

# Traffic Aware Power Saving Communication Assisted By Double-Faced Active RIS

Yuyan Zhou<sup>1</sup>, Yang Liu<sup>1\*</sup>, Qingqing Wu<sup>2</sup>, Qingjiang Shi<sup>3</sup>, and Jun Zhao<sup>4</sup>

1: School of Information and Communication Engineering, Dalian University of Technology

2: Department of Electronic Engineering, Shanghai Jiao Tong University

3: School of Software Engineering, Tongji University, and Shenzhen Research Institute of Big Data, Shenzhen, China

4: School of Computer Science and Engineering, Nanyang Technological University

Emails: zyy98@mail.dlut.edu.cn, yangliu\_613@dlut.edu.cn, qingqingwu@sjtu.edu.cn, shiqj@tongji.edu.cn, junzhao@ntu.edu.sg

**Abstract**—Despite its high energy and hardware efficiency, some defects of the reconfigurable intelligence surface (RIS) technology have come to be realized, including the severe fading loss and restricted-to-half-space coverage. This paper proposes a novel double-faced-active (DFA)-RIS structure to overcome these defects. Besides, we utilize this novel DFA-RIS to improve power saving of the communication system. Unlike traditional power saving literature, we aim at fulfilling queueing stability and long-term power minimization in a downlink system assisted by the DFA-RIS, with a realistic data arriving process taken into consideration. Enlightened by Lyapunov control theory, we propose an online optimization strategy that adaptively adjusts DFA-RIS configuration. Each online problem can be efficiently solved by leveraging alternative directional method of multipliers (ADMM) method. Numerical results demonstrate the effectiveness of our proposed Lyapunov-guided strategy and DFA-RIS' superiority over the classical passive RIS.

**Index Terms**—Reconfigurable intelligent surface (RIS), Lyapunov optimization, queue stability, analytic-based solution.

## I. INTRODUCTION

Reconfigurable intelligent surface (RIS) [1] has been cast with great attentions these days due to its capability of providing passive beamforming gain at relatively low energy and hardware cost. Many exciting progress of RIS technology can be found in [1] and the reference therein.

Along with research deepening, some undesirable characteristic of RIS has come to be realized very recently. One major shortcoming is the so-called *double-fading* loss unveiled by [2], which means the multiplicative attenuation loss experienced by the two-hop channels by way of RIS is generally several orders of magnitude larger than that of the direct channel. To deal with the double-fading issue, a novel *active*-RIS architecture [3]-[4] is proposed, which introduces amplifier into reflecting elements to enlarge the “reflected” signals. Besides the aforementioned double-fading effect, another predominant shortcoming is the restricted coverage since classic RIS can only serve the users lying in the same side as the transmitter (TX). To obtain a 360° full-space coverage, a new architecture named simultaneously transmitting and reflecting (STAR)-RIS [5] emerges lately

which allows a portion of incident signals penetrate through the surface. Similarly, a type of intelligent omni-surface (IOS) has also been proposed recently [6], which realizes simultaneous reflection and refraction of the incoming electromagnetic (EM) waves. Although active-RIS and STAR can partially overcome the shortcomings of classic passive RIS, perfect solution is still missing. In this paper, we propose a novel double-faced active (DFA)-RIS architecture that can simultaneously overcome the double-fading effect and half-space coverage restriction.

Besides proposing the novel DFA-RIS architecture, this paper also considers a more realistic data arriving model. In fact, as pointed by [7], user's communication rate is time-varying and the data packets' arrival is indeed a random process. Therefore, the typical power minimization problem, e.g., [8], which considers constant communication rate (or equivalently, signal-to-interference-plus-noise ratio (SINR)) constraints indeed overlooks the random nature of the user's data rate. Note that there exist a few works considering the random arriving data in their beamforming design [9]-[10]. For instance, the work [9] aims to maximize sum-rate subject to the latency of data queueing. The authors of [10] consider the beamforming design to reduce queueing latency. However, to the best of authors' knowledge, random traffic model is rarely considered in the RIS relevant literature.

The contribution of this paper is summarized as follows.

- A novel DFA-RIS architecture is proposed which can overcome the double-fading effect and simultaneously achieve full-space coverage.
- This paper considers utilizing DFA-RIS to minimize the long term power consumption while data queueing stability is guaranteed. We develop a Lyapunov based online optimization method with its variables being updated highly efficiently.
- Extensive numerical results verify the effectiveness of our proposed solution in queueing stability and power saving.

## II. DOUBLE-FACED ACTIVE RIS ARCHITECTURE

### A. Structure and Implementation

The construction of DFA-RIS is shown in Fig. 1. Two reflecting element arrays are installed on the opposite surfaces of the DFA-RIS. The incident signal is firstly amplified by a reflection amplifier (RA). Then the magnified signal is split

\* Corresponding author.

The work of Yang Liu was supported in part by Grant DUT20RC(3)029 and in part by the Open Research Project Programme of the State Key Laboratory of Internet of Things for Smart City (University of Macau) under Grant SKLIoTSC(UM)-2021-2023/ORP/GA01/2022.

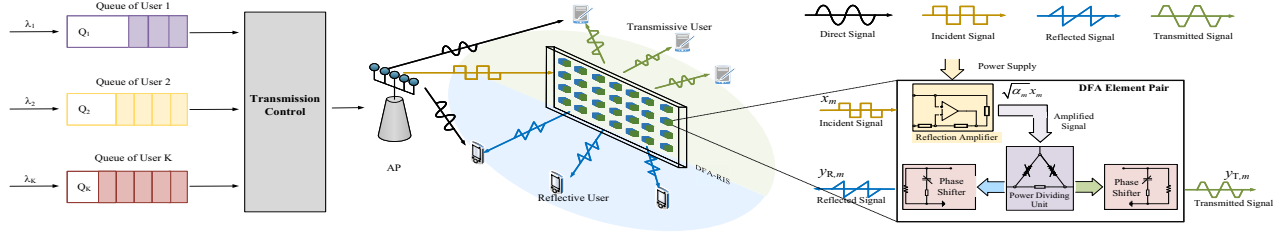


Fig. 1. Signal model of DFA-RIS and its deployment in the queueing-aware multiuser system.

into two portions via a tunable power dividing unit (PDU). The divided signals reach to their respective phase-shifters placed on opposite sides and propagate into two half-spaces.

