Energy-Efficient Power Allocation in Multi-User mmWave Systems With Rate-Splitting Multiple Access

Jared Everett*, Mohammad Reza Fasihi*, Igor Griva†, Brian L. Mark*
*Dept. of Electrical and Computer Engineering
†Dept. of Mathematical Sciences
George Mason University, Fairfax, Virginia, United States
Email: jared.everett@ieee.org, {mfasihi4, igriva, bmark}@gmu.edu

***Abstract*—We propose an energy-efficient power allocation algorithm for the multi-user millimeter-wave (mmWave) rate-splitting multiple access (RSMA) downlink with hybrid precoding and quality of service (QoS) constraints. The proposed scheme is applicable to the physical layer design of future wireless networks, such as the 6G cellular downlink, in which a transmitter equipped with multiple antennas must communicate unicast messages to multiple receivers simultaneously. First, we use a low-complexity design to define the analog and digital precoders in closed form. Second, we define an energy efficiency (EE) maximization problem to jointly optimize the power allocation among streams and the common stream rate allocation among users. We then solve the problem using a combination of Dinkelbach's algorithm and difference of convex functions (DC) programming methods. Simulation results show that the proposed RSMA scheme offers EE improvements over a comparable space division multiple access (SDMA) power allocation scheme in scenarios with perfect and imperfect channel state information at the transmitter. Lastly, we present extensive numerical experiments that suggest that the computational complexity of the proposed RSMA energy-efficient power allocation algorithm can be reduced using the interior-point method such that the computational efficiency of RSMA is comparable to that of SDMA.**

***Index Terms*—Rate-splitting multiple access (RSMA), millimeter wave (mmWave) communication, energy efficiency (EE), power allocation, hybrid precoding, 6G mobile communication**

I. Introduction

The International Telecommunications Union (ITU) identified sustainability as a foundational aspiration in their vision for International Mobile Telecommunications for 2030 and beyond (IMT-2030) [1], which will shape the development of sixth generation (6G) cellular networks and future wireless connectivity more broadly. This emphasis on sustainability brings into focus the importance of energy efficiency (EE) as a design metric in future physical layer (PHY) design. At the same time, rate-splitting multiple access (RSMA) has emerged in recent years as a powerful interference management strategy and multiple access scheme that has been proposed as a candidate technology for the 6G PHY [2]–[4]. RSMA offers a flexible framework that can generalize and softly bridge seemingly disparate multiple access schemes, including orthogonal multiple access (OMA), space division multiple access (SDMA), non-orthogonal multiple access (NOMA), and multicasting, while providing enhanced spectral efficiency (SE) and EE [2]. Ideally, a single multiple access scheme should be versatile enough to be used across all available spectrum in a future 6G system. Therefore, it is important to investigate whether RSMA can provide a flexible framework for 6G that is adaptable across all existing and future cellular bands. This includes millimeter-wave (mmWave) systems, which commonly use hybrid analog-digital precoding to reduce hardware cost and power consumption [5].

Initial research on the use of RSMA in mmWave systems with hybrid precoding has been promising. An RSMA strategy for the mmWave downlink with hybrid precoding was first proposed in [6], which showed that RSMA can be used to reduce channel training and feedback complexity. This strategy was later extended in [7] to a generalized mmWave RSMA hybrid precoding scheme for improved SE. In contrast, a low-complexity mmWave RSMA hybrid precoding scheme was proposed in [8], which achieved improved SE performance over SDMA in the presence of imperfect channel state information at the transmitter (CSIT) due to channel estimation error. The EE superiority of RSMA over SDMA and NOMA is well-established in sub-6 GHz spectrum [9], [10]. Recently, an energy-efficient mmWave RSMA hybrid precoding scheme was proposed in [11] that achieved higher EE performance than SDMA. All of these results suggest that RSMA is a promising technology in the mmWave regime.

Motivated by the EE superiority of RSMA in sub-6 GHz bands [9], [10] and recent results on the EE superiority of RSMA in mmWave hybrid precoding systems [11], this paper presents the first investigation of energy-efficient power allocation in the mmWave RSMA downlink. In this context, we first define a one-layer RSMA hybrid precoding scheme that uses a low-complexity, closed-form design for the analog and digital precoders. We then formulate the non-convex EE maximization problem to jointly optimize the power allocation among streams and the common stream rate allocation among

This work was supported in part by the U.S. National Science Foundation under Grants 2034616 and 2018631, and by resources provided by the Office of Research Computing at George Mason University (URL: https://orc.gmu.edu).

users. Based on this formulation, we propose a novel energy-efficient power allocation algorithm for RSMA that uses Dinkelbach's algorithm and difference of convex functions (DC) programming methods to achieve the optimal solution.

Importantly, we distinguish this work from prior work in [11], which studied a somewhat different EE problem in the mmWave RSMA downlink in which the entire digital precoder (i.e., both the direction and power allocation) was optimized using a successive convex approximation (SCA) approach. In contrast, the present paper assumes a low-complexity, closed-form design for the digital precoder and focuses on the problem of power allocation. This distinction is important from a computational complexity standpoint, as the energy-efficient power allocation algorithm proposed in the present paper has fewer optimization variables and is much more computationally efficient than the energy-efficient precoder optimization algorithm in [11]. Our simulation results show that the proposed RSMA scheme offers improvement in EE over a comparable space division multiple access (SDMA) power allocation scheme. Lastly, to further improve on computational efficiency, we present extensive numerical experiments that suggest that the computational complexity of the RSMA energy-efficient power allocation algorithm can be significantly reduced using the interior-point method [12].

The remainder of this paper is organized as follows. Sec. II introduces the system, channel, and power consumption models. Sec. III describes the RSMA hybrid precoder design. Sec. IV describes the EE maximization problem and proposed RSMA energy-efficient power allocation algorithm. Sec. V presents simulation results. Sec. VI gives concluding remarks.

