4] F. Sohrabi and W. Yu, "Hybrid digital and analog beamforming design for large-scale antenna arrays," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 501–513, Apr. 2016.

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

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

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

[8] X. Yu, J.-C. Shen, J. Zhang, and K. B. Letaief, "Alternating minimization algorithms for hybrid precoding in millimeter wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 485–500, Apr. 2016.

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

[10] Y. Li, B. Bakkaloglu, and C. Chakrabarti, "A comprehensive energy model and energy-quality evaluation of wireless transceiver front-ends," in *Proc. IEEE Workshop Signal Process. Syst. Design Implement.*, Athens, Greece, Nov. 2007, pp. 262–267.

[11] Z. Hasan, H. Boostanimehr, and V. K. Bhargava, "Green cellular networks: A survey, some research issues and challenges," *IEEE Commun. Surveys Tuts.*, vol. 13, no. 4, pp. 524–540, Nov. 2011.

[12] *Energy Efficiency Metrics of a Base Station Site*, Recommendation document ITU-T L.1350, Oct. 2016.

[13] A. Puglielli *et al.*, "Design of Energy- and cost-efficient massive MIMO arrays," *Proc. IEEE*, vol. 104, no. 3, pp. 586–606, Mar. 2016.

[14] E. Bjornson, M. Matthaiou, and M. Debbah, "Massive MIMO with non-ideal arbitrary arrays: Hardware scaling laws and circuit-aware design," *IEEE Trans. Wireless Commun.*, vol. 14, no. 8, pp. 4353–4368, Aug. 2015.

[15] H. Yan, S. Ramesh, T. Gallagher, C. Ling, and D. Cabric, "Performance, power, and area design trade-offs in millimeter-wave transmitter beamforming architectures," *IEEE Circuits Syst. Mag.*, vol. 19, no. 2, pp. 33–58, May 2019.

[16] M. Wen *et al.*, "A survey on spatial modulation in emerging wireless systems: Research progresses and applications," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 9, pp. 1949–1972, Sep. 2019.

[17] A. Zappone and E. Jorswieck, *Energy Efficiency in Wireless Networks via Fractional Programming Theory*. Boston, MA, USA: Now Foundations and Trends, 2015.

[18] L. Sboui, Z. Rezki, A. Sultan, and M.-S. Alouini, "A new relation between energy efficiency and spectral efficiency in wireless communications systems," *IEEE Wireless Commun.*, vol. 26, no. 3, pp. 168–174, Jun. 2019.

[19] C. Xiong, G. Y. Li, S. Zhang, Y. Chen, and S. Xu, "Energy- and spectral-efficiency tradeoff in downlink OFDMA networks," *IEEE Trans. Wireless Commun.*, vol. 10, no. 11, pp. 3874–3886, Nov. 2011.

[20] H. Quoc Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 61, no. 4, pp. 1436–1449, Apr. 2013.

[21] B. Xie, Z. Zhang, R. Q. Hu, G. Wu, and A. Papathanassiou, "Joint spectral efficiency and energy efficiency in FFR-based wireless heterogeneous networks," *IEEE Trans. Veh. Technol.*, vol. 67, no. 9, pp. 8154–8168, Sep. 2018.

[22] W. Hao, M. Zeng, Z. Chu, and S. Yang, "Energy-efficient power allocation in millimeter wave massive MIMO with non-orthogonal multiple access," *IEEE Wireless Commun. Lett.*, vol. 6, no. 6, pp. 782–785, Dec. 2017.

[23] X. Xiao, X. Tao, and J. Lu, "QoS-aware energy-efficient radio resource scheduling in multi-user OFDMA systems," *IEEE Commun. Lett.*, vol. 17, no. 1, pp. 75–78, Jan. 2013.

[24] D. W. K. Ng, E. S. Lo, and R. Schober, "Energy-efficient resource allocation in multi-cell OFDMA systems with limited backhaul capacity," *IEEE Trans. Wireless Commun.*, vol. 11, no. 10, pp. 3618–3631, Oct. 2012.

[25] J. Tang, D. K. C. So, E. Alsusa, and K. A. Hamdi, "Resource efficiency: A new paradigm on energy efficiency and spectral efficiency tradeoff," *IEEE Trans. Wireless Commun.*, vol. 13, no. 8, pp. 4656–4669, Aug. 2014.

[26] S. Han, I. Chih-Lin, Z. Xu, and C. Rowell, "Large-scale antenna systems with hybrid analog and digital beamforming for millimeter wave 5G," *IEEE Commun. Mag.*, vol. 53, no. 1, pp. 186–194, Jan. 2015.

