This article has been accepted for publication in IEEE Transactions on Signal Processing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TSP.2022.3221345

Spectral Efficiency Maximization For Double-Faced Active Reconfigurable Intelligent Surface

Yang Liu, Yanan Ma, Ming Li, Qingqing Wu, and Qingjiang Shi

***Abstract*—Although the recently emerging reconfigurable intelligent surface (RIS) has exhibited great potentials to enhance wireless communications, it has intrinsic defects as well, i.e., the double fading effect severely restricts its coverage and the conventional single-faced structure restrains its service within only half-space. To overcome these shortcomings, this paper proposes a novel double-faced active (DFA)-RIS structure. This new design can effectively magnify the impinging signal and refract it to fulfill 360° full-space coverage. Moreover, we study the joint design of transmit beamforming and the DFA-RIS configuration to maximize spectral efficiency (SE) in a multi-user multi-input-single-output system. Per-element power constraints (PEPCs) are considered to precisely reflect the limited magnifying capability of each RIS element's amplifier, which makes our problem have great number of constraints and hence become highly difficult. We successfully develop an analytic-based, highly computation-parallelizable and convergence guaranteed algorithm to resolve this challenge. Interestingly, our considered problem subsumes other state-of-the-art RIS structures, e.g. the single-faced active RIS and the passive reflective-transitive RIS, as special cases and therefore our proposal indeed provides a unifying solution. Extensive numerical results are presented to verify the efficiency of our solution and the advantageous performance of DFA-RIS.**

***Index terms*— Reconfigurable intelligent surface, joint beamforming, per-element power constraint, analytic-based solution.**

I. Introduction

Lately the reconfigurable intelligent surface (RIS), which is also widely known as intelligent reflecting surface (IRS), has been cast with great interest and been envisioned as a viable enhancement for the next generation wireless communication system [1], [2]. RIS is a planar array comprising a great number of reflecting elements that are made of meta-materials and can effectively adjust the phases of the impinging electromagnetic (EM) waveforms in a programmable manner. RIS is characterized by its high energy efficiency and low hardware cost. Its flexible phase-shifting control can bring about additional "passive" beamforming gain and boost communication performance from various aspects. Many exciting applications of RIS can be found in the survey [1], [2] and references therein.

A. Background

Although the RIS has exhibited great potentials in many applications, people have recently come to realize some intrinsic defects of the purely passive reflective RIS. One predominant shortcoming is the so-called *double fading* effect [3]. Double fading effect means the reflected signal, which experiences the accumulated attenuation through the cascaded TX-RIS-RX channel, is generally several orders of multitude smaller than that of the direct TX-RX signal. To alleviate the double fading loss, the RIS needs to be deployed very close to the either the transmitter (TX) or the receiver (RX) [4], [5], which severely restricts the "coverage" of the passive-RIS. For instance, the paper [6] has performed a comparison between the passive-RIS and the classical decode-and-forward relay and concluded that RIS can hardly beat the relaying scheme unless it has unrealistically tremendous number of elements. To combat the double fading curse, the work [7] proposes the active-RIS construction, where an amplifier is equipped by each RIS element to enlarge the "reflected" signal. Active-RIS can effectively compensate the double fading loss while still maintaining a low hardware cost (since RF chain is still unused). As demonstrated by the real field experiment results in [7], compared to the no-RIS baseline, the active-RIS can achieve a capacity gain that is several tens of times higher than that of the passive-RIS. Similar results have also been obtained in [4]. Within a typical indoor scenario on tens of meters scale, an active-RIS with identical transmit power with the TX can yield a receive signal-to-noise (SNR) $30-40$dB higher than that of the passive-RIS. The authors in [8] have shown that, in a single-user communication system, active-RIS can obtain above 25dB capacity gain over the passive-RIS. Very recently, emerging studies focusing on exploiting active-RIS to improve system performance, including energy efficiency, security and bit error rate (BER), have been conducted in [9]–[11].

Besides the aforementioned double fading effect, another major limitation of the classical reflective RIS is its "half-space" coverage. To be assisted by the RIS, both the TX and RX have to be located within the half-space that the RIS reflecting surface is facing up. To obtain a 360° coverage, [12] recently proposes a novel simultaneous transmitting and reflecting (STAR) construction. This novel structure splits the incident EM waves into two portions—one portion is reflected back against the surface and the other portion is allowed to penetrate through the surface and propagate into the backward half-space. The work [13] has established a general

This work is supported in part by the National Natural Science Foundation of China (Grant No. 61971088), in part by the Fundamental Research Funds for the Central Universities (Grant No. DUT20RC(3)029 and No. DUT20GJ214), in part by the Natural Science Foundation of Liaoning Province (Grant No. 2020-MS108), and in part by the Open Research Fund of National Mobile Communications Research Laboratory, Southeast University (Grant No. 2021D08), in part by SKL-IoTSC(UM)-2021-2023/ORP/GA30/2022.

Y. Liu is the corresponding author.

Y. Liu, Y. Ma and M. Li are with the School of Information and Communication Engineering, Dalian University of Technology, Dalian 116024, China, email: yangliu_613@dlut.edu.cn, mayanan@mail.dlut.edu.cn, mli@dlut.edu.cn

Q. Wu is with Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai 200240, China, email: wu.qq1010@gmail.com

Q. Shi is with the School of Software Engineering, Tongji University, Shanghai 201804, China, and also with the Shenzhen Research Institute of Big Data, Shenzhen 518172, China email: shiqj@tongji.edu.cn.

This article has been accepted for publication in IEEE Transactions on Signal Processing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TSP.2022.3221345

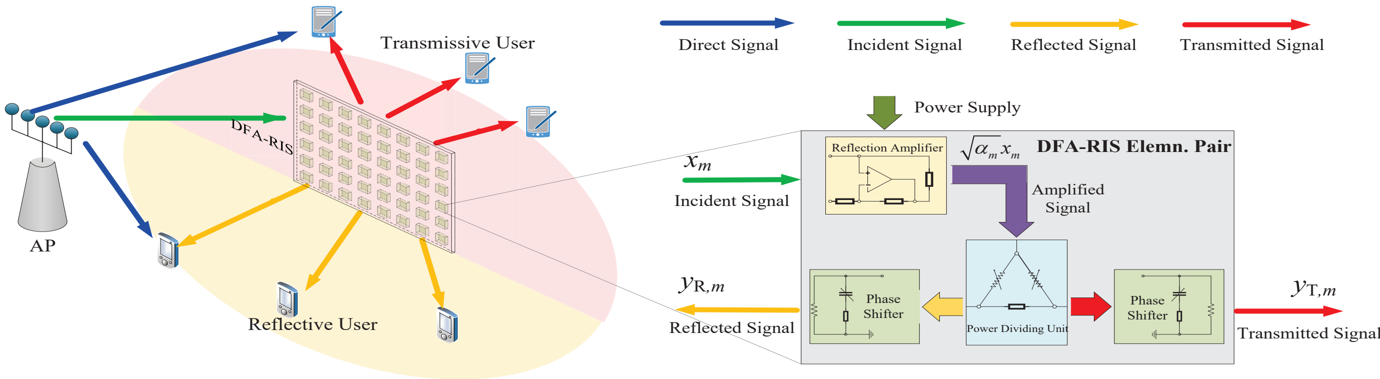


Fig. 1: Construction of DFA-RIS and its deployment in a multi-user MISO system.

hardware model of this novel STAR-RIS and showed it can achieve a higher diversity order than the passive-RIS. The research [14] investigates three possible operating protocols of STAR-RIS and considers the joint power-splitting and phase-shifting towards power minimization in a two-user network. Very interestingly, the key idea of STAR-RIS coincides with another emerging RIS structure named intelligent omni-surface (IOS) proposed in [15]–[18], where simultaneously reflecting and transmitting (refracting) are implemented using different hardware techniques. The study in [15] shows that IOS can improve the SE for SISO single-user system. The implementation technique of IOS is elaborated in [16]. The works [17], [18] investigate utilizing IOS to boost SE of a multi-user system under the assumption that no interference occurs between users (by conducting dirty-paper encoding or zero-forcing beamforming).

B. Motivation

The above discussions reflect the fact that the pure passive RIS, including the STAR-RIS [12]–[14] and IOS [15]–[18], suffers from severe double fading effect, and at the same time the conventional single-faced active-RIS [4], [5], [7]–[11] has the restrictive half-space coverage. These observations motivate us to explore a novel RIS architecture which can effectively magnify the "reflected" EM waves and at the same time fully cover all the 3D-space. This goal can actually be fulfilled by a novel double-faced active (DFA)-RIS architecture (as illustrated in Fig.1) that can amplify the impinging EM waves and simultaneously allow them pass through the intelligence surface. The novel structure of the DFA-RIS and its potential implementation schemes will be elaborated in Sec.II-A shortly. In addition to the classical reconfigurable phase-shifting capability, each DFA-RIS element can also tune the amplifying coefficient and the power distribution between the two possible propagation directions of the incoming EM waves. Therefore, the DFA-RIS cherishes stronger beamforming capability in contrast to other cutting-the-edge RIS techniques and hence achieves advantageous spacial multiplexing gain (as will be verified later). Along with its stronger beamforming capability, the configuration of the DFA-RIS has also become much more complicated. The tuning of its reconfigurable parameters needs to be carefully investigated.

C. Contributions

Based on the above novel DFA-RIS design, we proceed to investigate the SE maximization in a DFA-RIS assisted multi-input-single-output (MISO) multi-user system via jointly optimizing TX beamforming and DFA-RIS configuration. The contributions of this paper are specified as follows.

(i) Our proposed DFA-RIS structure wisely combines the advantage of the active-RIS [4], [5], [7]–[11] and the transmitting-type RIS [12]–[18]. Specifically, the DFA-RIS can effectively combat the double fading effect via magnifying impinging signals and also extend its service coverage to full 360° full-space via forwarding signals into the backward half-space.

(ii) Mathematically, our considered problem of jointly optimizing TX beamforming and DFA-RIS configuration to maximize SE subsumes the same task associated with the active-RIS [4], [5], [7]–[11], STAR-RIS [12]–[14] and IOS [15]–[18] as special cases (as will be clear in Sec.II-D). Therefore our developed algorithm provides a unifying solution to all the above RIS techniques.

(iii) We consider the PEPCs in the DFA-RIS configuration. This consideration has been missing in almost all existing literature on active-RIS beamforming design [4], [5], [7], [9]–[11] except for [8], where a hard limit is imposed for each element's amplifier. The great number (hundreds or even thousands) of element-wise PEPCs makes our problem highly challenging. We have successfully designed an analytic-based (not requiring numerical solvers like CVX [19]), highly computation parallelizable and convergence guaranteed algorithm to resolve this challenge.

(iv) The optimization of power-splitting coefficients is still an open problem for the control of STAR-RIS [12]–[14] and IOS [15]–[18]. The existing studies [12]–[18] generally assume fixed power-splitting and/or do not consider interference between multi-users in their beamforming design. Here, as a sub-problem, we propose a novel solution to optimize power-splitting coefficients for generic settings.

(v) Last but not least, extensive numerical results verify the efficiency of our proposed solution and demonstrate the advantageous performance of the DFA-RIS.

II. System Model and Problem Formulation

In this section, we will specify the architecture, possible implementation techniques and mathematical model of the newly proposed DFA-RIS architecture. Based on that, the problem of optimizing DFA-RIS' configuration will be formally introduced.

A. DFA-RIS' Architecture, Implementation and Signal Model

As shown in Fig.1, in the novel DFA-RIS design, "reflecting" elements are deployed on both sides of a plate. The two opposite element arrays are arranged in alignment. Each two co-located elements on the opposite faces form a binding pair with their back-ends connected by controlling circuits embedded within the plate, as shown in the right half of Fig.1. Similar to the active-RIS in [7], the incident signal is firstly fed into a reflection amplifier (RA) and magnified. The output of the RA is then forwarded into a tunable power dividing unit (PDU) and split into two portions, with each portion reaching one of the two phase-shifters installed on the opposite faces. The RA controls one element pair's transmit power, the tunable PDU adjusts the powers allocated to the two faces and the pair of phase-shifters configure the emitting signals' phases.

Next we elaborate potential implementation schemes of the proposed DFA-RIS architecture. Besides the phase-shifters that have been widely used in the passive RIS, realizations of the key ingredients—RA and tunable PDU can be referred to diverse up-to-date techniques. Specifically, RA is an equivalent active-load amplifier with negative input resistance and can effectively enlarge the impinging EM waves. For instance, the work [20] utilizes aperture-coupled microstrip patch to implement RA and builds reconfigurable reflect-array that can simultaneously amplify and phase-shift the incoming EM waves. The authors of [21] and [22] exploit tunnel diodes to implement RA and construct dual frequency-band and full-duplex bidirectional reflect-arrays having beam-steering capability. The authors of [23] adopt CMOS technology to implement RA and use it to construct a full-duplex active reconfigurable reflect-array in [24], which has high power gain and agile beamforming capability. Recently, the authors of [25] implement an RA with power gain up to 40dB and use it to build a Radio Frequency Identification (RFID) tag, which has successfully conducted communication of bandwidth 1MHz over 1.2km distance, as reported in [26]. PDU is a commonly used preliminary gradient in microwave circuit system. It divides one signal into two parts with each part having a predefined proportion of power. The ratio between the PDU's two outputs is called power dividing ratio (PDR). Conventional PDUs have fixed PDR. Within the past decade, tunable PDU technology has emerged and has made significant progress. The first PDU having tunable PDR is implemented in [27] by using tunable diodes. Tunable PDUs with very wide PDR range have already been implemented using various techniques within the last five years. For instance, the authors of [28] implement a tunable rat-rate coupler with its PDR ranging from -25dB to 25dB. By adopting synthesized transmission lines (STLs) technique, the works [29] and [30] implement novel PDUs with PDR range of $(-25, 25)$dB and $(-39, 29)$dB, respectively. Besides, the very recent work [31] realizes a novel complementary Π-type coupler and achieves a PDR ranging from -20.5dB to 21.3dB. Excitingly, tunable PDUs have been already been employed in antenna front-end circuits to empower antenna-arrays with beamsteering capability in [32] and [33].

Remark II.1. *Note that the aforementioned existing techniques, though not straight implementation schemes for DFA-RIS, serve as valuable reference. For instance, inspired by [21], we can leverage RA to make a bi-directional amplifier to enlarge the signal and send it to tunable PDU. Nevertheless, there are still a number of critical issues worth further studying to truly implement DFA-RIS. For instance, the potential feedback loop and oscillation issue when concatenating RA and PDU deserves careful investigation. These open issues in circuit design are left to future research.*

Mathematically, the signal flow of the aforementioned DFA-RIS can be modelled as follows. Suppose that the incident signal to the m-th reflecting pair is a complex scalar x_m. Borrowing the terminology coined in [12], the associated output signals at the two opposite phase-shifters are denoted as the *reflected* part $y_{\mathsf{R},m}$ and the *transmitted* part $y_{\mathsf{T},m}$, respectively, which are given as[1]

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

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

with α_m being the amplifying coefficient, ς_m being the power-splitting parameter, $\phi_{\mathsf{R},m} \triangleq e^{\jmath\theta_{\mathsf{R},m}}$ and $\phi_{\mathsf{T},m} \triangleq e^{\jmath\theta_{\mathsf{T},m}}$ being, respectively, phase-shifts of the reflected and transmitted signals and v_m being the thermal noise due to the hardware circuits [7].

Suppose that there are totally M reflecting element pairs. By packing all the M signals x_m's, $y_{\mathsf{R},m}$'s, $y_{\mathsf{T},m}$'s and the noise v_m's into vectors $\mathbf{x} \triangleq [x_1, \cdots, x_M]^T$, $\mathbf{y}_\mathsf{R} \triangleq [y_{\mathsf{R},1}, \cdots, y_{\mathsf{R},M}]^T$, $\mathbf{y}_\mathsf{T} \triangleq [y_{\mathsf{T},1}, \cdots, y_{\mathsf{T},M}]^T$ and $\mathbf{v} \triangleq [v_1, \cdots, v_M]^T$, respectively, we can describe the signal model of the DFA-RIS compactly as follows

$$\mathbf{y}_\mathsf{R} = \mathbf{\Phi}_\mathsf{R}\mathbf{E}_\mathsf{R}\mathbf{A}\big(\mathbf{x}+\mathbf{v}\big), \qquad \mathbf{y}_\mathsf{T} = \mathbf{\Phi}_\mathsf{T}\mathbf{E}_\mathsf{T}\mathbf{A}\big(\mathbf{x}+\mathbf{v}\big), \tag{2}$$

where the diagonal coefficient matrices are defined as $\mathbf{A} \triangleq \mathsf{Diag}\big([\sqrt{\alpha_1}, \cdots, \sqrt{\alpha_M}]\big)$, $\mathbf{E}_\mathsf{R} \triangleq \mathsf{Diag}\big([\varsigma_1, \cdots, \varsigma_M]\big)$, $\mathbf{E}_\mathsf{T} \triangleq \mathsf{Diag}\big(\big[\sqrt{1-\varsigma_1^2}, \cdots, \sqrt{1-\varsigma_M^2}\big]\big)$ and $\mathbf{\Phi}_\imath \triangleq \mathsf{Diag}(\boldsymbol{\phi}_\imath)$ with $\boldsymbol{\phi}_\imath \triangleq [e^{\jmath\theta_{\imath,1}}, \cdots, e^{\jmath\theta_{\imath,M}}]^T$, $\imath \in \{\mathsf{R}, \mathsf{T}\}$.

Considering the fact that RIS generally utilizes large amount of low-cost amplifiers to decrease hardware cost, each single amplifier generally has limited transmit power. This implies the per-element power constraint (PEPC) for each amplifier should be taken into account, which is given as

$$\mathbb{E}\{\alpha_m|x_m + v_m|^2\} \leq \mathsf{P}_m, \quad \forall m \in \mathcal{M}, \tag{3}$$

where $\mathcal{M} \triangleq \{1, \cdots, M\}$ and $\mathsf{P}_m = \rho\widetilde{\mathsf{P}}_m$ is the effective transmit radio power (TRP) limit with ρ being the energy conversion coefficient and $\widetilde{\mathsf{P}}_m$ being the actual power consumption. Note that per-antenna power constraint (PAPC) has always been widely adopted to precisely describe the

[1] We can employ polarization to isolate the input from the output on the reflecting element. In this case, there is a polarization change between the incoming and the "reflected" EM waves. That is, the RIS is limited to operating in one polarization.

This article has been accepted for publication in IEEE Transactions on Signal Processing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TSP.2022.3221345

per-antenna amplifier's behaviour in multi-antenna systems. See [34], [35] and references therein. Similar rationale is also applicable to the active RIS scenario. Besides, the total transmission power of DFA-RIS should be limited, which is determined by the capacity of the powerline and given as

$$\mathbb{E}\{\|\mathbf{y}_{\mathsf{R}}\|_2^2+\|\mathbf{y}_{\mathsf{T}}\|_2^2\}=\mathbb{E}\{\|\mathbf{A}(\mathbf{x}+\mathbf{v})\|_2^2\}\le \mathsf{P}_{\mathsf{R}}, \tag{4}$$

with P_{R} being the total effective TRP of the RIS.

