Multicell MIMO Communications Relying on Intelligent Reflecting Surfaces

Cunhua Pan, *Member, IEEE*, Hong Ren, *Member, IEEE*, Kezhi Wang, *Member, IEEE*,
Wei Xu, *Senior Member, IEEE*, Maged Elkashlan, *Member, IEEE*,
Arumugam Nallanathan, *Fellow, IEEE*,
and Lajos Hanzo, *Fellow, IEEE*

Abstract—Intelligent reflecting surfaces (IRSs) constitute a disruptive wireless communication technique capable of creating a controllable propagation environment. In this paper, we propose to invoke an IRS at the cell boundary of multiple cells to assist the downlink transmission to cell-edge users, whilst mitigating the inter-cell interference, which is a crucial issue in multicell communication systems. We aim for maximizing the weighted sum rate (WSR) of all users through jointly optimizing the active precoding matrices at the base stations (BSs) and the phase shifts at the IRS subject to each BS's power constraint and unit modulus constraint. Both the BSs and the users are equipped with multiple antennas, which enhances the spectral efficiency by exploiting the spatial multiplexing gain. Due to the non-convexity of the problem, we first reformulate it into an equivalent one, which is solved by using the block coordinate descent (BCD) algorithm, where the precoding matrices and phase shifts are alternately optimized. The optimal precoding matrices can be obtained in closed form, when fixing the phase shifts. A pair of efficient algorithms are proposed for solving the phase shift optimization problem, namely the Majorization-Minimization (MM) Algorithm and the Complex Circle Manifold (CCM) Method. Both algorithms are guaranteed to converge to at least locally optimal solutions. We also extend the proposed algorithms to the more general multiple-IRS and network MIMO scenarios. Finally, our simulation results confirm the advantages of introducing IRSs in enhancing the cell-edge user performance.

Index Terms—Intelligent reflecting surface (IRS), reconfigurable intelligent surfaces, manifold optimization, multicell communications, MIMO.

I. Introduction

NEXT-GENERATION wireless communication systems are expected to provide a 1000-fold increase in the network capacity over the operational system for satisfying the ever-increasing demand for higher data rates driven by emerging applications such as augmented reality (AR) and virtual reality (VR). To achieve this goal, promising techniques relying on massive multiple-input multiple-output (MIMO) solutions [1], millimeter wave (mmWave) communications [2] and ultra-dense cloud radio access networks (UD-CRAN) have been advocated [3]–[5]. By deploying a massive number of antennas at the base station (BS) for transmission over the millimeter-wave (mm-wave) bands, significant spectral efficiency improvements can be achieved by exploiting the joint benefits of a high spatial multiplexing gain and high bandwidth. However, escalating signal processing complexity, increased hardware costs as well as high power consumption are incurred by the associated high number of radio frequency (RF) chains operating in a high frequency band. These issues erode their practical benefits. Although the access points (AP) can be densely deployed in UD-CRAN systems for reducing the distance between the users and the APs, the limited fronthaul capacity becomes their performance bottleneck. Furthermore, these techniques have to operate in the face of unfavourable electromagnetic wave propagation, improving a high blockage probability.

As a remedy, intelligent reflecting surface (IRS) has been proposed as a revolutional technique of facilitating both spectrum- and energy-efficient communications through reconfiguring the wireless propagation environment [6], [7]. An IRS consists of a vast number of low-cost passive reflecting elements, each of which can independently adjust the phase shift of the signals incident upon it, and thus collaboratively creating favourable wireless transmission channels by innovatively harnessing the reflected signal. By properly tuning the phase shifts by using an IRS controller, the reflected

signals can be added constructively at the desired receiver for enhancing the received signal power, whilst destructively superimposing them at the non-intended receivers for reducing the co-channel interference. Although passive reflecting surfaces have already been used in radar systems, the phase shifts of passive elements cannot be changed once they were fabricated, and they are unable to control the wireless propagation channels. Fortunately, due to the recent advance in micro-electromechanical systems (MEMS) and metamaterials [8], the phase shifts can now be adjusted in real time, which results in near-instantaneously reconfigurable IRS possible. Although an IRS resembles the classic amplify-and-forward (AF) relay, the former has the advantage of lower power consumption, since it only reflects the signals passively without requiring active RF chains, while the latter necessitates active RF components for signal transmission. Hence, IRSs do not impose additional thermal noise on the reflected signals. Performance comparisons between AF relay and IRS were performed in [9], [10]. Given the limited functionality of IRSs, their phase shifters can be fabricated in a compact form. Hence, each IRS accommodates a large number of phase shifters and provides high beamforming gains. Furthermore, IRSs have the appealing advantages of light weight and small sizes, which can be readily installed at buildings facades, on the room-ceilings, on lamp posts, on road signs, etc. IRSs can also be integrated into the existing communication systems at a modest modification. However, to reap the aforementioned benefits promised of IRSs, the phase shifts have to be appropriately optimized along with the active beamforming weights at the BS. The main difficulty in optimizing the phase shifts is the non-convex unit modulus constraint imposed on the phase shifts. Although this kind of constraints have been studied both in hybrid digital/analog precoding [11], [12] and in constant-envelope precoding in massive MIMO systems [13], [14], these studies were only focused on the designs at the transmitter, which are not applicable for the joint active beamforming design of the BSs and of the passive beamforming design at the IRS.

Most recently, some initial efforts have been devoted to the transmitter design of IRS-assisted wireless communication systems, including the single-user case of [15]–[18], the downlink multiuser case of [19]–[23], wireless power transfer design of [24], mobile edge computing of [25], multicast scenario of [26] and the physical layer security design of [27]–[31]. However, the above-mentioned papers only studied the single-cell scenario, whilst there is a paucity of investigations on the multicell scenario in the existing literature. To mitigate the spectrum scarcity, different cells will reuse the same frequency resources, which causes severe inter-cell interference, especially for cell-edge users [32]. Hence, in this paper, we propose to employ an IRS at the cell boundary for assisting the cell-edge users of multicell systems as shown in Fig. 1, where the inter-cell interference can be alleviated with the aid of IRSs. Specifically, by carefully adjusting the phase shifts of the IRS's reflective elements, the inter-cell interference reflected by the IRS can be superimposed destructively on the direct interference impinging from the adjacent BS for minimizing the interference power
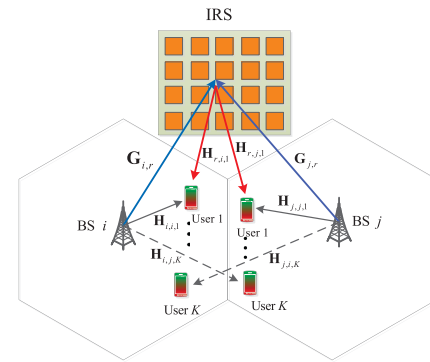

Fig. 1. An IRS-assisted multicell MIMO multiuser communication system.

at the receivers. This provides a higher degree of freedom for designing the beamforming/precoding at each BS for the users in its own cell. As a result, the active beamforming/precoding applied at each BS and the passive beamforming matrix of the IRS have to be jointly optimized. However, the resultant optimization problem is challenging to solve, since the optimization variables are highly coupled. Furthermore, all the existing contributions consider the single-antenna aided user scenario. However, owing to the rapid developments in antenna technology [33], the user equipment is also capable of accommodating multiple antennas for enhancing the received signal strength. Then, multiple data streams can be transmitted simultaneously, which boost the throughput. Therefore, in this paper, we consider the multiple-antenna aided user scenario. Given the complex mathematical data rate expression, the techniques conceived in [15]–[23], [27]–[30] cannot be directly applied. The multiple-antenna user case further complicates the optimization.

The main contributions of this paper can be summarized as follows:

1) To the best of our knowledge, this is the first attempt to explore the assistance of IRSs in enhancing the cell-edge performance in multicell MIMO communication systems. Specifically, we jointly optimize the active transmit precoding (TPC) matrices of all BSs and the phase shifts at the IRS for maximizing the weighted sum rate (WSR) of all users subject to each BS's power constraint and to the unit modulus constraint of the phase shifters. However, the objective function (OF) is not jointly concave over both the TPC matrices and the phase shifts, which are highly coupled. To tackle this challenging problem, we first reformulate the original problem into an equivalent one by exploiting the equivalence between the data rate and the weighted minimum mean-square error (WMMSE). Then, the block coordinate descent (BCD) algorithm is proposed for alternately optimizing the TPC matrices at the BSs and the passive beamforming at the IRS.

2) Given the fixed phase shifts, we derive the optimal TPC matrices in closed form by applying the classic Lagrangian multiplier method. Since the phase shift optimization problem is highly coupled with the various channel matrices and TPC matrices, this is quite

a challenge. By using sophisticated matrix manipulations and transformations, we successfully transform the phase shift optimization problem into a non-convex quadratically constrained quadratic program (QCQP) subject to unit modulus constraint. A pair of efficient iterative algorithms are proposed for solving this problem. The first one is the Majorization-Minimization (MM) Algorithm [34], where a closed-form solution can be obtained in each iteration. The second is based on the Complex Circle Manifold (CCM) Method [35], where we show that the unit modulus constraints of all phase shifters constitute a complex circle manifold. Both the MM algorithm and the CCM algorithm are guaranteed to obtain at least a locally optimal solution.

3) The proposed algorithms are also extended to the more general multiple-IRS and network MIMO scenarios.

4) Our simulation results show that the cell-edge performance can be significantly enhanced by employing IRSs compared to a conventional multicell system operating without IRSs. Moreover, it is also shown that the performance gain achieved by the IRS is indeed mainly due to the improving BS-IRS and IRS-user links. Furthermore, the location of IRSs should be carefully chosen. It is shown that deploying IRSs at the cell boundary achieves the highest gains for cell-edge users. Furthermore, simulation results also show that the IRSs should be deployed in the vicinity of the user clusters, and distributed IRS deployment has superior performance than the centralized deployment.

The remainder of this paper is organized as follows. In Section II, we present the system model of IRS-assisted multicell MIMO communication and formulate the WSR maximization problem. In Section III, we reformulate the original problem into a more tractable problem and the TPC matrices and passive beamforming phases are alternately optimized. In Section V, extensive simulation results are provided for quantifying the performance advantages of introducing IRSs into multicell systems. Finally, our conclusions are offered in Section VI.

Notations: For a complex value a, $\mathrm{Re}\{a\}$ represents the real part of a. Boldface lower case and upper case letters denote vectors and matrices, respectively. $\mathbb{C}^M$ denotes the set of $M \times 1$ complex vectors. $\mathbb{E}\{\cdot\}$ denotes the expectation operation. $\|\mathbf{x}\|_2$ denotes the 2-norm of vector $\mathbf{x}$. For two matrices $\mathbf{A}$ and $\mathbf{B}$, $\mathbf{A} \odot \mathbf{B}$ represents the Hadamard product of $\mathbf{A}$ and $\mathbf{B}$. $\|\mathbf{A}\|_F$, $\mathrm{Tr}\,(\mathbf{A})$ and $|\mathbf{A}|$ denote the Frobenius norm, trace operation and determinant of $\mathbf{A}$, respectively. $\nabla f_{\mathbf{x}}\,(\mathbf{x})$ denotes the gradient of the function f with respect to (w.r.t.) the vector $\mathbf{x}$. $\mathcal{CN}(\mathbf{0}, \mathbf{I})$ represents a random vector following the distribution of zero mean and unit variance matrix. $\arg\{\cdot\}$ means the extraction of phase information. $\mathrm{diag}(\cdot)$ denotes the diagonalization operation. $(\cdot)^*$, $(\cdot)^{\mathrm{T}}$ and $(\cdot)^{\mathrm{H}}$ denote the conjugate, transpose and Hermitian operators, respectively.

II. SYSTEM MODEL AND PROBLEM FORMULATION

A. System Model

We consider an IRS-aided multicell downlink MIMO model constituted by L macro cells, each of which has a single

base station (BS) that serves K cell-edge users. Each BS and each user is equipped with $N_t \geq 1$ and $N_r \geq 1$ transmit antennas (TAs) and receive antennas (RAs), respectively. Each cell-edge user suffers both from high attenuation from its serving BS and severe cochannel interference from its neighbouring BSs. To mitigate these, we propose to employ an IRS which has M reflection elements at the cell edge as shown in Fig. 1, which boost the useful signal power and mitigate the cochannel interference by carefully designing the phase shifts of the reflective elements.

The signal transmitted by the lth BS is given by

$$\mathbf{x}_l = \sum_{k=1}^K \mathbf{F}_{l,k}\mathbf{s}_{l,k}, \qquad (1)$$

where $\mathbf{s}_{l,k}$ is the $(d \times 1)$-element symbol vector transmitted to the kth user in its cell, satisfying $\mathbb{E}\left[\mathbf{s}_{l,k}\mathbf{s}_{l,k}^H\right] = \mathbf{I}_d$ and $\mathbb{E}\left[\mathbf{s}_{l,k}(\mathbf{s}_{i,j})^H\right] = \mathbf{0}$, for $\{l, k\} \neq \{i, j\}$, and $\mathbf{F}_{l,k} \in \mathbb{C}^{N_t \times d}$ is the linear TPC matrix used by the lth BS for transmitting its data vector $\mathbf{s}_{l,k}$ to the kth user. The baseband channels spanning from the nth BS to the kth user in the lth cell, as well as those from the IRS to the kth user in the lth cell, and the ones from the nth BS to the IRS are denoted by $\mathbf{H}_{n,l,k}$, $\mathbf{H}_{l,k}^r$ and $\mathbf{G}_n^r$, respectively. Let us denote the phase shift of the m-th reflection element of the IRS by $\theta_m \in [0, 2\pi]$. Thus the reflection operator simply multiplies the incident multi-path signals by $e^{j\theta_m}$ [1] at a single physical point and then forwards the combined signal to the users. Hence, the users will directly receive the desired signals from the BSs, plus the signals reflected by the IRS. However, we ignore the signal reflected more than once due to the severe path loss. Let us denote the diagonal phase-shifting matrix of the IRS as $\boldsymbol{\Phi} = \mathrm{diag}\left\{e^{j\theta_1}, \cdots, e^{j\theta_m}, \cdots, e^{j\theta_M}\right\}$. Then, the received signal vector at the kth user in the lth cell is given by

$$\mathbf{y}_{l,k} = \underbrace{\sum_{n=1}^L \mathbf{H}_{n,l,k}\mathbf{x}_n}_{\text{Siganls from BSs}} + \underbrace{\sum_{n=1}^L \mathbf{H}_{l,k}^r\boldsymbol{\Phi}\mathbf{G}_n^r\mathbf{x}_n}_{\text{Signals from the IRS}} + \mathbf{n}_{l,k}, \quad (2)$$

where $\mathbf{n}_{l,k}$ is the noise vector that satisfies $\mathcal{CN}\left(\mathbf{0}, \sigma^2\mathbf{I}_{N_r}\right)$.

