MASSIVE MIMO:

SIMULATIONS TO MITIGATE THE PILOT CONTAMINATION

by

Tianyuan Qiu

A thesis submitted to the Faculty of the University of Delaware in partial fulfillment of the requirements for the degree of Master of Science in Electrical and Computer Engineering

Spring 2017

MASSIVE MIMO:

SIMULATIONS TO MITIGATE THE PILOT CONTAMINATION

by

Tianyuan Qiu

Approved: ____
Xiang-Gen Xia, Ph.D.
Professor in charge of thesis on behalf of the Advisory Committee

Approved: ____
Kenneth E. Barner, Ph.D.
Chair of the Department of Electrical and Computer Engineering

Approved: ____
Babatunde A. Ogunnaike, Ph.D.
Dean of the College of Engineering

Approved: ____
Ann L. Ardis, Ph.D.
Senior Vice Provost for Graduate and Professional Education

ACKNOWLEDGMENTS

I wish to thank all of the little people (especially Chicken Little); my adviser, Xiang-Gen Xia; and my friends and colleagues Zhe Qu and Jie Zhang.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

In future mobile data traffic, increasing of global exponential mobile data traffic, supporting more devices simultaneously and larger variety of traffic types are the main problems that wireless communication systems face. However, the MIMO technology cannot solve these problems well. Thus, the Massive MIMO technology, which deploy a large excess of antennas at BSs as compared to the number of terminal served, was proposed in 2010 by Thomas L. Marzetta.

Massive MIMO is a multi-user MIMO technology where each BS is equipped with an array of M active antenna elements and utilizes these to communicate with K single-antenna terminals----over the same time and frequency band.

The benefits of the Massive MIMO technology include increasing the capacity, improving the radiated energy-efficiency, reducing latency on the air interface and increasing the robustness both to unintended man-made interference and to intentional jamming, etc. Nevertheless, the Massive MIMO technology also faces some challenges and pilot contamination is one of them.

In this thesis, firstly, I will introduce the background and significance of the Massive MIMO. Secondly, I will introduce the Massive MIMO system which will include the comparison between TDD mode and FDD mode, the introduction for the system model, the pilot-based channel estimation and the pilot contamination. Finally,

I will review and simulate some methods to mitigate the pilot contamination like Bayes channel estimation and pilot scheduling.

Chapter 1

INTRODUCTION TO MASSIVE MIMO

1.1 Background And Significance Of Massive MIMO

MIMO(multiple input & multiple output) use transmission diversity, spatial multiplex and beam forming technology to exploit space resources, which can enhance the spectral-efficiency and energy-efficiency. MIMO technology is a key to 4G. The speed of data transmission of 4G can achieve 1000 Mb/s. However, with the popularization of intelligent terminal and the development of internet business, the throughput of mobile data transmission in 2020 will be 100 times more than it is now[1]. Obviously, 4G cannot satisfy the demand for this dramatic growth. Modern mobile communication system needs to support more user equipment, higher speed of data transmission and shorter time delay of transmission.

Thomas L. Marzetta proposed a concept: Massive MIMO in 2010[2]. It is the extension over MIMO technology. Massive MIMO use antenna arrays with a few hundreds antennas, simultaneously serving many tens of terminals in the same time-frequency resource. The BS applies linear receive combining to discriminate the signal transmitted by each terminal from the interfering signals in the up link and precodes data for all served terminals in the down link. In order to receive accurate signal in the BS and create spatially multiplexed data streams for each terminal, the BS have to estimate the channel to get CSI with pilot sequence. Figure 1.1 illustrates the concept of a typical down-link Massive MIMO system. Since fading makes the channel responses vary over time and frequency, the estimation and data transmission must fit into a time/frequency block where the channel are approximately static. The dimensions of this block are essentially given by the coherence bandwidth B_c HZ and the coherence time T_c s,which fit $T = B_cT_c$ transmission symbols. Figure 1.2 shows the example of coherence interval. Overall, massive MIMO is an enabler for the development of future broadband (fixed and mobile) networks which will be energy-efficient, secure, and robust, and will use the spectrum efficiently[3].

Theoretically, with the increasing of the number of antennas in the BS, uncorrelated interference and noise will vanish; the matched filter will be optimal and the transmit power can be made arbitrarily small.

Economically, the antennas in the BS in Massive MIMO system are consist of many small-antenna units and amplifiers with low power consumption. Thus, Massive MIMO can degrade the requirements for hardware like amplifiers, etc[2].

Stably, the number of antennas in the BS is much more than the number of terminals, which means even if several antennas cannot work in the BS the system can still run well. The high degree of freedom of antennas can make system highly robust and reliable[4].

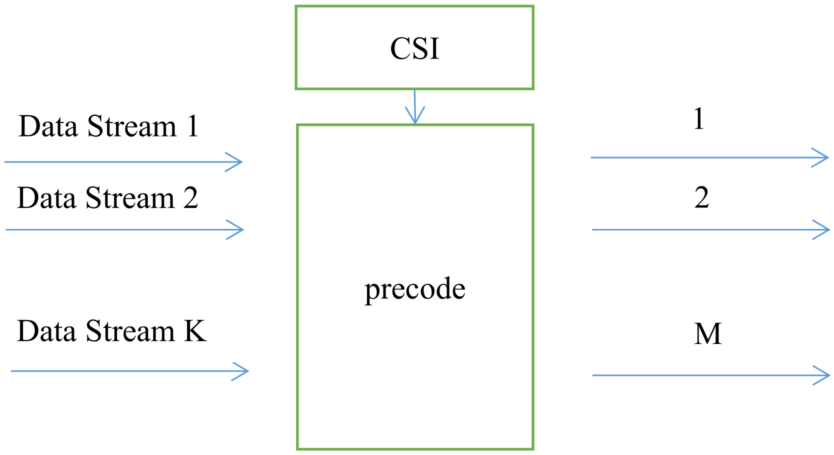

Figure 1.1: The Example Of A Down-Link Massive MIMO System.

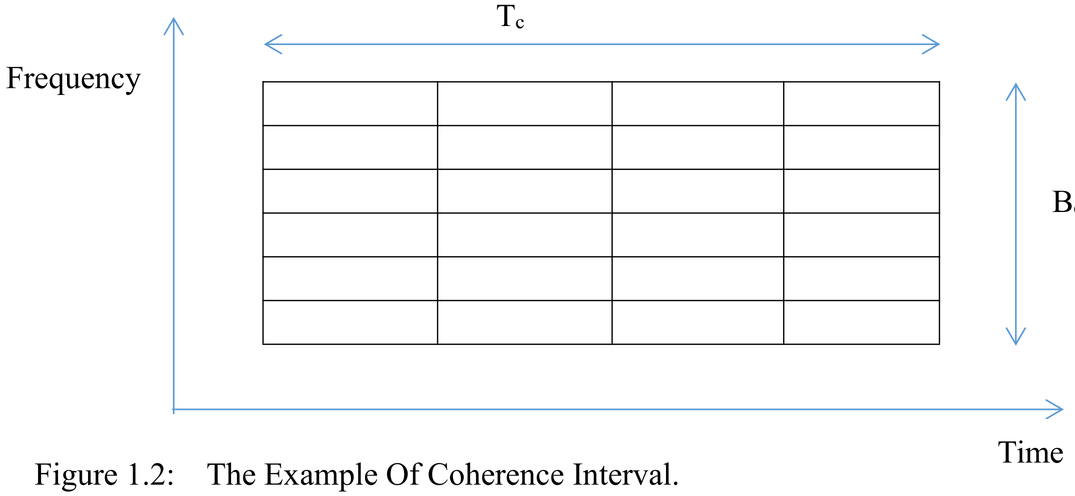

Figure 1.2: The Example Of Coherence Interval.

