Studienarbeit/Diplomarbeit

Block-Level Precoding via IRS-Aided Hybrid Transmitter

Zilu Zhao

Lehrstuhl für Digitale Übertragung
Dr.-Ing. Ali Bereyhi
Universität Erlangen-Nürnberg

Betreuer: Dr.-Ing. Ali Bereyhi

May 11, 2022

Themenstellung

Bitte die original Themenstellung hier einfügen.

Declaration

To the best of my knowledge and belief this work was prepared without aid from any other sources except where indicated. Any reference to material previously published by any other person has been duly acknowledged. This work contains no material which has been submitted or accepted for the award of any other degree in any institution.

Erlangen, May 11, 2022

Zilu Zhao
Sophienstrasse 91
Erlangen

Contents

Abstract

This article proposes a block-wise transmission scheme for a hybrid Analog-Digital system where IRS is used as analog unit. This scheme lowers the update rate of the analog unit by stacking several input vectors together. An iterative method is used to design the digital unit output signals and analog unit configurations. During each iteration, the digital unit outputs and IRS phase shifts are computed in turn in order to achieve a suboptimal point for the interference power under certain power constraints. We find out that the digital unit output can be computed by solving a convex optimization problem. Meanwhile, the designing of the analog unit configurationis non-convex. Thus, more general methods like the gradient descent methods and MM algortihm are used while designing the IRS configurations. By using numerical simulations, we reveal that trade-off can be made between required IRS number and the block length.

Glossary

Abbreviations

MIMO	multiple-input multiple-output
IRS	intelligent reflecting surface
HAD	Hybrid Analog-Digital
CSI	channel state information
RF	radio frequency
BS	base station
UT	user terminal
PAPR	peak-to-average-power-ratio
FI	full illumination
PI	partial illumination
SI	seperate illumination
MM	Majorize-minimization

Operators

$(\cdot)^*$	Conjugate
$(\cdot)^{-1}$	Matrix inversion
$(\cdot)^T$	Matrix transpose
$(\cdot)^H$	Hermitian matrix, conjugate transpose
$\|\|\cdot\|\|_{\text{fro}}$	Frobenius norm
$\mathscr{E}\{\cdot\}$	Expectation
$\text{tr}\{\cdot\}$	Trace

Symbols

K	Number of user terminals
L	Block length
M	Number of IRS elements
N	Number of RF chains
$\text{obj}(\cdot)$	Objective/cost function
σ_s^2	Signal variance/power
σ_f^2	Variance of the fading
σ_n^2	Noise variance/power
σ_e^2	Variance of the estimation error
T_s	Symbol interval

Chapter 1

Introduction

In the past few years, the Fifth Generation communication system (5G) which aims to provide higher data rate is becoming more and more widespread [1]. In order to achieve higher data rate, many efforts have been invested into the research of massive multiple-input multiple-output (MIMO) systems and millimeter wave communication systems. By using massive MIMO systems, the channel capacity can be significantly improved without the exponential increament of the transmittion power [2]. The millimeter communication technology has also gained interests of researchers recently. According to a research [3], the spacial correlation between the elements on an antenna array can be modeled by a sinc$(\cdot)$ function with angular frequency $\frac{2\pi}{\lambda_w}$ where λ_w is the wave length. Therefore, in order to acheive spatial uncorrelated transmission, the $\lambda_w/2$-spacing must be satisfied. Because of this spatial constraint, the millimeter wave technology allows the transmission systems to facilitate the use of large antenna arrays in a more compact form. Moreover, due to the directional beams created by the large antenna arrays, the systems implemented with millimeter wave technologies can combat the propagation loss caused by higer frequency with large array gains [1]. Despite all the advantages, This often leads to higher complexity, energy consumptions and cost [4].

Hybrid Analog-Digital (HAD) transmission is a rather recent technique for reducing the structural complexity in massive MIMO systems [5]. 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, for instance radio frequecy beamforming networks with buttler matrix[6]. Among those HAD architectures, the most considered one is a fully connected structure. However, such structures often require a large number of phase shifters. Those phase shifters are also energy-intensive [7]. Therefore, in this paper we consider a recent proposal which implements the analog units via intelligent reflecting surfaces (IRSs) [8]. We will mainly discuss 2-D IRS.

By using both digital units and analog units, the set of possible acheivable symbols always contains the set of the symbols generated by the system consists of only digital unit utilizing the same amount of RF chains. Therefore, some information is encoded into the output signal by the analog unit. As a result, the number of RF chains of a HAD system may be reduced while maintaining the same rate. This can lower the cost and complexity of base stations. According to [9], the update rate of analog unit can be set slower than the update rate of the digital unit for a system with a fully connected analog unit without causing any interference. It also revealed that HAD systems allows a trade-off between the number of RF-chains and the update rate. In this paper, we also try to investigate the trade-off between the number of passive antennas and the update rate in IRS-aided HAD transmission systems. Therefore, block fading is assumed and a block-wise precoding procedure will be discussed. And we will always assume that the channel state information (CSI) is always available at the base station (BS).

In the first part of this paper, the system model for the IRS aided HAD transmission system is discussed in detail. Based on channel model, we will give a formal representation of the precoding problem as an optimization problem. After that, we will introduce the idea of using iterative metod by optimizing the digital unit and the analog unit in turn iteratively to find a suboptimal solution for the previously optimization problem constructed explicitly in Chaper (3). A digital baseband precoding based on convex optimization method is introduced. The proposed analog beamforming scheme is formulized as a unit-modulus optimization problem. To address this problem, we develop two low-complexity alogorithms by means of method of gradient descent and MM algorithm. During the derivation, we also take the imperfect CSI into considerations. The performance of the proposed algorithms are further investigated through several numerical simulations. Motivated by the findings, we show that for symbol-wise precoding, interference free transmission is achievable at the base station if a large enough IRS is employed. We will also investigate numerically the relation between the number of passive antennas and the size of block length. Finally, we will give some conclusions and some future thoughts.

Chapter 2

System Model

2.1 Transmission Procedure

We consider downlink transmission in a multi-user MIMO system consisting of a 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

$$\mathbf{y}_i = \mathbf{H}\mathbf{a}_i + \mathbf{n}_i, \tag{2.1}$$

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 signals transmitted by 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_n^2.

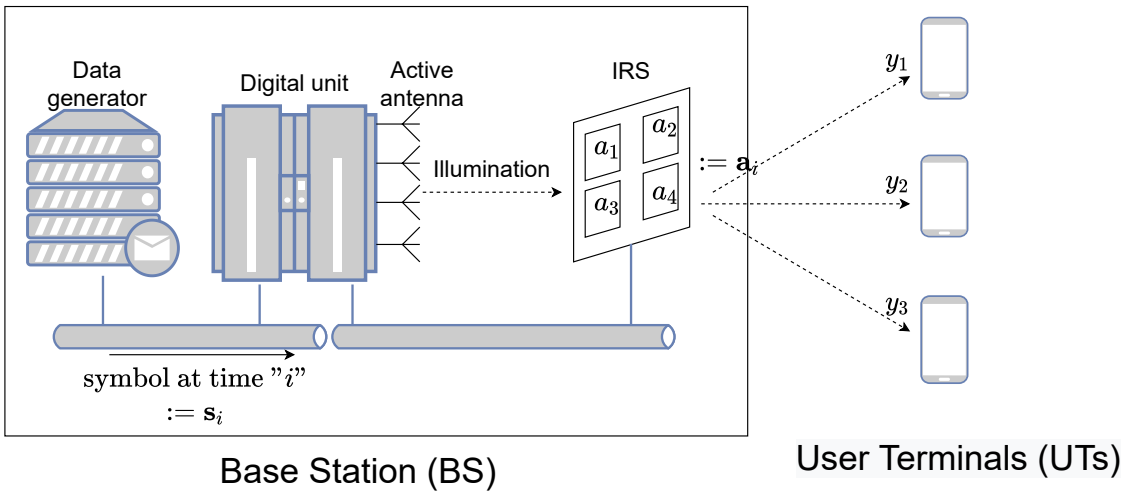


Figure 2.1: Transmitter Model

The BS employs an IRS-aided HAD transmitter whose architecture is demonstrated in Fig. (2.1).

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. (2.2).

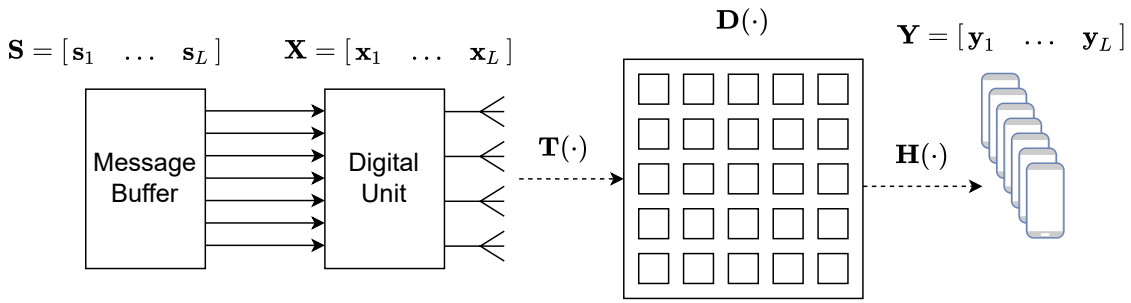


Figure 2.2: Signal diagram of the Base Station

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 this message block are contained by $\mathbf{D} = \text{diag}\{e^{j\beta_1}, \dots, e^{j\beta_M}\}$. 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 where only a simple disjoint decoder is required by the UTs which can be modeled as a diagonal matrix $\mathbf{F}_D$, i.e. $\tilde{\mathbf{Y}} = \mathbf{F}_D\mathbf{Y} = \mathbf{S}$.

Based on the above discussion, the matrix for the received signals $\mathbf{Y}$ can also be represented as

$$\mathbf{Y} = \mathbf{HDTX} + \mathbf{N}. \tag{2.2}$$

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, we have $L_X = L$ and $\mathbf{X} \in \mathbb{C}^{N \times L}$.

After the digital unit receives the entire intended message block from the buffer, it will computer the optimal RF chain output $\mathbf{X}$ and the IRS phase shifts $\mathbf{D}$. If we expand the matrix $\mathbf{Y}$ and $\mathbf{X}$ in Eq. (2.2) as column vectors, we would have

$$\begin{bmatrix} \mathbf{y}_1 & \dots & \mathbf{y}_L \end{bmatrix} = \mathbf{HDT} \begin{bmatrix} \mathbf{x}_1 & \dots & \mathbf{x}_L \end{bmatrix} + \mathbf{N}, \tag{2.3}$$

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 symbol time interval, one column of $\mathbf{X}$ is transmitted by the active antennas and this RF chain output vector is reflected by the IRS.

In block-wise precoding, we apply the same phase shift matrix $\mathbf{D}$ to all the active antenna output vectors contained in the same time block. In this way, the active antennas are updated once every symbol 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 of length L between transmission of vector $\mathbf{s}_i$ and the transmission of $\mathbf{x}_i$. Therefore, this system has two phases. The first phase is the computation phase that happens during the guard interval between two time blocks. During this phase, the buffer unit feeds the message matrix $\mathbf{S}$ to the digital unit. After that, also during the same guard interval, the digital unit calculates the suitable RF chain output symbol matrix $\mathbf{X}$ and the IRS configuration $\mathbf{D}$. The second phase is the transmission phase which should be as long as L symbol time intevals. During the transmission phase, the digital unit illuminates the matrix $\mathbf{X}$ calculated in the previous phase to the IRS with phase shifts described by matrix $\mathbf{D}$. If the computation phase/guard interval is instant, this procedure can be illustrated in Fig. (2.3).

In Fig. (2.3), 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 vectors from time interval $L+1$ to $2L$ are computed based on the buffered message vectors generated from time interval 1 to L. If the message vectors are generated at the rate R and $T_s = 1/R$. It can then be observed that the RF chain output also has the same update rate but with a delay of LT_s ralative to the corresponding messages. However, since the IRS is reconfigurated once per time block. Its update rate is only R/L.

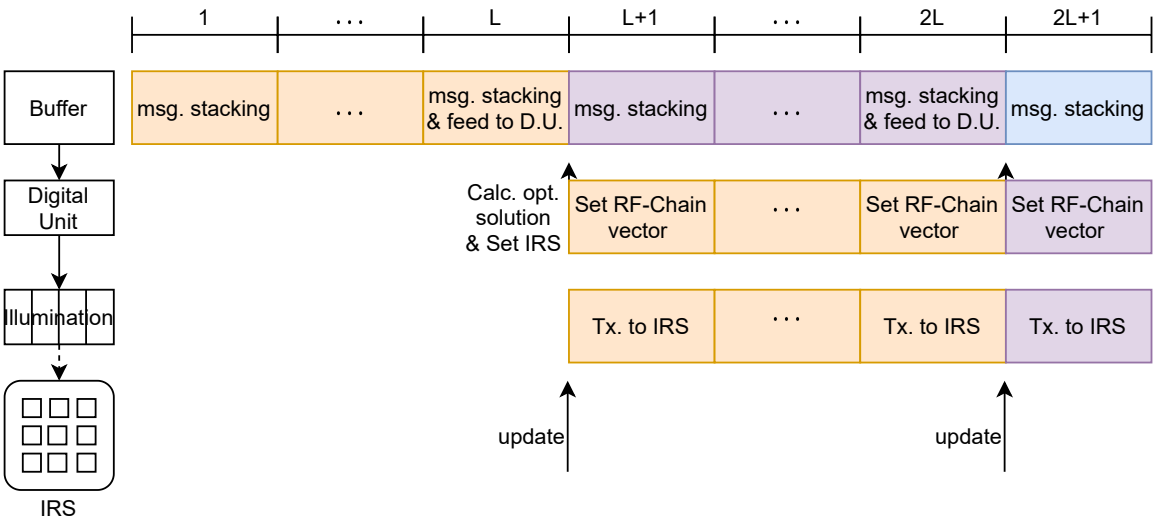


Figure 2.3: Transmitter Time Table

2.2 Channel Model

According to Eq. (2.1), the wireless channel from the BS to the k-th UT can be modeled as Eq. (2.4).

$$y_k = \mathbf{h}_k^T \mathbf{a} + n_k \tag{2.4}$$

The row vector $\mathbf{h}_k^T$ models the channel coefficients from the IRS to the k-th UT. If we take scattering and antenna spatial correlations into consideration, following [10], it can be further modeled in a more detailed form as

$$\mathbf{h}_k^T = \frac{1}{\sqrt{P}} \sum_{p=1}^{P} h_p \mathbf{h}_t^T(\theta_p^t, \phi_p^t), \tag{2.5}$$

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 which models the phase shifts of the IRS elements on the IRS relative to the reference IRS element. 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 respectivelz. Fig. (2.4) describes this spherical cordinate system in more details. The joint points in Fig. (2.4) represents the IRS reflecting elements.