B. DFA-RIS Assisted Communication System

This paper considers a downlink multi-user MISO system assisted by DFA-RIS, as shown in Fig.1. An access point (AP) equipped with N antennas serves K single-antenna mobile users, which are labelled by the index set $\mathcal{K}\triangleq\{1,\cdots,K\}$. The DFA-RIS divides the whole 3D-space into two halves, we denote the users lying within the same half-space together with the AP as *reflective* users and the remaining users in the other half-space as *transmissive* users. Without loss of generality, we denote the set of reflective users and transmissive users as the set $\mathcal{K}_{\mathsf{R}}\triangleq\{1,\cdots,K_{\mathsf{R}}\}$ and $\mathcal{K}_{\mathsf{T}}\triangleq\{K_{\mathsf{R}}+1,\cdots,K_{\mathsf{R}}+K_{\mathsf{T}}\}$, respectively, where $K=K_{\mathsf{R}}+K_{\mathsf{T}}$. The transmit signal at the AP can be represented as

$$\mathbf{x}=\sum\nolimits_{k=1}^{K}\mathbf{f}_k s_k, \tag{5}$$

where $\mathbf{f}_k$ and s_k denote the beam vector and the information symbol associated with the k-th user, respectively. Here we assume that all users' information symbols $\{s_k\}$ follow independent circularly symmetric complex Gaussian distribution with zero mean and unit variance, i.e. $\mathbb{E}\{|s_k|^2\}=1, \forall k\in\mathcal{K}$.

Here we denote the wireless channels connecting the AP and the DFA-RIS, the AP and the k-th user and the DFA-RIS and the k-th user as $\mathbf{G}\in\mathbb{C}^{M\times N}$, $\mathbf{h}_{\mathsf{d},k}\in\mathbb{C}^{N}$ and $\mathbf{h}_{\mathsf{r},k}\in\mathbb{C}^{M}$, respectively. To simplify notations, we introduce the category mapping $\imath(k):\mathcal{K}\to\{\mathsf{R},\mathsf{T}\}$ such that $\imath(k)=\mathsf{R}$ if $k\in\mathcal{K}_{\mathsf{R}}$ and $\imath(k)=\mathsf{T}$ if $k\in\mathcal{K}_{\mathsf{T}}$. Note that the wireless link $\mathbf{h}_{\mathsf{r},k}$ originates from exactly one of the two opposite faces of the DAF-RIS, according to the category $\imath(k)$ of user k. In the following, we assume that the channel state information (CSI) is known, which can be obtained via the state-of-the-art channel estimation techniques [36], [37]. The received signal at the k-th user is given as

$$\begin{aligned} y_k &= \mathbf{h}_{\mathsf{d},k}^H\mathbf{x}+\mathbf{h}_{\mathsf{r},k}^H\boldsymbol{\Phi}_{\imath(k)}\mathbf{E}_{\imath(k)}\mathbf{A}\mathbf{G}\mathbf{x}+\mathbf{h}_{\mathsf{r},k}^H\boldsymbol{\Phi}_{\imath(k)}\mathbf{E}_{\imath(k)}\mathbf{A}\mathbf{v}+n_k \\ &= \widetilde{\mathbf{h}}_k^H\mathbf{x}+\mathbf{h}_{\mathsf{r},k}^H\boldsymbol{\Phi}_{\imath(k)}\mathbf{E}_{\imath(k)}\mathbf{A}\mathbf{v}+n_k \\ &= \widetilde{\mathbf{h}}_k^H\mathbf{f}_k s_k+\sum\nolimits_{j\neq k}\widetilde{\mathbf{h}}_k^H\mathbf{f}_j s_j+\mathbf{h}_{\mathsf{r},k}^H\boldsymbol{\Phi}_{\imath(k)}\mathbf{E}_{\imath(k)}\mathbf{A}\mathbf{v}+n_k, \forall k\in\mathcal{K}, \end{aligned} \tag{6}$$

where $\widetilde{\mathbf{h}}_k$ is the effective channel of user k, i.e. $\widetilde{\mathbf{h}}_k\triangleq\mathbf{h}_{\mathsf{d},k}+\mathbf{G}^H\mathbf{A}\mathbf{E}_{\imath(k)}\boldsymbol{\Phi}_{\imath(k)}^*\mathbf{h}_{\mathsf{r},k}$ and n_k is the associated receive noise. It is reasonable to assume that the thermal noise $\mathbf{v}\sim\mathcal{CN}(\mathbf{0},\sigma_v^2\mathbf{I}_M)$ at the DFA-RIS is uncorrelated with all users' information symbols $\{s_k\}$. By defining $\mathbf{G}\triangleq[\widetilde{\mathbf{g}}_1,\cdots,\widetilde{\mathbf{g}}_M]^T$ with $\widetilde{\mathbf{g}}_m^T$ being the m-th row of $\mathbf{G}$, the element-wise power constraint in (3) can be explicitly written as

$$\alpha_m\sum\nolimits_{k=1}^{K}|\widetilde{\mathbf{g}}_m^H\mathbf{f}_k|^2+\alpha_m\sigma_v^2\le\mathsf{P}_m, \quad \forall m\in\mathcal{M}, \tag{7}$$

and the total power constraint in (4) reads

$$\sum\nolimits_{k=1}^{K}\|\mathbf{A}\mathbf{G}\mathbf{f}_k\|_2^2+\sigma_v^2\|\mathbf{A}\|_F^2\le\mathsf{P}_{\mathsf{R}}. \tag{8}$$

Denote $\mathbf{F}\triangleq[\mathbf{f}_1,\cdots,\mathbf{f}_K]$, $\boldsymbol{\alpha}\triangleq[\alpha_1,\cdots,\alpha_M]^T$, $\boldsymbol{\varsigma}\triangleq[\varsigma_1,\cdots,\varsigma_M]^T$ and $\boldsymbol{\phi}_\imath\triangleq[\phi_{\imath,1},\cdots,\phi_{\imath,M}]^T$ with $\imath\in\{\mathsf{R},\mathsf{T}\}$. Then the k-th user's signal to interference and noise ratio (SINR) is given as

$$\mathsf{SINR}_k(\mathbf{F},\boldsymbol{\alpha},\boldsymbol{\varsigma},\boldsymbol{\phi}_{\mathsf{R}},\boldsymbol{\phi}_{\mathsf{T}})=\frac{|\widetilde{\mathbf{h}}_k^H\mathbf{f}_k|^2}{\sum_{j\neq k}|\widetilde{\mathbf{h}}_k^H\mathbf{f}_j|^2+\sigma_v^2\|\mathbf{A}\mathbf{E}_{\imath(k)}\mathbf{h}_{\mathsf{r},k}\|_2^2+\sigma_k^2}, \tag{9}$$

where we have used the identity $\|\boldsymbol{\Phi}_{\imath(k)}\mathbf{A}\mathbf{E}_{\imath(k)}\mathbf{h}_{\mathsf{r},k}\|_2^2=\|\mathbf{A}\mathbf{E}_{\imath(k)}\mathbf{h}_{\mathsf{r},k}\|_2^2$ since $\phi_{\imath(k),m}$ has unit modulus. Therefore, the SE of the above multi-user MISO system is given as

$$\mathsf{R}(\mathbf{F},\boldsymbol{\alpha},\boldsymbol{\varsigma},\boldsymbol{\phi}_{\mathsf{R}},\boldsymbol{\phi}_{\mathsf{T}})\triangleq\sum\nolimits_{k=1}^{K}\log\Big(1+\mathsf{SINR}(\mathbf{F},\boldsymbol{\alpha},\boldsymbol{\varsigma},\boldsymbol{\phi}_{\mathsf{R}},\boldsymbol{\phi}_{\mathsf{T}})\Big). \tag{10}$$

C. Problem Formulation

The joint optimization of AP's beamforming $\mathbf{F}$, the amplifying coefficients $\boldsymbol{\alpha}$, the power-splitting coefficients $\boldsymbol{\varsigma}$ and the DFA-RIS phase-shifters $\boldsymbol{\phi}_{\mathsf{R}}$ and $\boldsymbol{\phi}_{\mathsf{T}}$ to maximize SE can be formulated as the following optimization problem

$$\begin{aligned} (\mathsf{P0}):\max_{\mathbf{F},\boldsymbol{\alpha},\boldsymbol{\varsigma},\boldsymbol{\phi}_{\mathsf{R}},\boldsymbol{\phi}_{\mathsf{T}}}&\sum\nolimits_{k=1}^{K}\log\Big(1+\mathsf{SINR}_k(\mathbf{F},\boldsymbol{\alpha},\boldsymbol{\varsigma},\boldsymbol{\phi}_{\mathsf{R}},\boldsymbol{\phi}_{\mathsf{T}})\Big) && \text{(11a)}\\ \text{s.t. }&\sum\nolimits_{k}\|\mathbf{f}_k\|_2^2\le\mathsf{P}_{\mathsf{AP}}, && \text{(11b)}\\ &\sum\nolimits_{k}\|\mathbf{A}\mathbf{G}\mathbf{f}_k\|_2^2+\sigma_v^2\|\mathbf{A}\|_F^2\le\mathsf{P}_{\mathsf{R}}, && \text{(11c)}\\ &\alpha_m\sum\nolimits_{k=1}^{K}|\widetilde{\mathbf{g}}_m^H\mathbf{f}_k|^2+\alpha_m\sigma_v^2\le\mathsf{P}_m,\ \forall m\in\mathcal{M}, && \text{(11d)}\\ &\varsigma_m\in[0,1],\quad\forall m\in\mathcal{M}, && \text{(11e)}\\ &|\phi_{\mathsf{R},m}|=1,|\phi_{\mathsf{T},m}|=1,\quad\forall m\in\mathcal{M}, && \text{(11f)} \end{aligned}$$

where $\mathsf{P}_{\mathsf{AP}}=\rho\widetilde{\mathsf{P}}_{\mathsf{AP}}$ is effective TRP limit of the AP, with ρ and $\widetilde{\mathsf{P}}_{\mathsf{AP}}$ being the energy conversion coefficient and actual power consumption, respectively, and (11f) assumes that continuous phase-shifters are utilized. The problem $(\mathsf{P0})$ is very challenging due to the highly non-convex objective/constraints and the large number of PEPCs in (11d).

D. Connections with Other Emerging RIS Techniques

In fact, the configuration of DFA-RIS in (P0) is closely related to those of other cutting-the-edge RIS techniques.

i) <u>*Connection with the active-RIS*</u>

By setting $\boldsymbol{\varsigma}=\mathbf{1}$ and $\mathcal{K}=\mathcal{K}_{\mathsf{R}}$ (alternatively, by setting $\boldsymbol{\varsigma}=\mathbf{0}$ and $\mathcal{K}=\mathcal{K}_{\mathsf{T}}$), (P0) reduces to the SE maximization using active-RIS proposed in [7] (without considering the bunch of PEPCs (11d)).

ii) <u>*Connection with the STAR-RIS/IOS*</u>

By setting $\boldsymbol{\alpha}=\mathbf{1}$, $\sigma_v=0$, $\mathsf{P}_{\mathsf{R}}=\infty$ and $\{\mathsf{P}_m=\infty\}$, (P0) reduces to the joint beamforming using STAR-RIS [14] or IOS [18] (still, with the PEPCs (11d) omitted).

To sum up, the SE maximizations associated with active-RIS [7], STAR-RIS [14] and IOS [18] are actually all sub-problems of ours. In the following, we will develop an efficient algorithm to solve (P0), which indeed provides a unifying solution to configure the active-RIS, STAR-RIS or IOS devices towards SE maximization.

$$\begin{aligned}
&\log\Big(1+\mathsf{SINR}_k\big(\mathbf{F},\boldsymbol{\alpha},\boldsymbol{\varsigma},\boldsymbol{\phi}_{\mathsf{R}},\boldsymbol{\phi}_{\mathsf{T}}\big)\Big) &\text{(12a)}\\
=&\log\Big(1+\big|\widetilde{\mathbf{h}}_k^H\mathbf{f}_k\big|^2\Big[\sum\nolimits_{j\neq k}\big|\widetilde{\mathbf{h}}_k^H\mathbf{f}_j\big|^2+\sigma_v^2\big\|\mathbf{A}\mathbf{E}_{\imath(k)}\mathbf{h}_{\mathsf{r},k}\big\|_2^2+\sigma_k^2\Big]^{-1}\Big) &\text{(12b)}\\
=&\max_{w_k\geq 0}\Big\{\log(w_k)-w_k\Big[\big|\widetilde{\mathbf{h}}_k^H\mathbf{f}_k\big|^2\Big(\sum\nolimits_{j\neq k}\big|\widetilde{\mathbf{h}}_k^H\mathbf{f}_j\big|^2+\sigma_v^2\big\|\mathbf{A}\mathbf{E}_{\imath(k)}\mathbf{h}_{\mathsf{r},k}\big\|_2^2+\sigma_k^2\Big)^{-1}+1\Big]^{-1}+1\Big\} &\text{(12c)}\\
=&\max_{w_k\geq 0,\beta_k}\Big\{\underbrace{\log(w_k)+1-w_k+2w_k\mathsf{Re}\big\{\beta_k^*\widetilde{\mathbf{h}}_k^H\mathbf{f}_k\big\}-w_k\big|\beta_k\big|^2\Big(\sum\nolimits_j\big|\widetilde{\mathbf{h}}_k^H\mathbf{f}_j\big|^2\Big)-w_k\big|\beta_k\big|^2\sigma_v^2\big\|\mathbf{A}\mathbf{E}_{\imath(k)}\mathbf{h}_{\mathsf{r},k}\big\|_2^2-w_k\big|\beta_k\big|^2\sigma_k^2}_{\triangleq\widetilde{\mathsf{R}}_k\left(w_k,\beta_k,\mathbf{F},\boldsymbol{\alpha},\boldsymbol{\varsigma},\boldsymbol{\phi}_{\mathsf{R}},\boldsymbol{\phi}_{\mathsf{T}}\right)}\Big\}. &\text{(12d)}
\end{aligned}$$

$$\begin{aligned}
&\widetilde{\mathsf{R}}\big(\mathbf{w},\boldsymbol{\beta},\mathbf{F},\boldsymbol{\alpha},\boldsymbol{\varsigma},\boldsymbol{\phi}_{\mathsf{R}},\boldsymbol{\phi}_{\mathsf{T}}\big)\\
=&\sum\nolimits_k\Big\{-w_k|\beta_k|^2\big(\sum\nolimits_j\big|\widetilde{\mathbf{h}}_k^H\mathbf{f}_j\big|^2\big)-\sigma_v^2w_k|\beta_k|^2\big\|\mathbf{A}\mathbf{E}_{\imath(k)}\mathbf{h}_{\mathsf{r},k}\big\|_2^2+2w_k\mathsf{Re}\{\beta_k^*\widetilde{\mathbf{h}}_k^H\mathbf{f}_k\}-w_k|\beta_k|^2\sigma_k^2+\log w_k-w_k+1\Big\}. \quad \text{(13)}
\end{aligned}$$

III. Algorithm Design

To address (P0), we first adopt the weighted mean square error (WMMSE) method [38] to convert the objective in (11a) into an equivalent variational form presented in (12), with $\mathbf{w}\triangleq[w_1,\cdots,w_K]^T$ and $\boldsymbol{\beta}\triangleq[\beta_1,\cdots,\beta_K]^T$ being the introduced intermediate variables. More details about WMMSE transformation can be found in [38] and are omitted here for brevity.

Utilizing the transformed rate function $\widetilde{\mathsf{R}}_k(\cdot)$ represented in (12d), we define the transformed SE function as $\widetilde{\mathsf{R}}(\,\cdot\,)\triangleq\sum_k\widetilde{\mathsf{R}}_k(\,\cdot\,)$ as shown in (13). The original problem (P0) has been equivalently converted to

$$\begin{aligned}
(\mathsf{P1}):&\max_{\mathbf{w},\boldsymbol{\beta},\mathbf{F},\boldsymbol{\alpha},\boldsymbol{\varsigma},\boldsymbol{\phi}_{\mathsf{R}},\boldsymbol{\phi}_{\mathsf{T}}}\ \widetilde{\mathsf{R}}\big(\mathbf{w},\boldsymbol{\beta},\mathbf{F},\boldsymbol{\alpha},\boldsymbol{\varsigma},\boldsymbol{\phi}_{\mathsf{R}},\boldsymbol{\phi}_{\mathsf{T}}\big) &\text{(14a)}\\
&\text{s.t.}\ \sum\nolimits_k\big\|\mathbf{f}_k\big\|_2^2\leq\mathsf{P}_{\mathsf{AP}}, &\text{(14b)}\\
&\sum\nolimits_k\big\|\mathbf{A}\mathbf{G}\mathbf{f}_k\big\|_2^2+\sigma_v^2\big\|\mathbf{A}\big\|_F^2\leq\mathsf{P}_{\mathsf{R}}, &\text{(14c)}\\
&\alpha_m\sum\nolimits_{k=1}^K\big|\widetilde{\mathbf{g}}_m^H\mathbf{f}_k\big|^2+\alpha_m\sigma_v^2\leq\mathsf{P}_m,\quad\forall m\in\mathcal{M}, &\text{(14d)}\\
&\varsigma_m\in[0,1],\quad\forall m\in\mathcal{M}, &\text{(14e)}\\
&\big|\phi_{\mathsf{R},m}\big|=1,\big|\phi_{\mathsf{T},m}\big|=1,\quad\forall m\in\mathcal{M}. &\text{(14f)}
\end{aligned}$$

In the following, we adopt the block descent ascent (BCA) methodology [39] to tackle (P1). According to the WMMSE method [38], when other variables are fixed, the optimal value of $\mathbf{w}$ and $\boldsymbol{\beta}$ can be determined analytically as follows

$$\beta_k^\star=\Big[\sum\nolimits_j\big|\widetilde{\mathbf{h}}_k^H\mathbf{f}_j\big|^2+\sigma_v^2\big\|\mathbf{A}\mathbf{E}_{\imath(k)}\mathbf{h}_{\mathsf{r},k}\big\|_2^2+\sigma_k^2\Big]^{-1}\big(\widetilde{\mathbf{h}}_k^H\mathbf{f}_k\big), \quad (15)$$

$$w_k^\star=\Big[|\beta_k|^2\big(\sum\nolimits_{j\neq k}\big|\widetilde{\mathbf{h}}_k^H\mathbf{f}_j\big|^2+\sigma_v^2\big\|\mathbf{A}\mathbf{E}_{\imath(k)}\mathbf{h}_{\mathsf{r},k}\big\|_2^2+\sigma_k^2\big)+\big|1-\beta_k^*\widetilde{\mathbf{h}}_k^H\mathbf{f}_k\big|^2\Big]^{-1}.$$

The update of other blocks of variables will be elaborated in the following.

