Energy-Efficient Reconfigurable Holographic Surfaces Operating in the Presence of Realistic Hardware Impairments

Qingchao Li, *Graduate Student Member, IEEE*, Mohammed El-Hajjar, *Senior Member, IEEE*,
Yanshi Sun, *Member, IEEE*, Ibrahim Hemadeh, *Senior Member, IEEE*,
Arman Shojaeifard, *Senior Member, IEEE*, and Lajos Hanzo, *Life Fellow, IEEE*

Abstract—Reconfigurable holographic surfaces (RHSs) constitute a promising technique of supporting energy-efficient communications. In this paper, we formulate the energy efficiency maximization problem of the switch-controlled RHS-aided beamforming architecture by alternately optimizing the holographic beamformer at the RHS, the digital beamformer, the total transmit power and the power sharing ratio of each user. Specifically, to deal with this challenging non-convex optimization problem, we decouple it into three sub-problems. Firstly, the coefficients of RHS elements responsible for the holographic beamformer are optimized to maximize the sum of the eigen-channel gains of all users by our proposed low-complexity eigen-decomposition (ED) method. Then, the digital beamformer is designed by the singular value decomposition (SVD) method to support multi-user information transfer. Finally, the total transmit power and the power sharing ratio are alternately optimized, while considering the effect of transceiver hardware impairments (HWI). We theoretically derive the spectral efficiency and energy efficiency performance upper bound for the RHS-based beamforming architectures in the presence of HWIs. Our simulation results show that the switch-controlled RHS-aided beamforming architecture achieves higher energy efficiency than the conventional fully digital beamformer and the hybrid beamformer based on phase shift arrays (PSA). Moreover, considering the effect of HWI in the beamforming design can bring about further energy efficiency enhancements.

Index Terms—Reconfigurable holographic surfaces (RHS), hybrid beamforming, energy efficiency, eigen-decomposition (ED), hardware impairment (HWI).

I. INTRODUCTION

DRIVEN by the ever-increasing demand for data sharing across mobile networks, sophisticated technologies, such as millimeter wave (mmWave) communications [1], [2], massive multiple-input-multiple-output (MIMO) systems [3],

Qingchao Li, Mohammed El-Hajjar and Lajos Hanzo are with the School of Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, U.K. (e-mail: qingchao.li@soton.ac.uk; meh@ecs.soton.ac.uk; lh@ecs.soton.ac.uk).

Yanshi Sun is with the School of Computer Science and Information Engineering, Hefei University of Technology, Hefei, 230009, China. (e-mail:sys@hfut.edu.cn).

Ibrahim Hemadeh and Arman Shojaeifard are with InterDigital, London EC2A 3QR, U.K. (e-mail: ibrahim.hemadeh@interdigital.com; arman.shojaeifard@interdigital.com).

[4], [5], [6], and ultra-dense networks (UDN) [7], [8] have been rolled out across the globe. Hence, at the time of writing research is also well under way for the exploration of next-generation communications [9].

In this context, holographic MIMO systems are expected to evolve toward an intelligent and software reconfigurable paradigm enabling ubiquitous communications between humans and mobile devices in support of improved spectral efficiency [10].

From the perspective of holographic MIMO channel models, Pizzo *et al.* [11] modelled the far-field small-scale fading of holographic MIMO systems as a zero-mean, spatially-stationary, and correlated Gaussian scalar random field, where a Fourier plane-wave based spectral representation is employed to describe the three-dimensional small-scale fading. Furthermore, the near-field small-scale fading of holographic MIMO systems is derived in [12], where a novel Fourier plane-wave stochastic channel model is conceived, which is mathematically tractable and consistent with the physics of wave propagation. In [13], Demir *et al.* proposed a spatial correlation model to characterize the correlated Rayleigh fading channel, considering both non-isotropic scattering and directive antennas. In [14], Zeng *et al.* considered a single-user holographic MIMO system model, where a reconfigurable refractive surface is deployed at the BS empowered by a single feed, and its downlink data rate is derived.

As a further advance, a reconfigurable holographic surface (RHS)-aided beamforming architecture is proposed in [15], [16]. Specifically, the RHS is composed of feeds and meta-material based radiation elements, in which the feeds connect the radio frequency (RF) chains and the RHS circuit. These schemes transform the input signals into electromagnetic waves, and the coefficients of the radiation elements are controlled electrically to achieve holographic beamforming. In [17], [18], [19], Deng *et al.* proposed an RHS-based hybrid beamforming scheme, where the digital beamformer and the holographic beamformer are realized at the BS and the RHS, respectively. Specifically, the digital beamformer relies on the state-of-the-art zero-forcing (ZF) precoding method, while the holographic beamforming is performed by configuring the amplitude-controlled RHS radiation elements. To maximize the achievable sum-rate, the digital beamformer and the holographic beamformer are alternately optimized. The simulation results showed that the proposed RHS-based hybrid beam-

former achieves better sum-rate than the state-of-the-art massive MIMO hybrid beamformer based on phase shift arrays. In [20], Hu *et al.* proposed a holographic beamforming scheme based on amplitude-controlled RHS elements having limited resolution, and it showed that the holographic beamforming scheme associated with 2-bit quantized amplitude resolution approached the sum-rate associated with continuous amplitude values. In [21], an amplitude-controlled RHS beamformer was harnessed for ultra-dense low-Earth-orbit (LEO) satellite communications to compensate for the severe path-loss of satellite communications.

The above treatises on RHS-aided systems have the following limitations. Firstly, the objective function in these works mainly focused on maximizing the data rate, whilst ignoring the optimization of the energy efficiency. However, according to the report from the Radiocommunications Sector of the International Telecommunication Union (ITU-R) M.2516-0, improving the energy efficiency is one of the key drivers for International Mobile Telecommunications (IMT) systems designed for sustainable development. Power-efficient solutions are needed both for the backhaul and for local access to make use of small-scale renewable energy sources. Secondly, perfect transceiver hardware is assumed, i.e. the practical signal impairments of the transceivers are ignored. However, the hardware impairments (HWI) may impose a performance floor at high transmit powers, both in terms of the achievable rate, outage probability and bit error ratio (BER) [22], [23], [24], [25], [26], [27], [28]. Hence the attainable performance improvement of wireless communication systems is limited by the HWI, which cannot be compensated upon implying increasing the transmit power. To deal with the above issues, our contributions in this paper can be summarized as follows:

- We formulate an energy efficiency maximization problem for the switch-controlled RHS-enabled wireless communication systems operating in the face of realistic transceiver HWIs. More explicitly, we optimize the holographic beamformer, the digital beamformer, the total transmit power, and the power allocation ratio of each user. We then solve the problem by decomposing it into three sub-problems. Firstly, in the holographic beamformer, the optimal ON-OFF state of each RHS radiation element is found based on our proposed low-complexity eigen-decomposition (ED) method to maximize the sum of the eigen-channel gains of all users. Then, the digital beamformer is designed based on the singular value decomposition (SVD) method to support multi-user information transfer. Finally, the optimal transmit power and the power sharing ratio are alternately optimized. Specifically, when the total transmit power is given, the proportion of power allocated for each user is optimized via our proposed power sharing algorithm, while considering the effect of HWIs. By contrast, when the power share of each user is determined, the total transmit power is optimized based on the gradient descent method.
- To investigate the effect of HWI, we theoretically derive the spectral efficiency and energy efficiency upper bound for the RHS-based beamforming architectures in the presence of HWIs.
- The simulation results show that the switch-controlled RHS-aided beamforming architecture achieves higher energy efficiency than the conventional digital beamformer and the hybrid beamformer based on a phase shift array. Furthermore, we show that considering the effect of HWIs in the beamforming design yields further energy efficiency enhancements.

The rest of this paper is organized as follows. In Section II we present the system model, while our reconfigurable holographic surface-based hybrid beamformer design is described in Section II. Our simulation results are presented in Section IV, while we conclude in Section V.

Notations: Vectors and matrices are denoted by boldface lower and upper case letters, respectively; $(\cdot)^{\mathrm{T}}$, $(\cdot)^{\dagger}$ and $(\cdot)^{\mathrm{H}}$ represent the operation of transpose, conjugate and Hermitian transpose, respectively; $|a|$ and $\angle a$ represent the amplitude and angle of the complex scalar a, respectively; $\mathbf{1}_n$ represent the $n \times 1$ one vector; $\mathbb{R}^{m \times n}$ and $\mathbb{C}^{m \times n}$ denotes the space of $m \times n$ real-valued and complex-valued matrices, respectively; a_n represents the nth element in vector $\mathbf{a}$, and $[\mathbf{A}]_{m,n}$ represents the (m, n)th element in matrix $\mathbf{A}$; $\mathbf{Diag}\{a_1, a_2, \cdots, a_N\}$ denotes a diagonal matrix with the diagonal elements being the elements of $a_1, a_2, \cdots, a_N$ in order; $\mathrm{rank}(\mathbf{A})$ denotes the rank of matrix $\mathbf{A}$; $\mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is a circularly symmetric complex Gaussian random vector with the mean $\boldsymbol{\mu}$ and the covariance matrix $\boldsymbol{\Sigma}$; $\mathbf{A} \succeq \mathbf{0}$ means $\mathbf{A}$ is a semi-definite matrix; $\mathbf{0}_n$ represents the $n \times 1$ zeros vector and $\mathbf{I}_n$ represents the $n \times n$ identical matrix; $\odot$ denotes Hadamard product operation, and $\mathcal{O}$ represents the time complexity.

II. SYSTEM MODEL

In this section, we briefly highlight the energy-efficient switch-controlled RHS-aided beamforming architecture of [15], [16], [17], [18], [19], [20], [21], followed by the channel model.

A. Reconfigurable Holographic Surface aided Beamforming Architecture

The physical principle of RHS is different from some other low-power aperture antenna designs, such as the dynamic metasurface antennas (DMA) [29], [30], [31], [32]. Specifically, the DMA consists of multiple microstrips and each microstrip is composed of numerous metamaterial elements. In the DMA, the information is beamformed by linearly combining the radiation observation from all metamaterial elements in each microstrip. By contrast, the RHS is a special leaky-wave antenna and the information is beamformed based on an exciting wave circuit and a radiation element configuration controller.

Referring to [15], [16], [17], [18], [19], [20], [21], the architecture of the RHS-based hybrid beamformer is shown in Fig. 1, including both the digital and the holographic beamformer.
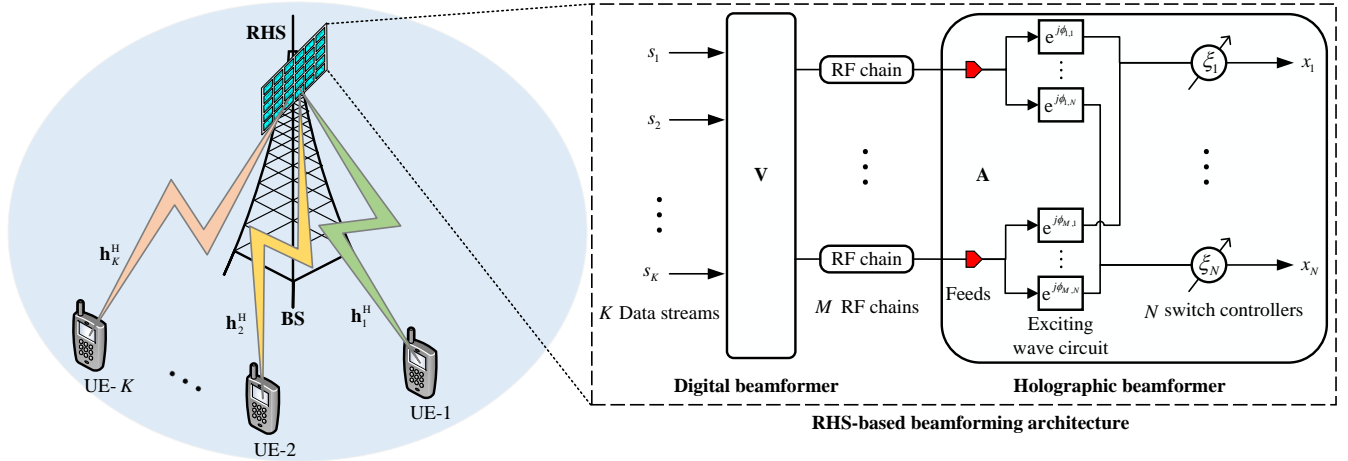
Fig. 1: System model of the RHS-based hybrid beamforming architecture.