We assume that the channel state information (CSI) of all channels is perfectly known at the BS, and the BS calculates the optimal phase shifts and sends them back to the IRS controller. Indeed, the assumption of having perfect CSI knowledge at the BS is idealistic because it is challenging to obtain the CSI in IRS-assisted communication systems. However, the algorithms developed allow us to derive the relevant performance upper bounds for realistic scenarios in the presence of realistic CSI errors. In addition, the proposed algorithms can provide insights into the performance gain provided by IRSs, which can inspire further research in this area. Recently, we have conceived a framework for the robust transmission design of an IRS-aided single-cell scenario [36] by considering both the bounded CSI error model and the statistical CSI error model associated with the cascaded channels.

---

[1] j is the imaginary unit.

Its extension to the multicell scenario will be studied in our future research.

Let us define $\bar{\mathbf{H}}_{n,l,k} \triangleq \mathbf{H}_{l,k}^r \boldsymbol{\Phi} \mathbf{G}_n^r + \mathbf{H}_{n,l,k}$, which can be regarded as the equivalent channel spanning from the nth BS to the kth user in the lth cell. By substituting (1) into (2), $\mathbf{y}_{l,k}$ can be written as

$$\mathbf{y}_{l,k} = \bar{\mathbf{H}}_{l,l,k} \mathbf{F}_{l,k} \mathbf{s}_{l,k} + \underbrace{\sum_{m=1,m \neq k}^{K} \bar{\mathbf{H}}_{l,l,k} \mathbf{F}_{l,m} \mathbf{s}_{l,m}}_{\text{Intra-cell interference}}$$
$$+ \underbrace{\sum_{n=1,n \neq l}^{L} \sum_{m=1}^{K} \bar{\mathbf{H}}_{n,l,k} \mathbf{F}_{n,m} \mathbf{s}_{n,m}}_{\text{Inter-cell interference}} + \mathbf{n}_{l,k}. \quad (3)$$

Then, the achievable data rate (nat/s/Hz) of the kth user in the lth cell is given by [4]

$$R_{l,k}(\mathbf{F}, \boldsymbol{\theta}) = \log \left| \mathbf{I} + \bar{\mathbf{H}}_{l,l,k} \mathbf{F}_{l,k} \mathbf{F}_{l,k}^{\mathrm{H}} \bar{\mathbf{H}}_{l,l,k}^{\mathrm{H}} \mathbf{J}_{l,k}^{-1} \right|, \quad (4)$$

where we have $\mathbf{F} = [\mathbf{F}_{l,k}, \forall l, k]$, $\boldsymbol{\theta} = [\theta_1, \cdots, \theta_M]$, and $\mathbf{J}_{l,k}$ is the interference-plus-noise covariance matrix:

$$\mathbf{J}_{l,k} = \sum_{m=1,m \neq k}^{K} \bar{\mathbf{H}}_{l,l,k} \mathbf{F}_{l,m} \mathbf{F}_{l,m}^{\mathrm{H}} \bar{\mathbf{H}}_{l,l,k}^{\mathrm{H}}$$
$$+ \sum_{n=1,n \neq l}^{L} \sum_{m=1}^{K} \bar{\mathbf{H}}_{n,l,k} \mathbf{F}_{n,m} \mathbf{F}_{n,m}^{\mathrm{H}} \bar{\mathbf{H}}_{n,l,k}^{\mathrm{H}} + \sigma^2 \mathbf{I}.$$

B. Problem Formulation

In this paper, we aim for maximizing the WSR of all the users by jointly optimizing the TPC matrices $\mathbf{F}$ at the BSs and the phase shifts $\boldsymbol{\theta}$ at the IRS, while guaranteeing the total power constraint at each BS. Specifically, the WSR maximization problem is formulated as:

$$\max_{\mathbf{F}, \boldsymbol{\theta}} \sum_{l=1}^{L} \sum_{k=1}^{K} \omega_{l,k} R_{l,k}(\mathbf{F}, \boldsymbol{\theta}) \quad (5a)$$

$$\text{s.t.} \sum_{k=1}^{K} \|\mathbf{F}_{l,k}\|_F^2 \leq P_{l,\max}, \quad l = 1, \cdots, L, \quad (5b)$$

$$0 \leq \theta_m \leq 2\pi, \quad m = 1, \cdots, M, \quad (5c)$$

where $\omega_{l,k}$ denotes the weighting factor representing the priority of the corresponding user. Due to the coupling effect between the TPC matrices $\mathbf{F}$ and the phase shifts $\boldsymbol{\theta}$, this optimization problem is difficult to solve. Additionally, the phase shift constraints in (5c) further aggravate the challenge. In the following, we provide a low-complexity algorithm for solving Problem (5).

III. LOW-COMPLEXITY ALGORITHM DEVELOPMENT

In this section, we first reformulate the original problem into a more tractable form. Then, the block coordinate descent (BCD) method is proposed for solving the formulated problem.

A. Reformulation of the Original Problem

In the following, we exploit the relationship between the data rate and the mean-square error (MSE) for the optimal decoding matrix. To reduce the decoding complexity, we consider a linear decoding matrix so that the estimated signal vector of each user is given by

$$\hat{\mathbf{s}}_{l,k} = \mathbf{U}_{l,k}^{\mathrm{H}} \mathbf{y}_{l,k}, \quad \forall l, k, \quad (6)$$

where $\mathbf{U}_{l,k} \in \mathbb{C}^{N_r \times d}$ is the decoding matrix for the kth user in the lth cell. Then, the MSE matrix of each user is given by

$$\mathbf{E}_{l,k} = \mathbb{E}_{\mathbf{s}, \mathbf{n}} \left[(\hat{\mathbf{s}}_{l,k} - \mathbf{s}_{l,k}) (\hat{\mathbf{s}}_{l,k} - \mathbf{s}_{l,k})^H \right] \quad (7)$$
$$= \left(\mathbf{U}_{l,k}^{\mathrm{H}} \bar{\mathbf{H}}_{l,l,k} \mathbf{F}_{l,k} - \mathbf{I} \right) \left(\mathbf{U}_{l,k}^{\mathrm{H}} \bar{\mathbf{H}}_{l,l,k} \mathbf{F}_{l,k} - \mathbf{I} \right)^{\mathrm{H}}$$
$$+ \sum_{m=1,m \neq k}^{K} \mathbf{U}_{l,k}^{\mathrm{H}} \bar{\mathbf{H}}_{l,l,k} \mathbf{F}_{l,m} \mathbf{F}_{l,m}^{\mathrm{H}} \bar{\mathbf{H}}_{l,l,k}^{\mathrm{H}} \mathbf{U}_{l,k}$$
$$+ \sum_{n=1,n \neq l}^{L} \sum_{m=1}^{K} \mathbf{U}_{l,k}^{\mathrm{H}} \bar{\mathbf{H}}_{n,l,k} \mathbf{F}_{n,m} \mathbf{F}_{n,m}^{\mathrm{H}} \bar{\mathbf{H}}_{n,l,k}^{\mathrm{H}} \mathbf{U}_{l,k}$$
$$+ \sigma^2 \mathbf{U}_{l,k}^{\mathrm{H}} \mathbf{U}_{l,k}, \quad \forall l, k. \quad (8)$$

Upon introducing a set of auxiliary matrices $\mathbf{W} = \{\mathbf{W}_{l,k} \succeq \mathbf{0}, \forall l, k\}$ and defining $\mathbf{U} = \{\mathbf{U}_{l,k}, \forall l, k\}$, Problem (5) can be reformulated as follows [4], [37]:

$$\max_{\mathbf{W}, \mathbf{U}, \mathbf{F}, \boldsymbol{\theta}} \sum_{l=1}^{L} \sum_{k=1}^{K} \omega_{l,k} h_{l,k}(\mathbf{W}, \mathbf{U}, \mathbf{F}, \boldsymbol{\theta}) \quad (9a)$$

$$\text{s.t.} \sum_{k=1}^{K} \|\mathbf{F}_{l,k}\|_F^2 \leq P_{l,\max}, \quad l = 1, \cdots, L, \quad (9b)$$

$$0 \leq \theta_m \leq 2\pi, \quad m = 1, \cdots, M, \quad (9c)$$

where $h_{l,k}(\mathbf{W}, \mathbf{U}, \mathbf{F}, \boldsymbol{\theta})$ is given by

$$h_{l,k}(\mathbf{W}, \mathbf{U}, \mathbf{F}, \boldsymbol{\theta}) = \log |\mathbf{W}_{l,k}| - \text{Tr}(\mathbf{W}_{l,k} \mathbf{E}_{l,k}) + d. \quad (10)$$

Note that compared to the original OF of Problem (5), the new OF in Problem (9) is in a more tractable form, although we have introduced more optimization variables. For a given phase shift $\boldsymbol{\theta}$, $h_{l,k}(\mathbf{W}, \mathbf{U}, \mathbf{F}, \boldsymbol{\theta})$ is a concave function for each set of the optimization matrices, when the other two are fixed. In the following, we propose the BCD algorithm for solving Problem (9). Specifically, we maximize the OF in (9) by alternately optimizing one set of optimization variables, while keeping the other variables fixed. Note that the decoding matrix $\mathbf{U}_{l,k}$ and the auxiliary matrix $\mathbf{W}_{l,k}$ are only related to $h_{l,k}(\mathbf{W}, \mathbf{U}, \mathbf{F}, \boldsymbol{\theta})$. In the following, we can derive the optimal solution for $\mathbf{U}_{l,k}$ and $\mathbf{W}_{l,k}$, when the other matrices are fixed. For given values of $\boldsymbol{\theta}$, $\mathbf{W}$, and $\mathbf{F}$, we can set the first-order derivative of $h_{l,k}(\mathbf{W}, \mathbf{U}, \mathbf{F}, \boldsymbol{\theta})$ with respect to $\mathbf{U}_{l,k}$ to zero, which gives the optimal $\mathbf{U}_{l,k}$:

$$\mathbf{U}_{l,k} = \left(\mathbf{J}_{l,k} + \bar{\mathbf{H}}_{l,l,k} \mathbf{F}_{l,k} \mathbf{F}_{l,k}^{\mathrm{H}} \bar{\mathbf{H}}_{l,l,k}^{\mathrm{H}} \right)^{-1} \bar{\mathbf{H}}_{l,l,k} \mathbf{F}_{l,k}. \quad (11)$$

Similarly, for given $\boldsymbol{\theta}$, $\mathbf{U}$, and $\mathbf{F}$, the optimal auxiliary matrix $\mathbf{W}_{l,k}$ can be obtained as follows:

$$\mathbf{W}_{l,k} = \mathbf{E}_{l,k}^{-1}, \quad (12)$$

where $\mathbf{E}_{l,k}$ is given in (8).

Let us now focus our attention on optimizing the TPC matrices $\mathbf{F}$ and phase shifts $\boldsymbol{\theta}$.

B. Optimizing the Precoding Matrices $\mathbf{F}$

In this subsection, we focus our attention on optimizing the TPC matrices $\mathbf{F}$, while fixing $\mathbf{W}, \mathbf{U}$ and $\boldsymbol{\theta}$. By substituting $\mathbf{E}_{l,k}$ into (10), the optimization over $\mathbf{F}$ can be decoupled among the different BSs. Specifically, by removing the constant terms, the TPC matrix optimization problem of the lth BS is given by

$$
\min_{\mathbf{F}_{l,k}, \forall k} \quad -\sum_{k=1}^{K} \omega_{l,k} \mathrm{Tr}\left(\mathbf{W}_{l,k}\mathbf{U}_{l,k}^{\mathrm{H}}\bar{\mathbf{H}}_{l,l,k}\mathbf{F}_{l,k}\right)
$$
$$
+ \sum_{k=1}^{K} \mathrm{Tr}\left(\mathbf{F}_{l,k}^{\mathrm{H}}\mathbf{A}_{l}\mathbf{F}_{l,k}\right)
$$
$$
- \sum_{k=1}^{K} \omega_{l,k}\mathrm{Tr}\left(\mathbf{W}_{l,k}\mathbf{F}_{l,k}^{\mathrm{H}}\bar{\mathbf{H}}_{l,l,k}^{\mathrm{H}}\mathbf{U}_{l,k}\right) \quad (13\mathrm{a})
$$
$$
\text{s.t.} \quad \sum_{k=1}^{K} \|\mathbf{F}_{l,k}\|_{F}^{2} \leq P_{l,\max}, \quad (13\mathrm{b})
$$

where $\mathbf{A}_{l}$ is

$$
\mathbf{A}_{l} = \sum_{n=1}^{L}\sum_{m=1}^{K} \omega_{n,m}\bar{\mathbf{H}}_{l,n,m}^{\mathrm{H}}\mathbf{U}_{n,m}\mathbf{W}_{n,m}\mathbf{U}_{n,m}^{\mathrm{H}}\bar{\mathbf{H}}_{l,n,m}. \quad (14)
$$

It can be readily verified that the above problem is a convex optimization problem, which can be transformed into a second order cone programming (SOCP) problem that can be efficiently solved by using standard optimization packages, such as CVX [38]. However, the computational complexity of solving an SOCP problem is high. To reduce the complexity, in the following we provide a near-optimal closed-form expression of the TPC matrices by using the Lagrangian multiplier method.

Following some further manipulations, the Lagrangian function of Problem (13) is written as

$$
\mathcal{L}\left(\mathbf{F}_{l,k}, \forall k, \lambda_{l}\right) = \sum_{k=1}^{K} \mathrm{Tr}\left(\mathbf{F}_{l,k}^{\mathrm{H}}\left(\mathbf{A}_{l}+\lambda_{l}\mathbf{I}\right)\mathbf{F}_{l,k}\right) - \lambda_{l}P_{l,\max}
$$
$$
- \sum_{k=1}^{K} \omega_{l,k}\mathrm{Tr}\left(\mathbf{W}_{l,k}\mathbf{U}_{l,k}^{\mathrm{H}}\bar{\mathbf{H}}_{l,l,k}\mathbf{F}_{l,k}\right)
$$
$$
- \sum_{k=1}^{K} \omega_{l,k}\mathrm{Tr}\left(\mathbf{W}_{l,k}\mathbf{F}_{l,k}^{\mathrm{H}}\bar{\mathbf{H}}_{l,l,k}^{\mathrm{H}}\mathbf{U}_{l,k}\right),
$$

where $\lambda_{l} \geq 0$ is the Lagrangian multiplier associated with the power constraint of the lth BS.

By setting the first-order derivative of $\mathcal{L}\left(\mathbf{F}_{l,k}, \forall k, \lambda_{l}\right)$ w.r.t. $\mathbf{F}_{l,k}$ to zero, we can obtain the optimal solution of $\mathbf{F}_{l,k}$ as follows:

$$
\mathbf{F}_{l,k}(\lambda_{l}) = \omega_{l,k}(\mathbf{A}_{l}+\lambda_{l}\mathbf{I})^{\dagger}\bar{\mathbf{H}}_{l,l,k}^{\mathrm{H}}\mathbf{U}_{l,k}\mathbf{W}_{l,k}, \quad (15)
$$

where $(\cdot)^{\dagger}$ denotes the matrix pseudoinverse. The value of λ_{l} should be chosen for ensuring that the following complementary slackness condition for the power constraint is satisfied:

$$
\lambda_{l}\left(\sum_{k=1}^{K}\|\mathbf{F}_{l,k}(\lambda_{l})\|_{F}^{2} - P_{l,\max}\right) = 0. \quad (16)
$$

In the following, we elaborate on how to obtain the optimal λ_{l}, which is divided into two cases: 1) $\mathbf{A}_{l}$ is full rank; 2) $\mathbf{A}_{l}$ is low rank.

