Intelligent Reflecting Surface-assisted MU-MISO Systems with Imperfect Hardware: Channel Estimation and Beamforming Design

Anastasios Papazafeiropoulos, Cunhua Pan, Pandelis Kourtessis, Symeon Chatzinotas, John M. Senior

***Abstract*—Intelligent reflecting surface (IRS), consisting of low-cost passive elements, is a promising technology for improving the spectral and energy efficiency of the fifth-generation (5G) and beyond networks. It is also noteworthy that an IRS can shape the reflected signal propagation. Most works in IRS-assisted systems have ignored the impact of the inevitable residual hardware impairments (HWIs) at both the transceiver hardware and the IRS while any relevant works have addressed only simple scenarios, e.g., with single-antenna network nodes and/or without taking the randomness of phase noise at the IRS into account. In this work, we aim at filling up this gap by considering a general IRS-assisted multi-user (MU) multiple-input single-output (MISO) system with imperfect channel state information (CSI) and correlated Rayleigh fading. In parallel, we present a general computationally efficient methodology for IRS reflect beamforming (RB) optimization. Specifically, we introduce an advantageous channel estimation (CE) method for such systems accounting for the HWIs. Moreover, we derive the uplink achievable spectral efficiency (SE) with maximal-ratio combining (MRC) receiver, displaying three significant advantages being: 1) its closed-form expression, 2) its dependence only on large-scale statistics, and 3) its low training overhead. Notably, by exploiting the first two benefits, we achieve to perform optimization with respect to the that can take place only at every several coherence intervals, and thus, reduces significantly the computational cost compared to other methods which require frequent phase optimization. Among the insightful observations, we highlight that uncorrelated Rayleigh fading does not allow optimization of the SE, which makes the application of an IRS ineffective. Also, in the case that the phase drifts, describing the distortion of the phases in the RBM, are uniformly distributed, the presence of an IRS provides no advantage. The analytical results outperform previous works and are verified by Monte-Carlo (MC) simulations.**

***Index Terms*—Intelligent reflecting surface (IRS), transceiver hardware impairments, channel estimation, achievable spectral efficiency, beyond 5G networks.**

I. Introduction

In the last decade, a variety of wireless technological advances, including millimeter-wave (mmWave) communication and massive multiple-input multiple-output (mMIMO) systems, have been proposed to achieve a 1000-fold capacity increase and ubiquitous wireless connectivity among a large number of devices [1]. Unfortunately, these technologies face practical limitations in terms of excessive energy consumption and hardware cost as well as guaranteed quality of service (QoS) in harsh propagation environments. For instance, mmWave communications exhibit high penetration/path loss while requiring expensive and energy-consuming transceivers. Similarly, mMIMO systems manifest low performance in poor scattering conditions while the number of active elements might render the energy usage prohibitive. Moreover, new challenging use cases will emerge with possibly similar shortcomings. As a result, future networks require radical paradigm shifts towards their energy sustainability, e.g., control to some extent over the propagation environment.

A disruptive technology, covering this gap, has emerged under the label intelligent reflecting surfaces (IRSs) or reconfigurable intelligent surfaces (RISs). An IRS consists of a meta-surface including a large number of reconfigurable passive elements that are able to function independently by inducing certain phase shifts on the impinging waves [2]. The smart adjustment of the phase shifts is managed by an attached controller and allows the coherent addition of the reflected signals to boost the desired signal at the receiver. The IRS design and applications have attracted a lot of significant research interest [2]–[9]. For example, in [3], the downlink of an IRS-assisted multi-user (MU) multiple-input single-output (MISO) communication system was studied by jointly optimizing the precoding and reflect beamforming matrices (RBMs), in order to minimize the transmit power at the base station (BS) with signal-to-interference-plus-noise ratio (SINR) constraints. In addition, the sum rate was maximized in [4] subject to a transmit power constraint. Furthermore, the outperformance of the IRS with respect to the decode-and-forward (DF) relay was presented in [5]. Also, the authors in [6] maximized the minimum user rate in the large number of antennas regime. Despite the fundamental design issues, applications regarding IRSs have started to emerge such as the maximization of the minimum secrecy rate for physical layer security [9].

In general, there are two approaches for phases optimization as the literature reveals. The first method is based on statistical channel state information (CSI) [6], [10]–[15] and the second method is based on instantaneous CSI [3], [4]. According to the second method, the phases are optimized at every

A. Papazafeiropoulos is with the Communications and Intelligent Systems Research Group, University of Hertfordshire, Hatfield AL10 9AB, U. K., and with SnT at the University of Luxembourg, Luxembourg. C. Pan with the School of Electronic Engineering and Computer Science at Queen Mary University of London, London E1 4NS, U.K. P. Kourtessis and John M.Senior are with the Communications and Intelligent Systems Research Group, University of Hertfordshire, Hatfield AL10 9AB, U. K. S. Chatzinotas is with the SnT at the University of Luxembourg, Luxembourg. A. Papazafeiropoulos was supported by the University of Hertfordshire's 5-year Vice Chancellor's Research Fellowship. S. Chatzinotas was supported by the National Research Fund, Luxembourg, under the projects RISOTTI. E-mails: tapapazaf@gmail.com, , c.pan@qmul.ac.uk, {p.kourtessis,j.m.senior}@herts.ac.uk, symeon.chatzinotas@uni.lu.

coherence interval since the corresponding expressions depend on small-scale channel fading. On the contrary, the first method includes expressions that depend on the large-scale statistics, which change at every several coherence intervals. Thus, the significance of the first method is noteworthy since it reduces considerably the signal overhead which can be prohibitive in the case of a large number of reflecting elements at the IRS. Moreover, it results in lower computational complexity. Especially, based on these observations, in high-speed scenarios with fast time-varying channels, it is more practical to design and adjust the IRS phase shifts according to the statistical CSI while the tuning of the IRS parameters, based on instantaneous CSI, would be more challenging since they would have to be updated more frequently. Furthermore, although the IRS does not consume ideally transmit power, its smart controller is power-consuming and its continuous overloading with operations in the case of instantaneous CSI would not be energy efficient.

Although most existing works with IRS-aided systems have relied on the knowledge of perfect CSI, this is a highly unrealistic assumption. In practice, systems have imperfect CSI. Especially, their passive elements make them energy efficient, but, contrary to conventional systems, this interesting feature does enable them to accomplish the channel estimation (CE) task by transmission/reception of pilot symbols. Hence, it is of paramount importance to take into account the CE before arriving at realistic conclusions. Among the fundamental works [7], [16]–[19], the authors in [17] proposed an ON/OFF channel estimation scheme that obtains one-by-one least squares (LS) estimates of all IRS-assisted channels for a single-user MISO system. Moving to MUs systems, finding more applications in contemporary systems, the authors in [18] exploited the sparsity of the channel and formulated a sparse channel matrix recovery problem for CE. In [19], the authors extended the model in [17] by assuming all IRS elements to be active during training while a number of sub-phases equal at least to the number of IRS elements are considered. This method provides better CE as the number of sub-phases increases, but the achievable rate worsens since the data transmission phase takes a smaller fraction of the coherence time due to excessive training overhead. Another drawback is that this method provides the estimates of the channels of the individual IRS elements while the covariance of the channel vector from all IRS elements to a specific user equipment (UE) is unknown.

On the other hand, prior literature of IRS-assisted systems has mostly assumed perfect hardware while practical applications are affected by unavoidable transceiver hardware impairments (T-HWIs) such as the in-phase/quadrature-phase (I/Q)-imbalance [20], the quantization noise in the analog-to-digital converters (ADCs), and the oscillator phase noise (PN) [21], [22]. Even if mitigation/compensation aRISlgorithms exist, T-HWIs cannot be completely removed [23], [24]. Basically, T-HWIs are divided into two main categories being the additive and multiplicative T-HWIs. In this work, we focus on the impact of the additive T-HWIs, while the study of multiplicative T-HWIs will be the topic for future work. In this direction, an examination of existing works with HWIs in the area of IRS-assisted systems shows that relevant studies are in their infancy [25]–[30]. In [25], only single-antennas nodes were considered, and the phase errors were assumed known (deterministic). In addition, the phase noise, induced by an IRS and, henceforth called IRS-HWIs, has been studied in [26], [27] in the case of perfect CSI, but no expectation was taken over the phase noise. Note that this phase noise, coming from the finite precision configuration of the phase drifts, is irrelevant with the phase noise coming from imperfect signal generation in local oscillators in standard antenna systems. Moreover, in [28], despite its randomness, again, no averaging of the phase noise was applied. Furthermore, only a single UE communication has been considered and only upper bounds on the channel capacities have been studied, which are not also obtained in closed forms. The authors in [29] provided the beamforming optimization by accounting for T-HWIs in a single-user setting, and in [30], the secrecy rate was derived. Notably, only a few works in IRS-aided systems have assumed correlated Rayleigh fading despite that this is normally the case in multi-antenna next-generation systems. Apart from that, most works perform RBM optimization with a high computational cost in every coherence interval.

A. Contribution

The previous observations motivate the topic of this work, which is the design/study of a general IRS-assisted MU-MISO system with imperfect CSI and HWIs at both the IRS and the transceiver while performing robust optimization. Notably, the introduction of HWIs increases the complexity/difficulty and demands substantial manipulations during the analysis of IRS-assisted systems. The main contributions are summarized as follows.

- Contrary to [26], we have assumed multiple antennas at the BS and multiple UEs as well as imperfect CSI. Also, compared to [28], we have considered correlated Rayleigh fading, multiple UEs, and closed-form lower bounds, which are more practical than any upper bounds. Both references have not addressed properly the impact of phase noise while, in [25], only deterministic phase noise was assumed. In [29] and [30], only the impact of T-HWIs was studied while only a single destination and perfect CSI are assumed. Notably, as far as the authors are aware, our work is the only one accounting for the randomness of the phase noise.
- Many previous works have assumed that the optimization of the RBM should take place at every coherence interval since the corresponding expressions depend on small-scale channel fading, while our proposed results, being dependent only on large-scale statistics are suggested to be optimized at every several coherence intervals. Thus, their significance is noteworthy since they reduce considerably the signal overhead which can be prohibitive in the case of a large number of reflecting elements at the IRS.
- We perform CE by means of linear minimum mean square error (LMMSE) while HWIs are taken into account. In parallel, we have assumed correlated Rayleigh fading. Our method provides analytical tractable expressions with low

overhead compared to previous works.[1]

- We derive the uplink achievable spectral efficiency (SE) (lower bound) of an IRS-assisted MU-MISO system with MRC, imperfect CSI, and HWIs in a closed-form dependent only on large-scale statistics (covariances).
- We optimize the achievable sum SE with respect to the RBM. As mentioned, contrary to other works that depend on small-scale statistics (e.g., see [8] where the stochastic successive convex approximation technique has been performed), our optimization can be performed quite efficiently by the project gradient ascent at every several coherence intervals since both the sum SE and the proposed algorithm require only the large-scale statistics. Notably, contrary to existing works such as [10], [11], based on statistical CSI, we achieved to provide the SE and the phases optimization in closed-form.
- We shed light on the degradation of the uplink sum SE of an IRS-aided MU-MISO system due to the presence of imperfect CSI, HWIs, and correlated fading. For example, we thoroughly examine how the probability density function (PDF) of the phase noise at the IRS and the severity of the T-HWIs affect the performance of the IRS-aided systems.

B. Paper Outline

The remainder of this paper is organized as follows. Section II presents the system model of an IRS-assisted MU-MISO system with correlated Rayleigh fading and HWIs. Section III provides the CE. Section IV presents the uplink sum SE and the optimization concerning the IRS RBM. The numerical results are placed in Section V, and Section VI concludes the paper.

C. Notation

Vectors and matrices are denoted by boldface lower and upper case symbols, respectively. The notations $(\cdot)^{\mathsf{T}}$, $(\cdot)^{\mathsf{H}}$, and $\mathrm{tr}(\cdot)$ represent the transpose, Hermitian transpose, and trace operators, respectively. The expectation operator is denoted by $\mathbb{E}[\cdot]$ (or $\mathbb{E}_x[\cdot]$ to denote expectation with respect to x) while $\mathrm{diag}(\mathbf{a})$ represents an $n \times n$ diagonal matrix with diagonal elements being the elements of vector $\mathbf{a}$. In the case of a matrix $\mathbf{A}$, $\mathrm{diag}(\mathbf{A})$ denotes a diagonal matrix with elements corresponding to the diagonal elements of $\mathbf{A}$. Also, $\arg(\cdot)$ and $\circ$ denote the argument function and the Hadamard product, respectively. Finally, $\mathbf{b} \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Sigma})$ represents a circularly symmetric complex Gaussian vector with zero mean and covariance matrix $\mathbf{\Sigma}$.

II. System Model

We consider an IRS-aided MU-MISO system as depicted in Fig. 1. In particular, a BS, equipped with M antennas, serves K single-antenna UEs by means of one IRS consisting of N passive reflecting elements introducing phase shifts onto the incoming signal waves. The phase-shifts are adjusted by a controller exchanging information with the BS through a backhaul link. Reasonably, the IRS is deployed in the line-of-sight (LoS) of the BS by assuming that both the BS and IRS are deployed at high altitude and their locations are fixed. Moreover, the proposed model assumes direct links between the BS and the UEs, but these could be neglected in certain scenarios. For example, in mmWave transmission, suggested by 5G and beyond systems, high penetration losses and resultant signal blockages do not allow the presence of an LoS component [31].

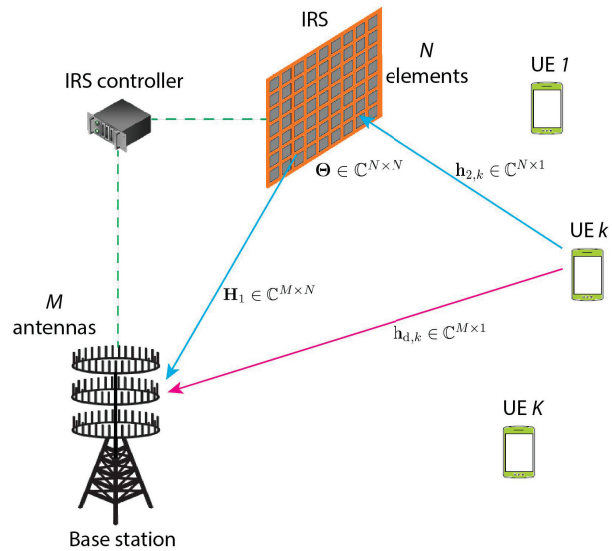


Fig. 1: An IRS-assisted uplink MU-MISO communication system with M BS antennas, N IRS elements, and K UEs.

A. Channel Model

We assume a time-varying narrowband channel, divided into coherence blocks, where each block has a duration τ_{c} channel uses.[2] Note that $\tau_{\mathrm{c}} = B_{\mathrm{c}} T_{\mathrm{c}}$ with B_{c} and T_{c} being the coherence bandwidth and the coherence time in Hz and s, respectively. Especially, we employ the standard time-division-duplex (TDD) protocol, where each block accounts for τ channel uses for the uplink training phase and $\tau_{\mathrm{u}} = \tau_{\mathrm{c}} - \tau$ channel uses for the uplink data transmission phase.

