HYBRID ANALOG-DIGITAL TRANSMITTER WITH PASSIVE ANTENNA ARRAY

Zilu Zhao, Ali Bereyhi

I. Introduction

Millimeter-wave communication is one of the technologies that enable the high-bandwidth communication. However, due to the small wave length, antenna arrays are often required to compensate for the high loss, which will increase the hardware complexity. In article [1], two methods are investigated, low-resolution ADCs and hybrid analog-digital (HAD) precoding. Receivers equipped low-resolution ADCs have low power consumption but require more RF chains. On the contrary, systems employing HAD architecture requires less RF chains than the antenna amount, but the performance is limited by the number of RF chains. Low complexity algorithms have also been developed for the HAD systems such as SIC-based hybrid precoding [2]. This precoding method avoid the SVD operations and is more energy efficient than the fully digital systems.

Intelligent reflecting surface (IRS) is a newly proposed reconfigurable communication component [3]. It contains a large number of passive antenna elements whose amplitude/phase is adjustable. Typical applications of IRS includes creating a virtual line-of-sight, providing extra flexibility in resource allocation and so on [4]. When the IRS used as a relay, some alternating algorithm can be used with semidefinite relaxation to offer a trade-off between performance and computational complexity [5]. It is also shown in [5] that the use of IRS is more cost efficient than conventional amplify-and-forward relay. For the purpose of resource allocation, IRS can be employed to degrade the channel quality of eavesdroppers [6] to enhance the physical layer security. In [7], algorithms for providing energy for the energy hungry units while keeping the quality of service for the information decoding receivers with IRS structures are provided.

Author is with IDC, FAU.

Beside being used as a passive relay, the IRS can also be utilized as a passive unit of the base station (BS). In this scenario, one or multiple RF chains are placed in the close proximity of the IRS serving as energy feeder or data feeder. With even only one RF chain, a variate of usages can be achieved by such IRS-aided transmitter structure. For example, IRS-based modulation, IRS-based transmitter and IRS-based encoding [8]. The idea of spatial modulation is to embed the information in both the choice of the constellation symbols and the choice of the antenna indeces [9]. Besides turning certain IRS regions on and off [8], two other index modulation methods are studied in [10] for receiver antennas, which are IRS-space shift keying and IRS-spatial modulation. The IRS-space shift keying method root the information purely into the selection of antenna index while IRS-spatial modulation embeds partial information also into the transmitted signal of the RF chain. By using this structure as an IRS-based transmitter, the IRS is only fed with carrier signals by the RF chain. Such application is investigated in [11] and a symbol wise precoding method is adopted to design the IRS configurations. Some analytical investigations for the single-RF chain IRS-based transmission systems are conducted in [12], and states that joint encoding of the RF chain and the IRS is necessary to achieve max rate. The IRS-aided hybrid transmitters use a digital unit connected to multiple RF chains and an IRS as the analog unit. In [13], point to point multiple-input multiple-output (MIMO) communication employing such IRS-aided hybrid transmission structure is studied, and two block-level precoding methods depending only on channel state information (CSI) are proposed. Multi-user scenario is investigated in [14], which transforms the precoding task into a generalized least-square-error (GLSE) problem to make both average power and peak power controllable.

The precoding methods can be generally divided into two categories, which are block level precoding and symbol level precoding [15]. In conventional block level precoding, the transmitter utilize the CSI without the knowledge of the data messages. However, in symbol level precoding, both CSI and knowledge of the data messages are taken into considerations to generate the precoding symbols. As is discussed above two convensional block level precoding method for IRS-aided hybrid transmission systems are discussed in [13], which are based on maximizing the mutual information and approximating the optimal fully digital precoder. The advantage of the block level precoding is low computational complexity and low update rate at the IRS unit. The analog units are often designed on block level because of practical reason [16], but symbol level precoding methods based on both CSI and information messages can be designed for digital units. Symbol level precoding introduced in [17] exploit the constructive interference in modulated

communication to push the received signal away from decision boundary and hence increase the performance. The article [18] utilizes the idea of constructive interference with block level precoding and employs such method in a conventional MIMO system. In this article, we wish to calculate the precoded symbols for the digital outputs and IRS configurations in one block jointly. In other words, we want to designed digital unit and the IRS based on the knowledge of both the CSI and information messages. Furthermore, Gaussian signaling is used. In a fully connected hybrid analog digital precoder, an alternating algorithm is suggested in [19], and some theoretical results about the block length and RF chain amount are also given.

This article investigates the precoding scheme in IRS aided HAD systems. The precoding task cannot be formulated as a convex problem. Hence, alternating optimization method is used to optimize the digital unit outputs and IRS configurations. In the meantime, we find out that the optimization of the digital unit output is a convex problem, while the optimization of the IRS configurations is non-convex. We use gradient descent and majorize-minimization (MM) algorithm to reconfigure the IRS. Since the passive antenna elements on the IRS are relative cheap, this article also studies the relation between IRS size and block length.

Notation: The $\delta(\cdot)$ is the discrete Dirac function which is evaluated to zero when it has non-zero argument and evaluate to one when its argument is zero.

II. Problem Formulation

Consider downlink transmission in a multi-user MIMO system where a BS equipped with M antenna elements simultaneously serves K single-antenna user terminals (UTs). The BS employs an IRS-aided HAD architecture whose digital unit contains N radio frequency (RF)-chains and whose analog unit is realized via a passive IRS with M reflecting elements.

A. Transmitter Architecture

The architecture of the transmitter is depicted in Fig. 1. The digital unit is modeled as a generic non-linear function $\Pi_\ell(\cdot) : \mathbb{C}^K \mapsto \mathbb{C}^M$, which determines the precoded signals in the base-band digital domain based on both information symbols and CSI. The digitally-precoded signal in the symbol interval ℓ, i.e., $\mathbf{x}_\ell \in \mathbb{C}^N$, can hence be written as

$$\mathbf{x}_\ell = \Pi_\ell(\mathbf{s}_\ell|\mathbf{H}), \tag{1}$$

and is assumed to satisfy

$$\mathbb{E}\left\{\left\|\mathbf{x}_\ell\right\|^2\right\} \leq P_{\text{avg}} \tag{2}$$

for some average transmit power P_{avg}. To avoid nonlinear distortion at the RF-chains, the digital unit further restricts the peak instantaneous power of the signal, i.e., at any given time interval ℓ,

$$\left|x_{k,\ell}\right|^2 \leq P_{\max} \tag{3}$$

for $k \in [K]$ and some peak power $P_{\max}$.

At the output of the digital unit, the signal $\mathbf{x}_\ell$ is up-converted to the RF domain and illuminated via the RF-chains towards the IRS. Since the digital and IRS unit are located closely, the illumination channel between them can be regarded as a noise-free linear transform whose transfer matrix is $\mathbf{T} \in \mathbb{C}^{M\times N}$ is given by [13]

$$\mathbf{T} = \left[\frac{\lambda_{\text{w}}\sqrt{\rho G^{\text{a}}(\theta^{\text{p}}_{m,n},\phi^{\text{p}}_{m,n})G^{\text{p}}(\theta^{\text{a}}_{m,n},\phi^{\text{a}}_{m,n})}}{4\pi r_{m,n}} e^{-j\frac{2\pi r_{m,n}}{\lambda_{\text{w}}}}\right]_{m,n}, \tag{4}$$

where λ_w is the wavelength, $\rho \in [0,1]$ denotes the power efficiency of the IRS. The function $G^{\text{a}}(\theta^{\text{p}}_{m,n},\phi^{\text{p}}_{m,n})$ and $G^{\text{p}}(\theta^{\text{a}}_{m,n},\phi^{\text{a}}_{m,n})$ model the antenna gain of the active antenna and the IRS element respectively. The variable $r_{m,n}$ measures the distance from the m-th IRS element to the n-th active antenna.

We assume here that the IRS elements reflect all the signals in front of them, therefore we have

$$G^{\text{p}}(\theta,\phi) = \begin{cases} 2 & \theta,\phi \in [0,\pi] \\ 0 & \text{otherwise} \end{cases}. \tag{5}$$

The active antenna gain following [13] is assumed to be

$$G^{\text{a}}(\theta,\phi) = \begin{cases} 2(1+\kappa)\cos^{\kappa}(\phi) & \phi \in [0,\pi/2] \\ 0 & \text{otherwise} \end{cases}. \tag{6}$$

The κ here is a normalization factor which ensures that the spherical surface integral of $G^{\text{a}}(\theta,\phi)$ is 4π. With larger κ, the beam from the active antenna will be more concentrated to its main direction.

At the symbol interval ℓ, the IRS receives the signal $\mathbf{r}_\ell \in \mathbb{C}^M$ which is given by

$$\mathbf{r}_\ell = \mathbf{T}\mathbf{x}_\ell. \tag{7}$$

Each reflecting element on the IRS reflects its received signal, i.e., $r_{m,\ell}$ for $m \in [1, M]$, with an adjustable phase-shift. As a result, the output signal of the transmitter, i.e., the signal reflected by the IRS, in the symbol interval ℓ is given by

$$\mathbf{z}_\ell = \mathbf{A}_\ell \mathbf{r}_\ell \tag{8}$$

where $\mathbf{A}_\ell = \text{Diag}\left\{e^{j\beta_{1,\ell}}, \ldots, e^{j\beta_{M,\ell}}\right\}$ with $\beta_{m,\ell}$ denoting the phase-shift applied by the reflecting element m in interval ℓ.

B. Downlink Transmission and CSI Acquisition

The signal received by UT k in symbol interval ℓ is given by

$$y_{k,\ell} = \mathbf{h}_{k,\ell}^{\mathsf{T}} \mathbf{z}_\ell + w_{k,\ell}, \tag{9}$$

where $\mathbf{h}_{k,\ell} \in \mathbb{C}^M$ denotes the vector of downlink channel coefficients from the BS to the k-th UT, and $w_{k,\ell}$ is additive white Gaussian noise (AWGN) with variance σ^2, i.e., $w_{k,\ell} \sim \mathcal{CN}(0, \sigma^2)$. Considering the transmitter architecture, the vector of received signals at the UTs, i.e., $\mathbf{y}_\ell = [y_{1,\ell}, \ldots, y_{K,\ell}]^{\mathsf{T}}$, is compactly represented in terms of the digitally-precoded signal $\mathbf{x}_\ell$ as

$$\mathbf{y}_\ell = \mathbf{H}_\ell^{\mathsf{T}} \mathbf{A}_\ell \mathbf{T} \mathbf{x}_\ell + \mathbf{w}_\ell \tag{10}$$

where the matrix $\mathbf{H}_\ell \in \mathbb{C}^{M \times K}$ is given by

$$\mathbf{H}_\ell = [\mathbf{h}_{1,\ell}, \ldots, \mathbf{h}_{K,\ell}] \tag{11}$$

and denotes the channel matrix between the BS to UTs. Moreover, the vector $\mathbf{w}_n = [w_{1,\ell}, \ldots, w_{K,\ell}]$ collects the samples of the AWGN processes in symbol interval ℓ.

The system is assumed to perform in the time division duplex (TDD) mode. This means that uplink and downlink transmissions are divided in time domain, and hence the uplink and downlink channels are reciprocal. Using this property, the BS acquires the CSI through an uplink training phase in the network that holds prior to data transmission [20]: at the beginning of each coherence time interval, the UTs transmit mutually-orthogonal and publicly-known pilot sequences in their uplink channels. The BS then uses the known pilot codebook to acquire the channel coefficients of each UT. In practical scenarios with receiver noise, quantization error and imperfect pilot sequences, the BSs assesses a distorted version of the CSI. We take this point into account by considering the expectation-based uncertainty model for the estimated CSI at the

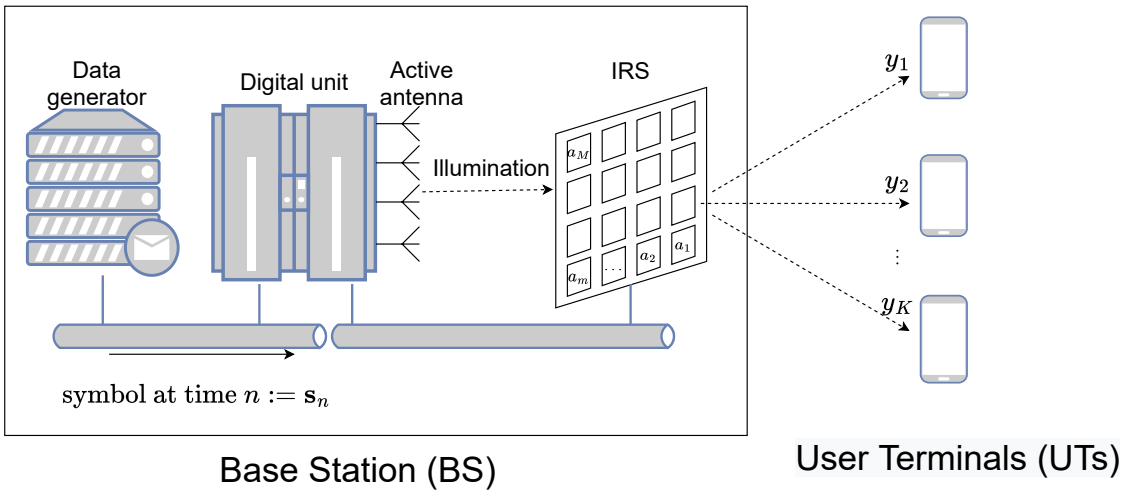

Fig. 1: Illustration of the system model

BS. Namely, we assume that the BS estimates the channel matrix as $\hat{\mathbf{H}}_n$ which is related to the true channel matrix as [21]

$$\mathbf{H}_\ell = \hat{\mathbf{H}}_\ell + \mathbf{E}_\ell, \tag{12}$$

for some random channel uncertainty process $\mathbf{E}_\ell \in \mathbb{C}^{M\times K}$ with independent columns whose column k is an independent and identical distribution (i.i.d.) Gaussian vector with mean zero and variance ξ_k^2, i.e., $\mathbf{e}_{k,\ell} \sim \mathcal{CN}\left(0, \xi_k^2\right)$ with $\mathbf{e}_{k,\ell}$ denoting the k-th column of $\mathbf{E}_\ell$.

