Intelligent Reflecting Surfaces: Sum-Rate Optimization Based on Statistical CSI

Andrea Abrardo, Davide Dardari, Marco Di Renzo

Intelligent Reflecting Surfaces: Sum-Rate Optimization Based on Statistical CSI

Andrea Abrardo, *Senior Member, IEEE*, Davide Dardari, *Senior Member, IEEE*,

and Marco Di Renzo, *Fellow, IEEE*

Abstract

In this paper, we consider a multi-user multiple-input multiple-output (MIMO) system aided by multiple intelligent reflecting surfaces (IRSs) that are deployed to increase the coverage and, possibly, the rank of the channel. We propose an optimization algorithm to configure the IRSs, which is aimed at maximizing the network sum-rate by exploiting only the statistical characterization of the environment, such as the distribution of the locations of the users and the distribution of the multipath channels. As a consequence, the proposed approach does not require the estimation of the instantaneous channel state information (CSI) for system optimization, thus significantly relaxing (or even avoiding) the need of frequently reconfiguring the IRSs, which constitutes one of the most critical issues in IRS-assisted systems. Numerical results confirm the validity of the proposed approach. It is shown, in particular, that IRS-assisted wireless systems that are optimized based on statistical CSI still provide large performance gains as compared to the baseline scenarios in which no IRSs are deployed.

Index Terms

Intelligent reflecting surfaces, statisticsl CSI, MU-MIMO, sum-rate.

I. INTRODUCTION

The path towards sixth generation (6G) wireless networks foresees the use of new frequency bands such as the terahertz (THz) spectrum, as well as the development of new transmission technologies in order to comply with the challenging requirements introduced by new mobile applications [1]–[4]. Among several promising transmission technologies, one technology that has recently received a significant attention is the intelligent reflecting surface (IRS) [5]. An IRS can be broadly defined as a man-made thin surface, which can be realized by using

inexpensive antenna elements or metamaterials, whose properties provide great flexibility on how the electromagnetic (EM) waves that impinge upon the surface are reflected, refracted, focused, etc. [6]–[9]. The most recent research activities on IRS are concerned with the design of nearly passive reconfigurable planar structures, i.e., surfaces, whose surface impedance can be appropriately configured according to specified communication needs [10]–[13]. These solutions are referred to as reconfigurable intelligent surfaces (RISs), and are capable of realizing multiple wave transformations (not only anomalous reflection, refraction, or focusing) [14].

IRSs have several applications, and, in particular, they are useful for enhancing the coverage in severe non-line-of-sight (NLOS) channel conditions by acting as intelligent mirrors [15]. More in general, IRSs are viewed as a technology enabler for realizing the so-called *smart radio environment*, i.e., a wireless system in which the environment (e.g., the channel) becomes a variable that can be optimized in addition to the parameters of the communication devices [16]. For instance, an IRS can be programmed to create "artificial" multipath in such a way that the channel rank, and so the channel capacity, in multiple-input multiple-output (MIMO)-aided systems is increased [17], [18]. In the presence of multiple users, IRSs can be optimized to maximize the signal-to-interference-noise ratio (SINR) or the achievable rate (AR) [19].

One of the main design and implementation challenges of IRS-assisted communication systems is the need of estimating the channel state information (CSI) of the transmitter-IRS and IRS-receiver channels, and the need of appropriately configuring (virtually in real-time) the operation of the IRS via a feedback channel. As recently shown in [20], the overhead associated with the channel estimation and reporting phases through a feedback channel may offset the performance gains introduced by an IRS, if the number of individually reconfigurable elements is too large and/or an inefficient channel estimation algorithm is employed. Furthermore, due to the dynamic behavior of the wireless channel and the mobility of the users, an IRS needs to have the capability of re-configuring itself at a time-scale that depends on the coherence time of the environment (e.g., the channel), which may be difficult to realize in some highly dynamic environments [21].

Motivated by these considerations, there is general a consensus that it is imperative to develop low-overhead IRS-assisted systems, which are designed and optimized without relying on the perfect knowledge of the CSI, while still providing performance gains with respect to legacy wireless networks in the absence of IRSs. The present paper tackles this open research issue, by proposing a *joint offline and online optimization design* for IRS-assisted communication systems. In the next two sub-sections, we first review the research works related to ours and

then summarize the main contributions and novelty of the proposed approach.

A. State of the art

A comprehensive state of the art review on IRS-assisted systems can be found in [14], [21], [22]. The present paper is focused on optimizing IRSs under the assumption that the perfect knowledge of the CSI is not available. In this section, therefore, we discuss only research works that are closely related to this main focus of the paper. Interested readers are referred to [14, Sec. V-J], [21, Table III] and [14, Sec. V-G], [21, Table V] for a detailed summary of the available research results on resource allocation and optimization, and channel estimation, respectively.

Based on the analysis of the state of the art (see, e.g., [21, Table III]), we evince that the vast majority of research works focused on optimizing the performance of IRS-assisted systems are based on the perfect knowledge of the CSI for all the available channels, as well as on the availability of appropriate feedback mechanisms to configure the operation of the IRSs based on the acquired CSI. This may make the channel estimation and signaling (feedback) overhead prohibitively high, especially in multi-user MIMO (MU-MIMO) systems [20]. More precisely, several channel estimation schemes have been proposed for different system configurations (single/multiple users, single/multiple antennas), as recently surveyed in [14, Sec. V-G]. The common denominator of most of the available schemes is to let the IRSs change the analog beamforming vectors during the channel estimation phase according to pre-designed reflection patterns that are exploited and appropriately optimized to facilitate and to make robust the channel estimation process. In spite of the relatively large number of research contributions available to date, the channel estimation in IRS-aided systems is still an open issue that is characterized by three major challenges: (i) the long training time, especially in MU-MIMO systems, which may not be tolerable in dynamic scenarios; (ii) the real-time reconfiguration of the reflection functionality of the IRS through a dedicated control channel with the base station (BS); and (iii) the need of ad hoc channel estimation and signaling protocols that make the deployment of IRSs non-transparent to existing communication protocols. In order to tackle these open research issues, some authors have recently started researching on IRS-aided systems that do not necessarily rely upon the perfect knowledge of the CSI, e.g., [23], [24], [25], [26], [27], [28].

In [23], the authors formulate an optimization problem for determining the optimal linear precoder, the power allocation matrix at the transmitter, and the phase shifts matrix at the IRS that maximize the minimum SINR of the system. The authors develop a deterministic approximation

for the parameters of the optimal linear precoder by using tools from random matrix theory. Based on the obtained deterministic approximations, an algorithm based on the projected gradient ascent method is proposed to solve the non-convex optimization problem for determining the phases that maximize the minimum SINR of the users. The developed algorithm requires the knowledge of only the large-scale channel statistics, including the slowly varying spatial correlation matrices, which can be estimated using the knowledge of only the mean angles and angular spreads, and the channel attenuation coefficients, which change slowly with time. A multiple-input single-output (MISO) downlink multiuser communication system is considered in [24], in which the phase shifts of a single IRS and the transmit beamforming/precoding vector at the access point are optimized by relaying on a two-timescale beamforming optimization algorithm. More precisely, the phase shifts applied by the IRS are optimized based on the statistical CSI of all the links, and the transmit beamforming/precoding vectors at the access point are designed by relying on the instantaneous CSI of the effective channels, given the optimized configuration of the IRS. The authors show that the proposed approach can reduce the channel training overhead and the complexity of the analog beamforming design over existing schemes that need the knowledge of the instantaneous CSI of all the available channels. It is shown that, even though the phase shifts are optimized by using statistical CSI, a significant performance gain in terms of sum-rate can be obtained, as compared with systems in the absence of an IRS. The authors of [25] study an IRS-assisted single-input single-output (SISO) broadcast channel, in which the IRS elements introduce random or deterministic phase rotations without requiring instantaneous CSI. The average sum-rate capacity of the considered system is achieved by opportunistic scheduling the user with the largest signal-to-noise ratio (SNR). Also, the authors show that the average sum-rate under the proposed scheme in a slow-fading environment is capable of approaching the performance of coherent beamforming as the number of users increases. In [26], the authors propose an IRS-enhanced MISO system with reflection pattern modulation, where the IRS can configure its reflection state for enhancing the received signal power via analog beamforming and simultaneously conveying its own information via reflection modulation. The authors formulate an optimization problem to maximize the average received signal power by jointly optimizing the digital beamforming at the access point and the analog beamforming at the IRS by assuming that the state information at the IRS is statistically known by the access point, and that the CSI is affected by channel estimation errors. A sub-optimal algorithm for solving the resulting optimization problem based on the alternating optimization method is proposed as well. The

authors of [27] investigate an IRS-aided multi-pair communication system, in which multi-pair users exchange information via an IRS. The authors derive an approximated expression of the achievable rate by assuming that only the knowledge of the statistical CSI is available. In [28], finally, the estimation of the local CSI and the need for a dedicated control channel are avoided by exploiting the frequency selectivity of the IRSs. It is shown that, if the IRSs are appropriately designed, different reflection properties can be obtained in an orthogonal frequency division multiplexing (OFDM) context by suitably changing the sub-carriers allocated to each user. The effectiveness of the method is validated in the far-field and near-field regimes.

B. Contribution of the paper

In the present paper, we consider a multi-IRS MU-MIMO system and we optimize the reconfigurable elements of the IRSs and the beamforming vectors at the BS and user equipments (UEs) so as to maximize the system sum-rate. The main novelty of the proposed approach lies in not requiring the instantaneous CSI for optimizing the IRSs, thus relaxing the need for their (real time) configuration. As far as the optimization of the IRSs is concerned, in fact, the proposed approach relies only on the statistical characterization of the environment, such as the distribution of the users' locations and the channel statistics. To this end, a two-phase optimization process is introduced, which encompasses an *offline* phase (long-term and sporadic) and an *online* (short-term and more frequent) phase. During the offline phase, the IRSs are optimized. During the online phase, the BS and the UEs are optimized. The main advantage of the proposed approach is that the statistical CSI can be either known a priori (e.g., one may know that some UEs are confined within a certain area) or can be learned occasionally during the operation of the system. In the former case, the IRSs need to be optimized once forever during the offline optimization phase, and no dedicated control channel is needed during the communication phase. In the latter case, the IRSs need to be optimized sporadically based on the *distribution* of the users' locations and the fading channels. In this case, the configuration of the IRS is not kept fixed forever but it is updated at a low rate and requires a low estimation and feedback overhead. As far as the online phase is concerned, on the other hand, the IRS is always kept fixed and is not part of the optimization process. During the online phase, the IRS is viewed as a *smart scatterer* that generates controlled multipath based on the statistical distribution of the channels and the users' locations. Hence, the IRS is implicitly embedded in the channel, and only the BS and the UEs need to be optimized by using conventional CSI acquisition and beamforming methods [29].

To substantiate the proposed approach, numerical results are illustrated for analyzing the coverage and channel rank improvement of the system sum-rate as a function of the statistical distribution of the users' locations, as well as the number of UEs and IRS. Monte Carlo simulations show that, even in the presence of moderate a priori knowledge about the scenario, the proposed approach leads to a significant performance improvement when the line-of-sight (LOS) is not available. Moreover, the numerical results indicate that increasing the number of IRSs deployed in the environment, while keeping constant the total surface area, increases the effective rank of the channel and reduces the interference at the UEs. These results make the proposed approach an appealing solution to facilitate the deployment of IRSs in wireless systems.

C. Comparison with the state of the art

Compared with the related research works discussed in Section I-B [23]-[28], the most closely related paper to ours is [24]. Therein, an IRS-assisted MISO system is considered, in which the beamforming optimization problem is split into a short-term and a long-term optimization phases. In the first phase, the IRS analog beamforming vector is assumed to be fixed and a conventional weighted minimum mean square error (WMMSE) algorithm is utilized to compute the BS digital beamforming vector with the goal of maximizing the system sum-rate. In the second phase, the IRS analog beamforming vector is optimized with the goal of maximizing the rate averaged with respect to the small-scale fading statistics. These latter statistics, together with the LOS channel coefficients, are estimated during a preceding transmission phase in which the IRS is in sensing mode and pilots and/or data symbols transmitted in the uplink and downlink are exploited for channel estimation. The optimization is performed by leveraging the close relationship between the rate and the SNR in MISO systems, and by exploiting such a relationship to derive the gradient of each user rate with respect to the IRS analog beamforming vector.