Notation: Boldface uppercase and lowercase letters denote matrices and column vectors, respectively. Superscripts $(\cdot)^T$ and $(\cdot)^H$ denote transpose and Hermitian operators, respectively. The trace of a square matrix is denoted by $\mathrm{tr}(\cdot)$. The Euclidean norm of a vector is denoted by $\|\cdot\|$. The expectation operator is denoted by $\mathbb{E}[\cdot]$. The identity matrix is denoted by $\mathbf{I}$. The set of complex numbers is denoted by $\mathbb{C}$.

II. System Model

A. System Model

We consider a single-cell multi-user mmWave downlink system shown in Fig. 1 in which a base station (BS) communicates with K users in the user set $\mathcal{K} = \{1, ..., K\}$. The BS is equipped with N_{T} antennas and N_{RF} radio frequency (RF) chains satisfying $N_{\mathrm{RF}} < N_{\mathrm{T}}$. Each user is equipped with a single antenna. For simplicity, we make the common assumption that the number of users equals the number of RF chains, i.e., $K = N_{\mathrm{RF}}$ [5], [6], [11].

We study the one-layer RSMA hybrid precoding system architecture first proposed in [6]. A hybrid precoding scheme is used at the BS, comprising an analog RF precoder $\mathbf{F}_{\mathrm{RF}} \in \mathbb{C}^{N_{\mathrm{T}} \times N_{\mathrm{RF}}}$ and a digital baseband precoder $\mathbf{F}_{\mathrm{BB}} \in \mathbb{C}^{N_{\mathrm{RF}} \times N_{\mathrm{S}}}$, where N_{S} is the number of streams to be transmitted. We consider a fully-connected architecture at the BS in which each RF chain is connected to all antennas through a group of phase shifters [5]. As in prior works [5]–[8], [11], we assume

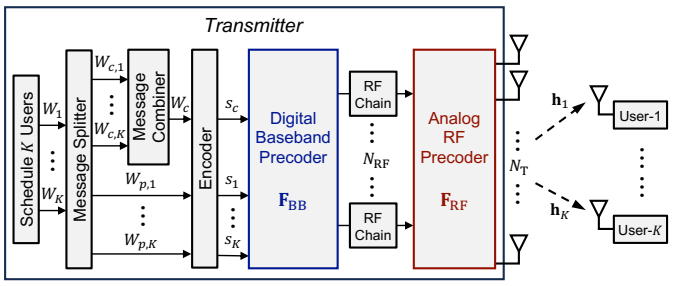

Fig. 1. Downlink multi-user mmWave RSMA system model [11].

the entries of the analog precoder have a constant modulus and quantized angle imposed by the phase shifting network, i.e., $[\mathbf{F}_{\mathrm{RF}}]_{m,n} = \frac{1}{\sqrt{N_{\mathrm{T}}}} e^{\varphi_{m,n}}$, where $\varphi_{m,n}$ is a quantized angle.

The BS communicates K messages, $W_1, ..., W_K$, to the K users. According to the one-layer rate-splitting strategy, the BS splits W_k, the message of user-k, into a common part and a private part, $W_{c,k}$ and $W_{p,k}$, respectively. The common parts of all K users' messages are combined into a single common message, $W_c = \{W_{c,1}, ..., W_{c,K}\}$, which is encoded into a common stream, s_c, using a shared codebook and decoded by all users. The private parts are independently encoded into private streams, $s_1, ..., s_K$, and decoded by their respective users. This results in $N_{\mathrm{S}} = K + 1$ streams at the input of the digital precoder. The streams are linearly precoded according to the hybrid precoding scheme. The resulting transmit signal is given by

$$\mathbf{x} = \sqrt{P_c}\mathbf{F}_{\mathrm{RF}}\mathbf{f}_{\mathrm{BB},c}s_c + \sum_{k\in\mathcal{K}} \sqrt{P_k}\mathbf{F}_{\mathrm{RF}}\mathbf{f}_{\mathrm{BB},k}s_k, \tag{1}$$

where P_c is the power allocation to the common stream, P_k is the power allocation to the private stream of user-k, $\mathbf{f}_{\mathrm{BB},j}$ is the digital precoder for stream j where $j \in \{c, 1, ..., K\}$, and $\mathbf{s} = [s_c, s_1, ..., s_K]^T$ is the stream vector for a given channel use. Assuming that $\mathbb{E}[\mathbf{s}\mathbf{s}^H] = \mathbf{I}$, the average transmit power constraint is written as

$$P_c + \sum_{k\in\mathcal{K}} P_k \le P_{\max}, \tag{2}$$

where $P_{\max}$ constrains the total average transmit power.

We adopt a narrowband block fading mmWave channel model as in [5]–[8], [11], [13]. The received signal at user-k is then written as

$$y_k = \underbrace{\sqrt{P_c}\mathbf{h}_k^H\mathbf{F}_{\mathrm{RF}}\mathbf{f}_{\mathrm{BB},c}s_c}_{\text{common stream}} + \underbrace{\sqrt{P_k}\mathbf{h}_k^H\mathbf{F}_{\mathrm{RF}}\mathbf{f}_{\mathrm{BB},k}s_k}_{\text{private stream}} + \underbrace{\sum_{j\in\mathcal{K},j\neq k} \sqrt{P_j}\mathbf{h}_k^H\mathbf{F}_{\mathrm{RF}}\mathbf{f}_{\mathrm{BB},j}s_j}_{\text{multi-user interference}} + \underbrace{n_k}_{\text{noise}}, \tag{3}$$

where $\mathbf{h}_k \in \mathbb{C}^{N_{\mathrm{T}}\times 1}$ is the channel vector from the BS to user-k, and $n_k \sim \mathcal{CN}(0, \sigma_{n,k}^2)$ is additive white Gaussian noise (AWGN) at user-k. Without loss of generality, we assume equal noise variance at the users, i.e., $\sigma_{n,k}^2 = \sigma_n^2$.