1) Digital beamformer architecture: Firstly, the data stream vector $\mathbf{s} = [s_1, s_2, \cdots, s_K]^{\mathrm{T}} \in \mathbb{C}^{K \times 1}$ is processed by a power allocation matrix $\mathbf{P} \in \mathbb{C}^{K \times K}$ and a digital beamformer $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_K] \in \mathbb{C}^{M \times K}$, where K is the number of data streams. Then, the resultant signal is up-converted to the carrier frequency by M RF chains, where we have $\mathbb{E}[|s_k|^2] = 1$, $\mathbf{P} = \mathbf{Diag}\{p_1, p_2, \cdots, p_K\}$ satisfying $\sum_{k=1}^{K} p_k = 1$, and $\|\mathbf{v}_k\|^2 = 1$.

2) Holographic beamformer architecture: In the RHS-based holographic beamformer, each RF chain is connected to a feed of the RHS, and the feeds convert the carrier frequency current into an electromagnetic wave, which propagates through the waveguide of the reconfigurable holographic surface and radiates the energy into the free space from the metamaterial radiation elements. Specifically, at the nth RHS element, the desired wave propagates in the target direction characterized by $\Psi_{\mathrm{obj}}(\mathbf{r}_n) = \mathrm{e}^{-j\mathbf{k}_n^{\mathrm{T}} \cdot \mathbf{r}_n}$, while the reference wave generated by the mth feed is $\Psi_{\mathrm{ref}}(\mathbf{r}_n^m) = \mathrm{e}^{-j\mathbf{k}_0^{\mathrm{T}} \cdot \mathbf{r}_n^m}$, where $\mathbf{k}_n$ is the desired directional propagation vector emerging from the nth RHS element, $\mathbf{k}_0$ is the propagation vector of the reference wave, $\mathbf{r}_n$ is the position vector of the nth RHS element, and $\mathbf{r}_n^m$ is the position vector representing the link spanning from the mth feed to the nth RHS element. As shown in Fig. 1, the holographic beamformer matrix, denoted as $\mathbf{A}$, is composed of the fixed exciting wave circuit matrix $\mathbf{\Phi} \in \mathbb{C}^{N \times M}$ and the diagonal reconfigurable controller matrix $\mathbf{F} \in \mathbb{C}^{N \times N}$, which can be represented as [15], [16], [17], [18], [19], [20], [21]:

$$\mathbf{A} = \mathbf{F}\mathbf{\Phi}. \tag{1}$$

In the exciting wave circuit, $[\mathbf{\Phi}]_{n,m} = \frac{1}{\sqrt{N}}\mathrm{e}^{j\phi_{n,m}}$ represents the exciting wave emanating from the mth feed to the nth RHS element. The holographic pattern in [15], [16], [17], [18], [19], [20], [21] is excited by the reference wave, i.e. $[\mathbf{\Phi}]_{n,m} = \frac{1}{\sqrt{N}}\mathrm{e}^{-j\mathbf{k}_0^{\mathrm{T}} \cdot \mathbf{r}_n^m}$. To avoid the impact of the feed and RHS element location on the beamforming, we design the exciting wave circuit based on the discrete Fourier transform (DFT) matrix as follows. In the exciting wave circuit between the mth feed and the nth RHS element, we apply a calibration

phase shifter of $\Delta\phi_{n,m} = \frac{(n-1)(m-1)}{N} + \mathbf{k}_0^{\mathrm{T}} \cdot \mathbf{r}_n^m$ to the reference wave, and then the generated exciting wave is given by

$$[\mathbf{\Phi}]_{n,m} = \frac{1}{\sqrt{N}}\mathrm{e}^{j\left(-\mathbf{k}_0^{\mathrm{T}} \cdot \mathbf{r}_n^m + \Delta\phi_{n,m}\right)} = \frac{1}{\sqrt{N}}\mathrm{e}^{j\frac{(n-1)(m-1)}{N}}. \tag{2}$$

Additionally, we denote the diagonal reconfigurable controller matrix as $\mathbf{F} = \mathbf{Diag}\{\xi_1, \xi_2, \cdots, \xi_N\}$, where ξ_n is the holographic pattern of the nth RHS element. We refer to [15], [16], [17], [18], [19], [20], [21] to configure the coefficients of each RHS element, where the RHS utilizes an amplitude-controlled method appropriately adjusting the radiation amplitude of the reference wave at each radiation element instead of conventional phase shifters. Specifically, the real part of the holographic pattern, i.e. $\mathrm{Re}[\Psi_{\mathrm{intf}}(\mathbf{r}_n^m)]$, is the cosine value of the phase difference between the desired object wave $\Psi_{\mathrm{obj}}(\mathbf{r}_n)$ and the reference wave $\Psi_{\mathrm{ref}}(\mathbf{r}_n^m)$. Thus, $\mathrm{Re}[\Psi_{\mathrm{intf}}(\mathbf{r}_n^m)]$ represents the amplitude of each radiation element. To avoid a negative value, $\mathrm{Re}[\Psi_{\mathrm{intf}}(\mathbf{r}_n^m)]$ is normalized to $[0, 1]$, and the coefficient of the nth RHS element is given by

$$\xi_n = \frac{\mathrm{Re}\left[\Psi_{\mathrm{intf}}(\mathbf{r}_n^m)\right] + 1}{2}, \tag{3}$$

where we have $\xi_n \in \mathbb{R}$ satisfying $0 \leq \xi_n \leq 1$. To reduce the power consumption, a switch-controlled method can be employed, which may be realized by PIN diodes, since PIN diodes have two states, i.e. ON and OFF [19]. Therefore, the diagonal reconfigurable controller matrix is designed as

$$\mathbf{F} = \mathbf{Diag}\{\xi_1, \xi_2, \cdots, \xi_N\}, \quad \xi_n \in \{0, 1\}. \tag{4}$$

B. Channel Model

As shown in Fig. 1, a base station (BS) relying on RHS-based hybrid beamforming supports K single-antenna user equipment (UE). We denote the large-scale fading and the link spanning from the metamaterial radiation elements to UE-k as υ_k and $\mathbf{h}_k^{\mathrm{H}} \in \mathbb{C}^{1 \times N}$, respectively. We adopt the mmWave channel model to characterize the propagation environment between the BS and each UE. Specifically, the mmWave

downlink channel of UE-k is assumed to be the sum of all propagation paths that are scattered in L_C clusters and each cluster contributes L_P paths, which can be expressed as

$$\mathbf{h}_k^{\mathrm{H}} = \sqrt{\frac{1}{L_C L_P}} \sum_{c=1}^{L_C} \sum_{p=1}^{L_P} \alpha_{c,p}^k \mathbf{a}^{\mathrm{H}}(\psi_{c,p}^k, \varphi_{c,p}^k), \qquad (5)$$

where $\alpha_{c,p}^k$ is the complex gain of the pth path in the cth cluster following $\alpha_{c,p}^k \sim \mathcal{CN}(0,1)$, and $\mathbf{a}^{\mathrm{H}}(\psi_{c,p}^k, \varphi_{c,p}^k)$ is given by

$$\begin{aligned}
&\mathbf{a}^{\mathrm{H}}(\psi_{c,p}^k, \varphi_{c,p}^k) \\
&= \Big[1, \cdots, \mathrm{e}^{-j\frac{2\pi}{\lambda}\left(\delta_x n_x \sin\psi_{c,p}^k \cos\varphi_{c,p}^k + \delta_y n_y \sin\psi_{c,p}^k \sin\varphi_{c,p}^k\right)}, \cdots, \\
&\quad \mathrm{e}^{-j\frac{2\pi}{\lambda}\left(\delta_x(N_x-1)\sin\psi_{c,p}^k \cos\varphi_{c,p}^k + \delta_y(N_y-1)\sin\psi_{c,p}^k \sin\varphi_{c,p}^k\right)}\Big],
\end{aligned} \tag{6}$$

where $n_x = 0, 1, \cdots, N_x-1$, $n_y = 0, 1, \cdots, N_y-1$, λ is the wave length of the carrier, δ_x and δ_y are the distance of the adjacent RHS radiation elements in the horizontal direction and vertical direction, respectively. Furthermore, $\psi_{c,p}^k$ and $\varphi_{c,p}^k$ are the elevation and azimuth angles of departure (AoD) from the RHS to UE-k in the pth path of the cth cluster, respectively. Within the cth cluster, the random variables $\psi_{c,p}^k$ and $\varphi_{c,p}^k$ have the uniformly distributed mean values of $\mu_{\psi_c^k}$ and $\mu_{\varphi_c^k}$, respectively, and have the angular spreads, i.e. standard deviations, of $\sigma_{\psi_c^k}$ and $\sigma_{\varphi_c^k}$, respectively. Furthermore, we define $\boldsymbol{\Upsilon} = \mathbf{Diag}\{v_1, v_2, \cdots, v_K\}$ and $\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, \cdots, \mathbf{h}_K]^{\mathrm{H}}$. At the receiver, we denote the receiver combining vector at UE-k by $\mathbf{u}_k^{\mathrm{H}}$ satisfying $\|\mathbf{u}_k\|^2 = 1$, and define $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_k]^{\mathrm{H}} \in \mathbb{C}^{K \times K}$. Although in multi-user systems typically each UE recovers its information independently, here we assume that the UEs can jointly recover their information to unveil the performance limits of the RHS-based multi-user systems. The joint multi-user detection scheme is also widely employed for the successive interference cancellation (SIC) operation of non-orthogonal multiple access (NOMA) systems [33], [34], [35], where the information recovery of each user may rely on the detection of other users, albeit it increases the channel estimation feedback overheads. Hence, the receiver combining vector $\mathbf{u}_k^{\mathrm{H}}$ at UE-k is designed based on the signals received from all UEs.

We denote the signals propagating from the metamaterial radiation elements by $\mathbf{x} = [x_1, x_2, \cdots, x_N]^{\mathrm{T}} \in \mathbb{C}^{N \times 1}$, which can be represented as

$$\begin{aligned}
\mathbf{x} = &\sqrt{\rho \varepsilon_{\mathrm{t}}} \sum_{k=1}^{K} \sqrt{p_k} \mathbf{F} \boldsymbol{\Phi} \mathbf{v}_k s_k \\
&+ \sqrt{\rho(1-\varepsilon_{\mathrm{t}})} \sum_{k=1}^{K} \sqrt{p_k} \mathbf{F} \boldsymbol{\Phi} \mathbf{v}_k \odot \boldsymbol{\eta}_k,
\end{aligned} \tag{7}$$

where ρ denotes the total transmit power at the BS. Furthermore, $\boldsymbol{\eta}_k \sim \mathcal{CN}(\mathbf{0}_M, \mathbf{I}_M)$ is the distortion of the information symbol s_k due to hardware impairments at the RF chains, resulting from power amplifier non-linearities, amplitude/phase imbalance in the In-phase/Quadrature mixers, phase noise in the local oscillator, sampling jitter and finite-resolution quantization in the analog-to-digital converters [36], at the BS. Finally, ε_{t} is the hardware quality factor of the RF-chains

at the BS satisfying $0 \leq \varepsilon_{\mathrm{t}} \leq 1$. Explicitly, a hardware quality factor of 1 indicates that the hardware is ideal, while 0 means that the hardware is completely inadequate. Therefore, the signal received by UE-k is given in (8), where $w_k \sim \mathcal{CN}(0, \sigma_w^2)$ represents the additive noise at UE-k. Furthermore, $\boldsymbol{\varsigma} = [\varsigma_1, \varsigma_2, \cdots, \varsigma_K]^{\mathrm{T}}$, with $\varsigma_k \sim \mathcal{CN}(0, 1)$ representing the distortion of the information symbol s_k due to the hardware impairments at UE-k, and the hardware quality factor of the UEs satisfying $0 \leq \varepsilon_{\mathrm{r}} \leq 1$.

III. RECONFIGURABLE HOLOGRAPHIC SURFACE BASED HYBRID BEAMFORMER DESIGN

In this section, we formulate the energy efficiency maximization problem of the switch-controlled RHS beamformer. Then, we decouple it into three sub-problems, including the holographic beamforming, digital beamforming, and power sharing ratio and total transmit power optimization.

