2022

POWER ALLOCATION AND MASSIVE MIMO CHANNEL MODELING FOR 5G WIRELESS COMMUNICATIONS

Zhangliang Chen

POWER ALLOCATION AND MASSIVE MIMO CHANNEL MODELING FOR 5G WIRELESS COMMUNICATIONS

by

ZHANGLIANG CHEN

Presented to the Faculty of the Graduate School of

The University of Texas at Arlington in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT ARLINGTON

May 2022

ACKNOWLEDGEMENTS

The first person I'd like to thank most is my supervising professor Dr. Qilian Liang, for constantly motivating and leading me as well as encouraging me. In addition, his invaluable advising during my doctoral studies is the treasure of my whole life. His enthusiasm and attitude toward research have deeply influenced me and helped me to do better research. Also, I want to thank Dr. Ioannis D. Schizas, Dr. Yan Wan , Dr. Yuze (Alice) Sun and Dr. Chenyun Pan for their interests in my research and for having time to attend my comprehensive exam and Dr. Ioannis D. Schizas, Dr. Yan Wan to be in my dissertation defense committee.

I would also like to thank all the teachers who taught me during my undergraduate and Ph.D. studies, first in China, then in the Unites States. I am grateful to Prof. Hongzhi Qi and Prof. Dong Ming for encouraging me and inspiring me to pursue my doctoral studies.

The great thank is expressed to my labmates, Dheeral, Chengchen and Swarada for their interest in my research and for the helpful discussions and invaluable comments.

Finally, I would like to express my deep gratitude to my parents who have encouraged me, inspired me and sponsored my undergraduate and graduate studies. I am also extremely grateful to my wife for her sacrifice, encouragement and patience.

May 5, 2022

ABSTRACT

POWER ALLOCATION AND MASSIVE MIMO CHANNEL MODELING FOR 5G WIRELESS COMMUNICATIONS

ZHANGLIANG CHEN, Ph.D.

The University of Texas at Arlington, 2022

Supervising Professor: Qilian Liang

To meet the demand for wireless communication transmission rate to reach a thousand times of the existing system in 2021, the fifth-generation (5G) mobile communication system was developed. Compared with Long-Term Evolution (LTE), the 5G mobile communication needs to be transmitted in wireless breakthrough innovation in technology to achieve the goal of ten times increase in spectrum efficiency and power efficiency. As a time series analysis tool based on vector autoregression, Granger causal analysis originated in the field of econometrics, and its staged generalized transfer entropy (TE) based on conditional co-information in information theory has been widely used in recent years in data analysis field.

Among them, further tapping the spatial multiplexing capability of multiple antennas is a key way to achieve 5G. Configure large-scale antenna arrays or multiple at the access point are interconnected by optical fibers to establish a large scale distributed Massive Multiple Input Multiple Output (MIMO) system. A significant improvement of the spectrum efficiency can be achieved by implementing the Massive MIMO system. Also, the Massive MIMO channel modeling is necessary both in the theoretical research phase

of Massive MIMO as well as in the application phase of the systems. Therefore, Massive MIMO channel modeling is an important content in the research.

This dissertation investigates the 5G channel forecasting and power allocation, as long as channel modeling by Massive MIMO.

The first part of this dissertation focuses some analysis of 5G channel forecasting and power allocation. Granger Causality has been derived to confirm the relationship between two random 5G channel coefficients. Transfer entropy is proposed to predict channel coefficients, and the accuracy of the prediction is computed by using root mean square error (RMSE) and Cramer-Rao lower bound (CRLB). Furthermore, an Inverse Water Filling (IWF) algorithm is applied to perform the power distribution based on the predicted channels.

The channel modeling for Massive MIMO communications at 28G Hz is mainly presented in the second part. The Indoor and through-wall (TW) Massive MIMO wireless communication scenario is used. We can consider the Line-of-sight (LOS) communication is achieved under indoor scenario and Non-Line-of-sight (NLOS) communication under TW scenario. The center frequency of 28G Hz is used on Transmission side and the actual RF signal is collected. Then, a Space Alternating Generalized Expectation maximization (SAGE) algorithm is used to extract large-scale parameters to obtain the millimeter wave Urban Micro (UMi) scale parameters and small-scale parameters of the LOS and NLOS of the channel on the basis of the actual channel data. Finally, the departure angle, angle extension of the transmitted signal, arrival angle, angle extension of the received signal, angular power spectrum, Doppler power spectrum and other parameters are used to model the channels.

Third, a forecasted channel system on the basis of LSTM neural network is devised to deal with the problem of gradient disappearance in the Recurrent Neural Network (RNN) specialized in dealing with time series problems. Forecasting was made using 5G channels simulated by the NYU Wireless Communications Simulator. Then, the designed arithmetic is performed and the performances are compared with the real channel, and small RMSE is obtained, which shows the high accuracy of the prediction. After that, based on the channel forecasted by using the LSTM network, the power allocation based on cooperative communication is derived, and compare the power results without cooperative communication. For the single-relay and multi-relay cooperative scenarios, a power allocation schema under the end goal to maximize the information transmission rate (ITR) at the destination node is proposed. The realization process of this scheme is constructing the information transmission rate function of the destination node under the condition of setting the total power transmitted is a fixed value by the node. The objective function can be considered as the optimization of a convex function, which can obtain the optimal solution by the Lagrangian formula. When the ITR of the destination node is the maximum, the source node and relay node achieved the optimal transmission power values. Therefore, system performance is improved by optimizing power allocation method of the transmitting nodes. By comparing with other schemes, it is verified that the power allocation scheme proposed in this dissertation has better performance and saves system resources.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

CHAPTER 1

INTRODUCTION

5G is a new generation of mobile communication system developed for the needs of mobile communication after 2020 [1]. According to the development law of mobile communication, 5G will have ultra-high spectrum utilization and energy efficiency, and increase the magnitude of LTE mobility or higher. Communicate in terms of transmission rate and resource utilization. Power efficiency has long played an important role in mobile communication devices. The high power efficiency of the device extends battery life and has been an important element of the mobile communications revolution.

At the same time, during the propagation process of wireless communication signals, from transmitter to receiver through a complex propagation environment, large-scale fading (includes path loss (PL) and shadow fading) and small scale fading are caused by some propagation mechanisms such as direct radiation, reflection, scattering, and diffraction [1]. Signal fading leads to distortion of the received signal, which affects the communication quality of the wireless communication [2]. Understanding propagation characteristics of wireless communication channels and their influence on wireless communication signals is crucial for the programming and testing of wireless communication systems. The wireless channel modeling is on the basis of fully understanding the propagation characteristics of wireless signals and characterizes the factors of channels through a series of parameters, which is an abstract simulation of the wireless propagation environment.

1.1 Goals of the Dissertation

1.1.1 Correlation Test Based on Granger Causality

For the things that have unclear causal relationships, the Granger Causality method can be used to perform statistical tests [3].

In this dissertation, the process of testing relationship between 5G Channels based on Granger Causality test is discussed. First, the causality of two random and relatively independent 5G channels was tested by Granger causality to test whether a causal relationship existed between the two channels. If there is no causal relationship between the channels, that is, the change of channel 1 does not affect the change of channel 2, or vice versa, the transfer entropy theory cannot be further applied. Therefore, the channel prediction described later is based on the causal relationship between the two channels.

1.1.2 5G Channel Forecasting Based on Transfer Entropy

In the experiment, we got a conclusion: 5G channel coefficients obey Gaussian distribution. Since Granger causality and transfer entropy are equivalent under Gaussian variables [4], we prove that transfer entropy can be used to predict 5G channels.

Transfer entropy was first proposed by Schreiber [5]. It is a time-asymmetric non-parametric information measure based on conditional co-information. Although it is quite different from Granger causality based on the ARMA model, it has a significant advantage that transfer entropy does not require model assumptions. In fact, both of them are essentially derived from Wiener's causal relationship construction. The key idea is to increase the weight of the new variable with respect to the historical information to reduce the uncertainty of the forecast [6] [7]. The metric used to measure this uncertainty change is causality.

Under the condition of Granger Causality is determined between two channels, transfer entropy algorithm is derived to forecast channels.

1.1.3 Power Allocation Based on Inverse Water-filling

The basis of the IWF algorithm is: under certain criteria, the transmit power is adaptively allocated according to the channel conditions. Generally, when the channel condition is good, more power will be allocated to the sub-channel, and when the channel condition is bad, less power will be allocated or even not allocated. The transmission rate is maximized through such a power allocation mechanism. Therefore, in order to be able to use the IWF algorithm for power allocation, the transmitter must know the channel state information (CSI). The IWF algorithm can effectively increase the channel capacity.

In this dissertation, the IWF algorithm is used on forecasted channels. The channel capacity based on IWF algorithm and equal gain (EG) algorithm is compared in this dissertation.

1.1.4 Channel Modeling for Massive MIMO at 28GHz Under Indoor and Through-Wall Scenarios

According to Shannon's formula $C = B \cdot log_2(1 + S/N)$, the capacity of the channel is directly proportional to the available bandwidth, and As the available bandwidth expands, so does the channel capacity [2]. Therefore, in the case of sufficient spectrum resources, increasing the transmission bandwidth is the most direct and effective method to increase the channel capacity and rate. However, the current low-frequency spectrum resources are occupied by traditional Bluetooth, wireless LAN and other services, which cannot provide more continuous bandwidth to achieve 5G requirements. Therefore, researchers and scholars have turned their attention to the millimeter wave band with abundant spectrum resources and most of the frequency bands have not been allocated. The millimeter frequency band (24-300GHz), as a core technical component of the 5G wireless communication system, allows to occupy more spectrum to support the larger data steam requirements of various multimedia services.

Although broadband millimeter wave communication can solve the needs of 5G high transmission rate through its high bandwidth advantages, due to the large difference between the channel characteristics of the high frequency band and the traditional band, there will be many problems to be solved in the realization of high frequency communication systems. And challenges. In the low- and mid-band communication system, there is usually no need to consider the influence of the atmosphere, rainfall, and the human body on its communication quality. However, the millimeter wave frequency band is high and the wavelength is small, not only the path loss is large, but also the additional loss caused by the atmospheric rainfall and other factors, which causes its propagation loss to be so high that it cannot be transmitted over a long distance. Moreover, the penetration ability of millimeter waves is poor. Research has shown that in millimeter wave communication, the blockage of the human body may also cause signal interruption, so millimeter waves cannot be used in traditional indoor to outdoor or outdoor to indoor communication scenarios [8]. The channel characteristics of millimeter wave have severely restricted the propagation range and use scenarios of millimeter wave communication systems. Although there will be many obstacles on the way to establish high-frequency communication systems, it will also bring many opportunities. For example, because the antenna size increases proportionally with the wavelength increasing, the short wavelength of the millimeter wave can be used to effectively reduce antenna size, and it is easy to implement large-scale antenna arrays, beamforming and other technologies to compensate for its larger propagation loss.

In order to make use of millimeter-band resources, it is first necessary to have an in-depth understanding of the propagation characteristics of this new frequency band and establish a channel model to determine the most suitable communication scenarios and link types for this frequency band. Traditional commonly used channel models are mainly divided into statistical channel models and deterministic channel models. The statistical channel model is mainly obtained by fitting the measured data, while the deterministic

channel model is often obtained by using ray tracing technology [9]. Compared with the deterministic model, the parameters of the statistical model are extracted by algorithms based on measured data, which can better reflect the propagation characteristics of wireless signals in the real environment. However, the current mobile communication wireless channel models are established through statistical measurements in the low frequency band, and cannot be directly applied to the high frequency millimeter wave system. Different from the traditional low frequency communication below 6GHz, millimeter wave communication has the characteristics of large path loss, easy to be affected by air and rain, poor diffraction ability, high scattering and very sensitive to dynamic environment. When building a millimeter wave mobile communication system, it is necessary to select a suitable communication frequency band and establish a millimeter wave 5G wireless channel transmission model through a large number of tests on different millimeter wave frequency bands, scenarios and links, which has very important and far-reaching significance.

In this part, there are two goals are achieved. 1. Use the 5G high-frequency channel mobile platform to generate and collect the 28GHz indoor LOS and through-wall NLOS millimeter-wave signals, showing the specific channel measurement planning scenarios and settings. At the transmitter and receiver, 64 (Uniform Rectangular Array, URA) matrix antennas are used to collect real communication data for the LOS and NLOS paths respectively, and the SAGE algorithm is used to extract large-scale parameters to obtain the UMi scale parameters and small-scale parameters of the LOS and NLOS of the channel in the scenario. 2. Perform path loss (PL) modeling on the measured data to obtain the millimeterwave path loss model and the corresponding shadow fading in this scenario, and obtain the Delay Spread (DS) and Angle Spread (AS) of the channel through large-scale parameter modeling And other statistical characteristics. By analyzing the power delay profile (PDP) and angular power spectrum of the LOS and NLOS paths, observe which azimuth and an-

gles the signal has the highest received power, and judge the degree of time dispersion and angular dispersion according to the distribution functions of DS and AS.

1.1.5 Channel Forecasting and Power Allocation Based on LSTM Network and Cooperative Communication

Channel forecasting and power allocation are very important techniques for wireless communication. If the state of the channel can be predicted, the power in the channel will be allocated reasonably, more power will be distributed on the sub-channel with good state, and less power will be distributed on the sub-channel with bad state, so as to realize the high efficiency of power allocation.

Wireless signal can be considered as a time series, and Recurrent Neural Network (RNN) is one of the most potential tools for time series modeling. The neural network model of RNN adds the features of time series. The hidden layer of the RNN model has feedback edges. The input of each hidden layer includes both the current sample features and the information brought by the previous time series. It can achieve high-accuracy predictions performance.

However, RNNs also have some drawbacks. For standard RNN networks, the time span of information that can be used in practice is very limited. When we use information from a relatively recent time point to solve the task of the current moment, RNN can effectively learn the information of the historical moment. However, when we need to use historical information that differs from the current moment information for a long time, the ability of RNN to learn information will be weakened, which is the problem of RNN's gradient disappearance. Long Short Term Memory (LSTM) can be considered as a special form of RNN network, which is superimposed a long term memory function on the RNN, the persistence of the RNN network can be maintained, which allows the long-term dependence of the model on the neural network to be realized. The biggest advantage here is that

the LSTM network itself can do the function of remembering information for a long time, which is independent of what it learns through data training. As explained in the previous paragraph, the vanishing gradient phenomenon is ubiquitous in traditional RNN networks. The LSTM network was born to solve this shortcoming. A long-term memory function that keeps information from decaying is superimposed on the traditional RNN network.

In this section, the final goal is deriving cooperative communication to achieve high efficient power allocation based on the forecasted 5G channel by using the LSTM network. In chapter 4, We compared the performance of power allocation based on cooperative communication with equal power distribution method.

1.2 Future Work

The scenario measured in the channel modeling part in this dissertation are indoor scenes and indoor through-wall scenes. The main propagation environment is only indoor objects such as desks and computers. It does not consider factors such as moving people and receiving users moving. In practice, users are basically moving, and the blockage of the human body also affects high frequency communications. Moreover, outdoor scenes also have some important influencing factors for high-frequency signal propagation such as foliage, rain attenuation, etc. Because the millimeter-wave antenna array has strong spatial directivity, before data transmission, the receiving and transmitting ends need to search for the DOA of signal propagation, and select the optimal spatial transmission direction to achieve the maximum gain. Therefore, in the next step, 5G ultra-dense networking and resource allocation will be a meaningful research direction.

Besides that, recently, academia and industry have been studying non-orthogonal multiple access (NOMA) technology to cope with the future demand for massive access and ultra-high capacity. In the traditional OMA method, wireless signals containing in-

formation of multiple users are allocated to mutually independent orthogonal domains for transmission, including but not limited to time domain, frequency domain, code domain, and the like. Since the OMA technology only allows each orthogonal resource to transmit information of one user, this limitation results in the cell throughput and the number of device connections not being too large. [10]. In addition, since only one user can be served on the same time-frequency resource, and the traditional power allocation strategy assigns power to users with strong channel conditions first, so user fairness and spectrum efficiency are low. However, NOMA is different from these traditional solutions. The core concept of power domain NOMA is to achieve multiple access in the power domain through different power levels on the same time-frequency resource block, which can allocate more to users with poor channel conditions [11]. In addition, the same resource block can also serve users with strong channel conditions, which greatly improves the spectrum efficiency and obtains a good balance between system throughput and user fairness. This also happens to meet the technical requirements of large-scale device connection in Internet of Thins (IoT) application scenarios [12]. NOMA uses Superposition Coding (SC) at the sending end to send out the information of different users at the same time, and adopts the Successive Interference Cancellation (SIC) at the user receiving end. For users with strong channel conditions, the signal is first The more powerful signal is used as the useful signal, and other signals are used as interference for demodulation, and then the useful signal obtained by demodulation is subtracted to obtain the user's information; while for users with weak channel conditions, the information of other users is used as The noise is directly demodulated. In addition, the advantages of NOMA have been verified in different wireless communication systems, such as broadcast channel, full duplex channel and physical layer security, etc [13]. Based on that, Massive MIMO NOMA and millimeter wave communications based on NOMA will be some good future research topics.

CHAPTER 2

POWER ALLOCATION FOR 5G WIRELESS COMMUNICATIONS

2.1 Granger Causality Test

Granger causality, as a method to measure the mutual influence between time series, has been favored in the past many years. In 1969, Granger proposed a definition of causality from the perspective of econometrics: With two time series x_t, y_t, consider the linear projection of x_t on the past values of x and y [14]:

$$x_t = \sum_{j=1}^{\infty} h_j x_{t-j} + \sum_{j=1}^{\infty} v_j y_{t-j} + \epsilon_t. \tag{2.1}$$

Where, for any positive integer k, $E\epsilon_t x_{t-k} = E\epsilon_t y_{t-k} = 0$ [15], ϵ_t is noise.

If for a given past value of all x, the past value of y contributes to predicting x, ie. there is at least one j_0 such that $vj_0 \neq 0$, then the variable y is the Granger sense reason for x [14].

According to this definition, in 1972 Sims proposed a proposition that there is no causal relationship which is, Let (x_t, y_t) be a zero-mean joint covariance stationary sequence, then y is not a necessary and sufficient condition for the Granger's cause of x [16]. There exists a vector moving average representation of the lower triangle:

$$\begin{pmatrix} x_t \\ y_t \end{pmatrix} = \begin{pmatrix} a(L) & 0 \\ b(L) & h(L) \end{pmatrix} \begin{pmatrix} \epsilon_t \\ u_t \end{pmatrix} \tag{2.2}$$

where ϵ_t and u_t are zero-ranging sequence-independent process, and for any integer $t, s, E\epsilon_t u_s = 0$, $a(L), b(L), h(L)$ is the non-negative exponent of L Polynomials on the side, ie., $a(L) = \sum_{j=0}^{\infty} a_j L^j, b(L) = \sum_{j=0}^{\infty} b_j L^j, h(L) = \sum_{j=0}^{\infty} h_j L^j$, in which L is a delay operator defined by $Lx_t = x_t - 1$ [14].

If y is not the Granger cause of x, Eq. 2.2 holds. Rewrite Eq. 2.2:

$$x_t = a(L)\epsilon_b$$

$$y_t = b(L)\epsilon_b + h(L)u_t$$

From the joint covariance stationarity of (x_t, y_t), the inverse polynomial $a^{-1}(L)$ of $a(L)$ exists, and it is non-negative power-universal to the delay operator L [14].

Let $b(L) = b(L)a^{-1}(L), e_t = h(L)u_t$, and then there are:

$$\begin{aligned} y_t =& b(L)a^{-1}(L)x_t + h(L)u_t \\ =& d(L)x_t + e_t \end{aligned} \tag{2.3}$$

This formula shows that the regression residual of y_t which is e_t on the current and past x (ie $x_t, x_{t-1}, x_{t-2},$) is not related to the future x_t [17]. In other words, given the current and past x_t, the future x_t does not affect y_t, i.e., Eq. 2.3 is a representation under the condition that y has no feedback effect on x. Observe that Granger's Causality definition assumes that future events cannot cause current or past events. Therefore, the real meaning of Granger's Causality is the 'preceding' relationship in time, not the causality in the usual sense [14].