C. Block-Level Analog Beamforming

While the digital unit can be updated at the symbol-level, the analog unit is updated once per a block of symbol intervals. From the implementational viewpoint, this means that the signals on the illuminators are updated once every symbol interval, but the phase-shifts at IRS are updated only once at the beginning of a transmission block. Fig. 2 depicts this block-level transmission strategy. In this figure, the table entries with the same color correspond to the message symbols of the same block. Assuming that the information symbol rate to be R, the RF-chains are updated at rate R while the IRS is tuned at the rate R/L.

Let $\mathbf{S}$ denote the collection of information symbols being transmitted in a particular block. This can be written as

$$\mathbf{S} = \left[\mathbf{s}_1, \mathbf{s}_2, \cdots, \mathbf{s}_L\right], \tag{13}$$

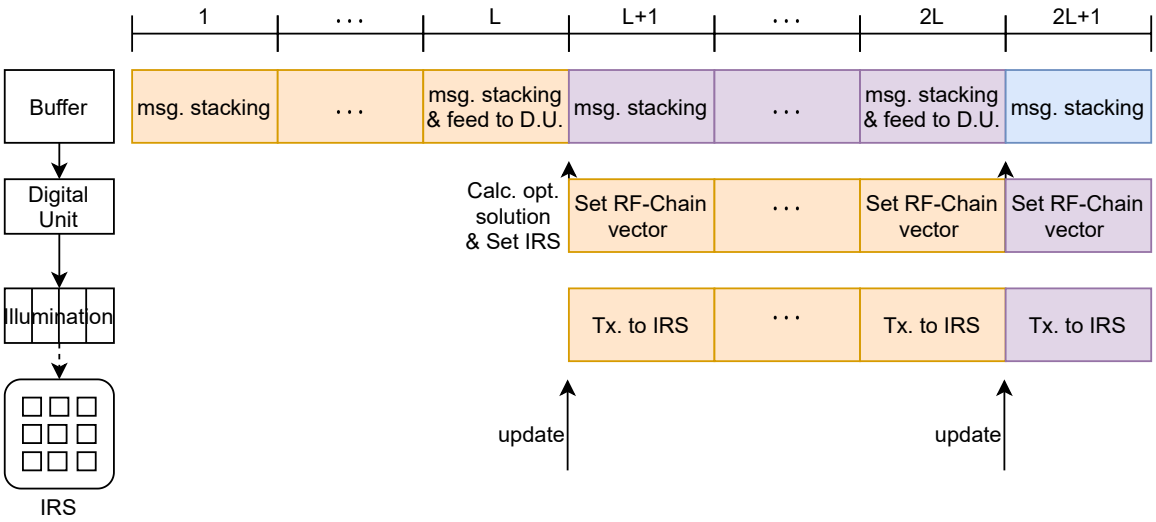

Fig. 2: Transmitter Time Table

with L being the block length. We assume that the block length L is smaller than the coherence time interval of the channel, as it is the case in practice. This means that the channel coefficients can be considered to remain unchanged during signal transmission in a single block. As a result, the matrix of signals received in this block at the UTs, i.e.,

$$\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2 \cdots \mathbf{y}_L], \tag{14}$$

is given by

$$\mathbf{Y} = \mathbf{H}^{\mathsf{T}}\mathbf{A}\mathbf{T}\mathbf{X} + \mathbf{W}, \tag{15}$$

where $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_L]$ denotes the digitally-precoded transmit signals in the block, and $\mathbf{W} = [\mathbf{w}_1, \ldots, \mathbf{w}_L]$. Here, we drop the time index ℓ for the analog phase-shifters and the channel coefficients as they remain unchanged during the block transmission.

D. Linear Reception

At the receiver side, we assume a simple post-processing strategy: as UT k receives the symbol $y_{k,\ell}$ in interval ℓ of the block, it takes its scaled version $\tilde{y}_{k,\ell} = F_k y_{k,\ell}$ as the soft estimate of the transmitted symbol $s_{k,\ell}$. The receiver gains of the UTs are updated based on channel statistics and hence remain unchanged within a coherence time interval. As the result, we can represent the matrix of soft estimates collected by the UTs within the transmission block as

$$\tilde{\mathbf{Y}} = \mathbf{F}\mathbf{Y}, \tag{16}$$

where $\mathbf{F} = \mathrm{Diag}\{F_1, \ldots, F_K\}$. In our transmission scheme, we assume that the elements of the matrix $\mathbf{F}$ compensate for the path loss.

E. Performance Metrics and Design Objectives

The main objective of this work is to design the digital and analog units oft the transmitter, i.e., $\Pi_\ell(\cdot)$ and $\mathbf{A}$, such that the system throughput is optimized. To this end, we need to introduce some performance measures for the evaluation. One commonly used designing criteria is achievable rate per user, which is bounded by [19]:

$$\bar{R} \geq \frac{LT_{\mathrm{s}}}{\tau + LT_{\mathrm{s}}} \log\left(1 + \frac{\sigma_s^2}{P_{\mathrm{Noise}} + P_{\mathrm{Interference}}}\right), \tag{17}$$

where T_{s} is symbol time, τ is the duration of guard interval and $P_{\mathrm{Interference}}$ is the average interference power after the decoding/compensation process. The power term P_{Noise} is the equivalent noise after reception. The variance σ_s^2 is the power of the information data.

For a system with fixed block length, $\Pi_\ell(\cdot)$ and $\mathbf{A}$ only have influence on $P_{\mathrm{Interference}}$. Thus, the residual sum of squares (RSS) can be adopted here to assess the distortion between the decoded signals and the intended messages. It is defined as

$$\mathrm{d}(\mathbf{S}, \tilde{\mathbf{Y}}) = ||\tilde{\mathbf{Y}} - \mathbf{S}||_{\mathrm{F}}^2 = \mathrm{tr}\{(\tilde{\mathbf{Y}} - \mathbf{S})^{\mathsf{H}}(\tilde{\mathbf{Y}} - \mathbf{S})\}. \tag{18}$$

It is desired to make $\tilde{\mathbf{Y}}$ similar to the coded message symbols $\mathbf{S}$. In the following two sections, we will try to develop an algorithm to configure the digital precoder Π_ℓ and analog precoder $\mathbf{A}$ jointly for arbitrary block length L to optimize the throughput.

III. Precoding via Least-Squares Method

In Eq.(18), $\mathbf{F}$ is used to compensate for the path loss which is defined in matrix form as

$$\mathbf{P}_{\mathrm{L}}^{1/2} = \begin{bmatrix} \sqrt{\bar{h}_1} & & \\ & \ddots & \\ & & \sqrt{\bar{h}_K} \end{bmatrix}, \tag{19}$$

where $\bar{h}_{1\ldots K}$ denotes the path loss between the BS and UTs. The decoder/compensator can then be designed as

$$\mathbf{F} = \mathbf{P}_{\mathrm{L}}^{-1/2}. \tag{20}$$

Thus, the compensated received signal is

$$\tilde{\mathbf{Y}} = \mathbf{P}_{\mathrm{L}}^{-1/2}[(\hat{\mathbf{H}}^{\mathsf{T}} + \mathbf{E}^{\mathsf{T}})\mathbf{A}\mathbf{T}\mathbf{X} + \mathbf{W}]. \tag{21}$$

To keep the further discussions brief, define the normalized matrices $\tilde{\mathbf{H}}$, $\tilde{\hat{\mathbf{H}}}$, $\tilde{\mathbf{E}}$ and $\tilde{\mathbf{W}}$ as Eq. (22).

$$\begin{aligned} \tilde{\mathbf{H}} =& \mathbf{P}_{\mathrm{L}}^{-1/2}\mathbf{H} \\ \tilde{\hat{\mathbf{H}}}^{\mathsf{T}} =& \mathbf{P}_{\mathrm{L}}^{-1/2}\hat{\mathbf{H}}^{\mathsf{T}} \\ \tilde{\mathbf{E}}^{\mathsf{T}} =& \mathbf{P}_{\mathrm{L}}^{-1/2}\mathbf{E}^{\mathsf{T}} \\ \tilde{\mathbf{W}} =& \mathbf{P}_{\mathrm{L}}^{-1/2}\mathbf{W} \end{aligned} \tag{22}$$

Because of the existence of random terms, statistics of Eq. (18) are used for the designing. Assume that the noise entries are independent with the estimation errors, we then have

$$\begin{aligned} & \mathcal{E}_{\mathbf{E}}\left\{\mathcal{E}_{\mathbf{W}}\{\mathrm{d}(\mathbf{S}, \tilde{\mathbf{Y}})\}\right\} \\ =& \left|\left|\tilde{\hat{\mathbf{H}}}^{\mathsf{T}}\mathbf{A}\mathbf{T}\mathbf{X} - \mathbf{S}\right|\right|_{\mathrm{F}}^2 + \mathrm{tr}\{\mathbf{X}^{\mathsf{H}}\mathbf{T}^{\mathsf{H}}\mathbf{A}^{\mathsf{H}}\mathbf{\Sigma}_{\tilde{\mathbf{E}}}\mathbf{A}\mathbf{T}\mathbf{X}\} + L\mathrm{tr}\{\mathbf{P}_{\mathrm{L}}^{-1}\}\sigma_{\mathrm{n}}^2, \end{aligned} \tag{23}$$

where

$$\mathbf{\Sigma}_{\tilde{\mathbf{E}}} = sum_{k=1}^{K}\sigma_{\mathrm{e},k}^2/\bar{h}_k \cdot \mathbf{I}_M. \tag{24}$$

The formula given in Eq. (23) is the expected distance between $\mathbf{S}$ and $\tilde{\mathbf{Y}}$, i.e. $\mathcal{E}_{\mathbf{W},\mathbf{E}}\{\mathrm{d}(\mathbf{S}, \tilde{\mathbf{Y}})\}$. From now on, instead of developing an algorithm that minimize Eq. (18) directly, we try to develop an algorithm that minimize Eq. (23) by optimizing $\mathbf{A}$ and $\mathbf{X}$.

A. Problem Formulation as a GLSE Precoder

Following [22], if we utilize the peak and total power limitation, the minimization of the expected distance between the intended message and the received symbols leads us to the design of a GLSE precoder. The optimization problem is given by

$$\begin{aligned} \underset{\mathbf{A},\mathbf{X}}{\text{minimize}} \quad & \left|\left|\tilde{\hat{\mathbf{H}}}^{\mathsf{T}}\mathbf{A}\mathbf{T}\mathbf{X} - \mathbf{S}\right|\right|_{\mathrm{F}}^2 + \mathrm{tr}\{\mathbf{X}^{\mathsf{H}}\mathbf{T}^{\mathsf{H}}\mathbf{A}^{\mathsf{H}}\mathbf{\Sigma}_{\tilde{\mathbf{E}}}\mathbf{A}\mathbf{T}\mathbf{X}\} + \lambda\mathrm{tr}\{\mathbf{X}\mathbf{X}^{\mathsf{H}}\} \\ \text{subject to} \quad & \max_{i,j}|x_{i,j}|^2 < P_{\mathrm{peak}}, \end{aligned} \tag{25}$$

where $P_{\mathrm{peak}} \geq 0$ is the peak power constraint. Another variable $\lambda \geq 0$ represents the regularizer that balences the trade-off between the expected distance and total transmission power. With larger λ, the system will have to put more emphasis on constraining the total transmission power.

On the contrary, with smaller λ, the system will pay more attention to minimize the expected distortion between received signals and intended signals.

The peak power constraint can be adjusted along with the regularizer to control the peak-to-average-power-ratio (PAPR). To keep the discussion brief, $\mathrm{obj}(\cdot)$ is used to denote the objective function in Eq. (25). It is defined as

$$\mathrm{obj}(\mathbf{A},\mathbf{X},\mathbf{S}) = \left|\left|\tilde{\hat{\mathbf{H}}}^{\mathsf{T}}\mathbf{A}\mathbf{T}\mathbf{X}-\mathbf{S}\right|\right|_{\mathrm{F}}^{2} + \mathrm{tr}\{\mathbf{X}^{\mathsf{H}}\mathbf{T}^{\mathsf{H}}\mathbf{A}^{\mathsf{H}}\mathbf{\Sigma}_{\tilde{\mathbf{E}}}\mathbf{A}\mathbf{T}\mathbf{X}\} + \lambda\mathrm{tr}\{\mathbf{X}\mathbf{X}^{\mathsf{H}}\} \tag{26}$$

The term $L\mathrm{tr}\{\mathbf{P}_{\mathrm{L}}^{-1}\}\sigma_{\mathrm{n}}^{2}$ is removed from the objective function, because it is a constant term that depends neither on $\mathbf{A}$ nor on $\mathbf{X}$.

IV. GLSE Precoding via Alternating Optimization

The joint optimization of the problem stated in (25) is non-trivial since it is generally non-convex. Alternating optimization method is adopted to obtain a suboptimal solution iteratively. In each iteration, the optimal RF chain outputs are first calculated while treating the IRS configurations as a fixed matrix, and then the system looks for a suitable (may not be optimal) phase-shifts configuration while treating $\mathbf{X}$ as a fixed matrix. The pseudo-code in Algorithm (1) depicts the general procedure.

Algorithm 1 General Procedure for finding ($\mathbf{A}$, $\mathbf{X}$)

1: $(\mathbf{A},\mathbf{X}) \leftarrow$ initial values
2: **while** terminate condition not met **do**
3: $\quad\mathbf{X} \leftarrow \arg\min_{\mathbf{X}}\ \mathrm{obj}(\mathbf{A},\mathbf{X},\mathbf{S})$ s.t. $\max_{i,j}|x_{i,j}|^2 < P_{\mathrm{peak}}$
4: $\quad\mathbf{A} \leftarrow \arg\min_{\mathbf{A}}\ \mathrm{obj}(\mathbf{A},\mathbf{X},\mathbf{S})$
5: **end while**

A termination condition for the loop should be chosen accordingly. For example, let the loop be executed until two consequent loops have their cost function difference smaller than a certain threshold or terminates when certain iteration time is reached. The step 3 in the algorithm (optimize with respect to $\mathbf{X}$) is quadratic programming, which is categorized as convex optimization problem. Therefore, an optimal $\mathbf{X}$ can always be obtained if $\mathbf{A}$ is given.

