Block-Level Precoding via IRS-Aided Hybrid Transmitter

Zilu Zhao, Ali Bereyhi

Abstract

This article proposes a transmission scheme for a hybrid Analog-Digital system where passive antenna array is used as analog unit. This scheme can lower the update rate of the analog unit by stacking several input vectors together. To determine the optimal phase shifts of the passive antennas, convex projection and gradient descent method are used and compared. This paper also shows experimentally that the number of passive antennas may affect the performance of this system.

I. Introduction

In the past few years, the Fifth Generation communication system (5G) is becoming more and more widespread. Massive multiple-input multiple-output (MIMO) systems are one of the crucial technologies that make ”5G” possible. It increases the channel capacity significantly without the exponential increament of the transmittion power by utilizing the diversity gain. This often leads to higher complexity and cost since each RF chain includes a power amplifier and an digital-to-analog converter. The high cost of the RF chains is a major holdback that limits the antenna number of base stations. For a time division duplex (TDM) system, if the base station has the perfect channel state information (CSI), reciprocity property can always be used. Therefore, only downlink is investigated in this paper.

Hybrid analog-digital (HAD) transmission is a rather recent technique for reducing the implementational complexity in MIMO systems. An HAD transmitter consists of a digital base-band unit and an analog unit which operates in the radio frequency (RF) domain. There are various technologies by which the analog unit can be implemented; see for instance radio frequecy beamforming networks with buttler matrix [1]. In this paper, we consider a recent proposal which implements the analog units via intelligent reflecting surfaces (IRSs).

Author is with IDC, FAU.

Intuitively, some information is encoded into the output signal by the analog unit. Therefore, the degree of freedom provided by RF chains can be reduced. In other words, the number of RF chains can be reduced. This can lower the cost and complexity of base stations. According to [**?**], the update rate of analog unit can be set slower than the update rate of the digital unit for a system with a general linear analog unit without causing any interference. To investigate this point in IRS-aided HAD transmission systems, block fading is assumed and a block-wise precoding procedure will be discussed. The IRS unit receives signals from the digital units and then reflects the signals with some tunable phase shifts. The digital unit is used to precode the user messages and tune the phase shifts of the IRS unit properly.

In the first part of this paper, the system model for the IRS-aided HAD transmission system is discussed in detail. Then in Section III, a digital baseband precoding scheme is introduced. The proposed analog beamforming scheme deals with a unit-modulus optimization problem. To address this problem, we develop three low-complexity alogorithms by means of convex projection technique, method of gradient descent and Majorize-minimization (MM) algorithm. Our derivation initially consider the ideal cas with perfect CSI at the transmitter. We then extend the derivations to more realistic scenarios with imperfect CSI. The performance of the proposed algorithms are further investigated through several numerical simulations. Motivated by the findings in the first part, we show that interference free transmission is achievable at the base station and a large enough IRS is employed. A proof for this result in the system with block length one will be given. Finally, in Section V, we will give some conclusions.

II. System Model

A. Transmission Procedure

We consider downlink transmission in a multi-user MIMO system consisting of a base station (BS) with M antenna elements and K user terminals (UTs). The vector of the received signals in this network in the i-th symbol time instance is given by

$$\mathbf{y}_i = \mathbf{H}\mathbf{a}_i + \mathbf{n}_i, \tag{1}$$

where $\mathbf{H} \in \mathbb{C}^{K\times M}$ is the wireless channel matrix from base station to all the UTs if the CSI is known. The receiving vector can be written as $\mathbf{y}_i = \begin{bmatrix} y_1 & \dots & y_K \end{bmatrix}^T$ with y_k denoting the signal at the k-th UT. Vector $\mathbf{a}_i \in \mathbb{C}^{M\times 1}$ represents the signal at the BS and $\mathbf{n}_i$ is a K by one noise

vector whose entries are i.i.d. drawn from Gaussian distribution with zero mean and variance σ^2.

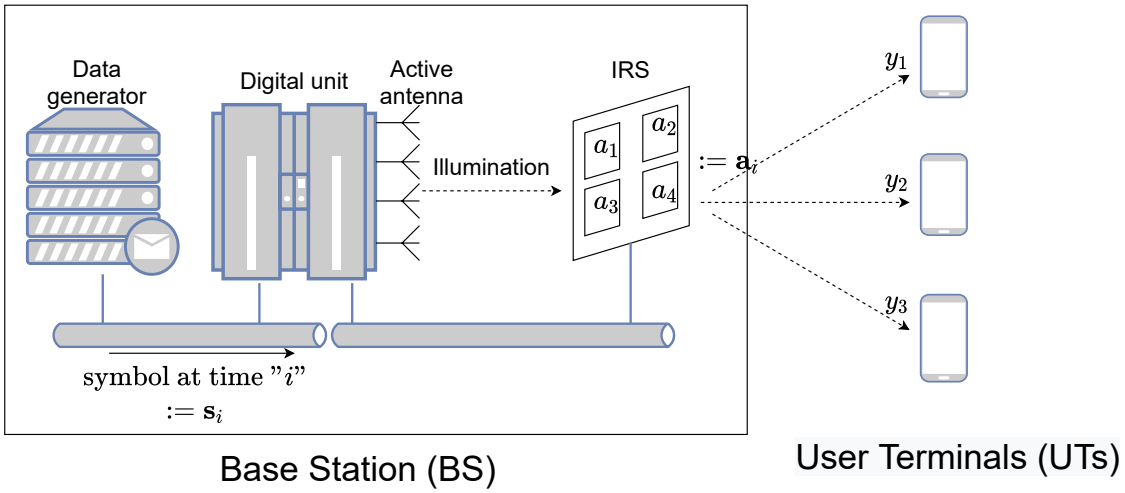


Fig. (1) Transmitter Model

The BS employs an IRS-aided HAD transmitter whose architecture is demonstrated in Fig. (1).

If we define the message symbol intended for the k-th UT as s_k, then the message vector for all the K UTs at the i-th time interval can be denoted as $\mathbf{s}_i = \begin{bmatrix} s_1 & \dots & s_K \end{bmatrix}^T$. For block-wise precoding with block size L, a buffer can be attached to the data generator, and the message for one entire block is represented as $\mathbf{S} = \begin{bmatrix} \mathbf{s}_1 & \dots & \mathbf{s}_L \end{bmatrix}$.

For further discussion, a more detailed illustration for the internal structure of the BS with N RF chains and M passive IRS elements is shown as Fig. (2).

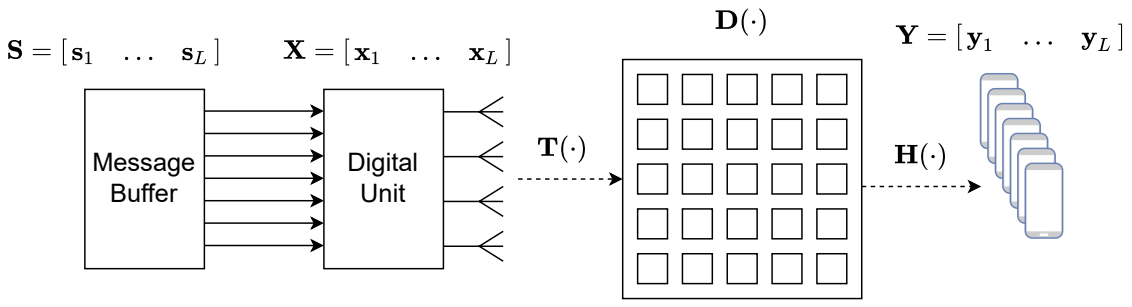


Fig. (2) Signal diagram of the Base Station

If the matrix $\mathbf{S}$ is the intended message for the UTs during one transmission time block,

we can define a matrix $\mathbf{X} \in \mathbb{C}^{N \times L_X}$ to be the RF chain output signal corresponding to that message matrix. Additionally, the phase shifts of the IRS for the message matrix is denoted by $\mathbf{D} = \mathrm{diag}\{e^{j\beta_1}, \ldots, e^{j\beta_M}\}$. The IRS is placed in the close proximity of the active antennas. Therefore, the illumination channel between them can be considered as fixed during the entire transmission. This channel can be denoted by a constant matrix $\mathbf{T} \in \mathbb{C}^{M \times N}$.

Since the UTs are disjoint and don't pocess the same computational power as the base station, we want to design a precoding procedure which requires only a simple disjoint decoder which can be modeled as a diagonal matrix $\mathbf{F}_D$, i.e. $\tilde{\mathbf{Y}} = \mathbf{F}_D\mathbf{Y} = \mathbf{S}$.

The received signal $\mathbf{Y}$ can also be represented as

$$\mathbf{Y} = \mathbf{HDTX} + \mathbf{N}. \tag{2}$$

In order to make the decodeded signal as close as possible to the intended message, $\mathbf{Y}$ and $\mathbf{S}$ must have the same dimension. Therefore, the matrix $\mathbf{X}$ must have as many column as matrix $\mathbf{S}$. Thus, $L_X = L$ and $\mathbf{X} \in \mathbb{C}^{N \times L}$.

After the digital unit receives the entire intended message block, it will computer the optimal RF chain output $\mathbf{X}$ and the IRS phase shift $\mathbf{D}$. If we expand the matrix $\mathbf{Y}$ and $\mathbf{X}$ in Eq. (2), we have

$$\begin{bmatrix} \mathbf{y}_1 & \ldots & \mathbf{y}_L \end{bmatrix} = \mathbf{HDT} \begin{bmatrix} \mathbf{x}_1 & \ldots & \mathbf{x}_L \end{bmatrix} + \mathbf{N}, \tag{3}$$

where $\mathbf{x}_i \in \mathbb{C}^{N \times 1}$ denotes the active antenna output vector corresponding to the i-th message vector. It can be observed that at every time interval, one column of X is transmitted by the active antennas and get reflected by the IRS. However, the same phase shift matrix $\mathbf{D}$ is applied to all the active antenna output vectors in the same time block. This indicates that the active antennas are updated once every time interval, but the IRS only updates once every time block.

As the buffer stacks the message symbol vectors into a matrix, it also introduces a delay. The message symbols won't be fed to the digital unit until L message vectors are generated to fill an entire time block. Thus, there is always a delay of an entire time block between vector $\mathbf{s}_i$ and $\mathbf{x}_i$. Therefore, this system has two phases. The first phase is the computation phase that happens between two time blocks. During this phase, the buffer unit feed the message matrix $\mathbf{S}$ to the digital unit. After that the digital unit calculates the suitable RF chain symbol matrix $\mathbf{X}$ and the IRS configuration $\mathbf{D}$. The second phase is the transmission phase that happens inside one time block. During the transmission phase, the digital unit will illuminate the matrix $\mathbf{X}$ calculated in

the previous phase to the IRS while the IRS is reconfigurated to matrix $\mathbf{D}$. If the computation phase is instant, this procedure can be illustrated in Fig. (3).

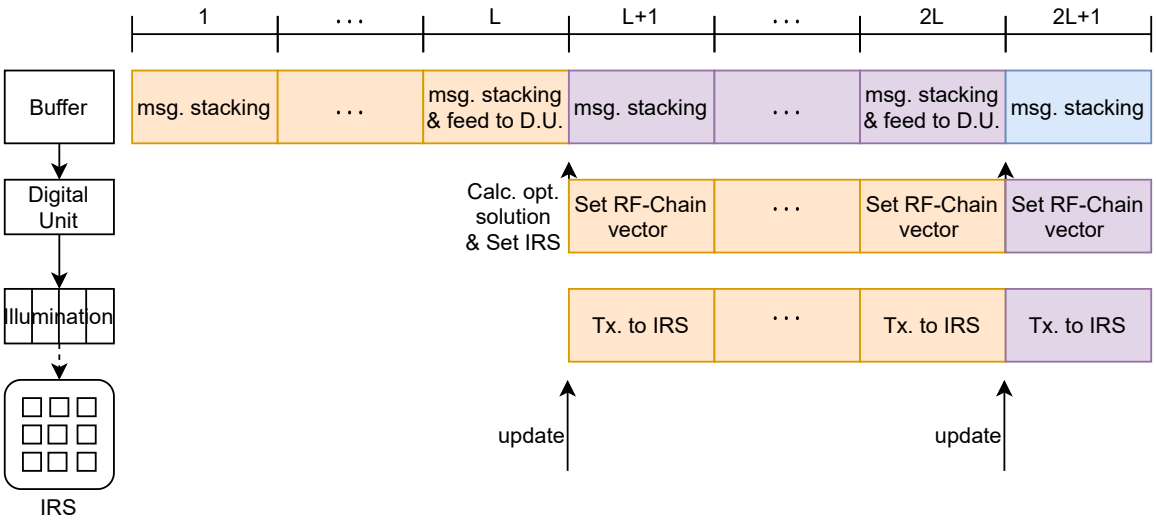


Fig. (3) Transmitter Time Table

The table entries with the same color indicates that they all correspond to the message symbols of the same block. For example, the RF chain output from time interval $L+1$ to $2L$ are computed based on the buffered message from time interval 1 to L. If the message vectors are generated at the rate R. It can then be observed that the RF chain output also has the same update rate. However, since the IRS is reconfigurated once per time block. Its update rate is only R/L.

B. Channel Model

According to Eq. (1), the wireless channel from the BS to the k-th UT can be modeled as,

$$y_k = \mathbf{h}_k^T \mathbf{a} + n_k. \tag{4}$$

The row vector $\mathbf{h}_k$ is the channel coefficient from the IRS to the k-th UT. If we take scattering into consideration, following [2], it can be further modeled as

$$\mathbf{h}_k^T = \frac{1}{\sqrt{P}} \sum_{p=1}^{P} h_p \mathbf{h}_t^T(\theta_p^t, \phi_p^t), \tag{5}$$

where $P \in \mathbb{N}$ is the number of effective channel paths while h_p is the coefficient that models the path loss and fading of the p-th path. The vector $\mathbf{h}_t(\theta_p^t, \phi_p^t) \in \mathbb{C}^{M\times 1}$ is the steering vector of the passive elements on the IRS. The directed angles $\theta_p^t \in [0, \pi]$ and $\phi_p^t \in [0, \pi]$ are the azimuthal

angle and polar angle of the departure beam. Fig. (4) describes this cordinates in more details. The points on the IRS are the reflection elements.