According to (8), the received signal-to-interference-plus-noise-ratio (SINR) of the information s_k at UE-k is given by

$$\gamma_k = \frac{\rho \varepsilon_{\mathrm{r}} \varepsilon_{\mathrm{t}} p_k \mathcal{G}_{k,k}}{\rho(1 - \varepsilon_{\mathrm{r}} \varepsilon_{\mathrm{t}}) p_k \mathcal{G}_{k,k} + \rho \sum_{k' \neq k} p_{k'} \mathcal{G}_{k,k'} + \sigma_w^2}, \qquad (9)$$

where $\mathcal{G}_{k1,k2} = \|\mathbf{u}_{k_1}^{\mathrm{H}} \sqrt{\boldsymbol{\Upsilon}} \mathbf{H} \mathbf{F} \boldsymbol{\Phi} \mathbf{v}_{k_2}\|^2$. Thus, the spectral efficiency, denoted as $\mathcal{S}$, is given by

$$\mathcal{S} = \sum_{k=1}^{K} \log_2(1 + \gamma_k), \qquad (10)$$

and the energy efficiency, denoted as $\mathcal{E}$, is given by

$$\mathcal{E} = \frac{B\mathcal{S}}{\mathcal{P}} = \frac{B \sum_{k=1}^{K} \log_2(1 + \gamma_k)}{\frac{1}{\zeta}\rho + P_{\mathrm{Syn}} + N_{\mathrm{RF}} P_{\mathrm{RF}} + N_{\mathrm{PS}} P_{\mathrm{PS}} + N_{\mathrm{SW}} P_{\mathrm{SW}}}, \tag{11}$$

where B is the system's bandwidth, $\mathcal{P} = \frac{1}{\zeta}\rho + P_{\mathrm{Syn}} + N_{\mathrm{RF}} P_{\mathrm{RF}} + N_{\mathrm{PS}} P_{\mathrm{PS}} + N_{\mathrm{SW}} P_{\mathrm{SW}}$ is total power required for achieving the spectral efficiency $\mathcal{S}$ with ζ being the efficiency of the power amplifiers, P_{Syn} being the power consumption of the frequency synthesizer, N_{RF}, N_{PS} and N_{SW} being the number of RF chains, phase shifters and switches, respectively. Finally, P_{RF}, P_{PS} and P_{SW} are the power consumption of each RF chain, phase shifter and switch, respectively.

Our objective is to maximize the energy efficiency by optimizing the total transmit power ρ, the power sharing ratio $\mathbf{P}$, the digital beamformer $\mathbf{V}$ and $\mathbf{U}$ at the transmitter and the receiver, respectively, as well as the holographic beamformer $\mathbf{F}$. Thus, the optimization problem in the switch-controlled holographic beamforming architecture can be formulated as

$$(\text{P1}) \quad \max_{\rho, \mathbf{P}, \mathbf{V}, \mathbf{U}, \mathbf{F}} \frac{B \sum_{k=1}^{K} \log_2(1 + \gamma_k)}{\mathcal{P}} \qquad (12)$$

$$\text{s.t.} \quad \rho \leq P_{\max}, \qquad (13)$$

$$\sum_{k=1}^{K} p_k = 1, \qquad (14)$$

$$\|\mathbf{v}_k\|^2 = 1, \|\mathbf{u}_k\|^2 = 1, \ k = 1, 2, \cdots, K, \qquad (15)$$

$$\xi_n \in \{0, 1\}, \ n = 1, 2, \cdots, N. \qquad (16)$$

$$y_k = \underbrace{\sqrt{\rho \varepsilon_{\mathrm{r}} \varepsilon_{\mathrm{t}} p_k \upsilon_k} \mathbf{u}_k^{\mathrm{H}} \mathbf{H} \mathbf{F} \boldsymbol{\Phi} \mathbf{v}_k s_k}_{\text{Desired signal of the UE-}k\text{ information}} + \underbrace{\sqrt{\rho \varepsilon_{\mathrm{r}} (1 - \varepsilon_{\mathrm{t}}) p_k \upsilon_k} \mathbf{u}_k^{\mathrm{H}} \mathbf{H} \mathbf{F} \boldsymbol{\Phi} (\mathbf{v}_k \odot \boldsymbol{\eta}_k)}_{\text{BS HWI distortion on the UE-}k\text{ information}} + \underbrace{\sqrt{\rho (1 - \varepsilon_{\mathrm{r}}) p_k \upsilon_k} \mathbf{u}_k^{\mathrm{H}} \mathbf{H} \mathbf{F} \boldsymbol{\Phi} \mathbf{v}_k \varsigma_k}_{\text{UE-}k\text{ HWI distortion on the UE-}k\text{ information}} +$$

$$\underbrace{\sum_{k' \neq k} \left(\sqrt{\rho \varepsilon_{\mathrm{r}} \varepsilon_{\mathrm{t}} p_{k'} \upsilon_k} \mathbf{u}_k^{\mathrm{H}} \mathbf{H} \mathbf{F} \boldsymbol{\Phi} \mathbf{v}_{k'} s_{k'} + \sqrt{\rho \varepsilon_{\mathrm{r}} (1 - \varepsilon_{\mathrm{t}}) p_{k'} \upsilon_k} \mathbf{u}_k^{\mathrm{H}} \mathbf{H} \mathbf{F} \boldsymbol{\Phi} (\mathbf{v}_{k'} \odot \boldsymbol{\eta}_{k'}) + \sqrt{\rho (1 - \varepsilon_{\mathrm{r}}) p_{k'} \upsilon_k} \mathbf{u}_k^{\mathrm{H}} \mathbf{H} \mathbf{F} \boldsymbol{\Phi} \mathbf{v}_{k'} \varsigma_{k'} \right)}_{\text{Inter-user interference}}$$

$$+ \underbrace{w_k}_{\text{Additive noise}} , \tag{8}$$

respectively, where $P_{\max}$ is the maximum transmit power budget. To solve problem (P1) efficiently, we decompose them into three sub-problems as follows.

A. Holographic Beamformer Design

In the holographic beamformer, we aim for maximizing the sum of the eigen-channel gains of all UEs based on sum-path-gain maximization (SPGM) criterion [37]. We denote the equivalent channel at the baseband corresponding to UE-k as $\mathbf{g}_k^{\mathrm{H}} = \mathbf{h}_k^{\mathrm{H}} \mathbf{F} \boldsymbol{\Phi}$. Thus, the optimization problem can be formulated as

$$(\text{P2}) \quad \max_{\mathbf{F}} \sum_{k=1}^{K} \|\mathbf{g}_k\|_2^2 \tag{17}$$

$$\text{s.t.} \quad \xi_n \in \{0, 1\}, \; n = 1, 2, \cdots, N, \tag{18}$$

for the switch-controlled holographic beamformer. In (17), we have

$$\sum_{k=1}^{K} \|\mathbf{g}_k\|_2^2 = \sum_{k=1}^{K} \left\| \mathbf{h}_k^{\mathrm{H}} \mathbf{F} \boldsymbol{\Phi} \right\|_2^2$$

$$= \boldsymbol{\xi}^{\mathrm{T}} \left(\sum_{k=1}^{K} \sum_{m=1}^{M} \frac{1}{N} \mathbf{q}_{k,m} \mathbf{q}_{k,m}^{\mathrm{H}} \right) \boldsymbol{\xi}, \tag{19}$$

where $\mathbf{q}_{k,m} = [\mathrm{e}^{-j\phi_{1,m}} h_{k,1}, \mathrm{e}^{-j\phi_{2,m}} h_{k,2}, \cdots, \mathrm{e}^{-j\phi_{N,m}} h_{k,N}]^{\mathrm{H}}$ and $\boldsymbol{\xi} = [\xi_1, \xi_2, \cdots \xi_N]^{\mathrm{T}}$. Upon introducing the matrix $\mathbf{Q} = \sum_{k=1}^{K} \sum_{m=1}^{M} \frac{1}{N} \mathbf{q}_{k,m} \mathbf{q}_{k,m}^{\mathrm{H}}$, the optimization problem (P2) can be written as

$$(\text{P3}) \quad \max_{\boldsymbol{\xi}} \boldsymbol{\xi}^{\mathrm{T}} \mathbf{Q} \boldsymbol{\xi} \tag{20}$$

$$\text{s.t.} \quad \boldsymbol{\xi} \in \mathbb{R}^{N \times 1}, \; \xi_n \in \{0, 1\}, \; n = 1, 2, \cdots, N. \tag{21}$$

Since $\mathbf{Q} \succeq \mathbf{0}$, the problem (P3) is a quadratic programming and the optimal solution is located at the vertex. Therefore, the optimization problem associated with the constraint $0 \leq \xi_n \leq 1$ and that with the constraint $\xi_n \in \{0, 1\}$ have the same optimal solution, which means that the switch-controlled RHS beamforming architecture leads to the same performance as the continuous amplitude-controlled RHS beamforming architecture. Due to the module-one constraint, (P3) is a non-convex optimization problem. Although (P3) can be solved by the semidefinite relaxation (SDR) method, it has high calculation complexity [38]. Therefore, we employ the low-complexity ED method as follows.

Upon defining $\boldsymbol{\xi} = \frac{1}{2}(\widetilde{\boldsymbol{\xi}} + \mathbf{1}_N)$ with $\widetilde{\xi}_n \in \{-1, 1\}$ in the

problem (P3), we can get

$$\boldsymbol{\xi}^{\mathrm{T}} \mathbf{Q} \boldsymbol{\xi} = \frac{1}{4} \left(\widetilde{\boldsymbol{\xi}}^{\mathrm{T}} \mathbf{Q} \widetilde{\boldsymbol{\xi}} + \mathbf{1}_N^{\mathrm{T}} \mathbf{Q} \widetilde{\boldsymbol{\xi}} + \widetilde{\boldsymbol{\xi}}^{\mathrm{T}} \mathbf{Q} \mathbf{1}_N + \mathbf{1}_N^{\mathrm{T}} \mathbf{Q} \mathbf{1}_N \right). \tag{22}$$

Since the value of $\mathbf{1}_N^{\mathrm{T}} \mathbf{Q} \mathbf{1}_N$ is independent of the variable $\widetilde{\boldsymbol{\xi}}$, the problem of maximizing $\boldsymbol{\xi}^{\mathrm{T}} \mathbf{Q} \boldsymbol{\xi}$ is equivalent to that of maximizing $\widetilde{\boldsymbol{\xi}}^{\mathrm{T}} \mathbf{Q} \widetilde{\boldsymbol{\xi}} + \mathbf{1}_N^{\mathrm{T}} \mathbf{Q} \widetilde{\boldsymbol{\xi}} + \widetilde{\boldsymbol{\xi}}^{\mathrm{T}} \mathbf{Q} \mathbf{1}_N$. Upon introducing the auxiliary variable $a \in \{-1, 1\}$, and defining $\widetilde{\boldsymbol{\xi}} = a\mathbf{z}$ and $\widetilde{\mathbf{z}} = [\mathbf{z}^{\mathrm{T}}, a]^{\mathrm{T}}$, the problem (P3) can be presented as

$$(\text{P4}) \quad \max_{\widetilde{\mathbf{z}}} \widetilde{\mathbf{z}}^{\mathrm{T}} \widetilde{\mathbf{Q}} \widetilde{\mathbf{z}} \tag{23}$$

$$\text{s.t.} \quad \widetilde{\mathbf{z}} \in \mathbb{R}^{N \times 1}, \; \widetilde{z}_n \in \{-1, 1\}, \; n = 1, 2, \cdots, N + 1, \tag{24}$$

where we have:

$$\widetilde{\mathbf{Q}} = \begin{bmatrix} \mathbf{Q} & \mathbf{Q} \mathbf{1}_N \\ \mathbf{1}_N^{\mathrm{T}} \mathbf{Q} & 0 \end{bmatrix}. \tag{25}$$

Due to the module-one constraint, (P4) is a non-convex optimization problem and ED method for solving (P4) is presented as follows.

