This article has been accepted for publication in IEEE Transactions on Wireless Communications. 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/TWC.2023.3247906

Joint Precoding for Active Intelligent Transmitting Surface Empowered Outdoor-to-Indoor Communication in mmWave Cellular Networks

Xie Xie, Chen He, *Member, IEEE*, Xue Ma, Feifei Gao, *Fellow, IEEE*, Zhu Han, *Fellow, IEEE*, and Z. Jane Wang, *Fellow, IEEE*

***Abstract*—Outdoor-to-indoor communications in millimeter-wave (mmWave) cellular networks have been one challenging research problem due to the severe attenuation and the high penetration loss caused by propagation characteristics of mmWave signals. We propose a viable solution to implement the outdoor-to-indoor mmWave communication with the aid of an active intelligent transmitting surface (active-ITS), where the active-ITS allows the incoming signal from an outdoor base station (BS) to pass through the surface and be received by indoor users (UEs) after shifting its phase and magnifying its amplitude. Then, the problem of joint precoding of the BS and active-ITS is investigated to maximize the weighted sum-rate (WSR) of the system. An efficient block coordinate descent (BCD) based algorithm is developed to solve it with the suboptimal solutions in nearly closed-forms. In addition, to reduce the size and hardware cost of active-ITSs, we provide a block-amplifying architecture to partially remove the circuit components for power-amplifying, where multiple transmissive-type elements (TEs) in each block share the same power amplifier. Simulations indicate that active-ITS has the potential of achieving a given performance with much fewer TEs compared to the passive-ITS under the same total system power consumption, which makes it suitable for application to the space-limited and aesthetic-needed scenario, and the performance degradation caused by the block-amplifying architecture is negligible.**

***Index Terms*—Active intelligent transmitting surfaces, millimeter-wave, power amplification.**

I. Introduction

Millimeter-wave (mmWave) cellular networks are capable of supplying the ever-increasing demand of rates for advanced fifth-generation communications thanks to their abundant available bandwidth. However, a fundamental challenge for mmWave communications is that the mmWave signal experience severe attenuation and a high penetration loss compared with the lower frequency bands, which makes mmWave signals highly vulnerable to obstacles [1]–[3]. The emerging technique of reconfigurable intelligent surfaces (RISs) have been proposed as a promising candidate for alleviating the unfavorable properties of mmWave signals. RIS is an ultra-thin metasurface comprising multiple programmable elements, which enables to achieve a high beamforming gain by smartly manipulating the incident signal for proactively customizing the radio propagation environment [4]–[7]. More importantly, RIS can significantly reduce the outage caused by the presence of random blockages by establishing virtual line-of-sight (LoS) links between base stations (BSs) and users (UEs), which can considerably enhance the reliability of mmWave communications, especially in harsh urban propagation environments [8]–[10]. These benefits have inspired a lot of work to investigate RIS-assisted mmWave communication networks and verify that RIS in favor of enhancing the signal strength, extending the service range, and improving the spectral- and energy-efficiency [10]–[14].

On the other hand, outdoor-to-indoor communication in mmWave cellular networks is a common communication scenario as most mobile data traffic is consumed indoors, however, which is challenging since BSs and UEs are located on the opposite side of building structures [15], e.g., walls and windows, while penetrating them leads a severe attenuation (measurements have shown that around 40 and 28 dB for tinted-glass and brick [16], respectively), which limits the feasibility for an outdoor BS communicates with indoor UEs inside the buildings. Relay-enabled system is a potential solution for outdoor-to-indoor mmWave communications but has some drawbacks, e.g., expensive hardware components and high signal processing complexity, while these shortcomings of relays can be overcome by utilizing the superiority of the RISs [17]–[19]. However, the widely studied RISs-assisted systems focused on traditional reflective-type RISs, also referred to intelligent reflecting surfaces (IRSs) [20], as objects hanging on walls or facades of buildings, which face a placement restriction: both BSs and UEs have to be located on the front side of an IRS, and hence IRSs can only achieve a half-space coverage and are not up to the outdoor-to-indoor communication. As a remedy, some works [21], [22] innovative deployed multiple IRSs in a cooperative multi-hop manner to bypass the obstacle. Nevertheless, this does not

This work is supported in part by the National Natural Science Foundation of China (NSFC) under Grant 62271392, in part by the Shaanxi Innovation Program under Grant 2023-CX-TD-04, and in part by the Shaanxi Key Research and Development Project under Grant 2021KWZ-07. The work of Zhu Han is partially supported by the U.S. National Science Foundation under Grant CNS-2107216, Grant CNS-2128368, Grant CMMI-2222810, Toyota and Amazon. *(Corresponding author: Chen He.)*

Xie Xie, Chen He, and Xue Ma are with the School of Information Science and Technology, Northwest University, Xi'an, 710069, China (e-mail: x_xie@stumail.nwu.edu.cn; chenhe@nwu.edu.cn; maxue@stumail.nwu.edu.cn).

Feifei Gao is with the Department of Automation, Tsinghua University, Beijing, 100084, China (e-mail: feifeigao@tsinghua.edu.cn).

Z. Han is with the Department of Electrical and Computer Engineering at the University of Houston, Houston, TX 77004 USA, and also with the Department of Computer Science and Engineering, Kyung Hee University, Seoul, South Korea (e-mail: hanzhu22@gmail.com).

Z. Jane Wang is with the Department of Electrical and Computer Engineering, The University of British Columbia, Vancouver, BC V6T 1Z4, Canada (e-mail: zjanew@ece.ubc.ca).

Authorized licensed use limited to: University of Houston. Downloaded on July 02,2023 at 13:19:36 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Wireless Communications. 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/TWC.2023.3247906

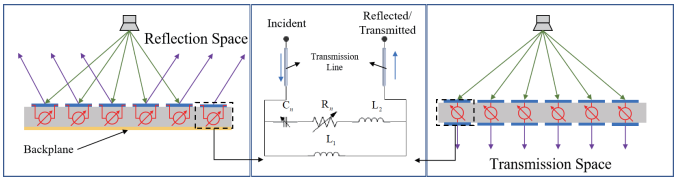

Fig. 1. The architecture schematic of the IRS and ITS.

apply to outdoor-to-indoor communications due to no matter where IRSs are placed, the signal after reflecting cannot bypass the wall to UEs behind the IRSs [17], [20].

To challenge this restriction and facilitate more flexible deployment of RIS, the novel concept of transmissive(refractive)-type RIS, also named intelligent transmitting surfaces (ITS), has been proposed [23]–[25]. ITS is a meta-material reflectionless surface structure, which is equipped with a large number of transmissive-type elements (TEs), and provides beam shaping, steering, and focus capabilities [25]. A key feature of ITS is that the incoming signal can pass through each TE, and then the refracted signal can be received by UEs behind the ITS. As shown in Fig. 1, the main difference between the architecture design of the IRS and ITS is that the reflective-type element (RE) of IRS reuses the patch for receiving and reflecting, while the TE consists of a pair of patches respectively for receiving and transmitting. In addition, the IRS needs a metal backplane to avoid signal energy leakage [14]. Both the RE and TE themselves can directly modify their responses by changing the corresponding tunable load [26]. Specifically, they can be accurately described by an equivalent lumped circuit model, where the waveguide can be regarded as a transmission line, L_1, L_2, C_n, and R_n denote the bottom layer inductance, top layer inductance, effective capacitance, and effective resistance, respectively. For a detailed analysis of this circuit model, please refer to [26], [27]. More importantly, different from the IRS that is hanging on walls or facades of buildings, ITSs can be flexibly deployed by coexisting with existing infrastructures in the middle of a communication environment, such as embedded in walls or windows between two different environments (outdoor-to-indoor, room-to-room, etc.) [17], [20], [28]. ITSs can fully unleash the potential of RISs on breaking the half-space (reflection space) limitation of signal propagation manipulated by IRS, and so that covers the backside space (transmission space). It is worth noting that by integrating IRS and ITS together, a highly flexible full-space manipulation of signal propagation can be achieved [23]. As an energy-efficient alternative to realize full-space coverage, the authors proposed a pair of novel prototypes, simultaneously transmitting and reflecting RISs (STAR-RISs) and reconfigurable intelligent Omni-surfaces (RIOSs) [28]–[30], where the surfaces can simultaneously transmit and reflect the incident signals. ITS or IRS can be regarded as a special full transmission or reflection mode of these concepts.

However, identically with the reflective cascaded channel, the transmissive cascaded channel in the ITS-empowered communication system will still suffer a multiplicative fading effect [31], which potentially causes the ITS to achieve a poor performance gain. A common approach to improve the performance gain of RISs is increasing the number of REs or TEs [32]. Intuitively speaking, this approach is feasible for the IRS but might not be practical for the ITS due to that the number of TEs is limited by the physical size of windows and the aesthetic appearance effect, which might be the price to be paid for that the ITS is embedded with existing building structures [20], [28]. Additionally, the substantial electromagnetic penetration loss when the mmWave signal impinges upon and penetrates the ITS is unavoidable [17]. These concerns imply that the performance gain reaped by the ITS might be much weaker than the IRS.

Inspired by the active-IRS [33]–[36] and active-RIOS [28], the deployment of an active-ITS will bring a notable leap forward for improving performance. The noticeable difference between the active- and passive- ITS is that the former can refract the incoming signals after amplifying amplitudes and shifting phases, rather than only refracting them after changing phases as done by the latter, which offers an extra degree of freedom (DoF) to reconfigure the incident signal. Identical to the principle of active REs, an active TE equips with an additional power amplifier for magnifying the signal amplitude [34]. As a result, the number of active TEs can be beneficially reduced compared with the passive-ITS with the same performance [36]–[38]. It should note that the fundamental operation mechanisms of the active-ITS and the conventional amplify-and-forward (AF) relay are quite different [35]. Specifically, the architecture design of the active-ITS is quite simple that does not need expensive and high power-consuming ratio-frequency (RF) chains, it directly refracts the incident signal with power amplification at the electromagnetic level [31]. In contrast, the conventional AF relays generally need RF chains to receive the signal first and then transmit it after power amplification [19]. Compared to the AF relay, although the active-ITS has no capability for signal processing, it has a shorter delay and allows us to avoid complex and expensive components to compensate for the multiplicative fading effect and achieve a similar performance [28].

Motivated by the above background, we aim to utilize the superiority of an active-ITS to alleviate the negative effects of the mmWave signal so that to implement the challenging but important outdoor-to-indoor communication system in the mmWave cellular network with high performance. Particularly, the main contributions are summarized as follows:

- To our best knowledge, this is the first work investigating outdoor-to-indoor mmWave communication with the aid of an active-ITS, where the incoming signal is refracted to the back side of the active-ITS after magnifying its amplitude and manipulating its phase.
- We investigate the joint optimization of the precoding matrices of both the BS and active-ITS with respect to the weighted sum-rate (WSR), and propose a suboptimal method to solve this non-convex problem. The original problem is first transformed into an equivalent form, then a block coordinate descent (BCD) based algorithm is developed to alternately solve the subproblems with respect to the linear precoding matrix of BS and transmissive coefficient of active-ITS.
- Particularly, the linear precoding matrix is optimized

Authorized licensed use limited to: University of Houston. Downloaded on July 02,2023 at 13:19:36 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Wireless Communications. 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/TWC.2023.3247906

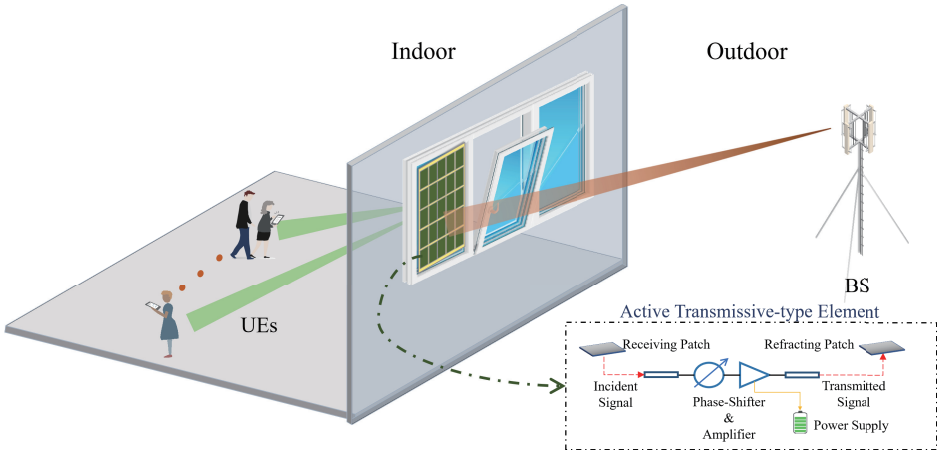

Fig. 2. A system model schematic of the active-ITS empowered outdoor-to-indoor mmWave communication.

by resorting to the dual-based method, while for the transmissive coefficient, by investigating its particular structures of the objective function and constraints, we decompose it into a pair of subproblems with respect to the amplification factor and phase-shifting. Then, an efficient iterative algorithm based on alternating optimization and price-mechanism is developed to solve them, which can significantly reduce the computational complexity.

- We extend the proposed algorithm so that it can be adopted into the block-amplifying ITS architecture with a low hardware complexity in which multiple TEs in each block share the same power amplifier.
- Simulation results demonstrate that the active-ITS can significantly improve the WSR performance of the outdoor-to-indoor mmWave communication system and the performance loss caused by the block-amplifying ITS architecture is negligible.

The paper is organized as follows. In Section II, we describe the active-ITS-empowered outdoor-to-indoor communications and formulate the WSR maximization problem. In Section III, we propose an efficient joint precoding algorithm to solve it. Then, in Section IV, the proposed algorithm is extended to solve the problem with the block-amplifying ITS architecture. Simulation results are presented in Section V, and finally, the paper is concluded in Section VI.

Notations: Scalars, vectors, and matrices are presented by lower-case, bold-face lower-case, and bold-face upper-case letters, respectively. $\mathcal{CN}(\mathbf{0}, \mathbf{I}_N)$ denotes the circularly symmetric complex Gaussian (CSCG) distribution with zero mean and covariance matrix $\mathbf{I}_N$, where $\mathbf{I}_N$ denotes an $N \times N$ identity matrix. $\circ$ denotes the Hadamard products. $\mathbf{A}^{\mathrm{H}}$ and $\mathrm{Tr}(\mathbf{A})$ denote the Hermitian and trace operators of matrix $\mathbf{A}$, respectively. $\Re\{a\}$ is the real part of a scalar a.

II. SYSTEM MODEL AND PROBLEM FORMULATION

As shown in Fig. 2, we consider the downlink of an active-ITS-empowered outdoor-to-indoor mmWave communication scenario. Thanks to the transmissive characteristic of the ITS, the incoming mmWave signal can pass through the surface when impinging upon it, and hence, the signal can be refracted from the outdoor BS to the indoor UEs through the ITS. To alleviate the severe attenuation of mmWave signals and greatly reduce the number of TEs for application to the space-limited window, we assume each active TE of the ITS can not only manipulate the phase but also magnify the amplitude of the incident signal with the aid of the power supply [34]. In addition, a smart controller is attached to the active-ITS and responsible to operate the active-ITS with the transmissive coefficients coordinated by the BS [23], [39]–[41].