W.l.o.g. in the following context, we will place the IRS in the first quadrant of the XZ-plane. We will 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

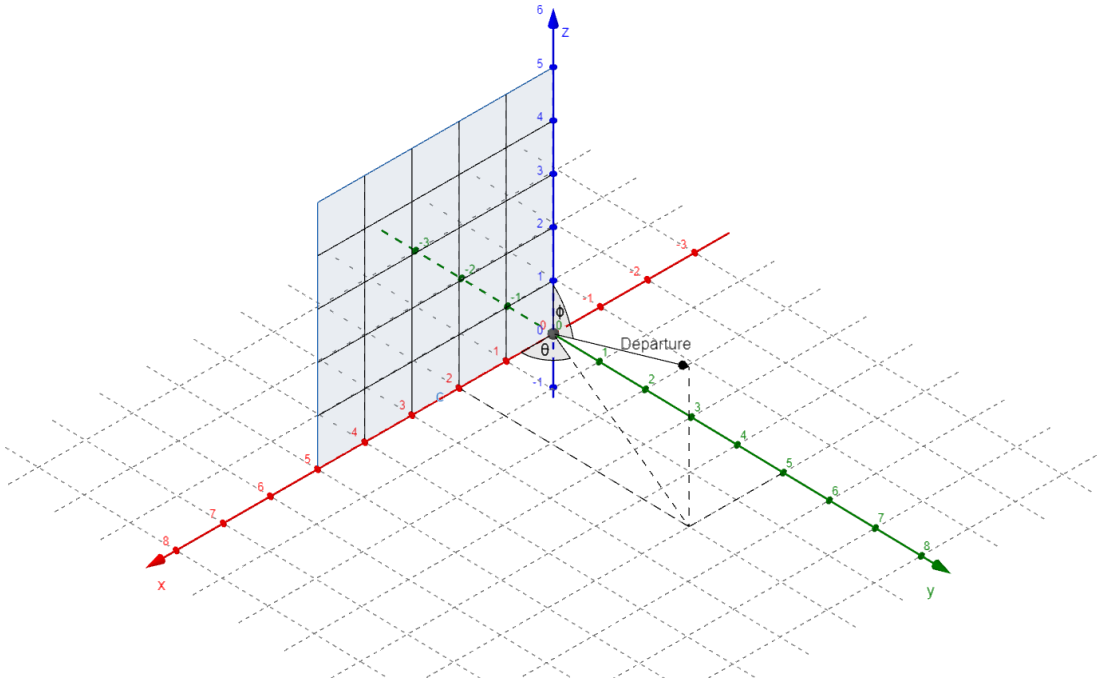


Figure 2.4: Angle of departure

as d. We will index the IRS elements from right to left, bottom to top. Based on these convensions, if $m \in \mathbb{N} \cap [1, M]$, and the m-th IRS element will have the coordinate

$$\begin{pmatrix} x = ((m-1) \bmod \sqrt{M}) \cdot d, \\ y = 0, \\ z = \lfloor (m-1)/\sqrt{M} \rfloor \cdot d \end{pmatrix}. \tag{2.6}$$

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 floor operator that evaluates to the largest interger that is smaller or equal to the number inside it.

Since the distance between the BS and the scattering cluster is often much larger than the wavelength, the departure beam from the any two IRS elements to the same scattering cluster can be considered as parallel rays. The phase shifts of the steering vector depend only on the relative positions of the IRS elements and the reference IRS element. In this article, we define the IRS element located at $(0,0,0)$ as the reference element. The steering vector can then be obtained from the following fig. (2.5).

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

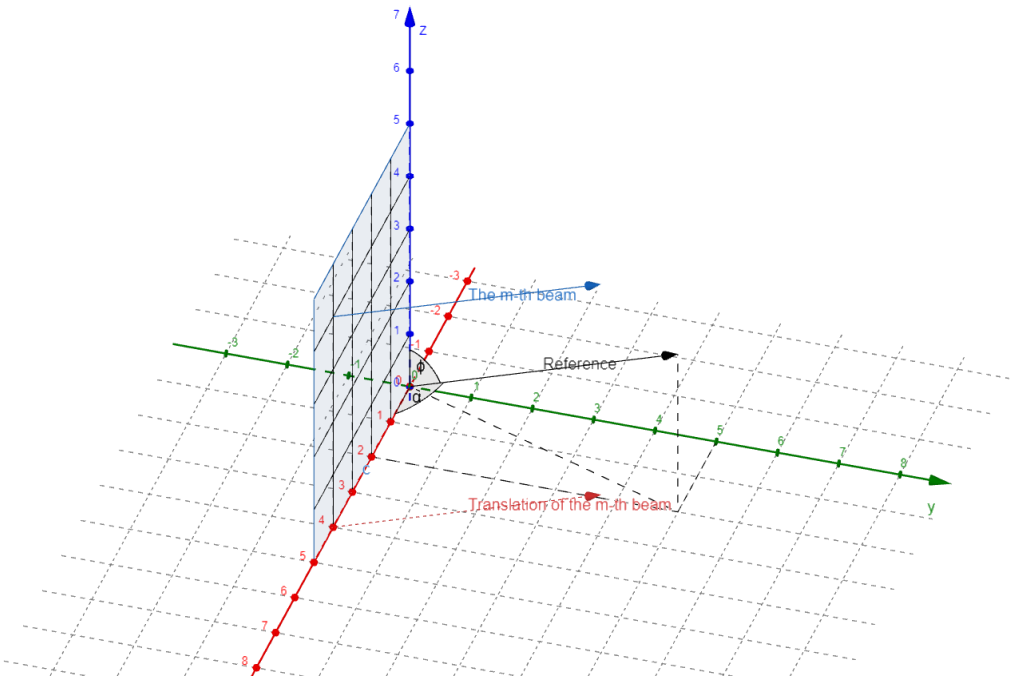


Figure 2.5: Length of the m-th beams

between the translated beam and the beam of the m-th element depends on the angle ϕ. Thus, we can get the following equations

$$\begin{aligned} & l_{\text{ref}} - l_{\text{tr},m} = d_{\text{x}} \cos(\alpha) = d_{\text{x}} \sin(\phi)\cos(\theta) \\ & l_{\text{tr},m} - l_m = d_{\text{z}} \cos(\phi) \\ \Rightarrow \\ & l_{\text{ref}} - l_m = d_{\text{x}} \sin(\phi)\cos(\theta) + d_{\text{z}} \cos(\phi), \end{aligned} \tag{2.7}$$

where l_{ref}, $l_{\text{tr},m}$ and l_m are the beam length from the IRS to the scattering object of the reference beam, the translated beam of the m-th element and the beam of the m-th element 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. (2.5), (2.6) and (2.7), the steering vector $\mathbf{h}_t^T(\theta_p^t, \phi_p^t)$ can be written as

$$\mathbf{h}_t(\theta_p^t, \phi_p^t) = \left[e^{j\frac{2\pi d}{\lambda_w}[((m-1) \bmod \sqrt{M}) \sin(\phi_p^t)\cos(\theta_p^t) + \lfloor (m-1)/\sqrt{M} \rfloor \cos(\phi_p^t)]} \right]_{m,1}. \tag{2.8}$$

Here, λ_w denotes the wavelength of the transmittion signal.

In Eq. (2.5), the path loss and the fading is included in the coefficient h_p. According to [10], 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

$$\begin{aligned} \bar{h}_k &= \left(\frac{\lambda_w}{4\pi l_{\text{ref}}}\right)^n \\ \tilde{h}_p &\sim \text{CN}(0, \sigma_f^2). \end{aligned} \tag{2.9}$$

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

2.3 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 XZ-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 XZ-plane is denoted as R_{d}. This setup of the active antennas relative to the IRS is show as Fig. (2.6).

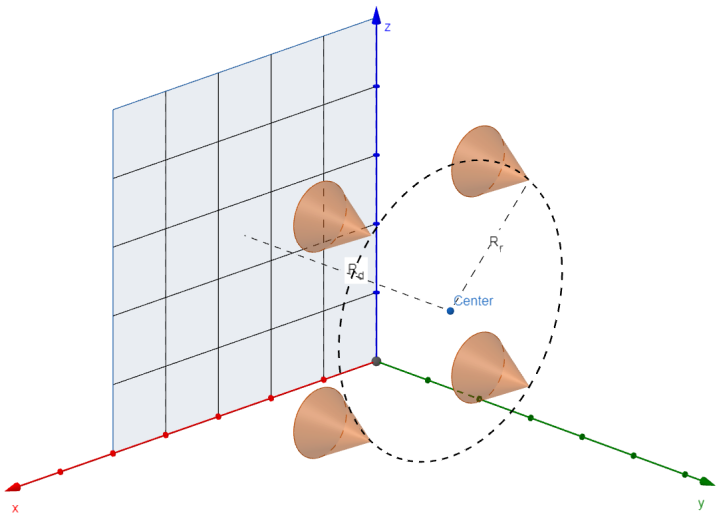


Figure 2.6: Placement of the active antennas

If we look at the antenna array from an angle such that the x-axis is pointing to the right and the z-axis is pointing up. The index of the active antennas are ordered according to the directed angle ranging in $[0, 2\pi]$ counterclock-wise. The active antennas with smaller angles are ordered before the antennas with larger angles. In Fig. (2.6), 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 will use a spherical/cartitian coordinate system originated at the active antenna to be studied with the following constraints. The beam direction from the active antenna 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. (2.7).

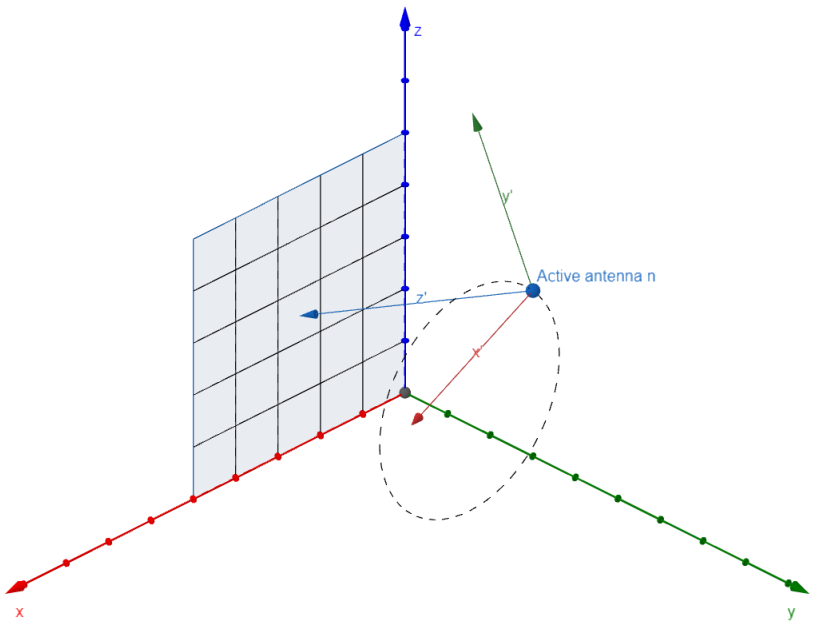


Figure 2.7: Auxilary coordination system of an active antenna

Under this $x'y'z'$-coordination system, the position of the m-th IRS element can also be represented by a spherical coordinatie 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 in the spherical coordinate originated at the active antenna to be studied. Fig. (2.8) illustrates an example coordinate of the m-th IRS element under the auxilary coordination system originated at 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, to make the discussion easier, we denote both of distance as $r_{m,n}$.

Based on the above discussion, and following [10], the matrix $\mathbf{T}$ is given by

$$\mathbf{T} = \left[\frac{\lambda_w \sqrt{\rho G^a(\theta^p_{m,n}, \phi^p_{m,n}) G^p(\theta^a_{m,n}, \phi^a_{m,n})}}{4\pi r_{m,n}} e^{-j\frac{2\pi r_{m,n}}{\lambda_w}} \right]_{m,n}, \tag{2.10}$$

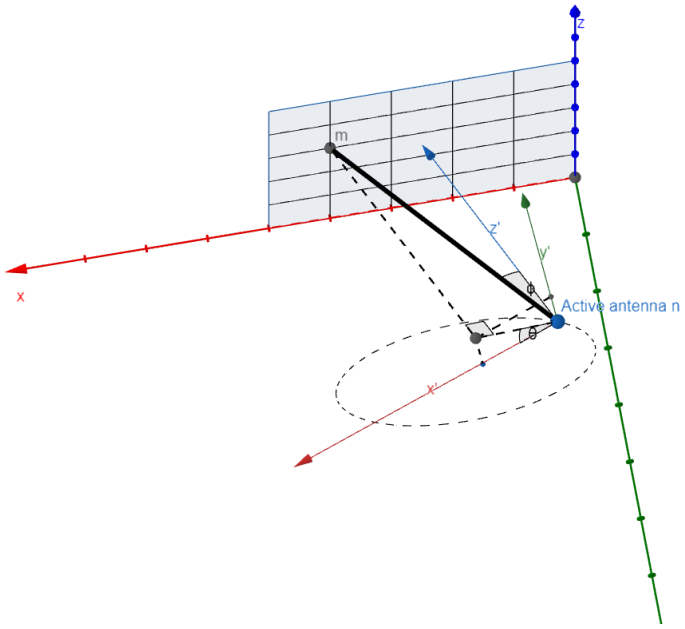


Figure 2.8: The coordination of the m-th IRS element

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 as well as the antenna gains of the different IRS elements. But the antenna pattern between the passive antennas and the active antennas may not be the same. By analysing Eq. (2.10), 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 [10].

