Passive Precoding and Power Allocation for Energy Efficient Reconfigurable Intelligent Surface Enhanced Millimeter Wave MU-MISO

Prabhat Raj Gautam, *Member, IEEE,* Li Zhang, *Senior Member, IEEE,* and Pingzhi Fan, *Fellow, IEEE*

***Abstract*—We consider narrowband downlink millimeter wave (mmWave) multiple-user multiple-input single-output (MU-MISO) communication assisted by several reconfigurable intelligent surfaces (RISs). Hybrid precoding is performed instead of conventional fully digital precoding to reduce the number of expensive and power-hungry radio frequency (RF) chains. Passive precoding. achieved by designing the reflecting coefficients of RIS, is performed together with transmit precoding and power allocation among different users to maximize energy efficiency of the system. By considering the use of zero-forcing (ZF) fully digital transmit precoder, the energy efficiency maximization problem is reduced to solving only passive precoding and power allocation which is achieved through alternating optimization technique. The passive precoding is attained through greedy optimization in phase, whereas power allocation is obtained by the method of Lagrange multipliers using Karush-Kuhn Tucker (KKT) conditions. The energy efficient passive precoding is further advanced through switching off of individual RIS elements accomplished through a low-complexity algorithm. Finally, hybrid ZF precoder is designed from the fully digital precoder. We show, through simulation results, that the proposed design delivers performance better than the existing solution, while entailing lower complexity.**

***Index Terms*—Reconfigurable Intelligent Surface (RIS), Millimeter wave (mmWave) communication, hybrid precoding, passive precoding, power allocation, energy efficiency.**

I. Introduction

MILLIMETER wave (mmWave) communication can help the future wireless communication technologies deliver high data rates in order of tens of gigabits-per-second by rendering large unused bandwidth [1]–[5]. The mmWave signals are susceptible to large path loss, owing to their small wavelengths [6], [7]. In order to recoup the huge propagation loss, massive multiple-input multiple-output (MIMO) are deployed to yield high array gain. Shadowing diminishes the average received power so severely that it becomes a daunting task to have a reliable communication in mmWave [8]. The user may frequently find itself with no connection to the transmitter, falling in the dead zone, particularly in the dense urban areas [9]. Reconfigurable Intelligent Surface (RIS) can establish a link between the base station (BS) and the receiver by creating a virtual line of sight (LoS) path to circumvent the blockade between the BS and the receiver [8], [10]. RIS accomplishes this by re-radiating the signal from the BS in the shape of a beam toward the receiver [11]. The RIS can also operate as centralized beamformer to increase channel gains where received power is low [8].

RIS is a metasurface consisting of a number of elements with the capability of reflecting the incident signal *smartly* without requiring any RF processing, decoding, encoding, or retransmission [10], [12], [13]. RIS introduces changes in the phases of the incident signals to control how electromagnetic wave behaves to beget a programmable, highly deterministic radio environments from highly probabilistic wireless environment [14]–[17]. RIS, thus gives rise to an *intelligent* radio channel that makes the communication between transmitter and receiver more reliable [15], [18]. The RISs are operated by evaluating the phase shift coefficients at the BS based on the channel state information (CSI), and controlled through a smart controller such as a field programmable gate array (FPGA) [19]. The design of phase shift coefficients of RIS, also known as passive precoding, can describe the phase response of RIS to achieve useful signal enhancement and/or interference suppression [20].

Ning *et al.* [21] propose to design passive precoding by sum-path-gain maximization (SPGM) with the view to maximize the sum of gains of different channel paths in an RIS-aided point-to-point MIMO. [21] uses alternating direction of method of multipliers (ADMM) in solving the SPGM problem to obtain RIS coefficients, and subsequently performs singular value decomposition (SVD) on the equivalent channel to construct active precoder. Wang *et al.* [22] carries out joint optimization of the transmitter precoder and RIS phase-shifts to maximize the received signal power in an mmWave MIMO environment assisted by multiple RISs. An optimal closed-form solution is obtained for single RIS case, whereas near-optimal analytical solution is derived for multi-RIS case by assuming near-orthogonality of different steering vectors.

[23] seeks to maximize the weighted sum-rate in an RIS-aided multiple user multiple-input single-output (MU-MISO) downlink system through the joint design of the precoder at the access point (AP) and the phases of the RIS elements by proposing a solution based on fractional programming. Pan *et al.* [19] consider the problem of maximizing weighted sum-rate of all users in a multi-cell MIMO having multi-antenna users, and solve the joint optimization of precoding at BS and RIS through alternating optimization. [19] optimizes the phase-shifts of RIS through majorization-minimization algorithm and complex

Prabhat Raj Gautam is currently with the Department of Computing, Imperial College London, London, UK and was previously with University of Leeds, Leeds, UK (e-mail: p.gautam@imperial.ac.uk).
Li Zhang is with the School of Electronic and Electrical Engineering, University of Leeds, Leeds, UK (email: l.x.zhang@leeds.ac.uk).
Pingzhi Fan is with the Key Lab of Information Coding and Transmission, Southwest Jiaotong University, Chengdu 610031, China (email: pzfan@swjtu.edu.cn).

circle manifold method, and determines the precoding matrices in closed form after computing the phase shifts of RIS. Zhang *et al.* [24] also aim to maximize weighted sum-rate of all users through joint optimization of precoding at BS and RIS in a cell free MIMO with multi-antenna users and multiple RISs. [24] uses fractional programming to sever the problem into multiple subproblems which are solved alternately. Zhang *et al.* [25] and Wang *et al.* [9] both consider joint optimization of active and passive precoding in point-to-point MIMO aided by a single RIS. [25] alternately optimizes the transmit covariance matrix and the phase-shift coefficients of the RIS. [9] capitalizes on the structure of mmWave channel and reconstructs the complex joint optimization problem into a simpler problem which is solved by a manifold optimization algorithm. The authors adopt manifold based optimization method [26] to generate hybrid precoder.

Zhao *et al.* [27] jointly optimize the precoding at the AP and discrete phase shifts of the RIS with the aim of minimizing AP's transmit power constrained to the outage probability for the users. Liu *et al.* [28] decouples the joint symbol-level precoding and reflection coefficients design problem in MU-MISO systems, and solves the symbol-level precoding and the reflection coefficients design subproblems by using algorithms based on gradient-projection and Riemannian conjugate gradient respectively. Yang *et. al.* [29] propose RIS-assisted Federated Spectrum Learning (FSL) framework by performing joint optimization of RIS phase shifts, user-RIS association and bandwidth allocation. The RIS phase shifts are optimized to maximize the achieved gain of each user served by the RIS, a matching game-based association scheme is utilized to associate the users with an RIS, and a bisection search method is used for dynamic bandwidth allocation.

[30] provides a comprehensive survey on resource allocation and energy efficiency strategies for RIS-aided MIMO systems. Various algorithms for optimal antenna selection, beamforming and power allocation are discussed as resource allocation strategies, while strategies like power minimization, dynamic power control, hybrid beamforming, and sleep/wake scheduling are explored for energy efficiency optimization. In [31] and [32], it is shown that the dynamic and careful selection of transmit antennas from the available ones can enhance energy efficiency while reducing hardware costs and complexity. Similar to selection of active antennas for improving energy efficiency, selection of active or "ON" RIS phase elements has been explored in [33], [34]. The RIS controller in [33], [34] dynamically configures ON/OFF status of RIS elements to maximize received SINR based on spectrum learning accomplished through convolutional neural network (CNN). Huang *et. al.* [35] propose an algorithm based on alternating maximization to maximize energy efficiency. In [35], two algorithms based on conjugate gradient search and sequential programming are presented to optimize the RIS phase shifts, and the power allocated to different users is optimized using the Dinklebach method. Wu *et al.* [36] aim to minimize transmit power through joint active and passive precoding design, subject to minimum SINR requirements at each user in both single-user and multiple-user MIMO communication assisted by an RIS with finite phase shifts. Yang *et al.* [37] propose an iterative solution that utilizes successive convex approximation to optimize RIS phase shifts and precoding at BS, while calculating the optimal power using a closed form solution to meet the goal of maximizing energy efficiency. In addition, the authors of [37] consider the dynamic turning on and off of RIS elements which is optimized through the dual problem in case of a single user, and through low-complexity search method in multiple users.

In this paper, we aim to design precoding at BS, power allocation to users and passive precoding at the RIS that optimize energy efficiency of the system in an RIS-aided mmWave MIMO. The mmWave MIMO employs large antenna arrays at the BS to counter huge path loss suffered by mmWave signals. As radio frequency chains operating at mmWave frequencies are very costly and power-consuming, hybrid precoding is used instead of fully digital precoding. The earlier works that aim to optimize energy efficiency through precoding at BS and passive precoding at RIS mostly involve fully digital precoder. In this work, we aim to determine hybrid precoder at the BS, power allocation to the users and passive precoding at RIS to maximize the energy efficiency of the system. We consider a situation in which a BS is communicating with multiple users in the mmWave channel. Communication is facilitated through the distributed placement of RISs. The contributions that we have made can be outlined as follows.

(i) We consider hybrid precoding at BS, power allocation to the users and passive precoding at RIS to maximize the energy efficiency of an mmWave multiple user multiple-input single-output (MU-MISO) system. The hybrid precoder which is a product of analog precoder and digital precoder makes the energy efficiency maximization problem more complicated. To simplify the problem, hybrid precoder is replaced by fully digital precoder from which hybrid precoder is computed at the end. As the optimization variables, fully digital precoder, power allocation matrix and passive precoding matrix are all coupled, the energy efficiency maximization problem is still complicated which is eased by choosing zero forcing (ZF) precoder as the fully digital precoder. The optimization problem is now only in terms of power allocation matrix and passive precoding matrix.

(ii) As both the optimization variables, power allocation matrix and passive precoding matrix are coupled, we can not optimize both at the same time to maximize energy efficiency. Thus, we take the route of alternating optimization technique to jointly optimize passive precoding matrix for the RIS and power allocation matrix. Specifically, passive precoding subproblem reduces into transmit power minimization subproblem which is translated into a boolean quadratic problem. The power allocation subproblem is solved by method of Lagrange multipliers using Karush-Kuhn-Tucker (KKT) conditions.

(iii) It is assumed that individual RIS elements can be switched off through the controller. To further push for energy efficient passive precoding, a computationally efficient strategy is proposed to switch off some of the RIS elements that leads to further enhancement in energy efficiency.

(iv) To evaluate hybrid precoder from fully digital precoder, analog precoder is first determined by minimizing the Euclidean distance between the hybrid precoder and the fully digital precoder using an algorithm based on truncated singular value decomposition (SVD). The digital precoder part of the hybrid ZF precoder is computed by taking pseudoinverse of the combined channel with analog precoder, and the power allocation is again calculated for the hybrid ZF precoder.

(v) Simulations are performed to investigate how different parameters affect the energy efficiency with the proposed method. The performance of the proposed method is also compared with the existing solution. The comparison evinces that the proposed method outstrips the performance of the existing solution.

A. Notations

$\mathbf{x}$ represents a vector, whereas $\mathbf{X}$ represents a matrix; $\mathbf{x}_i$ represents the i^{th} element of $\mathbf{x}$; $\mathbf{X}_{i,j}$ represents the $(i,j)^{th}$ element of $\mathbf{X}$; $\mathbf{X}_{.\mathbf{k}}$ is the k^{th} column of $\mathbf{X}$; $\mathbf{X}_{\mathbf{k}.}$ is the k^{th} row of $\mathbf{X}$; $\mathbf{X}_{:,m:n}$ is a submatrix of $\mathbf{X}$ with all rows and columns m to n; $\|\mathbf{x}\|_2$ is the $\ell 2$-norm of $\mathbf{x}$; $\|\mathbf{X}\|_F$ is the Frobenius norm of $\mathbf{X}$; $\mathbf{Tr}\left[\mathbf{X}\right]$ is the trace of $\mathbf{X}$; $\exp\left(\mathbf{X}\right)$ is a matrix whose $(i,j)^{th}$ entry is $\exp\left(\mathbf{X}_{i,j}\right)$, where $\exp\left(.\right)$ is the exponential operator; $\mathbf{X}^{-1}$, $\mathbf{X}^{\dagger}$ and $\mathbf{X}^{H}$ are the inverse, pseudoinverse and Hermitian transpose of $\mathbf{X}$ respectively; $\odot$ is the Hadamard product of two matrices; **diag** $\left(\mathbf{X}\right)$ is a column vector containing the diagonal elements of $\mathbf{X}$; **DIAG** $\left(a_1,\ldots,a_n\right)$ is an $n\times n$ diagonal matrix with $a_1,\ldots,a_n$ as the diagonal elements; **DIAG** $\left(\mathbf{x}\right)$ is a diagonal matrix with the elements of vector $\mathbf{x}$ as its diagonal elements; **BLKDIAG**$(\mathbf{A}_1,\mathbf{A}_2,\ldots,\mathbf{A}_n)$ is a block-diagonal matrix in which $\mathbf{A}_1,\mathbf{A}_2,\ldots,\mathbf{A}_n$ form the diagonal; $\mathbf{I}_N$ is an $N\times N$ identity matrix; $\mathbb{R}_+$ denotes the set of positive real numbers; $\Re\left(\cdot\right)$ represents the real part of the argument; $\mathbb{E}[.]$ is the expectation operator; $\sim$ means 'has the probability distribution of'; $\triangleq$ means 'is defined as'; $\mathcal{N}(\mu,\sigma^2)$ represents complex Gaussian variable with mean μ and variance σ^2; $\mathcal{CN}(\boldsymbol{\mu},\mathbf{C})$ represents complex Gaussian vector with mean $\boldsymbol{\mu}$ and covariance matrix $\mathbf{C}$.