A. Channel Model

Let M_t, M_r, and N denote the number of the transmitting antennas, the receiving antennas, and the TEs equipped by the BS, UEs, and active-ITS, respectively. The complicated uniform planar arrays (UPA) antenna configuration is employed at the BS, UEs, and active-ITS, which is more practical than uniform linear array (ULA) for RIS-assisted MIMO systems [12]. In addition, due to the low diffraction from objects of mmWave signals, mmWave channels are usually characterized according to the widely used Saleh-Valenzuela model [11]–[14]. Hence, the channel matrices from the BS to the active-ITS, from the active-ITS to the k-th UE, and from the BS to the k-th UE are denoted by $\mathbf{G} \in \mathbb{C}^{N \times M_t}$, $\mathbf{H}_k \in \mathbb{C}^{N \times M_r}$, and $\mathbf{B}_k \in \mathbb{C}^{M_t \times M_r}$, $\forall k$, respectively, which can be mathematically determined as follows

$$\mathbf{G} = \sqrt{\frac{M_t N}{L_{\mathrm{BI}}}} \sum_{l=1}^{L_{\mathrm{BI}}} \alpha_l \mathbf{a}_{\mathrm{ITS}}\left(\nu_l^{\mathrm{AOA}}, \vartheta_l^{\mathrm{AOA}}\right) \mathbf{a}_{\mathrm{BS}}^{\mathrm{H}}\left(\nu_l^{\mathrm{AOD}}, \vartheta_l^{\mathrm{AOD}}\right), \tag{1a}$$

$$\mathbf{H}_k = \sqrt{\frac{M_r N}{P_{\mathrm{IU}}}} \sum_{p=1}^{P_{\mathrm{IU}}} \beta_p \mathbf{a}_{\mathrm{UE}}\left(\nu_p^{\mathrm{AOA}}, \vartheta_p^{\mathrm{AOA}}\right) \mathbf{a}_{\mathrm{ITS}}^{\mathrm{H}}\left(\nu_p^{\mathrm{AOD}}, \vartheta_p^{\mathrm{AOD}}\right), \forall k, \tag{1b}$$

$$\mathbf{B}_k = \sqrt{\frac{M_r M_t}{Q_{\mathrm{BU}}}} \sum_{q=1}^{Q_{\mathrm{BU}}} \varpi_q \mathbf{a}_{\mathrm{UE}}\left(\vartheta_q^{\mathrm{AOA}}, \vartheta_q^{\mathrm{AOA}}\right) \mathbf{a}_{\mathrm{BS}}^{\mathrm{H}}\left(\nu_q^{\mathrm{AOD}}, \vartheta_q^{\mathrm{AOD}}\right), \forall k, \tag{1c}$$

where L_{BI}, P_{IU}, and Q_{BU} are the number of propagation paths between the BS and the active-ITS, between the active-

Authorized licensed use limited to: University of Houston. Downloaded on July 02,2023 at 13:19:36 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Wireless Communications. 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/TWC.2023.3247906

ITS and the k-th UE, and between the BS and the k-th UE, and where variables α_l, β_p, and ϖ_q denote the complex gains for links, with $l = 1$, $p = 1$, and $q = 1$ stand for the LoS path and the remaining are NLoS paths. $\mathbf{a}_{\text{BS}}\left(\nu_l^{\text{AOD}}, \vartheta_l^{\text{AOD}}\right)$ and $\mathbf{a}_{\text{ITS}}\left(\nu_p^{\text{AOD}}, \vartheta_p^{\text{AOD}}\right)$ are transmit array response vectors of the BS and the active-ITS, $\mathbf{a}_{\text{UE}}\left(\nu_l^{\text{AOA}}, \vartheta_l^{\text{AOA}}\right)$ and $\mathbf{a}_{\text{ITS}}\left(\nu_p^{\text{AOA}}, \vartheta_p^{\text{AOA}}\right)$ are receive array response vectors of the UEs and the active-ITS, respectively, where ν^{AOA} (ν^{AOD}) and ϑ^{AOA} (ϑ^{AOD}) denote the azimuth and elevation angles of arrival (departure), respectively. The typical array response vector for the UPA model can be mathematically expressed by

$$\mathbf{a}\left(\nu, \vartheta\right) = \frac{1}{\sqrt{HV}}\Big[1, \cdots, e^{j\frac{2\pi d}{\lambda}\left(h\sin(\vartheta)\sin(\nu)+v\cos(\vartheta)\right)}, \cdots, e^{j\frac{2\pi d}{\lambda}\left((H-1)\sin(\vartheta)\sin(\nu)+(V-1)\cos(\vartheta)\right)}\Big]^{\text{T}}, \quad (2)$$

where H and V denote the number of TEs (antennas) in the horizontal and vertical directions of the active-ITS (BS/UEs), and the element spacing is set to be $d = \lambda/2$. To unveil the theoretical upper bound of performance gain achieved by the active-ITS, we follow the assumption [18], [30], [39]–[41] that the full instantaneous channel state information (CSI) knowledge can be perfectly estimated [42] at the BS and all the calculations are performed at the BS.

B. Active-ITS Model

Similar to the widely-employed ideal active-IRS, we assume the ideal active-ITS that the phase and power factor can be controlled independently and continuously. Let $\mathbf{A} = \text{diag}\left(a_1, a_2, \cdots, a_N\right) \in \mathbb{R}^{N\times N}$ and $\mathbf{\Theta} = \text{diag}\left(e^{j\varphi_1}, e^{j\varphi_2}, \cdots, e^{j\varphi_N}\right) \in \mathbb{C}^{N\times N}$ represent the power amplification factor matrix and the transmissive phase-shifting matrix at the active-ITS with $\varphi_n \in [0, 2\pi), \forall n \in N$. It is worth noting that with the power amplification, a_n can be extended into greater than 1, and should be limited by a predetermined maximum value $a_n^{\max}$ that the amplifying circuit (active load) can provide [35]. As a result, we have the following constraint

$$0 \le a_n \le a_n^{\max}, \forall n \in N. \quad (3)$$

In practice, due to the circuit limitation, the power amplification factor generally depends on its corresponding phase [27]. In order to investigate the upper bound performance and get some insightful results, we first adopt the ideal active TE model, and in Section III-D, we further discuss the practical TE model that power amplification factor is phase-dependent.

C. System Model

The signal from the BS is given by $\mathbf{x} = \sum_{k=1}^{K}\mathbf{W}_k\mathbf{s}_k, \forall k$, where $\mathbf{s}_k \in \mathbb{C}^{s\times 1}$ represents s desired streams for the k-th UE and satisfies $\mathbf{s}_k \sim \mathcal{CN}\left(\mathbf{0}_s, \mathbf{I}_s\right)$, and $\mathbf{W}_k \in \mathbb{C}^{M_t\times s}$ denotes the linear precoding matrix. As a result, the signal received at the active-ITS is given by $\mathbf{y}_{\text{ITS}}^r = \mathbf{G}\mathbf{x} + \boldsymbol{\upsilon}$, where $\boldsymbol{\upsilon} \in \mathbb{C}^{N\times 1}$ is additive white Gaussian noise (AWGN) and satisfying $\boldsymbol{\upsilon} \sim \mathcal{CN}\left(\mathbf{0}_N, \delta^2\mathbf{I}_N\right)$. Then, by employing an active-ITS, the refracted signal can be denoted by

$$\mathbf{y}_{\text{ITS}}^t = \kappa\mathbf{A}\mathbf{\Theta}\left(\mathbf{G}\mathbf{x} + \boldsymbol{\upsilon}\right), \quad (4)$$

where κ characterizes the extent of loss corresponding to the absorption and the reflection of the power when the mmWave signal impinges upon and penetrates the ITS. The noise $\boldsymbol{\upsilon}$ after power amplifying cannot be negligible, which is significantly different from the passive-ITS [28].

The signal received by the k-th UE can be mathematically expressed as follows

$$\mathbf{y}_k = \left(\mathbf{B}_k^{\text{H}} + \kappa\mathbf{H}_k^{\text{H}}\mathbf{A}\mathbf{\Theta}\mathbf{G}\right)\sum_{i=1}^{K}\mathbf{W}_i\mathbf{s}_i + \kappa\mathbf{H}_k^{\text{H}}\mathbf{A}\mathbf{\Theta}\boldsymbol{\upsilon} + \mathbf{n}_k, \forall k, \quad (5)$$

where $\mathbf{n}_k \in \mathbb{C}^{M_r\times 1}$ represents the thermal noise at the k-th UE and satisfying $\mathbf{n}_k \sim \mathcal{C}\left(\mathbf{0}_{M_r}, \sigma_k^2\mathbf{I}_{M_r}\right)$.

For simplicity of presentation, we define the cascade channel between the BS and k-th UE is $\mathbf{Z}_k^{\text{H}} = \mathbf{B}_k^{\text{H}} + \kappa\mathbf{H}_k^{\text{H}}\mathbf{\Phi}\mathbf{G}$, where $\mathbf{\Phi} = \mathbf{A}\circ\mathbf{\Theta}$ denotes the transmissive coefficient matrix, with each diagonal element is denoted by $\Phi_{n,n} = a_n e^{j\varphi_n}, \forall n \in N$. Consequently, the signal-to-interference-plus-noise-ratio (SINR) at the k-th UE is given by

$$\mathbf{\Gamma}_k = \mathbf{Z}_k^{\text{H}}\mathbf{W}_k\mathbf{W}_k^{\text{H}}\mathbf{Z}_k\mathbf{V}_k^{-1}, \forall k, \quad (6)$$

where $\mathbf{V}_k$ denotes the interference-plus-noise covariance matrix and can be expressed by $\mathbf{V}_k = \sum_{i=1,i\neq k}^{K}\mathbf{Z}_k^{\text{H}}\mathbf{W}_i\mathbf{W}_i^{\text{H}}\mathbf{Z}_k + \kappa^2\delta^2\mathbf{H}_k^{\text{H}}\mathbf{\Phi}\mathbf{\Phi}^{\text{H}}\mathbf{H}_k + \sigma_k^2\mathbf{I}_{M_r}, \forall k$. Accordingly, the achievable data rate of the k-th UE is given by $\mathcal{R}_k = \log\left|\mathbf{I} + \mathbf{\Gamma}_{\boldsymbol{k}}\right|, \forall k$. The transmission and amplification power of the BS and active-ITS can be respectively determined by

$$P_{\text{BS}} = \mathbf{E}\left[\|\mathbf{x}\|_{\text{F}}^2\right] = \sum_{k=1}^{K}\|\mathbf{W}_k\|_{\text{F}}^2, \quad (7a)$$

$$P_{\text{ITS}} = \mathbf{E}\left[\left\|\mathbf{y}_{\text{ITS}}^t\right\|_{\text{F}}^2\right] = \sum_{k=1}^{K}\kappa^2\|\mathbf{\Phi}\mathbf{G}\mathbf{W}_k\|_{\text{F}}^2 + \delta^2\kappa^2\|\mathbf{\Phi}\|_{\text{F}}^2. \quad (7b)$$

D. Problem Formulation

In this paper, we aim to maximize the WSR by joint optimizing the linear precoding matrix at the BS, i.e., $\mathbf{W}$, and the transmissive coefficient matrix at the active-ITS, i.e., $\mathbf{\Phi}$. Based on the above discussions, the WSR maximization problem can be formulated as follows

$$\max_{\mathbf{W},\mathbf{\Phi}}\ \mathcal{R}\left(\mathbf{W}, \mathbf{\Phi}\right) = \sum_{k=1}^{K}\alpha_k\log\left|\mathbf{I} + \mathbf{\Gamma}_{\boldsymbol{k}}\right|, \quad (8a)$$

$$\text{s.t.}\ \sum_{k=1}^{K}\|\mathbf{W}_k\|_{\text{F}}^2 \le P_{\text{BS}}^{\max} - P_{\text{BS}}^{\text{circ}}, \quad (8b)$$

$$\sum_{k=1}^{K}\kappa^2\|\mathbf{\Phi}\mathbf{G}\mathbf{W}_k\|_{\text{F}}^2 + \delta^2\kappa^2\|\mathbf{\Phi}\|_{\text{F}}^2 \le P_{\text{ITS}}^{\max} - P_{\text{ITS}}^{\text{circ}} \quad (8c)$$

$$\left|\mathbf{\Phi}_{n,n}\right| \le a_n^{\max}, \forall n \in N, \quad (8d)$$

where $\boldsymbol{\alpha}$ denotes the weighting factor vector with each element α_k is a predetermined value depending on the fairness and required quality of service for applications. The constraints (8b) and (8c) ensure that the total power consumption at the BS and active-ITS should respectively not exceed the maximum

Authorized licensed use limited to: University of Houston. Downloaded on July 02,2023 at 13:19:36 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Wireless Communications. 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/TWC.2023.3247906

allowed power $P_{\mathrm{BS}}^{\max}$ and $P_{\mathrm{ITS}}^{\max}$, where $P_{\mathrm{BS}}^{\mathrm{circ}}$ and $P_{\mathrm{ITS}}^{\mathrm{circ}} = N\left(P_{\mathrm{SW}} + P_{\mathrm{DC}}\right)$ denote the static circuit power of the BS and active-ITS, with P_{SW} and P_{DC} are the power consumption for switching phase-shifting and direct current biasing power for power-amplifying of each active TE [35], [37].

Compared to the WSR maximization problem with the conventional passive-RIS [39]–[41], the noise term in (6), the additional constraints in (8c) and (8d), as well as the optimization of the power amplification factor matrix [33], [43] make Problem (8) even more challenging, and hence the existing algorithm cannot be directly employed to solve it.

III. The Proposed Suboptimal Joint Precoding Algorithm

The formulated problem is non-convex and arduous to tackle optimally, in this section, we propose an efficient joint precoding algorithm to handle it with suboptimal solutions. Particularly, to facilitate the solution development, we first reformulate the original problem as an equivalent but tractable form, and then a BCD-based method is provided to solve the transformed problem.

A. Problem Reformulation and Decomposition

First, to deal with the complexity of the objective function in the formulated WSR maximization problem, we transform it into an equivalent weighted minimum mean-square error (WMMSE) minimization problem [39], [40], [44]. Let us assume that the MMSE estimator $\mathbf{U} = \left\{\mathbf{U}_k \in \mathbb{C}^{M_r \times s}, \forall k\right\}$ are applied to UEs, and accordingly, the estimated signal vector is given by $\hat{\mathbf{s}}_k = \mathbf{U}_k^{\mathrm{H}}\mathbf{y}_k, \forall k$. Therefore, the MMSE matrix of the k-th UE is given by

$$\begin{aligned}\mathbf{E}_k =& \mathbb{E}_{\mathbf{x},\boldsymbol{\upsilon},\mathbf{n}}\left[\left(\hat{\mathbf{s}}_k - \mathbf{s}_k\right)\left(\hat{\mathbf{s}}_k - \mathbf{s}_k\right)^{\mathrm{H}}\right] \\ =& \sum_{i=1}^{K}\mathbf{U}_k^{\mathrm{H}}\mathbf{Z}_k^{\mathrm{H}}\mathbf{W}_i\mathbf{W}_i^{\mathrm{H}}\mathbf{Z}_k\mathbf{U}_k + \kappa^2\delta^2\mathbf{U}_k^{\mathrm{H}}\mathbf{H}_k^{\mathrm{H}}\boldsymbol{\Phi}\boldsymbol{\Phi}^{\mathrm{H}}\mathbf{H}_k\mathbf{U}_k \\ &+ \sigma_k^2\mathbf{U}_k^{\mathrm{H}}\mathbf{U}_k + \mathbf{I}_s - \mathbf{U}_k^{\mathrm{H}}\mathbf{Z}_k^{\mathrm{H}}\mathbf{W}_k - \mathbf{W}_k^{\mathrm{H}}\mathbf{Z}_k\mathbf{U}_k, \forall k. \quad (9)\end{aligned}$$

By introducing an auxiliary variable $\mathbf{F}_k \in \mathbb{C}^{s\times s}$ for the k-th UE and defining $\mathbf{F} = \{\mathbf{F}_k, \forall k\}$, Problem (8) can be equivalently transformed into

$$\max_{\mathbf{F},\mathbf{U},\mathbf{W},\boldsymbol{\Phi}} \sum_{k=1}^{K}\alpha_k h_k\left(\mathbf{F},\mathbf{U},\mathbf{W},\boldsymbol{\Phi}\right) \quad (10a)$$

$$\text{s. t. } (8\text{b}) - (8\text{d}), \quad (10b)$$

where $h_k\left(\mathbf{F},\mathbf{U},\mathbf{W},\boldsymbol{\Phi}\right) = \log|\mathbf{F}_k| - \mathrm{Tr}\left(\mathbf{F}_k\mathbf{E}_k\right) + s, \forall k$.

Although the above problem has been significantly simplified compared with Problem (8), it is still challenging to tackle. Fortunately, based on the fact that the objective function in (10a) is concave with respect to any one of the four variables (i.e., $\mathbf{U}$, $\mathbf{F}$, $\mathbf{W}$, and $\boldsymbol{\Phi}$) when the other three variables being fixed, which makes the problem more amenable. In the following, we provide an efficient BCD-based algorithm for solving Problem (10) in an iterative manner.

Note that the MMSE decoding matrix $\mathbf{U}$ and the auxiliary matrix $\mathbf{F}$ do not exist in any constraints in Problem (10), which implies that with fixed $\mathbf{W}$ and $\boldsymbol{\Theta}$, the calculation of $\mathbf{F}^\star$ and $\mathbf{U}^\star$ constitute a pair of unconstrained optimization problems. Therefore, the optimal solutions can be obtained by setting the first-order partial derivatives of the objective function of Problem (10) with respect to $\mathbf{U}_k, \forall k$ and $\mathbf{F}_k, \forall k$ to be zeros, respectively. After some matrix manipulations, the optimal closed-form solutions can be determined by

$$\mathbf{U}_k^\star = \left(\bar{\mathbf{V}}_k\right)^{-1}\mathbf{Z}_k^{\mathrm{H}}\mathbf{W}_k, \forall k, \quad (11)$$

where $\bar{\mathbf{V}}_k = \mathbf{V}_k + \mathbf{Z}_k^{\mathrm{H}}\mathbf{W}_k\mathbf{W}_k^{\mathrm{H}}\mathbf{Z}_k$, and

$$\mathbf{F}_k^\star = \left(\mathbf{E}_k^\star\right)^{-1}, \forall k, \quad (12)$$

where $\mathbf{E}_k^\star$ is determined by $\mathbf{E}_k^\star = \mathbf{I}_s - \mathbf{W}_k^{\mathrm{H}}\mathbf{Z}_k\bar{\mathbf{V}}_k^{-1}\mathbf{Z}_k^{\mathrm{H}}\mathbf{W}_k, \forall k$.