According to his above proposition, Sims proved a theorem that facilitates Granger Causality test as follows [18]. Assume that (x_t, y_t) is a zero-mean joint covariance stationary sequence, ϵ_t is a white noise sequence, considering the linear projection of y_t over the whole x process [14]:

$$y_t = \sum_{j=-\infty}^{\infty} b_j x_{t-j} + \epsilon_t$$

where, for any integer j, $E\epsilon_t x_{t-j} = 0$. Then y is not Granger's cause of x, i.e., in (1), for any j, $v_j = 0$, if and only if it is an arbitrary negative integer j, $b_j = 0$ [14].

Granger Causality can be tested using the following metrological methods. Let x_t, y_t be the covariance stationary sequences, set up a regression model of x_t for lags of y and x [14]:

$$x_t = c + \sum_{i=1}^{n} h_t y_{t-i} + \sum_{j=1}^{n} a_j x_{t-j} + \epsilon_t \tag{2.4}$$

where, c is constant.

Among them, the choice of lag period n is relatively arbitrary.Then the judgment that 'y is not the cause of x' is equivalent to performing an F-test on the null statistical hypothesis [14]:

$$H_0 : h_1 = h_2 = ...h_n = 0$$

Let SSR_1(Residual Sum of Squares) and SSR_0 represent the residual sum of squares of regression model(4) and the model when the null hypothesis H_0 holds [14]. Then, the test statistic

$$F = \frac{(SSR_0 - SSR_1)/n}{SSR_1/(N - 2n - 1)}$$

follows the F-distribution with the first degree of freedom n and the second degree of freedom $N - 2n - 1$ under the condition that H_0 holds. Where N is the number of sample data [14].

When the value of the above F statistic is greater than the critical $F_\alpha(n, N - 2n - 1)$ of the F distribution below the significance level $1 - \alpha$, y can be considered as the Granger cause of x under the confidence of $1 - \alpha$ [14].

We perform a Granger causality test on two random and relatively independent 5G channels in order to verify the existence of causality between the two channels. The channels are simulated from the New York University NYU WIRELESS center open source 5G channel simulator [14].

2.2 5G Channel Forecasting Based on Transfer Entropy

2.2.1 Introduction to Transfer Entropy

In this dissertation, we found that the 5G channel coefficient followed the Gaussian distribution [14]. Based on the equivalence between Granger causality and transfer entropy under Gaussian variables [4], we could use the transfer entropy to forecast the 5G channel coefficient [14].

Transfer Entropy is a time-asymmetry non-parametric information measure based on conditional co-information proposed by Schreiber [5]. Although it differs from Granger causality based on the vector auto-regression model, transfer entropy does not use any model assumptions [14]. Both are essentially derived from Wiener's construction of causality, that is, the addition of historical information of new variables, which reduces the uncertainty of the prediction of the target variable [6] [7]. And causality is a measure of this change in uncertainty [14].

The definition of transfer bribery: For two discrete random variables X and Y, the probability distribution functions are $p(x)$ and $p(y)$ respectively [14]. The joint probability of events x and y occurring at the same time is p(x,y), then the Shannon entropy H(x) is defined as [19].

$$H(x) = -\sum_{x} p(x) log p(x) \tag{2.5}$$

The conditional probability of Y with X is:

$$p(y|x) = p(x,y)/p(x) \tag{2.6}$$

In Eq.(6) if $p(y|x) = p(y)$ x and y are independent. Then $p(x,y) = p(y|x)p(x) = p(y)p(x)$ [14]

The joint entropy of X and Y is

$$H(X,Y) = -\sum_{x,y} p(x,y) log p(x,y) \tag{2.7}$$

The conditional entropy of X with Y is

$$H(X|Y) = -\sum_{x,y} p(x,y) log p(x|y) \tag{2.8}$$

The Mutual Information(MI) contained between X and Y is the output of the two systems as though they were independent as opposed to their 'actual' relationship [14]

$$\begin{aligned} M(X;Y) &= H(X) + H(Y) - H(X,Y) \\ &= -\sum_{x,y} p(x,y) log(p(x)p(y)) + \sum_{x,y} p(x,y) log(p(x,y)) \\ &= \sum_{x,y} p(x,y)[log(p(x,y)) - log(p(x)p(y))] \\ &= \sum_{x,y} p(x,y) log \frac{p(x,y)}{p(x)p(y)} \end{aligned} \tag{2.9}$$

However, MI is not effective at predicting future events from current data since it is symmetric, $M(X,Y) = M(Y,X)$. And it does not indicate which way the information is flowing [14]. These shortcomings may be remedied by time shifting one of the variables [14]. Transfer Entropy (TE) (Schreiber 2000 [5]) is based on rates of entropy change, it captures some of the dynamics of a system.

Suppose two systems which generate events. We define an entropy rate which is the amount of additional information required to represent the value of the next observation of one of the systems [20]:

$$h_1 = -\sum_{x_{n+1}} p(x_{n+1}, x_n, y_n) log_a p(x_{n+1}|x_n, y_n) \tag{2.10}$$

Suppose that value of observation x_{n+1} was not dependent on the current observation y_n:

$$h_2 = -\sum_{x_{n+1}} p(x_{n+1}, x_n, y_n) log_a p(x_{n+1}|x_n) \tag{2.11}$$

The quantity h_1 represents the entropy rate for the two systems, and h_2 represents the entropy rate assuming that x_{n+1} is independent of y_n.Thus, we get transfer entropy:

$$\begin{aligned} h_2 - h_1 &= -\sum_{x_{n+1},x_n,y_n} p(x_{n+1}, x_n, y_n) log_a p(x_{n+1}|x_n) \\ &+ \sum_{xn+1,x_n,y_n} p(x_{n+1}, x_n, y_n) log_a p(x_{n+1}|x_n, y_n) \\ &= \sum_{x_{n+1},x_n,y_n} p(x_{n+1}, x_n, y_n) log_a (\frac{p(x_{n+1}|x_n, y_n)}{p(x_{n+1}|x_n)}) \end{aligned} \tag{2.12}$$

There are actually two equations for the transfer entropy, because it has an inherent asymmetry in it.

$$T_{J\to I} = \sum_{x_{n+1},x_n,y_n} p(x_{n+1}, x_n, y_n) log_a (\frac{p(x_{n+1}|x_n, y_n)}{p(x_{n+1}|x_n)}) \tag{2.13}$$

$$T_{I\to J} = \sum_{y_{n+1},x_n,y_n} p(y_{n+1}, x_n, y_n) log_a (\frac{p(y_{n+1}|x_n, y_n)}{p(y_{n+1}|y_n)}) \tag{2.14}$$

Then with substitutions [14]:

$$\begin{aligned} p(x_{n+1}|x_n, y_n) &= p(x_{n+1}, x_n, y_n)/p(x_n, y_n) \\ p(x_{n+1}|x_n) &= p(x_{n+1}, x_n)/p(x_n) \end{aligned}$$

our equations become [20]

$$T_{J\to I} = \sum_{x_{n+1},x_n,y_n} p(x_{n+1},x_n,y_n)\cdot log(\frac{p(x_{n+1},x_n,y_n)\cdot p(x_n)}{p(x_n,y_n)\cdot p(x_{n+1},x_n)}) \tag{2.15}$$

$$T_{I\to J} = \sum_{y_{n+1},x_n,y_n} p(y_{n+1},x_n,y_n)\cdot log(\frac{p(y_{n+1},x_n,y_n)\cdot p(y_n)}{p(x_n,y_n)\cdot p(y_{n+1},y_n)}) \tag{2.16}$$

In Eq. 2.16 and 2.16, J represents 5G channel 1, and I represents 5G channel 2 [14]. We use TE algorithm to forecast the channels based on two equations [14]. Since the row 5G channels are too long to plot, we choose part of the forecasted channel randomly [14].

2.2.2 Forecasting Accuracy Analysis - RMSE And CRLB

2.2.2.1 Root Mean Squre Error of Forecasted Channel

First we apply the RMSE metric between real 5G channel and forecasted channel. The RMSE can be computed as [14]

$$RMSE = \sqrt{\sum_{n=1}^{N}((x_n - x_n')^2)/N} \tag{2.17}$$

where x_n is the observed value, which is the forecasted channel coefficient in our work [14]. x_n' is the true value, which represents the real 5G channel coefficients here. N is the total number of channel coefficient [14]. In order to avoid specialty, we proposed 2000 pairs of Real 5G channels as training groups and 1000 pairs of Real 5G channels as test groups to do the forecasting based on transfer entropy and calculated the average RMSE. We will show the RMSE performance in chapter 2.4 Simulations and Performance Analysis [14].

2.2.2.2 Cramer-Rao Lower Bound

Cramer-Rao Lower Bound (CRLB) can be used to calculate the best estimation accuracy that can be obtained in unbiased estimation, so it is often used to calculate the best estimation accuracy that can be achieved by theoretically, and to evaluate the performance of parameter estimation methods (Whether it is close to the lower limit of CRLB) [21].

The 5G channel follows a Gaussian distribution with a variance γ^2. Assume $x = A + \omega$, ω is a Gaussian noise, $\omega \sim N(0, \sigma^2)$ [14].

The pdf of x_m is [22]

$$f(x_m) = \frac{1}{\sqrt{2\pi(\gamma^2 + \sigma^2)}} exp\left[-\frac{(x_m - m)^2}{2(\gamma^2 + \sigma^2)}\right] \tag{2.18}$$

where x_m is a random channel coefficient, σ^2 is the variance of Gaussian noise ω [14].

Let $\mathbf{x} \triangleq [x_1, x_2, ..., x_M]$, then the pdf of $\mathbf{x}$ is [14]

$$\begin{aligned} f(\mathbf{x}) &= \prod_{m=1}^{M} f(x_m) \\ &= \prod_{m=1}^{M} \frac{1}{\sqrt{2\pi(\gamma^2 + \sigma^2)}} exp\left[-\frac{(x_m - m)^2}{2(\gamma^2 + \sigma^2)}\right] \end{aligned} \tag{2.19}$$

let

$$\theta \triangleq \gamma^2$$

then Eq. 2.19 can be expressed as [14]

$$f(\mathbf{x}) = \prod_{m=1}^{M} \frac{1}{\sqrt{2\pi(\theta + \sigma^2)}} exp\left[-\frac{(x_m - m)^2}{2(\theta + \sigma^2}\right] \tag{2.20}$$

Then find the logarithm to get the log likelihood function [14]

$$logf(x) = \sum_{m=1}^{M} \left[log\frac{1}{\sqrt{2\pi(\theta + \sigma^2)}}\right] + \sum_{m=1}^{M} \frac{-(x - m)^2}{2(\theta + \sigma^2)} \tag{2.21}$$

let

$$\begin{aligned}\frac{\partial}{\partial\theta}logf(x)|_{\theta=\hat{\theta}} &= \sum_{m=1}^{M}\left[\frac{-1}{2(\theta+\sigma^2)}+\frac{(x-m)^2}{2(\theta+\sigma^2)^2}\right] \\ &= 0 \end{aligned} \tag{2.22}$$

which has the unique solution [14]

$$\hat{\theta}(x)=\sum_{m=1}^{M}\sigma^2-(x-m)^2 \tag{2.23}$$

Since

$$\frac{\partial^2}{\partial\theta^2}logf(x)|_{\theta=\hat{\theta}}=\sum_{m=1}^{M}\frac{-1}{2(\theta+\sigma^2)}-\frac{(x-m)^2}{(\theta+\sigma^2)^3}<0 \tag{2.24}$$

this solution gives the unique maximum of $logf(x)$ [14]. The expectation of $\hat{\theta}(x)$ is [14]

$$\begin{aligned}E\left[\hat{\theta}(x)\right] &= \int_0^\infty [\sum_{m=1}^{M}\sigma^2-(x-m)^2]f(x_m)d_{x_m} \\ &= \int_0^\infty \left[\sum_{m=1}^{M}\sigma^2-(x-m)^2\right]\cdot \\ &\quad \frac{1}{\sqrt{2\pi(\theta+\sigma^2)}}exp\left[-\frac{(x_m-m)^2}{2(\theta+\sigma^2)}\right]d_{x_m} \\ &= \theta \end{aligned} \tag{2.25}$$

Therefore it's unbiased [14]. The Fisher's information for our case can be expressed as [14]

$$\begin{aligned}I_\theta &= -E_\theta\left[\frac{\partial^2}{\partial\theta^2}logf(x)\right] \\ &= -E_\theta\left[\sum_{m=1}^{M}\frac{-1}{2(\theta+\sigma^2)}-\frac{(x-m)^2}{(\theta+\sigma^2)^3}\right] \end{aligned} \tag{2.26}$$

So the Cramer-Rao lower bound (CRLB) is

$$Var_\theta[\hat{\theta}(x)]\geq\frac{1}{I_\theta}=\frac{1}{-E_\theta\left[\sum_{m=1}^{M}\frac{-1}{2(\theta+\sigma^2)}-\frac{(x-m)^2}{(\theta+\sigma^2)^3}\right]} \tag{2.27}$$

We will show the CRLB performance in Chapter 2.4 Simulations and Performance Analysis.

2.3 Power Allocation Based on forecasted channels by Inverse Water-Filling Algorithm

The Inverse Water-Filling(IWF) algorithm is based on a certain criterion, and adaptively allocates the transmission power according to the channel condition [14]. Usually when the channel condition is good, then the power is allocated more, and when the channel is poor, the power is allocated less, and when the channel is poor, the power is not allocated [14]. Thereby maximizing the transmission rate. To achieve the 'IWF' distribution of power, the transmitter must know the Channel State Information(CSI). When the transmitter knows the channel, the channel capacity can be increased [23].

Consider a $r \times 1$-dimensional zero-mean cyclic symmetric complex Gaussian signal vector $\tilde{s}$, where r is the rank of the transmission channel [14]. The vector is multiplied by the matrix $V(H = U\Sigma V^H)$ before transmission [14]. At the receiver, the received signal vector y is multiplied by U^H [24]. The effective input-output relationship for this system is given by [14]:

$$\begin{aligned} \tilde{y} &= \sqrt{\frac{E_s}{M_T}} U^H H V \tilde{s} + U^H n = \sqrt{\frac{E_s}{M_T}} U^H U \Sigma V^H \tilde{s} + U^H n \\ &= \sqrt{\frac{E_s}{M_T}} \Sigma \tilde{s} + \tilde{n} \end{aligned} \tag{2.28}$$

where $\tilde{y}$ is the received signal vector of the $r \times 1$ dimensional transform, and $\tilde{n}$ is the zero-mean cyclic symmetric complex Gaussian $r \times 1$ transform noise vector whose covariance matrix is $\xi\{\tilde{n}\tilde{n}_H\} = N_0 I_r$. The vector $\tilde{s}$ must satisfy $\xi\{\tilde{s}\tilde{s}_H\} = M_T$ to limit the total transmitted energy [14].

Then

$$\tilde{y}_i = \sqrt{\frac{E_s}{M_T}}\sqrt{\lambda_i}\tilde{s}_i + \tilde{n}_i, i = 1, 2, ..., r \tag{2.29}$$

The channel capacity is given by following [14]

$$C = \sum_{i=1}^{r} log_2(1 + \frac{E_s\gamma_i}{M_T N_0}\lambda_i) \tag{2.30}$$

where $\gamma_i = \xi\{|s_i|^2\}(i = 1, 2, ..., r))$ reflects the transmission energy of the i-th subchannel and satisfies $\sum_{i=1}^{r} \gamma_i = M_T$, E_s represents the channel gain on the s_{th} subchannel, λ_i is the Lagrange multiplier [14].

Variable energy can be allocated in the subchannel to maximize mutual information. Now the problem of maximizing energy becomes [14]

$$C = \max_{\sum_{i=1}^{r}} y_i = M_T \sum_{i=1}^{r} log_2(1 + \frac{E_s\gamma_i}{M_T N_0}\lambda_i) \tag{2.31}$$

Maximizing by Lagrangian method. The Optimal energy allocation policy is [14]

$$\gamma_i^{opt} = max\{(\mu - \frac{M_T N_0}{E_s\lambda_i}), 0\} \tag{2.32}$$

where

$$\sum_{i=1}^{r} \gamma_i^{opt} = M_T \tag{2.33}$$

We compared the channel capacity based on IWF algorithm and the equal gain algorithm [14]. We will show the simulations in Chapter 2.4 Simulations and Performance Analysis.

2.4 Simulations and Performance Analysis

2.4.1 Simulation Parameters of Simulated Channels

Table 2.1 shows the simulation parameters we used for simulated channels [14].

Table 2.1. Simulation Parameters

Frequency (GHz)	28.0	Bandwidth (MHz)	800
TXPower (dBm)	30.0	Environment	NLOS
Scenario	UMi	Pressure (mBar)	1013.25
Humidity	50	Temperature (Celsius)	20.0
RainRate (mm/hr)	0.0	Polarization	Co-Pol
Foliage	No	DistFol (m)	0.0
FoliageAttenuation (dB)	0	TxArrayType	ULA
RxArrayType	ULA	Num of TXElements	1
Num of RXElements	1	TXAziHPBW	10
TXElvHPBW	10	RXAziHPBW	10
RXElvHPBW	10		

Table 2.2. Granger Causality Test Results

	F-Value	Critical Value	Confidence Level
1 on 2	20.9127	0.229	0.95
2 on 1	97.0205	0.229	0.95

2.4.2 Granger causality Test Result

Table 2.2 shows the results of the Granger causality tests [14].

First, we did a Granger causality test of channel 1 on channel 2. In this simulation, the value of the F statistic is 20.9127, where the critical value from the F-distribution is 0.229, confidence level α is 0.95 [14]. Since the value of the F statistic is larger than the critical value from the F-distribution, we reject the null hypothesis that channel 2 does not Granger Cause channel 1 [14]. So this test proves that channel 2 is Granger cause of channel 1. Then we do the Granger causality test of channel 2 on channel 1 [14]. In this case, the value of the F statistic is 97.0205, where the critical value from the F-distribution is still 0.229, confidence level α is 0.95 [14]. So this test proves that channel 2 is also the Granger cause of channel 1 [14].

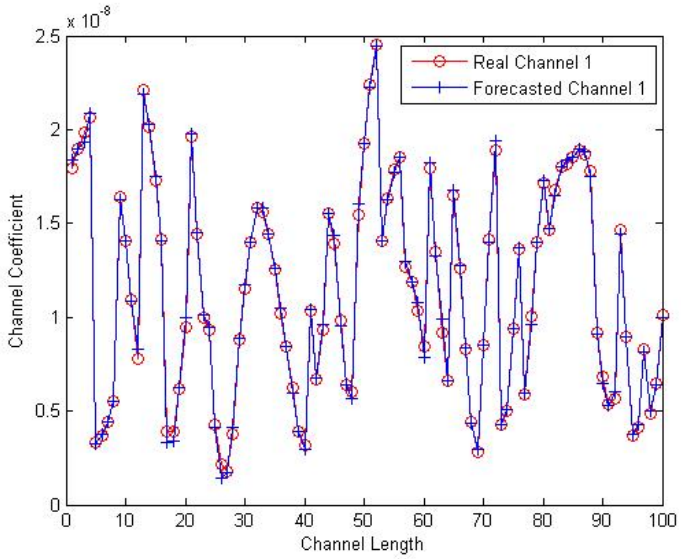

Figure 2.1. Forecasted 5G channel 1 based on transfer entropy.

So far, we verified the existence of Granger causality between the two channels. Based on this result, we use transfer entropy to forecast the two channels [14].

2.4.3 Forecasted Wireless Channels

In Fig. 2.1 and Fig. 2.2, we compare the real 5G channel coefficient with the predicted 5G channel coefficient [14]. Observe that the channels which are forecasted using the TE algorithm has high accuracy. But we still need a mathematical method to calculate the specific error between the two channels and prove the accuracy [14]. Therefore, for the comparison of real channels and prediction channels, we apply two metrics, i.e., RMSE and CRLB [14].

2.4.4 RMSE

In Fig. 2.3, we purposed the Box Jenkin's method and the Transfer Entropy method to forecast the real channels and compared their RMSE, which is the average RMSE of all 1000 pairs forecasted channels and 1000 pairs real 5G channels with different SNRs [14].