Note that the proposed DFA-RIS is implementable via existing basic analog components. RA is an analog device capable of magnifying the impinging electromagnetic (EM) waves effectively, which can be implemented via various technologies, e.g. microstrip patch [13] and tunnel diodes [14] to achieve high amplification gain. It is worth noting that RA has already been used in active RIS [3]. Tunable PDU is a device dividing its input signal into two parts with adjustable power dividing ratio (PDR), which is defined as the power ratio between its two output parts. Furthermore, tunable PDUs with wide PDR range have emerged. Via utilizing phase reconfigurable synthesized transmission lines and tunable varactors, the works [15] and [16] implement novel PDUs with PDR range of (-20.5, 21.3)dB and (-39, 29)dB, respectively.

### B. Signal Model

We proceed to elaborate the signal model of DFA-RIS. Suppose that the incident signal of the  $m$ -th pair of DFA-RIS is a complex scalar  $x_m$ . The two output signal, namely reflected and transmitted signal are respectively, given as

$$y_{R,m} = \phi_{R,m} \varsigma_m \sqrt{\alpha_m} (x_m + v_m), \quad (1a)$$

$$y_{T,m} = \phi_{T,m} \sqrt{1 - \varsigma_m^2} \sqrt{\alpha_m} (x_m + v_m), \quad (1b)$$

where  $\alpha_m$  is the amplifying coefficient,  $\varsigma_m$  and  $\sqrt{1 - \varsigma_m^2}$  are the PDRs for the reflected and transmitted signals, respectively. Denote  $\phi_{R,m} \triangleq e^{j\theta_{R,m}}$  and  $\phi_{T,m} \triangleq e^{j\theta_{T,m}}$  as the phase-shift for reflected and transmitted signals, respectively, and  $v_m$  is the thermal noise due to the hardware circuits [3].

By introducing the following notations,

$$\mathbf{E}_R \triangleq \text{Diag}([\varsigma_1, \dots, \varsigma_M]), \phi_i \triangleq [e^{j\theta_{i,1}}, \dots, e^{j\theta_{i,M}}], i \in \{R, T\},$$

$$\mathbf{A} \triangleq \text{Diag}([\sqrt{\alpha_1}, \dots, \sqrt{\alpha_M}]), \Phi_i = \text{Diag}(\phi_i), i \in \{R, T\},$$

$$\mathbf{E}_T \triangleq \text{Diag}([(1 - \varsigma_1^2)^{-\frac{1}{2}}, \dots, (1 - \varsigma_M^2)^{-\frac{1}{2}}]), \quad (2)$$

the reflected and transmitted signals of the DFA-RIS can be respectively expressed in a form of vectors

$$\mathbf{y}_R = \Phi_R \mathbf{E}_R \mathbf{A} (\mathbf{x} + \mathbf{v}), \quad \mathbf{y}_T = \Phi_T \mathbf{E}_T \mathbf{A} (\mathbf{x} + \mathbf{v}). \quad (3)$$

Besides, the power constraint for RIS reads

$$\mathbb{E}\{\|\mathbf{y}_R\|^2 + \|\mathbf{y}_T\|^2\} = \mathbb{E}\{\|\mathbf{A}(\mathbf{x} + \mathbf{v})\|^2\} \leq P_R, \quad (4)$$

where  $P_R$  represents the maximal power supply for the RIS.

## III. DFA-RIS AIDED QUEUEING COMMUNICATION

### A. System Model

As shown in Fig. 1, a typical downlink system is considered with the assistance of a DFA-RIS where an access point (AP) with  $N$  antennas serves  $K$  single-antenna users. Since

the DFA-RIS divides the entire space into two halves, the users located within the same half-space as the AP are denoted as *reflective* users and the remaining ones as *transmissive* users. The reflective and transmissive users are denoted by  $\mathcal{K}_R \triangleq \{1, \dots, K_R\}$  and  $\mathcal{K}_T \triangleq \{K_R+1, \dots, K_R+K_T\}$ , respectively, with  $\mathcal{K} \triangleq \mathcal{K}_R \cup \mathcal{K}_T \triangleq \{1, \dots, K\}$ .

The transmit signal of the AP can be represented as

$$\mathbf{x} = \sum_{k=1}^K \mathbf{f}_k s_k, \quad (5)$$

where  $s_k$  is the information symbol for the  $k$ -th user and  $\mathbf{f}_k \in \mathbb{C}^{N \times 1}$  represents the corresponding transmitting beamformer. For simplicity, we assume that  $s_k \sim \mathcal{CN}(0, 1)$  independently,  $\forall k \in \mathcal{K}$ . We denote the channels associated with the links of AP-RIS, AP-user  $k$ , and RIS-user  $k$  as  $\mathbf{G} \in \mathbb{C}^{M \times N}$ ,  $\mathbf{h}_{d,k} \in \mathbb{C}^{N \times 1}$  and  $\mathbf{h}_{r,k} \in \mathbb{C}^{M \times 1}$ , respectively. To simplify notations, we introduce the category mapping  $i(k) : \mathcal{K} \rightarrow \{R, T\}$ , such that  $i(k) = R$  if  $k \in \mathcal{K}_R$  and  $i(k) = T$  if  $k \in \mathcal{K}_T$ . Additionally, define  $\mathbf{C}_{i(k)} = \Phi_{i(k)} \mathbf{E}_{i(k)} \mathbf{A}$  as the overall beamforming operation of DFA-RIS. It can be expressed in a vector form  $\mathbf{c}_{i(k)} = \text{diag}(\mathbf{C}_{i(k)})$  equivalently. The channel state information (CSI) is assumed to be time-varying and known via channel estimation techniques, e.g., [11], on a time slot basis (see next subsection for the concept “slot”).

The received signal at the  $k$ -th user is given as

$$y_k = \mathbf{h}_{d,k}^H \mathbf{x} + \mathbf{h}_{r,k}^H \mathbf{C}_{i(k)} \mathbf{G} \mathbf{x} + \mathbf{h}_{r,k}^H \mathbf{C}_{i(k)} \mathbf{v} + n_k, \quad (6)$$

where  $n_k \sim \mathcal{CN}(0, \sigma_k^2)$  is receiving noise at user  $k$ . The SINR of the  $k$ -th user can be written

$$\text{SINR}_k(\mathbf{F}, \mathbf{C}) = \frac{|\mathbf{h}_k^H \mathbf{f}_k|^2}{\sum_{j \neq k} |\mathbf{h}_k^H \mathbf{f}_j|^2 + \sigma_v^2 \|\mathbf{C}_{i(k)} \mathbf{h}_{r,k}\|_2^2 + \sigma_k^2}, \quad (7)$$

where  $\mathbf{F} \triangleq [\mathbf{f}_1, \dots, \mathbf{f}_K]$  and  $\mathbf{h}_k^H \triangleq \mathbf{h}_{d,k}^H + \mathbf{h}_{r,k}^H \mathbf{C}_{i(k)} \mathbf{G}$ . Therefore, the achievable rate of the user  $k$  is given as

$$R_k(\mathbf{F}, \mathbf{C}) = \log(1 + \text{SINR}_k(\mathbf{F}, \mathbf{C})). \quad (8)$$