A. Optimizing The Beamformers $\mathbf{F}$

The update of $\mathbf{F}$ is reduced to solving the following problem (omitting irrelevant terms in the objective)

$$\begin{aligned}
(\mathsf{P2}):&\min_{\mathbf{F}}\sum_k\mathbf{f}_k^H\big(\sum_j w_j|\beta_j|^2\widetilde{\mathbf{h}}_j\widetilde{\mathbf{h}}_j^H\big)\mathbf{f}_k-2\sum_k\mathsf{Re}\big\{w_k\beta_k^*\widetilde{\mathbf{h}}_k^H\mathbf{f}_k\big\}\\
&\text{s.t.}\ \sum\nolimits_k\big\|\mathbf{f}_k\big\|_2^2\leq\mathsf{P}_{\mathsf{AP}}, &\text{(16a)}\\
&\sum\nolimits_k\big\|\mathbf{A}\mathbf{G}\mathbf{f}_k\big\|_2^2+\sigma_v^2\|\mathbf{A}\|_F^2\leq\mathsf{P}_{\mathsf{R}}, &\text{(16b)}\\
&\alpha_k\sum\nolimits_k\mathbf{f}_k^H\big(\widetilde{\mathbf{g}}_m\widetilde{\mathbf{g}}_m^H\big)\mathbf{f}_k+\alpha_m\sigma_v^2\leq\mathsf{P}_m,\forall m\in\mathcal{M}. &\text{(16c)}
\end{aligned}$$

To simplify notations, we first introduce the following notations

$$\begin{aligned}
&\mathbf{Q}_0\triangleq\mathbf{I}_K\otimes\big(\sum\nolimits_j w_j|\beta_j|^2\widetilde{\mathbf{h}}_j\widetilde{\mathbf{h}}_j^H\big),\quad\mathbf{Q}_2\triangleq\mathbf{I}_K\otimes\big(\mathbf{G}^H\mathbf{A}^2\mathbf{G}\big), &\text{(17a)}\\
&\mathbf{Q}_1\triangleq\mathbf{I}_{KN},\quad\mathbf{Q}_{m+2}\triangleq\mathbf{I}_k\otimes\big(\alpha_m\widetilde{\mathbf{g}}_m\widetilde{\mathbf{g}}_m^H\big),\forall m\in\mathcal{M}, &\text{(17b)}\\
&\bar{\mathsf{P}}_{m+2}\triangleq\mathsf{P}_m-\alpha_m\sigma_v^2,\bar{\mathsf{P}}_1\triangleq\mathsf{P}_{\mathsf{AP}},\ \bar{\mathsf{P}}_2\triangleq\mathsf{P}_{\mathsf{R}}-\sigma_v^2\big\|\mathbf{A}\big\|_F^2,\forall m, &\text{(17c)}\\
&\mathbf{f}=[\mathbf{f}_1^T,\cdots,\mathbf{f}_K^T]^T,\ \mathbf{q}_0\triangleq[w_1\beta_1\widetilde{\mathbf{h}}_1^T,\cdots,w_K\beta_K\widetilde{\mathbf{h}}_K^T]^T. &\text{(17d)}
\end{aligned}$$

Then the problem (P2) can be equivalently written as

$$\begin{aligned}
(\mathsf{P2}'):&\min_{\mathbf{f}}\ \mathbf{f}^H\mathbf{Q}_0\mathbf{f}-2\mathsf{Re}\big\{\mathbf{q}_0^H\mathbf{f}\big\} &\text{(18a)}\\
&\text{s.t.}\ \mathbf{f}^H\mathbf{Q}_m\mathbf{f}\leq\bar{\mathsf{P}}_m,\ \forall m\in\bar{\mathcal{M}}, &\text{(18b)}
\end{aligned}$$

where $\bar{\mathcal{M}}\triangleq\{1,\cdots,M+2\}$.

Obviously the problem (P2$'$) is a second order cone program (SOCP) and can be optimally solved via off-the-shelf numerical solvers, e.g. CVX [19]. However, since the number of reflecting elements M can be quite large in reality (several hundreds or even thousands), the complexity of solving (P2$'$) directly can be prohibitively high. This difficulty motivates us to explore more efficient solution. By introducing $M{+}2$ copies of $\mathbf{f}$, i.e. $\mathbf{u}_m=\mathbf{f}$, $\forall m\in\bar{\mathcal{M}}$, we decouple the $M{+}2$ constraints of (P2$'$) as follows

$$\begin{aligned}
(\mathsf{P3}):&\min_{\mathbf{f},\{\mathbf{u}_m\}}\ \mathbf{f}^H\mathbf{Q}_0\mathbf{f}-2\mathsf{Re}\big\{\mathbf{q}_0^H\mathbf{f}\big\} &\text{(19a)}\\
&\text{s.t.}\ \mathbf{u}_m^H\mathbf{Q}_m\mathbf{u}_m\leq\bar{\mathsf{P}}_m,\ \forall m\in\bar{\mathcal{M}}, &\text{(19b)}\\
&\mathbf{f}=\mathbf{u}_m,\ \forall m\in\bar{\mathcal{M}}. &\text{(19c)}
\end{aligned}$$

In the following, we adopt the ADMM methodology [40] to solve (P3). By relaxing the equality constraints (19c) and penalizing them in the objective, the augmented Lagrangian (AL) function of (P3) is given as

$$\begin{aligned}
\mathcal{L}\big(\mathbf{f},\mathbf{u},\boldsymbol{\lambda}\big)\triangleq&\ \mathbf{f}^H\mathbf{Q}_0\mathbf{f}-2\mathsf{Re}\big\{\mathbf{q}_0^H\mathbf{f}\big\} &\text{(20)}\\
&+\frac{\rho}{2}\sum\nolimits_{m=1}^{M+2}\big\|\mathbf{u}_m-\mathbf{f}\big\|_2^2+\sum\nolimits_{m=1}^{M+2}\mathsf{Re}\big\{\boldsymbol{\lambda}_m^H\big(\mathbf{f}-\mathbf{u}_m\big)\big\},
\end{aligned}$$

where $\mathbf{u}\triangleq[\mathbf{u}_1^T,\cdots,\mathbf{u}_{M+2}^T]^T$, $\boldsymbol{\lambda}\triangleq[\boldsymbol{\lambda}_1^T,\cdots,\boldsymbol{\lambda}_{M+2}^T]^T$ with each $\boldsymbol{\lambda}_m$ being introduced as the Lagrangian multiplier associated with the m-th constraint in (19c) and ρ is a positive constant. Next, following the ADMM framework [40], we separately update variables $\mathbf{u}$, $\mathbf{f}$ and $\boldsymbol{\lambda}$ in a block coordinate descent (BCD) manner to minimize the AL function $\mathcal{L}\big(\mathbf{f},\mathbf{u},\boldsymbol{\lambda}\big)$, which will be elaborated in the sequel.

With $\mathbf{f}$ and $\boldsymbol{\lambda}$ being fixed, the update of $\mathbf{u}$ towards minimizing $\mathcal{L}(\mathbf{f},\mathbf{u},\boldsymbol{\lambda})$ reduces to solving the following problem

$$(\mathsf{P4}): \min_{\{\mathbf{u}_m\}} \frac{\rho}{2}\sum_{m=1}^{M+2}\|\mathbf{u}_m-\mathbf{f}\|_2^2+\sum_{m=1}^{M+2}\mathsf{Re}\{\boldsymbol{\lambda}_m^H(\mathbf{f}-\mathbf{u}_m)\} \quad (21a)$$

$$\text{s.t. } \mathbf{u}_m^H\mathbf{Q}_m\mathbf{u}_m \le \bar{\mathsf{P}}_m,\ \forall m\in\bar{\mathcal{M}}, \quad (21b)$$

Clearly the above problem breaks down to $M+2$ independent small problems with the m-th small problem given as

$$(\mathsf{P4}_m): \min_{\mathbf{u}_m} \|\mathbf{u}_m\|_2^2-2\mathsf{Re}\{(\rho^{-1}\boldsymbol{\lambda}_m+\mathbf{f})^H\mathbf{u}_m\} \quad (22a)$$

$$\text{s.t. } \mathbf{u}_m^H\mathbf{Q}_m\mathbf{u}_m \le \bar{\mathsf{P}}_m, \quad (22b)$$

The above problem $(\mathsf{P4}_m)$ can be efficiently solved via the following lemma, which is proved in Appendix A.

Lemma 1. *Consider the following problem* $(\mathsf{P5})$*:*

$$(\mathsf{P5}): \min_{\mathbf{u}} \|\mathbf{u}\|_2^2-2\mathsf{Re}\{\mathbf{q}^H\mathbf{u}\} \quad (23a)$$

$$\text{s.t. } \mathbf{u}^H\mathbf{Q}\mathbf{u}\le \mathsf{P}_0, \quad (23b)$$

where the $n\times n$ matrix $\mathbf{Q}\succcurlyeq 0$. Assume that $\mathbf{Q}$ is of rank r and has eigenvalue decomposition $\mathbf{Q}=\mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^H$ with its eigenvalues $\mathsf{diag}(\boldsymbol{\Lambda})=[\lambda_1,\cdots,\lambda_n]$ being arranged in a descending order, i.e. $\lambda_1\ge\cdots\ge\lambda_n$ and define $\mathbf{p}\triangleq\mathbf{U}^H\mathbf{q}$ with p_i being the i-th entry of $\mathbf{p}$. Then the optimal solution $\mathbf{u}^\star$ to $(\mathsf{P5})$ can be determined in either of the two following cases:

- *CASE-I : if $\mathbf{q}^H\mathbf{Q}\mathbf{q}\le\mathsf{P}_0$, then $\mathbf{u}^\star=\mathbf{q}$;*
- *CASE-II : otherwise, $\mathbf{u}^\star=(\mathbf{I}+\mu^\star\mathbf{Q})^{-1}\mathbf{q}$ with the positive $\mu^\star$ being the unique solution to the following equation*

$$\sum_{i=1}^{r}\lambda_i|p_i|^2(1+\mu\lambda_i)^{-2}=\mathsf{P}_0. \quad (24)$$

Besides, the value of $\mu^\star$ can be bounded in the following

$$0<\lambda_1^{-1}\left(\sqrt{\sum_{i=1}^{r}\lambda_i|p_i|^2/\mathsf{P}_0}-1\right)<\mu^\star\le\sqrt{\mathsf{P}_0^{-1}\left(\sum_{i=1}^{r}|p_i|^2/\lambda_i\right)}. \quad (25)$$

Note that the value of $\mu^\star$ in CASE-II can be numerically determined very efficiently (e.g. by bisection search or Newton's method) within the range identified in (25). Via invoking Lemma 1, we can efficiently solve all the $M+2$ small problems $(\mathsf{P4}_m)$ in a parallel manner. Next, we investigate the update of $\mathbf{f}$, which is obtained via minimizing $\mathcal{L}(\mathbf{f},\mathbf{u},\boldsymbol{\lambda})$. With $\mathbf{f}$ and $\mathbf{u}$ being fixed, this sub-problem is reduced to

$$(\mathsf{P6}):\min_{\mathbf{f}}\mathbf{f}^H\mathbf{Q}_0\mathbf{f}+\frac{\rho}{2}\sum_m\|\mathbf{u}_m-\mathbf{f}\|_2^2-2\mathsf{Re}\{(\mathbf{q}_0-\sum_m\boldsymbol{\lambda}_m)^H\mathbf{f}\} \quad (26)$$

Notice (P6) is a non-constrained convex quadratic problem. Therefore its optimal solution can be readily obtained as

$$\mathbf{f}^\star=\left(\mathbf{Q}_0+\frac{\rho}{2}(M+2)\mathbf{I}_{KN}\right)^{-1}\left(\mathbf{q}_0+\frac{\rho}{2}\sum_{m=1}^{M+2}\mathbf{u}_m-\frac{1}{2}\sum_{m=1}^{M+2}\boldsymbol{\lambda}_m\right). \quad (27)$$

The update of Lagrangian dual variables $\boldsymbol{\lambda}$ follows a gradient ascent manner [40], which is given by

$$\boldsymbol{\lambda}_m^{(t+1)} := \boldsymbol{\lambda}_m^{(t)}+\rho(\mathbf{f}-\mathbf{u}_m),\ \forall m\in\bar{\mathcal{M}}. \quad (28)$$

Algorithm 1: ADMM to Solve (P2)

1 **Init.**: Generate feasible $\mathbf{F}^{(0)}$, $t:=0$;
2 **repeat**
3 Update $\mathbf{u}_m^{(t+1)}$ by solving $(\mathsf{P4}_m)$ `//In Parallel`;
4 Update $\mathbf{f}^{(t+1)}$ by (27);
5 Update $\boldsymbol{\lambda}^{(t+1)}$ by (28); $t++$;
6 **until** *convergence*;

Theorem 1. *Assume that the sequence of $\{\mathbf{u}^{(t)},\mathbf{f}^{(t)},\boldsymbol{\lambda}^{(t)}\}$ is generated by Algorithm 1. Then any limit point of $\{\mathbf{u}^{(t)},\mathbf{f}^{(t)},\boldsymbol{\lambda}^{(t)}\}$ is an optimal solution to* (P2).

Proof. This conclusion is indeed a direct implication of [41, Sec. 3.4, Prop. 4.2]. In fact, the variables $\mathbf{u}$ and $\mathbf{f}$ are connected by the following identity

$$\mathbf{u}^T=[\mathbf{u}_1^T,\cdots,\mathbf{u}_{M+2}^T]=\mathbf{f}^T[\mathbf{I}_{KN},\cdots,\mathbf{I}_{KN}]\triangleq\mathbf{f}^T\mathbf{J}^T. \quad (29)$$

Notice the fact that $\mathbf{J}^T\mathbf{J}=(M+2)\mathbf{I}_{KN}$ is of full rank. Therefore Prop. 4.2 of [41] can be invoked and the convergence result in the assertion can be obtained. More details are referred to [41]. □

B. Optimizing The Power Amplifier Coefficients $\boldsymbol{\alpha}$

We proceed to study the optimization of power amplifying coefficients $\boldsymbol{\alpha}$. Firstly we introduce the following notations

$$|\widetilde{\mathbf{h}}_k^H\mathbf{f}_j|^2=|\mathbf{f}_j^H\mathbf{h}_{\mathsf{d},k}+\mathbf{f}_j^H\mathbf{G}^H\boldsymbol{\Phi}_{\imath(k)}^*\mathbf{E}_{\imath(k)}\mathsf{Diag}(\mathbf{h}_{\mathsf{r},k})\sqrt{\boldsymbol{\alpha}}|^2 \quad (30a)$$

$$=|\xi_{k,j}+\boldsymbol{\eta}_{k,j}^H\sqrt{\boldsymbol{\alpha}}|^2, \quad (30b)$$

$$\|\mathbf{A}\mathbf{E}_{\imath(k)}\mathbf{h}_{\mathsf{r},k}\|_2^2=\sqrt{\boldsymbol{\alpha}}^H\mathsf{Diag}(|\boldsymbol{\kappa}_k|^2)\sqrt{\boldsymbol{\alpha}}, \quad (30c)$$

$$\mathsf{Re}\{\widetilde{\mathbf{h}}_{\mathsf{r},k}^H\mathbf{f}_k\}=\mathsf{Re}\{\boldsymbol{\tau}_k^H\sqrt{\boldsymbol{\alpha}}\}+\widetilde{c}_1, \quad (30d)$$

where $\sqrt{\cdot}$, $|\cdot|$ and $(\cdot)^2$ are all element-wise operations, $\widetilde{c}_1$ is a constant independent of $\boldsymbol{\alpha}$ and the newly introduced coefficients above are defined as

$$\xi_{k,j}\triangleq\mathbf{f}_j^H\mathbf{h}_{\mathsf{d},k},\boldsymbol{\eta}_{k,j}\triangleq\mathsf{Diag}(\mathbf{h}_{\mathsf{r},k}^*)\boldsymbol{\Phi}_{\imath(k)}\mathbf{E}_{\imath(k)}\mathbf{G}\mathbf{f}_j,\forall k,j\in\mathcal{K},$$
$$\boldsymbol{\kappa}_k\triangleq\mathbf{E}_{\imath(k)}\mathbf{h}_{\mathsf{r},k},\boldsymbol{\tau}_k\triangleq\mathsf{Diag}(\mathbf{h}_{\mathsf{r},k}^*)\boldsymbol{\Phi}_{\imath(k)}\mathbf{E}_{\imath(k)}\mathbf{G}\mathbf{f}_k,\forall k\in\mathcal{K}, \quad (31)$$

Following the above notations, we rearrange the objective $\widetilde{\mathsf{R}}(\cdot)$ into a function with respect to (w.r.t.) $\boldsymbol{\alpha}$ explicitly, which is given as

$$\begin{aligned}&\widetilde{\mathsf{R}}(\boldsymbol{\alpha}|\mathbf{w},\boldsymbol{\beta},\mathbf{F},\boldsymbol{\varsigma},\boldsymbol{\phi}_{\mathsf{R}},\boldsymbol{\phi}_{\mathsf{T}})=\\&\sqrt{\boldsymbol{\alpha}}^H\Big\{\sum_k\big[w_k|\beta_k|^2\big(\sigma_v^2\mathsf{Diag}(|\boldsymbol{\kappa}_k|^2)+\sum_j\boldsymbol{\eta}_{k,j}\boldsymbol{\eta}_{k,j}^H\big)\big]\Big\}\sqrt{\boldsymbol{\alpha}}\\&+2\mathsf{Re}\Big\{\sum_k\big[w_k|\beta_k|^2\sum_j\xi_{k,j}\boldsymbol{\eta}_{k,j}-w_k\beta_k^*\boldsymbol{\tau}_k\big]^H\sqrt{\boldsymbol{\alpha}}\Big\}+\widetilde{c}_2,\\&\triangleq\sqrt{\boldsymbol{\alpha}}^H\mathbf{P}_0\sqrt{\boldsymbol{\alpha}}+2\mathsf{Re}\{\mathbf{p}^H\sqrt{\boldsymbol{\alpha}}\}+\widetilde{c}_2.\end{aligned} \quad (32)$$

where $\mathbf{P}_0$ and $\mathbf{p}$ can be determined accordingly and $\widetilde{c}_2$ is a constant irrelevant of $\boldsymbol{\alpha}$.

At the same time, we introduce the following notations

$$\mathbf{G}\mathbf{F}=\mathbf{G}[\mathbf{f}_1,\cdots,\mathbf{f}_K]\triangleq[\widetilde{\mathbf{f}}_1,\cdots,\widetilde{\mathbf{f}}_M]^T, \quad (33)$$

i.e., $\widetilde{\mathbf{f}}_m^T$ is the m-th row of the matrix $\mathbf{GF}$. Then the total transmit power can be written as

$$\begin{aligned}&\sum_k\|\mathbf{A}\mathbf{G}\mathbf{f}_k\|_2^2+\sigma_v^2\|\mathbf{A}\|_F^2=\sum_m(\alpha_m\|\widetilde{\mathbf{f}}_m\|_2^2+\sigma_v^2\alpha_m) \quad (34)\\&=\sqrt{\boldsymbol{\alpha}}^H\Big[\mathsf{Diag}\big([\|\widetilde{\mathbf{f}}_1\|_2^2,\cdots,\|\widetilde{\mathbf{f}}_M\|_2^2]\big)+\sigma_v^2\mathbf{I}_M\Big]\sqrt{\boldsymbol{\alpha}}\triangleq\sqrt{\boldsymbol{\alpha}}\mathbf{P}_1\sqrt{\boldsymbol{\alpha}}.\end{aligned}$$

Similarly, the PEPC is given as

$$\alpha_m\left(\|\widetilde{\mathbf{f}}_m\|_2^2+\sigma_v^2\right) \leq \mathsf{P}_m, \quad \forall m \in \mathcal{M}. \tag{35}$$

Therefore, the amplifier coefficient optimization problem is given as follows:

$$(\mathsf{P7}) : \min_{\boldsymbol{\alpha}} \ \sqrt{\boldsymbol{\alpha}}^H \mathbf{P}_0 \sqrt{\boldsymbol{\alpha}} + 2\mathsf{Re}\{\mathbf{p}^H \sqrt{\boldsymbol{\alpha}}\}, \tag{36a}$$

$$\text{s.t. } \sqrt{\boldsymbol{\alpha}}^H \mathbf{P}_1 \sqrt{\boldsymbol{\alpha}} \leq \mathsf{P}_\mathsf{R}, \tag{36b}$$

$$\alpha_m\left(\|\widetilde{\mathbf{f}}_m\|_2^2+\sigma_v^2\right) \leq \mathsf{P}_m, \forall m \in \mathcal{M}, \tag{36c}$$

$$\boldsymbol{\alpha} \geq 0. \tag{36d}$$

The problem (P7) is still an SOCP, whose complexity is even higher than $(\mathsf{P2}')$. To develop low complexity solution, we consult to the majorization-minimization (MM) methodology [42], which largely coincides with the block successive upper-bound minimization (BSUM) framework in [43], [44]. We first rewrite the objective in (P7) via expanding all vectors into entries as follows

$$\begin{aligned}&\sqrt{\boldsymbol{\alpha}}^H \mathbf{P}_0 \sqrt{\boldsymbol{\alpha}} + 2\mathsf{Re}\{\mathbf{p}^H \sqrt{\boldsymbol{\alpha}}\} \\ =&\sum_{m=1}^M a_m\alpha_m + \sum_{m\neq n} b_{m,n}\sqrt{\alpha_m}\sqrt{\alpha_n} + \sum_{m=1}^M c_m\sqrt{\alpha_m},\end{aligned} \tag{37}$$

where the coefficients a_m's, $b_{m,n}$'s and c_m's can be determined accordingly. To apply MM methodology, we first introduce the following useful inequalities, which are proved in Appendix B.

