



HAL
open science

Energy Efficiency in RIS-Aided Wireless Networks: Active or Passive RIS?

Robert K Fotock, Alessio Zappone, Marco Di Renzo

► **To cite this version:**

Robert K Fotock, Alessio Zappone, Marco Di Renzo. Energy Efficiency in RIS-Aided Wireless Networks: Active or Passive RIS?. IEEE International Conference on Communications (ICC 2023), May 2023, Rome, Italy. 10.1109/icc45041.2023.10279284 . hal-04738915

HAL Id: hal-04738915

<https://hal.science/hal-04738915v1>

Submitted on 16 Oct 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Energy Efficiency in RIS-Aided Wireless Networks: Active or Passive RIS?

1st Robert K. Fotock
 CNIT
 Cassino, Italy
 rfotock@cnit.it

2nd Alessio Zappone
 CNIT & Univ. of Cassino and Southern Lazio
 Cassino, Italy
 alessio.zappone@unicas.it

3rd Marco Di Renzo
 CentraleSupélec-University
 Paris, France
 marco.di-renzo@universite-paris-saclay.fr

Abstract—This work addresses the comparison between active and passive RISs in wireless networks, with reference to the system energy efficiency (EE). Two provably convergent and computationally-friendly EE maximization algorithms are developed, which optimize the reflection coefficients of the RIS, the transmit powers, and the linear receive filters. Numerical results show the performance of the proposed methods and discuss the operating points in which active or passive RISs should be preferred from an energy-efficient perspective.

I. INTRODUCTION

Recently, reconfigurable intelligent surfaces (RISs) have emerged as one of the main technologies for future wireless networks, for their ability to provide many degrees of freedom with limited power consumption [1]–[4]. In particular, energy efficiency (EE) is a major requirement of future wireless networks, also because, according to recent studies, 5G has not achieved the 2000x EE increase that was sought after [5].

From an energy perspective, the nearly passive behavior of RISs has been recognized as a major advantage, but, on the other hand, it also limits the rate performance that a RIS-aided network can ensure [6], [7], especially when no direct path exists [8], [9]. For this reason, recently, the use of active RIS has started to be investigated, i.e. the RIS is equipped with analog amplifiers that allow it to increase the amplitude of the incoming signal [8]. The idea of active RIS was influenced by the related field of backscatter communications, where the use of active loads was proposed in [10]. In [11], the use of reflection-type amplifiers is proposed to implement an active RIS. In this context, a joint transmit beamforming and reflect precoding algorithm is proposed. A comparison between active and passive RIS in terms of rate performance is performed in [12], considering the optimization of the position of the RIS in both uplink and downlink communications. In [13] a comparative study on channel estimation and spectral efficiency for both purely passive and hybrid RIS architectures was carried out. A hardware architecture for active RIS with a single amplifier with variable gain was proposed in [14].

On the other hand, the use of active components for amplification purposes inevitably increases the power consumption

of the RIS, both due to the radio frequency power for signal amplification, and due to the additional hardware that is employed. Therefore, while it is clear that active RISs can increase the system spectral efficiency, it is not equally clear whether the same is true for the EE. Related works on the design of wireless networks aided by active RISs focus only on the system rate, e.g., [15]–[17], neglecting the EE. This work aims at filling this gap, considering the maximization of the EE in a multi-user RIS-aided wireless network and comparing active and passive RISs. Moreover, the comparison is performed considering a RIS capable of global reflection, i.e. a recently-proposed kind of RIS in which the constraint on the reflected power is not applied to each reflecting element individually, but rather to the complete surface [18]. Finally, EE optimization is tackled also with respect to the linear receive filters, which complicates the analysis.

II. SYSTEM MODEL AND PROBLEM FORMULATION

Let us consider the uplink of a multi-user system in which K single-antenna mobile terminals communicate with a base station with N_R antennas, through an RIS with N reflecting elements (Fig. 1). Let us denote by \mathbf{h}_k the $N \times 1$ channel from user k to the RIS, by \mathbf{G} the $N_R \times N$ channel from the RIS to the base station, by $\mathbf{\Gamma} = \text{diag}(\gamma_1, \dots, \gamma_N)$, the matrix whose diagonal contains the N RIS reflection coefficients, and by p_k and s_k the transmit power and information symbol of user k .

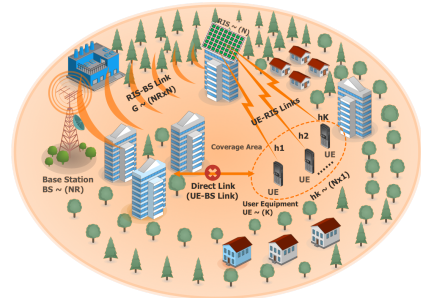


Fig. 1. Considered wireless network.

As already mentioned, most related works in the literature consider passive RISs with local reflection capabilities, i.e. they assume that the modulus of each RIS element is individually constrained to be less than a threshold, namely $|\gamma_n|^2 \leq P_R$

The work of R. K. Fotock and A. Zappone was supported by the European Commission through the H2020-MSCA-ITN-METAWIRELESS project, grant agreement 956256. The work of M. Di Renzo was supported in part by the European Commission through the H2020 ARIADNE project, grant agreement 871464 and the H2020 RISE-6G project, grant agreement 101017011.

for all $n = 1, \dots, N$, with $P_R \leq 1$. Instead, a more general scenario is that of active RISs with global reflection constraints, i.e. a constraint on the total power reflected by all of the RIS elements, and each element of the RIS is capable of amplifying the incoming signal thanks to the use of reflection-type amplifier implemented by different technologies, e.g. current inverters, current mirrors, or integrated circuits. In this case, the total power P_{out} departing from the RIS can be larger than the incoming power P_{in} , with the difference being the radio-frequency power consumed by the RIS for amplification purposes. Then, the power constraint for an active RIS is written as $P_{out} - P_{in} \leq P_{R,max}$, with $P_{R,max}$ the maximum power that can be provided by the RIS for signal amplification. Mathematically speaking, the signals that arrive at and depart from the RIS can be written as $\mathbf{y}_{in}^{RIS} = \sum_{k=1}^K \sqrt{p_k} \mathbf{h}_k s_k + \mathbf{n}$ and $\mathbf{y}_{out}^{RIS} = \sum_{k=1}^K \sqrt{p_k} \mathbf{\Gamma} \mathbf{h}_k s_k + \mathbf{\Gamma} \mathbf{n} = \sum_{k=1}^K \sqrt{p_k} \mathbf{H}_k \boldsymbol{\gamma} s_k + \mathbf{\Gamma} \mathbf{n}$, where $\mathbf{\Gamma} \mathbf{h}_k = \mathbf{H}_k \boldsymbol{\gamma}$ with \mathbf{H}_k a diagonal matrix having the vector \mathbf{h}_k as its diagonal, i.e., $\mathbf{H}_k = \text{diag}(h_k(1), \dots, h_k(N))$, $\boldsymbol{\gamma} = [\boldsymbol{\Gamma}(1,1), \dots, \boldsymbol{\Gamma}(N,N)]^T$ contains the RIS coefficients, p_k and s_k are the k -th user's transmit power and information symbol, and \mathbf{n} is the thermal noise introduced by the transmit amplifier of the RIS¹, modeled as a zero-mean, circularly symmetric Gaussian random variable with variance σ_{RIS}^2 . The signal received at the final destination is $\mathbf{r} = \sum_{k=1}^K \sqrt{p_k} s_k \mathbf{G} \mathbf{H}_k \boldsymbol{\gamma} + \mathbf{G} \mathbf{\Gamma} \mathbf{n} + \mathbf{w}$, which, after applying the linear filter \mathbf{c}_k , yields the SINR for user k