In the present paper, on the other hand, we consider a general multi-IRS and MU-MIMO communication system, in which the digital beamforming vectors at the BSs and UEs and the analog beamforming vectors at the IRSs are optimized with the goal of maximizing the average sum-rate. In addition, rather than considering a long-term optimization problem, we consider an *offline*-based approach in which only mild a priori knowledge on the network scenario is available. More specifically, we assume the a priori knowledge of only the statistical distribution of the users' locations in a given service area. This allows us to get rid of estimating the exact

LOS channel coefficients. Accordingly, the IRSs can be assumed to be nearly-passive and do not need to be equipped with any ad hoc circuits (or sensors) for estimating the CSI.

As far as the calculation of the average sum-rate is concerned, similar to [24], we resort to a Monte Carlo generation method for the channels. However, we include the generation of the users' locations as well. By capitalizing on the proposed approach, in the extreme case in which the network is static as a function of the users' locations and channel statistics, the IRSs need to be optimized only once during the offline phase. In this case, the IRSs act as passive repeaters.

As far as the optimization algorithm is concerned, we provide new contributions for the online and offline optimization phases. As for the online phase, we introduce a generalized version of the WMMSE algorithm for application to MU-MIMO systems. As for the offline phase, owing to the more involved expression of the per-user rate of MIMO systems with respect to MISO systems, we further generalize the WMMSE method for optimizing the analog beamforming vectors of the IRSs. It is worth noting that the WMMSE algorithm was recently used for optimizing IRSs [30]. Therein, however, the optimization is performed for a given instance of the channel (the instantaneous CSI is needed) and a single IRS is considered.

Finally, we emphasize that the terms offline and online optimization were recently used in [31]. Therein, however, the problem formulation and proposed approach are fundamentally different compared with ours. In [31], for example, during the online optimization phase, all the effective end-to-end channels between the transmitter and the user via each tile of the IRSs are needed. This does not apply to the proposed algorithm, since the individual knowledge of these channels is not necessary. Similar to [31], however, our proposed approach can be applied to IRSs that are made of multiple tiles, which is useful for reducing the optimization complexity.

D. *Organization of the paper*

This paper is structured as follows. In Section, II the channel and IRS physical models are introduced. In Section III, the offline and online IRS optimization problems are described and tackled. Numerical results referred to a typical indoor scenario are described in Section IV. Finally, concluding remarks are given in Section V.

II. DIRECT AND IRS-ASSISTED CHANNEL MODELS

We consider the downlink of a wireless system in which one BS equipped with M antennas serves N_u UEs that comprise L antennas each. In this network scenario, N_{IRS} IRSs are deployed

in some predefined locations for assisting the communication between the BS and the UEs. Each IRS comprises P_{IRS} nearly-passive reconfigurable scattering elements. In this setting, each IRS can be conveniently partitioned into K_t smaller tiles in order to reduce the design complexity (e.g., see [31]). In this case, the number of equivalent IRSs is $K = N_{IRS} \times K_t$, each comprising $P = P_{IRS}/K_t$ scattering elements. In the following, for ease of notation, we employ the notation K and P regardless of whether the IRSs are partitioned in tiles or not.

The objective of this section is to introduce the BS-UE and BS-IRS-UE channel models. For ease of writing, we formulate the channel models for a single UE.

A. Channel model for the direct link (BS-UE channel)

We consider a non-stationary narrow-band clustered channel model, typically used for application to millimeter wave (mmWave) communications [32], which accounts for the near-field and far-field effects of the EM propagation in the presence of large antenna arrays. We denote by f_0 the center frequency of the EM signal and by λ the corresponding wavelength. The channel is composed of a LOS component and a NLOS component that comprises a number of clustered paths, each corresponding to a macro-level scattering path. In particular, we assume that the center of each cluster $\mathbf{c}_q$ of multipath is located on an ellipse whose foci coincide with the antennas of the BS and UE. For simplicity, we assume that $\mathbf{c}_q$ are all located on the same ellipse. By denoting with $\mathbf{t}$ and $\mathbf{r}$ the center of the antenna arrays of the BS and UE, respectively, we have, by definition of ellipse, $|\mathbf{c}_q - \mathbf{t}| + |\mathbf{c}_q - \mathbf{r}| = |\mathbf{t} - \mathbf{r}|/\epsilon$, $\forall q$, ϵ being the ellipse eccentricity. Each cluster consists of a number of paths, and each path is represented by a point, which is denoted by $\mathbf{p}_{ql}$, that is randomly placed within a maximum (small) distance from the center of the cluster $\mathbf{c}_q$. Accordingly, all the paths of a given cluster reach the receiver within a limited angular spread, in agreement with typical mmWave channels. In mathematical terms, the channel complex gain can be formulated as

$$h_{mn} = X_d \beta_{mn}^{(\text{LOS})} \exp\left(-J\frac{2\pi |\mathbf{t}_m - \mathbf{r}_n|}{\lambda}\right) + \frac{\beta_{nm}^{(\text{NLOS})}}{N_p N_c} \sum_{q=1}^{N_c} \sum_{l=1}^{N_p} \alpha_{ql} \exp\left(-J2\pi\frac{|\mathbf{t}_m - \mathbf{p}_{ql}| + |\mathbf{r}_n - \mathbf{p}_{ql}|}{\lambda}\right)$$

(1)

where $\mathbf{t}_m$ denotes the position of the mth antenna-element of the BS, $\mathbf{r}_n$ denotes the position of the nth antenna-element of the UE, $\beta_{mn}^{(\text{LOS})} = \frac{\sqrt{G_T G_R}\lambda}{4\pi|\mathbf{t}_m - \mathbf{r}_n|}$ denotes the path-loss gain of the LOS, G_T denotes the antenna-element gain at the BS, G_R denotes the antenna-element gain at the UE, N_c denotes the number of clusters, N_p denotes the number of paths of each cluster,

$\alpha_{ql} \sim \mathcal{CN}(0, 1)$ denotes the random complex fading coefficient of the lth path of the qth cluster, and $\beta_{nm}^{(\text{NLOS})}$ denotes the path-loss gain of the channel in NLOS. According to [33], we have $\beta_{nm}^{(\text{NLOS})}(\text{dB}) = -\text{PL}_0 - 10\,n_e\,\log_{10} d$, where PL_0 denotes the path-loss (in dB) at the reference distance of one meter, d denotes the distance between the center of the antenna-array of the BS (**t**) and the center of the antenna-array of the UE (**r**), and n_e denotes the path-loss exponent. In particular, we consider the *ABG* channel model for the indoor office (IO) and the shopping mall (SM) NLOS scenairos at 28 GHz [33]. The main difference between the two scenarios is given by the path-loss exponents, which are equal to 3.83 and 3.21 for the IO and SM cases, respectively. Finally, the binary variable X_d allows us to model whether the LOS is present ($X_d = 1$) or is obstructed ($X_d = 0$).

B. Reflection model for the IRS

We consider an IRS that comprises P small unit cells (scattering elements), which are spaced $\lambda/2$ apart and whose area is A_c. We consider a locally-plane (with respect to the size of each unit cell)[1] incident wave, whose angle of incidence with respect to the IRS is $\Theta_{\text{inc}} = (\theta_{\text{inc}}, \phi_{\text{inc}})$, with $\phi_{\text{inc}} \in [0, 2\pi)$ and $\theta_{\text{inc}} \in [0, \pi)$ denoting the azimuth and elevation angles, respectively.

The local reflection coefficient of the pth unit cell towards the generic direction of scattering $\Theta = (\theta, \phi)$ is modeled as

$$r_p(\Theta_{\text{inc}}, \Theta) = \sqrt{F(\Theta_{\text{inc}})\,F(\Theta)}\,G_c\,b_p \tag{2}$$

where $F(\Theta)$ denotes the normalized power radiation pattern of each unit cell, which accounts for the possible non-isotropic behavior of each scattering element. As a first approximation, we assume that $F(\Theta)$ is frequency-independent within the bandwidth of interest. Also, G_c denotes the boresight gain of the unit cell, and $b_p = \rho_p \exp(J\phi_p)$ denotes the load reflection coefficient, with $\rho_p = |b_p|$ and ϕ_p being the amplitude and the phase shift introduced by the pth unit cell, respectively.[2] For analytical convenience, we collect the load reflection coefficients $b_p = \rho_p \exp(J\phi_p)$ in the vector $\mathbf{b} = [b_1, b_2, \ldots, b_P]$. It is worth mentioning that, especially for large

---

[1]The wave is assumed to be plane with respect to the area A_c, but it is spherical with respect to the size of the IRS.

[2]Depending on the specific implementation, the load reflection coefficient may have a different physical meaning [34].

IRSs, the angle of incidence Θ_{inc} can be different for each unit cell. Without loss of generality, the boresight gain is defined as $G_{\mathrm{c}} = A_{\mathrm{c}} \, 4\pi/\lambda^2$. In addition, a quite general model for $F(\Theta)$ is

$$
F(\Theta) = \begin{cases} \cos^q(\theta) & \theta \in [0, \pi/2] \,, \phi \in [0, 2\pi] \\ 0 & \text{otherwise} \end{cases}
\tag{3}
$$

which is compliant with the models used in, e.g., [17], [35], [36] for different values of the parameter q, and with the model recently derived in [31], which originates from the radar cross-section of small reflecting tiles. In the numerical results, without loss of generality, the following parameters are considered: $q = 0.57$, $A_{\mathrm{c}} = \frac{\lambda^2}{4}$ [28], [35].

C. Channel model for the indirect link (BS-IRS-UE channel)

In IRS-assisted communications, the near-field effects cannot be ignored if the size of the IRS is large and the transmission distances are short [37]. In this case, the plane wave approximation is not fulfilled anymore, and spherical waves need to be considered. Thus, the channel is non-stationary along the IRS. This implies that the angles of incidence and scattering are not sufficient to characterize the global response of the IRS, and the different transmission distances between the unit cells of the IRS and the antennas of the BS and UEs need to be taken into account.

In general, the IRSs are strategically deployed in order to ensure that they are in LOS with the BS and the UEs. In addition, they are more useful when the BS and the UEs are in NLOS [38]. Therefore, these assumptions are considered in the present paper and in the numerical results. In addition, second-order reflections between the IRS and the cluster of scatterers are ignored because they are expected not to contribute significantly due to the large path-loss experienced.

According to the IRS model in (2), the (complex) channel gain between the mth transmit antenna of the BS and the pth unit cell of the IRS can be written as

$$
s_{pm} = \frac{\sqrt{G_{\mathrm{T}} G_{\mathrm{c}} \, F(\Theta_{pm})} \lambda}{4\pi d_{pm}} \exp\left(-J \frac{2\pi d_{pm}}{\lambda}\right)
\tag{4}
$$

where d_{pm} and Θ_{pm} denote the distance between the mth transmit antenna of the BS and the pth unit cell of the IRS, and the angle from which the mth transmit antenna is viewed by the IRS, respectively. Similarly, the (complex) channel gain between the pth unit cell of the IRS and the nth receive antenna can be written as

$$
t_{np} = \frac{\sqrt{G_{\mathrm{R}} G_{\mathrm{c}} \, F(\Theta_{np})} \lambda}{4\pi d_{np}} \exp\left(-J \frac{2\pi d_{np}}{\lambda}\right)
\tag{5}
$$

where d_{np} and Θ_{np} denote the distance between the nth receive antenna and the pth unit cell of the IRS, and the angle from which the nth receive antenna is viewed by the IRS, respectively.

It is important to note that, differently from (1), the NLOS component of the channel is not present in (4) and (5), since it is related to the second- and higher-order reflections, which are typically negligible, as mentioned in previous text, and are hence ignored in this paper.

For generality, we consider IRSs that are designed obeying a local constraint (LC) and a global constraint (GC) design criteria [39]. In the first case, we assume $\rho_p \leq 1 \; \forall p$, i.e., each unit cell of the IRS scatters an amount of power that is smaller than or equal to the impinging power. In the second case, we assume $\sum_p \rho_p^2 \leq P$, i.e., the IRS as a whole scatters an amount of power that is smaller than or equal to the total impinging power. Also, we assume that the IRSs are non-lossy and that the phase shifts applied by the unit cells belong to a finite set, i.e., $b_p \in \mathcal{S}_{N_b} = \left\{ 1, e^{\frac{j2\pi}{M_l}}, e^{\frac{j4\pi}{M_l}}, \ldots, e^{\frac{j2(M_l-1)\pi}{M_l}} \right\}$, where $M_l = 2^{N_b}$ is the number of quantization levels.