Lemma 2. *Assume that α_m, α_n, $\alpha_{m,0}$ and $\alpha_{n,0}$ are arbitrary positive values. Then the following inequalities hold*

$$\sqrt{\alpha_m}\sqrt{\alpha_n} \leq 1/2\left(\sqrt{\alpha_{n,0}/\alpha_{m,0}}\alpha_m + \sqrt{\alpha_{m,0}/\alpha_{n,0}}\alpha_n\right); \tag{38a}$$

$$\sqrt{\alpha_m}\sqrt{\alpha_n} \geq 1/2\sqrt{\alpha_{m,0}}\sqrt{\alpha_{n,0}}\left(\log\alpha_m + \log\alpha_n\right) + \widetilde{c}_3; \tag{38b}$$

$$\sqrt{\alpha_m} \leq 1/2\sqrt{\alpha_{m,0}^{-1}}\alpha_m + 1/2\alpha_{m,0}; \tag{38c}$$

$$\sqrt{\alpha_m} \geq 1/2\sqrt{\alpha_{m,0}}\log\alpha_m - 1/2\sqrt{\alpha_{m,0}}\log\alpha_{m,0} + \sqrt{\alpha_{m,0}}, \tag{38d}$$

where

$$\widetilde{c}_3 \triangleq -1/2\sqrt{\alpha_{m,0}\alpha_{n,0}}\left(\log\alpha_{m,0} + \log\alpha_{n,0}\right) + \sqrt{\alpha_{m,0}\alpha_{n,0}}. \tag{39}$$

Inspired by the MM methodology, we can obtain an upper-bound of the objective in (37) by exploiting Lemma 2. Define the $\alpha_{m,0}$ and $\alpha_{n,0}$ as the latest value of α_m and α_n obtained in the last iteration, respectively. The cross terms $b_{m,n}\sqrt{\alpha_m}\sqrt{\alpha_n}$ can be upper-bounded in two ways according to the sign of its coefficient $b_{m,n}$, as specified in the following.

Case i): if $b_{m,n} \geq 0$, invoking (38a), we have

$$b_{m,n}\sqrt{\alpha_m\alpha_n} \leq b_{m,n}/2\left(\sqrt{\alpha_{n,0}/\alpha_{m,0}}\alpha_m + \sqrt{\alpha_{m,0}/\alpha_{n,0}}\alpha_n\right). \tag{40}$$

Case ii): if $b_{m,n} < 0$, by (38b), we obtain

$$b_{m,n}\sqrt{\alpha_m\alpha_n} \leq b_{m,n}/2\sqrt{\alpha_{m,0}\alpha_{n,0}}\left(\log\alpha_m + \log\alpha_n\right) + \widetilde{c}_6, \tag{41}$$

with $\widetilde{c}_6$ being some constant.

Similarly, for the square root term $c_m\sqrt{\alpha}_m$, we can upper-bound it in two cases:

Case i): if $c_m \geq 0$, according to (38c), we have

$$c_m\sqrt{\alpha}_m \leq c_m/2\sqrt{\alpha_{m,0}^{-1}}\alpha_m + \widetilde{c}_7; \tag{42}$$

Case ii): if $c_m < 0$, utilizing (38d), we obtain

$$c_m\sqrt{\alpha_m} \leq c_m/2\sqrt{\alpha_{m,0}}\log\alpha_m + \widetilde{c}_8. \tag{43}$$

with $\widetilde{c}_7$ and $\widetilde{c}_8$ being constants.

Via replacing all the addends $b_{m,n}\sqrt{\alpha_m}\sqrt{\alpha_n}$ and $c_m\sqrt{\alpha_m}$ in (37) by their corresponding upper-bounds developed in (40)-(43) and rearranging terms, we obtain the an upper-bound of (37) as follows

$$\sum_{m=1}^M \bar{a}_m\alpha_m - \sum_{m=1}^M \bar{b}_m\log\alpha_m, \tag{44}$$

where the coefficients $\bar{a}_m$'s and $\bar{b}_m$'s can be accordingly determined. Obviously $\bar{a}_m > 0$ and $\bar{b}_m \geq 0$, $\forall m \in \mathcal{M}$. In fact, for a specific m, if the upper-bounding in (41) or (43) has not been invoked, then the term $\bar{b}_m\log\alpha_m$ will not appear, i.e. $\bar{b}_m = 0$. Otherwise $\bar{b}_m > 0$. Note that $\mathbf{P}_1$ is diagonal as defined in (34). Therefore, by defining $\bar{c}_m \triangleq \|\widetilde{\mathbf{f}}_m\|_2^2+\sigma_v^2$, the total power constraint (36b) can be equivalently written as (45b). Besides, define $\bar{\mathsf{P}}_m \triangleq \mathsf{P}_m/(\|\widetilde{\mathbf{f}}_m\|_2^2+\sigma_v^2)$. Then adopting the MM methodology [42]–[44], instead of solving (36a) directly, we turn to optimize its upper-bound in (44), which reduces to the following problem

$$(\mathsf{P8}) : \min_{\boldsymbol{\alpha}} \ \sum_{m=1}^M \bar{a}_m\alpha_m - \sum_{m=1}^M \bar{b}_m\log\alpha_m \tag{45a}$$

$$\text{s.t. } \sum_{m=1}^M \bar{c}_m\alpha_m \leq \mathsf{P}_\mathsf{R}, \tag{45b}$$

$$0 \leq \alpha_m \leq \bar{\mathsf{P}}_m, \quad \forall m \in \mathcal{M}. \tag{45c}$$

Fortunately, the problem (P8) has an analytic solution that is identified via the following theorem (proved in Appendix C).

Theorem 2. *Define $\breve{\alpha}_m = \min\left\{\bar{\mathsf{P}}_m, \frac{\bar{b}_m}{\bar{a}_m}\right\}$, $\forall m \in \mathcal{M}$. The optimal solution $\boldsymbol{\alpha}^\star$ to (P8) can be determined in exactly one of the following two possible cases:*

- *CASE-I : if $\sum_m \bar{c}_m\breve{\alpha}_m \leq \mathsf{P}_\mathsf{R}$, then $\alpha_m^\star = \breve{\alpha}_m$, $\forall m \in \mathcal{M}$.*
- *CASE-II : if $\sum_m \bar{c}_m\breve{\alpha}_m > \mathsf{P}_\mathsf{R}$, then $\boldsymbol{\alpha}^\star$ is given as*

$$\alpha_m^\star = \left[\left(\bar{c}_m\nu^\star + \bar{a}_m\right)^{-1}\bar{b}_m\right]^{\bar{\mathsf{P}}_m}, \ \forall m \in \mathcal{M}, \tag{46}$$

where $[x]^{\bar{\mathsf{P}}_m} \triangleq \min\{x, \bar{\mathsf{P}}_m\}$ and $\nu^\star$ is the unique solution to the equality

$$\sum_m \bar{c}_m\alpha_m^\star = \mathsf{P}_\mathsf{R}. \tag{47}$$

Remark III.1. *Theorem 2 reminds us of the classical water-filling theorem, which identifies the optimal power allocation towards channel capacity maximization under a sum-power constraint [45]. Compared to the standard water-filling result [45], (P8) is much more challenging due to the more complicated objective (both linear and logarithm terms) and the element-wise power constraints.*

Note that the value of $\nu^\star$ in CASE-II of Theorem 2 should be identified numerically (e.g. by bisection search). Therefore one last remaining issue is that an upper-bound of $\nu^\star$ in CASE-II is still missing. The following lemma answers this question, which is proved in Appendix D.

Algorithm 2: MM Method to Solve (P7)

Init.: Given $\boldsymbol{\alpha}^{(0)}$ and ϵ_0; $t := 0$;
repeat
  Update $\{\bar{a}_m, \bar{b}_m, \bar{c}_m\}$ by (40)-(44);
  Update $\boldsymbol{\alpha}^{(t+1)}$ by Thm.2; $t++$;
until $\frac{\mathsf{obj}(\boldsymbol{\alpha}^{(t)}) - \mathsf{obj}(\boldsymbol{\alpha}^{(t+1)})}{|\mathsf{obj}(\boldsymbol{\alpha}^{t})|} \le \epsilon_0$;

Lemma 3. *Besides the definition of $\{\breve{\alpha}_m\}$ introduced in Theorem 2, we further define $\{\widetilde{\alpha}_m\}$ as any strictly feasible solution to (*P8*) and* $\mathsf{obj}(\boldsymbol{\alpha})$ *as the objective value of (*P8*) yielded by $\boldsymbol{\alpha}$. Then an upper bound of $\nu^\star$ in CASE-II of Theorem 2 is given by*

$$\nu^\star \le \frac{\mathsf{obj}(\breve{\boldsymbol{\alpha}}) - \mathsf{obj}(\widetilde{\boldsymbol{\alpha}})}{\mathsf{P}_\mathsf{R} - \sum_m \bar{c}_m \widetilde{\alpha}_m}. \tag{48}$$

The iterative method to solve (P7) based on MM method is summarized in Alg.2. Besides, the following theorem reveals the convergence of Alg.2 and its connection with (P7).

Theorem 3. *The iterate $\{\boldsymbol{\alpha}^{(t)}\}$ generated by Alg.2 yields a monotonic decreasing objective value of (*P7*). Besides, any limit point of $\{\boldsymbol{\alpha}^{(t)}\}$ is an optimal solution of (*P7*).*

Proof. Recall that Alg.2 iteratively updates $\boldsymbol{\alpha}$ via solving (P8) repeatedly, whose objective is constructed via the upper-bounds in (40)-(43) based on the latest iterate of $\boldsymbol{\alpha}^{(t)}$. In fact, the upper-bounding procedure (38) complies with conditions (A1)-(A4) in [44] (verifications are omitted due to space limit). Hence Alg.2 falls in the MM framework [42]–[44]. According to [43, Thm.1], $\{f(\boldsymbol{\alpha}^{(t)})\}$ monotonically decreases and converges to the optimum value. Besides, any limit point of $\{\boldsymbol{\alpha}^{(t)}\}$ is stationary, i.e. optimal solution to (P7). □

C. Optimizing The Power Splitting Coefficients $\boldsymbol{\varsigma}$

Next we discuss the update of the power-splitting factors $\boldsymbol{\varsigma}$. As before, we rewrite the objective $\widetilde{\mathsf{R}}(\boldsymbol{\varsigma}|\mathbf{w}, \boldsymbol{\beta}, \mathbf{F}, \boldsymbol{\alpha}, \boldsymbol{\phi}_\mathsf{R}, \boldsymbol{\phi}_\mathsf{T})$ in a form w.r.t. $\boldsymbol{\varsigma}$ explicitly. Denote

$$\begin{aligned} &\big|\widetilde{\mathbf{h}}_k^H \mathbf{f}_j\big|^2 = \big|\mathbf{h}_{\mathsf{d},k}^H \mathbf{f}_j + \mathbf{h}_{\mathsf{r},k}^H \boldsymbol{\Phi}_{\imath(k)} \mathbf{E}_{\imath(k)} \mathbf{A}\mathbf{G}\mathbf{f}_j\big|^2 \\ &= \big|\xi_{k,j} + \mathbf{f}_j^H \mathbf{G}^H \mathbf{A} \boldsymbol{\Phi}_{\imath(k)}^* \mathsf{Diag}(\mathbf{h}_{\mathsf{r},k}) \mathbf{e}_{\imath(k)}\big|^2 = \big|\xi_{k,j} + \boldsymbol{\chi}_{k,j}^H \mathbf{e}_{\imath(k)}\big|^2, \end{aligned} \tag{49}$$

where the newly introduced coefficients are defined as

$$\boldsymbol{\chi}_{k,j} \triangleq \mathsf{Diag}(\mathbf{h}_{\mathsf{r},k}^*) \boldsymbol{\Phi}_{\imath(k)} \mathbf{A}\mathbf{G}\mathbf{f}_j, \ \forall j, k \in \mathcal{K}, \tag{50a}$$

$$\mathsf{diag}(\mathbf{E}_{\imath(k)}) \triangleq \mathbf{e}_{\imath(k)} = \begin{cases} \boldsymbol{\varsigma}, & \text{if } k \in \mathcal{K}_\mathsf{R} \\ \sqrt{\mathbf{1} - \boldsymbol{\varsigma}^2}, & \text{if } k \in \mathcal{K}_\mathsf{T} \end{cases}. \tag{50b}$$

Based on (50), we further rewrite the following terms

$$\sum_{k,j} w_k |\beta_k|^2 \big|\widetilde{\mathbf{h}}_k^H \mathbf{f}_j\big|^2 = \sum_{\imath \in \{\mathsf{R},\mathsf{T}\}} \Big\{\mathbf{e}_\imath^H \mathbf{Q}_\imath \mathbf{e}_\imath + 2\mathsf{Re}\{\mathbf{q}_\imath^H \mathbf{e}_\imath\}\Big\}, \tag{51a}$$

$$\begin{aligned} &-2\sum_k \mathsf{Re}\big\{w_k \beta_k^* \big(\mathbf{h}_{\mathsf{d},k}^H \mathbf{f}_k + \mathbf{h}_{\mathsf{r},k}^H \boldsymbol{\Phi}_{\imath(k)} \mathbf{E}_{\imath(k)} \mathbf{A}\mathbf{G}\mathbf{f}_k\big)\big\} \\ &= -2\sum_k \mathsf{Re}\big\{w_k \beta_k^* \mathbf{f}_k^H \mathbf{G}^H \mathbf{A} \boldsymbol{\Phi}_{\imath(k)}^* \mathsf{Diag}\big(\mathbf{h}_{\mathsf{r},k}\big) \mathbf{e}_{\imath(k)}\big\} + \widetilde{c}_9 \\ &= -2\sum_{\imath \in \{\mathsf{R},\mathsf{T}\}} \mathsf{Re}\big\{\mathbf{d}_\imath^H \mathbf{e}_\imath\big\} + \widetilde{c}_9, \end{aligned} \tag{51b}$$

$$\sum_k \sigma_v^2 w_k |\beta_k|^2 \big\|\mathbf{E}_{\imath(k)} \mathbf{A} \mathbf{h}_{\mathsf{r},k}\big\|_2^2 = \sum_{\imath \in \{\mathsf{R},\mathsf{T}\}} \mathbf{e}_\imath^H \mathbf{D}_\imath \mathbf{e}_\imath, \tag{51c}$$

with the newly introduced coefficients above defined as

$$\mathbf{Q}_\imath \triangleq \sum_{k \in \mathcal{K}_\imath} w_k |\beta_k|^2 \Big(\sum_j \boldsymbol{\chi}_{k,j} \boldsymbol{\chi}_{k,j}^H\Big), \ \imath \in \{\mathsf{R},\mathsf{T}\}, \tag{52a}$$

$$\mathbf{q}_\imath \triangleq \sum_{k \in \mathcal{K}_\imath} w_k |\beta_k|^2 \Big(\sum_j \xi_{k,j} \boldsymbol{\chi}_{k,j}^H\Big), \ \imath \in \{\mathsf{R},\mathsf{T}\}, \tag{52b}$$

$$\mathbf{d}_\imath \triangleq \sum_{k \in \mathcal{K}_\imath} w_k \beta_k^* \mathsf{Diag}(\mathbf{h}_{\mathsf{r},k}^*) \boldsymbol{\Phi}_\imath \mathbf{A}\mathbf{G}\mathbf{f}_k, \imath \in \{\mathsf{R},\mathsf{T}\}, \tag{52c}$$

$$\mathbf{D}_\imath \triangleq \sum_{k \in \mathcal{K}_\imath} \sigma_v^2 w_k |\beta_k|^2 \mathsf{Diag}\big(\big|\mathbf{A}\mathbf{h}_{\mathsf{r},k}\big|^2\big), \imath \in \{\mathsf{R},\mathsf{T}\}. \tag{52d}$$

where $\widetilde{c}_9$ is a constant. Substituting (51) and (50) into (13), we obtain

$$\widetilde{\mathsf{R}}(\boldsymbol{\varsigma}|\mathbf{w}, \boldsymbol{\beta}, \mathbf{F}, \boldsymbol{\alpha}, \boldsymbol{\phi}_\mathsf{R}, \boldsymbol{\phi}_\mathsf{T}) = \sum_{\imath \in \{\mathsf{R},\mathsf{T}\}} \Big\{\mathbf{e}_\imath^H \mathbf{S}_\imath \mathbf{e}_\imath + 2\mathsf{Re}\{\mathbf{s}_\imath^H\} \mathbf{e}_\imath\Big\}, \tag{53}$$

with the coefficients in the above equation defined as

$$\mathbf{S}_\imath \triangleq \mathbf{Q}_\imath + \mathbf{D}_\imath, \ \mathbf{s}_\imath \triangleq \mathbf{q}_\imath - \mathbf{d}_\imath, \imath \in \{\mathsf{R},\mathsf{T}\}. \tag{54}$$

Therefore the update of $\boldsymbol{\varsigma}$ is meant to solve the problem

$$\begin{aligned} (\mathsf{P10}): \min_{\boldsymbol{\varsigma}} \ & \boldsymbol{\varsigma}^H \mathbf{S}_\mathsf{R} \boldsymbol{\varsigma} + \sqrt{\mathbf{1} - \boldsymbol{\varsigma}^2}^H \mathbf{S}_\mathsf{T} \sqrt{\mathbf{1} - \boldsymbol{\varsigma}^2} + 2\mathsf{Re}\big\{\mathbf{s}_\mathsf{R}^H \boldsymbol{\varsigma} + \mathbf{s}_\mathsf{T}^H \sqrt{\mathbf{1} - \boldsymbol{\varsigma}^2}\big\} \\ \mathsf{s.t.} \ & 0 \le \varsigma_m \le 1, \ \forall m \in \mathcal{M}. \end{aligned} \tag{55}$$