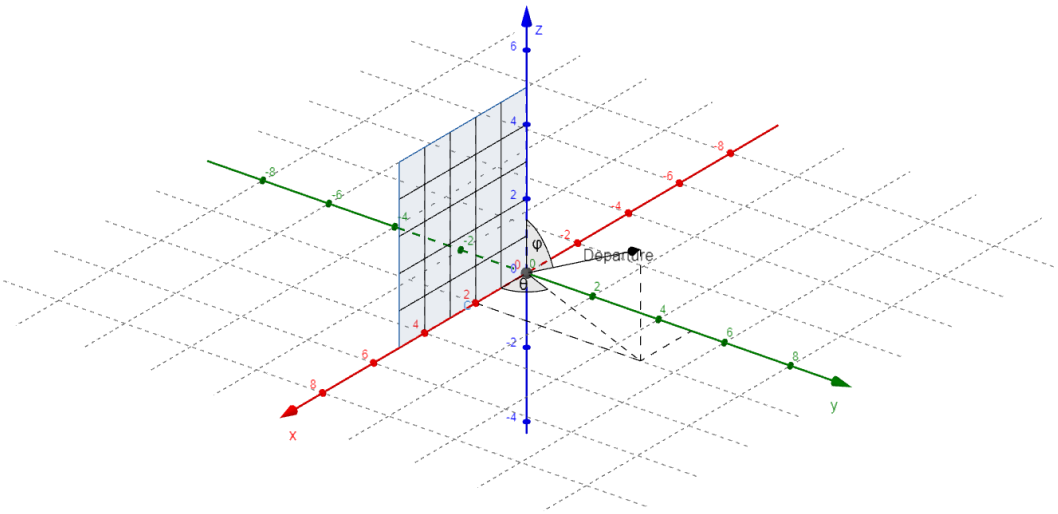


Fig. (4) Angle of departure

W.l.o.g. in the following context, we will place the IRS in the first quadrant of the X-Z plane. We will also assume that $\sqrt{M}$ is an even positive integer. Moreover, we assume that the distance between two neighboring IRS elements is constant and represented as d. We will index the IRS elements from right to left, bottom to top. Therefore, let $m \in \mathbb{N} \cap [1, M]$, and the m-th IRS element will have the coordinate

$$\begin{pmatrix} x = ((m-1) \mathrm{mod} \sqrt{M}) \cdot d, \\ y = 0, \\ z = \lfloor (m-1)/\sqrt{M} \rfloor \cdot d \end{pmatrix}. \tag{6}$$

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 half bracket $\lfloor \cdot \rfloor$ is the operator that evaluates to the largest interger that is smaller or equal to the number inside it.

If the phase of beam departing from the first IRS element is taken as the reference phase. Moreover, since the distance between the BS and the UT is often much larger than the wavelength, the departure beam from the any two IRS elements to the same scattering object can

be considered as parallel rays. The phase shifts of the steering vector depend only on the path length, which can be obtained from the following fig. (5).

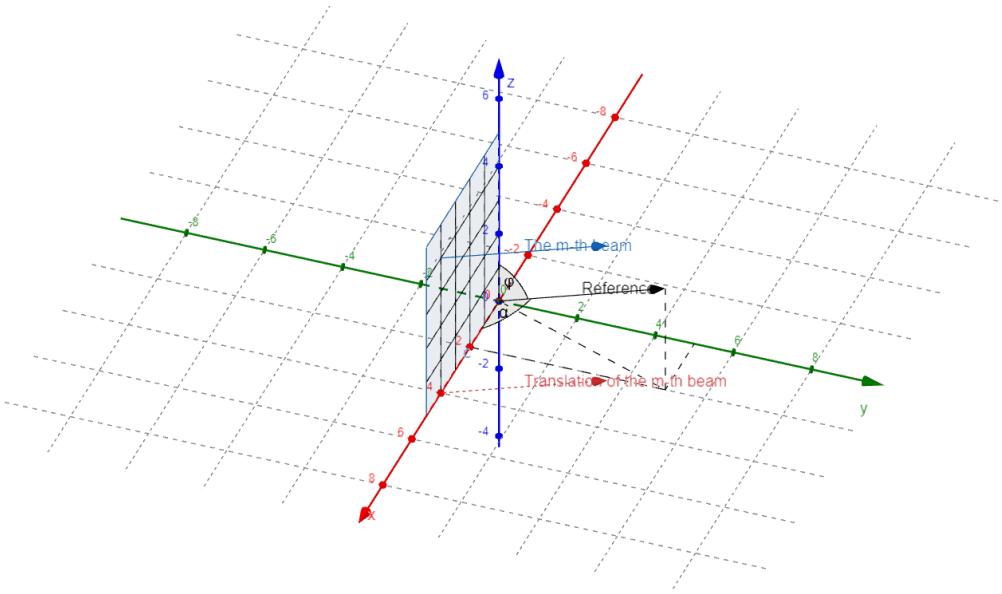


Fig. (5) Length of the m-th beams

In order to get the length different between the m-th beam and the reference beam, we first translate the m-th beam vertically down until the initial point is on the x-axis. The length different between the reference beam and the translated beam from IRS to the scattering object can be determined by the angle α, while the length difference between the translated beam and the beam of the m-th element depends on the angle ϕ. Thus, we can get the following equations

$$\begin{aligned} l_{ref} - l_{tr} &= d_x \cos(\alpha) = d_x \sin(\phi)\cos(\theta) \\ l_{tr} - l_m &= d_z \cos(\phi) \\ \Rightarrow \\ l_{ref} - l_m &= d_x \sin(\phi)\cos(\theta) + d_z \cos(\phi), \end{aligned} \tag{7}$$

where l_{ref}, l_{tr} and l_m are the beam length from the IRS to the scattering object of the reference beam, the translated beam and the m-th beam respectively. The value of d_x and d_z equal to the x and z coordinates of the m-th IRS element.

By combining the Eq. (5), (6) and (7), the steering vector $\mathbf{h}_t^T(\theta_p^t, \phi_p^t)$ can be written as

$$\mathbf{h}_t(\theta_p^t, \phi_p^t) = \left[e^{j\frac{2\pi d}{\lambda_w}[((m-1)\bmod\sqrt{M})\sin(\phi_p^t)\cos(\theta_p^t) + \lfloor (m-1)/\sqrt{M} \rfloor \cos(\phi_p^t)]} \right]_{m,1}. \tag{8}$$

Here, λ_w denotes the wavelength of the transmittion signal.

In Eq. (5), the path loss and the fading is included in the coefficient h_p. According to [2], h_p is modeled as $h_p = \sqrt{\bar{h}_k}\tilde{h}_p$, where $\bar{h}_k$ is the path loss and $\tilde{h}_p$ is the random fading which are given by

$$\begin{aligned} \bar{h}_k &= \left(\frac{\lambda_w}{4\pi l_{ref}}\right)^n \\ \tilde{h}_p &\sim \mathrm{CN}(0, \sigma_f^2). \end{aligned} \tag{9}$$

The variable $n \geq 2$ here is the path loss exponent.

C. Illumination Model

The matrix $\mathbf{T}$ describes the illumination channel from the active antennas to the IRS. In this article, the N active antennas are uniformly place on a circle parallel to the $x-z$ plane. The line determined by the center of the the active antenna arrays and the center of the IRS is also parallel to the y axis. The radius of the active antenna arrays is denoted as R_r, and the distance between one active antenna and the $x-z$ plane is denoted as R_d. This setup of the active antennas relative to the IRS is show as Fig. (6).

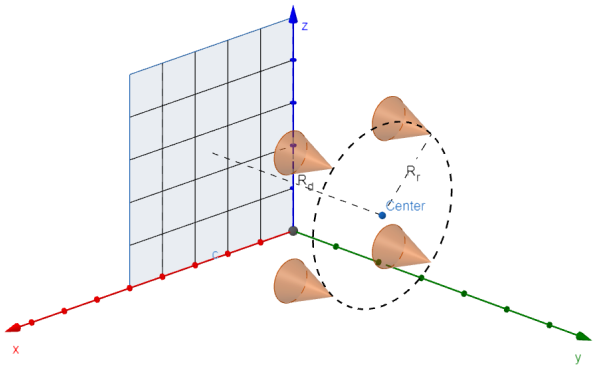


Fig. (6) Placement of the active antennas

The index of the active antennas are ordered according to the directed angle ranging in $[0, 2\pi]$ from the x-axis to the ray Center $\rightarrow$ Active-antenna. In Fig. (6), top left active antenna is labeled

as one, top right is labeled as two, bottom right is labeled as three and bottom left is labeled as four.

To study the antenna gain of a certain IRS element, we first create a spherical coordination system for each of the active antennas. The active antennas are placed at the origin of the coordination systems and the beam direction coincides with the z'-axis. The x'-axis is place on a plane parallel to the x-y plane and has a acute angle with the x-axis. This auxilary coordination system is illustrated in Fig. (7).

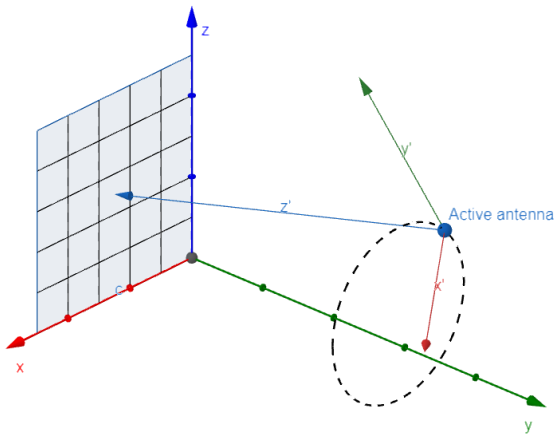


Fig. (7) Auxilary coordination system of an active antenna

Under this $x'y'z'$ coordination system, the position of the m-th IRS element is represented by a tuple $(r^p_{m,n}, \theta^p_{m,n}, \phi^p_{m,n})$, where $r^p_{m,n} \geq 0$ is the distance between the n-th active antenna. Similar to the previous definition, $\theta^p_{m,n}$ and $\phi^p_{m,n}$ are the azimuthal and polar angel of the m-th IRS element respectively. Fig. (8) illustrates an example coordinate of the m-th IRS element under the auxilary coordination system of the n-th active antenna.

The spherical coordinate of the active antenna relative to the IRS element can be obtained by similar manner. The xyz coordination system should first be translated to that IRS element. The relative coordinate of the active antenna can then be obtained with the same method, and can be characterized as tuple $(r^a_{m,n}, \theta^a_{m,n}, \phi^a_{m,n})$. Notice that $r^a_{m,n} = r^p_{m,n}$. Therefore, both of distance can be represented as $r_{m,n}$.

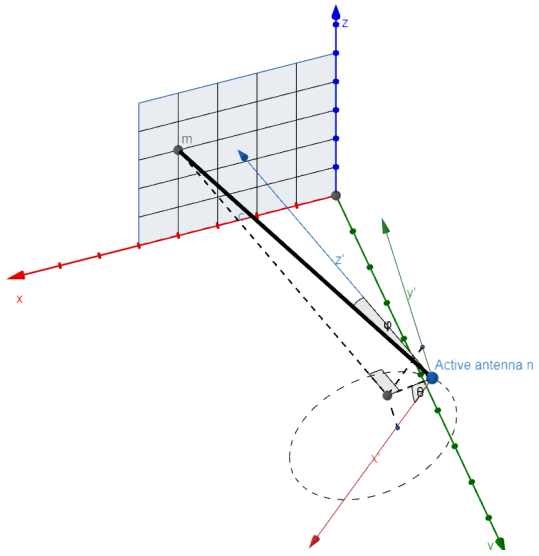


Fig. (8) The coordination of the m-th IRS element

Based on the above discussion, and following [2], the matrix $\mathbf{T}$ is given by

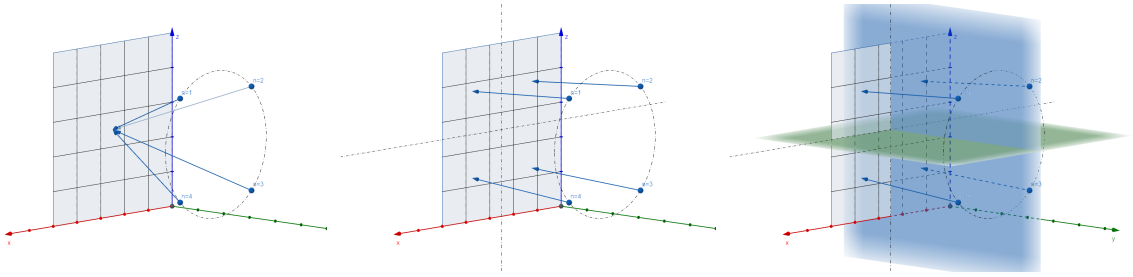
$$\mathbf{T} = \left[\frac{\lambda_w \sqrt{\rho G^a(\theta^p_{m,n},\phi^p_{m,n}) G^p(\theta^a_{m,n},\phi^a_{m,n})}}{4\pi r_{m,n}} e^{-j\frac{2\pi r_{m,n}}{\lambda_w}} \right]_{m,n}, \tag{10}$$

where $\rho \in [0,1]$ denotes the power efficiency of the IRS. The function $G^a(\theta^p_{m,n},\phi^p_{m,n})$ and $G^p(\theta^a_{m,n},\phi^a_{m,n})$ model the antenna gain of the active antenna and the IRS element respectively.

We assume that the antenna gains of all the active antennas have the same pattern. In addition to that, We also assume that the antenna gains of the different IRS elements also have the same pattern. By analysing Eq. (10), it can be observed that if the antenna gains and the wave length are fixed, matrix $\mathbf{T}$ is fully determined by the relative position and the orientation of the active antennas. To study the different designs of $\mathbf{T}$, we will investigate three illumination strategies mentioned in [2].