III. IRS Optimization Based on Statistical CSI

In this section, we formulate the problem statement and introduce the proposed (offline and online) optimization algorithms for sum-rate maximization.

A. Problem formulation

Let $\mathbf{s}_j = \left[s_j(1), s_j(2), \ldots, s_j(L) \right]^T$ be the complex vector comprising the L symbols[3] trans-mitted by the BS to the jth UE (also referred to as the jth stream) and $\mathbf{V}_j \in \mathbb{C}^{M \times L}$ be the corresponding precoding matrix. The transmitted vector can be formulated as

$$\mathbf{x}_j = \mathbf{V}_j \mathbf{s}_j \quad \in \mathbb{C}^{M \times 1} . \tag{6}$$

The information symbols are assumed to be zero-mean and independent, identically distributed (i.i.d.) random variables (RVs), i.e., $\mathbb{E}\left[\mathbf{s}_j \mathbf{s}_j^H \right] = \mathbf{I}_L$ and $\mathbb{E}\left[\mathbf{s}_j \mathbf{s}_i^H \right] = \mathbf{0}_L$ for $j \neq i$.

The signal received at the kth IRS can be formulated as

$$\mathbf{r}_{k,j} = \mathbf{S}_k \mathbf{x}_j \quad \in \mathbb{C}^{P \times 1} \tag{7}$$

where $\mathbf{S}_k \in \mathbb{C}^{P \times M}$ is the channel matrix between the BS and the kth IRS, whose elements are modeled according to (4). The signal reflected by the IRS is

$$\mathbf{r}'_{k,j} = \mathbf{B}_k \mathbf{r}_{k,j} \quad \in \mathbb{C}^{P \times 1} \tag{8}$$

---

[3]We assume that the number of symbols is equal to the number of receive antennas.

where $\mathbf{B}_k = \mathrm{diag}\,(\mathbf{b}_k) \in \mathbb{C}^{P \times P}$ is the diagonal matrix whose elements are the load reflection coefficients $\mathbf{b}_k \in \mathbb{C}^{P \times 1}$ introduced in Section II-B.

The signal received at the ith UE can be expressed as

$$\mathbf{z}_{i,k,j} = \mathbf{T}_{i,k}\mathbf{r}'_{k,j} = \mathbf{T}_{i,k}\mathbf{B}_k\mathbf{S}_k\mathbf{x}_j = \tilde{\mathbf{H}}_{i,k}\,(\mathbf{b}_k)\,\mathbf{x}_j \quad \in \mathbb{C}^{L \times 1} \tag{9}$$

where $\mathbf{T}_{i,k} \in \mathbb{C}^{L \times P}$ is the channel matrix between the kth IRS and the ith receiver, whose elements are modeled according to (5), and $\tilde{\mathbf{H}}_{i,k}\,(\mathbf{b}_k) = \mathbf{T}_{i,k}\mathbf{B}_k\mathbf{S}_k$.

In addition, the direct signal emitted by the BS and received by the ith UE is

$$\mathbf{w}_{i,j} = \bar{\mathbf{H}}_i\,\mathbf{x}_j \quad \in \mathbb{C}^{L \times 1} \tag{10}$$

where $\bar{\mathbf{H}}_i$ is the channel matrix between the BS and the ith UE.

Thus, considering only the jth stream and the kth IRS, the received signal at the ith UE is

$$\mathbf{y}_{i,k,j} = \mathbf{z}_{i,k,j} + \mathbf{w}_{i,j} + \mathbf{n}_i = \left(\bar{\mathbf{H}}_i + \tilde{\mathbf{H}}_{i,k}\,(\mathbf{b}_k)\right)\mathbf{x}_j + \mathbf{n}_i \quad \in \mathbb{C}^{L \times 1} \tag{11}$$

where $\mathbf{n}_i \in \mathbb{C}^{L \times 1}$ denotes the additive white Gaussian noise with distribution $\mathcal{CN}\left(0, \sigma_i^2\mathbf{I}_L\right)$.

Let $\mathbf{H}_i\,(\mathcal{B}) = \bar{\mathbf{H}}_i + \sum_{k=1}^{K} \tilde{\mathbf{H}}_{i,k}\,(\mathbf{b}_k)$ denote the total channel between the BS and the ith UE, where $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_K\}$ is the set of vectors containing the reflection coefficients of the K IRSs (analog beamforming vectors). The signal received at the ith UE in the presence of N_u concurrent transmitted streams and K IRSs is

$$\mathbf{y}_i = \sum_{j=1}^{N_u}\left(\sum_{k=1}^{K}\mathbf{z}_{i,k,j} + \mathbf{w}_{i,j}\right) + \mathbf{n}_i = \mathbf{H}_i\,(\mathcal{B})\,\mathbf{x}_i + \mathbf{H}_i\,(\mathcal{B})\left(\sum_{\substack{j=1 \\ j \neq i}}^{N_u}\mathbf{x}_j\right) + \mathbf{n}_i \quad \in \mathbb{C}^{L \times 1}. \tag{12}$$

Based on the MIMO interference channel in (12), the AR of the ith UE is [40]

$$R_i(\mathcal{V}, \mathcal{B}) = \log\det\left(\mathbf{I}_L + \mathbf{V}_i^H\mathbf{H}_i^H\,(\mathcal{B})\,\bar{\mathbf{J}}_i^{-1}\mathbf{H}_i\,(\mathcal{B})\mathbf{V}_i\right) \tag{13}$$

where $\bar{\mathbf{J}}_i = \sum_{j=1, j \neq i}^{N_u}\mathbf{H}_i\,(\mathcal{B})\mathbf{V}_j\mathbf{V}_j^H\mathbf{H}_i^H\,(\mathcal{B}) + \sigma^2\mathbf{I}_L$ is the interference-plus-noise covariance matrix and $\mathcal{V} = \left\{\mathbf{V}_1, \mathbf{V}_2, \ldots, \mathbf{V}_{N_u}\right\}$ is the set of precoding matrices of all the N_u UEs.

In the next two sections, we introduce the proposed approach for optimizing the digital beamforming vectors of the BS and the UEs (which occurs during the online phase), as well as the analog beamforming vectors of the IRSs (which occurs during the offline phase). The online phase is subsequent to the offline phase, but it is introduced first for ease of description.

B. BS and UEs beamforming optimization – Online phase

During the online phase, the analog beamforming vectors $\mathcal{B}$ of the IRSs are kept constant and are set equal to those computed during the offline phase described next. Hence, for notational simplicity, the dependence on $\mathcal{B}$ of the rate R_i in (13) and the channel matrices is omitted in this section. This implies that the impact of the IRSs is incorporated into the channels in a transparent manner and the IRSs operate as smart scatterers. Therefore, the system boils down to a classical MIMO broadcast channel, which can be optimized by relaying on well established methods and protocols for estimating the CSI of the BS-UE links. This case study has been investigated extensively and is, therefore, briefly discussed in this section. In particular, we propose to utilize the iterative WMMSE algorithm for optimizing the digital beamforming vectors of the BS and the UEs [40]. The WMMSE algorithm is chosen since it provides the basis for optimizing the analog beamforming vectors of the IRSs, which is described in the next section. In preparation for the next section, the main features of the WMMSE algorithm are briefly summarized.

In the online phase, the sum-rate maximization problem can be formulated as

$$\max_{\mathcal{V}} \sum_{i=1}^{N_u} \alpha_i R_i \left(\mathcal{V} \right) \tag{14}$$

$$\text{subject to} \quad \text{tr} \left(\mathbf{V}_i \mathbf{V}_i^H \right) \leq P_i, \quad i = 1, 2, \ldots, N_u \tag{14.a}$$

where P_i is the power budget of the *i*th UE and $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \ldots, \alpha_{N_u}]$ is a set of weights chosen to guarantee a given degree of fairness among the UEs, e.g., by assigning higher weights to UEs with weaker channels. The readers are referred to [41] for further information on $\boldsymbol{\alpha}$.

The WMMSE algorithm exploits the close relationship between the SINR and the mean square error (MSE), in order to find a locally optimal solution to the sum-rate problem in (14) [40]. To elaborate, let us introduce the MSE matrix $\mathbf{E}_i \left(\mathcal{V}, \mathbf{G}_i \right)$, for $i = 1, 2, \ldots, N_u$, as

$$\mathbf{E}_i \left(\mathcal{V}, \mathbf{G}_i \right) = \mathbb{E} \left\{ (\mathbf{s}_i - \tilde{\mathbf{s}}_i) (\mathbf{s}_i - \tilde{\mathbf{s}}_i)^H \right\} \quad \in \mathbb{C}^{L \times L} \tag{15}$$

where $\tilde{\mathbf{s}}_i = \mathbf{G}_i^H \mathbf{y}_i$ and $\mathbf{G}_i$ is the linear decoding matrix of the *i*th UE.

Problem (14) can be reformulated as the following weighted MSE minimization problem

$$\min_{\mathcal{V}, \mathcal{W}, \mathcal{G}} \sum_{i=1}^{N_u} \alpha_i \left\{ \text{tr} \left[\mathbf{W}_i \mathbf{E}_i \left(\mathcal{V}, \mathbf{G}_i \right) \right] - \log \det \left(\mathbf{W}_i \right) \right\} \tag{16}$$

$$\text{subject to} \quad \text{tr} \left(\mathbf{V}_i \mathbf{V}_i^H \right) \leq P_i, \quad i = 1, 2, \ldots, N_u$$

where $\mathbf{W}_i \succeq 0$ is the matrix of weights for the MSE of *i*th UE, and $\mathcal{W} = \left\{ \mathbf{W}_1, \mathbf{W}_2, \ldots, \mathbf{W}_{N_u} \right\}$ and $\mathcal{G} = \left\{ \mathbf{G}_1, \mathbf{G}_2, \ldots, \mathbf{G}_{N_u} \right\}$ are the sets of all weight and receive filter matrices, respectively.

The equivalence between problem (14) and (16) follows by recalling the relation between the MMSE covariance $\hat{\mathbf{E}}_i(\mathcal{V}) = \min_{\mathbf{G}_i} \mathbf{E}_i(\mathcal{V}, \mathbf{G}_i)$ and the rate R_i, i.e., $R_i = \log\det\left(\hat{\mathbf{E}}_i^{-1}(\mathcal{V})\right)$ and by the fact that the optimal solution of (16) is $\mathbf{W}_i = \hat{\mathbf{E}}_i^{-1}(\mathcal{V})$. Further details can be found in [40].

Problem (16) is non-convex. If all the optimization variables are fixed except one, however, it is a convex optimization problem in the remaining variables. Accordingly, the weighted MSE minimization problem in (16) can be solved by using an iterative block coordinate descent (BCD) algorithm [42]. To elaborate, let us denote by $\mathbf{G}_i^{(q+1)}$, $\mathbf{W}_i^{(q+1)}$ and $\mathbf{V}_i^{(q+1)}$ the optimization variables after the $(q+1)$th iteration. Then, the receive filter matrix is [40]

$$\mathbf{G}_i^{(q+1)} = \left(\mathbf{J}_i^{(q)}\right)^{-1}\mathbf{H}_i\mathbf{V}_i^{(q)} \tag{17}$$

where $\mathbf{J}_i^{(q)} = \sum_{j=1}^{N_u} \mathbf{H}_i\mathbf{V}_j^{(q)}\left(\mathbf{V}_j^{(q)}\right)^H\mathbf{H}_i^H + \sigma_i^2\mathbf{I}_L$, the weight matrix $\mathbf{W}_i^{(q+1)}$ is

$$\mathbf{W}_i^{(q+1)} = \left(\mathbf{E}_i^{(q+1)}\right)^{-1} \tag{18}$$

and the precoding filter at the BS $\mathbf{V}_{i,q+1}$ is

$$\mathbf{V}_i^{(q+1)} = \alpha_i\left(\mathbf{K}^{(q+1)} + \mu_i\mathbf{I}_M\right)^{-1}\mathbf{H}_i^H\mathbf{G}_i^{(q+1)}\mathbf{W}_i^{(q+1)} \tag{19}$$

where $\mathbf{K}^{(q+1)} = \sum_{j=1}^{N_u} \alpha_j\mathbf{H}_j^H\mathbf{G}_j^{(q+1)}\mathbf{W}_j^{(q+1)}\mathbf{G}_j^{(q+1)}\mathbf{H}_j$ and the Lagrange multiplier μ_i involved in the optimization problem is chosen so that the power constraint in (16) for the ith UE is fulfilled.