According to the constraint of (P4), we have $\|\widetilde{\mathbf{z}}\|^2 = N + 1$. We denote the eigen-decomposition of $\widetilde{\mathbf{Q}}$ as $\widetilde{\mathbf{Q}} = \sum_{n=1}^{N+1} \widetilde{\lambda}_n \widetilde{\boldsymbol{\nu}}_n \widetilde{\boldsymbol{\nu}}_n^{\mathrm{H}}$, where $\widetilde{\lambda}_1, \widetilde{\lambda}_2, \cdots, \widetilde{\lambda}_{N+1}$ are the eigen-values of $\widetilde{\mathbf{Q}}$ in a descending order, and $\widetilde{\boldsymbol{\nu}}_1, \widetilde{\boldsymbol{\nu}}_2, \cdots, \widetilde{\boldsymbol{\nu}}_{N+1}$ are the corresponding eigen-vectors. Based on the Rayleigh quotient, defined in [38], the optimal solution of $f(\mathbf{x}) = \frac{\mathbf{x}^{\mathrm{H}} \widetilde{\mathbf{Q}} \mathbf{x}}{\mathbf{x}^{\mathrm{H}} \mathbf{x}}$ is $\mathbf{x}^* = c\widetilde{\boldsymbol{\nu}}_1$, where c is an arbitrary non-zero constant and $\widetilde{\boldsymbol{\nu}}_1$ is the eigen-vector corresponding to the largest eigen-value of $\widetilde{\mathbf{Q}}$. Thus, based on the relaxation and projection method of [39], [40], the optimal solution of (P4), denoted as $\widetilde{\mathbf{z}}^*$, is $\widetilde{\mathbf{z}}^* = \mathrm{e}^{j\widetilde{\boldsymbol{\nu}}_1}$.

In the switch-controlled holographic beamforming architecture, we have $N_{\mathrm{RF}} = M$, $N_{\mathrm{PS}} = 0$, $N_{\mathrm{SW}} = N$. Then the total power consumption is $\mathcal{P} = \frac{1}{\zeta} \rho + P_{\mathrm{Syn}} + M P_{\mathrm{RF}} + N P_{\mathrm{SW}}$.

B. Digital Beamformer Design

Given the holographic beamformer $\mathbf{F}$, we can get the baseband equivalent channel as

$$\mathbf{G} = \sqrt{\boldsymbol{\Upsilon}} \mathbf{H} \mathbf{F} \boldsymbol{\Phi}. \tag{26}$$

Then the optimization problem can be formulated as

$$(\text{P5}) \quad \max_{\rho, \mathbf{P}, \mathbf{V}, \mathbf{U}} \sum_{k=1}^{K} \log_2 (1+$$

$$\left. \frac{\rho \varepsilon_{\mathrm{r}} \varepsilon_{\mathrm{t}} p_k \mathcal{G}_{k,k}}{\rho \left(1 - \varepsilon_{\mathrm{r}} \varepsilon_{\mathrm{t}}\right) p_k \mathcal{G}_{k,k} + \rho \sum_{k' \neq k} p_{k'} \mathcal{G}_{k,k'} + \sigma_w^2} \right) \tag{27}$$

$$\text{s.t.} \quad \sum_{k=1}^{K} p_k = 1, \tag{28}$$

$$\|\mathbf{v}_k\|^2 = 1, \|\mathbf{u}_k\|^2 = 1, \ k = 1, 2, \cdots, K. \tag{29}$$

To eliminate the inter-user interference, the singular value decomposition (SVD) can be employed for the digital beamformer as follows[1]. We denote the SVD of the baseband equivalent channel of $\mathbf{G}$ as

$$\mathbf{G} = \mathbf{U_G} \mathbf{\Lambda} \mathbf{V_G^H}, \tag{30}$$

where $\mathbf{\Lambda} = \{\lambda_1, \lambda_2, \cdots, \lambda_K\}$ is a diagonal matrix associated with diagonal elements that are the singular values in descending order, while $\mathbf{U_G} \in \mathbb{C}^{K \times K}$ and $\mathbf{V_G} \in \mathbb{C}^{N \times K}$ are complex unitary matrices. Thus, the digital precoding matrix $\mathbf{V}$ of the transmitter and the combining matrix $\mathbf{U^H}$ of the receiver are designed as $\mathbf{V} = \mathbf{V_G}$ and $\mathbf{U^H} = \mathbf{U_G^H}$, respectively.

C. Power Sharing Ratio and Total Transmit Power Optimization

According to (27) and (30), it may be readily shown that $\|\mathbf{u}_k^{\mathrm{H}} \mathbf{G} \mathbf{v}_k\|^2 = v_k \lambda_k^2$ and $\|\mathbf{u}_k^{\mathrm{H}} \mathbf{G} \mathbf{v}_{k'}\|^2 = 0$ if $k \neq k'$. Then (P1) can be expressed as

$$\text{(P6)} \quad \max_{\rho, \mathbf{P}} \quad \frac{B \sum_{k=1}^{K} \log_2 \left(1 + \frac{\rho \lambda_k^2 \varepsilon_{\mathrm{r}} \varepsilon_{\mathrm{t}} p_k}{\rho \lambda_k^2 (1 - \varepsilon_{\mathrm{r}} \varepsilon_{\mathrm{t}}) p_k + \sigma_w^2}\right)}{\frac{1}{\zeta} \rho + P_{\mathrm{Syn}} + M P_{\mathrm{RF}} + N P_{\mathrm{SW}}} \tag{31}$$

$$\text{s.t.} \quad \rho \leq P_{\max}, \tag{32}$$

$$\sum_{k=1}^{K} p_k = 1. \tag{33}$$

Since (P6) is a non-convex problem, we can decouple them into a pair of sub-problems and optimize them iteratively. Specifically, when the total transmit power ρ is given, the problems (P6) aims for optimizing the power sharing ratio $\mathbf{P}$, represented as

$$\text{(P6.a)} \quad \max_{\mathbf{P}} \quad \sum_{k=1}^{K} \log_2 \left(1 + \frac{\rho \lambda_k^2 \varepsilon_{\mathrm{r}} \varepsilon_{\mathrm{t}} p_k}{\rho \lambda_k^2 (1 - \varepsilon_{\mathrm{r}} \varepsilon_{\mathrm{t}}) p_k + \sigma_w^2}\right) \tag{34}$$

$$\text{s.t.} \quad \sum_{k=1}^{K} p_k = 1. \tag{35}$$

By contrast, when the power sharing ratio $\mathbf{P}$ is given, the problem (P6) aims for optimizing the total transmit power ρ, represented as

$$\text{(P6.b)} \quad \max_{\rho} \quad \frac{B \sum_{k=1}^{K} \log_2 \left(1 + \frac{\rho \lambda_k^2 \varepsilon_{\mathrm{r}} \varepsilon_{\mathrm{t}} p_k}{\rho \lambda_k^2 (1 - \varepsilon_{\mathrm{r}} \varepsilon_{\mathrm{t}}) p_k + \sigma_w^2}\right)}{\frac{1}{\zeta} \rho + P_{\mathrm{Syn}} + M P_{\mathrm{RF}} + N P_{\mathrm{SW}}} \tag{36}$$

$$\text{s.t.} \quad \rho \leq P_{\max}. \tag{37}$$

[1] Although it is more practical to employ the zero-forcing (ZF) method for multi-user systems, the SVD method can help us unveil the performance limit of the RHS-based multi-user systems.

By iteratively optimizing the power sharing ratio $\mathbf{P}$ and the total transmit power ρ until convergence, a nearly-optimal solution can be found. The details of the power sharing ratio optimization and the total transmit power optimization algorithm are given as follows.

1) Power sharing ratio optimization: When the hardware is ideal, i.e. $\varepsilon_{\mathrm{t}} = 1$ and $\varepsilon_{\mathrm{r}} = 1$, the problem (P6.a) can be solved by the state-of-the-art water-filling method [41]. By contrast, in the case of realistic non-ideal hardware, i.e. $\varepsilon_{\mathrm{t}} < 1$ or $\varepsilon_{\mathrm{r}} < 1$, the problem (P6.a) can be solved by using the Lagrange multiplier method of the calculus of variations [42]. Hence, we can formulate our problem as

$$J(\mathbf{P}) = \sum_{k=1}^{K} \log_2 \left(1 + \frac{\rho \lambda_k^2 \varepsilon_{\mathrm{r}} \varepsilon_{\mathrm{t}} p_k}{\rho \lambda_k^2 (1 - \varepsilon_{\mathrm{r}} \varepsilon_{\mathrm{t}}) p_k + \sigma_w^2}\right) + b \sum_{k=1}^{K} p_k, \tag{38}$$

where b is the Lagrange multiplier. Then, we take the partial derivative of $J(\mathbf{P})$ with respect to the power sharing variables to be optimized $p_1, p_2, \cdots, p_K$ and set them equal to zero, which results in

$$\frac{J(\mathbf{P})}{\partial p_k} = \frac{\rho \lambda_k^2 \varepsilon_{\mathrm{r}} \varepsilon_{\mathrm{t}} \sigma_w^2}{\ln 2 \cdot \left(\rho \lambda_k^2 p_k + \sigma_w^2\right) \left(\rho \lambda_k^2 (1 - \varepsilon_{\mathrm{r}} \varepsilon_{\mathrm{t}}) p_k + \sigma_w^2\right)} + b$$
$$= 0. \tag{39}$$

According to (39), p_k can be expressed as

$$p_k = \max \left\{f_k(b), 0\right\}, \tag{40}$$

where $f_k(b) = \frac{\sqrt{\varepsilon_{\mathrm{r}}^2 \varepsilon_{\mathrm{t}}^2 - \frac{\frac{4\rho}{\sigma_w^2} \lambda_k^2 (1 - \varepsilon_{\mathrm{r}} \varepsilon_{\mathrm{t}}) \varepsilon_{\mathrm{r}} \varepsilon_{\mathrm{t}}}{\ln 2 \cdot b}} - (2 - \varepsilon_{\mathrm{r}} \varepsilon_{\mathrm{t}})}{\frac{2\rho}{\sigma_w^2} \lambda_k^2 (1 - \varepsilon_{\mathrm{r}} \varepsilon_{\mathrm{t}})}$ and $\max\{x, 0\}$ represents the maximum value between x and 0. Furthermore, since $\sum_{k=1}^{K} p_k = 1$, we have

$$\sum_{k=1}^{K} \max \left\{f_k(b), 0\right\} = 1. \tag{41}$$

Algorithm 1 presents the details of deriving the power sharing ratio of $p_1, p_2, \cdots, p_K$, where the value of b in the equation $\sum_{k=1}^{K-\tau} f_k(b) = 1$ can be numerically calculated by the bisection method.

2) Total transmit power optimization: In (36), we denote the energy efficiency of the switch-controlled holographic beamformer as $\mathcal{E}$. Since the functions of $\mathcal{E}$ with respect to the transmit power ρ are concave, the optimal transmit power ρ can be found by the gradient descent method, which is given in Algorithm 2, where the gradient of $\mathcal{E}$ with respect to the transmit power ρ is formulated as:

$$\frac{\partial \mathcal{E}}{\partial \rho} = \frac{\partial \frac{B \sum_{k=1}^{K} \log_2 \left(1 + \frac{\rho \lambda_k^2 \varepsilon_{\mathrm{r}} \varepsilon_{\mathrm{t}} p_k}{\rho \lambda_k^2 (1 - \varepsilon_{\mathrm{r}} \varepsilon_{\mathrm{t}}) p_k + \sigma_w^2}\right)}{\frac{1}{\zeta} \rho + P_{\mathrm{Syn}} + M P_{\mathrm{RF}} + N P_{\mathrm{SW}}}}{\partial \rho}$$

$$= \frac{B}{\ln 2 \cdot \left(\frac{1}{\zeta} \rho + P_{\mathrm{Syn}} + M P_{\mathrm{RF}} + N P_{\mathrm{SW}}\right)^2} \cdot$$
$$\left(\sum_{k=1}^{K} \frac{\left(\frac{1}{\zeta} \rho + P_{\mathrm{Syn}} + M P_{\mathrm{RF}} + N P_{\mathrm{SW}}\right) \lambda_k^2 \varepsilon_{\mathrm{r}} \varepsilon_{\mathrm{t}} p_k \sigma_w^2}{\left(\rho \lambda_k^2 p_k + \sigma_w^2\right) \left(\rho \lambda_k^2 (1 - \varepsilon_{\mathrm{r}} \varepsilon_{\mathrm{t}}) p_k + \sigma_w^2\right)}\right)$$

Algorithm 1 Power allocation method conceived for maximizing the energy efficiency.

Input: Hardware quality factors ε_t and ε_r, transmit power ρ, baseband channel gain $\lambda_1, \lambda_2, \cdots, \lambda_K$, and additive noise power σ_w^2.

