

Low-Complexity Antenna Selection and Discrete Phase-Shifts Design in IRS-Assisted Multiuser Massive MIMO Networks

Zaid Abdullah, *Member, IEEE*, Gaojie Chen, *Senior Member, IEEE*, Sangarapillai Lambotharan, *Senior Member, IEEE*, and Jonathon A. Chambers, *Fellow, IEEE*

Abstract—We propose two novel antenna selection (AS) and discrete phase-shifts design (PSD) schemes for use in intelligent reflecting surface (IRS) assisted multiuser massive multiple-input multiple-output (mMIMO) networks. The first AS and PSD method aims at maximizing the gain of the channels; while the second method is an iterative sum-rate maximization (ISM) scheme that aims at maximizing the total achievable rate. For the AS part, we demonstrate that the ISM method achieves near optimal performance with much lower complexity compared to benchmark AS schemes, and can be utilized with any precoder at the mMIMO base station. For the PSD, our proposed successive-refinement optimization methods are not only efficient, but their complexities scale linearly with the number of elements at the IRS, making them highly attractive when dealing with large surfaces. A thorough complexity analysis for the proposed methods is carried out in terms of the number of floating point operations required for their implementations. Finally, extensive numerical results are provided and some key points are highlighted on the performance of the proposed schemes with both conjugate beamforming and zero-forcing precoders.

Index Terms—Antenna Selection, Massive MIMO, Intelligent Reflecting Surface, Passive Beamforming, Discrete Phase-Shifts Design

I. INTRODUCTION

THE deployment of multi-antenna base-stations (BSs) has proven its efficiency in achieving high data rates, and as a consequence, the concept of massive multiple-input multiple-output (mMIMO) has become a reality for the fifth generation (5G) of wireless communication networks. Although mMIMO can offer significant advantages, it is well known that BSs with a large number of active antenna elements suffer from a considerably high power consumption, since each antenna is connected to a separate power-demanding radio frequency (RF) chain. In fact, RF chains are responsible

This work was supported by the Engineering and Physical Sciences Research Council under Grants EP/R006377/1 and EP/R006385/1. (*Corresponding authors: Zaid Abdullah and Gaojie Chen*)

Z. Abdullah is with the Interdisciplinary Centre for Security, Reliability and Trust (SnT), University of Luxembourg, Luxembourg. Email: zaid.abdullah@uni.lu.

G. Chen is with the Institute for Communication Systems (ICS), 5GIC & 6GIC, University of Surrey, Guildford, Surrey, GU2 7XH, U.K. Email: gaojie.chen@surrey.ac.uk.

J. A. Chambers is with the School of Engineering, University of Leicester, Leicester, LE1 7RU, U.K. Email: jonathon.chambers@leicester.ac.uk.

S. Lambotharan is with the Wolfson School of Mechanical, Electrical and Manufacturing Engineering, Loughborough University, Loughborough LE11 3TU, U.K. Email: s.lambotharan@lboro.ac.uk.

for 50% – 80% of the total transceiving power consumption of communications systems [1]. In addition, adding more RF chains would increase the hardware cost and complexity of the system. One way to overcome the aforementioned challenges while maintaining the advantages of mMIMO is via applying antenna selection (AS) techniques. However, unlike conventional MIMO systems where the AS is carried out with the main focus being on enhancing the performance; designing AS in mMIMO is much more challenging as the computational complexity can become the main bottleneck and must be taken into account when designing such schemes. In fact, all works on AS in mMIMO have the same motivation, that is reducing the hardware complexity, cost, and power consumption that comes with massive antenna arrays; while the quality of any AS algorithm in mMIMO can be assessed by answering the following question ‘*how good is the trade-off between the required computational complexity and the achieved performance?*’.

There has been a considerable amount of work on AS in mMIMO in recent years. For example, the authors in [2] performed AS on measured channels, and they showed that AS can significantly reduce the hardware complexity and power consumption without large degradation in the performance. In [3], a low-complexity AS scheme was proposed to maximize the constructive interference using a conjugate-beamforming (CB) precoder. In our previous work in [4], user-centric and semiblind interference rejection AS schemes were proposed to maximize the signal-to-interference plus noise ratios (SINRs) of multiuser mMIMO with CB, and a similar methodology was utilized in our work in [5] alongside optimal power control to maximize the rates of cell-edge users. In [6] the authors proposed a branch-and-bound scheme for AS to maximize the MIMO channel capacity which can be achieved via dirty paper coding. Moreover, the authors in [7] proposed greedy AS schemes inspired by matching pursuit techniques; while in [8], the authors proposed a self-supervised learning based Monte-Carlo tree search AS algorithm to maximize the channel capacity. Furthermore, evolutionary methods have also been utilized for AS in mMIMO to maximize the channel capacity, for example, bio-inspired, quantum-inspired, and heuristic algorithms were proposed in [9] and [10]. Finally, the authors in [11]–[13] investigated joint beamforming and AS in large scale MIMO networks.

Moving beyond mMIMO, however, is not an easy challenge, as on one hand, increasing the number of active antenna

terminals comes at a huge price in terms of hardware complexity and power consumption; while on the other hand, the demand for higher data rates continues to increase rapidly and more than ever before [14]. To that end, the recently proposed intelligent reflecting surface (IRS) has emerged as a possible solution to satisfy the growth in demand for data rates in beyond 5G networks while maintaining low cost and power requirements [15]–[18]. In principle, IRS is similar to full-duplex amplify-and-forward relaying, with the main difference being that IRS cannot provide active amplification as it does not contain power amplifiers, but rather reflecting the incident electromagnetic waves on its planar surface with certain phase shifts, such that the direct-path signal and the one reflected by the IRS are constructively combined at the receiver end. Thus, with sufficiently large surfaces, it has the potential to provide reliable, cost-effective and energy efficient communication links [19]. Therefore, in this work our aim is to design efficient low-complexity methods for AS and phase-shifts design (PSD) for IRS-mMIMO networks.

To ensure efficient utilization of IRSs in multi-antenna communications systems, passive beamforming, i.e. the phase-shifts of reflecting elements at the IRS, should be properly designed. Conventional methods can include semidefinite relaxation (SDR) followed by randomization to approximate near optimal solutions [20]–[23]. However, such methods can suffer from extremely high complexity when the number of reflecting elements is large. Moreover, performing AS for mMIMO in the presence of IRSs is more challenging, as each antenna is influenced by the phase-shifts of the reflecting elements at the IRS. Therefore, complexity considerations become even more vital when designing AS in such scenarios. Thus far, the only work on AS with PSD in multiuser IRS-mMIMO was reported in [24], where the authors designed their schemes to maximize the MIMO channel capacity, which can only be obtained via dirty paper coding. Moreover, the authors assumed continuous phase-shifts, meaning that each reflecting element at the IRS can have any phase-shift value between 0 and 2π . However, designing AS and passive beamforming with such assumptions might not necessarily lead to a high performance in practical scenarios. In particular, dirty paper coding is known to suffer from extremely high complexity, and thus is unsuitable for mMIMO systems, where linear precoders are applied such as the zero-forcing (ZF) and CB. Furthermore, in practical scenarios, the IRS is expected to be implemented with only a few quantization levels [25], and as a result, it is more convenient to take discrete phase-shifts into account when designing such algorithms. In [26], the authors studied the problem of AS in an IRS-assisted network for the case of a single-antenna user and continuous phase-shifts. However, performing AS and PSD for multiple users is much more challenging and requires an entirely different design approach.

Few works have addressed the practical PSD in IRS, for example, the authors in [27] and [28] studied the power minimization problem of an IRS-assisted multiuser network under discrete phase-shifts, where they proposed a sub-optimal ‘successive refinement’ scheme, such that the phase of each reflecting element is optimized while keeping the phase-shifts of other elements fixed. Furthermore, the authors in [29] and

[30] solved the discrete phase-shifts design via a mixed-integer linear programming method, which can be tackled through the branch and bound (BaB) algorithm. However, such an algorithm can suffer from exponential complexity, and the authors further proposed a lower-complexity design scheme that is similar to the successive refinement method. Despite the fact that the proposed works in [27]–[30] have lower complexities than conventional SDR or BaB schemes, they still suffer from a large number of vector/matrix multiplications to optimize the phase of each reflecting element; while in our proposed methods, our design aims at avoiding all vector/matrix multiplications during the optimization of all reflecting elements. It is worth to point out that the considered objective function for the discrete PSD in our work is different from those in [27]–[30], as our aim is to maximize either the gain of the channels or the achievable sum-rate, rather than minimizing the total transmit power.

In this paper, we propose two novel AS and discrete PSD (AaP) schemes with a unified framework, while maintaining low computational complexities. In particular, the proposed schemes are designed in a way such that the amount of vector/matrix multiplications is minimized, thereby resulting in a dramatic complexity reduction. Moreover, a key advantage of the proposed schemes is that they can be applied with any type of active precoding at the mMIMO BS. We test the performance of our proposed methods, under perfect and imperfect channel state information (CSI), with two types of linear precoders ZF and CB. Our main contributions in this paper are summarized as follows

- We propose two novel AaP schemes for use in multiuser mMIMO-IRS networks. In particular, the first method is the maximum channels’ gain (MCG)-AaP scheme, where both AS and PSD are carried out to maximize the ℓ_1 norms of the effective channel matrix.
- The second method, called iterative sum-rate maximization (ISM)-AaP, is designed to maximize the total sum rate of the network. More specifically, for the AS, we follow a decremental approach where at each iteration the antenna that contributes the least to the total sum rate is discarded. Then, the PSD stage takes place where the phase of each reflecting element is optimized to maximize the total sum rate using the remaining set of available antennas. Moreover, the proposed design criterion is flexible in the sense that the PSD stage can be carried out after discarding any number of undesired antennas, and thus both AS and PSD can be applied either separately or in an alternating fashion.
- A thorough complexity analysis is provided to show that our schemes scale well when the number of antennas and reflecting elements become too large. In particular, and for the PSD, the complexities of our schemes increase linearly with the number of reflecting elements at the IRS, making them highly desirable when dealing with large IRS arrays.
- Numerical results demonstrate that the proposed MCG-AS is efficient when applied with a ZF precoder; while the ISM-AS achieves near optimal performance and can

TABLE I: List of acronyms

Acronym	Definition
AS	Antenna selection
AWGN	Additive white Gaussian noise
BS	Base station
CB	Conjugate beamforming
CE	Channel estimation
CSI	Channel state information
FLOP	Floating-point operation
IRS	Intelligent reflecting surface
ISM	Iterative sum-rate maximization
LoS	Line-of-sight
MCG	Maximum channel gain
MIMO	Multiple-input multiple-output
mMIMO	massive MIMO
NLoS	Non-LoS
PSD	Phase-shift design
SINR	Signal-to-interference plus noise ratio
SNR	Signal-to-noise ratio
ZF	Zero-forcing
AaP	AS and PSD

be applied with any type of active precoding. Moreover, we show that when the number of reflecting elements is small, CB outperforms the ZF precoder especially when the number of selected antennas is not too large.

The rest of this paper is organized as follows. The adopted system model is given in Section II. The proposed AS and discrete PSD schemes are introduced and explained in detail in Section III. The complexity requirement of the proposed schemes is thoroughly investigated in Section IV. Different numerical results are presented and discussed in Section V. Finally, conclusions are drawn in Section VI.

The list of acronyms used throughout this work is shown in Table I.

Notations: Matrices and vectors are represented by uppercase and lowercase boldface letters, respectively. \mathbf{A}^T , \mathbf{A}^H and $\text{Tr}(\mathbf{A})$ are the transpose, Hermitian transpose, and the trace of matrix \mathbf{A} , respectively. $[\mathbf{A}]_{m,n}$ represents the n th element of the m th row of \mathbf{A} . $[\mathbf{A}]_{i,:}$ and \mathbf{a}_i both refer to the i th row of \mathbf{A} (unless \mathbf{a}_i was defined clearly in the manuscript as a column vector) and used interchangeably as appropriate, while $[\mathbf{A}]_{:,j}$ represents the j th column of \mathbf{A} . $[\mathbf{A}]_{:,S}$ is a submatrix of \mathbf{A} which contains the column vectors of \mathbf{A} that belong to set S . S_j represents the j th element of S , while \underline{S}_j is a subset that has all but the j th element of S , i.e. $\underline{S}_j = S \setminus S_j$. Furthermore, $[\mathbf{x}]_i$ is the i th element of \mathbf{x} , and $[\mathbf{x}]_S$ is a subspace vector that has all the elements of \mathbf{x} that belong to S . $|S|$ represents the cardinality of S . $\text{diag}\{\mathbf{a}\}$ is a diagonal matrix whose diagonal contains the elements of \mathbf{a} , while $\text{diag}\{\mathbf{A}\}$ is a vector whose elements are the diagonal of \mathbf{A} . In addition, \mathbf{I}_N is the $N \times N$ identity matrix. $\mathbb{E}\{\cdot\}$ and $\mathbb{V}\{\cdot\}$ are the expectation and variance operators, respectively. Finally, $\|\mathbf{a}\|_1$ is the ℓ_1 norm of $\mathbf{a} \in \mathbb{C}^{N \times 1}$, and can be defined as $\|\mathbf{a}\|_1 = \sum_{i=1}^N |[a]_i|$.