C. IRS beamforming optimization – Offline phase

The offline phase encompasses the optimization of the IRSs. In contrast to the online phase that relies on instantaneous CSI, the offline phase relies only on channel long-term statistics, e.g., the distribution of the locations of the UEs in the given area. Therefore, it is executed only sporadically. In the extreme case that the IRSs operate as repeaters, it is executed only once.

The objective of the offline phase is to optimize the analog beamforming vectors of the IRSs, $\mathcal{B}$, that maximize the (statistical) average weighted sum-rate of the system. In particular, the average is computed with respect to all random phenomena that determine the channel matrices.

To elaborate, let us denote by Ω the set of random vectors ω that determine the channel status, including the locations of the UEs, the position of the antennas, the distribution of the clusters of multipath for each link, and the complex channel gains. Accordingly, the total channel matrices

depend on ω and can be formulated as $\mathbf{H}_i(\omega, \mathcal{B})$. Denoting by $f_\Omega(\omega)$ the joint probability density function of Ω, the optimization problem for computing $\mathcal{B}$ can be formulated as

$$\max_{\mathcal{B}} \mathbb{E}_\omega \left[\max_{\mathcal{V}(\omega)} R_{tot}(\mathcal{B}, \mathcal{V}(\omega)) \right] = \max_{\mathcal{B}} \int_{\omega \in \Omega} \max_{\mathcal{V}(\omega)} \sum_{i=1}^{N_u} \alpha_i R_i(\mathcal{V}(\omega), \mathcal{B}) \, f_\Omega(\omega) \, d\omega \tag{20}$$

subject to

$$\mathrm{tr}\left(\mathbf{V}_i(\omega) \ \mathbf{V}_i^H(\omega) \right) \le P_i, \quad \forall \omega, \quad i = 1, 2, \ldots, N_u \tag{20.a}$$

$$\mathrm{tr}\left(\mathbf{b}_k \mathbf{b}_k^H \right) \le 1, \quad k = 1, 2, \ldots, K \qquad \text{(GC case)} \tag{20.b.1}$$

or

$$b_{kp} \in \mathcal{S}_{N_b}, \quad k = 1, 2, \ldots, K \quad p = 1, \ldots, P \quad (\text{LC}_{N_b} \text{ case}) \tag{20.b.2}$$

where the dependence of $\mathcal{V}$ with respect to ω is made explicit in order to emphasize that $\mathbf{V}_i$ depends on ω. By direct inspection, we observe that GC in (20.b.1) is a convex constraint and LC_{N_b} in (20.b.2) is a non-convex constraint. Therefore, the LC setup is more difficult to study. In this section, therefore, we restrict the analysis to the GC setup. After solving the GC case study, a (suboptimal but low complexity) solution for the LC case study can be obtained by recalling that a local element-wise projection onto the unit circle of the solution obtained for the GC setup corresponds to the minimum distance projection on the feasibility set of the LC setup [29], [43]. Therefore, we solve the optimization problem in (20) under the GC constraint and we then obtain the solution under the LC constraint by applying an element-wise scaling to the solution of the GC setup. The performance offered by this approach is analyzed in the numerical results.

Even though the GC constraint is convex, problem (20) is non-convex and is, therefore, difficult to solve. To this end, the sum-rate maximization problem in (20) can be reformulated as a weighted MSE minimization problem. In particular, we have

$$\min_{\mathcal{B}} \int_{\omega \in \Omega} \min_{\mathcal{V}(\omega), \mathcal{W}(\omega), \mathcal{G}(\omega)} \sum_{i=1}^{N_u} \alpha_i \Big\{ \mathrm{tr}\left[\mathbf{W}_i(\omega) \, \mathbf{E}_i(\mathcal{V}(\omega), \mathcal{B}, \mathbf{G}_i(\omega)) \right]$$

$$- \log \det(\mathbf{W}_i(\omega)) \Big\} f_\Omega(\omega) \, d\omega \tag{21}$$

subject to

$$\mathrm{tr}\left(\mathbf{V}_i(\omega) \, \mathbf{V}_i^H(\omega) \right) \le P_i, \quad \forall \omega, \quad i = 1, 2, \ldots, N_u \tag{22}$$

$$\mathrm{tr}\left(\mathbf{b}_k \mathbf{b}_k^H \right) \le 1, \quad k = 1, 2, \ldots, K, \tag{23}$$

where $\mathbf{E}_i\left(\mathcal{V}\left(\omega\right),\mathcal{B}\left(\omega\right),\mathbf{G}_i\left(\omega\right)\right)$ is the $L\times L$ MSE matrix defined in (15).

From the statistical properties of the information vectors $\mathbf{s}_i$, we can elaborate (15) as

$$
\begin{aligned}
\mathbf{E}_i\left(\mathcal{V}\left(\omega\right),\mathcal{B},\mathbf{G}_i\left(\omega\right)\right) &= \mathbf{I}_L + \left(\mathbf{G}_i^H\left(\omega\right)\mathbf{H}_i\left(\omega,\mathcal{B}\right)\mathbf{V}_i\left(\omega\right)\right)\left(\mathbf{G}_i^H\left(\omega\right)\mathbf{H}_i\left(\omega,\mathcal{B}\right)\mathbf{V}_i\left(\omega\right)\right)^H \\
&\quad - \left(\mathbf{G}_i^H\left(\omega\right)\mathbf{H}_i\left(\omega,\mathcal{B}\right)\mathbf{V}_i\left(\omega\right)\right) - \left(\mathbf{G}_i^H\left(\omega\right)\mathbf{H}_i\left(\omega,\mathcal{B}\right)\mathbf{V}_i\left(\omega\right)\right)^H \\
&\quad + \sum_{\substack{j=1 \\ j\neq i}}^{N_u}\left(\mathbf{G}_i^H\left(\omega\right)\mathbf{H}_i\left(\omega,\mathcal{B}\right)\mathbf{V}_j\left(\omega\right)\right)\left(\mathbf{G}_i^H\mathbf{H}_i\left(\omega,\mathcal{B}\right)\mathbf{V}_j\left(\omega\right)\right)^H + \sigma_i^2\mathbf{G}_i^H\left(\omega\right)\mathbf{G}_i\left(\omega\right) \quad \in\mathbb{C}^{L\times L}
\end{aligned}
\tag{24}
$$

The optimization problem in (21) is constituted by an inner and an outer optimization sub-problems. To tackle both sub-problems, we employ an iterative BCD algorithm. More precisely, let us denote by $\mathcal{B}^{(q)}$, $\mathbf{G}_i^{(q)}\left(\omega\right)$, $\mathbf{W}_i^{(q)}\left(\omega\right)$ and $\mathbf{V}_i^{(q)}\left(\omega\right)$ the optimization variables after the qth iteration. As far as the inner optimization sub-problem is concerned, the set of variables $\mathcal{B}_q$ is fixed as a function of ω. Thus, the inner optimization sub-problem in (21) is formally the same as the optimization problem solved during the online phase. This implies that the single-step updates $\mathbf{G}_i^{(q+1)}\left(\omega\right)$, $\mathbf{W}_i^{(q+1)}\left(\omega\right)$ and $\mathbf{V}_i^{(q+1)}\left(\omega\right)$ can be computed from (17), (18), and (19).

As far as the outer optimization sub-problem in (21) is concerned, on the other hand, the solution is not straightforward. More precisely, the outer optimization sub-problem in (21) encompasses the computation of the optimal set of variables $\mathcal{B}$ while all the other optimization variables are kept fixed. Next, we introduce an iterative algorithm to solve this problem. For ease of notation, we drop the index of the iterations q and use the notation $\mathbf{G}_i\left(\omega\right)$, $\mathbf{W}_i\left(\omega\right)$, $\mathbf{V}_i\left(\omega\right)$.

To start with, we elaborate the expression of the MSE in (24). By recalling that $\mathbf{H}_i\left(\omega,\mathcal{B}\right) = \bar{\mathbf{H}}_i\left(\omega\right) + \sum_{k=1}^K \tilde{\mathbf{H}}_{i,k}\left(\omega,\mathbf{b}_k\right)$, we obtain

$$
\begin{aligned}
\mathbf{E}_i\left(\mathcal{V}\left(\omega\right),\mathcal{B},\mathbf{G}_i\left(\omega\right)\right) = & \\
& \sum_{j=1}^{N_u}\left[\mathbf{G}_i^H\left(\omega\right)\sum_{k=1}^K\tilde{\mathbf{H}}_{i,k}\left(\omega,\mathbf{b}_k\right)\mathbf{V}_j\left(\omega\right)\right]\left[\mathbf{G}_i^H\left(\omega\right)\sum_{k=1}^K\tilde{\mathbf{H}}_{i,k}\left(\omega,\mathbf{b}_k\right)\mathbf{V}_j\left(\omega\right)\right]^H \\
& + 2\mathfrak{R}\left[\sum_{j=1}^{N_u}\mathbf{G}_i^H\left(\omega\right)\left(\sum_{k=1}^K\tilde{\mathbf{H}}_{i,k}\left(\omega,\mathbf{b}_k\right)\right)\mathbf{V}_j\left(\omega\right)\mathbf{V}_j^H\left(\omega\right)\bar{\mathbf{H}}_i^H\left(\omega\right)\mathbf{G}_i\left(\omega\right)\right] \\
& - 2\mathfrak{R}\left[\mathbf{G}_i^H\left(\omega\right)\left(\sum_{k=1}^K\tilde{\mathbf{H}}_{i,k}\left(\omega,\mathbf{b}_k\right)\right)\mathbf{V}_i\left(\omega\right)\right] + \Upsilon \quad \in\mathbb{C}^{L\times L}
\end{aligned}
\tag{25}
$$

where $2\mathfrak{R}\left(\mathbf{A}\right) = \mathbf{A} + \mathbf{A}^H$, Υ is a constant, and $\mathfrak{R}(x)$ denotes the real part of x.

Let us consider the WMMSE minimization with respect to the load reflection coefficient, $\mathbf{b}_m$, of the mth IRS. To this end, we introduce the shorthand notation

$$\mathbf{A}_{i,m}(\omega) = \mathbf{G}_i^H(\omega)\,\mathbf{T}_{i,m}(\omega) \quad \in \mathbb{C}^{L\times P}$$

$$\mathbf{C}_{m,j}(\omega) = \mathbf{S}_m(\omega)\,\mathbf{V}_j(\omega) \quad \in \mathbb{C}^{P\times L}$$

$$\mathbf{D}_{i,m,j}(\omega) = \mathbf{S}_m(\omega)\,\mathbf{V}_j(\omega)\,\bar{\mathbf{H}}_i^H\mathbf{G}_i(\omega) \quad \in \mathbb{C}^{P\times L}$$

$$\mathbf{F}_{i,m,j}(\omega) = \mathbf{S}_m(\omega)\,\mathbf{V}_j(\omega)\left(\mathbf{G}_i^H(\omega)\sum_{\substack{k=1\\k\neq m}}^{K}\tilde{\mathbf{H}}_{i,k}(\omega,\mathbf{b}_k)\,\mathbf{V}_j(\omega)\right) \quad \in \mathbb{C}^{P\times L}. \tag{26}$$

From (9), (25) and (26), we can derive the contribution of $\mathbf{b}_m$ to the MSE as follows

$$\mathbf{E}_i^{(m)}(\mathcal{V}(\omega),\mathcal{B},\mathbf{G}_i(\omega)) = \sum_{j=1}^{N_u}\left[\mathbf{A}_{i,m}(\omega)\,\mathbf{B}_m\mathbf{C}_{m,j}(\omega)\right]\left[\mathbf{A}_{i,m}(\omega)\,\mathbf{B}_m\mathbf{C}_{m,j}(\omega)\right]^H$$

$$+ 2\Re\left[\sum_{j=1}^{N_u}\mathbf{A}_{i,m}(\omega)\,\mathbf{B}_m\mathbf{D}_{i,m,j}(\omega)\right] + 2\Re\left[\sum_{j=1}^{N_u}\mathbf{A}_{i,m}(\omega)\,\mathbf{B}_m\mathbf{F}_{i,m,j}(\omega)\right]$$

$$- 2\Re\left[\mathbf{A}_{i,m}(\omega)\,\mathbf{B}_m\mathbf{C}_{m,j}(\omega)\right] \quad \in \mathbb{C}^{L\times L}. \tag{27}$$

For ease of writing, we introduce the compact indexing notation $\mathbf{Z}^{\{n:m,l:p\}}$, which yields the submatrix extracted from the nth to the mth rows and from the lth to the pth columns of $\mathbf{Z}$. Furthermore, let us consider the mapping $\mathbf{Q}(\mathbf{Q}_1,\mathbf{Q}_2)$ between the matrices $\mathbf{Q}_1 \in \mathbb{C}^{L\times P}$ and $\mathbf{Q}_2 \in \mathbb{C}^{P\times L}$ and the matrices $\mathbf{Q} \in \mathbb{C}^{L\times PL}$ and $\mathbf{\Gamma}_m \in \mathbb{C}^{PL\times L}$ defined as

$$\begin{aligned}\mathbf{Q}^{\{l,(m-1)P+1:mP\}}(\mathbf{Q}_1,\mathbf{Q}_2) &= \mathbf{Q}_1^{\{l,:\}}\odot\left(\mathbf{Q}_2^{\{:,m\}}\right)^T \quad \text{for } l = 1,2,\ldots,L,\ m = 1,2,\ldots,L\\ \mathbf{\Gamma}_m^{\{(l-1)P+1:lP,l\}} &= \mathbf{b}_m \quad\quad\quad\quad\quad \text{for } l = 1,2,\ldots,L\end{aligned} \tag{28}$$

where $()^T$ denotes the transpose operation and $\odot$ denotes the Hadamard (element-wise) product.

