# Active Reconfigurable Intelligent Surface: Fully-Connected or Sub-Connected?

Kunzan Liu, Zijian Zhang, Linglong Dai, Shenheng Xu, and Fan Yang

***Abstract*— To overcome the "multiplicative fading effect" introduced by passive reconfigurable intelligent surface (RIS), the concept of active RIS has been recently proposed to amplify the radiated signals. However, the existing fully-connected architecture of active RIS consumes high power due to the additionally integrated active components. To address this issue, we propose the sub-connected architecture of active RIS. Different from fully-connected architecture, where each element integrates a dedicated power amplifier, in the sub-connected architecture, multiple elements control their phase shifts independently but share a same power amplifier, which significantly reduces the number of power amplifiers for power saving at the cost of fewer degrees of freedom (DoFs) for beamforming design. Fortunately, our analysis reveals that performance loss introduced by the sub-connected architecture is slight, indicating that it can achieve much higher energy efficiency (EE). Furthermore, we formulate the EE maximization problem in the active RIS-aided system for both architectures and develop a corresponding joint beamforming design. Simulation results verify the proposed sub-connected architecture as an energy-efficient realization of active RIS.**

***Index Terms*— Reconfigurable intelligent surface (RIS), sub-connected architecture, energy efficiency (EE), beamforming design.**

## I. Introduction

RECONFIGURABLE intelligent surface (RIS) has been envisioned as a promising candidate for future wireless communications. Composed of massive elements with controllable phase shifts, RIS is able to reconfigure wireless channels by reflecting the incident electromagnetic waves towards desired directions with high array gains. Benefiting from its low cost and power consumption, RIS has the potential to overcome blockages, improve channel capacity, and reduce transmit power for future communications [1]–[3].

However, RIS introduces the "multiplicative fading effect", i.e., the pathloss of the transmitter-RIS-receiver link is the multiplication (instead of the sum) of that of the transmitter-RIS link and RIS-receiver link, which makes the gain of reflection link much smaller than that of the direct link. This effect will result in a fatal problem that the expected capacity gain by employing RIS can hardly be achieved in many scenarios with strong direct links [4]. However, this problem is usually bypassed in many existing works by assuming that the direct link is blocked or very weak [2], which may not be accurate in practice. To overcome this "multiplicative fading effect", a novel concept called *active RIS* has been recently proposed, which can substantially improve the capacity regardless of whether the direct link is weak or not [4]. Specifically, in contrast to the existing passive RIS that reflects the signal only with phase-shift control, active RIS additionally integrates a power amplifier in each RIS element, which can actively amplify the reflected signals. Fundamentally, it has been proved that active RIS can transform the multiplicative fading into the additive fading [5]. Moreover, an active RIS element was fabricated and tested to verify its unique signal model [4]. In addition, the optimization of active RIS placement was investigated in [6], which indicates that active RIS can perform better than passive RIS under optimized placement in most practical scenarios. However, in existing contributions investigating active RIS, each active RIS element owns a dedicated power amplifier (we refer to it as the *fully-connected* architecture in this letter), which may result in high power consumption when the number of RIS elements is large. For instance, if we count the required power for a power amplifier as 10 mW [7], a 1000-element active RIS will consume 10 W, which approaches the typical transmit power of a base station (BS) [8].

To reduce the high power consumption of active RIS, we propose the *sub-connected* architecture as a novel realization of active RIS in this letter.[1] In contrast to the existing fully-connected architecture, where each RIS element is independently controlled by a phase-shift circuit and a power amplifier, in the sub-connected architecture, multiple RIS elements in a sub-array control their phase shifts independently but only share one power amplifier. In this way, the number of required power amplifiers can be significantly reduced to save power, but at the cost of fewer controllable degrees of freedom (DoFs) for beamforming design. Fortunately, our performance analysis reveals that, the performance loss of the proposed sub-connected architecture is slight, which enables the sub-connected architecture to enjoy a higher energy efficiency (EE). Furthermore, we formulate the EE maximization problem for both architectures in an active RIS-aided multiple-input multiple-output (MIMO) system, and develop a joint beamforming design to solve this problem by exploiting fractional programming and alternating optimization. Simulation results verify the benefits of the proposed sub-connected architecture as an energy-efficient realization of active RIS.[2]

*Notation:* $\mathbb{C}$ and $\mathbb{R}_+$ denote the set of complex and positive real numbers, respectively; $[L]$ represents the set of

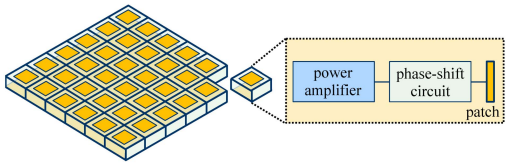
Manuscript received September 21, 2021; revised October 5, 2021; accepted October 10, 2021. Date of publication October 13, 2021; date of current version January 10, 2022. This work was supported in part by the National Key Research and Development Program of China (Grant No. 2020YFB1807201), in part by the National Natural Science Foundation of China (Grant No. 62031019), and in part by the European Commission through the H2020-MSCA-ITN META WIRELESS Research Project (Grant No. 956256). The associate editor coordinating the review of this letter and approving it for publication was Y. Deng. *(Corresponding author: Linglong Dai.)*

The authors are with the Beijing National Research Center for Information Science and Technology (BNRist), Tsinghua University, Beijing 100084, China, and also with the Department of Electronic Engineering, Tsinghua University, Beijing 100084, China (e-mail: lkz18@mails.tsinghua.edu.cn; zhangzj20@mails.tsinghua.edu.cn; daill@tsinghua.edu.cn; shxu@tsinghua.edu.cn; fan_yang@tsinghua.edu.cn).