Let $\mathbf{h}_{\mathrm{d},k} \in \mathbb{C}^{M\times 1}$, $\mathbf{H}_1 = [\mathbf{h}_{1,1} \ldots, \mathbf{h}_{1,N}] \in \mathbb{C}^{M\times N}$, and $\mathbf{h}_{2,k} \in \mathbb{C}^{N\times 1}$ be the direct channel between the BS and UE k, the LoS channel between the BS and the IRS with $\mathbf{h}_{1,i}$ for $i = 1, \ldots, N$ being its column vectors, and the channel between the IRS and UE k. The subscripts 1 and 2 correspond to the BS-IRS and IRS-UE k links, respectively. Although the majority of existing works, e.g., [3], [4], assumed independent Rayleigh model, we account for spatial correlation, which appears in practice and affects the performance [6]. Hence, $\mathbf{h}_{\mathrm{d},k}$ and $\mathbf{h}_{2,k}$ are described in terms of correlated Rayleigh

[1] Works such as [7], [18] do not provide analytical expressions. Also, previous CE methods with correlated fading do not allow the derivation of an optimizable achievable spectral efficiency (SE) being dependent on the RBM. In [19], only the estimated individual channels between each IRS element and each UE are obtained while the inter-element correlation is unknown.

[2] The extension to the wideband case could follow the lines of [16], [32] and could be the topic of future work.

fading distributions as

$$\mathbf{h}_{2,k} = \sqrt{\beta_{2,k}}\mathbf{R}_{\mathrm{IRS},k}^{1/2}\mathbf{z}_k, \quad (1)$$

$$\mathbf{h}_{\mathrm{d},k} = \sqrt{\beta_{\mathrm{d},k}}\mathbf{R}_{\mathrm{BS},k}^{1/2}\mathbf{z}_{\mathrm{d},k}, \quad (2)$$

where $\mathbf{R}_{\mathrm{IRS},k} \in \mathbb{C}^{N\times N}$ and $\mathbf{R}_{\mathrm{BS},k} \in \mathbb{C}^{M\times M}$ describe the deterministic Hermitian-symmetric positive semi-definite correlation matrices at the IRS and the BS respectively with $\mathrm{tr}\left(\mathbf{R}_{\mathrm{IRS},k}\right) = N$ and $\mathrm{tr}\left(\mathbf{R}_{\mathrm{BS},k}\right) = M$. Notably, the correlation matrices $\mathbf{R}_{\mathrm{IRS},k}$ and $\mathbf{R}_{\mathrm{BS},k}$ $\forall k$ are assumed to be known by the network. They can be obtained through the existing estimation methods (see e.g., [33]). Given that correlation models for IRS using meta-surfaces are not known, we adopt the correlation model in [34] for conventional antenna arrays.[3] Also, $\beta_{2,k}$ and $\beta_{\mathrm{d},k}$ denote the path-loss of the IRS-UE k and BS-UE k links, respectively. Furthermore, $\mathbf{z}_k \sim \mathcal{CN}\left(\mathbf{0}, \mathbf{I}_N\right)$ and $\mathbf{z}_{\mathrm{d},k} \sim \mathcal{CN}\left(\mathbf{0}, \mathbf{I}_N\right)$ describe the corresponding fast-fading vectors.

By taking into account that the IRS is designed to be installed in a location providing an LoS channel with the BS, the channel matrix $\mathbf{H}_1$ will likely have rank one, which results in performance gains only when $K = 1$ [6]. However, an MU scenario requires $\mathrm{rank}\left(\mathbf{H}_1\right) \geq K$. The higher rank could be achieved by placing the IRS close to the BS or by deterministic scattering between the BS and the IRS. The high rank LoS channel $\mathbf{H}_1$ can be obtained as

$$[\mathbf{H}_1]_{m,n} = \sqrt{\beta_1}\exp\big(j\frac{2\pi}{\lambda}\left(m-1\right)d_{\mathrm{BS}}\sin\theta_{1,n}\sin\phi_{1,n} + \left(n-1\right)d_{\mathrm{IRS}}\sin\theta_{2,m}\sin\phi_{2,m}\big), \quad (3)$$

where λ is the carrier wavelength, β_1 is the path-loss between the BS and IRS while d_{BS} and d_{IRS} are the inter-antenna separation at the BS and inter-element separation at the IRS, respectively [19]. Also, $\theta_{1,n}$, $\phi_{1,n}$ describe the elevation and azimuth LoS angles of departure (AoD) at the BS with respect to IRS element n, and $\theta_{2,n}$, $\phi_{2,n}$ describe the elevation and azimuth LoS angles of arrival (AoA) at the IRS. In practice, d_{BS} and d_{IRS} are known by construction while the angles, depending only on the locations, can be calculated when the locations are given according to [36]. It is worthwhile to mention that the estimation of the correlation matrices could also be obtained similar to $\mathbf{H}_1$ since the dependence of their expressions on the distances and the angles is similar.

B. Ideal Uplink Signal Model

The ideal received complex baseband signal vector by the BS is written as

$$\mathbf{y} = \sum_{k=1}^{K}\left(\mathbf{h}_{\mathrm{d},k} + \mathbf{H}_1\boldsymbol{\Theta}\mathbf{h}_{2,k}\right)x_k + \mathbf{w}, \quad (4)$$

where $\mathbf{w} \sim \mathcal{CN}\left(\mathbf{0}, \sigma^2\mathbf{I}_M\right)$ is the additive white Gaussian noise (AWGN) vector at the BS while $\boldsymbol{\Theta} = \mathrm{diag}\left(\alpha_1 e^{j\theta_1}, \ldots, \alpha_N e^{j\theta_N}\right) \in \mathbb{C}^{N\times N}$ is the RBM being diagonal and representing the response of the N elements with $\theta_n \in [0, 2\pi]$ and $\alpha_n \in [0, 1]$ denoting the phase and amplitude coefficient for element n, respectively. As commonly assumed due to recent advances towards lossless metasurfaces [37], [38], we set $\alpha_n = 1$ $\forall n$, i.e., we assume maximum signal reflection. For the sake of exposition, given the RBM, we denote the overall channel vector $\mathbf{h}_k = \mathbf{h}_{\mathrm{d},k} + \mathbf{H}_1\boldsymbol{\Theta}\mathbf{h}_{2,k}$, distributed as $\mathbf{h}_k \sim \mathcal{CN}\left(0, \mathbf{R}_k\right)$, where $\mathbf{R}_k = \beta_{\mathrm{d},k}\mathbf{R}_{\mathrm{BS},k} + \beta_{2,k}\mathbf{H}_1\boldsymbol{\Theta}\mathbf{R}_{\mathrm{IRS},k}\boldsymbol{\Theta}^{\mathsf{H}}\mathbf{H}_1^{\mathsf{H}}$.

[3]While writing this work, the authors in [35] presented a more suitable correlation for IRSs. Its thorough study is the topic of ongoing research.

C. Hardware Impairments

In this work, we consider two distinct types of HWIs in an IRS-assisted system: 1) the aggregate additive HWIs at the transceiver, and 2) the HWIs emerged from the passive elements of the IRS. Henceforth, we denote them T-HWIs and IRS-HWIs, respectively.

1) T-HWIs: The majority of papers in the IRS literature have relied on the unrealistic assumption of ideal transceiver hardware. Especially, next-generation antenna deployments with a large number of antennas such as a massive MIMO systems assisted by an IRS should be implemented with cheap hardware, in order to be cost-efficient as the number of antennas increases. However, cheaper hardware results in lower quality with more severe HWIs that are more power consuming and degrade further the system performance. Instead, we take into account the additive distortions at both the transmitter and the receiver being Gaussian distributed with average powers proportional to the average transmit and received signals, respectively [23]. We would like to mention that although Gaussian modeling for the HWIs can be assumed rudimentary, it is used widely because its tractability allows extracting primary conclusions, e.g., see the recent works [39], [40]. The Gaussianity results by means of the aggregate contribution of many impairments. Especially, let $p_k = \mathbb{E}\{|x_k|^2\}$ be the transmit power from UE k having transmit signal x_k and $\mathbf{h}_k \in \mathbb{C}^{M\times 1}$ be the channel vector of this UE. The additive transceiver distortions during the uplink are described in terms of conditional distributions with respect to the channel realizations as

$$\delta_{\mathrm{t},k} \sim \mathcal{CN}\left(0, \Lambda_k\right), \quad (5)$$

$$\boldsymbol{\delta}_{\mathrm{r}} \sim \mathcal{CN}\left(\mathbf{0}, \boldsymbol{\Upsilon}\right), \quad (6)$$

where $\Lambda_k = \kappa_{\mathrm{UE}}p_k$ and $\boldsymbol{\Upsilon} = \kappa_{\mathrm{BS}}\sum_{i=1}^{K}p_i\,\mathrm{diag}\left(|h_{i,1}|^2, \ldots, |h_{i,M}|^2\right)$ with $\mathbf{h}_i = [h_{i,1}, \ldots, h_{i,M}]^{\mathsf{T}}$. The variance $\boldsymbol{\Upsilon}$ can also be written as $\boldsymbol{\Upsilon} = \kappa_{\mathrm{BS}}\sum_{i=1}^{K}p_i\mathbf{I}_M \circ \mathbf{h}_i\mathbf{h}_i^{\mathsf{H}}$. The proportionality parameters κ_{UE} and κ_{BS} express the severity of the residual impairments at the transmitter and receiver side, and are met in applications in terms of the error vector magnitude (EVM) [41]. For example, the EVM at the BS is defined as

$$\mathrm{EVM}_{\mathrm{BS}} = \sqrt{\frac{\mathbb{E}[\|\boldsymbol{\delta}_{\mathrm{r}}\|_2^2]}{\mathbb{E}[\|\mathbf{y}\|_2^2]}} = \sqrt{\frac{\mathrm{tr}\left(\boldsymbol{\Upsilon}\right)}{\mathbb{E}[\|\mathbf{y}\|_2^2]}} = \sqrt{\kappa_{\mathrm{BS}}}, \quad (7)$$

where the expectations take place for a specific channel realization. Here, for the sake of simplicity, we assume that the parameter κ_{UE} is identical for all UEs.

2) IRS-HWIs: Taking into account that the reflection phases of the IRS passive elements cannot be configured with infinite precision, they can be modeled in terms of phase errors [42].[4] In particular, IRS-HWIs are In described by means of a random diagonal phase error matrix consisting of N random phase errors, i.e., $\widetilde{\mathbf{\Theta}} = \text{diag}\left(e^{j\tilde{\theta}_1}, \ldots, e^{j\tilde{\theta}_N}\right) \in \mathbb{C}^{N\times N}$, where $\tilde{\theta}_i, i = 1, \ldots, N$ are the random phase errors being i.i.d. randomly distributed in $[-\pi, \pi)$ according to a certain circular distribution. Also, we assume that the PDF of $\tilde{\theta}_i$ is symmetric and its mean direction is zero, i.e., $\arg\left(\mathbb{E}[\mathrm{e}^{j\tilde{\theta}_i}]\right) = 0$ [42]. The most widely used PDFs, being able to describe the phase noise are the uniform and the Von Mises distributions [42], where:

- the uniform distribution expresses completely lack of knowledge (random reflection) and its characteristic function (CF) denoted by m is 0,
- the Von Mises distribution with a zero-mean and concentration parameter $\kappa_{\tilde{\theta}}$, capturing the accuracy of the estimationThe, has a CF $m = \frac{\mathrm{I}_1(\kappa_{\tilde{\theta}})}{\mathrm{I}_0(\kappa_{\tilde{\theta}})}$ with $\mathrm{I}_p(\kappa_{\tilde{\theta}})$ being the modified Bessel function of the first kind and order p.

D. Realistic Uplink Signal Model with HWIs

Overall, the realistic received signal vector by the BS after having incorporated both the T-HWIs and IRS-HWIs in (4) is given by

$$\mathbf{y} = \sum_{k=1}^{K} \left(\mathbf{h}_{\mathrm{d},k} + \mathbf{H}_1 \mathbf{\Theta} \widetilde{\mathbf{\Theta}} \mathbf{h}_{2,k}\right)(x_k + \delta_{\mathrm{t},k}) + \boldsymbol{\delta}_{\mathrm{r}} + \mathbf{w}. \quad (8)$$

Now, given the RBM, the overall channel vector is $\mathbf{h}_k = \mathbf{h}_{\mathrm{d},k} + \mathbf{H}_1 \mathbf{\Theta} \widetilde{\mathbf{\Theta}} \mathbf{h}_{2,k}$, distributed as $\mathbf{h}_k \sim \mathcal{CN}(0, \mathbf{R}_k)$, where $\mathbf{R}_k = \beta_{\mathrm{d},k}\mathbf{R}_{\mathrm{BS},k} + \beta_{2,k}\mathbf{H}_1 \mathbf{\Theta} \widetilde{\mathbf{R}}_{\mathrm{IRS},k} \mathbf{\Theta}^{\mathsf{H}} \mathbf{H}_1^{\mathsf{H}}$ with $\widetilde{\mathbf{R}}_{\mathrm{IRS},k}$ given by

$$\widetilde{\mathbf{R}}_{\mathrm{IRS},k} = \mathbb{E}[\widetilde{\mathbf{\Theta}} \mathbf{R}_{\mathrm{IRS},k} \widetilde{\mathbf{\Theta}}^{\mathsf{H}}] \quad (9)$$

$$= \begin{bmatrix} r_{11} & r_{12}\mathbb{E}_{\tilde{\theta}}[e^{j\tilde{\theta}_1 - j\tilde{\theta}_2}] & \ldots & r_{1N}\mathbb{E}_{\tilde{\theta}}[e^{j\tilde{\theta}_1 - j\tilde{\theta}_N}] \\ r_{21}\mathbb{E}_{\tilde{\theta}}[e^{j\tilde{\theta}_2 - j\tilde{\theta}_1}] & r_{22} & \ldots & r_{2N}\mathbb{E}_{\tilde{\theta}}[e^{j\tilde{\theta}_2 - j\tilde{\theta}_N}] \\ \vdots & \vdots & \ddots & \vdots \\ r_{N1}\mathbb{E}_{\tilde{\theta}}[e^{j\tilde{\theta}_N - j\tilde{\theta}_1}] & r_{N2}\mathbb{E}_{\tilde{\theta}}[e^{j\tilde{\theta}_N - j\tilde{\theta}_2}] & \ldots & r_{NN} \end{bmatrix} \quad (10)$$

$$= \begin{bmatrix} r_{11} & m^2 r_{12} & \ldots & m^2 r_{1N} \\ m^2 r_{21} & r_{22} & \ldots & m^2 r_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ m^2 r_{N1} & m^2 r_{N2} & \ldots & r_{NN} \end{bmatrix} \quad (11)$$

$$= m^2 \mathbf{R}_{\mathrm{IRS},k} + \left(1 - m^2\right) \text{diag}\left(\mathbf{R}_{\mathrm{IRS},k}\right) \quad (12)$$

$$= m^2 \mathbf{R}_{\mathrm{IRS},k} + \left(1 - m^2\right) \mathbf{I}_N. \quad (13)$$

In (10), r_{ij} is the ijth element of the correlation matrix $\mathbf{R}_{\mathrm{IRS},k}$. Also, in (11), we have exploited that $\tilde{\theta}_i$ are i.i.d. distributed with a symmetric PDF while we have substituted with m the corresponding CF. The next equation is written in a compact form in terms of $\mathbf{R}_{\mathrm{IRS},k}$. In (13), we have accounted for that in correlated Rayleigh fading, we have $r_{ii} = 1, \ \forall i$. Notably, this is a very useful equation describing the dependence on the IRS-HWIs, i.e., the phase noises from the IRS elements.