The difficulty to solve (P10) lies in the terms containing $\mathbf{e}_\mathsf{T} = \sqrt{\mathbf{1} - \boldsymbol{\varsigma}^2}$. Here we still resort to the MM method. Firstly, for $\forall \imath \in \{\mathsf{R},\mathsf{T}\}$ we have

$$\begin{aligned} &\mathbf{e}_\imath^H \mathbf{S}_\imath \mathbf{e}_\imath \\ &= \big(\mathbf{e}_\imath - \mathbf{e}_{\imath,0}\big)^H \mathbf{S}_\imath \big(\mathbf{e}_\imath - \mathbf{e}_{\imath,0}\big) + 2\mathsf{Re}\big\{\mathbf{e}_{\imath,0}^H \mathbf{S}_\imath \big(\mathbf{e}_\imath - \mathbf{e}_{\imath,0}\big)\big\} + \mathbf{e}_{\imath,0}^H \mathbf{S}_\imath \mathbf{e}_{\imath,0} \\ &\le \lambda_{\max}\big(\mathbf{S}_\imath\big) \big\|\mathbf{e}_\imath - \mathbf{e}_{\imath,0}\big\|_2^2 + 2\mathsf{Re}\big\{\mathbf{e}_{\imath,0}^H \mathbf{S}_\imath \big(\mathbf{e}_\imath - \mathbf{e}_{\imath,0}\big)\big\} + \mathbf{e}_{\imath,0}^H \mathbf{S}_\imath \mathbf{e}_{\imath,0} \\ &= \lambda_{\max}\big(\mathbf{S}_\imath\big) \big\|\mathbf{e}_\imath\big\|_2^2 + 2\mathsf{Re}\big\{\big[\mathbf{S}_\imath \mathbf{e}_{\imath,0} - \lambda_{\max}\big(\mathbf{S}_\imath\big) \mathbf{e}_{\imath,0}\big]^H \mathbf{e}_\imath\big\} + \widetilde{c}_{10}, \end{aligned} \tag{56}$$

where $\lambda_{\max}(\mathbf{S}_\imath)$ denotes the maximal eigenvalue of $\mathbf{S}_\imath$, $\mathbf{e}_{\mathsf{R},0} = \boldsymbol{\varsigma}_0$ and $\mathbf{e}_{\mathsf{T},0} = \sqrt{1 - \boldsymbol{\varsigma}_0^2}$ with $\boldsymbol{\varsigma}_0$ being the latest value of $\boldsymbol{\varsigma}$. Therefore, replacing the two quadratic terms in (53) by the upper-bound developed in (56), we obtain

$$\begin{aligned} &\widetilde{\mathsf{R}}(\boldsymbol{\varsigma}\big|\mathbf{w}, \boldsymbol{\beta}, \mathbf{F}, \boldsymbol{\alpha}, \boldsymbol{\phi}_\mathsf{R}, \boldsymbol{\phi}_\mathsf{T}) \\ &\le \lambda_\mathsf{R} \|\mathbf{e}_\mathsf{R}\|_2^2 + \lambda_\mathsf{T} \|\mathbf{e}_\mathsf{T}\|_2^2 + 2\mathsf{Re}\{\mathbf{u}_\mathsf{R}^H \mathbf{e}_\mathsf{R}\} + 2\mathsf{Re}\{\mathbf{u}_\mathsf{T}^H \mathbf{e}_\mathsf{T}\} + \widetilde{c}_{11} \\ &= \sum_{m=1}^M \Big(\lambda_\mathsf{R} \varsigma_m^2 + \lambda_\mathsf{T} (1 - \varsigma_m^2) + 2\mathsf{Re}\big\{u_{\mathsf{R},m} \varsigma_m\big\} + 2\mathsf{Re}\big\{u_{\mathsf{T},m} \sqrt{1 - \varsigma_m^2}\big\}\Big). \end{aligned} \tag{57}$$

with the coefficients in the above being defined as

$$\lambda_\imath \triangleq \lambda_{\max}\big(\mathbf{S}_\imath\big), \ \mathbf{u}_\imath \triangleq \big(\mathbf{S}_\imath - \lambda_\imath \mathbf{I}_M\big) \mathbf{e}_{\imath,0} + \mathbf{s}_\imath, \ \imath \in \{\mathsf{R},\mathsf{T}\}. \tag{58}$$

To further convexify the objective, we linearize the concave term $-\lambda_\mathsf{T} \varsigma_m^2$ at the point of $\varsigma_{m,0}$ to obtain a tight upper-bound

$$-\lambda_\mathsf{T} \varsigma_m^2 \le -\lambda_\mathsf{T} \big[\varsigma_{m,0}^2 + 2\varsigma_{m,0}(\varsigma_m - \varsigma_{m,0})\big]. \tag{59}$$

Finally, for the last troublesome term $2\mathsf{Re}\{u_{\mathsf{T},m}\}\sqrt{1 - \varsigma_m^2}$ in the objective, we consider the following two possible cases

Case i): $\mathsf{Re}\{u_{\mathsf{T},m}\} < 0$. The term $2\mathsf{Re}\{u_{\mathsf{T},m}\}\sqrt{1 - \varsigma_m^2}$ is indeed convex. Hence we leave it unchanged.

Case ii): $\mathsf{Re}\{u_{\mathsf{T},m}\} \ge 0$. At this time, $2\mathsf{Re}\{u_{\mathsf{T},m}\}\sqrt{1 - \varsigma_m^2}$ is concave. To convexify it, we again linearize it at the point of $\varsigma_{m,0}$ to yield the following upper-bound

$$\begin{aligned} &2\mathsf{Re}\{u_{\mathsf{T},m}\}\sqrt{1 - \varsigma_m^2} \\ &\le 2\mathsf{Re}\{u_{\mathsf{T},m}\}\Big(\sqrt{1 - \varsigma_{m,0}^2} - \varsigma_{m,0}\sqrt{1 - \varsigma_{m,0}^2}^{-1}\big(\varsigma_m - \varsigma_{m,0}\big)\Big) \\ &= -2\mathsf{Re}\{u_{\mathsf{T},m}\}\varsigma_{m,0}\sqrt{1 - \varsigma_{m,0}^2}^{-1}\varsigma_m + \widetilde{c}_{12}. \end{aligned} \tag{60}$$

This article has been accepted for publication in IEEE Transactions on Signal Processing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TSP.2022.3221345

By applying (59) and (60) to (57), we obtain the following convex upper-bound

$$\widetilde{\mathsf{R}}(\boldsymbol{\varsigma}|\mathbf{w},\boldsymbol{\varsigma},\mathbf{F},\boldsymbol{\alpha},\boldsymbol{\phi}_{\mathsf{R}},\boldsymbol{\phi}_{\mathsf{T}}) \leq \sum\nolimits_m \left(\widehat{a}_m\varsigma_m^2+\widehat{b}_m\varsigma_m-\widehat{c}_m\sqrt{1-\varsigma_m^2}\right), \quad (61)$$

where the coefficients $\{\widehat{a}_m\}$, $\{\widehat{b}_m\}$ and $\{\widehat{c}_m\}$ can be determined accordingly. Obviously $\widehat{a}_m > 0$ and $\widehat{c}_m \geq 0$, $\forall m \in \mathcal{M}$. In fact, if Case ii) does not occur, the upper-bounding in (60) will not be conducted and therefore $\widehat{c}_m = 0$. Otherwise $\widehat{c}_m > 0$. Following the MM methodology, to deal with the non-convex (P10), we turn to optimize its convex upper-bound (61), which implies we need solve the problem

$$(\mathsf{P}11): \min_{\boldsymbol{\varsigma}} \ \sum\nolimits_m \left(\widehat{a}_m\varsigma_m^2+\widehat{b}_m\varsigma_m-\widehat{c}_m\sqrt{1-\varsigma_m^2}\right) \quad (62a)$$

$$\text{s.t. } 0 \leq \varsigma_m \leq 1, \ \forall m \in \mathcal{M}. \quad (62b)$$

Obviously, the problem (P11) naturally splits into M independent small problems with each one being given as

$$(\mathsf{P}11_m): \min_{0\leq\varsigma_m\leq 1} g_m(\varsigma_m) \triangleq \widehat{a}_m\varsigma_m^2+\widehat{b}_m\varsigma_m-\widehat{c}_m\sqrt{1-\varsigma_m^2}, \quad (63)$$

Fortunately, $(\mathsf{P}11_m)$ can be analytically solved as follows.

CASE-I: $\widehat{b}_m \geq 0$. Obviously, $g_m(\varsigma)$ is increasing in ς_m and therefore $\varsigma_m^\star = 0$.

CASE-II: $\widehat{b}_m < 0$. In this case, two possible sub-cases could occur.

Case i) If $\widehat{c}_m = 0$, then $\varsigma_m^\star = \min\left\{-\frac{\widehat{b}_m}{2\widehat{a}_m}, 1\right\}$.

Case ii) If $\widehat{c}_m > 0$, the derivative of $g_m(\varsigma_m)$ is given as

$$g_m'(\varsigma_m) = \varsigma_m\left(2\widehat{a}_m + \frac{\widehat{c}_m}{\sqrt{1-\varsigma_m^2}}\right) + \widehat{b}_m. \quad (64)$$

It can be easily verified that $g_m(\varsigma_m)$ first increases and then decreases when ς_m varies from 0 to 1. Therefore, the optimal $\varsigma_m^\star$ occurs somewhere within $(0,1)$ and can be identified via setting the derivative $g_m'(\varsigma_m)=0$, i.e.

$$\varsigma_m\left(2\widehat{a}_m + \frac{\widehat{c}_m}{\sqrt{1-\varsigma_m^2}}\right) = -\widehat{b}_m, \quad (65)$$

Note that the above equation has unique solution since the function on the left hand side increases monotonically from 0 to $+\infty$ when ς_m varies from 0 to 1.

Remark III.2. *The significance of the upper-bounding procedure in (56) lies in two folds: i) it eliminates the square-root components in the quadratic terms, which paves the way to further simplification (e.g. (57)); ii) more importantly, it removes all the cross terms containing two different power splitting coefficients and consequently decouples the objective, which makes the parallel computing possible in solving (*P11*).*

D. Optimizing the Phase Shifts $\boldsymbol{\phi}_{\mathsf{R}}$ and $\boldsymbol{\phi}_{\mathsf{T}}$

In this subsection, we consider the update of phase-shifters $\boldsymbol{\phi}_{\mathsf{R}}$ and $\boldsymbol{\phi}_{\mathsf{T}}$. Firstly, we introduce the following notations

$$\boldsymbol{\mu}_{k,j} \triangleq \mathsf{Diag}(\mathbf{h}_{\mathsf{r},k}^*)\mathbf{E}_{\imath(k)}\mathbf{A}\mathbf{G}\mathbf{f}_j, \ \forall k,j \in \mathcal{K}, \quad (66a)$$

$$\boldsymbol{\xi}_k \triangleq w_k\beta_k^*\mathsf{Diag}(\mathbf{h}_{\mathsf{r},k}^*)\mathbf{E}_{\imath(k)}\mathbf{A}\mathbf{G}\mathbf{f}_k, \forall k \in \mathcal{K}, \quad (66b)$$

Algorithm 3: SE Maximization using DFA-RIS

1: Initialize $\mathbf{F}^{(0)}$, $\boldsymbol{\alpha}^{(0)}$, $\boldsymbol{\varsigma}^{(0)}$, $\boldsymbol{\phi}_{\mathsf{R}}^{(0)}$ and $\boldsymbol{\phi}_{\mathsf{T}}^{(0)}$; $k=0$;
2: **repeat**
3: Update $\mathbf{F}^{(k+1)}$ via Alg.1;
4: Update $\boldsymbol{\alpha}^{(k+1)}$ by Alg.2;
5: Update $\boldsymbol{\varsigma}^{(k+1)}$ by solving $(\mathsf{P}11_m)$,$\forall m$;
6: Update $\boldsymbol{\phi}_{\mathsf{R}}^{(k+1)}$ and $\boldsymbol{\phi}_{\mathsf{T}}^{(k+1)}$ by (73);
7: Update $\boldsymbol{\beta}$ and $\mathbf{w}$ by (15); $k++$;
8: **until** *Convergence*

and simplify the expressions in the sequel

$$\left|\widetilde{\mathbf{h}}_k^H\mathbf{f}_j\right|^2 = \left|\xi_{k,j}+\mathbf{h}_{\mathsf{r},k}^H\boldsymbol{\Phi}_{\imath(k)}\mathbf{E}_{\imath(k)}\mathbf{A}\mathbf{G}\mathbf{f}_j\right|^2 \quad (67a)$$

$$=\left|\xi_{k,j}+\mathbf{f}_j^H\mathbf{G}^H\mathbf{A}\mathbf{E}_{\imath(k)}\mathsf{Diag}(\mathbf{h}_{\mathsf{r},k})\boldsymbol{\phi}_{\imath(k)}^*\right|^2 = \left|\xi_{k,j}+\boldsymbol{\mu}_{k,j}^H\boldsymbol{\phi}_{\imath(k)}\right|^2,$$

$$\mathsf{Re}\{w_k\beta_k\mathbf{f}_k^H\mathbf{G}^H\mathbf{A}\mathbf{E}_\imath\boldsymbol{\Phi}_{\imath(k)}^*\mathbf{h}_{\mathsf{r},k}\} \quad (67b)$$

$$=\mathsf{Re}\{w_k\beta_k\mathbf{f}_j^H\mathbf{G}^H\mathbf{A}\mathbf{E}_{\imath(k)}\mathsf{Diag}(\mathbf{h}_{\mathsf{r},k})\boldsymbol{\phi}_{\imath(k)}^*\} = \mathsf{Re}\{\boldsymbol{\xi}_k^H\boldsymbol{\phi}_{\imath(k)}^*\}.$$

Then $\mathsf{R}(\boldsymbol{\phi}_{\mathsf{R}},\boldsymbol{\phi}_{\mathsf{T}}|\mathbf{w},\boldsymbol{\beta},\mathbf{F},\boldsymbol{\alpha},\boldsymbol{\varsigma})$ can be rewritten concisely as

$$\mathsf{R}(\boldsymbol{\phi}_{\mathsf{R}},\boldsymbol{\phi}_{\mathsf{T}}|\mathbf{w},\boldsymbol{\beta},\mathbf{F},\boldsymbol{\alpha},\boldsymbol{\varsigma}) = \sum\nolimits_{\imath\in\{\mathsf{R},\mathsf{T}\}}\{\boldsymbol{\phi}_\imath^H\mathbf{P}_\imath\boldsymbol{\phi}_\imath+2\mathsf{Re}\{\mathbf{p}_\imath^H\boldsymbol{\phi}_\imath\}\}, \quad (68)$$

with the coefficients in the above being defined as

$$\mathbf{P}_\imath \triangleq \sum\nolimits_{k\in\mathcal{K}_\imath} w_k|\beta_k|^2\left(\sum\nolimits_{j\in\mathcal{K}}\boldsymbol{\mu}_{k,j}^*\boldsymbol{\mu}_{k,j}^T\right), \ \imath \in \{\mathsf{R},\mathsf{T}\}, \quad (69a)$$

$$\mathbf{p}_\imath \triangleq \sum\nolimits_{k\in\mathcal{K}_\imath} w_k|\beta_k|^2\left(\sum\nolimits_{j\in\mathcal{K}}\xi_{k,j}^*\mu_{k,j}^*\right) - \sum\nolimits_{k\in\mathcal{K}_\imath}\boldsymbol{\xi}_k^*, \forall\imath. \quad (69b)$$

Therefore, the update of phase-shifters is meant to solve

$$(\mathsf{P}12): \min_{\boldsymbol{\phi}_{\mathsf{R}},\boldsymbol{\phi}_{\mathsf{T}}} \sum\nolimits_{\imath\in\{\mathsf{R},\mathsf{T}\}}\left\{\boldsymbol{\phi}_\imath^H\mathbf{P}_\imath\boldsymbol{\phi}_\imath+2\mathsf{Re}\{\mathbf{p}_\imath^H\boldsymbol{\phi}_\imath\}\right\} \quad (70a)$$

$$\text{s.t. } |\phi_{\imath,m}| = 1, \ \imath \in \{\mathsf{R},\mathsf{T}\}, \ \forall m \in \mathcal{M}. \quad (70b)$$

The problem (P12) is non-convex due to (70b). We can still leverage MM method to overcome this difficulty. Assume that $\boldsymbol{\phi}_{\imath,0}$ is the latest value of $\boldsymbol{\phi}_\imath$. Following similar arguments in (56)

$$\begin{aligned}&\boldsymbol{\phi}_\imath^H\mathbf{P}_\imath\boldsymbol{\phi}_\imath\\ &\leq \lambda_{\max}(\mathbf{P}_\imath)\|\boldsymbol{\phi}_\imath-\boldsymbol{\phi}_{\imath,0}\|_2^2+2\mathsf{Re}\{\boldsymbol{\phi}_{\imath,0}^H\mathbf{P}_\imath(\boldsymbol{\phi}_\imath-\boldsymbol{\phi}_{\imath,0})\}+\boldsymbol{\phi}_{\imath,0}^H\mathbf{P}_\imath\boldsymbol{\phi}_{\imath,0}\\ &\overset{(a)}{=} 2\mathsf{Re}\{(\mathbf{P}_\imath\boldsymbol{\phi}_{\imath,0}-\lambda_{\max}(\mathbf{P}_\imath)\boldsymbol{\phi}_{\imath,0})^H\boldsymbol{\phi}_\imath\}+\widetilde{c}_{13}, \ \imath\{\mathsf{R},\mathsf{T}\}, \end{aligned} \quad (71)$$

where (a) utilizes the fact that $\|\boldsymbol{\phi}_\imath\|_2^2 = M$ due to (70b) and hence the quadratic terms goes into the constant term $\widetilde{c}_{13}$.

Replacing the original objective of (P12) with the upper-bound (71) and omitting constants, we turn to solve the following problem

$$(\mathsf{P}13): \min_{\boldsymbol{\phi}_{\mathsf{R}},\boldsymbol{\phi}_{\mathsf{T}}} \sum\nolimits_{\imath\in\{\mathsf{R},\mathsf{T}\}}\left\{2\mathsf{Re}\{(\mathbf{p}_\imath+\mathbf{P}_\imath\boldsymbol{\phi}_{\imath,0}-\lambda_{\max}(\mathbf{P}_\imath)\boldsymbol{\phi}_{\imath,0})^H\boldsymbol{\phi}_\imath\}\right\}$$

$$\text{s.t. } |\phi_{\imath,m}| = 1, \ \imath \in \{\mathsf{R},\mathsf{T}\}, \ \forall m \in \mathcal{M}. \quad (72)$$

The optimal solution to (P13) can be immediately obtained in a closed form given by

$$\boldsymbol{\phi}_\imath^\star = \exp\left(\jmath\angle\left(-[\mathbf{p}_\imath+\mathbf{P}_\imath\boldsymbol{\phi}_{\imath,0}-\lambda_{\max}(\mathbf{P}_\imath)\boldsymbol{\phi}_{\imath,0}]\right)\right), \imath\in\{\mathsf{R},\mathsf{T}\}. \quad (73)$$

The overall procedure to iteratively update different blocks of variables are summarized in Alg.3.

TABLE I: MATLAB Run Time to Solve (P2′) (in Sec.)

Alg.	M=200	M=400	M=800	M=1200	M=2000
SOCP	7.247	9.220	18.69	28.51	61.25
ADMM (per Iter.)	0.00085	0.00096	0.0011	0.0014	0.0020

IV. Discussions

A. Convergence

The following theorem (proved in Appendix E) characterizes the convergence behaviour of Alg.3.