Full illumination (FI) is an illumination strategy in which each active antenna illuminates the entire IRS. In the following text, it is assumed that the beams from the active antennas are all pointing at the center of the IRS. In FI, each IRS element will shift all the incoming beams with a same phase. However, in certain senarios, it may be desired to apply the phase shifts to the beams disjointly. Therefore, we have the following two illumination strategies.

Partial illumination (PI) is an illumination strategy, where we designate N sections on the IRS, such that each section is responsible for one active antenna. Moreover, the active antennas illuminate the centers of their designated sections. However, this strategy still cannot eliminate



(a) Full illumination (b) Partial illumination (c) Seperate illumination

Fig. (9) Illumination strategies

the interference if the antenna pattern is too wide or the distance between the active antennas and the IRS is too large.

Seperate illumination (SI) is an illumination strategy, where we physically prevent the active antennas from illuminating the sections that are not designated to them. The position and orientation design of SI is the same as PI. However, some electromagnetic-absorbing barriers are placed at the border of each section.

The above mentioned illumination strategies are illustrated in Fig. (9). The blue arrows indicate the orientations of the active antennas.

III. Problem Formalization

A. *Performance Measure*

In order to design the precoding algorithm, a performance measure is required. The squared frobenius norm is adopted here to measure the metric between the decoded signals and the intended messages. It is defined as

$$\mathrm{d}(\mathbf{S}, \tilde{\mathbf{Y}}) = ||\tilde{\mathbf{Y}} - \mathbf{S}||_{\mathrm{fro}}^2 = \mathrm{tr}\{(\tilde{\mathbf{Y}} - \mathbf{S})^H(\tilde{\mathbf{Y}} - \mathbf{S})\}, \tag{11}$$

where $\tilde{\mathbf{Y}} \in \mathbb{C}^{K\times L}$ is the decoded signal of the original received signal $\mathbf{Y}$.

This metric quantifies the total distortion of the received signals. The ideal value of it is zero. Therefore, we want to design an algorithm that minimize this metric.

B. *Near-Far Effect and Decoder*

The UTs often have different distance to the BS. As a result, the path loss factor $\bar{h}_k$ of different UTs are different and its value is very small. However, it is often the case that this factor can

be acquired by both BS and the corresponding UT. Thus, instead of transmitting signals with enormous power to ensure that received signal is the same as the intended signal, we can design a disjoint decoder and let the UTs to compensate for this. Although the statistics of $\mathbf{X}$, $\mathbf{D}$ and $\mathbf{S}$ are also obtainable at the UTs, these matrices are dependent on each other and the CSI. Therefore, it is hard to calculate the optimal decoder $\mathbf{F}_D^{opt}$. However, since these infomation are also known by the transmitter, we can consider the decoder matrix $\mathbf{F}_D$ as a known matrix while designing the precoder. We can also observe that the designing of the decoder matrix has little effect on the process of the designing of a precoder. W.l.o.g. define $\mathbf{P}_L^{1/2} \in \mathbb{C}^{K\times K}$ to be a diagonal path loss matrix

$$\mathbf{P}_L^{1/2} = \begin{bmatrix} \sqrt{\bar{h}_1} & & \\ & \ddots & \\ & & \sqrt{\bar{h}_K} \end{bmatrix}. \tag{12}$$

The decoder can then be designed as

$$\tilde{\mathbf{Y}} = \mathbf{P}_L^{-1/2}\mathbf{Y}. \tag{13}$$

C. Statistics of the Performance Measure

Most of time, the channel estimation is imperfect. The relation between real channel and estimated channel can be formulized in the following equation,

$$\mathbf{H} = \hat{\mathbf{H}} + \mathbf{e}, \tag{14}$$

where $\mathbf{H} \in \mathbb{C}^{K\times M}$ stands for the real channel, $\hat{\mathbf{H}} \in \mathbb{C}^{K\times M}$ is the estimated channel and their difference is modeled as a K by M complex random matrix $\mathbf{e}$. We assume here that each entry of the error matrix are i.i.d. and follows the two statistics in Eq. (15), which are known by the transmitter.

$$\begin{aligned} \mathcal{E}\{e_{k,l}\} &= 0; \\ \mathcal{E}\{e_{k,l}^* e_{k,l}\} &= \sigma_e^2. \end{aligned} \tag{15}$$

If we take the estimation error into consideration, the received signal and the decoded signal can then be written as

$$\begin{aligned} \mathbf{Y} =& (\hat{\mathbf{H}} + \mathbf{e})\mathbf{DTX} + \mathbf{N} \\ \tilde{\mathbf{Y}} =& \mathbf{P}_L^{-1/2}[(\hat{\mathbf{H}} + \mathbf{e})\mathbf{DTX} + \mathbf{N}]. \end{aligned} \tag{16}$$

To make the further discussion breif, let's define the matrices $\tilde{\mathbf{H}}$, $\tilde{\hat{\mathbf{H}}}$, $\tilde{\mathbf{e}}$ and $\tilde{\mathbf{N}}$ which are the corresponding matrices normalized to $\mathbf{P}_L^{-1/2}$as Eq. (17).

$$\begin{aligned}\mathbf{H} =&\mathbf{P}_L^{1/2}\tilde{\mathbf{H}}\\ \hat{\mathbf{H}} =&\mathbf{P}_L^{1/2}\tilde{\hat{\mathbf{H}}}\\ \mathbf{e} =&\mathbf{P}_L^{1/2}\tilde{\mathbf{e}}\\ \mathbf{N} =&\mathbf{P}_L^{1/2}\tilde{\mathbf{N}}\end{aligned} \tag{17}$$

By substituting Eq. (11) with Eq. (16), we can observe that the metric equation Eq. (11) contains a noise term and a estimation error term which are both random. Therefore, the precoding algorithm can only be designed base on the statistics of Eq. (11). Assume that the noise entries are independed with the estimation errors, we then have

$$\mathcal{E}_{\mathbf{N},\mathbf{e}}\{\mathrm{d}(\mathbf{S},\tilde{\mathbf{Y}})\} = \mathcal{E}_{\mathbf{e}}\cdot\mathcal{E}_{\mathbf{N}}\{\mathrm{d}(\mathbf{S},\tilde{\mathbf{Y}})\} \tag{18}$$

If the noise entry in the matrix $\mathbf{N}$ is i.i.d. and zero mean with variance σ_n^2. The mean value of the distance with respect to noise is given by

$$\begin{aligned}\mathcal{E}_{\mathbf{N}}\{\mathrm{d}(\mathbf{S},\tilde{\mathbf{Y}})\} =&\mathrm{tr}\{(\tilde{\mathbf{H}}\mathbf{DTX}-\mathbf{S})^H(\tilde{\mathbf{H}}\mathbf{DTX}-\mathbf{S})\}+\mathcal{E}_{\mathbf{N}}\{\mathrm{tr}\{\tilde{\mathbf{N}}^H\tilde{\mathbf{N}}\}\}\\ =&\mathrm{tr}\{(\tilde{\mathbf{H}}\mathbf{DTX}-\mathbf{S})^H(\tilde{\mathbf{H}}\mathbf{DTX}-\mathbf{S})\}+L\mathrm{tr}\{\mathbf{P}_L^{-1}\}\sigma_n^2.\end{aligned} \tag{19}$$

By performing the similar operation one more time, the expectation of Eq. (19) with respect to estimation error $\mathbf{e}$ is given by

$$\begin{aligned}&\mathcal{E}_{\mathbf{e}}\left\{\mathcal{E}_{\mathbf{N}}\{\mathrm{d}(\mathbf{S},\tilde{\mathbf{Y}})\}\right\}\\ =&\mathcal{E}_{\mathbf{e}}\{\mathrm{tr}\{((\tilde{\hat{\mathbf{H}}}+\tilde{\mathbf{e}})\mathbf{DTX}-\mathbf{S})^H((\tilde{\hat{\mathbf{H}}}+\tilde{\mathbf{e}})\mathbf{DTX}-\mathbf{S})\}\}+\mathrm{tr}\{\mathbf{P}_L^{-1}\mathbf{\Sigma}_N\}\\ =&\left|\left|\tilde{\hat{\mathbf{H}}}\mathbf{DTX}-\mathbf{S}\right|\right|_{\mathrm{fro}}^2+\mathrm{tr}\{\mathbf{X}^H\mathbf{T}^H\mathbf{D}^H\mathbf{\Sigma}_{\tilde{\mathbf{e}}}\mathbf{DTX}\}+L\mathrm{tr}\{\mathbf{P}_L^{-1}\}\sigma_n^2,\end{aligned} \tag{20}$$

where

$$\mathbf{\Sigma}_{\tilde{\mathbf{e}}} = \begin{bmatrix}\mathrm{tr}\{\mathbf{P}_L^{-1}\}\sigma_e^2 & & \\ & \ddots & \\ & & \mathrm{tr}\{\mathbf{P}_L^{-1}\}\sigma_e^2\end{bmatrix} \tag{21}$$

The formular given in Eq. (20) is the expected distance between $\mathbf{S}$ and $\tilde{\mathbf{Y}}$, i.e. $\mathcal{E}_{\mathbf{N},\mathbf{e}}\{\mathrm{d}(\mathbf{S},\tilde{\mathbf{Y}})\}$. From now on, instead of developing an algorithm that minimize Eq. (11) directly, we try to develop an algorithm that minimize Eq. (20) regarding to $\mathbf{D}$ and $\mathbf{X}$.

D. Problem Formalization as a GLSE Precoder

Following [3], 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

$$\begin{aligned} \underset{\mathbf{D},\mathbf{X}}{\text{minimize}} \quad & \left|\left|\tilde{\hat{\mathbf{H}}}\mathbf{DTX}-\mathbf{S}\right|\right|_{\text{fro}}^2 + \text{tr}\{\mathbf{X}^H\mathbf{T}^H\mathbf{D}^H\boldsymbol{\Sigma}_{\tilde{\mathbf{e}}}\mathbf{DTX}\} + \lambda\text{tr}\{\mathbf{X}\mathbf{X}^H\} \\ \text{subject to} \quad & \max_{i,j}|x_{i,j}|^2 < P_{\text{peak}}, \end{aligned} \tag{22}$$

where $P_{\text{peak}} \geq 0$ 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, while with smaller λ the system will lay more emphasis on minimizing 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, we define the objective function in Eq. (22) as

$$\text{obj}(\mathbf{D},\mathbf{X},\mathbf{S}) = \left|\left|\tilde{\hat{\mathbf{H}}}\mathbf{DTX}-\mathbf{S}\right|\right|_{\text{fro}}^2 + \text{tr}\{\mathbf{X}^H\mathbf{T}^H\mathbf{D}^H\boldsymbol{\Sigma}_{\tilde{\mathbf{e}}}\mathbf{DTX}\} + \lambda\text{tr}\{\mathbf{X}\mathbf{X}^H\} \tag{23}$$

If we neglect the power constraints by setting λ to 0, and letting P_{peak} tend to infinity, the optimization problem stated in Eq. (22) is equavalent to the direct optimization of Eq. (20), because the term $L\text{tr}\{\mathbf{P}_L^{-1}\}\sigma_n^2$ is a constant term that does not depend either on $\mathbf{D}$ nor on $\mathbf{X}$.

IV. IRS-Aided Hybrid Precoding

A. Joint Optimization

The joint optimization of the problem stated in (22) is non-trivial since it is generally non-convex. Alternating optimization methods is adopted to obtain a suboptimal solution iteratively. In each iteration, we first compute the optimal RF chain output while treating the IRS configuration as a fixed matrix, then we calculate the the suitable (may not be optimal) phase shift configuration for the IRS while treating $\mathbf{X}$ as a fixed matrix. The pseudo-code Algorithm (1) dipicts the general procedure.

A termiation condition for the loop should be chosen according to system requirements. For example, let the loop be executed until the difference between the cost function values of two consequent loops is smaller than a certain threshold. The step 3 in the algorithm (optimize with respect to $\mathbf{X}$) is quadratic programming, which is categorized as convex optimization problem.

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

1: $(\mathbf{D}, \mathbf{X}) \leftarrow$ initial values
2: **while** terminate condition not met **do**
3: $\quad \mathbf{X} \leftarrow \arg\min_{\mathbf{X}} \ \mathrm{obj}(\mathbf{D}, \mathbf{X})$
4: $\quad \mathbf{D} \leftarrow \arg\min_{\mathbf{D}} \ \mathrm{obj}(\mathbf{D}, \mathbf{X})$
5: **end while**

Therefore, an optimal $\mathbf{X}$ can always be obtained if $\mathbf{D}$ is given. However, the domain of $\mathbf{D}$ is not a convex set, which implies that the step 4 in the algorithm is also a non-convex problem. Thus, the optimal value in step 4 can genrally not be found. This leads to a problem that Algorithm (1) may not converge, because step 4 might find a suboptimal minimal for $\mathbf{D}$ that may lead to a larger cost value than the previous one.