The domain of $\mathbf{A}$ is not a convex set. This implies that the step 4 in the algorithm is a non-convex problem. Thus, the optimal value in step 4 can generally not be found. This leads

to a problem that Algorithm (1) may not converge, because in step 4, the system may find a suboptimal minimal for $\mathbf{A}$ that may lead to a larger cost than the previous iteration.

A. Optimization of the Digital Unit

When $\mathbf{A}$ is treated as a fixed matrix, the problem given by

$$\begin{aligned} \underset{\mathbf{X}}{\text{minimize}} \quad & \left|\left|\tilde{\hat{\mathbf{H}}}^{\mathsf{T}}\mathbf{ATX}-\mathbf{S}\right|\right|_{\mathrm{F}}^{2}+\mathrm{tr}\{\mathbf{X}^{\mathsf{H}}\mathbf{T}^{\mathsf{H}}\mathbf{A}^{\mathsf{H}}\mathbf{\Sigma}_{\tilde{\mathbf{E}}}\mathbf{ATX}\}+\lambda\mathrm{tr}\{\mathbf{XX}^{\mathsf{H}}\} \\ \text{subject to} \quad & \max_{i,j}|x_{i,j}|^{2}<P_{\text{peak}} \end{aligned} \tag{27}$$

belongs to the class of convex problem. Moreover, it is worth noticing that the columns (time intervals) of matrix $\mathbf{X}$ can be decoupled according to Eq. (28).

$$\begin{aligned} &\mathrm{tr}\{(\tilde{\hat{\mathbf{H}}}^{\mathsf{T}}\mathbf{ATX}-\mathbf{S})^{\mathsf{H}}(\tilde{\hat{\mathbf{H}}}^{\mathsf{T}}\mathbf{ATX}-\mathbf{S})\}+\mathrm{tr}\{\mathbf{X}^{\mathsf{H}}\mathbf{T}^{\mathsf{H}}\mathbf{A}^{\mathsf{H}}\mathbf{\Sigma}_{\tilde{\mathbf{E}}}\mathbf{ATX}\}+\lambda\mathrm{tr}\{\mathbf{X}^{\mathsf{H}}\mathbf{X}\} \\ &=\sum_{i=1}^{L}(\tilde{\hat{\mathbf{H}}}^{\mathsf{T}}\mathbf{AT}\boldsymbol{x_i}-\boldsymbol{s_i})^{\mathsf{H}}(\tilde{\hat{\mathbf{H}}}^{\mathsf{T}}\mathbf{AT}\boldsymbol{x_i}-\boldsymbol{s_i})+\mathbf{x}_i^{\mathsf{H}}\mathbf{T}^{\mathsf{H}}\mathbf{A}^{\mathsf{H}}\mathbf{\Sigma}_{\tilde{\mathbf{E}}}\mathbf{AT}\mathbf{x}_i+\lambda\boldsymbol{x_i}^{\mathsf{H}}\boldsymbol{x_i}. \end{aligned} \tag{28}$$

Therefore, this optimization problem can either be treated as one matrix convex optimization or L independent vector convex optimization problems. This implies that the block length has no influence over the RF chain outputs if $\mathbf{A}$ is fixed.

B. Optimization of the IRS Configuration

When treating $\mathbf{X}$ as a fixed matrix, the optimization problem is then formulated as

$$\underset{\mathbf{A}}{\text{minimize}} \quad \left|\left|\tilde{\hat{\mathbf{H}}}^{\mathsf{T}}\mathbf{ATX}-\mathbf{S}\right|\right|_{\mathrm{F}}^{2}+\mathrm{tr}\{\mathbf{X}^{\mathsf{H}}\mathbf{T}^{\mathsf{H}}\mathbf{A}^{\mathsf{H}}\mathbf{\Sigma}_{\tilde{\mathbf{E}}}\mathbf{ATX}\}. \tag{29}$$

It is generally non-convex problem, since the domain of $\mathbf{A}$ is not a convex set. Two methods are used here to obtain a suboptimal point, which are gradient descent and MM algorithm [23]. For further discussions, we denote the objective function in Eq. (29) by

$$\mathrm{obj}_{\mathbf{A}}(\mathbf{A})=\left|\left|\tilde{\hat{\mathbf{H}}}^{\mathsf{T}}\mathbf{ATX}-\mathbf{S}\right|\right|_{\mathrm{F}}^{2}+\mathrm{tr}\{\mathbf{X}^{\mathsf{H}}\mathbf{T}^{\mathsf{H}}\mathbf{A}^{\mathsf{H}}\mathbf{\Sigma}_{\tilde{\mathbf{E}}}\mathbf{ATX}\}. \tag{30}$$

1) Gradient descent method: Gradient descend is a general method to find the local optimal of the functions that maps one or multiple parameters to $\mathbb{R}$. The local minimal can be obtained by following the opposite gradient of the function iteratively from a random initial point. This can be understood graphically as following the steepest slope on a surface until reaching the trough.

In order to simplify the domain, define the phase vector

$$\boldsymbol{\beta} = \arg \cdot \mathrm{diag}_{\mathrm{vector}}(\mathbf{A}). \tag{31}$$

To keep the notation simple, we declare auxiliary matrices as follow

$$\begin{aligned} \mathbf{U} &= \mathbf{ATX}; \\ \mathbf{V} &= \mathbf{S}^{\mathsf{H}}\tilde{\hat{\mathbf{H}}}^{\mathsf{T}} \\ \mathbf{Q} &= \tilde{\hat{\mathbf{H}}}^{*}\tilde{\hat{\mathbf{H}}}^{\mathsf{T}} + \boldsymbol{\Sigma}_{\tilde{\mathbf{E}}}. \end{aligned} \tag{32}$$

Under this convention, the objective function can be expanded as

$$\mathrm{obj}_{\mathbf{A}}(\mathbf{A}) = \mathrm{tr}\{\mathbf{U}^{\mathsf{H}}\mathbf{QU} - (\mathbf{U}^{\mathsf{H}}\mathbf{V}^{\mathsf{H}} + \mathbf{VU}) + \mathbf{S}^{\mathsf{H}}\mathbf{S}\}. \tag{33}$$

Since the differential operation is linear, we first calculate the the derivative of the additive terms in Eq. (33) as

$$\begin{aligned} \frac{\partial \mathrm{tr}\{\mathbf{U}^{\mathsf{H}}\mathbf{QU}\}}{\partial \boldsymbol{\beta}} &= -2\mathrm{diag}_{\mathrm{vector}}(\Im(\mathbf{UU}^{\mathsf{H}}\mathbf{Q})) \\ \frac{\partial \mathrm{tr}\{\mathbf{U}^{\mathsf{H}}\mathbf{V}^{\mathsf{H}} + \mathbf{VU}\}}{\partial \boldsymbol{\beta}} &= -2\mathrm{diag}_{\mathrm{vector}}(\Im(\mathbf{UV})). \end{aligned} \tag{34}$$

From Eq. (33) and Eq. (34), the partial derivative of the objective function can be derived as

$$\frac{\partial \mathrm{obj}_{\mathbf{A}}(\mathbf{A})}{\partial \boldsymbol{\beta}} = -2\mathrm{diag}_{\mathrm{vector}}(\Im(\mathbf{UU}^{\mathsf{H}}\mathbf{Q} - \mathbf{UV})). \tag{35}$$

The auxiliary vector can be transformed back to the phase-shift matrix by Eq. (36).

$$\mathbf{A} = \mathrm{diag}_{\mathrm{matrix}}\left(\exp(j\boldsymbol{\beta})\right) \tag{36}$$

Combining with the designing of the digital unit, the pseudo-code of the precoding algorithm with gradient descent is shown as Algorithm (2).

In order to avoid the situation where the algorithm diverges, one can design the step size α in the following steps. At the beginning of each gradient descent loop (step 4 - 10 in Algorithm (2)), α is initialized to a relative large value. During each loop, the processor keeps the old value of $\mathbf{A}$ in the memory. If the value of the objective function does not decrease with the new $\mathbf{A}$, the processor will then load the old value of $\mathbf{A}$ from the memory and redo that gradient descent loop once again with halved step size.

The whole gradient descent loop terminates if the value of α is smaller than a certain threshold. In this way, $\mathbf{A}$ will always be assigned to a value that makes the objective function decrease.

Algorithm 2 Precoding ($\mathbf{A}$, $\mathbf{X}$) with gradient descent

$(\mathbf{A}, \mathbf{X}) \leftarrow$ initial values
while terminate condition not met **do**
  $\mathbf{X} \leftarrow \mathrm{convex}_{\mathrm{opt}}\{\mathrm{obj}(\mathbf{A}, \mathbf{X}, \mathbf{S}) \text{ s.t. } \max_{i,j} |x_{i,j}|^2 < P_{\mathrm{peak}}\}$
  while terminate condition not met **do**
    $\boldsymbol{\beta} \leftarrow \arg \cdot \mathrm{diag}_{\mathrm{vector}}(\mathbf{A})$
    $\boldsymbol{\beta}_{\mathrm{grad}} \leftarrow -2\mathrm{diag}_{\mathrm{vector}}(\Im(\mathbf{U}\mathbf{U}^{\mathsf{H}}\mathbf{Q} - \mathbf{U}\mathbf{V}))$
    $\alpha \leftarrow \mathrm{getStepSize}$
    $\boldsymbol{\beta}' = \boldsymbol{\beta} - \alpha\boldsymbol{\beta}_{\mathrm{grad}}$
    $\mathbf{A} \leftarrow \mathrm{diag}_{\mathrm{matrix}}\left(\exp(j\boldsymbol{\beta}')\right)$
  end while
end while

Moreover, if the termination threshold for α is small enough, $\mathbf{A}$ will always end up at the close proximity of a critical point after the termination of one entire gradient descent loop.

It is worth noticing that, though we can make the algorithm converge to a suboptimal point, it does not mean that the systems starting with different initial pairs $(\mathbf{A}, \mathbf{X})$ will converge to the same value.

2) Majorize-minimization (MM) algorithm: MM algorithm [23] is another general optimization method that can be used to optimize the IRS configurations. Similar to gradient descent, this algorithm also includes iterative procedures itself. However, when using MM algorithm we don't need to state the step size explicitly.

A special auxiliary function called surrogate function $\overline{\mathrm{obj}_{\mathbf{A}_0}} : \mathrm{domain}(\mathrm{obj}(\cdot)) \to \mathbb{R}$ is used in MM algorithm. It a less complicated function that need to be constructed at any point $\mathbf{A}_0$ of the objective function. The surrogate function constructed at a point must touch the objective function at that point. Meanwhile, it should always be larger or equal to the objective function at every point within its domain.

The procedure of optimizing $\mathbf{A}$ with MM algorithm can be iteratively described as follow. At the i-th MM iteration for optimizing the IRS, a surrogate function of a point $\mathbf{A}_i$ is constructed. We then find the optimal point $\mathbf{A}_{i+1}$ that minimize the surrogate function. After that, we proceed to the next MM iteration and optimize a new surrogate function tangent to point $\mathbf{A}_{i+1}$. We can

choose the initial point of $\mathbf{A}$ arbitrarily as long as it lies in the feasible set.

In oder to simplify the problem, we wish to construct a linear surrogate function free of quadratic terms. When expanded, the objective function in Eq. (33) is comprised of a quadratic term, two first order terms and one constant term with respect to $\mathbf{A}$.

By the definition of surrogate function, the objective function is always a surrogate function of itself. The additive property of the surrogate functions can also be proven by the definition. This property states that two surrogate functions are additive if they are both tangent to their objective functions at the same point. Based on these two properties, a linear surrogate function for the objective function Eq. (33) can be constructed as follow.

We first construct a linear surrogate function for the quadratic term of Eq. (33). The first order terms and constant term will serve as surrogate functions of themselves. The sum of those surrogate functions are the surrogate function for the whole objective function described by Eq. (33). Since the constant term does not contribute to the optimization, we can safely discard it.

Next we will give out the equivalent surrogate function.

To find the surrogate function, we use the auxiliary variables defined in Eq. (32). Furthermore, define

By substituting the variables according to Eq. (32), the quadratic term can be written as

$$\mathbf{X}^{\mathsf{H}}\mathbf{T}^{\mathsf{H}}\mathbf{A}^{\mathsf{H}}\left(\tilde{\hat{\mathbf{H}}}^{\mathsf{H}}\tilde{\hat{\mathbf{H}}}+\mathbf{\Sigma}_{\tilde{\mathbf{E}}}\right)\mathbf{A}\mathbf{T}\mathbf{X}=\mathbf{U}^{\mathsf{H}}\mathbf{Q}\mathbf{U}. \tag{37}$$

To construct the surrogate function for the quadratic term at an arbitrary point $\mathbf{A}_0$, we define an auxiliary matrix $\mathbf{P}$, such that $\mathbf{P}-\mathbf{Q}$ is positive definite. For the tangential point $\mathbf{A}_0$, we also define $\mathbf{U}_0=\mathbf{A}_0\mathbf{T}\mathbf{X}$. The surrogate for the quadratic term can be obtained by the following inequality [23].

$$\begin{aligned}
&\mathrm{tr}\left\{\mathbf{U}^{\mathsf{H}}\mathbf{Q}\mathbf{U}\right\}\\
=&\mathrm{tr}\left\{\mathbf{U}_0^{\mathsf{H}}\mathbf{Q}\mathbf{U}_0+(\mathbf{U}-\mathbf{U}_0)^{\mathsf{H}}\mathbf{Q}\mathbf{U}_0+\mathbf{U}_0^{\mathsf{H}}\mathbf{Q}(\mathbf{U}-\mathbf{U}_0)+(\mathbf{U}-\mathbf{U}_0)^{\mathsf{H}}\mathbf{Q}(\mathbf{U}-\mathbf{U}_0)\right\}\\
\leq&\mathrm{tr}\left\{\mathbf{U}_0^{\mathsf{H}}\mathbf{Q}\mathbf{U}_0+(\mathbf{U}-\mathbf{U}_0)^{\mathsf{H}}\mathbf{Q}\mathbf{U}_0+\mathbf{U}_0^{\mathsf{H}}\mathbf{Q}(\mathbf{U}-\mathbf{U}_0)+(\mathbf{U}-\mathbf{U}_0)^{\mathsf{H}}\mathbf{P}(\mathbf{U}-\mathbf{U}_0)\right\}\\
=&\mathrm{tr}\left\{\mathbf{U}^{\mathsf{H}}\mathbf{P}\mathbf{U}+\mathbf{U}^{\mathsf{H}}(\mathbf{Q}-\mathbf{P})\mathbf{U}_0+\mathbf{U}_0^{\mathsf{H}}(\mathbf{Q}-\mathbf{P})\mathbf{Y}+\mathbf{U}_0^{\mathsf{H}}(\mathbf{Q}-\mathbf{P})\mathbf{U}_0\right\}
\end{aligned} \tag{38}$$

The newly introduce quadratic term $\mathbf{U}^{\mathsf{H}}\mathbf{P}\mathbf{U}$ will be degraded to a constant term if we specify $\mathbf{P}$ to be $\mathbf{P}=\lambda_{\max}^{\mathbf{Q}}\mathbf{I}$, where $\lambda_{\max}^{\mathbf{Q}}$ is the largest eigenvalue of $\mathbf{Q}$.