Theorem 4. *The sequence* $\{\mathsf{R}(\mathbf{F}^{(t)}, \boldsymbol{\alpha}^{(t)}, \boldsymbol{\varsigma}^{(t)}, \boldsymbol{\phi}_{\mathsf{R}}^{(t)}, \boldsymbol{\phi}_{\mathsf{T}}^{(t)})\}$ *yielded by Alg.3 monotonically increases. Besides, any limit point of the solution iterates yielded by Alg.3 is stationary point of* (P0)*.*

B. Complexity

In this subsection, we briefly discuss the complexity of our solution. The complexity of solving (P2′) vis SOCP solver is $\mathcal{O}(M^{1.5}K^3N^3)$. Comparatively, the complexity of our proposed ADMM algorithm is $\mathcal{O}(C_1K^3N^3)$ with C_1 being the number of ADMM iterations, which usually takes a value from 600 to 1000 (see Sec.V). To update $\boldsymbol{\alpha}$, the complexity of solving (P7) is $\mathcal{O}(M^{3.5})$. In comparison, the closed-form solution by Thm.2 has negligible complexity. The major complexity of Alg.2 comes from the computation of the coefficients $\{\bar{a}_m, \bar{b}_m, \bar{c}_m\}$, which is $\mathcal{O}(M^2)$. Therefore, the overall complexity of Alg.2 is $\mathcal{O}(C_2M^2)$ with C_2 being the number of iterations, which generally takes a value no larger than 300 (see Sec.V). The updates of $\boldsymbol{\varsigma}$, $\boldsymbol{\phi}_{\mathsf{R}}$ and $\boldsymbol{\phi}_{\mathsf{T}}$ are all analytic based and therefore the associated complexity is negligible.

V. Numerical Results

In this section, we present numerical results to verify our proposals. In our experiment, the AP with $N = 6$ antennas is 120m apart from the DFA-RIS. The $N_{\mathsf{K_R}} = 3$ reflective users and $N_{\mathsf{K_T}} = 3$ transmissive users are located on a circle centered at the DFA-RIS with radius of R. Assume that the AP-RIS channel is Rician fading. The AP-user and RIS-user links follow Rayleigh fading. The fading exponents for the links of AP-user, AP-RIS and RIS-user are assumed to be -3.7, -2.8 and -3.1, respectively. The noise levels at the mobile users σ_k^2 and the RIS σ_v^2 are all set as -100dBm. The AP's TX power $\mathsf{P_{AP}}$ is usually fixed at 20dBm. $\mathsf{P_R}$ is set as 50% of $\mathsf{P_{AP}}$ in most cases. The PEPC limit is set as $\mathsf{P}_m = 2$mW, $\forall m \in \mathcal{M}$.

Firstly, Fig.2 and Fig.3 verify the convergence behaviours of our proposed ADMM based method to update $\mathbf{F}$, i.e. (P2′). According to our tests, setting the penalty coefficient ρ in the range of $[0.2, 1]$ generally leads to satisfactory convergence rate. Take $M = 1000$ as an example, the ADMM solution usually converges within 600 to 1000 iterations. The equality constraint $\mathbf{u}_m = \mathbf{f}$ are satisfied very well (with violation lower than 10^{-5}) within 800 iterations and the objective value generally achieves a precision of 10^{-3} within 1000 iterations.

Besides, in Table I we compare the time complexity of the SOCP and ADMM method when solving (P2′). We enumerate the Matlab run time for various numbers of reflecting elements (i.e. M) in Table I. The SOCP method is conducted by CVX. Combining the results in Fig.3 and Table I, we can conclude that when M is large (e.g. $M \geq 800$), ADMM method can achieve an efficiency 20 times higher than that of the CVX!

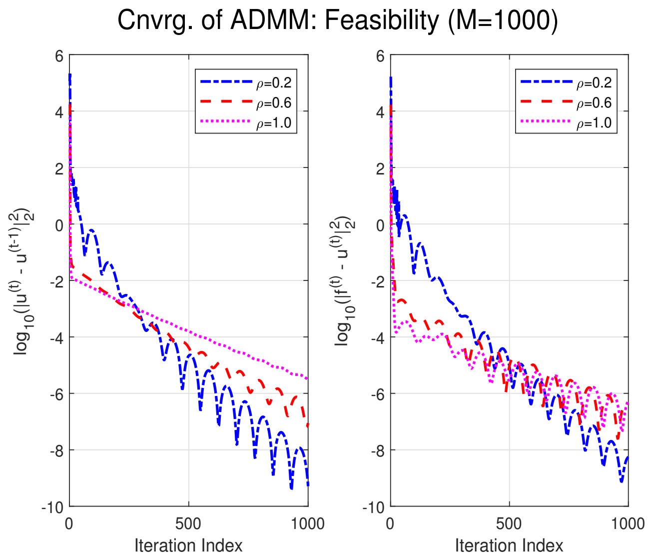


Fig. 2: Convergence of ADMM method: feasibility

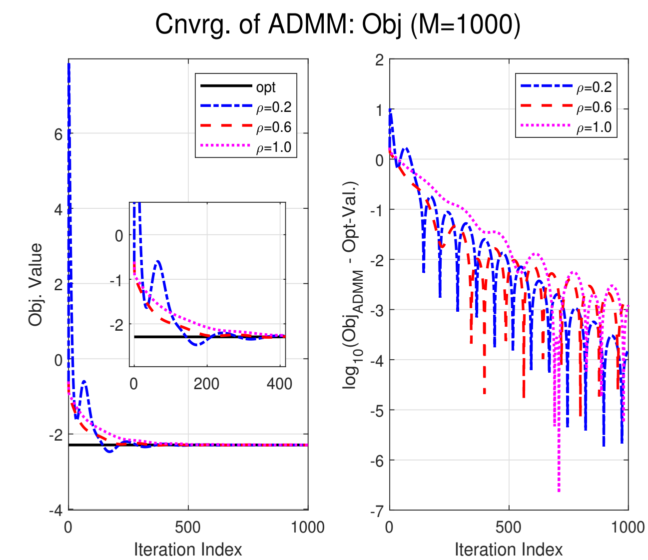


Fig. 3: Convergence of ADMM method: objective

Secondly, Fig.4 and Fig.5 test our proposed MM-based Alg.2 to update $\boldsymbol{\alpha}$ (i.e. (P7)). Alg.2 is invoked to iteratively solve (P7) for several random channels in Fig.4, where the true optimum values of (P7) obtained by CVX are also plotted (after normalization) for benchmark. The relative precisions of the objective iterates are plotted in the right half of Fig.4. We see that Alg.2 generally achieves precise objective values (within 1% loss) after 500 iterations. Fig.5 investigates the impact of the terminating threshold ϵ_0 in Alg.2 to the main solution Alg.3. As a benchmark, we also present the SOCP-based solution, which solves (P7) by CVX with other

This article has been accepted for publication in IEEE Transactions on Signal Processing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TSP.2022.3221345

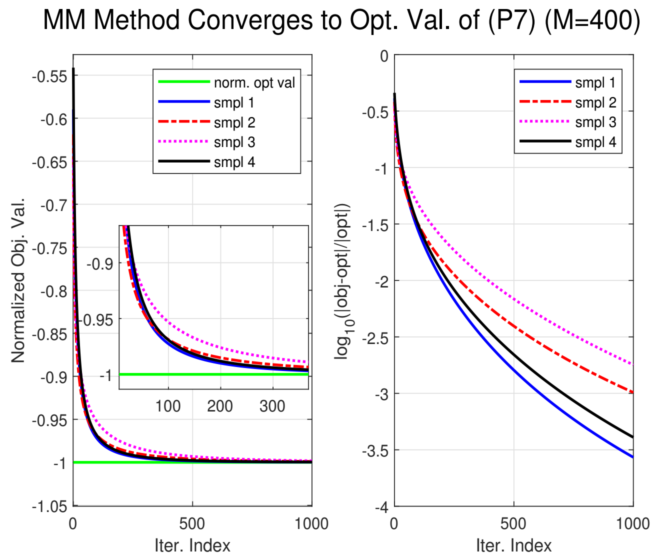


Fig. 4: Optimizing $\boldsymbol{\alpha}$: MM achieves the optimal solution

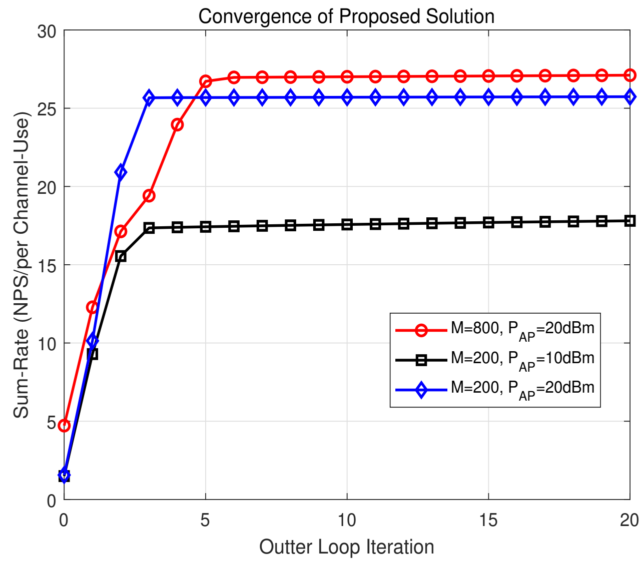


Fig. 6: Convergence of Alg.3

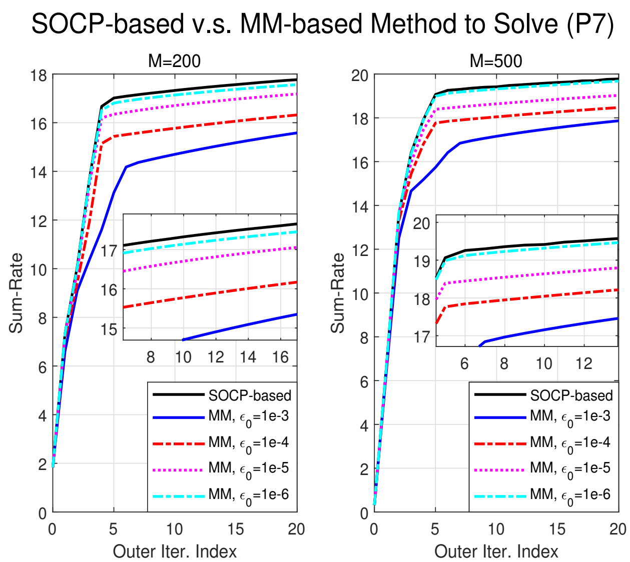


Fig. 5: Optimizing $\boldsymbol{\alpha}$: MM method v.s. SOCP

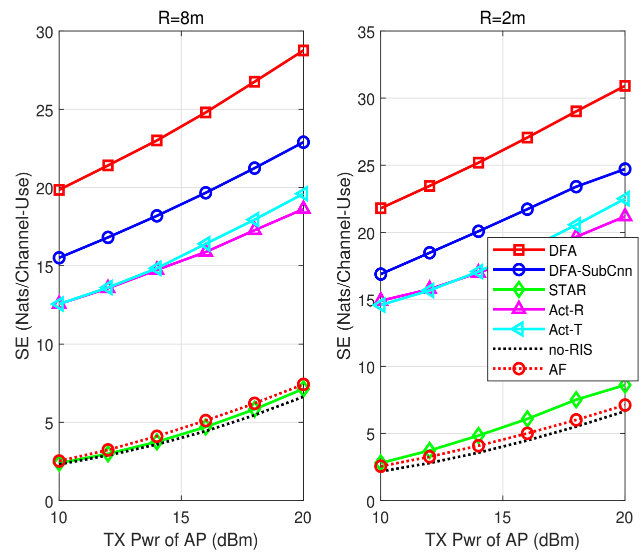


Fig. 7: SE of different RIS architectures

TABLE II: MATLAB Run Time to Update $\boldsymbol{\alpha}$ (in Sec.)

	M=200	M=400	M=800	M=1200	M=2000
SOCP (P7)	2.298	11.78	54.69	205.8	583.5
Thm.2 (P8)	0.0018	0.0050	0.0191	0.0416	0.1223
Num. (P8)	188.5	160.4	185.2	193.5	227.2
Avg. Time	0.3393	0.802	3.537	8.050	27.78

operations being the same. According to Fig.5, when ϵ_0 is chosen at 10^{-6}, our MM-based solution yields almost identical performance (within 1% loss) with that of the SOCP-based solution.

Next, Table II examines the time complexity of our Alg.2. For different M, the average Matlab run-time solving (P8) and the average number of times solving (P8) for each invoking of Alg.2 are shown in the 2nd and 3rd rows of Table II, respectively. The average time to optimize $\boldsymbol{\alpha}$ can hence be determined accordingly and presented in the 4-th row. In our test, the threshold ϵ_0 is set as 10^{-6}. The Matlab run-time for solving (P7) using CVX is presented in the 1st row. Clearly, when DFA-RIS is modestly large ($M \geq 400$), our proposed MM-based solution can achieve nearly identical performance with the CVX but with over 15 times higher efficiency!

Fig.6 illustrates the convergence behaviours of our proposed main solution Alg.3. As can be seen, under system settings, the objective value keeps increasing and usually obtains significant improvement during the first 5 iterations.

In Fig.7, we compare the DFA-RIS with other RIS architectures, i.e. the single-faced active-RIS [7] and the STAR-RIS [12] (equivalently the IOS [15]). We also extend sub-connected architecture [9] to our DFA-RIS, where every 10 element pairs are driven by one amplifier with power limit of 20mW and one power splitter. In the experiment, we fix $M = 200$ and test two scenarios with $\mathrm{R} = 2\mathrm{m}$ and $\mathrm{R} = 8\mathrm{m}$. The single-faced active-RIS faces to either the reflective users or the transmissive users. These two cases are labelled as act-RIS(R) and act-RIS(T), respectively. The DFA-RIS, sub-

This article has been accepted for publication in IEEE Transactions on Signal Processing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TSP.2022.3221345

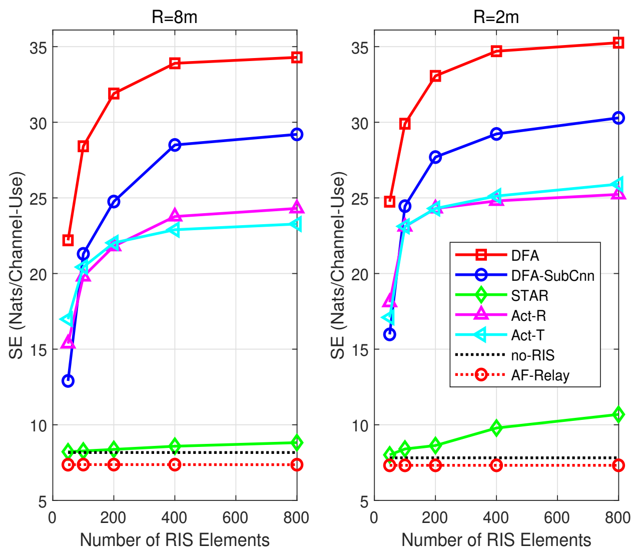


Fig. 8: The impact of number of RIS elements

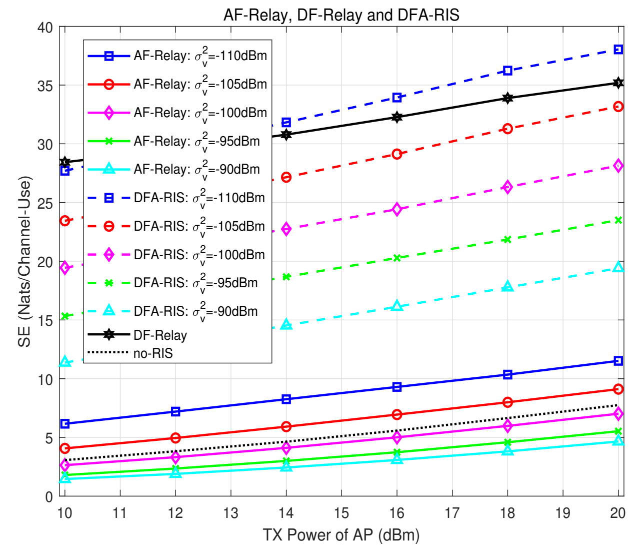


Fig. 9: Impact of receiving noise level (σ_v^2) at the RIS/relay

connected DFA and active-RIS all have identical total TX power limit. For fair comparison, the AP associated with STAR-RIS/no-RIS schemes has a TX power equal to the sum of those active-type RIS and their cooperating APs. We also test the half-duplex amplify-and-forward (AF) relay, which has 6 antennas, identical TX power and σ_v^2 as those of DFA-RIS. As shown in Fig.7, DFA-RIS obviously outperforms all other competitors. The sub-connected DFA-RIS achieves rather competing performance with much lower number of amplifiers and power dividers. The two singled-face active-RIS schemes are sub-optimal due to their half-space coverage limitation. Interestingly, when $R = 8$m, the STAR-RIS has negligible performance gain over the no-RIS, which is caused by the double-fading effect. The AF-relay yields negligible performance gain since its limited beamforming gain cannot effectively counteract its local receiving noise. The active AF-relay is even outperformed by the STAR-RIS when R = 2m.

In Fig.8, both P_{AP} and P_{R} are fixed at 20dBm and the number of RIS element (pairs) varies. The block-size of sub-connected DFA-RIS is still set as 10. As can be seen, when M increases from 50 to 200, the SE improvement is significant. When M is larger than 400, the SE improvement with growing M becomes marginal. Provided sufficiently close to the users, STAR-RIS yields an observable gain over the no-RIS case.

Fig.9 reflects the impact of the local receiving noise (σ_v^2) at the RIS/relay. The half-duplex decode-and-forward (DF) relay is also tested. The DF-relay first decodes the incoming signals perfectly and then retransmits. Both AF- and DF-relays have 6 antennas, while the DFA-RIS has $M = 100$ element pairs. All schemes all have identical TX powers. As shown in Fig.9, when σ_v^2 is high, the AF-relay's performance is even worse than no-RIS scheme. In fact, when σ_v^2 is large, the optimized beamformers of AF-relay tends to shut down its TX power to avoid introducing too much noise. Furthermore, since AF-relay is half-duplex, its performance is rather poor. In comparison, DFA-RIS performs much better due to its much stronger beamforming gain. Note that the DF-relay, although performing well, is indeed not practical since perfect decoding is generally too ideal to implement.

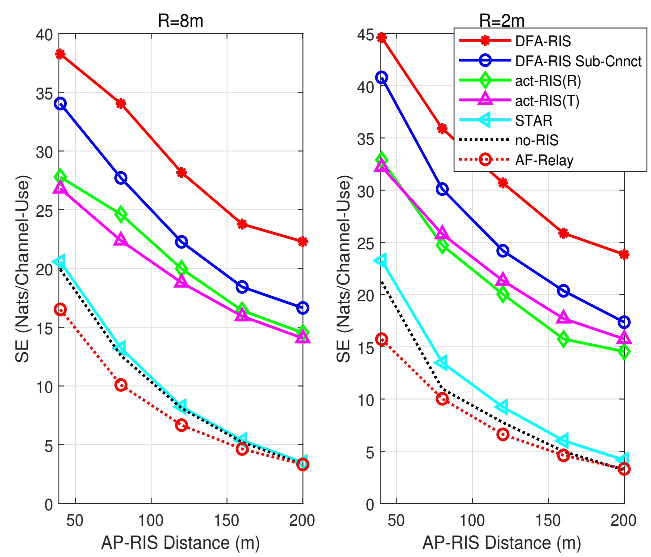


Fig. 10: Impact of AP-RIS distance

Finally, Fig.10 examines the impact of the AP-RIS distance. As before, the DFA-RIS still performs better than all other competing schemes.