1.2 Limiting Factors of Massive MIMO

1.2.1 channel reciprocity

Nowadays, the canonical Massive MIMO system operates in time-division duplex(TDD) mode, where the up link and down link transmission take place in the same frequency resource but separated in time[5]. The reason why Massive MIMO will use TDD mode will be explained later. TDD operation relies on channel reciprocity. There appears to be a reasonable consensus that the propagation channel itself is essentially reciprocal, unless the propagation is affected by materials with strange magnetic properties. However, the hardware chains in the BS and terminal transceivers may not be reciprocal between the up link and the down link. Calibration of the hardware chains does not seem to constitute a serious problem and there are calibration based solutions that have already been tested to some extent in practice [6,7]. Specifically, [6]treats reciprocity calibration for a 64-antenna system in some detail and claims a successful experimental implementation.

Note that calibration of the terminal up-link and down-link chains is not required in order to obtain the full beam forming gains of massive MIMO: if the BS equipment is properly calibrated then the array will indeed transmit a coherent beam to the terminal. (There will still be some mismatch within the receiver chain of the terminal but this can be handled by transmitting pilots through the beam to the terminal; the overhead for these supplementary pilots is very small.) Absolute calibration within the array is not required. Instead, as proposed in [6], one of the antennas can be treated as a reference and signals can be traded between the reference antenna and each of the other antennas to derive a compensation factor for that antenna. It may be possible entirely to forgo reciprocity calibration within the array; for example if the maximum phase difference between the up-link chain and the down-link chain were less than 60 degrees then coherent beam forming would still occur (at least with MRT beam forming) albeit with a possible 3 dB reduction in gain.

1.2.2 pilot contamination

Ideally every terminal in a Massive MIMO system is assigned an orthogonal up-link pilot sequence. However the maximum number of orthogonal pilot sequences that can exist is upper-bounded by the duration of the coherence interval divided by the channel delay-spread. In [2], for a typical operating scenario, the maximum number of orthogonal pilot sequences in a one millisecond coherence interval is estimated to be about 200. It is easy to exhaust the available supply of orthogonal pilot sequences in a multi-cellular system.

The effect of re-using pilots from one cell to another, and associated negative consequences, is termed “pilot contamination”. More specifically, when the service-array correlates its received pilot signal with pilot sequence associated with a particular terminal, it actually obtains a channel estimation that is contaminated by a linear combination of channels to the other terminals that share the same pilot sequence. Down-link beam forming based on the contaminated channel estimation results in interference that is directed to those terminals that share the same pilot sequence. Similar interference is associated with up-link transmissions of data. This directed interference grows with the number of service-antennas at the same rate as the desired signal [2].

Chapter 2

MASSIVE MIMO SYSTEM

2.1 TDD vs FDD operation

It is now well understood that Massive MIMO system needs to obtain information about up- and down-link channel when trying to maximize network throughput. In FDD mode, the links occupy the non-overlapping frequency bands that allows simultaneous transmission and reception in the BS[8]. The down-link channel may be estimated at the terminals using mutually orthogonal pilot transmissions from each BS antenna. The number of required time-frequency resource blocks is therefore proportional to M, where M is the number of antennas in the BS. Each terminal reports the $M \times 1$ channel vector over the up-link, consuming additional resources proportional to M. The estimation and feedback adds latency before the BS can use them for multiplexing the down-link data, which scales proportional to M. Therefore, in a Massive MIMO system, FDD is generally difficult to implement.

On the contrast, in TDD mode, the up-link and down-link transmissions take place in the same frequency resource but are separated in time. The estimation in the up-link channel are valid for the down-link channel as well because the channels are reciprocal, albeit with some calibration. There are several good reasons to operate in

TDD mode. Firstly, only the BS needs to know the channels to process the antennas coherently[5]. Secondly, the up-link estimation overhead is proportional to the number of terminals, but independently of M thus making the protocol fully scalable with respect to the number of service antennas. Furthermore, basic estimation theory tells us that the estimation quality (per antenna) cannot be reduced by adding more antennas in the BS ----in fact, the estimation quality improves with M if there is a known correlation structure between the channel responses over the array[9].

2.2 System Model

In this thesis, we consider a system model with L time-synchronized cells containing K single-antenna terminals each. The BS in each cell is equipped with an M-antenna array that communicates with the terminal over the same time-frequency interval($M \geq k$). Figure 2.1 shows the Massive MIMO system model in this thesis. In ideally condition, terminals at all cells use orthogonal pilots and with the increasing of M, uncorrelated interference and noise will vanish. The length of pilot sequences will be K x L at least. Generally, the value of L is big and the coherence time T_c is limited. Furthermore, the coherence time T_c will be small when terminals are in the process of rapid movement. Thus, the length of pilot sequence is limited.

Due to the limitation of the length of pilot sequence, every cells reuse the same orthogonal pilot sequence set in practice. Terminals at the same cell use the orthogonal pilot sequences. The number of antennas in the BS is much more than the number of terminals at one cell ($M \gg K$). BS can discriminate terminals within a cell because of abundant degrees of freedom. Thus, we can think that there is no interference among terminals at a cell and there is just one terminal at a cell[2].

Here, we can define $H_{ilk} = \beta_{ilk}^{1/2}h_{ilk}$, which denotes the M x 1 channel matrix from the k^{th} terminal at l^{th} cell to the BS at i^{th} cell.

$$h_{ilk} = R_{ilk}^{1/2}W_{ilk},$$

where W_{ilk} represents fast fading vector and $W_{ilk} \sim CN(0,1)$. W_{ilk} subjects to i.i.d. $R_{ilk} = E\{h_{ilk}h_{ilk}^H\}$ is the M x M correlation matrix of channel. β_{ilk} represents slow fading coefficient containing pathloss and shadow fading. In this thesis, we just consider the pathloss. Thus, β_{ilk} can be expressed as:

$$\beta_{ilk} = \alpha/d_{ilk}^{\gamma},$$

where d_{ilk} represents the distance from k^{th} terminal at l^{th} cell to the BS at i^{th} cell. The unit of d_{ilk} is m. γ is the index of pathloss and α is a constant, equaling to 1 in this thesis.

It will have interference only if terminals at different cells don't use orthogonal pilots. However, with the increasing of M at BS, the interference that terminals use non-orthogonal pilots can be ignored.

$$\lim\ (H_{ilk}^H H_{ilk})/M = \beta_i \text{ when M tends to infinity}$$

Now, the interference will exist only when terminals at different cells use the same pilots.

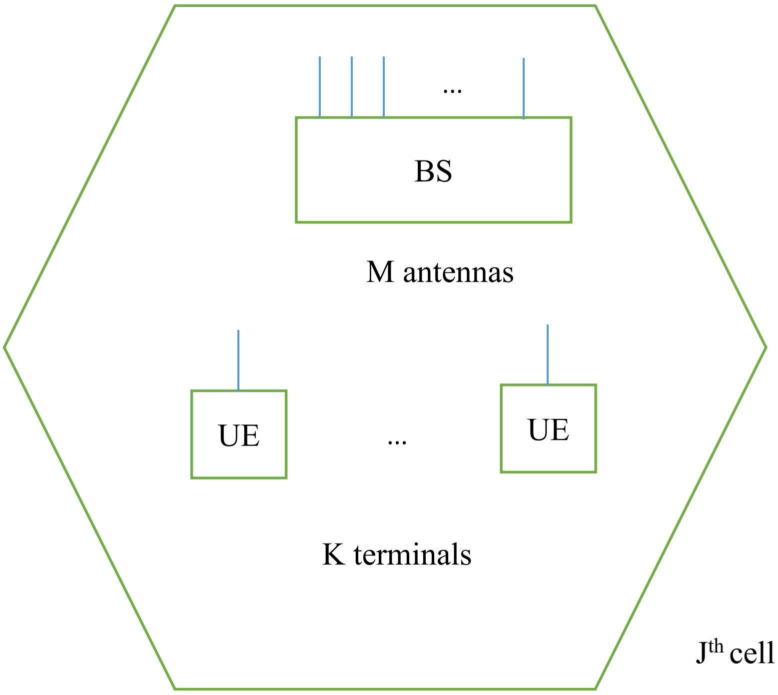

Figure 2.1: Massive MIMO System Model.