[27] X. Gao, L. Dai, S. Han, I. Chih-Lin, and R. W. Heath, "Energy-efficient hybrid analog and digital precoding for mmWave MIMO systems with large antenna arrays," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 998–1009, Apr. 2016.

[28] C. G. Tsinos, S. Maleki, S. Chatzinotas, and B. Ottersten, "On the energy-efficiency of hybrid analog–digital transceivers for single-and multi-carrier large antenna array systems," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 1980–1995, Sep. 2017.

[29] J. P. Gonzalez-Coma, J. Rodriguez-Fernandez, N. Gonzalez-Prelcic, L. Castedo, and R. W. Heath, "Channel estimation and hybrid precoding for frequency selective multiuser mmWave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 2, pp. 353–367, May 2018.

[30] L. N. Ribeiro, S. Schwarz, M. Rupp, and A. L. F. de Almeida, "Energy efficiency of mmWave massive MIMO precoding with low-resolution DACs," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 2, pp. 298–312, May 2018.

[31] N. N. Moghadam, G. Fodor, M. Bengtsson, and D. J. Love, "On the energy efficiency of MIMO hybrid beamforming for millimeter-wave systems with nonlinear power amplifiers," *IEEE Trans. Wireless Commun.*, vol. 17, no. 11, pp. 7208–7221, Nov. 2018.

[32] Z. Zheng and H. Gharavi, "Spectral and energy efficiencies of millimeter wave MIMO with configurable hybrid precoding," *IEEE Trans. Veh. Technol.*, vol. 68, no. 6, pp. 5732–5746, Jun. 2019.

[33] J. Zhang, Y. Huang, J. Wang, R. Schober, and L. Yang, "Power-efficient beam designs for millimeter wave communication systems," *IEEE Trans. Wireless Commun.*, vol. 8, no. 6, pp. 1319–1333, Nov. 2019.

[34] K. Ardah, G. Fodor, Y. C. B. Silva, W. C. Freitas, and A. L. F. de Almeida, "Hybrid analog-digital beamforming design for SE and EE maximization in massive MIMO networks," *IEEE Trans. Veh. Technol.*, vol. 69, no. 1, pp. 377–389, Jan. 2020.

[35] E. Bjrnson and E. Jorswieck, *Optimal Resource Allocation in Coordinated Multi-Cell Systems*. New York, NY, USA: Now, vol. 9, nos. 2–3. 2013, p. 381.

[36] E. Björnson, E. A. Jorswieck, M. Debbah, and B. Ottersten, "Multiobjective signal processing optimization: The way to balance conflicting metrics in 5G systems," *IEEE Signal Process. Mag.*, vol. 31, no. 6, pp. 14–23, Nov. 2014.

[37] J. Zhang *et al.*, "Aeronautical *Ad Hoc* networking for the Internet-above-the-clouds," *Proc. IEEE*, vol. 107, no. 5, pp. 868–911, May 2019.

[38] M. D. Renzo, A. Zappone, T. T. Lam, and M. Debbah, "System-level modeling and optimization of the energy efficiency in cellular networks—A stochastic geometry framework," *IEEE Trans. Wireless Commun.*, vol. 17, no. 4, pp. 2539–2556, Apr. 2018.

[39] R. Hu and T.-M. Lok, "Pareto optimality for the single-stream transmission in multiuser relay networks," *IEEE Trans. Wireless Commun.*, vol. 16, no. 10, pp. 6503–6513, Oct. 2017.

[40] R. Abu-alhiga and H. Haas, "Subcarrier-index modulation OFDM," in *Proc. IEEE 20th Int. Symp. Pers., Indoor Mobile Radio Commun.*, Tokyo, Japan, Sep. 2009, pp. 177–181.

[41] D. Tsonev, S. Sinanovic, and H. Haas, "Enhanced subcarrier index modulation (SIM) OFDM," in *Proc. IEEE GLOBECOM Workshops (GC Wkshps)*, Houston, HI, USA, Dec. 2011, pp. 728–732.

[42] T. F. Rahman, A. Habib, C. Sacchi, and M. El-Hajjar, "Mm-wave STSK-aided single carrier block transmission for broadband networking," in *Proc. IEEE Symp. Comput. Commun. (ISCC)*, Heraklion, Greece, Jul. 2017, pp. 507–514.

[43] S. Dang, J. P. Coon, and G. Chen, "Adaptive OFDM with index modulation for two-hop relay-assisted networks," *IEEE Trans. Wireless Commun.*, vol. 17, no. 3, pp. 1923–1936, Mar. 2018.

[44] S. Dang, J. Li, M. Wen, and S. Mumtaz, "Distributed processing for multi-relay assisted OFDM with index modulation," *IEEE Trans. Wireless Commun.*, vol. 18, no. 2, pp. 1318–1331, Feb. 2019.