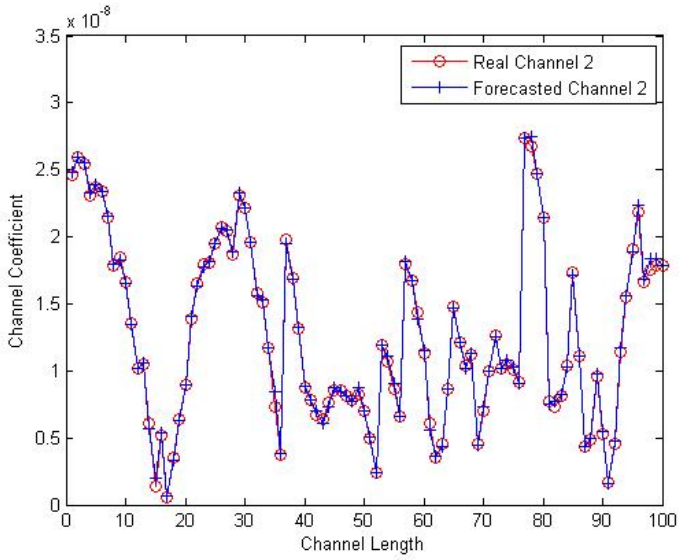

Figure 2.2. Forecasted 5G Channel 2 based on transfer entropy.

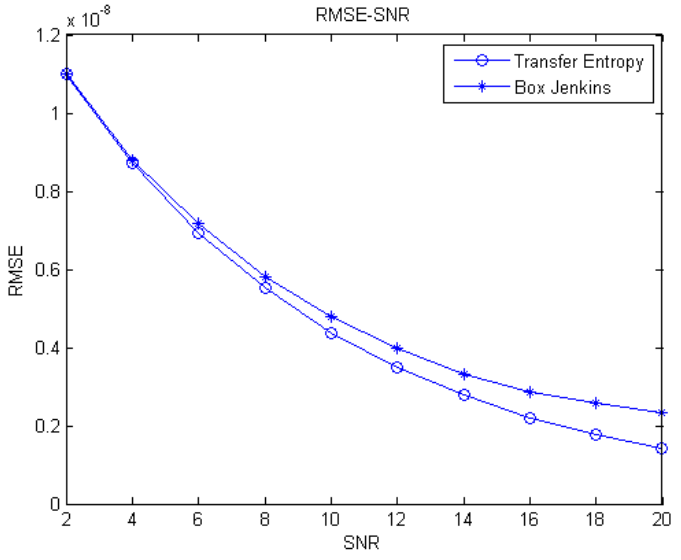

Figure 2.3. The average RMSE comparison between the real 5G channel and forecasted channel.

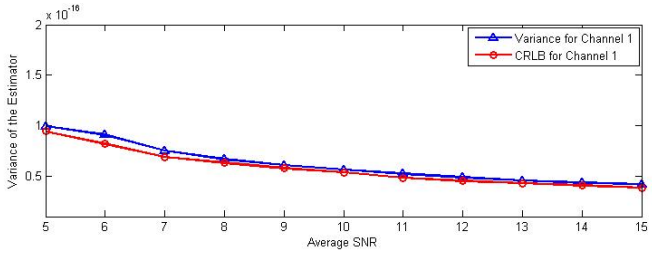

Figure 2.4. Variance of forecasted channel 1 with different value of SNRs.

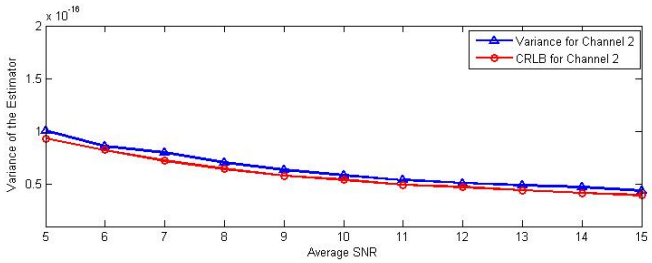

Figure 2.5. Variance of forecasted channel 2 with different value of SNRs.

1) We can observe from Fig. 2.1 and Fig. 2.2 that the average coefficient of the real 5G channel is about 1.5×10^{-8}. And in Fig. 2.3, the RMSE is really high when the SNR is low. So we can conclude that we compare the TE method and Box Jenkin's method when the SNR is low, the accuracy is insufficient [14].

2) However, with the SNR is increasing, the RMSE is going to horizontally and stable at about 1×10^{-9}. So this shows that the true RMSE of the real 5G channel and forecasted channel based on the TE algorithm is less than 6%. This result proves that TE is feasible to forecast 5G channels [14].

3) From Figure. 2.3 we can observe that under the same SNR, the RMSE between the forecasted channel and real channel by using the Box Jenkins method is larger than that of the Transfer Entropy method, which means that the TE method is more accurate in the forecasting under the same channel status [14].

2.4.5 CRAMER-RAO LOWER BOUND

Simulations: For real and forecasted 5G channel1 and channel 2, we run Monte Carlo simulations at each SNR value and applied $\hat{\theta}$ for total 4 channels respectively [14]. In Fig. 2.4 and Fig. 2.5, we plotted the variance of the estimator with different values of SNRs. Observe the following [14].

1) The actual variance of $\hat{\theta}$ is almost mach with the CRLB for different SNR value [25], which validate our result in Chapter 2.4.2: Our forecasted 5G channel is close to real 5G channel and TE is feasible to forecast 5G channels [14].

2) The actual variance of $\hat{\theta}$ reduces as SNR value increases, and tends to stable and horizontally, which is as we have shown the RMSE plots in Chapter 2.4.4 [14].

2.4.6 Channel Power Allocation

Simulations: We perform power allocation based on the IWF algorithm and Equal Gain(EG) algorithm for the forecasted 5G channel 1 and channel 2 and plot part of the results respectively [14]. Since the row 5G channels are too long to plot, we choose part of the forecasted channel randomly, so there are some variations in the channel coefficients compared with transfer entropy [14]. Fig. 2.6 - Fig. 2.9 shows the simulations. For each simulation, the total power is $1 \times 10^{-5}w$, which is $-20dBm$ to be allocated in the channel. And Fig 2.10 proposed the channel capacity comparison. Observe the following [14].

1) Fig. 2.6 and. 2.8 shows power allocation using the IWF algorithm. It can be seen from the figures that when the subchannel state is good, the much power is allocated to the subchannel, and when the subchannel is not good, the less power is allocated [14]. Some subchannels are very poor, and no power is allocated to the channel. Compared with the power allocation of the EG algorithm in Fig. 2.7 and Fig. 2.9, that is, the energy is evenly

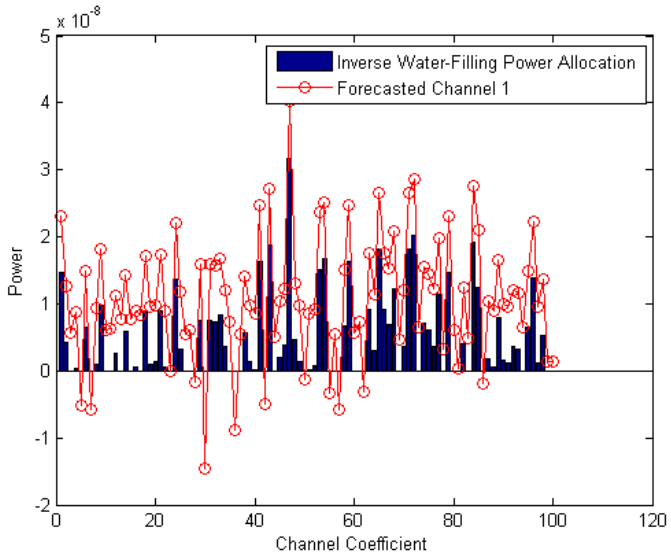

Figure 2.6. Forecasted 5G channel 1 power allocation based on inverse water-filling.

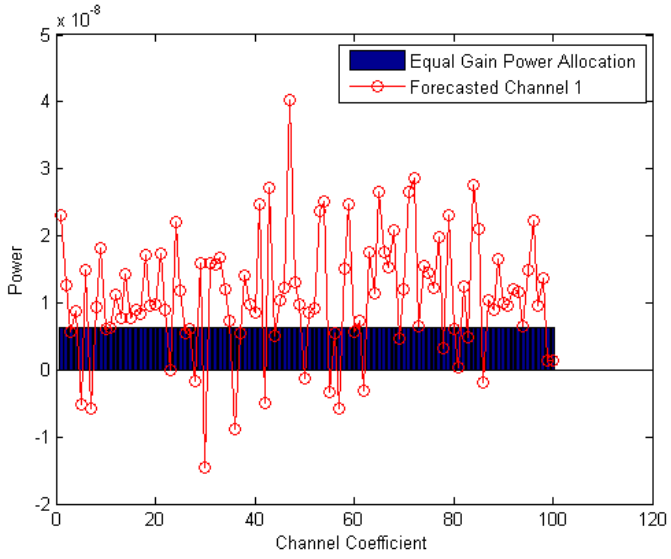

Figure 2.7. Forecasted 5G channel 1 power allocation based on equal gain.

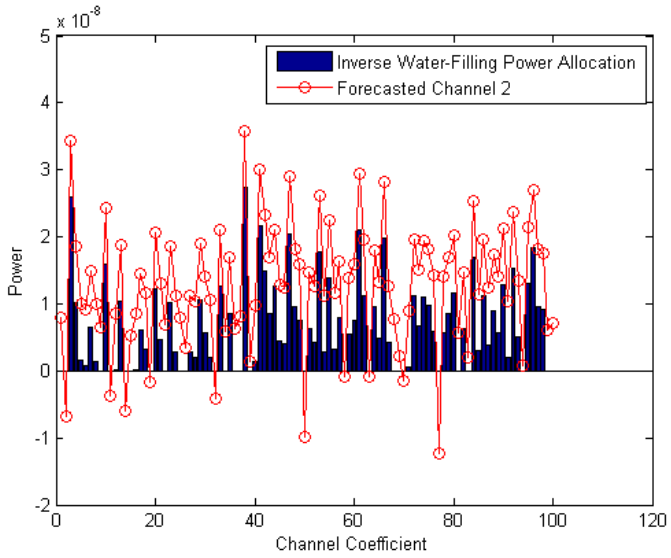

Figure 2.8. Forecasted 5G channel 2 power allocation based on inverse water-filling.

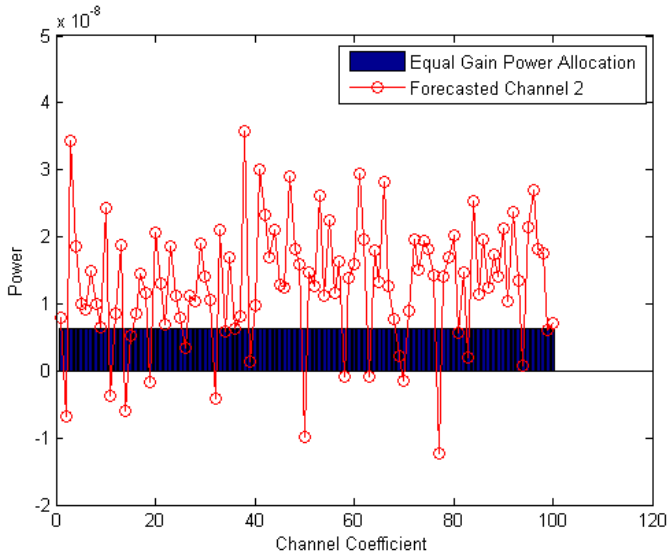

Figure 2.9. Forecasted 5G channel 2 power allocation based on equal gain.

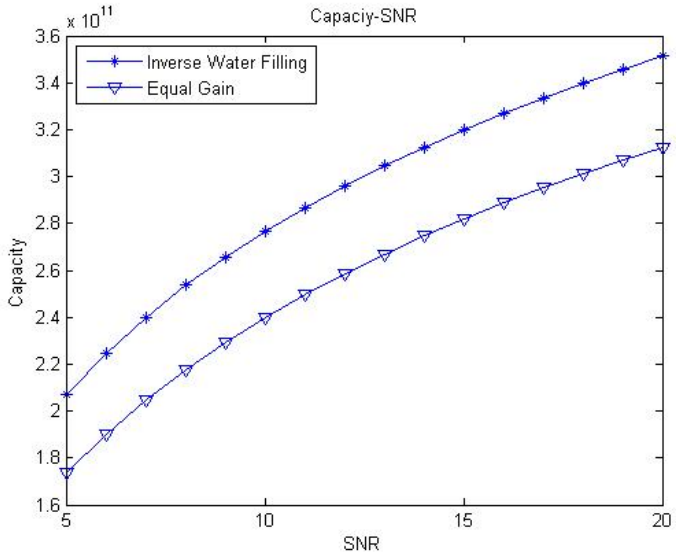

Figure 2.10. Average forecasted channel capacity comparison based on inverse water-filling and equal gain.

distributed on each subchannel regardless of the channel condition, and the IWF algorithm can greatly improve the energy efficiency and avoid the waste of power [14].

2)From Fig. 2.10, we can observe that at the same SNR, the channel capacity of the IWF is greater than the EG, where demonstrates the power allocation efficiency of the IWF algorithm working on the forecasted 5G channels [14].

CHAPTER 3

CHANNEL MODELING FOR MASSIVE MIMO AT 28GHz UNDER INDOOR AND THROUGH-WALL SCENARIOS

3.1 Introduction to Massive MIMO System and Millimeter Wave Communication

In recent years, the volume of mobile data services has grown almost exponentially, and it will reach the current thousand times by 2021 [26]. At the same time, with the increasing proportion of energy consumption of information technology systems, reducing the energy consumption of mobile communication network systems has gradually become an important goal of communication development [27]. The current LTE mobile communication system will be difficult to meet the future mobile communication needs for spectrum efficiency and energy consumption efficiency [28]. This is presenting great challenges to the spectrum of the fifth-generation mobile communication system (Fifth-generation, 5G) efficiency and energy efficiency.

How to further improve the spectrum efficiency and power efficiency of wireless mobile communications by an order of magnitude on the basis of LTE is the core of 5G. The development of 5G requires new changes in network system structure, networking technology and wireless transmission technology, fundamentally solve the problem of spectrum efficiency and power efficiency of mobile communication and achieve the dual goals of higher spectrum efficiency and green wireless communication [29]. The use of multiple antenna transmission and multiple antenna input (multiple-input multiple-output, MIMO) technology is a basic way to tap wireless space dimension resources, improve spectral efficiency and power efficiency, and has been one of the mainstream technologies researched and developed in the field of mobile communications for the past 20 years [30]. MIMO

technology can provide diversity gain, multiplexing gain and power gain. Diversity gain can improve the reliability of the system, multiplexing gain can support single-user spatial multiplexing and multi-user spatial multiplexing, and power gain can Improving the power efficiency of the system through beamforming [31]. At present, MIMO technology has been adopted by LTE (long term evolution, LTE), IEEE 802.11ac and other wireless communication standards [32].

The MIMO system can fully utilize all the characteristics of the space-time-frequency domain of the signal, and has the following advantages: (1) Utilize or reduce multipath fading: MIMO technology can make full use of various multi-path transmission/synthesis technologies to improve the performance of wireless communication systems. (2) Elimination of co-channel interference: MIMO systems can use adaptive beamforming technology or multi-user detection technology to effectively suppress or delete co-channel interference. (3) Improve the spectrum utilization rate: Because the array antenna can reduce the effects of co-channel interference and multi-path fading, the bit error rate (BER) can be reduced under a certain signal-to-interference and noise ratio (SINR), or the detection can be reduced under a certain BER. The MIMO system can suppress or eliminate co-channel interference and inter-symbol interference, and at the same time use the classification technology to improve the signal-to-noise ratio (SNR) of the received signal. Therefore, the transmit power of the base station and mobile terminal can be reduced to a certain extent, furthermore extending mobile terminal battery life, reducing the impact on the ecological environment, reducing the system's requirements for power control accuracy and devices [33].

Sub-6GHz and millimeter wave are two frequency band codes of the 5G network. Sub-6GHz refers to signals in the 450MHz-6GHz frequency band, while millimeter waves refer to waves with frequencies exceeding 24GHz. The characteristic of millimeter wave is faster transmission speed, theoretically can reach 10Gbps high-speed transmission. An-

other advantage of millimeter wave is higher bandwidth, it is easier to solve the problem of user network congestion to satisfy more throughput at the same time. This technology has a relatively small coverage area and is more suitable for applications in densely populated scenes such as stations, airports, and stadiums. However, compared to sub-6Ghz, millimeter waves have poorer penetration and are easily affected by the environment. The millimeter wave can be blocked by almost any obstacle in its propagation path. In view of the characteristics of millimeter wave, Massive MIMO is a better communication technology [34]. Compared with traditional MIMO technology, large-scale MIMO technology has some unique channel propagation characteristics, such as near-field effects, non-stationary characteristics of scattering clusters on the time axis and array axis, etc., which brings great challenges to practical applications [35]. Studying the propagation characteristics of massive MIMO wireless communication channels and proposing a channel model that can accurately describe the propagation characteristics of large-scale MIMO wireless communication channels, which is essential for the design and evaluation of 5G wireless communication systems [36].

3.2 Introduction to Massive MIMO Channel Characteristics and Channel Modeling

3.2.1 Massive MIMO Channel Characteristics

Understanding the basic propagation characteristics of wireless channels is the basis for channel modeling. However, the characteristics of high-frequency channels are different from those of traditional channels, such as rain attenuation, atmospheric absorption, and poor penetration performance [37]. These aspects are not suitable for direct modeling of high-frequency channels. This chapter analyzes the unique characteristics of high-frequency channels and lists the current commonly used channel models in industry and academia.

1. Large-scale fading characteristics

   The large-scale model is used to describe the change in the field strength of a signal after a long distance (hundreds to thousands of kilometers). It is generally on the scale of dozens or even hundreds of wavelengths. This is the long-term statistical average in seconds. the result of. Large-scale fading includes path loss and shadow fading, the former varies with the propagation distance, the latter is caused by the shadow of large obstacles (such as tall buildings, mountains, etc.), so large-scale models are divided into path loss models and shadow fading models.

   Path loss is caused by the dissipation of transmitter radiated power and the effects of propagation channels. The path loss model usually assumes that the path loss is the same at a given transmit-receive distance. Shadows are caused by obstacles between the transmitter and receiver. These obstacles attenuate the signal power through absorption, reflection, scattering, and diffraction. When the attenuation is very strong, the signal is blocked, and interruption may occur. Path loss refers to the average value of the transmitted signal power loss with distance, and shadow fading is a slow fluctuation around the average loss caused by obstacles [38].

   Shadow fading, at a given distance of the channel, the actual road loss often shows random changes, so only statistical models can be used to characterize this random fading. The most used is the log-normal shadow model. The measured data proves that the model can be accurate. To model the change in received power in indoor and outdoor wireless propagation environments, the expression of the path loss under this model is [39]:

$$PL(d)[dB] = \hat{PL}(d_0) + 10nlog(\frac{d}{d_0}) + X_\sigma \tag{3.1}$$

Where d_0 is the reference distance, X_σ is the zero-mean standard deviation of σdB Gaussian random variable, the values of σdB and n are calculated based on the actual channel measurement data, and the mean square error of the measured value and the estimated value is calculated based on linear recursion.

2. Small-scale fading characteristics

   Small-scale fading refers to the rapid fluctuation of radio signal amplitude, phase or multi-path delay after short-time or short-distance propagation. The fading is caused by the same transmission signal propagating along two or more paths and arriving with different propagation delays. The receiver generates the signal constructive interference and destructive interference. The multi-path effect is generally manifested in the rapid change in signal strength after short distance or short time propagation. On different multi-path signals, there is random frequency modulation caused by time varying Doppler frequency shift and time caused by multi-path propagation delay Spreading, usually using parameters such as delay spread, frequency spread and angle spread to describe the corresponding multi-path effect.

   Time dispersion, means that the multi-path signal arrives at the receiving end at different times, so that the received signal causes the spread of the digital signal waveform in the time domain. Important parameters describing the time dispersion characteristics of multi-path channels are average additional delay τ, root mean square delay extension σ, etc.

   Frequency dispersion, refers to the phenomenon of signal dispersion in frequency. The main reason is the relative movement between the mobile station and the base station. The change in the distance between the transceiver and the receiver causes the phase and frequency to change when the signal reaches the receiver, resulting in a Doppler shift effect.