1) Case I: $\mathbf{A}_{l}$ Is Full Rank: In this case, $\mathbf{A}_{l}$ is a positive definite matrix, which can be decomposed as $\mathbf{A}_{l} = \mathbf{Q}_{l}\boldsymbol{\Lambda}_{l}\mathbf{Q}_{l}^{\mathrm{H}}$ by using the singular value decomposition (SVD), where $\mathbf{Q}_{l}\mathbf{Q}_{l}^{\mathrm{H}} = \mathbf{Q}_{l}^{\mathrm{H}}\mathbf{Q}_{l} = \mathbf{I}_{N_{t}}$ and $\boldsymbol{\Lambda}_{l}$ is a diagonal matrix with positive diagonal elements. Then, we have

$$
f_{l}(\lambda_{l}) \triangleq \sum_{k=1}^{K} \mathrm{Tr}\left(\mathbf{F}_{l,k}(\lambda_{l})^{\mathrm{H}}\mathbf{F}_{l,k}(\lambda_{l})\right)
$$
$$
= \mathrm{Tr}\left((\boldsymbol{\Lambda}_{l}+\lambda_{l}\mathbf{I})^{-2}\mathbf{Z}_{l}\right) \quad (17)
$$
$$
= \sum_{i=1}^{N_{t}} \frac{[\mathbf{Z}_{l}]_{i,i}}{\left([\boldsymbol{\Lambda}_{l}]_{i,i}+\lambda_{l}\right)^{2}}, \quad (18)
$$

where $\mathbf{Z}_{l} = \sum_{k=1}^{L} \omega_{l,k}^{2}\mathbf{Q}_{l}^{\mathrm{H}}\bar{\mathbf{H}}_{l,l,k}^{\mathrm{H}}\mathbf{U}_{l,k}\mathbf{W}_{l,k}\mathbf{W}_{l,k}^{\mathrm{H}}\mathbf{U}_{l,k}^{\mathrm{H}}\bar{\mathbf{H}}_{l,l,k}\mathbf{Q}_{l}$, $[\mathbf{Z}_{l}]_{i,i}$ and $[\boldsymbol{\Lambda}_{l}]_{i,i}$ denote the ith diagonal element of matrix $\mathbf{Z}_{l}$ and matrix $\boldsymbol{\Lambda}_{l}$, respectively. It can be readily verified that $f_{l}(\lambda_{l})$ is a monotonically decreasing function. Hence, if $f_{l}(0) \leq P_{l,\max}$, then the optimal TPC matrix is given by $\mathbf{F}_{l,k}^{\mathrm{opt}} = \mathbf{F}_{l,k}(0)$. Otherwise, the optimal λ_{l} can be obtained by using the bisection based search method to find the solution of the following equation:

$$
f_{l}(\lambda_{l}) = \sum_{i=1}^{N_{t}} \frac{[\mathbf{Z}_{l}]_{i,i}}{\left([\boldsymbol{\Lambda}_{l}]_{i,i}+\lambda_{l}\right)^{2}} = P_{l,\max}. \quad (19)
$$

Since $f_{l}(\infty) = 0$, the solution of Equation (19) must exist, which is denoted as $\lambda_{l}^{\mathrm{opt}}$. Then, the optimal TPC matrix can be obtained as $\mathbf{F}_{l,k}^{\mathrm{opt}} = \mathbf{F}_{l,k}(\lambda_{l}^{\mathrm{opt}})$. To apply the bisection based search method, we have to find the upper bound of λ_{l}, which is given by

$$
\lambda_{l} < \sqrt{\frac{\sum_{i=1}^{N_{t}}[\mathbf{Z}_{l}]_{i,i}}{P_{l,\max}}} \triangleq \lambda_{l}^{\mathrm{ub}}. \quad (20)
$$

This can be proved as follows:

$$
f_{l}(\lambda_{l}^{\mathrm{ub}}) = \sum_{i=1}^{N_{t}} \frac{[\mathbf{Z}_{l}]_{i,i}}{\left([\boldsymbol{\Lambda}_{l}]_{i,i}+\lambda_{l}^{\mathrm{ub}}\right)^{2}} < \sum_{i=1}^{N_{t}} \frac{[\mathbf{Z}_{l}]_{i,i}}{\left(\lambda_{l}^{\mathrm{ub}}\right)^{2}} = P_{l,\max}.
$$
$$
(21)
$$

2) Case II: $\mathbf{A}_{l}$ Is Low Rank: In this case, the above method cannot be directly applied since the $\mathbf{Q}_{l}$ obtained by SVD is not a unitary matrix, hence the step in (17) cannot be applied. To resolve this issue, we first check whether $\lambda_{l} = 0$ is the optimal solution or not. If

$$
f_{l}(0) = \sum_{k=1}^{K} \mathrm{Tr}\left(\mathbf{F}_{l,k}(0)^{\mathrm{H}}\mathbf{F}_{l,k}(0)\right) \leq P_{l,\max}, \quad (22)
$$

then the optimal TPC matrix is given by $\mathbf{F}_{l,k}^{\mathrm{opt}} = \mathbf{F}_{l,k}(0)$, otherwise, the optimal λ_{l} is a positive value, which will be obtained as follows. Upon defining the rank of $\mathbf{A}_{l}$ as $r_{l} = \mathrm{rank}(\mathbf{A}_{l}) < N_{t}$ and using the SVD, we have

$$
\mathbf{A}_{l} = [\mathbf{Q}_{l,1}, \mathbf{Q}_{l,2}]\boldsymbol{\Lambda}_{l}[\mathbf{Q}_{l,1}, \mathbf{Q}_{l,2}]^{\mathrm{H}}, \quad (23)
$$

where $\mathbf{Q}_{l,1}$ contains the first r_l singular vectors corresponding to the r_l positive eigenvalues, and $\mathbf{Q}_{l,2}$ holds the last $N_t - r_l$ singular vectors corresponding to the $N_t - r_l$ zero-valued eigenvalues, $\mathbf{\Lambda}_l = \text{diag}\left\{\mathbf{\Lambda}_{l,1}, \mathbf{0}_{(N_t-r_l)\times(N_t-r_l)}\right\}$ with $\mathbf{\Lambda}_{l,1}$ denoting the diagonal matrix containing the first r_l positive eigenvalues. Upon defining $\mathbf{Q}_l \triangleq [\mathbf{Q}_{l,1}, \mathbf{Q}_{l,2}]$ and applying similar steps to those in (17) to (18), we have

$$f_l(\lambda_l) = \sum_{k=1}^{K} \text{Tr}\left(\mathbf{F}_{l,k}(\lambda_l)^{\text{H}}\mathbf{F}_{l,k}(\lambda_l)\right) \quad (24)$$

$$= \sum_{i=1}^{r_l} \frac{[\mathbf{Z}_l]_{i,i}}{\left([\mathbf{\Lambda}_l]_{i,i}+\lambda_l\right)^2} + \sum_{i=r_l+1}^{N_t} \frac{[\mathbf{Z}_l]_{i,i}}{\lambda_l^2}, \quad (25)$$

where $\mathbf{Z}_l$ is the same as that in Case I. It is plausible that $f_l(\lambda_l)$ is a monotonically decreasing function for $\lambda_l > 0$ and the optimal λ_l can be obtained by using the bisection based search method, where the lower bound of λ_l is set to a small positive value.

The overall algorithm to solve Problem (13) is summarized in Algorithm 1.

Algorithm 1 Bisection Search Method to Solve Problem (13)

1: Initialize the accuracy ε, the bounds λ_l^{lb} and λ_l^{ub};
2: If $f_l(0) \le P_{l,\max}$ holds, the optimal TPC matrix is given by $\mathbf{F}_{l,k}^{\text{opt}} = \mathbf{F}_{l,k}(0), \forall k$ and terminate; Otherwise, go to step 3;
3: Calculate $\lambda_l = \left(\lambda_l^{\text{lb}} + \lambda_l^{\text{ub}}\right)/2$;
4: If $f_l(\lambda_l) \le P_{l,\max}$, set $\lambda_l^{\text{ub}} = \lambda_l$. Otherwise, set $\lambda_l^{\text{lb}} = \lambda_l$;
5: If $\left|\lambda_l^{\text{lb}} - \lambda_l^{\text{ub}}\right| \le \varepsilon$, terminate. Otherwise, go to step 2.

C. Optimizing the Phase Shifts $\boldsymbol{\theta}$

In this subsection, we focus our attention on optimizing the phase shifts $\boldsymbol{\theta}$, while fixing $\mathbf{W}, \mathbf{U}$ and $\mathbf{F}$. By substituting $\mathbf{E}_{l,k}$ into (10) and ignoring the terms that are not related to the channels, the phase shift optimization problem is formulated as:

$$\min_{\boldsymbol{\theta}} \quad -\sum_{l=1}^{L}\sum_{k=1}^{K} \text{Tr}\left(\omega_{l,k}\mathbf{W}_{l,k}\mathbf{U}_{l,k}^{\text{H}}\bar{\mathbf{H}}_{l,l,k}\mathbf{F}_{l,k}\right)$$
$$+\sum_{l=1}^{L}\sum_{n=1}^{L}\sum_{m=1}^{K} \text{Tr}\left(\omega_{n,m}\mathbf{W}_{n,m}\mathbf{U}_{n,m}^{\text{H}}\right.$$
$$\left. \times \bar{\mathbf{H}}_{l,n,m}\mathbf{F}_l\bar{\mathbf{H}}_{l,n,m}^{\text{H}}\mathbf{U}_{n,m}\right)$$
$$-\sum_{l=1}^{L}\sum_{k=1}^{K} \text{Tr}\left(\omega_{l,k}\mathbf{W}_{l,k}\mathbf{F}_{l,k}^{\text{H}}\bar{\mathbf{H}}_{l,l,k}^{\text{H}}\mathbf{U}_{l,k}\right) \quad (26a)$$
$$\text{s.t. } 0 \le \theta_m \le 2\pi, \quad m = 1,\cdots,M, \quad (26b)$$

where $\mathbf{F}_l = \sum_{k=1}^{K}\mathbf{F}_{l,k}\mathbf{F}_{l,k}^{\text{H}}$.

By using $\bar{\mathbf{H}}_{l,n,m} = \mathbf{H}_{n,m}^r\boldsymbol{\Phi}\mathbf{G}_l^r + \mathbf{H}_{l,n,m}$, we have

$$\omega_{n,m}\mathbf{W}_{n,m}\mathbf{U}_{n,m}^{\text{H}}\bar{\mathbf{H}}_{l,n,m}\mathbf{F}_l\bar{\mathbf{H}}_{l,n,m}^{\text{H}}\mathbf{U}_{n,m}$$
$$= \omega_{n,m}\mathbf{W}_{n,m}\mathbf{U}_{n,m}^{\text{H}}\mathbf{H}_{n,m}^r\boldsymbol{\Phi}\mathbf{G}_l^r\mathbf{F}_l\mathbf{G}_l^{r\text{H}}\boldsymbol{\Phi}^{\text{H}}\mathbf{H}_{n,m}^{r\text{H}}\mathbf{U}_{n,m}$$
$$+\omega_{n,m}\mathbf{W}_{n,m}\mathbf{U}_{n,m}^{\text{H}}\mathbf{H}_{l,n,m}\mathbf{F}_l\mathbf{G}_l^{r\text{H}}\boldsymbol{\Phi}^{\text{H}}\mathbf{H}_{n,m}^{r\text{H}}\mathbf{U}_{n,m}$$

$$+\omega_{n,m}\mathbf{W}_{n,m}\mathbf{U}_{n,m}^{\text{H}}\mathbf{H}_{n,m}^r\boldsymbol{\Phi}\mathbf{G}_l^r\mathbf{F}_l\mathbf{H}_{l,n,m}^{\text{H}}\mathbf{U}_{n,m}$$
$$+\omega_{n,m}\mathbf{W}_{n,m}\mathbf{U}_{n,m}^{\text{H}}\mathbf{H}_{l,n,m}\mathbf{F}_l\mathbf{H}_{l,n,m}^{\text{H}}\mathbf{U}_{n,m} \quad (27)$$

and

$$\omega_{l,k}\mathbf{W}_{l,k}\mathbf{U}_{l,k}^{\text{H}}\bar{\mathbf{H}}_{l,l,k}\mathbf{F}_{l,k} = \omega_{l,k}\mathbf{W}_{l,k}\mathbf{U}_{l,k}^{\text{H}}\mathbf{H}_{l,l,k}\mathbf{F}_{l,k}$$
$$+\omega_{l,k}\mathbf{W}_{l,k}\mathbf{U}_{l,k}^{\text{H}}\mathbf{H}_{l,k}^r\boldsymbol{\Phi}\mathbf{G}_l^r\mathbf{F}_{l,k}. \quad (28)$$

By defining

$$\mathbf{B}_{n,m} \triangleq \omega_{n,m}\mathbf{H}_{n,m}^{r\text{H}}\mathbf{U}_{n,m}\mathbf{W}_{n,m}\mathbf{U}_{n,m}^{\text{H}}\mathbf{H}_{n,m}^r, \mathbf{C}_l \triangleq \mathbf{G}_l^r\mathbf{F}_l\mathbf{G}_l^{r\text{H}}$$

and

$$\mathbf{D}_{l,n,m} \triangleq \omega_{n,m}\mathbf{G}_l^r\mathbf{F}_l^{\text{H}}\mathbf{H}_{l,n,m}^{\text{H}}\mathbf{U}_{n,m}\mathbf{W}_{n,m}\mathbf{U}_{n,m}^{\text{H}}\mathbf{H}_{n,m}^r,$$

from (27) we have

$$\text{Tr}\left(\omega_{n,m}\mathbf{W}_{n,m}\mathbf{U}_{n,m}^{\text{H}}\bar{\mathbf{H}}_{l,n,m}\mathbf{F}_l\bar{\mathbf{H}}_{l,n,m}^{\text{H}}\mathbf{U}_{n,m}\right)$$
$$= \text{Tr}\left(\boldsymbol{\Phi}^{\text{H}}\mathbf{B}_{n,m}\boldsymbol{\Phi}\mathbf{C}_l\right) + \text{Tr}\left(\boldsymbol{\Phi}^{\text{H}}\mathbf{D}_{l,n,m}^{\text{H}}\right)$$
$$+\text{Tr}\left(\boldsymbol{\Phi}\mathbf{D}_{l,n,m}\right) + \text{const}_1, \quad (29)$$

where const_1 is a constant term that does not depend on $\boldsymbol{\Phi}$.