Once both $\mathbf{U}^\star$ and $\mathbf{F}^\star$ are determined, the remaining work is to optimize the linear precoding matrix $\mathbf{W}$ and the transmissive coefficient matrix $\boldsymbol{\Phi}$. Note that $\mathbf{W}$ and $\boldsymbol{\Phi}$ are intricately coupled in $\mathbf{E}_k, \forall k$, we consider decomposing Problem (10) into two subproblems with respect to these two variables. In particular, the two subproblems can be expressed by

$$\mathbf{W}^\star = \arg\min_{\mathbf{W}}\sum_{k=1}^{K}\alpha_k\,\mathrm{Tr}\left(\mathbf{F}_k\mathbf{E}_k\right), \text{s. t. } (8\text{b}) \text{ and } (8\text{c}), \quad (13a)$$

$$\boldsymbol{\Phi}^\star = \arg\min_{\boldsymbol{\Phi}}\sum_{k=1}^{K}\alpha_k\,\mathrm{Tr}\left(\mathbf{F}_k\mathbf{E}_k\right), \text{s. t. } (8\text{c}) \text{ and } (8\text{d}). \quad (13b)$$

In the rest of this section, we propose a pair of methods to solve the above two subproblems.

B. Linear Precoding Matrix Optimization

In this subsection, we focus on the subproblem state in (13a) for optimizing linear precoding matrix $\mathbf{W}^\star$ at the BS with the fixed transmissive coefficient matrix $\boldsymbol{\Phi}$. Since the part terms of the objective function in Problem (13a), i.e., $\mathrm{Tr}\left(\alpha_k\kappa^2\delta^2\mathbf{U}_k^{\mathrm{H}}\mathbf{H}_k^{\mathrm{H}}\boldsymbol{\Phi}\boldsymbol{\Phi}^{\mathrm{H}}\mathbf{H}_k\mathbf{U}_k\right)$, and $\mathrm{Tr}\left(\sigma_k^2\mathbf{U}_k^{\mathrm{H}}\mathbf{U}_k\right) + \mathbf{I}_s$ are irrelevant constants with respect to $\mathbf{W}$, and hence, by omitting them, the optimal solution of $\mathbf{W}$ can be determined by solving the following subproblem

$$\begin{aligned}\min_{\mathbf{W}} \quad & \sum_{k=1}^{K}\mathrm{Tr}\left(\mathbf{W}_k^{\mathrm{H}}\mathbf{Q}\mathbf{W}_k\right) - \sum_{k=1}^{K}\mathrm{Tr}\left(\alpha_k\mathbf{F}_k\mathbf{U}_k^{\mathrm{H}}\mathbf{Z}_k^{\mathrm{H}}\mathbf{W}_k\right) \\ & - \sum_{k=1}^{K}\mathrm{Tr}\left(\alpha_k\mathbf{F}_k\mathbf{W}_k^{\mathrm{H}}\mathbf{Z}_k\mathbf{U}_k\right), \quad (14a) \\ \text{s. t.} \quad & \sum_{k=1}^{K}\mathrm{Tr}\left(\mathbf{W}_k^{\mathrm{H}}\mathbf{W}_k\right) \le \hat{P}_{\mathrm{BS}}^{\max}, \quad (14b) \\ & \sum_{k=1}^{K}\mathrm{Tr}\left(\kappa^2\mathbf{W}_k^{\mathrm{H}}\mathbf{G}^{\mathrm{H}}\boldsymbol{\Phi}^{\mathrm{H}}\boldsymbol{\Phi}\mathbf{G}\mathbf{W}_k\right) \le \hat{P}_{\mathrm{ITS}}^{\max}, \quad (14c)\end{aligned}$$

where $\hat{P}_{\mathrm{BS}}^{\max} = P_{\mathrm{BS}}^{\max} - P_{\mathrm{BS}}^{\mathrm{circ}}$, $\hat{P}_{\mathrm{ITS}}^{\max} = P_{\mathrm{ITS}}^{\max} - P_{\mathrm{ITS}}^{\mathrm{circ}} - \mathrm{Tr}\left(\delta^2\kappa^2\boldsymbol{\Phi}^{\mathrm{H}}\boldsymbol{\Phi}\right)$, and $\mathbf{Q} = \sum_{k=1}^{K}\alpha_k\mathbf{Z}_k\mathbf{U}_k\mathbf{F}_k\mathbf{U}_k^{\mathrm{H}}\mathbf{Z}_k^{\mathrm{H}}$.

It can be proved that $\mathbf{Q}$ is a non-negative definite matrix, therefore the objective function in (14a) is a convex function. In addition, the constraints in (14b) and (14c) are convex with respect to $\mathbf{W}$, and so Problem (14) constitutes a convex optimization problem, which can be effectively solved by

Authorized licensed use limited to: University of Houston. Downloaded on July 02,2023 at 13:19:36 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Wireless Communications. 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/TWC.2023.3247906

employing convex solver toolbox, e.g., CVX [45]. Instead of relying on the generic solver with high computational complexity, we provide the following duality method to further proceed with Problem (14) efficiently. By introducing a pair of auxiliary variables $\lambda \geq 0$ and $\mu \geq 0$, the constraints in (14b) and (14c) can be combined as

$$\lambda \sum_{k=1}^{K} \mathrm{Tr}\left(\mathbf{W}_k^{\mathrm{H}} \mathbf{W}_k\right) + \mu \sum_{k=1}^{K} \mathrm{Tr}\left(\kappa^2 \mathbf{W}_k^{\mathrm{H}} \mathbf{G}^{\mathrm{H}} \mathbf{\Phi}^{\mathrm{H}} \mathbf{\Phi} \mathbf{G} \mathbf{W}_k\right) \leq P_{\mathrm{sum}}^{\max}, \quad (15)$$

where $P_{\mathrm{sum}}^{\max}$ is defined by $P_{\mathrm{sum}}^{\max} = \lambda \hat{P}_{\mathrm{BS}}^{\max} + \mu \hat{P}_{\mathrm{ITS}}^{\max}$. Consequently, Problem (14) can be recast by

$$g(\lambda, \mu) = \left\{ \min_{\mathbf{W}} \ (14\mathrm{a}), \ \mathrm{s.t.} \ (15) \right\}. \quad (16)$$

It is evident that a feasible solution to Problem (14) is also feasible for Problem (16). The relationship of the optimal values between the two problems is shown in the following lemma.

Lemma 1. *The optimal value of Problem* (16) *for any given pair of* $\{\lambda, \mu\}$*,* $\lambda \geq 0$*,* $\mu \geq 0$*, is an lower bound on the optimal value of Problem* (14)*. Furthermore, the optimal value of the dual problem of Problem* (16)*, i.e.,* $\max_{\lambda,\mu} g(\lambda, \mu)$*, is equal to that of Problem* (14)*.*

Proof. This lemma follows *Propositions* 4 in [46], and hence, is omitted here for brevity. □

With the fixed variables $\{\lambda, \mu\}$, introduce a Lagrangian multiplier $\varepsilon \geq 0$ associated with the combined constraint in (15), the corresponding Lagrangian function can be represented by

$$\mathcal{L}(\mathbf{W}, \varepsilon) = \sum_{k=1}^{K} \mathrm{Tr}\left(\mathbf{W}_k^{\mathrm{H}} \hat{\mathbf{Q}}(\varepsilon) \mathbf{W}_k\right) - \sum_{k=1}^{K} \mathrm{Tr}\left(\alpha_k \mathbf{F}_k \mathbf{W}_k^{\mathrm{H}} \mathbf{Z}_k \mathbf{U}_k\right) - \sum_{k=1}^{K} \mathrm{Tr}\left(\alpha_k \mathbf{F}_k \mathbf{U}_k^{\mathrm{H}} \mathbf{Z}_k^{\mathrm{H}} \mathbf{W}_k\right) - \varepsilon P_{\mathrm{sum}}^{\max}, \quad (17)$$

where $\hat{\mathbf{Q}}(\varepsilon) = \mathbf{Q} + \varepsilon\left(\lambda \mathbf{I}_{M_t} + \mu \kappa^2 \mathbf{G}^{\mathrm{H}} \mathbf{\Phi}^{\mathrm{H}} \mathbf{\Phi} \mathbf{G}\right)$. Accordingly, the Karush–Kuhn–Tucker (KKT) conditions are given by

$$\nabla_{\mathbf{W}_k} \mathcal{L}(\mathbf{W}, \varepsilon) = 2\partial \mathcal{L}\left(\tilde{\mathbf{W}}, \varepsilon\right) / \partial \mathbf{W}_k^* = \mathbf{0}, \forall k, \quad (18a)$$

$$\varepsilon f\left(\mathbf{W}_k(\varepsilon)\right) = 0, \forall k, \quad (18b)$$

where $f\left(\mathbf{W}_k(\varepsilon)\right) = \lambda \sum_{k=1}^{K} \mathrm{Tr}\left(\mathbf{W}_k^{\mathrm{H}} \mathbf{W}_k\right) + \mu \sum_{k=1}^{K} \mathrm{Tr}\left(\kappa^2 \mathbf{W}_k^{\mathrm{H}} \mathbf{G}^{\mathrm{H}} \mathbf{\Phi}^{\mathrm{H}} \mathbf{\Phi} \mathbf{G} \mathbf{W}_k\right) - P_{\mathrm{sum}}^{\max}$. From the first KKT condition in (18a), with a fixed ε, the optimal solution of $\mathbf{W}_k, \forall k$ is calculated by

$$\mathbf{W}_k(\varepsilon) = \alpha_k \left(\hat{\mathbf{Q}}(\varepsilon)\right)^{\dagger} \mathbf{Z}_k \mathbf{U}_k \mathbf{F}_k, \forall k, \quad (19)$$

where $(\cdot)^{\dagger}$ denotes the matrix pseudoinverse. Then, the dual variable ε should be determined for satisfying the second KKT condition in (18b). We search ε for two cases, i.e., $\varepsilon = 0$ and $\varepsilon > 0$. We first check the optimality of $\varepsilon = 0$, if $f\left(\mathbf{W}_k(0)\right) \leq 0$ holds, the optimal solution of $\mathbf{W}_k$ can be obtained by

$$\mathbf{W}_k = \mathbf{W}_k(0) = \alpha_k \mathbf{Q}^{\dagger} \mathbf{Z}_k \mathbf{U}_k \mathbf{F}_k, \forall k. \quad (20)$$

Algorithm 1 Iterative Subgradient algorithm

1: Initialize: λ^0, μ^0, and iteration index $p = 0$.
2: **repeat**
3: $\lambda^{p+1} \leftarrow \lambda^p + \xi f_{P_{\mathrm{BS}}}$, $\mu^{p+1} \leftarrow \mu^p + \xi f_{P_{\mathrm{ITS}}}$.
4: **until** $\left|\lambda^{p+1} - \lambda^p\right| \leq \epsilon$ and $\left|\mu^{p+1} - \mu^p\right| \leq \epsilon$.

Algorithm 2 Dual-based Algorithm for Calculating $\mathbf{W}$

1: Initialize: λ^0, μ^0, $\mathbf{W}^0$, ε^0, and iteration index $r = 0$.
2: Calculate the value of the function in (14a) as $v\left(\mathbf{W}^0\right)$.
3: **repeat**
4: Calculate $\mathbf{W}^{\mathrm{r}+1}$ and search $\varepsilon^{\mathrm{r}+1}$,
Update $\lambda^{\mathrm{r}+1}$ and $\mu^{\mathrm{r}+1}$ by employing Algorithm 1,
5: **until** $\left|v\left(\mathbf{W}^{\mathrm{r}+1}\right) - v\left(\mathbf{W}^{\mathrm{r}}\right)\right| / v\left(\mathbf{W}^{\mathrm{r}}\right) \leq \epsilon$, **Output:** $\mathbf{W}^{\star} = \mathbf{W}^{\mathrm{r}}$.

Otherwise, we have to find ε for ensuring $f\left(\mathbf{W}_k(\varepsilon)\right) = 0$ satisfied. Note that $f\left(\mathbf{W}_k(\varepsilon)\right)$ is a monotonically decreasing function of ε that enables a bisection search method to find ε.

Then, the remaining work is to solve the dual problem $\max_{\lambda,\mu} g(\lambda, \mu)$. We employ the iterative subgradient method [40] to find the solutions. First, the subgradient directions of the above problem are determined by

$$f_{P_{\mathrm{BS}}} = \sum_{k=1}^{K} \mathrm{Tr}\left(\mathbf{W}_k^{\mathrm{H}} \mathbf{W}_k\right) - \hat{P}_{\mathrm{BS}}^{\max}, \quad (21a)$$

$$f_{P_{\mathrm{ITS}}} = \sum_{k=1}^{K} \mathrm{Tr}\left(\kappa^2 \mathbf{W}_k^{\mathrm{H}} \mathbf{G}^{\mathrm{H}} \mathbf{\Phi}^{\mathrm{H}} \mathbf{\Phi} \mathbf{G} \mathbf{W}_k\right) - \hat{P}_{\mathrm{ITS}}^{\max}. \quad (21b)$$

Then, the values of λ and μ can be updated in an iterative manner, the details are summarized in Algorithm 1, where ξ denotes the step size of the subgradient algorithm, ϵ represents the tolerance, and the superscript stands for the number of iteration. The overall dual-based method for optimizing the linear precoding matrix at the BS is summarized in Algorithm 2, and the optimal solution $\mathbf{W}^{\star}$ can be obtained when the algorithm converged.

C. Transmissive Coefficient Matrix Optimization

Now, we consider solving the subproblem of optimizing the transmissive coefficient matrix at the active-ITS. With the fixed $\mathbf{W}^{\star}$ by employing Algorithm 2 and define $\tilde{\mathbf{W}} = \sum_{i=1}^{K} \mathbf{W}_i \mathbf{W}_i^{\mathrm{H}}$, after dropping the constant terms, Problem (13b) can be simplified expressed by

$$\begin{aligned} \min_{\mathbf{\Phi}} \ & \sum_{k=1}^{K} \alpha_k \mathrm{Tr}\left(\mathbf{F}_k \mathbf{U}_k^{\mathrm{H}} \mathbf{Z}_k^{\mathrm{H}} \tilde{\mathbf{W}} \mathbf{Z}_k \mathbf{U}_k\right) \\ & + \sum_{k=1}^{K} \alpha_k \kappa^2 \delta^2 \mathrm{Tr}\left(\mathbf{F}_k \mathbf{U}_k^{\mathrm{H}} \mathbf{H}_k^{\mathrm{H}} \mathbf{\Phi} \mathbf{\Phi}^{\mathrm{H}} \mathbf{H}_k \mathbf{U}_k\right) \\ & - \sum_{k=1}^{K} \alpha_k \mathrm{Tr}\left(\mathbf{F}_k \mathbf{U}_k^{\mathrm{H}} \mathbf{Z}_k^{\mathrm{H}} \mathbf{W}_k\right) \\ & - \sum_{k=1}^{K} \alpha_k \mathrm{Tr}\left(\mathbf{F}_k \mathbf{W}_k^{\mathrm{H}} \mathbf{Z}_k \mathbf{U}_k\right) \quad (22a) \\ \mathrm{s.t.} \ & (8\mathrm{c}) \ \text{and} \ (8\mathrm{d}). \quad (22b) \end{aligned}$$

Authorized licensed use limited to: University of Houston. Downloaded on July 02,2023 at 13:19:36 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Wireless Communications. 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/TWC.2023.3247906

By substituting $\mathbf{Z}_k$ into the objective function, we have $\mathbf{F}_k\mathbf{U}_k^{\mathrm{H}}\mathbf{Z}_k^{\mathrm{H}}\tilde{\mathbf{W}}\mathbf{Z}_k\mathbf{U}_k = \kappa^2\mathbf{F}_k\mathbf{U}_k^{\mathrm{H}}\mathbf{H}_k^{\mathrm{H}}\boldsymbol{\Phi}\mathbf{G}\tilde{\mathbf{W}}\mathbf{G}^{\mathrm{H}}\boldsymbol{\Phi}^{\mathrm{H}}\mathbf{H}_k\mathbf{U}_k + \kappa\mathbf{F}_k\mathbf{U}_k^{\mathrm{H}}\mathbf{B}_k^{\mathrm{H}}\tilde{\mathbf{W}}\mathbf{G}^{\mathrm{H}}\boldsymbol{\Phi}^{\mathrm{H}}\mathbf{H}_k\mathbf{U}_k + \kappa\mathbf{F}_k\mathbf{U}_k^{\mathrm{H}}\mathbf{H}_k^{\mathrm{H}}\boldsymbol{\Phi}\mathbf{G}\tilde{\mathbf{W}}\mathbf{B}_k\mathbf{U}_k + \mathbf{F}_k\mathbf{U}_k^{\mathrm{H}}\mathbf{B}_k^{\mathrm{H}}\tilde{\mathbf{W}}\mathbf{B}_k\mathbf{U}_k$ and $\mathbf{F}_k\mathbf{U}_k^{\mathrm{H}}\mathbf{Z}_k^{\mathrm{H}}\mathbf{W}_k = \kappa\mathbf{F}_k\mathbf{U}_k^{\mathrm{H}}\mathbf{H}_k^{\mathrm{H}}\boldsymbol{\Phi}\mathbf{G}\mathbf{W}_k + \mathbf{F}_k\mathbf{U}_k^{\mathrm{H}}\mathbf{B}_k^{\mathrm{H}}\mathbf{W}_k$.