II. System Model

We consider a downlink mmWave MU-MISO system with a base station serving K single-antenna users with the aid of R RISs. The BS employs N_t transmit antennas and each RIS has N_I elements. The transmit information signal $\mathbf{s}\in\mathbb{C}^{K\times 1}$, such that $\mathbb{E}\left[\mathbf{s}\mathbf{s}^H\right]=\mathbf{I}_K$, is precoded before transmission to produce transmitted signal as

$$\mathbf{x}=\sum_{k=1}^{K}\mathbf{F}_{.\mathbf{k}}\sqrt{p_k}s_k=\mathbf{F}\mathbf{P}^{1/2}\mathbf{s}, \tag{1}$$

where s_k is the symbol intended for the k^{th} user, $\mathbf{F}$ is the precoder, $\mathbf{F}_{.\mathbf{k}}$ is the k^{th} column of $\mathbf{F}$, and $\mathbf{P}=\mathbf{DIAG}(p_1,p_2,\ldots,p_K)$ is the power loading matrix with p_k being the power for the k^{th} user. The transmit power constraint forces the power of the transmitted signal to not exceed the maximum permitted power P_{max}, *i.e.*,

$$\mathbb{E}\left[\|\mathbf{x}\|^2\right]=\mathbf{Tr}\left[\mathbf{F}\mathbf{P}\mathbf{F}^H\right]\leq P_{max}. \tag{2}$$

We consider that there is no direct channel between the BS and the users which is justified in mmWave communication. The received signal at the k^{th} mobile station (MS) is sum of the signals received through various RISs,

$$\begin{aligned} y_k&=\left(\sum_{r=1}^{R}\mathbf{h}_{\mathrm{R}}^{r,k}\boldsymbol{\Phi}_r\mathbf{H}_{\mathrm{T}_r}\right)\mathbf{x}+n_k\\ &=\left(\sum_{r=1}^{R}\mathbf{h}_{\mathrm{R}}^{r,k}\boldsymbol{\Phi}_r\mathbf{H}_{\mathrm{T}_r}\right)\mathbf{F}\mathbf{P}^{1/2}\mathbf{s}+n_k, \end{aligned} \tag{3}$$

where $\mathbf{h}_{\mathrm{R}}^{r,k}\in\mathbb{C}^{1\times N_I}$ is the channel between the r^{th} RIS and the k^{th} MS, $\boldsymbol{\Phi}_r$ is a diagonal matrix with the phase shift coefficients of the r^{th} RIS along its diagonal, $\mathbf{H}_{\mathrm{T}_r}\in\mathbb{C}^{N_I\times N_t}$ is the channel between the BS and the r^{th} RIS, $n_k\in\mathbb{C}$ and $n_k\sim\mathcal{N}\left(0,\sigma_n^2\right)$ is the noise at the receiver of the k^{th} user. The receive vector can be formed by concatenating all y_ks vertically as $\mathbf{y}=\left[y_1,y_2,\ldots,y_K\right]^T$ which can be expressed as

$$\begin{aligned} \mathbf{y}&=\begin{bmatrix}\sum_{r=1}^{R}\mathbf{h}_{\mathrm{R}}^{r,1}\boldsymbol{\Phi}_r\mathbf{H}_{\mathrm{T}_r}\\ \vdots\\ \sum_{r=1}^{R}\mathbf{h}_{\mathrm{R}}^{r,K}\boldsymbol{\Phi}_r\mathbf{H}_{\mathrm{T}_r}\end{bmatrix}\mathbf{F}\mathbf{P}^{1/2}\mathbf{s}+\mathbf{n}\\ &=\left\{\sum_{r=1}^{R}\begin{bmatrix}\mathbf{h}_{\mathrm{R}}^{r,1}\\ \vdots\\ \mathbf{h}_{\mathrm{R}}^{r,K}\end{bmatrix}\boldsymbol{\Phi}_r\mathbf{H}_{\mathrm{T}_r}\right\}\mathbf{F}\mathbf{P}^{1/2}\mathbf{s}+\mathbf{n}\\ &=\left(\sum_{r=1}^{R}\mathbf{H}_{\mathrm{R}_{.r}}\boldsymbol{\Phi}_r\mathbf{H}_{\mathrm{T}_r}\right)\mathbf{F}\mathbf{P}^{1/2}\mathbf{s}+\mathbf{n}\\ &=\mathbf{H}_{\mathrm{R}}\boldsymbol{\Phi}\mathbf{H}_{\mathrm{T}}\mathbf{F}\mathbf{P}^{1/2}\mathbf{s}+\mathbf{n},\quad \text{where} \end{aligned}$$

$$\mathbf{H}_{\mathrm{R}}=\left[\mathbf{H}_{\mathrm{R}_{.1}},\ldots,\mathbf{H}_{\mathrm{R}_{.R}}\right], \tag{4a}$$

$$\begin{aligned}\boldsymbol{\Phi}&=\mathbf{BLKDIAG}\left(\boldsymbol{\Phi}_1,\ldots,\boldsymbol{\Phi}_R\right)\\ &=\mathbf{DIAG}\left(\boldsymbol{\phi}_1,\ldots,\boldsymbol{\phi}_{RN_I}\right),\end{aligned} \tag{4b}$$

$$\mathbf{H}_{\mathrm{T}}=\left[\mathbf{H}_{\mathrm{T}_1}^T,\ldots,\mathbf{H}_{\mathrm{T}_R}^T\right]^T, \tag{4c}$$

where $\mathbf{H}_{\mathrm{R}_{.r}}$ is the r^{th} column of $\mathbf{H}_{\mathrm{R}}$ and corresponds to the channel from r^{th} RIS to the K MSs, $\mathbf{n}\triangleq\left[n_1,n_2,\ldots,n_K\right]^T$ and $\mathbf{n}\sim\mathcal{CN}\left(\mathbf{0},\sigma_n^2\mathbf{I}\right)$. The channels from the BS to each RIS, and each RIS to all the MSs are considered narrowband and modeled by multi-path channel model. The channel from the BS to the r^{th} RIS and the channel from r^{th} RIS to the k^{th} user are respectively given by

$$\mathbf{H}_{T_r}=\sqrt{\frac{N_tN_I}{L_r}}\sum_{\ell=1}^{L_r}\alpha_\ell\mathbf{a}_{RIS}^r(\phi_{\ell,R}^{RIS,r},\theta_{\ell,R}^{RIS,r})\mathbf{a}_{BS}(\phi_\ell^{BS})^H, \tag{5a}$$

$$\mathbf{h}_R^{r,k}=\sqrt{\frac{N_I}{L_{r,k}}}\sum_{\ell=1}^{L_{r,k}}\beta_\ell\mathbf{a}_{RIS}^r(\phi_{\ell,T}^{RIS,r},\theta_{\ell,T}^{RIS,r})^H, \tag{5b}$$

where $\mathbf{a}_{RIS}^r(\cdot)$ and $\mathbf{a}_{BS}(\cdot)$ are antenna array response vectors of the r^{th} RIS and BS respectively. $\phi_{\ell,R}^{RIS,r}(\theta_{\ell,R}^{RIS,r})$, ϕ_ℓ^{BS} and $\phi_{\ell,T}^{RIS,r}(\theta_{\ell,T}^{RIS,r})$ are the azimuth (elevation) angles of arrival at the r^{th} RIS, azimuth angles of departure at the BS and azimuth

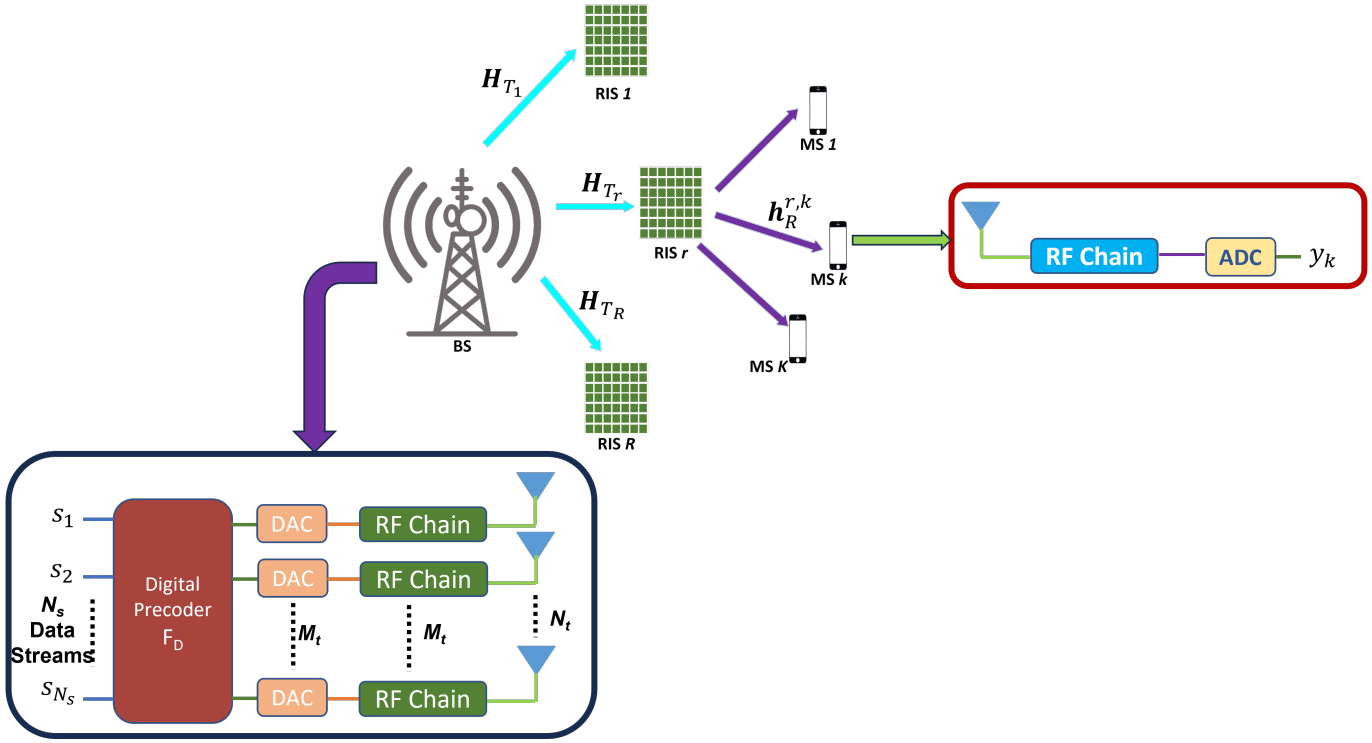


Fig. 1: mmWave MU-MISO communication enhanced by distributed RISs

(elevation) angles of departure at the r^{th} RIS respectively. The antenna array response vectors of ULA with N antennas and UPA with $N_y \times N_z$ antennas are given by

$$\mathbf{a}_{BS}(\phi) = \frac{1}{\sqrt{N}}\left[1 \quad e^{j\zeta d(\sin\phi)} \quad \dots \quad e^{j\zeta d((N-1)\sin\phi)}\right]^T, \tag{6a}$$

$$\mathbf{a}_{RIS}(\phi,\theta) = \frac{1}{\sqrt{N_y N_z}}\Big[1 \quad \dots \quad e^{j\zeta d(y\sin\phi\sin\theta + z\cos\theta)} \quad \dots \quad e^{j\zeta d((N_y-1)\sin\phi\sin\theta + (N_z-1)\cos\theta)}\Big]^T. \tag{6b}$$

where $\zeta = \dfrac{2\pi}{\lambda}$, λ being the carrier wavelength, and d is the distance between antenna elements.

A. Power Consumption of the system

The total spectral efficiency of the system is

$$\mathcal{R}_{tot} = \sum_{k=1}^{K} log_2(1 + \mathrm{SINR_k}), \tag{7}$$

where $\mathrm{SINR_k}$ is the SINR at the k^{th} user given by

$$\mathrm{SINR}_k = \frac{|\mathbf{H}_{k\cdot}\mathbf{F}_{\cdot k}|^2 p_k}{\sum\limits_{i=1, i\neq k}^{K} |\mathbf{H}_{i\cdot}\mathbf{F}_{\cdot i}|^2 p_i + \sigma_n^2}, \tag{8}$$

where $\mathbf{H}_{k\cdot} = \mathbf{h}_{\mathrm{R}}^k \mathbf{\Phi} \mathbf{H}_{\mathrm{T}}$ is the channel from the BS to the k^{th} MS where $\mathbf{h}_{\mathrm{R}}^k = \left[\mathbf{h}_{\mathrm{R}}^{1,k}, \dots, \mathbf{h}_{\mathrm{R}}^{r,k}\right]$. Mathematically, the total power consumption of the system with hybrid architecture at the BS is

$$P_{\mathrm{T_{hy}}} = \nu^{-1}\mathbf{Tr}(\mathbf{FPF}^H) + P_{com} + P_{MS}K + P_{RF}M_t + P_{PA}M_t + P_{PS}N_{PS} + P_n N_I^{on}, \tag{9}$$

where, ν is the efficiency of the transmit power amplifier, P_{com} is the common power of the BS, P_{MS} is the circuit power consumed by each MS, P_{RF} is the power consumed by each RF chain, P_{PA} is the power consumed by each power amplifier, P_{PS} is the power consumed by each phase-shifter, P_n is the power consumed by each element of RIS, N_{PS} is the total number of phase-shifters, and N_I^{on} is the total number of RIS elements that are switched on. $N_I^{on} = RN_I$ when all the elements are switched on. Since, the RF chains and power amplifiers are applied before analog precoder, the number of RF chains and power ampliers is M_t [38]. If the conventional MIMO architecture exists at the BS so that fully digital precoding is possible, the number of RF chains, $M_t = N_t$ and $N_{PS} = 0$ so that the total power consumption of the system is

$$P_{\mathrm{T_{FD}}} = \nu^{-1}\mathbf{Tr}(\mathbf{FPF}^H) + P_{com} + P_{MS}K + (P_{RF} + P_{PA})N_t + P_n N_I^{on}. \tag{10}$$

III. Problem Statement

The energy efficiency of the system is given by the ratio of spectral efficiency to the total power consumption as,

$$\eta = \frac{\sum\limits_{k=1}^{K} log_2(1 + \mathrm{SINR_k})}{P_{\mathrm{T}}} \quad bits/Hz/Joule, \tag{11}$$

where P_{T} is the total power consumption, $P_{\mathrm{T}} = P_{\mathrm{T_{hy}}}$ in case of hybrid precoding and $P_{\mathrm{T}} = P_{\mathrm{T_{FD}}}$ in case of fully digital precoding. The energy efficiency depends on the choice of $\mathbf{\Phi}$, $\mathbf{F}$ and $\mathbf{P}$. So, we design the precoder $\mathbf{F}$, $\mathbf{P}$ and $\mathbf{\Phi}$ to maximize the energy efficiency of the system subject to quality of service (QoS) constraint in terms of minimum guaranteed spectral efficiency $\mathcal{R}_{min_k}$ for the k^{th} user. Only the first term of the denominator depends on $\mathbf{F}$, $\mathbf{P}$ and $\mathbf{\Phi}$, so the rest of the terms is represented by $P_{rest} = P_{com} + P_{MS}K + P_{RF}M_t + P_{PA}M_t + P_{PS}N_{PS} + P_nN_I^{on}$ for simplicity.