Similarly, by defining $\mathbf{T}_{l,k} \triangleq \omega_{l,k}\mathbf{G}_l^r\mathbf{F}_{l,k}\mathbf{W}_{l,k}\mathbf{U}_{l,k}^{\text{H}}\mathbf{H}_{l,k}^r$, from (28) we have

$$\text{Tr}\left(\omega_{l,k}\mathbf{W}_{l,k}\mathbf{U}_{l,k}^{\text{H}}\bar{\mathbf{H}}_{l,l,k}\mathbf{F}_{l,k}\right) = \text{Tr}\left(\boldsymbol{\Phi}\mathbf{T}_{l,k}\right) + \text{const}_2, \quad (30)$$

where const_2 is a constant term that is independent of $\boldsymbol{\Phi}$.

By substituting (29) and (30) into the OF of Problem (26) and ignoring the constant terms, we have

$$\min_{\boldsymbol{\theta}} \quad \text{Tr}\left(\boldsymbol{\Phi}^{\text{H}}\mathbf{B}\boldsymbol{\Phi}\mathbf{C}\right) + \text{Tr}\left(\boldsymbol{\Phi}^{\text{H}}\mathbf{V}^{\text{H}}\right) + \text{Tr}\left(\boldsymbol{\Phi}\mathbf{V}\right) \quad (31a)$$
$$\text{s.t. } 0 \le \theta_m \le 2\pi, m = 1,\cdots,M, \quad (31b)$$

where $\mathbf{B}, \mathbf{C}$ and $\mathbf{V}$ are respectively given by

$$\mathbf{B} = \sum_{n=1}^{L}\sum_{m=1}^{K}\mathbf{B}_{n,m}, \mathbf{C} = \sum_{l=1}^{L}\mathbf{C}_l,$$

$$\mathbf{V} = \sum_{l=1}^{L}\sum_{n=1}^{L}\sum_{m=1}^{K}\mathbf{D}_{l,n,m} - \sum_{l=1}^{L}\sum_{k=1}^{K}\mathbf{T}_{l,k}.$$

Upon denoting the collection of diagonal elements of $\boldsymbol{\Phi}$ by $\boldsymbol{\phi} \triangleq \left[e^{j\theta_1},\cdots,e^{j\theta_m},\cdots,e^{j\theta_M}\right]^{\text{T}}$ and using the matrix identity of [39, Eq. (1.10.6)], we arrive at

$$\text{Tr}\left(\boldsymbol{\Phi}^{\text{H}}\mathbf{B}\boldsymbol{\Phi}\mathbf{C}\right) = \boldsymbol{\phi}^{\text{H}}\left(\mathbf{B}\odot\mathbf{C}^{\text{T}}\right)\boldsymbol{\phi}. \quad (32)$$

Let $\mathbf{v}$ be the collection of diagonal elements of matrix $\mathbf{V}$, given by $\mathbf{v} = \left[[\mathbf{V}]_{1,1},\cdots,[\mathbf{V}]_{M,M}\right]^{\text{T}}$. Then, we have

$$\text{Tr}\left(\boldsymbol{\Phi}\mathbf{V}\right) = \boldsymbol{\phi}^{\text{T}}\mathbf{v}, \text{Tr}\left(\boldsymbol{\Phi}^{\text{H}}\mathbf{V}^{\text{H}}\right) = \mathbf{v}^{\text{H}}\boldsymbol{\phi}^*. \quad (33)$$

Hence, Problem (31) can be rewritten as

$$\min_{\boldsymbol{\theta}} \quad \boldsymbol{\phi}^{\text{H}}\boldsymbol{\Xi}\boldsymbol{\phi} + \boldsymbol{\phi}^{\text{T}}\mathbf{v} + \mathbf{v}^{\text{H}}\boldsymbol{\phi}^* \quad (34a)$$
$$\text{s.t. } 0 \le \theta_m \le 2\pi, \quad m = 1,\cdots,M, \quad (34b)$$

where $\boldsymbol{\Xi} = \mathbf{B}\odot\mathbf{C}^{\text{T}}$. It can be readily verified that $\mathbf{B}$ and $\mathbf{C}^{\text{T}}$ are semidefinite matrices. Then, according to [39], the Hadamard product $\mathbf{B}\odot\mathbf{C}^{\text{T}}$ (or equivalently $\boldsymbol{\Xi}$) is also a semidefinite matrix.

Recall that $\phi_m = e^{j\theta_m}, \forall m$, and that $\boldsymbol{\phi} = [\phi_1, \cdots, \phi_M]^{\mathrm{T}}$. Then, Problem (34) can be equivalently rewritten as

$$\min_{\boldsymbol{\phi}} \ f(\boldsymbol{\phi}) \triangleq \boldsymbol{\phi}^{\mathrm{H}}\boldsymbol{\Xi}\boldsymbol{\phi} + 2\mathrm{Re}\left\{\boldsymbol{\phi}^{\mathrm{H}}\mathbf{v}^*\right\} \tag{35a}$$

$$\mathrm{s.t.} \ |\phi_m| = 1, \quad m = 1, \cdots, M. \tag{35b}$$

Due to the unit modulus constraint in (35b), Problem (35) is a non-convex optimization problem. In the following, we provide a pair of efficient algorithms for solving this problem.

1) Majorization-Minimization (MM) Algorithm: We adopt the MM algorithm [34] to solve Problem (35), which was originally introduced in [40]. Then, this method has been widely used in resource allocation for wireless communication networks [41]–[43]. The main idea is to solve a difficult problem by constructing a series of more tractable approximate subproblems. Specifically, let us denote the solution of the subproblem at the tth iteration by $\boldsymbol{\phi}^t$, and the OF value of Problem (35) at the tth iteration by $f(\boldsymbol{\phi}^t)$. Then, at the $(t+1)$st iteration, we have to introduce an upper bound [2] of the OF function based on the previous solution, which is denoted as $g(\boldsymbol{\phi}|\boldsymbol{\phi}^t)$. We solve the approximate subproblem with the aid of the new OF $g(\boldsymbol{\phi}|\boldsymbol{\phi}^t)$ at the $(t+1)$st iteration. If the OF $g(\boldsymbol{\phi}|\boldsymbol{\phi}^t)$ satisfies the following three conditions:

1) $g(\boldsymbol{\phi}^t|\boldsymbol{\phi}^t) = f(\boldsymbol{\phi}^t)$,
2) $\nabla_{\boldsymbol{\phi}} g(\boldsymbol{\phi}|\boldsymbol{\phi}^t)|_{\boldsymbol{\phi}=\boldsymbol{\phi}^t} = \nabla_{\boldsymbol{\phi}} f(\boldsymbol{\phi}^t)|_{\boldsymbol{\phi}=\boldsymbol{\phi}^t}$,
3) $g(\boldsymbol{\phi}|\boldsymbol{\phi}^t) \geq f(\boldsymbol{\phi})$,

then the sequence of the solutions obtained in each iteration will result in a monotonically decreasing OF $\{f(\boldsymbol{\phi}^t), t = 1, 2, \cdots\}$ and finally converge. The converged solution satisfies the Karush-Kuhn-Tucker (KKT) optimality conditions of Problem (35) [44]. The first two conditions represent that the OF $g(\boldsymbol{\phi}|\boldsymbol{\phi}^t)$ introduced and its first-order gradient should be the same as the original OF and its first-order gradient at point $\boldsymbol{\phi}^t$. The third condition means that the OF $g(\boldsymbol{\phi}|\boldsymbol{\phi}^t)$ constructed should represent the upper bound of the original OF. To make this algorithm work, the most important task is to find the OF $g(\boldsymbol{\phi}|\boldsymbol{\phi}^t)$, which should satisfy these three conditions and should be much more tractable than $f(\boldsymbol{\phi})$.

To this end, we first introduce the following lemma proposed in [45].

Lemma 1: For any given solution $\boldsymbol{\phi}^t$ at the tth iteration and for any feasible $\boldsymbol{\phi}$, we have

$$\boldsymbol{\phi}^{\mathrm{H}}\boldsymbol{\Xi}\boldsymbol{\phi} \leq \boldsymbol{\phi}^{\mathrm{H}}\mathbf{X}\boldsymbol{\phi} - 2\mathrm{Re}\left\{\boldsymbol{\phi}^{\mathrm{H}}\left(\mathbf{X} - \boldsymbol{\Xi}\right)\boldsymbol{\phi}^t\right\}$$
$$+ \left(\boldsymbol{\phi}^t\right)^{\mathrm{H}}\left(\mathbf{X} - \boldsymbol{\Xi}\right)\boldsymbol{\phi}^t \triangleq y(\boldsymbol{\phi}|\boldsymbol{\phi}^t), \tag{36}$$

where $\mathbf{X} = \lambda_{\max}\mathbf{I}_M$ and $\lambda_{\max}$ is the maximum eigenvalue of $\boldsymbol{\Xi}$. $\qquad\square$

Upon constructing the surrogate OF $g(\boldsymbol{\phi}|\boldsymbol{\phi}^t)$ as follows:

$$g(\boldsymbol{\phi}|\boldsymbol{\phi}^t) = y(\boldsymbol{\phi}|\boldsymbol{\phi}^t) + 2\mathrm{Re}\left\{\boldsymbol{\phi}^{\mathrm{H}}\mathbf{v}^*\right\}, \tag{37}$$

where $y(\boldsymbol{\phi}|\boldsymbol{\phi}^t)$ is defined in (36), it can be readily verified that $g(\boldsymbol{\phi}|\boldsymbol{\phi}^t)$ given in (37) satisfies the three conditions. Additionally, the OF $g(\boldsymbol{\phi}|\boldsymbol{\phi}^t)$ is more tractable than the original

OF $f(\boldsymbol{\phi})$. Specifically, the subproblem to be solved at the tth iteration is given by

$$\min_{\boldsymbol{\phi}} \ g(\boldsymbol{\phi}|\boldsymbol{\phi}^t) \tag{38a}$$

$$\mathrm{s.t.} \ |\phi_m| = 1, \quad m = 1, \cdots, M. \tag{38b}$$

Since $\boldsymbol{\phi}^{\mathrm{H}}\boldsymbol{\phi} = M$, we have $\boldsymbol{\phi}^{\mathrm{H}}\mathbf{X}\boldsymbol{\phi} = M\lambda_{\max}$, which is a constant. By removing the other constants, Problem (38) can be rewritten as follows:

$$\max_{\boldsymbol{\phi}} \ 2\mathrm{Re}\left\{\boldsymbol{\phi}^{\mathrm{H}}\mathbf{q}^t\right\} \tag{39a}$$

$$\mathrm{s.t.} \ |\phi_m| = 1, \quad m = 1, \cdots, M, \tag{39b}$$

where $\mathbf{q}^t = (\lambda_{\max}\mathbf{I}_M - \boldsymbol{\Xi})\boldsymbol{\phi}^t - \mathbf{v}^*$. The optimal solution of Problem (39) is given by

$$\boldsymbol{\phi}^{t+1} = e^{j\arg(\mathbf{q}^t)}. \tag{40}$$

Based on the above discussions, we provide the details of the MM algorithm in Algorithm 2. When the algorithm converges, we can obtain the optimal phase shift as $\boldsymbol{\theta}^\star = \arg(\mathbf{q}^t)$.

Algorithm 2 MM Algorithm

1: Initial the iteration number $t = 1$, the accuracy ε. Input the feasible solution $\boldsymbol{\phi}^0$. Calculate the value of the objective function in Problem (35) as $f(\boldsymbol{\phi}^1)$;
2: Calculate $\mathbf{q}^t = (\lambda_{\max}\mathbf{I}_M - \boldsymbol{\Xi})\boldsymbol{\phi}^t - \mathbf{v}^*$;
3: Update $\boldsymbol{\phi}^{t+1}$ in (40);
4: Calculate the objective function $f(\boldsymbol{\phi}^{t+1})$, if $\left|f(\boldsymbol{\phi}^{t+1}) - f(\boldsymbol{\phi}^t)\right|/f(\boldsymbol{\phi}^{t+1}) \leq \varepsilon$ holds, terminate; Otherwise, set $t \leftarrow t + 1$ and go to step 2.

---

2) Complex Circle Manifold (CCM) Method: In this subsection, we adopt the CCM method proposed in [35] for directly solving Problem (35). We first transform Problem (35) into the following equivalent problem

$$\min_{\boldsymbol{\phi}} \ \bar{f}(\boldsymbol{\phi}) \triangleq \boldsymbol{\phi}^{\mathrm{H}}(\boldsymbol{\Xi} + \alpha\mathbf{I}_M)\boldsymbol{\phi} + 2\mathrm{Re}\left\{\boldsymbol{\phi}^{\mathrm{H}}\mathbf{v}^*\right\} \tag{41a}$$

$$\mathrm{s.t.} \ |\phi_m| = 1, \quad m = 1, \cdots, M, \tag{41b}$$

where $\alpha > 0$ is a positive constant parameter, the value of which will be given in Theorem 1. Problem (35) is equivalent to Problem (41), since we have $\alpha\boldsymbol{\phi}^{\mathrm{H}}\boldsymbol{\phi} = \alpha M$. The parameter α can control the convergence of the CCM method, which will be discussed in Theorem 1.

The search space in Problem (41) can be regarded as the product of M complex circles[3], which is a sub-manifold of $\mathbb{C}^M$ given by

$$\mathcal{S}^M \triangleq \left\{\mathbf{x} \in \mathbb{C}^M : |x_l| = 1, l = 1, 2, \cdots, M\right\}, \tag{42}$$

where x_l is the lth element of vector $\mathbf{x}$.

The main idea of the CCM algorithm is to derive a gradient descent algorithm based on the manifold space defined in (42), which is similar to the concept of the gradient descent technique developed for the conventional optimization over the

---

[2]Please note that we consider the minimization problem here.

[3]Each complex circle is given by $\mathcal{S} \triangleq \{x \in \mathbb{C} : x^*x = \mathrm{Re}\{x\}^2 + \mathrm{Im}\{x\}^2 = 1\}$, which is a sub-manifold of $\mathbb{C}$ [35].

Algorithm 3 CCM Algorithm

1: Initial the iteration number $t = 1$, the accuracy ε. Input the feasible solution ϕ^1. Calculate the value of the OF in Problem (41) as $\bar{f}(\phi^1)$;

2: Calculate the Euclidean gradient η^t in (43);

3: Calculate the Riemannian gradient $\mathbf{P}_{\mathcal{T}_{\phi^t}\mathcal{S}^M}(\eta^t)$ in (44);

4: Update over the tangent space according to (45);

5: Update ϕ^{t+1} by retracting $\bar{\phi}^t$ to the complex circle manifold $\mathcal{S}^M$ according to (46);

6: Calculate the OF $\bar{f}(\phi^{t+1})$, if $|\bar{f}(\phi^{t+1}) - \bar{f}(\phi^t)|/\bar{f}(\phi^{t+1}) \le \varepsilon$ holds, terminate; Otherwise, set $t \leftarrow t + 1$ and go to step 2.
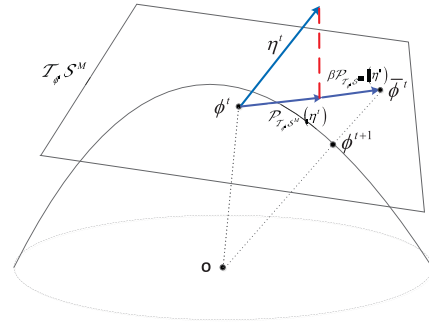

Fig. 2. Geometric interpretation of the CCM algorithm.