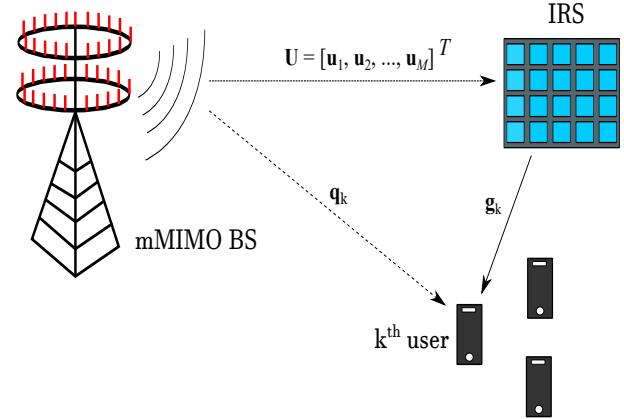


Fig. 1: The proposed multiuser mMIMO-IRS system model.

II. SYSTEM MODEL

We consider a single-cell downlink network operating in time division duplex. We assume a single BS equipped with N antennas is transmitting independent signals to K single-antenna users, with the help of an IRS with M reflecting elements, as shown in Fig. 1. The line-of-sight (LoS) paths between the BS and K users are assumed to be blocked, while the BS and all users have clear LoS links to the IRS. Let $\mathbf{Q} = [\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_K]^T \in \mathbb{C}^{K \times N}$ be the Rayleigh fading channel matrix between the BS and K users; $\mathbf{G} = [\mathbf{g}_1, \mathbf{g}_2, \dots, \mathbf{g}_K]^T \in \mathbb{C}^{K \times M}$ denotes the channel matrix between the IRS and K users and follows a Rician distribution with both LoS and non-LoS (NLoS) components; while $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_M]^T \in \mathbb{C}^{M \times N}$ denotes the Rician distributed channels between the BS and IRS. Similar to [6] and [22], we assume frequency-flat fading channels and perfect CSI is available at the BS.¹ Moreover, the reflecting elements at the IRS are controlled via the matrix $\Theta = \text{diag}\{[e^{j\theta_1}, e^{j\theta_2}, \dots, e^{j\theta_M}]\}$. Therefore, the effective channel matrix $\mathbf{H} \in \mathbb{C}^{K \times N}$ can be expressed as follows

$$\mathbf{H} = \mathbf{G}\Theta\mathbf{U} + \mathbf{Q} = \begin{bmatrix} \mathbf{g}_1\Theta\mathbf{U} + \mathbf{q}_1 \\ \mathbf{g}_2\Theta\mathbf{U} + \mathbf{q}_2 \\ \vdots \\ \mathbf{g}_K\Theta\mathbf{U} + \mathbf{q}_K \end{bmatrix} = \begin{bmatrix} \mathbf{h}_1^T \\ \mathbf{h}_2^T \\ \vdots \\ \mathbf{h}_K^T \end{bmatrix}. \quad (1)$$

We consider a practical scenario where each reflecting element can only have one of $L = 2^b$ quantization levels in α , where b is the number of quantization bits and α can be given as

$$\alpha = \left[1, e^{j2\pi/L}, e^{j4\pi/L}, \dots, e^{j2\pi(L-1)/L} \right]. \quad (2)$$

Assuming that all N antennas at the BS are activated, and by letting $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_K] \in \mathbb{C}^{N \times K}$ be any active precoding matrix (note that $\{\mathbf{w}_1, \dots, \mathbf{w}_K\}$ are column vectors), the transmitted signal from the BS is

$$\mathbf{s} = \sum_{k \in \mathcal{K}} \mathbf{w}_k x_k, \quad (3)$$

¹The performance of our proposed schemes against channel estimation errors will be evaluated in Section V.

where $\mathcal{K} = \{1, 2, \dots, K\}$ is the set of all users, and x_k is an information symbol with $\mathbb{E}\{|x_k|^2\} = 1$. Moreover, the precoding matrix must satisfy the constraint

$$\text{Tr}(\mathbf{W}\mathbf{W}^H) \leq P_t, \quad (4)$$

where P_t is the total transmit power. Then, and assuming that signals reflected from the IRS more than once have extremely small powers and thus can be neglected, the received signal at the k th user is

$$\begin{aligned} y_k &= \mathbf{h}_k^T \mathbf{s} + n_k \\ &= (\mathbf{g}_k \Theta \mathbf{U} + \mathbf{q}_k) \mathbf{w}_k x_k + \sum_{i \in \mathcal{K} \setminus k} (\mathbf{g}_k \Theta \mathbf{U} + \mathbf{q}_k) \mathbf{w}_i x_i + n_k, \end{aligned} \quad (5)$$

where $n_k \sim \mathcal{CN}(0, \sigma^2)$ represents the additive white Gaussian noise (AWGN). Therefore, the corresponding SINR for the k th user can be given as

$$\gamma_k = \frac{|\mathbf{g}_k \Theta \mathbf{U} + \mathbf{q}_k|^2 \mathbf{w}_k^2}{\sum_{i \in \mathcal{K} \setminus k} |\mathbf{g}_k \Theta \mathbf{U} + \mathbf{q}_k|^2 \mathbf{w}_i^2 + \sigma^2}. \quad (6)$$

For a system with K users, the spectral efficiency in bits/s/Hz can be defined as follows

$$\mathcal{R} = \sum_{k \in \mathcal{K}} \log_2(1 + \gamma_k). \quad (7)$$

It is clear that for any MIMO system, the performance is highly affected by the precoding scheme applied, and throughout this work, two different types of linear precoding are utilized, namely the ZF and CB, such that [31]

$$\mathbf{W} = \begin{cases} \delta_{\text{ZF}} \mathbf{H}^H (\mathbf{H} \mathbf{H}^H)^{-1}, & \text{for ZF} \\ \delta_{\text{CB}} \mathbf{H}^H, & \text{for CB} \end{cases} \quad (8)$$

where δ_{ZF} and δ_{CB} are scaling factors to ensure that the power constraint in (4) is met, and they can be given as [31]

$$\delta_{\text{ZF}} = \sqrt{P_t / \text{Tr}((\mathbf{H} \mathbf{H}^H)^{-1})}, \quad (9a)$$

$$\delta_{\text{CB}} = \sqrt{P_t / \text{Tr}(\mathbf{H} \mathbf{H}^H)}. \quad (9b)$$

It is worth to point out that for the ZF precoder, the inter-user interference, i.e. the term in the denominator of (6), is nulled. However, here we present a general formula for γ_k that holds for any precoding scheme.

Our main goal in this paper is to propose efficient methods to optimize Θ , and to select N_s out of the available N antennas at the BS. Note that for the considered multiuser mMIMO-IRS scenario, AS and discrete PSD are both in fact NP-hard problems. Therefore, our aim in this work is to provide sub-optimal, yet highly efficient AS and PSD schemes with low computational complexities.

III. PROPOSED ANTENNA SELECTION AND PHASE-SHIFT DESIGN METHODS

We consider two methods for the AS and PSD, the first one is the maximum channels' gain method, where the antennas and phase-shifts are selected/optimized to maximize the effective channels' gain. For the second method, we propose

an iterative AS and PSD to maximize the total sum-rate for any given active precoding matrix \mathbf{W} . Both methods aim at reducing the computational complexity by minimizing the number of vector/matrix multiplications as will be thoroughly explained in the following subsections.

A. Maximum channels gain AS and PSD (MCG-AaP) scheme

Here we present the ultra-low-complexity MCG-AaP method, and introduce our novel PSD scheme. We next formulate the corresponding optimization problem.

1) *Problem formulation:* for the MCG-AaP scheme, both the AS and PSD are carried out to maximize the effective gains of the channels in terms of the ℓ_1 norm, this optimization problem can be formulated as follows

$$\underset{\Delta, \Theta}{\text{maximize}} \quad \|(\mathbf{G}\Theta\mathbf{U} + \mathbf{Q})\Delta\|_1 \quad (10)$$

subject to

$$\sum_{n=1}^N [\Delta]_{n,n} = N_s, \quad (10a)$$

$$[\Delta]_{n,n} \in \{0, 1\}, \forall n \in \mathcal{N}, \quad (10b)$$

$$[\Theta]_{m,m} \in \{[\alpha]_1, [\alpha]_2, \dots, [\alpha]_L\}, \quad \forall m \in \mathcal{M}, \quad (10c)$$

where $\mathcal{N} = \{1, 2, \dots, N\}$, $\mathcal{M} = \{1, 2, \dots, M\}$, and Δ is a binary diagonal matrix that controls the selected antennas, such that

$$[\Delta]_{n,n} = \begin{cases} 1, & \text{if the } n\text{th antenna is selected} \\ 0, & \text{otherwise.} \end{cases} \quad (11)$$

It is clear that the optimization variables are coupled in this case, which makes the optimization difficult. Accordingly, we divide this problem into two sub-problems. At first, we select the antennas with maximum channels' norms under random phase-shifts. Then, we propose a novel low-complexity PSD scheme to achieve further channel gain enhancement for the selected antennas.

2) *Maximum channels' gain AS (MCG-AS):* we start by recasting the formulation of effective channel matrix as follows

$$\mathbf{H} = \begin{bmatrix} \mathbf{h}_1^T \\ \mathbf{h}_2^T \\ \vdots \\ \mathbf{h}_K^T \end{bmatrix} = \begin{bmatrix} \mathbf{g}_1 \Theta \mathbf{U} + \mathbf{q}_1 \\ \mathbf{g}_2 \Theta \mathbf{U} + \mathbf{q}_2 \\ \vdots \\ \mathbf{g}_K \Theta \mathbf{U} + \mathbf{q}_K \end{bmatrix} = \begin{bmatrix} \phi^T \Xi_1 + \mathbf{q}_1 \\ \phi^T \Xi_2 + \mathbf{q}_2 \\ \vdots \\ \phi^T \Xi_K + \mathbf{q}_K \end{bmatrix}, \quad (12)$$

where $\phi = \text{diag}\{\Theta\}$, and $\Xi_k = \text{diag}\{\mathbf{g}_k\}\mathbf{U} \in \mathbb{C}^{M \times N}$ is the cascaded channel, which can be represented in a matrix form as follows

$$\begin{aligned} \Xi_k &= \text{diag}\{\mathbf{g}_k\}\mathbf{U} \\ &= \begin{bmatrix} [\mathbf{G}]_{k,1} [\mathbf{U}]_{1,1} & [\mathbf{G}]_{k,1} [\mathbf{U}]_{1,2} & \dots & [\mathbf{G}]_{k,1} [\mathbf{U}]_{1,N} \\ [\mathbf{G}]_{k,2} [\mathbf{U}]_{2,1} & [\mathbf{G}]_{k,2} [\mathbf{U}]_{2,2} & \dots & [\mathbf{G}]_{k,2} [\mathbf{U}]_{2,N} \\ \vdots & & \ddots & \vdots \\ [\mathbf{G}]_{k,M} [\mathbf{U}]_{M,1} & [\mathbf{G}]_{k,M} [\mathbf{U}]_{M,2} & \dots & [\mathbf{G}]_{k,M} [\mathbf{U}]_{M,N} \end{bmatrix}. \end{aligned} \quad (13)$$

Note that the evaluation of Ξ_k is crucial to reduce the complexity during the PSD that follows the AS stage as will

be seen later.² Therefore, evaluating the channel vectors \mathbf{h}_k in terms of Ξ_k at this stage will result in avoiding any additional computations during the PSD stage.

After evaluating $\Xi_k (\forall k \in \mathcal{K})$, we generate a random phase-shift vector ϕ and obtain \mathbf{H} in (12). Then, the set of selected antennas, denoted as \mathcal{S} , with the maximum channels' gains can now be given as follows

$$\mathcal{S} = \arg \max_{N_s} \left\{ \left[\sum_{k \in \mathcal{K}} |[\mathbf{H}]_{k,1}|, \sum_{k \in \mathcal{K}} |[\mathbf{H}]_{k,2}|, \dots, \sum_{k \in \mathcal{K}} |[\mathbf{H}]_{k,N}| \right] \right\}, \quad (14)$$

where \max indicates that the N_s highest values of the argument will be identified. Next, we shift our attention to the optimization of ϕ .