B. Channel Model

To model the mmWave channel, we adopt a geometric channel model with L scatterers, where each scatterer contributes a single propagation path from the BS to the user [5]–[8], [11], [13]. The channel between the BS and user-k is modeled as

$$\mathbf{h}_k = \sqrt{\frac{G_k N_{\mathrm{T}}}{L_k}} \sum_{l=1}^{L_k} \alpha_{k,l} \mathbf{a}(\theta_{k,l}), \tag{4}$$

where G_k is the large-scale channel gain to user-k due to pathloss and shadow fading, L_k is the number of paths to user-k, $\alpha_{k,l}$ is the complex gain of path-l, $\theta_{k,l} \in [0, \pi]$ is the azimuth angle-of-departure (AoD) of path-l, and $\mathbf{a}(\theta_{k,l}) \in \mathbb{C}^{N_{\mathrm{T}} \times 1}$ is the array steering vector of path-l. The path gains are independent and identically distributed (i.i.d.) $\alpha_{k,l} \sim \mathcal{CN}(0, 1)$ and vary independently across different time slots.

While the algorithms and results in this paper can be applied to arbitrary antenna arrays, we consider a uniform linear array (ULA) at the BS and azimuth beamforming for simplicity. The model can easily be extended to a 2D array. Under the plane wave and balanced narrowband array assumptions, the array steering vector is written as

$$\mathbf{a}(\theta_{k,l}) = \frac{1}{\sqrt{N_{\mathrm{T}}}} [1, e^{j\frac{2\pi}{\lambda} d \cos\theta_{k,l}}, ..., e^{j(N_{\mathrm{T}}-1)\frac{2\pi}{\lambda} d \cos\theta_{k,l}}]^T, \tag{5}$$

where λ is the wavelength and d is the antenna spacing [6]. We assume $d = \lambda/2$. This model is extended to study the case of imperfect CSIT in Sec. V.

C. Power Consumption Model

Following standard methods for EE analysis [9]–[11], [13], the total power consumption at the BS consists of two parts: the flexible transmit power and the fixed circuit power. Power consumption at the user side is omitted since it is negligible compared to the power consumption at the BS [9]. The total power consumption is modeled as [9]–[11], [13]

$$P_{\text{total}} = \psi \left(P_c + \sum_{k \in \mathcal{K}} P_k \right) + P_{\text{cir}}, \tag{6}$$

where $\psi \geq 1$ accounts for the inefficiency of the power amplifier and P_{cir} is the fixed circuit power consumption. We model the fixed circuit power of a mmWave BS as [13]

$$P_{\text{cir}} = P_{\text{BB}} + N_{\text{RF}} P_{\text{RF}} + N_{\text{RF}} N_{\text{T}} P_{\text{PS}} + N_{\text{T}} P_{\text{PA}}, \tag{7}$$

where P_{BB}, P_{RF}, P_{PS}, and P_{PA} denote the power consumption of the baseband, RF chain, phase shifter, and power amplifier, respectively.

III. Hybrid Precoding Scheme

To achieve a low-complexity design, our hybrid precoding scheme uses a conventional RF beamsteering design for the analog precoder and a closed-form design for the digital precoders. The digital precoders are normalized such that power allocation among streams can be designed subsequently in Sec. IV.

A. Analog Precoder

A conventional RF beamsteering codebook design is used for the analog precoder [5]–[7], [11]. Let $\mathcal{F}$ represent the RF beamsteering codebook, with cardinality $|\mathcal{F}| = Q = 2^{B_{\text{RF}}}$. We define the codebook as $\mathcal{F} = \{\mathbf{a}(\theta_q) | \theta_q = \frac{\pi q}{Q}, q \in [1, Q]\}$, where the codewords have the same form as the array steering vector. Codeword selection in practical systems can be achieved using an efficient beam search algorithm with feedback [6]. The BS searches through beams in the codebook, and user-k feeds back the index of the codeword that gives the maximum received power using B_{RF} feedback bits. The BS then sets the codeword of user-k as

$$\mathbf{f}_{\text{RF},k} = \underset{\mathbf{f}_{\text{RF},k} \in \mathcal{F}}{\arg\max} |\mathbf{h}_k^H \mathbf{f}_{\text{RF},k}|^2. \tag{8}$$

The overall analog precoder matrix is then given by $\mathbf{F}_{\text{RF}} = [\mathbf{f}_{\text{RF},1}, ..., \mathbf{f}_{\text{RF},K}]$.

B. Digital Precoder

Zero-forcing (ZF) precoding is used to design the digital precoders for the private streams. For a given $\mathbf{F}_{\text{RF}}$, we define the effective channel of the digital precoder from the BS to user-k as

$$\mathbf{h}_{\text{eff},k} = \mathbf{F}_{\text{RF}}^H \mathbf{h}_k. \tag{9}$$

The overall effective channel from the BS to all users is then

$$\mathbf{H}_{\text{eff}} = [\mathbf{h}_{\text{eff},1}, ..., \mathbf{h}_{\text{eff},K}]. \tag{10}$$

We generate the ZF precoding matrix by computing the Moore-Penrose pseudoinverse of the Hermitian of the effective channel matrix, $\mathbf{H}_{\text{eff}}^H$, which is

$$\mathbf{H}_{\text{eff}}^+ = \left(\mathbf{H}_{\text{eff}} \mathbf{H}_{\text{eff}}^H \right)^{-1} \mathbf{H}_{\text{eff}}. \tag{11}$$

The digital precoder for the private stream of user-k is then

$$\mathbf{f}_{\text{BB},k} = \frac{\mathbf{H}_{\text{eff}}^+(k)}{\|\mathbf{F}_{\text{RF}} \mathbf{H}_{\text{eff}}^+(k)\|}, \quad k \in \{1, ..., K\}, \tag{12}$$

where $\mathbf{H}_{\text{eff}}^+(k)$ denotes the k-th column of $\mathbf{H}_{\text{eff}}^+$. The normalization in (12) satisfies $\|\mathbf{F}_{\text{RF}} \mathbf{f}_{\text{BB},k}\|^2 = 1$. The ZF precoder projects the private streams of each user onto a subspace orthogonal to the one spanned by all other users' effective channel vectors [14]. Under the assumption of perfect CSIT, ZF precoding ensures that the private stream intended for user-k is received at the intended user without interference from the other users' private streams, i.e., $\|\mathbf{h}_k^H \mathbf{F}_{\text{RF}} \mathbf{f}_{\text{BB},j}\|^2 = 0, \forall j \neq k$.