Remark 1: In the case of the uniform distribution, where the characteristic function is zero ($m = 0$), we obtain $\widetilde{\mathbf{R}}_{\mathrm{IRS},k} = \mathbf{I}_N$, i.e., there is no dependence on the phase noises. In such case, the overall covariance becomes $\mathbf{R}_k = \beta_{\mathrm{d},k}\mathbf{R}_{\mathrm{BS},k} + \beta_{2,k}\mathbf{H}_1\mathbf{H}_1^{\mathsf{H}}$, which obviously has no dependence on the RBM and cannot be optimized. Hence, the IRS cannot be exploited. Moreover, no knowledge of $\mathbf{R}_{\mathrm{IRS},k}$ is required at the BS. However, even in this case, the IRS still contributes with an additional signal to the receiver. Although it is expected to be weak, it is beneficial. Especially, when there is no direct signal. Note that these cases are very difficult to appear in practice. Specifically, regarding the independent Rayleigh assumption, it is uncommon to appear as mentioned in [35] while always there will be some knowledge and control in the reflection at the IRS, which means that the uniform distribution is not meaningful in practice (see Sec. II.C.2). On the contrary, any other circular PDF for the description of the phase errors allows studying the impact of these errors, and mostly, taking advantage of the IRS.

III. CE WITH HWIs

In practice, a BS does not have perfect CSI but estimates its channel by a TDD operation including an uplink training phase with pilot symbols [24]. Differently to conventional MISO with/without relay systems, the IRS implemented by means of passive elements, is not able to send pilots to the BS for CE or process the received pilot symbols from the UEs to obtain the corresponding estimated channels. Contrary to existing works providing separately the estimated direct and cascaded channels [19], we provide the estimate of the overall channel. For example, compared to [17] and [19], we perform the CE in a single phase instead of $N+1$ phases. In particular, the former is known as ON/OFF channel estimation and addresses a single UE setting while the latter assumes multiple UEs. Hence, the achievable SE in our case is higher since the pre-log factor in the SE is lower (lower training overhead). Also, we achieve to derive the covariance of the estimated cascaded channel vector while, in [19], it is assumed unknown.

The CE protocol assumes that the total uplink training phase has a duration of τ sec. Let the UEs transmit orthogonal pilot sequences. Especially, we denote by $\mathbf{x}_{p,k} = [x_{p,k,1}, \ldots, x_{p,k,\tau}]^{\mathsf{T}} \in \mathbb{C}^{\tau\times 1}$ the pilot sequence of UE k with $\mathbf{x}_{p,k}^{\mathsf{H}}\mathbf{x}_{p,l} = 0 \ \forall k \neq l$ and $\mathbf{x}_{p,k}^{\mathsf{H}}\mathbf{x}_{p,k} = \tau P$ joules, where $P = |x_{p,k,i}|^2, \ \forall k, i$ is the common average transmit power per UE during the training phase.

The received signal at the BS with T-HWIs and IRS-HWIs during the training period is given by

$$\mathbf{Y}^{\mathrm{tr}} = \sum_{i=1}^{K} \mathbf{h}_i \left(\mathbf{x}_{\mathrm{p},i}^{\mathsf{H}} + \boldsymbol{\delta}_{\mathrm{t},i}^{\mathsf{H}}\right) + \mathbf{\Delta}_{\mathrm{r}} + \mathbf{W}^{\mathrm{tr}}, \quad (14)$$

where $\boldsymbol{\delta}_{\mathrm{t},i} \in \mathbb{C}^{M\times 1} \sim \mathcal{CN}\left(\mathbf{0}, \kappa_{\mathrm{UE}} P \mathbf{I}_\tau\right)$ is the $\tau \times 1$ additive transmit HWI vector while $\mathbf{\Delta}_{\mathrm{r}} \in \mathbb{C}^{M\times\tau}$ is the additive receive HWI matrix where each column is distributed as $\boldsymbol{\delta}_{\mathrm{r}} \in \mathbb{C}^{M\times 1} \sim$

[4] We focus on the main IRS impairment, being the imperfection of phases configuration by assuming unity reflection amplitude, i.e., full signal reflection as in prior works e.g., see [2]–[4]. However, recently, it was suggested that the reflection amplitude can be phase-dependent due to hardware limitations [43], which requires a separate analysis and is left for future work.

$\mathcal{CN}(\mathbf{0}, \boldsymbol{\Upsilon})$ with $\boldsymbol{\Upsilon} = \kappa_{\rm BS} P \sum_{i=1}^{K} \mathbf{I}_M \circ \mathbf{h}_i \mathbf{h}_i^{\rm H}$. Note that the phase noise is hidden inside the expression of $\mathbf{h}_k$. In addition, $\mathbf{W}^{\rm tr} \in \mathbb{C}^{M \times \tau}$ is the AWGN matrix at the BS with independent columns, each one distributed as $\mathcal{CN}\left(\mathbf{0}, \sigma^2 \mathbf{I}_M\right)$.

The received training signal at the BS, given by (14), is multiplied by the transmitted training sequence from UE k to eliminate the interference caused by other UEs, and obtain

$$\mathbf{r}_k = \mathbf{h}_k + \sum_{i=1}^{K} \frac{\tilde{\delta}_{{\rm t},i}}{\tau P} \mathbf{h}_i + \frac{\tilde{\boldsymbol{\delta}}_{\rm r} + \mathbf{w}_k}{\tau P}, \tag{15}$$

where $\tilde{\delta}_{{\rm t},i} = \boldsymbol{\delta}_{{\rm t},i}^{\rm H} \mathbf{x}_{{\rm p},k}$, $\tilde{\boldsymbol{\delta}}_{\rm r} = \boldsymbol{\Delta}_{\rm r} \mathbf{x}_{{\rm p},k}$, and $\mathbf{w}_k = \mathbf{W}^{\rm tr} \mathbf{x}_{{\rm p},k}$.

Proposition 1: The LMMSE estimate of the overall channel $\mathbf{h}_k$ is given by

$$\hat{\mathbf{h}}_k = \mathbf{R}_k \mathbf{Q}_k \mathbf{r}_k, \tag{16}$$

where $\mathbf{Q}_k = \left(\mathbf{R}_k + \frac{\kappa_{\rm UE}}{\tau} \sum_{i=1}^{K} \mathbf{R}_i + \frac{\kappa_{\rm BS}}{\tau} \sum_{i=1}^{K} \mathbf{I}_M \circ \mathbf{R}_i + \frac{\sigma^2}{\tau P} \mathbf{I}_M\right)^{-1}$, and $\mathbf{r}_k$ is the noisy observation of the effective overall channel from UE k given by (15).

Proof: The proof is provided in Appendix A. ■

According to the property of orthogonality of LMMSE estimation, the overall perfect channel is given by

$$\mathbf{h}_k = \hat{\mathbf{h}}_k + \tilde{\mathbf{h}}_k, \tag{17}$$

where $\hat{\mathbf{h}}_k$ and $\tilde{\mathbf{h}}_k$ have zero mean and variances $\boldsymbol{\Psi}_k = \mathbf{R}_k \mathbf{Q}_k \mathbf{R}_k$ and $\tilde{\boldsymbol{\Psi}}_k = \mathbf{R}_k - \boldsymbol{\Psi}_k$, respectively. Contrary to conventional estimation theory concerning independent Gaussian noise, $\hat{\mathbf{h}}_k$ and $\tilde{\mathbf{h}}_k$ are neither independent nor jointly complex Gaussian vectors because the effective distortion noises are not Gaussian, e.g., $\mathbf{h}_k \tilde{\delta}_{{\rm t},k}$ is the product between two Gaussian variables. However, $\hat{\mathbf{h}}_k$ and $\tilde{\mathbf{h}}_k$ are uncorrelated and each of them has zero mean [24]. In the unrealistic case of perfect HWIs, the LMMSE estimator of the overall channel vector coincides with the optimal MMSE estimator. Notably, the CE can be easily generalized to include other fading models such as independent Rayleigh fading, where $\mathbf{R}_{{\rm IRS},k} = \mathbf{I}_N$ and $\mathbf{R}_{{\rm BS},k} = \mathbf{I}_M$.

Remark 2: A comparison with other CE schemes is difficult since the majority of works such as [18] does not yield analytical expressions, while the proposed method is indicated for future closed-form manipulations. Compared to [19] requiring $N+1$ subphases, our proposed method has a lower training overhead requiring only one phase and has achieved to obtain the estimated cascaded channel vector while, therein, only the estimated channel (scalar) concerning each element was obtained.

Remark 3: Generally, the covariances $\mathbf{R}_k$, $\boldsymbol{\Psi}_k$, and $\tilde{\boldsymbol{\Psi}}_k$ depend on both the T-HWIs and the IRS-HWIs. In the special case of uniformly distributed phase errors, these covariances do not depend on these errors or the reflect phase matrix $\boldsymbol{\Theta}$. Then, we can not take benefit from any IRS optimization to minimize the estimation error and achieve better estimation.

The study of the NMSE is insightful. Specifically, we define

$$\text{NMSE}_k = \frac{\text{tr}\left(\mathbb{E}[(\hat{\mathbf{h}}_k - \mathbf{h}_k)(\hat{\mathbf{h}}_k - \mathbf{h}_k)^{\rm H}]\right)}{\text{tr}\left(\mathbb{E}[\mathbf{h}_k \mathbf{h}_k^{\rm H}]\right)} \tag{18}$$

$$= 1 - \frac{\text{tr}(\boldsymbol{\Psi}_k)}{\text{tr}(\mathbf{R}_k)}. \tag{19}$$

The T-HWIs are found inside $\boldsymbol{\Psi}_k$ in terms of $\mathbf{Q}_k$ while the IRS-HWIs (phase noise) appear inside $\mathbf{R}_k$. From (19), we observe that an increase of the T-HWIs results in the increase of the NMSE_k. Moreover, according to Remark 1, if the phase errors are uniformly distributed, the NMSE does not depend on the RBM and the NMSE can not be optimized. A similar observation takes place if uncorrelated Rayleigh fading is assumed.

IV. Uplink Data Transmission with HWIs

In this section. we focus on the derivation of the uplink achievable sum SE of a practical IRS-aided MU-MISO setup with HWIs. The received signal by the BS can be written as

$$\mathbf{y} = \sum_{i=1}^{K} \mathbf{h}_i \left(s_i + \delta_{{\rm t},i}\right) + \boldsymbol{\delta}_{\rm r} + \mathbf{n}, \tag{20}$$

where $\mathbf{n} \sim \mathcal{CN}\left(\mathbf{0}, \sigma^2 \mathbf{I}_M\right)$ is the AWGN at the BS and phase noises are found inside the expression of $\mathbf{h}_i$, while $\delta_{{\rm t},i}$ and $\boldsymbol{\delta}_{\rm r}$ correspond to the T-HWIs.

A. Achievable SE

The BS estimates s_k from UE k by means of (20) in terms of linear single-user detection by applying the receive combining vector $\mathbf{v}_k$ as $\hat{s}_k = \mathbf{v}_k^{\rm H} \mathbf{y}$. Moreover, we exploit the use-and-then-forget (UatF) bound, suggested for systems with a large number of antennas ($M > 8$) [24], in order to obtain a closed-form expression of the SE. Note that this bound can be applied with different channel estimators (not only LMMSE) and decoders. Specifically, $\hat{s}_k$ can be rewritten as

$$\hat{s}_k = \underbrace{\sqrt{\rho_k} \mathbb{E}\left\{\mathbf{v}_k^{\rm H} \mathbf{h}_k\right\} s_k}_{\text{DS}_k} + \underbrace{\sqrt{\rho_k}\left(\mathbf{v}_k^{\rm H} \mathbf{h}_k - \mathbb{E}\left\{\mathbf{v}_k^{\rm H} \mathbf{h}_k\right\}\right) s_k}_{\text{BU}_k}$$
$$+ \sum_{i \neq k}^{K} \underbrace{\sqrt{\rho_i} \mathbf{v}_k^{\rm H} \mathbf{h}_i s_i}_{\text{MUI}_{ik}} + \sum_{i=1}^{K} \underbrace{\mathbf{v}_k^{\rm H} \mathbf{h}_i \delta_{{\rm i},k}}_{\text{TD}_i} + \underbrace{\mathbf{v}_k^{\rm H} \boldsymbol{\delta}_{\rm r}}_{\text{RD}_k} + \underbrace{\mathbf{v}_k^{\rm H}}_{\text{RN}_k}, \tag{21}$$

where DS_k, BU_k, MUI_{ik} express the desired signal (DS) part, the beamforming gain uncertainty (BU), and each term of the sum describing the MU interference (MUI). Also, TD_i, RD_k, and RN_k express the transmit distortion, the receive distortion, and the receiver AWGN noise. Next, by applying a standard bound technique assuming worst-case uncorrelated additive noise for the inter-user interference and the distortion noise [44], we derive a lower bound on the uplink average SE in bps/Hz, which is known as the use-and-then-forget bound in the massive MIMO (mMIMO) literature [24]. In particular, the achievable SE is given by

$$\text{SE}_k = \frac{\tau_{\rm c} - \tau}{\tau_{\rm c}} \log_2\left(1 + \gamma_k\right), \tag{22}$$

where the pre-log fraction expresses the percentage of samples per coherence block used for uplink data transmission and $\gamma_k = \frac{S_k}{I_k}$ is the uplink SINR with

$$S_k = |\text{DS}_k|^2, \tag{23}$$

$$I_k = \mathbb{E}\left\{|\text{BU}_k|^2\right\} + \sum_{i \neq k}^{K} \mathbb{E}\left\{\text{MUI}_{ik}\right\} + \sum_{i=1}^{K} \mathbb{E}\left\{|\text{TD}_i|^2\right\}$$
$$+ \mathbb{E}\left\{|\text{RD}_k|^2\right\} + \mathbb{E}\left\{|\text{RN}_k|^2\right\} \tag{24}$$

describing the desired signal power and the interference plus noise power. For the sake of further convenience, we denote by $\sigma_{\rm UE}^2 = \sum_{i=1}^K \mathbb{E}\left\{|{\rm TD}_i|^2\right\}$ and $\sigma_{\rm BS}^2 = \mathbb{E}\left\{|{\rm RD}_k|^2\right\}$ the variances of the additive transmit and receive HWIs from all transmit UEs and at the output of the decoder.

Generally, MRC and conventional MMSE decoders are the most common linear receivers for the uplink of next-generation systems such as mMIMO [45], [46]. However, the expectations in (23) and (24) cannot be derived in closed-form in the case of the optimal MMSE receiver except if the deterministic equivalent analysis is applied [21], [45], [46]. Also, the next step that includes the optimization with respect to reflection coefficients would be quite intractable. Hence, we focus on the derivation of a closed-form SINR by applying MRC decoding, which can be obtained even for a finite number of BS antennas. Thus, below we assume $\mathbf{v}_k = \hat{\mathbf{h}}_k$. However, given the higher performance expected by MMSE decoding, its application in the study of HWIs with statistical CSI according to the proposed methodology is the topic of ongoing work.

Theorem 1: Given the RBM $\mathbf{\Theta}$, the uplink achievable SINR of UE k with MRC decoding in an IRS-assisted MU-MISO system, accounting for imperfect CSI and HWIs, is given by

$$\gamma_k = \frac{\bar{S}_k}{\bar{I}_k}, 2 \tag{25}$$

where

$$\bar{S}_k = \rho_k |{\rm tr}\left(\mathbf{\Psi}_k\right)|^2, \tag{26}$$

$$\bar{I}_k = (1+\kappa_{\rm UE})\left(\sum_{i=1}^K \rho_i {\rm tr}(\mathbf{\Psi}_k \mathbf{R}_i) - \rho_k {\rm tr}\left(\mathbf{\Psi}_k^2\right)\right) + \rho_k \kappa_{\rm UE} |{\rm tr}(\mathbf{\Psi}_k)|^2 + \boldsymbol{\kappa}_{\rm BS}\left(\rho_k |{\rm tr}(\mathbf{I}_M \circ \mathbf{\Psi}_k)|^2 + \sum_{i=1}^K \rho_i {\rm tr}\left((\mathbf{I}_M \circ \mathbf{R}_i)\mathbf{\Psi}_k\right)\right) + \sigma^2 {\rm tr}(\mathbf{\Psi}_k). \tag{27}$$

Proof: The proof is provided in Appendix B. ■

Remark 4: Theorem 1 provides the uplink achievable SINR with MRC under imperfect CSI in closed-form. Notably, it shows the impact of the unavoidable HWIs. Especially, it depends directly on the T-HWIs by means of $\boldsymbol{\kappa}_{\rm BS}$ and $\boldsymbol{\kappa}_{\rm UE}$. The impact of the phase noise appears indirectly through the covariance matrices. Moreover, the expression of γ_k depends only on slowly-varying large-scale statistics.