3) *Maximum channels' gain PSD (MCG-PSD):* for the phase-shift design, we adopt a successive refinement method, such that we optimize the phase of one reflecting element while keeping the phases of the remaining $(M-1)$ elements fixed. This optimization problem can be expressed as follows

$$\underset{[\phi]_m}{\text{maximize}} \quad \sum_{k \in \mathcal{K}} \left\| \phi^T [\Xi_k]_{:, \mathcal{S}} + [\mathbf{q}_k]_{\mathcal{S}} \right\|_1 \quad (15)$$

subject to

$$[\phi]_m \in \{[\alpha]_1, [\alpha]_2, \dots, [\alpha]_L\}, \quad (15a)$$

where maximize means that the optimization will be carried out only on the j th entry of α , while the remaining elements are fixed. Note that only the columns/entries which correspond to \mathcal{S} in Ξ_k and \mathbf{q}_k are involved in the above optimization problem, and this is due to the fact that the phase-optimization is carried out only for the selected antennas.

Optimizing each of the M phase-shifts according to (15) is straight forward; we simply need to evaluate the objective function L times and chose the phase-shift value that leads to the maximum channel gain. However, this method is not efficient in terms of the required computational complexity, and our aim is to reduce the complexity of (15) without any degradation in the performance.

To achieve that, let us have a closer look at the k th row of \mathbf{H} , we can observe that

$$\begin{aligned} \mathbf{h}_k^T &= \phi^T \Xi_k + \mathbf{q}_k \\ &= \left[\sum_{m \in \mathcal{M}} [\phi]_m [\Xi_k]_{m,1} + [\mathbf{q}_k]_1, \sum_{m \in \mathcal{M}} [\phi]_m [\Xi_k]_{m,2} + [\mathbf{q}_k]_2 \right. \\ &\quad \left. \dots, \sum_{m \in \mathcal{M}} [\phi]_m [\Xi_k]_{m,N} + [\mathbf{q}_k]_N \right], \end{aligned} \quad (16)$$

it is clear that to optimize the phase of each reflecting element, we need to take into account all available antennas and users. Accordingly, we search over the L phase quantization levels in α to identify the optimal phase value. Considering the m th reflecting element, this can now be performed as follows

$$l_m = \arg \max_{l \in \mathcal{L}} \sum_{k \in \mathcal{K}} \left(\sum_{j \in \mathcal{S}} \left| [\mathbf{h}_k]_j + ([\alpha]_l - [\phi]_m) [\Xi_k]_{m,j} \right| \right), \quad (17)$$

²It was shown in [32] that the cascaded channel Ξ_k can be estimated at the BS with satisfactory accuracy, even when entirely passive IRSs are deployed. However, for semi-passive surfaces where one or few elements of the IRS are connected to the baseband via RF chains, the CSI for each link of the cascaded channels can be estimated at the IRS itself [33], and such implementation can reduce the required overhead for CSI acquisition.

Algorithm 1 The proposed MCG-AaP scheme

- 1: **input** $N_s, M, \alpha, \Xi_k (\forall k), \mathbf{Q}$
- 2: **generate** ϕ (any random phase-shift vector),
- 3: **evaluate** \mathbf{H} accord. to (12),
- 4: **identify** the set \mathcal{S} with the N_s highest ℓ_1 norms in \mathbf{H} accord. to (14),
- 5: **for** $m = 1 : M$
- 6: **find** l_m according to (17),
- 7: **update** \mathbf{H} accord. to (18), and set $[\phi^*]_m = [\alpha]_{l_m}$,
- 8: **end for**
- 9: **output** $[\mathbf{H}]_{:, \mathcal{S}}, \phi^*$

where $\mathcal{L} = \{1, 2, \dots, L\}$. Note that with (17), we can perform the search/optimization over the L quantization levels for all M reflecting elements at the IRS using only element-wise operations, and without any compromise in the performance compared to (15). Moreover, after optimizing the phase of each reflecting element, we update \mathbf{H} as follows

$$[\mathbf{H}]_{k,j} := [\mathbf{H}]_{k,j} + ([\alpha]_{l_m} - [\phi]_m) [\Xi_k]_{m,j}, \quad \forall k \in \mathcal{K}, \forall j \in \mathcal{S}, \quad (18)$$

and the optimal phase value for the m th reflecting element can now be obtained as $[\phi^*]_m = [\alpha]_{l_m}$. The steps of the MCG-AaP scheme are given in Algorithm 1.

B. Iterative sum-rate maximization AS and PSD (ISM-AaP) scheme

In general, selecting all antennas at once can often lead to a relatively poor performance. In contrast, iterative schemes can be much more efficient in terms of performance when designed properly, but they require higher complexities compared to non-iterative schemes. In this section, we propose an iterative AS and PSD scheme to maximize the total sum-rate of the network. In the following, we start by formulating the optimization problem and then thoroughly explain the proposed methods.

1) *Problem formulation:* analytically, and for a given pre-coding matrix \mathbf{W} , the optimization problem for maximizing the achievable sum rate via AS and PSD can be expressed as follows

$$\underset{\Delta, \phi}{\text{maximize}} \quad \sum_{k \in \mathcal{K}} \log_2 \left(1 + \frac{\left| (\phi^T \Xi_k + \mathbf{q}_k) \Delta \mathbf{w}_k \right|^2}{\sum_{i \in \mathcal{K} \setminus \{k\}} \left| (\phi^T \Xi_k + \mathbf{q}_k) \Delta \mathbf{w}_i \right|^2 + \sigma^2} \right) \quad (19)$$

subject to

$$[\phi]_m \in \{[\alpha]_1, [\alpha]_2, \dots, [\alpha]_L\}, \quad \forall m \in \mathcal{M}, \quad (19a)$$

$$\text{and constraints (10a) and (10b).} \quad (19b)$$

It is clear that the above optimization problem is more complex than that for the MCG case. In particular, both AS and discrete PSD are in general NP-hard problems. Moreover, any selection for the antennas is influenced by the phase shift vector ϕ , and therefore, one should take the phase-shifts into account when designing such an algorithm. Next we introduce our proposed iterative schemes for the joint AS and PSD.

2) *Iterative sum-rate maximization AS (ISM-AS):* the main idea behind iterative schemes is to select/discard one antenna

$$\zeta = \arg \max_{n \in \mathcal{S}} \sum_{k \in \mathcal{K}} \log_2 \left(1 + \frac{\left| \left(\phi^T [\boldsymbol{\Xi}_k]_{:, \mathcal{S}_n} + [\mathbf{q}_k]_{\mathcal{S}_n} \right) [\mathbf{w}_k]_{\mathcal{S}_n} \right|^2}{\sum_{i \in \mathcal{K} \setminus k} \left| \left(\phi^T [\boldsymbol{\Xi}_k]_{:, \mathcal{S}_n} + [\mathbf{q}_k]_{\mathcal{S}_n} \right) [\mathbf{w}_i]_{\mathcal{S}_n} \right|^2 + \sigma^2} \right), \quad (20)$$

$$\beta = \arg \max_{n \in \{1, \dots, |\mathcal{S}|\}} \sum_{k \in \mathcal{K}} \log_2 \left(1 + \frac{\left| [\boldsymbol{\Pi}]_{k,k} - \left([\mathbf{H}]_{k,n} [\mathbf{W}]_{n,k} \right) \right|^2}{\sum_{i \in \mathcal{K} \setminus k} \left| [\boldsymbol{\Pi}]_{k,i} - \left([\mathbf{H}]_{k,n} [\mathbf{W}]_{n,i} \right) \right|^2 + \sigma^2} \right). \quad (21)$$

at each iteration to maximize/minimize a given cost function. We follow a decremental approach, such that at each iteration we eliminate the antenna which contributes the least to the total sum-rate. Mathematically, and for a given \mathbf{W} and ϕ , this can be represented as shown in (20), where \mathcal{S} is the set of available antennas at any given iteration. To be more specific, we have $|\mathcal{S}| = N$ before eliminating any antenna (i.e. before the algorithm starts), while $|\mathcal{S}| = N_s$ after discarding the least desirable $(N - N_s)$ antennas. Moreover, ζ is the index of the discarded antenna at the current iteration. Note that the selection in (20) requires performing a very large number of vector and matrix multiplications, resulting in a considerably large computational complexity for a system with large number of antennas.

Accordingly, we aim at reducing the complexity without compromising the performance by minimizing the required amount of vector multiplications while selecting the antennas. Specifically, we generate a random phase-shift vector ϕ and obtain initial channel matrix \mathbf{H} according to (12).³ Then, we evaluate an initial precoding matrix \mathbf{W} according to (8). Moreover, by defining $\boldsymbol{\Pi} = [\boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \dots, \boldsymbol{\pi}_K]^T \in \mathbb{C}^{K \times K}$, such that

$$\begin{aligned} \boldsymbol{\pi}_k &= \mathbf{h}_k^T \mathbf{W} \\ &= \left[\sum_{n=1}^{|\mathcal{S}|} [\mathbf{h}_k]_n [\mathbf{w}_1]_n, \sum_{n=1}^{|\mathcal{S}|} [\mathbf{h}_k]_n [\mathbf{w}_2]_n, \dots, \sum_{n=1}^{|\mathcal{S}|} [\mathbf{h}_k]_n [\mathbf{w}_K]_n \right], \end{aligned} \quad (22)$$

where \mathcal{S} is used at this stage only for convenience to express a general case regardless of which iteration the algorithm is at, since discarding each antenna results in removing one column of \mathbf{H} and one row of \mathbf{W} as will be explained later. It follows that the SINR for the k th user in (6) can now equivalently be expressed as

$$\gamma_k = \frac{\left| [\boldsymbol{\Pi}]_{k,k} \right|^2}{\sum_{i \in \mathcal{K} \setminus k} \left| [\boldsymbol{\Pi}]_{k,i} \right|^2 + \sigma^2}. \quad (23)$$

Then, the least desirable antenna, denoted as β , can now be identified/discarded as shown in (21).⁴ Note that the search over all available antennas to discard the least favourite one is now carried out only through element-wise operations, which

³To ensure effective convergence of the proposed ISM-AS, and before we obtain initial precoding matrix \mathbf{W} , we maximize the channels' norms of the initial \mathbf{H} by optimizing the initial phase-shifts using (17), and \mathbf{H} is then updated after optimizing each reflecting element according to (18). It should be noted that at this stage we have $\mathcal{S} = \{1, \dots, N\}$, and this initial optimization is carried out only once before the iterative algorithm starts.

⁴Note that even when ZF precoding is applied, the interuser interference in (21), expressed as $\sum_{i \in \mathcal{K} \setminus k} \left| [\boldsymbol{\Pi}]_{k,i} - \left([\mathbf{H}]_{k,n} [\mathbf{W}]_{n,i} \right) \right|^2$, will not be zero when discarding any antenna, and our aim is to discard the antenna that leads to the minimum loss in the total sum rate.

results in a tremendous complexity reduction compared to (20) without any compromise in the performance. In other words, both (20) and (21) lead to the exact same solution, and the latter approach requires much lower complexity than the former one. After eliminating each antenna, we update $\boldsymbol{\Pi}, \boldsymbol{\Xi}_k, \mathbf{W}, \mathbf{H}$, and \mathcal{S} as follows

$$[\boldsymbol{\Pi}]_{i,j} := [\boldsymbol{\Pi}]_{i,j} - [\mathbf{H}]_{i,\beta} [\mathbf{W}]_{\beta,j}, \quad \forall i, j \in \mathcal{K}, \quad (24a)$$

$$\boldsymbol{\Xi}_k := \boldsymbol{\Xi}_k \setminus [\boldsymbol{\Xi}_k]_{:, \beta}, \quad \forall k \in \mathcal{K}, \quad (24b)$$

$$\mathbf{W} := \mathbf{W} \setminus [\mathbf{W}]_{\beta,:}, \quad (24c)$$

$$\mathbf{H} := \mathbf{H} \setminus [\mathbf{H}]_{:, \beta}, \quad (24d)$$

$$\mathcal{S} := \mathcal{S} \setminus \beta. \quad (24e)$$

It is noteworthy that in case the AS, PSD, and active precoding are carried out separately, then $\boldsymbol{\Pi}$ has to be evaluated only once. In contrast, for alternating optimization where the phase-shifts and active precoding are updated after discarding each v antennas ($v \geq 1$), then $\boldsymbol{\Pi}$ has to be re-evaluated every time \mathbf{W} is updated. However, when it comes to the effective channel matrix \mathbf{H} , it undergoes two different types of reformations. In particular, after eliminating each antenna, one column will be removed from \mathbf{H} according to (24d); while the same matrix will be updated according to (18) after optimizing the phase of each reflecting element as will be highlighted again in the following subsection.