3.2.2 Massive MIMO Channel Modeling

During the propagation process of wireless communication signals, from transmitter to receiver through a complex propagation environment, large-scale fading (including the path loss and shadow fading) and small-scale fading are caused by propagation mechanisms such as direct radiation, reflection, scattering, and diffraction [1]. Signal fading leads to distortion of the received signal, which affects the communication performance of the wireless communication system [2]. Understanding the propagation characteristics of wireless communication channels and their influence on wireless communication signals is crucial for the design and testing of wireless communication systems. The wireless channel model is based on fully understanding the propagation characteristics of wireless signals and characterizes the characteristics of the wireless channel through a series of parameters, which is an abstract simulation of the wireless propagation environment [12].

Up to now, several massive MIMO wireless communication channel models have been proposed. The classical independent and identically distributed Rayleigh fading channel is used as the channel model of massive MIMO communication system [37]. Since the channel coefficients are independent and identically distributed, the central limit theory and random matrix theory can be easily applied to the analysis of massive MIMO channel matrix [40]. Following are some popular Statistical and Analytical Channel Models.

1. 3GPP and WINNER II Models

   The geometry-based stochastic 3GPP and WINNER II Spatial channel model (SCM) [41] follows a system-level approach and is suitable for link-level or system-level simulation to estimate the actual channel (UE) between a base station (BS) and one or more user devices, which explains the empirical correlation between large-scale parameters. Large scale parameters represent omnidirectional RMS delay extension (DS), azimuth extension (AS), shadow fading (SF), and Rician K factor (for LOS channels), and show significant correlation between a given base and a cell phone

link [42]. According to the work, in the 3GPP model, The 3GPP model also specifies a spatial autocorrelation coefficient of +0.5 for the shadow fading experienced by the two MS, which is set to +0.5 for DS and AS, -0.6 for DS and SF, and -0.6 for AS and SF, respectively [39]. But does not specify the range of distance applicable to the shadow fading correlation. In WINNER SCM, the number of spatial interrelationships between two MS separated by distance dMS is modeled using a decay exponential function, which is parameterized using related distance parameters [42]. The correlation distance between two large-scale parameters is 0.37 (1/e), and the typical correlation distance of UMi scenarios ranges from 9 m to 14 m. In both SCMS, associated Gaussian random variables are used to generate large-scale parameters to recreate the combined statistics of measurements.

2. COST 2100 Models

   The COST 2100 model follows a cluster-level approach, where clusters (for example, scattering objects) are placed in an analog environment and can interact with one or more mobile terminals using the concept of visible regions. The visibility region defined in the COST 2100 model is a key concept in geometric and random propagation models and represents the spatial or temporal span over which a set of traveling multipath components appear on a universal radio terminal antenna [43] [44]. Spatial consistency is achieved by using visible regions associated with each cluster of multipath components [44]. One or more light clusters are assigned to the visible area and their size varies with the movement of the mobile terminal, thus allowing spatial consistency in the simulated environment. Spatial consistency refers to smooth channel conversions between closely separated mobile terminals that experience similar but slightly different scattering environments. Ignoring spatial consistency will overestimate the performance of spatial multi-antenna technology [45]. The COST 2100

model assumes Rayleigh fading of path gain amplitude in an NLOS environment to recreate the statistics of small fading [43].

3. Statistical Models

   The establishment of statistical channel model is mainly through the statistics of key channel parameters (such as path loss, delay, angle, etc.), because in practical applications, the characteristics of wireless channels in the same scenario are similar [46]. For example, the building structures and other environments of micro-cells in different cities are similar, so their channel propagation characteristics will also have similarities, so it is also called a stochastic modeling method.

   Parametric modeling is a statistical modeling method based on channel measurement. First, the scene of the measured channel is determined, and the measurement data is collected; then the measurement results are processed, and the distribution of related parameters such as delay and angle is extracted through an algorithm [46]. Thus, a channel model is established, which can be used in the same real propagation environment.

   Another approach to statistical modeling is the theoretical modeling approach based on physical propagation. The modeling method describes the characteristics of the signal and its scattering distribution by characterizing the statistical distribution of the scatterers, and obtaining parameters such as multipath delay, wave arrival angle and wave departure angle of the channel according to the basic propagation law of electromagnetic waves. Often used in the study of MIMO channels

4. Deterministic Channel Modeling

   The deterministic channel modeling method is to determine the geographic information in the actual propagation environment such as (buildings, trees, terrain information, etc.) in advance based on the actual map information, and then analyze and analyze the electromagnetic characteristics and geometric characteristics of the

environment. Predictive wireless propagation models. This method requires very accurate and detailed information about the dissemination environment. The accuracy of the model established by this method is related to the accuracy and detail of the environment. For example, not only the height, width and volume of the building should be described, but also the structure and materials should be described. The more detailed the description and the higher the accuracy, the closer the modeling results are to the actual propagation situation. Since deterministic channel modeling usually relies on the numerical calculation of solving Maxwell's equations or the calculation of geometric light theory [45], the computational complexity will increase greatly with the complexity of the propagation environment. Therefore, deterministic modeling methods are generally used for channel modeling in a smaller range such as indoors. Common deterministic channel modeling methods include (RayTracing, RT) and Finite-Difference Time-Domain (FDTD) methods. .

Deterministic channel modeling does not depend on the actual measurement activities, and as long as the specific information required for propagation can be obtained, the transmission characteristics of the signal can be forecasted more simply.

5. Semi-deterministic Channel Modeling

   In order to solve the computational complexity of the deterministic modeling method and the error of the statistical modeling method, a semi-deterministic modeling could be considered. This method absorbs the advantages of the first two methods, not only reduces the complexity of mathematical derivation requirements, but also has better results. accurate prediction of the actual signal. The most widely used modeling method in the semi-deterministic modeling method is the stochastic geometric modeling method. This modeling method is a simplification of RT in deterministic modeling. It does not need to provide accurate signal environment parameters. A fixed probability distribution is used, the locations of the scatterers are randomly se-

lected, and then a simplified RT method is used to obtain the actual channel impulse response.

The correlation matrix method is another commonly used modeling method in the semi-deterministic method, which reflects the correlation between the spatial channels. This method needs to obtain the path delay, Angel of Arrival (AOA), Angle of Departure (AOD), etc. Spatial parameters. Generally, parameters are obtained from measured data or channel statistical information, and then the channel spatial correlation matrix is directly calculated by using the geometric relationship between AOD, AOA and antenna [47]. The models established by the correlation matrix method include Weichselbreg model [48], VCR model 3GPPLTE channel model [49] and so on.

A measurement-based statistical indoor radio-channel impulse response (IR) model (SIRCIM) and outdoor mobile simulator (SMRCIM) were successfully implemented from many thousands of collected CIRs in factories at 1.3 GHz [50] [51], and from outdoor cellular channel PDPs [52] [53]. These CIR models were popular with industry in the early years of digital cellular and WiFi [54]. The SIRCIM and SMRCIM models were based on statistical and geometrical models to synthesize the phases and directions of arrival and departure in an IR model [55] [54].

Compared to the massive MIMO communication system with unified antenna array, because the Rayleigh fading channels are independent of each other and obey the Rayleigh distribution, this model can be better applied to the communication system of distributed antenna massive MIMO [56]. Compared with independent and identically distributed Rayleigh fading channels, the Kronecker channel model takes into account the correlation between antennas when it is used in massive MIMO communication systems, but does not consider the non-uniformity of the scattering clusters of massive MIMO communication channels on the antenna array axis. Kronecker's improved channel model uses

the Markov birth and death process to model the non-stationary characteristics of scattering clusters in massive MIMO communication channels [57], which makes up for the shortcomings of the traditional Kronecker channel model. Compared with the traditional Kronecker modeling, the Weichselberger modeling [58] is used in the massive MIMO to consider the mutual coupling between the transmitting and receiving antennas, but it also does not have the non-stationary characteristics of the scattering clusters on the array axis. Perform modeling [59]. The virtual channel characterization model uses the preset discrete Fourier transform to replace the unilateral correlation matrix for correlation channel modeling. This allows the accuracy of the channel model to increase when the TX and RX has more antennas [60]. However, the virtual channel characterization of the model is only applicable to massive MIMO systems equipped with one-dimensional linear antenna arrays and cannot be applied to massive MIMO wireless communication systems equipped with 2D Uniform Planar Array (UPA) [8].

The Regular-Shaped Geometry-based Stochastic Model (RS-GBSMs) of the massive MIMO wireless communication channel distributes the scattering clusters on the geometry of ellipse, multi-ring, and double-multi-ring to perform an accurate calculation of channel parameters to ensure the assumption of spherical waves [61]. For the scattering clusters in the massive MIMO channels, Markov birth-and-death processes are used to model non-stationary features in antenna arrays. The literature [48] proposed a massive MIMO wireless communication channel model based on the two-state Markov process, which uses the parabolic wave assumption to replace the plane wave. It can not only meet the linear angular offset of the multi-path component of the massive MIMO antenna array but also reduce the complexity of the model compared to the channel model using the spherical wave hypothesis. The Dimension GBSMs (D-GBSMs) based on the WINNER series channel model uses the spherical wave. It is assumed that the transmitter (TX) antenna array is divided into multiple sub-arrays at the same time using the visibility of the scatter-

ing clusters to simulate the non-stationary characteristics of the scattering clusters on the array axis, D-GBSMs based on the COST2100 channel model also uses the spherical wave assumption [62]. And whether the movement of the receiver (RX) enters the visible area of TX is used to simulate the non-stationary characteristics of scattering clusters on the array axis [63].

In addition to random channel models, deterministic channel models for massive MIMO wireless communication channels such as the ray-launching model and graph theory model [64] have also been proposed and applied to large-scale MIMO wireless communication channels.

3.3 Measurement and Collection of Experimental Data

3.3.1 Measuring platform

The software and hardware equipment used at the transmitting end of this measurement process are mainly a millimeter-wave vector signal generator, an arbitrary waveform generator, an up-converter, and an electronic computer and corresponding software and transmitting antenna. The hardware and software of the receiver mainly include down-converter, broadband digital receiver, microwave analog signal generator, waveform generator and receiving antenna. The Keysight measurement hardware used by the measurement platform is shown in Table 3.1 for specific models and signal information.

3.3.2 Measurement scenario

The location of this measurement is in a teaching hall. The measurement environment is 28GHz indoor scene and 28GHZ through-wall scenario. The indoor scenario is LOS, and the wall-through scenario is NLOS measurement scene plan as shown in Fig.c3.1 and Fig. 3.2.

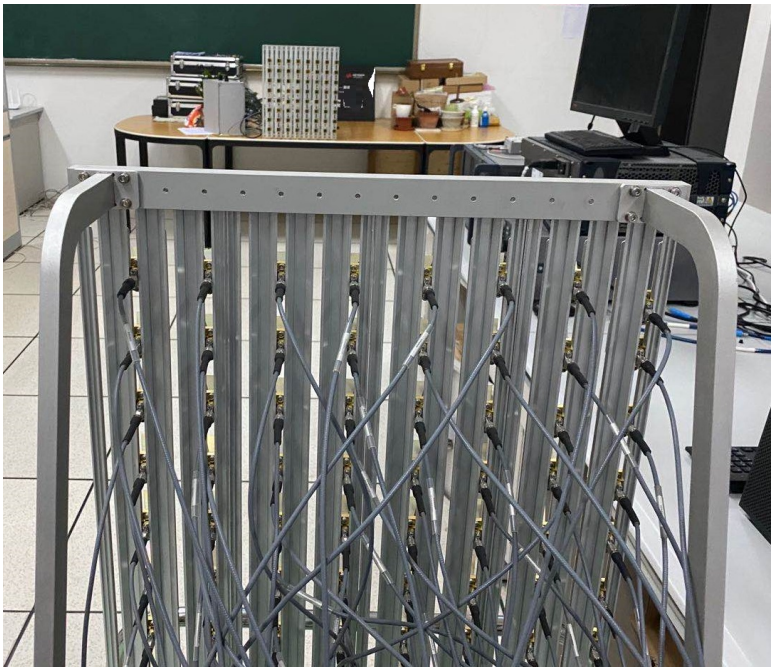

Figure 3.1. Indoor scenario measuring.

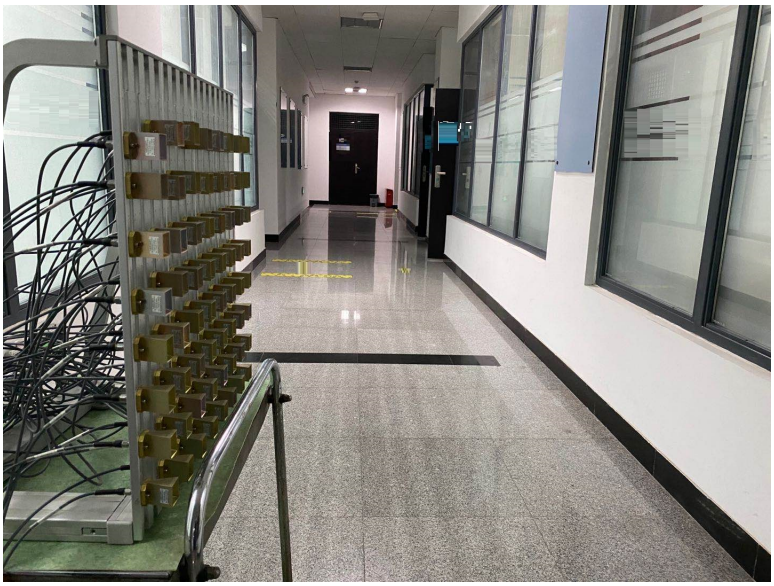

Figure 3.2. Through-wall scenario measuring.

Table 3.1. Signal Parameters and Hardware Information

Parameters	Value	Hardware Information
Central frequency [GHZ]	28	E8267D PSG VSG
Bandwidth [MHZ]	800	Keysight M8190A AWG
Sample rate [GHZ]	1.28	Keysight N9040B
TX array	64 URA	
RX array	64 URA	
Array element space	Half a wavelength	
Transmit power [dBm]	30	
Antenna directivity	Directional	

The two scenarios of this measurement are both used the TX using 64-URA (8x8) antenna, the transmitter antenna is placed on the table, the height is 1 meter. The RX antenna is also a 64-URA (8x8) antenna with a height of 1 meter. It is placed on an indoor table in the indoor scenario and is about ten meters away from the TX. The RX is placed on the table in the corridor in the wall-penetrating scenario, separated from TX by a wall.

3.4 Statistical channel modeling and parameter analysis

In this dissertation, through the obtained measured channel results, parametric modeling of large-scale parameters such as channel path loss, delay spread, angle spread, and shadow fading is derived to establish a GSCM. The GSCM modeling method separates the antenna and the propagation channel, and researchers can use different antenna array modes to obtain specific transmission channel models. The basic modeling steps are mainly divided into three stages [65].

The first stage is mainly the preparation before the measurement, which needs to determine the general model of the channel and the parameters to be measured. To formulate a detailed measurement plan, including selecting the wireless channel scenario, defining the measurement environment, planning the measurement path, and formulating the height

of the user and the antenna [66]. After that, the measurement time, link budget and other elements and the storage of measurement data and other specific matters must be considered. The second stage is the post-processing of the measurement data, usually using high-resolution parameter estimation algorithms, such as Expectation Maximization (EM) and other parameter estimation algorithms, to process and analyze the measurement data to extract the channel delay, wave polarization gain of AOA, wave departure angle, and Dense Multiple Components (DMC) parameters [67]. Through statistical analysis of the data, clustering processing of multiple path components using a family algorithm, observe the probability distribution function obeyed by different parameters and use the degree of fitting to select the optimal distribution function, then finally get the corresponding statistics parameters. The final stage is the generation of channels. Based on the statistical parameters and probability distribution function obtained in the previous stage, cluster and ray parameters are generated. Combined with the antenna array, the channel transmission matrix is obtained. Finally, the time-varying channel impulse response is observed [68].

In the measurement, $u(t)$ is the transmitted signal, the expression is

$$u(t) = \sum_{i=-\infty}^{\infty} a(t - iT_a) \tag{3.2}$$

Among them, $a(t)$ is the burst signal, the expression is

$$a(t) = \sum_{k=0}^{K-1} a_k p(t - kT_p) \tag{3.3}$$

The burst signal consists of a detection sequence $[a_1, a_2, \ldots, a_{K-1}]$ of length K, and the duration of the shaped pulse T_p is related to $T_a, T_a = KT_p$.

The received signal is actually composed of the superposition of multipath signals, where the contribution of the l_{th} path to the output signal can be expressed by the vector signal as the following formula

$$
\begin{aligned}
s(t;\theta_l) &= [s_1(t;\theta_l),\ldots,s_N(t;\theta_l)]^T \\
&= a_l exp\{j2\pi v_l t\} c_2(\Omega_{2;l}) c_1(\Omega_{1;l})^T u(t-\tau_l)
\end{aligned} \tag{3.4}
$$

Where θ_1 is the channel parameter set, including departure angle $\Omega_{1;l}$, arrival angle $\Omega_{2;l}$, delay τ_l, Doppler frequency shift v_l, and complex amplitude a_l. The expression of the received signal vector at the output of the antenna array is as follows

$$
Y(t) = [Y_1(t), Y_2(t), ..., Y_M(t)]^T = \sum_{i=1}^{L} s(t;\theta_l) + \sqrt{\frac{N_0}{2}} N(t) \tag{3.5}
$$

Among them, N_0 is a normal quantity, $N(t) = [N_1(t), N_2(t), \ldots, N_M(t)]^T$ is M-dimensional complex Gaussian white noise, which can be seen from the expression (4) of the received signal $Y(t)$, where θ_l is an unknown parameter, and others are known. The channel parameters are extracted from the measurement data, that is, the number of beams L and their parameters $\theta_l = [\Omega_{1,l}, \Omega_{2,l}, \tau_l, v_l, a_l], l = 1, 2, ..., L$ are estimated by observing the data $Y(t) = y(t)$; the is usually used to estimate the above parameters. The log-likelihood function of the parameter $\theta = [\theta_1, \theta_2, \ldots, \theta_L]$ under the given observation data $Y(t) = y(t)$ is

$$
L(\theta, y) = \frac{1}{N_0}\left[2\int_{PT} \Re\{s^H(t,\theta_l) y(t)\} dt - \int_{PT} \|s(t,\theta_l)\|^2 dt\right] \tag{3.6}
$$

Where $\Re\{\cdot\}$ represents the real part. MLE of θ refers to the vector that makes $\theta \to L(\theta; y)$ get the maximum value, which is

$$
\theta_{ML}(y) \in \arg\max_{\theta_l}\{L(\theta; y)\} \tag{3.7}
$$

The results of the dimension of θ is high, and the global maximum value of the nonlinear function $\theta \to L(\theta; y)$ cannot be expressed in a closed form, which makes the calculation of the value of $\hat{\theta}_{ML}(y)$ very complicated. In order to solve the complexity problem

of MLE calculation, the EM algorithm can be used to calculate MLE through an iterative idea, the SAGE algorithm is improved and developed on the basis of the EM algorithm, which is widely used in TX and RX parameter estimation of channel measurement. The process of using SAGE algorithm to estimate parameters is as follows [69].

Define an incomplete but measurable data set for the actual measurement data $Y(t)$, express $Y(t)$ as a complete but unobservable data set $X(t) = [X_1(t), X_2(t), \ldots, X_L(t)]$, and its conversion relationship is

$$Y(t) = \sum_{l=1}^{L} X_l(t) = \sum_{l=1}^{L} s(t;\theta_l) + \sum_{l=1}^{L} \sqrt{\frac{\beta_i N_0}{2}} N_l(t) \tag{3.8}$$

Since $X_1(t), X_2(t), \ldots, X_L(t)$ is independent of each other, the estimation of the parameter θ_L of the l_{th} path is unrelated with other paths.