Based on γ_k, provided by Theorem 1, the system (sum) SE in bps/Hz is obtained as

$$\mathcal{R} = \frac{\tau_{\rm c} - \tau}{\tau_{\rm c}} \sum_{i=1}^K \log_2\left(1+\gamma_i\right). \tag{28}$$

B. IRS Design Problem: Formulation and Solution

IRS-aided architectures require to design the corresponding RBM, found inside the covariance matrices, in order to maximize the sum SE given by (28). Hence, by resorting to the common assumption of infinite resolution phase shifters, herein, we formulate and solve the RB design problem under MRC and realistic conditions accounting for imperfect CSI and HWIs as

$$\begin{aligned} (\mathcal{P}1) \qquad & \max_{\mathbf{\Theta}} \quad \mathcal{R} \\ & \text{s.t} \quad |\phi_n| = 1, \quad n = 1, \ldots, N, \end{aligned} \tag{29}$$

with $\mathcal{R}$ given by (28) and $\phi_n = \exp\left(j\theta_n\right)$ are the elements of $\mathbf{\Theta}$. Obviously, $(\mathcal{P}1)$ is a non-convex maximization problem with respect to $\mathbf{\Theta}$ with a unit-modulus constraint regarding ϕ_n.

Remark 5: If the phase noise is uniformly distributed, the covariance matrices will not include the RBM according to Remark 3. Hence, the SINR/SE cannot be optimized, and the IRS does not serve its purpose.

Remark 6: Given that HWIs at the transceiver and IRS degrade the performance, the use of cheaper (lower quality) hardware will have a direct impact on the QoS. In such a case, a better RB design is suggested to compensate for the loss and improve the performance.

Taking the expression of γ_k into account, the optimization problem takes the form of a constrained maximization problem with a solution given by means of projected gradient ascent until converging to a stationary point as in [6]. At every step, we project the solution onto the closest feasible point satisfying the unit-modulus constraint concerning ϕ_n. In more detail, the procedure assumes the vectors $\mathbf{s}^i = [\phi_1^i, \ldots, \phi_N^i]^{\intercal}$ including the induced phases at step i. The next iteration point, resulting in the increase of $\mathcal{R}$ towards to its convergence, is given by

$$\tilde{\mathbf{s}}^{i+1} = \mathbf{s}^i + \mu \mathbf{q}^i, \tag{30}$$

$$\mathbf{s}^{i+1} = \exp\left(j \arg\left(\tilde{\mathbf{s}}^{i+1}\right)\right), \tag{31}$$

where μ is the step size and $\mathbf{q}^i$ describes the adopted ascent direction at step i. In particular, we have $[\mathbf{q}^i]_n = \frac{\partial \mathcal{R}}{\partial \phi_n^*}$, which is obtained by Proposition 2 below. The suitable step size is computed at each iteration by means of the backtracking line search [47]. The solution of the problem, described by (30) and (31), is found based on the projection problem $\min_{|\phi_n|=1, n=1,\ldots,N} \|\mathbf{s} - \tilde{\mathbf{s}}\|^2$ under the unit-modulus constraint. The outline of the algorithm is described by Algorithm 1.

Algorithm 1 Projected Gradient Ascent Algorithm for the IRS Design

1. **Initialisation**: $\mathbf{s}^0 = \exp\left(j\pi/2\right)\mathbf{1}_N$, $\mathbf{\Theta}^0 = {\rm diag}\left(\mathbf{s}^0\right)$, $\mathcal{R}^0 = f\left(\mathbf{\Theta}^0\right)$ given by (28); $\epsilon > 0$
2. **Iteration** i: **for** $i = 0, 1, \ldots,$ do
3. $[\mathbf{q}^i]_n = \frac{\partial \mathcal{R}}{\partial \phi_n^*}, n = 1, \ldots, N$, where $\frac{\partial \mathcal{R}}{\partial \phi_n^*}$ is given by Proposition 2;
4. **Find** μ by backtrack line search$(f\left(\mathbf{\Theta}^0\right), \mathbf{q}^i, \mathbf{s}^i)$ [47];
5. $\tilde{\mathbf{s}}^{i+1} = \mathbf{s}^i + \mu \mathbf{q}^i$;
6. $\mathbf{s}^{i+1} = \exp\left(j \arg\left(\tilde{\mathbf{s}}^{i+1}\right)\right)$; $\mathbf{\Theta}^{i+1} = \alpha {\rm diag}\left(\mathbf{s}^{i+1}\right)$;
7. $\mathcal{R}^{i+1} = f\left(\mathbf{\Theta}^{i+1}\right)$;
8. **Until** $\|\mathcal{R}^{i+1} - \mathcal{R}^i\|^2 < \epsilon$; **Obtain** $\mathbf{\Theta}^* = \mathbf{\Theta}^{i+1}$;
9. **end for**

The convergence of the proposed algorithm to a local maximum can be guaranteed because it is bounded due to the power constraint and it increases by setting $[\mathbf{q}^i]_n = \frac{\partial \mathcal{R}}{\partial \phi_n^*}$, where the backtracking line search is used to find a suitable step size.

Proposition 2: The derivative of $\mathcal{R}$ with respect to ϕ_n is provided by

$$\frac{\partial \mathcal{R}}{\partial \phi_n} = \sum_{k=1}^{K} \frac{\frac{\partial \bar{S}_k}{\partial \phi_n^*} \bar{I}_k - \bar{S}_k \frac{\partial \bar{I}_k}{\partial \phi_n^*}}{\ln(2) \bar{I}_k^2 \left(1 + \frac{\bar{S}_k}{\bar{I}_k}\right)}, \tag{32}$$

where $\bar{S}_k$, $\bar{I}_k$ follow by (26), (27) while

$$\frac{\partial \bar{S}_k}{\partial \phi_n^*} = 2\rho_k \operatorname{tr}(\mathbf{\Psi}_k) \mathbf{L}(\mathbf{R}_k, \mathbf{R}_k, \mathbf{I}_M), \tag{33}$$

$$\begin{aligned}
\frac{\partial \bar{I}_k}{\partial \phi_n^*} &= (1+\kappa_{\mathrm{UE}}) \Big(\sum_{i=1}^{K} \rho_i \big(\mathbf{L}\big(\mathbf{R}_i \mathbf{R}_k, \mathbf{R}_i \mathbf{R}_k, \mathbf{R}_i\big) \\
&+ \alpha \beta_{2,k} [\mathbf{H}_1^{\mathrm{H}} \mathbf{\Psi}_k \mathbf{H}_1 \mathbf{\Theta} \mathbf{R}_{\mathrm{IRS},k}]_{n,n} \big) \\
&- 2\rho_k \mathbf{L}\big(\mathbf{\Psi}_k \mathbf{R}_k, \mathbf{\Psi}_k \mathbf{R}_k, \mathbf{\Psi}_k\big) \Big) \\
&+ \boldsymbol{\kappa}_{\mathrm{BS}} \big(2\rho_k \operatorname{tr}\big(\mathbf{I}_M \circ \mathbf{\Psi}_k\big) \mathbf{L}\big(\mathbf{R}_k, \mathbf{R}_k, \mathbf{I}_M\big) \\
&+ \sum_{i=1}^{K} \rho_i \big(\alpha \beta_{2,k} [\mathbf{H}_1^{\mathrm{H}} \mathbf{\Psi}_k \mathbf{H}_1 \mathbf{\Theta} \mathbf{R}_{\mathrm{IRS},k}]_{n,n} \\
&+ \mathbf{L}\big(\big(\mathbf{I}_M \circ \mathbf{R}_i\big) \mathbf{R}_k, \big(\mathbf{I}_M \circ \mathbf{R}_i\big) \mathbf{R}_k, \mathbf{I}_M \circ \mathbf{R}_i \big) \big) \big) \\
&+ 2\rho_k \kappa_{\mathrm{UE}} \operatorname{tr}\big(\mathbf{\Psi}_k\big) \mathbf{L}\big(\mathbf{R}_k, \mathbf{R}_k, \mathbf{I}_M\big) + \sigma^2 \mathbf{L}\big(\mathbf{R}_k, \mathbf{R}_k, \mathbf{I}_M\big)
\end{aligned} \tag{34}$$

with $\mathbf{L}(\mathbf{A}, \mathbf{B}, \mathbf{C})$ given by (81) for any $\mathbf{A} \in \mathbb{C}^{N \times N}$, $\mathbf{B} \in \mathbb{C}^{N \times N}$, and $\mathbf{C} \in \mathbb{C}^{N \times N}$.

Proof: The proof of Proposition 2 is given in Appendix C. ■

The RBM beamforming design, based on the gradient ascent, results in an outstanding performance since the gradient ascent is obtained in a closed-from with low computational complexity based on simple matrix operations. In particular, the complexity of (32) is $\mathcal{O}\left(MN^2 + M^2N + M^2K\right)$. Obviously, it depends on all fundamental system parameters, i.e., K, M, and N but with a higher (square) dependence on M and N.

Remark 7: If we do not have a closed-form expression for the SE, we cannot apply the proposed method. Also, although the proposed algorithm, given by (30) and (31), does not provide a global optimum but a locally optimal solution due to the non-convexity of the initial optimization problem with respect to the phase shifts, it offers a good preliminary tool to study IRS-aided systems under realistic conditions in terms of imperfect CSI and HWIs.

Remark 8: The dependence of the algorithm on the large-scale channel statistics achieves a reduction in the signal exchange overhead between the IRS controller and the BS since it will take place every several coherence intervals defined by the variation of these statistics. On the contrary, on models, relying on the instantaneous CSI, the optimization should take place at every coherence interval, which results in large overhead, especially, when the IRS is large. We highlight that the proposed method can be exploited in both low-speed and fast-speed scenarios. The only difference is that in fast-speed scenarios, the large-scale statistics change faster. For this reason, in such cases, the optimization should take place more frequently. Also, the simple expression of $\frac{\partial \mathcal{R}}{\partial \phi_n^*}$ results in a significant decrease of the computational complexity. These two reasons make the proposed method quite beneficial.

Remark 9: The proposed methodology for the RB design, i.e., the optimization of the phase matrix after having derived the performance expression in terms of large-scale statistics, could also be used in the case of negligence of HWIs, where another property of an IRS-assisted system would be the main topic of study. In such a case, the optimization in terms of the derivative will be simplified even more since the additional terms, concerning the T-HWIs which depend on the overall channel (the optimization variables ϕ_n), will be omitted.

V. Numerical Results

In this section, we depict and discuss the analytical results corresponding to the uplink performance in terms of CE and achievable sum SE of an IRS-aided MU-MISO system with imperfect CSI and HWIs. Monte-Carlo (MC) simulations (10^3 independent channel realizations) represented by "✕" marks in Figs. 2 and 4 below, corroborate our analysis and the tightness of the UatF bound. For the sake of comparison, we have modified [19] to describe the uplink transmission.

A. Simulation Setup

We consider a uniform linear array (ULA) of M antennas ($M = 16$) at the BS, assisted by an IRS with a uniform planar array (UPA) of N elements ($N = 60$) that serve $K = 5$ UEs. The spatial correlation coefficient for the IRS between elements n and n' corresponding to UE k is given by [6].

$$[\mathbf{R}_{\mathrm{IRS},k}]_{n,n'} = \mathbb{E}\big[\exp\big(j \frac{2\pi}{\lambda} d_{\mathrm{IRS}}(n-n') \sin\phi_k \sin\theta_k\big)\big], \tag{35}$$

where $d_{\mathrm{IRS}} = 0.5\lambda$ while ϕ_k and θ_k express the elevation and azimuth angles for UE k, and are generated by the Laplace and the Von Mises distribution, respectively. In the former case, we assume that the mean angle of departure and the spread are 90° and 8°, respectively. The latter distribution is generated with mean angle of departure 0 and spread 0.2. Regarding the parameters for the channel matrix $\mathbf{H}_1$ between the BS and the IRS, we assume $d_{\mathrm{BS}} = 0.5\lambda$ and $\theta_{1,n}$, $\psi_{1,n}$ are uniformly distributed between 0 to π and 0 to 2π, respectively. Also, $\theta_{2,n} = \pi - \theta_{1,n}$, $\psi_{2,n} = \pi + \psi_{1,n}$. The correlation matrix $\mathbf{R}_{\mathrm{BS},k}$ is obtained similar to [45]. Moreover, the overall path loss for the IRS-assisted link is given by [3], [5], [6]

$$\bar{\beta}_k = C_1 C_2 d_{\mathrm{BS-IRS}}^{-\alpha_1} d_{\mathrm{IRS-UE}_k}^{-\alpha_2}, \tag{36}$$

where $\beta_1 = \frac{C_1}{d_{\mathrm{BS-IRS}}^{\alpha_1}}$ and $\beta_{2,k} = \frac{C_2}{d_{\mathrm{IRS-UE}_k}^{\alpha_2}}$ are the channel attenuation coefficients between the BS and the IRS and between the IRS and UE k, respectively. Note that α_1 and $d_{\mathrm{BS-IRS}}$ are the path-loss exponent and distance concerning the link BS-to-IRS, while α_2 and $d_{\mathrm{IRS-UE}_k}$ are the path-loss exponent and distance concerning the link IRS-to-UE k. The parameters values are chosen relied on the 3GPP Urban Micro (UMi) scenario from TR36.814 for a carrier frequency of 2.5 GHz and noise level -80 dBm, where the path losses for $\mathbf{h}_{2,k}$ and $\mathbf{H}_1$ are generated based on the NLOS and LOS versions [48]. Hence, we have $\alpha_1 = 2.2$ and $\alpha_2 = 3.67$ while $d_{\mathrm{BS-IRS}} = 8$ m and $d_{\mathrm{IRS-UE}_k} = 60$ m. Also, $C_1 = 26$ dB, $C_2 = 28$ dB by assuming 5 dBi antennas at the BS and IRS while each UE includes a single 0dBi antenna [5]. For $\beta_{\mathrm{d},k}$, we assume the same parameters as for $\beta_{2,k}$, but we also consider an additional penetration loss equal to 15 dB. In addition, we

assume that the coherence bandwidth is $B_{\rm c} = 200$ KHz and the coherence time is $T_{\rm c} = 1$ ms, i.e., each coherence block consists of $\tau_{\rm c} = 200$ samples. Also, we assume $P = 6$ dB and the same value for $p_i = \rho$, $\forall i$ during the uplink data transmission. Note that $\sigma^2 = -174 + 10\log_{10} B_{\rm c}$.

For the study of the T-HWIs, we assume that we have an Analog-to-Digital Converter (ADC) at the BS, which quantizes the received signal to a b bit resolution. As a consequence, the receive distortion can be written as $\kappa_{\rm BS} = 2^{-2b}/\left(1-2^{-2b}\right)$, which gives $\kappa_{\rm BS} = 0.258^2$, 0.126^2, and 0.062^2 for $b = 2, 3$, and 4 bits, respectively [49], [50]. In other words, a smaller resolution results in more severe distortion. The same value is used for $\kappa_{\rm UE}$. Note that the trend in 5G networks and beyond is the use of lower precision ADCs. Regarding, the IRS-HWIs, if the Von Mises PDF is assumed to model the phase noise, the concentration parameter is set to $\kappa_{\tilde{\theta}} = 2$. Unless otherwise stated, this set of parameters is used during the simulations.