B. Optimization of the Digital Unit

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

$$\begin{aligned} \underset{\mathbf{X}}{\text{minimize}} \quad & \left|\left|\tilde{\hat{\mathbf{H}}}\mathbf{D}\mathbf{T}\mathbf{X} - \mathbf{S}\right|\right|_{\text{fro}}^2 + \text{tr}\{\mathbf{X}^H\mathbf{T}^H\mathbf{D}^H\mathbf{\Sigma}_{\tilde{\mathbf{e}}}\mathbf{D}\mathbf{T}\mathbf{X}\} + \lambda\text{tr}\{\mathbf{X}\mathbf{X}^H\} \\ \text{subject to} \quad & \max_{i,j} |x_{i,j}|^2 < P_{\text{peak}} \end{aligned} \tag{24}$$

belongs to the class of convex problem. Moreover, it is worth noticing that the columns (time intervals) of matrix $\mathbf{X}$ can be decoupled as is shown in Eq. (25).

$$\begin{aligned} &\text{tr}\{(\tilde{\hat{\mathbf{H}}}\mathbf{D}\mathbf{T}\mathbf{X} - \mathbf{S})^H(\tilde{\hat{\mathbf{H}}}\mathbf{D}\mathbf{T}\mathbf{X} - \mathbf{S})\} + \text{tr}\{\mathbf{X}^H\mathbf{T}^H\mathbf{D}^H\mathbf{\Sigma}_{\tilde{\mathbf{e}}}\mathbf{D}\mathbf{T}\mathbf{X}\} + \lambda\text{tr}\{\mathbf{X}^H\mathbf{X}\} \\ &= \sum_{i=1}^{L}(\tilde{\hat{\mathbf{H}}}\mathbf{D}\mathbf{T}\boldsymbol{x_i} - \boldsymbol{s_i})^H(\tilde{\hat{\mathbf{H}}}\mathbf{D}\mathbf{T}\boldsymbol{x_i} - \boldsymbol{s_i}) + \mathbf{x}_i^H\mathbf{T}^H\mathbf{D}^H\mathbf{\Sigma}_{\tilde{\mathbf{e}}}\mathbf{D}\mathbf{T}\mathbf{x}_i + \lambda\boldsymbol{x_i}^H\boldsymbol{x_i}. \end{aligned} \tag{25}$$

Therefore, this optimization problem can either be treated as one matrix convex optimization or L independent vector convex optimization problems, which means the block-wise precoding has no influence over the design of the digital unit if $\mathbf{D}$ is fixed.

C. Optimization of the IRS Configuration

According to algorithm (1), the RF chain output matrix should be fixed when optimizing the IRS configuration $\mathbf{D}$. Thus $\mathbf{X}$ is independent to $\mathbf{D}$. The optimization problem is then formalized

as

$$\underset{\mathbf{D}}{\text{minimize}} \quad \left|\left|\tilde{\hat{\mathbf{H}}}\mathbf{DTX}-\mathbf{S}\right|\right|_{\text{fro}}^{2}+\text{tr}\{\mathbf{X}^{H}\mathbf{T}^{H}\mathbf{D}^{H}\mathbf{\Sigma}_{\tilde{\mathbf{e}}}\mathbf{DTX}\} \tag{26}$$

This problem is generally non-convex, since $\text{domain}(\mathbf{D})$ is not a convex set. Two methods are used here to obtain a suboptimal point, which are gradient descent and MM algorithm. For the further discussion, we denote the objective function in Eq. (26) by

$$\text{obj}_{\mathbf{D}}(\mathbf{D})=\left|\left|\tilde{\hat{\mathbf{H}}}\mathbf{DTX}-\mathbf{S}\right|\right|_{\text{fro}}^{2}+\text{tr}\{\mathbf{X}^{H}\mathbf{T}^{H}\mathbf{D}^{H}\mathbf{\Sigma}_{\tilde{\mathbf{e}}}\mathbf{DTX}\}. \tag{27}$$

1) Gradient descent method: Gradient descend is a general method of finding the local minimal 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 which is the steepest slope iteratively from a random initial point.

The domain of objective function of the problem in Eq. (26) is the set of complex diagonal matrices whose non-zero entries are on the unit circle. In order to simplify the domain, we define the phase vector

$$\boldsymbol{\beta}=\arg\cdot\text{diag}_{\text{vector}}(\mathbf{D}) \tag{28}$$

To keep the notation simple, we define

$$\begin{aligned}\mathbf{U}&=\mathbf{DTX};\\ \mathbf{V}&=\mathbf{S}^{H}\tilde{\hat{\mathbf{H}}}\\ \mathbf{Q}&=\tilde{\hat{\mathbf{H}}}^{H}\tilde{\hat{\mathbf{H}}}+\mathbf{\Sigma}_{\tilde{\mathbf{e}}}.\end{aligned} \tag{29}$$

Under this convention, the objective function can be expanded as

$$\text{obj}_{\mathbf{D}}(\mathbf{D})=tr\{\mathbf{U}^{H}\mathbf{QU}-(\mathbf{U}^{H}\mathbf{V}^{H}+\mathbf{VU})+\mathbf{S}^{H}\mathbf{S}\}. \tag{30}$$

The partial derivertive is a linear transformation, and the derivertive of the additive terms in Eq. (30) can be obtained as

$$\begin{aligned}\frac{\partial tr\{\mathbf{U}^{H}\mathbf{QU}\}}{\partial\boldsymbol{\beta}}&=-2\text{diag}_{\text{vector}}(\Im(\mathbf{UU^{H}Q}))\\ \frac{\partial tr\{\mathbf{U}^{H}\mathbf{V}^{H}+\mathbf{VU}\}}{\partial\boldsymbol{\beta}}&=-2\text{diag}_{\text{vector}}(\Im(\mathbf{UV})).\end{aligned} \tag{31}$$

From Eq. (30) and Eq. (31), the partial derivertive of the objective function can be derived as

$$\frac{\partial\text{obj}_{\mathbf{D}}(\mathbf{D})}{\partial\boldsymbol{\beta}}=-2\text{diag}_{\text{vector}}(\Im(\mathbf{UU^{H}Q}-\mathbf{UV})) \tag{32}$$

The auxilary vector can be transformed back to the phase shift matrix by Eq. (33).

$$\mathbf{D} = \mathrm{diag}_{\mathrm{matrix}}\left(\exp(j\boldsymbol{\beta})\right). \tag{33}$$

Based on the above discussion, if combined with the designing of the digital unit, the psuedo-code of the precoding algorithm with gradient descent is shown as Algorithm (2). In order to

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

1: $(\mathbf{D}, \mathbf{X}) \leftarrow$ initial values
2: **while** terminate condition not met **do**
3: $\quad \mathbf{X} \leftarrow \mathrm{convex}_{\mathrm{opt}}\{\left|\left|\tilde{\hat{\mathbf{H}}}\mathbf{DTX} - \mathbf{S}\right|\right|_{\mathrm{fro}}^2 + \mathrm{tr}\{\mathbf{X}^H\mathbf{T}^H\mathbf{D}^H\boldsymbol{\Sigma}_{\tilde{\mathbf{e}}}\mathbf{DTX}\} + \lambda\mathrm{tr}\{\mathbf{XX}^H\}\}$
4: $\quad$ **while** terminate condition not met **do**
5: $\quad\quad \boldsymbol{\beta} \leftarrow \arg\cdot\mathrm{diag}_{\mathrm{vector}}(\mathbf{D})$
6: $\quad\quad \boldsymbol{\beta}_{\mathrm{grad}} \leftarrow -2\mathrm{diag}_{\mathrm{vector}}(\Im(\mathbf{UU^HQ} - \mathbf{UV}))$
7: $\quad\quad \alpha \leftarrow \mathrm{getStepSize}$
8: $\quad\quad \boldsymbol{\beta}' = \boldsymbol{\beta} - \alpha\boldsymbol{\beta}_{\mathrm{grad}}$
9: $\quad\quad \mathbf{D} \leftarrow \mathrm{diag}_{\mathrm{matrix}}\left(\exp(j\boldsymbol{\beta}')\right)$
10: $\quad$ **end while**
11: **end while**

avoid the situation where the algorithm diverge, one can design the step size α as follow. At the beginning of each gradient descent loop (step 4 - 10 in Algorithm (2)) α is initiallized to a relative large value. During every loop the processor keeps the value of $\mathbf{D}$ in the memory. If the value of the objective function does not decrease with the new $\mathbf{D}$. The processor will then load the old value of $\mathbf{D}$ 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 $\mathbf{D}$ will always be assigned to a value that makes the objective function decrease. Moreover, if the threshold for α is small enough. $\mathbf{D}$ will always end up at the close proximity of a critical point after the termination of one entire gradient descent loop.

2) MM algorithm: MM algorithm is another general optimization method that can be used to optimize the IRS configuration. In MM algorithm, a less complicated surrogate function of each point of the objective function needs to be constructed, which touches the objective function at that point but is always larger or equal to every other point on it within the feasible domain.

The MM algorithm is like a variation of gradient descent. Both of the methods includes iterative procedures. However, when using MM algorithm we don't need to state the step size explicitly. The procedure of the MM algorithm is as follow. At the i-th MM iteration for optimizing the IRS, a surrogate function of a random initial point $\mathbf{D}_i$ is constructed. We then find the optimal point $\mathbf{D}_{i+1}$ that minimize the surrogate function. After that, we continue to the next MM iteration.

Following the variables defined in Eq. (29), we have

$$\mathbf{U} = \mathbf{DTX};$$

$$\mathbf{V} = \mathbf{S}^H\tilde{\hat{\mathbf{H}}}$$

$$\mathbf{Q} = \tilde{\hat{\mathbf{H}}}^H\tilde{\hat{\mathbf{H}}} + \mathbf{\Sigma}_{\tilde{\mathbf{e}}}$$

The complicated term in the objective function Eq. (27) is the quadratic term. By substituting the variables according to Eq. (29), the quadratic term can be written as

$$\mathbf{X}^H\mathbf{T}^H\mathbf{D}^H\left(\tilde{\hat{\mathbf{H}}}^H\tilde{\hat{\mathbf{H}}} + \mathbf{\Sigma}_{\tilde{\mathbf{e}}}\right)\mathbf{DTX} = \mathbf{U}^H\mathbf{QU} \tag{34}$$

Since the surrogate functions are additive if they touch the same point of the objective function, the surrogate function of the objective function at any point $\mathbf{D}_0$ can be obtained by constructing a surrogate function for the quadratic term at point $\mathbf{D}_0$ while letting the first order term be its own surrogate function.

To construct the surrogate function for the quadratic term, define an auxilary matrix $\mathbf{P}$, such that $\mathbf{P} - \mathbf{Q}$ is positive definite. For the tangential point $\mathbf{D}_0$, also define $\mathbf{U}_0 = \mathbf{D}_0\mathbf{TX}$. The surrogate for the quadratic term can be obtained by the following inequality.

$$\begin{aligned}
&\mathrm{tr}\left\{\mathbf{U}^H\mathbf{QU}\right\}\\
=&\mathrm{tr}\left\{\mathbf{U}_0^H\mathbf{QU}_0 + (\mathbf{U}-\mathbf{U}_0)^H\mathbf{QU}_0 + \mathbf{U}_0^H\mathbf{Q}(\mathbf{U}-\mathbf{U}_0) + (\mathbf{U}-\mathbf{U}_0)^H\mathbf{Q}(\mathbf{U}-\mathbf{U}_0)\right\}\\
\leq&\mathrm{tr}\left\{\mathbf{U}_0^H\mathbf{QU}_0 + (\mathbf{U}-\mathbf{U}_0)^H\mathbf{QU}_0 + \mathbf{U}_0^H\mathbf{Q}(\mathbf{U}-\mathbf{U}_0) + (\mathbf{U}-\mathbf{U}_0)^H\mathbf{P}(\mathbf{U}-\mathbf{U}_0)\right\}\\
=&\mathrm{tr}\left\{\mathbf{U}^H\mathbf{PU} + \mathbf{U}^H(\mathbf{Q}-\mathbf{P})\mathbf{U}_0 + \mathbf{U}_0^H(\mathbf{Q}-\mathbf{P})\mathbf{Y} + \mathbf{U}_0^H(\mathbf{Q}-\mathbf{P})\mathbf{U}_0\right\}
\end{aligned} \tag{35}$$

The newly introduce quadratic term $\mathbf{U}^H\mathbf{PU}$ will be degenerated 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 removing the terms that are independent of variable $\mathbf{D}$ from the surrogate function of the quadratic term and combining the first order term of the objective function Eq. (27). The equivalent surrogate function (not the true surrogate since the constant terms are removed) for the objective function at any point $\mathbf{D}_0$ reads

$$\overline{\mathrm{obj}_{\mathbf{D}_0}(\mathbf{D})} = \mathrm{tr}\left\{\mathbf{U}^H(\mathbf{Q}-\mathbf{P})\mathbf{U}_0 + \mathbf{U}_0^H(\mathbf{Q}-\mathbf{P})\mathbf{U} - \mathbf{VU} - \mathbf{U}^H\mathbf{V}^H\right\} \tag{36}$$

For simplicity, define

$$\mathbf{G}_0 = \mathbf{TXU}_0^H(\mathbf{Q}-\mathbf{P}) - \mathbf{TXV}. \tag{37}$$

By substituting Eq. (37) into Eq. (36), the latter becomes

$$\overline{\mathrm{obj}_{\mathbf{D}_0}(\mathbf{D})} = \mathrm{tr}\{\mathbf{G}_0\mathbf{D} + \mathbf{D}^H\mathbf{G}_0^H\} \tag{38}$$

According to previous discuss about the MM algorithm, the tangential point of the next MM iteration, namely $\mathbf{D}_1$ is obtained by minimizing the equivalent surrogate function described by Eq. (38). By introducing two auxilary constant terms $\mathbf{tr}\{\mathbf{D}^H\mathbf{D}\}$ and $tr\{\mathbf{G}_0\mathbf{G}_0^H\}$, we have

$$\begin{aligned}
\mathbf{D}_1 &= \arg\min_{\mathbf{D}}(\mathrm{tr}\{\mathbf{G}_0\mathbf{D} + \mathbf{D}^H\mathbf{G}_0^H\}) \\
\mathbf{D}_1 &= \arg\min_{\mathbf{D}}(\mathrm{tr}\{\mathbf{G}_0\mathbf{D} + \mathbf{D}^H\mathbf{G}_0^H + \mathbf{D}^H\mathbf{D} + \mathbf{G}_0\mathbf{G}_0^H\}) \\
\mathbf{D}_1 &= \arg\min_{\mathbf{D}}(\mathrm{tr}\{(\mathbf{D} - (-\mathbf{G}_0)^H)^H(\mathbf{D} - (-\mathbf{G}_0)^H)\}) \\
\mathbf{D}_1 &= \arg\min_{\mathbf{D}}(||\mathbf{D} - (-\mathbf{G}_0)||_{fro}^2) \\
\mathbf{D}_1 &= \exp((\arg(-\mathbf{G}_0))
\end{aligned} \tag{39}$$

The complete psuedo-code for the complete precoding with MM algorithm is illustrated as Algorithm (3) The convergence of Algorithm (3) is also guarenteed. Because of the definition of surrogate function, the value of the objective function in the MM iterations (step 4-13 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 befor the IRS optimization. Thus, Algorithm (3) also converge.