3) *Iterative sum-rate maximization PSD (ISM-PSD):* similar to the MCG case, we adopt a successive refinement method such that we optimize the phase of only one reflecting element at a time. However, the objective function considered in this case is maximizing the total sum-rate. This optimization problem can be expressed as follows

$$\underset{[\phi]_m}{\text{maximize}} \sum_{k \in \mathcal{K}} \log_2 \left(1 + \frac{\left| (\phi^T \boldsymbol{\Xi}_k + [\mathbf{q}_k]_{\mathcal{S}}) \mathbf{w}_k \right|^2}{\sum_{i \in \mathcal{K} \setminus k} \left| (\phi^T \boldsymbol{\Xi}_k + [\mathbf{q}_k]_{\mathcal{S}}) \mathbf{w}_i \right|^2 + \sigma^2} \right) \quad (25)$$

subject to

$$[\phi]_m \in \{[\alpha]_1, [\alpha]_2, \dots, [\alpha]_L\}. \quad (25a)$$

The above optimization problem can be solved by evaluating the objective function L times, and then set $[\phi^*]_m$ as the phase-shift value in α that achieved the highest sum-rate.⁵ However, this method can suffer from high computational complexity as will be seen later.

Accordingly, we aim at reducing the number of vector

⁵Note that since \mathbf{q}_k was not directly involved in the AS design stage, \mathcal{S} is utilized when dealing with these channels as can be seen in (25); while this is not the case for $\boldsymbol{\Xi}_k$ and \mathbf{W} which are updated according to (24b) and (24c), respectively, after discarding each antenna.

$$\mathbf{Z}_k = \boldsymbol{\Xi}_k \mathbf{W} = \begin{bmatrix} \sum_{n=1}^{|S|} [\boldsymbol{\Xi}_k]_{1,n} [\mathbf{w}_1]_n & \sum_{n=1}^{|S|} [\boldsymbol{\Xi}_k]_{1,n} [\mathbf{w}_2]_n & \dots & \sum_{n=1}^{|S|} [\boldsymbol{\Xi}_k]_{1,n} [\mathbf{w}_K]_n \\ \sum_{n=1}^{|S|} [\boldsymbol{\Xi}_k]_{2,n} [\mathbf{w}_1]_n & \sum_{n=1}^{|S|} [\boldsymbol{\Xi}_k]_{2,n} [\mathbf{w}_2]_n & \dots & \sum_{n=1}^{|S|} [\boldsymbol{\Xi}_k]_{2,n} [\mathbf{w}_K]_n \\ \vdots & \vdots & & \vdots \\ \sum_{n=1}^{|S|} [\boldsymbol{\Xi}_k]_{M,n} [\mathbf{w}_1]_n & \sum_{n=1}^{|S|} [\boldsymbol{\Xi}_k]_{M,n} [\mathbf{w}_2]_n & \dots & \sum_{n=1}^{|S|} [\boldsymbol{\Xi}_k]_{M,n} [\mathbf{w}_K]_n \end{bmatrix}, \quad (28)$$

$$l_m = \arg \max_{l \in \mathcal{L}} \sum_{k \in \mathcal{K}} \log_2 \left(1 + \frac{\left| [\boldsymbol{\Pi}]_{k,k} + ([\boldsymbol{\alpha}]_l - [\boldsymbol{\phi}]_m) [\mathbf{Z}_k]_{m,k} \right|^2}{\sum_{i \in \mathcal{K} \setminus k} \left| [\boldsymbol{\Pi}]_{k,i} + ([\boldsymbol{\alpha}]_l - [\boldsymbol{\phi}]_m) [\mathbf{Z}_k]_{m,i} \right|^2 + \sigma^2} \right), \quad (29)$$

and matrix multiplications involved in the optimization to its minimum. However, the current formulations of the SINRs in (23) and/or the objective function of (25) do not help our cause in reducing the complexity. In particular, we need to look at the SINRs from a different angle where we can represent the signal and interference terms for each user as a function of the M reflecting elements. Therefore, and recalling that $\mathbf{h}_k^T = \boldsymbol{\phi}^T \boldsymbol{\Xi}_k + [\mathbf{q}_k]_S$, we recast $\boldsymbol{\pi}_k$ in (22) as follows

$$\begin{aligned} \boldsymbol{\pi}_k &= (\boldsymbol{\phi}^T \boldsymbol{\Xi}_k + [\mathbf{q}_k]_S) \mathbf{W} = \boldsymbol{\phi}^T \mathbf{Z}_k + \mathbf{c}_k \\ &= \left[\sum_{m \in \mathcal{M}} [\boldsymbol{\phi}]_m [\mathbf{Z}_k]_{m,1} + [\mathbf{c}_k]_1, \sum_{m \in \mathcal{M}} [\boldsymbol{\phi}]_m [\mathbf{Z}_k]_{m,2} + [\mathbf{c}_k]_2 \right. \\ &\quad \dots, \left. \sum_{m \in \mathcal{M}} [\boldsymbol{\phi}]_m [\mathbf{Z}_k]_{m,K} + [\mathbf{c}_k]_K \right]. \end{aligned} \quad (26)$$

where $\mathbf{c}_k = [\mathbf{q}_k]_S \mathbf{W} \in \mathbb{C}^{1 \times K}$ contains the signal and interference values for the k th user through the direct link with the BS; while $\mathbf{Z}_k = [\boldsymbol{\Xi}_k \mathbf{w}_1, \boldsymbol{\Xi}_k \mathbf{w}_2, \dots, \boldsymbol{\Xi}_k \mathbf{w}_K] \in \mathbb{C}^{M \times K}$, which can be expressed in a matrix form as shown in (28), contains the signal and interference powers for the k th user via the IRS path before being multiplied or affected by $\boldsymbol{\phi}$. Note that we only need to evaluate $\{\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_K\}$ once for the optimization of all M reflecting elements. In particular, and considering the m th reflecting element, it is not difficult to see that the optimal phase can now be obtained as shown in (29), which does not involve any vector or matrix multiplications, resulting in a significant complexity reduction compared to the conventional successive refinement scheme given in (25), and without any compromise in the performance.

After optimizing the phase of each reflecting element, \mathbf{H} should be updated according to (18), while $\boldsymbol{\Pi}$ can be updated as follows

$$[\boldsymbol{\Pi}]_{i,j} := [\boldsymbol{\Pi}]_{i,j} + ([\boldsymbol{\alpha}]_{l_m} - [\boldsymbol{\phi}]_m) [\mathbf{Z}_i]_{m,j}, \quad \forall i, j \in \mathcal{K}, \quad (27)$$

and the optimal phase for the m th reflecting element can be obtained as $[\boldsymbol{\phi}^*]_m = [\boldsymbol{\alpha}]_{l_m}$. It is worth to highlight that $\{\mathbf{c}_1, \dots, \mathbf{c}_K\}$ in (26) are implicitly included in $\boldsymbol{\Pi}$, and we do not need to evaluate them since we only need $\{\mathbf{Z}_1, \dots, \mathbf{Z}_K\}$ to update $\boldsymbol{\Pi}$ after optimizing the phase of each reflecting element.

4) *The overall algorithm:* we initialize a random phase-shift vector and obtain an initial channel matrix. Then, we carry out an initial phase-shift optimization on $\boldsymbol{\phi}$ to maximize the channel norms of \mathbf{H} before evaluating \mathbf{W} . After that, $T = (N - N_s)/v$ iterations of AS and PSD are carried out for sum-rate maximization. In particular, and for each of the T iterations, we eliminate the least desirable v antennas according to (21), and then the M phase-shifts are optimized

Algorithm 2 The proposed ISM-AaP scheme

```

1: input  $v, M, N, N_s, \boldsymbol{\alpha}, \boldsymbol{\Xi}_k (\forall k), \boldsymbol{\phi}$  (any feasible solution),
    $\mathbf{H}$  (initial channel matrix),  $\mathcal{S} = \{1, 2, \dots, N\}$  (set of
   available antennas)
2: for  $m = 1 : M$  (initial phase-shifts optimization)
3:   find  $l_m$  according to (17),
4:   update  $\mathbf{H}$  accord. to (18), and set  $[\boldsymbol{\phi}]_m = [\boldsymbol{\alpha}]_{l_m}$ ,
5: end for
6: evaluate  $\mathbf{W}$  (initial precoding matrix)
7: for  $t = 1 : (N - N_s)/v$  (iterative AS and PSD)
8:   evaluate  $\boldsymbol{\Pi}$ ,
9:   for  $i = 1 : v$ 
10:    find  $\beta$  according to (21),
11:    update  $\boldsymbol{\Pi}, \boldsymbol{\Xi}_k, \mathbf{W}, \mathbf{H}, \mathcal{S}$  according to (24a)-(24e),
12:   end for
13:   evaluate  $\mathbf{Z}_k (\forall k \in \mathcal{K})$ ,
14:   for  $m = 1 : M$ 
15:     find  $l_m$  according to (29),
16:     update  $\boldsymbol{\Pi}$  accord. to (27),  $\mathbf{H}$  accord. to (18), and
        set  $[\boldsymbol{\phi}^*]_m = [\boldsymbol{\alpha}]_{l_m}$ ,
17:   end for
18:   evaluate  $\mathbf{W}$  accord. to (8) using the updated  $\mathbf{H}$ ,
19: end for

```

according to (29) for further performance enhancement. Note that when $v = 1$, $\boldsymbol{\phi}$ will be optimized after each antenna is eliminated; while if $v = N - N_s$, that means AS and PSD are carried out separately. Moreover, the precoding matrix \mathbf{W} is re-evaluated after each time $\boldsymbol{\phi}$ is optimized. The steps of the proposed ISM-AaP scheme are given in Algorithm 2.

IV. COMPLEXITY ANALYSIS FOR THE PROPOSED SCHEMES

In this section we present the complexity analysis of the proposed AS and PSD schemes in terms of number of floating point operations (FLOPs). In particular, we follow the analysis in [34], such that each operation (addition, multiplication, subtraction, or division) between two real numbers is equivalent to one FLOP. We also assume that for any $x \in \mathbb{R}$, finding \sqrt{x} and $\log_2(x)$ are each equivalent to 1 FLOP.

A. Complexity analysis of the MCG-AaP scheme

1) *Antenna selection:* evaluating $\boldsymbol{\Xi}_k (\forall k \in \mathcal{K})$ according to (13) requires MNK complex multiplications, which cor-

responds to $6MNK$ FLOPs.⁶ Then, obtaining each row in \mathbf{H} according to (12) requires NM complex multiplications and NM complex additions. As a result, obtaining \mathbf{H} takes $8MNK$ FLOPs. Finding the ℓ_1 norm for all columns in \mathbf{H} requires $N(5K - 1)$ FLOPs. Finally, sorting the N channel gain values before the selection requires $N \log_{10} N$ FLOPs. Therefore, the total complexity of the MCG-AS scheme is

$$\mathcal{C}_{AS}^{MCG} = NK(14M + 5) + N(\log_{10} N - 1), \quad (30)$$

and the complexity order of the MCG-AS can thus be given as $\mathcal{O}(NMK + N \log_{10} N)$.

2) *Phase-shifts design*: finding l_m for each reflecting element according to (17) requires $15N_sKL$ FLOPs. Then, updating \mathbf{H} according to (18) after optimizing each reflecting element requires $10KN_s$ FLOPs. Therefore, the total number of FLOPs required for optimizing all M reflecting elements with the necessary update of channel matrix is

$$\mathcal{C}_{PSD}^{MCG} = MN_s K(15L + 10), \quad (31)$$

and the corresponding complexity order can be simplified to $\mathcal{O}(MN_sKL)$. In contrast, optimizing the phase-shifts via the traditional method in (15) has a complexity order of $\mathcal{O}(M^2N_sKL)$. See Appendix A for details.