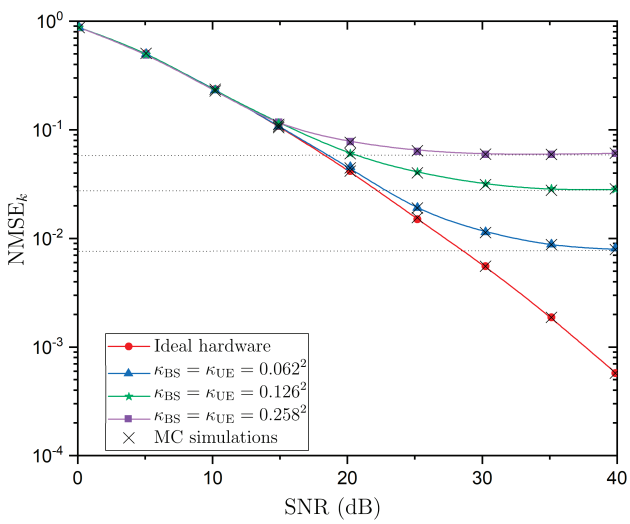


Fig. 2: NMSE of UE k versus the SNR of an IRS-assisted MIMO system with imperfect CSI ($M = 16$, $N = 60$, $K = 5$) for varying T-HWIs $\kappa_{\rm BS}$, $\kappa_{\rm UE}$ in the cases of uniform PDF for the phase noise or uncorrelated fading at the IRS (Analytical results and MC simulations).

Fig. 2 illustrates the relative estimation error per channel element, i.e., the normalized mean square error (NMSE) with respect to the uplink SNR for different values of the T-HWIs defining certain noise floors (asymptotic limits as $p \to \infty$). In addition, we show the result corresponding to perfect hardware. Obviously, this line decreases without bound. Moreover, it is shown that the error floors go higher with increasing the severity of T-HWIs. Even at mild values of T-HWIs, we observe that the NMSE approaches the corresponding floor after 20 dB, which means that IRS-assisted systems require a high SNR to operate since they are dependent on a conventional MU-MISO architecture. For the sake of exposition, we have assumed uniform phase noise or uncorrelated fading at the IRS, in order to avoid any RB optimization since the NMSE does not depend on $\mathbf{\Theta}$ in these cases (Remark 1). Optimization with respect to $\mathbf{\Theta}$ has been performed only in the case of the sum SE. Based on Remark 2, comparisons with other methods cannot be made at this stage. However, a comparison with respect to [19] takes place below (Fig. 3) in the case of achievable sum SE. Notably, MC simulations verify the analytical results.

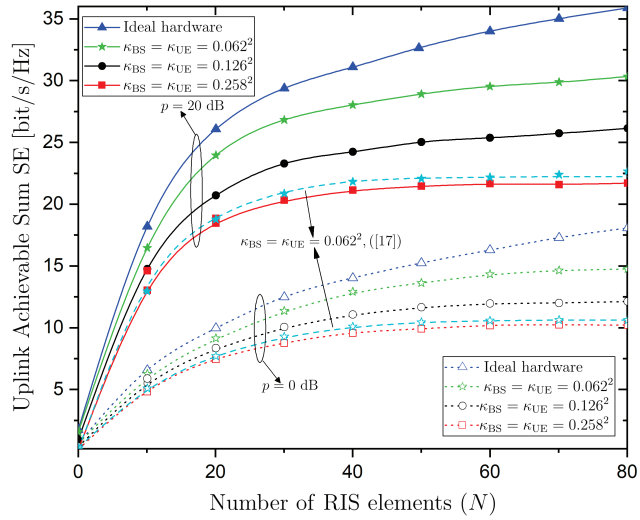


Fig. 3: Uplink achievable sum SE versus the number of IRS elements N of an IRS-assisted MIMO system with imperfect CSI ($M = 16$, $K = 5$) for varying T-HWIs $\kappa_{\rm BS}$, $\kappa_{\rm UE}$ and transmit power p.

In Fig. 3, we depict the achievable sum SE versus the number of IRS elements N for two different SNR values, $\rho = 0$ dB and $\rho = 20$ dB. Also, we have considered different values of T-HWIs. First, we observe an increase of $\mathcal{R}$ with N, which increases unboundedly in the case of ideal hardware, but saturates for imperfect hardware met in practice. As expected, the degradation is higher when T-HWIs are more severe, probably, in the case of cheaper hardware used for a cost-efficient implementation. Furthermore, the convergence speed is faster at the higher SNR group because the T-HWIs are power-dependent. Thus, the largest part of the gain is achieved at lower values of N. However, an increase of the IRS elements still allows for a further increase of $\mathcal{R}$. For the sake of comparison, we have considered the CE from [19] in terms of simulation ("dashed-star" lines) when $\kappa_{\rm BS} = \kappa_{\rm UE} = 0.258^2$. The achievable sum SE is much lower than the proposed method because of the high training overhead.

Fig. 4 illustrates the achievable sum SE with respect to the number of BS antennas M for varying T-HWIs and SNR. Notably, this figure resembles with the previous figure, i.e., $\mathcal{R}$ presents a similar dependence on M and N. Hence, when M grows large, the sum SE increases without limit, if perfect hardware is assumed while it appears ceilings in practice, where T-HWIs exist. In fact, lower hardware quality results in larger degradation. In addition, by increasing the SNR, the sum SE becomes larger. Also, the sum SE saturates faster in the instance of a larger SNR (20 dB). Thus, these two figures indicate that an IRS-assisted system performs better at higher SNR values, and with larger values of IRS elements and BS antennas. Note that the latter is further appealing since it agrees with the massive MIMO technology of which the implementation has already started. Moreover, the analytical results are accompanied by MC simulations showing the tightness and correctness of the lower bounds since they coincide.

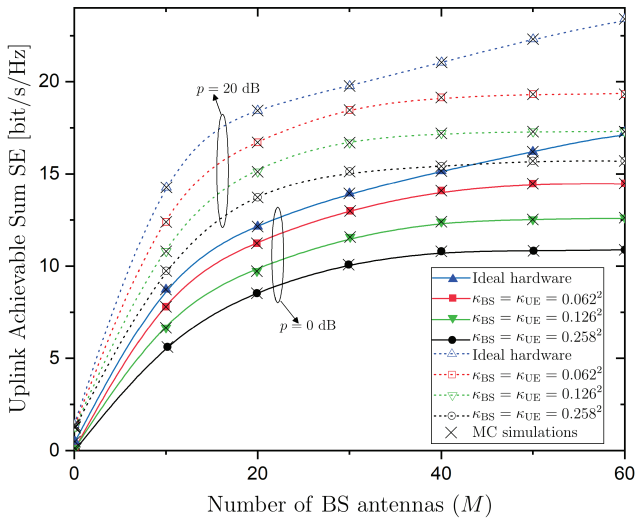


Fig. 4: Uplink achievable sum SE versus the number of BS antennas M of an IRS-assisted MIMO system with imperfect CSI ($N = 60$, $K = 5$) for varying T-HWIs $\kappa_{\rm BS}$, $\kappa_{\rm UE}$ and transmit power p (Analytical results and MC simulations).

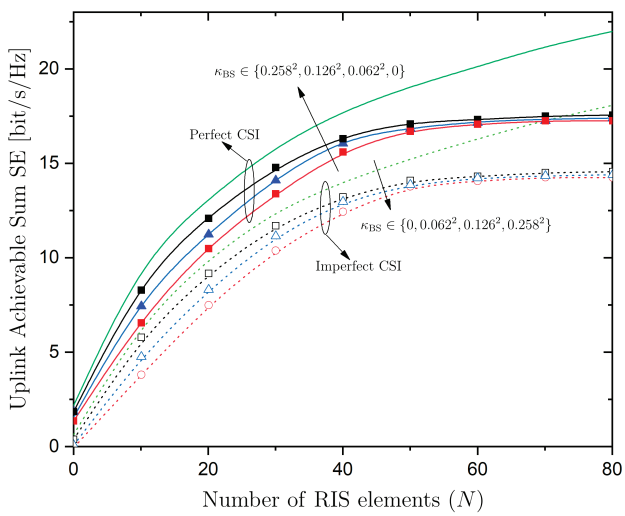


Fig. 5: Uplink achievable sum SE versus the number of IRS elements N of an IRS-assisted MIMO system ($M = 16$, $K = 5$) for varying BS distortion $\kappa_{\rm BS}$ and both cases of perfect and imperfect CSI.

Fig. 5 shows the achievable sum SE with respect to the number of IRS elements by varying the impact of the BS distortion $\kappa_{\rm BS}$ while the distortion at the UE side is assumed zero. Notably, the lines converge to the same finite limit as N increases, which means that the impact of $\kappa_{\rm BS}$ becomes negligible when N increases. In other words, despite the unavoidable existence of imperfect hardware, the IRS is suggested as N increases. Actually, it allows the use of low-cost hardware. Moreover, in the same figure, we have provided a comparison between perfect and imperfect CSI. We notice that, in both cases, the sum SE increases with the number of IRS elements while, again, the impact of $\kappa_{\rm BS}$ becomes negligible at large N. Obviously, in the realistic case of imperfect CSI, the achievable sum SE is lower, and the gap between the lines of ideal and imperfect hardware increases with larger N because the impact from the estimation error becomes larger.

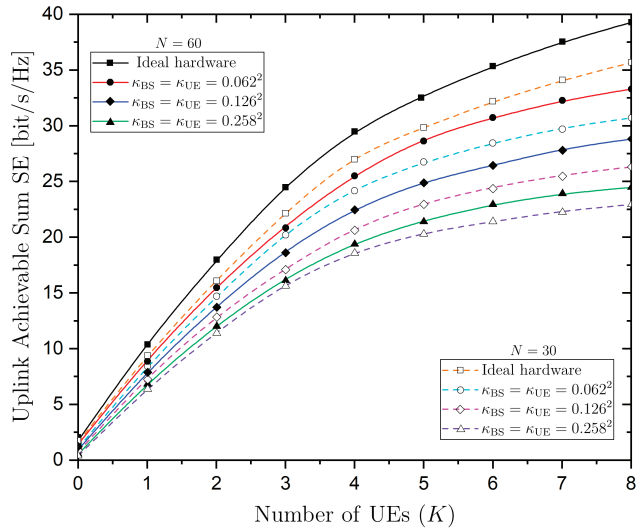


Fig. 6: Uplink achievable sum SE versus the number of UEs K of an IRS-assisted MIMO system ($M = 16$, $\rho = 20$ dB) for varying T-HWIs $\kappa_{\rm BS}$, $\kappa_{\rm UE}$ and IRS elements N.

Fig. 6 examines the achievable sum SE with respect to the number of UEs by varying the T-HWIs for different numbers of IRS elements. The sum SE increases with K almost linearly at the beginning but the gradient decreases as K increases. This result is reasonable since the increase of K increases the multi-user interference and the received distortion as described by (6). Also, a larger IRS in terms of N results in a larger sum SE as has been already shown.

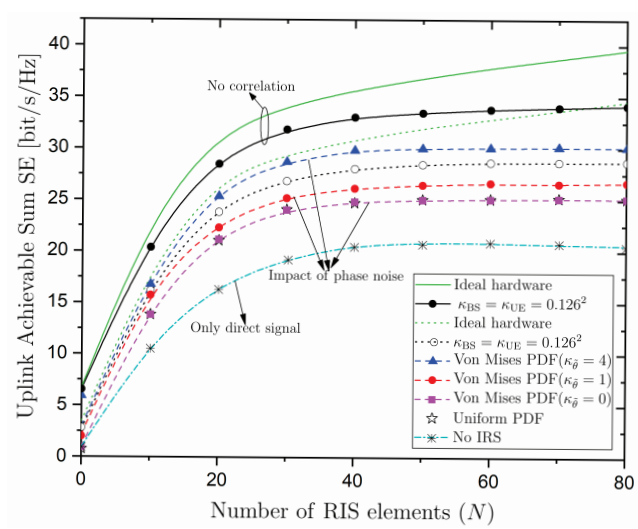


Fig. 7: Uplink achievable sum SE versus the number of IRS elements N of an IRS-assisted MIMO system with imperfect CSI ($M = 16$, $K = 5$) for varying IRS-HWIs and T-HWIs in the cases of correlated/non-correlated Rayleigh fading.

In Fig. 7, we show the impact of correlated Rayleigh fading on the achievable sum SE in the cases of correlation at both the BS and the IRS while varying the number of IRS elements. Also, we vary the quality of T-HWIs, and we show how for the same T-HWIs at the BS and UE sides, the sum SE

decreases in the case of correlated fading ("dot" lines) with comparison to no correlation ("solid" lines). Furthermore, this figure allows shedding light on the impact of phase noise at the IRS. In particular, when uniform phase noise is assumed ("star" symbols), $\mathcal{R}$ is the lowest because the design cannot take benefit (random fluctuations) from the IRS optimization since $\mathbf{R}_k$ does not depend on the phase matrix (see Rem. 5). However, in the case that the phase noise is distributed according to the Von Mises distribution ("dashed" lines), the achievable sum SE increases because the presence of the IRS becomes advantageous since it can adjust the phase shifts of its passive elements towards better performance. Especially, we have considered variation of the concentration parameter $\kappa_{\tilde{\theta}}$. As $\kappa_{\tilde{\theta}}$ decreases, the achievable sum SE decreases. Actually, we show that when $\kappa_{\tilde{\theta}} = 0$, the corresponding line coincides with the line describing the uniform PDF since the Von Mises PDF coincides with the uniform distribution in this case. Note that we have also depicted the performance in the absence of the IRS. From the figure, we can verify our observation in Rem. 1 explaining that the IRS contributes to the performance even in the worst-case IRS phase noise scenario since the line describing the case with no IRS is lower, i.e., the performance is worse.

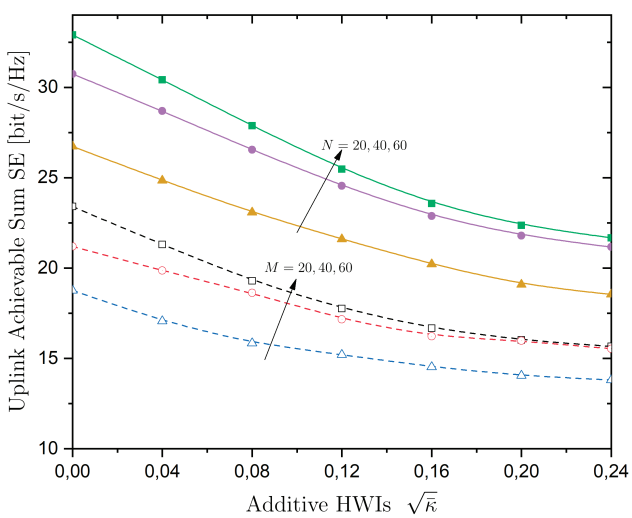
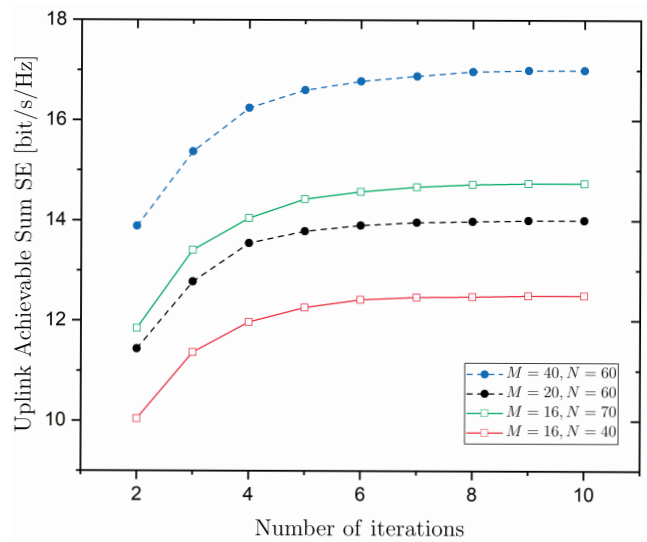
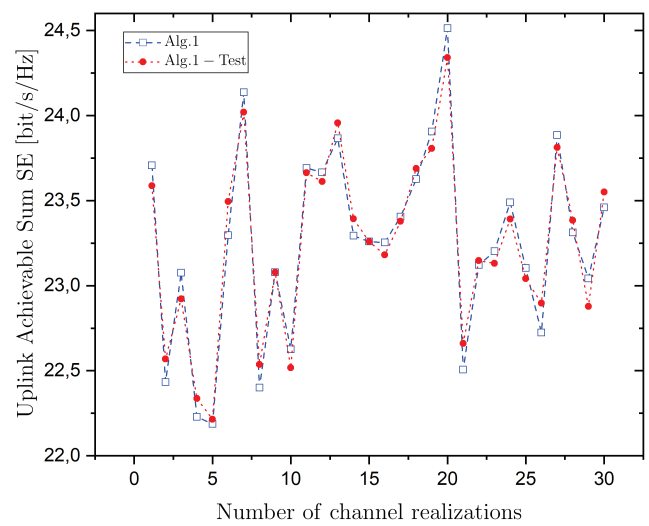


Fig. 8: Uplink achievable sum SE versus the T-HWIs quality $\sqrt{\bar{\kappa}}$ ($\rho = 20$ dB, $K = 5$) for varying BS antennas M and IRS elements N.