$$\text{SINR}_k = \frac{p_k |\mathbf{c}_k^H \mathbf{A}_k \boldsymbol{\gamma}|^2}{\mathbf{c}_k^H \mathbf{W} \mathbf{c}_k + \sum_{m \neq k} p_m |\mathbf{c}_k^H \mathbf{A}_m \boldsymbol{\gamma}|^2}, \quad (1)$$

with $\mathbf{A}_k = \mathbf{G} \mathbf{H}_k$ for all k and $\mathbf{W} = \sigma^2 \mathbf{I}_{N_R} + \sigma_{RIS}^2 \mathbf{G} \mathbf{\Gamma} \mathbf{\Gamma}^H \mathbf{G}^H$ is the covariance matrix of the overall colored noise at the receiver. Then, the power P_{in} that arrives at the RIS and the power P_{out} that departs from the RIS are given by

$$P_{in} = \sum_{k=1}^K p_k \|\mathbf{h}_k\|^2 + \sigma_{RIS}^2 N = \sum_{k=1}^K p_k \text{tr}(\mathbf{H}_k \mathbf{H}_k^H) + \sigma_{RIS}^2 N \quad (2)$$

$$P_{out} = \sum_{k=1}^K p_k \text{tr}(\mathbf{H}_k \boldsymbol{\gamma} \boldsymbol{\gamma}^H \mathbf{H}_k^H) + \sigma_{RIS}^2 \text{tr}(\boldsymbol{\gamma} \boldsymbol{\gamma}^H) \quad (3)$$

Then, the radio-frequency power consumed by the RIS is

$$\begin{aligned} P_{out} - P_{in} &= \sum_{k=1}^K p_k \text{tr}(\mathbf{H}_k (\boldsymbol{\gamma} \boldsymbol{\gamma}^H - \mathbf{I}_N) \mathbf{H}_k^H) + \sigma_{RIS}^2 \text{tr}(\boldsymbol{\gamma} \boldsymbol{\gamma}^H - \mathbf{I}_N) \\ &= \sum_{k=1}^K \left[p_k \text{tr}(\mathbf{H}_k (\boldsymbol{\gamma} \boldsymbol{\gamma}^H - \mathbf{I}_N) \mathbf{H}_k^H) + \frac{\sigma_{RIS}^2}{K} \text{tr}(\boldsymbol{\gamma} \boldsymbol{\gamma}^H - \mathbf{I}_N) \right] \\ &= \text{tr}((\boldsymbol{\gamma} \boldsymbol{\gamma}^H - \mathbf{I}_N) \mathbf{R}), \end{aligned} \quad (4)$$

wherein we have defined the positive definite, diagonal matrix

$$\mathbf{R} = \sum_{k=1}^K \left(p_k \mathbf{H}_k^H \mathbf{H}_k + \frac{\sigma_{RIS}^2}{K} \mathbf{I}_N \right) = \sum_{k=1}^K p_k \mathbf{H}_k^H \mathbf{H}_k + \sigma_{RIS}^2 \mathbf{I}_N. \quad (5)$$

¹Note that due to the presence of power amplifiers, an active RIS introduces a noise amplification effect.

Thus, when the RIS operates in the active regime, it must hold $P_{out} - P_{in} \geq 0$, which yields the condition $\text{tr}((\boldsymbol{\gamma} \boldsymbol{\gamma}^H - \mathbf{I}_N) \mathbf{R}) \geq 0$, which, in turn, can be readily expressed as $\boldsymbol{\gamma}^H \mathbf{R} \boldsymbol{\gamma} = \text{tr}(\mathbf{R} \boldsymbol{\gamma} \boldsymbol{\gamma}^H) \geq \text{tr}(\mathbf{R})$. Next, the total power consumption of the system is given by

$$\begin{aligned} P_{tot} &= \sum_{k=1}^K \mu_k p_k + p_k \text{tr}(\mathbf{H}_k (\boldsymbol{\gamma} \boldsymbol{\gamma}^H - \mathbf{I}_N) \mathbf{H}_k^H) \\ &\quad + \sigma_{RIS}^2 \text{tr}(\boldsymbol{\gamma} \boldsymbol{\gamma}^H - \mathbf{I}_N) + P_c \\ &= \text{tr}(\mathbf{R} \boldsymbol{\gamma} \boldsymbol{\gamma}^H) - \sigma_{RIS}^2 N + \sum_{k=1}^K p_k (\mu_k - \|\mathbf{h}_k\|^2) + P_c, \end{aligned} \quad (6)$$

with μ_k the inverse of the efficiency of user's k transmit amplifier, while $P_c = P_0 + N P_{c,n} + P_{0,RIS}$, with $P_{c,n}$ the static power consumed by each active RIS element to enable the reconfiguration, and $P_{0,RIS}$ accounting for all other sources of static power consumption of the active RIS. Thus, the resulting global EE in the active RIS scenario is given by

$$\text{GEE} = \frac{B \sum_{k=1}^K \log_2(1 + \text{SINR}_k)}{\text{tr}(\mathbf{R} \boldsymbol{\gamma} \boldsymbol{\gamma}^H) - \sigma_{RIS}^2 N + \sum_{k=1}^K p_k (\mu_k - \|\mathbf{h}_k\|^2) + P_c}, \quad (7)$$

with B the communication bandwidth. Finally, the GEE maximization problem in the case of an active RIS with global reflection constraints is written as

$$\max_{\boldsymbol{\gamma}, \mathbf{p}, \mathbf{C}} \text{GEE}(\boldsymbol{\gamma}, \mathbf{p}, \mathbf{C}) \quad (8a)$$

$$\text{s.t. } \text{tr}(\mathbf{R}) \leq \boldsymbol{\gamma}^H \mathbf{R} \boldsymbol{\gamma} \leq P_{R,max} + \text{tr}(\mathbf{R}) \quad (8b)$$

$$0 \leq p_k \leq P_{max,k} \quad \forall k = 1, \dots, K, \quad (8c)$$

wherein (8b) ensures that the RIS operates in the active regime and that the RIS power budget is not exceeded. Also, observe that (8b) is always feasible². In order to tackle the non-convex Problem (8), two methods are presented in Sections III and IV. In both cases, it is assumed that channel realizations are reliably estimated and available for resource allocation.