Weighted matched beamforming (MBF) is used to design the digital precoder for the common stream. To compensate for differences in pathloss, each user channel is weighted by the inverse of the large-scale path gain. This definition ensures that the expression $\mathbf{h}_k^H \mathbf{F}_{\text{RF}} \mathbf{f}_{\text{BB},c}$ is proportional to equal weighted MBF of the normalized user channels, as employed in [15]. We define the digital precoder for the common stream as

$$\mathbf{f}_{\text{BB},c} = \beta \sum_{k \in \mathcal{K}} \frac{\mathbf{h}_{\text{eff},k}}{G_k}, \tag{13}$$

where $\beta = 1/\| \sum_{k \in \mathcal{K}} \mathbf{F}_{\text{RF}} \mathbf{h}_{\text{eff},k} / G_k \|$ is a normalization scalar to ensure $\|\mathbf{F}_{\text{RF}} \mathbf{f}_{\text{BB},c}\|^2 = 1$.

IV. Proposed Energy-Efficient Power Allocation

A. Problem Formulation

We now define the optimization problem to find the power allocation among streams and common stream rate allocation among users that maximizes the EE subject to quality of service (QoS) constraints. We assume the receivers perform successive interference cancellation (SIC), i.e., each user first decodes the common stream, s_c, while treating interference from all private streams as noise. The common stream message estimate is then re-encoded and subtracted from the received signal. The residual signal is then used to decode the private stream of the intended user.

After hybrid precoding, the signal-to-interference-plus-noise ratio (SINR) of the common stream at user-k is given by

$$\gamma_{c,k} = \frac{\left|\mathbf{h}_k^H \mathbf{F}_{\text{RF}} \mathbf{f}_{\text{BB},c}\right|^2 P_c}{\left|\mathbf{h}_k^H \mathbf{F}_{\text{RF}} \mathbf{f}_{\text{BB},k}\right|^2 P_k + \sigma_n^2}. \tag{14}$$

Due to ZF precoding of the private streams, the common stream of user-k only experiences interference from the private stream of user-k. Assuming perfect SIC, the SINR of the private stream at user-k is given by

$$\gamma_{p,k} = \frac{\left|\mathbf{h}_k^H \mathbf{F}_{\text{RF}} \mathbf{f}_{\text{BB},k}\right|^2 P_k}{\sigma_n^2}. \tag{15}$$

Denote $\rho_{j,k} = \left|\mathbf{h}_k^H \mathbf{F}_{\text{RF}} \mathbf{f}_{\text{BB},j}\right|^2 / \sigma_n^2$, where $j \in \{c, 1, ..., K\}$ and $k \in \{1, ..., K\}$. Assuming Gaussian signaling, the achievable rate of the common stream at user-k is given by

$$R_{c,k} = \log_2\left(1 + \frac{\rho_{c,k} P_c}{\rho_{k,k} P_k + 1}\right). \tag{16}$$

To ensure that the common stream can be decoded by all the users, the common stream rate is limited to

$$R_c = \min\left(R_{c,1}, ..., R_{c,K}\right) \geq \sum_{k \in \mathcal{K}} C_k, \tag{17}$$

where C_k denotes the common stream rate allocation for user-k. This definition of R_c ensures successful SIC at all the users, which supports the perfect SIC assumption in (15) above. The achievable rate of the private stream at user-k is

$$R_{p,k} = \log_2\left(1 + \rho_{k,k} P_k\right). \tag{18}$$

In general, the total rate of user-k is defined as $R_k = C_k + R_{p,k}$. Although all scheduled users are required to decode the common stream when it is present, each individual user may or may not receive part of its message from the common stream for a given transmission, i.e., $C_k = 0$ is possible. Likewise, the private stream of a given user may be turned off for a given transmission, i.e., $R_{p,k} = 0$ is possible.

As in prior work [9]–[11], [13], [16], we define the EE of the system as the sum rate, R_{sum}, divided by the total power consumption defined in (6). Thus, the EE is

$$\eta_{\text{EE}} = \frac{R_{\text{sum}}}{P_{\text{total}}} = \frac{\sum_{k \in \mathcal{K}} \left(C_k + R_{p,k}\right)}{\psi\left(P_c + \sum_{k \in \mathcal{K}} P_k\right) + P_{\text{cir}}}. \tag{19}$$

Finally, the EE maximization problem is formulated as

$$\begin{aligned}
\max_{\mathbf{P},\mathbf{C}} \quad & \eta_{\text{EE}} && \text{(20a)}\\
\text{s.t.} \quad & \textstyle\sum_{i \in \mathcal{K}} C_i \leq R_{c,k}, \quad \forall k \in \mathcal{K}, && \text{(20b)}\\
& C_k + R_{p,k} \geq R_k^{\text{th}}, \quad \forall k \in \mathcal{K}, && \text{(20c)}\\
& P_c + \textstyle\sum_{k \in \mathcal{K}} P_k \leq P_{\max}, && \text{(20d)}\\
& P_j \geq 0, \quad \forall j \in \{c, 1, ..., K\}, && \text{(20e)}\\
& C_k \geq 0, \quad \forall k \in \mathcal{K}, && \text{(20f)}
\end{aligned}$$

where $\mathbf{P} = \{P_c, P_1, ..., P_K\}$ and $\mathbf{C} = \{C_1, ..., C_K\}$. Constraints (20b) ensure the common stream can be decoded by all users, (20c) are the users' quality of service (QoS) constraints, (20d) is the power constraint, and (20e) and (20f) constrain the power allocations and common stream rate allocations to be non-negative. Without loss of generality, we assume that all users have the same QoS threshold rate, i.e., $R_k^{th} = R^{th}, \forall k \in \mathcal{K}$.