(a)



(b)

Fig. 9: Uplink achievable sum SE of an IRS-assisted MIMO system with imperfect CSI versus: (a) the number of iterations N ($K = 5$, $p = 0$dB, $\kappa_{\text{BS}} = \kappa_{\text{UE}} = 0.126^2$) for varying BS antennas M and IRS elements N; (b) 30 channel realizations ($M = 16$, $N = 20$, $K = 5$, $p = 20$dB, $\kappa_{\text{BS}} = \kappa_{\text{UE}} = 0.126^2$).

Fig. 8 shows the performance of the achievable sum SE versus $\sqrt{\bar{\kappa}}$, where $\kappa_{\text{BS}} = \bar{\kappa}$ and $\kappa_{\text{UE}} = \bar{\kappa} + 0.03$. We have assumed that the distortion at the UE is larger due to its simpler hardware. The impact of IRS phase noise is not considered as we focus on the impact of the T-HWIs. The "solid" and "dashed" lines correspond to the variation of M BS antennas and N IRS elements, respectively. The horizontal axis starts from the case of no T-HWIs at the BS, i.e., when $\bar{\kappa} = 0$ and ends with severe additive HWIs. We observe that as T-HWIs increase, the performance decreases. Moreover, we observe that at severe T-HWIs the variations of the M and N do not affect the performance. Also, we notice that the number of IRS elements N affects more the performance (higher sum SE) than the number of BS antennas M while the same variation with respect to N has a larger impact since the gaps between the solid lines are larger.

In Fig. 9.(a), we show the convergence of the proposed algorithm, i.e., Algorithm 1. In particular, we have depicted the uplink achievable sum SE versus the number of iterations for various sets of BS antennas and IRS elements. Notably, the algorithm converges fast in all cases. For example, when $M = 20$ and $N = 60$, the algorithm converges in 7 iterations. Moreover, we notice that by increasing the IRS and BS sizes in terms of their elements and antennas, respectively, more iterations are required for convergence because the amount of optimization variables increases and the relevant search space is enlarged. On top of this, an increase in terms of BS antennas or IRS elements results in higher complexity of each iteration

of the proposed algorithm as mentioned in Sec. IV-B.

The non-convexity of the optimization problem suggests that its solution depends on the initial point, i.e., different initial points result in different locally optimal solutions. Fig. 9.(b) investigates this dependence on the initializations by accounting for 30 channel realizations. The initialization of Alg. 1 assumes that $\mathbf{s}^0 = \exp(j\pi/2)\,\mathbf{1}_N$ as mentioned in its description. "Alg. 1-Test" in the figure assumes the best initial point out of 100 random initial points for each channel instance. The figure shows that different initializations result in different solutions and that the sum SE in both cases is almost the same, which means that this phase shifts selection for initialization is a good choice.

VI. Conclusion

In this paper, not only we studied the impact of both T-HWIs and IRS-HWIs on a general IRS-assisted MU-MISO system with imperfect CSI and correlated Rayleigh fading, but we also proposed a novel optimization methodology regarding the optimization of the RBM with low computational cost, being quite useful in IRS-assisted systems that have a large number of elements. In particular, we obtained the LMMSE estimate of the channel with T-HWIs and IRS-HWIs. Moreover, we derived the uplink achievable sum SE with MRC in closed form, being dependent only on large-scale statistics, and performed high computationally efficient optimization with respect to the IRS RBM. In general, we provided a methodology resulting in analytical and tractable expressions being advantageous over previous works as shown by the simulation results. Furthermore, we evaluated the impact of HWIs at both the transceiver and the IRS on the system SE, and shed insightful light on their interplay with other system parameters towards efficient IRS design. Remarkably, this work opens new research directions for IRS-assisted systems such as the studies of energy efficiency and power scaling laws with imperfect CSI and T-HWIs.

Appendix A
Proof of Proposition 1

According to [51, Ch. 12] The LMMSE estimator of $\mathbf{h}_k$ is obtained by $\hat{\mathbf{h}}_k = \mathbf{F}\mathbf{r}_k$, where $\mathbf{F}$ is derived my minimizing $\mathrm{tr}\left(\mathbb{E}\left[\left(\hat{\mathbf{h}}_k - \mathbf{h}_k\right)\left(\hat{\mathbf{h}}_k - \mathbf{h}_k\right)^{\mathsf{H}}\right]\right)$ as

$$\mathbf{F} = \mathbb{E}\left[\mathbf{r}_k\mathbf{h}_k^{\mathsf{H}}\right]\left(\mathbb{E}\left[\mathbf{r}_k\mathbf{r}_k^{\mathsf{H}}\right]\right)^{-1}. \tag{37}$$

Given that the additive distortions and the receiver noise are uncorrelated with overall channel $\mathbf{h}_k$, the first term of (37) becomes

$$\mathbb{E}\left[\mathbf{r}_k\mathbf{h}_k^{\mathsf{H}}\right] = \mathbb{E}\Big[\big(\mathbf{h}_k + \sum_{i=1}^{K}\frac{\tilde{\delta}_{\mathrm{t},i}}{\tau P}\mathbf{h}_i + \frac{\tilde{\boldsymbol{\delta}}_{\mathrm{r}} + \mathbf{w}_k}{\tau P}\big)\mathbf{h}_k^{\mathsf{H}}\Big] \tag{38}$$

$$= \mathbb{E}\left[\mathbf{h}_k\mathbf{h}_k^{\mathsf{H}}\right] = \mathbf{R}_k. \tag{39}$$

Regarding the second term, we have

$$\mathbb{E}\left[\mathbf{r}_k\mathbf{r}_k^{\mathsf{H}}\right] = \mathbf{R}_k + \frac{\kappa_{\mathrm{UE}}}{\tau}\sum_{i=1}^{K}\mathbf{R}_i + \frac{\kappa_{\mathrm{BS}}}{\tau}\sum_{i=1}^{K}\mathbf{I}_M\circ\mathbf{R}_i + \frac{\sigma^2}{\tau P}\mathbf{I}_M, \tag{40}$$

where we have taken into account that the additive distortions and the receiver noise are uncorrelated with each other. Also, we have used that the variance of $\tilde{\boldsymbol{\delta}}_{\mathrm{r}}$ is $\tau P\boldsymbol{\Upsilon}$ with $\boldsymbol{\Upsilon} = \kappa_{\mathrm{BS}}P\sum_{i=1}^{K}\mathrm{tr}\left(\mathbf{R}_i\right)$. As a result, the LMMSE estimate is given by inserting (39) and (40) into (37) as

$$\hat{\mathbf{h}}_k = \mathbf{R}_k\mathbf{Q}_k\mathbf{r}_k. \tag{41}$$

Furthermore, the covariance matrix of the estimated channel is obtained as

$$\mathbb{E}\left[\hat{\mathbf{h}}_k\hat{\mathbf{h}}_k^{\mathsf{H}}\right] = \mathbf{R}_k\mathbf{Q}_k\mathbf{R}_k. \tag{42}$$

Appendix B
Proof of Theorem 1

For the derivation of γ_k for finite M, we recall each term of (23) and (24). Generally, we are going to apply a useful property suggesting that $\mathbf{x}^{\mathsf{H}}\mathbf{y} = \mathrm{tr}(\mathbf{y}\mathbf{x}^{\mathsf{H}})$ for any vectors $\mathbf{x}$, $\mathbf{y}$. First, we obtain the S_k given by (23). Specifically, the desired signal part DS_k (without ρ_k) is written as

$$\mathbb{E}\big[\hat{\mathbf{h}}_k^{\mathsf{H}}\mathbf{h}_k\big] = \mathrm{tr}\left(\mathbb{E}\left[\mathbf{h}_k\hat{\mathbf{h}}_k^{\mathsf{H}}\right]\right) \tag{43}$$

$$= \mathrm{tr}\left(\mathbb{E}\left[\mathbf{h}_k\mathbf{r}_k^{\mathsf{H}}\mathbf{Q}_k\mathbf{R}_k\right]\right) \tag{44}$$

$$= \mathrm{tr}\left(\boldsymbol{\Psi}_k\right), \tag{45}$$

where, in (44), we have used (16), while the last step is obtained by applying the expectation between $\mathbf{h}_k$ and $\mathbf{r}_k$ and by considering that the mean value of $\tilde{\delta}_{\mathrm{t},k}$ is zero.

Regarding the second-order moment in the denominator, expressing the MU interference part MUI_{ik} for $i \neq k$, it is written as

$$\mathbb{E}\big[\big|\hat{\mathbf{h}}_k^{\mathsf{H}}\mathbf{h}_i\big|^2\big] = \mathrm{tr}(\boldsymbol{\Psi}_k\mathbf{R}_i), \tag{46}$$

which relies on the independence between the two random vectors.

For the power of the beamforming uncertainty, we have

$$\mathbb{E}\left\{|\mathrm{BU}_k|^2\right\} = \mathbb{E}\big[\big|\hat{\mathbf{h}}_k^{\mathsf{H}}\mathbf{h}_k - \mathbb{E}\big[\hat{\mathbf{h}}_k^{\mathsf{H}}\mathbf{h}_k\big]\big|^2\big] \tag{47}$$

$$= \mathbb{E}\big[\big|\hat{\mathbf{h}}_k^{\mathsf{H}}\mathbf{h}_k\big|^2\big] - \big|\mathbb{E}\big[\hat{\mathbf{h}}_k^{\mathsf{H}}\mathbf{h}_k\big]\big|^2 \tag{48}$$

$$= \mathbb{E}\big[\big|\hat{\mathbf{h}}_k^{\mathsf{H}}\hat{\mathbf{h}}_k + \hat{\mathbf{h}}_k^{\mathsf{H}}\tilde{\mathbf{h}}_k\big|^2\big] - \big|\mathbb{E}\big[\hat{\mathbf{h}}_k^{\mathsf{H}}\hat{\mathbf{h}}_k\big]\big|^2 \tag{49}$$

$$= \mathbb{E}\big[|\hat{\mathbf{h}}_k^{\mathsf{H}}\tilde{\mathbf{h}}_k|^2\big] \tag{50}$$

$$= \mathrm{tr}(\boldsymbol{\Psi}_k\mathbf{R}_k) - \mathrm{tr}\left(\boldsymbol{\Psi}_k^2\right), \tag{51}$$

where in (49), we have substituted (17), and in (49), we have applied the property $\mathbb{E}\left[|X+Y|^2\right] = \mathbb{E}\left[|X|^2\right] + \mathbb{E}\left[|Y^2|\right]$, which holds between two independent random variables when one of them has zero mean value, e.g., $\mathbb{E}\left[X\right] = 0$. Equation (51) follows from the facts that $\mathbb{E}\big[|\hat{\mathbf{h}}_k^{\mathsf{H}}\tilde{\mathbf{h}}_k|^2\big] = \boldsymbol{\Psi}_k\left(\mathbf{R}_k - \boldsymbol{\Psi}_k\right)$ by taking advantage of the independence between the two random vectors. The derivation of the term σ^2_{UE}, corresponding to the additive transmit distortion from all UEs, is straightforward. Specifically, we have

$$\sigma^2_{\mathrm{UE}} = \sum_{i=1}^{K}\rho_i\kappa_{\mathrm{UE}}\mathbb{E}\big\{|\mathbf{v}_k^{\mathsf{H}}\mathbf{h}_i|^2\big\} \tag{52}$$

$$= \kappa_{\mathrm{UE}}\Big(\rho_k\mathbb{E}\big\{|\mathbf{v}_k^{\mathsf{H}}\mathbf{h}_k|^2\big\} + \sum_{i\neq k}^{K}\rho_i\mathbb{E}\left\{|\mathbf{v}_k^{\mathsf{H}}\mathbf{h}_i|^2\right\}\Big). \tag{53}$$

In (52), we have taken the expectation with respect to the transmit distortion for fixed channel realizations. The first part in (53) is obtained as

$$\mathbb{E}\left\{|\mathbf{v}_k^{\mathsf{H}}\mathbf{h}_k|^2\right\} = \mathbb{E}\big[\big|\hat{\mathbf{h}}_k^{\mathsf{H}}\hat{\mathbf{h}}_k + \hat{\mathbf{h}}_k^{\mathsf{H}}\tilde{\mathbf{h}}_k\big|^2\big] \tag{54}$$

$$= \mathbb{E}\big[|\hat{\mathbf{h}}_k^{\mathsf{H}}\hat{\mathbf{h}}_k|^2\big] + \mathbb{E}\big[|\hat{\mathbf{h}}_k^{\mathsf{H}}\tilde{\mathbf{h}}_k|^2\big] \tag{55}$$

$$= |\operatorname{tr}(\boldsymbol{\Psi}_k)|^2 + \operatorname{tr}(\boldsymbol{\Psi}_k\mathbf{R}_k) - \operatorname{tr}\left(\boldsymbol{\Psi}_k^2\right), \tag{56}$$

where we have used similar steps as before. Hence, inserting (56) into (53), and noticing that the second part of (53) is identical to MUI_{ik}, σ_{UE}^2 becomes

$$\sigma_{\mathrm{UE}}^2 = \kappa_{\mathrm{UE}}\big(\rho_k|\operatorname{tr}(\boldsymbol{\Psi}_k)|^2 + \sum_{i=1}^{K}\rho_i\operatorname{tr}(\boldsymbol{\Psi}_k\mathbf{R}_i) - \rho_k\operatorname{tr}\left(\boldsymbol{\Psi}_k^2\right)\big). \tag{57}$$