### B. Queueing Model

Cellular systems operate on a time slot basis which can be adjusted from  $62.5\mu\text{s}$  to  $1\text{ms}$  [19] in reality. The time axis is divided into consecutive time slots of period  $\Delta_{\text{slot}}$ , with each slot being indexed by  $t \in \{1, 2, \dots\}$ . The data from the upper layer arrive at the users randomly in the form of packets, which can be depicted as Poisson process [7]. Each user is allocated a dedicated buffer to store the data, which can be denoted as  $Q_k(t)$  within  $t$ -th time slot. The update of the buffer is given by the following dynamics

$$Q_k(t+1) = \max[Q_k(t) - R_k(t), 0] + A_k(t), \forall k \in \mathcal{K}, \quad (9)$$

where  $R_k(t)$  and  $A_k(t)$  denote the transmission rate and the number of arriving packets for the user  $k$ , during the  $t$ -th time slot, respectively. In the following, we introduce a definition to illustrate the description of queueing stability.

**Definition 1.** Queue Stability [18]: a data queue is strongly stable if  $\limsup_{T \rightarrow \infty} \frac{1}{T} \sum_{t=0}^{T-1} \mathbb{E}[Q_k(t)] < \infty, \forall k$  where the expectation  $\mathbb{E}\{\cdot\}$  is taken over the randomness of the data arrivals and channels.

The system is stable only if all the queues in the system are strongly stable. We aim to ensure all users' buffer stable with the known queueing dynamics (9).

### C. Power Consumption Model

The energy consumption of the AP during the  $t$ -th time slot is given as

$$P_{AP}(\mathbf{F}(t)) = \eta_{AP} \sum_{k=1}^K \|\mathbf{f}_k(t)\|_2^2 + P_{c,AP}, \quad (10)$$

where  $\eta_{AP}$  represents the inverse of the energy efficiency of the AP.  $P_{c,AP}$  accounts for the fixed power consumption required for control signaling and so on [12].

Similarly, the energy consumption of the DFA-RIS during the  $t$ -th time slot is given as

$$P_{DFA}(\mathbf{F}(t), \mathbf{C}(t)) = \eta_{DFA} \left[ \sum_k \sum_{i(k)} \|\mathbf{C}_{i(k)}(t) \mathbf{G} \mathbf{f}_k(t)\|_2^2 + \sigma_v^2 \sum_{i(k)} \|\mathbf{C}_{i(k)}(t)\|_F^2 \right] + MP_{c,DFA}, \quad (11)$$

where  $\eta_{DFA}$  stands for the inverse of the energy efficiency of the DFA-RIS.  $P_{c,DFA}$  represents the power consumption of the switch and control circuits at each reflecting element [4].

Based on the above discussion, the total power consumption of the whole system within the  $t$ -th time instant is given by

$$P_{total}(\mathbf{F}(t), \mathbf{C}(t)) = P_{AP}(\mathbf{F}(t)) + P_{DFA}(\mathbf{F}(t), \mathbf{C}(t)) + P_{user}, \quad (12)$$

where  $P_{user}$  is the power consumption of user equipment.

### D. Problem Formulation

We aim at minimizing the long-term power consumption while maintaining queues' stability via jointly optimizing the AP's beamformer and DFA-RIS configuration. In addition, the tunable parameters of the DFA-RIS, i.e., the amplifying coefficients  $\{\alpha_m\}$ , PDRs  $\{\varsigma_m\}$  and phase-shifts  $\{\theta_{R,m}, \theta_{T,m}\}$ , can be determined via the following equivalent identities

$$\alpha_m = |c_{R,m}|^2 + |c_{T,m}|^2, \forall m \in \mathcal{M}, \quad (13a)$$

$$\varsigma_m = [ |c_{R,m}|^2 / (|c_{T,m}|^2 + |c_{R,m}|^2) ]^{-\frac{1}{2}}, \forall m \in \mathcal{M}, \quad (13b)$$

$$\theta_{i,m} = \arctan(\text{Im}\{c_{i,m}\} / \text{Re}\{c_{i,m}\}), i \in \{R, T\}, \forall m \in \mathcal{M}. \quad (13c)$$

Based on the above discussion, the optimization problem can be equivalently formulated in an explicit form of  $\mathbf{C}$  instead of  $(\alpha, \varsigma, \theta_R, \theta_T)$  as follows

$$(P0) : \min_{\mathbf{F}(t), \mathbf{C}(t)} \overline{P_{total}} = \lim_{T \rightarrow \infty} \sum_{t=0}^{T-1} P_{total}(\mathbf{F}(t), \mathbf{C}(t)). \quad (14a)$$

$$\text{s.t.} \limsup_{T \rightarrow \infty} \frac{1}{T} \sum_{t=0}^{T-1} \mathbb{E}[Q_k(t)] < \infty, \quad \forall k, \quad (14b)$$

$$\sum_{k=1}^K \|\mathbf{f}_k(t)\|_2^2 \leq P_{AP}, \quad (14c)$$

$$\sum_{i(k)} \left[ \sum_k \|\mathbf{C}_{i(k)}(t) \mathbf{G} \mathbf{f}_k(t)\|_2^2 + \sigma_v^2 \|\mathbf{C}_{i(k)}(t)\|_F^2 \right] \leq P_R, \quad (14d)$$

where (14a) is defined as the long-term power consumption, (14d) is the power constraint for RIS with respect to  $\mathbf{C}$  and  $P_{AP}$  is the power budget of the AP.

Our problem (P0) is a highly challenging stochastic optimization problem, different from static beamforming design [8] due to the long-term expression (14a) and (14b).

## IV. ALGORITHM DEVELOPMENT

In this section, we will develop an online optimization strategy to resolve the stochastic optimization problem (P0).

### A. Lyapunov Optimization

Lyapunov optimization theory [18] is utilized to address the long-term expression (14a) and (14b). According to [18], define  $\mathbf{Q}(t) = [Q_1(t), Q_2(t), \dots, Q_K(t)]^T$  as a queue vector, to measure the congestion of a queueing system, *Lyapunov function* is proposed as

$$L[\mathbf{Q}(t)] = \frac{1}{2} \sum_{k=1}^K Q_k^2(t). \quad (15)$$

The change in *Lyapunov function* over one slot named conditional *Lyapunov drift* [18] is defined as

$$\Delta[\mathbf{Q}(t)] = \mathbb{E}\{L(\mathbf{Q}(t+1)) - L(\mathbf{Q}(t)) | \mathbf{Q}(t)\}, \quad (16)$$

where the expectation is taken over the randomness of data arrivals and channel realizations, along with AP's beamforming and RIS' configuration.