VI. Conclusion

This paper proposes the novel DFA-RIS architecture which can effectively combat the double-fading effect and also obtains 360° full-space coverage. The associated optimization of DFA-RIS configuration, which involves PEPCs, subsumes configurations the of many other RIS architectures as special cases and is very challenging. We have developed an analytic-based, highly computation-parallelizable and convergence guaranteed solution to resolve this problem. Extensive numerical results have verified the efficiency of our solution and the DFA-RIS' advantageous performance.

Appendix

A. Proof of Lemma 1

Proof. Obviously Slater's condition holds for (P5). Therefore its optimal solution can be obtained via solving the Karush-Kuhn-Tucker (KKT) conditions, which are given as follows

$$\mathbf{u} \triangleq \left(\mathbf{I}+\mu\mathbf{Q}\right)^{-1}\mathbf{q},\ \mu \geq 0, \mu\left(\mathbf{u}^H\mathbf{Q}\mathbf{u}-\mathsf{P}_0\right)=0, \mathbf{u}^H\mathbf{Q}\mathbf{u} \leq \mathsf{P}_0.$$

When $\mu = 0$, then $\mathbf{u} = \mathbf{q}$. At this time, KKT conditions hold if and only if the constraint $\mathbf{u}^H\mathbf{Q}\mathbf{x} \leq \mathsf{P}_0$ is satisfied. If this is true, then $\mathbf{u} = \mathbf{q}$ is the optimal solution. Otherwise, μ cannot be zero and therefore should be positive, which immediately implies that the power constraint (23b) is active. Therefore, by substituting $\mathbf{u} = \left(\mathbf{I}+\mu\mathbf{Q}\right)^{-1}\mathbf{q}$ into $\mathbf{u}^H\mathbf{Q}\mathbf{u} = \mathsf{P}_0$ and utilizing the eigenvalue value decomposition of $\mathbf{Q}$, we can obtain that

$$\mathbf{p}^H\left(\mathbf{I}+\mu\mathbf{\Lambda}\right)^{-2}\mathbf{\Lambda}\mathbf{p} = \mathsf{P}_0, \tag{74}$$

which is equivalent to

$$\sum_{i=1}^{r}\frac{\lambda_i|p_i|^2}{(1+\mu\lambda_i)^2} = \mathsf{P}_0. \tag{75}$$

Note that above equation has a solution since $\sum_{i=1}^{r}\lambda_i|p_i|^2 > \mathsf{P}_0$ (i.e. $\mathbf{q}^H\mathbf{Q}\mathbf{q} > \mathsf{P}_0$), which holds true due to the hypothesis (otherwise, the first case discussed previously would occur), and the solution is unique since the function on the left hand side above is monotonic in μ. Besides, starting from (75), we have the following inequalities

$$\mathsf{P}_0 = \sum_{i=1}^{r}\frac{\lambda_i|p_i|^2}{(1+\mu^\star\lambda_i)^2} < \sum_{i=1}^{r}\frac{\lambda_i|p_i|^2}{\mu^{\star 2}\lambda_i^2} = (\mu^\star)^{-2}\sum_{i=1}^{r}\frac{\lambda_i|p_i|^2}{\lambda_i}, \tag{76}$$

and

$$\frac{\sum_{i=1}^{r}\lambda_i|p_i|^2}{(1+\mu^\star\lambda_1)^2} \leq \sum_{i=1}^{r}\frac{\lambda_i|p_i|^2}{(1+\mu^\star\lambda_i)^2} = \mathsf{P}_0. \tag{77}$$

Rearranging the above two inequalities yield the bounds in (25). Therefore, Lemma 1 has been proved. □

B. Proof of Lemma 2

Proof. Recall the arithmetic-geometric mean inequality reads

$$\prod_{n=1}^{N} x_n^{\varpi_n} \leq \sum_{n=1}^{N}\varpi_n x_n, \tag{78}$$

where x_n and ϖ_n are nonnegative scalars and $\sum_n \varpi_n = 1$. By taking $N = 2$, $\varpi_1 = \varpi_2 = \frac{1}{2}$, $x_1 = \sqrt{\frac{\alpha_{m,0}}{\alpha_{n,0}}}\alpha_m$ and $x_2 = \sqrt{\frac{\alpha_{n,0}}{\alpha_{m,0}}}\alpha_n$ in (78), (38a) can be obtained. For (38b), we first note the well known inequality stands

$$x \geq \log x + 1. \tag{79}$$

Substituting $x = \sqrt{\frac{\alpha_m\alpha_n}{\alpha_{m,0}\alpha_{n,0}}}$ into the above inequality, we obtain

$$\begin{aligned}\sqrt{\alpha_m\alpha_n} &\geq \sqrt{\alpha_{m,0}\alpha_{n,0}}\left(\frac{1}{2}\log(\alpha_m\alpha_n) - \frac{1}{2}\log(\alpha_{m,0}\alpha_{n,0})+1\right)\\ &= \frac{1}{2}\sqrt{\alpha_{m,0}}\sqrt{\alpha_{n,0}}\left(\log\alpha_m + \log\alpha_n\right) + \widetilde{c}_3,\end{aligned} \tag{80}$$

with $\widetilde{c}_3$ being given in (39). Therefore, (38b) has been proved. (38c) and (38d) can be readily obtained via substituting $\alpha_n = \alpha_{n,0} = 1$ into (38a) and (38b), respectively, and rearranging the terms. □

C. Proof of Theorem 2

Proof. Firstly, we assume that $\bar{b}_m > 0$, $\forall m \in \mathcal{M}$. Notice that Slater's condition holds for (P8). In fact, as long as all α_m are chosen sufficiently small, all constraints of (P8) can be satisfied strictly. Therefore, strong duality holds and consequently the optimal solutions can be equivalently identified via checking KKT conditions. Introduce the Lagrangian multipliers λ_m, μ_m and ν associated with the primal constraints $\alpha_m \geq 0$, $\alpha_m \leq \bar{\mathsf{P}}_m$ and $\sum_m \bar{c}_m\alpha_m \leq \mathsf{P}_\mathsf{R}$, respectively, $\forall m \in \mathcal{M}$. Then the KKT conditions of (P8) are summarized in the following

$$0 \leq \alpha_m \leq \bar{\mathsf{P}}_m, \lambda_m \geq 0, \mu_m \geq 0, \nu \geq 0,\ \sum_m \bar{c}_m\alpha_m \leq \mathsf{P}_\mathsf{R}, \tag{81a}$$

$$\mu_m(\alpha_m - \bar{\mathsf{P}}_m) = 0,\ \nu\left(\sum_m \bar{c}_m\alpha_m - \mathsf{P}_\mathsf{R}\right) = 0, \forall m \in \mathcal{M}, \tag{81b}$$

$$\lambda_m\alpha_m = 0,\ \bar{a}_m - \bar{b}_m\alpha_m^{-1} + \mu_m + \bar{c}_m\nu - \lambda_m = 0, \forall m \in \mathcal{M}. \tag{81c}$$

In the following, we analyze the KKT conditions according to the value of ν, i.e. ν is zero or positive.

<u>CASE-I</u>: $\nu = 0$. Obviously $\alpha_m^\star > 0$, (otherwise the objective will go to infinity due to the logarithm term), $\forall m \in \mathcal{M}$. This implies $\lambda_m^\star = 0$, $\forall m$, which by (81c) further leads to

$$\mu_m^\star = \bar{b}_m/\alpha_m^\star - \bar{a}_m,\ \forall m \in \mathcal{M}. \tag{82}$$

In the following, we consider two possible sub-cases: $\mu_m^\star > 0$ or $\mu_m^\star = 0$.

Case i): If $\mu_m^\star > 0$, then by KKT condition, we immediately have $\alpha_m^\star = \bar{\mathsf{P}}_m$. Therefore, $\mu_m^\star = \frac{\bar{b}_m}{\bar{\mathsf{P}}_m} - \bar{a}_m > 0$. This requires $\bar{\mathsf{P}}_m < \frac{\bar{b}_m}{\bar{a}_m}$. Otherwise $\mu_m^\star$ could not be positive, which indeed implies this sub-case could not occur.

Case ii): If $\mu_m^\star = 0$, then $0 \leq \alpha_m^\star \leq \bar{\mathsf{P}}_m$. By (82), $0 = \mu_m^\star = \frac{\bar{b}_m}{\alpha_m^\star} - \bar{a}_m$, i.e. $\alpha_m^\star = \frac{\bar{b}_m}{\bar{a}_m}$. Note that at this time $\frac{\bar{b}_m}{\bar{a}_m} \leq \bar{\mathsf{P}}_m$ should stand. Otherwise this sub-case could not occur.

Summarizing the above two sub-cases, we readily obtain

$$\mu_m^\star = \left[\bar{b}_m/\bar{\mathsf{P}}_m - \bar{a}_m\right]_+, \alpha_m^\star = \min\left\{\bar{\mathsf{P}}_m, \bar{b}_m/\bar{a}_m\right\}, \forall m \in \mathcal{M}. \tag{83}$$

where $[x]_+ \triangleq \max\{x, 0\}$. Note that (83) will satisfy all KKT conditions in (81) except for the sum-power constraint $\sum_m \bar{c}_m\alpha_m^\star \leq \mathsf{P}_\mathsf{R}$. If the sum-power constraint is also satisfied, then $\boldsymbol{\alpha}^\star$ is indeed the optimal solution to (P8) (note that optimal solution is unique since the objective is strictly convex). However, if the sum-power constraint is not satisfied, then a contradiction has actually been reached. That actually implies that the hypothesis that $\nu^\star = 0$ from the very beginning is false, which further implies that $\nu^\star > 0$ (note that the optimal solution of (P8) obviously exists, i.e. these exits one unique solution to the KKT conditions in (81)).

<u>CASE-II</u>: $\nu^\star > 0$. This immediately implies that $\sum_m \bar{c}_m\alpha_m^\star = \mathsf{P}_\mathsf{R}$. Since $\alpha_m^\star > 0$, we have $\lambda_m^\star = 0$. Then by (81c) we obtain

$$\bar{c}_m^{-1}\mu_m^\star + \nu^\star = \bar{c}_m^{-1}\left(\bar{b}_m/\alpha_m^\star - \bar{a}_m\right),\ \forall m \in \mathcal{M}. \tag{84}$$

Analogous to the analysis for CASE-I, we consider two possible sub-cases in the following.

Case i): If $\mu_m^\star > 0$, then by KKT condition $\alpha_m^\star = \bar{\mathsf{P}}_m$. At this time, we have

$$\bar{c}_m^{-1}\mu_m^\star + \nu^\star = \bar{c}_m^{-1}\left(\bar{b}_m/\bar{\mathsf{P}}_m - \bar{a}_m\right). \tag{85}$$

This article has been accepted for publication in IEEE Transactions on Signal Processing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TSP.2022.3221345

Since, by hypothesis, $\mu_m^\star > 0$ should stand. Therefore, by (85), the following condition

$$0 < \nu^\star < \bar{c}_m^{-1}\left(\bar{b}_m/\bar{\mathsf{P}}_m - \bar{a}_m\right) \tag{86}$$

should be satisfied. Otherwise $\mu_m^\star$ cannot be positive, which implies that the current sub-case could not happen.

Case ii): if $\mu_m^\star = 0$, then $0 \le \alpha_m^\star \le \bar{\mathsf{P}}_m$. At this time, from (84) we obtain

$$\nu^\star = \bar{c}_m^{-1}\left(\bar{b}_m/\alpha_m^\star - \bar{a}_m\right). \tag{87}$$

Notice the fact that the feasible domain of α_m is the interval $[0, \bar{\mathsf{P}}_m]$. Therefore the right hand side of the above equation can only take value in the range $\left[\bar{c}_m^{-1}(\frac{\bar{b}_m}{\bar{\mathsf{P}}_m} - \bar{a}_m), +\infty\right)$. As a result, the equation (87) can hold only if

$$\nu^\star \ge \bar{c}_m^{-1}\left(\bar{b}_m/\bar{\mathsf{P}}_m - \bar{a}_m\right). \tag{88}$$

stands true. Otherwise, this sub-case could not occur indeed.

Summarizing the above two sub-cases and comparing the conditions in (86) and (88), we immediately conclude that

$$\alpha_m^\star = \left[\left(\nu^\star \bar{c}_m + \bar{a}_m\right)^{-1} \bar{b}_m\right]^{\bar{\mathsf{P}}_m}, \tag{89}$$

for CASE-II. As can be seen from (89), $\alpha_m(\nu)$ is a monotonically decreasing function in ν. Besides, $\alpha_m(0) = \min\left\{\bar{\mathsf{P}}_m, \bar{b}_m/\bar{a}_m\right\} = \breve{\alpha}_m$. Therefore, when ν varies from $+\infty$ to 0, the value of $\alpha_m(\nu)$ varies from 0 to $\breve{\alpha}_m$. Since, according to the hypothesis on CASE-II, $\sum_m \bar{c}_m \breve{\alpha}_m > \mathsf{P}_\mathsf{R}$. Consequently there exists a unique positive $\nu^\star$ satisfying $\sum_m \bar{c}_m \alpha_m^\star = \mathsf{P}_\mathsf{R}$.

Lastly, recall that we make the assumption that $\bar{b}_m > 0$, $\forall m \in \mathcal{M}$ from the very beginning. However, $\bar{b}_m = 0$ could happen for (P8). If there exists some index m with $\bar{b}_m = 0$, then, obviously, its associated optimal value $\alpha_m^\star = 0$. This conclusion in fact accommodates well to the solutions in (83) or (89) developed above for the all positive $\{\bar{b}_m\}$ case. Therefore the solution to (P8) has been fully identified. $\square$

D. Proof of Lemma 3

Proof. Consider the problem (P9) given as follows

$$(\mathsf{P9}): \min_{\boldsymbol{\alpha}} \sum_{m=1}^{M} \bar{a}_m \alpha_m - \sum_{m=1}^{M} \bar{b}_m \log \alpha_m \tag{90a}$$

$$\text{s.t. } 0 \le \alpha_m \le \bar{\mathsf{P}}_m,\ \forall m \in \mathcal{M}, \tag{90b}$$

which is actually a relaxation of (P8) obtained via omitting the sum-power constraint (45b). Obviously $\breve{\boldsymbol{\alpha}}$ is the optimal solution of (P9). Therefore, we have $\mathsf{obj}(\boldsymbol{\alpha}^\star) \ge \mathsf{obj}(\breve{\boldsymbol{\alpha}})$.

Now we rewrite (P8) into an equivalent form as follows

$$(\mathsf{P8}'): \min_{\boldsymbol{\alpha} \in \mathcal{A}} \sum_{m=1}^{M} \bar{a}_m \alpha_m - \sum_{m=1}^{M} \bar{b}_m \log \alpha_m \tag{91a}$$

$$\text{s.t. } \sum_{m=1}^{M} \bar{c}_m \alpha_m \le \mathsf{P}_\mathsf{R}, \tag{91b}$$

where the convex set $\mathcal{A}$ is defined as $\mathcal{A} \triangleq \{\boldsymbol{\alpha} | 0 \le \alpha_m \le \bar{\mathsf{P}}_m, \forall m \in \mathcal{M}\}$. The Lagrangian of (P8$'$) is given as

$$\mathcal{L}(\boldsymbol{\alpha}, \nu) = \sum_m \bar{a}_m \alpha_m - \sum_m \bar{b}_m \log \alpha_m + \nu\left(\sum_m \bar{c}_m \alpha_m - \mathsf{P}_\mathsf{R}\right). \tag{92}$$

Note that strong duality holds for (P8$'$) (since Slater's condition is obviously satisfied). Therefore the pair of optimal primal-dual variables $(\boldsymbol{\alpha}^\star, \mu^\star)$ forms a saddle point of the Lagrangian in (92) [46]. Then we have the following relations

$$\mathsf{obj}(\breve{\boldsymbol{\alpha}}) \le \mathsf{obj}(\boldsymbol{\alpha}^\star) = \mathcal{L}\left(\boldsymbol{\alpha}^\star, \nu^\star\right) \tag{93a}$$

$$\overset{(a)}{=} \min_{\boldsymbol{\alpha} \in \mathcal{A}} \left\{ \sum_m^M \bar{a}_m \alpha_m - \sum_m^M \bar{b}_m \log \alpha_m + \nu^\star\left(\sum_m \bar{c}_m \alpha_m - \mathsf{P}_\mathsf{R}\right)\right\} \tag{93b}$$

$$\overset{(b)}{\le} \sum_m \bar{a}_m \widetilde{\alpha}_m - \sum_m \bar{b}_m \log \widetilde{\alpha}_m + \nu^\star\left(\sum_m \bar{c}_m \widetilde{\alpha}_m - \mathsf{P}_\mathsf{R}\right), \tag{93c}$$

where (a) is due to the saddle point theorem [46] and (b) holds because $\widetilde{\boldsymbol{\alpha}}$ is an arbitrary strictly feasible solution. Notice that $\sum_m \bar{c}_m \widetilde{\alpha}_m < \mathsf{P}_\mathsf{R}$ according to our choice of $\widetilde{\boldsymbol{\alpha}}$. Rearranging the above inequality, we obtain the upper-bound in (48). $\square$

E. Proof of Theorem 4

Proof. We will show that Alg.3 falls in the MM framework [42]–[44]. In the subsequent discussion, we call the conditions (A1)-(A4) in [44] as "MM conditions" and any upper-bound satisfying the MM conditions as "MM upper-bound". Obviously, the update of $\mathbf{F}$ follows MM conditions. For $\boldsymbol{\alpha}$ update, the objective of (P8) is actually an MM upper-bound of the objective of (P7) (see the derivation of Lemma 2).