A. Choice of Precoder

Hybrid precoder is a combination of analog precoder $\mathbf{F}_{\mathrm{R}} \in \mathbb{C}^{N_t \times M_t}$ which is implemented by a network of phase-shifters and low-dimensional digital precoder $\mathbf{F}_{\mathrm{D}} \in \mathbb{C}^{M_t \times K}$ so that the hybrid precoder is $\mathbf{F} = \mathbf{F}_{\mathrm{R}}\mathbf{F}_{\mathrm{D}}$. The optimization problem for maximizing energy efficiency is

$$\begin{aligned}&\mathbf{F}_{\mathrm{R}}, \mathbf{F}_{\mathrm{D}}, \mathbf{\Phi}, \mathbf{P} \\ &= \underset{\mathbf{F}_{\mathrm{R}}, \mathbf{F}_{\mathrm{D}}, \mathbf{\Phi}, \mathbf{P}}{\arg\max} \frac{\sum_{k=1}^{K} log_2\left(1 + \frac{\left|\mathbf{H}_{k.}\mathbf{F}_{\mathrm{R}}\mathbf{F}_{\mathrm{D}_{.k}}\right|^2 p_k}{\sum_{i=1, i\neq k}^{K}\left|\mathbf{H}_{i.}\mathbf{F}_{\mathrm{R}}\mathbf{F}_{\mathrm{D}_{.i}}\right|^2 p_i + \sigma_n^2}\right)}{\nu^{-1}\mathbf{Tr}(\mathbf{F}_{\mathrm{R}}\mathbf{F}_{\mathrm{D}}\mathbf{P}\mathbf{F}_{\mathrm{R}}\mathbf{F}_{\mathrm{D}}^H) + P_{rest}} \qquad (12a)\end{aligned}$$

$$\text{s.t.} \quad log_2\left(1 + \frac{\left|\mathbf{H}_{k.}\mathbf{F}_{\mathrm{R}}\mathbf{F}_{\mathrm{D}_{.k}}\right|^2 p_k}{\sum_{i=1, i\neq k}^{K}\left|\mathbf{H}_{i.}\mathbf{F}_{\mathrm{R}}\mathbf{F}_{\mathrm{D}_{.i}}\right|^2 p_i + \sigma_n^2}\right) \geq \mathcal{R}_{min_k}, \qquad (12b)$$

$$\mathbf{Tr}(\mathbf{F}_{\mathrm{R}}\mathbf{F}_{\mathrm{D}}\mathbf{P}\mathbf{F}_{\mathrm{D}}^H\mathbf{F}_{\mathrm{R}}^H) \leq P_{max}, \qquad (12c)$$

$$|\phi_r| = 1 \quad \forall\ r = 1, 2, \ldots, RN_I, \qquad (12d)$$

$$\left|\mathbf{F}_{\mathrm{R}_{i,j}}\right| = 1 \quad \forall i = 1, \ldots, N_t;\ j = 1, \ldots, M_t, \qquad (12e)$$

where $\mathbf{F}_{\mathrm{D}_{.k}}$ is the k^{th} column of $\mathbf{F}_{\mathrm{D}}$, (12b) represents the QoS constraint guaranteeing minimum spectral efficiency $\mathcal{R}_{min_k}$ for k^{th} user, (12c) represents transmit power constraint, (12d) represents phase-only altering capability of RIS with ϕ_r being the r^{th} diagonal element of $\mathbf{\Phi}$, and (12e) represents constant modulus constraint imposed by phase-shifters on analog precoder.

Clearly, the optimization problem (12) is non-convex and very challenging. One of the difficulties is introduced by the hybrid precoder which has two optimization variables $\mathbf{F}_{\mathrm{R}}$ and $\mathbf{F}_{\mathrm{D}}$ coupled. To tackle this challenge, instead of trying to directly optimize $\mathbf{F}_{\mathrm{R}}$ and $\mathbf{F}_{\mathrm{D}}$, we first determine fully-digital precoder $\mathbf{F}_{\mathrm{FD}}$ and determine $\mathbf{F}_{\mathrm{R}}$ and $\mathbf{F}_{\mathrm{D}}$ later from $\mathbf{F}_{\mathrm{FD}}$.

Even with $\mathbf{F}_{\mathrm{R}}\mathbf{F}_{\mathrm{D}}$ in optimization problem (12) replaced with $\mathbf{F}_{\mathrm{FD}}$, the optimization is still complicated as the optimization variables $\mathbf{F}_{\mathrm{FD}}$, $\mathbf{P}$ and $\mathbf{\Phi}$ are still coupled in both the objective function and transmit power constraint as $\mathbf{H}$ depends on $\mathbf{\Phi}$. To simplify the optimization problem, we choose $\mathbf{F}_{\mathrm{FD}}$ to be a ZF precoder which is calculated as the right inverse of the equivalent channel, *i.e.*, $\mathbf{F}_{\mathrm{FD}} = \mathbf{H}^{\dagger} = \mathbf{H}^H\left(\mathbf{H}\mathbf{H}^H\right)^{-1}$. It is guaranteed to exist if $N_t \geq K$ and $RN_I \geq K$ [35]. The reason behind the choice of ZF precoder stems from the fact that application of ZF precoder makes interference zero, and SINR for the k^{th} user reduces to the ratio of useful power to the noise power. Hence, SINR for the k^{th} user looks like

$$\mathrm{SINR}_k = log_2\left(1 + \frac{p_k}{\sigma_n^2}\right). \qquad (13)$$

The precoder needs to satisfy transmit power constraint (2), which is accomplished through the appropriate choice of $\mathbf{P}$. Since the ZF precoder is used, the problem for determining $\mathbf{\Phi}$ and $\mathbf{P}$ to maximize energy efficiency can now be stated as

$$\mathbf{\Phi}, \mathbf{P} = \underset{\mathbf{\Phi}, \mathbf{P}}{\arg\max} \frac{\sum_{k=1}^{K} log_2(1 + \frac{p_k}{\sigma_n^2})}{\nu^{-1}\mathbf{Tr}\left[\left(\mathbf{H}_{\mathrm{R}}\mathbf{\Phi}\mathbf{H}_{\mathrm{T}}\right)^{\dagger}\mathbf{P}\left(\mathbf{H}_{\mathrm{R}}\mathbf{\Phi}\mathbf{H}_{\mathrm{T}}\right)^{\dagger^H}\right] + P_{rest}} \qquad (14a)$$

$$\text{s.t.} \quad p_k \geq \sigma_n^2\left(2^{\mathcal{R}_{min_k}/10} - 1\right), \qquad (14b)$$

$$\mathbf{Tr}\left[\left(\mathbf{H}_{\mathrm{R}}\mathbf{\Phi}\mathbf{H}_{\mathrm{T}}\right)^{\dagger}\mathbf{P}\left(\mathbf{H}_{\mathrm{R}}\mathbf{\Phi}\mathbf{H}_{\mathrm{T}}\right)^{\dagger^H}\right] \leq P_{max}, \qquad (14c)$$

$$|\phi_r| = 1 \quad \forall\ r = 1, 2, \ldots, RN_I, \qquad (14d)$$

where the constraint (14b) represents the QoS constraint for user k and the constraint (14c) represent the transmit power constraint.

IV. Maximization Of Energy Efficiency

In this section, we solve the problem (14) of maximizing energy efficiency. It is not possible to optimize both $\mathbf{\Phi}$ and $\mathbf{P}$ at the same time. The route of alternating optimization [39], [40] is chosen where $\mathbf{\Phi}$ and $\mathbf{P}$ are optimized alternately in a series of iterations until convergence is achieved. We begin with optimization of $\mathbf{\Phi}$ for fixed $\mathbf{P}$, and follow it with optimization of $\mathbf{P}$ keeping $\mathbf{\Phi}$ fixed.

A. Passive Precoding for Energy Efficiency

The subproblem for determination of RIS phase-shift matrix or passive precoding matrix is defined when $\mathbf{P}$ is held constant. It can be seen that with fixed $\mathbf{P}$, maximizing the objective function (14a) is equivalent to minimizing the first term of the denominator. Hence if $\mathbf{P}$ is kept constant, the problem (14) reduces to

$$\mathbf{\Phi} = \underset{\mathbf{\Phi}}{\arg\min}\ \mathbf{Tr}\left[\left(\mathbf{H}_{\mathrm{R}}\mathbf{\Phi}\mathbf{H}_{\mathrm{T}}\right)^{\dagger}\mathbf{P}\left(\mathbf{H}_{\mathrm{R}}\mathbf{\Phi}\mathbf{H}_{\mathrm{T}}\right)^{\dagger^H}\right], \qquad (15a)$$

$$\text{s.t.} \quad \mathbf{Tr}\left[\left(\mathbf{H}_{\mathrm{R}}\mathbf{\Phi}\mathbf{H}_{\mathrm{T}}\right)^{\dagger}\mathbf{P}\left(\mathbf{H}_{\mathrm{R}}\mathbf{\Phi}\mathbf{H}_{\mathrm{T}}\right)^{\dagger^H}\right] \leq P_{max}, \qquad (15b)$$

$$|\phi_r| = 1 \quad \forall\ r = 1, 2, \ldots, RN_I. \qquad (15c)$$

We can simplify the objective function (15a) as

$$\begin{aligned}
&\mathbf{Tr}\left[\left(\mathbf{H}_{\mathrm{R}}\mathbf{\Phi}\mathbf{H}_{\mathrm{T}}\right)^{\dagger}\mathbf{P}\left(\mathbf{H}_{\mathrm{R}}\mathbf{\Phi}\mathbf{H}_{\mathrm{T}}\right)^{\dagger^{H}}\right]\\
&\overset{(a)}{=}\mathbf{Tr}\left[\mathbf{H}_{\mathrm{T}}^{\dagger}\mathbf{\Phi}^{H}\mathbf{H}_{\mathrm{R}}^{\dagger}\mathbf{P}\mathbf{H}_{\mathrm{R}}^{H^{\dagger}}\mathbf{\Phi}\mathbf{H}_{\mathrm{T}}^{H^{\dagger}}\right]\\
&\overset{(b)}{=}\mathbf{Tr}\left[\mathbf{\Phi}^{H}\left(\mathbf{H}_{\mathrm{R}}^{H}\mathbf{P}^{-1}\mathbf{H}_{\mathrm{R}}\right)^{\dagger}\mathbf{\Phi}\left(\mathbf{H}_{\mathrm{T}}\mathbf{H}_{\mathrm{T}}^{H}\right)^{\dagger}\right]\\
&=\mathbf{Tr}\left[\mathbf{\Phi}^{H}\mathbf{C}_{1}\mathbf{\Phi}\mathbf{C}_{2}\right]\\
&\overset{(c)}{=}\boldsymbol{\phi}^{H}\mathbf{C}_{1}\odot\mathbf{C}_{2}^{T}\boldsymbol{\phi}\\
&=\boldsymbol{\phi}^{H}\mathbf{C}\boldsymbol{\phi},
\end{aligned}$$

where the reason behind (a) is $(\mathbf{AB})^{\dagger}=\mathbf{B}^{\dagger}\mathbf{A}^{\dagger}$, $\mathbf{\Phi}^{\dagger}=\mathbf{\Phi}^{H}$, $(\mathbf{AB})^{\dagger^{H}}=(\mathbf{AB})^{H^{\dagger}}=\left(\mathbf{B}^{H}\mathbf{A}^{H}\right)^{\dagger}$; the reason behind (b) is $\mathbf{Tr}\left[\mathbf{AB}\right]=\mathbf{Tr}\left[\mathbf{BA}\right]$; $\mathbf{C}_{1}\triangleq\left(\mathbf{H}_{\mathrm{R}}^{H}\mathbf{P}^{-1}\mathbf{H}_{\mathrm{R}}\right)^{\dagger}$, $\mathbf{C}_{2}\triangleq\left(\mathbf{H}_{\mathrm{T}}\mathbf{H}_{\mathrm{T}}^{H}\right)^{\dagger}$; $\boldsymbol{\phi}=\mathbf{diag}\left(\mathbf{\Phi}\right)$ is a column vector; the reason behind (c) is the matrix identity $\mathbf{Tr}\left[\mathbf{\Phi}\mathbf{C}_{1}\mathbf{\Phi}^{H}\mathbf{C}_{2}\right]=\boldsymbol{\phi}\left(\mathbf{C}_{1}\odot\mathbf{C}_{2}^{T}\right)\boldsymbol{\phi}^{H}$ for the diagonal matrix $\mathbf{\Phi}$ [41] and $\mathbf{C}\triangleq\mathbf{C}_{1}\odot\mathbf{C}_{2}^{T}$ is a Hermitian matrix. Thus, the passive precoding subproblem can be written as

$$\boldsymbol{\phi}^{\star}=\arg\min_{\mathbf{\Phi}}\boldsymbol{\phi}^{H}\mathbf{C}\boldsymbol{\phi}, \tag{16a}$$

$$\text{s.t.}\quad \boldsymbol{\phi}^{H}\mathbf{C}\boldsymbol{\phi}\leq P_{max}, \tag{16b}$$

$$|\boldsymbol{\phi}_{r}|=1\quad\forall\ r=1,2,\ldots,RN_{I}. \tag{16c}$$

The optimization problem (16) is non-convex because of constraint (16c). We solve for $\boldsymbol{\phi}$ by starting with an initial value and optimizing over each element $\boldsymbol{\phi}_i$ at a time. The objective function (16a) can be written as an optimization problem in $\boldsymbol{\phi}_i$ as [42]

$$\begin{aligned}
\min_{\boldsymbol{\phi}_{i}}\quad &\boldsymbol{\phi}_{i}^{*}\mathbf{C}_{i,i}\boldsymbol{\phi}_{i}+2\Re\left(\boldsymbol{\phi}_{i}\sum_{\substack{j=1\\j\neq i}}^{RN_{I}}\boldsymbol{\phi}_{j}^{*}\mathbf{C}_{j,i}\right)\\
\text{s.t.}\quad &|\boldsymbol{\phi}_{i}|=1,\quad\forall\ i=1,2,\ldots,RN_{I},
\end{aligned} \tag{17}$$

where $\mathbf{C}_{i,i}$ is a real quantity. We can rewrite the optimization problem over $\boldsymbol{\phi}_i$ as

$$\begin{aligned}
\boldsymbol{\phi}_{i}^{\star}=\arg\max_{\boldsymbol{\phi}_{i}}\ &\Re\left(\boldsymbol{\phi}_{i}\sum_{j=1}^{RN_{I}}\boldsymbol{\phi}_{j}^{*}\mathbf{b}_{j}^{i}\right)\\
\text{s.t.}\quad &|\boldsymbol{\phi}_{i}|=1,\quad\forall\ i=1,2,\ldots,RN_{I},
\end{aligned} \tag{18}$$

where $\mathbf{b}_{j}^{i}$ is the j^{th} element of $\mathbf{b}^{i}$ and given by

$$\mathbf{b}_{j}^{i}=\begin{cases}-\mathbf{C}_{j,i} & j\neq i\\ -\dfrac{1}{2}\mathbf{C}_{i,i} & j=i\end{cases} \tag{19}$$