To minimize the power consumption while stabilizing all queues meanwhile, we seek to minimize *drift-plus-penalty* function [18] for each time slot  $t$ , which is given as

$$\Delta_V[\mathbf{Q}(t)] = \Delta[\mathbf{Q}(t)] + V\mathbb{E}\{P_{total} | \mathbf{Q}(t)\}, \quad (17)$$

where  $V$  is a weighting factor to balance the trade-off between power consumption and queue stabilization. Since the drift-plus-penalty in (17) is difficult to solve directly, we proceed to optimize the upper bound of it, which is given as

$$\Delta_V[\mathbf{Q}(t)] \leq \bar{D} - \sum_k Q_k(t) \mathbb{E}\{R_k(t) | \mathbf{Q}(t)\} + V\mathbb{E}\{P_{total} | \mathbf{Q}(t)\}. \quad (18)$$

where  $\bar{D} = \sum_k Q_k(t) A_{k,\max} + \sum_k \frac{R_{k,\max}^2 + A_{k,\max}^2}{2}$  is a positive finite value with  $\{R_{k,\max}\}$  and  $\{A_{k,\max}\}$  being the upper bounds for  $\{R_k(t)\}$  and  $\{A_k(t)\}$ , under mild assumption that wireless channels and data arrivals are uniformly bounded.

For each time slot  $t$ , adopting *opportunistic expectation minimization* technique, we optimize the instant objective in the expectations of (18), which is formulated as

$$(P1_t) : \min_{\mathbf{F}(t), \mathbf{C}(t)} VP_{total}(\mathbf{F}(t), \mathbf{C}(t)) - \sum_k Q_k(t) R_k(\mathbf{F}(t), \mathbf{C}(t)), \quad (23a)$$

$$\text{s.t.} \sum_{k=1}^K \|\mathbf{f}_k(t)\|_2^2 \leq P_{AP}, \quad (23b)$$

$$\sum_{i(k)} \left( \sum_k \|\mathbf{C}_{i(k)}(t) \mathbf{G} \mathbf{f}_k(t)\|_2^2 + \sigma_v^2 \|\mathbf{C}_{i(k)}(t)\|_F^2 \right) \leq P_R. \quad (23c)$$

### B. Low Complexity Solution

In the following, we drop the index  $t$  for simplicity. To address (P1<sub>t</sub>), we adopt the weighted minimum mean square error (WMMSE) method [17] to convert the expression  $R_k(\mathbf{F}, \mathbf{C})$  into a more tractable form (24) shown at the bottom of this page, with  $\mathbf{w} \triangleq [w_1, \dots, w_K]^T$  and  $\beta \triangleq [\beta_1, \dots, \beta_K]^T$  being the introduced variables [17]. Therefore, the original problem is equivalently transformed into

$$(P1_t') : \min_{\mathbf{F}, \mathbf{C}} VP_{total}(\mathbf{F}, \mathbf{C}) - \sum_k Q_k(t) \tilde{R}_k(\mathbf{F}, \mathbf{C}) \quad (25a)$$

$$\begin{aligned} & \log(1 + \text{SINR}_k(\mathbf{F}, \mathbf{C})) \\ &= \max_{w_k \geq 0, \beta_k} \left\{ \log(w_k) - w_k + 2w_k \text{Re}\{\beta_k^* \mathbf{h}_k^H \mathbf{f}_k\} - w_k |\beta_k|^2 \sum_{j=1}^K |\mathbf{h}_k^H \mathbf{f}_j|^2 - w_k |\beta_k|^2 \sigma_v^2 \|\mathbf{C}_{i(k)} \mathbf{h}_{r,k}\|_2^2 - w_k |\beta_k|^2 \sigma_k^2 + 1 \right\}. \end{aligned} \quad (24)$$

$\triangleq \tilde{R}_k(\mathbf{F}, \mathbf{C})$

s.t. (23b) – (23c). (25b)

In the following, we adopt the block descent decent (BCD) method to tackle (P1'). The update of each block will be elaborated in the following.

### 1) Optimizing Auxiliary Variables $\mathbf{w}$ and $\beta$

When other variables are fixed, following WMMSE method [17], the update of  $\mathbf{w}$  and  $\beta$  are respectively given as

$$\beta_k^* = \left( \sum_{j=1}^K |\mathbf{h}_k^H \mathbf{f}_j|^2 + \sigma_v^2 \|\mathbf{C}_{i(k)} \mathbf{h}_{r,k}\|_2^2 + \sigma_k^2 \right)^{-1} \mathbf{h}_k^H \mathbf{f}_k, \quad (26)$$

$$\mathbf{w}_k^* = (1 - \beta_k^* \mathbf{h}_k^H \mathbf{f}_k + |\beta_k|^2 (\sum_{j \neq k} |\mathbf{h}_k^H \mathbf{f}_j|^2 + \sigma_v^2 \|\mathbf{C}_{i(k)} \mathbf{h}_{r,k}\|_2^2 + \sigma_k^2))^{-1}.$$

### 2) Optimizing The DFA-RIS' Configuration $\mathbf{C}$

We now proceed to optimize RIS' beamforming  $\mathbf{C}$ . Via introducing the notations being defined in (27) as below

$$\xi_{k,j} = \mathbf{f}_j^H \mathbf{h}_{d,k}, \quad \varpi_{k,j} = \text{Diag}(\mathbf{h}_{r,k}^* \mathbf{G} \mathbf{f}_j), \quad (27)$$

$$\kappa_k = \mathbf{h}_{r,k}, \quad \tau_k = \text{Diag}(\mathbf{h}_{r,k}^* \mathbf{G} \mathbf{f}_k), \quad \forall k, j,$$

we rewrite the expression  $\tilde{\mathbf{R}}_k(\mathbf{F}, \mathbf{C})$  with respect to  $\mathbf{C}$  as

$$\begin{aligned} \tilde{\mathbf{R}}_k &= \mathbf{c}_{i(k)}^H \left\{ \omega_k |\beta_k|^2 [\sigma_v^2 \text{diag}(\|\kappa_k\|^2) + \sum_j \varpi_{k,j} \varpi_{k,j}^H] \right\} \mathbf{c}_{i(k)} \\ &\quad + 2\text{Re}\{[\omega_k |\beta_k|^2 \sum_j \xi_{k,j} \varpi_{k,j} - \omega_k \beta_k^* \tau_k] \mathbf{c}_{i(k)}\} + z_k \\ &\triangleq \mathbf{c}_{i(k)}^H \mathbf{Z}_{k,i(k)} \mathbf{c}_{i(k)} + 2\text{Re}\{\mathbf{z}_{k,i(k)}^H \mathbf{c}_{i(k)}\} + z_k, \end{aligned} \quad (28)$$

where  $\mathbf{Z}_{k,i(k)}$  and  $\mathbf{z}_{k,i(k)}$  are defined accordingly and  $z_k$  is a constant irrelevant of  $\mathbf{C}$ .