3) *Total complexity*: since this approach is non-iterative and does not require active precoding, it follows that the total complexity of the MCG-AaP can be given as follows

$$\mathcal{C}_{total}^{MCG} = \mathcal{C}_{AS}^{MCG} + \mathcal{C}_{PSD}^{MCG}. \quad (32)$$

B. Complexity analysis of the ISM-AaP scheme

First, we need to include the complexity of the initial phase-optimization to maximize the channel norms to ensure good convergence behaviour. Obtaining $\Xi_k (\forall k \in \mathcal{K})$ and \mathbf{H} (under random phase-shifts) require $6MNK$ and $8MNK$ FLOPs, respectively. Then, and following the analysis of the MCG-PSD, it is not difficult to see that the complexity of initial phase-optimization and the corresponding channels' update is $MNK(15L + 10)$. Therefore, the initial complexity for obtaining Ξ , \mathbf{H} , and carrying out the initial phase-shifts optimization and channels' update is

$$\mathcal{C}_{init}^{ISM} = MNK(15L + 24). \quad (33)$$

Now we shift our attention to the complexity analysis of the iterative ISM scheme. In particular, since this method requires evaluating the precoding weights, we start by evaluating the complexities of CB and ZF precoders. Then we follow to evaluate the complexities of AS and PSD. Moreover, we take into account that the precoding matrix must be updated after discarding each v antennas, meaning that \mathbf{W} has to be evaluated $(T+1)$ times. However, only the first T evaluations of \mathbf{W} are required for the AS and PSD, and the last one is carried out specifically for the data transmission, and is not utilized for either AS or PSD purposes, and thus is neglected in our analysis.

⁶Note that when a passive IRS is employed, the BS can estimate the cascaded channel Ξ_k at once without performing the matrix multiplication in (13). Thus, our analysis in this section represents the worst-case scenario in terms of complexity requirements.

1) *Complexity of CB*: to evaluate δ_{CB} , we first need to obtain the channel cross-correlation matrix $\mathbf{R} = \mathbf{H}\mathbf{H}^H$, which requires $\sum_{t=0}^{T-1} K^2(8N - 8tv - 2)$ FLOPs. The complexity for finding $\sqrt{P_t/\text{tr}(\mathbf{R})}$ is $2KT$ FLOPs given that the trace of a square matrix is the summation of its diagonal. Finally, obtaining the CB weights $\mathbf{W} = \delta_{CB}\mathbf{H}^H$ requires $\sum_{t=0}^{T-1} 2K(N - tv)$ FLOPs. Therefore, the total complexity of the CB for any arbitrary number of AS and PSD iterations T can be given as

$$\mathcal{C}_{\mathbf{W}}^{CB} = \sum_{t=0}^{T-1} \left[K^2(8N - 8tv - 2) + 2K(N - tv) \right] + 2KT. \quad (34)$$

2) *Complexity of ZF*: obtaining the channel cross-correlation matrix \mathbf{R} requires $\sum_{t=0}^{T-1} K^2(8N - 8tv - 2)$ FLOPs; while finding \mathbf{R}^{-1} requires $\frac{8}{3}K^3T$ FLOPs [3]. Evaluating $\sqrt{P_t/\text{tr}(\mathbf{R}^{-1})}$ requires $2KT$ FLOPs; while $\mathbf{A} = \delta_{ZF}\mathbf{H}^H$ requires $\sum_{t=0}^{T-1} 2K(N - tv)$ FLOPs. Finally, finding the ZF precoding weights $\mathbf{W} = \mathbf{A}\mathbf{R}^{-1}$ requires $\sum_{t=0}^{T-1} (N - tv)(8K^2 - 2K)$ FLOPs. Therefore, the total complexity for the ZF precoder for any number of AS and PSD iterations T can be given as

$$\begin{aligned} \mathcal{C}_{\mathbf{W}}^{ZF} = & \sum_{t=0}^{T-1} \left[K^2(8N - 8tv - 2) + 8K^2(N - tv) \right] \\ & + \left(\frac{8}{3}K^3 + 2K \right) T. \end{aligned} \quad (35)$$

3) *Antenna selection*: once an initial precoding matrix \mathbf{W} is evaluated, $T = (N - N_s)/v$ iterations are carried out for AS and PSD. As a result, matrix $\Pi = [\pi_1, \dots, \pi_K]^T$ has to be evaluated T times according to (22), which requires $\sum_{t=0}^{T-1} K^2(8N - 8tv - 2)$ FLOPs. Eliminating the least desirable $N - N_s$ antennas according to (21) requires $\sum_{l=1}^{N-N_s} (14K^2 + 3K)(N - l + 1)$ FLOPs. Finally, updating Π according to (24a) after discarding each antenna requires $8K^2$ FLOPs. Therefore, the total complexity for the AS can be given as

$$\begin{aligned} \mathcal{C}_{AS}^{ISM} = & \sum_{l=1}^{N-N_s} (14K^2 + 3K)(N - l + 1) + \sum_{t=0}^{T-1} K^2(8N - 8tv - 2) \\ & + 8K^2(N - N_s), \end{aligned} \quad (36)$$

and the corresponding complexity order of the ISM-AS can be upper bounded to $\mathcal{O}(N(N - N_s)K^2)$.

4) *Phase-shifts design*: at first, $\mathbf{Z}_k (\forall k \in \mathcal{K})$ has to be evaluated for each iteration in t , which requires $\sum_{t=1}^T K^2M(8N - 8tv - 2)$ FLOPs. Moreover, finding l_m according to (29) requires $(16K^2 + 3K)TML$ FLOPs. Then, updating Π according to (27) requires $10K^2MT$ FLOPs; while updating \mathbf{H} according to (18) requires $\sum_{t=1}^T 10MK(N - tv)$ FLOPs. Therefore, the total complexity for the PSD can be given as

$$\begin{aligned} \mathcal{C}_{PSD}^{ISM} = & \sum_{t=1}^T \left[K^2M(8N - 8tv - 2) + 10MK(N - tv) \right] \\ & + MKT(16KL + 10K + 3L), \end{aligned} \quad (37)$$

and the corresponding complexity order can be expressed as $\mathcal{O}(K^2MN_s)$, assuming that $N_s > L$, and $T = 1$; while the conventional approach in (25) has a complexity order of $\mathcal{O}(M^2K^2N_sL)$. See Appendix B for details.

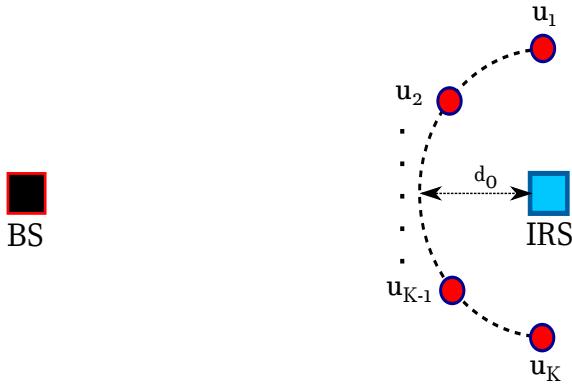


Fig. 2: A top-view of the adopted simulation setup.

5) *Total complexity*: the overall complexity required for the ISM-Ap can be given as follows

$$\mathcal{C}_{total}^{ISM} = \mathcal{C}_{init}^{ISM} + \mathcal{C}_{W}^{\rho} + \mathcal{C}_{AS}^{ISM} + \mathcal{C}_{PSD}^{ISM}, \quad (38)$$

where $\rho \in \{ZF, CB\}$, depending on the type of precoding utilized.

V. NUMERICAL RESULTS AND DISCUSSIONS

We start by introducing the properties of different channels used in this work. In particular, \mathbf{Q} was modelled as Rayleigh fading with $\mathbf{q}_k \sim \mathcal{CN}(\mathbf{0}, d_{BS,k}^{-\bar{\alpha}} \mathbf{I}_N)$, where $d_{BS,k}$ is the distance between the BS and k th user, and $\bar{\alpha}$ is the path-loss exponent for the NLoS links. In contrast, \mathbf{G} follows a Rician distribution such that

$$\mathbf{g}_k = \beta^{\text{LoS}} \mathbf{g}_k^{\text{LoS}} + \beta^{\text{NLoS}} \mathbf{g}_k^{\text{NLoS}}, \quad (39)$$

where $\mathbf{g}_k^{\text{LoS}}$ contains the deterministic LoS components for the k th user with a constant variance of $d_{IRS,k}^{-\bar{\alpha}}$, with $d_{IRS,k}$ being the distance between the IRS and k th user, and $\bar{\alpha}$ is the path-loss exponent for the LoS channels; in contrast, $\mathbf{g}_k^{\text{NLoS}} \sim \mathcal{CN}(\mathbf{0}, d_{IRS,k}^{-\bar{\alpha}} \mathbf{I}_M)$ contains the NLoS Rayleigh distributed channels. In addition, we have $\beta^{\text{LoS}} = \sqrt{\frac{K_{\text{Rician}}}{K_{\text{Rician}} + 1}}$ and $\beta^{\text{NLoS}} = \sqrt{\frac{1}{K_{\text{Rician}} + 1}}$, where K_{Rician} is the Rician- K factor. Similarly, \mathbf{U} follows a Rician distribution such that $\mathbf{u}_m = \beta^{\text{LoS}} \mathbf{u}_m^{\text{LoS}} + \beta^{\text{NLoS}} \mathbf{u}_m^{\text{NLoS}}$, where $\mathbf{u}_m^{\text{LoS}}$ contains the deterministic components with a constant variance of $d_{BS,IRS}^{-\bar{\alpha}}$, and $\mathbf{u}_m^{\text{NLoS}} \sim \mathcal{CN}(\mathbf{0}, d_{BS,IRS}^{-\bar{\alpha}} \mathbf{I}_N)$ is the Rayleigh distributed channel vector between the BS and m th reflecting element. In our work, we assume $\bar{\alpha} = 2.5$ and $\bar{\alpha} = 3.5$, while σ^2 was set to -80 dBm. Unless stated otherwise, we also assume $N = M = 64$, $K = 6$, $b = 3$, and $K_{\text{Rician}} = 5$ dB.

The simulation setup is shown in Fig. 2, where the BS is located at the origin of a 2D plane such that $(x_{BS}, y_{BS}) = (0, 0)$, while $(x_{IRS}, y_{IRS}) = (80, 0)$ meters. Moreover, the K users all have equal distances to the IRS, and lie evenly on half a circle of radius $d_0 = 30$ meters centred at (x_{IRS}, y_{IRS}) . In particular, the location of the 1st user, denoted as u_1 in Fig. 2, is fixed at $(80, 30)$, and similarly, the K th user is located at $(80, -30)$.

A. Efficiency of our proposed AS schemes

We compare the performance of our AS algorithms with two schemes. The first one is the random-based AS (R-AS);

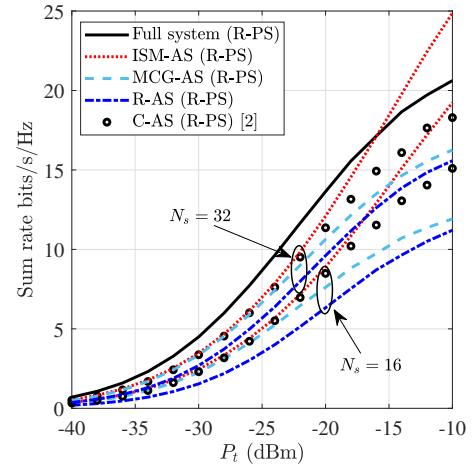


Fig. 3: Achievable rates of different AS schemes with CB and random phase-shifts when $K = 6$, $N = M = 64$, $v = N - N_s$, and $\sigma^2 = -80$ dBm.

while the second scheme is the convex optimization-based AS (C-AS) utilized in [2], where the authors stated that it can provide near optimal performance. Moreover, to ensure a fair comparison, we generate random phase-shifts (R-PS) to obtain \mathbf{H} , which is then utilized in all different AS schemes (ISM-AS, MCG-AS, R-AS, and C-AS), and we skip the PSD part to focus only on the performance of the proposed AS schemes and avoid any kind of bias in the comparison. Moreover, we set $v = N - N_s$, meaning that for our iterative ISM-AS scheme, the active precoding and AS will be carried out separately.

As demonstrated in Fig. 3, the proposed ISM-AS significantly outperforms all other AS methods and for different number of selected antennas. In addition, when $N_s = 32$ and the transmit power is greater than -16 dBm, the ISM-AS outperforms even the full-system case where all antennas are activated at the BS. This is due to the fact that with CB, some antennas can be harmful as they cause more interference than they contribute with useful signal power [4], [35]. For example, when the transmit power is -10 dBm and the number of selected antennas is 32, the ISM-AS outperforms the full system, C-AS, MCG-AS and R-AS by 4.26, 6.57, 8.61, and 9.28 bits/s/Hz, respectively.