B. Proposed Solution

Problem (20) is a non-convex fractional program, which is hard to solve directly. Therefore, we adopt Dinkelbach's algorithm to transform the fractional problem into a series of subproblems in parametric subtractive form [17]. These subproblems are then solved iteratively using a DC programming method [18]. This process results in a two-layer algorithm with outer and inner iterations.

First, by transforming the objective function into parametric subtractive form, problem (20) is rewritten as

$$\begin{aligned}
\max_{\mathbf{P},\mathbf{C}} \quad & \sum_{k \in \mathcal{K}} \left(C_k + R_{p,k}\right) - \lambda^{(\kappa-1)} \left(\psi\left(P_c + \sum_{k \in \mathcal{K}} P_k\right) + P_{\text{cir}}\right)\\
\text{s.t.} \quad & \text{(20b)}, \text{(20c)}, \text{(20d)}, \text{(20e)}, \text{(20f)}, && \text{(21)}
\end{aligned}$$

where κ is the outer iteration index and $\lambda^{(\kappa)}$ is a non-negative parameter. Starting from $\lambda^{(0)} = 0$, this parameter is updated iteratively by $\lambda^{(\kappa)} = \eta_{\text{EE}}^{(\kappa)}$, where $\eta_{\text{EE}}^{(\kappa)}$ is the solution to (19) using the updated values of $\mathbf{P}$ and $\mathbf{C}$ after solving (21) for outer iteration κ. Defining $\varepsilon^{\star(\kappa)}$ as the maximum value of the objective function after solving (21) for iteration κ, it is shown in [17] that $\lambda^{(\kappa)}$ increases while $\varepsilon^{\star(\kappa)}$ decreases with each iteration. When $\varepsilon^{\star(\kappa)} = 0$, $\lambda^{(\kappa)}$ is maximized, which is also the maximum EE solution to the original problem (20). This forms the outer iteration of our proposed algorithm.

Next, we address the subproblem of how to solve (21) for a given $\lambda^{(\kappa)}$. The challenge lies in the non-convex constraint (20b), since the objective and the remaining constraints are concave. Constraint (20b) can be equivalently expressed as

$$\begin{aligned}
\sum_{i \in \mathcal{K}} C_i \leq \min_{k \in \mathcal{K}} \{ & \log_2(\rho_{c,k} P_c + \rho_{k,k} P_k + 1)\\
& - \log_2(\rho_{k,k} P_k + 1)\}. \qquad \text{(22)}
\end{aligned}$$

We create a common term inside the minimization for all k users. This term can be moved outside the minimization, which yields

$$\sum_{i\in\mathcal{K}} C_i \le \min_{k\in\mathcal{K}}\Big\{\log_2(\rho_{c,k}P_c+\rho_{k,k}P_k+1) + \sum_{j\neq k}\log_2(\rho_{j,j}P_j+1)\Big\} - \sum_{j=1}^{K}\log_2(\rho_{j,j}P_j+1). \quad (23)$$

This constraint is equivalently expressed by the following K per-user constraints

$$\sum_{i\in\mathcal{K}} C_i \le \underbrace{\log_2(\rho_{c,k}P_c+\rho_{k,k}P_k+1)+\sum_{j\neq k}\log_2(\rho_{j,j}P_j+1)}_{f_{1,k}(\mathbf{P})} - \underbrace{\sum_{j=1}^{K}\log_2(\rho_{j,j}P_j+1)}_{f_2(\mathbf{P})}, \quad \forall k\in\mathcal{K}, \quad (24)$$

where both functions $f_{1,k}(\mathbf{P})$ and $f_2(\mathbf{P})$ are concave. Thus, the expression $f_{1,k}(\mathbf{P})-f_2(\mathbf{P})$ is a DC function. Furthermore, we note that $f_2(\mathbf{P})$ is common to all K constraints in (24).

The DC programming method from [18] is used to generate a sequence $\{\mathbf{P}^{(\delta)}\}$ of improved feasible solutions for inner iteration δ. The gradient of $f_2(\mathbf{P})$ at $\mathbf{P}$ is given by

$$\nabla f_2(\mathbf{P}) = \frac{1}{\ln 2}\left[\frac{\rho_{1,1}}{\rho_{1,1}P_1+1},...,\frac{\rho_{K,K}}{\rho_{K,K}P_K+1}\right]^T. \quad (25)$$

Initialized from a feasible $\{\mathbf{P}^{(0)}\}$, $\{\mathbf{P}^{(\delta)}\}$ is obtained as the optimal solution to the following problem at inner iteration δ:

$$\begin{aligned}\max_{\mathbf{P},\mathbf{C}} \quad & \sum_{k\in\mathcal{K}}(C_k+R_{p,k}) - \lambda^{(\kappa-1)}\left(\psi(P_c+\sum_{k\in\mathcal{K}}P_k)+P_{\text{cir}}\right)\\ \text{s.t.} \quad & \textstyle\sum_{i\in\mathcal{K}} C_i \le f_{1,k}(\mathbf{P}) - f_2(\mathbf{P}^{(\delta-1)})\\ & \quad - \langle\nabla f_2(\mathbf{P}^{(\delta-1)}), \mathbf{P}-\mathbf{P}^{(\delta-1)}\rangle,\ \forall k\in\mathcal{K},\\ & (20c),(20d),(20e),(20f), \end{aligned} \quad (26)$$

where $\langle\cdot,\cdot\rangle$ denotes the inner product. Problem (26) is convex and can be efficiently solved using commercial solvers. This forms the inner iteration of our proposed algorithm.