III. FIRST PROPOSED APPROACH

The first optimization method is based on the alternating optimization algorithm applied to the variables $\boldsymbol{\gamma}$, \mathbf{p} , and \mathbf{C} .

1) *Optimization of $\boldsymbol{\gamma}$* : With respect to $\boldsymbol{\gamma}$, the problem is

$$\max_{\boldsymbol{\gamma}} \frac{\sum_{k=1}^K \log_2 \left(1 + \frac{p_k |\mathbf{c}_k^H \mathbf{A}_k \boldsymbol{\gamma}|^2}{\mathbf{c}_k^H \mathbf{W} \mathbf{c}_k + \sum_{m \neq k} p_m |\mathbf{c}_k^H \mathbf{A}_m \boldsymbol{\gamma}|^2} \right)}{\text{tr}(\mathbf{R} \boldsymbol{\gamma} \boldsymbol{\gamma}^H) + P_{c,eq}} \quad (9a)$$

$$\text{s.t. } \text{tr}(\mathbf{R}) \leq \boldsymbol{\gamma}^H \mathbf{R} \boldsymbol{\gamma} \leq P_{R,max} + \text{tr}(\mathbf{R}) \quad (9b)$$

where $P_{c,eq} = \sum_k p_k (\mu_k - \|\mathbf{h}_k\|^2) + P_c - \sigma_{RIS}^2 N$, and we recall that \mathbf{W} depends on the RIS vector $\boldsymbol{\gamma}$. Problem (9) is more challenging than EE optimization with passive RISs, because: 1) the RIS vector $\boldsymbol{\gamma}$ appears also at the denominator of the objective (9a), which makes Problem (9) a fractional program; 2) the RIS vector $\boldsymbol{\gamma}$ appears also in the receiver noise covariance matrix \mathbf{W} ; 3) the first inequality in (9b) is a non-convex constraint.

²Indeed, since \mathbf{R} is a diagonal matrix, the left-hand-side of (8b) can be always fulfilled by choosing $|\gamma_n| = 1$ for all $n = 1, \dots, N$, which falls back into the passive RIS regime.

In order to deal with the fractional nature of Problem (9), the framework of fractional programming can be employed [19]. However, it can not be used directly, because the numerator of (9a) is not a concave function. This issue will be addressed by resorting to the framework of sequential programming [20], and in particular to the sequential fractional programming method [19]. To begin with, let us express the numerator of (9a), and in particular the term $\mathbf{c}_k^H \mathbf{W} \mathbf{c}_k$, as a function of the vector $\boldsymbol{\gamma}$ rather than the matrix $\boldsymbol{\Gamma}$. To this end, defining $\mathbf{u}_k = \mathbf{G}^H \mathbf{c}_k$, $\tilde{\mathbf{U}}_k = \text{diag}(|u_1|^2, \dots, |u_N|^2)$, and plugging the expression of \mathbf{W} we obtain $\mathbf{c}_k^H \mathbf{W} \mathbf{c}_k = \sigma^2 \|\mathbf{c}_k\|^2 + \sigma_{\text{RIS}}^2 \boldsymbol{\gamma}^H \tilde{\mathbf{U}}_k \boldsymbol{\gamma}$. Next, in order to apply the sequential fractional programming method, a concave lower-bound of the numerator of (9a) must be found. To this end we apply the bound $\log_2 \left(1 + \frac{x}{y}\right) \geq \log_2 \left(1 + \frac{\bar{x}}{\bar{y}}\right) + \frac{\bar{x}}{\bar{y}} \left(\frac{2\sqrt{\bar{x}} - x + y}{\sqrt{\bar{x}} + \bar{x} + \bar{y}} - 1\right)$, which holds for any x, y, \bar{x} and \bar{y} , and holds with equality whenever $x = \bar{x}$ and $y = \bar{y}$. Indeed, denoting by $\bar{\boldsymbol{\gamma}}$ any feasible vector of RIS reflection coefficients and applying the bound to each summand of the numerator of (9a), we obtain

$$\begin{aligned} R_k &= \log_2(1 + \text{SINR}_k) \quad (10) \\ &\geq \log_2 \left(1 + \frac{p_k |\mathbf{c}_k^H \mathbf{A}_k \bar{\boldsymbol{\gamma}}|^2}{\sigma^2 \|\mathbf{c}_k\|^2 + \sigma_{\text{RIS}}^2 \bar{\boldsymbol{\gamma}}^H \tilde{\mathbf{U}}_k \bar{\boldsymbol{\gamma}} + \sum_{m \neq k} p_m |\mathbf{c}_k^H \mathbf{A}_m \bar{\boldsymbol{\gamma}}|^2}\right) \\ &+ \frac{p_k |\mathbf{c}_k \mathbf{A}_k \bar{\boldsymbol{\gamma}}|^2}{\sigma^2 \|\mathbf{c}_k\|^2 + \sigma_{\text{RIS}}^2 \bar{\boldsymbol{\gamma}}^H \tilde{\mathbf{U}}_k \bar{\boldsymbol{\gamma}} + \sum_{m \neq k} p_m |\mathbf{c}_k^H \mathbf{A}_m \bar{\boldsymbol{\gamma}}|^2} \left(\frac{2 |\mathbf{c}_k^H \mathbf{A}_k \bar{\boldsymbol{\gamma}}|}{|\mathbf{c}_k^H \mathbf{A}_k \bar{\boldsymbol{\gamma}}|}\right. \\ &\left. - \frac{\sigma^2 \|\mathbf{c}_k\|^2 + \sigma_{\text{RIS}}^2 \bar{\boldsymbol{\gamma}}^H \tilde{\mathbf{U}}_k \bar{\boldsymbol{\gamma}} + \sum_{m=1}^K p_m |\mathbf{c}_k^H \mathbf{A}_m \bar{\boldsymbol{\gamma}}|^2}{\sigma^2 \|\mathbf{c}_k\|^2 + \sigma_{\text{RIS}}^2 \bar{\boldsymbol{\gamma}}^H \tilde{\mathbf{U}}_k \bar{\boldsymbol{\gamma}} + \sum_{m=1}^K p_m |\mathbf{c}_k^H \mathbf{A}_m \bar{\boldsymbol{\gamma}}|^2} - 1\right) = \bar{R}_k \end{aligned}$$