From (28), we obtain the identity

$$\mathbf{Q}_1\mathbf{B}_m\mathbf{Q}_2 = \mathbf{Q}(\mathbf{Q}_1,\mathbf{Q}_2)\,\mathbf{\Gamma}_m \tag{29}$$

from which (27) can be reformulated as

$$
\mathbf{E}_i^{(m)} \left(\mathcal{V} \left(\omega \right), \mathcal{B}, \mathbf{G}_i \left(\omega \right) \right) =
$$

$$
\sum_{j=1}^{N_u} \left[\mathbf{Q} \left(\mathbf{A}_{i,m} \left(\omega \right), \mathbf{C}_{m,j} \left(\omega \right) \right) \mathbf{\Gamma}_m \right] \left[\mathbf{Q} \left(\mathbf{A}_{i,m} \left(\omega \right), \mathbf{C}_{m,j} \left(\omega \right) \right) \mathbf{\Gamma}_m \right]^H
$$

$$
+ 2 \Re \left[\sum_{j=1}^{N_u} \mathbf{Q} \left(\mathbf{A}_{i,m} \left(\omega \right), \mathbf{D}_{i,m,j} \left(\omega \right) \right) \mathbf{\Gamma}_m \right] + 2 \Re \left[\sum_{j=1}^{N_u} \mathbf{Q} \left(\mathbf{A}_{i,m} \left(\omega \right), \mathbf{F}_{i,m,j} \left(\omega \right) \right) \mathbf{\Gamma}_m \right]
$$

$$
- 2 \Re \left[\mathbf{Q} \left(\mathbf{A}_{i,m} \left(\omega \right), \mathbf{C}_{m,j} \left(\omega \right) \right) \mathbf{\Gamma}_m \right] \quad \in \mathbb{C}^{L \times L} . \tag{30}
$$

From mathematical convenience, we introduce the notation

$$
\tilde{\mathbf{M}}_m \left(\omega \right) = \sum_{i=1}^{N_u} \alpha_i \mathbf{W}_i \left(\omega \right) \sum_{j=1}^{N_u} \mathbf{Q}^H \left(\mathbf{A}_{i,m} \left(\omega \right), \mathbf{C}_{m,j} \left(\omega \right) \right) \mathbf{Q} \left(\mathbf{A}_{i,m} \left(\omega \right), \mathbf{C}_{m,j} \left(\omega \right) \right) \quad \in \mathbb{C}^{PL \times PL}
$$

$$
\tilde{\mathbf{U}}_m^H \left(\omega \right) = \sum_{i=1}^{N_u} \alpha_i \mathbf{W}_i \left(\omega \right) \sum_{j=1}^{N_u} \mathbf{Q} \left(\mathbf{A}_{i,m} \left(\omega \right), \mathbf{C}_{m,j} \left(\omega \right) \right) - \mathbf{Q} \left(\mathbf{A}_{i,m} \left(\omega \right), \mathbf{D}_{i,m,j} \left(\omega \right) \right)
$$

$$
+ \mathbf{Q} \left(\mathbf{A}_{i,m} \left(\omega \right), \mathbf{F}_{i,m,j} \left(\omega \right) \right) \in \mathbb{C}^{PL \times L} \tag{31}
$$

with $\mathbf{M}_m \left(\omega \right) \in \mathbb{C}^{P \times P}$ and $\mathbf{u}_m \left(\omega \right) \in \mathbb{C}^{P \times 1}$ defined as

$$
\begin{aligned}
\mathbf{M}_m \left(\omega \right) &= \sum_{l=1}^{L} \tilde{\mathbf{M}}_m \left(\omega \right)^{\{(l-1)P+1:lP,(l-1)P+1:lP\}} &&\text{for } l = 1, 2, \ldots, L \\
\mathbf{u}_m \left(\omega \right) &= \sum_{l=1}^{L} \tilde{\mathbf{U}}_m \left(\omega \right)^{\{(l-1)P+1:lP,l\}} &&\text{for } l = 1, 2, \ldots, L .
\end{aligned} \tag{32}
$$

Based on these definitions, a closed-form expression for the gradient of the WMMSE in (21) with respect to $\mathbf{b}_m$ is

$$
\nabla_{\mathbf{b}_m} \left[\int_{\omega \in \Omega} \sum_{i=1}^{N_u} \alpha_i \operatorname{tr} \left(\mathbf{W}_i \left(\omega \right) \mathbf{E}_i \left(\mathcal{V} \left(\omega \right), \mathcal{B}, \mathbf{G}_i \left(\omega \right) \right) \right) f_\Omega \left(\omega \right) \right]
$$

$$
= 2 \left(\int_{\omega \in \Omega} \mathbf{M}_m \left(\omega \right) f_\Omega \left(\omega \right) \right) \mathbf{b}_m - 2 \int_{\omega \in \Omega} \mathbf{u}_m \left(\omega \right) f_\Omega \left(\omega \right) . \tag{33}
$$

Therefore, the analog beamforming vectors of the IRSs at the $(q+1)$th iteration, $\mathbf{b}_m^{(q+1)}$, can be formulated in analytical form. More precisely, let $\bar{\mathbf{M}}_m^{(q)}$ and $\bar{\mathbf{u}}_m^{(q)}$ denote the statistical expectation, with respect to the random variables ω, of $\mathbf{M}_m^{(q)} \left(\omega \right)$ and $\mathbf{u}_m^{(q)} \left(\omega \right)$ evaluated at the qth iteration

$$
\bar{\mathbf{M}}_m^{(q)} = \int_{\omega \in \Omega} \mathbf{M}_m^{(q)} \left(\omega \right) f_\Omega \left(\omega \right) \tag{34}
$$

$$
\bar{\mathbf{u}}_m^{(q)} = \int_{\omega \in \Omega} \mathbf{u}_m^{(q)} \left(\omega \right) f_\Omega \left(\omega \right) .
$$

We eventually obtain

$$\mathbf{b}_m^{(q+1)} = \left(\bar{\mathbf{M}}_m^{(q)} + \mu_m \right)^{-1} \bar{\mathbf{u}}_m^{(q)} \tag{35}$$

where the Lagrange multipliers μ_m is chosen so that the GC constraint in (21) is fulfilled.

The proposed approach requires the analytical expression of the probability density function $f_\Omega(\omega)$. If $f_\Omega(\omega)$ is available in closed-form, then the integrals in (34) can be computed numerically. In general, however, $f_\Omega(\omega)$ is seldom available in closed-form because of the large number of optimization variables involved. In this latter case, the typical approach used to overcome this issue lies in using Monte Carlo methods based on a repeated random sampling of all the optimization variables. The proposed algorithm is summarized in Algorithm 1, where $\{\omega_n\}$, for $n = 1, \ldots, N_s$, denote the N_s random samples of the Monte Carlo method.

Algorithm 1: Proposed algorithm for (offline) sum-rate maximization

Initialize:

Generate an initial precoding matrix $\mathbf{V}^{(0)}$ and IRS analog beamforming vectors $\mathbf{b}_k^{(1)}$, $k = 1, \ldots, K$, that fulfill the power constraint;

Set an arbitrarily small value ϵ;

$q \leftarrow 1, \Delta \leftarrow 1$;

Generate the Monte Carlo samples ω_n, $\forall n = 1, 2, \ldots, N_s$, from $f_\Omega(\omega)$

while $\Delta > \epsilon$ **do**

    for $n = 1, \ldots, N_s$ **do**

        for $i = 1, \ldots, N_u$ **do**

            Compute $\mathbf{G}_i^{(q)}(\omega_n)$, $\mathbf{W}_i^{(q)}(\omega_n)$ and $\mathbf{V}_i^{(q)}(\omega_n)$ according to (17), (18) and (19);

        for $k = 1, \ldots, K$ **do**

            $\bar{\mathbf{M}}_k^{(q)} \leftarrow 0$

            $\bar{\mathbf{u}}_k^{(q)} \leftarrow 0$

            Compute $\mathbf{M}_k^{(q)}(\omega_n)$ and $\mathbf{u}_k^{(q)}(\omega_n)$ according to (32);

            $\bar{\mathbf{M}}_k^{(q)} \leftarrow \bar{\mathbf{M}}_k^{(q)} + \mathbf{M}_k^{(q)}(\omega_n)$;

            $\bar{\mathbf{u}}_k^{(q)} \leftarrow \bar{\mathbf{u}}_k^{(q)} + \mathbf{u}_k^{(q)}(\omega_n)$;

    for $k = 1, \ldots, K$ **do**

        Compute $\mathbf{b}_k^{(q+1)}$ according to (35).

    $\Delta \leftarrow \|\mathbf{b}^{(q+1)} - \mathbf{b}^{(q)}\|$;

    $q \leftarrow q + 1$

D. Convergence of the proposed offline optimization algorithm

The proposed offline optimization algorithm minimizes the average weighted MSE. In particular, all the optimization variables are updated simultaneously. From [44], it follows that the

proposed algorithm is guaranteed to converge to a stationary point of (21). An important property about the convergence of Algorithm 1 is stated in the following theorem.

Theorem 1 *Any stationary point of* (21) *is also a stationary point of* (20).

Proof: Let $\hat{\mathcal{B}}$, $\hat{\mathcal{V}}(\omega)$, $\hat{\mathcal{W}}(\omega)$, and $\hat{\mathcal{G}}(\omega)$ denote the optimization variables of Algorithm 1 at convergence. Since the WMMSE is guaranteed to converge to a stationary point of (21), the unconstrained variable set $\mathcal{G}(\omega)$ fulfills the MMSE conditions $\mathbf{E}_i\left(\hat{\mathcal{V}}(\omega), \hat{\mathcal{B}}, \hat{\mathbf{G}}_i(\omega)\right) = \hat{\mathbf{E}}_i\left(\hat{\mathcal{V}}(\omega), \hat{\mathcal{B}}\right)$ and $\hat{\mathbf{W}}_i(\omega) = \hat{\mathbf{E}}_i^{-1}\left(\hat{\mathcal{V}}(\omega), \hat{\mathcal{B}}\right)$. Let us define

$$\Theta_1\left(\mathcal{B}, \mathcal{V}(\omega), \mathcal{G}(\omega), \mathcal{W}(\omega)\right) = \int_{\omega \in \Omega} \sum_{i=1}^{N_u} \alpha_i \Big\{ \mathrm{tr}\left[\mathbf{W}_i(\omega) \mathbf{E}_i\left(\mathcal{V}(\omega), \mathcal{B}, \mathbf{G}_i(\omega)\right)\right]$$

$$- \log \det\left(\mathbf{W}_i(\omega)\right) \Big\} f_\Omega(\omega)\, d\omega \qquad (36)$$

$$\Theta_2\left(\mathcal{B}, \mathcal{V}(\omega)\right) = - \int_{\omega \in \Omega} \sum_{i=1}^{N_u} \alpha_i R_i\left(\mathcal{V}(\omega), \mathcal{B}\right)\, f_\Omega(\omega)\, d\omega. \qquad (37)$$