Suppose the complete data set $X_(t)$ is an observable value, and the observation value $X_l(t) = x_l(t)$ within the observation time PT. From equation (6), the log-likelihood function of the parameter θ_L of the l_{th} path corresponding to $x_l(t)$ is

$$L(\theta_i; x_l) = \frac{1}{\beta_l N_0} \left[2 \int_{PT} \Re\{s^H(t;\theta_l) x_l(t)\} dt - \int_{PT} \|s(t;\theta_l\|^2 dt \right] \tag{3.9}$$

Find the MLE of θ_L from the hypothetical $x_l(t)$ as

$$\theta_{L,ML}(x_l) \in \arg\max_{\theta_l} \{L(\theta; x_l)\} \tag{3.10}$$

In practice, $X_l(t)$ is unobservable. The idea of the SAGE algorithm is to calculate the conditional expected value of $X_l(t)$ based on when $Y_(t) = y(t)$ is not completely observed data and the last estimated value of θ, $\hat{\theta}'$, and use this value as the observation value of $X_l(t)$, then estimate the parameter set again by MLE [70]. This step is the E step of the SAGE algorithm, as shown in the following formula

$$
\begin{aligned}
x_i(t;\hat{\theta'}) &= E_{\tilde{\theta}}\left[X_l(t)|Y(t)=y(t)\right] \\
&= y(t) - \sum_{i=1,j\neq l}^{L} s(t;\hat{\theta'})
\end{aligned} \tag{3.11}
$$

The M step of the SAGE algorithm is

$$
\bar{\theta}_{L,ML}(x_l) \in \arg\max_{\bar{\theta}_l}\{z(\bar{\theta}_l;x_l)\}
$$

$$
a_{lML}(x_l) \in \frac{z(\bar{\theta}_i;x_i|_{\bar{\theta}_l=\bar{\theta}_{lML}(x_l)})}{|c_2(\hat{\Omega}_{2,l})_{ML}|\cdot|c_1(\hat{\Omega}_{1,l})_{ML}|\cdot P\cdot T_{sc}} \tag{3.12}
$$

The M step of the SAGE algorithm is improved on the basis of the EM algorithm. The parameter matrix in each path is divided into 6 parameter subsets $\{\tau_l, a_l\}, \{\theta_{2,l}, a_l\}, \{\phi_{2,l}, a_l\}$, $\{\theta_{1,l}, a_l\}, \{\phi_{1,l}, a_l\}$, and $\{v_l, a_l\}$, all parameters are updated in order, and each time only one parameter can be updated at a time, then the next guidance parameter is updated in sequence until iteratively converges. The update equation of each parameter in θ_l is as follows

$$
\hat{\tau_l}^{"} = \arg\max_{\tau_l} z(\hat{\phi}'_{1,l}, \hat{\theta}'_{1,l}, \hat{\phi}'_{2,l}, \hat{\theta}'_{2,l}, \hat{\tau_l}', \hat{v}'_{1,l}, \hat{x}_l)
$$

$$
\hat{\theta}^{"}_{2,l} = \arg\max_{\theta_{2,l}} z(\hat{\phi}'_{1,l}, \hat{\theta}'_{1,l}, \hat{\phi}'_{2,l}, \hat{\theta}'_{2,l}, \hat{\tau_l}^{"}, \hat{v}'_{1,l}, \hat{x}_l)
$$

$$
\hat{\phi}^{"}_{2,l} = \arg\max_{\theta_{2,l}} z(\hat{\phi}'_{1,l}, \hat{\theta}'_{1,l}, \hat{\phi}'_{2,l}, \hat{\theta}^{"}_{2,l}, \hat{\tau_l}^{"}, \hat{v}'_{1,l}, \hat{x}_l)
$$

$$
\hat{\theta}^{"}_{1,l} = \arg\max_{\theta_{1,l}} z(\hat{\phi}'_{1,l}, \hat{\theta}'_{1,l}, \hat{\phi}^{"}_{2,l}, \hat{\theta}^{"}_{2,l}, \hat{\tau_l}^{"}, \hat{v}'_{1,l}, \hat{x}_l)
$$

$$
\hat{\phi}^{"}_{1,l} = \arg\max_{\theta_{1,l}} z(\hat{\phi}'_{1,l}, \hat{\theta}^{"}_{1,l}, \hat{\phi}^{"}_{2,l}, \hat{\theta}^{"}_{2,l}, \hat{\tau_l}^{"}, \hat{v}'_{1,l}, \hat{x}_l)
$$

$$\hat{v}_l^{"} = \arg\max_{v_l} z(\hat{\phi}_{1,l}^{"}, \hat{\theta}_{1,l}^{"}, \hat{\phi}_{2,l}^{"}, \hat{\theta}_{2,l}^{"}, \hat{\tau}_l^{"}, \hat{v}_{1,l}^{'}, \hat{x}_l)$$

$$\hat{a}_l^{"} = \left[|c_2(\hat{\phi}_{2,l}^{"}, \hat{\theta}_{2,l}^{"})| \cdot |(\hat{\phi}_{1,l}^{"}, \hat{\theta}_{1,l}^{"})| IPT_{sc}\right]^{-1}$$

$$\cdot z(\hat{\phi}_{1,l}^{'}, \hat{\theta}_{1,l}^{"}, \hat{\phi}_{2,l}^{"}, \hat{\theta}_{2,l}^{"}, \hat{\tau}_l^{"}, \hat{v}_{1,l}^{'}, \hat{x}_l) \tag{3.13}$$

It can be seen from Eq. 3.13 that the θ_l under each path is estimated, and the MLE of θ can be obtained by bringing it into the equation, that is, the parameters of the channel impulse response can be determined.

Channel parameter analysis is mainly to obtain the main channel statistical characteristics from the parameters extracted from the channel, including the basic parameter distribution fitting method, basic correlation analysis, as well as calculating the channel characteristics from the delay domain and the angle domain.

3.4.1 Basic Distribution Fitting Method

The results of the channel parameter fitting are usually expressed using the probability density function (PDF) or cumulative distribution function (CDF) of the channel parameter following the distribution function containing the parameters. According to the measured discrete data points, certain criteria (for example, commonly used ML, Least Square (LS), Minimum Mean Square Error (MMSE) or Uniformly Minimum Variance Unbiased Estimation (UMVUE)) to determine the parameters of the distribution function and describe the statistical characteristics of parameters [71]. In channel parameter estimation, according to the n observations $x_i(in)$ of the random variable X, the specific form of the PDF is generally assumed first, but the specific parameter value is unknown and the channel parameter is estimated. The purpose is to use known measurements to estimate unknown parameters to determine the specific expression of the PDF[25]. For example,

assuming that the PDF of a certain parameter follows a normal distribution, through a set of measurement data $x_1, x_2, \ldots, x_n$, the expression of the mean and variance of the normal distribution can be obtained as

$$\hat{\mu} \triangleq E[X] = \frac{1}{n}\sum_{i=1}^{n} x_i, \hat{\sigma}^2 \triangleq Var[X] = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{\mu})^2 \tag{3.14}$$

In the process of channel statistical parameter estimation, are commonly applied include log-normal distribution function, Gaussian distribution, Rayleigh distribution, Rice distribution, etc. Also some angular distribution functions are derived such as Uniform distribution, Winding Normal distribution and Laplacian distribution [72].

The channel impulse response (CIR) can be obtained from the measurement data, and the channel's Power Delay Profile (PDP), Delay Spread (DS), root mean square delay spread, power azimuth spectrum (PAS), angular spread (AS), and path loss (PL)can be estimated [72].

3.4.2 Path Parameter Fitting

Establishing the relationship between path loss and distance is a very important part of channel modeling. From the definition of path loss, we can calculate the difference under the conditions that the transmission power, measured distance, received power, and channel frequency are all known. The path loss $PL(di)$ of the different distance measurement points di is simply expressed as

$$PL(d_i) = P_{Tx} - P_{Rx}(d_i) + G_t - G_T \tag{3.15}$$

Using the actual pass loss data calculated from the measurement points at different distances and the same distance measurement point, the slope and intercept in different pass loss models can be calculated using linear regression analysis

$$\begin{bmatrix} PL(d_1) \\ PL(d_2) \\ \vdots \\ PL(d_n) \end{bmatrix} = \begin{bmatrix} 10log_{10}(d_1) \\ 10log_{10}(d_2) \\ \vdots \\ 10log_{10}(d_n) \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} \tag{2}$$

The model to get the path loss is

$$\overline{PL}(d) = \alpha 10log_{10}(d) + \beta \tag{3.16}$$

3.4.3 Shadow Fading Variance

According to the fading characteristics of wireless channels, large-scale pass loss is usually modeled as a linear model, and the actual pass loss exhibits random fluctuations. Multiple measurements of shadow fading are obtained by the deviation between the N actual measured path loss values $PL(d)$ and the expected path loss $\overline{PL}(d)$

$$SF(d_n) = PL(d_n) - \overline{PL}(d_n) \tag{3.17}$$

From equation (12), the variance of shadow fading is

$$\sigma_{SF} = \sqrt{\frac{1}{N}\sum_{n=1}^{N} |SF(d_n)|^2} \tag{3.18}$$

In PL modeling, the slope and intercept are often found by minimizing shadow fading [73].

3.4.4 PDP and DS

The multi-path delay power spectrum of the channel describes the average power at the multi-path delay, and the impulse response obtained from the measurement data is easily calculated

$$P(t,\tau) = |h(t,\tau)|^2 \tag{3.19}$$

The root mean square delay spread is the square root of the second moment of the PDP, which represents the spread of the additional delay around the average delay. Its expression is

$$DS = \sqrt{\frac{\sum_{l=1}^{L} p_i \tau_i^2}{\sum_{l=1}^{L} p_L} - \tau_0^2} \tag{3.20}$$

Where L is the total number of paths estimated by the SAGE algorithm, and τ_0 is the average additional delay, which is the statistical average of the PDP.

3.4.5 PAS and AS

The high-resolution parameter estimation algorithm can obtain multiple path parameters in the actual channel, such as angle, delay, gain in each polarization direction, and so on. The angle includes Angle of Arrival (AOA), Angel of Departure (AOD), Zenith Angles of Arrival (ZOA), Zenith Angles of Departure (ZOD), etc., the angle field is analyzed by PAS and AS indicators [74].

PAS mainly describes the distribution of power in the angle domain. The main calculation method is to select an observation angle and then accumulate other dimensions, such as time delay[26]. The winding process is usually used to calculate AS, and the specific process is explained by taking AOD as an example [48].

First introduce a small offset Δ to all the angle information, then $AOD_i(\Delta) = AOD(i) + \Delta$. After that, wind the new AOD, as shown in equation (23), for the azimuth angle in the horizontal direction, the winding range is $[-\pi, \pi)$, and for the elevation angle, the range is $[-\pi/2, \pi/2)$.

$$AOD_i(\Delta) = mod(AOD_i(\Delta) + pi, 2\pi) - \pi \tag{3.21}$$

Then the angle distribution function is estimated by PAS as

$$pas(AOD_i'(\Delta)) = \frac{PAS(AOD_i'(\Delta))}{\sum_{i=1}^{N_x} PAS(AOD_i'(\Delta))} \tag{3.22}$$

The PAS in equation (21) is the normalized PAS, the average value of the adjusted angle is calculated and removed from the angle, and then the second winding process is obtained

$$\begin{aligned} &AOD_i^*(\Delta) = mod(AOD_i(\Delta) - \\ &(\sum_{i=1}^{N_x} AOD_i'(\Delta) pas(AOD_i'(\Delta))) + \pi, 2\pi) - \pi \end{aligned} \tag{3.23}$$

Then calculate the second order matrix of the angle

$$\sigma(\Delta) = \sqrt{\sum_{i=1}^{N_x} [AOD_i'(\Delta)]^2 pas(AOD_i'(\Delta))} \tag{3.24}$$

AS is the smallest angular second-order matrix $\sigma(\Delta)$ that can be obtained by changing Δ, which is

$$AS = \min_{\Delta} \sigma(\Delta) \tag{3.25}$$

In the actual measurement, the mean and variance of AOD can be obtained according to the AOD measured from multiple positions, and the same process can be applied to AOA and other angles [64].

3.5 Simulations and Performance Analysis

In this section, the simulations of measured data and modeled channel is derived, which includes distribution of received data, AOA, AOD, PDP, small scale PDP, PL and distribution fitting.

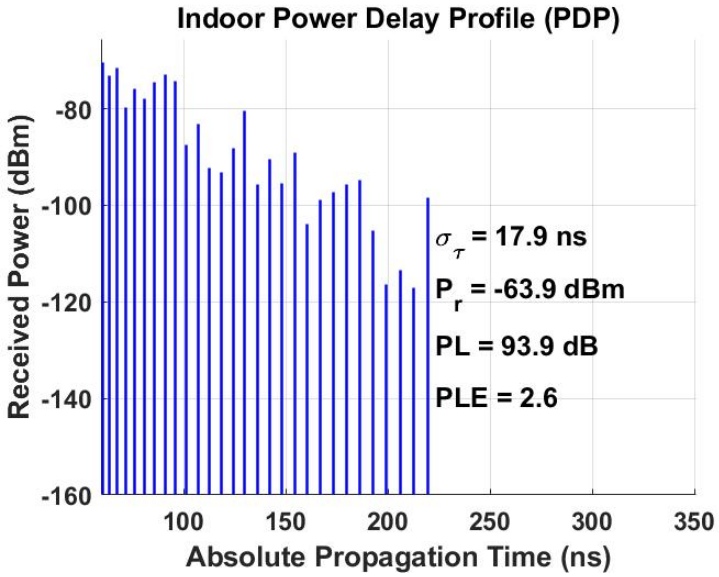

Figure 3.3. PDP of received indoor signal.

3.5.1 Indoor Scenario

Due to the direct propagation, reflection, diffraction and scattering of the signal propagation process, the multi-path phenomenon of the signal is generated. PDP is usually used to indicate the distribution of the received signal power with the multi-path arrival delay. It can be seen intuitively that the main path can be distinguished. Arrival delay is the main power distribution of multi-path. Fig. 3.3 and 3.4 shows the PDP and PDP with strongest power of indoor scenario.

It can be seen from the Fig. 3.3 and 3.4 that the power of PDP with strongest power is significantly higher than that of PDP, and the arrival delay is also smaller than that of PDP. This result is consistent with the LOS propagation characteristics. The smaller the arrival delay, the stronger the received signal power. However, the path loss is also higher than PDP, due to the high energy, it is easily affected by fading.

The RMSEs for Indoor and Through-wall scenario signals are listed in Table II. It demonstrates that Log-normal is the distribution that fits the channel coefficient data best.

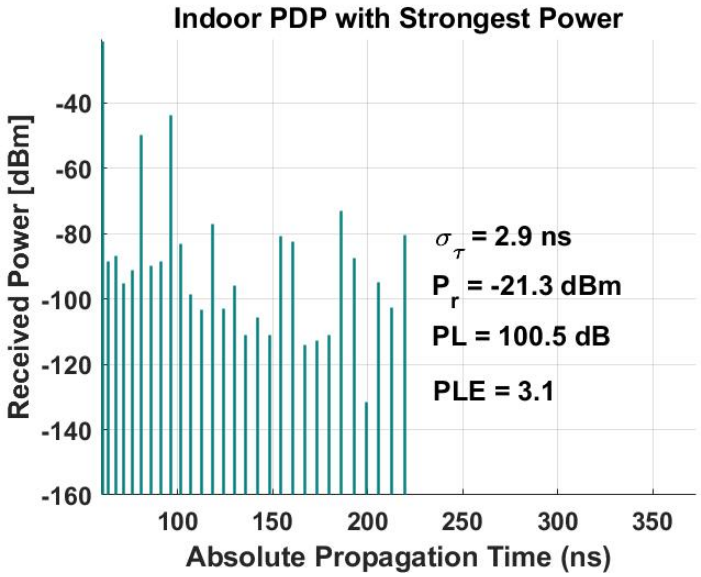

Figure 3.4. PDP with strongest power of received indoor signal.

Table 3.2. RMSE Comparison among Different Distribution Fittings

Scenario	Log-normal	Log-logistic	Weibull	Rayleigh
Indoor	3.351	4.756	6.881	7.285
Through-wall	3.183	4.336	6.706	7.564

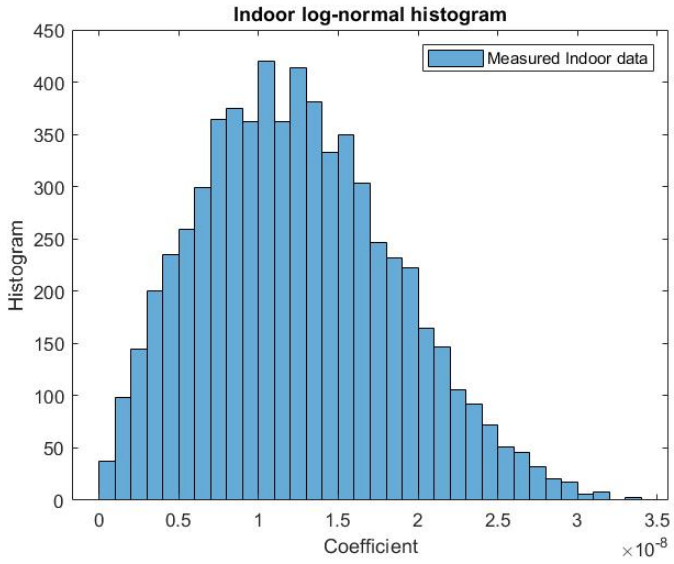

Figure 3.5. Histogram of received indoor signal.

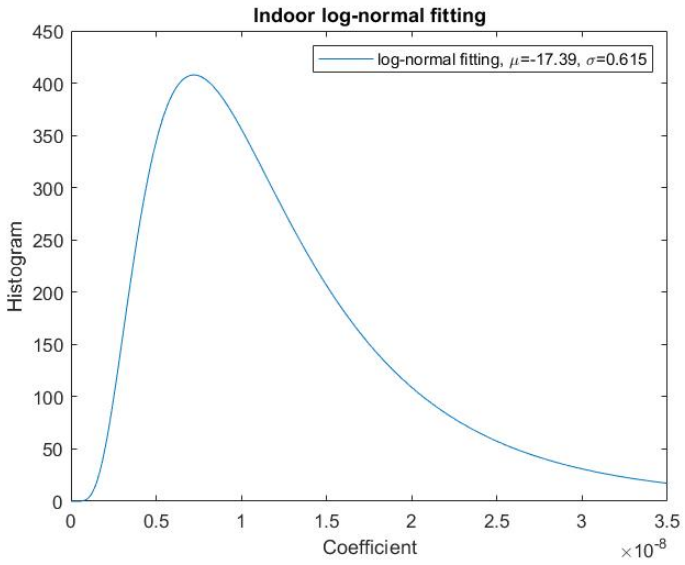

Figure 3.6. Log-normal fitting of received indoor signal.

The histogram of received indoor signal is shown in Fig.3.5 and 3.6, displays the log-normal curve fitting of received signal. The Root Mean Square Error (RMSE) of fitting is 3.351, which is smaller than other distribution fitting we have tried when processing the data. The RMSEs for Indoor and Through-wall scenario signals are listed in Table II. It demonstrates that Log-normal is the distribution that fits the channel coefficient data best.

The power delay spectrum is used to describe the relationship between power and delay, and from the angle power spectrum, we can see the strength of the received signal at each angle. Fig. 3.7 and 3.8 shows the AOA and AOD power spectrum of the received signal in the indoor scene. Because it is indoor LOS propagation, and we are using URA TX antenna and RX antenna, the propagation distance is close, it can be observed that the angle of AOA and AOD tends to be directional.

Fig. 3.9 proposes the small-scale PDP of received signal under indoor scenario. If we look it as a 2-D figure, it is close to the PDP of the signal. The PL of received signal is

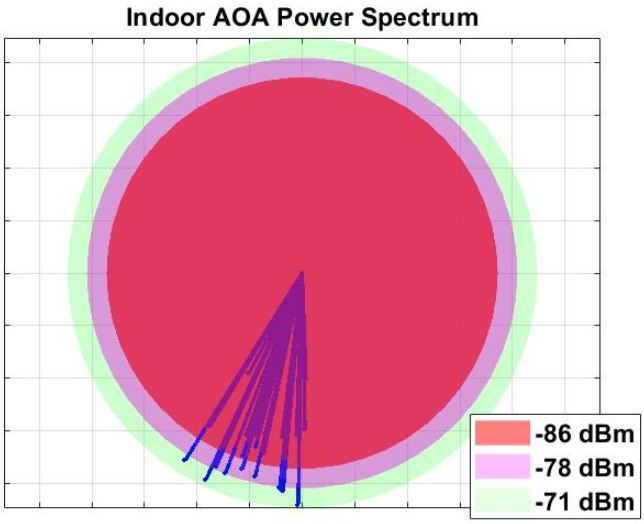

Figure 3.7. AOA power spectrum of received indoor signal.