By adding the first order term of the objective function Eq. (30) and removing constant terms, the equivalent surrogate function (not the true surrogate since the constant terms are removed) for the objective function at any point $\mathbf{A}_0$ reads

$$\overline{\mathrm{obj}_{\mathbf{A}_0}(\mathbf{A})} = \mathrm{tr}\left\{\mathbf{U}^{\mathsf{H}}(\mathbf{Q}-\mathbf{P})\mathbf{U}_0 + \mathbf{U}_0^{\mathsf{H}}(\mathbf{Q}-\mathbf{P})\mathbf{U} - \mathbf{V}\mathbf{U} - \mathbf{U}^{\mathsf{H}}\mathbf{V}^{\mathsf{H}}\right\}. \tag{39}$$

For simplicity, define a constant

$$\mathbf{G}_0 = \mathbf{T}\mathbf{X}\mathbf{U}_0^{\mathsf{H}}(\mathbf{Q}-\mathbf{P}) - \mathbf{T}\mathbf{X}\mathbf{V}. \tag{40}$$

By substituting Eq. (40) into Eq. (39), the latter becomes

$$\overline{\mathrm{obj}_{\mathbf{A}_0}(\mathbf{A})} = \mathrm{tr}\{\mathbf{G}_0\mathbf{A} + \mathbf{A}^{\mathsf{H}}\mathbf{G}_0^{\mathsf{H}}\} \tag{41}$$

The tangential point $\mathbf{A}_1$ for starting the next MM iteration is obtained by minimizing the equivalent surrogate function described by Eq. (41). By introducing two auxiliary constant terms $\mathbf{tr}\{\mathbf{A}^{\mathsf{H}}\mathbf{A}\}$ and $tr\{\mathbf{G}_0\mathbf{G}_0^{\mathsf{H}}\}$, we have [23]

$$\begin{aligned}
\mathbf{A}_1 =& \arg\min_{\mathbf{A}}(\mathrm{tr}\{\mathbf{G}_0\mathbf{A} + \mathbf{A}^{\mathsf{H}}\mathbf{G}_0^{\mathsf{H}}\}) \\
=& \arg\min_{\mathbf{A}}(\mathrm{tr}\{\mathbf{G}_0\mathbf{A} + \mathbf{A}^{\mathsf{H}}\mathbf{G}_0^{\mathsf{H}} + \mathbf{A}^{\mathsf{H}}\mathbf{A} + \mathbf{G}_0\mathbf{G}_0^{\mathsf{H}}\}) \\
=& \arg\min_{\mathbf{A}}(\mathrm{tr}\{(\mathbf{A} - (-\mathbf{G}_0)^{\mathsf{H}})^{\mathsf{H}}(\mathbf{A} - (-\mathbf{G}_0)^{\mathsf{H}})\}) \\
=& \arg\min_{\mathbf{A}}(||\mathbf{A} - (-\mathbf{G}_0)||_{\mathrm{F}}^2) \\
=& \exp((\arg(-\mathbf{G}_0))
\end{aligned} \tag{42}$$

The complete pseudo-code for the MM algorithm based precoding is illustrated as Algorithm (3). The convergence of Algorithm (3) is also guaranteed. Due to the definition of surrogate function, the value of the objective function in the MM iterations (step 4 to step 13 in Algorithm (3)) must form a monotonically decreasing sequence. Therefore, the value of the objective function after the IRS optimization must be smaller than the value before. Thus, Algorithm (3) also converge. However, similar to the gradient descent method, Algorithm (3) may converge to different values if the initial points are chosen differently

V. Numerical Investigations

It is proven in [19] that by using Jensen's inequality, the average rate per user is bounded by

$$\bar{R} \geq \frac{LT_{\mathrm{s}}}{\tau + LT_{\mathrm{s}}} \log\left(1 + \frac{\sigma_s^2}{\sigma_{\tilde{n}_{\max}}^2 + P_{\mathrm{Interference}}}\right), \tag{43}$$

Algorithm 3 Precoding ($\mathbf{A}$, $\mathbf{X}$) with MM algorithm

$(\mathbf{A}, \mathbf{X}) \leftarrow$ initial values
while terminate condition not met **do**
  $\mathbf{X} \leftarrow \mathrm{convex}_{\mathrm{opt}}\{\left|\left|\tilde{\hat{\mathbf{H}}}^{\mathsf{T}}\mathbf{A}\mathbf{T}\mathbf{X} - \mathbf{S}\right|\right|_{\mathrm{F}}^{2} + \mathrm{tr}\{\mathbf{X}^{\mathsf{H}}\mathbf{T}^{\mathsf{H}}\mathbf{A}^{\mathsf{H}}\mathbf{\Sigma}_{\tilde{\mathbf{E}}}\mathbf{A}\mathbf{T}\mathbf{X}\} + \lambda\mathrm{tr}\{\mathbf{X}\mathbf{X}^{\mathsf{H}}\}\}$
  while terminate condition not met **do**
    $\mathbf{A}_0 \leftarrow \mathbf{A}$
    $\mathbf{U}_0 \leftarrow \mathbf{A}_0\mathbf{T}\mathbf{X}$
    $\mathbf{V} \leftarrow \mathbf{S}^{\mathsf{H}}\tilde{\hat{\mathbf{H}}}^{\mathsf{T}}$
    $\mathbf{Q} \leftarrow \tilde{\hat{\mathbf{H}}}^{*}\tilde{\hat{\mathbf{H}}}^{\mathsf{T}} + \mathbf{\Sigma}_{\tilde{\mathbf{E}}}$
    $\mathbf{P} \leftarrow \lambda_{\max}^{\mathbf{Q}}\mathbf{I}$
    $\mathbf{G}_0 \leftarrow \mathbf{T}\mathbf{X}\mathbf{U}_0^{\mathsf{H}}(\mathbf{Q} - \mathbf{P}) - \mathbf{T}\mathbf{X}\mathbf{V}$
    $\mathbf{A}_1 \leftarrow \exp((\arg(-\mathbf{G}_0))$
    $\mathbf{A} \leftarrow \mathbf{A}_1$
  end while
end while

where T_{s} is symbol time, τ is the duration of guard interval and $P_{\mathrm{Interference}}$ is the average interference power after the decoding/compensation process. The variance $\sigma^2_{\tilde{n}_{\max}}$ is the maximal power of the normalized noise, while σ_s^2 is the power of the intended signal.

Another value of interest is the PAPR. In our simulations, the chain-wise PAPR of every RF chain is calculated by dividing the chain-wise peak power by the average power of transmitted signals in each corresponding chain. The average PAPR is calculated by averaging the chain-wise PAPR over all the RF chains.

A. Channel Model

If we take scattering and antenna spatial correlations into consideration, the channel coefficients can be further modeled in a more detailed form as [13]

$$\mathbf{h}_k^{\mathsf{T}} = \frac{1}{\sqrt{P}}\sum_{p=1}^{P} h_p \mathbf{h}_{\mathrm{t}}^{\mathsf{T}}(\theta_p^{\mathsf{T}}, \phi_p^{\mathsf{T}}), \tag{44}$$

where $P \in \mathbb{N}$ is the number of effective channel paths, while $h_p \in \mathbb{C}$ is the coefficient that models the path loss and fading of the p-th path. The vector $\mathbf{h}_{\mathrm{t}}(\theta_p^{\mathsf{T}}, \phi_p^{\mathsf{T}}) \in \mathbb{C}^{M\times 1}$ is the steering vector which models the phase-shifts of the IRS elements relative to the reference element. The

directed angles $\theta_p^{\mathsf{T}} \in [0, \pi]$ and $\phi_p^{\mathsf{T}} \in [0, \pi]$ are the azimuthal and polar angle of the departure beam respectively. Fig. (3) describes this spherical cordinate system closely. The joint points in Fig. (3) represent the IRS reflecting elements.

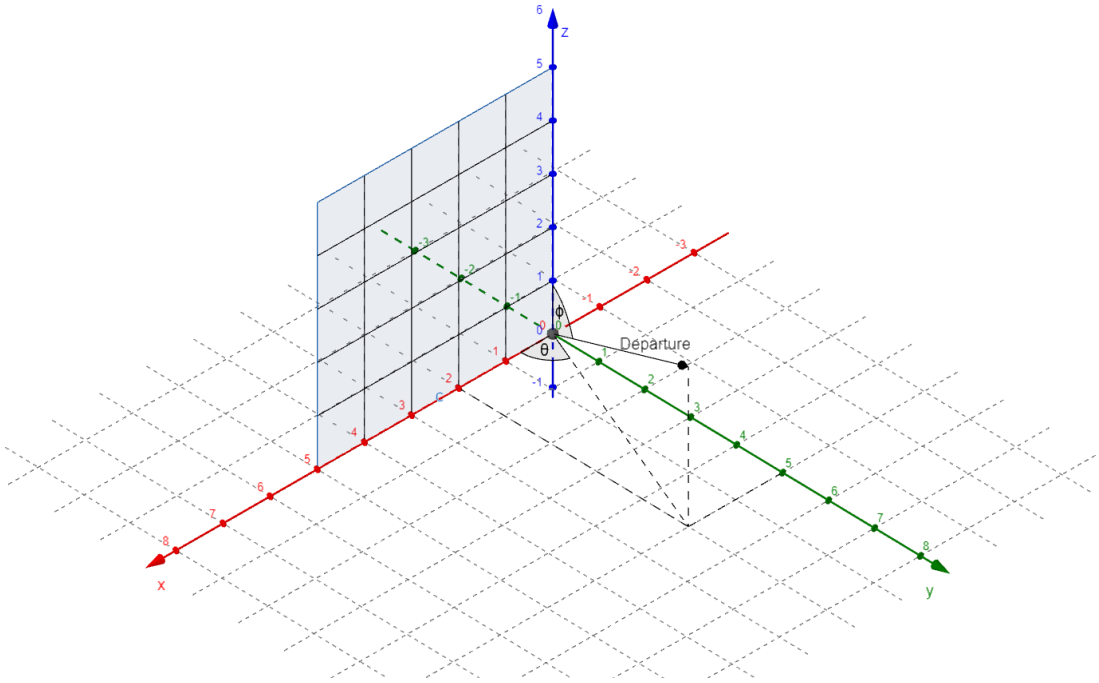

Fig. 3: Angle of departure

W.l.o.g., in the following context, the IRS is placed in the first quadrant of the xz-pane. We also assume that $\sqrt{M}$ is a positive integer. Moreover, we assume that the distances between any two neighboring IRS elements are constant and are denoted as d. The IRS elements are indexed from right to left (along x-axis, 0 to $+\infty$), bottom to top (along z-axis, 0 to $+\infty$). Based on these conventions, if $m \in \mathbb{N} \cap [1, M]$, the m-th IRS element will have the coordinate

$$\begin{bmatrix} x = ((m-1) \bmod \sqrt{M}) \cdot d, \\ y = 0, \\ z = \lfloor (m-1)/\sqrt{M} \rfloor \cdot d \end{bmatrix}. \tag{45}$$

The mod operator in the above equation is the modulo operation which is evaluated to the remainder of the division of $(m-1)$ by $\sqrt{M}$. The lower half bracket $\lfloor \cdot \rfloor$ is the floor operator that evaluates to the largest interger that is smaller or equal to the number inside it.

The steering vector $\mathbf{h}_{\mathrm{t}}^{\mathsf{T}}(\theta_p^{\mathsf{T}}, \phi_p^{\mathsf{T}})$ is described by

$$\mathbf{h}_{\mathrm{t}}(\theta_p^{\mathsf{T}}, \phi_p^{\mathsf{T}}) = \left[e^{j \frac{2\pi d}{\lambda_{\mathrm{w}}} [((m-1) \bmod \sqrt{M}) \sin(\phi_p^{\mathsf{T}}) \cos(\theta_p^{\mathsf{T}}) + \lfloor (m-1)/\sqrt{M} \rfloor \cos(\phi_p^{\mathsf{T}})]} \right]_{m,1}, \tag{46}$$

l_{ref}	λ_{w}	σ_f^2	$\theta_p^{\mathrm{t}}, \phi_p^{\mathrm{t}}$	d	n	P	K
$[50, 150]$	0.01	1	$[0, \pi]$	$\lambda_{\mathrm{w}}/2$	3	8	8

TABLE I: Parameters for channel $\mathbf{H}$

where λ_{w} denotes the wavelength of the transmitted signal.

In Eq. (44), the path loss and the fading is included in the coefficient h_p. According to [13], h_p is modeled as $h_p = \sqrt{\bar{h}_k}\tilde{h}_p$, where $\bar{h}_k$ is the path loss of the k-th UT and $\tilde{h}_p$ is the random fading of the p-th path. Their mathematical descriptions are given by

$$\begin{aligned} \bar{h}_k &= \left(\frac{\lambda_{\mathrm{w}}}{4\pi l_{\mathrm{ref}}}\right)^n \\ \tilde{h}_p &\sim \mathrm{CN}(0, \sigma_f^2). \end{aligned} \tag{47}$$

The variable $n \in \mathbb{R}$ is the path loss exponent and it is greater or equal than two.