Since the optimization problems in (21) and (20) have the same constraints in terms of the optimization variables $\mathcal{B}$ and $\mathcal{V}(\omega)$, Theorem 1 can be proved by showing that Θ_1 and Θ_2 have the same gradient, computed with respect to $\mathcal{B}$ and $\mathcal{V}(\omega)$, in correspondence of the convergence points $\hat{\mathcal{B}}$, $\hat{\mathcal{V}}(\omega)$, $\hat{\mathcal{W}}(\omega)$, $\hat{\mathcal{G}}(\omega)$. In detail, with the aid of some notable results on the differentiation of complex-valued matrices, the following gradients are obtained

$$\nabla_{\mathbf{b}_m}\left\{\mathrm{tr}\left[\hat{\mathbf{W}}_i(\omega) \hat{\mathbf{E}}_i\left(\hat{\mathcal{V}}(\omega), \hat{\mathcal{B}}\right)\right]\right\} = \hat{\mathbf{W}}_i^T(\omega) \nabla_{\mathbf{b}_m}\left[\hat{\mathbf{E}}_i\left(\hat{\mathcal{V}}(\omega), \hat{\mathcal{B}}\right)\right] \qquad (38)$$

$$\nabla_{\mathbf{b}_m}\left[\det\left(\hat{\mathbf{E}}_i\left(\hat{\mathcal{V}}(\omega), \hat{\mathcal{B}}\right)\right)\right] = \det\left(\hat{\mathbf{E}}_i\left(\hat{\mathcal{V}}(\omega), \hat{\mathcal{B}}\right)\right)\left[\hat{\mathbf{E}}_i^T\left(\hat{\mathcal{V}}(\omega), \hat{\mathcal{B}}\right)\right]^{-1}$$

$$\cdot \nabla_{\mathbf{b}_m}\left[\left(\hat{\mathbf{E}}_i\left(\hat{\mathcal{V}}(\omega), \hat{\mathcal{B}}\right)\right)\right].$$

Since $R_i\left(\hat{\mathcal{V}}(\omega), \hat{\mathcal{B}}\right) = -\log \det\left(\hat{\mathbf{E}}_i\left(\hat{\mathcal{V}}(\omega), \hat{\mathcal{B}}\right)\right)$, from (38) and (36) we obtain

$$\nabla_{\mathbf{b}_m}\left[\Theta_1\left(\hat{\mathcal{B}}, \hat{\mathcal{V}}(\omega), \hat{\mathcal{G}}(\omega), \hat{\mathcal{W}}(\omega)\right)\right] = \int_{\omega \in \Omega} \sum_{i=1}^{N_u} \alpha_i \hat{\mathbf{W}}_i^T \nabla_{\mathbf{b}_m}\left[\hat{\mathbf{E}}_i\left(\hat{\mathcal{V}}(\omega), \hat{\mathcal{B}}\right)\right] f_\Omega(\omega)\, d\omega$$

$$\nabla_{\mathbf{b}_m}\left[\Theta_2\left(\hat{\mathcal{B}}, \hat{\mathcal{V}}(\omega)\right)\right] = \int_{\omega \in \Omega} \sum_{i=1}^{N_u} \alpha_i \left[\hat{\mathbf{E}}_i^T\left(\hat{\mathcal{V}}(\omega), \hat{\mathcal{B}}\right)\right]^{-1} \nabla_{\mathbf{b}_m}\left[\hat{\mathbf{E}}_i\left(\hat{\mathcal{V}}(\omega), \hat{\mathcal{B}}\right)\right] f_\Omega(\omega)\, d\omega.$$

$$(39)$$

Since $\hat{\mathbf{W}}_i(\omega) = \hat{\mathbf{E}}_i^{-1}\left(\hat{\mathcal{V}}(\omega), \hat{\mathcal{B}}\right)$, we have $\nabla_{\mathbf{b}_m}\Theta_1 = \nabla_{\mathbf{b}_m}\Theta_2$. By using a similar procedure, it is possible to prove the equality $\nabla_{\mathbf{V}_k}\Theta_1 = \nabla_{\mathbf{V}_k}\Theta_2$. This concludes the proof. ∎
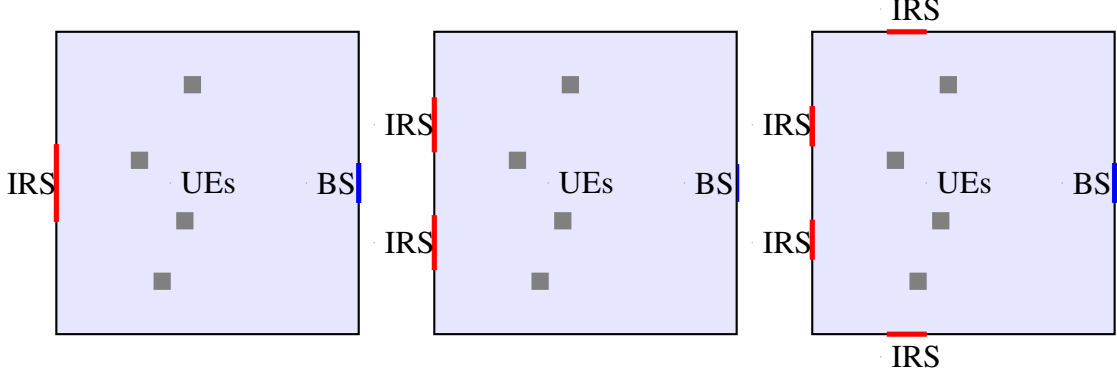
Fig. 1: Layout of the considered indoor scenario (aerial view) for N_u = 4 UEs and N_{IRS} = 1, 2, 3, 4.

IV. NUMERICAL RESULTS AND DISCUSSION

In this section, we illustrate some numerical results in order to analyze the performance of the proposed optimization algorithm, and, in particular, to assess the impact of not using instantaneous CSI for optimizing the IRSs.

A. Simulation scenario

We consider an indoor environment with dimensions 30 m × 30 m along the x-axis and y-axis, respectively, as depicted in Fig. 1. The UEs are equipped with $L = 4$ vertically polarized antennas and are deployed randomly on the left-half of the room, i.e., in an area of 15 m × 30 m, at a height of 1 m. The BS is equipped with $M = 16$ vertically polarized antennas and is placed on the opposite side of the room at the position $(15, 30)$ m and at a height of 2 m. The 16 antennas of the BS are arranged on the $y - z$ plane with 8 antennas along the y-axis and 2 antennas along the z-axis, while the 4 antennas of the UEs are arranged horizontally along the y-axis. The antenna-element gains at the BS and at the UE are set to 3 dB. The variance σ_i^2 of the additive zero-mean Gaussian noise and the maximum power budget P_i (assumed the same for all UEs), are set equal to -97 dBm and 0 dBm, respectively.

B. IRS deployment

The IRSs are placed on the walls around the UEs. The size of each IRS is $0.428/\sqrt{N_{IRS}}$ m × $0.214/\sqrt{N_{IRS}}$ m. This ensures that, regardless of the number of IRSs, the total area covered by the IRSs is 0.428 m × 0.214 m, for a fair comparison. The inter-distance between the scattering elements of the IRS is $\lambda/2$, where $\lambda = 0.011$ m corresponds to a carrier frequency of 28 GHz. Accordingly, the total number of scattering elements that is available in the considered

environment is $P_{IRS} = 3200$. To ensure an affordable computational complexity for Algorithm 1, each IRS is partitioned into $K_t = 64$ tiles each comprising $P = 50$ scattering elements. As a matter of fact, the complexity of the offline optimization algorithm is mainly determined by the computation of $\mathbf{b}_m^{(q+1)}$ in (35), which requires the inversion of a $P \times P$ matrix whose complexity is $O(P^3)$. As a function of K_t, on the other hand, the complexity of Algorithm 1 increases only linearly. As for the power constraints of the IRSs, both GC and LC_{N_b} case studies are considered.

C. Channel model

As for the wireless channel, we assume an NLOS channel model for the direct path, i.e., $X_d = 0$ in (1). This represents the most interesting network scenario for using the IRSs. If a strong LOS path is available, in fact, the impact and contribution of the IRSs are usually less significant if the transmission distances are large and the IRSs have a small size [38]. As for the NLOS contribution to the channel, we consider $N_c = 5$ and $N_p = 10$. More specifically, the centers of the clusters are positioned on an ellipse as discussed in Section II-A with an eccentricity $\epsilon = 0.5$, and the N_p scatterers associated to each cluster are randomly distributed around the clusters' centers such that the angle spread of the signals received by the UEs is 15°. These values are in agreement with typical NLOS multipath models for mmWave communications, e.g., see [45].

D. IRS optimization

In order to assess the effectiveness of the proposed joint online-offline optimization algorithm, we compare the sum-rate of two network setups: (i) the IRSs are optimized according to Algorithm 1 (denoted by OPT) and (ii) the IRSs act as non-controllable scatterers (denoted by NON-OPT). In this latter case, the unit cells of the IRSs operate as diffusive radiating elements whose reflection coefficient has unit amplitude and whose phases are independently and uniformly distributed in $[0, 2\pi]$. Also, the weights α in (14) are set equal to 1.

As for the statistics of the locations of the UEs, we consider three scenarios that are denoted by UD, UD-1m, and UD-0m. In the UD scenario, the UEs are deployed uniformly at random within a given service area. This corresponds to the scenario with the minimal a priory information about the locations of the UEs, and, therefore, the analog beamforming vectors of the IRSs are optimized by averaging over a probability density function $f_\Omega(\omega)$ that accounts for the largest uncertainty. In the UD-1m scenario, the locations of the UEs are assumed to be known with an uncertainty of 0.5 m around the actual location of each UE. In other words, the UEs

are distributed within a disk of diameter 1 m that is centered at their actual locations. In the *UD*-0m scenario, the locations of the UEs are assumed to be perfectly known. This case study corresponds to the scenario in which the maximal a priori CSI is necessary and the IRSs need to be reconfigured more often based on the UEs' locations. It is worth recalling that the IRS-assisted links are determined only by the positions of the UEs. Therefore, the *UD* scenario corresponds to the full CSI setup for optimizing the analog beamforming vectors of the IRSs.

The simulation results are obtained by averaging the sum-rate over 10,000 network realizations and by setting $N_s = 1,000$ in Algorithm 1.
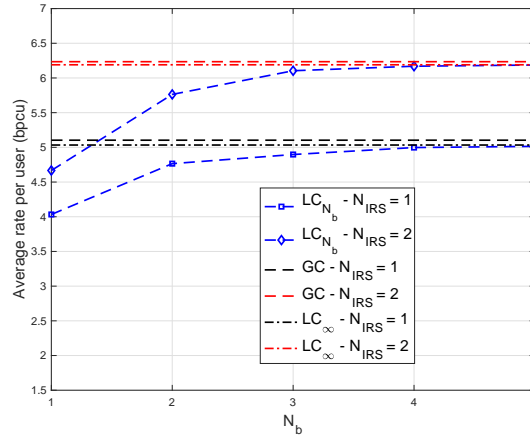
Fig. 2: Performance of the OPT scheme for the LC, GC, and LC$_\infty$ constraints as a function of N_b ($N_{IRS} = 1, 2$ and $N_u = 1$).

E. Local vs. global power constraint and phase quantization noise

In Fig. 2, we investigate the impact of the phase quantization noise on the sum-rate, by assuming a single UE scenario ($N_u = 1$), the OPT setup, and the IO channel model. The constraints LC$_{N_b}$, GC, and LC$_\infty$ are considered, where the latter constraint is referred to the case study with no phase quantization noise. We observe that the LC constraint provides, in the considered case study, similar performance as the GC constraint, even if the number of quantization bits N_b is low. For example, the performance difference is less than 2% if $N_b = 3$. Similar results are obtained for all considered network scenarios. For simplicity, therefore, only the curves that correspond to the GC constraint are reported in the following figures.