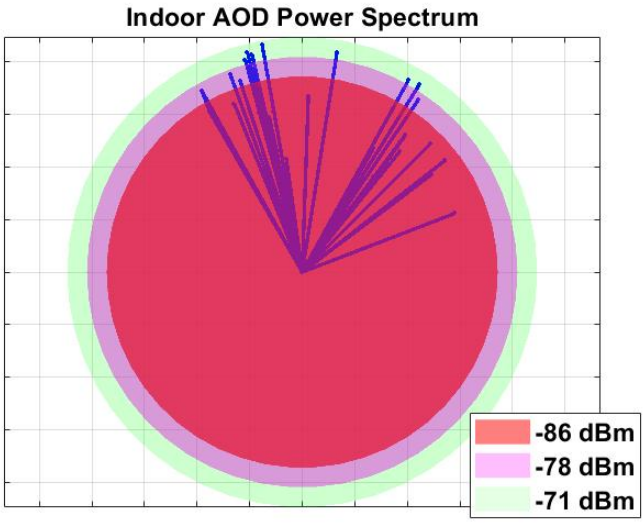

Figure 3.8. AOD power spectrum of received indoor signal.

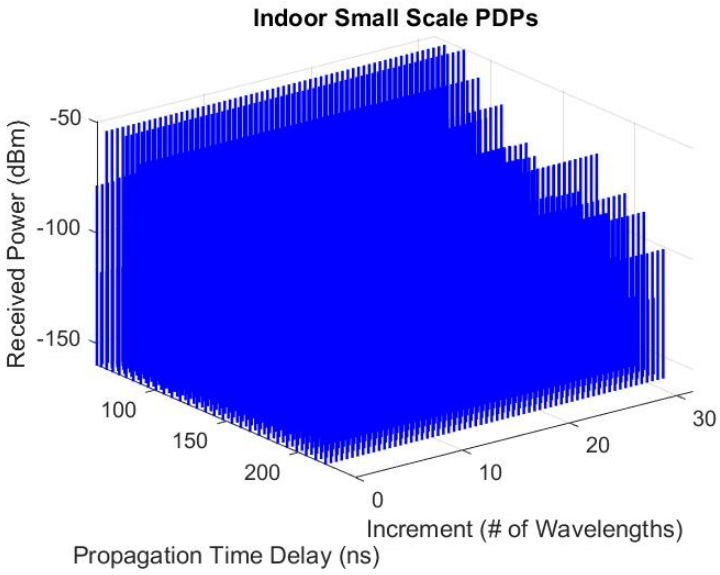

Figure 3.9. Small scale PDP of received indoor signal.

shown in Fig. 3.10, it is very clear to observe that the PL of strongest signal is less than the average. This is also consistent with the propagation characteristics of LOS.

3.5.2 Through-wall Scenario

The simulations of through-wall signal are displayed as following. From the figures we can observe the power of received signal is lower than indoor scenario due to multi times reflection.

The characteristics of NLOS propagation and LOS propagation are different. NLOS transmission has no direct path, and its propagation distance, main multipath and LOS are different. Compared with PDP in LOS scenes, the number of resolvable multipaths in NLOS scenes is smaller and the signal strength is weaker. Followed as Fig. 3.11 and 3.12 shown.

The histogram of received signal and fitting curve is displayed in Fig. 3.13 and 3.14. If we compare them with indoor received signal, we can observe they are roughly

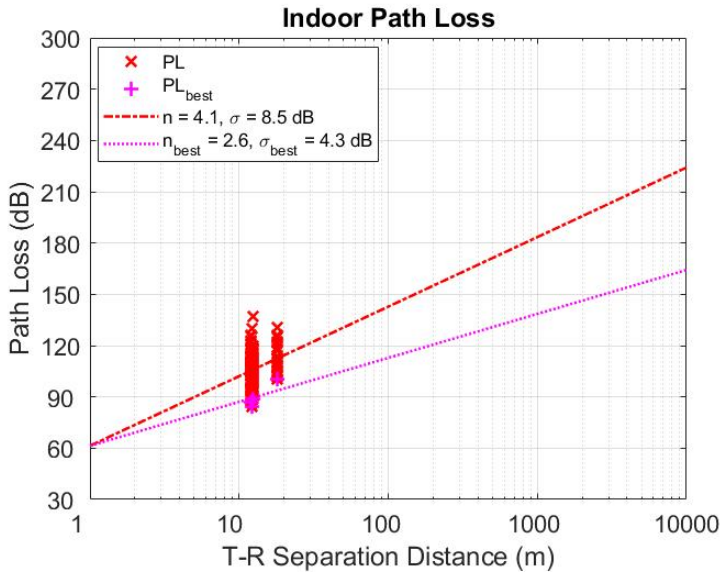

Figure 3.10. Path loss of received indoor signal.

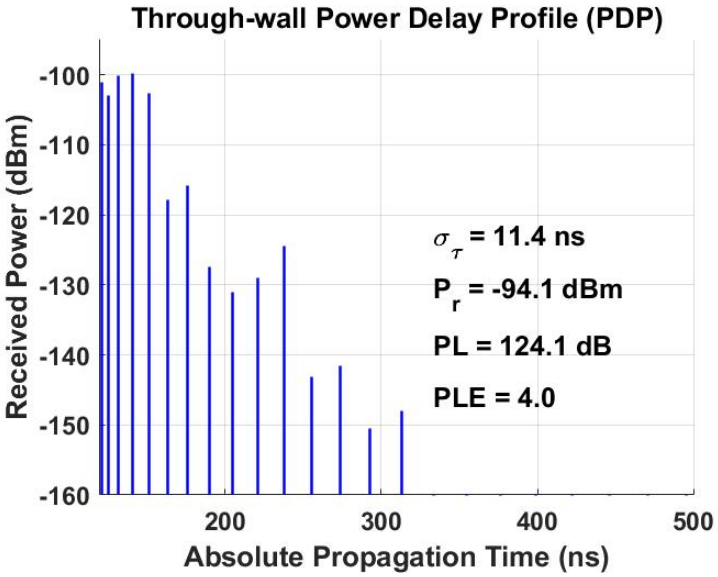

Figure 3.11. PDP of received through-wall signal.

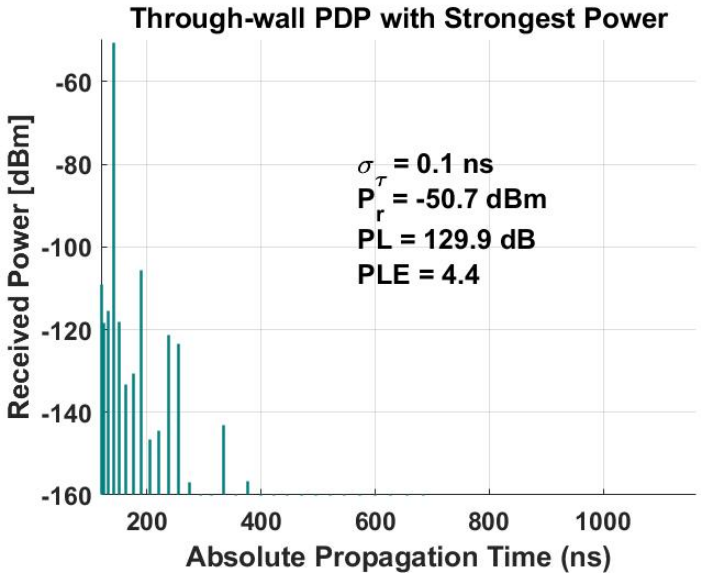

Figure 3.12. PDP with strongest power of received through-wall signal.

Table 3.3. Estimated parameters for log-normal

Central frequency	Parameter Scenario	Indoor	Through-wall
5*28 GHz	μ	17.39	18.58
	σ	0.615	0.628
	ϵ_μ	0.075	0.279
	ϵ_σ	0.059	0.164
	RMSE	3.351	3.183

similar, only the parameters are different. The RMSE of log-normal fitting for through-wall received signal is 3.183.

Table. 3.3 shows the comparison of parameters of log-normal fittings for indoor scenario and through wall scenario.

Compared Fig. 3.15 with Fig. 3.7 and Fig. 3.16 with Fig. 3.8, due to there is almost no direct received signal in through-wall scenario, the AOA of it is closer to omnidirectional. And the angle spread of AOD is also larger than indoor scenario.

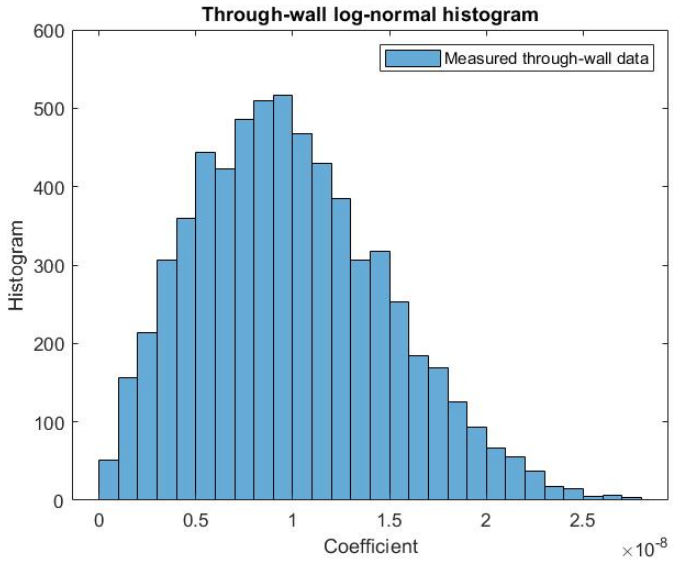

Figure 3.13. Histogram of received through-wall signal.

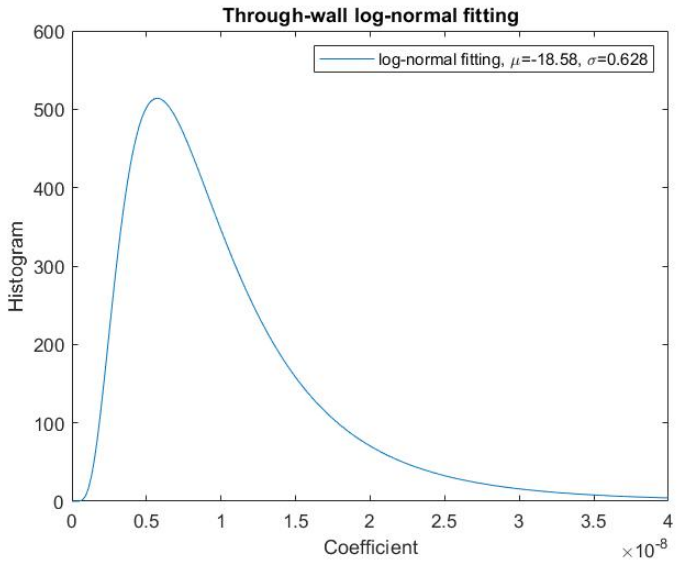

Figure 3.14. Log-normal fitting of received through-wall signal.

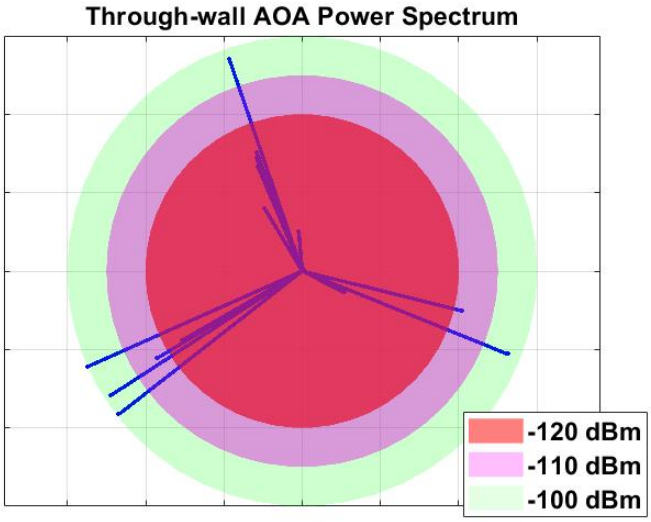

Figure 3.15. AOA power spectrum of received through-wall signal.

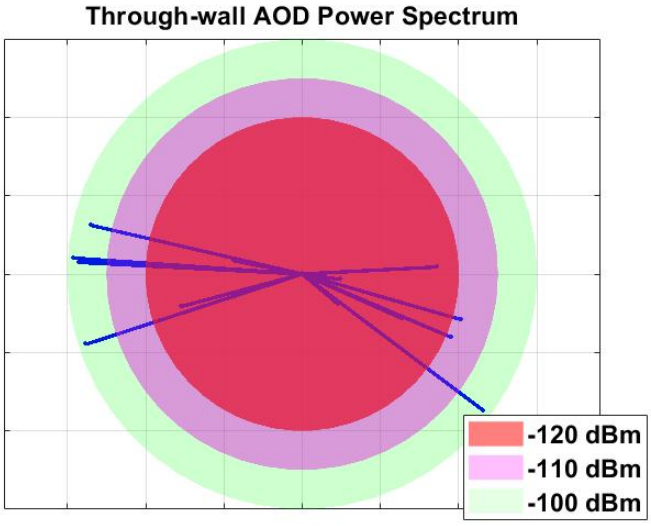

Figure 3.16. AOD power spectrum of received through-wall signal.

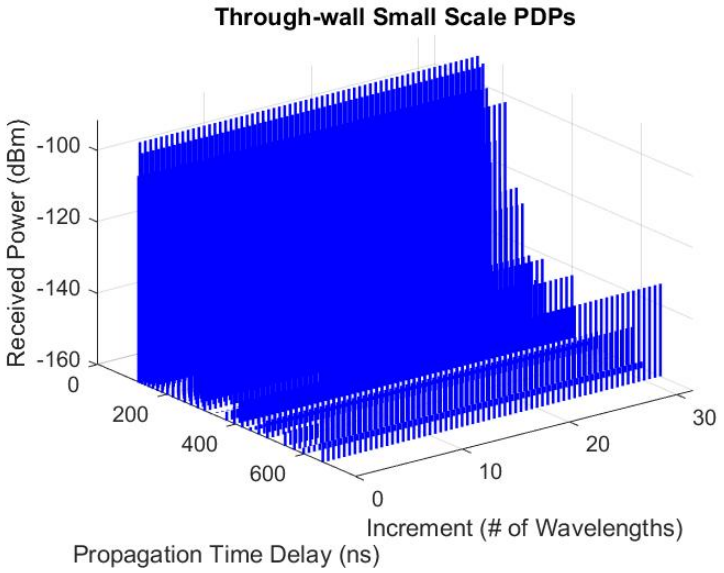

Figure 3.17. Small scale PDP of received through-wall signal.

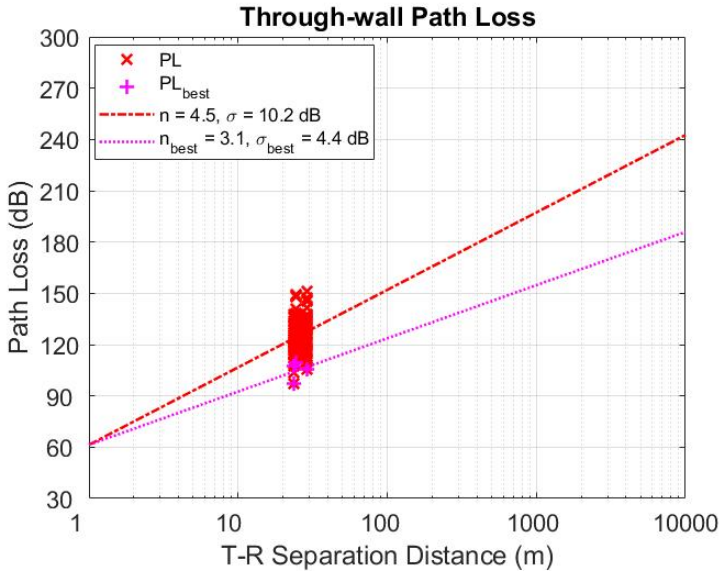

Figure 3.18. Path loss of received through-wall signal.

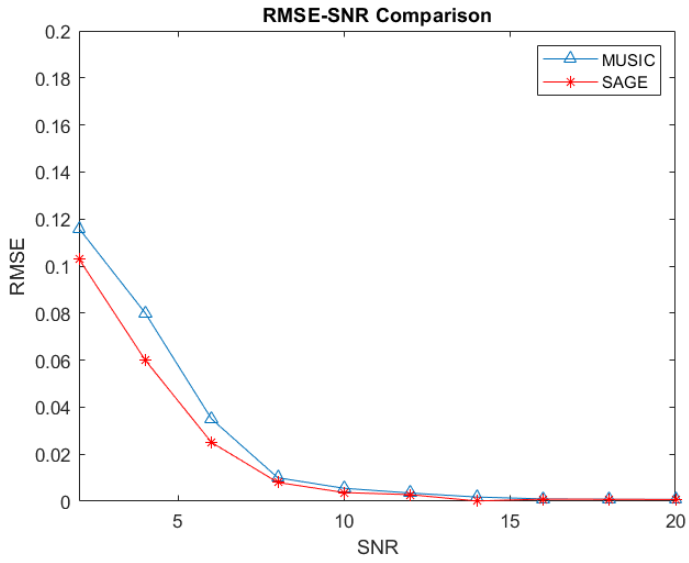

Figure 3.19. RMSE-SNR comparison of MUSIC and SAGE method.

Fig. 3.17 and 3.18 is the small-scale PDP and PL under through-wall scenario. It is easy to observe that both shows the fading and PL of through-wall is stronger than indoor scenario.

Fig. 3.19 presents the RMSE-SNR comparison of a common subspace-based estimation method, MUSIC, and the SAGE method which is derived in this dissertation. From the figure we can observe that when the SNR is the same, the RMSE of SAGE is smaller than the MUSIC method, which illustrates the better performance of SAGE method.

CHAPTER 4

Channel Forecasting and Power Allocation Based on LSTM Network and Cooperative Communication

4.1 Channel Forecasting Based on LSTM Network

4.1.1 Introduction to Recurrent Neural Networks

Time series is a series of data points obtained by a certain element according to the time sequence index, which reflects the law of the change of this element over time [75]. Time series analysis is precisely through the learning of these changing laws to predict possible future trends. Time series data exists in various fields such as meteorology, economy, medicine, electric power, and transportation, and of course, the field of wireless communication is also included. Time series is to collect and record observed things in the order of time. The data usually has the three characteristics of large data scale, high dimensionality, and constant updating [76]. These data are usually viewed as a whole, rather than looking at individual values one-sidedly. We can briefly understand the time series from different angles. First of all, from the perspective of predicting the step size, there are long, medium and short points. Of course, there is no difference between the better and the worse for the prediction step length, and it can be selected according to actual needs. From the point of view of statistical characteristics, time series data includes stationary series and non-stationary series [77]. In practical applications, time series data are mostly non-stationary series. Therefore, research on time series forecasting is also mainly focused on non-stationary series [78]. From the perspective of the composition of the sequence, the time series can be divided into the following parts: the trend part, which usually determines the trend of the sequence: seasonal changes, that is, some cyclical changes: cyclical

changes, often manifested as oscillating phenomena: random changes, that is, random factor interference [79].

Traditional linear time series models mainly include exponential smoothing, ARMA and so on. These methods can achieve ideal results under the premise that the sequence is stationary, but if these prediction methods for stationary sequences are used in the analysis of non-stationary time series, wrong conclusions will often be obtained. With the advancement and improvement of various machine learning algorithms, many nonlinear time series prediction models based on machine learning algorithms have emerged. For time series forecasting of some data, neural networks can achieve better results than traditional statistical-based methods. This benefits from the neural network's strong fitting ability and good generalization ability for nonlinear functions. It is also during this period of rapid development of neural networks that a large number of time series prediction studies based on artificial neural networks have emerged [80]. In these studies, neural networks have indeed demonstrated excellent processing capabilities for nonlinear sequences.

With the improvement of computing power, deep neural networks began to appear in time series forecasting. The main characteristics of deep learning are deep levels and many neurons, which guarantees the ability to express nonlinearity. As known, Convolutional Neural Network (CNN) is the most general and useful neural network, which is the most popular method for image processing in recent years. There are also Recurrent Neural Network (RNN), along with the evolved Long Short-Term Memory (LSTM) network, had been proven that the processing of time series achieves remarkable results. [81].