By defining $\mathbf{D}_k = \alpha_k\mathbf{H}_k\mathbf{U}_k\mathbf{F}_k\mathbf{U}_k^{\mathrm{H}}\mathbf{H}_k^{\mathrm{H}}$, $\mathbf{T} = \kappa^2\mathbf{G}\tilde{\mathbf{W}}\mathbf{G}^{\mathrm{H}}$, and $\mathbf{L}_k = \kappa\alpha_k\mathbf{H}_k\mathbf{U}_k\mathbf{F}_k\mathbf{U}^{\mathrm{H}}\mathbf{B}_k^{\mathrm{H}}\tilde{\mathbf{W}}\mathbf{G}^{\mathrm{H}}$, we have

$$\begin{aligned}&\alpha_k\,\mathrm{Tr}\left(\mathbf{F}_k\mathbf{U}_k^{\mathrm{H}}\mathbf{Z}_k^{\mathrm{H}}\tilde{\mathbf{W}}\mathbf{Z}_k\mathbf{U}_k\right)\\&= \mathrm{Tr}\left(\boldsymbol{\Phi}^{\mathrm{H}}\mathbf{D}_k\boldsymbol{\Phi}\mathbf{T}\right) + \mathrm{Tr}\left(\boldsymbol{\Phi}^{\mathrm{H}}\mathbf{L}_k\right) + \mathrm{Tr}\left(\mathbf{L}_k^{\mathrm{H}}\boldsymbol{\Phi}\right) + c_{1,k}, \quad (23a)\\&\alpha_k\,\mathrm{Tr}\left(\mathbf{F}_k\mathbf{U}_k^{\mathrm{H}}\mathbf{H}_k^{\mathrm{H}}\boldsymbol{\Phi}\boldsymbol{\Phi}^{\mathrm{H}}\mathbf{H}_k\mathbf{U}_k\right) = \mathrm{Tr}\left(\boldsymbol{\Phi}^{\mathrm{H}}\mathbf{D}_k\boldsymbol{\Phi}\right), \quad (23b)\end{aligned}$$

where $c_{1,k} = \mathrm{Tr}\left(\mathbf{F}_k\mathbf{U}_k^{\mathrm{H}}\mathbf{B}_k^{\mathrm{H}}\tilde{\mathbf{W}}\mathbf{B}_k\mathbf{U}_k\right)$.

Then, by defining $\mathbf{O}_k = \alpha_k\kappa\mathbf{H}_k\mathbf{U}_k\mathbf{F}_k\mathbf{W}_k^{\mathrm{H}}\mathbf{G}$, we have $\alpha_k\,\mathrm{Tr}\left(\mathbf{F}_k\mathbf{U}_k^{\mathrm{H}}\mathbf{Z}_k^{\mathrm{H}}\mathbf{W}_k\right) = \mathrm{Tr}\left(\mathbf{O}_k^{\mathrm{H}}\boldsymbol{\Phi}\right) + c_{2,k}$, where $c_{2,k} = \mathrm{Tr}\left(\mathbf{F}_k\mathbf{U}_k^{\mathrm{H}}\mathbf{B}_k^{\mathrm{H}}\mathbf{W}_k\right)$. Similarly, by defining $\bar{P}_{\mathrm{ITS}}^{\max} = \zeta\left(P_{\mathrm{ITS}}^{\max} - P_{\mathrm{ITS}}^{\mathrm{circ}}\right)$, constraint (8c) can be recast as

$$\mathrm{Tr}\left(\boldsymbol{\Phi}^{\mathrm{H}}\boldsymbol{\Phi}\mathbf{T}\right) + \delta^2\kappa^2\,\mathrm{Tr}\left(\boldsymbol{\Phi}^{\mathrm{H}}\boldsymbol{\Phi}\right) \le \bar{P}_{\mathrm{ITS}}^{\max}.$$

Consequently, Problem (22) can be transformed into the following equivalent form

$$\begin{aligned}\min_{\boldsymbol{\Phi}}\ &\mathrm{Tr}\left(\boldsymbol{\Phi}^{\mathrm{H}}\mathbf{D}\boldsymbol{\Phi}\mathbf{T}\right) + \delta^2\kappa^2\,\mathrm{Tr}\left(\boldsymbol{\Phi}^{\mathrm{H}}\mathbf{D}\boldsymbol{\Phi}\right),\\&- \mathrm{Tr}\left(\boldsymbol{\Phi}^{\mathrm{H}}\mathbf{C}\right) - \mathrm{Tr}\left(\mathbf{C}^{\mathrm{H}}\boldsymbol{\Phi}\right), \quad (24a)\\\text{s.t. }&\text{(24a) and (8d)}, \quad (24b)\end{aligned}$$

where $\mathbf{D} = \sum_{k=1}^{K}\mathbf{D}_k$ and $\mathbf{C} = \sum_{k=1}^{K}\mathbf{O}_k - \sum_{k=1}^{K}\mathbf{L}_k$.

Note that the above problem is still difficult to tackle. In the following, we transform Problem (24) into an equivalent form by employing some further algebraic manipulations. We define a sequence of equalities as follows $\mathrm{Tr}\left(\boldsymbol{\Phi}^{\mathrm{H}}\mathbf{D}\boldsymbol{\Phi}\mathbf{T}\right) = \boldsymbol{\phi}^{\mathrm{H}}\boldsymbol{\Omega}\boldsymbol{\phi}$, $\mathrm{Tr}\left(\boldsymbol{\Phi}^{\mathrm{H}}\mathbf{D}\boldsymbol{\Phi}\right) = \boldsymbol{\phi}^{\mathrm{H}}\hat{\mathbf{D}}\boldsymbol{\phi}$, $\mathrm{Tr}\left(\boldsymbol{\Phi}^{\mathrm{H}}\mathbf{D}\boldsymbol{\Phi}\right) = \boldsymbol{\phi}^{\mathrm{H}}\hat{\mathbf{T}}\boldsymbol{\phi}$, $\mathrm{Tr}\left(\boldsymbol{\Phi}^{\mathrm{H}}\boldsymbol{\Phi}\right) = \boldsymbol{\phi}^{\mathrm{H}}\phi$, $\mathrm{Tr}\left(\boldsymbol{\Phi}^{\mathrm{H}}\mathbf{C}\right) = \boldsymbol{\phi}^{\mathrm{H}}\mathbf{c}$, and $\mathrm{Tr}\left(\boldsymbol{\Phi}\mathbf{C}^{\mathrm{H}}\right) = \mathbf{c}^{\mathrm{H}}\boldsymbol{\phi}$, where $\boldsymbol{\Omega} = \mathbf{D}\circ\mathbf{T}^{\mathrm{T}}$, $\boldsymbol{\phi} = \mathrm{Vecd}\left(\boldsymbol{\Phi}\right)$, $\hat{\mathbf{D}} = \delta^2\kappa^2\left(\mathbf{D}\circ\mathbf{I}_N^{\mathrm{T}}\right)$, $\hat{\mathbf{T}} = \mathbf{I}_N\circ\mathbf{T}^{\mathrm{T}}$, $\boldsymbol{\phi} = \mathrm{Vecd}\left(\boldsymbol{\Phi}\right)$ and $\mathbf{c} = \mathrm{Vecd}\left(\mathbf{C}\right)$, with $\mathrm{Vecd}\left(\mathbf{X}\right)$ forms a vector out of the diagonal of its matrix argument. The above equalities given above follow from the properties in [47, Theorem 1.11]. Additionally, the constraint $|\phi_n| \le a_n^{\max}$ is convex and can be recast by $\boldsymbol{\phi}^{\mathrm{H}}\mathbf{e}_n\mathbf{e}_n^{\mathrm{T}}\boldsymbol{\phi} \le \left(a_n^{\max}\right)^2$, where $\mathbf{e}_n$ is the elementary vector with one in the n-th position and zeros elsewhere. Consequently, by defining $\hat{\boldsymbol{\Omega}} = \boldsymbol{\Omega} + \hat{\mathbf{D}}$ and $\bar{\mathbf{T}} = \hat{\mathbf{T}} + \delta^2\kappa^2\mathbf{I}_N$, Problem (24) can be reformulated by

$$\begin{aligned}\min_{\boldsymbol{\phi}}\ &\boldsymbol{\phi}^{\mathrm{H}}\hat{\boldsymbol{\Omega}}\boldsymbol{\phi} - \boldsymbol{\phi}^{\mathrm{H}}\mathbf{c} - \mathbf{c}^{\mathrm{H}}\boldsymbol{\phi}, \quad (25a)\\\text{s.t. }&\boldsymbol{\phi}^{\mathrm{H}}\bar{\mathbf{T}}\boldsymbol{\phi} \le \bar{P}_{\mathrm{ITS}}^{\max}, \quad (25b)\\&\boldsymbol{\phi}^{\mathrm{H}}\mathbf{e}_n\mathbf{e}_n^{\mathrm{T}}\boldsymbol{\phi} \le \left(a_n^{\max}\right)^2, \forall n \in N. \quad (25c)\end{aligned}$$

Since Problem (25) is a convex quadratic constraint quadratic programming (QCQP) problem, the optimal solution of $\boldsymbol{\phi}$ can be determined by employing the Lagrangian multiplier method, i.e.,

$$\boldsymbol{\phi}^{\star} = \left(\hat{\boldsymbol{\Omega}} + \vartheta^{\star}\bar{\mathbf{T}} + \sum_{n=1}^{N}\tau_n^{\star}\mathbf{e}_n\mathbf{e}_n^{\mathrm{T}}\right)^{-1}\mathbf{c}, \quad (26)$$

where $\vartheta^{\star}$ and $\tau_n^{\star}, \forall n$ denote the searched optimal Lagrangian multipliers associated with the constraints in (25b) and (25c). However, for a given accuracy tolerance ϵ, the total computational complexity for obtaining the transmissive coefficient $\boldsymbol{\phi}^{\star}$ by employing the Lagrangian multiplier method is $\mathcal{O}\left(\log\left(\frac{1}{\epsilon}\right)\sqrt{N+1}\left(3N+1\right)N^3\right) \approx \mathcal{O}\left(N^{4.5}\right)$, which is mainly due to the matrix inversion and the two-dimensional grid search for multipliers. To reduce the computational complexity so that the algorithm can be adopted to real-time communications systems in practice, hereinafter, we provide a computationally efficient algorithm (on the order of $\mathcal{O}\left(N^2\right)$) by investigating the particular structures of the objective function and constraints in Problem (25). In addition, the proposed algorithm can be easily extended to optimize the transmissive coefficient $\boldsymbol{\phi}$ under the more practical TE models [26], [40], which is to be discussed in Section III-D.

Here, our objective is to obtain the transmissive coefficient vector $\boldsymbol{\phi} = \mathbf{a}\circ\boldsymbol{\theta}$, which consists of amplification factor $\mathbf{a}$ and phase-shifting $\boldsymbol{\theta}$, where $\mathbf{a} = \mathrm{Vecd}\left(\mathbf{A}\right)$ and $\boldsymbol{\theta} = \mathrm{Vecd}\left(\boldsymbol{\Theta}\right)$, respectively. In addition, since $\mathrm{Tr}\left(\boldsymbol{\Phi}^{\mathrm{H}}\boldsymbol{\Phi}\right) = \mathbf{a}^{\mathrm{T}}\mathbf{a}$, the terms in Problem (24) can be recast as $\mathrm{Tr}\left(\boldsymbol{\Phi}^{\mathrm{H}}\boldsymbol{\Phi}\right) = \sum_{n=1}^{N}a_n^2$, $\mathrm{Tr}\left(\boldsymbol{\Phi}^{\mathrm{H}}\mathbf{D}\boldsymbol{\Phi}\right) = \sum_{n=1}^{N}a_n^2D_{\mathrm{n,n}}$, and $\mathrm{Tr}\left(\boldsymbol{\Phi}^{\mathrm{H}}\boldsymbol{\Phi}\mathbf{T}\right) = \sum_{n=1}^{N}a_n^2T_{\mathrm{n,n}}$, which have no impact on updating $\boldsymbol{\theta}$. As a result, the amplification power budget constraint in (24a) can be recast by $\sum_{n=1}^{N}a_n^2\left(T_{\mathrm{n,n}} + \delta^2\kappa^2\right) \le \bar{P}_{\mathrm{ITS}}^{\max}$.

Based on the above discussion, Problem (24) can be equivalently transformed as follows

$$\begin{aligned}\min_{\mathbf{a},\boldsymbol{\theta}}\ &\left(\mathbf{a}^{\mathrm{T}}\circ\boldsymbol{\theta}^{\mathrm{H}}\right)\boldsymbol{\Omega}\left(\mathbf{a}\circ\boldsymbol{\theta}\right) + \delta^2\kappa^2\sum_{n=1}^{N}a_n^2D_{\mathrm{n,n}}\\&- \left(\mathbf{a}^{\mathrm{T}}\circ\boldsymbol{\theta}^{\mathrm{H}}\right)\mathbf{c} - \mathbf{c}^{\mathrm{H}}\left(\mathbf{a}\circ\boldsymbol{\theta}\right), \quad (27a)\\\text{s.t. }&\sum_{n=1}^{N}a_n^2\left(T_{\mathrm{n,n}} + \delta^2\kappa^2\right) \le \bar{P}_{\mathrm{ITS}}^{\max}, \quad (27b)\\&|\theta_n| = 1, \forall n \in N, \quad (27c)\\&0 \le a_n \le a_n^{\max}, \forall n \in N. \quad (27d)\end{aligned}$$

Although Problem (27) has been significantly simplified compared with Problem (24), it is still non-convex due to the constant modulus constraint (27c). As a result, the dual-based algorithm cannot be employed here due to the dual gap is not zero. Therefore, we transform the above challenging problem by employing the price mechanism [48] as follows:

$$\begin{aligned}&\min_{\mathbf{a},\boldsymbol{\theta}} h\left(\eta,\mathbf{a},\boldsymbol{\theta}\right) = f\left(\mathbf{a},\boldsymbol{\theta}\right) + \eta g\left(\mathbf{a}\right), \quad (28a)\\&\text{s.t. (27c) and (27d)}. \quad (28b)\end{aligned}$$

where $\eta \ge 0$ is the introduced price of the function $g\left(\mathbf{a}\right) = \sum_{n=1}^{N}a_n^2\left(T_{\mathrm{n,n}} + \delta^2\kappa^2\right)$, which is the left hand of constraint in (27b), and $f\left(\mathbf{a},\boldsymbol{\theta}\right)$ is the objective function of Problem (27). Then, for a given η, we propose an efficient pair-wise alternate sequential optimization (ASO) algorithm [18], [49] to obtain the high-quality suboptimal solutions $\mathbf{a}$ and $\boldsymbol{\theta}$ [43].