Details of the energy-efficient power allocation algorithm for our proposed mmWave RSMA downlink scheme are summarized in Algorithm 1. For the inner iteration, the convergence of the DC programming method to the global optimal solution was shown in [18]. For the outer iteration, i.e., the fractional programming, convergence to the stationary and optimal solution was shown in [17]. Therefore, the proposed algorithm always converges to the optimal solution.

V. Simulation Results

In this section, we simulate the proposed mmWave RSMA scheme and compare it to SDMA in an outdoor urban small cell deployment scenario. System parameters common to all simulations are shown in Table I. Users are randomly placed in a cell with a minimum 2D distance of 10 m and a cell radius of 100 m following a uniform distribution by area. Large-scale path gains, G_k, use the pathloss, shadow fading, and line-of-sight (LOS) probability model from 3GPP TR 38.901 [19] for an urban microcell street canyon (UMi-SC) scenario assuming outdoor users at ground level, i.e., user antenna height of 1.5 m, and a BS antenna height of 10 m. Channels are randomly generated according to (4), where the AoDs are uniformly distributed over $[0,\pi]$. As in prior works on mmWave RSMA, we consider $L=15$ channel paths per user [6], [7], [11]. As in [6], [11], we assume full rank $\mathbf{F}_{\text{RF}}$, which is required to compute the ZF digital precoders for the private streams. A noise power spectral density of -174 dBm/Hz is assumed. Power consumption model parameters for a mmWave BS are taken from [13].

For a fair comparison, we define a baseline SDMA hybrid precoding scheme with energy-efficient power allocation as follows. The SDMA scheme uses the analog precoder defined in Sec. III-A and the ZF digital precoder defined in (12), equivalent to the RSMA private streams. The resulting EE maximization problem for SDMA is equivalent to (20) by setting $P_c=0$ and $C_k=0$, which makes constraints (20b) and (20f) unnecessary. This problem is then solved using Dinkelbach's algorithm, where the inner subproblem is convex.

Fig. 2 shows simulation results when perfect CSIT is assumed for scenarios with $K=2$ and $K=4$ users. All

Algorithm 1 Energy-Efficient Power Allocation Algorithm

Initialize $\varepsilon\leftarrow 10^{-6}, \lambda^{(0)}\leftarrow 0$
repeat {Outer iteration, κ}
  Initialize feasible power $\mathbf{P}^{(0)}$
  repeat {Inner iteration, δ}
    Solve (26) and denote the optimal values $\mathbf{P}^\star, \mathbf{C}^\star$
    Update $\mathbf{P}^{(\delta)}\leftarrow\mathbf{P}^\star$
  until $\mathbf{P}^{(\delta)}$ converges
  Compute $\varepsilon^{\star(\kappa)}\leftarrow$ $\sum_{k\in\mathcal{K}}(C_k+R_{p,k})-\lambda^{(\kappa-1)}(\psi(P_c+\sum_{k\in\mathcal{K}}P_k)+P_{\text{cir}})$
  Update $\lambda^{(\kappa)}\leftarrow\eta_{\text{EE}}^{(\kappa)}$
until $\varepsilon^{\star(\kappa)}\le\varepsilon$

TABLE I
System Parameters

Parameter	Value
Carrier Frequency	28 GHz
Bandwidth	50 MHz
Pathloss Model	3GPP UMi-SC
Number of BS Transmit Antennas (N_{T})	32
Number of BS Transmit RF Chains (N_{RF})	K
Number of Channel Paths Per User (L)	15
QoS threshold rate per user (R^{th})	0.5 bps/Hz
RF Codebook Feedback Bits (B_{RF})	5
PA Inefficiency Factor (ψ)	$1/0.38$
Baseband Power Consumption (P_{BB})	200 mW
RF Chain Power Consumption (P_{RF})	160 mW
PA Power Consumption (P_{PA})	40 mW
Phase Shifter Power Consumption (P_{PS})	20 mW

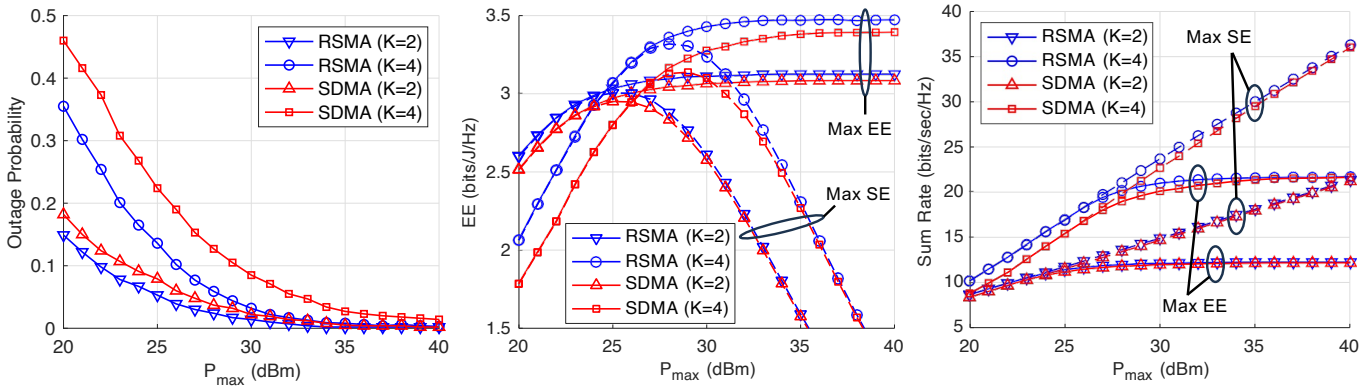

Fig. 2. Simulation results of RSMA and SDMA for $K = 2$ and $K = 4$ users with perfect CSIT vs. maximum transmit power constraint ($P_{\max}$): (a) Outage probability (p_{out}), (b) Weighted average EE (η_{EE}), (c) Weighted average SE (R_{sum}).

simulation results are averaged over 1000 channel realizations. These results help to show the trade-off between EE, SE, and QoS, the latter being closely related to outage probability.