Facts have proved that the RNN networks in deep learning can predict the future state of time series very well. It can acquire multi-level features through learning like a universal neural network model, and on this basis, it can also evolve into high-level features through the learned basic features, and finally discover the distributed characteristics of the data , The bigger feature is that its neurons have added memory function [82]. The

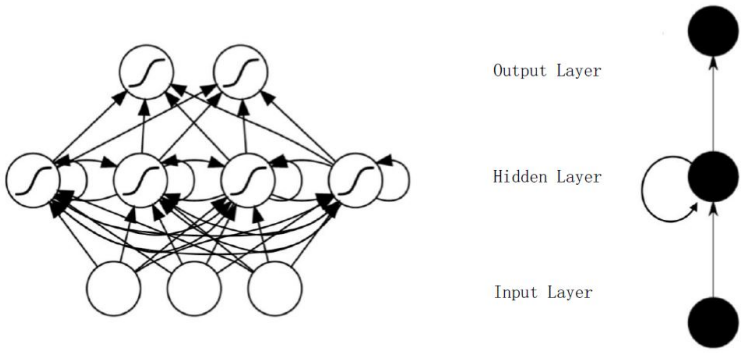

Figure 4.1. Structure diagram of RNN.

hidden neurons of the RNN increase the interconnection, which realizes the memory of information, and is not easy to lose in the prediction, so that some influencing factors can be used, and the model can avoid over-fitting the training data [83].

4.1.2 The Recurrent Neural Network

4.1.2.1 The Structure of RNN

The structure of the RNN is shown in Fig. 4.1, it also consists of input layer, hidden layer and output layer. But what it is more perfect than the general neural network is that for the nodes of each hidden layer, their input is not only from the input layer, but also from the output of the node of the previous hidden layer.

The detailed structure of RNN model is shown in Fig. 4.2. Among them, the initial input from the input layer at time t is expressed by x_t, s_t is the output of the hidden layer, and y_t is the output of the output layer. According to Fig. 4.2, the value of s_t not only depends on x_t, but also is related to s_{t-1}, this is the structural factor of the RNN model. Based on this, we will forecast y_t , by using both the data of x_t and the signal of x_{t-1} . The structure of one RNN neuron and the training model process are shown in Fig. 4.3:

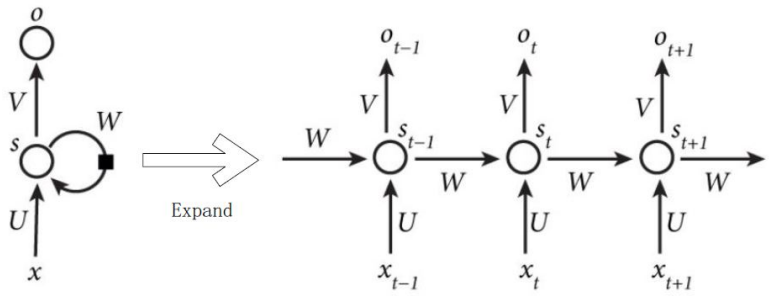

Figure 4.2. Expanded structure of RNN.

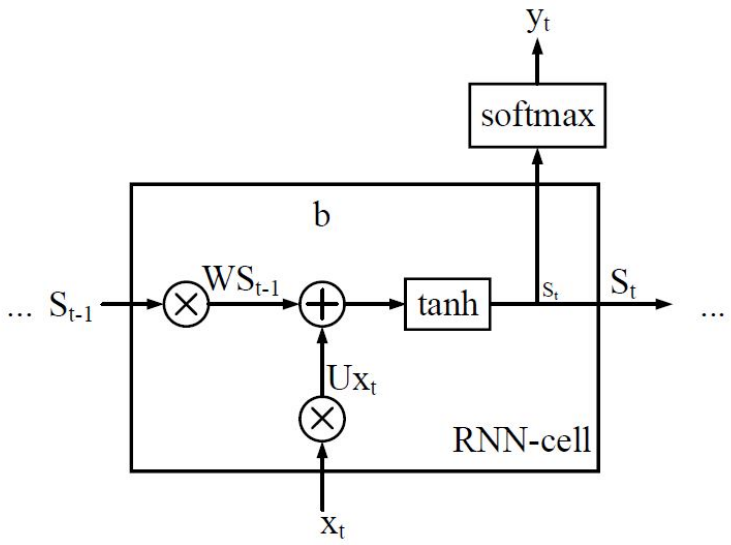

Figure 4.3. Structure of one RNN Unit.

4.1.2.2 The Training Principle of RNN

The difference between the training method of RNN network and the traditional neural network training method is that the training parameters (W, U, V) in RNN are shared. And in the gradient descent method used by RNN, the output of each step is not only related to the current output, but also related to the memory state of the previous steps [84]. The backpropagation training algorithm used in RNNs unrolls the neural network in time, redefining the connections in the network.

After expanding the RNN network, it can be seen that each layer has the same weight matrix (W, U, V), and the update process of the weight matrix is as follows:

Suppose $x(t)$ is the input time series, $u(t)$ is the intermediate calculation result, the final output result is $\hat{y}(t)$, and the true value of the training sample is represented by $y(t)$, where

$$\begin{aligned} x(t) &= (x_1(t), x_2(t), ...x_k(t))' \\ y(t) &= (y_1(t), y_2(t), ...y_k(t))' \\ t &= 1, 2, ...T \end{aligned} \tag{4.1}$$

The final output error of the model can be expressed as:

$$E = \sum_{t=1}^{T} ||y(t) - \hat{y(t)}||^2 = \sum_{t=1}^{T} E(t) \tag{4.2}$$

The output of the model can be obtained through the forward propagation algorithm, and then the error is back propagated along the time series. The error term at each time point can be expressed as:

$$\delta_j = (y_j(T) - \hat{y}_j(T)) \left. \frac{\partial f(u)}{\partial u} \right|_{u=S_j(T)} \tag{4.3}$$

$$\delta_i(T) = \left[\sum_{j=1}^{L} \delta_j(T) \right] \left. \frac{\partial f(u)}{\partial u} \right|_{u=S_i(T)} \tag{4.4}$$

where $\delta_j(T)$ is the error of the output unit at time T, $\delta_i(T)$ is the error of the intermediate state $u_i(T)$ at time T, the errors of $\delta_j(T)$ and $\delta_i(T)$ are calculated as follows:

$$\delta_j(t) = \left(y_j(T) - \hat{y}_j(T) + \sum_{i=1}^{N} \delta_i(t+1) V'_{ij} \right) \left. \frac{\partial f(u)}{\partial u} \right|_{u=S_j(t)} \tag{4.5}$$

$$\delta_i(t) = \left(\sum_{i=1}^{N} \delta_i(t+1) W_{ij} + \sum_{j=1}^{L} \delta_j(t) V_{ij} \right) \left. \frac{\partial f(u)}{\partial u} \right|_{u=S_i(t)} \tag{4.6}$$

By using the backpropagation algorithm, the update process of the weight matrix is as follows:

$$W_{ij} = W_{ij} - \eta \sum_{t=1}^{T} \delta_i(t)\mu_j(t-1), \mu_j(t-1) = 0, t = 1 \tag{4.7}$$

$$U_{ij} = U_{ij} - \eta \sum_{t=1}^{T} \delta_i(t)x_j(t) \tag{4.8}$$

$$V_{ij} = V_{ij} - \eta \sum_{t=1}^{T} \delta_i(t)u_j(t) \tag{4.9}$$

$$V'_{ij} = V'_{ij} - \eta \sum_{t=1}^{T} \delta_i(t)\hat{y}_j(t-1), \hat{y}_j(t-1) = 0, t = 1 \tag{4.10}$$

Among them, W_{ij} is the weight matrix between neurons in the hidden layer, U_{ij} is the weight matrix between the input layer and the hidden layer, V_{ij} is the weight matrix between the hidden layer and the output layer, V'_{ij} is the weight matrix of output layer to the hidden layer.

But at the same time, the standard RNN also has its own shortcomings. If the input sample spans a long time, the gradient will disappear. This phenomenon will cause the loss of information at a time far away from the current time. That is, the standard RNN can only remember for a short period of time. The long-term sample information has a low utilization rate. Therefore, people proposed a special RNN model - LSTM model, proposed by Hochreiter and Schmidhuber, and was revised and carried forward by many later scholars. Through the learning of nonlinear network, it can not only obtain approximate solutions of complex functions, but also extract input samples to grasp the essential characteristics. Furthermore, the gradient vanishing problem common to traditional RNNs is also solved by LSTM through the long-term memory function, in addition to achieve long-term memory of information [85].

4.1.3 The Long Short-Term Memory Network

4.1.3.1 Introduction and Structure of LSTM Network Model

Long short-term memory network is a variant of recurrent neural network [86], which is mainly used to solve long-term time-dependent problems. For example, in image description, speech recognition, and natural language processing, the LSTM model has performed well and is widely used in academic and industry fields.

Fig. 4.4 shows the structure of a neuron in an LSTM network. The LSTM network is essentially a unidirectional chain structure, and the internal structure of each neuron on the chain is the same, which is the same as the structure of RNN. But the difference of LSTM is that three new structures and special memory units are added to solve the gradient vanishing. The three gate structures are input gate, forget gate and output gate [87]. The gate structure contains the Sigmoid function layer, which can compress the value between 0-1, which helps to remember useful information for a long time or forget redundant information. When the previous data is needed, the activation function will be multiplied by 0 to get 0 output, and this part of the information will not be passed to the next neuron as input. Similarly, when encountering information data that needs to be memorized, it will be multiplied by 1 and still save itself, and then passed to the next neuron as input information. It is through this selective memory or forgetting of information that the LSTM network can save a part of useful information for a long time, and the basis for this phenomenon is the newly added three gate structures and memory units, and the problem of gradient vanishing can be solved.

In Fig. 4.4, x_t indicates the input at time t, or we can call it the CSI at time t. σ is the Sigmoid function, it is an activation function. $\otimes$ represents the calculation of element multiplication, and the forget gate, input gate and output gate is established by the combination of σ and $\otimes$ together, which are respectively shown from left to right. The

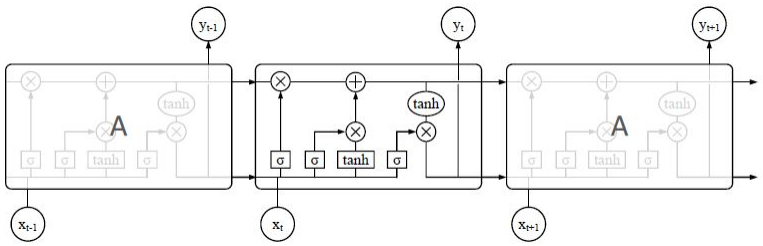

Figure 4.4. Structure diagram of LSTM.

information passing through the forget gate will be decided whether it needs to be forgotten, the signal passing through the input gate will be retained and added to the memory unit as a component, and the output of the next hidden state will be determined by the output gate.

4.1.3.2 The Training Principle of LSTM

The training method of the LSTM model covers the calculation process of forward calculation and back propagation. The specific calculation process is as follows:

1. Forward calculation process:
   (a) Forget gate: The input vector at time t and the output vector of the output layer at time $t-1$ jointly determine the output state of the forgetting gate at time t. The calculation method is shown in equation (4.11).

$$f_t = \sigma\left(W_f \cdot (y_{t-1}, x_t) + b_f\right) \tag{4.11}$$

where: f_t is the output vector at time t, σ is the activation function, W_f is the weight matrix, concated by matrix W_{fy} and matrix W_{fx}, y_{t-1}, x_t are the output vector of the output layer at time $t-1$ and the input vector of the input layer at time t, respectively. b_f is the bias term.

Among them, W_f can be presented as:

$$(W_f)\begin{pmatrix} y_{t-1} \\ x_t \end{pmatrix} = (W_{fy}W_{fx})\begin{pmatrix} y_{t-1} \\ x_t \end{pmatrix} = W_{fy}y_{t-1} + W_{fx}x_t \tag{4.12}$$

(b) Input gate: The input gate can be represented by the equation (4.13), and the meanings represented by the symbols in the formula are similar to those of the forget gate, which will not be repeated here.

$$i_t = \sigma\left(W_i \cdot (y_{t-1}, x_t) + b_i\right) \tag{4.13}$$

(c) Memory unit: The state of the memory unit at the current moment can be calculated in two steps. First, the output at time $t-1$ and the input at time t are used to calculate the state $\hat{c}_t$ of the current input unit. Its calculation expression can be expressed as:

$$\hat{c}_t = tanh\left(W_c \cdot (y_{t-1}, x_t) + b_c\right) \tag{4.14}$$

Then calculate the memory cell state c_t at the current moment:

$$c_t = f * c_{t-1} + i_t * \hat{c}_t \tag{4.15}$$

The $*$ symbol in the formula is the multiplication operation between each element. The unit state $\hat{c}_t$ input by the LSTM at the current moment and the long-term memory c_{t-1} are combined through the memory unit.

(d) Output gate:

$$o_t = \sigma\left(W_o \cdot (y_{t-1}, x_t) + b_o\right) \tag{4.16}$$

$$y_t = o_t * tanh(c_t) \tag{4.17}$$

Table 4.1. Learning Parameters

Module	Weight Matrix	Bias Item
Output gate	W_o	b_o
Forget gate	W_f	b_f
Input gate	W_i	b_i
Memory unit	W_c	b_c

y_t represents the final output of the LSTM model at the current time, that is, the spectral prediction value at time t. It can be seen from equation (4.17) that the output of the output gate and the state of the memory unit at the current moment jointly determine the value of y_t.

2. Back propagation process of error:

   Similar to the RNN model, the calculation of the error back propagation of the LSTM model also uses the value of the error term σ, and according to the calculated error term, the gradient descent method is used to update the weights. In the training process of the LSTM model, there are 4 groups of weights that the model needs to learn, and the specific learning parameters are shown in Table 4.1:

   The weight matrix plays a different role in the derivation process of backpropagation, so it is divided as follows:

$$\begin{aligned} W_o &= [W_{oy}, W_{ox}] \\ W_f &= [W_{fy}, W_{fx}] \\ W_i &= [W_{iy}, W_{ix}] \\ W_c &= [W_{cy}, W_{cx}] \end{aligned} \tag{4.18}$$

Assuming that the error (loss function) is E, and the output at time t is y_t, then the error term δ_t of the output layer at time t is:

$$\delta_t = \frac{\partial E}{\partial y_t} \tag{4.19}$$

For the above four weight matrices, there are four weighted inputs, which correspond to f_t, i_t, c_t and o_t at time t, respectively. Let their corresponding error terms be δ_{ft}, δ_{it}, δ_{ct} and δ_{ot}, respectively, then:

$$\begin{aligned} \delta_{o_t}^T &= \frac{\partial E}{\partial (W_o \cdot (y_{t-1}, x_t) + b_o)} \\ &= \frac{\partial E}{\partial y_t} \cdot \frac{\partial y_t}{\partial o_t} \cdot \frac{\partial o_t}{\partial (W_o \cdot (y_{t-1}, x_t) + b_o)} \\ &= \delta_t^T * tanh(c_t) * o_t * (1 - o_t) \end{aligned} \tag{4.20}$$

In the same way, it can be known from the chain rule:

$$\begin{aligned} \delta_{f_t}^T &= \frac{\partial E}{\partial (W_f \cdot (y_{t-1}, x_t) + b_f)} \\ &= \delta_t^T * o_t * (1 - tanh(c_t)^2) * c_{t-1} * f_t * (1 - f_t) \end{aligned} \tag{4.21}$$

$$\begin{aligned} \delta_{i_t}^T &= \frac{\partial E}{\partial (W_i \cdot (y_{t-1}, x_t) + b_i)} \\ &= \delta_t^T * o_t * (1 - tanh(c_t)^2) * \hat{c}_t * i_t * (1 - i_t) \end{aligned} \tag{4.22}$$

$$\begin{aligned} \delta_{\hat{c}t}^T &= \frac{\partial E}{\partial (W_c \cdot (y_{t-1}, x_t) + b_c)} \\ &= \delta_t^T * o_t * (1 - tanh(c_t)^2) * i_t * (1 - \hat{c}_t^2) \end{aligned} \tag{4.23}$$

The error propagation along the time direction is to calculate the error term δ_{t-1} at time t_1, and the expression of δ_{t-1} is:

$$\delta_{t-1} = \delta_{o_t}^T W_{oy} + \delta_{f_t}^T W_{fy} + \delta_{i_t}^T W_{iy} + \delta_{\hat{c}_t}^T W_{cy} \tag{4.24}$$

The calculation formula for the propagation of the error term to the upper layer can be expressed as:

$$\begin{aligned}\delta_{t-1}^{l} &= \frac{\partial E}{\partial(W \cdot (y_t, x_t) + b)} \\ &= (\delta_{o_t}^T W_{ox} + \delta_{f_t}^T W_{fx} + \delta_{i_t}^T W_{ix} + \delta_{\hat{c}_t}^T W_{cx}) * f'(W \cdot (y_t, x_t) + b) \end{aligned} \tag{4.25}$$

where $W \cdot (y_t, x_t) + b)$ represents the weighted input of layer $l-1$, and f is the activation function of layer $l-1$.

Taking the output gate as an example, the calculated error term is used to calculate the gradient of each weighting matrix and the gradient of the bias term at time t. The calculation can be expressed as:

$$\frac{\partial E}{\partial W_{oy,t}} = \frac{\partial E}{\partial(W_o \cdot (y_{t-1}, x_t) + b_o)} \cdot \frac{\partial(W_o \cdot (y_{t-1}, x_t) + b_o)}{\partial W_{oy,t}} = \delta_{o_t} y_{t-1}^T \tag{4.26}$$

$$\frac{\partial E}{\partial W_{ox,t}} = \frac{\partial E}{\partial(W_o \cdot (y_{t-1}, x_t) + b_o)} \cdot \frac{\partial(W_o \cdot (y_{t-1}, x_t) + b_o)}{\partial W_{ox,t}} = \delta_{o_t} x_{t-1}^T \tag{4.27}$$

$$\frac{\partial E}{\partial b_{o,t}} = \frac{\partial E}{\partial(W_o \cdot (y_{t-1}, x_t) + b_o)} \cdot \frac{\partial(W_o \cdot (y_{t-1}, x_t) + b_o)}{\partial b_{o,t}} = \delta_{o_t} \tag{4.28}$$

The final gradient is the sum of the gradients at each time:

$$\frac{\partial E}{\partial W_{oy}} = \sum_{j=1}^{t} \delta_{o_j} y_{j-1}^T \tag{4.29}$$

$$\frac{\partial E}{\partial W_{ox}} = \sum_{j=1}^{t} \delta_{o_j} x_j^T \tag{4.30}$$

$$\frac{\partial E}{\partial b_o} = \sum_{j=1}^{t} \delta_{oj} \tag{4.31}$$

In the same way, the weight matrix and the gradient of the bias term of the forget gate, input gate and memory unit can be obtained.

4.2 Power Allocation Based on Cooperative Communication

In cooperative communication, nodes cooperate with each other, which can not only expand the communication range, but also improve the information transmission rate of the system. The research on resource allocation in cooperative communication mainly focuses on power allocation. By rationally allocating the power of nodes, the system performance can be improved while the power consumption can be reduced [88].

When a relay node cooperates with other nodes to transmit information, it should not only consider improving the system performance, but also consider when to cooperate, with whom and how to cooperate. Compared with the straight forward communication link, cooperative communication has better transmission quality and channel capacity, but the structure of cooperative communication is complex and the amount of calculation is large. The direct transmission link has a simple structure and low complexity. Due to the limited wireless communication resources, the main goal of cooperative communication is to improve the performance of the communication network and allocate power reasonably under the premise that the total power of the wireless communication system is constrained.

The main structure of this section is arranged as follows. The first part expounds the system model of the cooperative communication network; the second part mainly introduces several typical power allocation algorithms; the third and fourth parts study the objective function of maximizing the information transmission rate and optimize the node power allocation in the single relay scenario; Section V studies the channel capacity of AF and DF transport protocol systems

4.2.1 System Model of Cooperative Communication Network

In cooperative communication, system model is the most basic research content. According to different classification standards, the following simple cooperative communication network system models are introduced respectively.