When the ZF precoder is adopted at the BS, the ISM-AS provides the exact same performance as the C-AS as demonstrated in Fig. 4, and for all considered values of P_t and N_s . However, the ISM scheme requires much lower complexity than the C-AS, especially when the number of antennas at the BS is large as will be seen later. Moreover, the ultra-low complexity MCG-AS becomes near optimal when $K = 6$ and the number of selected antennas is relatively large, in this case when $N_s = 32$, which was not the case with CB. Note that when ZF precoder is applied, the optimal performance is obtained when all antennas are activated, which was not the case with CB. This is due to the fact that the interuser interference is completely nulled when the ZF precoder is applied, and therefore activating more antennas will always result in higher achievable rates.

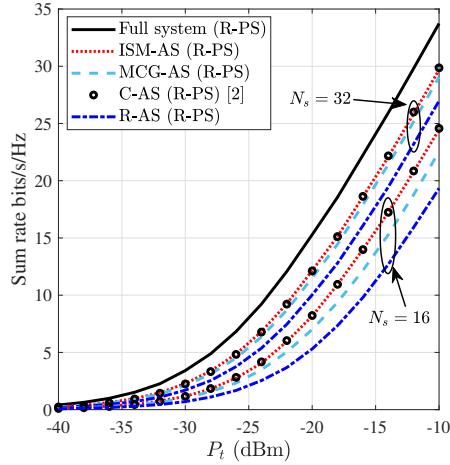


Fig. 4: Achievable rates of different AS schemes with ZF precoder and random phase-shifts when $K = 6$, $N = M = 64$, $v = N - N_s$, and $\sigma^2 = -80$ dBm.

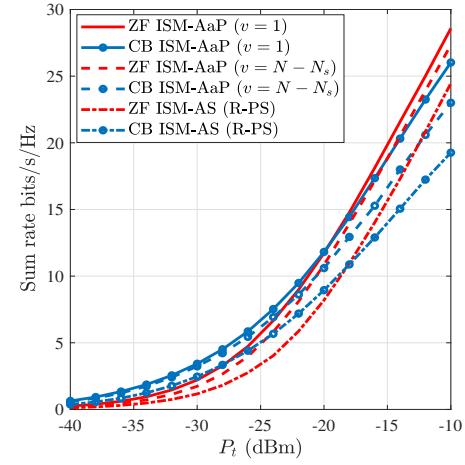


Fig. 5: Achievable rates of ISM-AaP for CB and ZF precoders when $K = 6$, $N = M = 64$, $N_s = 16$, $b = 3$, and $\sigma^2 = -80$ dBm.

It should be noted that the main motive behind AS design is to enhance the energy efficiency (EE) performance of wireless networks. In general, both AS and transmit power play key roles in the EE performance of mMIMO systems. However, our focus in this work is on the AS part. Nonetheless, joint AS and power control design for energy-efficient mMIMO-IRS networks will be the topic of future investigations.

B. Performance evaluation of the proposed PSD

Fig. 5 demonstrates the efficiency of the proposed ISM-PSD scheme. In particular, and for CB, the ISM-PSD provides a significant 6.8 and 3.8 bits/s/Hz gain compared to the random phase-shift case, for $v = 1$ and $v = N - N_s$, respectively, given that $P_t = -10$ dBm; while for the ZF precoder, a gain of 4.13 and 2.89 bits/s/Hz is obtained compared to random phase-shifts, when $v = 1$ and $v = N - N_s$, respectively, and under the same power budget of -10 dBm. We can conclude that when CB is applied, setting v to a small value (i.e. updating the phase-shifts after discarding one or few antennas) is beneficial in terms of performance, although that would come at the price of increased complexity; while for ZF, carrying the PSD after discarding all $N - N_s$ undesired antennas can still provide high performance. Note that the results shown in Fig. 5 highlight the gain obtained via the proposed PSD only, as all different curves adopt the same AS scheme.

Fig. 6 demonstrates the efficiency of the MCG-PSD scheme. It is clear that, and similar to the AS case, when the total transmit power is relatively high, the MCG-PSD works well with the ZF precoder but not with CB. This is due to the fact that maximizing the channel gain does not contribute much to the total sum-rate when CB is applied, as the interference among users becomes the main bottleneck and needs to be taken into consideration. However, with the ZF, the ultra-low-complexity MCG PSD can provide a decent gain of 3 bits/s/Hz compared to random phase-shifts when $P_t = -10$ dBm.

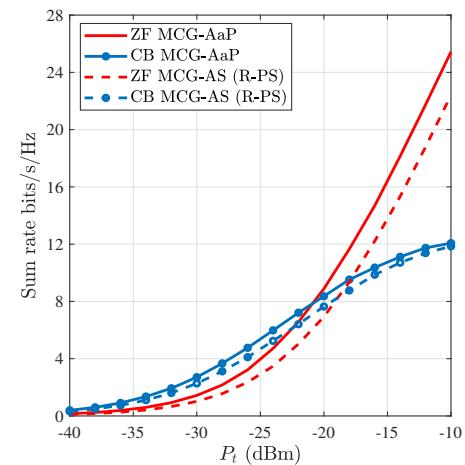


Fig. 6: Achievable rates of MCG-AaP for CB and ZF precoders when $K = 6$, $N = M = 64$, $N_s = 16$, $b = 3$, and $\sigma^2 = -80$ dBm.

C. Performance of the proposed schemes with different number of reflecting elements, quantization levels, and users

Fig. 7 shows the performance of the proposed ISM-AaP for a wide range of number of reflecting elements and different number of quantization levels. The results indicate that when the BS has a small number of RF chains, CB outperforms the ZF when the number of reflecting elements at the IRS is not very large, regardless of the number of quantization levels available at the IRS. In contrast, when M is large, ZF becomes the better option and regardless of the number of available RF chains at the BS. Moreover, when the number of IRS elements is small, having a large number of quantization levels at each reflecting element does not provide much gain. For example, when $N_s = M = 16$ and ZF precoder is applied, having 2 quantization levels per reflecting element can provide a sum-rate that is only 0.4 bits/s/Hz less than that provided

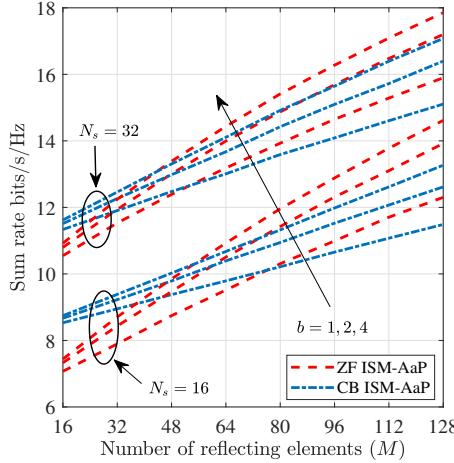


Fig. 7: Sum rate vs. number of reflecting elements under ISM-AaP scheme for CB and ZF precoders when $K = 6$, $N = 64$, $P_t = -20$ dBm, $v = N - N_s$ and $\sigma^2 = -80$ dBm.

by reflecting elements with 16 quantization levels; while the same performance gap becomes 2.3 bits/s/Hz when the IRS is equipped with 128 reflecting elements.

Fig. 8 highlights the performance of different AS and PSD schemes with ZF and CB and for a wide range of number of users. Our results indicate that for both different AS and PSD schemes, when the transmit power is limited (in this case -20 dBm), CB outperforms the ZF unless the number of users is very small (less than 6). In contrast, when the transmit power is relatively high (in this case -10 dBm), ZF can provide a significant performance gain compared to CB, as long as the number of active users is less than the number of RF chains at the BS. In addition, the ZF MCG-AaP is no longer near optimal in this case, which demonstrates the importance of the ISM-AaP scheme even when the ZF precoder is applied. For example, when the number of users is 12 and $P_t = -10$ dBm, the ZF ISM-AaP outperforms the ZF MCG-AaP and CB ISM-AaP by 9 and 9.8 bits/s/Hz, respectively. However, when the number of users is equal or close to the number of selected antennas, CB becomes the better choice as the ZF suffers from a dramatic performance loss and regardless of the AS and PSD schemes utilized, this is due to the fact that most of the available power will be used to null the interference among users.

It is worth noting that the proposed discrete PSD schemes in this work are sub-optimal, and thus are not guaranteed to converge to a global solution. However, our main focus for the design of discrete phase-shifts is to reduce the cost of implementing such methods to a considerably low value.

D. Robustness against imperfect CSI

In this subsection, we investigate the robustness of our proposed schemes against channel estimation (CE) errors. Note that despite the passive nature of IRSs, estimation of wireless channels with satisfactory accuracy can still be achieved as demonstrated in [32], [36]–[38] and the references therein.

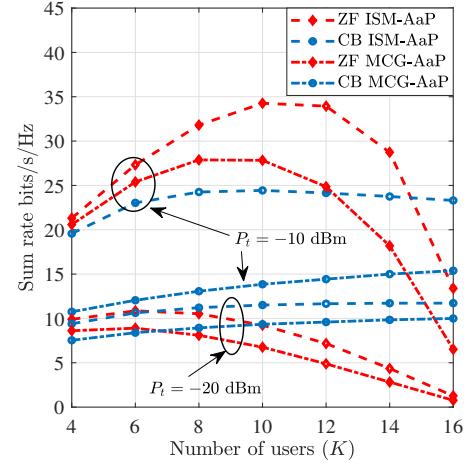


Fig. 8: Sum rate vs. number of users for CB and ZF precoders when $N = M = 64$, $N_s = 16$, $v = N - N_s$, $b = 3$ and $\sigma^2 = -80$ dBm.

However, estimating each of the two links \mathbf{u}_m and \mathbf{g}_k ($m \in \mathcal{M}$, $k \in \mathcal{K}$), of the cascaded channel Ξ_k separately can be challenging when dealing with passive IRSs. Instead, one can obtain an estimate of Ξ_k altogether using orthogonal pilot sequences sent by the users [32].⁷ Moreover, estimating the direct links between each user and BS \mathbf{q}_k ($k \in \mathcal{K}$) can be performed while the IRS is “switched off” (i.e. working in absorbing mode rather than reflecting mode). As a result, we introduce two independent CE errors, one for the direct links, and one for the cascaded channels as follows

$$\hat{\mathbf{q}}_k = \sqrt{1 - \sigma_e} \mathbf{q}_k + \sqrt{\sigma_e} \tilde{\mathbf{q}}_k, \quad (40a)$$

$$\hat{\Xi}_k = \sqrt{1 - \sigma_e} \Xi_k + \sqrt{\sigma_e} \tilde{\Xi}_k, \quad (40b)$$

where $\tilde{\mathbf{q}}_k$ and $\tilde{\Xi}_k$ are the complex Gaussian CE errors [39], which are assumed to be uncorrelated with \mathbf{q}_k and Ξ_k , but have the same statistical properties as of \mathbf{q}_k and Ξ_k in terms of mean and variance values (see appendix C), and $\sigma_e \in [0, 1]$ accounts for the estimation accuracy. Then, we can represent the channel coefficients for the k th user as

$$\hat{\mathbf{h}}_k = \phi^T \hat{\Xi}_k + \hat{\mathbf{q}}_k. \quad (41)$$

As Fig. 9 demonstrates, both the ZF and CB suffer from performance degradation when dealing with imperfect CSI. However, the CB is more robust against CE errors compared to the ZF, as the latter is known to be more sensitive to imperfect CSI and unmodeled interference. Moreover, the ISM-AaP scheme maintains its superiority over the MCG-AaP even for imperfect CSI, and for both CB and ZF precoders.

E. Computational complexity

Here we highlight the complexity requirement for the proposed AS and PSD schemes. We first demonstrate the complexities of AS and PSD separately, for a wide range of

⁷Note that our proposed AaP schemes require the knowledge of only the cascaded channel of the IRS (Ξ_k), and thus can be implemented even when the IRS is entirely passive.

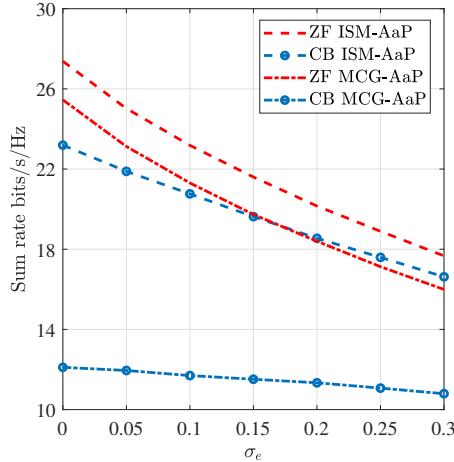


Fig. 9: Sum rate vs. estimation accuracy for CB and ZF precoders when $N = M = 64$, $N_s = 16$, $K = 6$, $v = N - N_s$, $b = 3$, $P_t = -10$ dBm and $\sigma^2 = -80$ dBm.