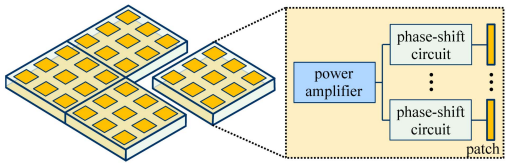
Digital Object Identifier 10.1109/LCOMM.2021.3119696

[1]Simulation codes are provided to reproduce the results in this letter: http://oa.ee.tsinghua.edu.cn/dailinglong/publications/publications.html

[2]This kind of sub-connected architecture has also been widely used for hybrid precoding in millimeter-wave (mmWave) communication system.




(a) Active RIS with existing fully-connected architecture.




(b) Active RIS with proposed sub-connected architecture.

Fig. 1. Comparison between different architectures for active RIS.

integers $\{1, 2, \ldots, L\}$; $\boldsymbol{X}^*, \boldsymbol{X}^T, \boldsymbol{X}^H$, and $\boldsymbol{X}^\dagger$ denote the conjugate, transpose, conjugate transpose, and pseudo inverse of matrix $\boldsymbol{X}$, respectively; $\|\boldsymbol{X}\|$ denotes the Frobenius norm of matrix $\boldsymbol{X}$; $\otimes$ and $\odot$ denote the Kronecker and Hadamard product, respectively; $\mathrm{diag}(\cdot)$ is the diagonal operation; $\mathbf{0}_L$ and $\mathbf{1}_L$ denotes the $L \times 1$ zero vector and all-ones vector, respectively.

## II. System Model

In this section, we propose the sub-connected architecture for active RIS. Then, we consider a general signal model to investigate the EE of the active RIS-aided MIMO system, and provide analysis on the sub-connected architecture.

### A. The Proposed Sub-Connected Architecture for Active RIS

Active RIS has been recently proposed to overcome the "multiplicative fading effect". In contrast to existing relays that integrate RF chains and have the signal processing ability, active RISs only utilize power amplifiers and phase-shift circuits that reflect and control the signals [4]. The investigation of active RIS provides good supports to the general concept of holographic MIMO surfaces [9].

Specifically, as shown in Fig. 1 (a), for the existing architecture of active RIS, despite the phase-shift circuit for phase-shift control in passive RIS, each RIS element also independently integrates a power amplifier to amplify the radiated signals. Intuitively, we call this architecture as the fully-connected architecture due to the pairwise appearance of the phase-shift circuit and the power amplifier. However, when the number of RIS elements is large, the fully-connected architecture would consume much power due to the large number of power amplifiers required. To reduce the high power consumption of active RIS, we propose the sub-connected architecture for active RIS, as shown in Fig. 1 (b). Different from the fully-connected architecture, multiple RIS elements in a sub-array are served by different phase-shift circuits but a same power amplifier, and thus these RIS elements can control the phase shifts independently but share the same amplification factor.[3] As another benefit, the sub-connected architecture also avoids the bulky circuits in practical implementation, since the hardware components are partially removed.

For comparison, let us assume each power amplifier is used to serve $T$ RIS elements in the sub-connected architecture. Then, the existing fully-connected architecture can be interpreted as a special case of the sub-connected architecture when $T = 1$. Moreover, with the same number of RIS elements, the number of power amplifiers in the sub-connected architecture becomes only $1/T$ of that of the fully-connected architecture, so the power consumed by power amplifiers can be significantly reduced. Following the example in Section I and assuming that $T = 10$, the 1000-element active RIS with the fully-connected architecture will consume 10 W for the required power of power amplifiers, while the sub-connected architecture can reduce this value to only 1 W. Meanwhile, as a trade-off, the number of controllable DoFs of active RIS also decreases, which may degrade the capability of active RIS to some extent and will be investigated in the following.

[3]For hardware implementation, circulators can be used to separately amplify and reflect the signals from a same power amplifier.

### B. Signal Model

Let us consider an active RIS-aided MIMO system, where an $N$-element active RIS assists the transmission from an $M$-antenna BS to $K$ single-antenna users. Specifically, for both architectures, we denote $L = N/T$ as the number of required power amplifiers. Then, the precoding matrix $\boldsymbol{\Psi} \in \mathbb{C}^{N\times N}$ at the active RIS can be written as $\boldsymbol{\Psi} = \mathrm{diag}(\boldsymbol{\psi}) = \mathrm{diag}(\boldsymbol{\Theta}\boldsymbol{\Gamma}\boldsymbol{a})$, where $\boldsymbol{\Theta} \in \mathbb{C}^{N\times N}$ denotes the phase-shift diagonal matrix, which is the same as that for passive RIS [2], and $\boldsymbol{a} \in \mathbb{R}_+^{L\times 1}$ denotes the amplification factor vector. Moreover, $\boldsymbol{\Gamma} \in \mathbb{C}^{N\times L}$ is defined as an indicator matrix representing the connection relationship between the power amplifiers and the phase-shift circuits. Without loss of generality, we assume $\boldsymbol{\Gamma} = \boldsymbol{I}_L \otimes \mathbf{1}_T$.