Defining $\bar{L}_k = \sigma^2 \|\mathbf{c}_k\|^2 + \sigma_{\text{RIS}}^2 \bar{\boldsymbol{\gamma}}^H \tilde{\mathbf{U}}_k \bar{\boldsymbol{\gamma}} + \sum_{m \neq k} p_m |\mathbf{c}_k^H \mathbf{A}_m \bar{\boldsymbol{\gamma}}|^2$, $\bar{A}_k = \log_2(1 + p_k |\mathbf{c}_k^H \mathbf{A}_k \bar{\boldsymbol{\gamma}}|^2 / \bar{L}_k)$, $\bar{B}_k = p_k |\mathbf{c}_k^H \mathbf{A}_k \bar{\boldsymbol{\gamma}}|^2 / \bar{L}_k$, $\bar{D}_k = 2 / |\mathbf{c}_k^H \mathbf{A}_k \bar{\boldsymbol{\gamma}}|$, $\bar{E}_k = 1 / (\bar{L}_k + p_k |\mathbf{c}_k^H \mathbf{A}_k \bar{\boldsymbol{\gamma}}|^2)$, and $\bar{F}_k = \bar{E}_k \sigma^2 \|\mathbf{c}_k\|^2 + 1$, we can write \bar{R}_k as

$$\begin{aligned} \bar{R}_k &= \bar{A}_k + \bar{B}_k \left(\bar{D}_k |\mathbf{c}_k^H \mathbf{A}_k \bar{\boldsymbol{\gamma}}| - \bar{F}_k \right. \\ &\left. - \bar{E}_k \left(\sigma_{\text{RIS}}^2 \bar{\boldsymbol{\gamma}}^H \tilde{\mathbf{U}}_k \bar{\boldsymbol{\gamma}} + \sum_{m=1}^K p_m |\mathbf{c}_k^H \mathbf{A}_m \bar{\boldsymbol{\gamma}}|^2 \right) \right) \quad (11) \end{aligned}$$

Let us observe that the term $-\sigma_{\text{RIS}}^2 \bar{\boldsymbol{\gamma}}^H \tilde{\mathbf{U}}_k \bar{\boldsymbol{\gamma}}$ is concave and thus the only non-concave term in \bar{R}_k is $|\mathbf{c}_k^H \mathbf{A}_k \bar{\boldsymbol{\gamma}}|$. However, since $|\mathbf{c}_k^H \mathbf{A}_k \bar{\boldsymbol{\gamma}}|$ is convex in $\boldsymbol{\gamma}$, it is lower-bounded by its first-order Taylor expansion around any point $\bar{\boldsymbol{\gamma}}$, which yields the following concave lower-bound of the generic summand in the numerator of (9a)

$$R_k \geq \bar{R}_k \geq \bar{A}_k + \bar{B}_k \left(\bar{D}_k \left(|\mathbf{c}_k^H \mathbf{A}_k \bar{\boldsymbol{\gamma}}| \right. \right. \quad (12)$$

$$\left. \left. + \Re \left\{ \frac{\mathbf{A}_k^H \mathbf{c}_k \mathbf{c}_k^H \mathbf{A}_k \bar{\boldsymbol{\gamma}}}{|\mathbf{c}_k^H \mathbf{A}_k \bar{\boldsymbol{\gamma}}|} (\boldsymbol{\gamma} - \bar{\boldsymbol{\gamma}}) \right\} \right) \right) \quad (13)$$

$$- \bar{E}_k \left(\sigma_{\text{RIS}}^2 \bar{\boldsymbol{\gamma}}^H \tilde{\mathbf{U}}_k \bar{\boldsymbol{\gamma}} + \sum_{m=1}^K p_m |\mathbf{c}_k^H \mathbf{A}_m \bar{\boldsymbol{\gamma}}|^2 \right) - \bar{F}_k = \tilde{R}_k$$

Finally, it remains to deal with the first, non-convex inequality constraint in (9b). This is accomplished by employing the sequential approximation framework. Specifically, since $\boldsymbol{\gamma}^H \mathbf{R} \boldsymbol{\gamma}$ is a convex function of $\boldsymbol{\gamma}$, it is lower-bounded by its first-order Taylor expansion around any point $\bar{\boldsymbol{\gamma}}$, i.e., $\boldsymbol{\gamma}^H \mathbf{R} \boldsymbol{\gamma} \geq \bar{\boldsymbol{\gamma}}^H \mathbf{R} \bar{\boldsymbol{\gamma}} + 2\Re\{\bar{\boldsymbol{\gamma}}^H \mathbf{R} (\boldsymbol{\gamma} - \bar{\boldsymbol{\gamma}})\}$. Thus, in each iteration of the sequential method, the problem to be solved is

$$\max_{\boldsymbol{\gamma}} \frac{\sum_{k=1}^K \tilde{R}_k(\boldsymbol{\gamma})}{\text{tr}(\mathbf{R} \boldsymbol{\gamma} \boldsymbol{\gamma}^H) + P_{c,eq}} \quad (14a)$$

$$\text{s.t. } \boldsymbol{\gamma}^H \mathbf{R} \boldsymbol{\gamma} \leq P_{R,max} + \text{tr}(\mathbf{R}) . \quad (14b)$$

$$\bar{\boldsymbol{\gamma}}^H \mathbf{R} \bar{\boldsymbol{\gamma}} + 2\Re\{\bar{\boldsymbol{\gamma}}^H \mathbf{R} (\boldsymbol{\gamma} - \bar{\boldsymbol{\gamma}})\} \geq \text{tr}(\mathbf{R}) . \quad (14c)$$

The objective (14a) has a concave numerator and a convex denominator, since \tilde{R}_k is concave and \mathbf{R} is positive definite, while (14c) is a linear constraint. Thus, (14) is a pseudo-concave maximization subject to convex constraints, which can be solved by fractional programming methods.