TABLE I

COMPUTATIONAL COMPLEXITY COMPARISON FOR TWO DIFFERENT ALGORITHMS TO FIND THE PHASE SHIFTS

Algorithms	MM Alg.	CCM Alg.
Complexity	$\mathcal{O}(M^3 + T_{MM}M^2)$	$\mathcal{O}(M^3 + T_{CCM}M^2)$

Euclidean space. The main steps of the CCM algorithm is composed of four main steps in each iteration t:

1) *Gradient in Euclidean Space:* We first have to find the search direction and the most common search direction for a minimization problem is to move in the direction opposite to the gradient of $\bar{f}(\phi^t)$, which is given by

$$\eta^t = -\nabla_\phi \bar{f}(\phi^t) = -2(\boldsymbol{\Xi} + \alpha\mathbf{I}_M)\phi^t - 2\mathbf{v}^*. \quad (43)$$

2) *Riemannian gradients:* Since we optimize over the manifold space, we have to find the Riemannian gradient [12]. The Riemannian gradient of $\bar{f}(\phi^t)$ at the current point $\phi^t \in \mathcal{S}^M$ is in the tangent space $\mathcal{T}_{\phi^t}\mathcal{S}^M$.[4] Specifically, the Riemannian gradient of $\bar{f}(\phi^t)$ at ϕ^t can be obtained by projecting the search direction η^t in the Euclidean space onto $\mathcal{T}_{\phi^t}\mathcal{S}^M$ by using the projection operator, which can be calculated as follows [12]:

$$\mathbf{P}_{\mathcal{T}_{\phi^t}\mathcal{S}^M}(\eta^t) = \eta^t - \mathrm{Re}\{\eta^{t*} \odot \phi^t\} \odot \phi^t. \quad (44)$$

3) *Update over the tangent space:* Update the current point ϕ^t on the tangent space $\mathcal{T}_{\phi^t}\mathcal{S}^M$:

$$\bar{\phi}^t = \phi^t + \beta\mathbf{P}_{\mathcal{T}_{\phi^t}\mathcal{S}^M}(\eta^t), \quad (45)$$

where β is a constant step size that will be discussed in Theorem 1.

4) *Retraction operator:* In general, the $\bar{\phi}^t$ obtained is not in $\mathcal{S}^M$, i.e. we have $\bar{\phi}^t \notin \mathcal{S}^M$. Hence, it has to be mapped into the manifold $\mathcal{S}^M$ by using the retraction operator[5] as follows

$$\phi^{t+1} = \bar{\phi}^t \odot \frac{1}{|\bar{\phi}^t|}. \quad (46)$$

Note that both ϕ^{t+1} and ϕ^t belongs to $\mathcal{S}^M$, which satisfies the unit constant modulus constraints. The details of the CCM algorithm are presented in Algorithm 3. The CCM algorithm is also illustrated geometrically in Fig. 2.

The following theorem provides guidance for the choices of parameters α and β to guarantee the convergence of the CCM algorithm.

---

[4]The tangent space of $\mathcal{S}$ at point z_m is defined as $\mathcal{T}_{z_m}\mathcal{S} = \{x \in \mathbb{C} : \mathrm{Re}\{x^*z_m\} = 0\}$. Then, the tangent space $\mathcal{T}_{\mathbf{z}}\mathcal{S}^M$ is the product of these M tangent space $\mathcal{T}_{z_m}\mathcal{S}$ given by $\mathcal{T}_{\mathbf{z}}\mathcal{S}^M = \mathcal{T}_{z_1}\mathcal{S} \times \mathcal{T}_{z_2}\mathcal{S} \cdots \times \mathcal{T}_{z_M}\mathcal{S}$.

[5]The retraction operator normalizes each element of $\bar{\phi}^t$ to be unit.

Theorem 1 [35]: Let $\lambda_{\boldsymbol{\Xi}}$ and $\lambda_{\boldsymbol{\Xi}+\alpha\mathbf{I}_M}$ be the largest eigenvalue of matrices $\boldsymbol{\Xi}$ and $\boldsymbol{\Xi} + \alpha\mathbf{I}_M$, respectively. If α and β are chosen to satisfy the following conditions,

$$\alpha \ge \frac{M}{8}\lambda_{\boldsymbol{\Xi}} + \|\mathbf{v}\|_2, \quad 0 < \beta < \frac{1}{\lambda_{\boldsymbol{\Xi}+\alpha\mathbf{I}}}, \quad (47)$$

then the CCM algorithm generates a non-increasing sequence $\{\bar{f}(\phi^t), t = 1, 2, \cdots\}$, and finally converges to a finite value. $\square$

3) Complexity Analysis: In this part, we analyze the complexity of both proposed methods in solving Problem (35).

Let us now analyze the complexity of the MM algorithm. At the beginning of the MM algorithm, we have to calculate $\lambda_{\max}$, i.e. the maximum eigenvalue of $\boldsymbol{\Xi}$. The associated complexity is given by $\mathcal{O}(M^3)$. For each iteration of the MM algorithm, the main complexity lies in the calculation of $\mathbf{q}^t$ in Step 2, the complexity of which is $\mathcal{O}(M^2)$. Let us denote the number of iterations required for the MM algorithm to converge by T_{MM}. Then, the total complexity of the MM algorithm is given by $C_{\mathrm{MM}} = \mathcal{O}(M^3 + T_{MM}M^2)$.

We then analyze the complexity of the CCM algorithm. At the start of the CCM algorithm, we have to find the range of α and β to guarantee the convergence of the CCM algorithm, which relies on calculating the largest eigenvalue of the matrices $\boldsymbol{\Xi}$ ($\lambda_{\boldsymbol{\Xi}}$), as shown in Theorem 1. Its complexity order is given by $\mathcal{O}(M^3)$. For each iteration of the CCM algorithm, the complexity mainly depends on the calculation of the Euclidean gradient η^t, which is given by $\mathcal{O}(M^2)$. Let us denote the total number of iterations required by the CCM algorithm to converge by T_{CCM}. Then, the total complexity of the CCM algorithm is given by $C_{\mathrm{CCM}} = \mathcal{O}(M^3 + T_{CCM}M^2)$.

The complexity of these algorithms is summarized in Table I. It can be observed that the complexity mainly depends on the number of iterations required for convergence. The simulation results of Section V will compare their convergence speed.

D. Overall Algorithm to Solve Problem (5)

Based on the above analysis, we provide the detailed description of the BCD algorithm conceived for solving Problem (5) in Algorithm 4. In Step 5, we have to apply two algorithms for solving Problem (35) to find the phase shifts $\boldsymbol{\theta}^{(n+1)}$. Both the MM algorithm and the CCM algorithm can guarantee to yield a monotonically decreasing OF value of Problem (35) compared to the previous phase solution, i.e., $f(\boldsymbol{\phi}^{(n+1)}) < f(\boldsymbol{\phi}^{(n)})$. It can be readily verified that the OF value of Problem (9) monotonically increases in each step of Algorithm 4. Additionally, due to the power constraints, the OF value has an upper bound. Hence, Algorithm 4 is guaranteed to converge.

Algorithm 4 Block Coordinate Descent Algorithm

1: Initialize iterative number $n = 1$, maximum number of iterations $n_{\max}$, feasible $\mathbf{F}^{(1)}$, $\boldsymbol{\theta}^{(1)}$, error tolerance ε, calculate the OF value of Problem (5), denoted as $\mathrm{Obj}(\mathbf{F}^{(1)}, \boldsymbol{\theta}^{(1)})$;

2: Given $\mathbf{F}^{(n)}$ and $\boldsymbol{\theta}^{(n)}$, calculate the optimal decoding matrices $\mathbf{U}^{(n)}$ in (11);

3: Given $\mathbf{F}^{(n)}$, $\mathbf{U}^{(n)}$ and $\boldsymbol{\theta}^{(n)}$, calculate the optimal auxiliary matrices $\mathbf{W}^{(n)}$;

4: Given $\mathbf{U}^{(n)}$, $\mathbf{W}^{(n)}$ and $\boldsymbol{\theta}^{(n)}$, calculate the optimal precoding matrices $\mathbf{F}^{(n+1)}$ by solving Problem (13) with the Lagrangian multiplier method in Subsection III-B;

5: Given $\mathbf{U}^{(n)}$, $\mathbf{W}^{(n)}$ and $\mathbf{F}^{(n+1)}$, calculate the optimal $\boldsymbol{\theta}^{(n+1)}$ by solving Problem (35) with the algorithms developed in Subsection III-C;

6: If $n \geq n_{\max}$ or

$$\frac{\left| \mathrm{Obj}(\mathbf{F}^{(n+1)}, \boldsymbol{\theta}^{(n+1)}) - \mathrm{Obj}(\mathbf{F}^{(n)}, \boldsymbol{\theta}^{(n)}) \right|}{\mathrm{Obj}(\mathbf{F}^{(n)}, \boldsymbol{\theta}^{(n)})} < \varepsilon,$$

terminate. Otherwise, set $n \leftarrow n + 1$ and go to step 2.

Let us now analyze the complexity of the BCD algorithm. In Step 2, the complexity of computing the decoding matrices $\mathbf{U}^{(n)}$ is $\mathcal{O}(LKN_r^3)$. In Step 3, the complexity of calculating the auxiliary matrices $\mathbf{W}^{(n)}$ is given by $\mathcal{O}(LKd^3)$. In Step 4, we have to calculate the TPC matrices $\mathbf{F}^{(n+1)}$. The detailed analysis is provided as follows. For any pair of complex matrices $\mathbf{X} \in \mathbb{C}^{m \times n}, \mathbf{Y} \in \mathbb{C}^{n \times p}$, the complexity of computing $\mathbf{XY}$ is $\mathcal{O}(mnp)$ [46]. We assume that $N_t > N_r > d$. Hence, the complexity of computing the matrices $\{\mathbf{A}_{l,k}, \forall l, k\}$ in (14) is given by $\mathcal{O}(LKN_t^2 d)$. The complexity of calculating $\mathbf{F}_{l,k}$ in (15) is given by $\mathcal{O}(LKN_t^3)$. The SVD decomposition of $\{\mathbf{A}_l, \forall l\}$ is given by $\mathcal{O}(LN_t^3)$. The complexity of calculating $\{\mathbf{Z}_l\}$ is given by $\mathcal{O}(L^2 N_t^2 N_r)$. The complexity of evaluating the Lagrangian multipliers $\{\lambda_l, \forall l\}$ can be ignored. Hence, the overall complexity of calculating the TPC matrices $\mathbf{F}^{(n+1)}$ is given by $\mathcal{O}(\max\{LKN_t^3, L^2 N_t^2 N_r\})$. The complexity of calculating the optimal $\boldsymbol{\theta}^{(n+1)}$ is given in Table I, while the complexity of each algorithm is denoted by $C_i, i = $ MM, CCM. Then, the overall complexity of the BCD algorithm is given by

$$C_{\mathrm{BCD,i}} = \mathcal{O}(\max\{LKN_t^3, L^2 N_t^2 N_r, C_i\}), \quad i = \mathrm{MM, CCM},$$
$$\text{(48)}$$

where $C_{\mathrm{BCD,i}}$ denotes the overall complexity of the BCD algorithm, when the phase shifts are obtained by using method i, $i = \mathrm{MM, CCM}$.

IV. EXTENSION TO OTHER SCENARIOS

A. Network MIMO

In network MIMO, multiple BSs in different cells cooperate with each other and send the same data to each user. In this scenario, the antennas of all BSs form a giant antenna array and jointly serve each user, where the inter-cell interference can be effectively mitigated [3]–[5]. It should be emphasized that compared to the model in Section II, the data should be shared among multiple BSs, which incurs increased information exchange overhead.

Let $\mathbf{F}_{i,l,k}$ be the precoding matrix of the ith BS for the kth user in the lth cell, and $\mathbf{F}_{l,k} = \left[\mathbf{F}_{i,l,k}^{\mathrm{H}}, \forall i \right]^{\mathrm{H}} \in \mathbb{C}^{LN_t \times d}$ be the overall precoding matrix from all BSs to the user. Define $\mathbf{G}^r = [\mathbf{G}_i^r, \forall i] \in \mathbb{C}^{M \times LN_t}$ as the overall channel from all the BSs to the IRS, and $\mathbf{H}_{l,k} = [\mathbf{H}_{i,l,k}, \forall i] \in \mathbb{C}^{N_r \times LN_t}$ the direct channel from all the BSs to the kth user in the lth cell. Let $\bar{\mathbf{H}}_{l,k} \triangleq \mathbf{H}_{r,l,k}\boldsymbol{\Phi}\mathbf{G}^r + \mathbf{H}_{l,k}$ be the equivalent channel spanning from all the BSs to the kth user in the lth cell.

Then, the signal received at the kth user in the lth cell is given by

$$\mathbf{y}_{l,k} = \bar{\mathbf{H}}_{l,k}\mathbf{F}_{l,k}\mathbf{s}_{l,k} + \sum_{m=1,m \neq k}^{K} \bar{\mathbf{H}}_{l,k}\mathbf{F}_{l,m}\mathbf{s}_{l,m}$$
$$+ \sum_{i=1,i \neq l}^{L} \sum_{m=1}^{K} \bar{\mathbf{H}}_{l,k}\mathbf{F}_{i,m}\mathbf{s}_{i,m} + \mathbf{n}_{l,k}. \quad \text{(49)}$$

The data rate of the kth user in the lth cell is given by

$$R_{l,k}(\mathbf{F}, \boldsymbol{\theta}) = \log \left| \mathbf{I} + \bar{\mathbf{H}}_{l,k}\mathbf{F}_{l,k}\mathbf{F}_{l,k}^{\mathrm{H}}\bar{\mathbf{H}}_{l,k}^{\mathrm{H}}\mathbf{J}_{l,k}^{-1} \right|, \quad \text{(50)}$$

where $\mathbf{J}_{l,k}$ is given by

$$\mathbf{J}_{l,k} = \sum_{m=1,m \neq k}^{K} \bar{\mathbf{H}}_{l,k}\mathbf{F}_{l,m}\mathbf{F}_{l,m}^{\mathrm{H}}\bar{\mathbf{H}}_{l,k}^{\mathrm{H}}$$
$$+ \sum_{i=1,i \neq l}^{L} \sum_{m=1}^{K} \bar{\mathbf{H}}_{l,k}\mathbf{F}_{i,m}\mathbf{F}_{i,m}^{\mathrm{H}}\bar{\mathbf{H}}_{l,k}^{\mathrm{H}} + \sigma^2 \mathbf{I}.$$

The weighted sum rate problem is the same as in (5), except that the power constraint for each BS is formulated as follows:

$$\sum_{l=1}^{L} \sum_{k=1}^{K} \|\mathbf{F}_{i,l,k}\|_F^2 \leq P_{i,\max}, \quad i = 1, \cdots, L. \quad \text{(51)}$$

The optimization problem formulated for the case of Network MIMO can be similarly solved by using the methods of Section III, details of which are omitted for simplicity.