Furthermore, let $\boldsymbol{h}_k \in \mathbb{C}^{M\times 1}$, $\boldsymbol{G} \in \mathbb{C}^{N\times M}$, and $\boldsymbol{f}_k \in \mathbb{C}^{N\times 1}$ denote the channels from the BS to user $k$, from the BS to the active RIS, and from the active RIS to user $k$, respectively. Then, the signal $y_k$ received at user $k$ can be modeled as

$$y_k = \left(\boldsymbol{h}_k^H + \boldsymbol{f}_k^H \boldsymbol{\Psi}\boldsymbol{G}\right) \sum\nolimits_{j=1}^{K} \boldsymbol{w}_j s_j + \boldsymbol{f}_k^H \boldsymbol{\Psi}\boldsymbol{z} + n_k, \quad (1)$$

where $s_k$ is the normalized transmitted symbol for user $k$ and $\boldsymbol{w}_k \in \mathbb{C}^{M\times 1}$ is the corresponding beamforming vector at the BS; $\boldsymbol{z} \sim \mathcal{CN}\left(\mathbf{0}_N, \sigma_z^2 \boldsymbol{I}_N\right)$ and $n_k \sim \mathcal{CN}\left(0, \sigma^2\right)$ denote the introduced dynamic noise at the active RIS and the additive white Gaussian noise (AWGN) at user $k$, respectively.

From (1), the signal-to-interference-plus-noise ratio (SINR) at user $k$ can be denoted as

$$\mathrm{SINR}_k = \frac{\left|\boldsymbol{H}_k^H \boldsymbol{w}_k\right|^2}{\sum_{j=1, j\neq k}^{K} \left|\boldsymbol{H}_k^H \boldsymbol{w}_j\right|^2 + \left\|\boldsymbol{f}_k^H \boldsymbol{\Psi}\right\|^2 \sigma_z^2 + \sigma^2}, \quad (2)$$

where $\boldsymbol{H}_k^H = \boldsymbol{h}_k^H + \boldsymbol{f}_k^H \boldsymbol{\Psi}\boldsymbol{G}$ is the equivalent channel from the BS to user $k$. Then, the spectrum efficiency (SE) of this system is given by

$$R = \sum\nolimits_{k=1}^{K} \log_2\left(1 + \mathrm{SINR}_k\right). \quad (3)$$

As for the power consumption, it comprises the transmit power of the BS and RIS, as well as the power consumed by all components of the system, which is expressed as

$$P = \xi \sum\nolimits_{k=1}^{K} \|\boldsymbol{w}_k\|^2 + \zeta\left(\sum\nolimits_{k=1}^{K} \|\boldsymbol{\Psi}\boldsymbol{G}\boldsymbol{w}_k\|^2 + \|\boldsymbol{\Psi}\|^2 \sigma_z^2\right) + KW_{\mathrm{U}} + W_{\mathrm{BS}} + NW_{\mathrm{PS}} + LW_{\mathrm{PA}}, \quad (4)$$

where $\xi$ and $\zeta$ are the inverse of energy conversion coefficients at the BS and the active RIS, respectively; $W_{\mathrm{U}}$ and $W_{\mathrm{BS}}$

denote the dissipated power consumed at each user and the BS, respectively; $W_{\mathrm{PS}}$ and $W_{\mathrm{PA}}$ constitutes the hardware static power of active RIS, corresponding to phase-shift circuit and power amplifier, respectively. Note that we adopt an ideal assumption, that the power amplifier operates in its linear region with no constraints on the incident signal power. We also assume that the amplification factor can be any positive number bigger or smaller than 1, which is independent of the incident signal power[4] [8], [10].

In this letter, we focus on the EE of this system, which is defined as $\eta = R/P$. To illustrate the superiority of the proposed sub-connected architecture of active RIS, we will provide theoretical analysis in the following Subsection II-C.

### C. Performance Analysis

To observe more insights, similar to [4], we analyze the EE by considering a simplified single-input single-output (SISO) system. Focusing on the capability of the sub-connected architecture, we also adopt the assumptions in [4] and [5] that, the direct link is neglected and reflection links are line-of-sight (LoS). Our result is stated in the following **Lemma 1**.

**Lemma 1:** *The maximum EE of the active RIS-aided SISO system is achieved by the optimal amplification factors* $a_1 = \cdots = a_L$ *with any given* $T$.

*Proof:* See Appendix A. ∎

The reason behind **Lemma 1** is that, the amplitudes of channels associated with all RIS elements are the same in the far-field. Thus, compared with fully-connected architecture, the reduced DoFs in sub-connected architecture has no impact on beamforming design.[5] Particularly, in SISO system, since the optimal amplification factors are the same for all elements, only $L = 1$ power amplifier is sufficient for active RIS, which can significantly reduce the power consumption and indicate the practicability of the proposed sub-connected architecture.

For further investigations on active RIS, based on **Lemma 1**, we provide asymptotic analysis stated in **Lemma 2**.

**Lemma 2:** *As* $N \to \infty$, *let us denote the asymptotic signal-to-noise ratio (SNR) in SISO system assisted by active RIS and full-duplex amplify-and-forward (FD-AF) relay as* $\gamma_{\text{active RIS}}$ and $\gamma_{\mathrm{FD-AF}}$, *respectively. It holds that* $\gamma_{\text{active RIS}} \approx \frac{\pi^2}{16}\gamma_{\mathrm{FD-AF}}$.