2) *Optimization of \mathbf{p}* : Defining $a_{k,m} = |\mathbf{c}_k^H \mathbf{A}_m \boldsymbol{\gamma}|^2$, for all m and k , $d_k = \mathbf{c}_k^H \mathbf{W} \mathbf{c}_k$, $P_{c,eq} = \text{tr}(\mathbf{R} \boldsymbol{\gamma} \boldsymbol{\gamma}^H) - \sigma_{\text{RIS}}^2 N + P_c$, and $\mu_{k,eq} = \mu_k - \|\mathbf{h}_k\|^2$, the power optimization problem is

$$\max_{\mathbf{p}} \frac{\sum_{k=1}^K \log_2 \left(1 + \frac{p_k a_{k,k}}{d_k + \sum_{m \neq k} p_m a_{k,m}}\right)}{\sum_{k=1}^K \mu_{k,eq} p_k + P_{c,eq}} \quad (15a)$$

$$\text{s.t. } 0 \leq p_k \leq P_{max,k}, \forall k = 1, \dots, K \quad (15b)$$

Since the numerator of (15a) is not a concave function of \mathbf{p} , the objective in (15a) is not a pseudo-concave function and thus it is computationally unfeasible to solve (15) by fractional programming [19]. Then, we resort to the sequential fractional programming method [19], in order to derive a pseudo-concave lower-bound of (15a), which can be maximized by fractional programming. To this end, let us express (15a) as

$$\begin{aligned} \text{GEE}(\mathbf{p}) &= \frac{\sum_{k=1}^K \log_2 \left(d_k + \sum_{k=1}^K p_k a_{k,k} \right)}{\underbrace{\sum_{k=1}^K \mu_{k,eq} p_k + P_{c,eq}}_{g_1(\mathbf{p})}} \quad (16) \\ &- \frac{\sum_{k=1}^K \log_2 \left(d_k + \sum_{m \neq k} p_m a_{k,m} \right)}{\underbrace{\sum_{k=1}^K \mu_{k,eq} p_k + P_{c,eq}}_{g_2(\mathbf{p})}} . \end{aligned}$$

Since the numerator of $g_2(\mathbf{p})$ is concave, a pseudo-concave lower-bound of $\text{GEE}(\mathbf{p})$, denoted by $\widetilde{\text{GEE}}(\mathbf{p})$, is obtained replacing the numerator of $g_2(\mathbf{p})$ with its first-order Taylor expansion around any feasible point $\bar{\mathbf{p}}$ [19]. In each iteration of the sequential method, the function $\widetilde{\text{GEE}}(\mathbf{p})$ is maximized, with the constraint $p_k \in [0, P_{max,k}]$, for all $k = 1, \dots, K$, which is efficiently solved by fractional programming.

3) *Optimization of \mathbf{C}* : The optimization of the receive filters in \mathbf{C} affects only the numerator of the GEE. Moreover, it can be decoupled over the users, thus reducing to maximizing the individual rate of the users. The solution to this problem is well-known to be the linear MMSE receiver, which for

the case at hand, is expressed according to the formula $c_k = \sqrt{p_k} \mathbf{M}_k^{-1} \mathbf{A}_k \gamma$, with $\mathbf{M}_k = \sum_{m \neq k} p_m \mathbf{A}_m \gamma \gamma^H \mathbf{A}_m^H + \mathbf{W}$.

IV. SECOND PROPOSED APPROACH

While the previous approach considers the alternating optimization of three variables, namely, the transmit powers \mathbf{p} , the receive filters $\mathbf{c}_1, \dots, \mathbf{c}_K$, and the RIS reflection coefficients γ , a different strategy is that of embedding the optimal expression of the linear MMSE filters into the GEE, and optimizing the resulting expression with respect to \mathbf{p} and γ . While this is expected to yield better performance, an intuition that will be confirmed by numerical results, it also leads to a more challenging optimization problem, due to the more involved expression of the EE function. To elaborate, accounting for the fact that RIS amplification colors the noise at the receiver, the sum-rate with linear MMSE filtering is written as

$$\begin{aligned} \text{SR}_{\text{MMSE}} &= \sum_{k=1}^K \log_2 \left| \mathbf{W} + \sum_{m=1}^K p_m \mathbf{A}_m \gamma \gamma^H \mathbf{A}_m^H \right| \\ &\quad - \sum_{k=1}^K \log_2 \left| \mathbf{W} + \sum_{m \neq k} p_m \mathbf{A}_m \gamma \gamma^H \mathbf{A}_m^H \right| \end{aligned} \quad (17)$$

Thus, the GEE maximization problem can be formulated as

$$\max_{\gamma, \mathbf{p}} \frac{\text{SR}_{\text{MMSE}}(\gamma, \mathbf{p})}{\text{tr}(\mathbf{R}\gamma\gamma^H) - \sigma_{\text{RIS}}^2 N + \sum_{k=1}^K p_k (\mu_k - \|\mathbf{h}_k\|^2) + P_{c,eq}} \quad (18a)$$

$$\text{s.t. } \text{tr}(\mathbf{R}) \leq \text{tr}(\mathbf{R}\gamma\gamma^H) \leq P_{R,max} + \text{tr}(\mathbf{R}) \quad (18b)$$

$$p_k \in [0, P_{max,k}], \forall k = 1, \dots, K, \quad (18c)$$

where we have exploited that $\gamma^H \mathbf{R} \gamma = \text{tr}(\mathbf{R}\gamma\gamma^H)$. Problem (18) can be tackled by alternating optimization of γ and \mathbf{p} as discussed in the rest of this section.