V. Numerical Simulations

In this section, we are going to study the performance of the IRS-aided precoding procedure discussed above.

A. Performance Measure

It is proven in [4] that by using Jensen's inequality the average rate per user is bounded by

$$\bar{R} \geq \frac{LT_s}{\tau + LT_s}\log\left(1 + \frac{\sigma_s^2}{\sigma_{\tilde{n}_{\max}}^2 + P_{\text{Interference}}}\right) \tag{40}$$

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

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

where T_s is symbol time, τ is the time for 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 performance measure is the peak to average power ratio (PAPR). In our simulation the chain-wise PAPR of every RF chain is calculated by dividing its peak power by the average power of all its transmitted signals. The average PAPR is calculated by average the chain-wise PAPR over all the RF chains.

B. Simulation Setup

In order to generate a channel matrix with near-far effect, 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 pathes. The intended messages for them has a power of $\sigma_s^2 = 1$. The angles from the BS to the scatterings are constant. The fading parameters is generated once every block. To make the further discussion easier, we assume that the distance between the UTs and the BS are deterministic and known by all the UTs and the BS, and the matrix $\tilde{\mathbf{e}}$ which stands for the normalized esimation error

l_{ref}	λ_w	σ_f^2	θ_p^t, ϕ_p^t	d	n	P	K
$[50, 150]$	0.01	1	$[0, \pi]$	$\lambda_w/2$	3	8	8

TABLE (I) Parameters for channel $\mathbf{H}$

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

is comprised of i.i.d. entries. The default setting for the channel is concluded as the following TABLE (I).

The illumination channel is generated based on the antenna pattern of both active antennas and IRS elements. We assume here that the IRS element reflect all the signals in front of it, therefore we have

$$G^p(\theta, \phi) = \begin{cases} 2 & \theta, \phi \in [0, \pi] \\ 0 & \text{otherwise} \end{cases}. \tag{41}$$

The active antenna gain following [2] is assumed to be

$$G^a(\theta, \phi) = \begin{cases} 2(1+\kappa)\cos^{\kappa}(\phi) & \phi \in [0, \pi/2] \\ 0 & \text{otherwise} \end{cases}. \tag{42}$$

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

Besides the antenna pattern, the relative position and the orientation of the active antennas also matter. These parameters discussed above are concluded in TABLE (II) [2]. We also set the maximal normalized noise to be $\sigma^2_{\tilde{n}_{\max}} = 0.1$

C. Bechmark Method

We use a fully-connected analog beamforming network (FC-ABFN) as a benchmark structure [5]. The structure before the RF chains are similar to the structure of IRS aided HAD precoder. In the benchmark structure, each RF chain is connected to a seperate group of phase shifters through a power divider. Then each shifted signal is combined with the corresponding signal

from other groups. More specifically, for a fully connected precoder with N RF chains and M analog units the analog beamforming network is characterized as

$$\mathbf{F}_{\mathrm{RF}} = \mathbf{F}_{\mathrm{PC}}\mathbf{F}_{\mathrm{PS}}\mathbf{F}_{\mathrm{PD}} \tag{43}$$

where $\mathbf{F}_{\mathrm{PC}} \in \mathbb{C}^{MN\times N}$ models the power combiner that connects the beamformer with the passive antennas, $\mathbf{F}_{\mathrm{PS}} \in \mathbb{C}^{MN\times MN}$ is the diagonal phase shift matrix of the beamformer, and $\mathbf{F}_{\mathrm{PD}} \in \mathbb{C}^{M\times MN}$ models the the power divider that couples the RF chain with the beamformer. The power divider is realized by Wilkinson power dividers [6]. Therefore, it can be modeled as a matrix with block diagonal structure [5]

$$\mathbf{F}_{\mathrm{PD}} = \sqrt{\frac{1}{L_{\mathrm{D}}M}}\begin{bmatrix} \mathbf{1}_M & \dots & \mathbf{0}_M \\ \vdots & \ddots & \vdots \\ \mathbf{0}_M & \dots & \mathbf{1}_M \end{bmatrix}, \tag{44}$$

where L_D represents the power loss in the power divider [6], the vectors $\mathbf{1}_M \in \mathbb{N}^M$ and $\mathbf{0}_M \in \mathbb{N}^M$ represents all one vector and all zero vector of size M. The phase shifts matrix is defined similar to the IRS matrix which is a diagonal matrix of entries with absolute value one. The matrix for the power combiner is represented as

$$\mathbf{F}_{\mathrm{PC}} = \sqrt{\frac{1}{L_C N}}\begin{bmatrix}\mathrm{diag}(\mathbf{1}_M) & \dots & \mathrm{diag}(\mathbf{1}_M)\end{bmatrix}, \tag{45}$$

where L_C is the power loss introduced by the power combiner [5]. The channel can then be represented as

$$\mathbf{Y} = \mathbf{H}\mathbf{F}_{\mathrm{PC}}\mathbf{F}_{\mathrm{PS}}\mathbf{F}_{\mathrm{PD}}\mathbf{X} + \mathbf{N}. \tag{46}$$

By comparing Eq. (46) with Eq. (2), we set L_D and L_C such that the total transmitted power of the passive antenna of the benchmark structure equals to the full illumination strategy.

D. Rate and PAPR

By using the above mentioned setup, we compare the result for full illumination in gradient descent and MM algorithm and partial illumination in MM algorithm with estimation error equal to zero. The PAPR is adjusted by increasing the peak power constraint while the regularizer λ is fixed at 0.5. For block length $L = 1$ the result is shown as Fig. (10).

It can be observed by comparing the red line and the yellow line that partial illumination strategy performs slightly worse that the full illumination strategy. One reason is because of the

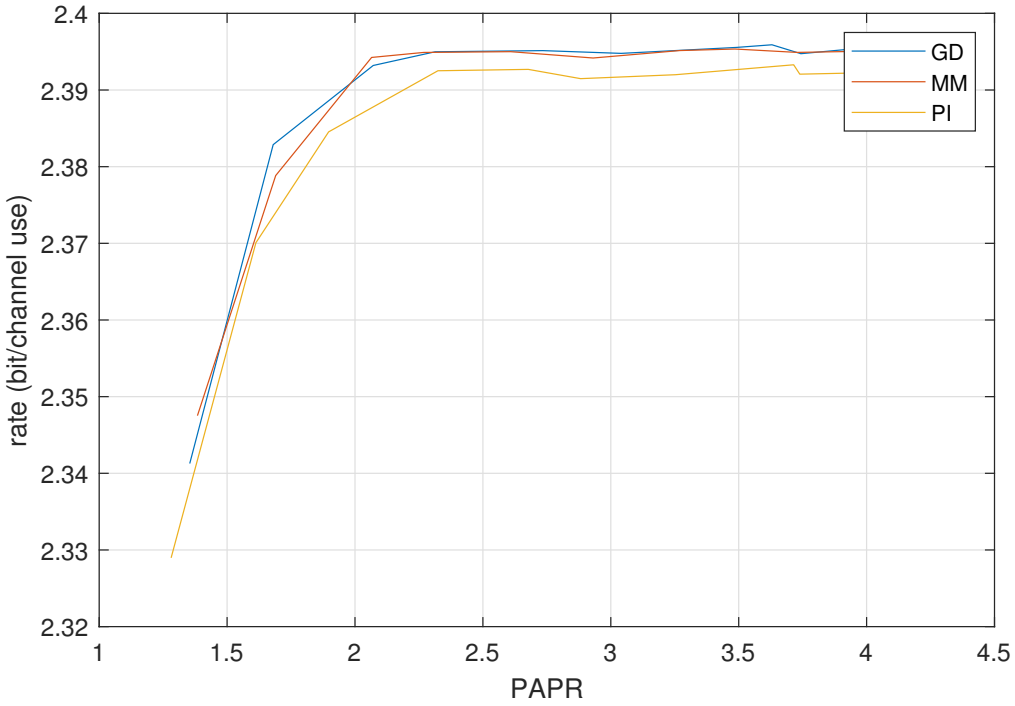


Fig. (10) Rate-PAPR with block length $L = 1$, guard interval $\tau = 0$ and regularizer $\lambda = 0.5$

spillover effect caused by the antenna pattern. In full illumination, the IRS catches more power than the one adopting the partial illumination strategy.

For larger block length, we simulated the situation where $L = 2$ and $L = 4$ which are shown as Fig. (11) and Fig. (12).

It can be seen here that with guard interval $\tau = 0$ the maximal achieveable rate of system with block length larger than one always smaller than the rate with block length one. This can be explained by Eq. (25). For block lengths L_B, let's assume a fixed stream of data with length L_{tot} which is an integer multiple of it. Following Eq. (23) the cost function for the optimization over the whole stream can be written as

$$\underset{\mathbf{X}_{1\dots L_{tot}/L_B},\mathbf{D}_{1\dots L_{tot}/L_B}}{\text{minimize}} \quad \sum_{j=1}^{L_{tot}/L_B} \text{obj}(\mathbf{D}_j, \mathbf{X}_j, \mathbf{S}_j) \tag{47}$$

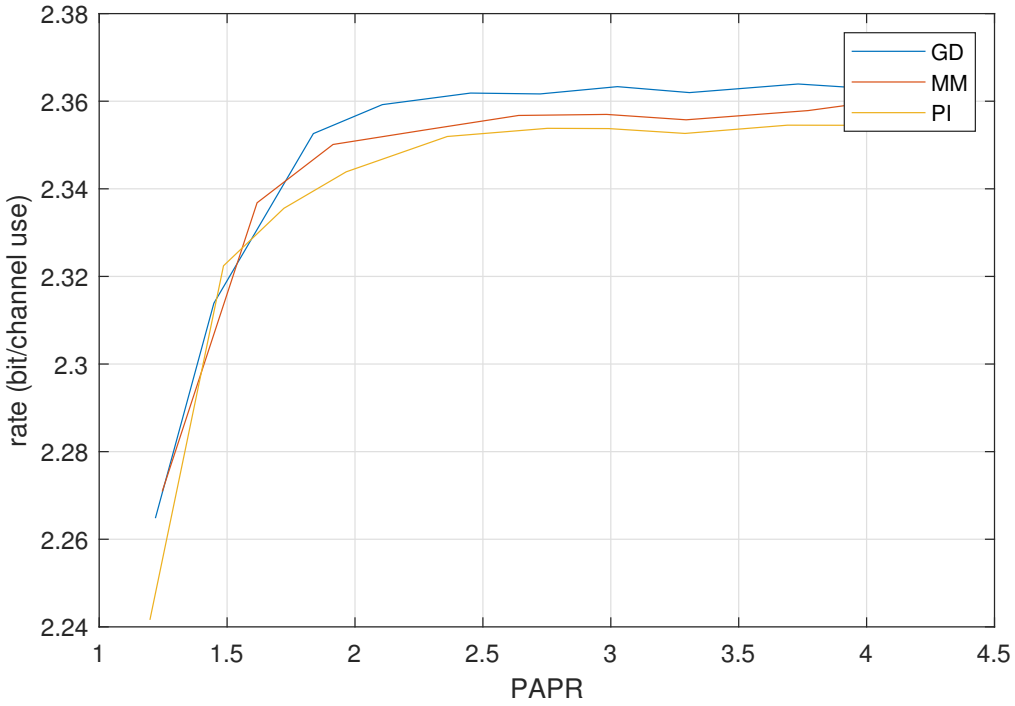


Fig. (11) Rate-PAPR with block length $L = 2$, guard interval $\tau = 0$ and regularizer $\lambda = 0.5$

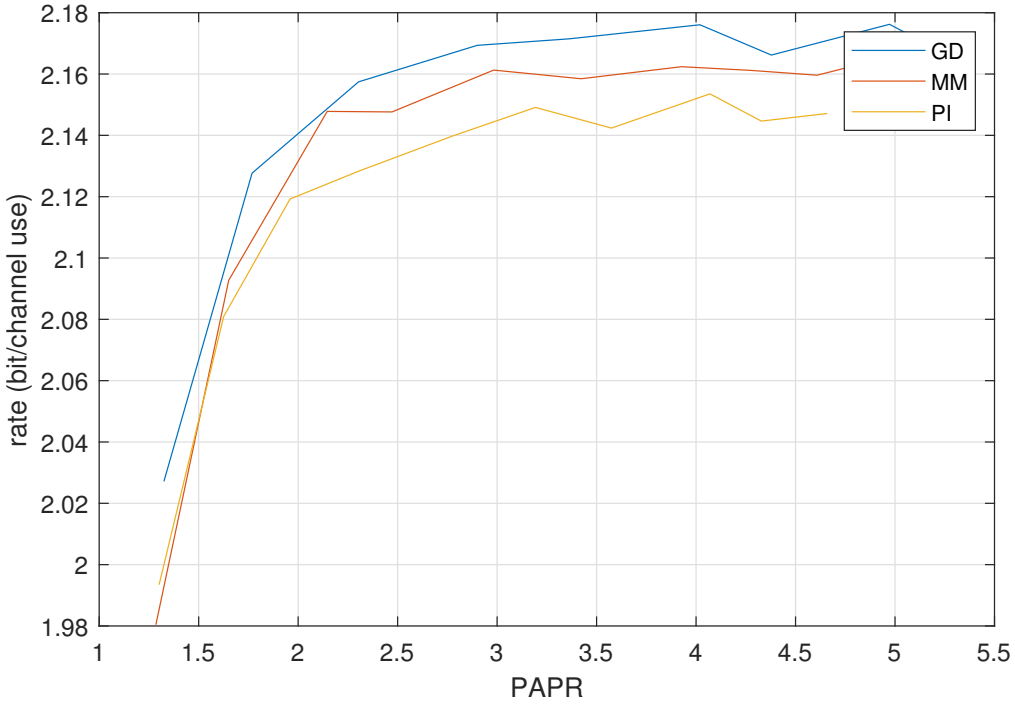


Fig. (12) Rate-PAPR with block length $L = 4$, guard interval $\tau = 0$ and regularizer $\lambda = 0.5$

According to Eq. (25), the equation Eq. (47) is equivalent to

$$\underset{\mathbf{x}_{1\dots L_{tot}/L_B},\mathbf{D}_{1\dots L_{tot}/L_B}}{\text{minimize}} \quad \sum_{j=1}^{L_{tot}/L_B} \sum_{i=1}^{L_B} \text{obj}(\mathbf{D}_j, \mathbf{x}_{i\cdot j}, \mathbf{s}_{i\cdot j}) \tag{48}$$

where $\mathbf{x}_k$ and $\mathbf{s}_k$ denote the k-th data in the stream. From Eq. (48). We know that for all $L_B \in \mathbb{N}$, any solution of Eq. (48) lies in the feasible set of the same problem described in Eq. (48) but with $L_B = 1$. Therefore, the symbolwise transmission must performs better than the blockwise transmission when the guard interval is zero. Similarly, we can see from the graph that the achieveable rate decreases as the length increases. This cannot be proven with the upper method, since a close method for finding the optimal solution is unknown. Moreover, we can even construct a special data stream such that the system with larger block length outperforms the system with smaller block length. However, from the simulation results we learn that these cases are rare. If the large block length is an integer multiple of the smaller block length, we can then prove that the system with smaller block length always performs better than the one with larger block length for the same data stream. In conclusion, the system with smaller block length tend to perform better than the system with larger block length at the cost of higher computational load and higher update rate for the IRS.