1: Initial $\tau = 0$.
2: **repeat**
3:     Numerically calculate b in the equation $\sum\limits_{k=1}^{K-\tau} f_k(b) = 1$.
4:     Update $p_k = f_k(b)$ for $k = 1, 2, \cdots, K-\tau$.
5:     Update $p_k = 0$ for $k = K-\tau+1, K-\tau+2\cdots, K$.
6:     $\tau \leftarrow \tau + 1$.
7: **until** $p_k \geq 0$ for $k = 1, 2, \cdots, K$.

Output: The optimal power allocation ratio, denoted as $\mathbf{P}^* = \mathbf{Diag}\{p_1^*, p_2^*, \cdots, p_K^*\}$.

Algorithm 2 Gradient descent method conceived for optimizing the total transmit power.

Input: Hardware quality factors ε_t and ε_r, power allocation ratio $p_1, p_2, \cdots, p_K$, baseband channel gain $\lambda_1, \lambda_2, \cdots, \lambda_K$, additive noise power σ_w^2, iteration step ι, precision error ϵ and maximum transmit power budget $P_{\max}$.

1: Initial $\tau = 0$, and transmit power $\rho^{(\tau)} = P_{\max}$.
2: **if** $\frac{\partial \mathcal{E}}{\partial \rho^{(\tau)}} \geq 0$
3:     The optimal transmit power $\rho^* = P_{\max}$.
4: **else**
5: **repeat**
6:     $\rho^{(\tau+1)} = \rho^{(\tau)} + \iota \cdot \frac{\partial \mathcal{E}}{\partial \rho^{(\tau)}}$.
7:     $\tau \leftarrow \tau + 1$.
8: **until** $\left| \frac{\partial \mathcal{E}}{\partial \rho^{(\tau)}} \right| < \epsilon$.

Output: The optimal transmit power ρ^*.

$$-\frac{1}{\zeta} \sum_{k=1}^{K} \ln \left(1 + \frac{\rho \lambda_k^2 \varepsilon_r \varepsilon_t p_k}{\rho \lambda_k^2 (1 - \varepsilon_r \varepsilon_t) p_k + \sigma_w^2} \right) \right). \qquad (42)$$

D. Theoretical Analysis

Firstly, the impact of hardware impairments on the spectral efficiency and energy efficiency is analyzed. Then, we present the time complexity of our proposed hybrid beamformer design.

1) Impact of Hardware Impairments: We investigate the impact of the hardware impairments on the transceiver's spectral efficiency and energy efficiency. According to (9) and (10), we can get that in the high signal-to-noise ratio (SNR) region, the spectral efficiency is limited by the hardware impairments at the transceivers, which results in a performance floor. This is due to the fact that increasing the transmit power increases both the desired signal and the signal distortion resulting from the HWI. Specifically, when the hardware of the transceiver is imperfect, the spectral efficiency attained in the high-SNR region tends to saturation at:

$$\mathcal{S} = K \log_2 \left(1 + \frac{\varepsilon_r \varepsilon_t}{1 - \varepsilon_r \varepsilon_t} \right), \qquad (43)$$

which can be obtained by setting $\rho \to \infty$ in (9). Furthermore, according to (11) and (43), we can formulate the energy efficiency in the high-SNR region as

$$\mathcal{E} = \frac{BK \log_2 \left(1 + \frac{\varepsilon_r \varepsilon_t}{1 - \varepsilon_r \varepsilon_t} \right)}{\frac{1}{\zeta} \rho + P_{\mathrm{Syn}} + N_{\mathrm{RF}} P_{\mathrm{RF}} + N_{\mathrm{PS}} P_{\mathrm{PS}} + N_{\mathrm{SW}} P_{\mathrm{SW}}}. \qquad (44)$$

By optimizing the transmit power ρ to maximize the energy efficiency, we can formulate the energy efficiency upper bound of the switch-controlled holographic beamforming architecture, which is given by

$$\overline{\mathcal{E}} = \frac{BK \log_2 \left(1 + \frac{\varepsilon_r \varepsilon_t}{1 - \varepsilon_r \varepsilon_t} \right)}{P_{\mathrm{Syn}} + M P_{\mathrm{RF}} + N P_{\mathrm{SW}}}. \qquad (45)$$

2) Computational complexity: The computational complexity of the proposed hybrid beamformer design depends on the number of iterations in the alternating maximization, denoted as T_a, and on the computational complexity required to solve each sub-problem. As for the sub-problem of holographic beamformer design, it can be seen that the optimization of the RHS coefficients depends on the number of RHS elements N, which can be seen to scale with N^3 [38]. As for the sub-problem of digital beamformer design, each SVD process in (30) requires NK^2 floating-point operations [43]. Moreover, the sub-problem of the power sharing ratio optimization in Algorithm 1 updates the power ratio factors at most K times, and in each update the equation $\sum\limits_{k=1}^{K-\tau} f_k(b) = 1$ is numerically calculated [44]. Since it only depends on the single variable of b, the sub-problem of the power sharing ratio optimization in Algorithm 1 has the complexity order of $\mathcal{O}(K)$. Since the sub-problem of the total transmit power optimization in Algorithm 2 iterates τ times and each iteration solves a single-variable gradient update, the sub-problem of the total transmit power optimization in Algorithm 2 has the complexity order of $\mathcal{O}(\tau)$. Since $N > K$, the overall computational complexity of our proposed hybrid beamformer design is $\mathcal{O}(T_a(N^3 + \tau))$.

IV. NUMERICAL AND SIMULATION RESULTS

In this section, the simulation results characterizing the energy efficiency of the switch-controlled RHS beamformers is presented. Similarly to [45], we denote the average large-scale fading of all users as $\overline{\upsilon}$, and the large-scale fading of the kth user as $\upsilon_k = \beta_k \overline{\upsilon}$, where the factors $\beta_1, \beta_2, \cdots, \beta_K$ are uniformly distributed in [0.5, 1.5]. Unless otherwise specified, the simulation parameters are set as follows: the bandwidth $B = 20$MHz, the number of users is $K = 8$, the number of RF chains is $M = 8$, the number of RHS elements is $N = 16 \times 16$, $\overline{\upsilon} = -80$dB, $\sigma_w^2 = -90$dBm, and the iterative times of the alternating optimization is $\tau = 4$. Similarly to [45], the mmWave propagation channel parameters are set as: the number of clusters is $L_C = 8$, the number of paths in each cluster is $L_P = 10$. The mean values of the elevation angles $\mu_{\psi_c^k}$ are randomly distributed in $[0°, 180°]$, the mean values of the azimuth angles $\mu_{\varphi_c^k}$ are randomly distributed in $[0°, 360°]$, the elevation angle spreads are $\sigma_{\psi_{c,p}^k} = 7.5°$, and the azimuth angle spreads are $\sigma_{\varphi_{c,p}^k} = 7.5°$, for $c = 1, 2, \cdots, L_C$, $p =$

$1, 2, \cdots, L_P$ and $k = 1, 2, \cdots, K$. According to [46], [47], the power consumption parameters are set as: the efficiency of the power amplifiers is $\zeta = 0.39$, the power consumption of the frequency synthesizer is $P_{\text{Syn}} = 2$W, while the power consumption of the RF chains, phase shifters and switches are $P_{\text{RF}} = 1$W, $P_{\text{PS}} = 30$mW and $P_{\text{SW}} = 1$mW, respectively.

For comparison, we employ the following state-of-the-art beamforming schemes as benchmarks:

1) *Fully digital beamformer:* The signals are processed by a pure digital beamformer having N RF chains. Thus, N RF chains dissipate power. Thus, the total power consumption in the fully digital beamformer is $\mathcal{P} = \frac{1}{\zeta}\rho + P_{\text{Syn}} + N P_{\text{RF}}$.

2) *Fully-connected phase shift array (PSA) beamformer:* The signals are processed by a digital beamformer having M RF chains, followed by an analogue beamformer equipped with N antennas, where the M RF chains and the N antennas are fully-connected by an $N \times M$ PSA. The analogue beamformer is designed based on the right singular matrix of the channel [48], [49], [46]. Specifically, we denote the right singular matrix as $\mathbf{V} \in \mathbb{C}^{N \times M}$. The coefficient of the phase shifter connecting the mth RF chain and the nth transmit antenna, denoted as $f_{n,m}^{(\text{FPSA})}$, is designed as $f_{n,m}^{(\text{FPAS})} = \frac{1}{\sqrt{N}} e^{j\angle V_{n,m}}$. Thus, a total of M RF chains and MN phase shifters consume power during signal processing. Thus, the total power consumption in the fully-connected PSA beamformer is $\mathcal{P} = \frac{1}{\zeta}\rho + P_{\text{Syn}} + M P_{\text{RF}} + MN P_{\text{PS}}$.

3) *Sub-connected PSA beamformer:* The signals are processed by a digital beamformer having M RF chains, followed by an analogue beamformer associated with N antennas, where the M RF chains and the N antennas are 'sub-connected', i.e. each RF chain is connected to $\frac{N}{M}$ antennas by an $\frac{N}{M} \times 1$ PSA. The analogue beamformer is designed based on the right singular matrix of the channel [48], [49]. The coefficient of the phase shifter connecting the mth RF chain and the nth transmit antenna, denoted as $f_{n,m}^{(\text{FPSA})}$, is designed as $f_{n,m}^{(\text{SPAS})} = \frac{1}{\sqrt{N}} e^{j\angle V_{n,m}}$ if the mth RF chain and the nth transmit antenna are connected, and $f_{n,m}^{(\text{SPAS})} = 0$ otherwise. Thus, a total of M RF chains and N phase shifters consume power. Thus, the total power consumption in the sub-connected PSA beamformer is $\mathcal{P} = \frac{1}{\zeta}\rho + P_{\text{Syn}} + M P_{\text{RF}} + N P_{\text{PS}}$.

4) *PSA with switches beamformer:* To further reduce the number of phase shifters activated, the method of PAS with switches is proposed in [48], [49]. Based on the sub-connected PSA circuit, each phase shifter is connected to a switch. Some phase shifters associated with low-magnitude element in the right singular matrix are deactivated by turning off the corresponding switch. Specifically, the state of the switch connecting the mth RF chain and the nth transmit antenna, denoted as $f_{n,m}^{(\text{S})}$, is defined as

$$f_{n,m}^{(\text{S})} = \begin{cases} 1, & \text{if } V_{n,m} > a_{(\kappa)} \\ 0, & \text{if } V_{n,m} \leq a_{(\kappa)} \end{cases}, \tag{46}$$

where $f_{n,m}^{(\text{S})} = 1$ represents that the switch is turned on, while $f_{n,m}^{(\text{S})} = 1$ indicates that the switch is turned off. Furthermore, $a_{(\kappa)}$ is the threshold, which ensures that $(1-\kappa)N$ phase shifters are activated while the remaining κN phase shifters are deactivated. Thus, a total of M RF chains, $(1-\kappa)N$ phase shifters and N switches consume power. Thus, the total power consumption in the PSA with switches beamformer is $\frac{1}{\zeta}\rho + P_{\text{Syn}} + M P_{\text{RF}} + (1-\kappa)N P_{\text{PS}} + N P_{\text{SW}}$.
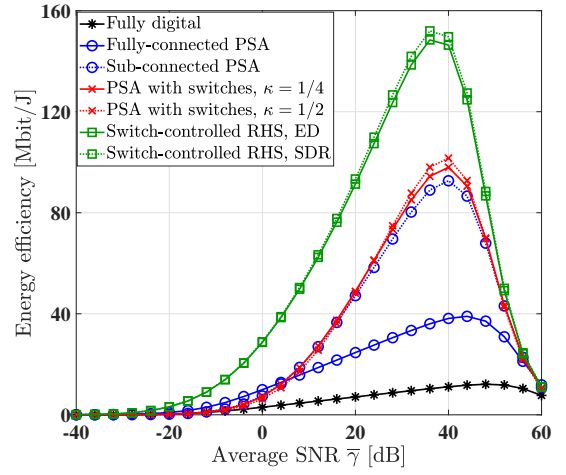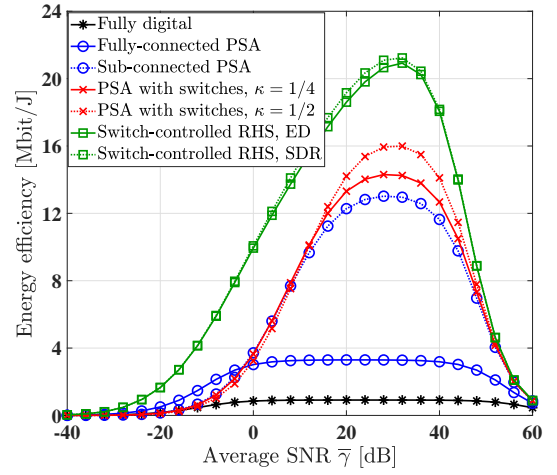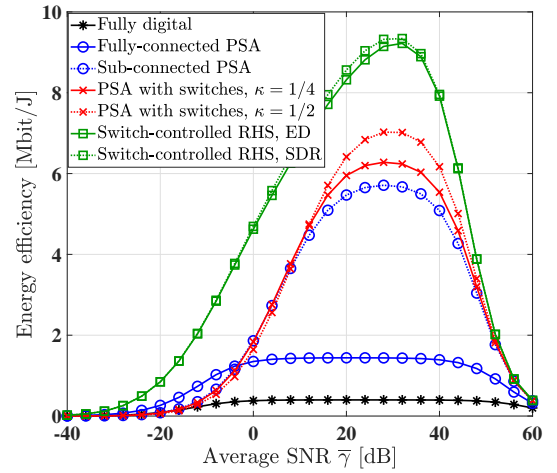
Explicitly, the fully digital and fully-connected PSA schemes have the highest power consumption. The power consumption is reduced by the sub-connected PSA beamformer. Furthermore, the PSA with switches scheme and the switch-controlled RHS beamforming scheme have very low power consumption due to the employment of low-power switches allowing us to reduce the number of phase shifters.