B. Multiple-IRS Scenario

Assume that the system has A IRSs, each of which has M reflection elements. The baseband channels spanning from the ath IRS to the kth user in the lth cell, and the ones from the ith BS to the ath IRS are denoted by $\mathbf{H}_{a,l,k}^r$ and $\mathbf{G}_{i,a}^r$,

respectively. The diagonal phase-shifting matrix of the ath IRS is denoted by $\boldsymbol{\Phi}_a = \text{diag}\left\{e^{j\theta_{a,1}}, \cdots, e^{j\theta_{a,m}}, \cdots, e^{j\theta_{a,M}}\right\}$. Then, the received signal vector at the kth user in the lth cell is given by

$$\mathbf{y}_{l,k} = \sum_{n=1}^{L} \mathbf{H}_{n,l,k}\mathbf{x}_n + \sum_{n=1}^{L}\sum_{a=1}^{A} \mathbf{H}_{a,l,k}^r \boldsymbol{\Phi}_a \mathbf{G}_{n,a}^r \mathbf{x}_n + \mathbf{n}_{l,k},$$

$$(52)$$

where $\mathbf{H}_{n,l,k}$, $\mathbf{x}_n$ and $\mathbf{n}_{l,k}$ are defined in Section II.

By defining $\mathbf{H}_{l,k}^r = \left[\mathbf{H}_{1,l,k}^r, \cdots, \mathbf{H}_{A,l,k}^r\right]$, $\boldsymbol{\Phi} = \text{diag}\left\{\boldsymbol{\Phi}_1, \cdots, \boldsymbol{\Phi}_A\right\}$ and $\mathbf{G}_n^r = \left[\mathbf{G}_{n,1}^{r\text{H}}, \cdots, \mathbf{G}_{n,A}^{r\text{H}}\right]^{\text{H}}$, (52) can be rewritten as

$$\mathbf{y}_{l,k} = \sum_{n=1}^{L} \mathbf{H}_{n,l,k}\mathbf{x}_n + \sum_{n=1}^{L} \mathbf{H}_{l,k}^r \boldsymbol{\Phi} \mathbf{G}_n^r \mathbf{x}_n + \mathbf{n}_{l,k}, \quad (53)$$

which is the same as (2). Hence, the derivations for the single-IRS scenario are directly applicable.

V. SIMULATION RESULTS

In this section, simulation results are provided for validating the benefits of employing IRSs for improving WSR of multicell systems. The large-scale path loss in dB is given by

$$\text{PL} = \text{PL}_0 - 10\alpha\log_{10}\left(\frac{d}{d_0}\right), \quad (54)$$

where PL_0 is the path-loss at the reference distance d_0, d is the link distance, α is the path-loss exponent. In our simulations, we set $\text{PL}_0 = -30$ dB and $d_0 = 1$ m. Due to extensive obstacles and scatterers, the path-loss exponent between the BS and the users is given by $\alpha_{\text{BU}} = 3.75$. The heights of BSs, IRSs, and users are assumed to be 30 m, 10 m, and 1.5 m, respectively. By carefully choosing the location of the IRS, the IRS-aided link has a higher probability of experiencing nearly free-space path loss. Then, we set the path-loss exponents of the BS-IRS link and of the IRS-user link to $\alpha_{\text{BI}} = \alpha_{\text{IU}} \triangleq \alpha_{\text{IRS}} = 2.2$. For the direct channel from the BSs to users, the small-scale fading is assumed to be Rayleigh fading due to extensive scatters. However, for the IRS-related channels, the small-scale fading is assumed to be Rician fading. In specific, the small-scale channel can be modeled as

$$\tilde{\mathbf{H}} = \sqrt{\frac{\beta}{\beta+1}}\tilde{\mathbf{H}}^{\text{LoS}} + \sqrt{\frac{1}{\beta+1}}\tilde{\mathbf{H}}^{\text{NLoS}}, \quad (55)$$

where β is the Rician factor, $\tilde{\mathbf{H}}^{\text{LoS}}$ is the deterministic line of sight (LoS), and $\tilde{\mathbf{H}}^{\text{NLoS}}$ is the non-LoS (NLoS) component that is Rayleigh fading. The LoS component $\tilde{\mathbf{H}}^{\text{LoS}}$ is given by $\tilde{\mathbf{H}}^{\text{LoS}} = \mathbf{a}_{D_r}\left(\vartheta^{AoA}\right)\mathbf{a}_{D_t}^H\left(\vartheta^{AoD}\right)$, where $\mathbf{a}_{D_r}\left(\vartheta^{AoA}\right)$ is defined as

$$\mathbf{a}_{D_r}\left(\vartheta^{AoA}\right) = \left[1, e^{j\frac{2\pi d}{\lambda}\sin\vartheta^{AoA}}, \cdots, e^{j\frac{2\pi d}{\lambda}(D_r-1)\sin\vartheta^{AoA}}\right]^T$$

$$(56)$$

and

$$\mathbf{a}_{D_t}\left(\vartheta^{AoD}\right) = \left[1, e^{j\frac{2\pi d}{\lambda}\sin\vartheta^{AoD}}, \cdots, e^{j\frac{2\pi d}{\lambda}(D_t-1)\sin\vartheta^{AoD}}\right]^T.$$
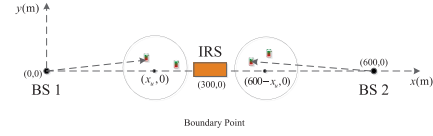
$$(57)$$

Fig. 3. The simulated two-cell IRS-aided MIMO communication scenario.

In (56) and (57), D_r and D_t are the number of antennas/elements at the receiver side and transmitter side, respectively, d is the antenna separation distance, λ is the wavelength, ϑ^{AoD} is the angle of departure and ϑ^{AoA} is the angle of arrival. It is assumed that ϑ^{AoD} and ϑ^{AoA} are randomly distributed within $[0, 2\pi]$. For simplicity, we set $d/\lambda = 1/2$. Unless otherwise stated, we set the simulation parameters as follows: Channel bandwidth of 10 MHz, noise power density of -174 dBm/Hz, number of transmit antennas of $N_t = 4$, number of receive antennas of $N_r = 2$, number of data streams of $d = 2$, number of reflection elements of $M = 50$, maximum BS power of $P_{l,\max} = 1$ W, $\forall l$, Rician factor of $\beta = 3$, error tolerance of $\varepsilon = 10^{-6}$, and weighting factor of $\omega_{l,k} = 1, \forall l, k$. The x coordinate of the center point of the first circle is given by $x_u = 280$ m, which means that the users are located at the edge of their corresponding cells. The following results are obtained by averaging over 200 independent channel generations. In Step 5 of the BCD algorithm, if the MM method is used, the BCD algorithm is denoted as BCD-MM. Similar definition holds for BCD-CCM. The step parameters α and β in the CCM algorithm are set based on Theorem 1.

A. Two-Cell Scenario

In order to obtain more insights about the benefits of deploying IRS, we first consider a two-cell communication network with a single IRS shown in Fig. 3, in which there are two BSs located at $(0,0)$ and $(600,0)$,[6] respectively. By default, the IRS is deployed at the boundary point between two cells, the coordinate of which is $(300,0)$. Two users in the first cell are uniformly and randomly placed in a circle centered at $(x_u, 0)$ with radius 20 m, while two users in the second cell are also uniformly and randomly distributed in a circle centered at $(600 - x_u, 0)$ with radius 20 m. Note that these two circles are symmetric w.r.t. the boundary point.

1) Convergence Behavior of BCD Algorithm: We first study the convergence behaviour of the BCD algorithm in Algorithm 4. Fig. 4 shows the WSR versus the number of iterations for various number of phase shifts, i.e., for $M = 10, 20$ and 40. Both the BCD-MM and BCD-CCM algorithms are tested. It can be observed from this figure that both the BCD-MM and BCD-CCM have a very similar convergence speed and converged value. Having more phase shifts leads to a slightly slower convergence speed. This is due to the fact that more optimization variables are involved, and more iterations are required for convergence. However, for different values

---

[6]We only illustrate the horizontal plane of the system, where the height of various devices are not shown.
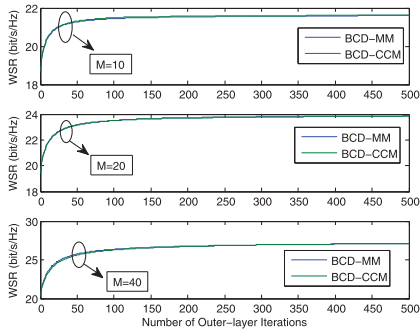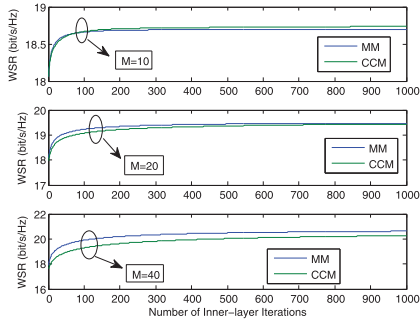
Fig. 4. Convergence behaviour of the BCD algorithm.

Fig. 5. Convergence behaviour of the MM and CCM algorithm.

of M, the proposed algorithms converge within 200 iterations, which confirm the practical benefits of our algorithms.

2) Convergence Behavior of the MM and CCM Algorithms: In each iteration of the BCD algorithm, we have to use the MM or CCM algorithm for finding the phase shifts of the IRS. Fig. 5 shows the convergence performance of the MM and CCM algorithms for the first iteration of the BCD algorithm. It can be seen from Fig. 5 that the MM algorithm converges a little faster than the CCM algorithm, which implies having a lower computational complexity for the MM algorithm based on the complexity analysis of Table I. As expected, the number of iterations required for the convergence of the two algorithms increase with the number of phase shifts, since more variables have to be optimized. For different values of M, the MM algorithm and CCM algorithm may converge to different values. However, as seen from Fig. 4, the final WSR value obtained by the BCD algorithm by using different algorithms to update the phase shifts is almost the same.

We then compare our proposed algorithms to the following benchmark schemes:

1) **RandPhase**: We assume that the phase for each reflection element is uniformly and independently generated from $[0, 2\pi]$. We only have to optimize the TPC matrices, which can be obtained by skipping Step 5 of the BCD algorithm.

2) **No-IRS**: Set the IRS related channel matrices to zero matrices, i.e., $\mathbf{H}_{r,l,k} = \mathbf{0}$, $\mathbf{G}_{n,r} = \mathbf{0}$, $\forall n, l, k$. Then, use the BCD algorithm to find the optimal TPC matrices by removing Step 5 for the phase shift update.

3) Impact of the Number of Phase Shifts: Fig. 6 compares the WSR performance of various algorithms versus the number
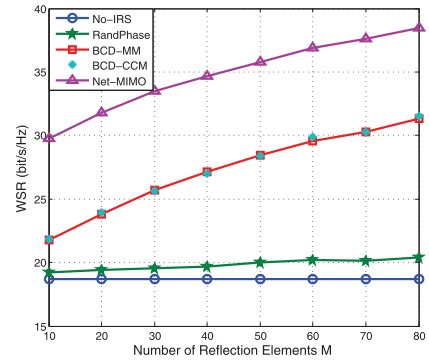
Fig. 6. Achievable WSR versus the number of phase shifts M.

of phase shifts M. The performance of Network MIMO scheme (with legend 'Net-MIMO') proposed in Section IV-A is also compared. We can observe that both the BCD-MM algorithm and BCD-CCM algorithm have similar performances over the entire range of M, and both of them significantly outperform the other two benchmark schemes. The performance gain becomes quite pronounced upon increasing M. Specifically, when $M = 10$, the performance gain over the No-IRS is only 2 bit/s/Hz, while the performance gain increases up to 13 bit/s/Hz when $M = 80$. This is mainly attributed to two reasons. Firstly, the signal power received at the IRS can be enhanced by increasing M, leading to a higher array gain. On the other hand, by appropriately designing the phase shifts, the reflected signal power received by the users increases with M. Hence, the proposed IRS-assisted system can exploit not only the array gain, but also the reflecting beamforming gain at the IRS. More importantly, the IRS is a passive reflection device, hence installing more passive reflecting elements is both energy-efficient and economical since the IRS does not require active radio frequency chains and power amplifiers as in conventional transmitters. These results demonstrate that introducing IRSs into wireless communications enhances the system performance, and it is a promising technique for future networks. It is seen that the performance of the RandPhase algorithm is slightly better than that of the No-IRS scheme. This is because the reflected signals have not been carefully beamed towards the receivers. By contrast, for the proposed algorithms, both the direct signals and reflected signals are superposed more constructively, while the multicell interference signals are added destructively. As expected, the WSR achieved by the Network MIMO is significantly higher than that of the system studied in Section II (denoted as 'Coordinated beamforming'). However, this performance gain is attained at the cost of the heavy information exchange associated with Network MIMO, where the data streams of all the users should be exchanged. By contrast, only the CSI has to be shared among the BSs, the amount of which is much lower than that of the data.

4) Impact of the IRS-Related Path Loss Exponent: In the above examples, the path loss exponents of the IRS-related links is set as $\alpha_{\text{IRS}} = 2.2$, since we assume that the location of the IRS can be appropriately chosen for ensuring that a free space BS-IRS link and IRS-user link can be established.
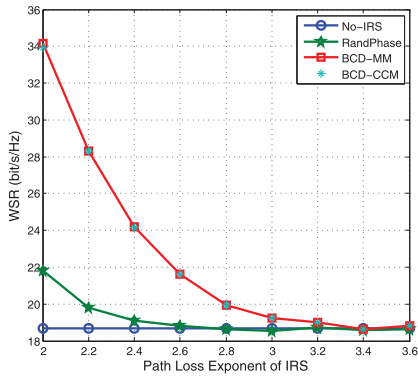
Fig. 7. Achievable WSR versus IRS-related path loss exponent.

Fig. 8. Achievable WSR versus the location of the IRS x_{IRS}.

Fig. 9. Achievable WSR versus the location of the UE x_u.

However, in some practical scenarios, it may not be feasible to find such ideal places. Hence, it is intriguing to investigate the performance gain that can be achieved by our proposed algorithms when the IRS-related links experience rich scattering fading with higher value of α_{IRS}. To this end, we plot Fig. 7 to show the impact of the IRS-related path-loss exponent. As expected, the WSR achieved by the proposed algorithms decreases upon increasing α_{IRS}, and finally converges to the same WSR as achieved by the No-IRS scheme. This is because upon increasing α_{IRS}, the signal attenuation associated with the IRS-related links becomes larger, and the signal received from the IRS is weaker, hence more negligible. However, when α_{IRS} is very small, significant performance gains can be achieved by our proposed algorithms over the No-IRS scheme. For example, for a free-space channel associated with $\alpha_{\mathrm{IRS}} = 2$, the performance gain is up to 14.5 bit/s/Hz. Hence, for multicell systems, the performance gain of IRS-assisted systems may be attributed to the favourable channel conditions of the BS-IRS link and IRS-user link. This provides an important engineering design insight, where the IRS should be deployed in an obstacle-free scenario, such as the ceiling for indoor use or advertisement panels for outdoor use. Otherwise, the performance gain brought about by the IRS is marginal. Fig. 7 also shows that if the phase shifts are not optimized, the performance of an IRS-aided system may even be worse than that operating without the IRS, i.e. the WSR achieved by the RandPhase algorithm is equal to or lower than that of the No-IRS scheme. This emphasizes the importance of jointly optimizing the TPC matrices and the phase shifts at the IRS.