In order to compare the IRS structure to the benchmark structure, we simulate block-wise precoding of the fully connected HAD system with block length $L = 2$ and $L = 4$ as Fig. (13) and Fig. (14).

They both performs better than the IRS architecture with the same block length respectively with higher achieveable rate and less PAPR. However, IRS structure is much easier to implement in both hardware and software aspects. In IRS structure, the power divider and power combiner are not neccessary. Only 64 phase shifters are required in IRS instead of 256 phase shifters in fully connected HAD systems, which reduce the computational load greatly.

E. Regularizer

The transmission power is controlled by the regularizer λ. It balances the emphasis between the interference power and the signal power transmitted by the RF chains. In order to study its behavior, we set the peak power constraint to infinity and plot the rate-λ curves in Fig. (18).

As is mentioned before, the power of the intended symbol is 0.1 and the normalized noise has a max value of 0.1. Therefore, for the best case where the interference power is zero, we

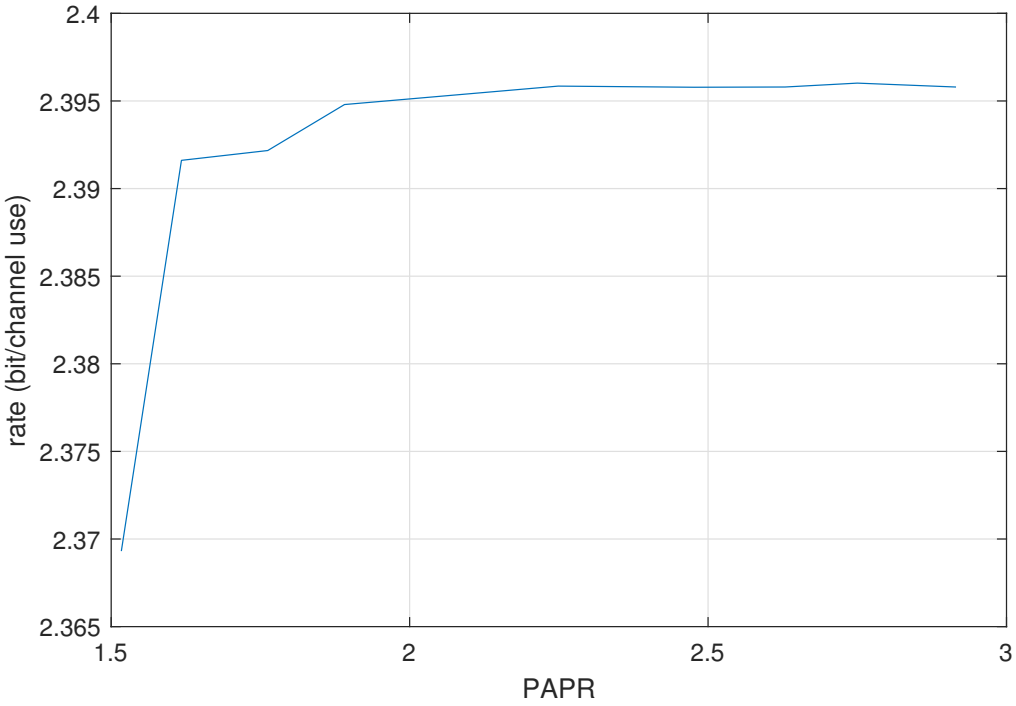


Fig. (13) Rate-PAPR of fully connected system with block length $L = 2$, guard interval $\tau = 0$ and regularizer $\lambda = 0.5$

would have $rate = \log_2 11 \approx 3.45943$. From Fig. (18) it can be observed that for $L \leq 4$ the curves approach this rate limit as λ tend to zero. This phenomenon was also discussed in [4] with a general analog unit. It is stated that for a HAD system with general analog unit, the interference power $||\mathbf{HF}_{\text{GA}}\mathbf{X} - \mathbf{S}||^2_{\text{fro}}$ can never be made exactly to zero if the block length is larger than the number of RF chains and the the channel matrix $\mathbf{H}$ and $\mathbf{S}$ have full rank. The matrix $\mathbf{F}_{\text{GA}} \in \mathbb{C}^{M\times N}$ in the statement represents the 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 explain the curve standing for $L = 8$. We can also explain the curve for $L = 1$ by using Theorem (1).

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

For the simple case where $\mathbf{H}$ is a row vector, namely when there is only one user $K = 1$.

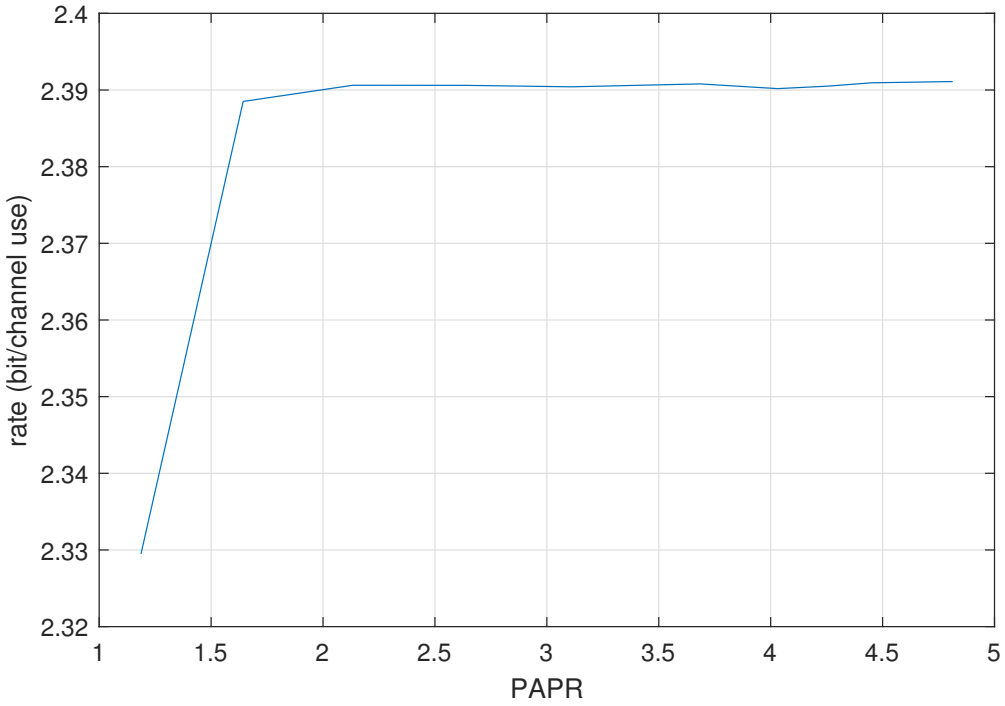


Fig. (14) Rate-PAPR of fully connected system with block length $L = 4$, guard interval $\tau = 0$ and regularizer $\lambda = 0.5$

Let $\mathbf{h}^T = H$, then we try to find out the lower bound of probability $Pr(\exists \mathbf{d}, (\mathbf{h}^T\mathbf{d} = 0))$.

This can be translated into geometry representation. Each entry in vector $\mathbf{h}^T$ is an vector whose length is a random variable following standard gaussian distribution and phase is uniformly distributed between 0 to 2π. The vector $\mathbf{d}$ represents the rotation operations which are applied to the columns of $\mathbf{h}^T$ with an arbitary angle. Therefore, the proposition $\exists \mathbf{d}, (\mathbf{h}^T\mathbf{d} = 0)$ can be interpreted as finding an rotation angle for each of the entries in vector $\mathbf{h}^T$ such that these rotated entries add up to zero. Following triangular inequality, we have

$$\begin{aligned} &\exists \mathbf{d}, (\mathbf{h}^T\mathbf{d} = 0) \\ \Leftrightarrow &\max(|h_1|, \ldots, |h_M|) \leq |h_1| + \cdots + |h_M| - \max(|h_1|, \ldots, |h_M|) \end{aligned} \tag{49}$$

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

$$\begin{aligned} P_1 =& Pr(\exists \mathbf{d}, (\mathbf{h}^T\mathbf{d} = 0)) \\ =& 1 - (M \times Pr(|h_1| > |h_2| + \cdots + |h_M|)) \end{aligned} \tag{50}$$

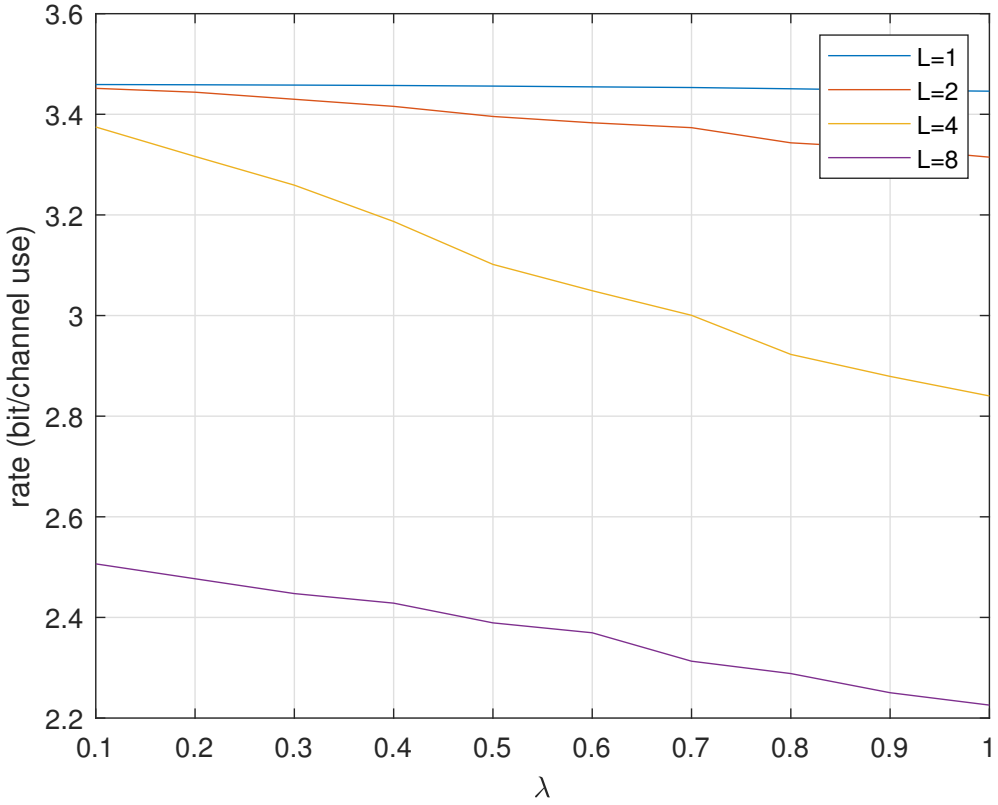


Fig. (15) Rate-λ behavior of MM-algorithm with guard interval $\tau = 0$

The properties of F-distribution can be used to determin the upper bound of $Pr(|h_1| > |h_2| + \cdots + |h_M|)$.

$$\begin{aligned}
&M \times Pr(|X| > |X_1| + \cdots + |X_{M-1}|) \\
\leq &M \times Pr(|X|^2 > |X_1|^2 + \cdots + |X_{M-1}|^2) \\
\leq &M \times Pr(|X|^2 + |Y|^2 > |X_1|^2 + \cdots + |X_{M-1}|^2) \\
= &M \times Pr\left(\frac{2}{M-1} > \frac{(|X_1|^2 + \cdots + |X_{M-1}|^2)/(M-1)}{(|X|^2 + |Y|^2)/2}\right) \\
= &M \times F\left(\frac{2}{M-1}; M-1, 2\right) = M \times I_{\frac{1}{2}}\left(\frac{M-1}{2}, 1\right) \\
= &M \times \left(\frac{1}{2}\right)^{\frac{M-1}{2}}
\end{aligned} \tag{51}$$

The first less equal sign holds because $|X|^2 > (|X_1|^2 + \cdots + |X_{M-1}|^2)^2 > |X_1|^2 + \cdots + |X_{M-1}|^2$.