Authorized licensed use limited to: University of Houston. Downloaded on July 02,2023 at 13:19:36 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Wireless Communications. 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/TWC.2023.3247906

More specifically, we further express the terms of the objective function $h(\eta, \mathbf{a}, \boldsymbol{\theta})$ as follows

$$\begin{aligned}
&\left(\mathbf{a}^{\mathrm{T}} \circ \boldsymbol{\theta}^{\mathrm{H}}\right) \boldsymbol{\Omega} \left(\mathbf{a} \circ \boldsymbol{\theta}\right) \\
&= \sum_{i=1, i\neq n}^{N} \left(\mathbf{a}^{\mathrm{T}} \circ \boldsymbol{\theta}^{\mathrm{H}}\right) \boldsymbol{\Omega}_{:,i} a_i \theta_i + \left(\mathbf{a}^{\mathrm{T}} \circ \boldsymbol{\theta}^{\mathrm{H}}\right) \boldsymbol{\Omega}_{:,n} a_n \theta_n \\
&= \sum_{i=1, i\neq n}^{N} a_n \theta_n^* \Omega_{n,i} a_i \theta_i + \sum_{j=1, j\neq n}^{N} \sum_{i=1, i\neq n}^{N} a_j \theta_j^* \Omega_{j,i} a_i \theta_i \\
&\quad + a_n \theta_n^* \Omega_{n,n} a_n \theta_n + \sum_{i=1, i\neq n}^{N} a_i \theta_i^* \Omega_{i,n} a_n \theta_n, \qquad (29)
\end{aligned}$$

and $\left(\mathbf{a}^{\mathrm{T}} \circ \boldsymbol{\theta}^{\mathrm{H}}\right) \mathbf{c} = a_n \theta_n^* c_n + \sum_{i=1, i\neq n}^{N} a_i \theta_i^* c_i$. It can be readily verified that $\boldsymbol{\Omega} = \mathbf{D} \circ \mathbf{T}^{\mathrm{T}}$ is a positive semi-definite matrix due to $\mathbf{D}$ and $\mathbf{T}^{\mathrm{T}}$ are semi-definite matrices, and hence, we have $\Omega_{n,i} = \Omega_{i,n}^*$. Similarly, the constraint on the amplification power budget in (27b) can be expressed by $a_n^2 \left(T_{\mathrm{n,n}} + \delta^2 \kappa^2\right) \leq \tilde{P}_{\mathrm{ITS}}^{\max}$ where $\tilde{P}_{\mathrm{ITS}}^{\max} = \bar{P}_{\mathrm{ITS}}^{\max} - \sum_{i=1, i\neq n}^{N} a_i^2 \left(T_{i,i} + \delta^2 \kappa^2\right)$.

Clearly, the objective function $h(\eta, \mathbf{a}, \boldsymbol{\theta})$ can be recast as an equivalent function with respect to the transmissive coefficients of n-th active TE, i.e., a_n and θ_n, which is given by $h(\eta, a_n, \theta_n) = 2\Re\left\{a_n \theta_n^* \left(\sum_{i\neq n}^{N} \Omega_{n,i} a_i \theta_i - c_n\right)\right\} + a_n^2 \Omega_{n,n} + \delta^2 \kappa^2 a_n^2 D_{\mathrm{n,n}} + \eta a_n^2 \left(T_{\mathrm{n,n}} + \delta^2 \kappa^2\right) + \chi_n$, where χ_n is a constant value with respect to the n-th pair of optimization variables $\{a_n, \theta_n\}$, which is given by $\chi_n = \sum_{j\neq n}^{N} \sum_{i\neq n}^{N} a_j \theta_j^* \Omega_{j,i} a_i \theta_i + \delta^2 \kappa^2 \sum_{i\neq n}^{N} a_i^2 D_{\mathrm{i,i}} + 2\Re\left\{\sum_{i\neq n}^{N} a_i \theta_i^* c_i\right\} + \eta \sum_{i\neq n}^{N} a_i^2 \left(T_{i,i} + \delta^2 \kappa^2\right)$. Consequently, we can only investigate the following problem for sequentially optimizing a pair of values $\{a_n, \theta_n\}$ while fixing the remaining $N-1$ pairs. By omitting the constant term χ_n in $h(\eta, a_n, \theta_n)$, which has no impact on optimizing $\{a_n, \theta_n\}$, and defining $e_n = c_n - \sum_{i\neq n}^{N} \Omega_{n,i} a_i \theta_i$, we have the following surrogate problem with respect to $\{a_n, \theta_n\}$, and takes the forms

$$\begin{aligned}
\min_{a_n, \theta_n} \; & -2\Re\left\{a_n \theta_n^* e_n\right\} + a_n^2 \Omega_{n,n} + \delta^2 \kappa^2 a_n^2 D_{\mathrm{n,n}} \\
& + \eta a_n^2 \left(T_{\mathrm{n,n}} + \delta^2 \kappa^2\right), \qquad (30a) \\
\text{s.t. } & (27c) \text{ and } (27d). \qquad (30b)
\end{aligned}$$

Note that θ_n and a_n only constrained by (27c) and (27d), respectively, we decompose Problem (30) into two subproblems

$$\theta_n = \arg\max_{\theta_n} \; \Re\left\{a_n \theta_n^* e_n\right\}, \text{ s.t. } (27c), \qquad (31a)$$

$$a_n = \arg\min_{a_n} \; (30a), \text{ s.t. } (27d). \qquad (31b)$$

An equivalent expression for Problem (31a) is given by

$$\max_{\varphi_n} \; \cos\left(-\varphi_n + \angle e_n\right), \quad \text{s.t. } (27c). \qquad (32a)$$

The optimal solution to the above problem is

$$\varphi_n = \angle e_n. \qquad (33)$$

Note that with the obtained solution of φ_n by employing (33), the value of $\cos\left(-\varphi_n + \angle e_n\right)$ can be calculated as 1, and hence, the term of $2\Re\left\{a_n \theta_n^* e_n\right\}$ in (30a) reduce to $2a_n |e_n|$.

Algorithm 3 Pair-wise ASO Algorithm

1: **Initialize** η^l, $\mathbf{a}_0^{(0)} \triangleq \mathbf{a}^l$, $\boldsymbol{\theta}_0^{(0)} \triangleq \boldsymbol{\theta}^l$, and inner-iteration index $m = 0$. With η^l, let the initial value of (29a) in l-th iteration as $\rho\left(\mathbf{a}_0^{(0)}, \boldsymbol{\theta}_0^{(0)}\right)$.
2: **repeat**
3: **Sequentially Optimize:** $\mathbf{a}_n^{(m+1)}$ and $\boldsymbol{\theta}_n^{(m+1)}$, $n = 1, \cdots, N$, by using (34) and (33);
4: **Calculate** $\rho\left(\mathbf{a}_N^{(m+1)}, \boldsymbol{\theta}_N^{(m+1)}\right)$;
5: **until** $\left|\rho\left(\mathbf{a}_N^{(m+1)}, \boldsymbol{\theta}_N^{(m+1)}\right) - \rho\left(\mathbf{a}_N^{(m)}, \boldsymbol{\theta}_N^{(m)}\right)\right| \leq \epsilon$, **output** $\mathbf{a}^{l+1} \triangleq \mathbf{a}_N^m$, $\boldsymbol{\theta}^{l+1} \triangleq \boldsymbol{\theta}_N^m$.

Algorithm 4 Price-mechanism-based Algorithm

1: **Initialize** $\mathbf{a}^0$, $\eta^{up,0}$, $\eta^{low,0}$, and iteration index $l = 0$.
2: **Calculate:** $\left(\mathbf{a}\left(\eta^l\right)\right)^{l+1}$ and $\left(\boldsymbol{\theta}\left(\eta^l\right)\right)^{l+1}$ by using Algorithm 3;
3: **Calculate:** $\eta^{l+1} = \frac{\eta^{up,l} + \eta^{low,l}}{2}$ and $g\left(\eta^{l+1}\right)$, and then **Update:** $\eta^{up,l+1}, \eta^{low,l+1}$;
4: If $\left|h^{l+1} - h^l\right| \leq \epsilon$, **terminate and output** $\eta^\star \triangleq \eta^l$, $\mathbf{a}^\star \triangleq \mathbf{a}^l$, $\boldsymbol{\theta}^\star \triangleq \boldsymbol{\theta}^l$. **Otherwise**, go to step 2.

Consequently, the solution to Problem (31b) is given in a semi-closed-form expression as follows

$$a_n = \min\left\{\frac{|e_n|}{\delta^2 \kappa^2 D_{n,n} + \Omega_{n,n} + \eta\left(T_{n,n} + \delta^2 \kappa^2\right)}, a_n^{\max}\right\}. \qquad (34)$$

Based on the above discussions, a pair of $\{a_n, \theta_n\}$ can be conducted by successively optimizing with the other $N-1$ pairs being fixed, and then repeating until converging. The details of the pair-wise ASO algorithm in l-th iteration are summarized in Algorithm 3, where $\boldsymbol{\theta}_n^{(m+1)}$ and $\mathbf{a}_n^{(m+1)}$ denote the optimized vectors after updating the n-th TE at the m-th inner-iteration of the l-th iteration for optimizing the transmissive coefficients. It can be readily verified Algorithm 3 is guaranteed to converge and obtain a high-quality suboptimal solution [50]. Then, the remaining work is to find the price $\eta^\star$ so that the complementary slackness condition for constraint (27b) is satisfied, i.e., $\eta\left(g\left(\mathbf{a}\left(\eta\right)\right) - \bar{P}_{\mathrm{ITS}}^{\max}\right) = 0$. Since $g\left(\mathbf{a}\left(\eta\right)\right)$ is a non-increasing function of η, the optimal solution of η can be obtained by employing the bisection search method [48]. Searching the optimal price $\eta^\star$ is similar to searching the multiplier $\varepsilon^\star$ when updating $\mathbf{W}$ in (20), and it is omitted for simplicity. Based on the above discussion, the details of the price-mechanism-based algorithm for alternating optimizing $\eta^\star$, $\mathbf{a}^\star$, and $\boldsymbol{\theta}^\star$ are provided in Algorithm 4.

D. Extension to Practical Active TE Models

In the above discussion, we follow the widely-employed ideal active TE model that the phase shift and amplification factor are perfectly decoupled, and hence they can be tuned independently and continuously. However, limited by hardware implementation, the assumption is impractical. Specifically, in practice, the phase shift of the TE generally can only take a finite number of discrete values from the discrete phase

Authorized licensed use limited to: University of Houston. Downloaded on July 02,2023 at 13:19:36 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Wireless Communications. 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/TWC.2023.3247906

set for economic reasons [40]. In addition, the amplification factor is phase-dependent due to the circuit limitation, which introduces the additional constraint between the amplification factor and phase shift [26]. Hence, we further discuss the two practical active TE models. It is worth noting that the proposed Algorithm 3 can be easily extended to optimize $\boldsymbol{\phi}$ under the two models after some modifications.

1) Discrete phase shift TE model: Let b denote the number of bits to represent the resolution levels of the active-ITS, the discrete phase set is represented by $\mathcal{D} = \left\{ \frac{2\pi g}{2^b} \middle| g = 0, 1, 2, \cdots, 2^b - 1 \right\}$. A widely employed way to obtain the discrete phases of TEs is projecting the continuous phase φ_n obtained by (33) to the nearest discrete phase in the set $\mathcal{D}$, i.e., $\varphi_n^{\mathcal{D}} = \arg\min_{\boldsymbol{\varphi}_n^{\mathcal{D}} \in \mathcal{D}} \left| \varphi_n^{\mathcal{D}} - \varphi_n \right|$.

2) Phase-dependent amplification factor TE model: In this case, the amplification factor and the phase shift of the n-th TE are coupled together, and hence the amplification factor becomes a function with respect to its phase, which is referred to $\hat{a}_n(\varphi_n)$, determined by the equivalent RLC circuit [27]. Under this model, by defining $\hat{\boldsymbol{\phi}} = \hat{\mathbf{a}} \circ \boldsymbol{\theta}$, Problem (28) can be reformulated by

$$\min_{\hat{\boldsymbol{\phi}}, \boldsymbol{\varphi}} \sum_{n=1}^{N} \left(\hat{a}_n(\varphi_n) \right)^2 \left(\delta^2 \kappa^2 D_{\mathrm{n,n}} + \eta T_{\mathrm{n,n}} + \eta \delta^2 \kappa^2 \right) + \hat{\boldsymbol{\phi}}^{\mathrm{H}} \boldsymbol{\Omega} \hat{\boldsymbol{\phi}} - \hat{\boldsymbol{\phi}}^{\mathrm{H}} \mathbf{c} - \mathbf{c}^{\mathrm{H}} \hat{\boldsymbol{\phi}}, \tag{35a}$$

$$\text{s.t. } \hat{\phi}_n = \hat{a}_n(\varphi_n) e^{j\varphi_n}, -\pi \leq \varphi_n \leq \pi, \forall n \in N. \tag{35b}$$

By employing the same operation as the transformation of Problem (28) into Problem (30), and defining $\bar{\Omega}_{n,n} = \Omega_{n,n} + \delta^2 \kappa^2 D_{n,n} + \eta \left(T_{n,n} + \delta^2 \kappa^2 \right)$, the above problem can be reformulated by

$$\max_{\varphi_n} 2\hat{a}_n(\varphi_n) |e_n| \cos(\angle e_n - \varphi_n) - \left(\hat{a}_n(\varphi_n) \right)^2 \bar{\Omega}_{n,n} \tag{36a}$$

$$\text{s.t. } -\pi \leq \varphi_n \leq \pi. \tag{36b}$$

An efficient way to solve the above non-linear optimization problems is employing the approximate model to fit the function in (36a) at the trust region [26]. However, to our best knowledge, for the mmWave band, the exact form of $\hat{a}_n(\varphi_n)$ is unknown yet. Nevertheless, it is known that $\hat{a}_n(\varphi_n)$ generally has a similar shape to [26, Fig. 3], where the minimum amplitude occurs near zero-phase and the maximum amplitude achieves at the phases of $\pm\pi$. To be specific, based on the fact that the objective function achieves maximum when φ_n slightly deviates away from $\angle e_n$, the trust region that encloses the optimal solution of Problem (36) can be determined in a similar way as that of [26]. Then, by choosing the proper value for the radius of the trust region and fitting a quadratic function through three points over the trust region, a closed-form approximate solution can be obtained [26].

E. Convergence and Complexity Analysis

The detailed description of the proposed BCD-based joint precoding algorithm is summarized in Algorithm 5, we update one of the variables while the others are fixed. Algorithm 5 guarantees to converge which can be proved as follows. For each updating step, with the fixed other variables, it can be readily proved that the value of $\mathcal{R}$ is monotonically increasing after each step. Additionally, since the power budget constraints at the BS and active-ITS, $\mathcal{R}$ is upper bounded by a finite value, thus Algorithm 5 guarantees to converge.

Algorithm 5 BCD-based Joint Precoding Algorithm

1: **Initialize** $\mathbf{W}^{(0)}$, $\mathbf{A}^{(0)}$, $\boldsymbol{\Theta}^{(0)}$,and iteration index $t = 0$.
2: **repeat**
3: **Update** $\mathbf{U}^{(\mathrm{t}+1)}$ by using (11);
4: **Update** $\mathbf{F}^{(\mathrm{t}+1)}$ by using (12);
5: **Update** $\mathbf{W}^{(\mathrm{t}+1)}$ by employing Algorithm 2;
6: **Update** $\boldsymbol{\Theta}^{(\mathrm{t}+1)}$ and $\mathbf{A}^{(\mathrm{t}+1)}$ by employing Algorithm 4;
7: **until** $\left| \mathcal{R}^{(\mathrm{t}+1)} - \mathcal{R}^{(\mathrm{t})} \right| < \varepsilon$, **output** $\mathbf{W}^{\star} \triangleq \mathbf{W}^{(\mathrm{t})}$, $\mathbf{A}^{\star} \triangleq \mathbf{A}^{(\mathrm{t})}$, and $\boldsymbol{\Theta}^{\star} \triangleq \boldsymbol{\Theta}^{(\mathrm{t})}$.

Now we briefly analyze the total computational complexity of Algorithm 5. The complexities for updating $\mathbf{U}$ and $\mathbf{F}$ in Steps 3, 4 are $\mathcal{O}\left(KM_r^3\right)$ and $\mathcal{O}\left(Ks^3\right)$, respectively. In Step 5, the major computational cost of Algorithm 2 is on calculating the inverse operation, which is $\mathcal{O}\left(M_t^3\right)$, and the number of inner-iterations required for employing Algorithm 2 is $\mathcal{I}_1$. In Step 6, the complexity of Algorithm 3 is $\mathcal{O}\left(N^2\right)$ for each inner-iteration, and the number of required inner-iterations is $\mathcal{I}_2$. Therefore, the total complexity of Algorithm 5 is $\mathcal{O}\left(\mathcal{I}_0 \max\left\{ \mathcal{I}_1 \log\left(\frac{\varepsilon^{up} - \varepsilon^{low}}{\epsilon} \right) KM_t^3, \mathcal{I}_2 \log\left(\frac{\eta^{up} - \eta^{low}}{\epsilon} \right) N^2 \right) \right\}$, where $\mathcal{I}_0$ is the total number of iterations of Algorithm 5, and $\log\left(\frac{\varepsilon^{up} - \varepsilon^{low}}{\epsilon} \right)$ and $\log\left(\frac{\eta^{up} - \eta^{low}}{\epsilon} \right)$ are the required iterations for searching the ϵ-optimal solutions of multipliers and price factor, respectively.