2.3 up-link pilot transmission

The aim of pilot transmission is BS can acquire CSI by pilot sequences. BS can receive accurate signals and precode signals on the down link based on the CSI. The terminals transmit pilots to their BSs respectively. Here, we take an example of a BS at i^{th} cell. The received signal from the BS at i^{th} cell can be expressed as:

$$Y_i = \sum_{l=1}^{L}\sum_{k=1}^{K} H_{ilk}s_{ilk}+N_i,$$

where s_{lk} is the pilot transmitted from k^{th} terminal at l^{th} cell. It is the 1 x T vector and $s_{lk}s_{lk}^H = T$. N_i is additive white Gaussian noise received in the BS at i^{th} cell. It is the M x T vector. H_{ilk} is the M x 1 channel matrix from k^{th} terminal at l^{th} cell to the BS at i^{th} cell. Thus, from the above equation, we can find that the received signal contains the interfering signal from other cells and the noise.

2.4 pilot-based channel estimation

According to 2.3, because we can assume that each cell just has one terminal and each terminal use the same pilots, we can simplify the Y_i to:

$$Y_i = H_i s + \sum_{l \neq i}^{L} H_l s + N_i$$

where s is the same pilot that each terminal use and the length of the pilot is T. $ss^H = T$. We let H_i be the target channel and the estimation of Hi is $\hat{H}_i$.

2.4.1 LS estimation

The LS approach to channel estimation seeks to minimize the squared error between the received pilot sequence and its noise-and-interference free version [8]. The cost function of LS is:

$$J_{LS} = (Y_i - s\hat{H}_i)H(Y_i - s\hat{H}_i) .$$

Taking partial derivative of J_{LS} with respect to $\hat{H}i$ and let the result equals to 0:

$$dJ_{LS}/d\hat{H}_i = 0.$$

Thus, we can get the LS estimator is:

$$G_i^{LS} = s^H(ss^H)^{-1} = s^H/T.$$

The result of the estimation can be expressed as:

$$\hat{H}_i^{LS} = Y_i G_i^{LS} = H_i + \sum_{l \neq i}^{L} H_l + N_i s^H/T$$

The LS estimator has low complexity and treats the channel coefficient as a deterministic variable to obtain a "best-fit" estimate from the observed pilot signal. It makes no prior assumption about the channel statistics, and due to its simple implementation it the most common approach to channel estimation in practice.

2.4.2 MMSE estimation

The cost function of MMSE is:

$$J_{MMSE} = E(H_i-\hat{H}_i)(H_i-\hat{H}_i)^H.$$

Taking partial derivative of J_{MMSE} with respect to $\hat{H}i$ and let the result equals to 0:

$$dJ_{MMSE}/d\hat{H}_i = 0.$$

We can get the estimator of MMSE is:

$$G_i^{MMSE} = (s^H R_{H_iH_i}s + \sum_{l\neq i}^{L} s^H R_{H_lH_i}s + \sigma_n^2 I_T)^{-1} s^H R_{H_iH_i}s .$$

Thus, the result of MMSE can be expressed as:

$$\hat{H}_i^{MMSE} = YiG_i^{MMSE}$$

$$\hat{H}_i^{MMSE} = Yi(s^H R_{H_iH_i}s + \sum_{l\neq i}^{L} s^H R_{H_lH_i}s + \sigma_n^2 I_T)^{-1} s^H R_{H_iH_i}s ,$$

where $R_{XY} = E\{X^HY\}$ represents the covariance matrix of X and Y. When X = Y, the matrix is autocovariance matrix.

We observe that theoretically, the MMSE estimator has significantly higher implementation and processing complexity than the LS estimator. It requires the knowledge of all cross-channel covariance matrices in all BSs, that must be estimated prior to MMSE channel estimation. In practice, this imposes significant overhead and additional latency on the system. To reduce this overhead, we can assume that the interference from terminals located more than a few cells away to be negligible and not estimate the corresponding cross-channel matrices, at the cost of slightly poorer estimator performance. In terms of processing requirements, the complexity of M x M matrix inversion required during evaluation of MMSE estimate is proportional to the cube of array size, and may be especially problematic for Massive MIMO system[8].

2.5 pilot contamination

the channel estimates obtained above are useful only within the coherence interval, after which the channel must be estimated again. Moreover, the maximum number of mutually orthogonal pilot sequences is fundamentally limited by T which must be smaller than number of coherent time-frequency elements, $T \leq N_{och}$. Within a cell, the K terminals always use orthogonal pilot sequences to eliminate intra-cell pilot interference($T \geq K$). However, depending on the value of N_{och}, these sequences may have to be reused in other cells, which leads to inter-cell interference[8]. From the equation:

$$Y_i = \sum_{l=1}^{L} H_l s + N_i$$

we observed the effect of this interference on the received pilot signal, known as the "pilot contamination" effect. The Figure 2.2 illustrates the principle of pilot contamination. In particular, we found that the received pilot signal is contaminated with the transmissions from terminals at l^{th} cell, which $l \neq i$, reusing the same sequence. The worst-case pilot contamination occurs when each cell reuses the same set of mutually orthogonal pilot sequences.

During the transmission of up-link data, the signal transmitted from the k^{th} terminal at l^{th} cell to the BS is x_{lk}.it is a scalar and $x_{lk}^H x_{lk} = \sigma_x^2$, so the signals received at i^{th} cell is:

$$Y_i = H_i x_i + \sum_{l \neq i}^{L} H_l x_l + N_i.$$

It is M x 1 vector and Ni is the white Gaussian noise received at i^{th} cell. N_i subjects to $CN(0,\sigma_n^2)$. If we use MF detector, we can get the detected signal is:

$$\hat{H}_i^H Y_i = \hat{H}_i^H (H_i x_i + \sum_{l \neq i}^{L} H_l x_l + N_i),$$

where $\hat{H}_i H$ is the conjugate transposition of estimated channel. If we use the LS estimator, we could get:

$$\hat{H}_i^H Y_i = (H_i + \sum_{l \neq i}^{L} H_l + Nis^*(s^T s^*)^{-1})^H (H_i x_i + \sum_{l \neq i}^{L} H_l x_l + N_i).$$

Thus, the available signal power is:

$$\|(H_i + \sum_{l \neq i}^{L} H_l + Nis^*(s^T s^*)^{-1})^H H_i x_i\|_F^2.$$

According to $E\{s^H s\} = \sigma_s^2$, the average available signal power is:

$$P_u = \{\|(H_i + \sum_{l \neq i}^{L} H_l + N_i s^*(s^T s^*)^{-1})^H H_i x_i\|_F^2\}$$

$$= E\{|(H_i^H H_i x_i + \sum_{l \neq i}^{L} H_l H_i^H x_i + (N_i s^*(s^T s^*)^{-1})^H) H_i x_i|^2\}$$

$$= E\{|(R_{ii} x_i + \sum_{l \neq i}^{L} H_l H_i^H x_i + (N_i s^*(s^T s^*)^{-1})^H) H_i x_i|^2\},$$

Where R_{ii} is the covariance matrix of the channel. Because there is only one antenna at the terminal, the R_{ii} is actually a scalar:

$$R_{ii} = E\{H_i^H H_i\} = M\beta_{ii}.$$

The power of $H_i^H H_l$ is $E\{|H_i^H H_l|\}$ and H_l is not related to H_i, so:

$$E\{\sum_{l\neq i}^{L} H_l H_i^H x_i\} = 0.$$

And $E\{|N_i s^*(s^T s^*)^{-1})^H)H_i x_i|^2\} = (\sigma_n^2\sigma_x^2/T\sigma_s^2)M\beta_{ii}$.