Besides, RIS's power consumption can be rewritten as

$$\sum_{i(k)} \left( \sum_k \|\mathbf{C}_{i(k)} \mathbf{G} \mathbf{f}_k\|_2^2 + \sigma_v^2 \|\mathbf{c}_{i(k)}\|_2^2 \right) \triangleq \mathbf{c}_R^H \mathbf{P} \mathbf{c}_R + \mathbf{c}_T^H \mathbf{P} \mathbf{c}_T, \quad (29)$$

where  $\tilde{\mathbf{f}}_m$  and  $\mathbf{P}$  are defined as

$$\tilde{\mathbf{F}} = \mathbf{G} \mathbf{F} = [\tilde{\mathbf{f}}_1, \tilde{\mathbf{f}}_2, \dots, \tilde{\mathbf{f}}_m]^T, \quad \mathbf{P} = \text{Diag}(\|\tilde{\mathbf{f}}_1\|_2^2, \dots, \|\tilde{\mathbf{f}}_m\|_2^2) + \sigma_v^2 \mathbf{I}_M. \quad (30)$$

Therefore, the update of  $\mathbf{C}$  is given as follows

$$(\text{P2}) : \min_{\mathbf{x}} \mathbf{x}^H \mathbf{Z} \mathbf{x} - 2\text{Re}\{\mathbf{z}^H \mathbf{x}\}, \quad (31a)$$

$$\text{s.t. } \|\mathbf{x}\|_2^2 \leq P_R, \quad (31b)$$

with the notations being defined as

$$\begin{aligned} \mathbf{S}_{k,i(k)} &= \mathbf{V} \mathbf{P} + \sum_k Q_k \mathbf{Z}_{k,i(k)}, \quad \mathbf{s}_{k,i(k)} = \sum_k Q_k \mathbf{z}_{k,i(k)}, \\ \mathbf{x}_{i(k)} &= \mathbf{P}^{\frac{1}{2}} \mathbf{c}_{i(k)}, \quad \mathbf{x} = [\mathbf{x}_R^T, \mathbf{x}_T^T]^T, \quad \mathbf{Z}_{i(k)} = \mathbf{P}^{-\frac{1}{2}} \mathbf{S}_{i(k)} \mathbf{P}^{-\frac{1}{2}}, \\ \mathbf{z}_{i(k)} &= \mathbf{P}^{-\frac{1}{2}} \mathbf{s}_{i(k)}, \quad \mathbf{Z} = \text{diag}\{\mathbf{Z}_R, \mathbf{Z}_T\}, \quad \mathbf{z} = -[\mathbf{z}_R^T, \mathbf{z}_T^T]^T. \end{aligned} \quad (32)$$

Note that although the problem (P2) is a second order cone program (SOCP) and can be solved via numerical solvers, e.g. CVX [20], we can develop highly efficient solution via the following lemma.

**Lemma 1.** Consider the following optimization problem

$$(\text{P3}) : \min_{\mathbf{x} \in \mathbb{C}^d} \mathbf{x}^H \mathbf{Z} \mathbf{x} - 2\text{Re}\{\mathbf{z}^H \mathbf{x}\}, \quad (33a)$$

$$\text{s.t. } \|\mathbf{x}\|_2^2 \leq P, \quad (33b)$$

where the  $d \times d$  matrix  $\mathbf{Z} \succ 0$ . Assume that  $\mathbf{Z}$  has eigenvalue decomposition  $\mathbf{Z} = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^H$ , with its eigenvalues  $\text{diag}(\mathbf{\Lambda}) = [\lambda_1, \dots, \lambda_d]$  being arranged in descending order, i.e.,  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_d > 0$ . Define  $\mathbf{b} \triangleq \mathbf{U}^H \mathbf{z}$ . The optimal solution  $\mathbf{x}^*$  can be determined in either of the following two cases.

- If  $\|\mathbf{Z}^{-1} \mathbf{z}\|_2^2 \leq P$ , then  $\mu^* = 0$  and the optimal solution  $\mathbf{x}^*$  is given as

$$\mathbf{x}^* = \mathbf{Z}^{-1} \mathbf{z}. \quad (34)$$

- Otherwise,  $\mu^*$  is positive. The optimal value  $\mu^*$  and  $\mathbf{x}^*$  is determined via the following equality

$$\mathbf{b}^H (\mathbf{\Lambda} + \mu^* \mathbf{I})^{-2} \mathbf{b} = P, \quad \mathbf{x}^* = (\mathbf{Z} + \mu^* \mathbf{I})^{-1} \mathbf{z}. \quad (35)$$

The proof is omitted here due to the space limit.

By virtue of Lemma 1, the problem (P2) can be efficiently solved.

### 3) Optimizing The AP's Beamformer $\mathbf{F}$

The update of the beamformer  $\mathbf{F}$  is meant to solve the following problem

$$(\text{P4}) : \min_{\mathbf{f}} \mathbf{f}^H \mathbf{D} \mathbf{f} - 2\text{Re}\{\mathbf{q}^H \mathbf{f}\} \quad (36a)$$

$$\text{s.t. } \|\mathbf{f}\|_2^2 \leq P_{AP}, \quad (36b)$$

$$\mathbf{f}^H \mathbf{D}_1 \mathbf{f} \leq P_1, \quad (36c)$$

with the coefficients introduced above being defined as

$$\mathbf{q} = [Q_1 w_1 \beta_1 \mathbf{h}_1^T, \dots, Q_K w_K \beta_K \mathbf{h}_K^T]^T, \quad \mathbf{f} = [\mathbf{f}_1^T, \dots, \mathbf{f}_K^T]^T,$$

$$\mathbf{D}_1 = \sum_{i(k)} \mathbf{G}^H \mathbf{C}_{i(k)}^H \mathbf{C}_{i(k)} \mathbf{G}, \quad P_1 = P_R - \sigma_v^2 \sum_{i(k)} \|\mathbf{c}_{i(k)}\|_2^2,$$

$$\mathbf{D}_3 = \mathbf{I}_K \otimes \sum_k Q_k w_k |\beta_k|^2 \mathbf{h}_k \mathbf{h}_k^H, \quad \mathbf{D} = \mathbf{D}_2 + \mathbf{D}_3,$$

$$\mathbf{D}_2 = \mathbf{I}_K \otimes (\eta_{AP} \mathbf{V} \mathbf{I}_N + \eta_{DFA} \mathbf{V} \mathbf{D}_1). \quad (37)$$