5) Impact of the IRS Location: Denote the coordinate of the IRS as $(x_{\mathrm{IRS}}, 0)$. In Fig. 8, we study the impact of the IRS location by moving the IRS from $x_{\mathrm{IRS}} = 50$ m (cell center of the first cell) to $x_{\mathrm{IRS}} = 300$ m (cell boundary). It may be observed again that both the proposed algorithms achieve the similar performance, and drastically improves the WSR performance over the other benchmark schemes. It is interesting to observe that the WSR achieved by the proposed algorithms first decreases with x_{IRS} (50 m$< x_{\mathrm{IRS}} < 150$ m), and then increases for $x_{\mathrm{IRS}} > 150$ m. This becomes plausible upon considering a special case, where the IRS lies on the line between the BS and the user central point. Let us denote the distance between the BS and the IRS by d, and that
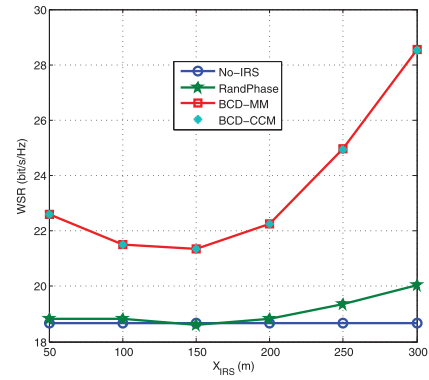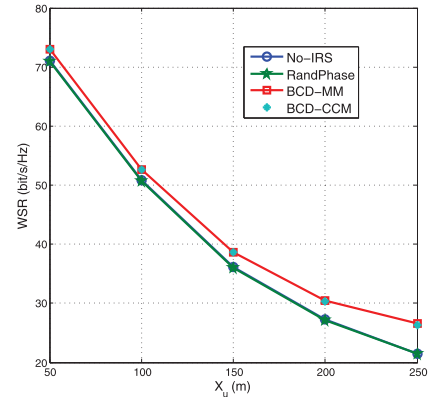
between the BS and the user central point by D. By ignoring the small-scale fading, the large-scale channel gain of the combined channel from the IRS may be approximated by

$$\mathrm{PL}_{\mathrm{IRS}} = 2\mathrm{PL}_0 - 10\alpha_{\mathrm{IRS}}\log_{10}(d) - 10\alpha_{\mathrm{IRS}}\log_{10}(\mathrm{D\text{-}d}),$$

$$(58)$$

which achieves its minimum value at $d^\star = D/2$. Hence, the combined channel gain achieves its minimum value when the IRS is located at the middle point, which is consistent with the simulation results of Fig. 8. Due to the strong BS-IRS link, the WSR performance gain achieved by our proposed algorithms over the No-IRS is 4 bit/s/Hz at $x_{\mathrm{IRS}} = 50$ m. However, this performance gain doubles when the IRS moves to the boundary of these two cells. This performance is partly due to the favourable IRS-user channel link. The other important reason is that we can optimize the phase shifts of the IRS to make the equivalent channel spanning from the inter-cell BS to the users approach zero matrices. Specifically, we can optimize $\boldsymbol{\Phi}$ to let $\bar{\mathbf{H}}_{n,l,k}, n \neq l$ approach zero matrices. This alleviates the severe inter-cell interference for the cell-edge users, which significantly enhances the system performance. Additionally, deploying the IRS at the cell center for Cell 1 is only beneficial for the users in Cell 1, while all the users will benefit from the IRS, when positioning it at the cell boundary. This means that for multicell communication systems, significant performance gains can be obtained when
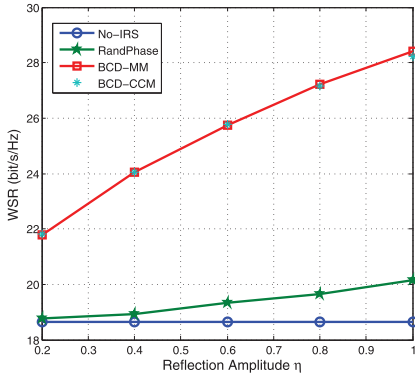
Fig. 10.   Achievable WSR versus the reflection amplitude η.

Fig. 11.   Individual data rate under two sets of weights.

the IRS is employed at the cell boundary, which mitigates the inter-cell interference. Furthermore, the phase shifts should be carefully designed. Otherwise, the performance may in fact become inferior to that without IRS, e.g., $x_{\mathrm{IRS}} = 150$ m.

6) Impact of the User Location: In Fig. 9, we compare the WSR achieved by all schemes versus the horizontal distance between BS 1 and the first circle central point, i.e., x_u. Since the users are randomly positioned in this circle, this is equivalent to varying the locations of the users. It is again observed that the proposed algorithms achieve almost the same performance and achieve superior performance over the other two benchmark schemes. Additionally, the performance gap increases with x_u, because the users receive strong reflected signals from the IRS, when the users approach the cell edge. This means that the IRS mitigates the inter-cell interference.

7) Impact of the Reflection Amplitude: Due to the absorption and parasitic reflection of the phase shifters, there may be a signal power loss at the IRS. Then, in Fig. 10, we study the impact of the reflection amplitude on the system performance. Specifically, the phase-shift matrix of the IRS is rewritten as $\Phi = \eta \mathrm{diag} \left\{ e^{j\theta_1}, \cdots, e^{j\theta_m}, \cdots, e^{j\theta_M} \right\}$, where the reflection amplitudes of all the elements are the same as η. As expected, the WSR achieved by the IRS-aided scheme increases with η due to the reduced power loss. The reflection amplitude has a substantial impact on the system performance. Specifically, when η increases from 0.2 to 1, the WSR increases by about 6 bit/s/Hz.

8) Impact of the Weights: As mentioned in our problem formulation, the weights can be used for controlling the fairness among the users. To be more explicit, we provide an example for illustrating this point. For clarity, the index of the jth user in the ith cell is denoted as $2(i-1)+j$. For example, the index of the second user in the second cell is 4. The coordinates for the four users (two in each cell) are respectively given by $(100, 0), (250, 0), (350, 0)$ and $(500, 0)$, which indicate that the first user is closer to BS 1 than the second user, and the third user is closer to BS 2 than the fourth user. Two sets of weights are tested: 1) $\omega_k = 0.5, \forall k$; 2) $\omega_1 = 0.15, \omega_2 = 0.85, \omega_3 = 0.3, \omega_4 = 0.7$. In Fig. 11, the individual data rates achieved under two sets of weights are illustrated. For the case of the equal weights, the first user and the third user have higher data rate than the other two users,
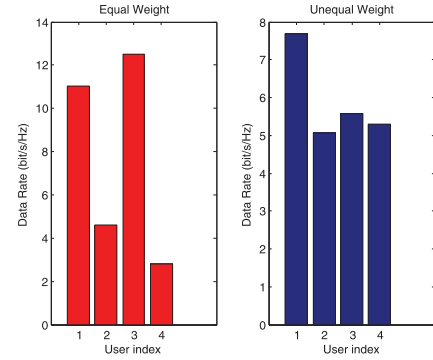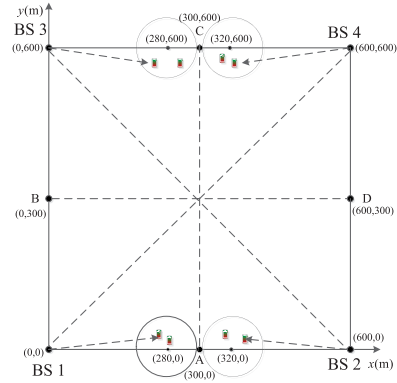

Fig. 12.   The simulated four-cell IRS-aided MIMO communication scenario.

since they are closer to the BSs. To guarantee rate-fairness amongst the users, for the case of unequal weights, a more balanced data rate distribution can be achieved by assigning higher weights to the users having low channel gains.

B. Four-Cell Scenario

Finally, in order to study the beneficial impact of IRS deployment on the system's performance, we consider the four-cell scenario of Fig. 12, where the coordinates of the four BSs are given by $(0, 0), (600, 0), (0, 600)$ and $(600, 600)$, respectively. Additionally, the coordinates of the user distribution center in the four cells are $(280, 0), (320, 0), (280, 600)$ and $(320, 600)$, respectively. The circle radius is also 20 m. Four points (i.e., A,B,C,D) are located at the middle of the corresponding two BSs. The number of antennas at each BS is set to 2, and each cell has three users.

1) Single-IRS Case: We first study the single-IRS scenario, when the number of phase shifts at the IRS is 50. Three IRS schemes are considered: 1) **Scheme-1**: As in the case of two-cell scenario, the IRS moves from BS 1 to BS 2; 2) **Scheme-2**: The IRS moves from point B to point D; 3) **Scheme-3**: The IRS moves from BS 3 to BS 2. The WSR achieved by various schemes is shown in Fig. 13. It is seen from this figure that Scheme 1 achieves its maximum WSR at the cell boundary point with $x_{\mathrm{IRS}} = 300\ m$, which implies that the IRS should be deployed at the cell edge to benefit the users in the first and second cells. This conclusion is
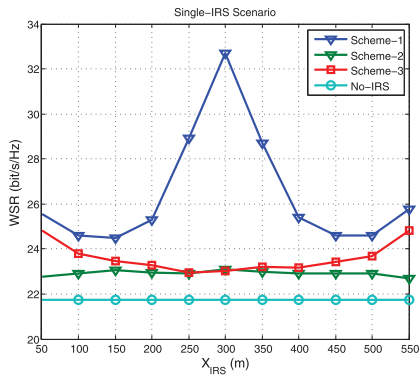
Fig. 13.   Achievable WSR versus various IRS deployment schemes for single-IRS case.
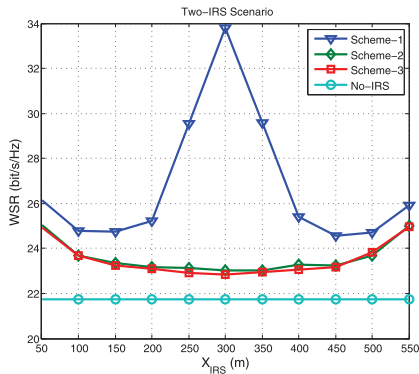

Fig. 14.   Achievable WSR versus various IRS deployment schemes for two-IRS case.

consistent with that of the two-cell scenario shown in Fig. 8. It is also shown that Scheme-1 has the best performance for any locations of the IRS. The reason may be that the IRS in Scheme-1 is more close to the first and second users. However, the users in the third and fourth cells are far away from the IRS in Scheme-1, thus the benefits of the IRS for these users are marginal. This motivates the deployment of more IRSs in the system. Again, the WSR achieved by the various schemes is higher than that without IRS, which demonstrates the benefits of installing IRSs in multicell networks.

2) Two-IRS Case: In this case, two IRSs are deployed, each of which has 25 phase shifts. Hence, the total number of phase shifters is equal to that of the single-IRS case. Three schemes are considered: 1) **Scheme-1**: IRS 1 moves from BS 1 to BS 2, and IRS moves from BS 3 to BS 4; 2) **Scheme-2**: IRS 1 moves from BS 1 to BS 4, and IRS 2 moves from BS 3 to BS 2; 3) **Scheme-3**: IRS 1 moves from point B to point D, and IRS 2 moves from point C to point A. The WSR achieved by various schemes is shown in Fig. 14. Similar trends have been observed to the single-IRS case. For example, Scheme-1 performs the best, and achieves its highest WSR when the IRS is located at points A and C, respectively. The reason is that the IRSs are closer to the users in these two points. By comparing Fig. 13 and Fig. 14, when $X_{\text{IRS}} = 300\ m$, the WSR of the Scheme-1 in the two-IRS scenario is higher than that in the single-IRS scenario, which means that the distributed IRS deployment is more beneficial than centralized

deployment. In general, the number of IRSs depends on the number of user clusters. It is expected that in the vicinity of each user cluster, there is at least one IRS.

VI. Conclusion

In this paper, we have enhanced the cell-edge user performance of multicell communication systems by employing an IRS at the cell boundary. Specifically, by carefully tuning the phase shifts, the inter-cell interference reflected by IRS can be added destructively to that directly received from the adjacent BSs, which alleviates the inter-cell interference received by the cell-edge users. We studied the WSR maximization problem by jointly optimizing the active TPC matrices at the BSs and passive shifts at the IRSs, while guaranteeing each BS's power constraint and unit-modulus constraint at the IRS. To tackle this non-convex problem, the BCD algorithm was used for optimizing them in an alternating manner. The optimal TPC matrices were obtained in closed form, and a pair of efficient algorithms were provided for solving the challenging phase shift optimization problem. Our simulation results verified that the proposed algorithms achieve significant performance gains over their conventional counterpart operating without incorporating an IRS. Furthermore, the location of IRS should be carefully chosen to guarantee a favourable BS-IRS link and IRS-user link. When the IRSs are deployed in the vicinity of user clusters, distributed IRS deployment is shown to be advantageous over the centralized deployment.

References

[1] W. Zhang *et al.*, "Large-scale antenna systems with UL/DL hardware mismatch: Achievable rates analysis and calibration," *IEEE Trans. Commun.*, vol. 63, no. 4, pp. 1216–1229, Apr. 2015.

[2] J. G. Andrews *et al.*, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, Jun. 2014.

[3] C. Pan, M. Elkashlan, J. Wang, J. Yuan, and L. Hanzo, "User-centric C-RAN architecture for ultra-dense 5G networks: Challenges and methodologies," *IEEE Commun. Mag.*, vol. 56, no. 6, pp. 14–20, Jun. 2018.

[4] C. Pan, H. Zhu, N. J. Gomes, and J. Wang, "Joint precoding and RRH selection for user-centric green MIMO C-RAN," *IEEE Trans. Wireless Commun.*, vol. 16, no. 5, pp. 2891–2906, May 2017.

[5] C. Pan, H. Ren, M. Elkashlan, A. Nallanathan, and L. Hanzo, "Robust beamforming design for ultra-dense user-centric C-RAN in the face of realistic pilot contamination and limited feedback," *IEEE Trans. Wireless Commun.*, vol. 18, no. 2, pp. 780–795, Feb. 2019.

[6] M. D. Renzo *et al.*, "Smart radio environments empowered by reconfigurable AI meta-surfaces: An idea whose time has come," *EURASIP J. Wireless Commun. Netw.*, vol. 2019, no. 1, p. 129, Dec. 2019.