Fig. 2 and Fig. 3 characterize the spectral efficiency and energy efficiency of different beamforming schemes versus the average SNR $\overline{\gamma}$ associated with different hardware quality factors, respectively. The relationship between the average SNR and the transmit power is $\overline{\gamma} = \frac{\rho\overline{v}}{\sigma_w^2}$, which means that the transmit power is fixed without being optimized, when the average SNR is given. This shows that the fully digital scheme and the fully-connected PSA scheme achieve better spectral efficiency than the switch-controlled RHS beamformer. This is due to the fact that more RF chains and more phase shifters are deployed in the fully digital scheme and in the fully-connected PSA scheme, respectively. Furthermore, the switch-controlled RHS beamforming scheme has a comparable spectral efficiency to that of the sub-connected PSA and the PSA with switches. By contrast, in terms of the energy efficiency, the switch-controlled RHS beamformer has the lowest power consumption. Next comes the PSA associated with switches. Fig. 2 and Fig. 3 show that in the switch-controlled RHS beamformer, the ED method almost has the same performance to the SDR optimization method, while at much lower complexity. Furthermore, it shows that when the transceivers have imperfect hardware quality, i.e. $\varepsilon_t < 1$ and $\varepsilon_r < 1$, the spectral efficiency is limited in the high-SNR region by the hardware impairments and saturates at a constant value of $K \log_2 \left(1 + \frac{\varepsilon_r \varepsilon_t}{1 - \varepsilon_r \varepsilon_t} \right) = 11.79$ bit/s/Hz when $\varepsilon_t = \varepsilon_r = 0.8$ and at $K \log_2 \left(1 + \frac{\varepsilon_r \varepsilon_t}{1 - \varepsilon_r \varepsilon_t} \right) = 5.151$ bit/s/Hz when $\varepsilon_t = \varepsilon_r = 0.6$, which agrees with our analysis in (43). Moreover, for the switch-controlled holographic beamforming architecture, according to (45), the energy efficiency upper bound is $\frac{BK \log_2 \left(1 + \frac{\varepsilon_r \varepsilon_t}{1 - \varepsilon_r \varepsilon_t} \right)}{P_{\text{Syn}} + M P_{\text{RF}} + N P_{\text{SW}}} = 22.99$ Mbit/J when $\varepsilon_t = \varepsilon_r = 0.8$ and $\frac{BK \log_2 \left(1 + \frac{\varepsilon_r \varepsilon_t}{1 - \varepsilon_r \varepsilon_t} \right)}{P_{\text{Syn}} + M P_{\text{RF}} + N P_{\text{SW}}} = 10.04$ Mbit/J when $\varepsilon_t = \varepsilon_r = 0.6$.

Fig. 4 shows the energy efficiency of the switch-controlled RHS-based beamformers versus the average SNR along with the hardware quality factors of $\varepsilon_t = \varepsilon_r = 0.8$ and $\varepsilon_t = \varepsilon_r = 0.6$, respectively. The legend '*Ignore HWI*' means that the HWIs are ignored in the power sharing and the state-of-the-art water-filling method is employed. By contrast, the legend '*Consider HWI*' means that the HWIs are considered in the
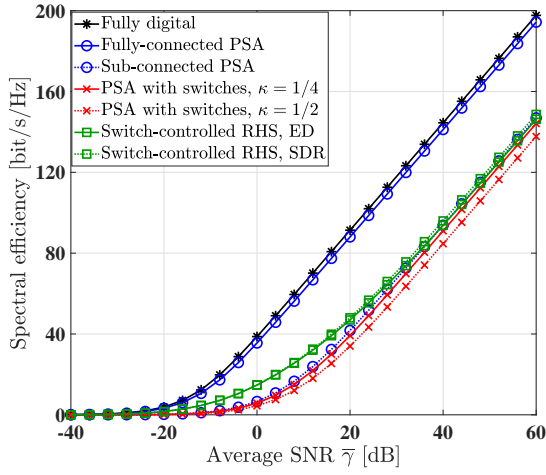
(a) Hardware quality factor $\varepsilon_{\mathrm{t}} = \varepsilon_{\mathrm{r}} = 1$.
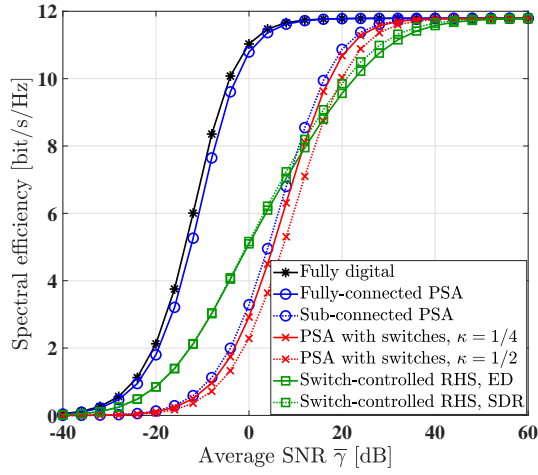
(b) Hardware quality factors $\varepsilon_{\mathrm{t}} = \varepsilon_{\mathrm{r}} = 0.8$.
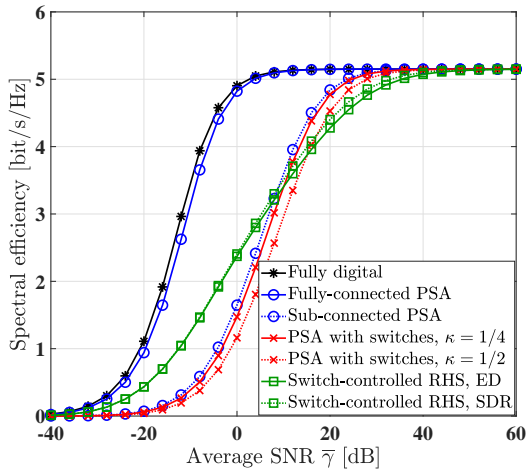
(c) Hardware quality factors $\varepsilon_{\mathrm{t}} = \varepsilon_{\mathrm{r}} = 0.6$.

Fig. 2: Comparison of the spectral efficiency versus the average SNR $\overline{\gamma}$ in various beamforming schemes.

(a) Hardware quality factors $\varepsilon_{\mathrm{t}} = \varepsilon_{\mathrm{r}} = 1$.

(b) Hardware quality factors $\varepsilon_{\mathrm{t}} = \varepsilon_{\mathrm{r}} = 0.8$.

(c) Hardware quality factors $\varepsilon_{\mathrm{t}} = \varepsilon_{\mathrm{r}} = 0.6$.

Fig. 3: Comparison of the energy efficiency versus the average SNR $\overline{\gamma}$ in various beamforming schemes.

power sharing and Algorithm 1 is employed. This shows that increased energy efficiency can be attained when the effect of HWIs is taken into account. Specifically, the maximum energy efficiency in the switch-controlled RHS beamformer

scheme can be achieved when the transmit power is $\rho = 32$dB. When the hardware quality is imperfect as indicated by the factor of $\varepsilon_{\mathrm{t}} = \varepsilon_{\mathrm{r}} = 0.6$ and the HWIs are ignored in the power sharing, approximately 8.7 Mbit/J and 8.9 Mbit/J of
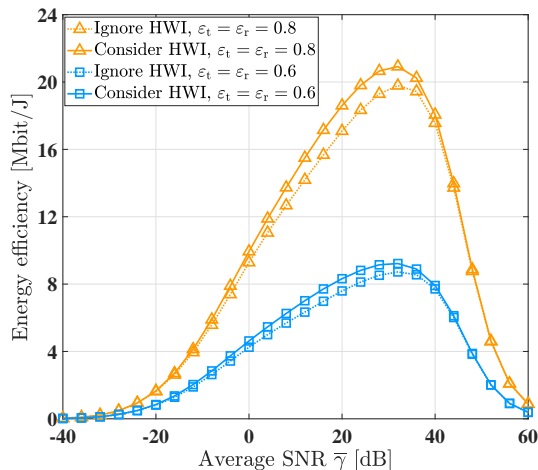
Fig. 4: Comparison of the energy efficiency versus the average SNR $\overline{\gamma}$ in the switch-controlled RHS-based beamforming scheme, with the effect of HWI is considered or ignored in the beamforming algorithms.

maximum energy efficiency can be achieved by the ED method and by the SDR method, respectively. By contrast, when the HWIs are explicitly considered in the power sharing design, approximately 9.2 Mbit/J and 9.3 Mbit/J of maximum energy efficiency can be achieved by the ED method and by the SDR method, respectively. This means that about 1.5 Mbit/J energy efficiency gain can be achieved by considering the HWIs in the power sharing design.

Fig. 5 compares the energy efficiency versus the maximum transmit power budget $P_{\max}$ of various beamformers. Fig. 5 (a) shows that when the hardware quality is ideal, the switch-controlled RHS beamformer outperforms the benchmarks. Furthermore, the energy efficiency saturates when the maximum transmit power budget is $P_{\max} \approx 30$dBm, which means that the optimal energy efficiency is achieved when $P_{\max} \approx 30$dBm. Fig. 5 (b) and Fig. 5 (c) show that the switch-controlled RHS beamformer performs constantly better than other beamforming schemes, regardless of the hardware quality.