To again utilize Lemma 1 and develop efficient solution, we introduce a copy  $\mathbf{y}$  of  $\mathbf{f}$  to deal with the coupling constraints (36b) and (36c). To address the constraint  $\mathbf{y} = \mathbf{f}$ , following the framework of ADMM [21], we turn to optimizing the augmented Lagrangian (AL) function which is given as

$$(\text{P5}) : \min_{\mathbf{f}, \mathbf{y}} \mathbf{f}^H \mathbf{D} \mathbf{f} - 2\text{Re}\{\mathbf{q}^H \mathbf{f}\} + \frac{\rho}{2} \|\mathbf{f} - \mathbf{y}\|^2 + \text{Re}\{\boldsymbol{\lambda}^H (\mathbf{f} - \mathbf{y})\}, \quad (38a)$$

$$\text{s.t. } \mathbf{f}^H \mathbf{I}_{KN} \mathbf{f} \leq P_{AP}, \quad (38b)$$

$$\mathbf{y}^H \mathbf{D}_1 \mathbf{y} \leq P_1, \quad (38c)$$

with  $\boldsymbol{\lambda}$  being the Lagrangian multiplier associated with the constraint  $\mathbf{y} = \mathbf{f}$  and  $\rho$  being a predefined positive constant.

According to the ADMM framework [21], we solve the problem (P5) via alternatively updating variables  $\mathbf{y}$ ,  $\mathbf{f}$  and  $\boldsymbol{\lambda}$ . With  $(\mathbf{f}, \boldsymbol{\lambda})$  fixed, the optimization of  $\mathbf{y}$  is reduced to

$$(\text{P6}) : \min_{\mathbf{y}} \|\mathbf{y}\|_2^2 - 2\text{Re}\{\mathbf{q}_1^H \mathbf{y}\}, \quad (39a)$$

$$\text{s.t. } \mathbf{y}^H \mathbf{D}_1 \mathbf{y} \leq P_1, \quad (39b)$$

where  $\mathbf{q}_1 \triangleq \mathbf{f} + \rho^{-1} \boldsymbol{\lambda}$ . Obviously, (P6) can be efficiently solved via taking advantage of Lemma 1.

When  $(\mathbf{y}, \boldsymbol{\lambda})$  is fixed, the update of  $\mathbf{f}$  is given as

$$(\text{P7}) : \min_{\mathbf{f}} \mathbf{f}^H \mathbf{W} \mathbf{f} - 2\text{Re}\{\mathbf{w}^H \mathbf{f}\}, \quad (40a)$$

$$\text{s.t. } \|\mathbf{f}\|_2^2 \leq P_{AP}, \quad (40b)$$

where  $\mathbf{W} = \mathbf{D} + \frac{\rho}{2} \mathbf{I}$  and  $\mathbf{w} = \mathbf{q} + \frac{\rho}{2} \mathbf{y} - \frac{\boldsymbol{\lambda}}{2}$ . This problem be solved by resorting to Lemma 1, again.

The update of Lagrangian dual variables  $\boldsymbol{\lambda}$  is shown as

$$\boldsymbol{\lambda}^{(n+1)} = \boldsymbol{\lambda}^{(n)} + \rho (\mathbf{f} - \mathbf{y}). \quad (41)$$

The analytic based solution to optimize  $\mathbf{f}$  is summarized in Alg. 1. To conclude, the overall low complexity solution to the problem (P0) is summarized in Alg. 2.

**Theorem 1.** Assume that the sequence of  $\{\boldsymbol{\lambda}^{(n)}, \mathbf{f}^{(n)}, \mathbf{y}^{(n)}\}$  is generated by Algorithm 1. Then any limit point of  $\{\mathbf{f}^{(n)}\}$  is an optimal solution to the original problem (P4).

*Proof.* Note that the problem (P4) is convex. The hypothesis of [[22], Sec. 3.4, Prop. 4.2] is satisfied since  $\mathbf{f}$  and  $\mathbf{y}$  are connected by  $\mathbf{I}_{KN} \mathbf{f} = \mathbf{y}$  with  $\mathbf{I}_{KN}^T \mathbf{I}_{KN}$  being invertible. Therefore, the convergence result can be readily obtained following the statement of [[22], Sec. 3.4, Prop. 4.2].

## V. SIMULATION RESULTS

In this section, we evaluate the performance of our proposed algorithm by numerical results. In the experiment network,

---

**Algorithm 1** ADMM Method to Solve (P4)

---

- 1: Initialize  $\mathbf{y}^{(0)}$ ,  $\mathbf{f}^{(0)}$  and  $\boldsymbol{\lambda}^{(0)}$  set  $n = 0$ ;
  - 2: **repeat**
  - 3:   Update  $\mathbf{y}^{(n+1)}$  by solving the problem (P6);
  - 4:   Update  $\mathbf{f}^{(n+1)}$  by solving the problem (P7);
  - 5:   Update  $\boldsymbol{\lambda}^{(n+1)}$  by (41),  $n++$ ;
  - 6: **until** Convergence
- 

---

**Algorithm 2** Low Complexity Power Minimization

---

- 1: Set time slot  $t = 1$ ;
  - 2: **repeat**
  - 3:   Initialize  $\mathbf{F}^{(0)}$ , set  $n = 0$ ;
  - 4:   **repeat**
  - 5:     Update  $\boldsymbol{\beta}^{(n+1)}$  and  $\mathbf{w}^{(n+1)}$  by (26);
  - 6:     Update  $\mathbf{C}^{(n+1)}$  by invoking Lemma 1;
  - 7:     Update  $\mathbf{F}^{(n+1)}$  via Alg. 1;
  - 8:   **until** Convergence
  - 9:   Update  $Q_k(t+1) = \max[Q_k(t) - R_k(t), 0] + A_k(t)$ ;  $t++$ ;
  - 10: **until** All data queues become stable;
- 