*Proof:* See Appendix B. ∎

**Lemma 2** demonstrates that, with the optimal amplification factors described in **Lemma 1** and the asymptotically large $N$, the SNR in active RIS-aided system is approximately $\frac{\pi^2}{16}$ of that of the FD-AF-aided system. However, the FD-AF integrates energy-hungry radio-frequency (RF) chains (e.g. 250 mW per RF chain [11]), which makes active RIS with the sub-connected architecture more energy-efficient, since only $L = 1$ power amplifier and phase-shift circuits (e.g. 10 mW per circuit [8]) are utilized.

## III. Joint Beamforming Design

In this section, we extend our discussions to the more general MIMO case. Specifically, we formulate the EE maximization problem for both architectures, and then propose a joint beamforming design to tackle the non-convexity of the formulated problem.

### A. Problem Formulation

For notation simplicity, let us denote $\boldsymbol{W} = \left[\boldsymbol{w}_1^T, \ldots, \boldsymbol{w}_K^T\right]^T$ and $\boldsymbol{\Theta} = \mathrm{diag}\,(\boldsymbol{\theta})$. Based on the developed signal model in Subsection II-B, the EE maximization problem can be formulated as

$$\max_{\boldsymbol{W},\boldsymbol{\Theta},\boldsymbol{a}} \quad \eta = \frac{R}{P} \tag{5a}$$

$$\text{s.t. } \mathrm{C}_1: \ \xi \sum\nolimits_{k=1}^{K} \|\boldsymbol{w}_k\|^2 + W_{\mathrm{BS}} \le P_{\mathrm{BS}}^{\max}, \tag{5b}$$

$$\mathrm{C}_2: \ \zeta\left(\sum\nolimits_{k=1}^{K} \|\boldsymbol{\Psi}\boldsymbol{G}\boldsymbol{w}_k\|^2 + \|\boldsymbol{\Psi}\|^2 \sigma_z^2\right) + NW_{\mathrm{PS}} + LW_{\mathrm{PA}} \le P_{\mathrm{RIS}}^{\max}, \tag{5c}$$

$$\mathrm{C}_3: \ |\theta_n| = 1, \ \forall n \in [N], \tag{5d}$$

$$\mathrm{C}_4: \ a_l \ge 0, \ \forall l \in [L], \tag{5e}$$

where $\mathrm{C}_1$ and $\mathrm{C}_2$ guarantee the maximum total power consumed at the BS and the active RIS, denoted by $P_{\mathrm{BS}}^{\max}$ and $P_{\mathrm{RIS}}^{\max}$, respectively; $\mathrm{C}_3$ and $\mathrm{C}_4$ constrain the feasible sets of the phase-shift matrix $\boldsymbol{\Theta}$ and the amplification factor vector $\boldsymbol{a}$, respectively.

Due to the non-convexity of problem (5), we extend the algorithm in [2] for joint optimization, where the problem is firstly reformulated by fractional programming, and then solved by alternating optimization.

### B. Preprocessing

To tackle the fractional objective function (5a), we first apply the Dinkelbach's algorithm to transform (5a) into an equivalent form. Specifically, the optimal EE $\eta^{\mathrm{opt}}$ satisfies

$$\max_{\boldsymbol{W},\boldsymbol{\Theta},\boldsymbol{a}} \left(R - \eta^{\mathrm{opt}} P\right) = 0, \tag{6}$$

which indicates that the optimal $\eta^{\mathrm{opt}}$ can be solved by alternately optimizing

$$\max_{\boldsymbol{W},\boldsymbol{\Theta},\boldsymbol{a}} \quad f\left(\boldsymbol{W},\boldsymbol{\Theta},\boldsymbol{a}\right) = R - \eta P \quad \text{s.t. } \mathrm{C}_1, \ \mathrm{C}_2, \ \mathrm{C}_3, \ \mathrm{C}_4. \tag{7}$$

Since the new objective function $f$ in (7) is still nonconvex, by applying fractional programming, we can introduce the auxiliary variables $\boldsymbol{\mu} \in \mathbb{C}^{K\times 1}$ and $\boldsymbol{\nu} \in \mathbb{C}^{K\times 1}$ to equivalently reformulate the problem (7) as

$$\max_{\boldsymbol{W},\boldsymbol{\Theta},\boldsymbol{a},\boldsymbol{\mu},\boldsymbol{\nu}} \quad g\left(\boldsymbol{W},\boldsymbol{\Theta},\boldsymbol{a},\boldsymbol{\mu},\boldsymbol{\nu}\right) \quad \text{s.t. } \mathrm{C}_1, \ \mathrm{C}_2, \ \mathrm{C}_3, \ \mathrm{C}_4, \tag{8}$$

where

$$\begin{aligned} &g\left(\boldsymbol{W},\boldsymbol{\Theta},\boldsymbol{a},\boldsymbol{\mu},\boldsymbol{\nu}\right) \\ &= \sum\nolimits_{k=1}^{K}\Big[\ln\left(1+\mu_k\right) - \mu_k + 2\sqrt{1+\mu_k}\mathrm{Re}\left\{\nu_k^* \boldsymbol{H}_k^H \boldsymbol{w}_k\right\} \\ &\quad - |\nu_k|^2\left(\sum\nolimits_{j=1}^{K}\left|\boldsymbol{H}_k^H\boldsymbol{w}_j\right|^2 + \left\|\boldsymbol{f}_k^H\boldsymbol{\Psi}\right\|^2\sigma_z^2 + \sigma^2\right)\Big] - \eta P. \end{aligned}$$

Then, according to the convergence proved in [2], the locally optimal solution to problem (8) can be obtained by alternating optimization provided in the following Subsection III-C.