The optimal solution to the problem (18) is obtained when

$$\boldsymbol{\phi}_{i}=\frac{\left(\sum_{j=1}^{RN_{I}}\boldsymbol{\phi}_{j}^{*}\mathbf{b}_{j}^{i}\right)^{*}}{\left|\sum_{j=1}^{RN_{I}}\boldsymbol{\phi}_{j}^{*}\mathbf{b}_{j}^{i}\right|}. \tag{20}$$

Algorithm 1 Passive Precoding using Greedy Algorithm

Require: $\mathbf{C},\boldsymbol{\phi}_{init}$.

Set $\delta=1$, $\epsilon=10^{-3}$, a very small value. Iteration index, m is set to 1.
while $\delta>\epsilon$ **do**
  for i = 1 to RN_I **do**
    Set $\mathbf{b}^{i(m)}=-\mathbf{C}_{.i}$
    Set $\mathbf{b}_{i}^{i(m)}=\dfrac{1}{2}\mathbf{b}_{i}^{i(m)}$
    Compute $\boldsymbol{\phi}_{i}^{(m)}$ using (20).
  end for
  Compute $\delta=\left|\boldsymbol{\phi}^{(m)}-\boldsymbol{\phi}^{(m-1)}\right|$.
  $m\leftarrow m+1$.
end while
return $\boldsymbol{\phi}$.

B. Power allocation optimization for energy efficiency

Once we determine $\mathbf{\Phi}$, we have the equivalent channel, and we can determine the precoder using (12). The total transmit power can then be expressed as

$$\mathbf{Tr}\left[\mathbf{FPF}^{H}\right]=\sum_{k=1}^{K}\left|\mathbf{F}_{.\mathbf{k}}\right|^{2}p_{k}=\sum_{k=1}^{K}a_{k}p_{k},$$

where $a_{k}\triangleq\left|\mathbf{F}_{.\mathbf{k}}\right|^{2}$. The power optimization subproblem can be written as

$$\{p_{k}\}_{k=1}^{K}=\arg\max_{\{p_{k}\}_{k=1}^{K}}\frac{\sum_{k=1}^{K}log_{2}\left(1+\frac{p_{k}}{\sigma_{n}^{2}}\right)}{\nu^{-1}\sum_{k=1}^{K}a_{k}p_{k}+P_{rest}} \tag{21a}$$

$$\text{s.t.}\quad p_{k}\geq p_{min_{k}}, \tag{21b}$$

$$\sum_{k=1}^{K}a_{k}p_{k}\leq P_{max}, \tag{21c}$$

where $p_{min_{k}}\triangleq\sigma_{n}^{2}\left(2^{\mathcal{R}_{min_{k}}/10}-1\right)$. We use following notations to make further computations tractable.

$$P_{\mathrm{T}}\left(p_{k}\right)\triangleq\nu^{-1}\sum_{k=1}^{K}a_{k}p_{k}+P_{rest}, \tag{22a}$$

$$\mathcal{R}\left(p_{k}\right)\triangleq\sum_{k=1}^{K}log_{2}\left(1+\frac{p_{k}}{\sigma_{n}^{2}}\right), \tag{22b}$$

$$f(p_{k})=\frac{\mathcal{R}\left(p_{k}\right)}{P_{\mathrm{T}}\left(p_{k}\right)}. \tag{22c}$$

The Kuhn-Tucker Lagrangian for problem (21) is formed as

$$\mathcal{L}=f\left(p_{k}\right)-\sum_{k=1}^{K}\mu_{k}\left(p_{min_{k}}-p_{k}\right)-\lambda\left(\sum_{k=1}^{K}a_{k}p_{k}-P_{max}\right), \tag{23}$$

where $\{\mu_k\}_{k=1}^K$ and λ are Lagrangian multipliers. The Karush Kuhn-Tucker (KKT) conditions are

$$\frac{\partial \mathcal{L}}{\partial p_k} = \frac{\partial f(p_k)}{\partial p_k} + \mu_k - \lambda a_k \quad \text{where} \tag{24a}$$

$$\frac{\partial f(p_k)}{\partial p_k} = \frac{1}{ln2} \cdot \frac{1}{p_k + \sigma_n^2} \cdot \frac{1}{P_T(p_k)} - \frac{\nu^{-1} a_k \mathcal{R}(p_k)}{P_T(p_k)^2}, \tag{24b}$$

$$\frac{\partial \mathcal{L}}{\partial \mu_k} = -(p_{min_k} - p_k) \tag{24c}$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} = -\left(\sum_{k=1}^K a_k p_k - P_{max}\right) \tag{24d}$$

The stationary point is given by equating (24a) to 0,

$$\frac{1}{ln2} \cdot \frac{1}{p_k + \sigma_n^2} \cdot \frac{1}{P_T(p_k)} - \frac{\nu^{-1} a_k \mathcal{R}(p_k)}{P_T(p_k)^2} + \mu_k - \lambda a_k = 0, \tag{25}$$

which gives

$$p_k = \frac{1}{ln2} \cdot \frac{P_T(p_k)}{P_T(p_k)^2 \{\lambda a_k - \mu_k\} + \nu^{-1} a_k \mathcal{R}(p_k)} - \sigma_n^2, \tag{26a}$$

$$\mu_k = \lambda a_k + \frac{\nu^{-1} a_k \mathcal{R}(p_k)}{P_T(p_k)^2} - \frac{1}{ln2} \cdot \frac{1}{p_k + \sigma_n^2} \cdot \frac{1}{P_T(p_k)}. \tag{26b}$$

The complementary slackness equations are

$$\mu_k \frac{\partial \mathcal{L}}{\partial \mu_k} = \mu_k (p_{min_k} - p_k) = 0 \tag{27a}$$

$$\lambda \frac{\partial \mathcal{L}}{\partial \lambda} = \lambda \left(\sum_{k=1}^K a_k p_k - P_{max}\right) = 0 \tag{27b}$$

$$p_k \frac{\partial \mathcal{L}}{\partial p_k} = p_k \left(\frac{\partial f(p_k)}{\partial p_k} + \mu_k - \lambda a_k\right) = 0 \tag{27c}$$

Substituting the value of $\frac{\partial f(p_k)}{\partial p_k}$ in (27c),

$$\lambda a_k p_k = \frac{1}{ln2} \cdot \frac{p_k}{p_k + \sigma_n^2} \cdot \frac{1}{P_T(p_k)} - \frac{\nu^{-1} a_k \mathcal{R}(p_k)}{P_T(p_k)^2} p_k + \mu_k p_k.$$

Taking $\sum_{k=1}^K$ on the both sides,

$$\lambda \sum_{k=1}^K a_k p_k = \frac{1}{ln2} \cdot \frac{1}{P_T(p_k)} \sum_{k=1}^K \frac{p_k}{p_k + \sigma_n^2} - \frac{\nu^{-1} \mathcal{R}(p_k)}{P_T(p_k)^2} \sum_{k=1}^K a_k p_k + \sum_{k=1}^K \mu_k p_k.$$

Substituting $\mu_k p_k = \mu_k p_{min_k}$ and $\lambda \sum_{k=1}^K a_k p_k = \lambda P_{max}$ from (27a) and (27b) respectively,

$$\lambda = \frac{1}{P_{max}} \left[\frac{1}{ln2} \cdot \frac{1}{P_T(p_k)} \sum_{k=1}^K \frac{p_k}{p_k + \sigma_n^2} - \frac{\nu^{-1} \mathcal{R}(p_k)}{P_T(p_k)^2} \sum_{k=1}^K a_k p_k + \sum_{k=1}^K \mu_k p_{min_k}\right] \tag{28}$$

We develop an iterative algorithm to determine $\{p_k\}_{k=1}^K$ using the developments so far which is summarized in Algorithm 2.

Algorithm 2 Iterative power design using Kuhn Tucker Lagrangian Method

Require: $a_k, p_{min_k}, P_{max}, p_{k_{init}}, P_{rest}, \sigma_n^2, \nu$.

1: Set $\mu_k = 0$
2: Choose initial value of $p_k^{(0)} = p_{k_{init}}$. Calculate $P_{\mathrm{T}}^{(0)}(p_k)$ using (22a) with $p_k = p_k^{(0)}$.
3: Set $\epsilon = 10^{-4}$, a very small value, and set $i \leftarrow 1$.
4: **repeat**
5: Compute $\lambda^{(i)}$ using (28).
6: Compute $p_k^{(i)}$ using (26a). If $p_k^{(i)} < p_{min_k}$, set $p_k^{(i)} = p_{min_k}$.
7: Compute $\delta = \sqrt{\sum_{k=1}^K \left|p_k^{(i)} - p_k^{(i-1)}\right|^2}$
8: Set $i \leftarrow i + 1$.
9: **until** $\delta < \epsilon$
10: **return** p_k.

C. Alternating Optimization

We begin with an initial random value for $\mathbf{\Phi}$, so we have the equivalent channel $\mathbf{H} = \mathbf{H}_{\mathrm{R}} \mathbf{\Phi} \mathbf{H}_{\mathrm{T}}$. We also calculate a valid initial value for $\mathbf{P}$ as

$$\mathbf{P} = \frac{p^0}{K} \mathbf{I}, \quad \text{where} \quad p^0 = \frac{P_{max}}{\mathbf{Tr}\left(\mathbf{H}^\dagger \mathbf{H}^{\dagger H}\right)}, \tag{29}$$

which ensures that the transmit power is P_{max}. Then we alternately optimize $\mathbf{\Phi}$ and $\{p_k\}_{k=1}^K$ until $\{p_k\}_{k=1}^K$ converge. After convergence is attained, we strategically switch off the RIS elements that help us increase energy efficiency. The details of the RIS elements switching off strategy is given in the next subsection. The alternating optimization algorithm to maximize the energy efficiency is summarized in Algorithm 3.

D. Low-complexity Strategy for Switching Off RIS elements for Energy Efficiency

We have considered that RISs are equipped with switches for each element which can be switched on and off using a controller. The works in [37], [43] also employ the switching off of some RIS elements to promote energy RIS. After optimizing the RIS and power allocation for different users, we strategically switch off the RIS elements which when switched off increase the energy efficiency. If we set out to determine such switches exhaustively, we would require to find new equivalent channel, ZF precoder, power allocation for each user and finally determine the energy efficiency to check which RIS element is best suited to be switched off first. Then, we would be repeating the same procedure to determine the second RIS element to switch off. In this section, we propose a strategy to switch off the RIS elements so as to increase energy efficiency.

To determine $\mathbf{\Phi}$ we minimize the transmit power. To develop an order to switch off RIS elements, we sort ϕ_is in descending order in terms of their contribution to the term $\boldsymbol{\phi}^H \mathbf{C} \boldsymbol{\phi}$, the equivalent expression for transmit power of the system. It will be in the same order we will switch off the elements of the RIS. To determine the sorting order in terms of contribution to the transmit power, we compute

$$d_i = \boldsymbol{\phi}_{-i}^H \mathbf{C}_{-i,-i} \boldsymbol{\phi}_{-i}, \tag{30}$$

Algorithm 3 Alternating Optimization Based Algorithm for Energy Efficiency

Require: $\mathbf{H}_{\mathrm{R}}, \mathbf{H}_{\mathrm{T}}, p_{min_k}, P_{max}, P_{rest}, P_n\sigma_n^2, \nu$.