So, we can get the Pu:

$$P_u = M^2\beta_{ii}^2\sigma_x^2+M\sigma_x^2\beta_{ii}\sum_{l\neq i}^{L} \beta_{li}+(\sigma_n^2\sigma_x^2/T\sigma_s^2)M\beta_{ii}.$$

where σ_x^2 is the variance of up-link data signals, σ_s^2 is the variance of up-link pilot signals and σ_n^2 is the variance of noise. β_{li} is the large-scale fading between terminals at l^{th} cell to the BS at i^{th} cell. In the Massive MIMO system, the number of antenna at BS is big. We can think M^2>>M. Thus, Pu can be approximately expressed as:

$$P_u \approx M^2\beta_{ii}^2\sigma_x^2.$$

Similarly, we can get the approximate interfering power is:

$$P_i \approx M^2\sigma_x^2\sum_{l\neq i}^{L} \beta_{li}^2.$$

The average power of noise is:

$$P_n \approx \sigma_n^2(M\beta_{ii}+M\sum_{l\neq i}^{L} \beta_{li}+\sigma_n^2/\sigma_s^2).$$

So, we can get the statistical average of SINR of the BS at i^{th} cell:

$$SINR_i^{UL} = P_u/(P_i+P_n) \approx M^2\beta_{ii}^2\sigma_x^2/(M^2\sigma_x^2\sum_{l\neq i}^{L} \beta_{li}^2+\sigma_n^2(M\beta_{ii}+M\sum_{l\neq i}^{L} \beta_{li}+\sigma_n^2/\sigma_s^2))$$

Thus, according to the analysis above, we can conclude that with the increasing of the number of M, SINR will increase a little. However, when M tends to infinite, SINR will not change with the M.

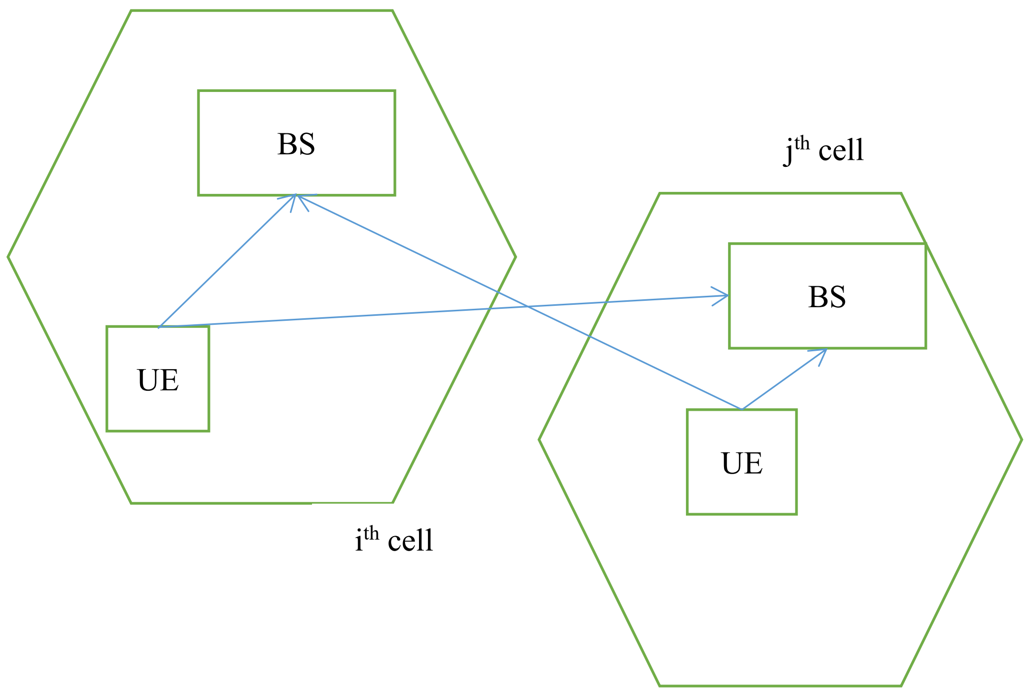

Figure 2.2: The Example Of Pilot Contamination.

2.6 the results of simulation

In this section, firstly, I will prove that the interference that terminals use non-orthogonal pilots can be ignored with the increasing of M at BS. In this simulation, we assume that channel from terminal i and terminal j to the same base station are H_i and H_j, where H_i and H_j are subject to complex Gaussian distribution. Other parameters of this simulation are showed in Table 2.1.

Table 2.1: The parameters of simulations

parameter	value
the number of cells (L)	2
diameter of cell	1000m
the number of terminals at each cell (K)	1
coefficient of slow fading of the target terminal (β_1)	1
coefficient of slow fading of the interfering terminal (β_2)	0.1
length of the pilot sequence (T)	20
SNR	10dB

The Figure 2.3shows the results of simulation:

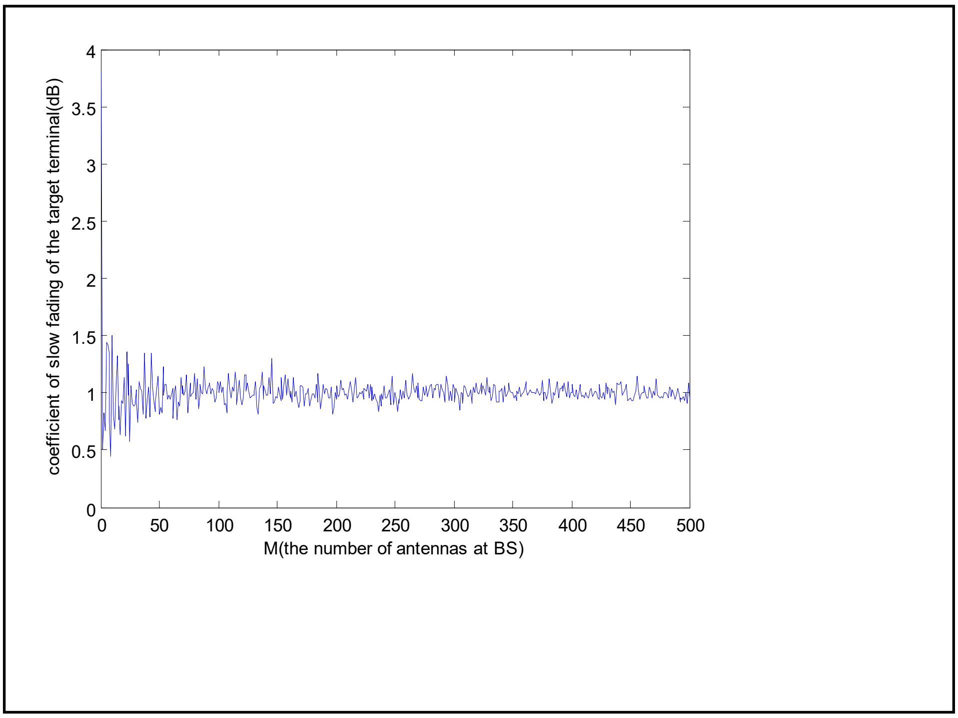

Figure 2.3: the simulation of progressive orthogonality of two channels

Form the figure 2.3 the curve represents the equation lim $(H_{ilk}{}^{H}H_{ilk})/M = \beta_i$. From the parameters of simulation we can know that $\beta_i = 1$. We can see that the channels are progressive orthogonal.

And then, I will simulate the effect of the number of antennas at BS on pilot contamination. I assume the terminals at different cells will use the same pilot sequence and in this situation the pilot contamination will be the heaviest. In the simulation, I will use LS estimation and the standard of evaluation will be MSE of estimated channel. The parameters have been shown in Table 2.1.