[4]When the amplification factor of active RIS is smaller than 1, active RIS still has a benefit compared with passive RIS, i.e., it can provide an extra DoF in amplitude for beamforming design.

[5]The optimal amplification factors for MIMO system do not have close-form expressions, and the analysis is left for future work.

**Algorithm 1** Proposed Joint Beamforming Design

**Inputs:** Channel matrices and system parameters.
**Output:** EE $\eta$ under optimized variables.
1: Initialize $\boldsymbol{W}$, $\boldsymbol{\Theta}$, $\boldsymbol{a}$, $\boldsymbol{\mu}$, and $\boldsymbol{\nu}$;
2: **while** no convergence of $f$ **do**
3: Update $\boldsymbol{\mu}^{\text{opt}}$ and $\boldsymbol{\nu}^{\text{opt}}$ by (9a) and (9b), respectively;
4: Update $\boldsymbol{W}^{\text{opt}}$ by solving (11);
5: Update $\boldsymbol{\Psi}^{\text{opt}}$ by solving (13);
6: Update $\boldsymbol{\Theta}^{\text{opt}}$ and $\boldsymbol{a}^{\text{opt}}$ by (15a) and (15b), respectively;
7: **end while**
8: Update $\eta$ by (5a) using optimized $\boldsymbol{W}$, $\boldsymbol{\Theta}$, and $\boldsymbol{a}$.

### C. Proposed Joint Beamforming Design

For clarity, we first summarize the proposed joint beamforming design in **Algorithm 1**, where the variables $\boldsymbol{W}$, $\boldsymbol{\Theta}$, $\boldsymbol{a}$, $\boldsymbol{\mu}$, and $\boldsymbol{\nu}$ are alternately optimized with the remained variables fixed. Specifically, the detailed closed-form expression of the optimal solution to problem (8) is given as follows.

*1) Optimal Auxiliary Variables:* For all $k \in [K]$, by setting $\partial g/\partial \mu_k$ and $\partial g/\partial \nu_k$ to 0, we obtain the optimal solution as

$$\mu_k^{\text{opt}} = \frac{\rho_k}{2}\left(\rho_k + \sqrt{\rho_k^2+4}\right), \tag{9a}$$

$$\nu_k^{\text{opt}} = \frac{\sqrt{1+\mu_k}\boldsymbol{H}_k^H\boldsymbol{w}_k}{\sum_{j=1}^K\left|\boldsymbol{H}_k^H\boldsymbol{w}_j\right|^2 + \left\|\boldsymbol{f}_k^H\boldsymbol{\Psi}\right\|^2\sigma_z^2+\sigma^2}, \tag{9b}$$

where $\rho_k = \text{Re}\left\{\nu_k^*\boldsymbol{H}_k^H\boldsymbol{w}_k\right\}$.

*2) Optimal BS Beamforming:* Let us denote

$$\widetilde{P}_{\text{BS}}^{\max} = \xi^{-1}\left(P_{\text{BS}}^{\max} - W_{\text{BS}}\right),$$
$$\widetilde{P}_{\text{RIS}}^{\max} = \zeta^{-1}\left(P_{\text{RIS}}^{\max} - NW_{\text{PS}} - LW_{\text{PA}}\right)$$

as the transmit power of the BS and the active RIS, respectively. For the optimal beamforming at the BS, the subproblem with respect to $\boldsymbol{W}$ can be rewritten as

$$\begin{aligned}\max_{\boldsymbol{W}}\ & \text{Re}\left\{\boldsymbol{u}^H\boldsymbol{W}\right\} - \boldsymbol{W}^H\boldsymbol{S}\boldsymbol{W}\\ \text{s.t. } \text{C}_1:\ & \boldsymbol{W}^H\boldsymbol{W} \le \widetilde{P}_{\text{BS}}^{\max},\\ \text{C}_2:\ & \boldsymbol{W}^H\boldsymbol{T}\boldsymbol{W} \le \widetilde{P}_{\text{RIS}}^{\max} - \|\boldsymbol{\Psi}\|^2\sigma_z^2,\end{aligned} \tag{11}$$

where

$$\boldsymbol{S} = \boldsymbol{I}_K \otimes \left(\eta\xi\boldsymbol{I}_M + \eta\zeta\boldsymbol{G}^H\boldsymbol{\Psi}^H\boldsymbol{\Psi}\boldsymbol{G} + \sum_{k=1}^K|\nu_k|^2\boldsymbol{H}_k\boldsymbol{H}_k^H\right),$$
$$\boldsymbol{u} = \left[\boldsymbol{u}_1^T,\ldots,\boldsymbol{u}_K^T\right]^T,\quad \boldsymbol{u}_k = 2\sqrt{1+\mu_k}\nu_k\boldsymbol{H}_k,$$
$$\boldsymbol{T} = \boldsymbol{I}_K \otimes \left(\boldsymbol{G}^H\boldsymbol{\Psi}^H\boldsymbol{\Psi}\boldsymbol{G}\right).$$

Observe that the problem (11) is a standard quadratic constraint quadratic programming (QCQP) problem, and thus the optimal solution $\boldsymbol{W}^{\text{opt}}$ can be solved by existing methods like alternating direction method of multipliers (ADMM).