[45] E. Başar, Ü. Aygölü, E. Panayirci, and H. Vincent Poor, "Orthogonal frequency division multiplexing with index modulation," *IEEE Trans. Signal Process.*, vol. 61, no. 22, pp. 5536–5549, Nov. 2013.

[46] E. Basar, "On multiple-input multiple-output OFDM with index modulation for next generation wireless networks," *IEEE Trans. Signal Process.*, vol. 64, no. 15, pp. 3868–3878, Aug. 2016.

[47] M. Wen *et al.*, *Index Modulation for 5G Wireless Communications*. Cham, Switzerland: Springer, 2017.

[48] E. P. Simon, L. Ros, H. Hijazi, J. Fang, D. P. Gaillot, and M. Berbineau, "Joint carrier frequency offset and fast time-varying channel estimation for MIMO-OFDM systems," *IEEE Trans. Veh. Technol.*, vol. 60, no. 3, pp. 955–965, Mar. 2011.

[49] M. Wen, X. Cheng, M. Ma, B. Jiao, and H. V. Poor, "On the achievable rate of OFDM with index modulation," *IEEE Trans. Signal Process.*, vol. 64, no. 8, pp. 1919–1932, Apr. 2016.

[50] B. Zheng, M. Wen, E. Basar, and F. Chen, "Multiple-input multiple-output OFDM with index modulation: Low-complexity detector design," *IEEE Trans. Signal Process.*, vol. 65, no. 11, pp. 2758–2772, Jun. 2017.

[51] A. A. M. Saleh and R. Valenzuela, "A statistical model for indoor multipath propagation," *IEEE J. Sel. Areas Commun.*, vol. 5, no. 2, pp. 128–137, Feb. 1987.

[52] Y. Yang, S. Dang, M. Wen, S. Mumtaz, and M. Guizani, "Mobile millimeter wave channel tracking: A Bayesian beamforming framework against DOA uncertainty," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2019, pp. 1–6.

[53] Y. Yang, S. Dang, M. Wen, S. Mumtaz, and M. Guizani, "Bayesian beamforming for mobile millimeter wave channel tracking in the presence of DOA uncertainty," *IEEE Trans. Commun.*, vol. 68, no. 12, pp. 7547–7562, Dec. 2020.

[54] M. Haenggi, J. G. Andrews, F. Baccelli, O. Dousse, and M. Franceschetti, "Stochastic geometry and random graphs for the analysis and design of wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 7, pp. 1029–1046, Sep. 2009.

[55] S. Cui, A. J. Goldsmith, and A. Bahai, "Energy-efficiency of MIMO and cooperative MIMO techniques in sensor networks," *IEEE J. Sel. Areas Commun.*, vol. 22, no. 6, pp. 1089–1098, Aug. 2004.
[56] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, Mar. 2014.
[57] E. Karipidis, N. D. Sidiropoulos, and Z.-Q. Luo, "Quality of service and max-min fair transmit beamforming to multiple cochannel multicast groups," *IEEE Trans. Signal Process.*, vol. 56, no. 3, pp. 1268–1279, Mar. 2008.
[58] C. W. Tan, M. Chiang, and R. Srikant, "Maximizing sum rate and minimizing MSE on multiuser downlink: Optimality, fast algorithms and equivalence via max-min SINR," *IEEE Trans. Signal Process.*, vol. 59, no. 12, pp. 6127–6143, Dec. 2011.
[59] R. T. Marler and J. S. Arora, "Survey of multi-objective optimization methods for engineering," *Structural Multidisciplinary Optim.*, vol. 26, no. 6, pp. 369–395, Apr. 2004.
[60] J. B. Rao and A. O. Fapojuwo, "On the tradeoff between spectral efficiency and energy efficiency of homogeneous cellular networks with outage constraint," *IEEE Trans. Veh. Technol.*, vol. 62, no. 4, pp. 1801–1814, May 2013.
[61] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[62] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Trans. Evol. Comput.*, vol. 6, no. 2, pp. 182–197, Apr. 2002.

Yan Yang (Member, IEEE) received the B.Sc. degree in electronics engineering from the University of Electronic Science and Technology of China in 1990, and the M.Sc. degree in signal processing from Sichuan University in 1997, and the D.Sc. degree in signal processing from the Institute of Acoustic, Chinese Academy of Science, China, in 2004.