Fig. 2a shows outage probability. We define an outage event to occur when the K per-user QoS constraints in (20c) cannot be met for a given $P_{\max}$. The outage probability is defined as

$$p_{\text{out}} = \Pr\left(\bigcup_{k\in\mathcal{K}} \{R_k < R^{\text{th}}\}\right), \tag{27}$$

where R_k is the total rate of user-k. This occurs when a solution in the feasible set of problem (20) is not found for a given scheme and channel realization. The outage probability is used to calculate the weighted average EE and weighted average SE in Fig. 2b and Fig. 2c, respectively. We observe from Fig. 2a that RSMA has smaller outage probability than SDMA. Thus, RSMA is better at meeting the prescribed QoS, particularly at low values of $P_{\max}$.

Since there are channel realizations for which an outage event occurs using the SDMA scheme but a solution exists using the RSMA scheme, a simple average of the feasible solutions is insufficient to capture the advantages of RSMA over SDMA. Therefore, we use a weighted average metric for EE and SE. The weighted average EE is computed as $(1 - p_{\text{out}})\overline{\eta_{\text{EE}}}$, where $\overline{\eta_{\text{EE}}}$ is the average EE over the channel realizations for which outage did not occur and a solution was obtained. The weighted average SE is computed similarly.

Fig. 2b and Fig. 2c show the weighted average EE and weighted average SE, respectively. We compare the EE performance of the proposed EE-maximized schemes ("Max EE") to the EE performance when the SE of the system is maximized ("Max SE"). The latter results are generated using the same algorithms as EE maximization, except we set $\psi = 0$ and $P_{\text{cir}} = 1$ so the objective function reduces to R_{sum}. We observe that the EE performance of the EE-maximized scheme is non-decreasing as $P_{\max}$ increases, which is consistent with our design. In contrast, the EE of the SE-maximized scheme decreases at high $P_{\max}$. This behavior occurs because the SE-maximized scheme uses all available power to maximize R_{sum}, which may sacrifice EE. In contrast, for the EE-maximized scheme, the BS transmits below the maximum power constraint, i.e., $(P_c + \sum_{k\in\mathcal{K}} P_k) < P_{\max}$, when $P_{\max}$ is large to maximize EE.

Furthermore, we also observe from Fig. 2b that the proposed EE-maximized RSMA scheme achieves higher EE than SDMA across the entire range of $P_{\max}$. This is also true for the SE-maximized scheme for small $P_{\max}$. The EE performance improvement of RSMA over SDMA is larger at small $P_{\max}$ and for $K = 4$ users, largely due to the QoS constraints and the higher outage probability of SDMA. Combined with Fig. 2c, these results show that RSMA is able to outperform SDMA in terms of both EE and SE in this regime. However, the SE performance of both schemes tends to converge at high $P_{\max}$, where conventional ZF precoding performs well.

Fig. 3 shows the impact of imperfect CSIT on the proposed RSMA scheme and baseline SDMA scheme for $K = 2$ users. The estimated channel matrix at the BS can be modeled as [16]

$$\hat{\mathbf{H}} = \xi\mathbf{H} + \sqrt{1-\xi^2}\mathbf{E}, \tag{28}$$

where $\mathbf{H} = [\mathbf{h}_1, ..., \mathbf{h}_K]$ is the actual channel matrix, $\xi \in [0, 1]$ is the CSIT accuracy with $\xi = 1$ representing perfect CSIT, and $\mathbf{E}$ is the error matrix with i.i.d. $\mathcal{CN}(0, 1)$ entries. In Fig. 3, the perfect CSIT scenario (i.e., $\xi = 1$) is compared to two scenarios with imperfect CSIT (i.e., $\xi = 0.9$ and $\xi = 0.8$). The EE performance of both RSMA and SDMA decreases as CSIT degrades. However, it is clear that the achievable EE performance of our proposed RSMA scheme is higher than SDMA under imperfect CSIT, which demonstrates the robustness and advantage of our proposed scheme.

Next, we investigate methods to reduce the computational complexity of the mmWave RSMA energy-efficient power allocation algorithm. To start, the reduced-complexity RSMA algorithm uses Dinkelbach's algorithm to address the fractional programming problem in (20). However, instead of applying the DC programming method, we solve the non-convex inner subproblem (21) directly using the well-known interior-point method [12] and a commercial solver (i.e., lines 4-7 in

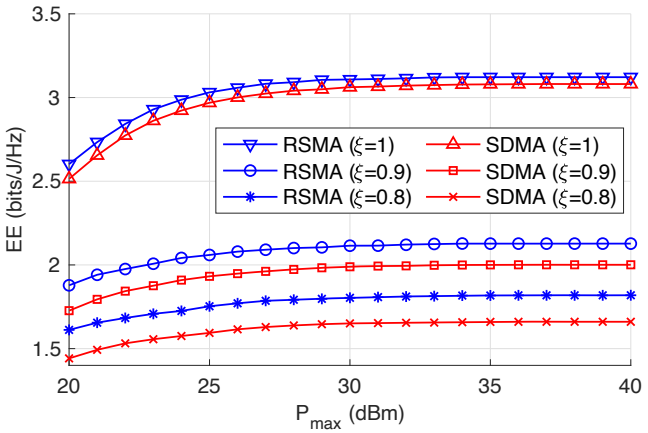

Fig. 3. Weighted average EE (η_{EE}) vs. maximum transmit power constraint ($P_{\max}$) of RSMA and SDMA for $K = 2$ users with imperfect CSIT.

TABLE II
MEAN / MEDIAN ALGORITHM RUN TIMES. PROBLEM INSTANCES WHERE NOT ALL ALGORITHMS OBTAINED A FEASIBLE SOLUTION ARE EXCLUDED.