*3) Optimal Active RIS Precoding:* For the optimal RIS precoding, for ease of notation, let us denote $\alpha_{k,j} = \boldsymbol{h}_k^H\boldsymbol{w}_j$ and $\boldsymbol{\beta}_j = \boldsymbol{G}\boldsymbol{w}_j$. Then, $\boldsymbol{H}_k^H\boldsymbol{w}_j$ can be rewritten as

$$\boldsymbol{H}_k^H\boldsymbol{w}_j = \alpha_{k,j} + \boldsymbol{f}_k^H\text{diag}\left(\boldsymbol{\beta}_j\right)\boldsymbol{\psi}. \tag{12}$$

By substituting (12), we can formulate the active RIS precoding optimization problem as

$$\begin{aligned}\max_{\boldsymbol{\Theta},\boldsymbol{a}}\ & \text{Re}\left\{\boldsymbol{\psi}^H\boldsymbol{v}\right\} - \boldsymbol{\psi}^H\boldsymbol{Q}\boldsymbol{\psi}\\ \text{s.t. } \text{C}_2:\ & \boldsymbol{\psi}^H\boldsymbol{R}\boldsymbol{\psi} \le \widetilde{P}_{\text{RIS}}^{\max},\\ & \text{C}_3,\ \text{C}_4,\end{aligned} \tag{13}$$

TABLE I
SIMULATION PARAMETERS

| Parameters | Values |
|---|---|
| BS position | $(0\,\text{m}, -20\,\text{m})$ |
| Active RIS position | $(100\,\text{m}, 5\,\text{m})$ |
| Dissipated power of each user $W_{\text{U}}$ | 10 dBm |
| Dissipated power of BS $W_{\text{BS}}$ | 6 dBW |
| Hardware static power of phase-shift circuit $W_{\text{PS}}$ | 10 dBm |
| Hardware static power of power amplifier $W_{\text{PA}}$ | 10 dBm |
| Maximum total power consumption $P_{\text{BS}}^{\max}$, $P_{\text{RIS}}^{\max}$ | 9 dBW |
| Inverse of energy conversion coefficients $\xi$, $\zeta$ | 1.1 |
| Noise power $\sigma^2$, $\sigma_z^2$ | $-80$ dBm |

where

$$\begin{aligned}\boldsymbol{Q} =& \sum_{k=1}^K\left(|\nu_k|^2\sigma_z^2\text{diag}\left(\boldsymbol{f}_k\odot\boldsymbol{f}_k^*\right) + \eta\zeta\text{diag}\left(\boldsymbol{\beta}_k\odot\boldsymbol{\beta}_k^*\right)\right)\\ &+\sum_{k=1}^K|\nu_k|^2\sum_{j=1}^K\text{diag}\left(\boldsymbol{\beta}_j^*\right)\boldsymbol{f}_k\boldsymbol{f}_k^H\text{diag}\left(\boldsymbol{\beta}_j\right) + \eta\zeta\sigma_z^2\boldsymbol{I}_N,\\ \boldsymbol{v} =& \sum_{k=1}^K\text{diag}\left(\boldsymbol{f}_k^H\right)\left(2\sqrt{1+\mu_k}\nu_k^*\boldsymbol{\beta}_k - |\nu_k|^2\sum_{j=1}^K\alpha_{k,j}^*\boldsymbol{\beta}_j\right),\\ \boldsymbol{R} =& \sum_{k=1}^K\text{diag}\left(\boldsymbol{\beta}_k\odot\boldsymbol{\beta}_k^*\right) + \sigma_z^2\boldsymbol{I}_N.\end{aligned}$$

Again, observe that the problem (13) is also a QCQP problem, so it can be solved by existing methods. By considering the constraints $\text{C}_3$ and $\text{C}_4$, the associated phase-shift matrix $\boldsymbol{\Theta}^{\text{opt}}$ and amplification factor vector $\boldsymbol{a}^{\text{opt}}$ are given by

$$\boldsymbol{\Theta}^{\text{opt}} = \text{diag}\left(\exp\left(j\arg\left(\boldsymbol{\psi}^{\text{opt}}\right)\right)\right), \tag{15a}$$

$$\boldsymbol{a}^{\text{opt}} = \boldsymbol{\Gamma}^{\dagger}\text{diag}\left(\exp\left(-j\arg\left(\boldsymbol{\psi}^{\text{opt}}\right)\right)\right)\boldsymbol{\psi}^{\text{opt}}. \tag{15b}$$

## IV. SIMULATION RESULTS

For simulations, we consider a 256-element passive RIS or active RIS with fully-connected/sub-connected architecture is used for the transmission from a 6-antenna BS to 4 mobile single-antenna users moving together along the $x$-axis. For comparison, we adopt the existing beamforming design for passive RIS-aided system [8], and assume the maximum total power consumptions of passive RIS and active RIS-aided system are the same. We also add a benchmark by extending the beamforming for the fully-connected active RIS in [4]. To extend the existing scheme [4] to our proposed sub-connected active RIS, we can modify it by forcing each amplification factor to be the average of all amplification factors in the sub-array that it belongs to. The other key simulation parameters are summarized in Table I, which is based on 3GPP propagation environments or experimental results [7], [8].