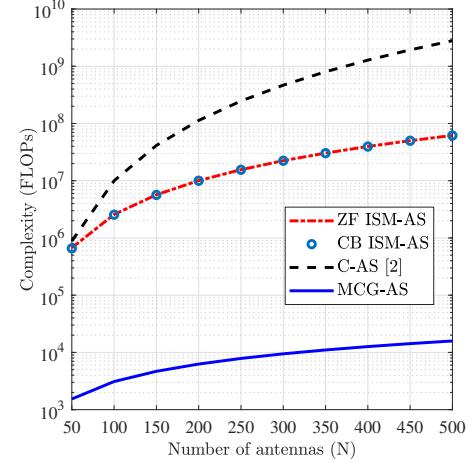


Fig. 10: Computational complexity vs number of BS antennas for different AS schemes when $N_s = N/4$, $v = N - N_s$, and $K = 6$.

N and M values, respectively. Then, we show the complexity of the overall design.

For the AS, the type of precoding utilized is irrelevant when MCG-AS is applied, as this method requires only the knowledge of \mathbf{H} . In contrast, the ISM-AS requires the knowledge of both \mathbf{H} and \mathbf{W} . Therefore, the complexity of the ISM-AS is the sum of \mathcal{C}_w and \mathcal{C}_{AS}^{ISM} . We compare the complexities of our methods with the C-AS in [2], which relies on interior-point methods [7], and has a complexity order of $\mathcal{O}(N^{3.5})$ [24]. Finally, for the MCG-AS, we focus only on the computations required to perform the AS after obtaining \mathbf{H} to ensure a fair comparison among different schemes. Note that for our own AaP methods, we show the exact complexities required for their implementations based on our analysis in Section IV.

As demonstrated in Fig. 10, the MCG requires the least complexity; however, that comes at the price of a degraded performance when CB is adopted, or when ZF is applied and the number of users is relatively large. In contrast, the C-AS has the highest complexity, especially when the number of antennas at the BS is large; while the ISM-AS scheme has lower complexity than the C-AS method and achieves near optimal performance regardless of the precoding scheme utilized, which is not the case for both C-AS and the MCG-AS schemes.

For the PSD, we show the complexity as a function of M . In particular, the complexity of the proposed MCG-PSD is given in (31); while the complexity of the proposed ISM-PSD is the sum of \mathcal{C}_{PSD}^{ISM} and \mathcal{C}_{init}^{ISM} after subtracting the complexity required for finding \mathbf{H} , which accounts for $14MNK$. We compare the required complexities of our proposed implementation methods with that required by implementing conventional MCG and ISM (cMCG and cISM) successive refinement schemes based on our complexity analysis in Appendix A and Appendix B.

As Fig. 11 demonstrates, the proposed design methods achieve significant complexity reduction compared to conven-

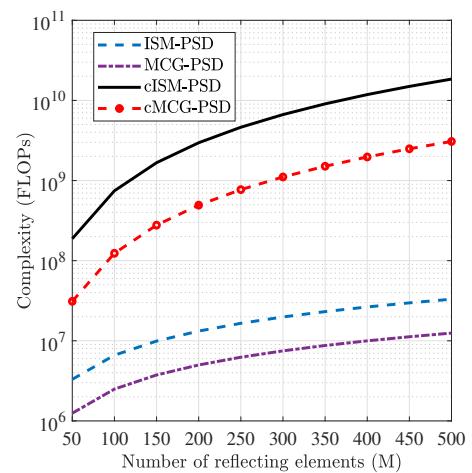


Fig. 11: Computational complexity vs number of IRS elements for different PSD schemes when $N = 64$, $N_s = 32$, $v = N - N_s$, $b = 3$, and $K = 6$.

tional implementation methods and without any compromise in the performance. More importantly, the complexity of the proposed ISM and MCG schemes scale linearly with M , which makes the proposed designing methods in this paper highly attractive when dealing with large IRS arrays. Finally, the total complexity is shown in Fig. 12, which demonstrates that the proposed AS with PSD schemes scale well when both the number of antennas at the BS and the number of reflecting elements at the IRS are large. Moreover, we can observe that the complexity of the MCG-PSD becomes dominant for the MCG-AaP scheme, since optimizing the phase of each reflecting element affects all available antennas resulting in higher complexity compared to the MCG-AS, which works only on the effective channel matrix \mathbf{H} , and thus does not involve the M reflecting elements in the selection process.

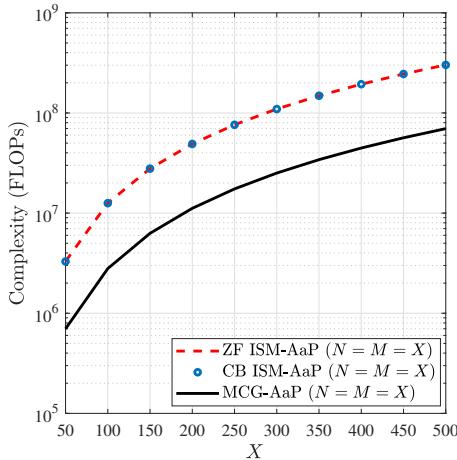


Fig. 12: Computational complexity vs number of antennas/reflecting elements for different AS and PSD schemes when $N = M = X$, $N_s = N/4$, $v = N - N_s$, $b = 3$, and $K = 6$.

VI. CONCLUSION

We proposed two novel AS and PSD schemes in multiuser mMIMO-IRS networks, and thoroughly evaluated their performances under both CB and ZF precoders. The proposed algorithms were designed in a way such that matrix/vector multiplications were minimized, and the AS and PSD were carried out using only element-wise operations, and without compromising the performance. In particular, for the PSD, our proposed schemes provided a significant performance gain compared to random phase-shifts, and their complexities scaled linearly with the number of reflecting elements at the IRS; while for the AS, the ISM approach provided near optimal performance with reduced complexity compared to other efficient AS methods found in the literature, and can be utilized with any type of active precoding at the mMIMO BS.

APPENDIX A

To optimize each of the M reflecting elements according to (15), we start by evaluating $\|\phi^T[\Xi_k]_{:,s} + [\mathbf{q}_k]_s\|_1$, which requires $(8MN_s + 5N_s - 1)$ FLOPs. Then, summing over all $k \in \mathcal{K}$ and for all L possible values of $[\phi]_m$ results in $(8MN_sKL + 5N_sKL)$ FLOPs. Therefore, the total number of FLOPs required for the conventional MCG (cMCG)-PSD to optimize all M reflecting elements is

$$C_{PSD}^{cMCG} = 8M^2N_sKL + 5MN_sKL. \quad (42)$$

and the corresponding complexity order is $\mathcal{O}(M^2N_sKL)$.

APPENDIX B

Evaluating the objective function of (25) for each of the L values of $[\phi]_m$ in α at any given iteration t requires $MK^2(8N - 8tv) + K^2(8N - 8tv + 4) + (3K - 1)$ FLOPs. As a result, the total number of FLOPs required for optimizing all M phase-shifts with an arbitrary number of iterations T ,

utilizing the conventional ISM (cISM)-PSD approach in (25), is

$$C_{PSD}^{cISM} = \sum_{t=1}^T \left[M^2K^2L(8N - 8tv) + K^2ML(8N - 8tv + 4) \right] + 3MKLT, \quad (43)$$

and the corresponding complexity order when $T = 1$ can be reduced to $\mathcal{O}(M^2K^2N_sL)$.

APPENDIX C

Here we present the derivation of statistical mean and variance values for the cascaded channels.

For the m th link between the k th user and the IRS, we have

$$\mathbb{E}\{[\mathbf{g}_k]_m\} = \mathbb{E}\{\beta^{\text{LoS}}[\mathbf{g}_k]^{\text{LoS}}_m + \beta^{\text{NLoS}}[\mathbf{g}_k]^{\text{NLoS}}_m\} = \beta^{\text{LoS}}\mathbb{E}\{[\mathbf{g}_k]^{\text{LoS}}_m\} + \beta^{\text{NLoS}}\mathbb{E}\{[\mathbf{g}_k]^{\text{NLoS}}_m\} = 0, \quad (44)$$

where $\mathbb{E}\{[\mathbf{g}_k]^{\text{LoS}}_m\} = 0$ since we assume random phases for the deterministic LoS links similar to [40]. Moreover, the variance for the same channel is

$$\begin{aligned} \mathbb{V}\{[\mathbf{g}_k]_m\} &= \mathbb{V}\{\beta^{\text{LoS}}[\mathbf{g}_k]^{\text{LoS}}_m + \beta^{\text{NLoS}}[\mathbf{g}_k]^{\text{NLoS}}_m\} \\ &= \mathbb{V}\{\beta^{\text{LoS}}[\mathbf{g}_k]^{\text{LoS}}_m\} + \mathbb{V}\{\beta^{\text{NLoS}}[\mathbf{g}_k]^{\text{NLoS}}_m\} \\ &\quad + \underbrace{2\beta^{\text{LoS}}\beta^{\text{NLoS}}\text{cov}([\mathbf{g}_k]^{\text{LoS}}_m, [\mathbf{g}_k]^{\text{NLoS}}_m)}_{=0} \\ &= (\beta^{\text{LoS}})^2 d_{\text{IRS},k}^{-\bar{\alpha}} + (\beta^{\text{NLoS}})^2 d_{\text{IRS},k}^{-\bar{\alpha}} = \beta_g^{(k)}, \end{aligned} \quad (45)$$

where $\text{cov}(x, y)$ is the covariance between x and y , and is equal to zero when the two random variables are uncorrelated. Similarly, for links between the IRS and BS, we can write

$$\mathbb{E}\{[\mathbf{U}]_{m,n}\} = 0, \quad (46)$$

$$\mathbb{V}\{[\mathbf{U}]_{m,n}\} = (\beta^{\text{LoS}})^2 d_{\text{BS},\text{IRS}}^{-\bar{\alpha}} + (\beta^{\text{NLoS}})^2 d_{\text{BS},\text{IRS}}^{-\bar{\alpha}} = \beta_U, \quad (47)$$

and we drop the indices of (m, n) in β_U as the distance between any antenna at the BS and any element at the IRS is assumed to be equal for all antennas/reflecting-elements. Finally, for the cascaded channel we have

$$\mathbb{E}\{[\Xi_k]_{m,n}\} = \mathbb{E}\{[\mathbf{g}_k]_m[\mathbf{U}]_{m,n}\} = \mathbb{E}\{[\mathbf{g}_k]_m\}\mathbb{E}\{[\mathbf{U}]_{m,n}\} = 0, \quad (48)$$

$$\begin{aligned} \mathbb{V}\{[\Xi_k]_{m,n}\} &= \mathbb{V}\{[\mathbf{g}_k]_m[\mathbf{U}]_{m,n}\} \\ &= (\mathbb{E}\{[\mathbf{g}_k]_m\})^2 \mathbb{V}\{[\mathbf{U}]_{m,n}\} + (\mathbb{E}\{[\mathbf{U}]_{m,n}\})^2 \mathbb{V}\{[\mathbf{g}_k]_m\} \\ &\quad + \mathbb{V}\{[\mathbf{g}_k]_m\}\mathbb{V}\{[\mathbf{U}]_{m,n}\} \\ &= \beta_g^{(k)}\beta_U. \end{aligned} \quad (49)$$

REFERENCES

- [1] L. M. Correia, D. Zeller, O. Blume, D. Ferling, Y. Jading, I. Góðor, G. Auer, and L. Van Der Perre, "Challenges and enabling technologies for energy aware mobile radio networks," *IEEE Commun. Mag.*, vol. 48, no. 11, pp. 66–72, Nov. 2010.
- [2] X. Gao, O. Edfors, F. Tufvesson, and E. G. Larsson, "Massive MIMO in real propagation environments: Do all antennas contribute equally?" *IEEE Trans. Commun.*, vol. 63, no. 11, pp. 3917–3928, Nov. 2015.
- [3] P. V. Amadori and C. Masouros, "Interference-driven antenna selection for massive multiuser MIMO," *IEEE Trans. Veh. Technol.*, vol. 65, no. 8, pp. 5944–5958, Aug. 2015.
- [4] Z. Abdullah, C. C. Tsimenidis, G. Chen, M. Johnston, and J. A. Chambers, "Efficient low-complexity antenna selection algorithms in multi-user massive MIMO systems with matched filter precoding," *IEEE Trans. Veh. Technol.*, vol. 69, no. 3, pp. 2993–3007, Mar. 2020.