Initialize $\mathbf{\Phi}$ with $\mathbf{\Phi}^{(0)} = \mathbf{DIAG}\left(\exp\left(j\boldsymbol{\psi}\right)\right)$, where $\boldsymbol{\psi}$ is an RN_I-length vector and $\boldsymbol{\psi}_{i,j}$ are random phase angles.
Initialize $\mathbf{P}$ with $\mathbf{P}^0$ calculated using (29).
Compute $\eta^{(0)}$ using (11). Set $\epsilon = 10^{-3}$, a very small value, and initialize iteration index $i \leftarrow 1$.
repeat
  Compute $\mathbf{C} = \mathbf{C}_1^{(i)} \odot \mathbf{C}_2^T$, where $\mathbf{C}_1^{(i)} = \left(\mathbf{H}_{\mathrm{R}}^H \mathbf{P}^{(i-1)^{-1}} \mathbf{H}_{\mathrm{R}}\right)^{\dagger}$ and $\mathbf{C}_2 = \left(\mathbf{H}_{\mathrm{T}}\mathbf{H}_{\mathrm{T}}^H\right)^{\dagger}$.
  Compute $\mathbf{\Phi}^{(i)} = \mathbf{DIAG}\left(\boldsymbol{\phi}^{(i)}\right)$ where $\boldsymbol{\phi}^{(i)}$ is calculated using Algorithm 1.
  Compute $\{p_k\}_{k=1}^K$ using Algorithm 2. Set $\mathbf{P}^{(i)} = \mathbf{DIAG}\left(p_1, \ldots, p_K\right)$.
  Compute $\eta^{(i)}$ using (11).
  Compute $\delta^{(i)} = \left|\eta^{(i)} - \eta^{(i-1)}\right|$
  Set $i \leftarrow i + 1$.
until $\delta^{(i)} < \epsilon$
Set $\mathbf{P} = \mathbf{P}^{(i)}$
Compute $\mathbf{F}_{\text{FD}} = \mathbf{H}^{\dagger}$.
if $\mathbf{Tr}\left[\mathbf{F}_{\text{FD}}\mathbf{P}\mathbf{F}_{\text{FD}}^H\right] \leq P_{max}$ **then**
  Compute $\mathbf{C} = \mathbf{C}_1 \odot \mathbf{C}_2^T$, where $\mathbf{C}_1 = \left(\mathbf{H}_{\mathrm{R}}^H \mathbf{P}^{-1} \mathbf{H}_{\mathrm{R}}\right)^{\dagger}$ and $\mathbf{C}_2 = \left(\mathbf{H}_{\mathrm{T}}\mathbf{H}_{\mathrm{T}}^H\right)^{\dagger}$.
  Obtain $\mathbf{\Phi}$ and $\{p_k\}_{k=1}^K$ using Algorithm 4.
  Obtain hybrid precoder $\mathbf{F}$ from $\mathbf{F}_{\text{FD}}$ using Algorithm 5.
  Calculate $\{p_k\}_{k=1}^K$ using Algorithm 4 with $\mathbf{F}$ as precoder.
  if $\mathbf{Tr}\left[\mathbf{F}\mathbf{P}\mathbf{F}^H\right] \leq P_{max}$ **then**
    return $\mathbf{\Phi}$ and $\mathbf{P} = \mathbf{DIAG}\left(p_1, \ldots, p_K\right)$.
  else Declare infeasibility.
  end if
else Declare infeasibility.
end if

where $\boldsymbol{\phi}_{-i}$ is the vector formed by removing i^{th} element from $\boldsymbol{\phi}$ and $\mathbf{C}_{-i,-i}$ is the matrix formed by removing i^{th} row and the i^{th} column from $\mathbf{C}$. The contribution of $\boldsymbol{\phi}_i$ to the transmit power is given by $\boldsymbol{\phi}^H\mathbf{C}\boldsymbol{\phi} - d_i$. Thus, RIS element '$i$' corresponding to the lowest d_i contributes the highest to the transmit power. Hence, if we arrange the $\boldsymbol{\phi}_i$s in ascending order in terms of the d_i values we will get $\boldsymbol{\phi}_i$s in descending order in terms of contribution to the transmit power.

We now have the order to switch off RIS elements. We calculate the energy efficiency when i^{th} element is switched off. If it is greater than the previous value of energy efficiency with all elements switched on, we switch off the element. Then we proceed to check if the energy efficiency increases when the "next in queue" RIS element is switched off. If the energy efficiency increases we proceed to check for next RIS element, else we do not switch off the element and stop the process of switching off the RIS elements. The RIS On-Off procedure is summarized in Algorithm 4.

Algorithm 4 RIS Elements Switching Off Strategy for Energy Efficiency

Require: $\mathbf{H}_{\mathrm{R}}, \mathbf{H}_{\mathrm{T}}, \mathbf{C}, p_{min_k}, \boldsymbol{\phi}, \{p_k\}_{k=1}^K, P_{max}, P_{rest}, P_n, \sigma_n^2, \nu$.

Set $\boldsymbol{\phi}^{(0)} = \boldsymbol{\phi}$ Compute $d_i,\ i = 1, 2, \ldots, RN_I$ using (30).
Sort d_i in ascending order and store the sorting index order in $sort\_d$.
Set $\delta = 1$ and initialize iteration index $i \leftarrow 1$.
while $\delta > 0$ **do**
  $\ell = sort\_d(i)$
  Set $\boldsymbol{\phi}^{(i)} = \boldsymbol{\phi}^{(i-1)}$ and $\boldsymbol{\phi}_{\ell}^{(i)} = 0$.
  Compute $\mathbf{\Phi}^{(i)} = \mathbf{DIAG}\left(\boldsymbol{\phi}^{(i)}\right)$ and $\mathbf{H}^{(i)} = \mathbf{H}_{\mathrm{R}}\mathbf{\Phi}^{(i)}\mathbf{H}_{\mathrm{T}}$.
  $\mathbf{F}^{(i)} = \mathbf{H}^{(i)^{\dagger}}$, $a_k^{(i)} = \left|\mathbf{F}_{\mathbf{.k}}^{(\mathbf{i})}\right|^2$.
  Compute $\left\{p_k^{(i)}\right\}_{k=1}^K$ using Algorithm 2. Set $\mathbf{P}^{(\mathbf{i})} = \mathbf{DIAG}\left(p_1^{(i)}, \ldots, p_K^{(i)}\right)$
  If $\mathbf{Tr}\left[\mathbf{F}^{(i)}\mathbf{P}^{(i)}\mathbf{F}^{(i)^H}\right] > P_{max}$, go to step 15.
  Compute $\eta^{(i)}$ using (11).
  Compute $\delta = \eta^{(i)} - \eta^{(i-1)}$
  Set $i \leftarrow i + 1$.
end while
Revert to the previous values of p_k and $\boldsymbol{\phi}$, *i.e.*, $\{p_k\}_{k=1}^K = \left\{p_k^{(i-1)}\right\}_{k=1}^K$, $\boldsymbol{\phi} = \boldsymbol{\phi}^{(i-1)}$
return $\{p_k\}_{k=1}^K$ and $\boldsymbol{\phi}$.

E. Hybrid Precoder from fully digital precoder

The hybrid precoder is computed by minimizing the Euclidean distance between their fully digital counterpart, $\mathbf{F}_{\text{FD}}$. The hybrid precoding problem [44] may be stated as

$$\underset{\mathbf{F}_{\mathrm{R}}, \mathbf{F}_{\mathrm{D}}}{\arg\min} \quad \left\|\mathbf{F}_{\text{FD}} - \mathbf{F}_{\mathrm{R}}\mathbf{F}_{\mathrm{D}}\right\|_F^2 \tag{31a}$$

$$\text{s.t.} \quad \left|\mathbf{F}_{\mathrm{R}_{i,j}}\right| = 1,\ \forall i, j. \tag{31b}$$

The precoding problem (31) does not have the transmit power constraint because it is satisfied through the choice of $\mathbf{P}$. We determine $\mathbf{F}_{\mathrm{R}}$ and $\mathbf{F}_{\mathrm{D}}$ in two separate stages. If $\mathbf{F}_{\mathrm{R}}$ is known, $\mathbf{F}_{\mathrm{D}}$ is computed as

$$\mathbf{F}_{\mathrm{D}} = \left(\mathbf{H}\mathbf{F}_{\mathrm{R}}\right)^{\dagger}. \tag{32}$$

In fact, $\mathbf{F}_{\mathrm{D}}$ can also be computed as $\mathbf{F}_{\mathrm{D}} = \mathbf{F}_{\mathrm{R}}{}^{\dagger}\mathbf{F}_{\text{FD}}$ which we substitute in the objective function (31a) to get the analog precoding subproblem as [45],

$$\underset{\mathbf{F}_{\mathrm{R}}}{\arg\max} \quad \mathbf{Tr}\left[\mathbf{F}_{\mathrm{R}}^H\mathbf{F}_{\text{FD}}\mathbf{F}_{\text{FD}}^H\mathbf{F}_{\mathrm{R}}\right] \tag{33a}$$

$$\text{s.t.} \quad \left|\mathbf{F}_{\mathrm{R}_{i,j}}\right| = 1, \quad \forall i, j, \tag{33b}$$

$$\mathbf{F}_{\mathrm{R}}^H\mathbf{F}_{\mathrm{R}} = N_t\mathbf{I}_{\mathrm{M}_t}. \tag{33c}$$

The constraint (33c) makes sense as $\mathbf{F}_{\mathrm{R}}^H\mathbf{F}_{\mathrm{R}} \approx N_t\mathbf{I}_{\mathrm{M}_t}$ in mmWave MIMO as N_t is high. To solve problem (33), we

define $\bar{\mathbf{F}}_{\mathrm{R}} \triangleq \frac{1}{\sqrt{N_t}}\mathbf{F}_{\mathrm{R}}$ and rewrite the optimization problem (32) in terms of $\bar{\mathbf{F}}_{\mathrm{R}}$ as

$$\max_{\bar{\mathbf{F}}_{\mathrm{R}}} \quad \mathbf{Tr}\left[\bar{\mathbf{F}}_{\mathrm{R}}^{H}\left(\mathbf{F}_{\mathrm{FD}}\mathbf{F}_{\mathrm{FD}}^{H}\right)\bar{\mathbf{F}}_{\mathrm{R}}\right] \tag{34a}$$

$$\text{s.t.} \quad \left|\bar{\mathbf{F}}_{\mathrm{R}_{i,j}}\right| = \frac{1}{\sqrt{N_t}}, \ \forall i,j, \tag{34b}$$

$$\bar{\mathbf{F}}_{\mathrm{R}}^{H}\bar{\mathbf{F}}_{\mathrm{R}} = \mathbf{I}_{\mathrm{M}_t}, \tag{34c}$$

We compute $\mathbf{F}_{\mathrm{R}}$ as $\mathbf{F}_{\mathrm{R}} = \sqrt{N_t}\bar{\mathbf{F}}_{\mathrm{R}}$ after solving the problem (34). We use the iterative truncated SVD-based procedure proposed in [46] to solve the problem (33). The outline of the SVD-based algorithm is presented in Algorithm 5.

Algorithm 5 SVD-based Iterative Trace Maximization Method

Require: $\mathbf{F}_{\mathrm{FD}}, M_t$.

1: Initialize $\bar{\mathbf{F}}_{\mathrm{R}}^{(0)} = \frac{1}{\sqrt{N_t}}\exp\left(j\boldsymbol{\Theta}\right)$, where $\boldsymbol{\Theta}$ is $N_t \times M_t$ matrix and $\boldsymbol{\Theta}_{i,j}$ are random phase angles, and set $k = 1$.
2: **repeat**
3: Compute $\mathbf{A}^{(k)} = \bar{\mathbf{F}}_{\mathrm{R}}^{(k-1)^H}\left(\mathbf{F}_{\mathrm{FD}}\mathbf{F}_{\mathrm{FD}}^{H}\right)$.
4: Compute truncated SVD : $\mathbf{A}^{(k)} = \mathbf{U}^{(k)}\mathbf{S}^{(k)}\mathbf{V}_t^{(k)^H}$.
5: Compute $\mathbf{D}^{(k)} = \mathbf{V}_t^{(k)}\mathbf{U}^{(k)^H}$.
6: Compute $\bar{\mathbf{F}}_{\mathrm{R}}^{(k)} = \frac{1}{\sqrt{N_t}}\exp\left(j\angle\left(\mathbf{D}^{(k)}\right)\right)$.
7: $k \leftarrow k+1$.
8: **until** convergence, or $k \geq iter_{max}$, where $iter_{max}$ is the maximum of number of iterations.
9: Compute $\mathbf{F}_{\mathrm{R}} = \sqrt{N_t}\bar{\mathbf{F}}_{\mathrm{R}}^{(k)}$.
10: Calculate $\mathbf{F}_{\mathrm{D}}$, using (32).
11: **return** $\mathbf{F} = \mathbf{F}_{\mathrm{R}}\mathbf{F}_{\mathrm{D}}$.

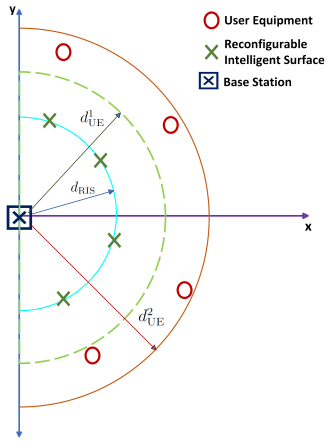


Fig. 2: The arrangement of BS, RISs and UEs for performance analysis.

V. Complexity of the Proposed Algorithm

The complexity of Algorithm 1 is $\mathcal{O}\left(\left(RN_I\right)^2\right)$ [42]. Similar to the Dinklebach's algorithm to determine $\left\{p_k\right\}_{k=1}^{k=K}$ in [35], the complexity of Algorithm 2 that computes $\left\{p_k\right\}_{k=1}^{k=K}$ is $N_p\mathcal{O}\left(K^q\right)$, where N_p is the number of iterations and $1 \leq q \leq 4$. Thus, the alternating optimization has complexity of $N_{AO}\mathcal{O}\left(\left(RN_I\right)^2 + N_pK^q\right)$ where N_{AO} is the number of iterations required during alternating optimization. The alternating optimization is followed by RIS switching off strategy which has a complexity of $RN_I\mathcal{O}\left(\left(K\left(RN_I\right)^2 + KRN_IN_t + N_pK^q + KN_t^2\right)\right)$. On the other hand, the hybrid precoding has complexity of $N_{Hy}\mathcal{O}\left(N_t^2K + 2N_tK^2\right)$ [46], where N_{Hy} is the number of iterations.

VI. Simulation Results

An mmWave MIMO with a single BS, R RISs and K single-antenna UEs is considered. The BS is placed at $(0, 0, z_B)$, where $z_B = 10$ m. All the RISs are placed on the positive side of x-axis at a distance d_{RIS} m from the BS. The exact placement of the RISs, however, is chosen randomly. The UEs are randomly placed on the positive side of x-axis between the distance d^1_{UE} m and d^2_{UE} m from the BS such that $d^2_{\mathrm{UE}} > d^1_{\mathrm{UE}} > d_{\mathrm{RIS}}$. We take $d^2_{\mathrm{UE}} = 180$ m, $d^1_{\mathrm{UE}} = 144$ m and $d_{\mathrm{RIS}} = 80$ m. The topview of the arrangement of BS, RISs and UEs looks like Fig. 2. The height of RIS is chosen to be 10 m, whereas the height of every UE is fixed to 1.8 m.