IV. The Proposed Block-Amplifying Architecture of Active-ITS

In this section, we provide a block-amplifying architecture of active-ITS to partially remove the circuit components for power-amplifying, which is beneficial for reducing the surface size and hardware cost. More specifically, as illustrated in Fig. 3, for the block-amplifying architecture, the total active TEs are divided into several blocks, and the TEs in each block are assumed to share the same power amplifier. In practice, the phase shifts of elements of the passive-RIS are generally controlled by row or column due to the circuit limitation and implementation cost [51]. Therefore, from the hardware implementation point of view, TEs can be partitioned by row or column, and then deploy a circuit for power amplification, where TEs in each row(column) share the same amplification factor. Although this block partition strategy might not be optimal for performance, it facilitates the practical hardware implementation of the active-ITS.

Discussion of the inevitable performance degradation: Compared with element-amplifying architecture, the block-amplifying architecture can reduce the scale of the power-amplifying circuit, which makes it further suitable for application to the space-limited scenario. In addition, by partially removing the circuit components for power amplifying, the total static circuit power consumption of the block-amplifying architecture reduces to $P_{\mathrm{ITS}}^{\mathrm{circ}} = NP_{\mathrm{SW}} + RP_{\mathrm{DC}}$, where R ($R \leq N$) denotes the number of blocks. Clearly, the

Authorized licensed use limited to: University of Houston. Downloaded on July 02,2023 at 13:19:36 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Wireless Communications. 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/TWC.2023.3247906

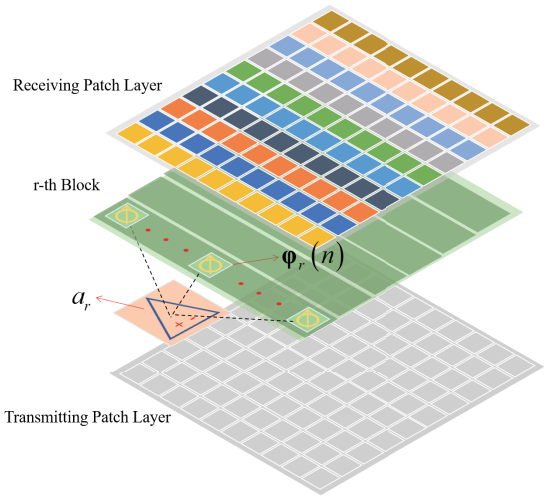

Fig. 3. Block-amplifying architecture of the active-ITS.

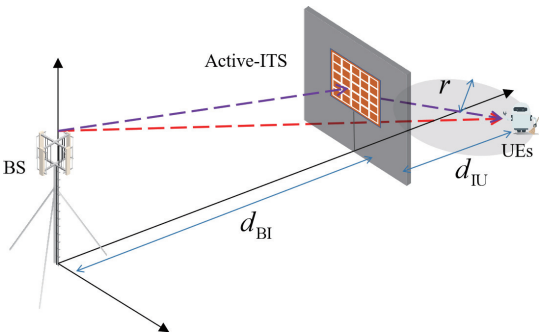

Fig. 4. The simulated active-ITS empowered outdoor-to-indoor mmWave communication scenario.

power consumption of the block-amplifying architecture is less than that of the element-amplifying architecture, i.e., $N\left(P_{\mathrm{SW}}+P_{\mathrm{DC}}\right)$, and according to constraint (8c), the more power will be used to perform the amplitude-magnifying, while it can counteract some inevitable performance loss caused by such an architecture.

Hereinafter, we modify the proposed algorithm to maximize the WSR of the system with block-amplifying ITS. Note that the variables $\mathbf{U}$, $\mathbf{F}$, $\mathbf{W}$, and $\boldsymbol{\Theta}$ can be directly obtained as that under element-amplifying architecture, and hence, the remaining work is to optimize the power amplifying factor matrix $\mathbf{A}$. In the following, we extend the proposed ASO algorithm from the pair-wise form into the block-wise form, which can optimize $\mathbf{A}$ in a block-by-block manner.

Remark: *From a mathematical point of view, since the element-amplifying architecture can be interpreted as a special case ($R=N$) of the block-amplifying architecture, the proposed ASO Algorithm under the block-wise form is a superset of the pair-wise form.*

Particularly, we partition the power amplifying factor $\mathbf{a}$ into R blocks, i.e., $\mathbf{a}=\left[\mathbf{a}_1^{\mathrm{T}},\mathbf{a}_2^{\mathrm{T}},\cdots,\mathbf{a}_R^{\mathrm{T}}\right]$, where $\mathbf{a}_r\in\mathbb{R}^{\hat{N}_r\times 1}$ is the power amplifying factor vector corresponding to $\hat{N}_r$ TEs in the r-th block, which is given by $\mathbf{a}_r=\left[\mathbf{a}_r\left(1\right),\mathbf{a}_r\left(2\right),\cdots,\mathbf{a}_r\left(\hat{N}_r\right)\right]^{\mathrm{T}}=a_r\mathbf{1}_{\hat{N}_r}^{\mathrm{T}}$, where $\mathbf{a}_r\left(m\right)$ is the $\left(\sum_{l=1}^{r-1}\hat{N}_l+m\right)$-th element of $\mathbf{a}$ with $r=2,3,\cdots,R$ and $m=1,2,\cdots,\hat{N}_r$. Here, let us revisit Problem (27). First, $\left(\mathbf{a}^{\mathrm{T}}\circ\boldsymbol{\theta}^{\mathrm{H}}\right)\boldsymbol{\Omega}\left(\mathbf{a}\circ\boldsymbol{\theta}\right)$ can be expanded as the following form

$$\begin{aligned}&\left(\mathbf{a}_r^{\mathrm{T}}\circ\boldsymbol{\theta}_r^{\mathrm{H}}\right)\boldsymbol{\Omega}_{r,r}\left(\mathbf{a}_r\circ\boldsymbol{\theta}_r\right)+\sum_{i\neq r}^{R}\sum_{j=1}^{R}\left(\mathbf{a}_j^{\mathrm{T}}\circ\boldsymbol{\theta}_j^{\mathrm{H}}\right)\boldsymbol{\Omega}_{j,i}\left(\mathbf{a}_i\circ\boldsymbol{\theta}_i\right)\\&+\sum_{i\neq r}^{R}\left(\left(\mathbf{a}_i^{\mathrm{T}}\circ\boldsymbol{\theta}_i^{\mathrm{H}}\right)\boldsymbol{\Omega}_{i,r}\left(\mathbf{a}_r\circ\boldsymbol{\theta}_r\right)+\left(\mathbf{a}_r^{\mathrm{T}}\circ\boldsymbol{\theta}_r^{\mathrm{H}}\right)\boldsymbol{\Omega}_{r,i}\left(\mathbf{a}_i\circ\boldsymbol{\theta}_i\right)\right),\end{aligned}\tag{37}$$

where $\boldsymbol{\Omega}=\left[\boldsymbol{\Omega}_1,\cdots,\boldsymbol{\Omega}_R\right]$ with $\boldsymbol{\Omega}_r=\left[\boldsymbol{\Omega}_{1,r},\cdots,\boldsymbol{\Omega}_{R,r}\right]^{\mathrm{T}}$, and $\boldsymbol{\theta}=\left[\boldsymbol{\theta}_1^{\mathrm{T}},\cdots,\boldsymbol{\theta}_R^{\mathrm{T}}\right]^{\mathrm{T}}$ with $\boldsymbol{\theta}_r=\left[\boldsymbol{\theta}_r\left(1\right),\cdots,\boldsymbol{\theta}_r\left(\hat{N}_r\right)\right]^{\mathrm{T}}$ where $\boldsymbol{\theta}_r\left(m\right)$ denotes the $\left(\sum_{l=1}^{r-1}\hat{N}_l+m\right)$-th element of $\boldsymbol{\theta}$ with $r=2,\cdots,R$ and $m=1,\cdots,\hat{N}_r$.

Further, we have $\left(\mathbf{a}_i^{\mathrm{T}}\circ\boldsymbol{\theta}_i^{\mathrm{H}}\right)\boldsymbol{\Omega}_{i,r}\left(\mathbf{a}_r\circ\boldsymbol{\theta}_r\right)=a_ra_i\left(\mathbf{1}_{\hat{N}_r}^{\mathrm{T}}\circ\boldsymbol{\theta}_r^{\mathrm{H}}\right)\boldsymbol{\Omega}_{r,i}\left(\mathbf{1}_{\hat{N}_i}\circ\boldsymbol{\theta}_i\right)=a_ra_i\boldsymbol{\theta}_r^{\mathrm{H}}\boldsymbol{\Omega}_{r,i}\boldsymbol{\theta}_i$, and $\boldsymbol{\Omega}_{i,r}=\boldsymbol{\Omega}_{r,i}^{\mathrm{H}}$. Similarly, we have $\left(\mathbf{a}_r^{\mathrm{T}}\circ\boldsymbol{\theta}_r^{\mathrm{H}}\right)\boldsymbol{\Omega}_{r,r}\left(\mathbf{a}_r\circ\boldsymbol{\theta}_r\right)=a_r^2\boldsymbol{\theta}_r^{\mathrm{H}}\boldsymbol{\Omega}_{r,r}\boldsymbol{\theta}_r$ and $\left(\mathbf{a}^{\mathrm{T}}\circ\boldsymbol{\theta}^{\mathrm{H}}\right)\mathbf{c}=a_r\boldsymbol{\theta}_r^{\mathrm{H}}\mathbf{c}_r+\sum_{i\neq r}^{R}a_i\boldsymbol{\theta}_i^{\mathrm{H}}\mathbf{c}_i$. By defining $\mathbf{d}=\mathrm{Vecd}\left(\mathbf{D}\right)$ and $\mathbf{t}=\mathrm{Vecd}\left(\mathbf{T}\right)$, and then partition the two vector into R blocks, i.e., $\mathbf{d}=\left[\mathbf{d}_1^{\mathrm{T}},\mathbf{d}_2^{\mathrm{T}},\cdots,\mathbf{d}_R^{\mathrm{T}}\right]$ and $\mathbf{t}=\left[\mathbf{t}_1^{\mathrm{T}},\mathbf{t}_2^{\mathrm{T}},\cdots,\mathbf{t}_R^{\mathrm{T}}\right]$, we have $\mathrm{Tr}\left(\boldsymbol{\Phi}^{\mathrm{H}}\mathbf{D}\boldsymbol{\Phi}\right)=a_r^2\sum_{p=1}^{\hat{N}_r}\mathbf{d}_r\left(p\right)+\sum_{i\neq r}^{R}a_i^2\sum_{q=1}^{\hat{N}_i}\mathbf{d}_i\left(q\right)$ and $\mathrm{Tr}\left(\boldsymbol{\Phi}^{\mathrm{H}}\boldsymbol{\Phi}\mathbf{T}\right)=a_r^2\sum_{p=1}^{\hat{N}_r}\mathbf{t}_r\left(p\right)+\sum_{i\neq r}^{R}a_i^2\sum_{q=1}^{\hat{N}_i}\mathbf{t}_i\left(q\right)$. Similarly, we have $\mathrm{Tr}\left(\boldsymbol{\Phi}^{\mathrm{H}}\boldsymbol{\Phi}\right)=a_r^2\hat{N}_r+\sum_{i\neq r}^{R}a_i^2\hat{N}_i$.

Consequently, Problem (27) can be reformulated by

$$\min_{a_r} f\left(a_r\right)=a_r^2m_r+2\Re\left\{a_rz_r\right\}+\chi_r,\tag{38a}$$

$$\mathrm{s.t.}\ a_r^2g_r\leq\hat{P}_{\mathrm{ITS}}^{\max},0\leq a_r\leq a_r^{\max},\tag{38b}$$

where $m_r=\boldsymbol{\theta}_r^{\mathrm{H}}\boldsymbol{\Omega}_{r,r}\boldsymbol{\theta}_r+\delta^2\kappa^2\sum_{p=1}^{\hat{N}_r}\mathbf{d}_r\left(p\right)$, $z_r=\sum_{i\neq r}^{R}a_i\boldsymbol{\theta}_r^{\mathrm{H}}\boldsymbol{\Omega}_{r,i}\boldsymbol{\theta}_i-\boldsymbol{\theta}_r^{\mathrm{H}}\mathbf{c}_r$, $g_r=\sum_{p=1}^{\hat{N}_r}\mathbf{t}_r\left(p\right)+\kappa^2\delta^2\hat{N}_r$, $\hat{P}_{\mathrm{ITS}}^{\max}=\bar{P}_{\mathrm{ITS}}^{\max}-\sum_{i\neq r}^{R}a_i^2\sum_{q=1}^{\hat{N}_i}\mathbf{t}_i\left(q\right)-\kappa^2\delta^2\sum_{i\neq r}^{R}a_i^2\hat{N}_i$, and $\chi_r=\sum_{i\neq r}^{R}\sum_{j=1}^{R}a_ja_i\boldsymbol{\theta}_j^{\mathrm{H}}\boldsymbol{\Omega}_{j,i}\boldsymbol{\theta}_i+2\Re\left\{\sum_{i\neq r}^{R}a_i\boldsymbol{\theta}_i^{\mathrm{H}}\mathbf{c}_i\right\}+\kappa^2\delta^2\sum_{i\neq r}^{R}a_i^2\sum_{q=1}^{\hat{N}_i}\mathbf{d}_i\left(q\right)$ denotes a constant value with respect to a_r and does not affect its solution.

Note that the above problem can be solved similarly to Problem (31b) and hence it is omitted for brevity.

Based on the above discussions, $\mathbf{a}_r,\forall r$ can be conducted by successively optimizing with the other $R-1$ blocks being fixed, and then repeating until the convergence is attained [50].

V. Simulation Result

In this section, numerical results are provided to evaluate the WSR achieved by the proposed BCD-based joint precoding algorithm under the two architectures of active-ITSs.

We consider a three-dimensional setup as illustrated in Fig. 4. The BS, ITS, and UEs are assumed in a straight line, and their heights of them are 3m, 6m, and 1.5m, respectively, and the total horizontal distances $d_{\mathrm{BU}}=80$ m. The $K=4$ UEs are uniformly and randomly distributed in a circle centered at $(0\,\mathrm{m},70\,\mathrm{m})$ with a radius of 10 m. For the array response

Authorized licensed use limited to: University of Houston. Downloaded on July 02,2023 at 13:19:36 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Wireless Communications. 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/TWC.2023.3247906

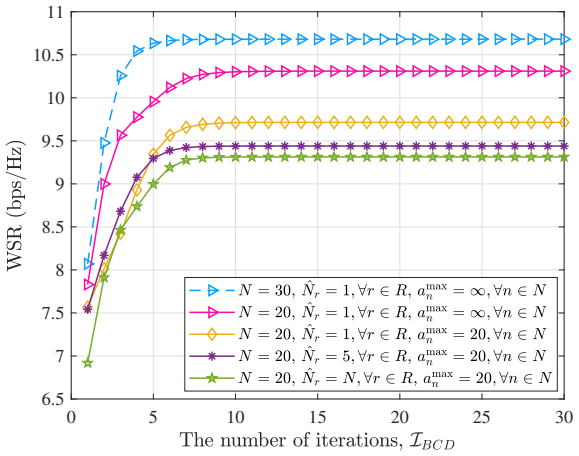

Fig. 5. The WSR against the number of iterations.

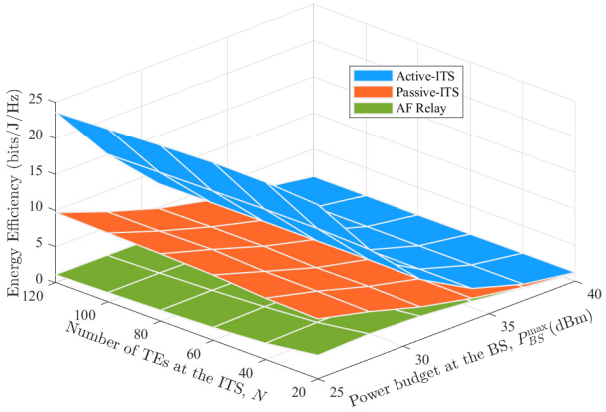

Fig. 6. The energy efficiency achieved by active/passive-ITS and AF relay.

vector, the azimuth angles of arrival and departure are uniformly distributed in the interval $[-\pi, \pi)$. The elevation angles of arrival and departure are uniformly distributed in the interval $\left[\frac{-\pi}{2}, \frac{\pi}{2}\right)$. The complex gains of LOS paths $\alpha_l, \forall l$ and $\beta_p, \forall p$ are independently distributed with $\mathcal{CN}\left(0, 10^{-0.1\mathrm{PL}}\right)$, where the distance-dependent path loss PL for both the LoS and the NLoS paths can be modeled by $\mathrm{PL} = \mathrm{PL}_0 + 10b\log_{10}(D) + \zeta$ dB, where D is the individual link distance and ζ denotes lognormal shadow fading following $\zeta \sim \mathcal{N}\left(0, \sigma_\zeta^2\right)$. As the real-world channel measurements for the carrier frequency of 28 GHz suggested in [3, Table I], the parameters are set as $\mathrm{PL}_0 = 61.4$, $b = 2$, $\sigma_\zeta = 5.8$ for LoS paths, and $\mathrm{PL}_0 = 72$, $b = 2.92$, $\sigma_\zeta = 8.7$ for NLoS paths. The bandwidth $B = 251$ MHz and the noise power is $\sigma^2 = -90$ dBm.