In full illumination (FI), 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.

In partial illumination (PI), 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 the interference if the antenna pattern is too wide or the distance between the active antennas and the IRS is too large.

In seperate illumination (SI), 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. (2.9) to Fig. (2.11). The blue arrows indicate the orientations of the active antennas.

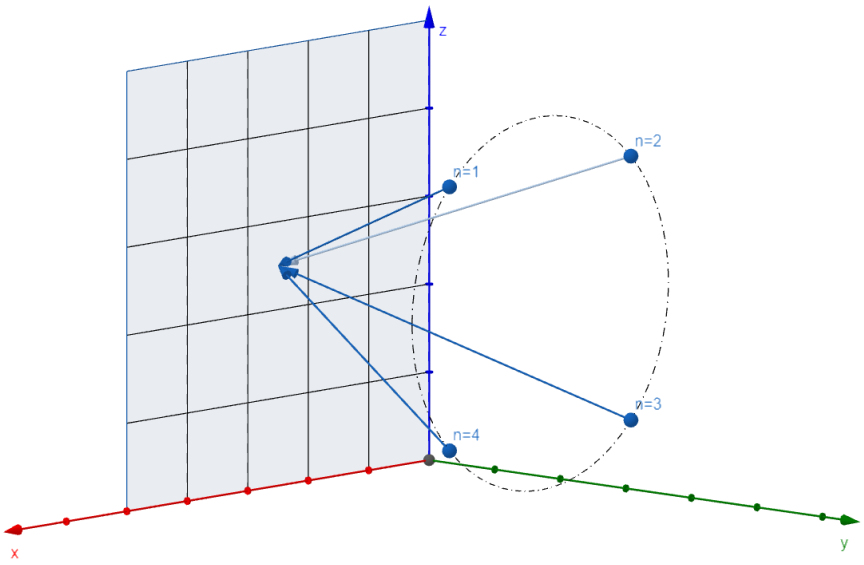


Figure 2.9: Full illumination

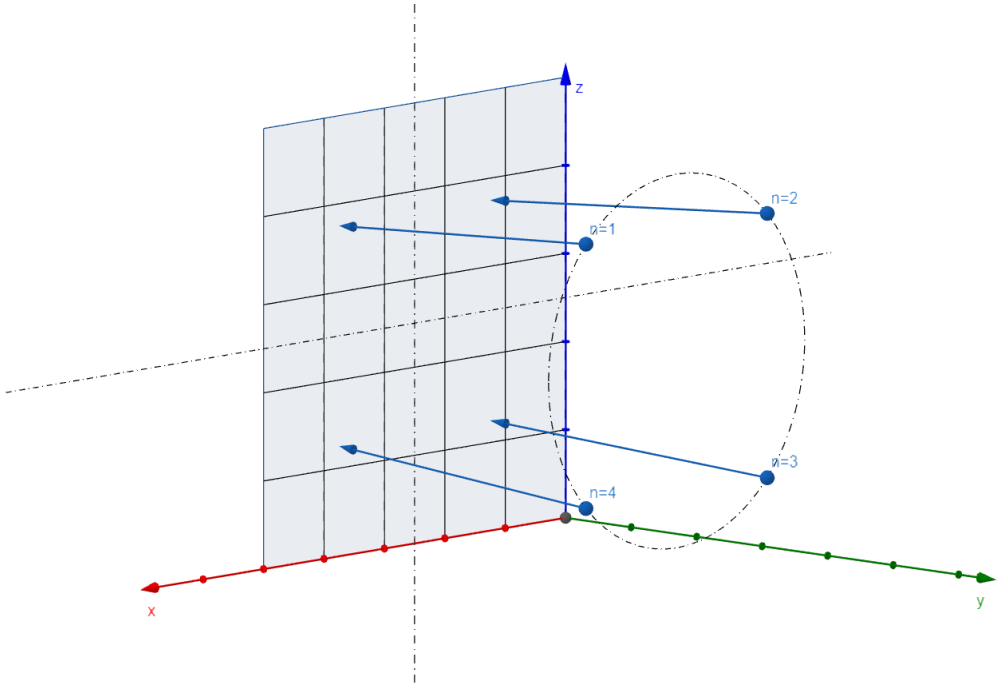


Figure 2.10: Partial illumination

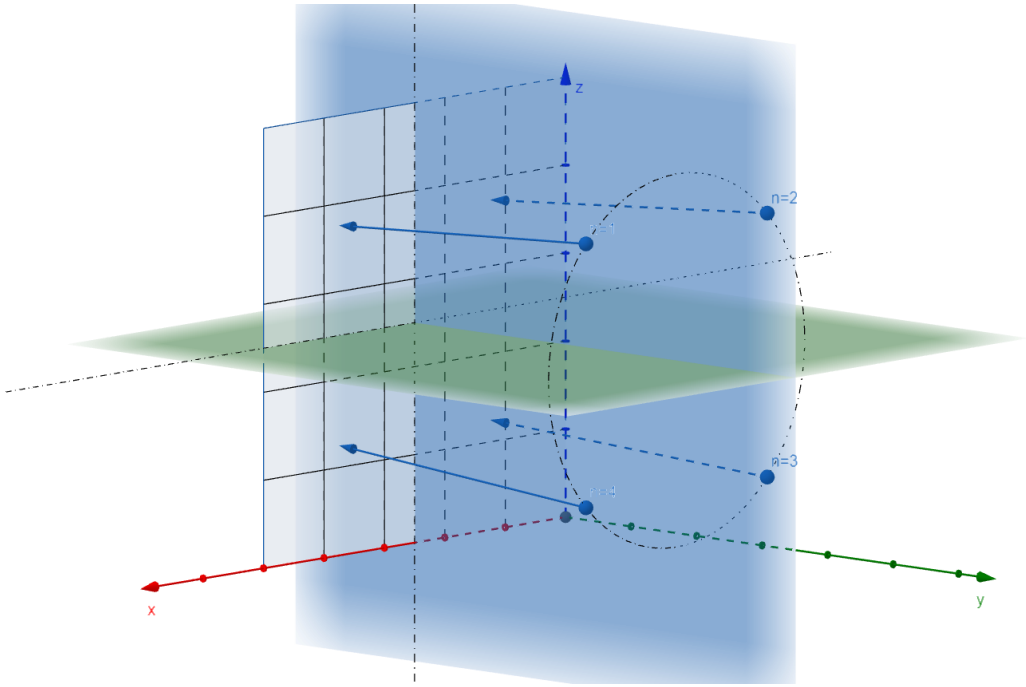


Figure 2.11: Seperate illumination

Chapter 3

Problem Formulation

3.1 Performance Measure

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

$$\mathrm{d}(\mathbf{S}, \tilde{\mathbf{Y}}) = ||\tilde{\mathbf{Y}} - \mathbf{S}||_{\mathrm{fro}}^2 = \mathrm{tr}\{(\tilde{\mathbf{Y}} - \mathbf{S})^H(\tilde{\mathbf{Y}} - \mathbf{S})\}, \tag{3.1}$$

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

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.

3.2 Near-Far Effect and Decoder

The UTs often have different distances to the BS. As a result, the path loss factor $\bar{h}_k$ of different UTs are different and their values are often very small because of the exponential decay based on the distance. In this article, we assume that the relative distances between UTs and BS 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. W.l.o.g. define $\mathbf{P}_L^{1/2} \in \mathbb{C}^{K \times K}$ to be a diagonal path loss matrix

$$\mathbf{P}_L^{1/2} = \begin{bmatrix} \sqrt{\bar{h}_1} & & \\ & \ddots & \\ & & \sqrt{\bar{h}_K} \end{bmatrix}. \tag{3.2}$$

The decoder can then be designed as

$$\tilde{\mathbf{Y}} = \mathbf{P}_L^{-1/2}\mathbf{Y}. \tag{3.3}$$

3.3 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,

$$\mathbf{H} = \hat{\mathbf{H}} + \mathbf{e}, \tag{3.4}$$

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. (3.5), which are known by the transmitter.

$$\begin{aligned} \mathscr{E}\{e_{k,l}\} &= 0; \\ \mathscr{E}\{e_{k,l}^* e_{k,l}\} &= \sigma_e^2. \end{aligned} \tag{3.5}$$

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

$$\begin{aligned} \mathbf{Y} &= (\hat{\mathbf{H}} + \mathbf{e})\mathbf{DTX} + \mathbf{N} \\ \tilde{\mathbf{Y}} &= \mathbf{P}_L^{-1/2}[(\hat{\mathbf{H}} + \mathbf{e})\mathbf{DTX} + \mathbf{N}]. \end{aligned} \tag{3.6}$$

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. (3.7).

$$\begin{aligned} \mathbf{H} &= \mathbf{P}_L^{1/2}\tilde{\mathbf{H}} \\ \hat{\mathbf{H}} &= \mathbf{P}_L^{1/2}\tilde{\hat{\mathbf{H}}} \\ \mathbf{e} &= \mathbf{P}_L^{1/2}\tilde{\mathbf{e}} \\ \mathbf{N} &= \mathbf{P}_L^{1/2}\tilde{\mathbf{N}} \end{aligned} \tag{3.7}$$

By substituting Eq. (3.1) with Eq. (3.6), we can observe that the metric equation Eq. (3.1) contains a noise term and an estimation error term which are both random and unobtainable at both the BS and UT side. Therefore, the precoding algorithm can

only be designed base on the statistics of Eq. (3.1). Assume that the noise entries are independed with the estimation errors, we then have

$$\mathscr{E}_{\mathbf{N},\mathbf{e}}\{\mathrm{d}(\mathbf{S},\tilde{\mathbf{Y}})\} = \mathscr{E}_{\mathbf{e}} \cdot \mathscr{E}_{\mathbf{N}}\{\mathrm{d}(\mathbf{S},\tilde{\mathbf{Y}})\} \tag{3.8}$$

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

$$\begin{aligned}\mathscr{E}_{\mathbf{N}}\{\mathrm{d}(\mathbf{S},\tilde{\mathbf{Y}})\} =& \mathrm{tr}\{(\tilde{\mathbf{H}}\mathbf{DTX}-\mathbf{S})^H(\tilde{\mathbf{H}}\mathbf{DTX}-\mathbf{S})\} + \mathscr{E}_{\mathbf{N}}\{\mathrm{tr}\{\tilde{\mathbf{N}}^H\tilde{\mathbf{N}}\}\} \\ =& \mathrm{tr}\{(\tilde{\mathbf{H}}\mathbf{DTX}-\mathbf{S})^H(\tilde{\mathbf{H}}\mathbf{DTX}-\mathbf{S})\} + L\mathrm{tr}\{\mathbf{P}_L^{-1}\}\sigma_n^2.\end{aligned} \tag{3.9}$$

By performing the similar operation one more time, the expectation of Eq. (3.9) with respect to estimation error $\mathbf{e}$ is given by

$$\begin{aligned}&\mathscr{E}_{\mathbf{e}}\left\{\mathscr{E}_{\mathbf{N}}\{\mathrm{d}(\mathbf{S},\tilde{\mathbf{Y}})\}\right\} \\ =&\mathscr{E}_{\mathbf{e}}\{\mathrm{tr}\{((\tilde{\tilde{\mathbf{H}}}+\tilde{\mathbf{e}})\mathbf{DTX}-\mathbf{S})^H((\tilde{\tilde{\mathbf{H}}}+\tilde{\mathbf{e}})\mathbf{DTX}-\mathbf{S})\}\} + L\mathrm{tr}\{\mathbf{P}_L^{-1}\}\sigma_n^2 \\ =&\left\|\tilde{\tilde{\mathbf{H}}}\mathbf{DTX}-\mathbf{S}\right\|_{\mathrm{fro}}^2 + \mathrm{tr}\{\mathbf{X}^H\mathbf{T}^H\mathbf{D}^H\mathbf{\Sigma}_{\tilde{\mathbf{e}}}\mathbf{DTX}\} + L\mathrm{tr}\{\mathbf{P}_L^{-1}\}\sigma_n^2,\end{aligned} \tag{3.10}$$

where

$$\mathbf{\Sigma}_{\tilde{\mathbf{e}}} = \begin{bmatrix}\mathrm{tr}\{\mathbf{P}_L^{-1}\}\sigma_e^2 & & \\ & \ddots & \\ & & \mathrm{tr}\{\mathbf{P}_L^{-1}\}\sigma_e^2\end{bmatrix} \tag{3.11}$$

The formular given in Eq. (3.10) is the expected distance between $\mathbf{S}$ and $\tilde{\mathbf{Y}}$, i.e. $\mathscr{E}_{\mathbf{N},\mathbf{e}}\{\mathrm{d}(\mathbf{S},\tilde{\mathbf{Y}})\}$. From now on, instead of developing an algorithm that minimize Eq. (3.1) directly, we try to develop an algorithm that minimize Eq. (3.10) by designing $\mathbf{D}$ and $\mathbf{X}$.

3.4 Problem Formulization as a GLSE Precoder

Following [11], 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

$$\begin{aligned}\underset{\mathbf{D},\mathbf{X}}{\text{minimize}} \quad & \left\|\tilde{\tilde{\mathbf{H}}}\mathbf{DTX}-\mathbf{S}\right\|_{\mathrm{fro}}^2 + \mathrm{tr}\{\mathbf{X}^H\mathbf{T}^H\mathbf{D}^H\mathbf{\Sigma}_{\tilde{\mathbf{e}}}\mathbf{DTX}\} + \lambda\mathrm{tr}\{\mathbf{X}\mathbf{X}^H\} \\ \text{subject to} \quad & \max_{i,j}|x_{i,j}|^2 < P_{\mathrm{peak}},\end{aligned} \tag{3.12}$$

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. (3.12) as

$$\text{obj}(\mathbf{D},\mathbf{X},\mathbf{S}) = \left|\left|\tilde{\tilde{\mathbf{H}}}\mathbf{DTX} - \mathbf{S}\right|\right|_{\text{fro}}^2 + \text{tr}\{\mathbf{X}^H\mathbf{T}^H\mathbf{D}^H\boldsymbol{\Sigma}_{\bar{\mathbf{e}}}\mathbf{DTX}\} + \lambda\text{tr}\{\mathbf{X}\mathbf{X}^H\} \tag{3.13}$$

If we neglect the power constraints by setting λ to 0, and letting P_{peak} tend to infinity, the optimization problem stated in Eq. (3.12) is equavalent to the direct optimization of Eq. (3.10). The term $L\text{tr}\{\mathbf{P}_L^{-1}\}\sigma_n^2$ is removed from the objective function, because it is a constant term that depends neither on $\mathbf{D}$ nor on $\mathbf{X}$.