The BS is assumed to be equipped with uniform linear array(ULA) having N_t antennas, whereas the RIS utilizes uniform planar array (UPA) with $N_I = N_I^y \times N_I^z$ antennas. The number of RF chains at the BS, M_t is taken equal to K. All the antenna elements are separated by a distance of half wavelength. The noise power $\sigma_n^2 = -90$ dBm. The number of paths in the BS-RIS link and RIS-UE link is chosen to be $L = 7$ including the line of sight (LOS) path. The complex gain of the LOS path is $\alpha_1 \sim \mathcal{CN}(0, 10^{-\kappa})$ where κ indicates the path loss given by [47]

$$\kappa = a + 10b\ log_{10}(d_{TR}) + \xi, \tag{35}$$

where d_{TR} denotes the distance between the transmitter and the receiver, and $\xi \sim \mathcal{N}(0, \sigma_\xi^2)$. The values $a = 61.4, b = 2$, and $\sigma_\xi = 5.8$ dB are taken based on the results from real-world measurements [47].

A. Impact of Strategic Switching Off of RIS elements on Performance

In this section, we demonstrate how the strategic switching off of RIS elements helps enhance the energy efficiency. We consider two cases, one with a single RIS and the other with two RIS's with varying number of RIS elements to demonstrate the effect of switching off strategy on performance. We only consider the fully digital precoding at the transmitter. The energy efficiency of the system and the average number of switched ON elements per RIS are plotted as a function of the number of RIS elements per RIS in Fig. 3. The proposed method without using the switching off strategy is labeled 'All ON' and the proposed method with switching off strategy is labeled 'Switching Off Strategy' in the figure. It can be seen that the switching off strategy contributes to the energy efficiency, and its impact is more significant when there is only one RIS deployed.

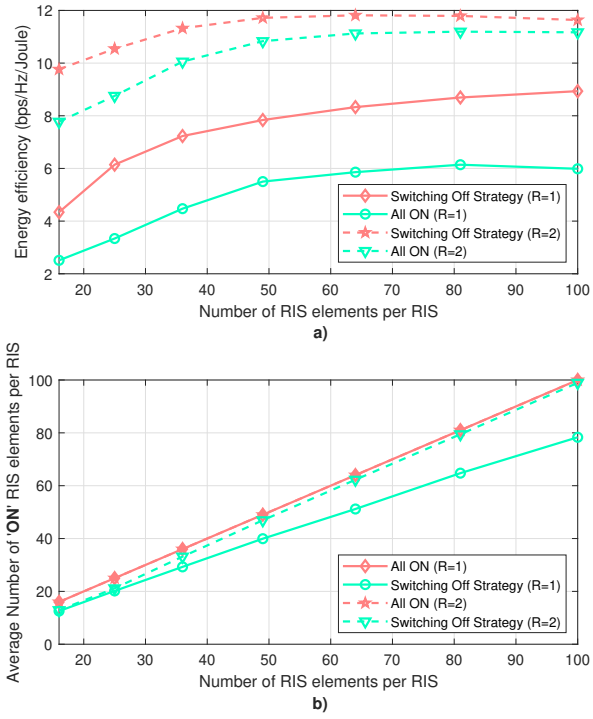


Fig. 3: Impact of switching off strategy *a*) Energy efficiency versus Number of RIS elements per RIS, *b*) Average Number of 'ON' RIS elements per RIS versus Number of RIS elements per RIS at $P_{max} = 40$ dBm, $\mathcal{R}_{min_k} = 4.01$ bps/Hz.

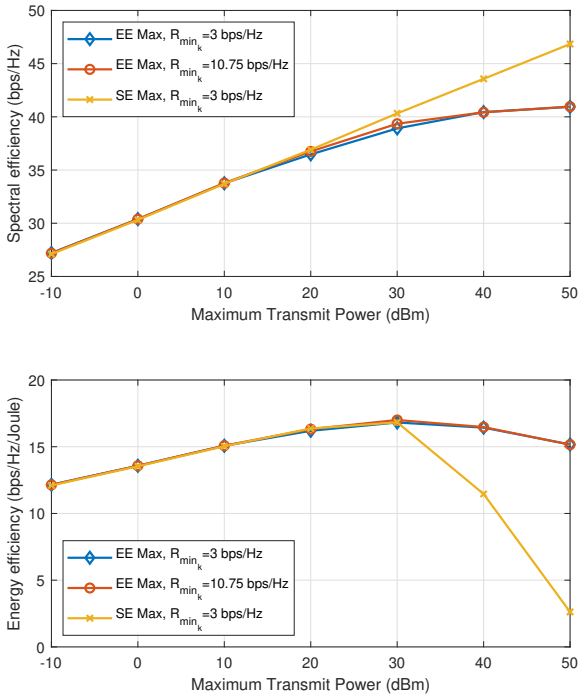


Fig. 4: Tradeoff between Spectral efficiency and Energy efficiency, $N_t = 32$, $N_I = 36$

B. Energy efficiency and Spectral efficiency tradeoff

In this section, we show that there is a tradeoff between energy efficiency and spectral efficiency. We establish this through the simulation results in Fig. 4. The "**EE Max**" in the figure represents the proposed method of maximizing energy efficiency and "**SE Max**" in the figure represents the method of maximizing sum spectral efficiency. $\mathcal{R}_{min_k}$ refers to the minimum spectral efficiency for the k^{th} user. We can realize that replacing P_T in (11) by 1 gives spectral efficiency of the system. Thus, the method of maximimizing spectral efficiency of the system can be developed in a way similar to that of maximizing energy efficiency. As transmit power is increased in Fig. 4, there is an increase in SINR, leading to increase in system spectral efficiency. Energy efficiency will also increase at the beginning because of increase in spectral efficiency. However, when the increase in power is higher, the increase in system's spectral efficiency will not be sufficient to increase energy efficiency. Energy efficiency will start to decrease instead.

From the figures, we can see that even though spectral efficiency is increasing with increase in power, energy efficiency starts decreasing after a point, 30 dBm in this case. This tells us that there is trade-off between system's spectral efficiency and energy efficiency.

C. Performance Analysis of Proposed Method

1) Hybrid Precoding's Performance: We compare the performance of the proposed hybrid precoding scheme with other existing hybrid precoding methods in Fig. 5 and Fig. 6. The chosen hybrid precoding techniques for comparison construct hybrid precoders from fully digital precoder by minimizing the Euclidean distance between the hybrid precoder and the optimal fully digital precoder. In Fig. 5 and Fig. 6, MO represents manifold optimization based hybrid precoder [48], MBCD represents hybrid precoder based on modified form of block coordinate descent method [45], and HD-LSR represents hybrid design by least squares relaxation (HD-LSR) [49].

For the hybrid precoders under consideration, the optimal fully digital precoder $\mathbf{F}_{\text{opt}}$ is constructed first to maximize energy efficiency using the proposed method. From $\mathbf{F}_{\text{FD}}$, analog precoder $\mathbf{F}_{\text{R}}$ is constructed using respective algorithms and the digital part of the hybrid precoder $\mathbf{F}_{\text{D}}$ is computed using the obtained analog precoder $\mathbf{F}_{\text{R}}$ as $\mathbf{F}_{\text{D}} = (\mathbf{H}\mathbf{F}_{\text{R}})^{\dagger}$. $\mathbf{F}_{\text{D}}$ may also be computed as $\mathbf{F}_{\text{D}} = \mathbf{F}_{\text{R}}^{\dagger}\mathbf{F}_{\text{FD}}$. Finally, the hybrid precoder is normalized and multiplied by a suitable factor so that the Frobenius norm of the hybrid precoder is same as that of the fully digital precoder as,

$$\mathbf{F}_{Hy} = \|\mathbf{F}_{\text{FD}}\sqrt{\mathbf{P}}\|_F \frac{\mathbf{F}_{\text{R}}\mathbf{F}_{\text{D}}}{\|\mathbf{F}_{\text{R}}\mathbf{F}_{\text{D}}\|_F}, \tag{36}$$

where $\|.\|_F$ represents Frobenius norm.

It can be seen in Fig. 5 and Fig. 6 that the performance of the proposed hybrid precoding method is better than the existing

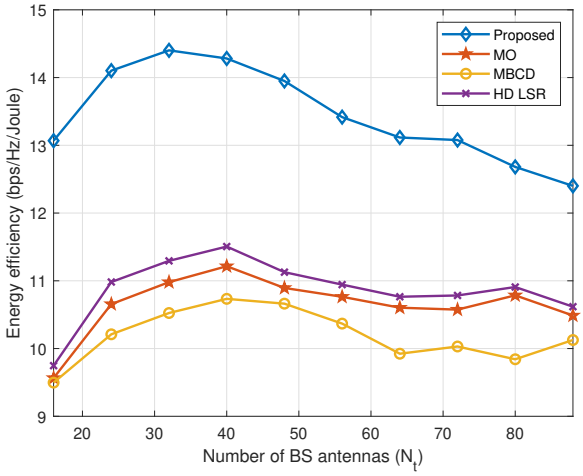


Fig. 5: Energy Efficiency Vs Number of BS antennas, $R = 2$, N_I = 36, $K = 8$, P_{max}= 30 dBm, $\mathcal{R}_{min_k}$ = 3.7 bps/Hz.

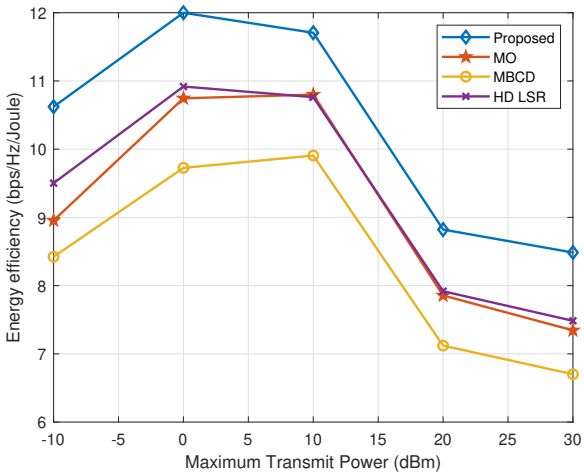


Fig. 6: Energy Efficiency Vs Maximum Transmit Power, N_t = 64, K = 8, R = 2, N_I = 36, $\mathcal{R}_{min_k}$ = 2.67 bps/Hz.

ones. In this paper, we not only determine the hybrid precoder but also the optimal power for each user to optimize energy efficiency, given the hybrid precoder. This also contributes hugely into achieving better energy efficiency performance in comparison to existing hybrid precoding methods.

2) Performance of the proposed method Vs existing method: We compare the performances of the proposed scheme with the existing method [35] and "BS Precoding Only" (both digital and hybrid) at the BS. For "BS Precoding Only" case, we choose the phase-shifts of RISs randomly and then determine the fully digital precoder that maximizes the energy efficiency in similar way as the proposed algorithm. We determine hybrid precoder from the fully digital precoder in a similar way as described in VI-C1. The curves for the proposed method are plotted with both the fully digital precoder and the proposed hybrid precoder, whereas the existing method's performance is plotted with only the fully digital precoder.

The energy efficiency and spectral efficiency performances

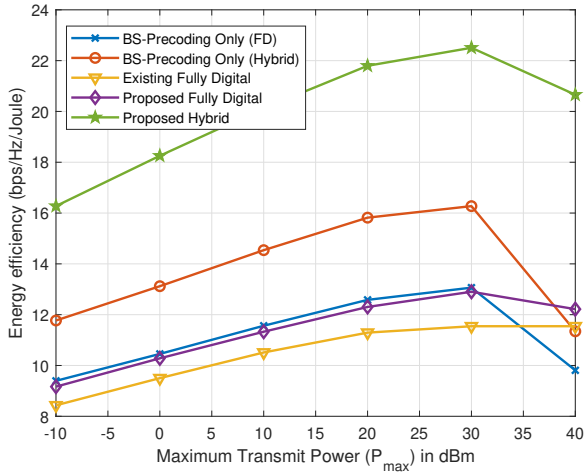


Fig. 7: Energy efficiency versus Maximum transmit power, $N_t = 64$, $R = 4$, $N_I = 36$, $K = 8$, $\mathcal{R}_{min_k} = 3$ bps/Hz.

as a function of maximum transmit power are portrayed in Fig. 7 and Fig. 8 respectively. The energy efficiency increases with maximum transmit power before gaining a maximum value and finally decreasing. With the increase in maximum transmit power, the BS can allocate more power to each user which naturally increases the spectral efficiency as proven by Fig. 8 as well. This, however, comes with increase in transmit power which lowers energy efficiency once the growth in spectral efficiency is eclipsed by the increase in transmit power.

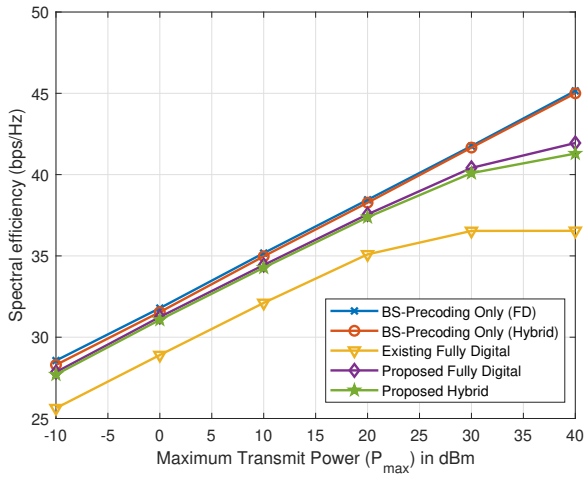


Fig. 8: Spectral efficiency versus Maximum transmit power, $N_t = 64$, $R = 4$, $N_I = 36$, $K = 8$, $\mathcal{R}_{min_k} = 3$ bps/Hz.