Before proceeding, we notice the following useful results

Lemma 4. *The following facts hold true*

- Fact-1 *: Suppose that $u(\mathbf{x})$ and $f(\mathbf{x})$ are continuously differentiable in $\mathbf{x}$ and $\mathbf{x}(\mathbf{y})$ is continuously differentiable in $\mathbf{y}$. Denote $\mathbf{x}^{(t)} = \mathbf{x}(\mathbf{y}^{(t)})$. If $u(\mathbf{x}|\mathbf{x}^{(t)})$ and $f(\mathbf{x})$ satisfy MM conditions w.r.t. $\mathbf{x}$ at $\mathbf{x}^{(t)}$. Then $u(\mathbf{y}|\mathbf{y}^{(t)})$ and $f(\mathbf{y})$ still satisfy MM conditions w.r.t. $\mathbf{y}$ at $\mathbf{y}^{(t)}$.*
- Fact-2 *: if $u_1(\mathbf{x}|\mathbf{x}_0)$ is an MM upper-bound of $f(\mathbf{x})$ at $\mathbf{x}_0$ and $u_2(\mathbf{x}|\mathbf{x}_0)$ is an MM upper-bound of $u_1(\mathbf{x}|\mathbf{x}_0)$ at $\mathbf{x}_0$. Then $u_2(\mathbf{x}|\mathbf{x}_0)$ is still an MM upper-bound of $f(\mathbf{x})$ at $\mathbf{x}_0$.*

Proof. To see Fact-1, we can easily verify the following facts

$$u(\mathbf{y}^{(t)}|\mathbf{y}^{(t)}) = u(\mathbf{x}(\mathbf{y}^{(t)})|\mathbf{x}(\mathbf{y}^{(t)})) = u(\mathbf{x}^{(t)}|\mathbf{x}^{(t)}) = f(\mathbf{x}^{(t)}), \tag{94a}$$

$$u(\mathbf{y}|\mathbf{y}^{(t)}) = u(\mathbf{x}(\mathbf{y})|\mathbf{x}(\mathbf{y}^{(t)})) = u(\mathbf{x}|\mathbf{x}^{(t)}) \ge f(\mathbf{x}) = f(\mathbf{y}), \forall \mathbf{y}. \tag{94b}$$

Besides, following the chain rule of derivative, we have

$$\nabla_\mathbf{y} u(\mathbf{y}|\mathbf{y}^{(t)}) = \nabla_\mathbf{y} u(\mathbf{x}(\mathbf{y})|\mathbf{x}(\mathbf{y}^{(t)})) = \nabla_\mathbf{y} \mathbf{x}(\mathbf{y}) \nabla_\mathbf{x} u(\mathbf{x}|\mathbf{x}^{(t)}). \tag{95}$$

Therefore, substituting $\mathbf{y}$ with the value of $\mathbf{y}^{(t)}$ into the above

$$\begin{aligned} \nabla_\mathbf{y} u(\mathbf{y}^{(t)}|\mathbf{y}^{(t)}) &= \nabla_\mathbf{y} \mathbf{x}(\mathbf{y}^{(t)}) \nabla_\mathbf{x} u(\mathbf{x}(\mathbf{y}^{(t)})|\mathbf{x}^{(t)}) \\ &= \nabla_\mathbf{y} \mathbf{x}(\mathbf{y}^{(t)}) \nabla_\mathbf{x} f(\mathbf{x}(\mathbf{y}^{(t)})) = \nabla_\mathbf{y} f(\mathbf{y}^{(t)}). \end{aligned} \tag{96}$$

Hence Fact-1 has been clarified. For Fact-2, $u_2(\mathbf{x}_0|\mathbf{x}_0) = u_1(\mathbf{x}_0|\mathbf{x}_0) = f(\mathbf{x}_0)$, $u_2(\mathbf{x}|\mathbf{x}_0) \ge u_1(\mathbf{x}|\mathbf{x}_0) \ge f(\mathbf{x})$, $\forall \mathbf{x}$, and $\nabla u_2(\mathbf{x}_0) = \nabla u_1(\mathbf{x}_0) = \nabla f(\mathbf{x}_0)$. Therefore Fact-2 has also been proved. $\square$

For the update of $\boldsymbol{\varsigma}$, we rewrite the terms $\mathbf{e}_\imath^H \mathbf{S}_\imath \mathbf{e}_\imath$ in a form of the second order Taylor expansion and upper-bound its second order terms (see (56)). Therefore, the upper-bound in (56) has identical first order and zero-th order behaviour with the original objective function at $\boldsymbol{\varsigma}_0$. This implies (57) is indeed an MM upper-bound w.r.t $\boldsymbol{\varsigma}$ and $\mathbf{e}_\mathsf{T}$, respectively. By Fact-2 and Fact-1, we see that (56) is actually an MM upper-bound w.r.t. $\boldsymbol{\varsigma}$. Besides, since (59-60) are obtained via linearizing smooth concave functions, the MM conditions are satisfied. Via invoking Fact-2 for multiple times, we conclude that the objective of (P11) is indeed an MM upper-bound. For the update of $(\boldsymbol{\phi}_{\mathsf{R}}, \boldsymbol{\phi}_{\mathsf{T}})$, following identical arguments as for (56), (71) is also an MM upper-bound. Besides, the updates of $\mathbf{w}$ and $\boldsymbol{\beta}$ also comply with MM conditions, as proved in [43, Sec.8.1].

Therefore, we have clarified that Alg.3 actually follows the MM methodology. According to [43, Thm.1], the obtained objective sequence $\{\widetilde{\mathsf{R}}(\mathbf{w}^{(t)}, \boldsymbol{\beta}^{(t)}, \mathbf{F}^{(t)}, \boldsymbol{\alpha}^{(t)}, \boldsymbol{\varsigma}^{(t)}, \boldsymbol{\phi}_{\mathsf{R}}^{(t)}, \boldsymbol{\phi}_{\mathsf{T}}^{(t)})\}$ keeps increasing after each update of any block of variables. Since $\widetilde{\mathsf{R}}(\mathbf{w}^{(t)}, \boldsymbol{\beta}^{(t)}, \mathbf{F}^{(t)}, \boldsymbol{\alpha}^{(t)}, \boldsymbol{\varsigma}^{(t)}, \boldsymbol{\phi}_{\mathsf{R}}^{(t)}, \boldsymbol{\phi}_{\mathsf{T}}^{(t)}) = \mathsf{R}(\mathbf{F}^{(t)}, \boldsymbol{\alpha}^{(t)}, \boldsymbol{\varsigma}^{(t)}, \boldsymbol{\phi}_{\mathsf{R}}^{(t)}, \boldsymbol{\phi}_{\mathsf{T}}^{(t)})$ (proved in [38]), the monotonic decreasing objective value of (P0) has been proved. Besides, from the development of our MM-based solution, it can be readily verified that the updates of the blocks $\boldsymbol{\alpha}$, $\boldsymbol{\varsigma}$, $(\boldsymbol{\phi}_{\mathsf{R}}, \boldsymbol{\phi}_{\mathsf{T}})$, $\boldsymbol{\beta}$ and $\mathbf{w}$ all have unique solution and all the optimization variables obviously have compact feasible domains. Therefore, via invoking [43, Thm.2], we conclude that any limit point of the solution iterates yielded by Alg.3 is stationary point of the original problem (P0). □

REFERENCES

[1] C. Pan and *et al.*, "Reconfigurable intelligent surfaces for 6G systems: Principles, applications, and research directions," *IEEE Commun. Mag.*, vol. 59, no. 6, pp. 14–20, 2021.

[2] 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.

[3] 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.

[4] E. Basar and H. V. Poor, "Present and future of reconfigurable intelligent surface-empowered communications," *IEEE Signal Process. Mag.*, vol. 38, no. 6, pp. 146–152, Nov. 2021.

[5] C. You and R. Zhang, "Wireless communication aided by intelligent reflecting surface: Active or passive?" *IEEE Wireless Commun. Lett.*, vol. 10, no. 12, pp. 2659–2663, Dec. 2021.

[6] E. Bjornson, O. Ozdogan, and E. G. Larsson, "Intelligent reflecting surface versus decode-and-forward: How large surfaces are needed to beat relaying?" *IEEE Wireless Commun. Lett.*, vol. 9, no. 2, pp. 244–248, Oct. 2020.

[7] 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?," Mar. 2021. arXiv: 2103.15154.

[8] 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.

[9] K. Liu, Z. Zhang, L. Dai, S. Xu, and F. Yang, "Active reconfigurable intelligent surface: Fully-connected or sub-connected?" *IEEE Commun. Lett.*, 2021, Early Access.

[10] M. H. Khoshafa, T. M. N. Ngatched, M. H. Ahmed, and A. R. Ndjiongue, "Active reconfigurable intelligent surfaces-aided wireless communication system," *IEEE Commun. Lett.*, vol. 25, no. 11, pp. 3699–3703, Nov. 2021.

[11] R. A. Tasci, F. Kilinc, E. Basar, and G. C. Alexandropoulos, "An amplifying RIS architecture with a single power amplifier: Energy efficiency and error performance analysis," Nov. 2021. arXiv: 2111.09855.

[12] 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," Mar. 2021. arXiv: 2103.09104.

[13] J. Xu, Y. Liu, X. Mu, and O. A. Dobre, "STAR-RISs: Simultaneous transmitting and reflecting reconfigurable intelligent surfaces," *IEEE Commun. Lett.*, vol. 25, no. 9, pp. 3134–3138, Sep. 2021.

[14] X. Mu, Y. Liu, L. Guo, J. Lin, and R. Schober, "Simultaneously transmitting and reflecting (STAR) RIS aided wireless communications," *IEEE Trans. Wireless Commun.*, Sep. 2021, Early Access.

[15] S. Zeng and *et al.*, "Reconfigurable intelligent surfaces in 6G: Reflective, transmissive, or both?" *IEEE Commun. Lett.*, vol. 25, no. 6, pp. 2063–2067, Jun. 2021.

[16] ——, "Intelligent omni-surfaces: Reflection-refraction circuit model, full-dimensional beamforming, and system implementation," Jul. 2022. arXiv: 2206.00204.

[17] ——, "Beyond intelligent reflecting surfaces: Reflective-transmissive metasurface aided communications for full-dimensional coverage extension," *IEEE Trans. Veh. Technol.*, vol. 69, no. 11, pp. 13 905–13 909, Nov. 2020.

[18] S. Zhang and *et al.*, "Intelligent omni-surface: Ubiquitous wireless transmission by reflective-transmissive metasurface," Jun. 2021. arXiv: 2011.00765.

[19] M. Grant and S. Boyd, *CVX: Matlab software for disciplined convex programming, version 2.1*, http://cvxr.com/cvx, Mar. 2014.

[20] K. K. Kishor and S. V. Hum, "An amplifying reconfigurable reflectarray antenna," *IEEE Trans. Ant. Propag.*, vol. 60, no. 1, pp. 197–205, 2012.

[21] 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.*, vol. 65, no. 11, pp. 4198–4207, Nov. 2017.

[22] S. Khaledian, F. Farzami, D. Erricolo, and B. Smida, "A full-duplex bidirectional amplifier with low dc power consumption using tunnel diodes," *IEEE Microw. Wireless Compon. Lett.*, vol. 27, no. 11, pp. 1125–1127, Dec. 2017.

[23] N. Landsberg and E. Socher, "Design and measurements of 100 ghz reflectarray and transmitarray active antenna cells," *IEEE Trans. Ant. Propag.*, vol. 65, no. 12, pp. 6986–6997, Dec. 2017.

[24] ——, "A low-power 28-nm cmos fd-soi reflection amplifier for an active f-band reflectarray," *IEEE Trans. Microw. Theory Techn.*, vol. 65, no. 10, pp. 3910–3921, Oct. 2017.

[25] F. Amato, H. M. Torun, and G. D. Durgin, "Rfid backscattering in long-range scenarios," *IEEE Trans. Wireless Commun.*, vol. 17, no. 4, pp. 2718–2725, Apr. 2018.

[26] F. Amato, C. W. Peterson, B. P. Degnan, and G. D. Durgin, "Tunneling rfid tags for long-range and low-power microwave applications," *IEEE J. Radio Freq. Identificat.*, vol. 2, no. 2, pp. 93–103, Jun. 2018.

[27] K.-K. M. Cheng and S. Yeung, "A novel rat-race coupler with tunable power dividing ratio, ideal port isolation, and return loss performance," *IEEE Trans. Microw. Theory Techn.*, vol. 61, no. 1, pp. 55–66, Jan. 2013.

[28] Y. Wu, L. Jiao, B. Zhang, M. Li, W. Wang, and Y. Liu, "A new coupler structure with phase-controlled power divisions of extremely-wide tunable ranges and arbitrary phase differences," *IEEE Access*, vol. 6, pp. 10 121–10 130, 2018.

[29] 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.

[30] ——, "A coupler with wide power division ratio tuning range and flexible coupling direction," *IEEE Microw. Wireless Comp. Lett.*, vol. 31, no. 2, pp. 121–124, Feb. 2021.

[31] H. Peng, P. Lei, H. Yang, S. Zhao, and X. Ding, "Pi-type reconfigurable coupler based on a complementary tunable method," *J. Electr. Waves App.*, 2021.

[32] Y. Xiao, F. Lin, H. Ma, X. Tan, and H. Sun, "A planar balanced power divider with tunable power-dividing ratio," *IEEE Trans. Microw. Theory Techn.*, vol. 65, no. 12, pp. 4871–4882, Dec. 2017.

[33] F. Lin, "A planar balanced power divider with tunable power-dividing ratio," *IEEE Trans. Indust. Electr.*, vol. 65, no. 8, pp. 6515–6526, Aug. 2018.

[34] W. Yu and T. Lan, "Transmitter optimization for the multi-antenna downlink with per-antenna power constraints," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 2646–2660, 2007.

[35] H. Shen, W. Xu, A. Lee Swindlehurst, and C. Zhao, "Transmitter optimization for per-antenna power constrained multi-antenna downlinks: An SLNR maximization methodology," *IEEE Trans. Signal Process.*, vol. 64, no. 10, pp. 2712–2725, 2016.

[36] J. Chen, Y.-C. Liang, H. V. Cheng, and W. Yu, *Channel estimation for reconfigurable intelligent surface aided multi-user MIMO systems*, 2019. arXiv: 1912.03619.

[37] Z. Wang, L. Liu, and S. Cui, "Channel estimation for intelligent reflecting surface assisted multiuser communications: Framework, algorithms, and analysis," *IEEE Trans. Wireless Commun.*, Early Access.

[38] 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, 2011.

[39] D. P. Bertsekas, *Nonlinear Programming, 2nd ed.* Athena Scientific, Belmont, MA, 1999.

[40] 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.

[41] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*. Athena Scientific, Belmont, MA., 1997.

[42] Y. Sun, P. Babu, and D. P. Palomar, "Majorization minimization algorithms in signal processing, communications, and machine learning," *IEEE Trans. Signal Process.*, vol. 65, no. 3, pp. 794–816, Feb. 2017.

[43] M. Razaviyayn, M. Hong, and Z.-Q. Luo, "A unified convergence analysis of block successive minimization methods for nonsmooth optimization," *SIAM J. Optim.*, vol. 23, no. 2, pp. 1126–1153, 2013.

[44] M. Hong and *et. al.*, "A unified algorithmic framework for block-structured optimization involving big data: With applications in machine learning and signal processing," *IEEE Signal Process. Mag.*, vol. 33, no. 1, pp. 57–77, 2016.

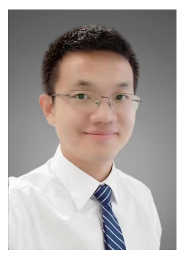
[45] T. M. Cover and J. A. Thomas, *Elements of Information Theory, 2nd Ed.* John Wiley & Sons. Inc., 2006.

[46] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York: Cambridge University Press, 2004.

Yang Liu received the Ph.D. degree in Electrical Engineering from Lehigh University, PA, USA, in 2016. Before that he received the B.E. and M.E. degrees in Electrical Engineering from Beijing University of Posts and Telecommunications (BUPT), Beijing, China, in 2007 and 2010, respectively. In 2016, He worked as a senior engineer in Marvell Semiconductor Ltd., Santa Clara, CA, USA. From 2016 to 2019, he worked as a research engineer in Huawei Technologies Co., Ltd., Shanghai, China. From Nov. 2019 to April 2020, he worked as a postdoctoral research fellow in Nanyang Technological University (NTU), Singapore. He is currently an associate professor in the School of Information and Communication Engineering, Dalian University of Technology, Dalian, China. His research interests include error correction coding, multi-antenna wireless communication systems, intelligent reflective surface and wireless sensor networks.

Yanan Ma received the B.S. degree in electronics information engineering from Dalian University of Technology, Dalian, China, in 2020. She is now studying toward the M.S. degree with the School of Information and Communication Engineering, Dalian University of Technology, Dalian, China. Her current research interests are focused on signal processing, reconfigurable intelligent surface, and massive MIMO systems.

Ming Li (S'05-M'11-SM'17) received the M.S. and Ph.D. degrees in electrical engineering from the State University of New York at Buffalo (SUNY-Buffalo), Buffalo, in 2005 and 2010, respectively. From Jan. 2011 to Aug. 2013, he was a Post-Doctoral Research Associate with the Signals, Communications, and Networking Research Group, Department of Electrical Engineering, SUNY-Buffalo. From Aug. 2013 to June 2014, Dr. Li joined Qualcomm Technologies Inc. as a Senior Engineer. Since June 2014, he has been with the School of Information and Communication Engineering, Dalian University of Technology, Dalian, China, where he is presently an Associate Professor. His current research interests are in the general areas of communication theory and signal processing with applications to intelligent reflecting surface, mmWave communications, secure wireless communications, cognitive radios and networks.



Qingqing Wu (S'13-M'16-SM'21) received the B.Eng. and the Ph.D. degrees in Electronic Engineering from South China University of Technology and Shanghai Jiao Tong University (SJTU) in 2012 and 2016, respectively. From 2016 to 2020, he was a Research Fellow in the Department of Electrical and Computer Engineering at National University of Singapore. His current research interest includes intelligent reflecting surface (IRS), unmanned aerial vehicle (UAV) communications, and MIMO transceiver design. He has coauthored more than 100 IEEE journal papers with 26 ESI highly cited papers and 8 ESI hot papers, which have received more than $15,000$ Google citations. He was listed as the Clarivate ESI Highly Cited Researcher in 2022 and 2021, the Most Influential Scholar Award in AI-2000 by Aminer in 2021 and World's Top 2% Scientist by Stanford University in 2020 and 2021.

Dr. Wu was the recipient of the IEEE Communications Society Asia Pacific Best Young Researcher Award and Outstanding Paper Award in 2022, the IEEE Communications Society Young Author Best Paper Award in 2021, the Outstanding Ph.D. Thesis Award of China Institute of Communications in 2017, the Outstanding Ph.D. Thesis Funding in SJTU in 2016, the IEEE ICCC Best Paper Award in 2021, and IEEE WCSP Best Paper Award in 2015. He was the Exemplary Editor of IEEE Communications Letters in 2019 and the Exemplary Reviewer of several IEEE journals. He serves as an Associate Editor for IEEE Transactions on Communications, IEEE Communications Letters, IEEE Wireless Communications Letters, IEEE Open Journal of Communications Society (OJ-COMS), and IEEE Open Journal of Vehicular Technology (OJVT). He is the Lead Guest Editor for IEEE Journal on Selected Areas in Communications on "UAV Communications in 5G and Beyond Networks", and the Guest Editor for IEEE OJVT on "6G Intelligent Communications" and IEEE OJ-COMS on "Reconfigurable Intelligent Surface-Based Communications for 6G Wireless Networks". He is the workshop co-chair for IEEE ICC 2019-2022 workshop on "Integrating UAVs into 5G and Beyond", and the workshop co-chair for IEEE GLOBECOM 2020 and ICC 2021 workshop on "Reconfigurable Intelligent Surfaces for Wireless Communication for Beyond 5G". He serves as the Workshops and Symposia Officer of Reconfigurable Intelligent Surfaces Emerging Technology Initiative and Research Blog Officer of Aerial Communications Emerging Technology Initiative. He is the IEEE Communications Society Young Professional Chair in Asia Pacific Region.

Qingjiang Shi received his Ph.D. degree in electronic engineering from Shanghai Jiao Tong University (SJTU), Shanghai, China, in 2011. From September 2009 to September 2010, he visited Prof. Z.-Q. (Tom) Luo's research group at the University of Minnesota, Twin Cities. In 2011, he worked as a Research Scientist at Bell Labs China. From 2012, He was with the School of Information and Science Technology at Zhejiang Sci-Tech University. From Feb. 2016 to Mar. 2017, he worked as a research fellow at Iowa State University, USA. From Mar. 2018, he is currently a full professor with the School of Software Engineering at Tongji University. He is also with the Shenzhen Research Institute of Big Data. His interests lie in algorithm design and analysis with applications in machine learning, signal processing and wireless networks. So far he has published more than 70 IEEE journals and filed about 30 national patents.

Dr. Shi was an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING. He was the recipient of the Huawei Outstanding Technical Achievement Award in 2021, the Huawei Technical Cooperation Achievement Transformation Award (2nd Prize) in 2022, the Golden Medal at the 46th International Exhibition of Inventions of Geneva in 2018, the First Prize of Science and Technology Award from China Institute of Communications in 2017, the National Excellent Doctoral Dissertation Nomination Award in 2013, the Shanghai Excellent Doctorial Dissertation Award in 2012, and the Best Paper Award from the IEEE PIMRC'09 conference.