Unless otherwise stated, the other parameters are set as follows: the paths $L_{\mathrm{BI}} = 4$, $P_{\mathrm{IU}} = 4$, $Q_{\mathrm{BU}} = 4$, the distance $d_{\mathrm{BI}} = 70$ m, the circuit power consumption for phase-shifting and power-amplifying are $P_{\mathrm{SW}} = -10$ dBm and $P_{\mathrm{DC}} = -5$ dBm [35], the power budgets of the BS and active-ITS are $P_{\mathrm{BS}}^{\max} = 25$ dBm and $P_{\mathrm{ITS}}^{\max} = 15$ dBm, the number of TEs of $N = 4 \times 5$, the number of antennas of $M_t = 2 \times 3$ and $M_r = 2 \times 2$, the weighting factor of $\alpha_k = 1$, the penetration efficiency of $\kappa = 0.8$, the power conversion efficiency of $\zeta = 1$, the static circuit power at the BS of $P_{\mathrm{BS}}^{\mathrm{circ}} = 50$ dBm, and the target accuracy of $\varepsilon = 10^{-3}$.

A. Performance Comparison

1) Convergence behavior: We first show the convergence of the proposed algorithm in Fig. 5. It can be observed that the WSR converges to corresponding stationary points after a few iterations. Due to the number of TEs of the active-ITS is small, the convergence speed is quick. In addition, the impacts of the maximum amplification factor at each TE, and the number of TEs at each block on the convergence speed are slight.

2) Energy efficiency comparison among the active/passive ITS and AF relay: Specifically, the energy efficiency is determined by $\mathcal{E} = \frac{\mathcal{R}}{P_{\mathrm{ITS/AF}}^{a/p} + P_{\mathrm{BS}}^{\mathrm{circ}} + P_{\mathrm{BS}}^{\max}}$. For the active-ITS, $P_{\mathrm{ITS}}^{a} = P_{\mathrm{ITS}}^{\max} + P_{\mathrm{ITS}}^{\mathrm{circ}}$, and for the passive-ITS, $P_{\mathrm{ITS}}^{\max} = 0$, and $P_{\mathrm{ITS}}^{\mathrm{circ}} = NP_{\mathrm{SW}}$. While for the AF relay, $P_{\mathrm{AF}} = P_{\mathrm{AF}}^{\max} + NP_{\mathrm{RF}}^{\mathrm{circ}}$ with $P_{\mathrm{RF}}^{\mathrm{circ}} = 15$ dBm being the static power to supply single energy-intensive RF chain [28]. It can be seen from Fig. 6, the active/passive-ITS achieve higher energy efficiencies than the AF relay. More specifically, the increased N has no discernible influence on the energy efficiency of the active/passive-ITS, while significantly impacting that of the AF relay. This can be attributed to that the static circuit power consumption of each TE is small, while each RF chain is power-hungry. Additionally, in the large $P_{\mathrm{BS}}^{\max}$ region, both types of ITS have almost the same energy efficiency. This is because as $P_{\mathrm{BS}}^{\max}$ increases, the incident power at the active-ITS becomes large, resulting in less amplification gain at the active-ITS. It can conclude from the figure, that active-ITS is a low implementation-cost and high energy-efficiency option for mmWave outdoor-to-indoor communication.

3) Impact of system parameters: Hereinafter, we investigate the impact of key system parameters on the WSR. We consider different architectures of the active-ITS as well as other benchmark schemes, as indicated by the following terminology

- **EA-ITS-ASO/SDR/CVX:** The schemes denote the performance achieved by the proposed Algorithm 3, semi-definite relaxation (SDR) technique [33], [36], and CVX toolbox [45] to solve Problem (25) under the element-amplifying architecture of active-ITS, with the active load can provide an unlimited gain (i.e., $a^{\max} = \infty$);
- **BA-ITS, 10/15/20:** All TEs are partitioned by row that each block consists of $\hat{N}_r = 5, \forall r$ TEs and the maximum value of amplification factor $a^{\max} = 10/15/20$ dB;
- **IA-ITS, 10/15/20:** All TEs share the same amplifier (i.e., $R = 1$) that they have identical amplitudes [36] with the maximum amplification factor $a^{\max} = 10/15/20$ dB;
- **AF Relay/Passive-ITS:** Exploiting the AF relay or the conventional passive-ITS in the place of the active-ITS.
- **Without-ITS:** Only exists the direct links.

For fair comparisons, we assume the power budget at the BS for the Passive-ITS and Without-ITS schemes are $P_{\mathrm{ITS}}^{\max} + P_{\mathrm{BS}}^{\max}$, then all ITS-related schemes have the same total power consumption. All the simulation results as follows are averaged over 500 independent channel realizations.

Authorized licensed use limited to: University of Houston. Downloaded on July 02,2023 at 13:19:36 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Wireless Communications. 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/TWC.2023.3247906

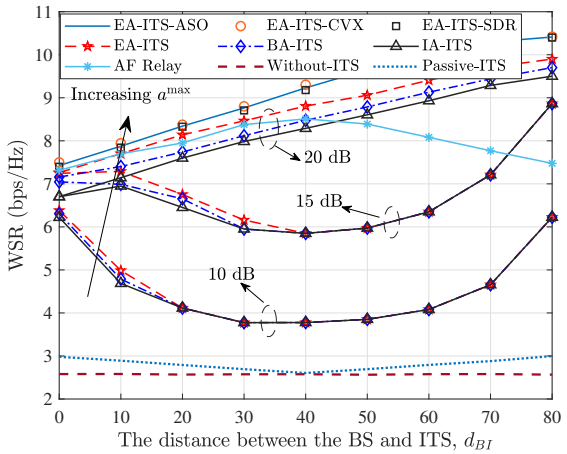

Fig. 7. The WSR against the distance between the BS and ITS.

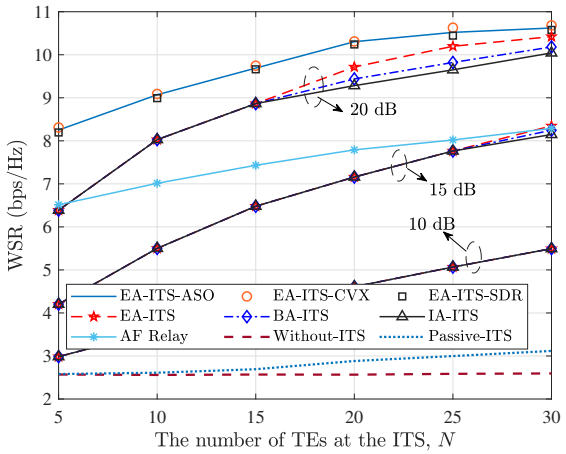

Fig. 8. The WSR against the number of TEs at the ITS.

In Fig. 7, we investigate the WSR for different locations of the ITS by varying d_{BI}. One important observation is that $a^{\max}$ is one of the dominant parameters for boosting the WSR. More specifically, even for the active-ITS schemes with $a^{\max} = 10$ dB, the WSR has been enhanced compared to that of Passive-ITS scheme, but the trends of the corresponding curves are still similar to that of Passive-ITS scheme. While for the active-ITS scheme with $a^{\max} = 20$ dB, with increasing d_{BI}, the WSR monotonically increased, its trend is different from Passive-ITS scheme. This observation can be explained as follows. With increasing d_{BI}, the channel gain of $\mathbf{G}$ becomes weaker, resulting in less incident signal power at the active-ITS, and according to the constraint (8c), the active-ITS can provide a higher amplification gain, which compensates for the attenuation caused by the multiplicative path loss fading effect and contributes to increasing the WSR. However, since the amplification factor is limited by the circuit loads, the attenuation cannot be compensated well for the active-ITS with a small value of $a^{\max}$. This interesting observation indicates that the active-ITS should be deployed closer to UEs to fully unleash their potential on achieving a better WSR performance, especially for the active-ITS with a small value of $a^{\max}$. Intuitively speaking, the deployment strategy meets the requirements of a practical outdoor-to-indoor communication scenario where the active-ITS is usually very close to UEs, i.e., $d_{\mathrm{IU}} \ll d_{\mathrm{BI}}$. In addition, it can observe that the proposed algorithm achieves almost the same WSR compared to SDR and CVX, while the proposed algorithm has a much smaller computational complexity, which demonstrates its superiority. Moreover, it can observe that the active-ITS schemes perform constantly better than Passive-ITS scheme in all locations, which implies that the active-ITS brings a significant performance gain due to the amplified signal is only attenuated once. More importantly, it can be seen that the WSR of EA-ITS, 20 scheme outperforms that of BA/IA-ITS, 20 schemes. This observation indicates the importance of carefully optimizing the power amplification factor at the active-ITS for enhancing the WSR. While for the schemes with small $a^{\max}$, the performance degradation caused by the block architecture is slight due to amplification factors of TEs can easily reach the maximum value.

The WSR performances versus the number of TEs are shown in Fig. 8. It illustrates that with the increasing number of TEs, both curves of the active-ITS and Passive-ITS schemes monotonically increase due to more TEs introducing more DoFs to proactively configure signals which yields a higher WSR performance. The usage of active ITSs can greatly reduce the number of TEs compared with the passive-ITS for reaping a given performance level, and hence greatly decreases both the size and the complexity of ITS, which is beneficial for embedding in space-limited and aesthetic-needed building structures. It can be explained by that when equipped with a small number of active TEs in the active-ITS, according to the constraint in (8c), a larger power amplification factor will be allocated at each active TE. In addition, it can observe that with a small number of TEs, the performance gap between EA-ITS and BA-ITS is slight, while it in the large N region also is acceptable. This is because, for a small number of TEs, the amplifying factor can easily be the maximum value. In this case, constraint (8c) is inactive but constraint (8d) is always active and dominant in the performance. As the number of TEs increases, constraint (8c) will be active, resulting in a small amplification factor at each TE, and become the dominant parameter for boosting the WSR. Moreover, compared to the element-amplifying architecture, the block-amplifying architecture reduces the power left for the static circuit, and hence it has more power to amplify the signal, which can compensate for some of the loss of performance and reduce the inevitable performance degradation.

We plot the WSR performance against the power budget at the BS in Fig. 9. As expected, it can be observed that all the WSR performances of all schemes increase as $P_{\mathrm{BS}}^{\max}$ increase. More specifically, it can be concluded from Fig. 9, the active-ITS scheme can significantly reduce the power consumption compared to Passive-ITS scheme when achieving a given performance level. In addition, it can be observed that all active-ITS schemes have almost the same WSR in the large $P_{\mathrm{BS}}^{\max}$ region, this is because as $P_{\mathrm{BS}}^{\max}$ increases, the

Authorized licensed use limited to: University of Houston. Downloaded on July 02,2023 at 13:19:36 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Wireless Communications. 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/TWC.2023.3247906

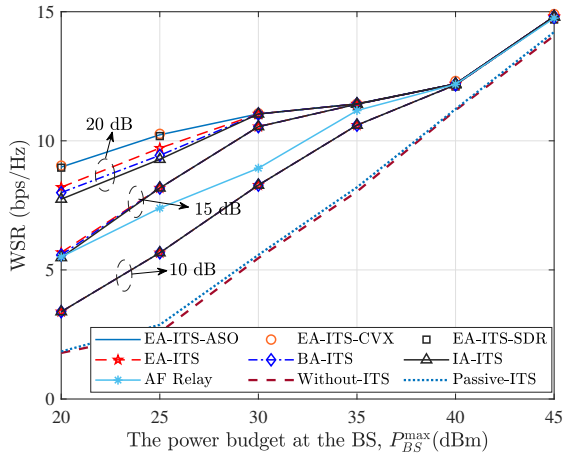

Fig. 9. The WSR against the power budget at the BS.

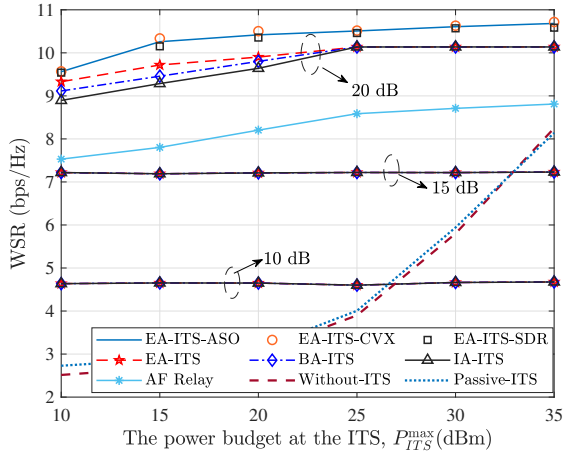

Fig. 10. The WSR against the amplifying power budget at the active-ITS.

incident power at the active-ITS become large, resulting in less amplification gain at the active-ITS. This is also the reason that when $P_{\mathrm{BS}}^{\max}$ is weak, the WSR is sensitive to the maximum value of the amplification factor.

Fig. 10 illustrates the WSR against the amplifying power budget. One important observation is that for the EA-ITS, 20 scheme, as amplifying power increases, the WSR first increases, and then tends to converge. This is because in the large $P_{\mathrm{ITS}}^{\max}$ region, the constraint (8c) will be inactive while constraint (8d) dominant the performance, that it, all TEs reach the maximum value of the amplification factor. This is also the reason that the curves with $a^{\max} = 10/15$ do not change as $P_{\mathrm{ITS}}^{\max}$ increases. Therefore, it can be concluded that the amplification gain provided by the active-ITS is limited by the maximum value of the amplification factor [35].

VI. Conclusion

In this paper, an active-ITS-empowered outdoor-to-indoor mmWave communication system was investigated, where the active-ITS could greatly reduce the number of TEs while maintaining a given performance compared with the passive-ITS. To jointly optimize the linear precoding matrix of the BS and transmissive coefficients of the active-ITS, we formulated a WSR maximization problem, and then a computational efficiently BCD-based joint precoding algorithm was proposed for solving it. In order to further reduce the size and hardware cost of active-ITS, a block-amplifying architecture was proposed to partially remove the hardware components for power amplifying. And then we extended the proposed algorithm into the block-wise form for optimizing the transmissive coefficients of the active-ITS under the block-amplifying architecture. Simulation results demonstrated that the active-ITS could significantly enhance the system performance, and the performance degradation caused by the block-amplifying ITS architecture was negligible.

References

[1] R. W. Heath, N. González-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, "An Overview of Signal Processing Techniques for Millimeter Wave MIMO Systems," *IEEE J. Sel. Top. Signal Process.*, vol. 10, no. 3, pp. 436–453, Apr. 2016.

[2] M. Di Renzo, "Stochastic Geometry Modeling and Analysis of Multi-Tier Millimeter Wave Cellular Networks," *IEEE Trans. Wireless Commun.*, vol. 14, no. 9, pp. 5038–5057, Sep. 2015.

[3] M. R. Akdeniz, Y. Liu, M. K. Samimi, S. Sun, S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter Wave Channel Modeling and Cellular Capacity Evaluation," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1164–1179, Jun. 2014.

[4] X. Hu, C. Zhong, Y. Zhu, X. Chen, and Z. Zhang, "Programmable Metasurface-Based Multicast Systems: Design and Analysis," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 8, pp. 1763–1776, Aug. 2020.

[5] M. A. ElMossallamy, H. Zhang, L. Song, K. G. Seddik, Z. Han, and G. Y. Li, "Reconfigurable Intelligent Surfaces for Wireless Communications: Principles, Challenges, and Opportunities," *IEEE Trans. Cogn. Commun. Netw.*, vol. 6, no. 3, pp. 990–1002, Sep. 2020.

[6] M. Di Renzo, A. Zappone, M. Debbah, M. S. 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 J. Sel. Areas Commun.*, vol. 38, no. 11, pp. 2450–2525, Nov. 2020.

[7] Y. Chen, Y. Wang, J. Zhang, P. Zhang, and L. Hanzo, "Reconfigurable Intelligent Surface (RIS)-Aided Vehicular Networks: Their Protocols, Resource Allocation, and Performance," *IEEE Veh. Technol. Mag.*, vol. 17, no. 2, pp. 26–36, Jun. 2022.

[8] G. Zhou, C. Pan, H. Ren, K. Wang, M. Elkashlan, and M. D. Renzo, "Stochastic Learning-Based Robust Beamforming Design for RIS-Aided Millimeter-Wave Systems in the Presence of Random Blockages," *IEEE Trans. Veh. Technol.*, vol. 70, no. 1, pp. 1057–1061, Jan. 2021.