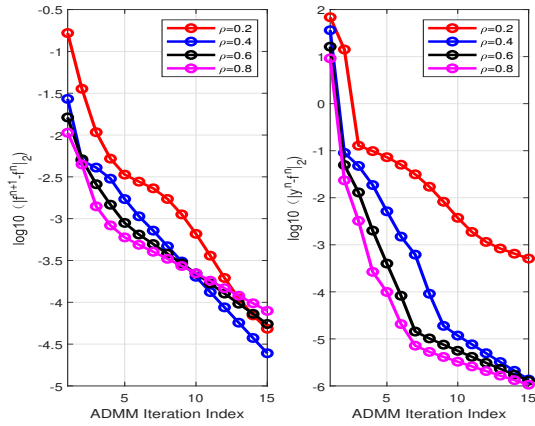


Fig. 2. Convergence of ADMM method.

one AP equipped with  $N = 6$  antennas is 100 meters (ms) apart from one DFA-RIS device with  $M = 100$  elements. The network contains  $K_R = 2$  and  $K_T = 2$  users, which are distributed within a circle centered at the DFA-RIS with the radius of 8m. The heights of the AP, RIS and users are set as 4.5m, 2.5m and 1.5m, respectively. The large-scale fading coefficients are modeled as  $F_k = -30 + 10\kappa \log_{10}(d_k/d_0) + N_k$  dB, with  $\kappa$  being the fading exponent,  $d_0 = 1$  m and the shadow fading coefficient  $N_k \sim \mathcal{N}(0, 2)$ . The fading exponents  $\kappa$  for the links of AP-user, AP-RIS and RIS-user are assumed to be -3.5, -3.1 and -2.8, respectively. The AP-user and RIS-user links follow Rayleigh fading channels while the AP-RIS channel follows Rician fading distribution with Rician factor being 5dB. Besides, Poisson process is adopted to simulate all users' data arrival procedure, i.e.  $A_k(t) \sim \text{Pois}(\bar{\lambda}_k)$  with  $\bar{\lambda}_k$  being the average data rate. The average arrival rates of all users are set as the same, i.e.,  $\bar{\lambda}_k = \bar{\lambda}, \forall k$ , unless stated explicitly. Other parameter settings are specified in Table I.

Fig. 2 examines the convergence behaviour of our proposed

TABLE I: Simulation Parameters

Parameters	Values
Inverse of the emission efficiency of the AP $\eta_{\text{AP}}$	1.2
Inverse of the emission efficiency of the RIS $\eta_{\text{DFA}}$	1.2
Dissipated power of the AP $P_{\text{c,AP}}$	2W
Dissipated power of each reflecting element $P_{\text{c,DFA}}$	5mW
Dissipated power of the users $P_{\text{c,k}}$	10dBm
Noise power at the DFA and the users $\sigma_k^2, \sigma_{\text{DFA}}^2, \forall k$	-90dBm
Maximal transmit power at the AP $P_{\text{AP}}$	20dBm
Maximal power supply at the DFA-RIS $P_{\text{R}}$	16dBm

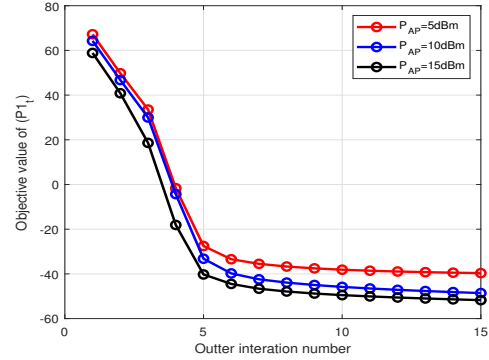


Fig. 3. Solving the online sub-problem.

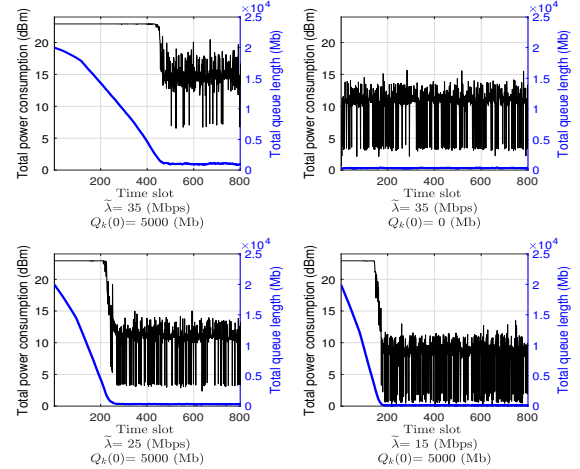


Fig. 4. Power consumption and queue length iterates.

ADMM method to optimize  $\mathbf{f}$ . We first normalize<sup>1</sup> the coefficients before solving (P4). It is observed that our proposed method converges within 15 iterations.

Fig. 3 demonstrates the overall convergence in solving the online problem (P1<sub>t</sub>). As suggested by the figure, our proposed algorithm keeps decreasing when the iteration loop proceeds and converges within the 15 iterations.

Fig. 4 illustrates power expenditure and the total queue length (i.e., sum of all queues) in the long run. It can be seen from all the sub-figures except for the upper-right one, that the power consumption decreases along with the congestion being mitigated. Furthermore, lower data arrival rate leads to lower consumption and shorter queue lengths after reaching

<sup>1</sup>The term “normalize” here means scaling (36a) and (36c) of (P4) such that the maximal eigenvalues of  $\mathbf{D}$  and  $\mathbf{D}_1$  are both 1.



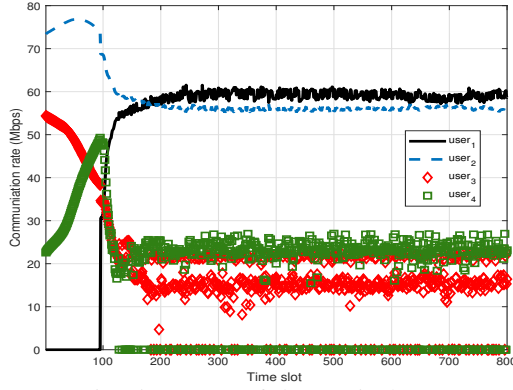


Fig. 5. Impact of data arrival rates.

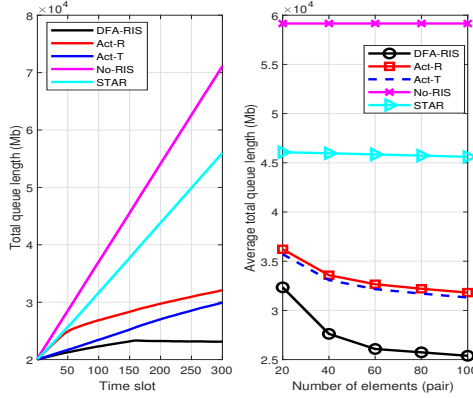


Fig. 6. Performance of different RIS techniques.

stabilization. Combining the sub-figures in the first line, we can conclude that, compared with occupied initial buffer cases, the queue lengths maintain small from the very beginning and power consumption is slightly smaller with clear initial buffer.