Chapter 4

Precoding Methods

4.1 Joint Optimization

The joint optimization of the problem stated in (3.12) 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 shifts configuration for the IRS while treating $\mathbf{X}$ as a fixed matrix. The pseudo-code Algorithm (1) dipicts the general procedure.

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: $\quad \mathbf{X} \leftarrow \arg\min_{\mathbf{X}} \text{obj}(\mathbf{D}, \mathbf{X})$
4: $\quad \mathbf{D} \leftarrow \arg\min_{\mathbf{D}} \text{obj}(\mathbf{D}, \mathbf{X})$
5: **end while**

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 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 $\mathbf{D}$ is given.

The domain of $\mathbf{D}$ 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 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.

4.2 Optimization of the Digital Unit

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

$$\begin{aligned} \underset{\mathbf{X}}{\text{minimize}} \quad & \left|\left|\tilde{\tilde{\mathbf{H}}}\mathbf{DTX}-\mathbf{S}\right|\right|_{\text{fro}}^2+\text{tr}\{\mathbf{X}^H\mathbf{T}^H\mathbf{D}^H\boldsymbol{\Sigma}_{\tilde{\mathbf{e}}}\mathbf{DTX}\}+\lambda\text{tr}\{\mathbf{X}\mathbf{X}^H\} \\ \text{subject to} \quad & \max_{i,j}|x_{i,j}|^2<P_{\text{peak}} \end{aligned} \tag{4.1}$$

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. (4.2).

$$\begin{aligned} &\text{tr}\{(\tilde{\tilde{\mathbf{H}}}\mathbf{DTX}-\mathbf{S})^H(\tilde{\tilde{\mathbf{H}}}\mathbf{DTX}-\mathbf{S})\}+\text{tr}\{\mathbf{X}^H\mathbf{T}^H\mathbf{D}^H\boldsymbol{\Sigma}_{\tilde{\mathbf{e}}}\mathbf{DTX}\}+\lambda\text{tr}\{\mathbf{X}^H\mathbf{X}\} \\ &=\sum_{i=1}^{L}(\tilde{\tilde{\mathbf{H}}}\mathbf{DT}\boldsymbol{x_i}-\boldsymbol{s_i})^H(\tilde{\tilde{\mathbf{H}}}\mathbf{DT}\boldsymbol{x_i}-\boldsymbol{s_i})+\mathbf{x}_i^H\mathbf{T}^H\mathbf{D}^H\boldsymbol{\Sigma}_{\tilde{\mathbf{e}}}\mathbf{DTx}_i+\lambda\boldsymbol{x_i}^H\boldsymbol{x_i}. \end{aligned} \tag{4.2}$$

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

4.3 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, when designing the IRS configuration in iterative method, $\mathbf{X}$ is independent to $\mathbf{D}$. The optimization problem is then formulized as

$$\underset{\mathbf{D}}{\text{minimize}} \quad \left|\left|\tilde{\tilde{\mathbf{H}}}\mathbf{DTX}-\mathbf{S}\right|\right|_{\text{fro}}^2+\text{tr}\{\mathbf{X}^H\mathbf{T}^H\mathbf{D}^H\boldsymbol{\Sigma}_{\tilde{\mathbf{e}}}\mathbf{DTX}\} \tag{4.3}$$

This problem is generally non-convex, since the domain of $(\mathbf{D})$ is not a convex set. Two methods are used here to obtain a suboptimal point, which are gradient descent and MM algorithm [12]. For the further discussion, we denote the objective function in Eq. (4.3) by

$$\text{obj}_{\mathbf{D}}(\mathbf{D})=\left|\left|\tilde{\tilde{\mathbf{H}}}\mathbf{DTX}-\mathbf{S}\right|\right|_{\text{fro}}^2+\text{tr}\{\mathbf{X}^H\mathbf{T}^H\mathbf{D}^H\boldsymbol{\Sigma}_{\tilde{\mathbf{e}}}\mathbf{DTX}\}. \tag{4.4}$$

4.3.1 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. (4.3) 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

$$\boldsymbol{\beta} = \arg \cdot \mathrm{diag}_{\mathrm{vector}}(\mathbf{D}). \tag{4.5}$$

To keep the notation simple, we define

$$\begin{aligned} \mathbf{U} &= \mathbf{DTX}; \\ \mathbf{V} &= \mathbf{S}^H \tilde{\tilde{\mathbf{H}}} \\ \mathbf{Q} &= \tilde{\tilde{\mathbf{H}}}^H \tilde{\tilde{\mathbf{H}}} + \boldsymbol{\Sigma}_{\tilde{\tilde{\mathbf{e}}}}. \end{aligned} \tag{4.6}$$

Under this convention, the objective function can be expanded as

$$\mathrm{obj}_{\mathbf{D}}(\mathbf{D}) = tr\{\mathbf{U}^H\mathbf{Q}\mathbf{U} - (\mathbf{U}^H\mathbf{V}^H + \mathbf{V}\mathbf{U}) + \mathbf{S}^H\mathbf{S}\}. \tag{4.7}$$

The partial derivertive is a linear transformation, and the derivertive of the additive terms in Eq. (4.7) can be obtained as

$$\begin{aligned} \frac{\partial\, tr\{\mathbf{U}^H\mathbf{Q}\mathbf{U}\}}{\partial \boldsymbol{\beta}} &= -2\mathrm{diag}_{\mathrm{vector}}(\Im(\mathbf{U}\mathbf{U}^{\mathbf{H}}\mathbf{Q})) \\ \frac{\partial\, tr\{\mathbf{U}^H\mathbf{V}^H + \mathbf{V}\mathbf{U}\}}{\partial \boldsymbol{\beta}} &= -2\mathrm{diag}_{\mathrm{vector}}(\Im(\mathbf{U}\mathbf{V})). \end{aligned} \tag{4.8}$$

From Eq. (4.7) and Eq. (4.8), the partial derivertive of the objective function can be derived as

$$\frac{\partial \mathrm{obj}_{\mathbf{D}}(\mathbf{D})}{\partial \boldsymbol{\beta}} = -2\mathrm{diag}_{\mathrm{vector}}(\Im(\mathbf{U}\mathbf{U}^{\mathbf{H}}\mathbf{Q} - \mathbf{U}\mathbf{V})). \tag{4.9}$$

The auxilary vector can be transformed back to the phase shift matrix by Eq. (4.10).

$$\mathbf{D} = \mathrm{diag}_{\mathrm{matrix}}\left(\exp(j\boldsymbol{\beta})\right) \tag{4.10}$$

Based on the above discussion, when 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 avoid the situation where the algorithm diverge, one can design the step

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

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. 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 $(\mathbf{D}, \mathbf{X})$ will converge to the same value.

4.3.2 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 [12].

The MM algorithm is like a variant 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. (4.6), we have

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

By substituting the variables according to Eq. (4.6), the quadratic term can be written as

$$\mathbf{X}^H\mathbf{T}^H\mathbf{D}^H\left(\tilde{\tilde{\mathbf{H}}}^H\tilde{\tilde{\mathbf{H}}} + \boldsymbol{\Sigma}_{\tilde{\tilde{\mathbf{e}}}}\right)\mathbf{DTX} = \mathbf{U}^H\mathbf{QU}. \tag{4.11}$$

Two surrogate functions are additive if they are both tangent to the objective function at the same point. The surrogate function of the objective function at any point $\mathbf{D}_0$ can then be obtained by constructing a surrogate function for the quadratic term at point $\mathbf{D}_0$ and we add this partial surrogate function with the first order terms.

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

$$\begin{aligned}&\mathrm{tr}\left\{\mathbf{U}^H\mathbf{QU}\right\}\\ =&\mathrm{tr}\left\{\mathbf{U}_0^H\mathbf{QU}_0 + (\mathbf{U}-\mathbf{U}_0)^H\mathbf{QU}_0 + \mathbf{U}_0^H\mathbf{Q}(\mathbf{U}-\mathbf{U}_0) + (\mathbf{U}-\mathbf{U}_0)^H\mathbf{Q}(\mathbf{U}-\mathbf{U}_0)\right\}\\ \leq&\mathrm{tr}\left\{\mathbf{U}_0^H\mathbf{QU}_0 + (\mathbf{U}-\mathbf{U}_0)^H\mathbf{QU}_0 + \mathbf{U}_0^H\mathbf{Q}(\mathbf{U}-\mathbf{U}_0) + (\mathbf{U}-\mathbf{U}_0)^H\mathbf{P}(\mathbf{U}-\mathbf{U}_0)\right\}\\ =&\mathrm{tr}\left\{\mathbf{U}^H\mathbf{PU} + \mathbf{U}^H(\mathbf{Q}-\mathbf{P})\mathbf{U}_0 + \mathbf{U}_0^H(\mathbf{Q}-\mathbf{P})\mathbf{Y} + \mathbf{U}_0^H(\mathbf{Q}-\mathbf{P})\mathbf{U}_0\right\}\end{aligned} \tag{4.12}$$

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. (4.4). 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

$$\overline{\mathrm{obj}_{\mathbf{D}_0}(\mathbf{D})} = \mathrm{tr}\left\{\mathbf{U}^H(\mathbf{Q}-\mathbf{P})\mathbf{U}_0 + \mathbf{U}_0^H(\mathbf{Q}-\mathbf{P})\mathbf{U} - \mathbf{VU} - \mathbf{U}^H\mathbf{V}^H\right\}. \tag{4.13}$$

For simplicity, define a constant

$$\mathbf{G}_0 = \mathbf{TXU}_0^H(\mathbf{Q}-\mathbf{P}) - \mathbf{TXV}. \tag{4.14}$$

By substituting Eq. (4.14) into Eq. (4.13), the latter becomes

$$\overline{\text{obj}_{\mathbf{D}_0}(\mathbf{D})} = \text{tr}\{\mathbf{G}_0\mathbf{D} + \mathbf{D}^H\mathbf{G}_0^H\} \tag{4.15}$$

According to previous discussion 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. (4.15). By introducing two auxilary constant terms $\mathbf{tr}\{\mathbf{D}^H\mathbf{D}\}$ and $tr\{\mathbf{G}_0\mathbf{G}_0^H\}$, we have

$$\begin{aligned}
\mathbf{D}_1 &= \arg\min_{\mathbf{D}}(\text{tr}\{\mathbf{G}_0\mathbf{D} + \mathbf{D}^H\mathbf{G}_0^H\}) \\
&= \arg\min_{\mathbf{D}}(\text{tr}\{\mathbf{G}_0\mathbf{D} + \mathbf{D}^H\mathbf{G}_0^H + \mathbf{D}^H\mathbf{D} + \mathbf{G}_0\mathbf{G}_0^H\}) \\
&= \arg\min_{\mathbf{D}}(\text{tr}\{(\mathbf{D} - (-\mathbf{G}_0)^H)^H(\mathbf{D} - (-\mathbf{G}_0)^H)\}) \\
&= \arg\min_{\mathbf{D}}(||\mathbf{D} - (-\mathbf{G}_0)||_{fro}^2) \\
&= \exp((\arg(-\mathbf{G}_0))
\end{aligned} \tag{4.16}$$

The complete psuedo-code for the precoding with MM algorithm is illustrated as Algorithm (3). The convergence of Algorithm (3) is also guarenteed. Because of the

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

1: $(\mathbf{D}, \mathbf{X}) \leftarrow$ initial values
2: **while** terminate condition not met **do**
3: $\quad\mathbf{X} \leftarrow \text{convex}_{\text{opt}}\{\left|\left|\tilde{\tilde{\mathbf{H}}}\mathbf{DTX} - \mathbf{S}\right|\right|_{\text{fro}}^2 + \text{tr}\{\mathbf{X}^H\mathbf{T}^H\mathbf{D}^H\mathbf{\Sigma}_{\tilde{\mathbf{e}}}\mathbf{DTX}\} + \lambda\text{tr}\{\mathbf{XX}^H\}\}$
4: $\quad$**while** terminate condition not met **do**
5: $\quad\quad\mathbf{D}_0 \leftarrow \mathbf{D}$
6: $\quad\quad\mathbf{U}_0 \leftarrow \mathbf{D}_0\mathbf{TX}$
7: $\quad\quad\mathbf{V} \leftarrow \mathbf{S}^H\tilde{\tilde{\mathbf{H}}}$
8: $\quad\quad\mathbf{Q} \leftarrow \tilde{\tilde{\mathbf{H}}}^H\tilde{\tilde{\mathbf{H}}} + \mathbf{\Sigma}_{\tilde{\mathbf{e}}}$
9: $\quad\quad\mathbf{P} \leftarrow \lambda_{max}^{\mathbf{Q}}\mathbf{I}$
10: $\quad\quad\mathbf{G}_0 \leftarrow \mathbf{TXU}_0^H(\mathbf{Q}-\mathbf{P}) - \mathbf{TXV}$
11: $\quad\quad\mathbf{D}_1 \leftarrow \exp((\arg(-\mathbf{G}_0))$
12: $\quad\quad\mathbf{D} \leftarrow \mathbf{D}_1$
13: $\quad$**end while**
14: **end while**

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 the IRS optimization. Thus, Algorithm (3) also converge.

Chapter 5

Numerical Simulations

In this chapter, we are going to study the performance of the IRS-aided precoding procedure discussed above. It can be observed that during the optimization process, the algorithm is designed based on Eq. (3.12) which is independent of channel-wise noise variance and channel-wise estimation error variance. Therefore, in the simulations, for simplicity, we will use maximal normalized variance for both noise and estimation error.