F. Sum-rate optimization

In Fig. 3-(left), we compare the OPT and NON-OPT configurations of the IRSs, by assuming the *UD*-1m scenario and the IO channel model. The curves show the sum-rate as a function
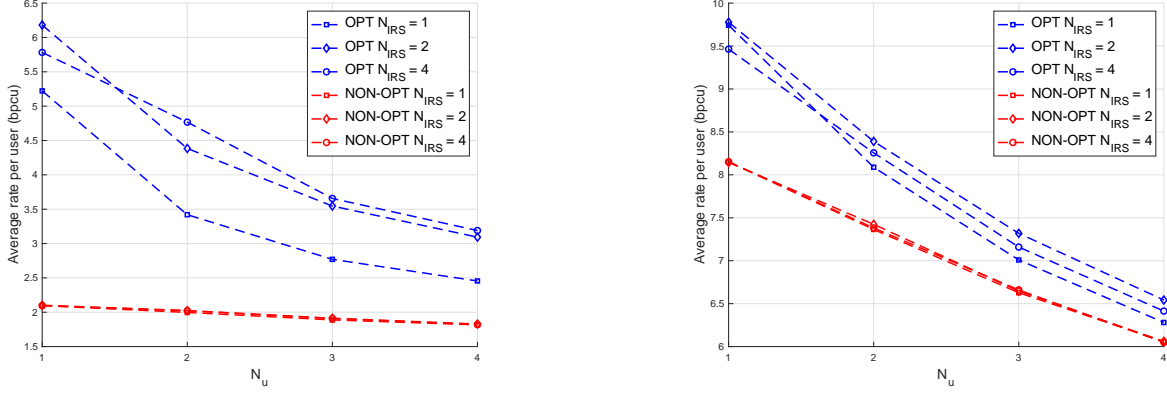
Fig. 3: Comparison between the OPT and NON-OPT schemes. (left) UD-1m setup and IO channel model; (right) UD-1m setup and SM channel model.

of the number of users N_u for $N_{IRS} = 1$, $N_{IRS} = 2$, and $N_{IRS} = 4$. We observe that the OPT configuration outperforms the NON-OPT configuration, which substantiates the effectiveness of deploying IRSs even if the locations of the UEs are not perfectly known. It is worth noting that the average rate per user decreases with the increase of the number of UEs. This can be attributed to a stronger interference in the presence of multiple UEs. As far as the impact of the number of IRSs is concerned, we observe that, in general, the average rate per user increases with N_{IRS} even if the total surface area covered by the IRSs is kept unchanged, i.e., deploying a larger number but smaller in size IRSs yields better performance. This is attributed to the fact that an higher number of IRSs located in different positions allow to increase the multipath, and, hence, the rank of the channel matrices as shown in subsequent results.

Slightly different results are, on the other hand, obtained in the SM channel model, as illustrated in Fig. 3-(right). In this case, in fact, the contribution of the NLOS multipath undergoes a smaller path-loss, and, therefore, fewer IRSs may be needed for achieving good performance.

In Fig. 4, we show the average rate per user that is achievable in the UD-0m and UD scenarios by assuming the IO channel model. By comparing Fig. 3 with Fig. 4, we observe, as expected, that the UD-0m setup outperforms the UD-1m setup, but at the cost of reconfiguring the IRSs more frequently. Likewise, the UD scenario offers the worse performance but with the main advantage of reducing the rate at which the IRSs need to be reconfigured. Nevertheless, it is interesting to note that the OPT case still outperforms the NON-OPT case even if the locations of the UEs are known with a quite large uncertainty.
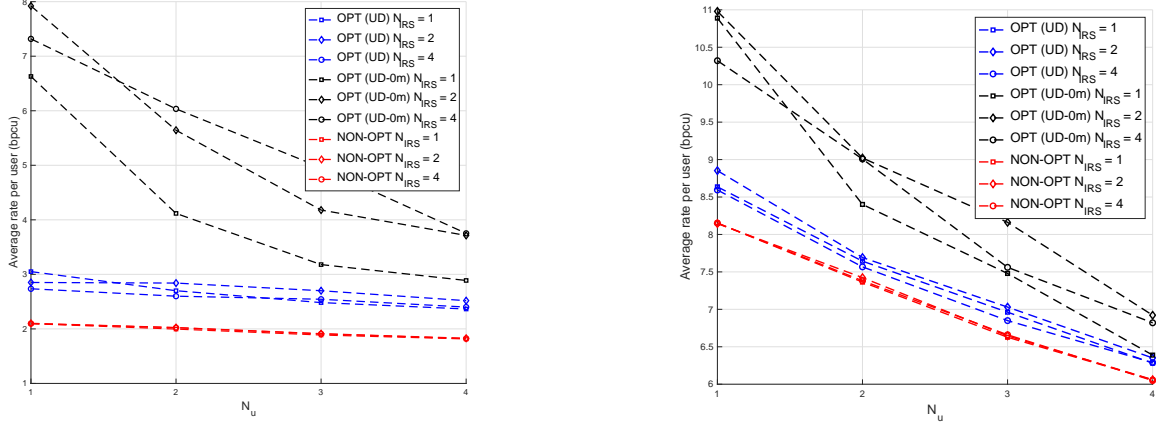
Fig. 4: Comparison between the OPT and NON-OPT schemes. (left) UD-1m vs. UD setups and IO channel model; (right) UD-0m vs. UD setups and IO channel model.

G. Rank of the IRS-assisted channel

In Fig. 5, we analyze the effective rank of the IRS-assisted channel, by assuming $N_u = 1$ and the OPT configuration of the IRSs in the three scenarios UD, UD-1m, UD-0m. Given the channel matrix in (12), we compute $R = \sum |\lambda_i| / \max |\lambda_i|$ as a function of the number of IRSs, where λ_i are the singular values of the channel matrix in (12). The parameter R provides one with information on how well conditioned the IRS-assisted channel is. In other words, the larger R is, the larger the capacity and the number of information streams that can be transmitted simultaneously are. In particular, $R = 1$ indicates that a singular value is predominant, and, thus, the rank of the channel is, in practice, one. If, on the other hand, all singular values are almost the same, the channel capacity is maximized. In the considered system setup, $R \leq 4$, since four antennas are available at the UEs. The results in Fig. 5 show that increasing the number of IRSs increases the effective rank of the channel and, thus, the system sum-rate. The effective rank is larger in the SM channel model because of the more favorable propagation conditions.

H. Analysis of the equivalent array factors of the IRSs

In order to get engineering insights on how the IRSs shape the wireless channel for increasing the sum-rate, Figs. 6 and 7 report the array factor of the BS and the equivalent array factor of the IRSs for some given locations of the UEs. More precisely, the equivalent array factor of an IRS, corresponding to some given locations of the UEs, is obtained by considering an equivalent antenna array virtually located at the IRS and whose array factor is obtained from the
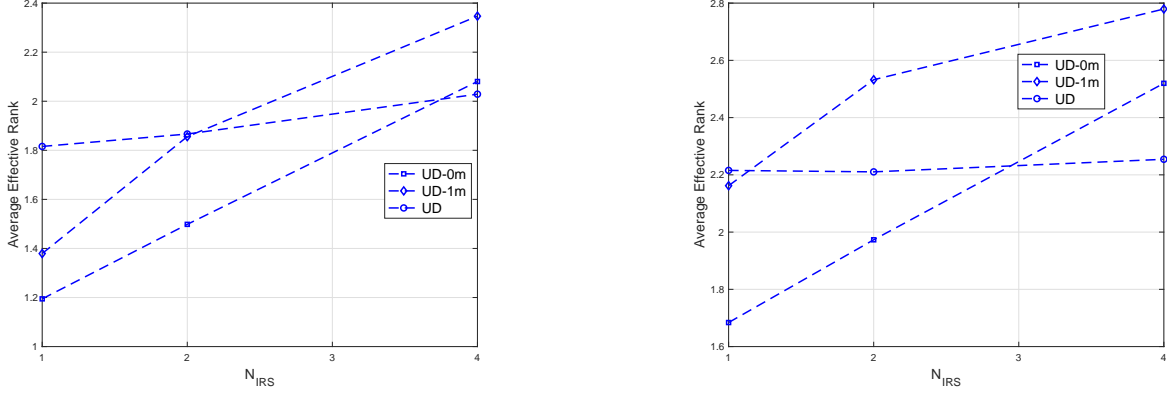
Fig. 5: Average effective rank $E(R)$ of the IRS-assisted channel as a function of N_{IRS} for the UD, UD-1m, UD-0m setups and $N_u = 1$. (left) IO channel model; (right) SM channel model.

digital beamforming vector of the BS, the channel response of the BS-IRS link, and the analog beamforming vector of the IRS [28]. It is worth noting that the equivalent array factor of each IRS may change even if the IRSs are computed during the offline phase and are not updated during the online phase. This is because the array factor of the BS, the wireless channels, and the locations of the UEs are different during the online phase.

In Fig. 6, we illustrate the obtained equivalent array factors by assuming $N_{IRS} = 4$ and $N_u = 2$ (left figure) and $N_{IRS} = 2$ and $N_u = 4$ (right figure) in the UD-1m scenario. The equivalent array factors correspond to a single stream emitted by the BS. The equivalent array factors of the IRSs that correspond to each UE in the system are depicted in a different color. When the number of IRSs is larger than the number of UEs (left figure), the proposed optimization algorithm leads to a configuration in which each UE receives the signals reflected by the two closest IRSs. In fact, we note that the equivalent array factor of the BS associated to each UE points towards two IRSs, which, in turn, reflect the signal towards a single UE. When the number of UEs is larger than the number of IRSs (right figure), on the contrary, the proposed optimization algorithm configures the IRSs in such a way that they reflect the signals mainly towards their respective nearest UEs (UE-1 and UE-2 in the figure), whereas the farthest UEs (UE-3 and UE-4 in the figure) are served directly by the BS through the uncontrolled multipath from the environment.

Finally, Fig. 7 shows the equivalent array factors in the UD scenario. In this case, in particular, the IRSs are optimized with the only a priori information that the UEs can be located anywhere in the grey shadowed area (uncertainty region). The figure reports the results that correspond to the system setup $N_{IRS} = 4$ and $N_u = 4$ for a generic realization of the positions of the UEs. In this case, we observe that the IRSs are optimized in order to diffuse the reflected signals towards
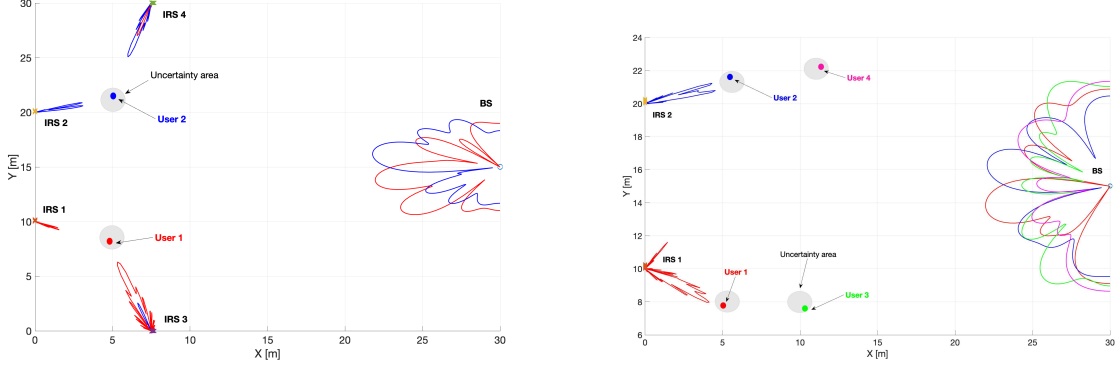
Fig. 6: BS and IRS equivalent array factors. (left) $N_{IRS} = 4$, $N_u = 2$, and UD-1m; (right) $N_{IRS} = 2$, $N_u = 4$, and UD-1m.
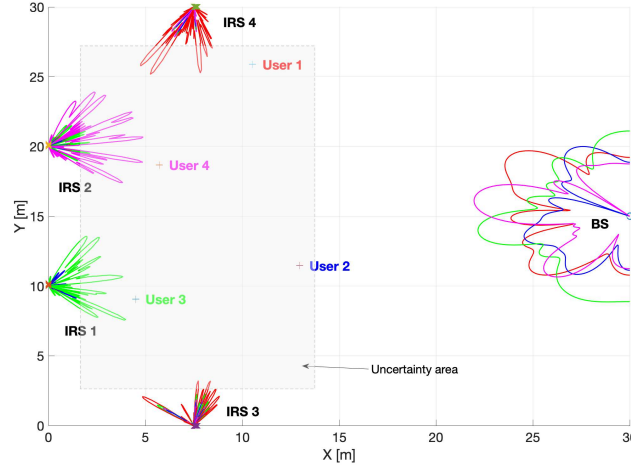

Fig. 7: BS and IRS equivalent array factors for $N_{IRS} = 4$ and $N_u = 4$. The location uncertainty of the UEs is the grey region.

nearby UEs while, at the same time, reducing the interference towards more distant UEs. Let us consider, for instance, UE-3 (in green) and UE-4 (in magenta). We observe that they are primarily served by IRS-1 and IRS-2, respectively. As far as UE-3 and UE-4 are concerned, in addition, we observe that IRS-3 and IRS-4 are configured in such a way they create little interference towards them, i.e., the equivalent array factors of IRS-3 and IRS-4 have notches towards the directions of UE-3 and UE-4. It is also interesting to note that UE-2 is, on the other hand, mainly served by the nearby BS through the uncontrolled multipath from the environment.