[9] P. Wang, J. Fang, L. Dai, and H. Li, "Joint Transceiver and Large Intelligent Surface Design for Massive MIMO mmWave Systems," *IEEE Trans. Wireless Commun.*, vol. 20, no. 2, pp. 1052–1064, Feb. 2021.

[10] Y. Chen, Y. Wang, and L. Jiao, "Robust Transmission for Reconfigurable Intelligent Surface Aided Millimeter Wave Vehicular Communications With Statistical CSI," *IEEE Trans. Wireless Commun.*, vol. 21, no. 2, pp. 928–944, Feb. 2022.

[11] Y. Xiu, J. Zhao, E. Basar, M. D. Renzo, W. Sun, G. Gui, and N. Wei, "Uplink Achievable Rate Maximization for Reconfigurable Intelligent Surface Aided Millimeter Wave Systems With Resolution-Adaptive ADCs," *IEEE Wireless Commun. Lett.*, vol. 10, no. 8, pp. 1608–1612, Aug. 2021.

[12] H. Niu, Z. Chu, F. Zhou, C. Pan, D. W. K. Ng, and H. X. Nguyen, "Double Intelligent Reflecting Surface-Assisted Multi-User MIMO Mmwave Systems With Hybrid Precoding," *IEEE Trans. Veh. Technol.*, vol. 71, no. 2, pp. 1575–1587, Feb. 2022.

[13] C. Feng, W. Shen, J. An, and L. Hanzo, "Joint Hybrid and Passive RIS-Assisted Beamforming for MmWave MIMO Systems Relying on Dynamically Configured Subarrays," *IEEE Internet Things J.*, vol. 9, no. 15, pp. 13 913–13 926, Aug. 2022.

[14] V. Jamali, A. M. Tulino, G. Fischer, R. R. Müller, and R. Schober, "Intelligent Surface-Aided Transmitter Architectures for Millimeter-Wave Ultra Massive MIMO Systems," *IEEE Open J. Commun. Soc.*, vol. 2, pp. 144–167, Dec. 2021.

Authorized licensed use limited to: University of Houston. Downloaded on July 02,2023 at 13:19:36 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Wireless Communications. 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/TWC.2023.3247906

[15] D. W. K. Ng, M. Breiling, C. Rohde, F. Burkhardt, and R. Schober, "Energy-Efficient 5G Outdoor-to-Indoor Communication: SUDAS Over Licensed and Unlicensed Spectrum," *IEEE Trans. Wireless Commun.*, vol. 15, no. 5, pp. 3170–3186, May. 2016.

[16] H. Zhao, R. Mayzus, S. Sun, M. Samimi, J. K. Schulz, Y. Azar, K. Wang, G. N. Wong, F. Gutierrez, and T. S. Rappaport, "28 GHz Millimeter Wave Cellular Communication Measurements for Reflection and Penetration Loss in and Around Buildings in New York City," in *2013 IEEE Int. Conf. Commun. (ICC)*, Jun. 2013, pp. 5163–5167.

[17] M. Nemati, B. Maham, S. R. Pokhrel, and J. Choi, "Modeling RIS Empowered Outdoor-to-Indoor Communication in mmWave Cellular Networks," *IEEE Trans. Commun.*, vol. 69, no. 11, pp. 7837–7850, Nov. 2021.

[18] Q. Wu and R. Zhang, "Beamforming Optimization for Wireless Network Aided by Intelligent Reflecting Surface With Discrete Phase Shifts," *IEEE Trans. Commun.*, vol. 68, no. 3, pp. 1838–1851, Mar. 2020.

[19] N. T. Nguyen, D. Vu, K. Lee, and M. Juntti, "Hybrid Relay-Reflecting Intelligent Surface-Assisted Wireless Communications," *IEEE Trans. Veh. Technol.*, vol. 71, no. 6, pp. 6228–6244, Jun. 2022.

[20] E. Basar and H. V. Poor, "Present and Future of Reconfigurable Intelligent Surface-Empowered Communications [Perspectives]," *IEEE Signal Process. Mag.*, vol. 38, no. 6, pp. 146–152, Nov. 2021.

[21] W. Mei and R. Zhang, "Distributed Beam Training for Intelligent Reflecting Surface Enabled Multi-Hop Routing," *IEEE Wireless Commun. Lett.*, vol. 10, no. 11, pp. 2489–2493, Nov. 2021.

[22] B. Zheng, C. You, and R. Zhang, "Double-IRS Assisted Multi-User MIMO: Cooperative Passive Beamforming Design," *IEEE Trans. Wireless Commun.*, vol. 20, no. 7, pp. 4513–4526, Jul. 2021.

[23] S. Zeng, H. Zhang, B. Di, Y. Tan, Z. Han, H. V. Poor, and L. Song, "Reconfigurable Intelligent Surfaces in 6G: Reflective, Transmissive, or Both?" *IEEE Commun. Lett.*, vol. 25, no. 6, pp. 2063–2067, Jun. 2021.

[24] K. Liu, Z. Zhang, L. Dai, and L. Hanzo, "Compact User-Specific Reconfigurable Intelligent Surfaces for Uplink Transmission," *IEEE Trans. Commun.*, vol. 70, no. 1, pp. 680–692, Jan. 2022.

[25] C. Pfeiffer and A. Grbic, "Metamaterial Huygens' Surfaces: Tailoring Wave Fronts with Reflectionless Sheets," *Phys. Rev. Lett.*, vol. 110, no. 19, pp. 197 401–197 401, May. 2013.

[26] S. Abeywickrama, R. Zhang, Q. Wu, and C. Yuen, "Intelligent Reflecting Surface: Practical Phase Shift Model and Beamforming Optimization," *IEEE Trans. Commun.*, vol. 68, no. 9, pp. 5849–5863, Sep. 2020.

[27] 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," *Digit. Commun. Netw.*, vol. 8, no. 2, pp. 87–104, Apr. 2022.

[28] Y. Chen, Y. Wang, Z. Wang, and P. Zhang, "Robust Beamforming for Active Reconfigurable Intelligent Omni-surface in Vehicular Communications," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 10, pp. 3086–3103, Oct. 2022.

[29] X. Mu, Y. Liu, L. Guo, J. Lin, and R. Schober, "Simultaneously Transmitting And Reflecting (STAR) RIS Aided Wireless Communications," *IEEE Trans. Wireless Commun.*, vol. 21, no. 5, pp. 3083–3098, May 2022.

[30] S. Zhang, H. Zhang, B. Di, Y. Tan, M. Di Renzo, Z. Han, H. Vincent Poor, and L. Song, "Intelligent Omni-Surfaces: Ubiquitous Wireless Transmission by Reflective-Refractive Metasurfaces," *IEEE Trans. Wireless Commun.*, vol. 21, no. 1, pp. 219–233, Jan. 2022.

[31] Z. Zhang, L. Dai, X. Chen, C. Liu, F. Yang, R. Schober, and H. Vincent Poor, "Active RIS vs. Passive RIS: Which Will Prevail in 6G?" *IEEE Trans. Commun.*, pp. 1–1, 2022.

[32] S. Zhang and R. Zhang, "Intelligent Reflecting Surface Aided Multi-User Communication: Capacity Region and Deployment Strategy," *IEEE Trans. Commun.*, vol. 69, no. 9, pp. 5790–5806, Sep. 2021.

[33] L. Dong, H.-M. Wang, and J. Bai, "Active Reconfigurable Intelligent Surface Aided Secure Transmission," *IEEE Trans. Veh. Technol.*, vol. 71, no. 2, pp. 2181–2186, Feb. 2022.

[34] 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, Sep. 2021.

[35] 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. 4962–4975, Aug. 2021.

[36] G. Chen, Q. Wu, C. He, W. Chen, J. Tang, and S. Jin, "Active IRS Aided Multiple Access for Energy-Constrained IoT Systems," *IEEE Trans. Wireless Commun.*, pp. 1–1, 2022.

[37] 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 Commun. Lett.*, vol. 26, no. 5, pp. 1150–1154, May 2022.

[38] 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.

[39] C. Pan, H. Ren, K. Wang, W. Xu, M. Elkashlan, A. Nallanathan, and L. Hanzo, "Multicell MIMO Communications Relying on Intelligent Reflecting Surfaces," *IEEE Trans. Wireless Commun.*, vol. 19, no. 8, pp. 5218–5233, Aug. 2020.

[40] M. Hua, Q. Wu, D. W. K. Ng, J. Zhao, and L. Yang, "Intelligent Reflecting Surface-Aided Joint Processing Coordinated Multipoint Transmission," *IEEE Trans. Commun.*, vol. 69, no. 3, pp. 1650–1665, Mar. 2021.

[41] X. Xie, C. He, H. Luan, Y. Dong, K. Yang, F. Gao, and Z. J. Wang, "A Joint Optimization Framework for IRS-Assisted Energy Self-Sustainable IoT Networks," *IEEE Internet Things J.*, vol. 9, no. 15, pp. 13 767–13 779, Aug. 2022.

[42] E. Shtaiwi, H. Zhang, S. Vishwanath, M. Youssef, A. Abdelhadi, and Z. Han, "Channel Estimation Approach for RIS Assisted MIMO Systems," *IEEE Trans. Cogn. Commun. Netw.*, vol. 7, no. 2, pp. 452–465, Jun. 2021.

[43] M.-M. Zhao, Q. Wu, M.-J. Zhao, and R. Zhang, "Exploiting Amplitude Control in Intelligent Reflecting Surface Aided Wireless Communication With Imperfect CSI," *IEEE Trans. Commun.*, vol. 69, no. 6, pp. 4216–4231, Jun. 2021.

[44] Q. Shi, M. Razaviyayn, Z. Luo, and C. He, "An Iteratively Weighted MMSE Approach to Distributed Sum-Utility Maximization for a MIMO Interfering Broadcast Channel," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4331–4340, Sep. 2011.

[45] M. Grant and S. Boyd, "CVX: Matlab Software for Disciplined Convex Programming, Version 2.1," Mar. 2014.

[46] L. Zhang, R. Zhang, Y.-C. Liang, Y. Xin, and H. V. Poor, "On Gaussian MIMO BC-MAC Duality With Multiple Transmit Covariance Constraints," *IEEE Trans. Inf. Theory*, vol. 58, no. 4, pp. 2064–2078, Apr. 2012.

[47] X. Zhang, *Matrix Analysis and Applications*. Cambridge University Press, 2017.

[48] C. Pan, H. Ren, K. Wang, M. Elkashlan, A. Nallanathan, J. Wang, and L. Hanzo, "Intelligent Reflecting Surface Aided MIMO Broadcasting for Simultaneous Wireless Information and Power Transfer," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 8, pp. 1719–1734, Jun. 2020.

[49] S. Zhang and R. Zhang, "Capacity Characterization for Intelligent Reflecting Surface Aided MIMO Communication," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 8, pp. 1823–1838, Aug. 2020.

[50] G. Cui, X. Yu, G. Foglia, Y. Huang, and J. Li, "Quadratic Optimization With Similarity Constraint for Unimodular Sequence Synthesis," *IEEE Trans. Signal Process.*, vol. 65, no. 18, pp. 4756–4769, Sep. 2017.

[51] J. Wang, H. Wang, Y. Han, S. Jin, and X. Li, "Joint Transmit Beamforming and Phase Shift Design for Reconfigurable Intelligent Surface Assisted MIMO Systems," *IEEE Trans. Cogn. Commun. Netw.*, vol. 7, no. 2, pp. 354–368, Jun. 2021.

Xie Xie received the B.Eng. and the M.S. degrees from Northwest University, Xi'an, China, in 2017 and 2020, respectively. He is currently working toward the Ph.D. degree at Northwest University, Xi'an, China. His research interests are in the area of wireless communications and signal processing.

Authorized licensed use limited to: University of Houston. Downloaded on July 02,2023 at 13:19:36 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Wireless Communications. 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/TWC.2023.3247906

Chen He (Member, IEEE) received the B.Eng. degree (*summa cum laude*) from McMaster University in 2007, and the M.A.Sc. and Ph.D. degrees from The University of British Columbia (UBC), Vancouver, in 2009 and 2014, respectively, all in electrical and computer engineering.

He was a Research Engineer with Blackberry Limited, Canada, and a Post-Doctoral Research Fellow with UBC. Since 2016, he has been a Full Professor at Northwest University, China. His research interests are the area of wireless communications, signal processing, and quantum information science. He is serving as an Associate Editor for the IEEE SIGNAL PROCESSING LETTERS.

Xue Ma received the B.Eng. Degree in communication engineering from Northwest University, Xi'an, China, in 2020, where she is currently pursuing the Ph.D. degree in computer science. Her research interests are in the area of UAV millimeter-wave wireless communications and reinforcement learning.

Feifei Gao (Fellow, IEEE) received the B.Eng. degree from Xi'an Jiaotong University, Xi'an, China, in 2002, the M.Sc. degree from McMaster University, Hamilton, ON, Canada, in 2004, and the Ph.D. degree from the National University of Singapore, Singapore, in 2007.

Since 2011, he has been with the Department of Automation, Tsinghua University, Beijing, China. His research interests include signal processing for communications, array signal processing, convex optimizations, and artificial intelligence assisted communications. He has authored/coauthored more than 150 refereed IEEE journal articles and more than 150 IEEE conference proceeding papers that are cited more than 12400 times in Google Scholar. He has served as an Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, the IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING (Lead Guest Editor), the IEEE TRANSACTIONS ON COGNITIVE COMMUNICATIONS AND NETWORKING, the IEEE SIGNAL PROCESSING LETTERS (Senior Editor), the IEEE COMMUNICATIONS LETTERS (Senior Editor), the IEEE WIRELESS COMMUNICATIONS LETTERS, and CHINA COMMUNICATIONS. He has also served as the Symposium Co-Chair for 2019 IEEE Conference on Communications (ICC), 2018 IEEE Vehicular Technology Conference Spring (VTC), 2015 IEEE ICC, 2014 IEEE Global Communications Conference (GLOBECOM), and 2014 IEEE Vehicular Technology Conference Fall (VTC), as well as Technical Committee Members for more than 50 IEEE conferences.

Zhu Han (Fellow, IEEE) received the B.S. degree in electronic engineering from Tsinghua University, in 1997, and the M.S. and Ph.D. degrees in electrical and computer engineering from the University of Maryland, College Park, in 1999 and 2003, respectively.

From 2000 to 2002, he was an R&D Engineer of JDSU, Germantown, Maryland. From 2003 to 2006, he was a Research Associate at the University of Maryland. From 2006 to 2008, he was an assistant professor at Boise State University, Idaho. Currently, he is a John and Rebecca Moores Professor in the Electrical and Computer Engineering Department as well as in the Computer Science Department at the University of Houston, Texas. Dr. Han's main research targets on the novel game-theory related concepts critical to enabling efficient and distributive use of wireless networks with limited resources. His other research interests include wireless resource allocation and management, wireless communications and networking, quantum computing, data science, smart grid, security and privacy. Dr. Han received an NSF Career Award in 2010, the Fred W. Ellersick Prize of the IEEE Communication Society in 2011, the EURASIP Best Paper Award for the JOURNAL ON ADVANCES IN SIGNAL PROCESSING in 2015, IEEE Leonard G. Abraham Prize in the field of Communications Systems (Best Paper Award in IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS (JSAC)) in 2016, and several best paper awards in IEEE conferences. Dr. Han was an IEEE Communications Society Distinguished Lecturer from 2015-2018, AAAS fellow since 2019, and ACM distinguished Member since 2019. Dr. Han is a 1% highly cited researcher since 2017 according to Web of Science. Dr. Han is also the winner of the 2021 IEEE Kiyo Tomiyasu Award, for outstanding early to mid-career contributions to technologies holding the promise of innovative applications, with the following citation: "for contributions to game theory and distributed management of autonomous communication networks."

Z. Jane Wang (Fellow, IEEE) received the B.Sc. degree in electrical engineering from Tsinghua University, China, in 1996, and the M.Sc. and Ph.D. degrees in electrical engineering from the University of Connecticut, in 2000 and 2002, respectively.

She has been a Research Associate with the Electrical and Computer Engineering Department, University of Maryland, College Park. Since 2004, she has been with the Department Electrical and Computer Engineering, The University of British Columbia, Canada, where she is currently a Professor. Her research interests include statistical signal processing theory and applications. She co-received the EURASIP Journal on Applied Signal Processing (JASP) Best Paper Award in 2004, and the IEEE Signal Processing Society Best Paper Award in 2005. She was the Editor-in-Chief for the IEEE SIGNAL PROCESSING LETTERS. She is serving as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING, the IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, and the IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING.

Authorized licensed use limited to: University of Houston. Downloaded on July 02,2023 at 13:19:36 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.