Fig. 6 compares the energy efficiency versus the maximum transmit power budget $P_{\max}$ of the switch-controlled RHS-based beamformer, when the effect of HWIs is either considered or ignored in the beamformer design. Explicitly, considering the effect of HWIs can bring about further energy efficiency enhancements. Furthermore, when $P_{\max} > 20$dBm, the energy efficiency actually decays upon increasing the maximum transmit power budget if the deleterious effect of HWIs is ignored in the power sharing design. This is due to the fact that the effect of HWIs becomes more grave upon increasing the transmit power. Moreover, Fig. 6 shows that when $P_{\max} > 20$dBm, considering the HWIs in the power sharing design can bring about 1.4 Mbit/J and 1.7 Mbit/J energy efficiency performance gain, when the hardware quality factors are $\varepsilon_t = \varepsilon_r = 0.8$ and $\varepsilon_t = \varepsilon_r = 0.6$, respectively.
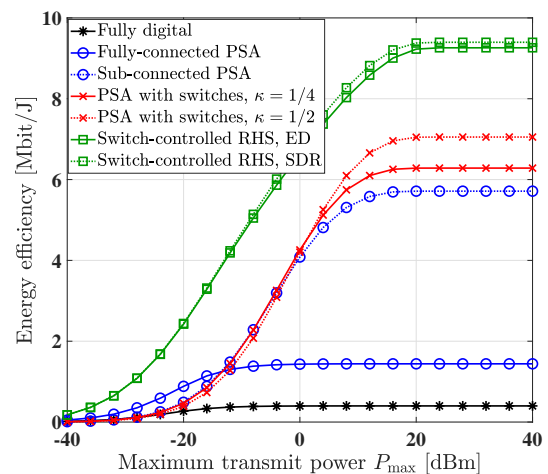
Fig. 7 compares the energy efficiency versus the hardware quality factors of various beamformers, where $\varepsilon = \varepsilon_t = \varepsilon_r$. This shows that the energy efficiency increases with the
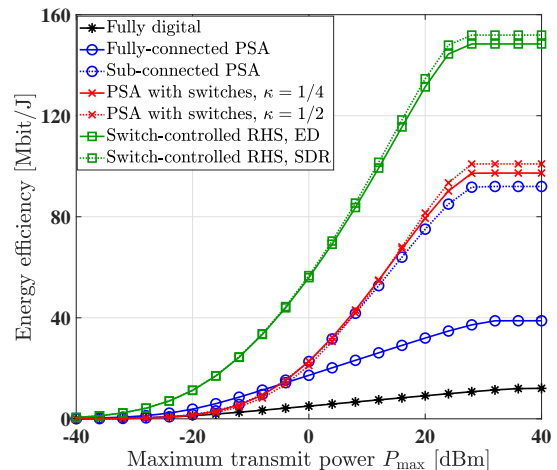
(a) Hardware quality factors $\varepsilon_t = \varepsilon_r = 1$.
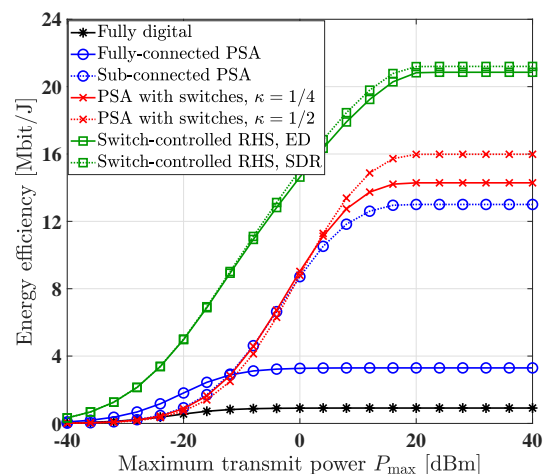
(b) Hardware quality factors $\varepsilon_t = \varepsilon_r = 0.8$.

(c) Hardware quality factors $\varepsilon_t = \varepsilon_r = 0.6$.

Fig. 5: Comparison of the energy efficiency versus the maximum transmit power budget $P_{\max}$ in various beamforming schemes.

improvement of the hardware quality factors. Moreover, the switch-controlled RHS beamformer consistently outperforms the other schemes.
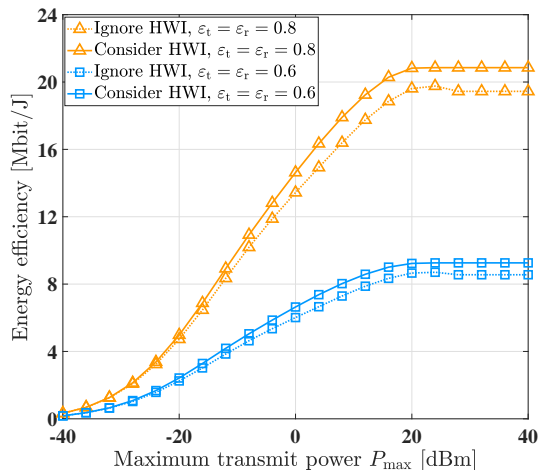
Fig. 6: Comparison of the energy efficiency versus the maximum transmit power budget $P_{\max}$ in the RHS-based beamforming scheme, with the effect of HWI is considered or ignored in the beamforming algorithms.
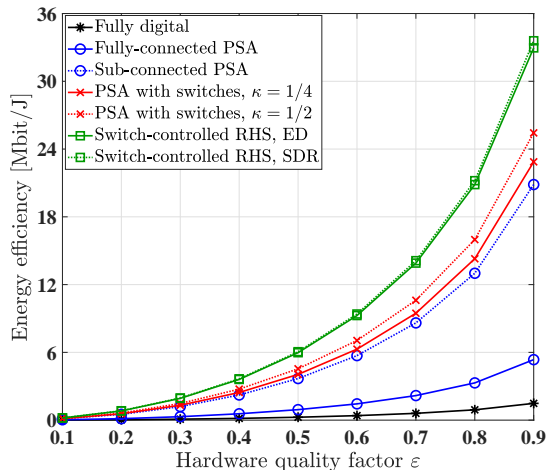

Fig. 7: Comparison of the energy efficiency versus the hardware quality factors in various beamforming schemes, where $\varepsilon = \varepsilon_{\mathrm{t}} = \varepsilon_{\mathrm{r}}$.
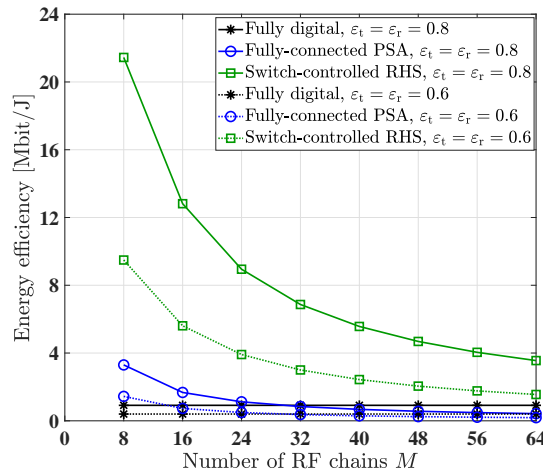

Fig. 8: Comparison of the energy efficiency versus the number of RF chains for various beamforming schemes.

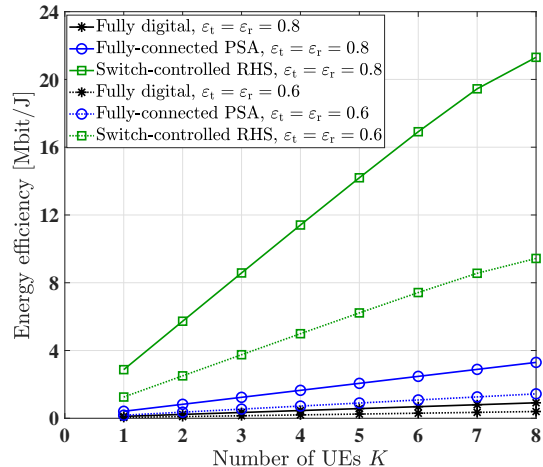Fig. 9: Comparison of the energy efficiency versus the number of UEs for various beamforming schemes.

Fig. 8, Fig. 9 and Fig. 10 compare the energy efficiency versus the number of RF chains, number of UEs and number of RHS elements in various beamformers, respectively. Firstly, Fig. 8 shows that the energy efficiency decreases upon deploying more RF chains due to the power consumption of RF chains. Furthermore, the switch-controlled RHS beamforming architectures have higher energy efficiency than the fully digital beamforming architecture and fully-connected PSA beamforming architecture. Fig. 9 portrays that the energy efficiency increases almost linearly with the number of UEs. Fig. 10 shows that with the increase of the number of RHS elements, the energy efficiency of the switch-controlled RHS beamforming architecture is much higher than that of the fully digital beamforming architecture and fully-connected PSA beamforming architecture. Furthermore, the energy efficiency of the fully digital beamforming architecture, the fully-connected PSA beamforming architecture was reduced

upon increasing the number of RHS elements due to the high power consumption of the RF chains and phase shifters. By contrast, the energy efficiency of the switch-controlled RHS beamforming architecture remains almost unchanged, as the number of RHS elements increases due to the lower power consumption of the switches.

In Fig. 11, the energy efficiency versus the number of iterations τ used by the alternating optimization algorithm is presented for the different hardware quality factors. It shows that the convergence speed in our proposed alternating optimization algorithm is excellent. Furthermore, the alternating optimization algorithm in the switch-controlled RHS-aided beamforming architecture converges when $\tau = 2$.

V. CONCLUSIONS

The problem of maximizing the energy efficiency for switch-controlled RHS beamforming architectures was formulated by optimizing the digital beamformer, holographic beamformer, the total transmit power and the power sharing ratio of users, while considering the effect of HWIs.
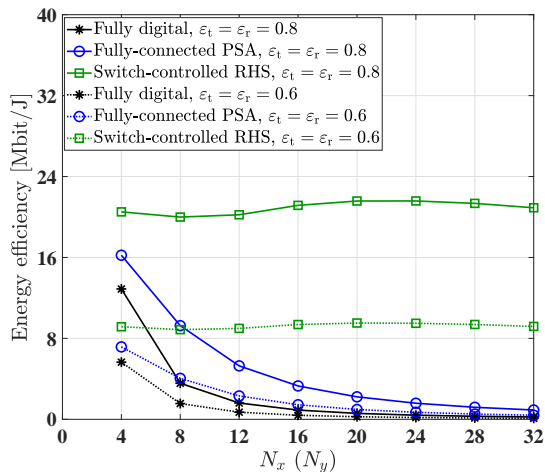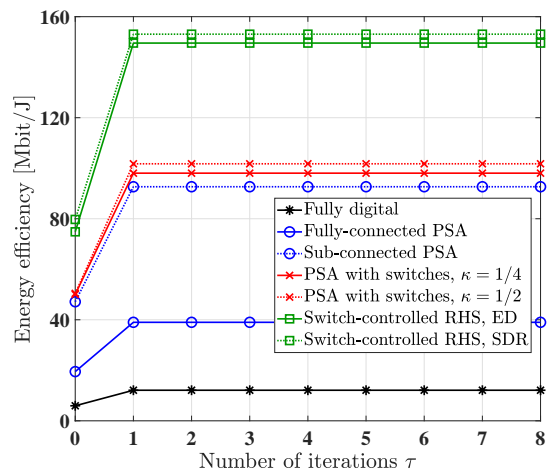
Fig. 10: Comparison of the energy efficiency versus the number of RHS elements for various beamforming schemes.

Our simulation results demonstrated that the switch-controlled RHS beamformers outperform the conventional fully digital beamformer and hybrid beamformer based on the PSA scheme in terms of the energy efficiency attained. The energy-efficient RHS beamforming architecture constitutes a promising option for the practical holographic MIMO implementation.

REFERENCES

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

[2] W. Tang, X. Chen, M. Z. Chen, J. Y. Dai, Y. Han, M. Di Renzo, S. Jin, Q. Cheng, and T. J. Cui, "Path loss modeling and measurements for reconfigurable intelligent surfaces in the millimeter-wave frequency band," *IEEE Trans. Commun.*, vol. 70, no. 9, pp. 6259–6276, 2022.

[3] J. Ma, S. Zhang, H. Li, F. Gao, and S. Jin, "Sparse bayesian learning for the time-varying massive MIMO channels: Acquisition and tracking," *IEEE Trans. Commun.*, vol. 67, no. 3, pp. 1925–1938, 2018.

[4] H. He, R. Wang, W. Jin, S. Jin, C.-K. Wen, and G. Y. Li, "Beamspace channel estimation for wideband millimeter-wave MIMO: A model-driven unsupervised learning approach," *IEEE Trans. Wireless Commun.*, vol. 22, no. 3, pp. 1808–1822, 2022.

[5] J. Du, S. Ye, L. Jin, X. Li, H. Q. Ngo, and O. A. Dobre, "Tensor-based joint channel estimation for multi-way massive MIMO hybrid relay systems," *IEEE Trans. Veh. Technol.*, vol. 71, no. 9, pp. 9571–9585, 2022.

[6] X. Li, M. Zhang, H. Chen, C. Han, L. Li, D.-T. Do, S. Mumtaz, and A. Nallanathan, "UAV-enabled multi-pair massive MIMO-NOMA relay systems with low-resolution ADCs/DACs," *IEEE Trans. Veh. Technol.*, vol. 73, no. 2, pp. 2171–2186, 2024.

[7] C. Pan, H. Ren, M. Elkashlan, A. Nallanathan, and L. Hanzo, "Robust beamforming design for ultra-dense user-centric C-RAN in the face of realistic pilot contamination and limited feedback," *IEEE Trans. Wireless Commun.*, vol. 18, no. 2, pp. 780–795, 2018.

[8] C. Pan, H. Mehrpouyan, Y. Liu, M. Elkashlan, and N. Arumugam, "Joint pilot allocation and robust transmission design for ultra-dense user-centric TDD C-RAN with imperfect CSI," *IEEE Trans. Wireless Commun.*, vol. 17, no. 3, pp. 2038–2053, 2018.