The energy efficiency plot of proposed method in Fig. 9 shows that the energy efficiency increases as the number of RISs grows and starts to decline after reaching the maximum. The higher number of RISs helps boost the spectral efficiency. But the power consumed is also higher with the larger number of RISs. The energy efficiency surges with the increasing number of RISs as long as the increase in spectral efficiency overcomes the increase in power consumed by the RISs. The energy efficiency plot as a function of number of RIS elements

per RIS in Fig. 10 is also similar to the Fig. 9. The increase in the number of RIS elements per RIS helps enhance the spectral efficiency but the power consumed by the RISs also increases at the same time. As a result, the energy efficiency saturates at some point and starts decreasing as the number of elements per RIS increases further.

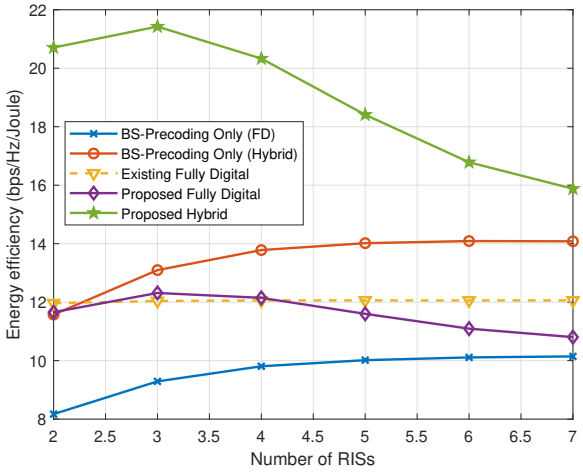


Fig. 9: Energy efficiency versus Number of RISs, $P_{max} = 40$ dBm, $\mathcal{R}_{min_k} = 3$ bps/Hz.

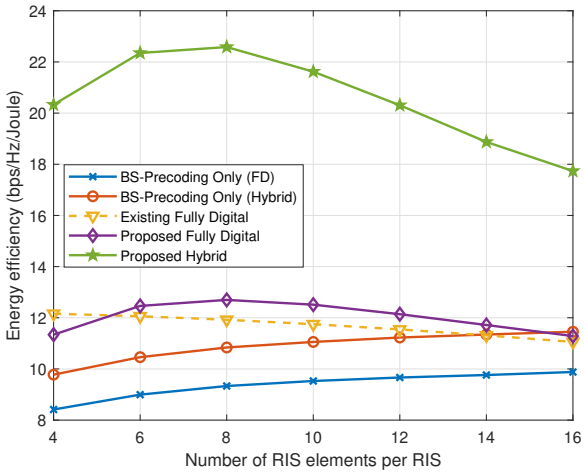


Fig. 10: Energy efficiency versus Number of elements per each RIS, $N_t = 64$, $R = 4$, $K = 8$,$P_{max} = 40$ dBm, $\mathcal{R}_{min_k} = 3$ bps/Hz.

The performance plots in Fig. 9 and Fig.10 can be explained better via channel hardening which is a phenomenon where a wireless channel behaves like a deterministic channel rather than a random one. The increase in number of RIS elements in a MIMO system results in channel hardening [50]. As a result, the spectral efficiency becomes more stable and does not increase much even as the RIS elements grow. Thus, energy efficiency increases with growing RIS elements in the beginning as spectral efficiency is increasing rapidly. But after a certain point, the channel starts to behave more deterministically, and spectral efficiency only grows by a small amount as the number of RIS elements increase. The power consumption increased because of the increase in number of RIS elements makes sure energy efficiency starts decreasing.

The Fig. 11 shows how energy efficiency varies when the QoS constraint in the form of minimum spectral efficiency $\mathcal{R}_{min_k}$ is increased. The minimum power required for each MS, p_{min_k} increases with $\mathcal{R}_{min_k}$. As long as the value of p_{min_k} is smaller enough, it does not have a say in the power allocated for each MS. But when p_{min_k} reaches a large enough value, it dictates the power allocated for each MS. Hence, it increases the spectral efficiency while decreasing the energy efficiency on the other hand, at higher values of $\mathcal{R}_{min_k}$. Thus, when $\mathcal{R}_{min_k}$ increases from 36.17 bps/Hz to 38.18 bps/Hz, there is a sudden rise in spectral efficiency which comes at the cost of abrupt rise in power consumed, causing a dip in energy efficiency. The decline is more pronounced in existing method compared to the proposed solution.

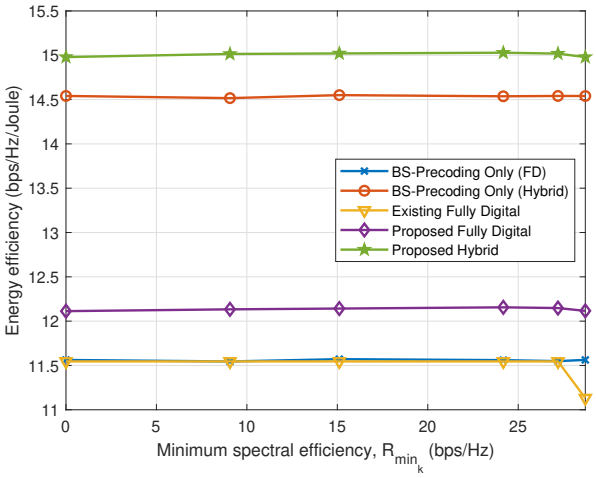


Fig. 11: Energy efficiency versus Minimum rate for each user, $N_t = 64$, $R = 4$, $N_I = 36$, $K = 8$, $P_{max} = 40$ dBm.

The Fig. 12 depicts the effect of the number of users on energy efficiency. There is an amplification in the sum spectral efficiency with the increase in number of users. The increase in power consumed by users is significantly lower than the increase in sum spectral efficiency, thus, increasing the energy efficiency with the number of users.

The increase in BS transmit antennas definitely enhances the spectral efficiency. But there is also an augmentation in the number of power amplifiers and the number of RF chains. There is no rise in the number of RF chains and power amplifiers but only a rise in the number of phase shifters where hybrid precoding is applied . The multiplication in the amount of power consumed outweighs the amplification of spectral efficiency. As a result, the energy efficiency decreases with the rising number of BS antennas as shown by Fig. 13.

The Fig. 14 exhibits the energy efficiency performance as the distance between BS and RIS is increased, keeping the distance between BS and the MS constant. As the distance between BS and RIS is increased, except the increase in pathloss between the BS and RIS (and decrease in pathloss between RIS and MS) all the parameters remains intact. There is no direct increase

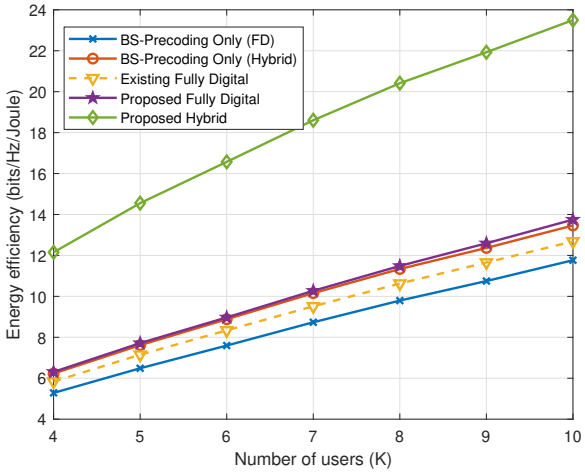


Fig. 12: Energy efficiency versus Number of users, $N_t = 64$, $R = 4$, $N_I = 36$, $P_{max} = 40$ dBm, $\mathcal{R}_{min_k} = 3$ bps/ Hz.

or decrease in power consumption as distance between BS and RIS is increased except through precoder and power allotted to different users which depend on the channel. As expected the energy efficiency remains roughly the same except some small changes.

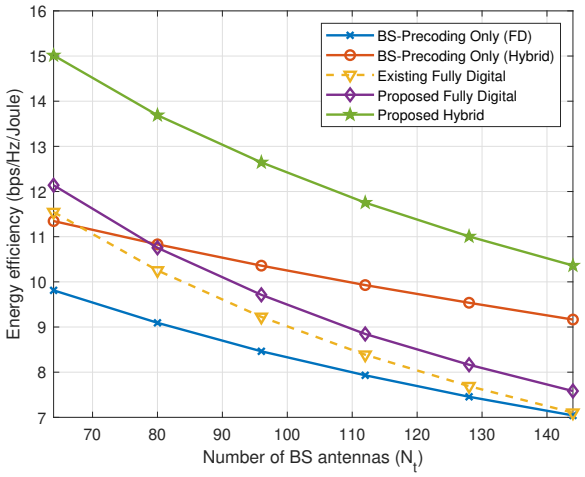


Fig. 13: Energy efficiency versus Number of BS transmit antennas, $R = 4$, $N_I = 36$, $K = 8$, $P_{max} = 40$ dBm, $\mathcal{R}_{min_k} = 3$ bps/ Hz.

From all the performance comparison plots, we can deduce that hybrid precoding provides better energy efficiency compared to fully digital precoding which is expected because of the fewer number of RF chains employed by the hybrid precoding. In general, the proposed method performs better than the existing method and the BS-Only precoding.

The proposed method, as well as the existing approach in [35], employs a ZF precoder and jointly optimizes the phase shift matrix $\mathbf{\Phi}$ and the power allocation matrix $\mathbf{P}$ to maximize energy efficiency via alternating optimization. In both cases, the subproblem for determining $\mathbf{\Phi}$ is formulated as a binary quadratic problem. However, the dimensionality of this subproblem in the proposed method is reduced to RN_I, in contrast to the higher dimensionality $(RN_I)^2$ formulation in [35], thereby achieving significant computational savings. To solve this subproblem, the proposed method adopts a greedy algorithm in phase, whereas [35] employs two alternative strategies: gradient descent search and sequential fractional programming. For the power allocation subproblem, the proposed method utilizes the method of Lagrange multipliers with KKT conditions, while [35] adopts the Dinkelbach algorithm. Furthermore, the proposed framework introduces an RIS switching-off strategy that selectively deactivates certain reflecting elements to further enhance energy efficiency, resulting in updated values of $\mathbf{\Phi}$ and $\mathbf{P}$. In addition, a hybrid precoding scheme is proposed, which provides additional improvements in energy efficiency.

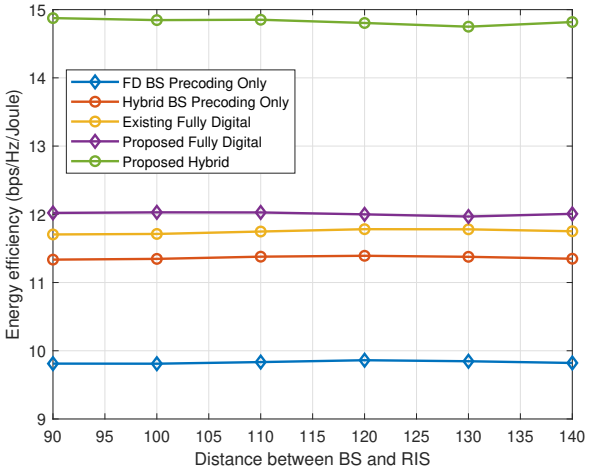


Fig. 14: Energy Efficiency Vs Distance between BS and RIS, $N_t = 64$, $R = 4$, $N_I = 36$, $K = 8$, $P_{max} = 40$ dBm, $\mathcal{R}_{min_k} = 3$ bps/ Hz.

VII. Conclusion

In this paper, an mmWave MU-MISO communication enhanced by the distributed RISs is considered, with the purpose of maximizing energy efficiency. The energy efficiency maximization problem is shrunk into a simplified problem of determining RIS phase shift coefficients and power allocated to the users with the use of fully digital ZF precoding at the BS. An algorithm that alternately optimizes RIS phase shift coefficients and the power allocation is proposed, opting for an alternating optimization route. A greedy algorithm in phase is used to determine RIS phase shift matrix, and an iterative algorithm based on the method of Lagrange multipliers using KKT conditions is developed to optimize power allocation matrix. Moreover, the energy efficiency is aggrandized by switching off some of the RIS elements by proposing a computationally efficacious RIS Switching Off strategy. The impacts of different parameters on energy efficiency are studied through simulations. The simulation results divulge that the proposed algorithm manages to achieve better energy efficiency as compared to the existing solution. It would be interesting to explore the energy efficiency maximization problem by

including the selection of active or "ON" RIS elements in the optimization problem itself, and we would investigate it in our future work.

REFERENCES

[1] W. H. Chin, Z. Fan, and R. Haines, "Emerging technologies and research challenges for 5G wireless networks," *IEEE Wireless Communications*, vol. 21, no. 2, pp. 106–112, 2014.

[2] A. L. Swindlehurst, E. Ayanoglu, P. Heydari, and F. Capolino, "Millimeter-wave massive MIMO: The next wireless revolution?" *IEEE Communications Magazine*, vol. 52, no. 9, pp. 56–62, Sep. 2014.

[3] T. S. Rappaport, J. N. Murdock, and F. Gutierrez, "State of the art in 60-ghz integrated circuits and systems for wireless communications," *Proceedings of the IEEE*, vol. 99, no. 8, pp. 1390–1436, 2011.

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

[5] A. Ghosh, T. A. Thomas, M. C. Cudak, R. Ratasuk, P. Moorut, F. W. Vook, T. S. Rappaport, G. R. MacCartney, S. Sun, and S. Nie, "Millimeter-wave enhanced local area systems: A high-data-rate approach for future wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1152–1163, 2014.

[6] C.-X. Wang, F. Haider, X. Gao, X.-H. You, Y. Yang, D. Yuan, H. M. Aggoune, H. Haas, S. Fletcher, and E. Hepsaydir, "Cellular architecture and key technologies for 5g wireless communication networks," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 122–130, 2014.

[7] T. E. Bogale and L. B. Le, "Beamforming for multiuser massive MIMO systems: Digital versus hybrid analog-digital," in *2014 IEEE Global Communications Conference*, 2014, pp. 4066–4071.

[8] M. A. ElMossallamy, H. Zhang, L. Song, K. G. Seddik, Z. Han, and G. Y. Li, "Reconfigurable intelligent surfaces for wireless communications: Principles, challenges, and opportunities," *IEEE Transactions on Cognitive Communications and Networking*, vol. 6, no. 3, pp. 990–1002, 2020.