5.1 Performance Measure

It is proven in [9] that by using Jensen's inequality, the average rate per user is bounded by

$$\bar{R} \geq \frac{LT_s}{\tau + LT_s} \log\left(1 + \frac{\sigma_s^2}{\sigma_{\tilde{n}_{\max}}^2 + P_{\text{Interference}}}\right), \tag{5.1}$$

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

Another performance measure 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.

5.2 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 have the same power, which is denoted as $\sigma_s^2 = 1$. 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

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 5.1: Parameters for channel $\mathbf{H}$

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

Table 5.2: Parameters for fading and path loss

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 setting for the channel is concluded as Table (5.1).

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

$$G^p(\theta, \phi) = \begin{cases} 2 & \theta, \phi \in [0, \pi] \\ 0 & \text{otherwise} \end{cases}. \tag{5.2}$$

The active antenna gain following [10] is assumed to be

$$G^a(\theta, \phi) = \begin{cases} 2(1+\kappa)\cos^\kappa(\phi) & \phi \in [0, \pi/2] \\ 0 & \text{otherwise} \end{cases}. \tag{5.3}$$

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 are concluded in Table (5.2) [10]. We also set the maximal normalized noise to be $\sigma^2_{\tilde{n}_{\max}} = 0.1$

5.3 Bechmark Method

We use a fully-connected analog beamforming network (FC-ABFN) as a benchmark structure [13]. The structure before and including 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 added with the corresponding signal from other groups with a power combiner. More specifically, for a fully connected precoder with N RF chains and M analog units the analog beamforming network is characterized as

$$\mathbf{F}_{\mathrm{RF}} = \mathbf{F}_{\mathrm{PC}}\mathbf{F}_{\mathrm{PS}}\mathbf{F}_{\mathrm{PD}} \tag{5.4}$$

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 shifts 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 [14]. Therefore, it can be modeled as a matrix with block diagonal structure [13]

$$\mathbf{F}_{\mathrm{PD}} = \sqrt{\frac{1}{L_{\mathrm{D}}M}}\begin{bmatrix} \mathbf{1}_M & \dots & \mathbf{0}_M \\ \vdots & \ddots & \vdots \\ \mathbf{0}_M & \dots & \mathbf{1}_M \end{bmatrix}, \tag{5.5}$$

where L_D represents the power loss in the power divider [14], 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

$$\mathbf{F}_{\mathrm{PC}} = \sqrt{\frac{1}{L_C N}}\begin{bmatrix} \mathrm{diag}(\mathbf{1}_M) & \dots & \mathrm{diag}(\mathbf{1}_M) \end{bmatrix}, \tag{5.6}$$

where L_C is the power loss introduced by the power combiner [13]. The channel can then be represented as

$$\mathbf{Y} = \mathbf{H}\mathbf{F}_{\mathrm{PC}}\mathbf{F}_{\mathrm{PS}}\mathbf{F}_{\mathrm{PD}}\mathbf{X} + \mathbf{N}. \tag{5.7}$$

By analizing Eq. (5.7) with Eq. (2.2), 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 so that we can have a fair comparison.

5.4 Rate and PAPR

By using the above mentioned setup, we compare the result for full illumination combined with gradient descent and MM algorithm and partial illumination combined with MM algorithm. Here, we assume that the etimation error is zero. 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. (5.1).

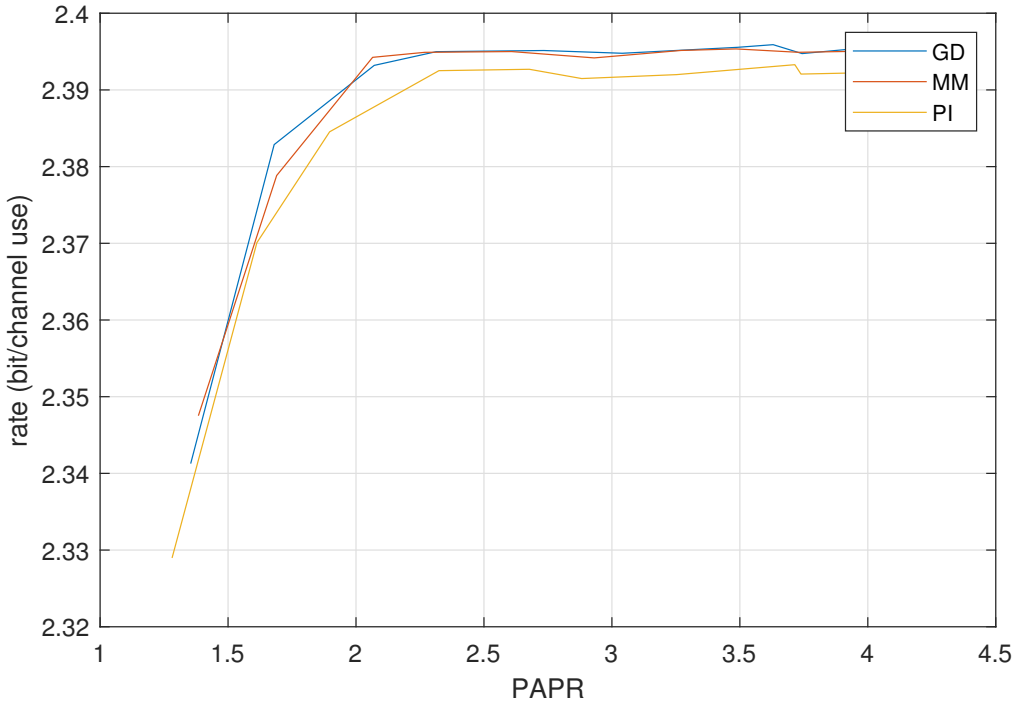


Figure 5.1: Rate-PAPR with block length $L = 1$, guard interval $\tau = 0$ and regularizer $\lambda = 0.5$

It can be observed by comparing the red line and the yellow line that partial illumination strategy performs slightly worse that 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.

For larger block length, we simulated the situations where $L = 2$ and $L = 4$ which are shown as Fig. (5.2) and Fig. (5.3).

It can be seen here that with guard interval $\tau = 0$, the systems with larger block lengths tend to have worse rate perform than the systems with smaller block lengths. If we compare two systems where one system has a block length larger than one the other has a block length equal to one, this can be explained by Eq. (4.2). For block length L_B, let's assume a fixed stream of data with length L_{tot} which is an integer multiple of it. Following Eq. (3.13), the cost function for the optimization over the whole stream can be written as

$$\underset{\mathbf{X}_{1\dots L_{tot}/L_B}, \mathbf{D}_{1\dots L_{tot}/L_B}}{\text{minimize}} \quad \sum_{j=1}^{L_{tot}/L_B} \text{obj}(\mathbf{D}_j, \mathbf{X}_j, \mathbf{S}_j) \tag{5.8}$$

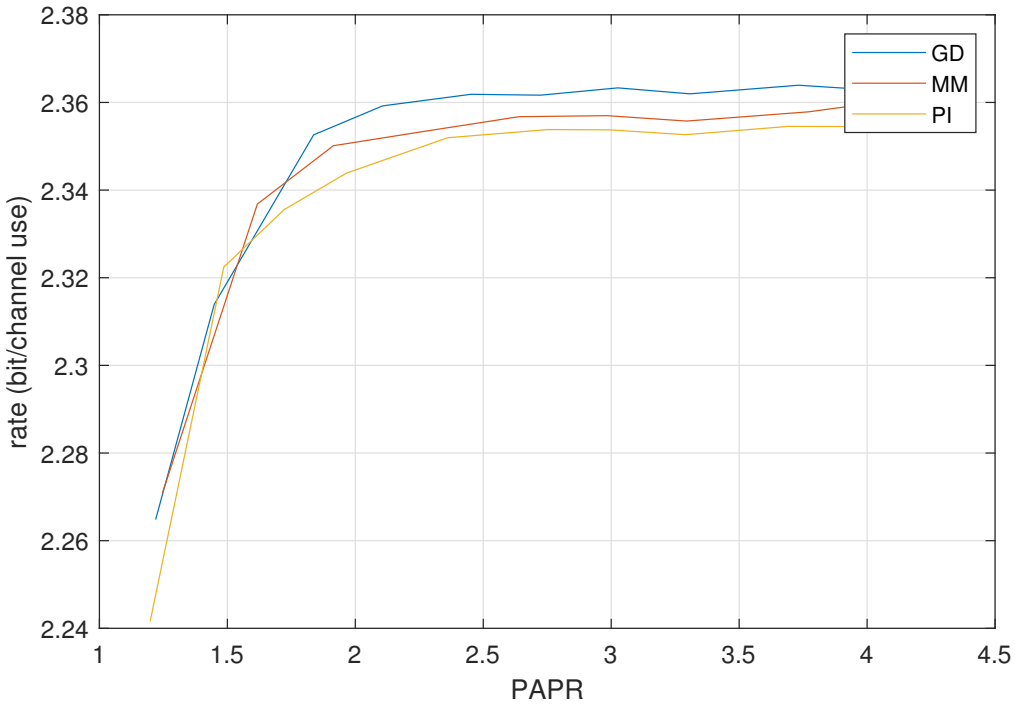


Figure 5.2: Rate-PAPR with block length $L = 2$, guard interval $\tau = 0$ and regularizer $\lambda = 0.5$

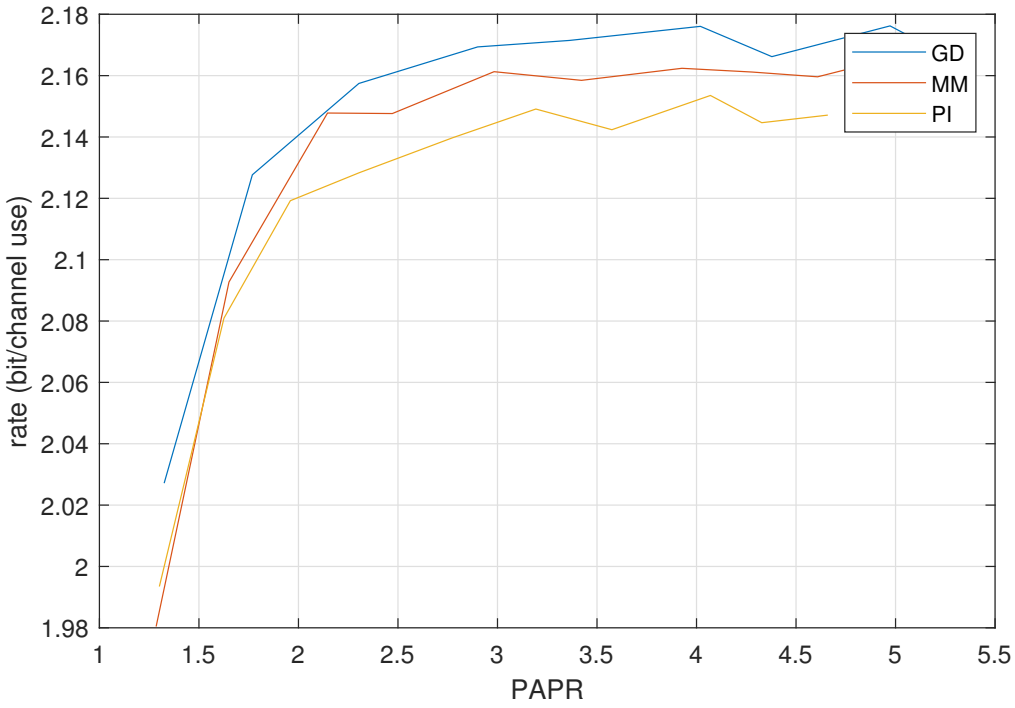


Figure 5.3: Rate-PAPR with block length $L = 4$, guard interval $\tau = 0$ and regularizer $\lambda = 0.5$

According to Eq. (4.2), the equation Eq. (5.8) is equivalent to

$$\underset{\mathbf{x}_{1\dots L_{tot}/L_B},\mathbf{D}_{1\dots L_{tot}/L_B}}{\text{minimize}} \quad \sum_{j=1}^{L_{tot}/L_B} \sum_{i=1}^{L_B} \text{obj}(\mathbf{D}_j, \mathbf{x}_{i\cdot j}, \mathbf{s}_{i\cdot j}) \tag{5.9}$$

where $\mathbf{x}_k$ and $\mathbf{s}_k$ denote the k-th data in the stream. From Eq. (5.9). We know that for all $L_B \in \mathbb{N}$, any solution of Eq. (5.9) with $L_B \geq 1$ lies in the feasible set of the same problem described by the same equation but with $L_B = 1$. Therefore, the symbol-wise transmission must performs better or equal to the block-wise transmission when the guard interval is zero. However, the this explaination can not be used to explain the rate difference between two strict block-wise precoding systems with different block lengths. Moreover, we can even construct a special data stream artificially 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. We can expain the phenomenon in the view of degrees of freedom. The larger the block length the less the degrees of freedom. In reality, the symbols and phase shifts we use are often discrete. Therefore, the signal stream range for the systems with smaller degrees of freedom to achieve zero interference is smaller than the range of the system with larger degrees of freedom. Thus, it is more probable for the systems with smaller block lengths to generate a data stream that is closer to the intended data stream.

For two systems with different block lengths, if one system has a block length that is an integer multiple of the other, we can then again 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. However, symbol-wise precoding also means 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. (5.4) and Fig. (5.5).

They both performs better than the IRS architecture. 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. Moreover, only 64 phase shifters are required in IRS instead of 256 phase shifters in fully connected HAD systems, which reduce the computational load greatly.