We show the SE and EE comparison versus user position in Fig. 2 (a) and (b), respectively. We can observe that both passive and active RIS can enhance the system. We also find that the SE of the proposed sub-connected architecture decreases by about 11% compared to the fully-connected architecture [4], while the EE has a 22% increment. We also depict the relationship between the SE and EE versus the maximum total power consumption $P_{\text{BS}}^{\max}$ with users fixed at $(150\,\text{m}, 0\,\text{m})$, as shown in Fig. 2 (c) and (d), respectively. The results also coincide with our previous analysis. In Table II, we further provide the maximized EE under different values of $T$ when the users are fixed at $(100\,\text{m}, 0\,\text{m})$, which shows

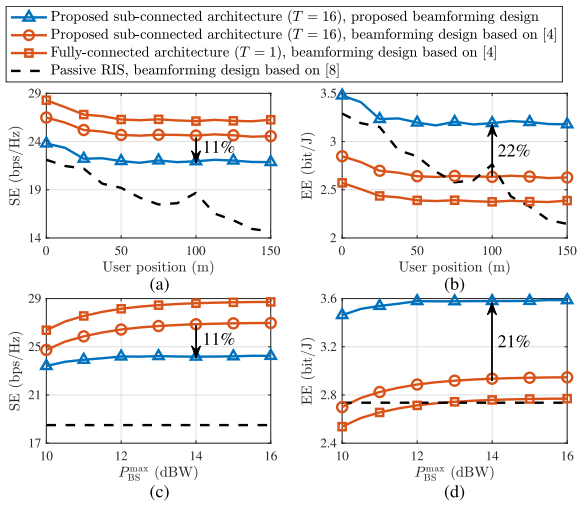



Fig. 2. SE and EE comparison.

TABLE II
EE UNDER DIFFERENT VALUES OF $T$

| **Value of $T$** | 1 | 4 | 16 | 64 | 256 |
|---|---|---|---|---|---|
| **EE (bit/J)** | 2.638 | 3.056 | 3.217 | 3.252 | 3.265 |

that the EE increases as $T$ increases, and the optimized $T$ in this simulation setting is $T^{\mathrm{opt}} = 256$. Thus, the proposed sub-connected architecture is preferred to realize a more energy-efficient active RIS.

## V. CONCLUSION

In this letter, we have proposed the sub-connected architecture to reduce the power consumption of active RIS. In the sub-connected architecture, multiple RIS elements share one power amplifier to reduce power consumption at the cost of fewer DoFs for beamforming design, while we have theoretically demonstrated that this performance loss is slight, which makes it enjoy a higher EE. Furthermore, we have formulated the EE maximization problem in the active RIS-aided system, and developed the corresponding joint beamforming design. Simulation results have shown that the EE of sub-connected architecture is $22\%$ higher than the fully-connected architecture, which verifies the proposed sub-connected architecture as an energy-efficient realization of active RIS.

## APPENDIX A
## PROOF OF LEMMA 1

In a SISO system, the channel matrices degenerate to vectors $\boldsymbol{g} \triangleq \boldsymbol{G} \in \mathbb{C}^{N\times 1}$ and $\boldsymbol{f} \triangleq \boldsymbol{f}_k \in \mathbb{C}^{N\times 1}$, and the beamforming vector degenerates to a scalar $w \triangleq \boldsymbol{w}_k$. Thus, the problem (5) can be reformulated as

$$\max_{w,\boldsymbol{\Theta},\boldsymbol{a}} \ \eta = \frac{\log_2\left(1+\frac{\left|\boldsymbol{f}^H\boldsymbol{\Psi}\boldsymbol{g}w\right|^2}{\left\|\boldsymbol{f}^H\boldsymbol{\Psi}\right\|^2\sigma_z^2+\sigma^2}\right)}{\xi\left|w\right|^2+\zeta\left(\left\|\boldsymbol{\Psi}\boldsymbol{g}w\right\|^2+\left\|\boldsymbol{\Psi}\right\|^2\sigma_z^2\right)+C_T} \tag{16a}$$

$$\text{s.t. } \mathrm{C}_1: \ \left|w\right|^2 \le \widetilde{P}_{\mathrm{BS}}^{\max}, \tag{16b}$$

$$\mathrm{C}_2: \ \left\|\boldsymbol{\Psi}\boldsymbol{g}w\right\|^2+\left\|\boldsymbol{\Psi}\right\|^2\sigma_z^2 \le \widetilde{P}_{\mathrm{RIS}}^{\max}, \tag{16c}$$

$$\mathrm{C}_3,\ \mathrm{C}_4,$$

where $C_T$ denotes a scalar with respect to $T$.

With the assumption that the reflection links are far-field and LoS links, the channels $\boldsymbol{g}$ and $\boldsymbol{f}$ can be expressed as

$$\boldsymbol{g} = \left|g_1\right|\mathbf{1}_N \odot \exp\left(j\arg\left(\boldsymbol{g}\right)\right), \quad \boldsymbol{f} = \left|f_1\right|\mathbf{1}_N \odot \exp\left(j\arg\left(\boldsymbol{f}\right)\right).$$

By exploiting Lagrangian dualities, we obtain

$$\left|w^{\mathrm{opt}}\right|^2 = \widetilde{P}_{\mathrm{BS}}^{\max}, \quad \left\|\boldsymbol{\Psi}^{\mathrm{opt}}\right\|^2 = \frac{\widetilde{P}_{\mathrm{RIS}}^{\max}}{\left|g_1\right|^2\widetilde{P}_{\mathrm{BS}}^{\max}+\sigma_z^2}. \tag{17}$$