[9] P. Wang, J. Fang, L. Dai, and H. Li, "Joint transceiver and large intelligent surface design for massive MIMO mmwave systems," *IEEE Transactions on Wireless Communications*, vol. 20, no. 2, pp. 1052–1064, 2021.

[10] Q. Wu, S. Zhang, B. Zheng, C. You, and R. Zhang, "Intelligent reflecting surface-aided wireless communications: A tutorial," *IEEE Transactions on Communications*, vol. 69, no. 5, pp. 3313–3351, 2021.

[11] E. Bj"ørnson, O. "Øzdogan, and E. G. Larsson, "Reconfigurable intelligent surfaces: Three myths and two critical questions," *IEEE Communications Magazine*, vol. 58, no. 12, pp. 90–96, 2020.

[12] E. Basar, "Transmission through large intelligent surfaces: A new frontier in wireless communications," in *2019 European Conference on Networks and Communications (EuCNC)*, 2019, pp. 112–117.

[13] E. Basar, M. Di Renzo, J. De Rosny, M. Debbah, M.-S. Alouini, and R. Zhang, "Wireless communications through reconfigurable intelligent surfaces," *IEEE Access*, vol. 7, pp. 116 753–116 773, 2019.

[14] C. Liaskos, S. Nie, A. Tsioliaridou, A. Pitsillides, S. Ioannidis, and I. Akyildiz, "A new wireless communication paradigm through software-controlled metasurfaces," *IEEE Communications Magazine*, vol. 56, no. 9, pp. 162–169, 2018.

[15] M. D. Renzo, M. Debbah, D.-T. Phan-Huy, A. Zappone, M.-S. Alouini, C. Yuen, V. Sciancalepore, G. C. Alexandropoulos, J. Hoydis, H. Gacanin *et al.*, "Smart radio environments empowered by reconfigurable ai meta-surfaces: An idea whose time has come," *EURASIP Journal on Wireless Communications and Networking*, vol. 2019, no. 1, pp. 1–20, 2019.

[16] M. Di Renzo, K. Ntontin, J. Song, F. H. Danufane, X. Qian, F. Lazarakis, J. De Rosny, D.-T. Phan-Huy, O. Simeone, R. Zhang, M. Debbah, G. Lerosey, M. Fink, S. Tretyakov, and S. Shamai, "Reconfigurable intelligent surfaces vs. relaying: Differences, similarities, and performance comparison," *IEEE Open Journal of the Communications Society*, vol. 1, pp. 798–807, 2020.

[17] M. Di Renzo, A. Zappone, M. Debbah, M.-S. Alouini, C. Yuen, J. de Rosny, and S. Tretyakov, "Smart radio environments empowered by reconfigurable intelligent surfaces: How it works, state of research, and the road ahead," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 11, pp. 2450–2525, 2020.

[18] X. Yuan, Y.-J. A. Zhang, Y. Shi, W. Yan, and H. Liu, "Reconfigurable-intelligent-surface empowered wireless communications: Challenges and opportunities," *IEEE Wireless Communications*, vol. 28, no. 2, pp. 136–143, 2021.

[19] C. Pan, H. Ren, K. Wang, W. Xu, M. Elkashlan, A. Nallanathan, and L. Hanzo, "Multicell MIMO communications relying on intelligent reflecting surfaces," *IEEE Transactions on Wireless Communications*, vol. 19, no. 8, pp. 5218–5233, 2020.

[20] Q. Wu and R. Zhang, "Towards smart and reconfigurable environment: Intelligent reflecting surface aided wireless network," *IEEE Communications Magazine*, vol. 58, no. 1, pp. 106–112, 2020.

[21] B. Ning, Z. Chen, W. Chen, and J. Fang, "Beamforming optimization for intelligent reflecting surface assisted MIMO: A sum-path-gain maximization approach," *IEEE Wireless Communications Letters*, vol. 9, no. 7, pp. 1105–1109, 2020.

[22] P. Wang, J. Fang, X. Yuan, Z. Chen, and H. Li, "Intelligent reflecting surface-assisted millimeter wave communications: Joint active and passive precoding design," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 12, pp. 14 960–14 973, 2020.

[23] H. Guo, Y.-C. Liang, J. Chen, and E. G. Larsson, "Weighted sum-rate maximization for reconfigurable intelligent surface aided wireless networks," *IEEE Transactions on Wireless Communications*, vol. 19, no. 5, pp. 3064–3076, 2020.

[24] Z. Zhang and L. Dai, "A joint precoding framework for wideband reconfigurable intelligent surface-aided cell-free network," *IEEE Transactions on Signal Processing*, vol. 69, pp. 4085–4101, 2021.

[25] S. Zhang and R. Zhang, "Capacity characterization for intelligent reflecting surface aided MIMO communication," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 8, pp. 1823–1838, 2020.

[26] H. Kasai, "Fast optimization algorithm on complex oblique manifold for hybrid precoding in millimeter wave mimo systems," in *2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, 2018, pp. 1266–1270.

[27] M.-M. Zhao, A. Liu, and R. Zhang, "Outage-constrained robust beamforming for intelligent reflecting surface aided wireless communication," *IEEE Transactions on Signal Processing*, vol. 69, pp. 1301–1316, 2021.

[28] R. Liu, M. Li, Q. Liu, and A. L. Swindlehurst, "Joint symbol-level precoding and reflecting designs for irs-enhanced mu-miso systems," *IEEE Transactions on Wireless Communications*, vol. 20, no. 2, pp. 798–811, 2021.

[29] B. Yang, X. Cao, C. Huang, C. Yuen, M. Di Renzo, Y. L. Guan, D. Niyato, L. Qian, and M. Debbah, "Federated spectrum learning for reconfigurable intelligent surfaces-aided wireless edge networks," *IEEE Transactions on Wireless Communications*, vol. 21, no. 11, pp. 9610–9626, 2022.

[30] N. Baskar, P. Selvaprabhu, V. B. Kumaravelu, S. Chinnadurai, V. Rajamani, V. Menon U, and V. Kumar C, "A survey on resource allocation and energy efficient maximization for IRS-aided MIMO wireless communication," *IEEE Access*, vol. 12, pp. 85 423–85 454, 2024.

[31] A. Bereyhi, S. Asaad, and R. R. Mueller, "Stepwise transmit antenna selection in downlink massive multiuser MIMO," in *WSA 2018; 22nd International ITG Workshop on Smart Antennas*, 2018, pp. 1–8.

[32] S. Asaad, A. M. Rabiei, and R. R. Müller, "Massive MIMO with antenna selection: Fundamental limits and applications," *IEEE Transactions on Wireless Communications*, vol. 17, no. 12, pp. 8502–8516, 2018.

[33] B. Yang, X. Cao, C. Huang, C. Yuen, L. Qian, and M. D. Renzo, "Intelligent spectrum learning for wireless networks with reconfigurable intelligent surfaces," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 4, pp. 3920–3925, 2021.

[34] B. Yang, X. Cao, C. Huang, Y. L. Guan, C. Yuen, M. Di Renzo, D. Niyato, M. Debbah, and L. Hanzo, "Spectrum-learning-aided reconfigurable intelligent surfaces for "green" 6g networks," *IEEE Network*, vol. 35, no. 6, pp. 20–26, 2021.

[35] C. Huang, A. Zappone, G. C. Alexandropoulos, M. Debbah, and C. Yuen, "Reconfigurable intelligent surfaces for energy efficiency in wireless communication," *IEEE Transactions on Wireless Communications*, vol. 18, no. 8, pp. 4157–4170, 2019.

[36] Q. Wu and R. Zhang, "Beamforming optimization for wireless network aided by intelligent reflecting surface with discrete phase shifts," *IEEE Transactions on Communications*, vol. 68, no. 3, pp. 1838–1851, 2020.

[37] Z. Yang, M. Chen, W. Saad, W. Xu, M. Shikh-Bahaei, H. V. Poor, and S. Cui, "Energy-efficient wireless communications with distributed reconfigurable intelligent surfaces," *IEEE Transactions on Wireless Communications*, vol. 21, no. 1, pp. 665–679, 2022.

[38] V. Venkateswaran and R. Krishnan, "Hybrid analog and digital precoding: From practical RF system models to information theoretic bounds," in *2016 IEEE Globecom Workshops (GC Wkshps)*. IEEE, 2016, pp. 1–6.

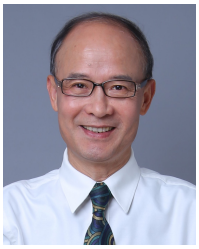
[39] I. Csiszár, "Information geometry and alternating minimization procedures," *Statistics and decisions*, vol. 1, pp. 205–237, 1984.

[40] J. C. Bezdek and R. J. Hathaway, "Some notes on alternating optimization," in *Advances in Soft Computing — AFSS 2002*, N. R. Pal and

M. Sugeno, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2002, pp. 288–300.
[41] X.-D. Zhang, *Matrix analysis and applications*. Cambridge University Press, 2017.
[42] I. Waldspurger, A. d'Aspremont, and S. Mallat, "Phase recovery, MaxCut and complex semidefinite programming," *Math. Program.*, vol. 149, no. 1-2, pp. 47–81, 2015. [Online]. Available: https://doi.org/10.1007/s10107-013-0738-9
[43] T. Guo, X. Li, M. Mei, Z. Yang, J. Shi, K.-K. Wong, and Z. Zhang, "Joint communication and sensing design in coal mine safety monitoring: 3-d phase beamforming for ris-assisted wireless networks," *IEEE Internet of Things Journal*, vol. 10, no. 13, pp. 11 306–11 315, 2023.
[44] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Transactions on Wireless Communications*, vol. 13, no. 3, pp. 1499–1513, March 2014.
[45] P. R. Gautam and L. Zhang, "Hybrid precoding for millimeter wave MIMO: Trace optimization approach," *IEEE Access*, vol. 10, pp. 66 874–66 885, 2022.
[46] P. R. Gautam, L. Zhang, and P. Fan, "Hybrid MMSE precoding for millimeter wave MU-MISO via trace maximization," *IEEE Transactions on Wireless Communications*, vol. 23, no. 3, pp. 1999–2010, 2024.
[47] M. R. Akdeniz, Y. Liu, M. K. Samimi, S. Sun, S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1164–1179, 2014.
[48] X. Yu, J. Shen, J. Zhang, and K. B. Letaief, "Alternating minimization algorithms for hybrid precoding in millimeter wave MIMO systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 3, pp. 485–500, April 2016.
[49] C. Rusu, R. Mèndez-Rial, N. González-Prelcic, and R. W. Heath, "Low complexity hybrid precoding strategies for millimeter wave communication systems," *IEEE Transactions on Wireless Communications*, vol. 15, no. 12, pp. 8380–8393, 2016.
[50] A. Bereyhi, S. Asaad, C. Ouyang, R. R. Müller, R. F. Schaefer, and H. V. Poor, "Channel hardening of IRS-aided multi-antenna systems: How should IRSs scale?" *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 8, pp. 2321–2335, 2023.

Prabhat Raj Gautam (M'23) received the Master of Technology degree in communications engineering from Indian Institute of Technology (IIT) Delhi, India in 2015, and the Ph.D. degree in Electronic and Electrical Engineering from University of Leeds, UK in 2023. He is currently a Research Fellow at the Department of Computing, Imperial College London. His research focuses on wireless communications, with particular interest in millimeter-wave communications, reconfigurable intelligent surface (RIS)-enhanced communications, and integrated sensing and communication (ISAC). He received the Best Student Paper Award at the International Symposium on Wireless Personal Multimedia Communications (WPMC2019), in 2019. He was also awarded the F.W. Carter Prize for the best Ph.D. thesis at the University of Leeds.

Li Zhang (SM'12) Li Zhang is an Associate Professor in Communications at the School of Electronic and Electrical Engineering, University of Leeds. She leads the Wireless Communication Group within the Institute of Communication and Power Networks (ICaPNet). Her research focuses on wireless communications, including massive MIMO, mmWave and THz communications, Non-Terrestrial Networks (NTNs), mobile edge computing and 5G/6G systems. Her group is also actively investigating wireless communication challenges in high-speed railway systems. She has served on the Technical Programme Committees of most major IEEE communications conferences since 2006 and is currently an associate editor for several IEEE journals. She has been a member of the UK EPSRC Peer Review College since 2006 and has served on grant review panels for both EPSRC and Royal Society. She also regularly reviews grant proposals for international funding bodies, including the Qatar National Research Fund, and agencies in Denmark and France, as well as book proposals for various academic publishers. She has acted as a PhD examiner for numerous universities within and outside the UK and has served as an external examiner for MSc programmes at other UK institutions. She received a Nuffield Award for newly appointed lecturers in 2005, became a Fellow of the Higher Education Academy in 2006, and was elevated to IEEE Senior Member in 2012.



Pingzhi Fan (M'93-SM'99-F'15) received the M.Sc. degree in computer science from Southwest Jiaotong University, China, in 1987, and the Ph.D. degree in electronic engineering from Hull University, U.K., in 1994. He is currently a distinguished professor of Southwest Jiaotong University (SWJTU), honorary dean of the SWJTU-Leeds Joint School (2015-), honorary professor of the University of Nottingham (Ningbo, 2025), and a visiting professor of Leeds University, UK (1997-). He is a recipient of the UK ORS Award (1992), the National Science Fund for Distinguished Young Scholars (1998, NSFC), IEEE VT Society Jack Neubauer Memorial Award (2018), IEEE SP Society SPL Best Paper Award (2018), IEEE VT Society Best Magazine Paper Award (2023), and several IEEE conference best paper awards. He served as chief scientist of a National 973 Plan Project (MoST, 2012.1-2016.12). He also served as general chair or TPC chair of a number of IEEE conferences, including VTC2016Spring, ITW2018, IWSDA2022, PIMRC2023, as well as the coming VTC2025Fall, ISIT2026, ICC2028. His research interests include high mobility wireless communications, multiple access techniques, ISAC, signal design and coding, etc. He is an IEEE VTS Distinguished Speaker (2022-2028), a fellow of IEEE, IET, CIE and CIC.