	RSMA Alg. 1	**RSMA Int-pt.**	**SDMA Int-pt.**
K = 2	0.598 s / 0.415 s	0.022 s / 0.021 s	0.033 s / 0.021 s
K = 3	1.479 s / 0.690 s	0.025 s / 0.023 s	0.018 s / 0.017 s
K = 4	2.574 s / 1.051 s	0.026 s / 0.024 s	0.018 s / 0.017 s
K = 5	4.222 s / 1.479 s	0.030 s / 0.027 s	0.019 s / 0.017 s

Algorithm 1 are replaced by an interior-point algorithm). This approach has the advantage that it improves the computational efficiency of RSMA, but has the disadvantage that it may lead to a (sub-optimal) local maximum solution.

Extensive numerical experiments showed that the reduced-complexity RSMA algorithm performed comparably to Algorithm 1 in terms of EE. Algorithmic performance was evaluated by running 1000 random i.i.d. channel realizations each for normalized $P_{\max} \in \{25, 26, ..., 40\}$ dBm, $K \in \{2, 3, 4, 5\}$ users, and perfect CSIT. Mean and median computation times on a single core of an Intel Xeon Gold 6240R central processing unit (CPU) are reported in Table II. These results suggest that the RSMA energy-efficient power allocation problem in (21) is simple enough that a commercial solver using the interior-point method can find a near-optimal solution in practice. Furthermore, the computation time of RSMA was reduced by roughly one order of magnitude for $K = 2$ and two orders of magnitude for $K = 5$, making the computation time of RSMA similar to SDMA. Therefore, although it was not rigorously proven, we conclude from empirical observation that the EE gains of RSMA over SDMA demonstrated in this paper can likely be realized in practice without a significant increase in computational complexity.

VI. Conclusion

In this paper, we studied for the first time the problem of energy-efficient power allocation for the multi-user mmWave RSMA downlink with hybrid precoding and QoS constraints. Using low-complexity, closed-form design for the analog and digital precoders, we defined the RSMA EE maximization problem to jointly optimize the power allocation and common stream rate allocation. We then proposed a novel energy-efficient power allocation algorithm for RSMA that combined Dinkelbach's algorithm and DC programming methods to achieve the optimal solution. Simulation results showed that the EE of the proposed RSMA scheme exceeded that of SDMA in scenarios with perfect and imperfect CSIT. Lastly, we presented extensive numerical experiments that suggest that RSMA can achieve a computational efficiency comparable to that of SDMA using the well-known interior point method. Rigorous proof of this empirical observation is left as a topic for future investigation.

References

[1] *Framework and overall objectives of the future development of IMT for 2030 and beyond*, ITU Rec. ITU-R M.2160-0, Nov. 2023.

[2] Y. Mao, O. Dizdar, B. Clerckx, R. Schober, P. Popovski, and H. V. Poor, "Rate-splitting multiple access: Fundamentals, survey, and future research trends," *IEEE Commun. Surv. and Tut.*, vol. 24, no. 4, pp. 2073–2126, 4th Quart. 2022.

[3] B. Clerckx, Y. Mao, E. Jorswieck, J. Yuan, D. Love, E. Erkip, and D. Niyato, "A primer on rate-splitting multiple access: Tutorial, myths, and frequently asked questions," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 5, pp. 1265–1308, May 2023.

[4] O. Dizdar, Y. Mao, W. Han, and B. Clerckx, "Rate-splitting multiple access: A new frontier for the PHY layer of 6G," in *IEEE VTC-2020-Fall*, 2020.

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

[6] M. Dai and B. Clerckx, "Multiuser millimeter wave beamforming strategies with quantized and statistical CSIT," *IEEE Trans. Wireless Commun.*, vol. 16, no. 11, pp. 7025–7038, Nov. 2017.

[7] Z. Li, S. Yang, and T. Clessienne, "A general rate splitting scheme for hybrid precoding in mmWave systems," in *IEEE ICC*, 2019.

[8] O. Kolawole, A. Papazafeiropoulos, and T. Ratnarajah, "A rate-splitting strategy for multi-user millimeter-wave systems with imperfect CSIT," in *IEEE SPAWC*, 2018.

[9] Y. Mao, B. Clerckx, and V. Li, "Energy efficiency of rate-splitting multiple access, and performance benefits over SDMA and NOMA," in *15th Int. Symp. on Wireless Commun. Syst.*, 2018.

[10] ——, "Rate-splitting for multi-antenna non-orthogonal unicast and multicast transmission: Spectral and energy efficiency analysis," *IEEE Trans. Commun.*, vol. 67, no. 12, pp. 8754–8770, Dec. 2019.

[11] J. Everett and B. L. Mark, "Energy efficiency of multi-user mmWave rate-splitting multiple access with hybrid precoding," in *IEEE ICC Workshops*, 2024.

[12] R. Byrd, M. Hribar, and J. Nocedal, "An interior point algorithm for large-scale nonlinear programming," *SIAM Journal on Optimmization*, vol. 9, no. 4, pp. 877–900, 1999.

[13] 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, 2017.

[14] R. Heath Jr. and A. Lozano, *Foundations of MIMO Communication*. Cambridge, UK: Cambridge University Press, 2019.

[15] M. Dai, B. Clerckx, D. Gesbert, and G. Caire, "A rate splitting strategy for massive MIMO with imperfect CSIT," *IEEE Trans. Wireless Commun.*, vol. 15, no. 7, pp. 4611–4624, Jul. 2016.

[16] X. Gao, L. Dai, S. Han, C.-L. I, and R. 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.

[17] W. Dinkelbach, "On nonlinear fractional programming," *Management Science*, vol. 13, no. 7, pp. 492–498, Mar. 1967.

[18] H. Kha, H. Tuan, and H. Nguyen, "Fast global optimal power allocation in wireless networks by local D.C. programming," *IEEE Trans. Wireless Commun.*, vol. 11, no. 2, pp. 510–515, Feb. 2012.

[19] 3GPP TR 38.901, "Study on channel model for frequencies from 0.5 to 100 GHz," Release 18, ver. 18.0.0, Mar. 2024.