Finally, using the Cauchy–Schwarz inequality, we arrive at

$$\left|\boldsymbol{f}^H\boldsymbol{\Psi}\boldsymbol{g}\right|^2 = \left|\sum\nolimits_{n=1}^{N} f_n^* g_n \psi_n\right|^2 \le \left\|\boldsymbol{\Psi}\right\|^2\left\|\boldsymbol{f}^*\odot\boldsymbol{g}\right\|^2, \tag{18}$$

with the equality holds if and only if

$$\boldsymbol{\Theta} = \mathrm{diag}\left(\exp\left(j\arg\left(\boldsymbol{f}-\boldsymbol{g}\right)\right)\right), \quad a_1 = \cdots = a_L. \tag{19}$$

## APPENDIX B
## PROOF OF LEMMA 2

For asymptotic analysis, let us assume $\boldsymbol{g} \sim \mathcal{CN}(\mathbf{0}_N, \varsigma_g^2\boldsymbol{I}_N)$ and $\boldsymbol{f} \sim \mathcal{CN}(\mathbf{0}_N, \varsigma_f^2\boldsymbol{I}_N)$. When $N \to \infty$, based on the amplification factors for RIS in **Lemma 1**, we have [4]

$$\gamma_{\text{active RIS}} \to \frac{\pi^2}{16}N\frac{\widetilde{P}_{\mathrm{BS}}^{\max}\widetilde{P}_{\mathrm{RIS}}^{\max}\varsigma_g^2\varsigma_f^2}{\widetilde{P}_{\mathrm{BS}}^{\max}\varsigma_g^2\sigma^2+\widetilde{P}_{\mathrm{RIS}}^{\max}\varsigma_f^2\sigma_z^2+\sigma^2\sigma_z^2}. \tag{20}$$

As for the FD-AF-aided SISO system, the asymptotic SNR is provided as [3, Prop. 3]

$$\gamma_{\text{FD-AF}} \to \frac{N^2\widetilde{P}_{\mathrm{BS}}^{\max}\widetilde{P}_{\mathrm{RIS}}^{\max}\varsigma_g^2\varsigma_f^2}{N\widetilde{P}_{\mathrm{BS}}^{\max}\varsigma_g^2\sigma^2+N\widetilde{P}_{\mathrm{RIS}}^{\max}\varsigma_f^2\sigma_z^2+\sigma^2\sigma_z^2}. \tag{21}$$

Therefore,

$$\frac{\gamma_{\text{FD-AF}}}{\gamma_{\text{active RIS}}} = \frac{16}{\pi^2}\left(1+\frac{\sigma^2\sigma_z^2\left(1-1/N\right)}{\widetilde{P}_{\mathrm{BS}}^{\max}\varsigma_g^2\sigma^2+\widetilde{P}_{\mathrm{RIS}}^{\max}\varsigma_f^2\sigma_z^2+\sigma^2\sigma_z^2}\right) \approx \frac{16}{\pi^2}.$$

## REFERENCES

[1] E. Basar, M. Di Renzo, J. De Rosny, M. Debbah, M.-S. Alouini, and R. Zhang, "Wireless communications through reconfigurable intelligent surfaces," *IEEE Access*, vol. 7, pp. 116753–116773, 2019.
[2] Z. Zhang and L. Dai, "A joint precoding framework for wideband reconfigurable intelligent surface-aided cell-free network," *IEEE Trans. Signal Process.*, vol. 69, pp. 4085–4101, 2021.
[3] Q. Wu and R. Zhang, "Intelligent reflecting surface enhanced wireless network via joint active and passive beamforming," *IEEE Trans. Wireless Commun.*, vol. 18, no. 11, pp. 5394–5409, Nov. 2019.
[4] Z. Zhang *et al.*, "Active RIS vs. passive RIS: Which will prevail in 6G?" 2021, *arXiv:2103.15154*. [Online]. Available: http://arxiv.org/abs/2103.15154
[5] E. Basar and H. V. Poor, "Present and future of reconfigurable intelligent surface-empowered communications," 2021, *arXiv:2105.00671*. [Online]. Available: http://arxiv.org/abs/2105.00671
[6] C. You and R. Zhang, "Wireless communication aided by intelligent reflecting surface: Active or passive?" 2021, *arXiv:2106.10963*. [Online]. Available: http://arxiv.org/abs/2106.10963
[7] L. E. Larson, "Radio frequency integrated circuit technology for low-power wireless communications," *IEEE Pers. Commun.*, vol. 5, no. 3, pp. 11–19, Jun. 1998.
[8] C. Huang, A. Zappone, G. C. Alexandropoulos, M. Debbah, and C. Yuen, "Reconfigurable intelligent surfaces for energy efficiency in wireless communication," *IEEE Trans. Wireless Commun.*, vol. 18, no. 8, pp. 4157–4170, Aug. 2019.
[9] C. Huang *et al.*, "Holographic MIMO surfaces for 6G wireless networks: Opportunities, challenges, and trends," *IEEE Wireless Commun.*, vol. 27, no. 5, pp. 118–125, Oct. 2020.
[10] K. K. Kishor and S. V. Hum, "An amplifying reconfigurable reflectarray antenna," *IEEE Trans. Antennas Propag.*, vol. 60, no. 1, pp. 197–205, Jan. 2011.
[11] P. V. Amadori and C. Masouros, "Low RF-complexity millimeter-wave beamspace-MIMO systems by beam selection," *IEEE Trans. Commun.*, vol. 63, no. 6, pp. 2212–2223, Jun. 2015.