The term σ_{BS}^2, concerning the additive receive distortion at the BS, is obtained as

$$\sigma_{\mathrm{BS}}^2 = \boldsymbol{\kappa}_{\mathrm{BS}}\mathbb{E}\big\{\hat{\mathbf{h}}_k^{\mathsf{H}}\big(\sum_{i=1}^{K}\rho_i\mathbf{I}_M\circ\mathbf{h}_i\mathbf{h}_i^{\mathsf{H}}\big)\hat{\mathbf{h}}_k\big\} \tag{58}$$

$$= \underbrace{\boldsymbol{\kappa}_{\mathrm{BS}}\mathbb{E}\big\{\hat{\mathbf{h}}_k^{\mathsf{H}}\big(\rho_k\mathbf{I}_M\circ\mathbf{h}_k\mathbf{h}_k^{\mathsf{H}}\big)\hat{\mathbf{h}}_k\big\}}_{\mathcal{I}_1}$$

$$+ \underbrace{\boldsymbol{\kappa}_{\mathrm{BS}}\mathbb{E}\big\{\hat{\mathbf{h}}_k^{\mathsf{H}}\big(\sum_{i\neq k}^{K}\rho_i\mathbf{I}_M\circ\mathbf{h}_i\mathbf{h}_i^{\mathsf{H}}\big)\hat{\mathbf{h}}_k\big\}}_{\mathcal{I}_2}, \tag{59}$$

where, in (58), we have taken the expectation with respect to the receive distortion for fixed channel realizations, we have accounted for MRC, i.e, $\mathbf{v}_k = \mathbf{h}_k$, and we have used the Hadamard product to write the diagonal matrix. In the next equation, we have simply split the sum and denoted the two parts as $\mathcal{I}_1$ and $\mathcal{I}_2$. In the case of the former part, we have

$$\mathcal{I}_1 = \boldsymbol{\kappa}_{\mathrm{BS}}\mathbb{E}\big\{\operatorname{tr}\big(\big(\rho_k\mathbf{I}_M\circ\big(\hat{\mathbf{h}}_k\hat{\mathbf{h}}_k^{\mathsf{H}} + \tilde{\mathbf{h}}_k\tilde{\mathbf{h}}_k^{\mathsf{H}}\big)\big)\hat{\mathbf{h}}_k\hat{\mathbf{h}}_k^{\mathsf{H}}\big)\big\} \tag{60}$$

$$= \underbrace{\boldsymbol{\kappa}_{\mathrm{BS}}\mathbb{E}\big\{\operatorname{tr}\big(\big(\rho_k\mathbf{I}_M\circ\hat{\mathbf{h}}_k\hat{\mathbf{h}}_k^{\mathsf{H}}\big)\hat{\mathbf{h}}_k\hat{\mathbf{h}}_k^{\mathsf{H}}\big)\big\}}_{\mathcal{I}_{11}}$$

$$+ \underbrace{\boldsymbol{\kappa}_{\mathrm{BS}}\mathbb{E}\big\{\operatorname{tr}\big(\big(\rho_k\mathbf{I}_M\circ\tilde{\mathbf{h}}_k\tilde{\mathbf{h}}_k^{\mathsf{H}}\big)\hat{\mathbf{h}}_k\hat{\mathbf{h}}_k^{\mathsf{H}}\big)\big\}}_{\mathcal{I}_{12}}, \tag{61}$$

where, at first, we have used (17) and exploited that $\hat{\mathbf{h}}_k$ and $\tilde{\mathbf{h}}_k$ are uncorrelated. Next, we have split (61) into $\mathcal{I}_{11}$ and $\mathcal{I}_{12}$. For $\mathcal{I}_{11}$, we exploit that the diagonal matrix can be written as $\mathbf{I}_M\circ\hat{\mathbf{h}}_k\hat{\mathbf{h}}_k^{\mathsf{H}} = \sum_{m=1}^{M}|\mathbf{e}_m^{\mathsf{H}}\hat{\mathbf{h}}_k|^2\mathbf{e}_m\mathbf{e}_m^{\mathsf{H}}$, where $\mathbf{e}_m$ is the mth column of $\mathbf{I}_M$ [49]. Hence, we have

$$\mathcal{I}_{11} = \boldsymbol{\kappa}_{\mathrm{BS}}\rho_k\sum_{m=1}^{M}\mathbb{E}\big\{\operatorname{tr}\big(\mathbf{e}_m\mathbf{e}_m^{\mathsf{H}}\hat{\mathbf{h}}_k\hat{\mathbf{h}}_k^{\mathsf{H}}\mathbf{e}_m\mathbf{e}_m^{\mathsf{H}}\hat{\mathbf{h}}_k\hat{\mathbf{h}}_k^{\mathsf{H}}\big)\big\} \tag{62}$$

$$= \boldsymbol{\kappa}_{\mathrm{BS}}\rho_k\sum_{m=1}^{M}\mathbb{E}\big\{|\hat{\mathbf{h}}_k^{\mathsf{H}}\mathbf{e}_m\mathbf{e}_m^{\mathsf{H}}\hat{\mathbf{h}}_k|^2\big\} \tag{63}$$

$$= \boldsymbol{\kappa}_{\mathrm{BS}}\rho_k\sum_{m=1}^{M}\big(|\operatorname{tr}\left(\mathbf{e}_m\mathbf{e}_m^{\mathsf{H}}\boldsymbol{\Psi}_k\right)|^2 + \operatorname{tr}\left(\mathbf{e}_m\mathbf{e}_m^{\mathsf{H}}\boldsymbol{\Psi}_k\mathbf{e}_m\mathbf{e}_m^{\mathsf{H}}\boldsymbol{\Psi}_k\right)\big) \tag{64}$$

$$= \boldsymbol{\kappa}_{\mathrm{BS}}\rho_k\big(|\operatorname{tr}\left(\mathbf{I}_M\circ\boldsymbol{\Psi}_k\right)|^2 + \operatorname{tr}\left(\left(\mathbf{I}_M\circ\boldsymbol{\Psi}_k\right)\boldsymbol{\Psi}_k\right)\big), \tag{65}$$

where, in (64), we have used [49, Lemma 2]. In the next equation, we have reverted the matrix expansion. More easily, in the case of $\mathcal{I}_{12}$, we have

$$\mathcal{I}_{12} = \boldsymbol{\kappa}_{\mathrm{BS}}\rho_k\sum_{m=1}^{M}\mathbb{E}\big\{\operatorname{tr}\big(\mathbf{e}_m\mathbf{e}_m^{\mathsf{H}}\tilde{\mathbf{h}}_k\tilde{\mathbf{h}}_k^{\mathsf{H}}\mathbf{e}_m\mathbf{e}_m^{\mathsf{H}}\hat{\mathbf{h}}_k\hat{\mathbf{h}}_k^{\mathsf{H}}\big)\big\} \tag{66}$$

$$= \boldsymbol{\kappa}_{\mathrm{BS}}\rho_k\operatorname{tr}\big(\big(\mathbf{I}_M\circ\big(\mathbf{R}_k - \boldsymbol{\Psi}_k\big)\big)\boldsymbol{\Psi}_k\big), \tag{67}$$

where, in the first equation, we have used the diagonal matrix expansion, and in the second equation, we have reverted this expansion after taking advantage of the independence between $\tilde{\mathbf{h}}_k$ and $\hat{\mathbf{h}}_k$. Substitution of (65) and (67) into (61) gives $\mathcal{I}_1$ after simple algebraic manipulations as

$$\mathcal{I}_1 = \boldsymbol{\kappa}_{\mathrm{BS}}\rho_k\big(|\operatorname{tr}\left(\mathbf{I}_M\circ\boldsymbol{\Psi}_k\right)|^2 + \operatorname{tr}\left(\left(\mathbf{I}_M\circ\mathbf{R}_k\right)\boldsymbol{\Psi}_k\right)\big). \tag{68}$$

Regarding $\mathcal{I}_2$, it follows that

$$\mathcal{I}_2 = \boldsymbol{\kappa}_{\mathrm{BS}}\mathbb{E}\big\{\operatorname{tr}\big(\big(\sum_{i\neq k}^{K}\rho_i\mathbf{I}_M\circ\mathbf{h}_i\mathbf{h}_i^{\mathsf{H}}\big)\hat{\mathbf{h}}_k\hat{\mathbf{h}}_k^{\mathsf{H}}\big)\big\} \tag{69}$$

$$= \boldsymbol{\kappa}_{\mathrm{BS}}\sum_{i\neq k}^{K}\rho_i\operatorname{tr}\left(\mathbb{E}\big\{\left(\mathbf{I}_M\circ\mathbf{h}_i\mathbf{h}_i^{\mathsf{H}}\right)\hat{\mathbf{h}}_k\hat{\mathbf{h}}_k^{\mathsf{H}}\big\}\right) \tag{70}$$

$$= \boldsymbol{\kappa}_{\mathrm{BS}}\sum_{i\neq k}^{K}\rho_i\operatorname{tr}\left(\left(\mathbf{I}_M\circ\mathbf{R}_i\right)\boldsymbol{\Psi}_k\right). \tag{71}$$

In the first equality, $\mathcal{I}_2$ has been written in terms of the trace. Next, we have exchanged the orders among summation, trace, and expectations since they are linear operators. In (71), given the independence between indices i and k, we have computed the separate expectations. Having obtained $\mathcal{I}_1$ and $\mathcal{I}_2$, we replace them in (59) to obtain σ_{BS}^2 as

$$\sigma_{\mathrm{BS}}^2 = \boldsymbol{\kappa}_{\mathrm{BS}}\big(\rho_k|\operatorname{tr}\left(\mathbf{I}_M\circ\boldsymbol{\Psi}_k\right)|^2 + \sum_{i=1}^{K}\rho_i\operatorname{tr}\left(\left(\mathbf{I}_M\circ\mathbf{R}_i\right)\boldsymbol{\Psi}_k\right)\big). \tag{72}$$

In the case of $\mathbb{E}\left\{|\mathrm{RN}_k|^2\right\}$, we easily obtain

$$\mathbb{E}\left\{|\mathrm{RN}_k|^2\right\} = \sigma^2\operatorname{tr}(\boldsymbol{\Psi}_k). \tag{73}$$

Use of (45), (51), (57), (72), and (73) concludes the proof by resulting in $\bar{S}_k$ and $\bar{I}_k$.

Appendix C
Proof of Proposition 2

We aim at finding the gradient of $\mathcal{R}$ with respect to $\phi_n, n = 1,\ldots,N$. We use the facts that $\frac{\partial\mathcal{R}}{\partial\phi_n^*} = \frac{\frac{\partial\gamma_k}{\partial\phi_n^*}}{\ln(2)\left(1+\frac{S_k}{I_k}\right)}$, which requires the derivation of $\frac{\partial\gamma_k}{\partial\phi_n^*}$. A closer observation of Theorem 1, providing γ_k, reveals that it is a fraction consisting of terms including functions of traces. Hence, the standard quotient rule derivative gives

$$\frac{\partial\gamma_k}{\partial\phi_n^*} = \frac{\frac{\partial S_k}{\partial\phi_n^*}I_k - S_k\frac{\partial I_k}{\partial\phi_n^*}}{I_k^2}, \tag{74}$$

where the partial derivatives follow. Henceforth, for the sake of simplicity, we replace the notation for the partial derivative

with respect to ϕ_n^* by $(\cdot)'$. Specifically, in the case of S_k', we obtain

$$S_k' = \rho_k \left(|\operatorname{tr}(\mathbf{\Psi}_k)|^2 \right)' \tag{75}$$
$$= 2\rho_k \operatorname{tr}(\mathbf{\Psi}_k) \operatorname{tr}(\mathbf{\Psi}_k'), \tag{76}$$

where (76) includes a simple derivative. Since all the terms in $\mathbf{\Psi}_k$ depend on ϕ_n^*, we have

$$\operatorname{tr}(\mathbf{\Psi}_k') = \operatorname{tr}(\mathbf{R}_k'\mathbf{Q}_k\mathbf{R}_k + \mathbf{R}_k\mathbf{Q}_k'\mathbf{R}_k + \mathbf{R}_k\mathbf{Q}_k\mathbf{R}_k') \tag{77}$$
$$= \operatorname{tr}\big(\mathbf{R}_k'\mathbf{Q}_k\mathbf{R}_k - \mathbf{R}_k\mathbf{Q}_k\left(\mathbf{Q}_k^{-1}\right)'\mathbf{Q}_k\mathbf{R}_k + \mathbf{R}_k\mathbf{Q}_k\mathbf{R}_k'\big), \tag{78}$$

where the derivative of $\mathbf{Q}_k$, being an inverse matrix, is obtained by [52, Eq. 40] while $\left(\mathbf{Q}_k^{-1}\right)'$ is obtained as

$$\left(\mathbf{Q}_k^{-1}\right)' = \mathbf{R}_k' + \frac{\kappa_{\mathrm{UE}}}{\tau}\sum_{i=1}^{K}\mathbf{R}_i' + \frac{\kappa_{\mathrm{BS}}\sum_{i=1}^{K}(\mathbf{I}_M \circ \mathbf{R}_i)'}{\tau}. \tag{79}$$

Substitution of (79) into (78) gives

$$\begin{aligned}\operatorname{tr}(\mathbf{\Psi}_k') &= \operatorname{tr}(\mathbf{R}_k'\mathbf{Q}_k\mathbf{R}_k) - \operatorname{tr}\left(\left(\mathbf{R}_k\mathbf{Q}_k\right)^2\mathbf{R}_k'\right)\\ &-\frac{\kappa_{\mathrm{UE}}}{\tau}\sum_{i=1}^{K}\operatorname{tr}\big(\left(\mathbf{R}_k\mathbf{Q}_k\right)^2\mathbf{R}_i'\big) - \frac{\kappa_{\mathrm{BS}}}{\tau}\sum_{i=1}^{K}\operatorname{tr}\big(\left(\mathbf{R}_k\mathbf{Q}_k\right)^2\left(\mathbf{I}_M\circ\mathbf{R}_i\right)'\big)\\ &+\operatorname{tr}(\mathbf{R}_k\mathbf{Q}_k\mathbf{R}_k').\end{aligned} \tag{80}$$

To proceed further, for the sake of exposition, let the matrices $\mathbf{A}, \mathbf{B}, \mathbf{C} \in \mathbb{C}^{M\times M}$, we have denoted

$$\begin{aligned}\mathbf{L}(\mathbf{A},\mathbf{B},\mathbf{C}) &= \alpha\beta_{2,i}[\mathbf{H}_1^{\mathsf{H}}\mathbf{A}\mathbf{Q}_k\mathbf{H}_1\mathbf{\Theta}\mathbf{R}_{\mathrm{IRS},i}]_{n,n}\\ &-\left(1+\frac{\kappa_{\mathrm{UE}}}{\tau}\right)\alpha\beta_{2,k}[\mathbf{H}_1^{\mathsf{H}}\mathbf{A}\mathbf{Q}_k\mathbf{C}\mathbf{Q}_k\mathbf{A}\mathbf{H}_1\mathbf{\Theta}\mathbf{R}_{\mathrm{IRS},k}]_{n,n}\\ &-\alpha\frac{\kappa_{\mathrm{BS}}}{\tau}\sum_{i=1}^{K}\beta_{2,i}[\mathbf{H}_1^{\mathsf{H}}\mathbf{A}\mathbf{Q}_k\mathbf{C}\mathbf{Q}_k\mathbf{A}\mathbf{H}_1\mathbf{\Theta}\mathbf{R}_{\mathrm{IRS},i}]_{n,n}\\ &+\alpha\beta_{2,k}[\mathbf{H}_1^{\mathsf{H}}\mathbf{B}\mathbf{Q}_k\mathbf{H}_1\mathbf{\Theta}\mathbf{R}_{\mathrm{IRS},k}]_{n,n}.\end{aligned} \tag{81}$$

Also, we are going to use the following useful lemma.