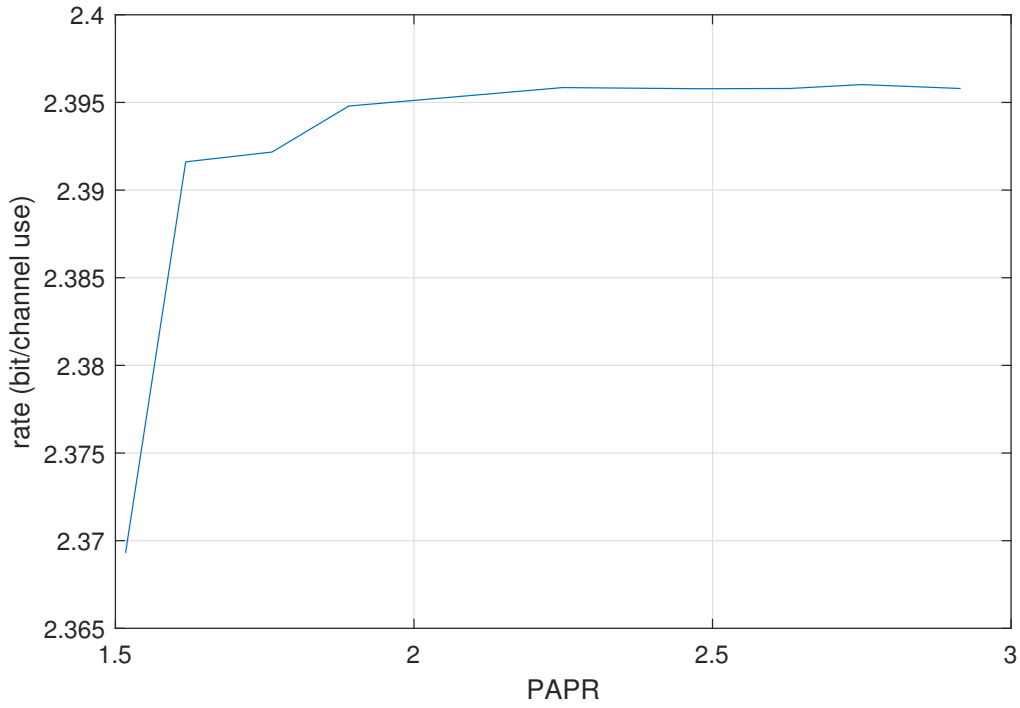


Figure 5.4: Rate-PAPR of fully connected system with block length $L = 2$, guard interval $\tau = 0$ and regularizer $\lambda = 0.5$

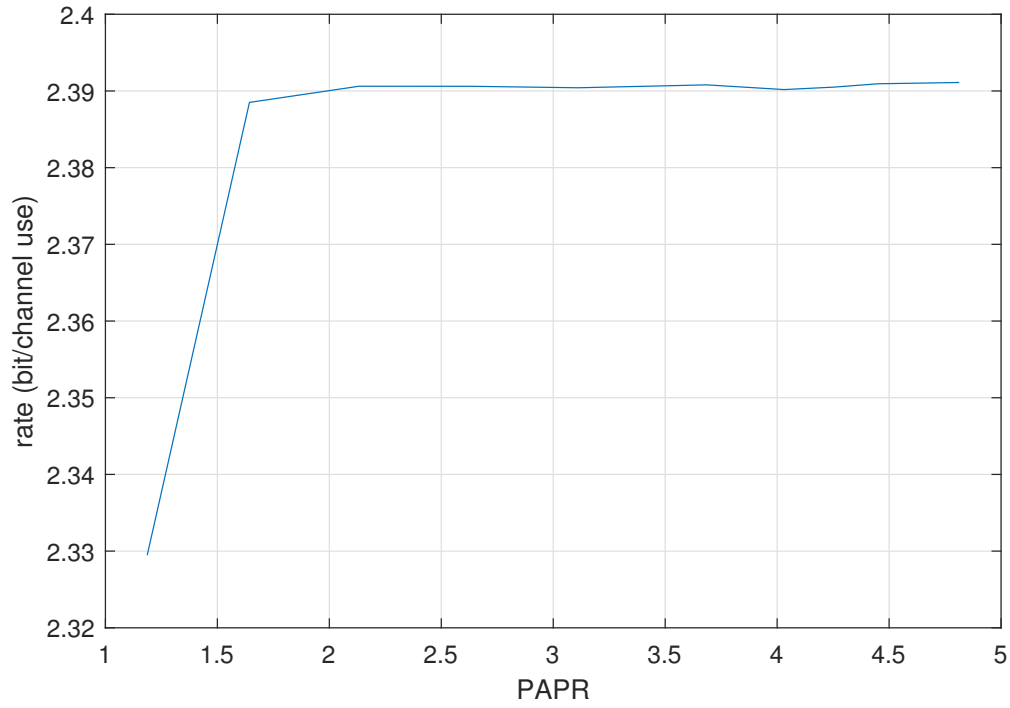


Figure 5.5: Rate-PAPR of fully connected system with block length $L = 4$, guard interval $\tau = 0$ and regularizer $\lambda = 0.5$

5.5 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. (5.9).

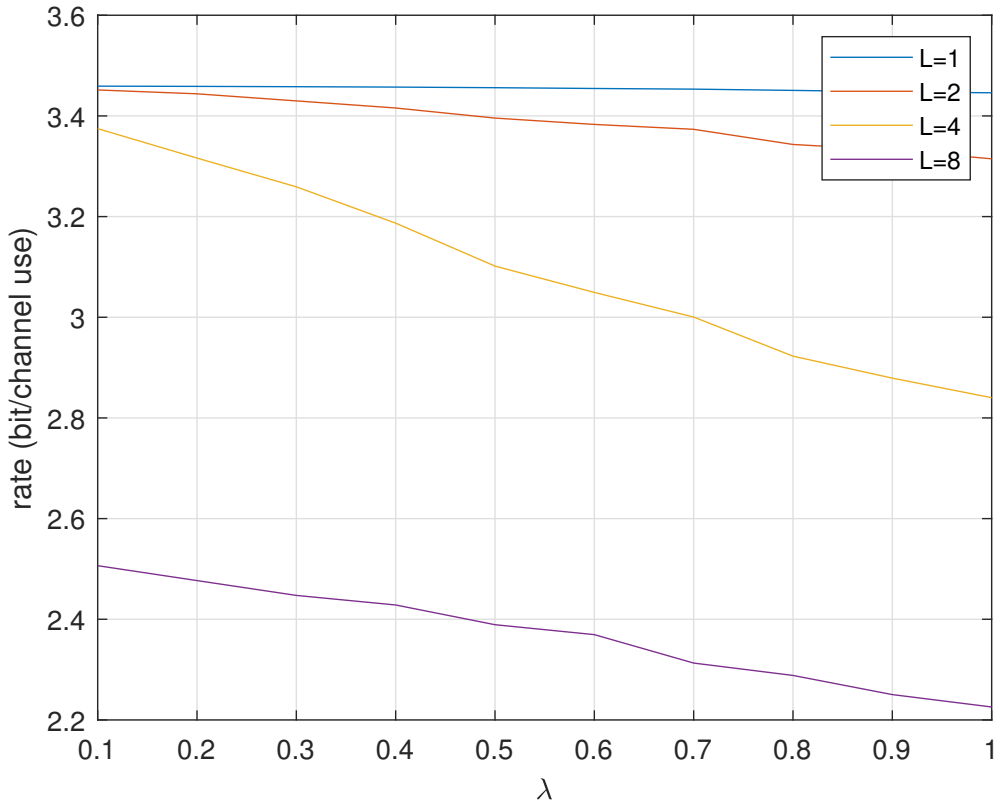


Figure 5.6: Rate-λ behavior of MM-algorithm with guard interval $\tau = 0$

As is mentioned before in the setups, the power of the intended symbol is 1 and the normalized noise has a max value of 0.1. Therefore, for the best case where the interference power is zero, we would have $rate = \log_2 11 \approx 3.45943$. From Fig. (5.9) 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 [9] with a general/fully connected analog unit. It is stated that for a HAD system with fully connected beamforming network, the interference power $\left|\left|\mathbf{H}\mathbf{F}_{\text{GA}}\mathbf{X} - \mathbf{S}\right|\right|_{\text{fro}}^2$ 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 statement can also be used to explain the purple curve standing for

$L = 8$. Moreover, we can also explain the curve for $L = 1$ by using Theorem (1). For the theorem, we assume that the channel estimation error is zero.

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 $K = 1$. Define $\mathbf{h}^T = \mathbf{H}$. We then 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 complex entry in vector $\mathbf{h}^T$ 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 vector $\mathbf{d}$ represents the rotation operations which are applied to the columns of $\mathbf{h}^T$ with controllable angles. 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

$$\begin{aligned} &\exists \mathbf{d}, (\mathbf{h}^T \mathbf{d} = 0) \\ \Leftrightarrow &\max(|h_1|, \ldots, |h_M|) \leq |h_1| + \cdots + |h_M| - \max(|h_1|, \ldots, |h_M|) \end{aligned} \tag{5.10}$$

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

$$\begin{aligned} P_1 =& Pr(\exists \mathbf{d}, (\mathbf{h}^T \mathbf{d} = 0)) \\ =& 1 - (M \times Pr(|h_1| > |h_2| + \cdots + |h_M|)) \end{aligned} \tag{5.11}$$

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

$$\begin{aligned} & M \times Pr(|X| > |X_1| + \cdots + |X_{M-1}|) \\ \leq & M \times Pr(|X|^2 > |X_1|^2 + \cdots + |X_{M-1}|^2) \\ \leq & M \times Pr(|X|^2 + |Y|^2 > |X_1|^2 + \cdots + |X_{M-1}|^2) \\ = & M \times Pr\left(\frac{2}{M-1} > \frac{(|X_1|^2 + \cdots + |X_{M-1}|^2)/(M-1)}{(|X|^2 + |Y|^2)/2}\right) \\ = & M \times F\left(\frac{2}{M-1}; M-1, 2\right) = M \times I_{\frac{1}{2}}\left(\frac{M-1}{2}, 1\right) \\ = & M \times \left(\frac{1}{2}\right)^{\frac{M-1}{2}} \end{aligned} \tag{5.12}$$

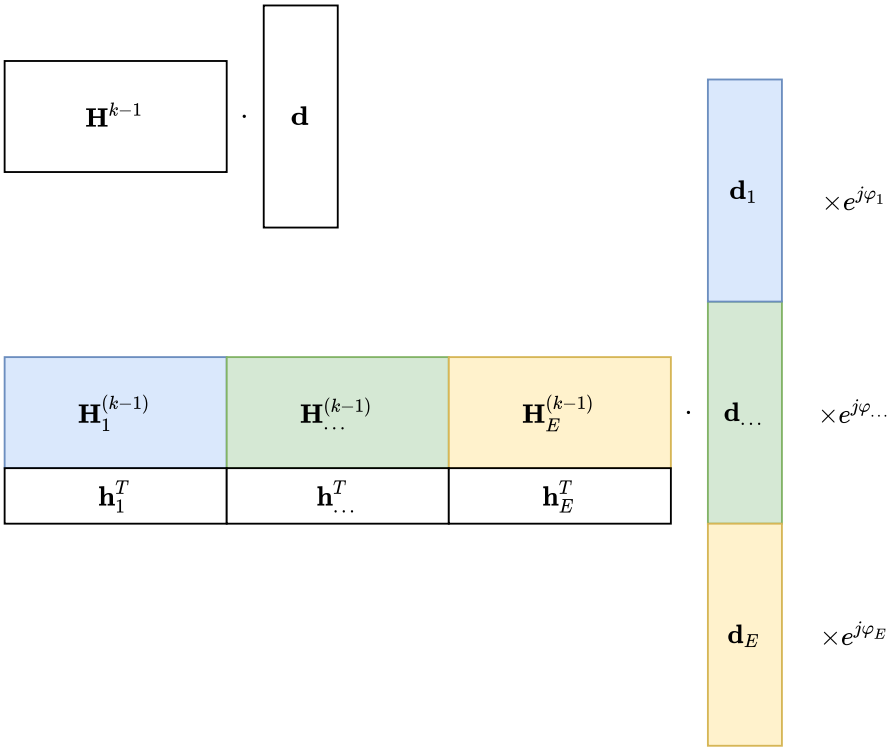


Figure 5.7: Idea of construction

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 $\mathbf{d}$ 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.

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 $K = k-1$. Thus, 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 $(k-1)$-rows submatrices with smaller column numbers and sectorize $\mathbf{d}$ into several vecrtors correspondingly. This is illustrated as Fig. (5.7).

We only need to find one solution for this problem and the conditions for that solution to exist can serve as sufficient conditions for $\exists \mathbf{d}, (\mathbf{H}^{(k)}\mathbf{d} = 0)$. Therefore, we first find $\mathbf{d}_1, \dots, \mathbf{d}_E$ such that $\forall i \in \{1, \dots, E\}, \mathbf{H}_i^{(k-1)}\mathbf{d}_i = \mathbf{0}$. Then for the last row of matrix $\mathbf{H}^{(k)}$, we know that $\mathbf{h}_i^T \cdot \mathbf{d}_i$ is 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 (5.7) stay $\mathbf{0}$.

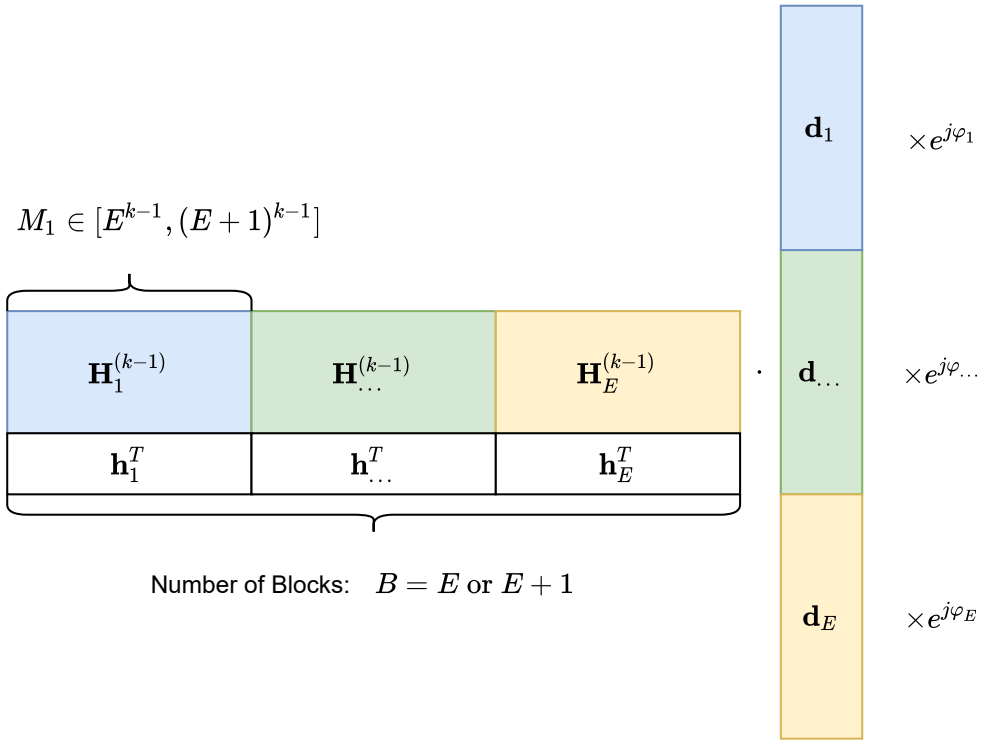


Figure 5.8: Submatrix Construction

Now the problem is transformed back to the previous one, where we want to find rotation angles $\varphi_1, \dots, \varphi_E$ such that

$$\begin{bmatrix} \mathbf{h}_1^T \mathbf{d}_1 & \dots & \mathbf{h}_E^T \mathbf{d}_E \end{bmatrix} \cdot \begin{bmatrix} e^{j\varphi_1} \\ \vdots \\ e^{j\varphi_E} \end{bmatrix} = 0 \tag{5.13}$$

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 Fig. (5.8). It is worth noticing that the sectorization method is not unique, the only requirement is that the the block column length is within the range of $[E^{k-1}, (E+1)^{k-1}]$. After the sectorization, there are either E or $E+1$ blocks in total.