1) *Optimization of γ* : The problem can be stated as

$$\max_{\gamma} \frac{\text{SR}_{\text{MMSE}}(\gamma)}{\text{tr}(\mathbf{R}\gamma\gamma^H) + P_{c,eq}} \quad (19a)$$

$$\text{s.t. } \text{tr}(\mathbf{R}) \leq \text{tr}(\mathbf{R}\gamma\gamma^H) \leq P_{R,max} + \text{tr}(\mathbf{R}) \quad (19b)$$

To begin with, let us observe that the noise covariance matrix \mathbf{W} can be expressed as $\mathbf{W} = \sigma^2 \mathbf{I}_{N_R} + \sigma_{\text{RIS}}^2 \mathbf{G} \Gamma \Gamma^H \mathbf{G}^H = \sigma^2 \mathbf{I}_{N_R} + \sigma_{\text{RIS}}^2 \mathbf{G} \text{diag}(\gamma \gamma^H) \mathbf{G}^H$. Then, defining $\mathbf{X} = \gamma \gamma^H$, the numerator of the objective in (19a) can be expressed as

$$\text{SR}_{\text{MMSE}}(\mathbf{X}) = \quad (20)$$

$$\begin{aligned} &\sum_{k=1}^K \log_2 \left| \underbrace{\sigma^2 \mathbf{I}_{N_R} + \sigma_{\text{RIS}}^2 \mathbf{G} \text{diag}(\mathbf{X}) \mathbf{G}^H + \sum_{m=1}^K p_m \mathbf{A}_m \mathbf{X} \mathbf{A}_m^H}_{G_1(\mathbf{X})} \right| - \\ &\sum_{k=1}^K \log_2 \left| \underbrace{\sigma^2 \mathbf{I}_{N_R} + \sigma_{\text{RIS}}^2 \mathbf{G} \text{diag}(\mathbf{X}) \mathbf{G}^H + \sum_{m \neq k} p_m \mathbf{A}_m \mathbf{X} \mathbf{A}_m^H}_{G_2(\mathbf{X})} \right|, \end{aligned}$$

where $\text{diag}(\mathbf{X})$ is the diagonal of the matrix \mathbf{X} , which is a linear function of \mathbf{X} .³ Thus, being the difference of two

³Observe also that $\text{diag}(\mathbf{X}) = \mathbf{X} \odot \mathbf{I}_N$, with \odot denoting the component-wise product.

concave functions, a concave lower-bound of (20) is found linearizing the negative sum in (20) around any point $\bar{\mathbf{X}}$. Specifically, it holds

$$\begin{aligned} \text{SR}_{\text{MMSE}}(\mathbf{X}) &\geq G_1(\mathbf{X}) - G_2(\bar{\mathbf{X}}) \\ &\quad - \Re \{ \text{tr}(\nabla G_2(\bar{\mathbf{X}})^H (\mathbf{X} - \bar{\mathbf{X}})) \} = \widetilde{\text{SR}}_{\text{MMSE}}(\mathbf{X}), \end{aligned} \quad (21)$$

with

$$\nabla G_2(\mathbf{X}) = \sum_{k=1}^K \left(\sigma_{\text{RIS}}^2 (\mathbf{G}^H \mathbf{T}_k^{-1} \mathbf{G}) \odot \mathbf{I}_N + \sum_{m \neq k} p_m \mathbf{A}_m^H \mathbf{T}_k^{-1} \mathbf{A}_m \right), \quad (22)$$

and $\mathbf{T}_k = \sigma^2 \mathbf{I}_{N_R} + \sigma_{\text{RIS}}^2 \mathbf{G} \text{diag}(\mathbf{X}) \mathbf{G}^H + \sum_{m \neq k} p_m \mathbf{A}_m \mathbf{X} \mathbf{A}_m^H$. Then, the problem to be solved in each iteration of the sequential method can be formulated as

$$\max_{\mathbf{X} \succeq \mathbf{0}} \frac{\widetilde{\text{SR}}_{\text{MMSE}}(\mathbf{X})}{\text{tr}(\mathbf{R}\mathbf{X}) + P_{c,eq}} \quad (23a)$$

$$\text{s.t. } \text{tr}(\mathbf{R}) \leq \text{tr}(\mathbf{R}\mathbf{X}) \leq P_{R,max} + \text{tr}(\mathbf{R}), \quad (23b)$$

where we have lifted the rank-one constraint on \mathbf{X} , thus employing the semidefinite relaxation method [21]. It can be seen that (23a) has a concave numerator and a convex denominator, and thus is a pseudo-concave function. Then, since the constraints of Problem (23) are affine, the problem can be globally solved by standard fractional programming tools. Upon convergence of the sequential procedure, if the convergence point \mathbf{X}^* has unit-rank, then \mathbf{X}^* is also feasible for the original problem. Otherwise, a feasible solution can be obtained by randomization techniques [21].

2) *Optimization of \mathbf{p}* : The problem can be stated as

$$\max_{\mathbf{p}} \text{GEE}_{\text{MMSE}} = \frac{\text{SR}_{\text{MMSE}}(\mathbf{p})}{\sum_{k=1}^K \mu_{k,eq} p_k + P_{c,eq}} \quad (24a)$$

$$\text{s.t. } p_k \in [0, P_{max,k}] \forall k = 1, \dots, K, \quad (24b)$$

which can be tackled by sequential fractional programming by linearizing the negative term in the numerator of (24a). Then, in each iteration of the sequential fractional programming algorithm, the surrogate problem to be solved is

$$\max_{\mathbf{p}} \widetilde{\text{GEE}}_{\text{MMSE}}(\mathbf{p}), \text{ s.t. } p_k \in [0, P_{max,k}], \forall k = 1, \dots, K, \quad (25)$$

with the objective given by

$$\begin{aligned} \widetilde{\text{GEE}}_{\text{MMSE}}(\mathbf{p}) &= \frac{\sum_{k=1}^K \log_2 \left| \mathbf{W} + \sum_{m=1}^K p_m \mathbf{A}_m \gamma \gamma^H \mathbf{A}_m^H \right|}{\sum_{k=1}^K \mu_{k,eq} p_k + P_{c,eq}} \\ &\quad - \frac{\sum_{k=1}^K \log_2 \left| \mathbf{W} + \sum_{m \neq k} \bar{p}_m \mathbf{A}_m \gamma \gamma^H \mathbf{A}_m^H \right|}{\sum_{k=1}^K \mu_{k,eq} p_k + P_{c,eq}} - \frac{(\nabla_{\mathbf{p}} F(\bar{\mathbf{p}}))^T (\mathbf{p} - \bar{\mathbf{p}})}{\sum_{k=1}^K \mu_{k,eq} p_k + P_{c,eq}}, \end{aligned}$$

wherein $\bar{\mathbf{p}}$ is the point around which the first-order Taylor expansion is computed, and, for any $i = 1, \dots, K$, it holds

$$\frac{\partial F}{\partial p_i} = \sum_{k \neq i} \text{tr} \left(\left(\mathbf{W} + \sum_{m \neq k} p_m \mathbf{A}_m \gamma \gamma^H \mathbf{A}_m^H \right)^{-1} \mathbf{A}_i \gamma \gamma^H \mathbf{A}_i^H \right).$$

3) *Convergence and complexity*: Both methods are provably convergent, which follows from the convergence of the alternating optimization and sequential optimization methods [22]. Moreover, they both enjoy polynomial complexity in the number of variables. Specifically, the computational complexity of the method from Section⁴ III is $\mathcal{C} = \mathcal{O}(I(I_\gamma(N+1)^\alpha + I_p(K+1)^\beta))$, wherein I_γ and I_p are the number of iterations that the sequential method applied to γ and \mathbf{p} requires to converge; I is the number of iterations that the outer alternating optimization requires to converge; $(N+1)^\alpha$ is the complexity of each RIS optimization and $(K+1)^\beta$ is the complexity of each users' powers optimization⁵. A similar complexity analysis holds for the method from Section IV. Due to the involved mathematical structure of the problems at hand, the specific values of α and β , as well as of the iterations numbers I_γ , I_p , I , are not known in closed-form. Similarly, it is difficult to relate the choice of the initialization points to the convergence rate. Nevertheless, the theoretical convergence claim and complexity formulas remain valid for any choice of the initialization points.