V. CONCLUSION

In this paper, we have introduced an optimization algorithm in order to maximize the network sum-rate of a MU-MIMO system in the presence of multiple IRSs. The proposed algorithm consists of an offline phase and an online phase. During the offline phase, the IRSs are configured based on an optimization procedure that relies exclusively on the statistical characterization of

the environment, such as the distribution of the users' locations and the channel statistics. In this phase, in particular, instantaneous CSI is not necessary. During the online phase, on the other hand, only the beamforming vectors at the BS and at the users are optimized, whereas the configuration of the IRSs is kept unchanged. The main advantage of the proposed approach lies in not requiring the estimation of the instantaneous CSI for optimizing the IRSs, thus significantly reducing the overhead associated with the optimization of IRS-assisted wireless systems. The obtained numerical results confirm the validity of the proposed approach. Notably, IRS-assisted wireless systems that are optimized based solely on long-term CSI still provide large performance gains as compared to wireless systems in the absence of IRSs.

REFERENCES

[1] K. David and H. Berndt, "6G vision and requirements: Is there any need for beyond 5G?" *IEEE Vehicular Technology Magazine*, vol. 13, no. 3, pp. 72–80, Sep. 2018.

[2] E. Björnson, L. Sanguinetti, H. Wymeersch, J. Hoydis, and T. L. Marzetta, "Massive MIMO is a reality - what is next?: Five promising research directions for antenna arrays," *Digital Signal Processing*, vol. 94, pp. 3–20, November 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1051200419300776

[3] 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 Vehicular Technology Magazine*, vol. 14, no. 3, pp. 28–41, Sep. 2019.

[4] T. S. Rappaport, Y. Xing, O. Kanhere, S. Ju, A. Madanayake, S. Mandal, A. Alkhateeb, and G. C. Trichopoulos, "Wireless communications and applications above 100 GHz: Opportunities and challenges for 6G and beyond," *IEEE Access*, vol. 7, pp. 78 729–78 757, 2019.

[5] C. Liaskos, S. Nie, A. Tsioliaridou, A. Pitsillides, S. Ioannidis, and I. Akyildiz, "A new wireless communication paradigm through software-controlled metasurfaces," *IEEE Communications Magazine*, vol. 56, no. 9, pp. 162–169, Sep. 2018.

[6] V. S. Asadchy, M. Albooyeh, S. N. Tcvetkova, A. Díaz-Rubio, Y. Ra'di, and S. A. Tretyakov, "Perfect control of reflection and refraction using spatially dispersive metasurfaces," *Phys. Rev. B*, vol. 94, p. 075142, Aug 2016. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevB.94.075142

[7] C. L. Holloway, E. F. Kuester, J. A. Gordon, J. O'Hara, J. Booth, and D. R. Smith, "An overview of the theory and applications of metasurfaces: The two-dimensional equivalents of metamaterials," *IEEE Antennas and Propagation Magazine*, vol. 54, no. 2, pp. 10–35, April 2012.

[8] Y. Ra'di, V. S. Asadchy, and S. A. Tretyakov, "Tailoring reflections from thin composite metamirrors," *IEEE Transactions on Antennas and Propagation*, vol. 62, no. 7, pp. 3749–3760, July 2014.

[9] I. A. Buriak, V. O. Zhurba, G. S. Vorobjov, V. R. Kulizhko, O. K. Kononov, and O. Rybalko, "Metamaterials: Theory, classification and application strategies (review)," *Journal of Nano- and Electronic Physics*, vol. 8, no. 4, 2016. [Online]. Available: https://search.proquest.com/docview/1853386963?accountid=9652

[10] C. L. Holloway, M. A. Mohamed, E. F. Kuester, and A. Dienstfrey, "Reflection and transmission properties of a metafilm: with an application to a controllable surface composed of resonant particles," *IEEE Transactions on Electromagnetic Compatibility*, vol. 47, no. 4, pp. 853–865, Nov 2005.

[11] N. Kaina, M. Dupré, G. Lerosey, and M. Fink, "Shaping complex microwave fields in reverberating media with binary tunable metasurfaces," *Scientific Reports*, vol. 4, no. 1, p. 6693, 2014. [Online]. Available: https://doi.org/10.1038/srep06693

[12] J. P. Turpin, J. A. Bossard, K. L. Morgan, D. H. Werner, , and P. L. Werner, "Reconfigurable and tunable metamaterials: A review of the theory and applications," *International Journal of Antennas and Propagation*, vol. 2014, 2014. [Online]. Available: https://doi.org/10.1155/2014/429837.

[13] L. Dai, B. Wang, M. Wang, X. Yang, J. Tan, S. Bi, S. Xu, F. Yang, Z. Chen, M. Di Renzo, and L. Hanzo, "Reconfigurable intelligent surface-based wireless communication: Antenna design, prototyping and experimental results," *IEEE Access*, vol. 8, pp. 45 913–45 923, Dec 2020.

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

[15] M. Di Renzo, F. Habibi Danufane, X. Xi, J. de Rosny, and S. Tretyakov, "Analytical modeling of the path-loss for reconfigurable intelligent surfaces – Anomalous mirror or scatterer ?" *21st International Workshop on Signal Processing Advances in Wireless Communications*, May 2020.

[16] M. D. Renzo, M. Debbah, D.-T. Phan-Huy, A. Zappone, M.-S. Alouini, C. Yuen, V. Sciancalepore, G. C. Alexandropoulos, J. Hoydis, H. Gacanin, J. d. Rosny, A. Bounceur, G. Lerosey, and M. Fink, "Smart radio environments empowered by reconfigurable AI meta-surfaces: an idea whose time has come," *EURASIP Journal on Wireless Communications and Networking*, vol. 2019, no. 1, p. 129, 2019. [Online]. Available: https://doi.org/10.1186/s13638-019-1438-9

[17] Ö. Özdogan, E. Björnson, and E. G. Larsson, "Intelligent reflecting surfaces: Physics, propagation, and pathloss modeling," *IEEE Wireless Communications Letters*, vol. 9, no. 5, pp. 581–585, May 2020.

[18] D. Dardari, "Communicating with large intelligent surfaces: Fundamental limits and models," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 11, pp. 2526–2537, Nov 2020.

[19] S. Hu, F. Rusek, and O. Edfors, "Beyond massive MIMO: The potential of data transmission with large intelligent surfaces," *IEEE Transactions on Signal Processing*, vol. 66, no. 10, pp. 2746–2758, May 2018.

[20] A. Zappone, M. D. Renzo, F. Shams, X. Qian, and M. Debbah, "Overhead-aware design of reconfigurable intelligent surfaces in smart radio environments," *IEEE Transactions on Wireless Communications*, to appear, 2020.

[21] Q. Wu, S. Zhang, B. Zheng, C. You, and R. Zhang, "Intelligent reflecting surface aided wireless communications: A tutorial," *arXiv:2007.02759*, 2020.

[22] Y. Liu, X. Liu, X. Mu, T. Hou, J. Xu, Z. Qin, M. D. Renzo, and N. Al-Dhahir, "Reconfigurable intelligent surfaces: Principles and opportunities," *arXiv:2007.03435*, 2020.

[23] Q.-U.-A. Nadeem, A. Kammoun, A. Chaaban, M. Debbah, and M.-S. Alouini, "Asymptotic max-min SINR analysis of reconfigurable intelligent surface assisted miso systems," *IEEE Transactions on Wireless Communications*, vol. 19, pp. 7748–7764, Dec. 2020.

[24] M.-M. Zhao, Q. Wu, M.-J. Zhao, and R. Zhang, "Intelligent reflecting surface enhanced wireless network: Two-timescale beamforming optimization," *IEEE Transactions on Wireless Communications*, to appear, 2020.

[25] Q.-U.-A. Nadeem, A. Chaaban, and M. Debbah, "Opportunistic beamforming using an intelligent reflecting surface without instantaneous CSI," *IEEE Wireless Communications Letters*, to appear, 2020.

[26] S. Lin, B. Zheng, G. C. Alexandropoulos, M. Wen, M. Di Renzo, and F. Chen, "Reconfigurable intelligent surfaces with reflection pattern modulation: Beamforming design and performance analysis," *IEEE Transactions on Wireless Communications*, to appear, 2020.

[27] Z. Peng, T. Li, C. Pan, H. Ren, W. Xu, and M. Di Renzo, "Analysis and optimization for IRS-aided multi-pair communications relying on statistical csi," *arXiv:2007.11704*, 2020.

[28] D. Dardari and D. Massari, "Using metaprisms for performance improvement in wireless communications," *arXiv:2003.13505*, 2020.

[29] A. Abrardo, G. Fodor, and M. Moretti, "Distributed digital and hybrid beamforming schemes with MMSE-SIC receivers for the MIMO interference channel," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 7, pp. 6790–6804, May 2019.

[30] C. Pan, H. Ren, K. Wang, M. Elkashlan, A. Nallanathan, J. Wang, and L. Hanzo, "Intelligent reflecting surface aided MIMO broadcasting for simultaneous wireless information and power transfer," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 8, pp. 1719–1734, Aug. 2020.

[31] M. Najafi, V. Jamali, R. Schober, and V. H. Poor, "Physics-based Modeling and Scalable Optimization of Large Intelligent Reflecting Surfaces," *arXiv e-prints*, p. arXiv:2004.12957, Apr. 2020.

[32] S. Wu, C.-X. Wang, H. Haas, el Hadi M. Aggoune, M. M. Alwakeel, and B. Ai, "A non-stationary wideband channel model for massive MIMO communication systems," *IEEE Transactions on Wireless Communications*, vol. 14, no. 3, March 2015.

[33] "5G channel model for bands up to 100 GHz," *Report*, pp. 1–107, October 2016. [Online]. Available: https://prepareforchange.net/wp-content/uploads/2018/12/5G_Channel_Model_for_bands_up_to100_GHz2015-12-6.pdf

[34] G. Gradoni and M. Di Renzo, "End-to-end mutual coupling aware communication model for reconfigurable intelligent surfaces: An electromagnetic-compliant approach based on mutual impedances," *arXiv:2009.02694*, 2020.

[35] S. W. Ellingson, "Path Loss in Reconfigurable Intelligent Surface-Enabled Channels," *arXiv:1912.06759*, Dec. 2019.

[36] W. Tang, M. Z. Chen, X. Chen, J. Y. Dai, Y. Han, M. Di Renzo, Y. Zeng, S. Jin, Q. Cheng, and T. J. Cui, "Wireless communications with reconfigurable intelligent surface: Path loss modeling and experimental measurement," *IEEE Transactions on Wireless Communications*, 2020, early access.

[37] F. H. Danufane, M. Di Renzo, J. de Rosny, and S. Tretyakov, "On the path-loss of reconfigurable intelligent surfaces: An approach based on Green's theorem applied to vector fields," *arXiv:2007.13158*, 2020.

[38] N. S. Perovic, L. N. Tran, M. Di Renzo, and M. F. Flanagan, "Achievable rate optimization for mimo systems with reconfigurable intelligent surfaces," *arXiv:2008.09563*, 2020.

[39] V. S. Asadchy, M. Albooyeh, S. N. Tcvetkova, A. Díaz-Rubio, Y. Ra'di, and S. A. Tretyakov, "Perfect control of reflection and refraction using spatially dispersive metasurfaces," *Phys. Rev. B*, vol. 94, p. 075142, Aug 2016. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevB.94.075142

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

[41] J. Wang, Q. Peng, Y. Huang, H.-M. Wang, and X. You, "Convexity of weighted sum rate maximization in noma systems," *IEEE Communications Letters*, vol. 24, no. 9, pp. 1323–1327, September 2017.

[42] D. P. Bertsekas, *Nonlinear programming*. Athena Scientific, 1999.

[43] H. Gauch, T. Kim, M. Bengtsson, and M. Skoglund, "Subspace estimation and decomposition for large millimeter-wave MIMO systems," *IEEE Journal on Selectect Topics in Signal Processing*, vol. 10, no. 3, pp. 528–542, April 2016.

[44] M. V. Solodov, "On the convergence of constrained parallel variable distribution algorithm," *SIAM J. Optimization*, vol. 8, no. 1, pp. 187–196, February 1998.

[45] S. Buzzi and C. D'Andrea, "Energy efficiency and asymptotic performance evaluation of beamforming structures in doubly massive MIMO mmWave systems," *IEEE Transactions on Green Communications and Networking*, vol. 2, no. 2, pp. 385–396, 2018.