The figure 2.4 shows the result of simulation:

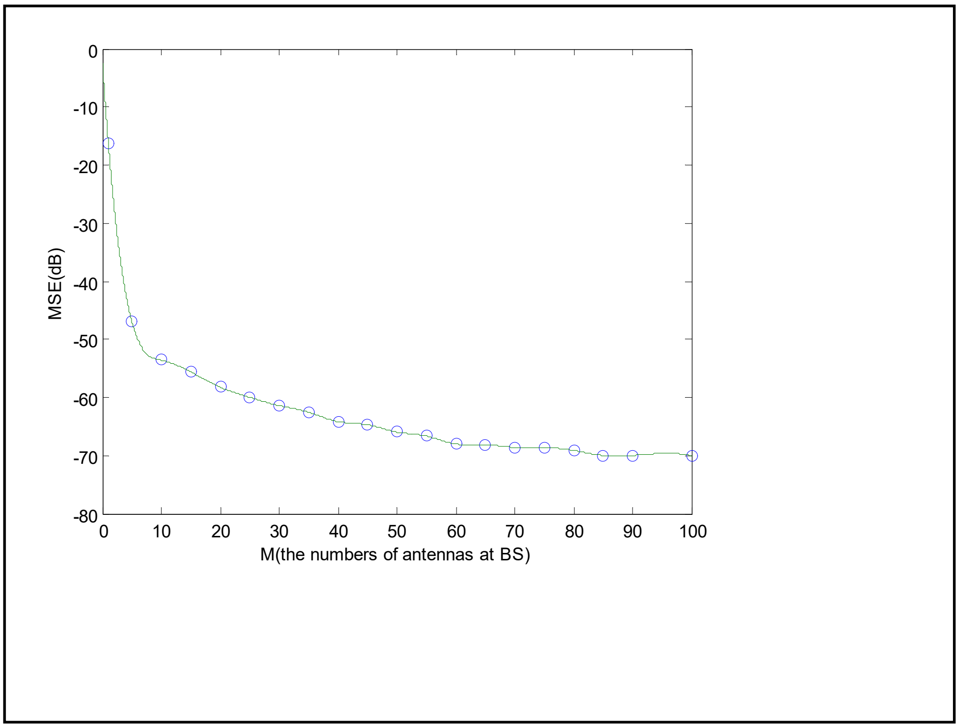

Figure 2.4: effect of number of antennas on pilot contamination with LS estimation

Form the figure 2.4, I can find that when the number of antennas at BS is few, increasing the number of antennas at BS will help system mitigate the pilot contamination. However, after the number of antennas at BS equals to 60, increasing numbers of antennas will be helpless to mitigate pilot contamination.

Chapter3

METHODS TO MITIGATE PILOT CONTAMINATION

3.1 Bayes channel estimation

3.1.1 signal model

We consider a setup with L cells. At each cell, BS has M antennas. All these cells reuse the frequency resources. There is no interference between terminals at each cell. Every cell have one terminal. Actually, every cell are equivalent. We can set the l^{th} cell as the target cell and the pilot sequence $\mathbf{s}_l$ with T length:

$$s_l = [s_{l1}\ s_{l2}\ \ldots\ s_{lT}].$$

The pilot sequences satisfy the equation $|s_l s_l^H| = T$. Because pilot sequences are orthogonal we can get $|s_l s_j^H| = 0$. We can express the channel matrix from terminal at j^{th} cell to the target cell as h_j, which is a M x 1 vector. Thus, during the transmission of pilot signal on up link, received pilot signals by BS at the target cell can be denoted:

$$Y_l = h_l s_l + \sum_{j \neq l}^{L} h_j s_j + N_l, \qquad (1)$$

Where "l" is the target cell, h_l is the local channel matrix and h_j is the channel matrix from other cells to the target cell. N_l is the white Gaussian noise received by l^{th} cell, which is the M x T vector.

We can denote the h_l:

$$h_l = R_l^{1/2} h_{wl}$$

$$R_l = E\{h_l h_l^H\}.$$

Where R_l is the M x M covariance matrix of channel and h_{wl} is a M x 1 vector which express the small-scale fading from the terminal to the BS.($h_{wl} \sim CN(0,1)$)

From the equation (1), we can know that the first item of the equation is the expected signal by the target cell and the second item is the overlap of the pilot sequences transmitted from other cells. If other cell use pilot sequences which are not orthogonal to the target cell's, it will cause pilot contamination.

We can simplify the equation(1):

$$Y_l = [h_1\ h_2\ \ldots\ h_L] \begin{bmatrix} s_1 \\ s_2 \\ \ldots \\ s_L \end{bmatrix} + N_l,$$

Furthermore, we can express the component of the equation above as:

$$h = \begin{bmatrix} h_1 \\ h_2 \\ \ldots \\ h_L \end{bmatrix},$$

Where h is a ML x 1 vector. Similarly, Y_l and N_l can be denoted by $y = \mathrm{vec}(Y_l)$, $n = \mathrm{vec}(N_l)$, which are MT x 1 vectors.

Also, we can define a matrix:

$$S = [s_1^T \otimes I_M\ s_2^T \otimes I_M\ \ldots\ s_L^T \otimes I_M],$$

where I_M is a M x M identity matrix and S is a MT x ML matrix. $\otimes$ denote Kronecker product. Thus, we can get the received signal at l^{th} BS is:

$$Y = Sh + n.$$

3.1.2 theory of estimation

To get the accurate estimation, Bayes channel estimation maximum a posterior [9]with known prior distribution information of to be estimated variables. Here, we can express the target channel and interference channel of Bayes estimation at the l^{th} cell as:

$$h^{Bay} = (h_1, h_2, ..., h_L)^{Bay} = \text{argmax } p(h_1, h_2, ..., h_L|Y).$$

According to Bayes theorem, the conditional probability of $h_1, h_2, ..., h_L$ can be expressed as:

$$p(h_1, h_2, ..., h_L|Y) = p(h_1, h_2, ..., h_L)p(Y|h_1, h_2, ..., h_L)/p(Y)$$

$P(Y) = 1$ because Y can be known in the BS. Thus, the equation above can be equivalent to:

$$p(h_1, h_2, ..., h_L|Y) = p(h_1, h_2, ..., h_L)p(Y|h_1, h_2, ..., h_L).$$

In this model, we can assume that there is a long distance among terminals at each cell. Thus, we can think the channels between each terminal and BS are independent. The jointly probability density distribution function is:

$$P(h_1, h_2, ..., h_L) = p(h_1)p(h_2)...p(h_L),$$

where h_l subject to Rayleigh fading, $h_l \sim CN(0,R_l)$. Now, we can get the PDF of h_l and jointly PDF of all channels are:

$$p(h_l) = \exp(-h_l^H R_l^{-1} h_l)/\{pi^M(\det R_l)^M\} \text{ and}$$

$$p(h_1, h_2, ..., h_L) = \exp(-\sum_{l=1}^{L} -h_l^H R_l^{-1} h_l)/\{pi^{ML}(\det R_1 \det R_2 \ldots \det R_L)^M\}.$$

We can define two diagonal matrix: $R = \mathrm{diag}(R_1\ R_2\ \ldots\ R_L) \begin{bmatrix} R_1 & 0 & 0 & 0 \\ 0 & R_2 & 0 & 0 \\ 0 & 0 & O & 0 \\ 0 & 0 & 0 & R_4 \end{bmatrix}$ and

$$\bar{R} = R^{-1} = \mathrm{diag}(R_1^{-1}\ R_2^{-1}\ \ldots\ R_L^{-1}) = \begin{bmatrix} 1/R_1 & 0 & 0 & 0 \\ 0 & 1/R_2 & 0 & 0 \\ 0 & 0 & O & 0 \\ 0 & 0 & 0 & 1/R_L \end{bmatrix}.$$