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

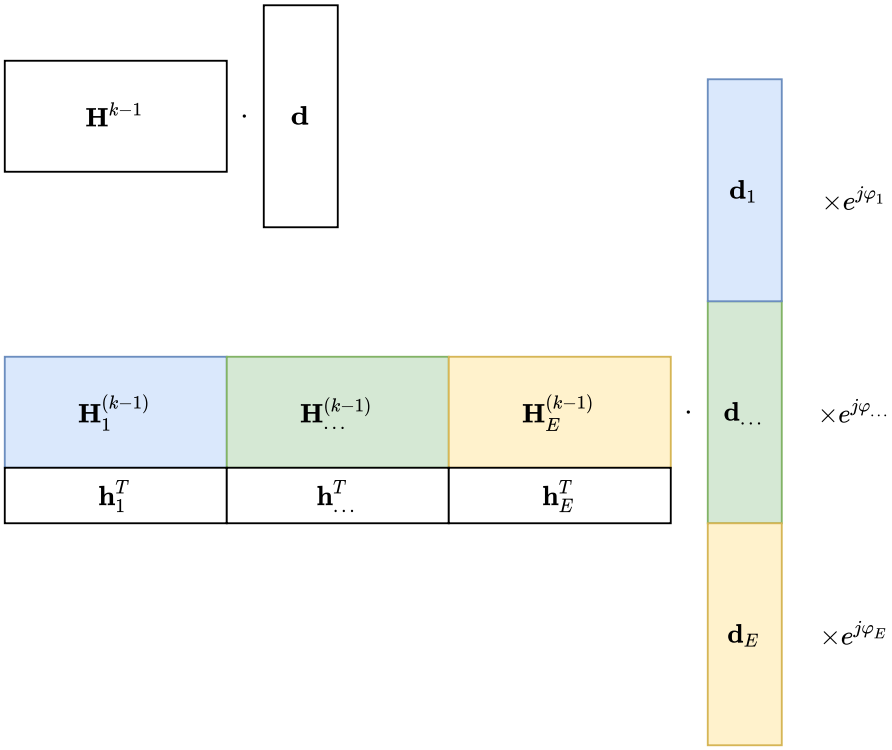


Fig. (16) Idea of construction

$\mathbf{d}$ can be constructed recursively. We try to construct a sufficient condition for the existence of the vector $\mathbf{d}$ that fulfills the requirements and then we try to prove that probability that this sufficient condition not hold tend to zero. Suppose we know how to construct $\mathbf{d}$ if $K = k - 1$. For sake of simplicity, we denote the channel matrix in that condition as $\mathbf{H}^{(k-1)}$, and we have $\mathbf{H}^{(k-1)} \cdot \mathbf{d} = \mathbf{0}$. 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 divide $\mathbf{H}^{(k)}$ into several $(k-1)$-row submatrix with smaller column numbers and sectorize $\mathbf{d}$ into several vecrtors with corresponding columns. This is illustrated as Fig. (16).

To find the sufficient condition, we only need to find one solution for this problem. Therefore, we first find $\mathbf{d}_1, \dots, \mathbf{d}_E$ such that $\forall i \in \{1, \dots, E\}, \mathbf{H}_i \mathbf{d}_i = \mathbf{0}$. Then for the last row of matrix $\mathbf{H}^{(k)}$, we know that $\mathbf{h}_i^T \cdot \mathbf{d}_i$ is a scalar value and for each subvector $\mathbf{d}_i$ there is at least one degree of freedom left. We can apply an rotation $e^{j\varphi_i}$ to each of the vectors $\mathbf{d}_i$, and the colored parts in figure (16) stay $\mathbf{0}$.

Now the problem is transformed back to the previous one, where we want to find rotation

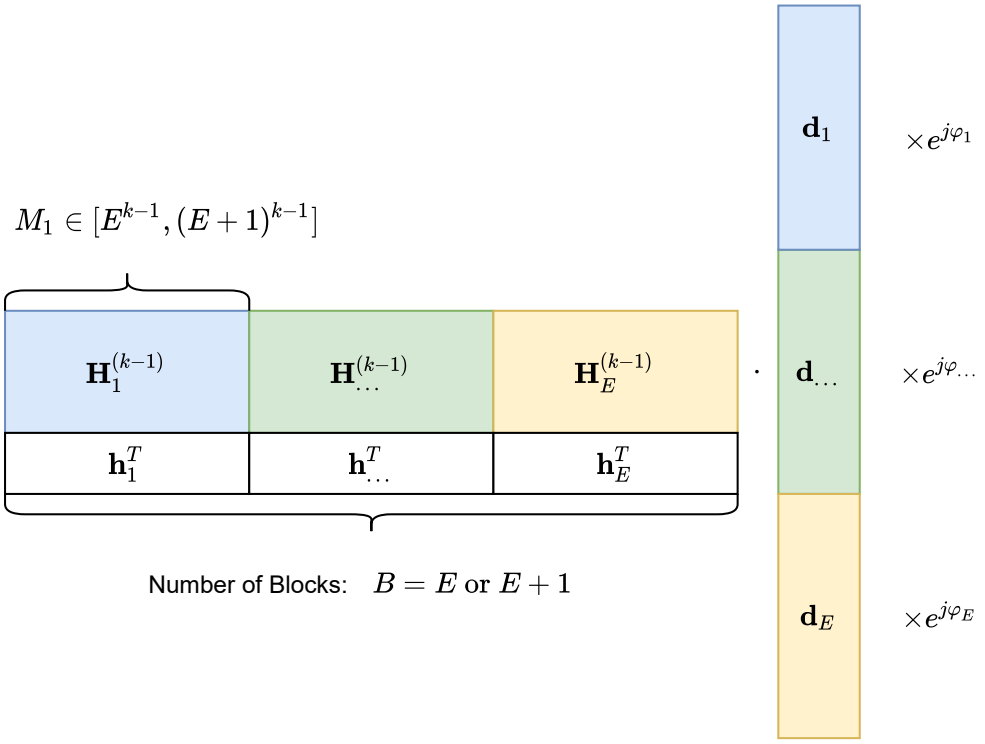


Fig. (17) Submatrix Construction

angles $\varphi_1, \ldots, \varphi_E$ such that

$$\begin{bmatrix} \mathbf{h}_1^T\mathbf{d}_1 & \ldots & \mathbf{h}_E^T\mathbf{d}_E \end{bmatrix} \cdot \begin{bmatrix} e^{j\varphi_1} \\ \vdots \\ e^{j\varphi_E} \end{bmatrix} = 0 \tag{52}$$

This is the general idea for the proof.

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 figure (17). It is worth noticing that the sectorization method is not unique, we only require that the the block column length is within the range of $[E^{k-1}, (E+1)^{k-1}]$ and there are E or $E+1$ blocks in total. The proposition that $\mathbf{H}^{(k)}$ cannot be rotated to an all-zero matrix implies that either one of the colored blocks cannot be rotated to zero or the final row cannot be made zero. The probability that the colored part cannot be made zero can be calculated by using recursion. So let's consider the last row of matrix $\mathbf{H}^{(k)}$ first. Define

$$\begin{aligned} &B \in \{E, E+1\} : \text{block number} \\ &M_{1\ldots B} \in [E^{k-1}, (E+1)^{k-1}] : \text{column number the blocks.} \end{aligned} \tag{53}$$

This is also illustrated in figure (17). Each block in the last row $\mathbf{h}_i^T$ will be multiplied with corrspondence vector block $\mathbf{d}_i$,which create 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 normlize the variance to 1. Define

$$\begin{aligned} &\forall b \in \{1, \ldots, B\}, Y_b = \mathbf{h}_b^T\mathbf{d}_b \sim CN(0, M_b) \\ &\forall b \in \{1, \ldots, B\}, X_b \sim CN(0, 1) \end{aligned} \tag{54}$$

Then we have,

$$\begin{aligned} &\sum_{b=1}^{B} Pr(|Y_b| > |Y_1| + \cdots + |Y_{b-1}| + |Y_{b+1}| + \cdots + |Y_B|) \\ \leq &\sum_{b=1}^{B} Pr(|Y_b| > \sqrt{M_{min}}(\frac{|Y_1|}{\sqrt{M_i}} + \cdots + \frac{|Y_{b-1}|}{\sqrt{M_{b-1}}} + \frac{|Y_{b+1}|}{\sqrt{M_{b+1}}} + \cdots + \frac{|Y_B|}{\sqrt{M_B}})) \\ = &\sum_{b=1}^{B} Pr(\frac{\sqrt{M_b}}{\sqrt{M_{min}}}\frac{|Y_b|}{\sqrt{M_b}} > |X_1| + \cdots + |X_{B-1})| \\ \leq &B \times Pr(\frac{\sqrt{M_{max}}}{\sqrt{M_{min}}}|X| > |X_1| + \cdots + |X_{B-1}|) \\ \leq &B \times \left(\frac{M_{max}}{M_{max} + M_{min}}\right)^{\frac{B-1}{2}} \end{aligned} \tag{55}$$

For larger M, we have

$$\lim_{M \to +\infty} \frac{M_{max}}{M_{max} + M_{min}} \leq \lim_{E \to +\infty} \frac{(E+1)^{k-1}}{2 \times E^{k-1}} = \frac{1}{2} \leq \frac{2}{3} \tag{56}$$

Therefore,

$$Pr\left(\nexists(\varphi_1, \cdots, \varphi_B). \begin{bmatrix} \mathbf{h}_1^T\mathbf{d}_1 & \ldots & \mathbf{h}_B^T\mathbf{d}_B \end{bmatrix} \cdot \begin{bmatrix} e^{j\varphi_1} \\ \vdots \\ e^{j\varphi_B} \end{bmatrix} = 0\right) \leq (N+1) \times \left(\frac{2}{3}\right)^{\frac{E-1}{2}} \tag{57}$$

The complete probability can then be obtained

$$\begin{aligned} P_k &= Pr(\nexists \mathbf{d}, (\mathbf{H}^{(k)T}\mathbf{d} = 0)) \\ &\leq (E+1) \times P_{k-1} + (E+1) \times \left(\frac{2}{3}\right)^{\frac{E-1}{2}} \end{aligned} \tag{58}$$

Based on this inequality, one can define an auxilary sequence recursively

$$\begin{aligned} P_1' &= (E+1) \times \left(\frac{2}{3}\right)^{\frac{E-1}{2}} \\ P_k' &= (E+1) \times P_{k-1} + (E+1) \times \left(\frac{2}{3}\right)^{\frac{E-1}{2}} \end{aligned} \tag{59}$$

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

$$P'_k = \frac{(E+1)^{k+1} - (E+1)}{E} \left(\frac{2}{3}\right)^{\frac{E-1}{2}} \tag{60}$$

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

For the precoding optimization problem with block length $L = 1$ and RF chain number $N = 1$. Define $\mathbf{H}' = \mathbf{H} \times \mathrm{diag}_{\mathrm{matrix}}\{x\mathbf{t}\}$, and the problem is transformed to finding vector $\mathbf{d} = \mathrm{diag}_{\mathrm{vector}}\{\mathbf{D}\}$ such that $\mathbf{s} = \mathbf{H}'\mathbf{d}$. In order to make the left side zero, define

$$\begin{aligned} \mathbf{H}'' &= \begin{bmatrix} \mathbf{H}', & \mathbf{s} \end{bmatrix} \\ \mathbf{d}' &= \begin{bmatrix} \mathbf{d}_{aux} \\ d_{M+1} \end{bmatrix} \end{aligned} \tag{61}$$

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 $\mathbf{D} = \mathrm{diag}_{\mathrm{matrix}}\{-d_{M+1}^{-1}\mathbf{d}_{aux}\}$. For larger RF chain number $N > 1$, we can always transform it back to the case where $N = 1$ by setting the transmitted signal all the RF chains other than the first one to zero. However, the asymptotic behavior of block length larger than one is still unclear.

The relation between the average RF chain power and the regularizer lambda is shown as Fig. (18). It can be observed from the figure that for relative small block lengths, the structure with small L performs better. However, for relative large block lengths the power behavior is very similar. We can find the explaination for this by considering the objective function. For smaller block lengths, the interference power can be made small easily, as a result, the dominant part of the objective function is the term for total RF chain power. Thus, the system with small L will put more emphasis on reducing the RF chain power while the system with larger L will put more emphasis on miniming the interference power.

F. Rate, Block Length and IRS Element Number

Now we set the guard interval to be the same length as the symbol interval $\tau = T_s$. We also introduce the normalized channel estimation error matrix whose entries are i.i.d. with variance $\sigma_e^2 \in \{0, 0.01, 0.04\}$. The peak power constraint is also removed. The Fig. (19) shows that the achieveable rate increases as the number of IRS elements increases and the rate decreases as the variance of the estimation increases which are expected.