B. Simulation Setup

In order to simulate the channel correlations, we place 8 UTs at the range of $50 + 14 \times k$ meters, where $k \in \{0, \ldots, 7\}$. Each BS-UT channel consists of 8 effective paths. The intended messages for the UTs have the same power $\sigma_s^2 = 1$. The angles from the BS to the scatterings are constant during the whole transmission.

Since block fading is assumed, the fading parameters are generated once every block. To make the further discussion easier, we assume that the distances between the UTs and the BS are deterministic and known by all the UTs and the BS. The default settings for the channel are concluded as Table (I).

The matrix $\mathbf{T}$ describes the illumination channel from the active antennas to the IRS as is defined in Eq.(4). In the simulations, N horn-shape active antennas are uniformly placed on a circle parallel to the xz-pane. The center of the the active antenna arrays and the center of the IRS form a line that is also parallel to the y-axis. The radius of the active antenna arrays is denoted as R_{r}, and the distance between this antenna array and the xz-pane is denoted as R_{d}. This array setup is exemplified in Fig. (4). To investigate the effect of interference on the IRS, two illumination strategies based on the orientations of active antennas are used.

In full illumination (FI), all active antenna illuminates toward the center of IRS. In this way, the IRS receives the most power as well as most interference. This illumination strategy is shown in Fig. (5).

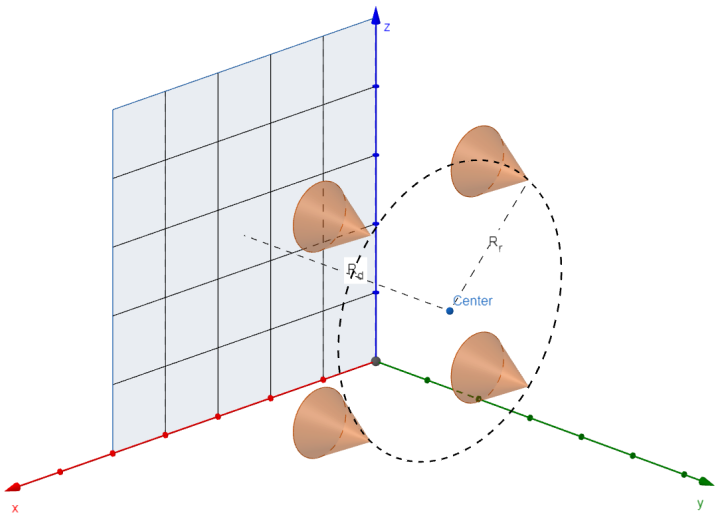

Fig. 4: Placement of the active antennas

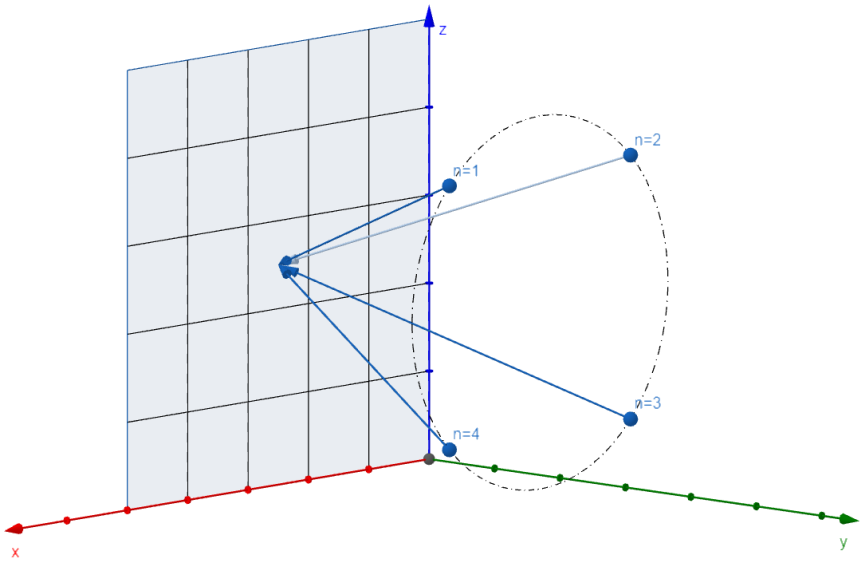

Fig. 5: Full illumination

ρ	R_{d}	R_{r}	κ
1	$\mathrm{S1}: \frac{4d\sqrt{M}}{\sqrt{\pi}}, \mathrm{S2}: \frac{4d\sqrt{M}}{\sqrt{N\pi}}$	$\mathrm{S1}: 2d, \mathrm{S2}: \frac{d\sqrt{2M}}{4}$	49

TABLE II: Parameters for fading and path loss

In partial illumination (PI), we designate N sections on the IRS, such that each section is responsible for one active antenna. Each active antenna illuminates the signals to the center of its designated section. Fig. (6) depicts this scenario.

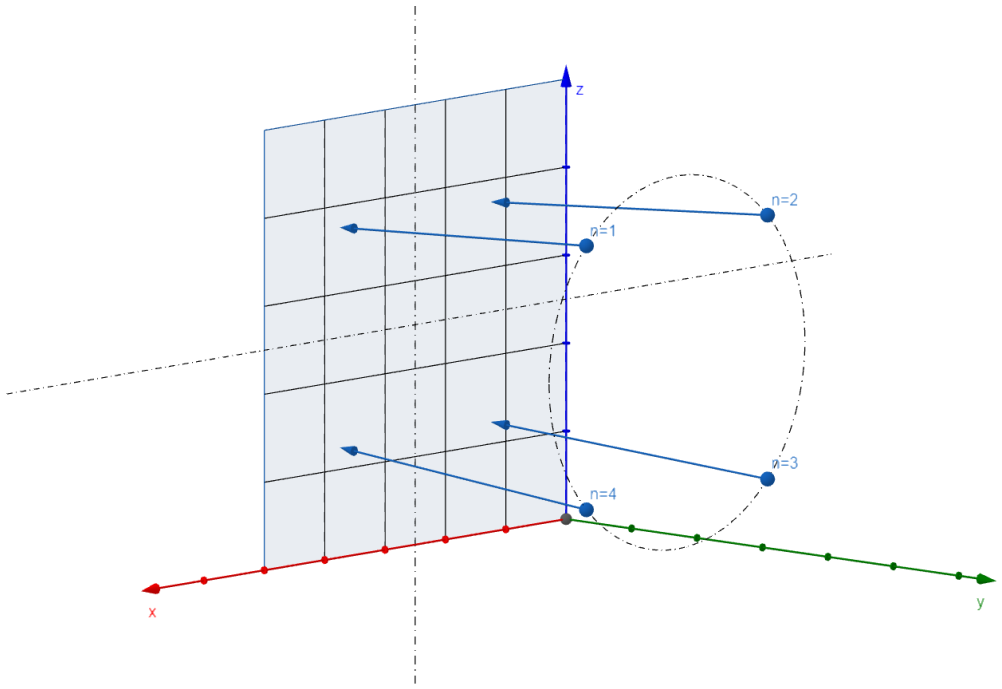

Fig. 6: Partial illumination

The illumination parameters are concluded in Table (II) [13].

C. Rate and PAPR

By using the above mentioned setup, and setting IRS antenna element amount to 64, we plot the rate performance for full illumination with gradient descent and MM algorithm. Partial illumination based on MM algorithm is also simulated for comparison. It is assume here that the estimation error is zero and the maximal normalized noise is set to be $\sigma^2_{\tilde{n}_{\max}} = 0.001$ such that it has a similar scale as the interference power. The PAPR is adjusted by increasing the peak

power constraint while the regularizer λ is fixed to be 0.5. For block length $L = 1$, the result is shown as Fig. (7).

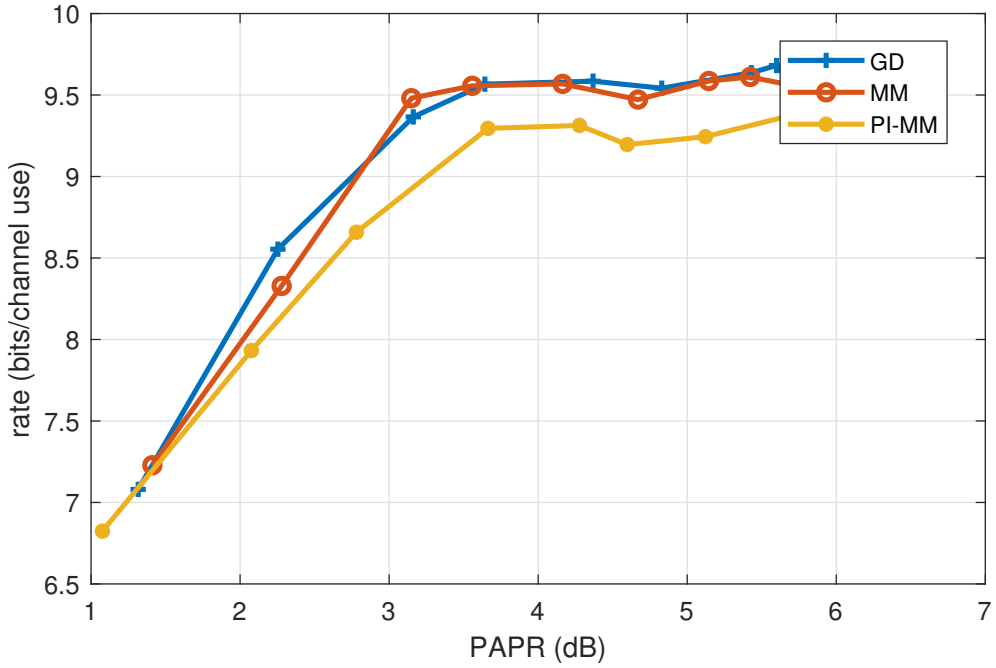

Fig. 7: Rate-PAPR with block length $L = 1$, guard interval $\tau = 0$ and regularizer $\lambda = 0.5$

It can be observed by comparing the red line and the yellow line that partial illumination strategy performs slightly worse than the full illumination strategy. One reason is the spillover effect caused by the antenna pattern. In full illumination, the IRS receives more power than the partial illumination strategy. In fact, it can be proven that given enough power and enough IRS size, symbol-wise distortion-free transmission is always achievable.

Theorem 1. *If* $\mathbf{H}$ *is a complex normal Gaussian random matrix with independent entries and the length of the uni-modular vector* $\mathbf{d}$ *can be arbitrary long, then it is almost certain that there is a* $\mathbf{d}$ *such that* $\mathbf{H} \cdot \mathbf{d} = \mathbf{0}$.

Proof. See appendix. □

For the precoding optimization problem with block length $L = 1$ and number of RF chains $N = 1$. Define $\mathbf{H}' = \mathbf{H} \times \mathrm{diag}_{\mathrm{matrix}}\{x\mathbf{t}\}$, and the problem is transformed to finding vector

$\mathbf{d} = \mathrm{diag}_{\mathrm{vector}}\{\mathbf{A}\}$ such that $\mathbf{s} = \mathbf{H}'\mathbf{d}$. To use the theorem, define

$$\mathbf{H}'' = \begin{bmatrix}\mathbf{H}', & \mathbf{s}\end{bmatrix}$$
$$\mathbf{d}' = \begin{bmatrix}\mathbf{d}_{\mathrm{aux}} \\ d_{M+1}\end{bmatrix}. \tag{48}$$

According the the theorem above, a solution for $\mathbf{d}'$ can be found such that $\mathbf{0} = \mathbf{H}''\mathbf{d}'$. Therefore the IRS parameters should be $\mathbf{A} = \mathrm{diag}_{\mathrm{matrix}}\{-d_{M+1}^{-1}\mathbf{d}_{\mathrm{aux}}\}$. For the cases where more RF chains are used $N > 1$, we can always transform it back to the case where $N = 1$ by muting all the RF chains except the first one.

For increased block length, where $L = 2$, the simulation results are shown in Fig. (8).

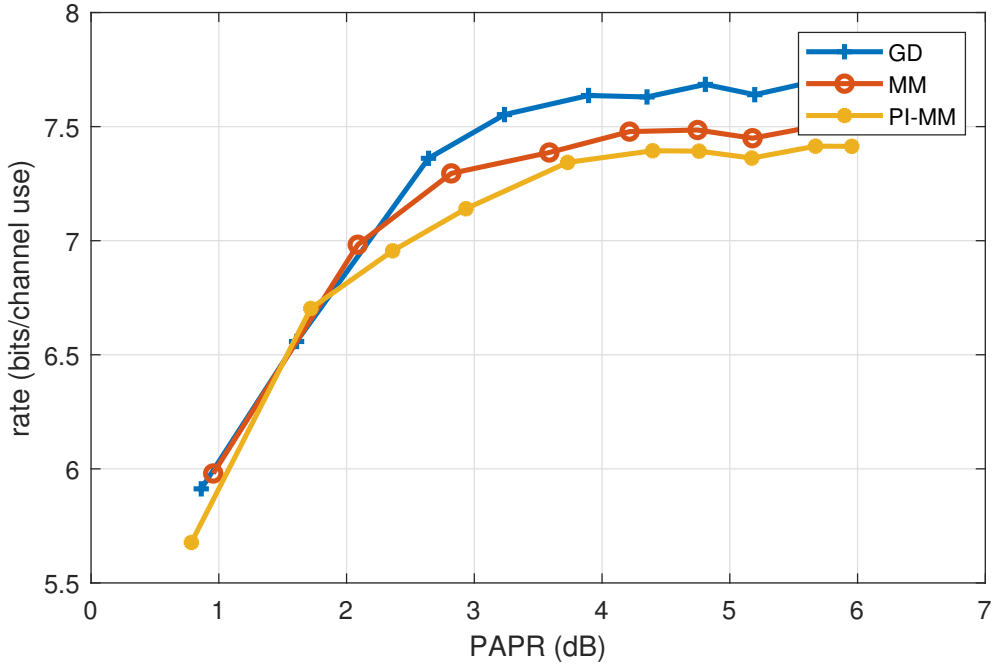

Fig. 8: Rate-PAPR with block length $L = 2$, guard interval $\tau = 0$ and regularizer $\lambda = 0.5$

It can be seen here that with guard interval $\tau = 0$, the systems with larger block lengths tend to have worse rate performance compared to the systems with smaller block lengths. This can be explained by code space density at the receivers side. Let's assume a fixed stream of code and messages with length L_{tot}. If we have two options for block lengths, L_1 and L_2. Suppose L_{tot} is dividable by L_1 and L_2.Following Eq. (28), we can see that the received code at the UTs side is actually generated by any combination of $(\mathbf{x}_i, \ldots, \mathbf{x}_{L_{\mathrm{tot}}}, \mathbf{A}_i, \ldots, \mathbf{A}_{L_{\mathrm{tot}}})$ such that $\mathbf{A}_i = \mathbf{A}_j$ if i

and j belong to the same time block. There are less constraints for the scheme where smaller block length is used. thus, the system which adopts a smaller block length should have a denser code stream space at the receiver side, and it is more likely for the transmitter to find a code stream that is closer to the intended messages.