We also need to define a scalar:

$$B = pi^{ML}(\det R_1 \det R_2 \ldots \det R_L)^M = pi^{ML}\det R.$$

From above, we can rewrite :

$$p(h) = \exp(-h^H\bar{R}h)/B$$

After the transmission of pilot signal S, the probability of y received in the BS is:

$$p(y|S,h) = \exp\{-(y-Sh)^H(y-Sh)/\sigma_n^2\}/(pi\sigma_n^2)^{ML}$$

Thus, we can conclude that:

$$p(h|y) = p(h_1\ h_2\ \ldots\ h_L|Y)=\exp(-(h^H\bar{R}h+(y-Sh)^H(y-Sh))/\sigma_n^2)/AB,$$

where A is a scalar and is equal to $(pi\sigma_n^2)^{ML}$. σ_n^2 is the average power of noise received in the target BS. According to Maximum A Posterior(MAP) criterion[10], we can denote the Bayes estimation as:

$$h^{Bay} = \mathrm{argmax}(p(h|y)).$$

Because A and B are scalar, the equation above can be equivalent to:

$$h^{bay} = \mathrm{argmax}\ \exp(-(h^H\bar{R}h+(y-Sh)^H(y-Sh))/\sigma_n^2)/AB$$

$$=\arg\min \left(h^H\bar{R}h+(y-Sh)^H(y-Sh))/\sigma_n^2\right) = \arg\min (Q(h)),$$

where $Q(h) = h^H\bar{R}h+(y-Sh)^H(y-Sh))/\sigma_n^2$. To get the result of Bayes estimation, we need to calculate the $dQ(h)/dh$:

$$dQ(h)/dh = d(h^H\bar{R}h+(y-Sh)^H(y-Sh))/\sigma_n^{2)}/dh$$

$$= \bar{R}h - S^Hy/\sigma_n^2 + S^HSh/\sigma_n^2.$$

Let the equation above equals to 0. We can get the result of Bayes estimation:

$$h^{Bay} = (S^HS + \sigma_n^2\bar{R})^{-1}S^Hy = TS^Hy,$$

where T equals to $(S^HS + \sigma_n^2\bar{R})^{-1}$. When we calculate T, we need to operate matrix inversion twice. To avoid the irreversible condition, we can transform T to:

$$T = (S^HS + \sigma_n^2\bar{R})^{-1} = (\sigma_n^2 I_{ML} + RS^HS)^{-1}R.$$

Finally, we can get the result of Bayes estimation:

$$h^{Bay} = (\sigma_n^2 I_{ML} + RS^HS)^{-1}RS^Hy.$$

3.1.3 the calculation of covariance matrix

From 3.1.2, we can find that the key to get the result of Bayes estimation is the calculation of covariance matrix. Covariance matrix of channel refers to angles-of-arrival and multipath transmission of signal.

1. Multipath transmission

Because of the reflection and scattering, signals transmitted from terminals will have multipath to get to BS. Here we assume that antennas at BS will use Uniform Linear Array(ULA) and the interval among each antennas is longer than half wavelength. We set signals have P independent paths to get to BS. The angle-of-arrival of each path is Θ_{lp}. Thus, we can get the channel vector from terminals to the target BS at l^{th} cell:

$$h_l = (1/P)^{1/2} \sum_{p=1}^{P} a(\Theta_{lp})\alpha_{lp},$$

where $\alpha_{lp} \sim CN(0,\sigma_l^2)$ and represents fading coefficient of h_l, which is independent with p and only dependent on pathloss β_l:

$$\beta_l = \alpha/d_l^{\gamma}$$

where d_l (m)represents the distance between terminals and the target BS. γ represents the pathloss index and α equals to 1.

Here we can simplify Θ_{lp} to Θ, which will not affect on the results. Thus, angle region vector $a(\Theta)$can be expressed as[11]:

$$a(\Theta) = \begin{bmatrix} 1 \\ \exp(-j2pi(D/\lambda)\cos(\theta)) \\ O \\ \exp(-j2pi((M-1)D/\lambda)\cos(\theta)) \end{bmatrix},$$

where M is the number of antennas at BS. D is the interval among each adjacent antennas. λ is the wavelength. $D \leq \lambda/2$.

2. Angle-of-arrival and covariance matrix

The information of angle distribution of multipath signals transmitted from terminals to BS is covariance matrix of channel. It contains average angle-of-arrival and angle spread value of multipath signals. Different location of terminals will have an effect on average angle-of-arrival and angle spread value.

The channel between terminals and mth antenna can be denoted:

$$h_m = \beta^{1/2}/P^{1/2}\left(\sum_{p=1}^{P} \exp(-j2pi(D/\lambda)(m-1)\cos(\theta_p) - j\varphi_p)\right),$$

where θ_p represents the angle of p^{th} path to m^{th} antenna and φ_p represents the phase of p^{th} path. Note that φ_p is uniform distribution over [-pi, pi) and phases of each path are independent. Generally, there are two conditions of distribution of angle-of-arrival of signals: uniform distribution and normal distribution.

A. Uniform distribution

We assume that angle-of-arrival region subjects to uniform distribution and the average value is Θ’. Angle region spread is Θ_Δ. Thus, we can get the PDF of angle-of-arrival is:

$$f(\Theta) = 1/2\Theta_\Delta,\ \Theta \in [\Theta' \text{-} \Theta_\Delta,\ \Theta' + \Theta_\Delta].$$

Therefore, we can get the component of covariance at m^{th} row and n^{th} column:

$$R_{m,n} = \beta/2\Theta_\Delta \int_{-\theta\Delta}^{\theta\Delta} \exp(-j2pi(D/\lambda)(m-n)\cos(\theta+\theta')),$$

where λ is the wavelength of signal and D is the interval among adjacent antennas at BS.

When angle-of-arrival region subject to uniform distribution, if the angle-of-arrival region of target terminal and interfering terminal are misaligned, the target signal and interfering signal are orthogonal over angle region. Thus, we can discriminate target signal from interfering signal by Bayes estimation so that the results of channel estimation will be more exact.

B. Normal distribution

We assume that angle-of-arrival region subjects to normal distribution. The average value of angle is Θ' and variance is σ_Θ^2. Thus, we can get the PDF of angle-of-arrival is:

$$f(\Theta) = 1/(2pi)^{1/2}\sigma_\Theta e^{-(\Theta-\Theta')2/2\sigma\Theta 2}.$$

So, the component of covariance at mth row and nth column is:

$$R_{m,n} = E\{h_m h_n^H\}$$

$$R_{m,n} = (\beta/(2*pi)^{1/2}\sigma_\Theta)\int_{-\infty}^{\infty} \exp(-j2pi(m-n)(D/\lambda)\cos(\theta+\theta')\theta^2/2\sigma_\theta^2).$$

When the angle-of-arrival region subjects to normal distribution, two arbitrary angle-of-arrival regions will always be overlapped theoretically because the range is (∞ , - ∞). However, the variance of angle-of-arrival has an effect on the value of angle-of-arrival. Thus, it will be difficult to discriminate target signal from interfering signal by Bayes estimation.

3.1.4 the results of simulation

In this simulation, I assume that different cells will use the same pilot sequence. I will focus on the performance of Bayes estimation and compare it against the performance of LS estimation to check whether Bayes estimation can be more effective than LS estimation to mitigate pilot contamination. Table 3.1 shows the parameters of simulation:

Table 3.1: The parameters of simulation about the performance of Bayes estimation

parameter	value
the number of cells (L)	2
radius of cell	1000m
the number of terminals at each cell (K)	1
Index of path fading (γ)	3.8
the number of multipath (P)	30
intervals between antennas (D)	$\lambda/2$
length of pilot sequence (T)	20
SNR	10db
location of terminals	uniform distribution

Figure 3.1 shows the result of simulation:

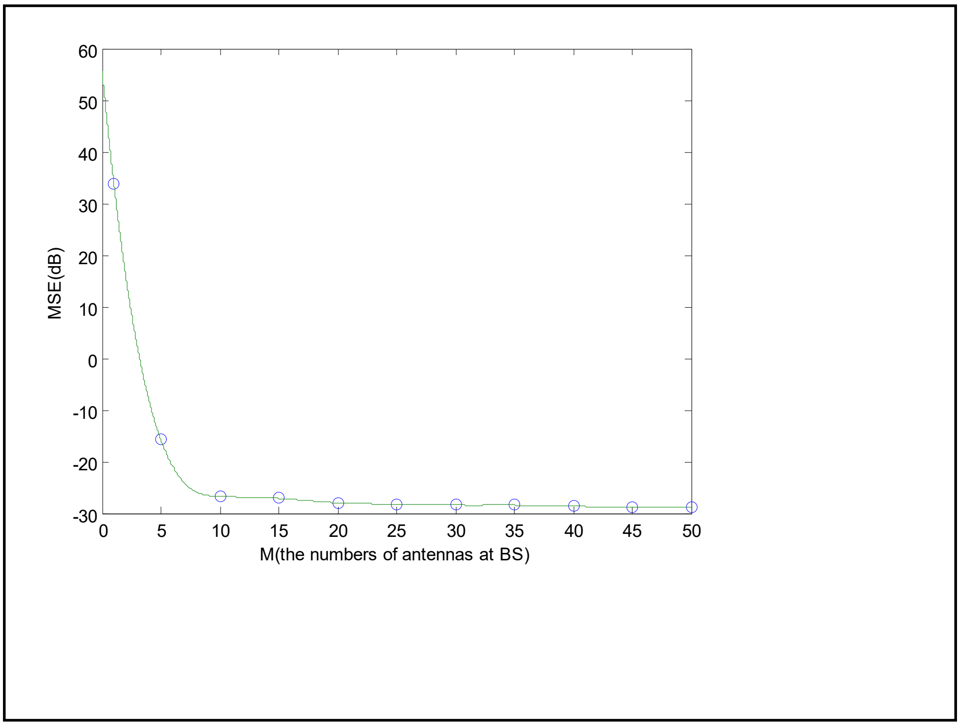

Figure 3.1: The performance of Bayes estimation

In this simulation, the angle-of-arrival is subject to uniform distribution. The average value of angle-of-arrival of the terminal at 1cell is 0 and the value at 2 cell is pi/3. Angle region spreads are all pi/12. Thus, the angle region of these two terminals are orthogonal.

From the figure 3.1, I can find that when pilot contamination exists, the performance of Bayes estimation is better than LS estimation(the performance of LS estimation has been simulated in section2.6). Because Bayes estimation can distinguish the target channel and the interference channel according to the angle-of-arrival, path loss and the information of noise. Additional, the angle-of-arrival is not overlapping. Thus, the performance of Bayes estimation is better than LS estimation and it will be helpful to mitigate the pilot contamination. And then, when the number of antennas at BS is few, increasing the number of antennas at BS will be helpful to mitigate the pilot contamination. However, after the number of antennas at BS is 10, the effect of number of antennas at BS is almost vanished.

3.2 pilot scheduling

3.2.1 system model

Pilot scheduling solve the problem that how to allocate pilot for terminals reasonably. The goal of scheduling pilot is minimizing the channel estimation error so that we can mitigate pilot contamination. We will propose some pilot scheduling algorithms including method of exhaustion and degradation based greedy algorithm.

Here, we consider a setup with L time-synchronized cells. BSs have M antennas and there are K single-antenna terminals at each cell (M>>K). The system operate TDD mode so that BS can utilize the estimation from up link to acquire CSI of down link. We assume to use fast fading channel model. h_{ijk} represents the M x 1 channel vector from k^{th} terminal at j^{th} cell to the BS at i^{th} cell. The channel vector subjects to normal distribution with the average value 0 and M x M covariance matrix R_{ijk}. To simplify symbol, H_{ij} can be represented as the M x K channel matrix from K terminals at j^{th} cell to the BS at i^{th} cell.

We let p represent SNR of pilot sequences and assume the transmission power of each pilot sequence is same. T represents the length of the pilot sequence. Thus, the pilot signal received at i^{th} cell can be represented as:

$$Y_i = p^{1/2}\sum_{j=1}^{L} H_{ij}D_j+N_i$$

It is a M x T vector. K x T vector D_j represents the pilot signals transmitted from K terminals at j^{th} cell. $N_i \sim CN(0,1)$.

If every cells use the same pilot sequences and received signals will be operated simple matched filter processing, we can get $D_j^H D_j = I$ because of the orthogonal pilot. Thus, we can get the estimation:

$$\hat{H}_{ii} = Y_i(1/p^{1/2})D_j^H = \sum_{j=1}^{L} H_{ij} + (1/p^{1/2})\check{N}_i,$$

where $\check{N}_i = N_i D_i^H$. The equation above is same with the result estimated by LS.

3.2.2 scheduling model

In this system, L cells use the same pilot. However, the pilot scheduling will has an effect on the result of channel estimation. Based on this reason, we let $D_j = P_jD$. D is a T x T channel vector containing many orthogonal pilot sequences. Scheduling matrix P_j is a K x T matrix vector and the value of elements in this matrix is 0 or 1. P_j represents the status of allocated pilots by terminals at j^{th} cell. $P_jP_j^T = I_K$.

$$Y_i = p^{1/2}\sum_{j=1}^{L} H_{ij}P_jD+N_i.$$

Thus, the result estimated by LS can be written as:

$$\hat{H}_{ii}^{LS} = Y_i(1/p^{1/2})D^HP_i^T = H_{ii} + \sum_{j\neq i}^{L} H_{ij}P_jP_i^T + (1/p^{1/2})\check{N}_i.$$

Through the observation of the above equation, we can find that the result of channel estimation is dependent on the scheduling matrix P_j. We let K_j represent the k^{th} terminal at i^{th} cell(i,k) and μ(i,k) represent the pilots allocated for k^{th} terminal at i^{th} cell.

If we have known the second order statistics R_{ijk}, we can use Bayes estimator to get the result of estimation:

$$\hat{H}_{ii}^{Bay} = (\sigma_n^2 I_{ML} + R_iS^HS)\text{-}1R_iS^H\mathrm{Vec}(p^{1/2}\sum_{j=1}^{L} H_{ij}P_jD+N_i).$$

From the equation above, covariance matrix R_i contains the information of the target channel and interfering channel. μ(i,k) will have an effect on the interfering channel part. We can observe that the result of channel estimation will be dependent on μ(i,k). Thus, it is significant to study pilot scheduling for improve the accuracy of result of channel estimation.

In this paper, we define a normalized mean square error of channel estimation as effectiveness function:

$$\Theta = \sum_{K_1}^{K_T} \sum_{j=1}^{|U_{K_j}|} MSE_I(U_{kj},K_j),$$

where $K = \{K_1, K_2, ..., K_j,..., K_T\}$ represents the pilot set containing all pilot sequences used by terminals. U_{K_j} represents the set containing terminals who use the pilot K_j and U_{K_j} represents the number of terminals in set U_{K_j} .

The principle of scheduling pilot is minimizing the sum of mean square error of the target channel estimation(or we can say minimize the effectiveness function) by using the prior information of covariance matrix and noise of all cells. Thus, considering the scheduling matrix is P_j, the best allocation of pilot can be expressed as:

$$\text{Min} \{ \sum_{K_1}^{K_T} \sum_{j=1}^{|U_{K_j}|} MSE_I(U_{kj},K_j)\}.$$

3.2.3 algorithm of pilot scheduling

3.2.3.1 random pilot scheduling

Random pilot scheduling[9] is simple to operate. It does not need to consider the correlation of spatial information of received signals at BS through up-link transmission. When system operate the data transmission, system pick a terminal from each interfering cell randomly. And then, system make this terminal and the terminals from the target cell as a group and allocate the same pilot sequences to this group at the next data transmission.

This kind of pilot scheduling is easy to realize and will not add complexity of system. However, when the system has pilot contamination, the random pilot scheduling will be effected heavily.

3.2.3.2 degradation based greedy algorithm

This algorithm is based on a degradation method proposed by C.Hellings[12]. During every iterations, one group has been allocated with pilot sequences and another one hasn't which we could call it free group. When system operate the initialization, terminals at a cell will be allocated with pilot sequences While terminals at other cells will be free. And then, system will allocate pilot sequences for the terminals at the rest (L-1) cells.

We will take an example of i^{th} cell:

The first step: calculating the best pilot of every terminals.

System will calculate the effectiveness of every allocated pilots for terminals at ith cell. The pilot that will have a minimum effectiveness will be defined as the best pilot of terminal.

The second step: calculating the second best pilot and degradation of every terminals.