The proposition that $\mathbf{H}^{(k)}$ cannot be column-wise rotated to an all-zero matrix by using the above mention method implies the the sufficient conditions do not hold. This means 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

$$\begin{aligned} &B \in \{E, E+1\} : \text{block number} \\ &M_{1\ldots B} \in [E^{k-1}, (E+1)^{k-1}] : \text{column number the blocks.} \end{aligned} \tag{5.14}$$

as is also illustrated in Fig. (5.8). Each block $\mathbf{h}_i^T$ in the last row will be multiplied with corrspondence 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 normlize the variance to 1. This is done in Eq. (5.15) and Eq. (5.16) Define

$$\begin{aligned} &\forall b \in \{1, \ldots, B\}, Y_b = \mathbf{h}_b^T \mathbf{d}_b \sim CN(0, M_b) \\ &\forall b \in \{1, \ldots, B\}, X_b \sim CN(0, 1) \end{aligned} \tag{5.15}$$

Then we have,

$$\begin{aligned} &\sum_{b=1}^{B} Pr(|Y_b| > |Y_1| + \cdots + |Y_{b-1}| + |Y_{b+1}| + \cdots + |Y_B|) \\ \leq &\sum_{b=1}^{B} Pr(|Y_b| > \sqrt{M_{min}}(\frac{|Y_1|}{\sqrt{M_i}} + \cdots + \frac{|Y_{b-1}|}{\sqrt{M_{b-1}}} + \frac{|Y_{b+1}|}{\sqrt{M_{b+1}}} + \cdots + \frac{|Y_B|}{\sqrt{M_B}})) \\ = &\sum_{b=1}^{B} Pr(\frac{\sqrt{M_b}}{\sqrt{M_{min}}} \frac{|Y_b|}{\sqrt{M_b}} > |X_1| + \cdots + |X_{B-1})| \\ \leq &B \times Pr(\frac{\sqrt{M_{max}}}{\sqrt{M_{min}}} |X| > |X_1| + \cdots + |X_{B-1}|) \\ \leq &B \times \left(\frac{M_{max}}{M_{max} + M_{min}}\right)^{\frac{B-1}{2}} \end{aligned} \tag{5.16}$$

For larger M, we have

$$\lim_{M \to +\infty} \frac{M_{max}}{M_{max} + M_{min}} \leq \lim_{E \to +\infty} \frac{(E+1)^{k-1}}{2 \times E^{k-1}} = \frac{1}{2} \leq \frac{2}{3} \tag{5.17}$$

Therefore,

$$Pr\left(\nexists(\varphi_1, \cdots, \varphi_B). \begin{bmatrix} \mathbf{h}_1^T \mathbf{d}_1 & \ldots & \mathbf{h}_B^T \mathbf{d}_B \end{bmatrix} \cdot \begin{bmatrix} e^{j\varphi_1} \\ \vdots \\ e^{j\varphi_B} \end{bmatrix} = 0\right) \leq (N+1) \times \left(\frac{2}{3}\right)^{\frac{E-1}{2}} \tag{5.18}$$

The complete probability can then be obtained

$$\begin{aligned} P_k &= Pr(\nexists \mathbf{d}, (\mathbf{H}^{(k)T}\mathbf{d} = 0)) \\ &\leq (E+1) \times P_{k-1} + (E+1) \times \left(\frac{2}{3}\right)^{\frac{E-1}{2}}. \end{aligned} \tag{5.19}$$

Based on this inequality, one can define an auxilary sequence for the upperbond of the probabilities recursively

$$\begin{aligned} P'_1 &= (E+1) \times \left(\frac{2}{3}\right)^{\frac{E-1}{2}} \\ P'_k &= (E+1) \times P_{k-1} + (E+1) \times \left(\frac{2}{3}\right)^{\frac{E-1}{2}}. \end{aligned} \tag{5.20}$$

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 as

$$P'_k = \frac{(E+1)^{k+1} - (E+1)}{E} \left(\frac{2}{3}\right)^{\frac{E-1}{2}}. \tag{5.21}$$

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 $\mathbf{H}' = \mathbf{H} \times \text{diag}_{\text{matrix}}\{x\mathbf{t}\}$, and the problem is transformed to finding vector $\mathbf{d} = \text{diag}_{\text{vector}}\{\mathbf{D}\}$ such that $\mathbf{s} = \mathbf{H}'\mathbf{d}$. In order to make the left side zero, define

$$\begin{aligned} \mathbf{H}'' &= \begin{bmatrix} \mathbf{H}', & \mathbf{s} \end{bmatrix} \\ \mathbf{d}' &= \begin{bmatrix} \mathbf{d}_{aux} \\ d_{M+1} \end{bmatrix}. \end{aligned} \tag{5.22}$$

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 $\mathbf{D} = \text{diag}_{\text{matrix}}\{-d_{M+1}^{-1}\mathbf{d}_{aux}\}$. 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 lengths larger than one is still unclear.

The relation between the average RF chain power and the regularizer lambda is shown as Fig. (5.9). 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 behaviors are 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

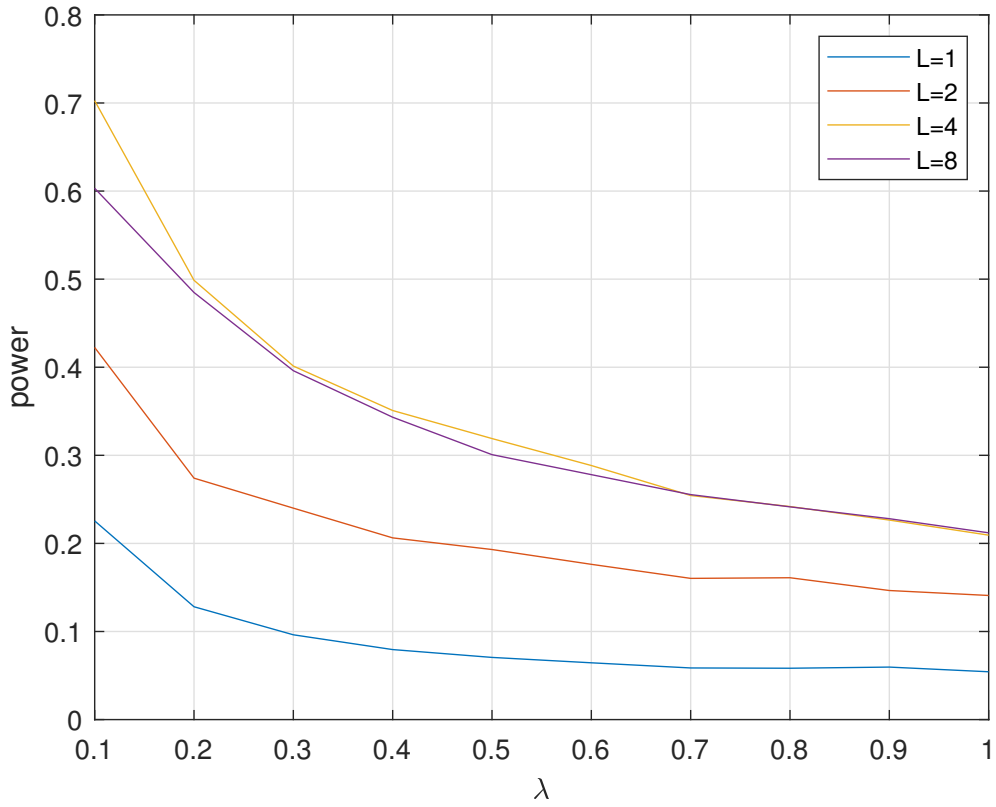


Figure 5.9: Average RF chain power-λ with guard interval $\tau = 0$

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 minimizing the interference power.

5.6 Rate, Block Length and IRS Element Number

Now we set the guard interval to be the same as the symbol interval, and we have $\tau = T_s$. The normalized channel estimation errors are no longer set to zero they are now matrice 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. (5.10) 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.

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. (5.11). In this figure, the rate first increases and then decreases. This phenomenon is caused by the non-zero guard interval.

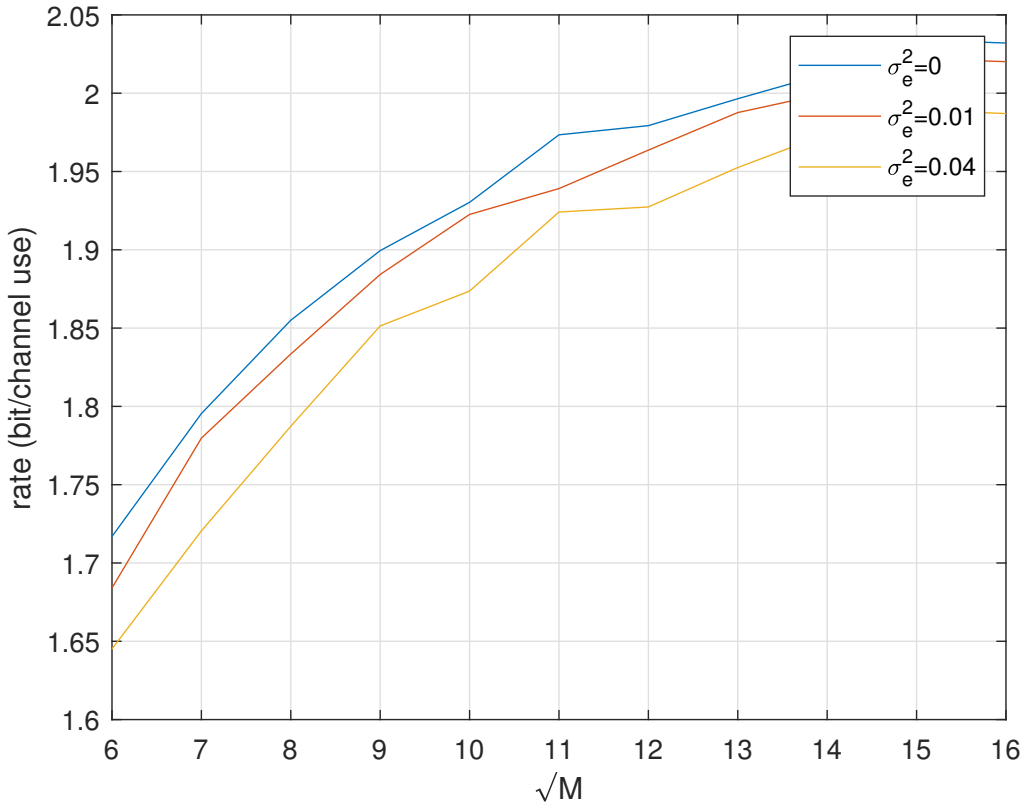


Figure 5.10: Rate-$\sqrt{M}$ with guard interval $\tau = T_s$ and $L = 10$

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 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 these two points, we can predict that for larger τ, the peak of the rate-L plot will shift to right since the term $\frac{LT_S}{\tau+LT_S}$ 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. (5.12).

At last it would also be 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. (5.12), we calculate the rate of each transmission block, and use the block rate to calculated the averaged minimal IRS size for reaching a certain rate.