- [5] Z. Abdullah, C. C. Tsimenidis, M. Alageli, M. Johnston, G. Chen, and J. Chambers, "Cell-edge-aware antenna selection and power allocation in massive MIMO systems," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Hawaii, USA, Dec. 2019, pp. 1–6.
- [6] Y. Gao, H. Vinck, and T. Kaiser, "Massive MIMO antenna selection: Switching architectures, capacity bounds, and optimal antenna selection algorithms," *IEEE Trans. Signal Process.*, vol. 66, no. 5, pp. 1346–1360, Mar. 2018.
- [7] M. O. Mendonça, P. S. Diniz, T. N. Ferreira, and L. Lovisolo, "Antenna selection in massive MIMO based on greedy algorithms," *IEEE Trans. Wireless Commun.*, vol. 19, no. 3, pp. 1868–1881, Mar. 2020.
- [8] J. Chen, S. Chen, Y. Qi, and S. Fu, "Intelligent massive MIMO antenna selection using Monte Carlo tree search," *IEEE Trans. Signal Process.*, vol. 67, no. 20, pp. 5380–5390, Oct. 2019.
- [9] Z. Abdullah, C. C. Tsimenidis, and M. Johnston, "Tabu search vs. bio-inspired algorithms for antenna selection in spatially correlated massive MIMO uplink channels," in *Proc. IEEE European Signal Process. Conf. (EUSIPCO)*, Budapest, Hungary, Sep. 2016, pp. 41–45.
- [10] ———, "Quantum-inspired tabu search algorithm for antenna selection in massive MIMO systems," in *Proc. IEEE Wireless Commun. Net. Conf. (WCNC)*, Barcelona, Spain, Apr. 2018, pp. 1–6.
- [11] P. V. Amadori and C. Masouros, "Large scale antenna selection and precoding for interference exploitation," *IEEE Trans. Commun.*, vol. 65, no. 10, pp. 4529–4542, Oct. 2017.
- [12] M. Tian, W. Sun, P. Zhang, L. Huang, and Q. Li, "Joint beamforming design and receive antenna selection for large-scale MIMO wiretap channels," *IEEE Trans. Veh. Technol.*, vol. 69, no. 3, pp. 2716–2730, Mar. 2020.
- [13] M. S. Ibrahim, A. Konar, and N. D. Sidiropoulos, "Fast algorithms for joint multicast beamforming and antenna selection in massive MIMO," *IEEE Trans. Signal Process.*, vol. 68, pp. 1897–1909, Mar. 2020.
- [14] K. B. Letaief, W. Chen, Y. Shi, J. Zhang, and Y.-J. A. Zhang, "The roadmap to 6G: AI empowered wireless networks," *IEEE Commun. Mag.*, vol. 57, no. 8, pp. 84–90, Aug. 2019.
- [15] 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.
- [16] Z. Zhang, Y. Xiao, Z. Ma, M. Xiao, Z. Ding, X. Lei, G. K. Karagiannidis, and P. Fan, "6G wireless networks: Vision, requirements, architecture, and key technologies," *IEEE Veh. Technol. Mag.*, vol. 14, no. 3, pp. 28–41, Sep. 2019.
- [17] N. Rajatheva, I. Atzeni, E. Bjornson, A. Bourdoux, S. Buzzi, J.-B. Dore, S. Erikuçuk, M. Fuentes, K. Guan, Y. Hu *et al.*, "White paper on broadband connectivity in 6G," Available at: <http://arxiv.org/abs/2004.14247>, 2020.
- [18] Z. Abdullah, G. Chen, S. Lambotharan, and J. A. Chambers, "A hybrid relay and intelligent reflecting surface network and its ergodic performance analysis," *IEEE Wireless Commun. Lett.*, vol. 9, no. 10, pp. 1653–1657, Oct. 2020.
- [19] Q. Wu and R. Zhang, "Towards smart and reconfigurable environment: Intelligent reflecting surface aided wireless network," *IEEE Commun. Mag.*, vol. 58, no. 1, pp. 106–112, Jan. 2020.
- [20] 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.
- [21] G. Zhou, C. Pan, H. Ren, K. Wang, M. Di Renzo, and A. Nallanathan, "Robust beamforming design for intelligent reflecting surface aided MISO communication systems," *IEEE Wireless Commun. Lett.*, vol. 9, no. 10, pp. 1658–1662, Oct. 2020.
- [22] Q. Wu and R. Zhang, "Intelligent reflecting surface enhanced wireless network via joint active and passive beamforming," *IEEE Trans. Wireless Commun.*, vol. 18, no. 11, pp. 5394–5409, Nov. 2019.
- [23] Z. Abdullah, G. Chen, S. Lambotharan, and J. A. Chambers, "Optimization of intelligent reflecting surface assisted full-duplex relay networks," *IEEE Wireless Commun. Lett.*, vol. 10, no. 2, pp. 363–367, Feb. 2021.
- [24] J. He, K. Yu, Y. Shi, Y. Zhou, W. Chen, and K. B. Letaief, "Reconfigurable intelligent surface assisted massive MIMO with antenna selection," Available at: <https://arxiv.org/abs/2009.07546>, 2020.
- [25] P. Xu, G. Chen, Z. Yang, and M. Di Renzo, "Reconfigurable intelligent surfaces assisted communications with discrete phase shifts: How many quantization levels are required to achieve full diversity?" *IEEE Wireless Commun. Lett.*, vol. 10, no. 2, pp. 358–362, Feb. 2021.
- [26] R. Sarvendranath and A. K. R. Chavva, "Low-complexity joint antenna selection and beamforming for an IRS assisted system," in *Proc. IEEE Wireless Commun. Net. Conf. (WCNC)*, Nanjing, China, Mar. 2021, pp. 1–6.
- [27] Q. Wu and R. Zhang, "Beamforming optimization for wireless network aided by intelligent reflecting surface with discrete phase shifts," *IEEE Trans. Commun.*, vol. 68, no. 3, pp. 1838–1851, Mar. 2020.
- [28] C. You, B. Zheng, and R. Zhang, "Channel estimation and passive beamforming for intelligent reflecting surface: Discrete phase shift and progressive refinement," *IEEE J. Sel. Areas Commun.*, Nov. 2020.
- [29] B. Di, H. Zhang, L. Song, Y. Li, Z. Han, and H. V. Poor, "Hybrid beamforming for reconfigurable intelligent surface based multi-user communications: Achievable rates with limited discrete phase shifts," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 8, pp. 1809–1822, Aug. 2020.
- [30] B. Di, H. Zhang, L. Li, L. Song, Y. Li, and Z. Han, "Practical hybrid beamforming with finite-resolution phase shifters for reconfigurable intelligent surface based multi-user communications," *IEEE Trans. Veh. Technol.*, vol. 69, no. 4, pp. 4565–4570, Apr. 2020.
- [31] A. H. Mehana and A. Nosratinia, "Diversity of MIMO linear precoding," *IEEE Trans. Inf. Theory*, vol. 60, no. 2, pp. 1019–1038, Feb. 2014.
- [32] Q.-U.-A. Nadeem, H. Alwazani, A. Kamoun, A. Chaaban, M. Debbah, M.-S. Alouini *et al.*, "Intelligent reflecting surface-assisted multi-user MISO communication: Channel estimation and beamforming design," *IEEE Open J. Commun. Soc.*, vol. 1, pp. 661–680, June 2020.
- [33] A. Taha, M. Alrabeiah, and A. Alkhateeb, "Enabling large intelligent surfaces with compressive sensing and deep learning," *IEEE Access*, vol. 9, pp. 44 304–44 321, Mar. 2021.
- [34] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge University Press, 2004.
- [35] R. Hamdi, E. Driouch, and W. Ajib, "Resource allocation in downlink large-scale MIMO systems," *IEEE Access*, vol. 4, pp. 8303–8316, Nov. 2016.
- [36] Z. Wang, L. Liu, and S. Cui, "Channel estimation for intelligent reflecting surface assisted multiuser communications: Framework, algorithms, and analysis," *IEEE Trans. Wireless Commun.*, vol. 19, no. 10, pp. 6607–6620, Oct. 2020.
- [37] B. Zheng and R. Zhang, "Intelligent reflecting surface-enhanced OFDM: Channel estimation and reflection optimization," *IEEE Wireless Commun. Lett.*, vol. 9, no. 4, pp. 518–522, Apr. 2019.
- [38] B. Zheng, C. You, and R. Zhang, "Intelligent reflecting surface assisted multi-user OFDMA: Channel estimation and training design," *IEEE Trans. Wireless Commun.*, vol. 19, no. 12, pp. 8315–8329, Dec. 2020.
- [39] S. Hong, C. Pan, H. Ren, K. Wang, K. K. Chai, and A. Nallanathan, "Robust transmission design for intelligent reflecting surface-aided secure communication systems with imperfect cascaded CSI," *IEEE Trans. Wireless Commun.*, vol. 20, no. 4, pp. 2487–2501, Apr. 2021.
- [40] 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.



Zaid Abdullah (Member, IEEE) Received his BSc degree in electrical engineering from Mosul University, Mosul, Iraq, in 2012, and the MSc and PhD degrees in communications and signal processing from Newcastle University, Newcastle-upon-Tyne, UK, in 2014 and 2019, respectively. Between October 2018 and May 2021, he was a Post-Doctoral Research Assistant at the University of Leicester, UK. In June 2021, he joined the Interdisciplinary Centre for Security, Reliability and Trust (SnT), University of Luxembourg, Luxembourg, as a Research Associate.

His research interests are mainly focused on massive MIMO, intelligent reflecting surfaces, low-complexity algorithm design, hybrid relaying networks, and satellite communications.



Gaojie Chen (S'09 – M'12 – SM'18) received the B.Eng. and B.Ec. Degrees in electrical information engineering and international economics and trade from Northwest University, China, in 2006, and the M.Sc. (Hons.) and Ph.D. degrees in electrical and electronic engineering from Loughborough University, Loughborough, U.K., in 2008 and 2012, respectively. After graduation, he took up academic and research positions at DT Mobile, Loughborough University, University of Surrey, University of Oxford and University of Leicester, U.K. He is currently an Assistant Professor with the Institute for Communication Systems, 5GIC & 6GIC, University of Surrey, U.K. His current research interests include information theory, wireless communications, cooperative communications, cognitive radio, the Internet of Things, secrecy communications, and random geometric networks. He received the Exemplary Reviewer Certificates of the IEEE WIRELESS COMMUNICATIONS LETTERS in 2018 and the IEEE TRANSACTIONS ON COMMUNICATIONS in 2019 and Exemplary Editor of the IEEE COMMUNICATIONS LETTERS in 2021. He serves as an Associate Editor for the IEEE COMMUNICATIONS LETTERS, IEEE WIRELESS COMMUNICATIONS LETTERS, IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS - MACHINE LEARNING IN COMMUNICATIONS AND NETWORKS and *Electronics Letters* (IET).



Sangarapillai Lambotharan (Senior Member, IEEE) received the Ph.D. degree in signal processing from Imperial College London, U.K., in 1997. He was a Visiting Scientist with the Engineering and Theory Centre, Cornell University, USA, in 1996. Until 1999, he was a Post-Doctoral Research Associate with Imperial College London. From 1999 to 2002, he was with the Motorola Applied Research Group, U.K., where he investigated various projects, including physical link layer modeling and performance characterization of GPRS, EGPRS, and UTRAN. He was with King's College London and Cardiff University as a Lecturer and a Senior Lecturer, respectively, from 2002 to 2007. He is currently a Professor of Digital Communications and the Head of the Signal Processing and Networks Research Group, Wolfson School of Mechanical, Electrical and Manufacturing Engineering, Loughborough University, U.K. His current research interests include 5G networks, MIMO, blockchain, machine learning, and network security. He has authored 250 journal articles and conference papers in these areas. He is a Fellow of IET and Senior Member of IEEE. He serves as an Associate Editor for the IEEE Transactions on Signal Processing and IEEE Transactions on Communications.



Jonathon A. Chambers (Fellow, IEEE) received the Ph.D. and D.Sc. degrees in signal processing from the Imperial College of Science, Technology and Medicine, Imperial College London, London, U.K., in 1990 and 2014, respectively. He served as the Head of the School of Engineering, University of Leicester from 2017 to 2019, and became an Emeritus Professor of Signal and Information Processing in 2020. He is also an International Honorary Dean of the College of Intelligent Systems Science and Engineering, Harbin Engineering University, China. He has advised almost 100 researchers through to Ph.D. graduation and published more than 600 conference papers and journal articles. His research interests include adaptive signal processing and machine learning and their applications, including in navigation and wireless communications.