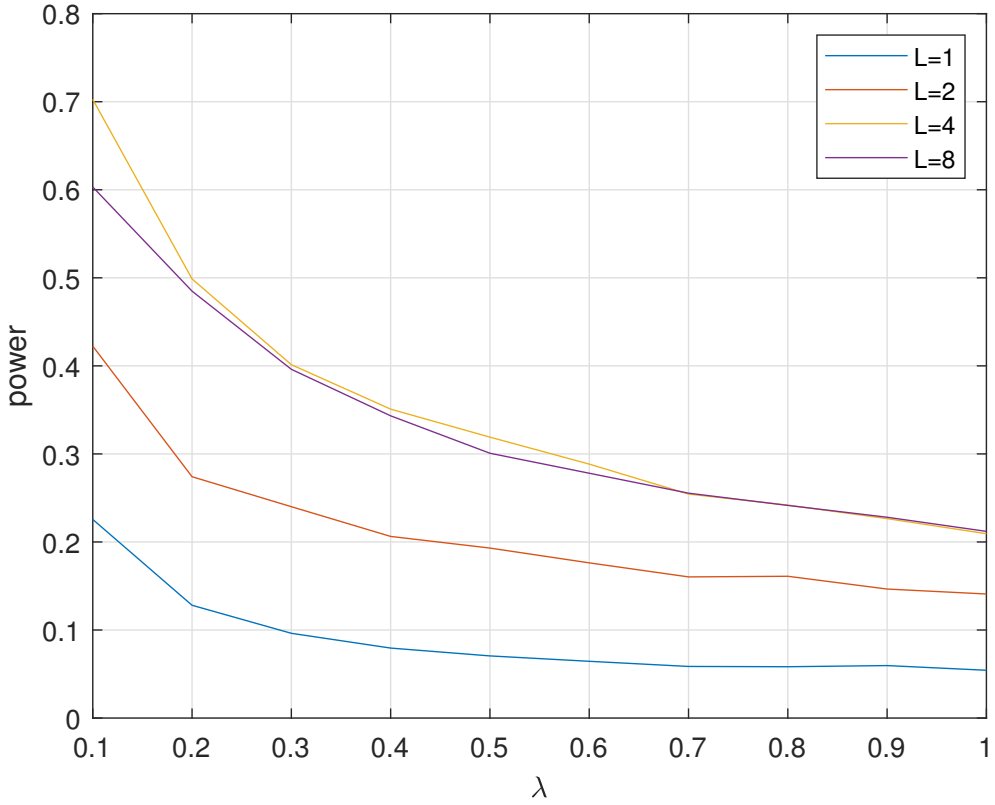


Fig. (18) Average RF chain power-λ with guard interval $\tau = 0$

Next, we want to study the rate-block length behavior. The IRS element number is set to $M = 144$, and the simulation result is plotted in Fig. (20). 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 and the dominant part is the guard interval. It can be seen that the rate at $L = 1$ of the estimation error free curve is 1.73 which is one half of $\log_2 11$. This is consistent with Theorem (1). As the block length increases, the term $\frac{LT_S}{\tau+LT_S}$ tend to 1 and the dominant part becomes the term $\log(1+SINR)$. Based on thess two points we can predict that for larger τ, the peak of the rate-L plot will shift to right since the term $\frac{LT_S}{\tau+LT_S}$ monotonically increases as L increases while the term $\log(1+SINR)$ decreases as L increases. The complete relation of rate, $\sqrt{M}$ and L can be shown as a mesh surface, which is shown in Fig. (21).

At last it is also interesting to know the minimal required IRS element to achieve a certain rate while the block length is increasing. Instead of directly ploting by using Fig. (21), we calculate the rate of each transmission block, and use the block rate to calculated the averaged minimal

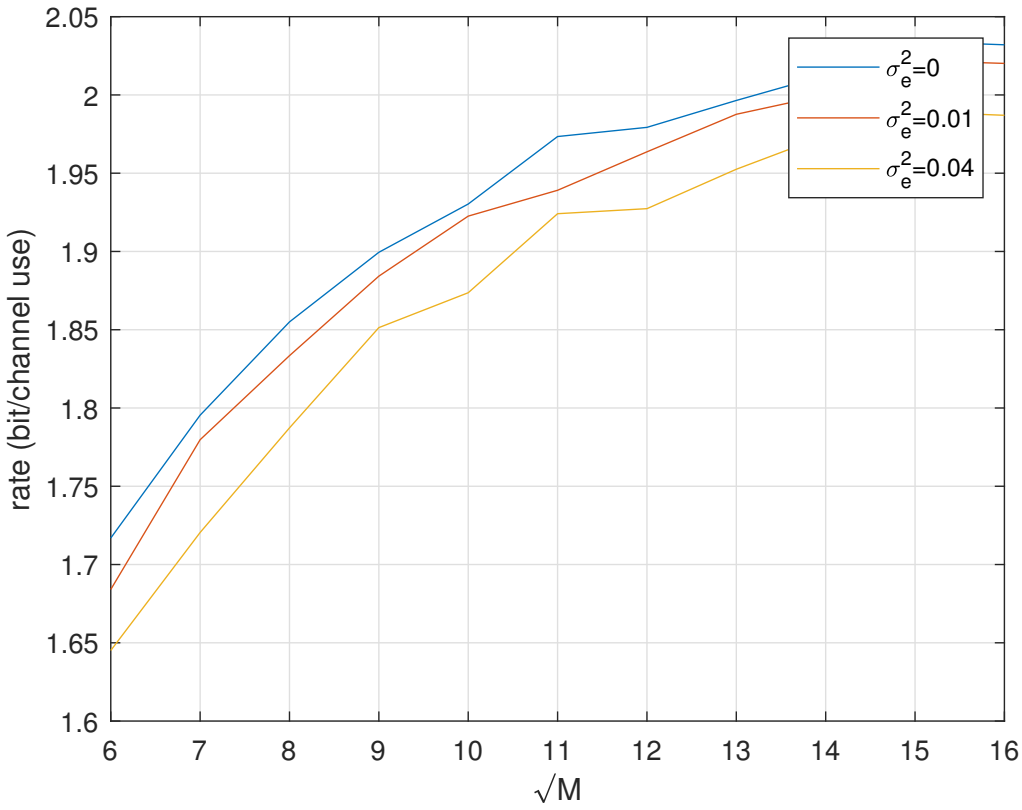


Fig. (19) Rate-$\sqrt{M}$ with guard interval $\tau = T_s$ and $L = 10$

IRS size for reaching a certain rate. For rate threshold 1.8, the averaged minimal M versus L is plotted as Fig. (22). It can be directlty read from the graph that influence of the estimation error grows faster as L increases. One reason is that the value of the error term $||\tilde{\mathbf{e}}\mathbf{DTX}||^2_{\text{fro}}$ grows larger as the block length increases. The growth rate of M is faster than linear. However, for L increasing from 2 to 20 the minimal required IRS element is only 4 times larger.

VI. Conclusions

This paper discusses the performance of a Hybrid A-D transmitter system with passive antenna array. In order to lower the update rate of the passive antennas, data vectors of multiple time steps can be stacked into a matrix. By doing so, the system will suffer a performance loss. However, from the experiments, if $L < N$, this loss can be compensated by increasing the number of passive antennas. Although gradient descend method can only reach a suboptimal solution, the mean RSS can still reach zero if the number of passive antennas is high. There are still some

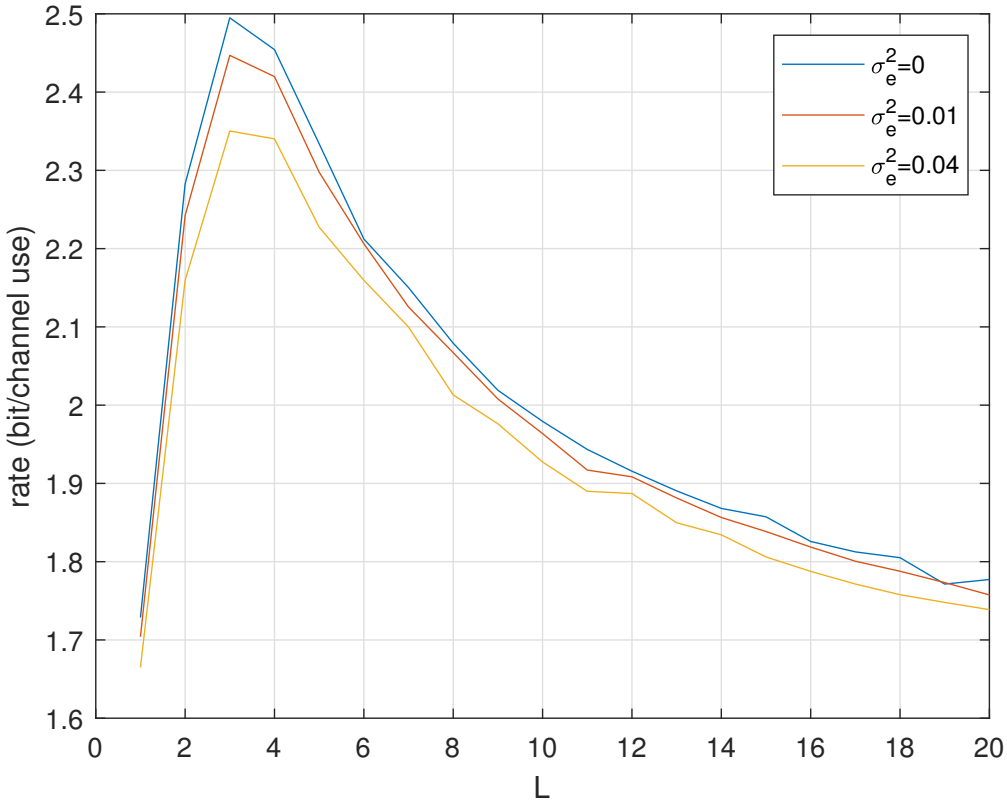
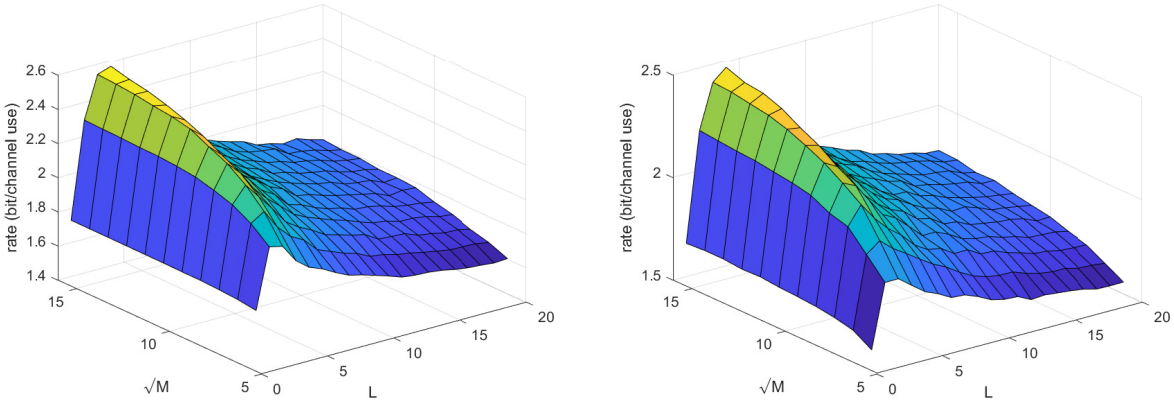


Fig. (20) Rate-L with guard interval $\tau = T_s$ and $M = 144$



(a) $\sigma_e^2 = 0$ (b) $\sigma_e^2 = 0.04$

Fig. (21) Full relation between rate, $\sqrt{M}$ and L with $\tau = 1$

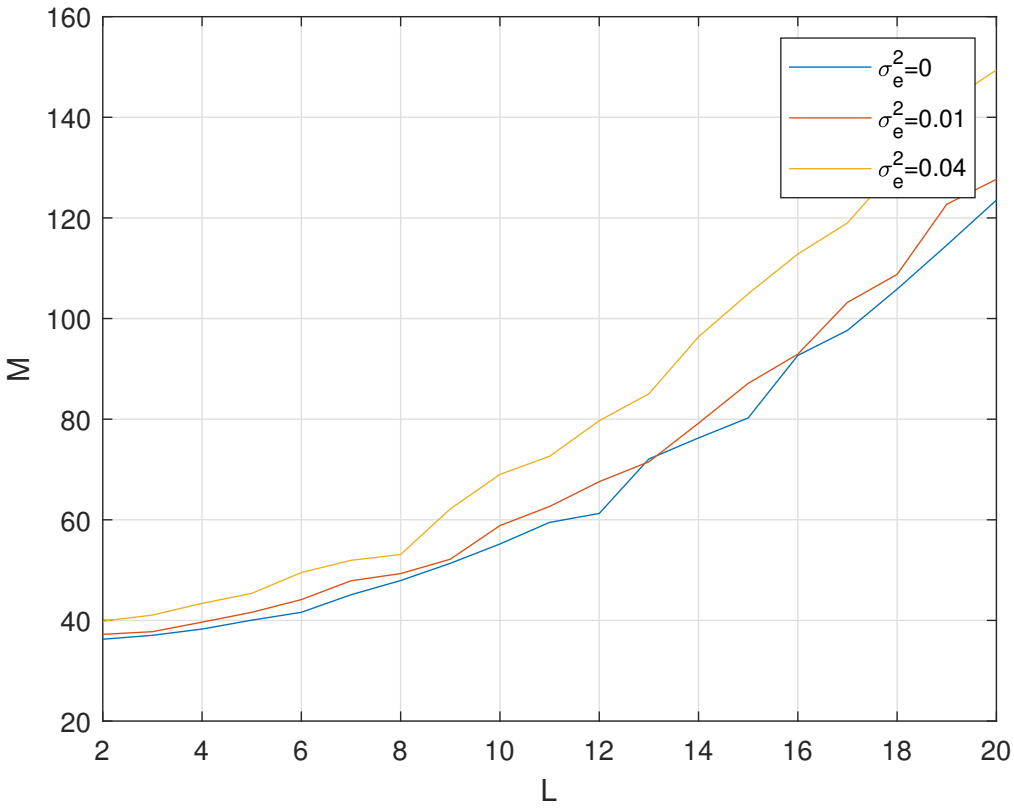


Fig. (22) Averaged minimal M-L with rate_threshold$= 1.8$

work to be done. The asymptotic behavior of large passive antenna number M is proven only for the case where block length $L = 1$. We need to varify whether the mean RSS will still tend to zero if M tends to zero with larger block lengths. We only consider the Gaussian channel here. It may be better if we investigate the design of the precoding method by using Kronecker channel model.

REFERENCES

[1] V. Venkateswaran, F. Pivit, and L. Guan, "Hybrid rf and digital beamformer for cellular networks: Algorithms, microwave architectures, and measurements," *IEEE Transactions on Microwave Theory and Techniques*, vol. 64, no. 7, pp. 2226–2243, 2016.

[2] 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, 2021.

[3] 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.

[4] M. A. Sedaghat, B. Gade, R. R. Müller, and G. Fischer, "A novel hybrid analog-digital transmitter for multi-antenna base stations," pp. 1714–1718, 2017.

[5] A. Garcia-Rodriguez, V. Venkateswaran, P. Rulikowski, and C. Masouros, "Hybrid analog–digital precoding revisited under realistic rf modeling," *IEEE Wireless Communications Letters*, vol. 5, no. 5, pp. 528–531, 2016.

[6] D. M. Pozar, *Microwave engineering*. John wiley & sons, 2011.