V. NUMERICAL RESULTS

We consider an instance of the multi-user network described in Section II, with $K = 4$, $N_R = 4$, $N = 100$, $B = 20$ MHz, $P_0 = 40$ dBm, $P_{0,RIS} = 30$ dBm, $P_{c,n} = 0$ dBm, $P_{R,max} = 10$ dBW. The noise power spectral density is -174 dBm/Hz, and a noise figure of 10 dB has been considered. The mobile users are randomly placed in an area with radius 100 m around the RIS, and the base station is placed 50 m away from the RIS. The users have a random height in $[0, 5]$ m, while the RIS and base station have heights of 15 m and 10 m, respectively. A power decay factor $\eta = 4$ has been considered, while the fading component of all channels follows the Rice model, with factors $K_t = 4$ for the channel from the RIS to the base station and $K_r = 2$ for the channels from the mobile users to the RIS.

Fig. 2 shows the GEE achieved by:

- (a) the resource allocation obtained by maximizing the GEE through the method from Section III.
- (b) the resource allocation obtained by maximizing the GEE through the method from Section IV.
- (c) the resource allocation obtained by maximizing the spectral efficiency through the method from Section⁶ III.
- (d) the resource allocation obtained by maximizing the spectral efficiency through the method from Section IV.

⁴We neglect the complexity of computing the receive matrix \mathbf{C} , which can be accomplished by the closed-form expressions in Section III-3.

⁵Pseudo-concave maximizations with n variables can be restated as concave maximizations with $n+1$ variables [19], and the complexity of a concave maximization is polynomial in the number of variables with exponent in the interval $[1, 4]$. Thus, α and β belong to the interval $[1, 4]$ [23].

⁶Both the methods from Sections III and IV can be specialized to perform sum-rate maximization by simply setting $\mu_k = 0$, for all $k = 1, \dots, K$.

- (e) uniform power allocation and random RIS phases.

As anticipated, the method from Section IV significantly outperforms the approach from Section III, thanks to the exploitation of the mathematical structure of the optimal receive filter, rather than simply updating it within the alternating maximization algorithm. Moreover, a huge gain is obtained compared to Case (e), in which the system radio resources are not optimized. As expected, the value of the GEE saturates after the point at which P_{max} is large enough to allow reaching the maximum of the GEE function. Indeed, after this point, changing the resource allocation would only decrease the GEE.

Fig. 3 compares the GEE obtained by the GEE-maximizing method from Section IV to that obtained by a passive RIS versus the power consumption $P_{c,n}$ of each active RIS element. As for the passive RIS scenario, the optimization is performed as a special case of the algorithm from Section IV, by setting $P_{R,max} = 0$, which leads to setting $|\gamma_n| = 1$ for all $n = 1, \dots, N$, and leaves the phases of the RIS reflection coefficients as optimization variables. Also, for the passive RIS case, fixed values of $P_{c,n}^{(p)} = 0$ dBm and $P_{0,RIS}^{(p)} = 20$ dBm have been considered. The results show the operating regime in which the active RIS is more energy-efficient than the passive one. Eventually, as $P_{c,n}$ of each active RIS element increases, the passive RIS becomes more energy-efficient than the active one. Moreover, the crossing point is encountered for lower values of $P_{c,n}$ when the RIS is equipped with more active elements N . Thus, there is a trade-off between the GEE and the number of active RIS elements.

Finally, Table I analyzes the convergence time of the proposed methods. Let us denote by $T_{1,p}$ and $T_{2,p}$ the convergence time of the algorithms from Sections III and IV when applied with a passive RIS. Similarly, let us denote by $T_{1,a}$ and $T_{2,a}$ the convergence time of the algorithms from Sections III and IV when applied with an active RIS. Simulations have been performed by a workstation equipped with an AMD Ryzen 9, 5950x, 16-core processor. All convex optimization problems have been solved the CVX software package. Convergence is declared when the absolute error between the GEE values in two successive iterations is lower than $\epsilon = 10^{-6}$. The results show that applying the proposed methods with active RIS can take longer than with a passive RIS. This is expected since the formulation of the active scenario leads to a more general and complex mathematical expression of the GEE and problem constraints. Similarly, the use of the method from Section IV can take longer than the one from Section III, which is also expected due to the more involved expression of the objective function that we obtain by embedding the optimal expression of the receive filters into the SINR formula. This leads to a larger number of iterations required for convergence and, possibly, to larger values of the exponents α and β in the computational complexity formulas derived in Section IV-3. Thus, there is a trade-off between the first and second proposed methods. The former achieves a lower GEE value, but requires a lower computational complexity than the latter.

TABLE I
COMPARISON BETWEEN CONVERGENCE TIME OF THE PROPOSED ALGORITHMS WITH ACTIVE AND PASSIVE RIS.

P_{max} [dBW]	$T_{1,a}/T_{1,p}$	$T_{2,a}/T_{2,p}$	$T_{2,p}/T_{1,p}$	$T_{2,a}/T_{1,a}$
-40	2.26	1.88	12.58	10.45
-20	3.34	2.58	17.49	13.52
0	6.13	1.90	26.67	20.64
20	11.98	8.67	40.45	29.26

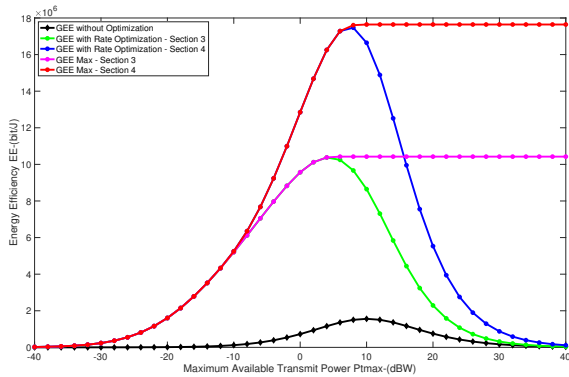


Fig. 2. Achieved GEE versus P_{max} . $K = 4$, $N_R = 4$, $N = 100$.