From 2014 to 2015, he was a Visiting Scholar with the Georgia Institute of Technology, Atlanta, USA. He is currently an Associate Professor with the State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing, China. His current research interests include wireless communications, signal processing, and machine learning for communications. He was a recipient of the Best Paper Award from the IEEE ComComAp'2019 and the First Research Award from the Science and Technology of China Railways Society in 2007 and 2014. He has served as a Reviewer for various journals, including IEEE INTERNET OF THINGS JOURNAL, IEEE WIRELESS COMMUNICATIONS LETTERS, IEEE ACCESS, *IEEE Network Magazine*, and *IEEE Wireless Communications Magazine*. He is also an active participant in the Working Party 5A (WP 5A) and the International Telecommunication Union (ITU), and a Technical Specialist for the Research Item 1.11 of the Resolution 236 World Radiocommunication Conference (WRC-15).

Shuping Dang (Member, IEEE) received the B.Eng. degree (Hons.) in electrical and electronic engineering from The University of Manchester and the B.Eng. degree in electrical engineering and automation from Beijing Jiaotong University in 2014 via a joint '2+2' dual-degree program, and the D.Phil. degree in engineering science from the University of Oxford in 2018. He joined the Research and Development Center, Huanan Communication Company Ltd., after graduating from the University of Oxford. He is currently working as a Post-Doctoral Fellow with the Computer, Electrical and Mathematical Science and Engineering Division, King Abdullah University of Science and Technology (KAUST). His current research interests include novel modulation schemes, cooperative communications, terahertz communications, and 6G wireless network design. He was a co-recipient of the Best Paper Award for work presented at 2019 19th IEEE International Conference on Communication Technology. He serves as a Reviewer for a number of key journals in communications and information science, including IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, IEEE TRANSACTIONS ON COMMUNICATIONS, IEEE WIRELESS COMMUNICATIONS LETTERS, IEEE COMMUNICATIONS LETTERS, and IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY. He is recognized as the Exemplary Reviewer of IEEE COMMUNICATIONS LETTERS in 2019 and 2020.

Miaowen Wen (Senior Member, IEEE) received the Ph.D. degree from Peking University, Beijing, China, in 2014.

From 2019 to 2021, he was with the Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong, as a Post-Doctoral Research Fellow. He is currently an Associate Professor with the South China University of Technology, Guangzhou, China. He has published two books and more than 110 journal articles. His research interests include a variety of topics in the areas of wireless and molecular communications. He was a recipient of the IEEE ComSoc Asia-Pacific Outstanding Young Researcher Award in 2020, and the Best Paper Award from the IEEE ITST'12, the IEEE ITSC'14, the IEEE ICNC'16, and the IEEE ICCT'19. He was the Winner of the Data Bakeoff Competition (Molecular MIMO) at the IEEE Communication Theory Workshop (CTW) 2019, Selfoss, Iceland. He served as a Guest Editor for the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS and the IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING. He is also serving as an Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS, the IEEE TRANSACTIONS ON MOLECULAR, BIOLOGICAL AND MULTI-SCALE COMMUNICATIONS, and the IEEE COMMUNICATIONS LETTERS, and a Guest Editor for the IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING (Special Issue on advanced signal processing for local and private 5G networks).

Mohsen Guizani (Fellow, IEEE) received the B.S. (Hons.) and M.S. degrees in electrical engineering and the M.S. and Ph.D. degrees in computer engineering from Syracuse University, Syracuse, NY, USA, in 1984, 1986, 1987, and 1990, respectively. He is currently a Professor with the Computer Science and Engineering Department, Qatar University, Qatar. Previously, he served in different academic and administrative positions at the University of Idaho, Western Michigan University, the University of West Florida, the University of Missouri-Kansas City, the University of Colorado Boulder, and Syracuse University. He is the author of nine books and more than 600 publications in refereed journals and conferences. His research interests include wireless communications and mobile computing, computer networks, mobile cloud computing, security, and smart grid. He is also a Senior Member of ACM. He also served as a member, the chair, and the general chair for a number of international conferences. He was a recipient of the 2017 IEEE Communications Society Wireless Technical Committee (WTC) Recognition Award, the 2018 AdHoc Technical Committee Recognition Award for his contribution to outstanding research in wireless communications and Ad Hoc sensor networks, and the 2019 IEEE Communications and Information Security Technical Recognition (CISTC) Award for outstanding contributions to the technological advancement of security. Throughout his career, he received three teaching awards and four research awards. He was the Chair of the IEEE Communications Society Wireless Technical Committee and the TAOS Technical Committee. He has guest edited a number of special issues in IEEE journals and magazines. He is also the Editor-in-Chief of the *IEEE Network*. He serves on the Editorial Board for several international technical journals, and the Founder and the Editor-in-Chief of *Wireless Communications and Mobile Computing* Journal (Wiley). He served as an IEEE Computer Society Distinguished Speaker and is also an IEEE ComSoc Distinguished Lecturer.