[9] Z. Huang, X. Cheng, and X. Yin, "A general 3D non-stationary 6G channel model with time-space consistency," *IEEE Trans. Commun.*, vol. 70, no. 5, pp. 3436–3450, 2022.

[10] C. Huang, S. Hu, G. C. Alexandropoulos, A. Zappone, C. Yuen, R. Zhang, M. Di Renzo, and M. Debbah, "Holographic MIMO surfaces for 6G wireless networks: Opportunities, challenges, and trends," *IEEE Wireless Commun.*, vol. 27, no. 5, pp. 118–125, 2020.

[11] A. Pizzo, T. L. Marzetta, and L. Sanguinetti, "Spatially-stationary model for holographic MIMO small-scale fading," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 9, pp. 1964–1979, 2020.

(a) Hardware quality factors $\varepsilon_\mathrm{t} = \varepsilon_\mathrm{r} = 1$.
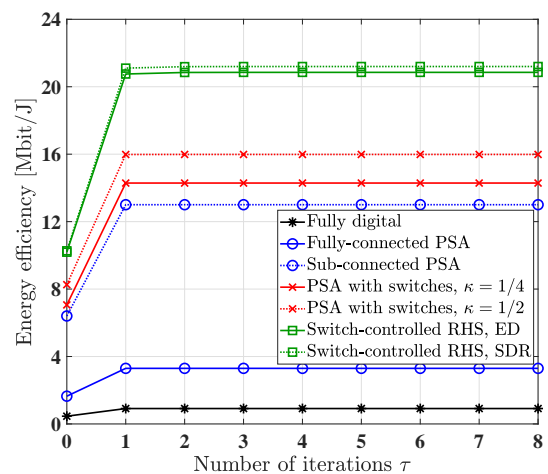
(b) Hardware quality factors $\varepsilon_\mathrm{t} = \varepsilon_\mathrm{r} = 0.8$.
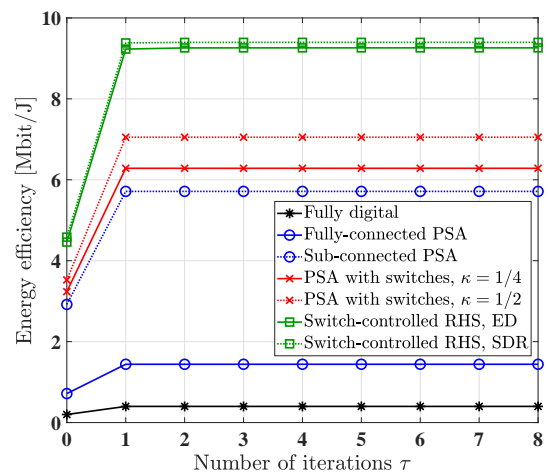
(c) Hardware quality factors $\varepsilon_\mathrm{t} = \varepsilon_\mathrm{r} = 0.6$.

Fig. 11: Comparison of the energy efficiency versus the number of iterations for various beamforming schemes.

[12] A. Pizzo, L. Sanguinetti, and T. L. Marzetta, "Fourier plane-wave series expansion for holographic MIMO communications," *IEEE Trans. Wireless Commun.*, vol. 21, no. 9, pp. 6890–6905, 2022.

[13] Ö. T. Demir, E. Björnson, and L. Sanguinetti, "Channel modeling and channel estimation for holographic massive MIMO with planar arrays," *IEEE Wireless Commun. Lett.*, vol. 11, no. 5, pp. 997–1001, 2022.

[14] S. Zeng, H. Zhang, B. Di, H. Qin, X. Su, and L. Song, "Reconfigurable refractive surfaces: An energy-efficient way to holographic MIMO," *IEEE Commun. Lett.*, vol. 26, no. 10, pp. 2490–2494, 2022.

[15] R. Deng, B. Di, H. Zhang, D. Niyato, Z. Han, H. V. Poor, and L. Song, "Reconfigurable holographic surfaces for future wireless communications," *IEEE Wireless Commun.*, vol. 28, no. 6, pp. 126–131, 2021.

[16] R. Deng, Y. Zhang, H. Zhang, B. Di, H. Zhang, and L. Song, "Reconfigurable holographic surface: A new paradigm to implement holographic radio," *IEEE Veh. Technol. Mag.*, vol. 18, no. 1, pp. 20–28, 2023.

[17] R. Deng, B. Di, H. Zhang, Y. Tan, and L. Song, "Reconfigurable holographic surface: Holographic beamforming for metasurface-aided wireless communications," *IEEE Trans. Veh. Technol.*, vol. 70, no. 6, pp. 6255–6259, 2021.

[18] R. Deng, B. Di, H. Zhang, and L. Song, "HDMA: Holographic-pattern division multiple access," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 4, pp. 1317–1332, 2022.

[19] R. Deng, B. Di, H. Zhang, Y. Tan, and L. Song, "Reconfigurable holographic surface-enabled multi-user wireless communications: Amplitude-controlled holographic beamforming," *IEEE Trans. Wireless Commun.*, vol. 21, no. 8, pp. 6003–6017, 2022.

[20] X. Hu, R. Deng, B. Di, H. Zhang, and L. Song, "Holographic beamforming for ultra massive MIMO with limited radiation amplitudes: How many quantized bits do we need?" *IEEE Commun. Lett.*, vol. 26, no. 6, pp. 1403–1407, 2022.

[21] R. Deng, B. Di, H. Zhang, H. V. Poor, and L. Song, "Holographic MIMO for leo satellite communications aided by reconfigurable holographic surfaces," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 10, pp. 3071–3085, 2022.

[22] P. Yang, L. Yang, and S. Wang, "Performance analysis for RIS-aided wireless systems with imperfect CSI," *IEEE Wireless Commun. Lett.*, vol. 11, no. 3, pp. 588–592, 2021.

[23] Z. Peng, T. Li, C. Pan, H. Ren, and J. Wang, "RIS-aided D2D communications relying on statistical CSI with imperfect hardware," *IEEE Commun. Lett.*, vol. 26, no. 2, pp. 473–477, 2021.

[24] T.-H. Vu and S. Kim, "Performance analysis of full-duplex two-way RIS-based systems with imperfect CSI and discrete phase-shift design," *IEEE Commun. Lett.*, vol. 27, no. 2, pp. 512–516, 2022.

[25] Q. Li, M. El-Hajjar, I. Hemadeh, D. Jagyasi, A. Shojaeifard, and L. Hanzo, "Performance analysis of active RIS-aided systems in the face of imperfect CSI and phase shift noise," *IEEE Trans. Veh. Technol.*, vol. 72, no. 6, pp. 8140–8145, 2023.

[26] J. Wang, S. Gong, Q. Wu, and S. Ma, "RIS-aided MIMO systems with hardware impairments: Robust beamforming design and analysis," *IEEE Trans. Wireless Commun.*, vol. 22, no. 10, pp. 6914–6929, 2023.

[27] Q. Li, M. El-Hajjar, Y. Sun, I. Hemadeh, A. Shojaeifard, Y. Liu, and L. Hanzo, "Achievable rate analysis of the STAR-RIS aided NOMA uplink in the face of imperfect CSI and hardware impairments," *IEEE Trans. Commun.*, vol. 71, no. 10, pp. 6100–6114, 2023.

[28] Q. Li, M. El-Hajjar, Y. Sun, I. Hemadeh, Y. Tsai, A. Shojaeifard, and L. Hanzo, "Achievable rate analysis of intelligent omni-surface assisted NOMA holographic MIMO systems," *IEEE Trans. Veh. Technol.*, 2024.

[29] L. You, J. Xu, G. C. Alexandropoulos, J. Wang, W. Wang, and X. Gao, "Energy efficiency maximization of massive MIMO communications with dynamic metasurface antennas," *IEEE Trans. Wireless Commun.*, vol. 22, no. 1, pp. 393–407, 2022.

[30] H. Jiang, L. You, J. Wang, W. Wang, and X. Gao, "Hybrid RIS and DMA assisted multiuser MIMO uplink transmission with electromagnetic exposure constraints," *IEEE J. Sel. Top. Sign. Proces.*, vol. 16, no. 5, pp. 1055–1069, 2022.

[31] N. Shlezinger, G. C. Alexandropoulos, M. F. Imani, Y. C. Eldar, and D. R. Smith, "Dynamic metasurface antennas for 6G extreme massive MIMO communications," *IEEE Wireless Commun.*, vol. 28, no. 2, pp. 106–113, 2021.

[32] H. Wang, N. Shlezinger, Y. C. Eldar, S. Jin, M. F. Imani, I. Yoo, and D. R. Smith, "Dynamic metasurface antennas for MIMO-OFDM receivers with bit-limited ADCs," *IEEE Trans. Commun.*, vol. 69, no. 4, pp. 2643–2659, 2020.

[33] Y. Wang, J. Wang, D. W. K. Ng, R. Schober, and X. Gao, "A minimum error probability NOMA design," *IEEE Trans. Wireless Commun.*, vol. 20, no. 7, pp. 4221–4237, 2021.

[34] J. A. de Carvalho, D. B. da Costa, L. Yang, G. C. Alexandropoulos, R. Oliveira, and U. S. Dias, "User fairness in wireless powered communication networks with non-orthogonal multiple access," *IEEE Wireless Commun. Lett.*, vol. 10, no. 1, pp. 189–193, 2020.

[35] L. Dai, B. Wang, Z. Ding, Z. Wang, S. Chen, and L. Hanzo, "A survey of non-orthogonal multiple access for 5G," *IEEE Commun. Surv. Tut.*, vol. 20, no. 3, pp. 2294–2323, 2018.

[36] E. Björnson, J. Hoydis, L. Sanguinetti *et al.*, "Massive mimo networks: Spectral, energy, and hardware efficiency," *Foundations and Trends® in Signal Processing*, vol. 11, no. 3-4, pp. 154–655, 2017.

[37] 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 Commun. Lett.*, vol. 9, no. 7, pp. 1105–1109, 2020.

[38] C.-Y. Chi, W.-C. Li, and C.-H. Lin, *Convex optimization for signal processing and communications: from fundamentals to applications*. CRC press, 2017.

[39] Y. Zhang, C. Zhong, Z. Zhang, and W. Lu, "Sum rate optimization for two way communications with intelligent reflecting surface," *IEEE Commun. Lett.*, vol. 24, no. 5, pp. 1090–1094, 2020.

[40] G. Zhou, C. Pan, H. Ren, K. Wang, M. Elkashlan, and M. Di Renzo, "Stochastic learning-based robust beamforming design for RIS-aided millimeter-wave systems in the presence of random blockages," *IEEE Trans. Veh. Technol.*, vol. 70, no. 1, pp. 1057–1061, 2021.

[41] D. Tse and P. Viswanath, *Fundamentals of wireless communication*. Cambridge university press, 2005.

[42] J. R. Hampton, *Introduction to MIMO communications*. Cambridge university press, 2013.

[43] X. Li, S. Wang, and Y. Cai, "Tutorial: Complexity analysis of singular value decomposition and its variants," *arXiv preprint arXiv:1906.12085*, 2019.

[44] A. Ben-Tal and A. Nemirovski, *Lectures on modern convex optimization: analysis, algorithms, and engineering applications*. SIAM, 2001.

[45] W. Ni and X. Dong, "Hybrid block diagonalization for massive multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 64, no. 1, pp. 201–211, 2015.

[46] S. Payami, N. M. Balasubramanya, C. Masouros, and M. Sellathurai, "Phase shifters versus switches: An energy efficiency perspective on hybrid beamforming," *IEEE Wireless Commun. Lett.*, vol. 8, no. 1, pp. 13–16, 2018.

[47] A. Garcia-Rodriguez, V. Venkateswaran, P. Rulikowski, and C. Masouros, "Hybrid analog-digital precoding revisited under realistic RF modeling," *IEEE Wireless Commun. Lett.*, vol. 5, no. 5, pp. 528–531, 2016.

[48] S. Payami, M. Ghoraishi, and M. Dianati, "Hybrid beamforming for large antenna arrays with phase shifter selection," *IEEE Trans. Wireless Commun.*, vol. 15, no. 11, pp. 7258–7271, 2016.

[49] S. Payami, M. Ghoraishi, M. Dianati, and M. Sellathurai, "Hybrid beamforming with a reduced number of phase shifters for massive MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 67, no. 6, pp. 4843–4851, 2018.