Lemma 1: Let $\mathbf{A} \in \mathbb{C}^{M\times M}$ be independent of $\mathbf{\Theta}$ and $\mathbf{R}_k = \beta_{2,k}\mathbf{H}_1\mathbf{\Theta}\ \mathbf{R}_{\mathrm{IRS},k}\mathbf{\Theta}^{\mathsf{H}}\mathbf{H}_1^{\mathsf{H}}$, then

$$\operatorname{tr}\left(\mathbf{A}\frac{\partial\mathbf{R}_k}{\partial\phi_n^*}\right) = \alpha\beta_{2,k}[\mathbf{H}_1^{\mathsf{H}}\mathbf{A}\mathbf{H}_1\mathbf{\Theta}\mathbf{R}_{\mathrm{IRS},k}]_{n,n}. \tag{82}$$

Proof: We have

$$\operatorname{tr}\left(\mathbf{A}\frac{\partial\mathbf{R}_k}{\partial\phi_n^*}\right) = \sum_{i,j}[\mathbf{A}]_{ij}\frac{\partial[\mathbf{R}_k]_{ji}}{\partial\phi_n^*} \tag{83}$$
$$= \alpha\beta_{2,k}\sum_{i,j}[\mathbf{A}]_{ij}[\mathbf{H}_1\mathbf{\Theta}\mathbf{R}_{\mathrm{IRS},k}]_{jn}[\mathbf{H}_1^{\mathsf{H}}]_{in}^{\mathsf{T}} \tag{84}$$
$$= \alpha\beta_{2,k}[\mathbf{H}_1^{\mathsf{H}}\mathbf{A}\mathbf{H}_1\mathbf{\Theta}\mathbf{R}_{\mathrm{IRS},k}]_{nn}, \tag{85}$$

since $\frac{\partial[\mathbf{R}_k]_{ji}}{\partial\phi_n^*} = \alpha\beta_{2,k}[\mathbf{H}_1\mathbf{\Theta}\mathbf{R}_{\mathrm{IRS},k}]_{jn}[\mathbf{H}_1^{\mathsf{H}}]_{in}^{\mathsf{T}}$. ■

By exploiting Lemma 1 for each term of (80) and that the trace of the transpose of a matrix equals the trace of this matrix, after several algebraic manipulations, we obtain

$$\begin{aligned}\operatorname{tr}(\mathbf{\Psi}_k') &= \alpha\beta_{2,k}\big[\mathbf{H}_1^{\mathsf{H}}\mathbf{R}_k\mathbf{Q}_k\big(2\mathbf{I}_M - \left(1+\frac{\kappa_{\mathrm{UE}}}{\tau}\right)\mathbf{R}_k\mathbf{Q}_k\big)\mathbf{H}_1\mathbf{\Theta}\mathbf{R}_{\mathrm{IRS},k}\big]_{n,n}\\ &-\alpha\frac{\kappa_{\mathrm{BS}}}{\tau}\sum_{i=1}^{K}\beta_{2,i}\big[\mathbf{H}_1^{\mathsf{H}}\left(\mathbf{R}_k\mathbf{Q}_k\right)^2\mathbf{H}_1\mathbf{\Theta}\mathbf{R}_{\mathrm{IRS},i}\big]_{n,n}\end{aligned} \tag{86}$$
$$= \mathbf{L}(\mathbf{R}_k,\mathbf{R}_k,\mathbf{I}_M), \tag{87}$$

which completes the derivation of S_k' after its insertion into (76).

The computation of I_k' consists of the sum of the derivatives of different terms, requiring separate manipulations. Thus, we start by the computation of the derivative of the first term in (27) as

$$(\operatorname{tr}(\mathbf{\Psi}_k\mathbf{R}_i))' = \operatorname{tr}(\mathbf{\Psi}_k'\mathbf{R}_i + \mathbf{\Psi}_k\mathbf{R}_i') \tag{88}$$
$$= \mathbf{L}(\mathbf{R}_i\mathbf{R}_k, \mathbf{R}_i\mathbf{R}_k, \mathbf{R}_i) + \alpha\beta_{2,k}[\mathbf{H}_1^{\mathsf{H}}\mathbf{\Psi}_k\mathbf{H}_1\mathbf{\Theta}\mathbf{R}_{\mathrm{IRS},k}]_{n,n}, \tag{89}$$

where we have applied Lemma 1.

Moreover, we have

$$\left(\operatorname{tr}\left(\mathbf{\Psi}_k^2\right)\right)' = 2\operatorname{tr}(\mathbf{\Psi}_k'\mathbf{\Psi}_k) \tag{90}$$
$$= 2\mathbf{L}(\mathbf{\Psi}_k\mathbf{R}_k, \mathbf{\Psi}_k\mathbf{R}_k, \mathbf{\Psi}_k). \tag{91}$$

In addition, we have

$$\left(|\operatorname{tr}(\mathbf{I}_M\circ\mathbf{\Psi}_k)|^2\right)' = 2\operatorname{tr}(\mathbf{I}_M\circ\mathbf{\Psi}_k)\operatorname{tr}(\mathbf{I}_M\circ\mathbf{\Psi}_k)' \tag{92}$$
$$= 2\operatorname{tr}(\mathbf{I}_M\circ\mathbf{\Psi}_k)\mathbf{L}(\mathbf{R}_k,\mathbf{R}_k,\mathbf{I}_M). \tag{93}$$

Furthermore, we obtain

$$(\operatorname{tr}((\mathbf{I}_M\circ\mathbf{R}_i)\mathbf{\Psi}_k))' = \operatorname{tr}\left(\left((\mathbf{I}_M\circ\mathbf{R}_i)'\mathbf{\Psi}_k + (\mathbf{I}_M\circ\mathbf{R}_i)\mathbf{\Psi}_k'\right)\right) \tag{94}$$
$$\begin{aligned}&= \alpha\beta_{2,k}[\mathbf{H}_1^{\mathsf{H}}\mathbf{\Psi}_k\mathbf{H}_1\mathbf{\Theta}\mathbf{R}_{\mathrm{IRS},k}]_{n,n}\\ &+\mathbf{L}((\mathbf{I}_M\circ\mathbf{R}_i)\mathbf{R}_k, (\mathbf{I}_M\circ\mathbf{R}_i)\mathbf{R}_k, \mathbf{I}_M\circ\mathbf{R}_i),\end{aligned} \tag{95}$$

where we have used Lemma 1. Equations (89)-(95) give I_k', which together with S_k' provide $\frac{\partial\gamma_k}{\partial\phi_n^*}$.

References

[1] F. Boccardi *et al.*, "Five disruptive technology directions for 5G," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 74–80, 2014.

[2] E. Basar *et al.*, "Wireless communications through reconfigurable intelligent surfaces," *IEEE Access*, vol. 7, pp. 116 753–116 773, 2019.

[3] Q. Wu and R. Zhang, "Intelligent reflecting surface enhanced wireless network via joint active and passive beamforming," *IEEE Trans. Wireless Commun.*, vol. 18, no. 11, pp. 5394–5409, 2019.

[4] C. Pan *et al.*, "Multicell MIMO communications relying on intelligent reflecting surfaces," *IEEE Trans. Wireless Commun.*, 2020.

[5] E. Björnson, Ö. Özdogan, and E. G. Larsson, "Intelligent reflecting surface versus decode-and-forward: How large surfaces are needed to beat relaying?" *IEEE Wireless Commun. Lett.*, vol. 9, no. 2, pp. 244–248, 2019.

[6] Q. Nadeem *et al.*, "Asymptotic max-min SINR analysis of reconfigurable intelligent surface assisted MISO systems," *IEEE Trans. Wireless Commun.*, 2020.

[7] A. M. Elbir *et al.*, "Deep channel learning for large intelligent surfaces aided mm-wave massive MIMO systems," *IEEE Wireless Commun. Lett.*, 2020.

[8] H. Guo *et al.*, "Weighted sum-rate maximization for reconfigurable intelligent surface aided wireless networks," *IEEE Trans. Wireless Commun.*, vol. 19, no. 5, pp. 3064–3076, 2020.

[9] J. Chen *et al.*, "Intelligent reflecting surface: A programmable wireless environment for physical layer security," *IEEE Access*, vol. 7, pp. 82 599–82 612, 2019.

[10] Y. Jia, C. Ye, and Y. Cui, "Analysis and optimization of an intelligent reflecting surface-assisted system with interference," *IEEE Trans. Wireless Commun.*, vol. 19, no. 12, pp. 8068–8082.

[11] M.-M. Zhao *et al.*, "Intelligent reflecting surface enhanced wireless network: Two-timescale beamforming optimization," *IEEE Trans. Wireless Commun.*, 2020.

[12] K. Zhi *et al.*, "Power scaling law analysis and phase shift optimization of RIS-aided massive MIMO systems with statistical CSI," *arXiv preprint arXiv:2010.13525*, 2020.

[13] A. Papazafeiropoulos *et al.*, "Asymptotic analysis of max-min weighted SINR for IRS-assisted MISO systems with hardware impairments," *IEEE Wireless Commun. Lett.*, pp. 1–1, 2021.

[14] T. Van Chien *et al.*, "Outage probability analysis of IRS-assisted systems under spatially correlated channels," *IEEE Wireless Commun. Lett.*, vol. 10, no. 8, pp. 1815–1819, 2021.

[15] A. Papazafeiropoulos *et al.*, "Coverage probability of distributed IRS systems under spatially correlated channels," vol. 10, no. 8, pp. 1722–1726.

[16] 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, 2019.

[17] D. Mishra and H. Johansson, "Channel estimation and low-complexity beamforming design for passive intelligent surface assisted MISO wireless energy transfer," in *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019, pp. 4659–4663.

[18] Z.-Q. He and X. Yuan, "Cascaded channel estimation for large intelligent metasurface assisted massive MIMO," *IEEE Wireless Commun. Lett.*, vol. 9, no. 2, pp. 210–214, 2019.

[19] Q. Nadeem *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, 2020.

[20] J. Qi and S. Aïssa, "Analysis and compensation of I/Q imbalance in MIMO transmit-receive diversity systems," *IEEE Trans. Commun.*, vol. 58, no. 5, pp. 1546–1556, 2010.

[21] A. K. Papazafeiropoulos, "Impact of general channel aging conditions on the downlink performance of massive MIMO," *IEEE Trans. Veh. Tech.*, vol. 66, no. 2, pp. 1428–1442, 2017.

[22] A. Papazafeiropoulos, B. Clerckx, and T. Ratnarajah, "Rate-splitting to mitigate residual transceiver hardware impairments in massive MIMO systems," *IEEE Trans. Veh. Tech.*, vol. 66, no. 9, pp. 8196–8211, 2017.

[23] T. Schenk, *RF imperfections in high-rate wireless systems: impact and digital compensation*. Springer Science & Business Media, 2008.

[24] E. Björnson *et al.*, "Massive MIMO networks: Spectral, energy, and hardware efficiency," *Foundations and Trends® in Signal Processing*, vol. 11, no. 3-4, pp. 154–655, 2017.

[25] D. Li, "Ergodic capacity of intelligent reflecting surface-assisted communication systems with phase errors," *IEEE Commun. Lett.*, vol. 24, no. 8, pp. 1646–1650, 2020.

[26] Z. Xing and R. Wang, "Achievable rate analyses and phase shift optimizations on intelligent reflecting surface with hardware impairments," *arXiv preprint arXiv:2005.14411*, 2020.

[27] X. Qian *et al.*, "Beamforming through reconfigurable intelligent surfaces in single-user MIMO systems: SNR distribution and scaling laws in the presence of channel fading and phase noise," *CoRR*, vol. abs/2005.07472, 2020. [Online]. Available: https://arxiv.org/abs/2005.07472

[28] Y. Liu *et al.*, "Beamforming designs and performance evaluations for intelligent reflecting surface enhanced wireless communication system with hardware impairments," *arXiv preprint arXiv:2006.00664*, 2020.

[29] H. Shen *et al.*, "Beamforming optimization for IRS-aided communications with transceiver hardware impairments," *IEEE Trans. Commun.*, pp. 1–1, 2020.

[30] G. Zhou *et al.*, "Secure wireless communication in RIS-aided MISO systems with hardware impairments," *arXiv preprint arXiv:2012.12733*, 2020.

[31] T. S. Rappaport *et al.*, "Wireless communications and applications above 100 GHz: Opportunities and challenges for 6G and beyond," *IEEE Access*, vol. 7, pp. 78 729–78 757, 2019.

[32] Y. Yang *et al.*, "Intelligent reflecting surface meets OFDM: Protocol design and rate maximization," *IEEE Trans. Commun.*, vol. 68, no. 7, pp. 4522–4535, 2020.

[33] D. Neumann, M. Joham, and W. Utschick, "Covariance matrix estimation in massive MIMO," *IEEE Signal Process. Lett.*, vol. 25, no. 6, pp. 863–867, 2018.

[34] A. Kammoun *et al.*, "A generalized spatial correlation model for 3D MIMO channels based on the Fourier coefficients of power spectrums," *IEEE Trans. Signal Process.*, vol. 63, no. 14, pp. 3671–3686, 2015.

[35] E. Björnson and L. Sanguinetti, "Rayleigh fading modeling and channel hardening for reconfigurable intelligent surfaces," *IEEE Wireless Commun. Lett.*, vol. 10, no. 4, pp. 830–834, 2021.

[36] X. Hu *et al.*, "Location information aided multiple intelligent reflecting surface systems," *IEEE Trans. Commun.*, vol. 68, no. 12, pp. 7948–7962, 2020.

[37] A. Epstein and G. V. Eleftheriades, "Synthesis of passive lossless metasurfaces using auxiliary fields for reflectionless beam splitting and perfect reflection," *Phys. rev. lett.*, vol. 117, no. 25, p. 256103, 2016.

[38] T. Badloe, J. Mun, and J. Rho, "Metasurfaces-based absorption and reflection control: Perfect absorbers and reflectors," *J. of Nanomaterials*, vol. 2017, 2017.

[39] P. Zhang *et al.*, "Physical layer authentication for massive MIMO systems with hardware impairments," *IEEE Trans. Wireless Commun.*, vol. 19, no. 3, pp. 1563–1576, 2020.

[40] A. Papazafeiropoulos *et al.*, "Scalable cell-free massive MIMO systems: Impact of hardware impairments," *accepted in IEEE Trans. Veh. Tec.*

[41] H. Holma and A. Toskala, *LTE for UMTS: Evolution to LTE-Advanced*, Wiley, Ed., 2011.

[42] M.-A. Badiu and J. P. Coon, "Communication through a large reflecting surface with phase errors," *IEEE Wireless Commun. Lett.*, vol. 9, no. 2, pp. 184–188, 2020.

[43] S. Abeywickrama *et al.*, "Intelligent reflecting surface: Practical phase shift model and beamforming optimization," *IEEE Trans. Commun.*, vol. 68, no. 9, pp. 5849–5863.

[44] B. Hassibi and B. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inform. Theory*, vol. 49, no. 4, pp. 951–963, 2003.

[45] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?" *IEEE J. Select. Areas Commun.*, vol. 31, no. 2, pp. 160–171, 2013.

[46] A. K. Papazafeiropoulos and T. Ratnarajah, "Deterministic equivalent performance analysis of time-varying massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 14, no. 10, pp. 5795–5809, 2015.

[47] S. Boyd, S. P. Boyd, and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

[48] 3GPP, "Further advancements for E-UTRA physical layer aspects (Release 9)," 3GPP TS 36.814, Tech. Rep., 2010.

[49] E. Björnson, M. Matthaiou, and M. Debbah, "Massive MIMO with non-ideal arbitrary arrays: Hardware scaling laws and circuit-aware design," *IEEE Trans. Wireless Commun.*, vol. 14, no. 8, pp. 4353–4368, 2015.

[50] A. Papazafeiropoulos and T. Ratnarajah, "Towards a realistic assessment of multiple antenna HCNs: Residual additive transceiver hardware impairments and channel aging," *IEEE Trans. Veh. Tech.*, vol. 66, no. 10, pp. 9061–9073, 2017.

[51] S. M. Kay, *Fundamentals of statistical signal processing: Estimation theory*. Upper Saddle River: Prentice Hall PTR, 1993.

[52] K. B. Petersen and M. S. Pedersen, "The matrix cookbook, nov 2012," *URL http://www2. imm. dtu. dk/pubdb/p. php*, vol. 3274, p. 14, 2012.