D. Regularizer

The transmission power is controlled by the regularizer λ. It balances the emphasis between the interference power and the RF chains signal power.

In order to study the behavior of regularizer, we set the peak power constraint to infinity and plot the rate-λ curves in Fig. (9). The default illumination strategy is FI. And the channel estimation error is not included.

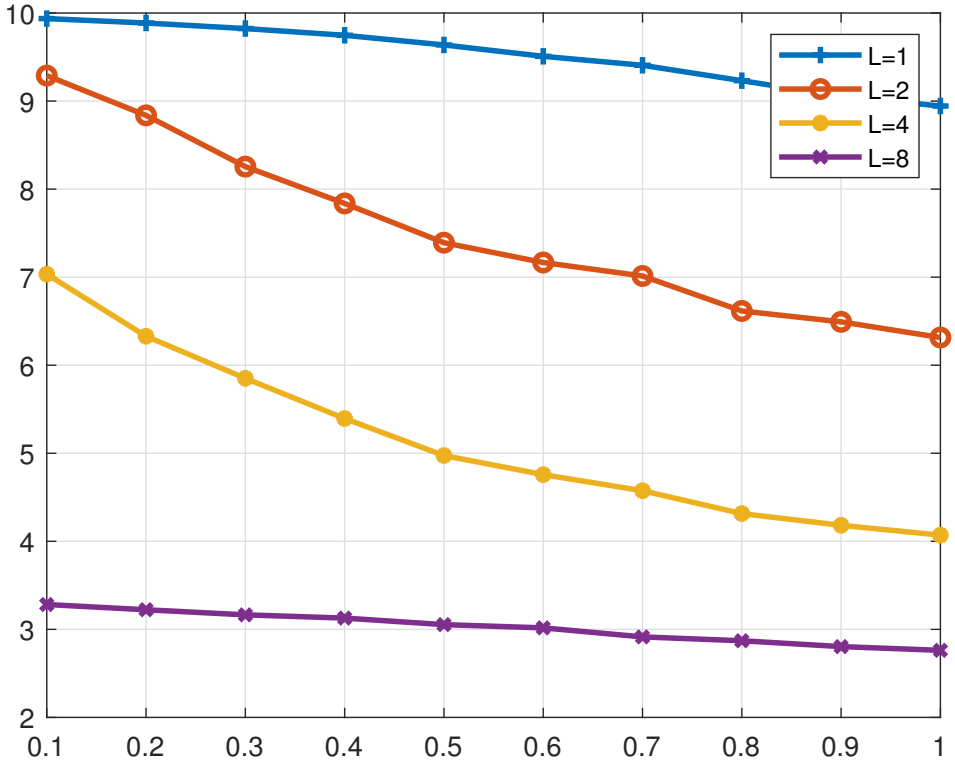

Fig. 9: Rate-λ behavior of MM-algorithm with guard interval $\tau = 0$

Since $\sigma_{\mathrm{s}}^2 = 1$ and $\sigma_{\tilde{\mathrm{n}}_{\max}}^2 = 0.001$, we would have $rate = \log_2 1001 \approx 9.97$ if the distortion power is zero. Fig. (9) shows that for $L = 1$, the curve approaches this rate limit as λ tend to

zero. This matches the Theorem (1) discussed above. Similar phenomenon was researched in [19]. It states that for full rank $\mathbf{H}$ and $\mathbf{S}$, the interference power $||\mathbf{H}\mathbf{F}_{\mathrm{GA}}\mathbf{X}-\mathbf{S}||_{\mathrm{F}}^2$ can never be made exactly to zero if the block length is larger than the number of RF chains. The matrix $\mathbf{F}_{\mathrm{GA}} \in \mathbb{C}^{M\times N}$ in the statement represents the fully connected analog unit and it can be any complex matrix of size $M \times N$. Since the IRS structure is a special case of general HAD system, this statement can also be used to explain the purple curve standing for $L = 8$. However, it is also stated in [19] that if the block length is smaller than the RF chain number, the distortion in a fully connected transmission system can always be made to zero. Since the optimization method we use here is sub-optimal, the validity of this proposition in IRS-aided transmission systems remains unknown.

The relation between the average RF chain power and the regularizer lambda is shown as Fig. (10). For relative small block lengths, the system with small L performs better. However, for

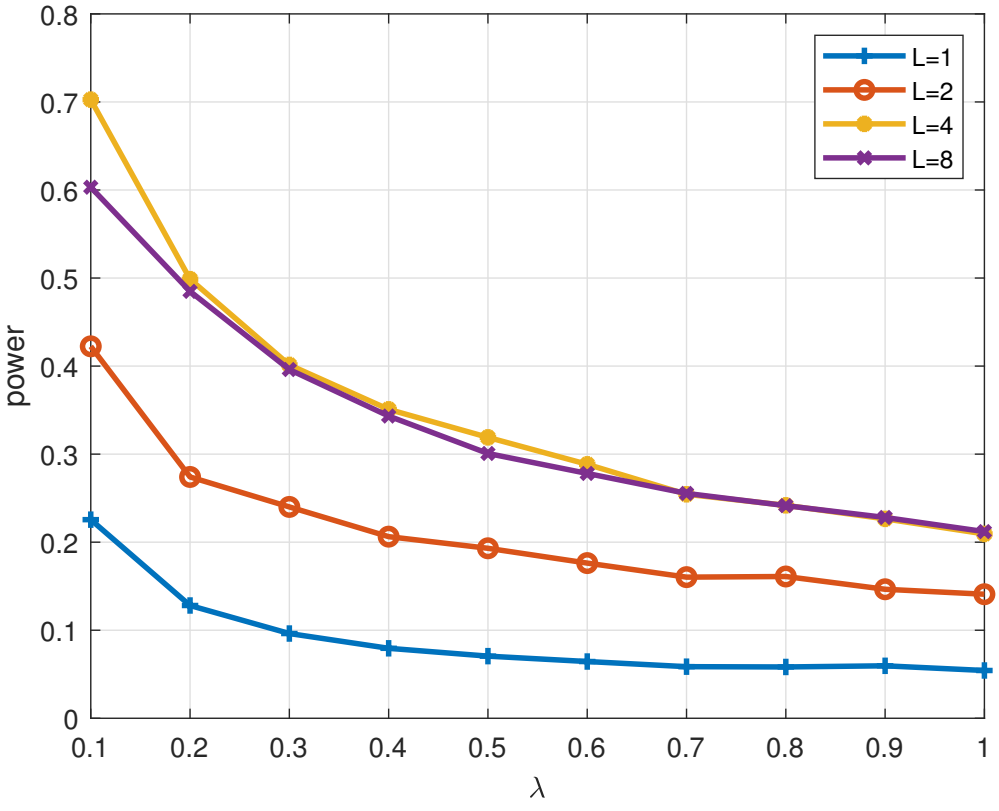

Fig. 10: Average RF chain power-λ with guard interval $\tau = 0$

systems with relative large block lengths, the power curves are close to each other. The reason

for this can be explained by analyzing the objective function.

For smaller block lengths, the interference power can be made very small. As a result, the dominant part of the objective function is the term for transmission power. Thus, the system with small L will put more emphasis on reducing the RF chain power. However, systems with larger L will concentrate on minimizing the interference power.

E. Rate, Block Length and Number of IRS Elements

Now, set the guard interval to $\tau = T_{\mathrm{s}}$ and remove the peak power constraint. We assumes that the normalized channel estimation error matrix is comprised of i.i.d. entries [21], i.e.

$$\forall k \in \mathbb{N} \cap [1, K], \sigma_{\mathrm{e},k}^2 / \bar{h}_k = \sigma_{\tilde{\mathrm{e}}}^2, \tag{49}$$

where $\sigma_{\tilde{\mathrm{e}}}^2$ is the variance of the entries of the normalized estimation error matrix.

The Fig. (11) shows that the achievable rate increases as the number of IRS elements increases and the rate decreases as the variance of the estimation error increases, which are expected.

Next, we want to study the rate-block length behavior. The number of IRS elements is set to $M = 144$, and the simulation results are plotted in Fig. (12). In this figure, the rate first increases and then decreases. This phenomenon is caused by the non-zero guard interval.

For smaller block length, the interference is small relative to the noise power. As a result, the dominant part is the guard interval. It can be seen that the at $L = 1$ the rate of the estimation error free curve is $4.9(\mathrm{bits/channeluse})$ which is close to one half of $\log_2 1001$. This is consistent with Theorem (1). As the block length increases, the term $\frac{LT_{\mathrm{s}}}{\tau + LT_{\mathrm{s}}}$ tends to 1 and the dominant part becomes the term $\log(1 + SINR)$.

Based on these two points, we can predict that for larger τ, the peak of the rate-L plot will shift to right since the term $\frac{LT_{\mathrm{s}}}{\tau + LT_{\mathrm{s}}}$ monotonically increases as L increases while the term $\log(1 + SINR)$ decreases as L increases. The complete relation of rate, $\sqrt{M}$ and L are shown as a mesh surface in Fig. (13).

At last it would also be interesting to know the minimal required IRS elements to achieve a certain rate with increasing block length. Instead of directly plotting by reading Fig. (13), we calculate the rate of each transmission block, and use this value to obtain the averaged minimal IRS size for reaching a certain rate.

For rate threshold $2.1(\mathrm{bits/channeluse})$, the averaged minimal M versus L is plotted as Fig. (14). It shows that the minimal required IRS size has an approximately quadratical growth rate

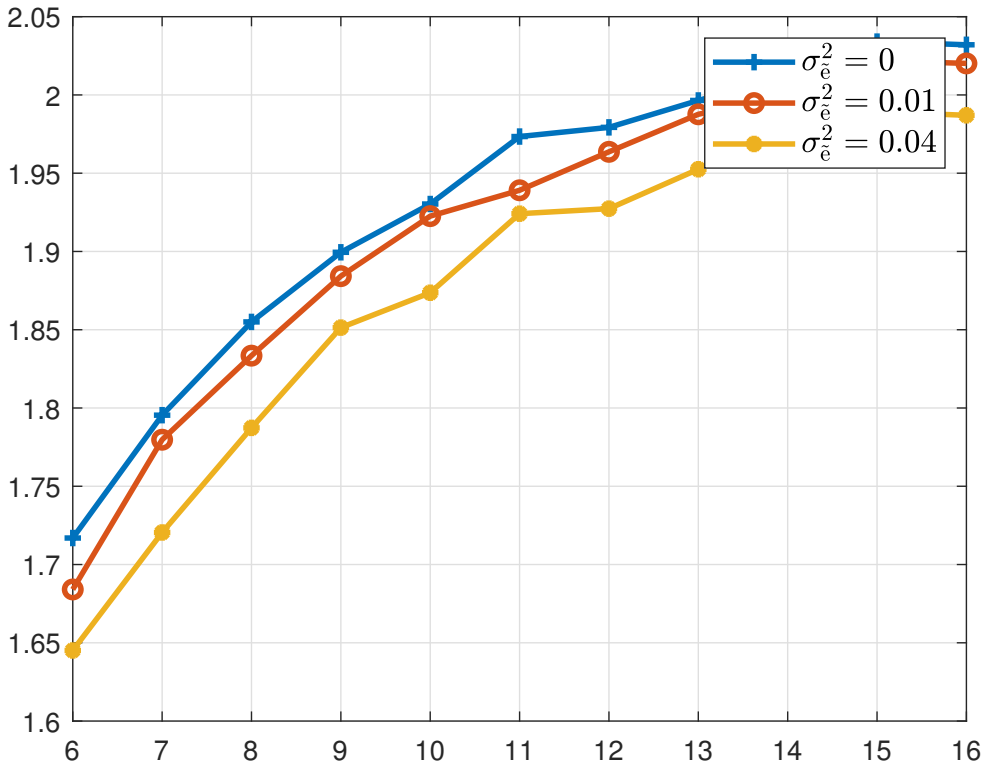

Fig. 11: Rate-$\sqrt{M}$ with guard interval $\tau = T_s$ and $L = 10$

as the block length if the threshold rate is achievable. This means we can trade the IRS update rate with IRS size efficiently. However, if both higher data rate and lower update rate are desired, we may have to increase the number of RF chains as is suggested in [19].

VI. Appendix

A. *Proof for Theorem 1*

For the simple case where $\mathbf{H}$ is a row vector, namely when there is only one user $K = 1$, define $\mathbf{h}^{\mathsf{T}} = \mathbf{H}$. We then try to find out the lower bound of probability $\Pr(\exists \mathbf{d}, (\mathbf{h}^{\mathsf{T}}\mathbf{d} = 0))$.