[7] Q. Wu and R. Zhang, "Towards smart and reconfigurable environment: Intelligent reflecting surface aided wireless network," 2019, *arXiv:1905.00152*. [Online]. Available: http://arxiv.org/abs/1905.00152

[8] T. J. Cui, M. Q. Qi, X. Wan, J. Zhao, and Q. Cheng, "Coding metamaterials, digital metamaterials and programmable metamaterials," *Light, Sci. Appl.*, vol. 3, no. 10, p. e218, Oct. 2014.

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

[10] M. Di Renzo *et al.*, "Reconfigurable intelligent surfaces vs. Relaying: Differences, similarities, and performance comparison," 2019, *arXiv:1908.08747*. [Online]. Available: http://arxiv.org/abs/1908.08747

[11] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, Jr., "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, Mar. 2014.

[12] X. Yu, J.-C. Shen, J. Zhang, and K. B. Letaief, "Alternating minimization algorithms for hybrid precoding in millimeter wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 485–500, Apr. 2016.

[13] S. K. Mohammed and E. G. Larsson, "Constant-envelope multi-user precoding for frequency-selective massive MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 2, no. 5, pp. 547–550, Oct. 2013.

[14] S. K. Mohammed and E. G. Larsson, "Single-user beamforming in large-scale MISO systems with per-antenna constant-envelope constraints: The doughnut channel," *IEEE Trans. Wireless Commun.*, vol. 11, no. 11, pp. 3992–4005, Nov. 2012.

[15] Q. Wu and R. Zhang, "Intelligent reflecting surface enhanced wireless network: Joint active and passive beamforming design," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2018, pp. 1–6.

[16] X. Yu, D. Xu, and R. Schober, "MISO wireless communication systems via intelligent reflecting surfaces," 2019, *arXiv:1904.12199*. [Online]. Available: http://arxiv.org/abs/1904.12199

[17] Y. Yang, B. Zheng, S. Zhang, and R. Zhang, "Intelligent reflecting surface meets OFDM: Protocol design and rate maximization," 2019, *arXiv:1906.09956*. [Online]. Available: http://arxiv.org/abs/1906.09956

[18] Y. Han, W. Tang, S. Jin, C.-K. Wen, and X. Ma, "Large intelligent surface-assisted wireless communication exploiting statistical CSI," *IEEE Trans. Veh. Technol.*, vol. 68, no. 8, pp. 8238–8242, Aug. 2019.

[19] Q. Wu and R. Zhang, "Intelligent reflecting surface enhanced wireless network via joint active and passive beamforming," 2018, *arXiv:1810.03961*. [Online]. Available: http://arxiv.org/abs/1810.03961

[20] C. Huang, A. Zappone, G. C. Alexandropoulos, M. Debbah, and C. Yuen, "Reconfigurable intelligent surfaces for energy efficiency in wireless communication," *IEEE Trans. Wireless Commun.*, vol. 18, no. 8, pp. 4157–4170, Aug. 2019.

[21] H. Guo, Y.-C. Liang, J. Chen, and E. G. Larsson, "Weighted sum-rate optimization for intelligent reflecting surface enhanced wireless networks," 2019, *arXiv:1905.07920*. [Online]. Available: http://arxiv.org/abs/1905.07920

[22] Q.-U.-U. Nadeem, A. Kammoun, A. Chaaban, M. Debbah, and M.-S. Alouini, "Asymptotic max-min SINR analysis of reconfigurable intelligent surface assisted MISO systems," 2019, *arXiv:1903.08127*. [Online]. Available: http://arxiv.org/abs/1903.08127

[23] A. Taha, M. Alrabeiah, and A. Alkhateeb, "Enabling large intelligent surfaces with compressive sensing and deep learning," 2019, *arXiv:1904.10136*. [Online]. Available: http://arxiv.org/abs/1904.10136

[24] C. Pan *et al.*, "Intelligent reflecting surface aided MIMO broadcasting for simultaneous wireless information and power transfer," 2019, *arXiv:1908.04863*. [Online]. Available: http://arxiv.org/abs/1908.04863

[25] T. Bai, C. Pan, Y. Deng, M. Elkashlan, A. Nallanathan, and L. Hanzo, "Latency minimization for intelligent reflecting surface aided mobile edge computing," 2019, *arXiv:1910.07990*. [Online]. Available: http://arxiv.org/abs/1910.07990

[26] G. Zhou, C. Pan, H. Ren, K. Wang, and A. Nallanathan, "Intelligent reflecting surface aided multigroup multicast MISO communication systems," 2019, *arXiv:1909.04606*. [Online]. Available: http://arxiv.org/abs/1909.04606

[27] X. Yu, D. Xu, and R. Schober, "Enabling secure wireless communications via intelligent reflecting surfaces," 2019, *arXiv:1904.09573*. [Online]. Available: http://arxiv.org/abs/1904.09573

[28] M. Cui, G. Zhang, and R. Zhang, "Secure wireless communication via intelligent reflecting surface," *IEEE Wireless Commun. Lett.*, vol. 8, no. 5, pp. 1410–1414, Oct. 2019.

[29] H. Shen, W. Xu, S. Gong, Z. He, and C. Zhao, "Secrecy rate maximization for intelligent reflecting surface assisted multi-antenna communications," *IEEE Commun. Lett.*, vol. 23, no. 9, pp. 1488–1492, Sep. 2019.

[30] J. Chen, Y.-C. Liang, Y. Pei, and H. Guo, "Intelligent reflecting surface: A programmable wireless environment for physical layer security," 2019, *arXiv:1905.03689*. [Online]. Available: http://arxiv.org/abs/1905.03689

[31] S. Hong, C. Pan, H. Ren, K. Wang, and A. Nallanathan, "Artificial-Noise-Aided secure MIMO wireless communications via intelligent reflecting surface," 2020, *arXiv:2002.07063*. [Online]. Available: http://arxiv.org/abs/2002.07063

[32] D. Gesbert, S. Hanly, H. Huang, S. S. Shitz, O. Simeone, and W. Yu, "Multi-cell MIMO cooperative networks: A new look at interference," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 9, pp. 1380–1408, Dec. 2010.

[33] T. Huang, Y. Yu, and L. Yi, "Design of highly isolated compact antenna array for MIMO applications," *Int. J. Antennas Propag.*, vol. 2014, pp. 1–5, Nov. 2014.

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

[35] K. Alhujaili, V. Monga, and M. Rangaswamy, "Transmit MIMO radar beampattern design via optimization on the complex circle manifold," *IEEE Trans. Signal Process.*, vol. 67, no. 13, pp. 3561–3575, Jul. 2019.

[36] G. Zhou, C. Pan, H. Ren, K. Wang, and A. Nallanathan, "A framework of robust transmission design for IRS-aided MISO communications with imperfect cascaded channels," 2020, *arXiv:2001.07054*. [Online]. Available: http://arxiv.org/abs/2001.07054

[37] Q. Shi, M. Razaviyayn, Z.-Q. Luo, and C. He, "An iteratively weighted MMSE approach to distributed sum-utility maximization for a MIMO interfering broadcast channel," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4331–4340, Sep. 2011.

[38] M. Grant and S. Boyd. (Mar. 2014). *CVX: MATLAB Software for Disciplined Convex Programming, Version 2.1.* [Online]. Available: http://cvxr.com/cvx

[39] X.-D. Zhang, *Matrix Analysis and Applications*. Cambridge, U.K.: Cambridge Univ. Press, 2017.

[40] B. R. Marks and G. P. Wright, "A general inner approximation algorithm for nonconvex mathematical programs," *Oper. Res.*, vol. 26, no. 4, pp. 681–683, Aug. 1978.

[41] D. T. Ngo, S. Khakurel, and T. Le-Ngoc, "Joint subchannel assignment and power allocation for OFDMA femtocell networks," *IEEE Trans. Wireless Commun.*, vol. 13, no. 1, pp. 342–355, Jan. 2014.

[42] L. Venturino, N. Prasad, and X. Wang, "Coordinated scheduling and power allocation in downlink multicell OFDMA networks," *IEEE Trans. Veh. Technol.*, vol. 58, no. 6, pp. 2835–2848, Jul. 2009.

[43] A. Zappone and E. Jorswieck, "Energy efficiency in wireless networks via fractional programming theory," *Found. Trends Commun. Inf. Theory*, vol. 11, nos. 3–4, pp. 185–396, 2015.

[44] C. Pan, W. Xu, W. Zhang, J. Wang, H. Ren, and M. Chen, "Weighted sum energy efficiency maximization in ad hoc networks," *IEEE Wireless Commun. Lett.*, vol. 4, no. 3, pp. 233–236, Jun. 2015.

[45] J. Song, P. Babu, and D. P. Palomar, "Sequence design to minimize the weighted integrated and peak sidelobe levels," *IEEE Trans. Signal Process.*, vol. 64, no. 8, pp. 2051–2064, Apr. 2016.

[46] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

Cunhua Pan (Member, IEEE) received the B.S. and Ph.D. degrees from the School of Information Science and Engineering, Southeast University, Nanjing, China, in 2010 and 2015, respectively.

From 2015 to 2016, he was a Research Associate with the University of Kent, U.K. He held a Post-Doctoral Researcher position at the Queen Mary University of London, U.K., from 2016 and 2019, where he is currently a Lecturer (an Assistant Professor). His research interests mainly include intelligent reflection surface (IRS), machine learning, UAV, the Internet of Things, and mobile edge computing. He serves as a TPC member for numerous conferences, such as ICC and GLOBECOM, and the Student Travel Grant Chair for ICC 2019. He also serves as an Editor of the IEEE WIRELESS COMMUNICATIONS LETTERS and IEEE ACCESS.

Hong Ren (Member, IEEE) received the B.S. degree from Southwest Jiaotong University, Chengdu, China, in 2011, and the M.S. and Ph.D. degrees from Southeast University, Nanjing, China, in 2014 and 2018, respectively, all in electrical engineering. From October 2016 to January 2018, she was a Visiting Student with the School of Electronics and Computer Science, University of Southampton, U.K. She is currently a Post-Doctoral Scholar with the School of Electronic Engineering and Computer Science, Queen Mary University of London, U.K. Her research interests lie in the areas of communication and signal processing, including cooperative transmission, the Internet of Things, and ultrareliability and low-latency communications.

Kezhi Wang (Member, IEEE) received the B.E. and M.E. degrees from the School of Automation, Chongqing University, China, in 2008 and 2011, respectively, and the Ph.D. degree in engineering from the University of Warwick, U.K., in 2015. He was a Senior Research Officer with the University of Essex, U.K. He is currently a Senior Lecturer with the Department of Computer and Information Sciences, Northumbria University, U.K. His research interests include wireless communications and machine learning.

Wei Xu (Senior Member, IEEE) received the B.Sc. degree in electrical engineering and the M.S. and Ph.D. degrees in communication and information engineering from Southeast University, Nanjing, China in 2003, 2006, and 2009, respectively. From 2009 to 2010, he was a Post-Doctoral Research Fellow with the Department of Electrical and Computer Engineering, University of Victoria, Victoria, BC, Canada. He is currently a Professor with the National Mobile Communications Research Laboratory, Southeast University. He is also an Adjunct Professor with the University of Victoria, and a Distinguished Visiting Fellow of the Royal Academy of Engineering, U.K. He has coauthored more than 100 refereed journal articles in addition to 36 domestic patents and four U.S. patents granted. His research interests include cooperative communications, information theory, signal processing, and machine learning for wireless communications. He received the best paper awards from the IEEE MAPE 2013, the IEEE/CIC ICCC 2014, the IEEE Globecom 2014, the IEEE ICUWB 2016, WCSP 2017, and ISWCS 2018. He was a co-recipient of the First Prize of the Science and Technology Award in Jiangsu Province, China, in 2014. He received the Youth Science and Technology Award of China Institute of Communications in 2018. He was an Editor of the IEEE COMMUNICATIONS LETTERS from 2012 to 2017. He is also an Editor of the IEEE TRANSACTIONS ON COMMUNICATIONS and IEEE ACCESS.

Maged Elkashlan (Member, IEEE) received the Ph.D. degree in electrical engineering from The University of British Columbia, Vancouver, BC, Canada, in 2006.

From 2007 to 2011, he was with the Commonwealth Scientific and Industrial Research Organization (CSIRO), Australia. During this time, he held visiting appointments at the University of New South Wales and the University of Technology Sydney. In 2011, he joined the School of Electronic Engineering and Computer Science, Queen Mary University of London, U.K. His research interests fall into the broad areas of communication theory and statistical signal processing. He received the best paper awards at the IEEE International Conference on Communications (ICC) in 2016 and 2014, the International Conference on Communications and Networking in China (CHINACOM) in 2014, and the IEEE Vehicular Technology Conference (VTC-Spring) in 2013. He is currently serving as an Editor of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS and the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY.

Arumugam Nallanathan (Fellow, IEEE) was an Assistant Professor with the Department of Electrical and Computer Engineering, National University of Singapore, from August 2000 to December 2007. He was with the Department of Informatics, King's College London, from December 2007 to August 2017, where he was a Professor of wireless communications, from April 2013 to August 2017, and has been a Visiting Professor, since September 2017. He has also been a Professor of wireless communications and the Head of the Communication Systems Research (CSR) Group, School of Electronic Engineering and Computer Science, Queen Mary University of London, since September 2017. He has published nearly 500 technical articles in scientific journals and international conferences. His research interests include artificial intelligence for wireless systems, beyond 5G wireless networks, the Internet of Things (IoT), and molecular communications. He is a co-recipient of the best paper awards presented at the IEEE International Conference on Communications 2016 (ICC'2016), the IEEE Global Communications Conference 2017 (GLOBECOM'2017), and the IEEE Vehicular Technology Conference 2018 (VTC'2018). He is an IEEE Distinguished Lecturer. He has been selected as a Web of Science Highly Cited Researcher in 2016. He is an Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS. He was an Editor of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS from 2006 to 2011, the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY from 2006 to 2017, the IEEE WIRELESS COMMUNICATIONS LETTERS, and the IEEE SIGNAL PROCESSING LETTERS. He has served as the Chair for the Signal Processing and Communication Electronics Technical Committee of the IEEE Communications Society and the Technical Program Chair and a member of Technical Program Committees in numerous IEEE conferences. He received the IEEE Communications Society SPCE Outstanding Service Award 2012 and the IEEE Communications Society RCC Outstanding Service Award 2014.

Lajos Hanzo (Fellow, IEEE) received the master's and Ph.D. degrees from the Technical University (TU) of Budapest in 1976 and 1983, respectively, the D.Sc. degree from the University of Southampton in 2004, and the Honorary Doctorates from the TU of Budapest and The University of Edinburgh in 2009 and 2015, respectively. He has served several terms as the Governor of both IEEE ComSoc and VTS. He has published over 1900 contributions at IEEE Xplore, 19 Wiley-IEEE Press books, and has helped the fast-track career of 123 Ph.D. students. Over 40 of them are Professors at various stages of their careers in academia and many of them are leading scientists in the wireless industry. He is a fellow of the Royal Academy of Engineering (REng), and a fellow of the IET and EURASIP. He is also a Foreign Member of the Hungarian Academy of Sciences and a former Editor-in-Chief of the IEEE PRESS.