Fig. 5 shows the impact of the data arrival rates. We set  $\tilde{\lambda}_1 = \tilde{\lambda}_2 = 45\text{Mbps}$  while  $\tilde{\lambda}_3 = \tilde{\lambda}_4 = 15\text{Mbps}$  with  $Q_k(0) = 5000\text{Mb}, \forall k$ . Before all queues stay stable, user 3 and 4 aim to alleviate the crowd queues in the first 100 slots. When reaching stability, all users' communication rates are in line with their respective data arrival rates which confirms the effectiveness of our Lyapunov guided online optimization strategy.

Fig. 6 compares the queue length using various RIS architectures. Both the act-RIS(R) and act-RIS(T) scenario where the RIS is facing to the AP or the opposite direction, respectively, are considered for single-faced active RIS. The left sub-figure shows that the queue can keep stable only if DFA-RIS is employed, confirming DFA-RIS' superiority compared to all other competing schemes. The right part shows that when the number of RIS' element  $M$  increases, the reduction of the stabilized queue lengths is significant.

## VI. CONCLUSIONS

In this paper, we consider a power minimization problem where the realistic data arriving process is considered with the assistance of our proposed novel DFA-RIS architecture that can magnify the incident signal and achieve full-space coverage at the same time. We successfully develop analytic-based solution to deal with the per-slot sub-problems illuminated by the

Lyapunov optimization framework. Extensive numerical results verify our proposals.

## REFERENCES

- [1] Q. Wu, S. Zhang, B. Zheng, C. You, and R. Zhang, "Intelligent reflecting surface aided wireless communications: A tutorial," *IEEE Trans. Commun.*, vol. 69, no. 5, pp. 3313-3351, May. 2021.
- [2] M. Najafi, V. Jamali, R. Schober, and H. V. Poor, "Physics-based modeling and scalable optimization of large intelligent reflecting surfaces," *IEEE Trans. Commun.*, vol. 69, no. 4, pp. 2673-2691, Apr. 2021.
- [3] Z. Zhang, L. Dai, X. Chen, C. Liu, F. Yang, R. Schober, and H. V. Poor, "Active RIS vs. passive RIS: Which will prevail in 6G?," Jan. 2022. [Online]. Available: <http://arxiv.org/abs/2103.15154>.
- [4] R. Long, Y.-C. Liang, Y. Pei, and E.G. Larsson, "Active reconfigurable intelligent surface-aided wireless communications," *IEEE Trans. Wireless Commun.*, vol. 20, no. 8, pp. 4962-4957, Aug. 2021.
- [5] Y. Liu, X. Mu, J. Xu, R. Schober, Y. Hao, H. V. Poor, and L. Hanzo, "STAR: Simultaneous transmission and reflection for 360° coverage by intelligent surfaces," *IEEE Wireless Commun.*, vol. 26, no. 1, pp. 167-171, Jan. 2022.
- [6] S. Zeng and *et al.*, "Intelligent omni-surfaces: reflection-refraction circuit model, full-dimensional beamforming, and system implementation," *IEEE Trans. Commun.*, 2022, Early Access.
- [7] 3GPP, "TR 36.885 v14.0.0: Study on LTE-based V2X services," ETSI Technical Specification, Technical Report Release 14, Jun. 2016.
- [8] R. Li, B. Guo, M. Tao, Y. -F. Liu, and W. Yu, "Joint design of hybrid beamforming and reflection coefficients in RIS-aided mmWave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 70, no. 4, pp. 2404-2416, Apr. 2022.
- [9] I. M. Braga Jr., R. P. Antonioli, G. Fodor, Y. C. B. Silva, C. F. M. e Silva, and W. C. Freitas Jr., "Joint resource allocation and transceiver design for sum-rate maximization under latency constraints in multicell MU-MIMO systems," *IEEE Trans. Commun.*, vol. 69, no. 7, pp. 3036-3040, Sep. 2021.
- [10] G. Venkatraman, A. Tolli, M. Juntti, and L. Tran., "Traffic aware resource allocation schemes for multi-cell MIMO-OFDM systems," *IEEE Trans. Commun.*, vol. 64, no. 11, pp. 2730-2745, Jun. 2016.
- [11] L. Wei, C. Huang, G. C. Alexandropoulos, C. Yuen, Z. Zhang, and M. Debbah, "Channel estimation for RIS-empowered multi-user MISO wireless communications," *IEEE Trans. Commun.*, vol. 69, no. 6, pp. 4144-4157, Jun. 2021.
- [12] E. Björnson, J. Hoydis, and L. Sanguinetti, "Massive MIMO networks: Spectral, energy, and hardware efficiency," *Foundat. Trends Signal Process.*, vol. 11, no. 3-4, pp. 154-655, 2017.
- [13] K. K. Kishor and S. V. Hum, "An amplifying reconfigurable reflectarray antenna," *IEEE Trans. Ant. Propag.*, vol. 60, no. 1, pp. 197-205, 2012.
- [14] F. Farzami, S. Khaledian, B. Smida, and D. Erricolo, "Reconfigurable dual-band bidirectional reflection amplifier with applications in Van Atta array," *IEEE Trans. Microw. Theory Techn.*, no. 65, no. 11, pp. 4198-4207, Nov. 2017.
- [15] H. N. Chu and T.-G. Ma, "Tunable directional coupler with very wide tuning range of power division ratio," *IEEE Microw. Wireless Comp. Lett.*, vol. 29, no. 10, pp. 652-654, Oct. 2019.
- [16] H. Peng, P. Lei, H. Yang, S. Zhao, X. Ding, "II-type reconfigurable coupler based on a complementary tunable method," *J. Electr. Waves App.*, 2021.
- [17] Q. Shi, M. Razaviyayn, Z.-Q. Luo, and C. He, "An iteratively weighted MMSE approach to distributed sum-utility maximization for a MIMO interfering broadcast channel," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4331-4340, Sep. 2011.
- [18] M. J. Neely, "Stochastic network optimization with application to communication and queueing systems," *Synth. Lect. Commun. Netw.*, vol. 3, no. 1, pp. 1-211, Jan. 2010.
- [19] 3GPP, "TR 38.211 v15.2.0: Physical Channels and Modulation," ETSI, Technical Specification Release 15, Jun. 2018.
- [20] M. Grant and S. Boyd, *CVX: Matlab software for disciplined convex programming, version 2.1*, <http://cvxr.com/cvx>, Mar. 2014.
- [21] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. and Trends in Machine Learning*, vol. 3, no. 1, pp. 1-122, 2011.
- [22] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*. Athena Scientific, Belmont, MA., 1997.