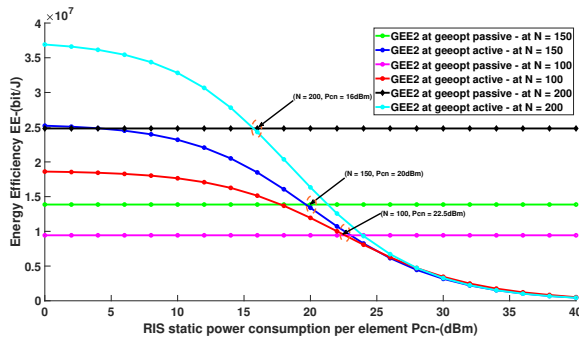


Fig. 3. Achieved GEE versus $P_{c,r}$ for the active and passive RIS scenarios. $K = 4$, $N_R = 4$, $N = 100, 150, 200$.

VI. CONCLUSIONS

This work has proposed two provably convergent algorithms with polynomial complexity for EE maximization in a wireless network aided by an active RIS. The RIS coefficients, the mobile users' transmit powers, and linear receive filters have been jointly optimized. Numerical results show the merits of the proposed algorithms and highlight a trade-off between the EE of active and passive RISs, in terms of the number of RIS elements and the additional power consumption due to the presence of the active-load hardware.

REFERENCES

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

[2] Q. Wu and R. Zhang, "Towards smart and reconfigurable environment: Intelligent reflecting surface aided wireless network," *IEEE Communications Magazine*, 2019.

[3] M. Di Renzo *et al.*, "Smart radio environments empowered by reconfigurable AI meta-surfaces: An idea whose time has come," *EURASIP Journal on Wireless Communications and Networking*, vol. 129, 2019.

[4] C. Huang *et al.*, "Holographic MIMO surfaces for 6G wireless networks: Opportunities, challenges, and trends," *IEEE Wireless Communications*, vol. 27, no. 5, pp. 118–125, 2019.

[5] D. López-Pérez, A. De Domenico, N. Piovesan, G. Xinli, H. Bao, S. Qitao *et al.*, "A survey on 5G radio access network energy efficiency: Massive MIMO, lean carrier design, sleep modes, and machine learning," *IEEE Communications Surveys Tutorials*, 2022.

[6] M. Dajer, Z. Ma, L. Piazzi, N. Prasad, X.-F. Qi, B. Sheen, J. Yang, and G. Yue, "Reconfigurable intelligent surface: design the channel – a new opportunity for future wireless networks," *Digital Communications and Networks*, vol. 8, no. 2, pp. 87–104, 2022. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2352864821000912>

[7] L. Dong, H.-M. Wang, and J. Bai, "Active reconfigurable intelligent surface aided secure transmission," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 2, pp. 2181–2186, 2022.

[8] R. Long, Y.-C. Liang, Y. Pei, and E. G. Larsson, "Active reconfigurable intelligent surface-aided wireless communications," *IEEE Transactions on Wireless Communications*, vol. 20, no. 8, pp. 4962–4975, 2021.

[9] Z. Yigit, E. Basar, M. Wen, and I. Altunbas, "Hybrid reflection modulation," 11 2021.

[10] S. Khaledian, F. Farzami, H. Soury, B. Smida, and D. Erricolo, "Active two-way backscatter modulation: An analytical study," *IEEE Transactions on Wireless Communications*, vol. 18, no. 3, pp. 1874–1886, 2019.

[11] 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?" 2021. [Online]. Available: <https://arxiv.org/abs/2103.15154>

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

[13] R. Schroeder, J. He, and M. Juntti, "Passive ris vs. hybrid ris: A comparative study on channel estimation," in *2021 IEEE 93rd Vehicular Technology Conference (VTC2021-Spring)*, 2021, pp. 1–7.

[14] R. A. Tasci, F. Kilinc, E. Basar, and G. C. Alexandropoulos, "A new ris architecture with a single power amplifier: Energy efficiency and error performance analysis," *IEEE Access*, vol. 10, pp. 44 804–44 815, 2022.

[15] Z. Peng, X. Liu, C. Pan, L. Li, and J. Wang, "Multi-pair D2D communications aided by an active RIS over spatially correlated channels with phase noise," *IEEE Communications Letters*, vol. 11, no. 10, pp. 2090–2094, 2022.

[16] K. Zhi, C. Pan, H. Ren, K. K. Chai, and M. Elkashlan, "Active RIS versus passive RIS: Which is superior with the same power budget?" *IEEE Communications Letters*, vol. 26, no. 5, pp. 1150–1154, 2022.

[17] K. Liu *et al.*, "Active reconfigurable intelligent surface: Fully-connected or sub-connected?" *IEEE Comm. Lett.*, vol. 26, no. 1, 2022.

[18] M. Di Renzo, F. Danufane, and S. Tretyakov, "Communication models for reconfigurable intelligent surfaces: From surface electromagnetics to wireless networks optimization," *Proceedings of the IEEE*, vol. 110, no. 9, pp. 1164–1209, 2022.

[19] A. Zappone and E. A. Jorswieck, "Energy efficiency in wireless networks via fractional programming theory," *Found. and Trends® in Commun. and Inf. Theory*, vol. 11, no. 3-4, pp. 185–396, 2015.

[20] B. R. Marks and G. P. Wright, "A general inner approximation algorithm for non-convex mathematical programs," *Operations Research*, vol. 26, no. 4, pp. 681–683, 1978.

[21] Z.-Q. Luo, W.-K. Ma, A. M.-C. So, Y. Ye, and S. Zhang, "Semidefinite relaxation of quadratic optimization problems," *IEEE Signal Processing Magazine*, vol. 27, no. 3, pp. 20–34, 2010.

[22] A. Aubry, A. D. Maio, A. Zappone, M. Razaviyayn, and Z.-Q. Luo, "A new sequential optimization procedure and its applications to resource allocation for wireless systems," *IEEE Transactions on Signal Processing*, vol. 66, no. 24, pp. 6518 – 6533, 2018.

[23] R. P. Brent and P. Zimmermann, *Modern Computer Arithmetic*. Cambridge University Press, 2010.