This can be translated into geometry representations. Each complex entry in vector $\mathbf{h}^{\mathsf{T}}$ is an 2-D vector whose length is a random variable following standard Gaussian distribution and phase is uniformly distributed between 0 to 2π. From this point of view, the uni-modular vector $\mathbf{d}$ represents the rotation operations which are applied to the columns of $\mathbf{h}^{\mathsf{T}}$ with controllable angles. Therefore, the proposition $\exists \mathbf{d}, (\mathbf{h}^{\mathsf{T}}\mathbf{d} = 0)$ can be interpreted as finding an rotation angle

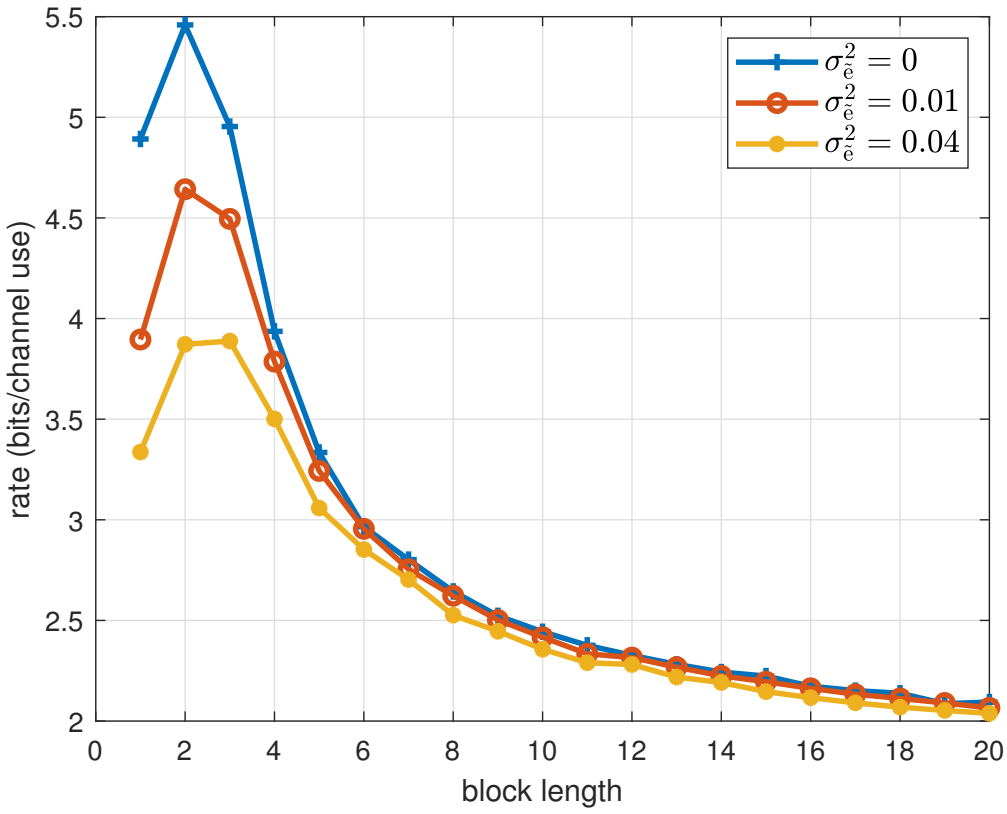

Fig. 12: Rate-L with guard interval $\tau = T_s$ and $M = 144$

for each of the entries in vector $\mathbf{h}^T$ such that these rotated entries add up to zero. Following triangular inequality, we can say that there exists a vector $\mathbf{d}$, such that $\mathbf{h}^T\mathbf{d} = 0$. This means

$$\max(|h_1|, \ldots, |h_M|) \leq |h_1| + \cdots + |h_M| - \max(|h_1|, \ldots, |h_M|). \tag{50}$$

It can also be observed that $|h_1| > |h_2| + \cdots + |h_M|$ and $|h_2| > |h_1| + \cdots + |h_M|$ will never happen simultaneously. Therefore, we can write the above probability in negation form while taking the independence assumption into consideration

$$\begin{aligned} P_1 =& \Pr(\exists \mathbf{d}, (\mathbf{h}^T\mathbf{d} = 0)) \\ =& 1 - (M \times \Pr(|h_1| > |h_2| + \cdots + |h_M|)) \end{aligned} \tag{51}$$

The properties of F-distribution can be used to determine the upper bound of $\Pr(|h_1| > |h_2| +$

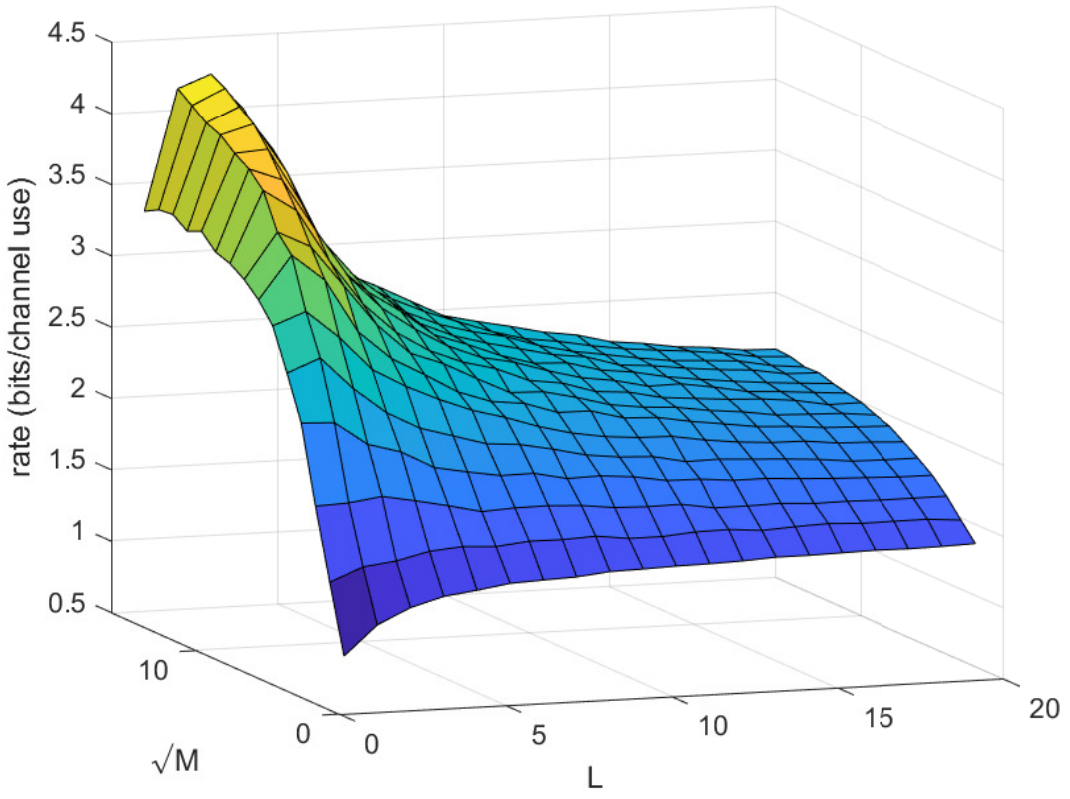

Fig. 13: Impact of IRS size and Block size ($\tau = 1$)

$\cdots + |h_M|)$.

$$
\begin{aligned}
& M \times \Pr(|X| > |X_1| + \cdots + |X_{M-1}|) \\
\leq & M \times \Pr(|X|^2 > |X_1|^2 + \cdots + |X_{M-1}|^2) \\
\leq & M \times \Pr(|X|^2 + |Y|^2 > |X_1|^2 + \cdots + |X_{M-1}|^2) \\
= & M \times \Pr\left(\frac{2}{M-1} > \frac{(|X_1|^2 + \cdots + |X_{M-1}|^2)/(M-1)}{(|X|^2 + |Y|^2)/2}\right) \\
= & M \times F\left(\frac{2}{M-1}; M-1, 2\right) = M \times I_{\frac{1}{2}}\left(\frac{M-1}{2}, 1\right) \\
= & M \times \left(\frac{1}{2}\right)^{\frac{M-1}{2}},
\end{aligned}
\tag{52}
$$

where $F(x; z_1, z_2)$ is the cumulative distribution function with z_1 and z_2 degrees of freedom and $I_x(a, b)$ is the regularized incomplete beta function. The first less equal sign holds because $|X|^2 > (|X_1|^2 + \cdots + |X_{M-1}|^2)^2 > |X_1|^2 + \cdots + |X_{M-1}|^2$.

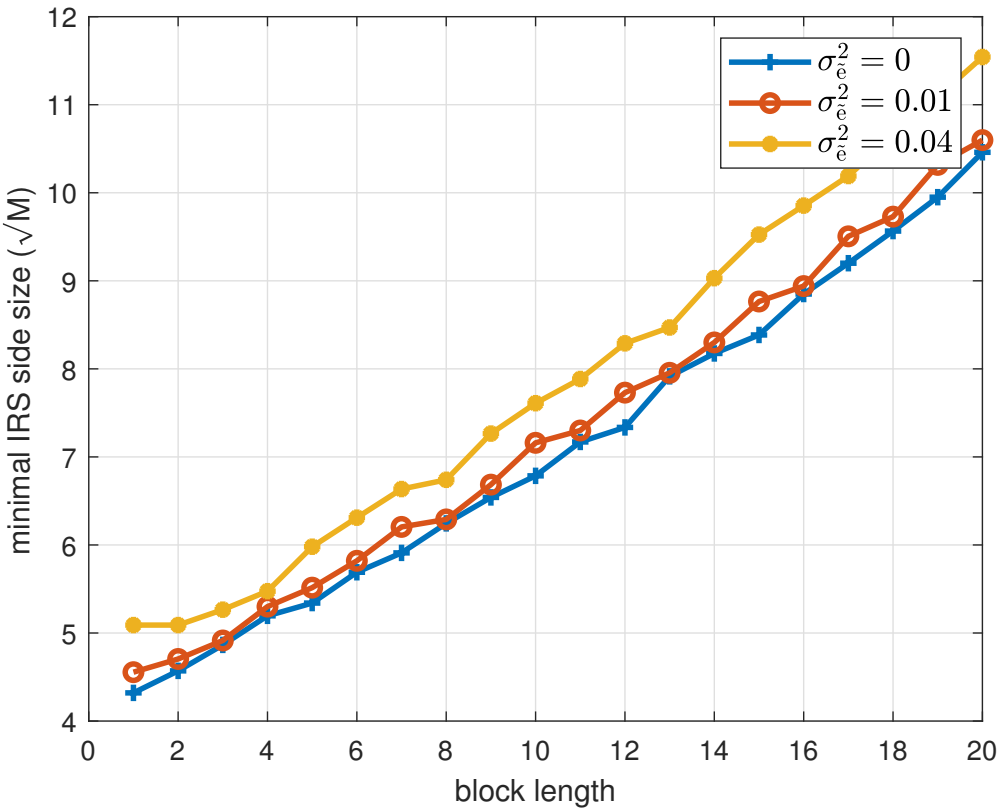

Fig. 14: Averaged minimal M-L with rate_threshold$= 1.8$

For the multi-user case where $K \neq 1$, the matrix $\mathbf{H}$ has more than one rows. A solution for $\mathbf{d}$ can be constructed recursively.

We first try to construct a sufficient condition for the existence of such vector $\mathbf{d}$. Then we try to prove that the probability for this sufficient condition not to hold tends to zero.

For sake of simplicity, we denote the channel matrix as $\mathbf{H}^{(k)}$ if $K = k$. Suppose we know how to construct $\mathbf{d}$ if $K = k-1$ such that $\mathbf{H}^{(k-1)} \cdot \mathbf{d} = \mathbf{0}$. This also implies that $(\mathbf{H}^{(k-1)} \cdot \mathbf{d}) \cdot e^{j\varphi} = \mathbf{0}$, where φ is an arbitary number.

Based on these assumptions and discussions, for the case where $K = k$, we can sectorize $\mathbf{H}^{(k)}$ as Fig. (15) and sectorize $\mathbf{d}$ into several vecrtors correspondingly.

We only need to find one solution for this problem, and the condition for that solution to exist can serve as sufficient condition for $\exists \mathbf{d}, (\mathbf{H}^{(k)}\mathbf{d} = 0)$.

One instance of $\mathbf{d}$ may be constructed via the following steps. We first find $\mathbf{d}_1, \ldots, \mathbf{d}_B$ such that $\forall i \in \{1, \ldots, B\}, \mathbf{H}_i^{(k-1)}\mathbf{d}_i = \mathbf{0}$. Then for the last row of matrix $\mathbf{H}^{(k)}$, we know that $\mathbf{h}_i^T \cdot \mathbf{d}_i$ is

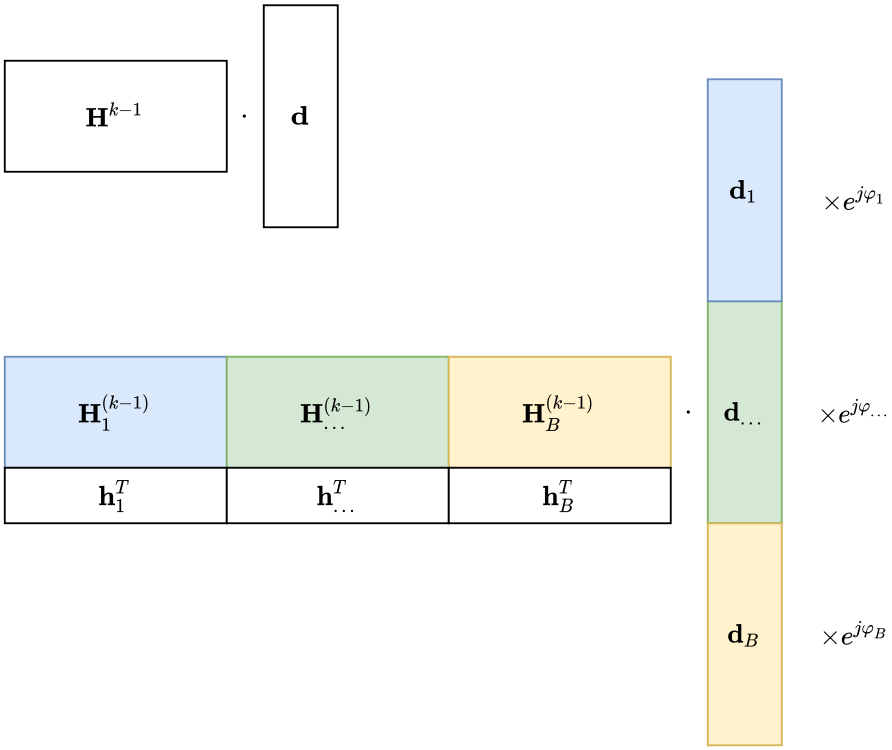

Fig. 15: Idea of construction

a scalar value. At the same time, for each subvector $\mathbf{d}_i$, we can still apply an additional rotation $e^{j\varphi_i}$ to each of the vectors $\mathbf{d}_i$.