System will calculate the effectiveness when terminals are allocated with pilots besides the best pilot. The pilot besides the best pilot that will have a minimum effectiveness will be defined as the second best pilot. And then, system needs to calculate the degradation of effectiveness function of terminals allocated the best pilot and the second best pilot respectively.

The third step: we define the terminal that has a maximum degradation as the most sensitive terminal. System need to allocate this terminal with the best pilot.

The forth step: repeating the first step to the third step until all terminals are allocated with pilots.

In order to calculate the effectiveness of allocated pilot μ (i,k), we define U_{ik} represents the terminal set that use this pilot and the instantaneous value of μ(i,k) is:

$$\Theta'ik = \sum_{j=1}^{|U_{K_j}|} MSE_I(U_{kj},K_j),$$

where $K_j = \mu$(i,k). We define F represents the set that the terminals haven't been allocated with pilots. At every cells, for every terminals that haven't been allocated with pilots, we have $(i,k) \in F$. We define P_i represents the pilots that can be used at i^{th} cell. Thus, we can express the best pilot as:

$$p_{ik}^* = \text{argmin } \Theta'_{ik}(K_1,...,K_p \cup \{(i,k)\},...,K_T),$$

where $p \in P_i$.

When terminals are allocated with the best pilot and the second best pilot, the degradation can be expressed as:

$$d_{ik} = \Theta'_{ik}(K_1,...,K_p \cup \{(i,k)\},...,K_T)\text{-argmin } \Theta'_{ik}(K_1,...,K_p \cup \{(i,k)\},...,K_T),$$

where $p \in P_i$ and $p \neq p_{ik}^*$.

We define the terminals (i*,k*)that have maximum degradation is the most sensitive terminal:

$$(i^*,k^*) = \mathrm{argmax}\ d_{ik},$$

where $(i,k) \in F$.

Thus, the terminal (i^*,k^*) will be allocated with the best pilot:

$$K_{p_{i^*k^*}^*} \leftarrow K_{p_{i^*k^*}^*} \cup \{(i^*,k^*)\}.$$

We need to remove the terminal (i^*,k^*) from the set P_{i^*}:

$$P_{i^*} \leftarrow P_{i^*} \backslash \{p^*_{i^*k^*}\}.$$

We also need to remove the terminal (i^*,k^*) from the set F:

$$F \leftarrow F \backslash (i^*,k^*).$$

Repeating these steps until the set F become a empty set.

3.2.4 the results of simulation

In this section, I will simulate the effect of pilot scheduling by degradation based greedy algorithm on the mitigation of pilot contamination and I will focus on the comparison between the performance of system by Bayes estimation with pilot scheduling and without pilot scheduling. Additionally, I will simulate the effect of the number of antennas at BS on the performance of mitigation of pilot contamination by degradation based greedy algorithm. Table 3.2 shows the parameters of this simulation:

Table 3.2: The Parameters Of Simulation About The Performance Of System With Pilot Scheduling Based On Greedy Algorithm.

parameter	value
the number of cells (L)	2
radius of cell	1000m
the number of terminals at each cell (K)	10
Index of path fading (γ)	3.8
the number of multipath (P)	30
interval between antennas (D)	$\lambda/2$
length of pilot sequences (T)	20
SNR	10db
location of terminals	uniform distribution

Figure 3.2 shows the result of simulation:

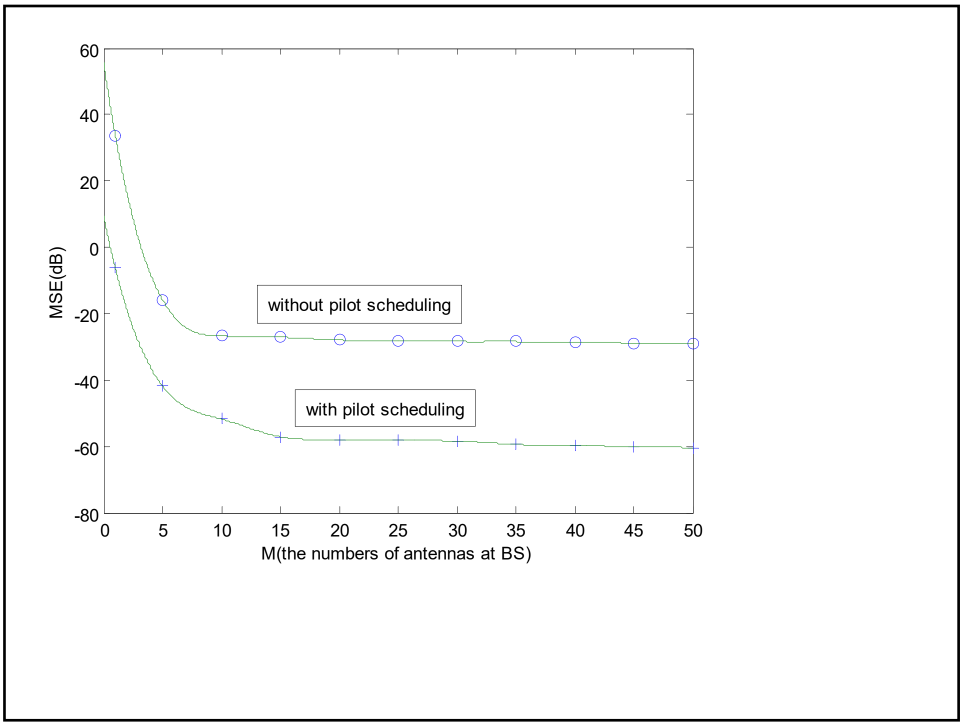

Figure 3.2: The Comparison Between The Performance Of System By Bayes Estimation With Pilot Scheduling And Without Pilot Scheduling.

From the figure 3.2, I can find that the performance of system by Bayes estimation with pilot scheduling is better than that without pilot scheduling. However, after the number of antennas is 15, increasing the number of antennas at BS will be helpless for mitigation of pilot contamination.

REFERENCES

[1] Baker M. From LTE-advanced to the future. Communications Magazine, IEEE,2012, 50(2): pp 116-120.

[2] Marzetta T L. Noncooperative cellular wireless with unlimited numbers of base station antennas. Wireless Communications, IEEE Transactions on, 2010, 9(11): pp 3590-3600.

[3] Erik G. Larsson, Ove Edfors, Fredrik Tufvesson, and Thomas L. Marzetta. Massive MIMO For Next Generation Wireless Systems. January 23, 2014, pp 2-3.

[4] Marzetta T L. Massive MIMO: An Introduction. Bell Labs Technical Journal, 2015, 20: pp 11-22.

[5] Emil Bjornson, Erik G Larsson, and Thomas L. Marzetta. Massive MIMO: Ten Myths and One Critical Question. August 18, 2015.

[6] C. Shepard, H. Yu, N. Anand, L. E. Li, T. L. Marzetta, R. Yang, and L. Zhong, “Argos: Practical many-antenna base stations,” in ACM Int. Conf. Mobile Computing and Networking (MobiCom), Istanbul, Turkey, Aug. 2012.

[7] F. Kaltenberger, J. Haiyong, M. Guillaud, and R. Knopp, “Relative channel reciprocity calibration in MIMO/TDD systems,” in Proc. of Future Network and Mobile Summit, 2010.

[8] Vidit Saxena. Pilot Contamination and Mitigation Techniques in Massive MIMO Systems. October 24, 2014, pp 28-29.

[9] H.Yin, D. Gesbert, M. Filippou, and Y. Liu. A coordinated approach to channel estimation in large-scale multiple-antenna systems. IEEE, 2013, pp 264-273.

[10] Gray R M, Davisson L D. An introduction to statistical signal processing. Cambridge University Press, 2004.

[11] Tsai J A, Buehrer R M, and Woerner B D. The impact of AOA energy distribution on the spatial fading correlation of linear antenna array. Vehicular Technology Conference, 2002, pp 933-937.

[12] Hellings C, Utschick W, and Joham M. Power minimization in parallel vector broadcast channels with separate linear precoding. Proceedings of the Nineteenth European Signals Processing Conference, Barcelona, Spain, 2011.