For rate threshold 1.8, the averaged minimal M versus L is plotted as Fig. (5.13). 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

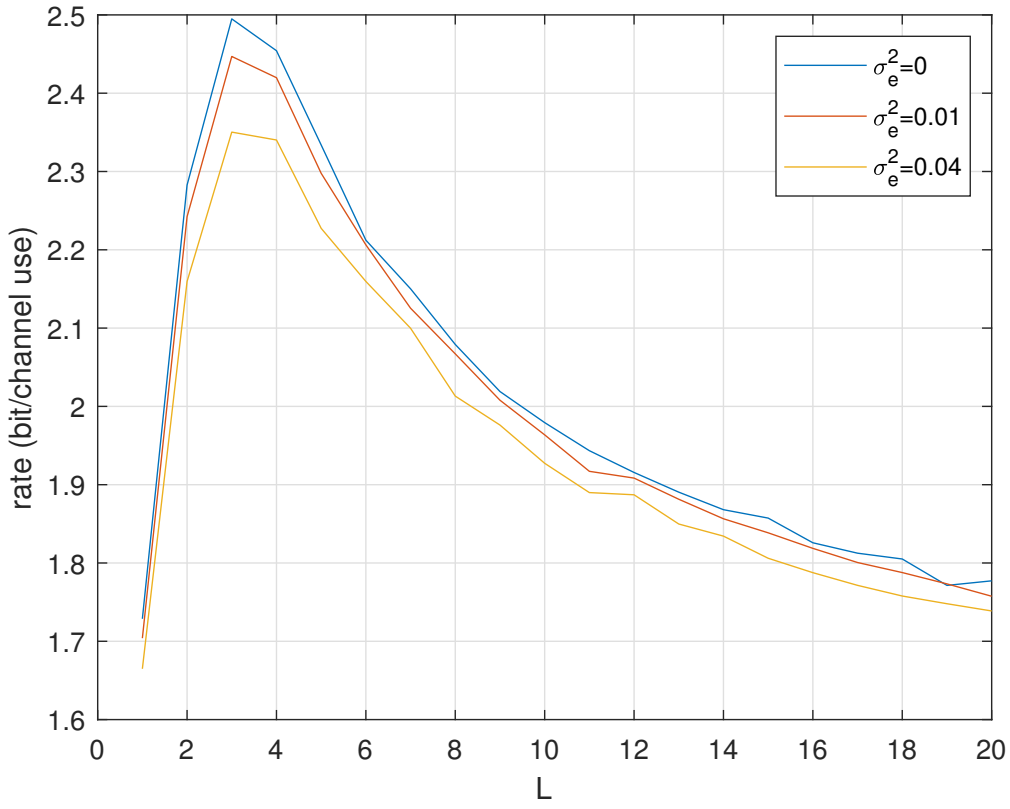
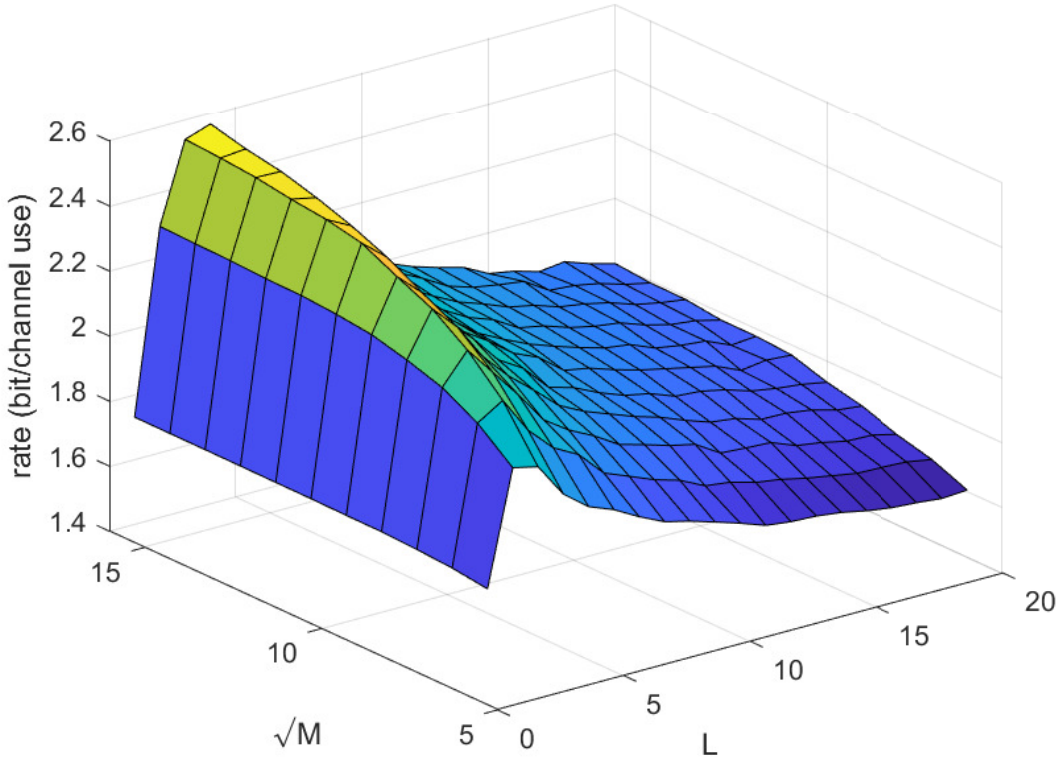
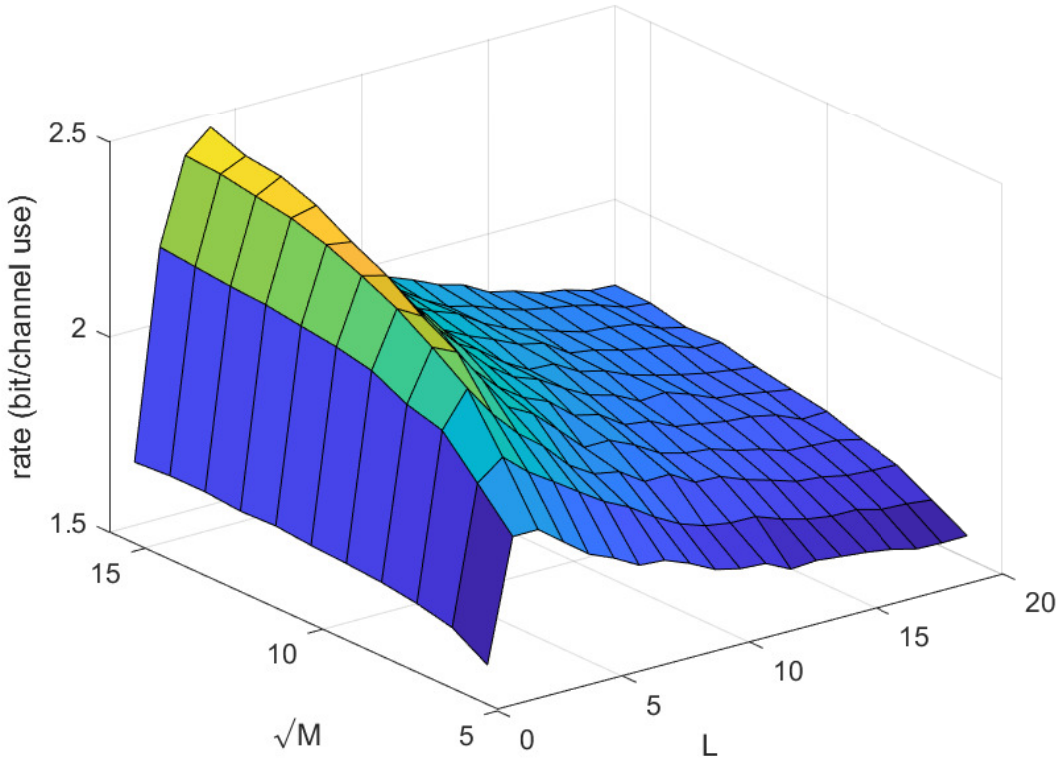


Figure 5.11: Rate-L with guard interval $\tau = T_s$ and $M = 144$

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.



(a) $\sigma_e^2 = 0$



(b) $\sigma_e^2 = 0.04$

Figure 5.12: Full relation between rate, $\sqrt{M}$ and L with $\tau = 1$

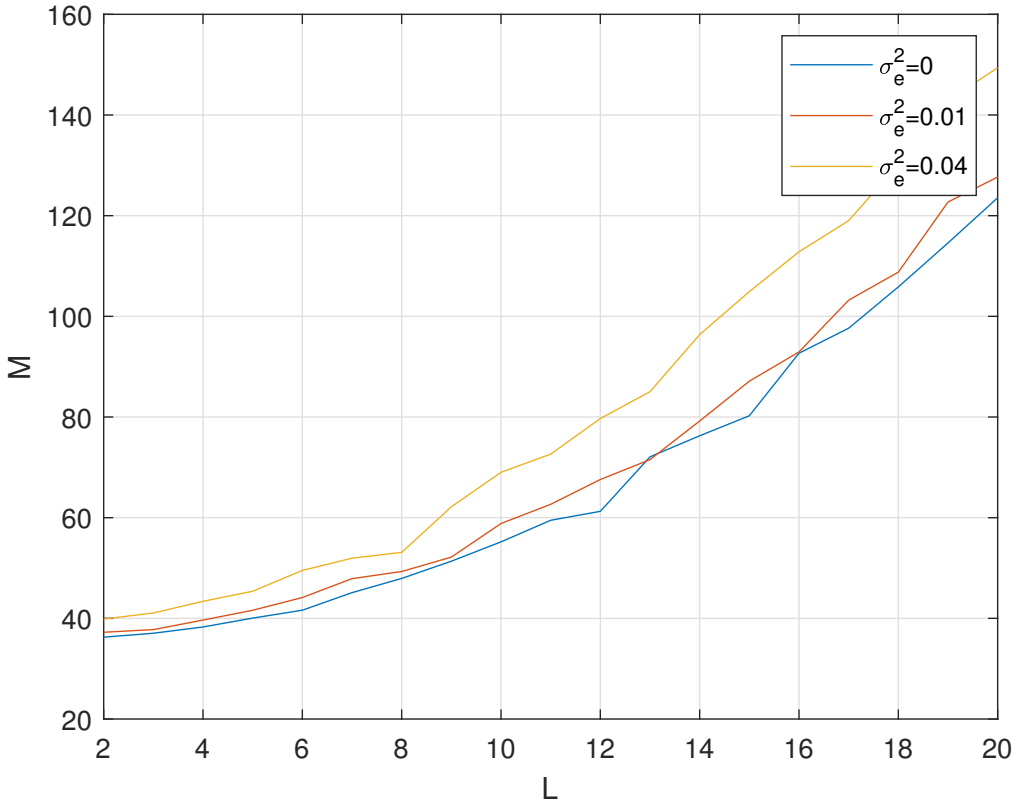


Figure 5.13: Averaged minimal M-L with rate_threshold= 1.8

Chapter 6

Conclusions

In this article, we investigate a block-wise precoding IRS aided transmission systems. The precoding task can be formulized as a GLSE problem. The RF chain output and IRS phase shifts are computed in turn during each iteration to find a suboptimal point for the interference power under certain power constraints. The RF chain output is computed with convex optimization methods, while the IRS configuration can be computed with gradient descent or MM algorithm.

From the simulations, we find out full illumination strategy is superior to the partial illumination strategy in terms of rate and PAPR. We also find out that gradient descent is slightly better than MM algorithm. It is proven in this article that for symbol-wise transmission, the interference power can be made to zero if we have really large IRS size. However, we have not yet proven the asymptotic behavior for the systems with block legnth is larger than one. From the observation of the Simulation results, the asymptotic behavior may hold.

Although symbol-wise systems have smaller interference, block-wise systems have better rate if guard interval is large. We also show that the influence of the estimation error increases as the block length increases. Despite these drawbacks, it can also be observed that the minimal required number for IRS elements increases not so much even if we use a relative big block length.

In the future, we may think of a precoding method that will compute the RF chain output and IRS configurations jointly by using gradient descent methods. This may be done by transforming the RF chain output into an auxilary variable such that the peak power constraint is included implicitly. Assume that x is one entry of $\mathbf{X}$, we can design the auxilary variable for it as Eq.(6.1).

$$x' = \frac{x}{|x|} \ln\left(\frac{2P_{\text{peak}}}{|x|} - 1\right) \tag{6.1}$$

After doing the joint optimization, we can transform the auxilary variable back via Eq. (6.2).

$$x = \frac{x'}{|x'|} \frac{2P_{\text{peak}}}{e^{|x'|} + 1} \tag{6.2}$$

Bibliography

[1] Wonil Roh, Ji-Yun Seol, Jeongho Park, Byunghwan Lee, Jaekon Lee, Yungsoo Kim, Jaeweon Cho, Kyungwhoon Cheun, and Farshid Aryanfar. Millimeter-wave beamforming as an enabling technology for 5G cellular communications: theoretical feasibility and prototype results. *IEEE Communications Magazine*, 52(2):106–113, 2014.

[2] Erik G. Larsson, Ove Edfors, Fredrik Tufvesson, and Thomas L. Marzetta. Massive MIMO for next generation wireless systems. *IEEE Communications Magazine*, 52(2):186–195, 2014.

[3] Emil BjÃ¶rnson and Luca Sanguinetti. Rayleigh Fading Modeling and Channel Hardening for Reconfigurable Intelligent Surfaces. *IEEE Wireless Communications Letters*, 10(4):830–834, 2021.

[4] Shuangfeng Han, Chih-lin I, Zhikun Xu, and Corbett Rowell. Large-scale antenna systems with hybrid analog and digital beamforming for millimeter wave 5G. *IEEE Communications Magazine*, 53(1):186–194, 2015.

[5] Linglong Dai, Xinyu Gao, Jinguo Quan, Shuangfeng Han, and Chih-Lin I. Near-optimal hybrid analog and digital precoding for downlink mmWave massive MIMO systems. In *2015 IEEE International Conference on Communications (ICC)*, pages 1334–1339, 2015.

[6] Vijay Venkateswaran, Florian Pivit, and Lei Guan. Hybrid RF and Digital Beamformer for Cellular Networks: Algorithms, Microwave Architectures, and Measurements. *IEEE Transactions on Microwave Theory and Techniques*, 64(7):2226–2243, 2016.

[7] Xinyu Gao, Linglong Dai, Shuangfeng Han, Chih-Lin I, and Robert W. Heath. Energy-Efficient Hybrid Analog and Digital Precoding for MmWave MIMO Systems With Large Antenna Arrays. *IEEE Journal on Selected Areas in Communications*, 34(4):998–1009, 2016.

[8] A. Bereyhi, V. Jamali, R. R. Mueller, G. Fischer, R. Schober, and A. M. Tulino. PAPR-Limited Precoding in Massive MIMO Systems with Reflect- and Transmit-Array Antennas. pages 1690–1694, 2019.

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

[10] Vahid Jamali, Antonia M. Tulino, Georg Fischer, Ralf R. Mueller, and Robert Schober. Intelligent Surface-Aided Transmitter Architectures for Millimeter-Wave Ultra Massive MIMO Systems. *IEEE Open Journal of the Communications Society*, 2:144–167, 2021.

[11] Ali Bereyhi, Mohammad Ali Sedaghat, Ralf R. Mueller, and Georg Fischer. GLSE Precoders for Massive MIMO Systems: Analysis and Applications. *IEEE Transactions on Wireless Communications*, 18(9):4450–4465, 2019.

[12] Surya Prakash Sankuru and Prabhu Babu. Designing unimodular sequence with good auto-correlation properties via Block Majorization-Minimization method. *Signal Processing*, 176:107707, 2020.

[13] Adrian Garcia-Rodriguez, Vijay Venkateswaran, Pawel Rulikowski, and Christos Masouros. *IEEE Wireless Communications Letters*, (5):528–531, 2016.

[14] David M Pozar. *Microwave engineering*. John wiley & sons, 2011.