Now the problem is transformed back to the simple case, where we want to find rotation angles $\varphi_1, \dots, \varphi_B$ such that

$$\begin{bmatrix} \mathbf{h}_1^\mathsf{T}\mathbf{d}_1 & \dots & \mathbf{h}_B^\mathsf{T}\mathbf{d}_B \end{bmatrix} \cdot \begin{bmatrix} e^{j\varphi_1} \\ \vdots \\ e^{j\varphi_B} \end{bmatrix} = 0 \tag{53}$$

By finding appropriate $\mathbf{d}_i$ and φ_i, the rotation vector $\mathbf{d}$ is obtained. This is the general idea for the proof. Next, we will provide a detailed proof for this theorem.

Suppose $\mathbf{H}^{(k)}$ has M columns. We can find at least one $E \in \mathbb{N}$, such that $M \in [E^k, (E+1)^k]$. We then sectorize the matrix $\mathbf{H}^{(k)}$ with the scheme shown as Fig. (16). It is worth noticing that the sectorization method is not unique, the only requirement is that the number of columns in each sub-matrix lies within the range of $[E^{k-1}, (E+1)^{k-1}]$. After the sectorization, there are either E or $E+1$ blocks in total.

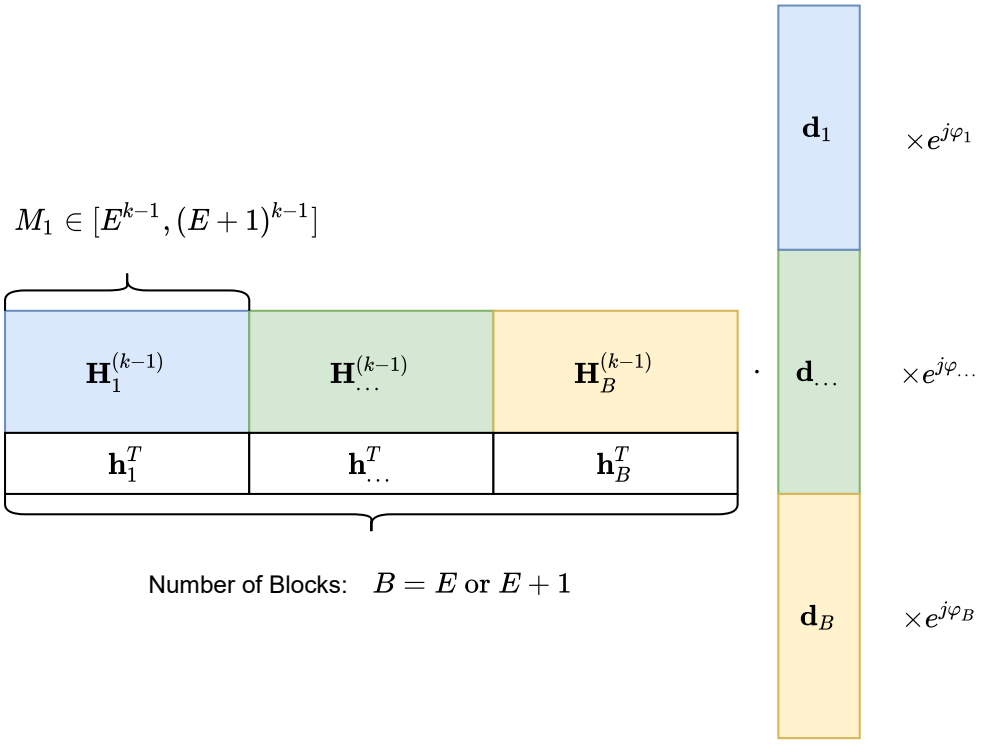

Fig. 16: Submatrix Construction

If the rotation vector $\mathbf{d}$ cannot be found with the idea illustrated in Fig (15), then one of the two sufficient conditions does not hold, i.e., either one of the colored submatrix $\mathbf{H}_i^{(k-1)}$ cannot be rotated to zero or the last row of $\mathbf{H}^{(k)}$ cannot be made zero. Recursive method can be used to calculate the probability that one of the colored sub-matrices cannot be made zero. We focus on analyzing the last row of matrix $\mathbf{H}^{(k)}$ first. Define

$$\begin{aligned} &B \in \{E, E+1\} : \text{number of blocks} \\ &M_{1\dots B} \in [E^{k-1}, (E+1)^{k-1}] : \text{number of columns in one block.} \end{aligned} \tag{54}$$

as is also illustrated in Fig. (16). Each block $\mathbf{h}_i^T$ in the last row will be multiplied with correspondent vector block $\mathbf{d}_i$,which creates a complex Gaussian random variable $\mathbf{h}_i^T\mathbf{d}_i \sim CN(0, M_i)$. The F-distribution is defined on the normal Gaussian variables. Thus, we should first normalize the variance to 1. This is done in Eq. (55) and Eq. (56) Define

$$\begin{aligned} &\forall b \in \{1, \dots, B\}, Y_b = \mathbf{h}_b^{\mathsf{T}}\mathbf{d}_b \sim \mathrm{CN}(0, M_b) \\ &\forall b \in \{1, \dots, B\}, X_b \sim \mathrm{CN}(0, 1). \end{aligned} \tag{55}$$

Then we have

$$\begin{aligned}
&\sum_{b=1}^{B}\Pr\left(|Y_b|>|Y_1|+\cdots+|Y_{b-1}|+|Y_{b+1}|+\cdots+|Y_B|\right)\\
\leq&\sum_{b=1}^{B}\Pr\left(|Y_b|>\sqrt{M_{\min}}\left(\frac{|Y_1|}{\sqrt{M_i}}+\cdots+\frac{|Y_{b-1}|}{\sqrt{M_{b-1}}}+\frac{|Y_{b+1}|}{\sqrt{M_{b+1}}}+\cdots+\frac{|Y_B|}{\sqrt{M_B}}\right)\right)\\
=&\sum_{b=1}^{B}\Pr\left(\frac{\sqrt{M_b}}{\sqrt{M_{\min}}}\frac{|Y_b|}{\sqrt{M_b}}>|X_1|+\cdots+|X_{B-1}|\right)\\
\leq&B\times\Pr\left(\frac{\sqrt{M_{\max}}}{\sqrt{M_{\min}}}|X|>|X_1|+\cdots+|X_{B-1}|\right)\\
\leq&B\times\left(\frac{M_{\max}}{M_{\max}+M_{\min}}\right)^{\frac{B-1}{2}}.
\end{aligned}\tag{56}$$

For larger M, we have

$$\lim_{M\to+\infty}\frac{M_{\max}}{M_{\max}+M_{\min}}\leq\lim_{E\to+\infty}\frac{(E+1)^{k-1}}{2\times E^{k-1}}=\frac{1}{2}\leq\frac{2}{3}.\tag{57}$$

Therefore,

$$\Pr\left(\nexists(\varphi_1,\cdots,\varphi_B).\left[\mathbf{h}_1^{\mathsf{T}}\mathbf{d}_1\quad\ldots\quad\mathbf{h}_B^{\mathsf{T}}\mathbf{d}_B\right]\cdot\begin{bmatrix}e^{j\varphi_1}\\\vdots\\e^{j\varphi_B}\end{bmatrix}=0\right)\leq(N+1)\times\left(\frac{2}{3}\right)^{\frac{E-1}{2}}.\tag{58}$$

The bound for the complete probability can then be obtained

$$\begin{aligned}
P_k&=\Pr(\nexists\mathbf{d},(\mathbf{H}^{(k)\mathsf{T}}\mathbf{d}=0))\\
&\leq(E+1)\times P_{k-1}+(E+1)\times\left(\frac{2}{3}\right)^{\frac{E-1}{2}}.
\end{aligned}\tag{59}$$

Based on this inequality, one can define an auxiliary sequence for the upper bound of the probabilities iteratively

$$\begin{aligned}
P'_1&=(E+1)\times\left(\frac{2}{3}\right)^{\frac{E-1}{2}}\\
P'_k&=(E+1)\times P_{k-1}+(E+1)\times\left(\frac{2}{3}\right)^{\frac{E-1}{2}}.
\end{aligned}\tag{60}$$

By using mathematical induction, it can be proven that $\forall k\in\mathbb{N}, P_k<P'_k$. This sequence can also be written in analytic form by solving difference equation as

$$P'_k=\frac{(E+1)^{k+1}-(E+1)}{E}\left(\frac{2}{3}\right)^{\frac{E-1}{2}}.\tag{61}$$

It can be observed that this upper bound will converge to zero if E tend to infinity.

REFERENCES

[1] A. Alkhateeb, J. Mo, N. Gonzalez-Prelcic, and R. W. Heath, "Mimo precoding and combining solutions for millimeter-wave systems," *IEEE Communications Magazine*, vol. 52, no. 12, pp. 122–131, 2014.

[2] X. Gao, L. Dai, S. Han, I. Chih-Lin, and R. W. Heath, "Energy-efficient hybrid analog and digital precoding for mmwave mimo systems with large antenna arrays," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 4, pp. 998–1009, 2016.

[3] Q. Wu, S. Zhang, B. Zheng, C. You, and R. Zhang, "Intelligent reflecting surface-aided wireless communications: A tutorial," *IEEE Transactions on Communications*, vol. 69, no. 5, pp. 3313–3351, 2021.

[4] Q. Wu and R. Zhang, "Towards smart and reconfigurable environment: Intelligent reflecting surface aided wireless network," *IEEE Communications Magazine*, vol. 58, no. 1, pp. 106–112, 2019.

[5] ——, "Intelligent reflecting surface enhanced wireless network via joint active and passive beamforming," *IEEE Transactions on Wireless Communications*, vol. 18, no. 11, pp. 5394–5409, 2019.

[6] D. Xu, X. Yu, Y. Sun, D. W. K. Ng, and R. Schober, "Resource allocation for secure irs-assisted multiuser miso systems," in *2019 IEEE Globecom Workshops (GC Wkshps)*. IEEE, 2019, pp. 1–6.

[7] D. Xu, V. Jamali, X. Yu, D. W. K. Ng, and R. Schober, "Optimal resource allocation design for large irs-assisted swipt systems: A scalable optimization framework," *IEEE Transactions on Communications*, vol. 70, no. 2, pp. 1423–1441, 2022.

[8] M. Di Renzo, A. Zappone, M. Debbah, M.-S. Alouini, C. Yuen, J. De Rosny, and S. Tretyakov, "Smart radio environments empowered by reconfigurable intelligent surfaces: How it works, state of research, and the road ahead," *IEEE journal on selected areas in communications*, vol. 38, no. 11, pp. 2450–2525, 2020.

[9] M. Di Renzo, H. Haas, and P. M. Grant, "Spatial modulation for multiple-antenna wireless systems: A survey," *IEEE Communications Magazine*, vol. 49, no. 12, pp. 182–191, 2011.

[10] E. Basar, "Reconfigurable intelligent surface-based index modulation: A new beyond mimo paradigm for 6g," *IEEE Transactions on Communications*, vol. 68, no. 5, pp. 3187–3196, 2020.

[11] R. Liu, M. Li, Q. Liu, A. L. Swindlehurst, and Q. Wu, "Intelligent reflecting surface based passive information transmission: A symbol-level precoding approach," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 7, pp. 6735–6749, 2021.

[12] R. Karasik, O. Simeone, M. Di Renzo, and S. S. Shitz, "Beyond max-snr: Joint encoding for reconfigurable intelligent surfaces," in *2020 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2020, pp. 2965–2970.

[13] V. Jamali, A. M. Tulino, G. Fischer, R. R. Müller, and R. Schober, "Intelligent surface-aided transmitter architectures for millimeter-wave ultra massive mimo systems," *IEEE open journal of the communications society*, vol. 2, pp. 144–167, 2020.

[14] A. Bereyhi, V. Jamali, R. R. Müller, G. Fischer, R. Schober, and A. M. Tulino, "Papr-limited precoding in massive mimo systems with reflect-and transmit-array antennas," in *2019 53rd Asilomar Conference on Signals, Systems, and Computers*. IEEE, 2019, pp. 1690–1694.

[15] S. Domouchtsidis, C. G. Tsinos, S. Chatzinotas, and B. Ottersten, "Constant envelope mimo-ofdm precoding for low complexity large-scale antenna array systems," *IEEE Transactions on Wireless Communications*, vol. 19, no. 12, pp. 7973–7985, 2020.

[16] A. Li, C. Masouros, and F. Liu, "Hybrid analog-digital precoding for interference exploitation," in *2018 26th European Signal Processing Conference (EUSIPCO)*. IEEE, 2018, pp. 812–816.

[17] R. Liu, H. Li, and M. Li, "Symbol-level hybrid precoding in mmwave multiuser miso systems," *IEEE Communications Letters*, vol. 23, no. 9, pp. 1636–1639, 2019.

[18] A. Li, C. Shen, X. Liao, C. Masouros, and A. L. Swindlehurst, "Practical interference exploitation precoding without symbol-by-symbol optimization: A block-level approach," *arXiv preprint arXiv:2202.09830*, 2022.

[19] M. A. Sedaghat, B. Gade, R. R. Müller, and G. Fischer, "A novel hybrid analog-digital transmitter for multi-antenna base stations," in *2017 25th European Signal Processing Conference (EUSIPCO)*. IEEE, 2017, pp. 1714–1718.

[20] J. Jose, A. Ashikhmin, P. Whiting, and S. Vishwanath, "Channel estimation and linear precoding in multiuser multiple-antenna tdd systems," *IEEE Transactions on Vehicular Technology*, vol. 60, no. 5, pp. 2102–2116, 2011.

[21] P. Zetterberg, "Experimental investigation of tdd reciprocity-based zero-forcing transmit precoding," *EURASIP Journal on Advances in Signal Processing*, vol. 2011, pp. 1–10, 2011.

[22] A. Bereyhi, M. A. Sedaghat, R. R. Müller, and G. Fischer, "Glse precoders for massive mimo systems: Analysis and applications," *IEEE Transactions on Wireless Communications*, vol. 18, no. 9, pp. 4450–4465, 2019.

[23] S. P. Sankuru and P. Babu, "Designing unimodular sequence with good auto-correlation properties via block majorization-minimization method," *Signal Processing*, vol. 176, p. 107707, 2020.