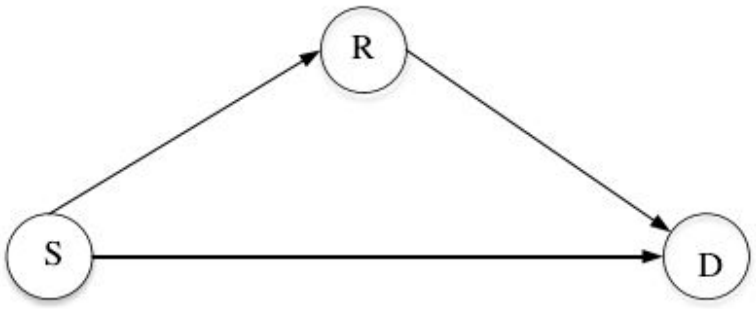

Figure 4.5. Single-relay two-hop system model.

1. Single-relay two-hop system model

   In the system model of Fig. 4.5, the number of source nodes, relay nodes and destination nodes are all 1. The signal sent by the source node S is forwarded by the relay node R, and finally transmitted to the destination node D. The single-relay two-hop network model has a simple structure and is relatively easy to implement [89].

2. Multi-hop relay system model

   As shown in Fig. 4.6, R is the relay node which processes the signals sent by the source node S in order, and then forwards them to the destination node D. The relay node forwards once, and the network model is a single relay network model, as shown in Fig. 4.5. The relay node forwards two or more times, and the model is a multi-hop relay network model, as shown in Fig. 4.6. When the distance between the source node and the destination node is relatively far, the model reduces the impact of path loss on transmission performance, and expands the coverage radius of the system.

3. One-way multi-relay system model

   As shown in Fig.4.7, the source node S forwards the signal through multiple cooperative relay nodes R, and finally transmits the signal to the destination node D. Single-relay cooperative transmission and multi-relay cooperative transmission depend on the number of relay nodes. When there is only one relay in the links, the model is

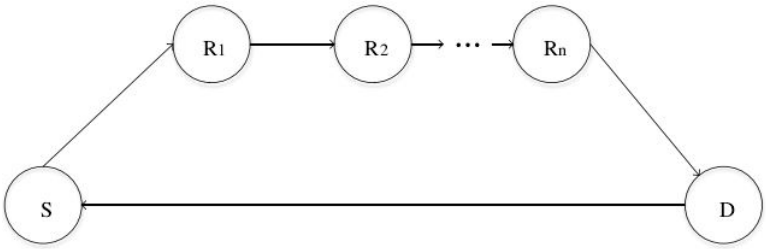

Figure 4.6. Multi-hop relay system model.

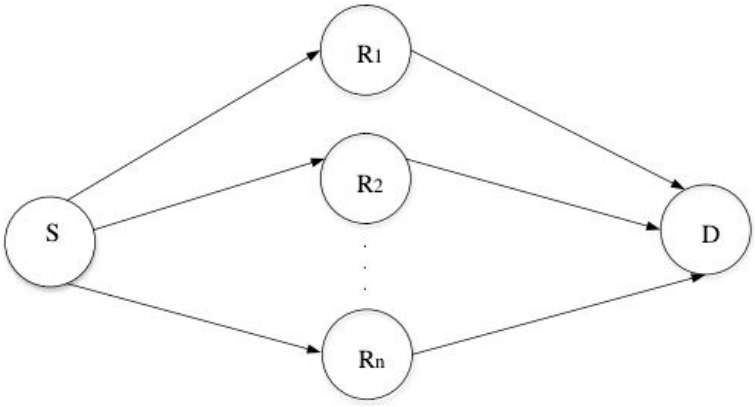

Figure 4.7. One-way multi-relay system model.

called single-relay cooperative communication; and the model is called multi-relay cooperative communication when there are multiple relays exist. The cooperative transmission of multiple relay nodes can improve the system performance and also obtain diversity gain.

4. Multi-source and multi-relay system model

   The commonly used cooperative communication network model consists of a single source node, multiple relay nodes and a single destination node, but the system in the actual scenario is dominated by the multi-source and multi-relay model. In actual communication, multiple source nodes may transmit information at the same time, and multiple access schemes need to be adopted, such as time division multiple access, frequency division multiple access, code division multiple access, and space

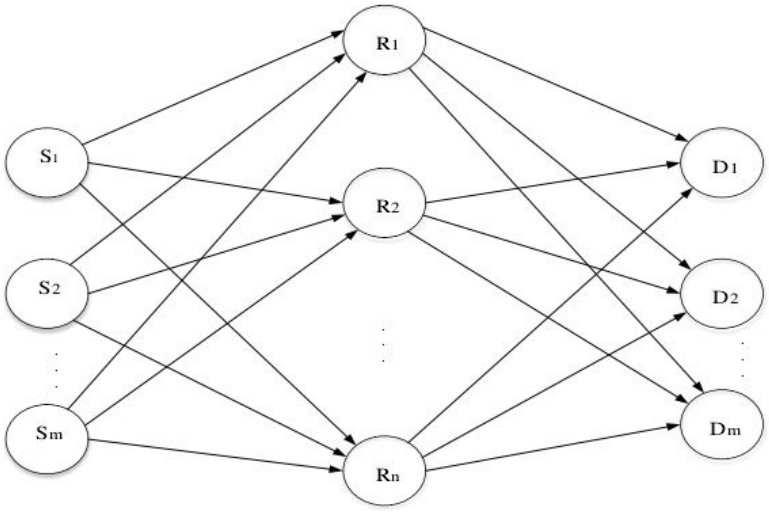

Figure 4.8. Multi-source and multi-relay system model.

division multiple access. Single-source node transmission and multi-source node transmission depend on the number of source nodes. When the number of source nodes is 1, the model is called a single-source communication model; when the number of source nodes is multiple, the model is called a multi-source communication model. As shown in Fig. 4.8, compared with the single-source relay network, the multi-source relay network is more in line with the actual communication scenario. Multiple source nodes jointly use the same set of cooperative relay nodes to complete communication transmission, and make the relay nodes reasonably allocate resources to the source nodes.

4.2.2 Power Allocation Algorithm

1. Equal Power Allocation Algorithm In the wireless cooperative communication system, in the first transmission link, the node S transmits information to node R and D directly. In the second transmission link, the cooperative node R processes the received information according to a certain protocol, and then transmits the processed

information to node D. P_S is the transmit power of the source node, P_{Ri} is the transmit power of the i_{th} relay node, P_0 is the total power constraint between the source node and the relay node, and the power constraint condition is $P_S + \sum_i^N P_{Ri} \leq P_0$. The equal power distribution is expressed by the following formula:

$$P_S = P_{R1} = P_{R2} = P_{Ri} = \frac{1}{N+1} P_0 \tag{4.32}$$

The equal power distribution algorithm is to distribute the total power equally to all nodes, and the algorithm complexity is low. But when the link state between the node S and the node R is better, the overall performance of the algorithm is not high. For the single relay cooperative communication system to be proposed in this dissertation, the transmission power of the source node and the relay node is $P_S = P_R = \frac{1}{2} P_0$

2. Power Allocation Algorithm Based on Minimum Outage Probability

   With the in-depth study of cooperative communication, power allocation has gradually become the main research part of cooperative communication technology, and the system outage probability is an important evaluation factor for an relatively optimised power allocation scheme. The power distribution method with the minimum outage probability refers to the calculation of the optimal allocation factor of the power of the node S and the node R under the premise of constraining the total power of the system, so as to minimize the outage probability of the system.

3. Power Allocation Algorithm Based on Maximizing Channel Capacity

   By analyzing the three-node two-hop model, it is verified that the relay channel can improve the channel capacity from information theory, which provides a foundation for the implementation of cooperative diversity technology. The allocation algorithm based on maximizing channel capacity is based on the premise of constraining the to-

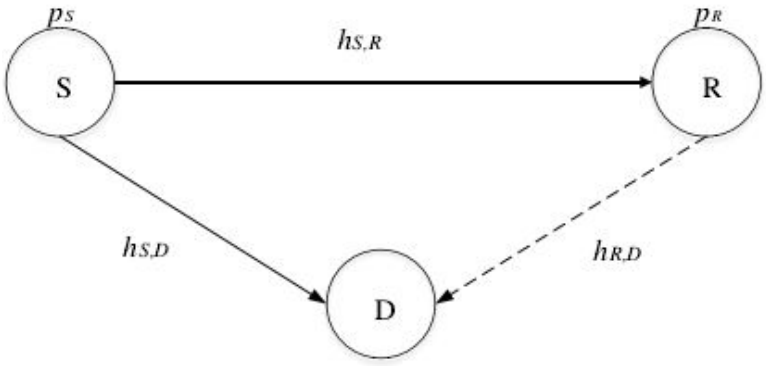

Figure 4.9. Cooperative Communication System Model.

tal power of nodes, and establishing the objective function according to the optimization objective of the maximum channel capacity C. The power distribution method obtains the optimal distribution factor between nodes by calculating the objective function of the maximum channel capacity, reasonably allocates the total power of the system, and optimizes the performance of the system

4.2.3 Modeling of Power Allocation Algorithm for Cooperative Communication

4.2.3.1 Cooperative Communication System Model

1. System Model

   As shown in Fig. 4.9, the model includes three nodes, which is source node S, relay node R and destination node D. There are two transmission links in the communication with different transmission time: the first link at time one, node S sends signals to the node R and the node D directly; at the second link at time two, the node R transmits the signal to the node D through DF transmission [90]. The node D cooperates and decodes the signals received from the node S and node R to obtain diversity gain. In the system model, the channel coefficients of the three links are $h_{S,D}$, $h_{S,R}$, and $H_{r,d}$, respectively. The channel gains of the three links are expressed as:

$$H_{S,D} = |h_{S,D}|^2, H_{S,R} = |h_{S,R}|^2, H_{R,D} = |h_{R,D}|^2 \tag{4.33}$$

The noises of the three links are independent of each other and are Gaussian white noise with 0 mean and variance σ^2. The noise of the three links is expressed as:

$$\sigma_{S,D}^2 = \sigma_{S,R}^2 = \sigma_{R,D}^2 = \sigma^2 \tag{4.34}$$

2. Channel Model

   In the first transmission link, the signals $y_{S,R}$ and $Y_{s,d}$ received by the node R and node D are:

$$\begin{aligned} y_{S,D} &= \sqrt{p_S} h_{S,D} x_1 + Z_1 \\ y_{S,R} &= \sqrt{p_S} h_{S,R} x_1 + Z_2 \end{aligned} \tag{4.35}$$

   Among them, p_S is the transmission power from the node S, x_1 is the signal sent by the node S, Z_1 and Z_2 are Gaussian white noise with 0 mean and variance as σ^2.

   In the second transmission link, the signal $y_{R,D}$ received by the node D is expressed as:

$$y_{R,D} = \sqrt{p_R} h_{R,D} x_2 + Z_1 + Z_2 \tag{4.36}$$

   Among them, p_R is the transmission power from node R, and x_2 is the signal after the node R processes the received signal.

3. Information Rate of Cooperative Communication System

   In the first time slot, the information transmission rates $R_{S,D}$ and $R_{R,D}$ obtained by the node R and node D are expressed as:

$$\begin{aligned} R_{S,D} &= \frac{1}{2} log_2(1 + \gamma_{S,D}) \\ R_{S,R} &= \frac{1}{2} log_2(1 + \gamma_{S,R}) \end{aligned} \tag{4.37}$$

Among them, $\frac{1}{2}$ in the equation is because the whole transmission process includes two transmission time slots. The S-D link has a SNR of $\gamma_{S,D} = \frac{p_S H_{S,D}}{\sigma^2}$, and the S-R link has a SNR of $\gamma_{S,R} = \frac{p_S H_{S,R}}{\sigma^2}$.

In the second time slot, the information transmission rate $R_{R,D}$ obtained by the destination node is expressed as:

$$R_{R,D} = \frac{1}{2} log_2(1 + \gamma_{R,D}) \tag{4.38}$$

Where the SNR of R-D link is $\gamma_{R,D} = \frac{p_R H_{R,D}}{\sigma^2}$.

4.2.3.2 Power Allocation

Power distribution is to calculate the optimal power allocation factor with the goal of maximizing the QoS performance at the destination node under the condition of constraining the total power of the node. In the problem of power allocation, this dissertation takes the information transmission rate as the optimization goal to reduce power consumption when the system performance is optimal.

1. Define Variables

   The node power allocation vector is expressed as:

   $$P = [p_S, p_R] \tag{4.39}$$

   Compare the information transmission rate of the relay node and the destination node, and express the minimum value $f(p_S, p_R)$ as

   $$f(p_S, p_R) = \min(R_{S,D} + R_{R,D}, R_{S,D}) \tag{4.40}$$

2. Target Optimization

   In the system model, the proposed power allocation scheme should not only ensure

the maximum information transmission rate of the destination node, but also minimize the power consumption of the entire transmission process and reduce the power consumption overhead. The optimization objective function is represented by the following equation:

$$\max_{p_S, p_R} f(p_S, p_R) \tag{4.41}$$

The constraints are expressed as:

$$p_S + p_R \leq p_{tot} \tag{4.42}$$

$$log^2(1 + p_S H_{S,D}) + log^2(1 + p_R H_{R,D}) \geq log^2(1 + p_S H_{S,R}) \tag{4.43}$$

$$p_S \geq 0, p_R \geq 0 \tag{4.44}$$

Among them, $\sigma^2 = 1$, equation (4.41) is the optimization objective function, equation (4.42) is the constraint condition of the transmission power from node S and node R, equation (4.43) is the maximum value reached by the destination node rate, equation (4.44) is the transmission power values of node S and node R are greater than 0.

On the foundation of above three constraints, the optimization state is described as, under the premise of ensuring that the power of the node is positive and constraining the total transmit power of the node, the goal is to maximize the information transmission rate of the destination node, so as to obtain the optimal relationship between the node S and node R.

4.2.4 The Power Allocation Algorithm for Cooperative Communication

In order to decline the squander of resources and improve the system performance, a power allocation scheme is proposed in the single-relay cooperative communication model. The power allocation scheme is to establish a convex optimization function of the maximum information transmission rate of the destination node under the constraint of ensuring the total transmit power, and use the Lagrangian function to optimize the transmission power of node S and node R.

The steps of the algorithm are as follows:

Step1: The Lagrange equations for optimizing objective equation (4.40) to (4.42) are expressed as:

$$\begin{aligned} L(p_S, p_R) =& f(p_S, p_R) + \lambda[log_2(1 + p_S H_{S,D}) + log_2((1 + p_R H_{R,D}) \\ & - log_2(1 + p_S H_{S,R})] + \mu(p_S + p_R - p_{tot}) \end{aligned} \tag{4.45}$$

where $\lambda \geq 0$ is the Lagrange multiplier under power constraints.

Step2: According to the Karush–Kuhn–Tucker condition, set the partial derivatives of $L(p_S, p_R)$ to p_S and p_R to 0 respectively. The optimal solutions p_S^* and p_R^* should satisfy the following equations:

$$\frac{\partial L}{\partial p_S^*} = \frac{1}{2ln2}\left[\frac{H_{S,D} + 2\lambda H_{S,D}}{(1 + p_S H_{S,D})} - \frac{2\lambda H_{S,R}}{(1 + p_S H_{S,R}})\right] + \mu \tag{4.46}$$

$$\frac{\partial L}{\partial p_R^*} = \frac{H_{R,D}(1 + 2\lambda)}{2ln2(1 + p_R H_{R,D})} + \mu \tag{4.47}$$

Step3: In the optimization equation (4.41), under the condition that the total power of the node is constrained, when the information transmission rate of the destination node

reaches the maximum, the optimal power p_S^* and p_R^* of the source node and the relay node are respectively:

$$p_S^* = \min\left(\frac{1+2\lambda}{2\mu ln2} - \frac{1}{H_{R,D}, p_{tot}}\right) \tag{4.48}$$

$$p_R^* = \max\left(\frac{H_{R,D}(1+2\lambda)}{2ln2(1+p_R H_{r,d})} + \mu\right)^+ \tag{4.49}$$

Where, $(\cdot)^+ = \max(\cdot, 0)$

Equations (4.46) to (4.49) are the optimal solutions of the optimization problem (4.41), and the proof process is as follows:

In order to reduce the computational complexity of the optimization objective equation (4.41), I adopts the Lagrange dual method, which is expressed by the following formula:

$$g(\lambda, \mu) = \max_{p_S, p_R} L(p_S, p_R) \tag{4.50}$$

The dual function of equation (4.44) is expressed as:

$$\min_{\lambda,\mu} g(\lambda, \mu) \tag{4.51}$$

The convex dual function of $L(p_S, p_R)$ obtained from equation (4.50) is $g(\lambda, \mu)$. Then, the convex dual function of equation (4.51) is solved by the sub-gradient method, and its formula is:

$$\Delta\lambda = log_2(1 + p_S H_{S,D} + log_2(1 + p_R H_{R,D}) - log_2(1 + p_S H_{S,R}) \tag{4.52}$$

Therefore, the optimization objective equation (4.41) ensures that the information transmission rate of the destination node reaches the maximum value under the condition of constraining the total power value of the node. The optimal power distribution between the source node and the relay node obtained by the equation (4.48) and equation (4.49) is the optimal solution of the optimization equation (4.41), and the proof is completed.

$$\Delta\mu = p_S + p_R - p_{tot} \tag{4.53}$$

4.3 Simulations and Performance Analysis

4.3.1 5G Channel Forecasting Based on LSTM Network

In this section, the simulation of Massive MIMO channel forecasting based on LSTM network is derived. The parameters of simulated channels are the same as in Table 2.1, except for the number of TX and RX antenna is 64. We selected 10000 channels data, and performed average value of 10000 times forecasting. Each data set contains more than six thousand observations. During the forecasting, first 90 % of observations is used to train the model, and last 10 % data is used to test. Through experiments, we selected the optimal LSTM network parameters, used a two-layer neural network with 330 hidden units, and performed 2000 iterations of training, specifying an initial learning rate of 0.005, and multiplying it by a factor of 0.2 after 1000 iterations of training to reduce the learning rate. At the same time, to prevent the gradient exploding, the gradient threshold is set to 1. Before plug data into training, we did standardization for all data, and de-standardization for comparing with the actual data after forecasting. Fig. 4.10 presents the RMSE and loss of training model.

From the figure we can observe that the loss function of model is closing to 0 with the iteration increasing, so we can consider the model to have converged.

Then, we forecasted values for multiple time steps in the future, one time step each time slot, and updated the network with each forecasting. After each step of forecasting, use the previous forecasted value as the input to the function. Again, the testing data set was standardized by using training model's data parameters. Afterwards, the RMSE was calculated based on the standardized data. Fig. 4.11 presents the concatenation of actual

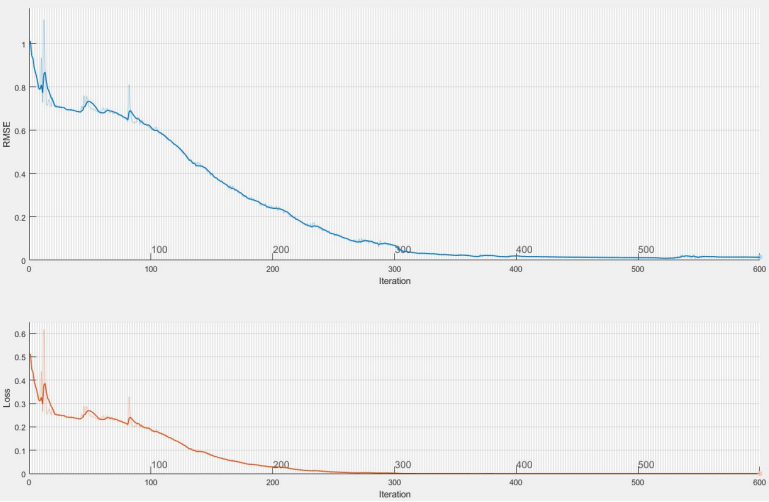

Figure 4.10. RMSE and loss of training model.

channel and forecasted channel. The comparison of forecasting accuracy and RMSE is performed in Fig. 4.12. So far, the RMSE is about 0.16.

However, because the CSI has been obtained, the actual value of the time step between each forecasting can be obtained, so if the actual value is used to replace the forecasted value to update the network state, the following forecating of the test data set will be more accurate. Likewise, we make forecasting for each time step. The difference is that for each forecasting, the actual values at the previous time step are used to forecast the next time step, not the forecasted values. By this optimization, RMSE decresed to about 0.05, a better forecasting performance can be obtained. The updated comparison and RMSE is revealed in Fig. 4.13.

We can conclude that our optimized model parameters for LSTM network can achieve a relatively high forecasting accuracy, and forecastings are more accurate when the network state is updated with observations rather than predictions.

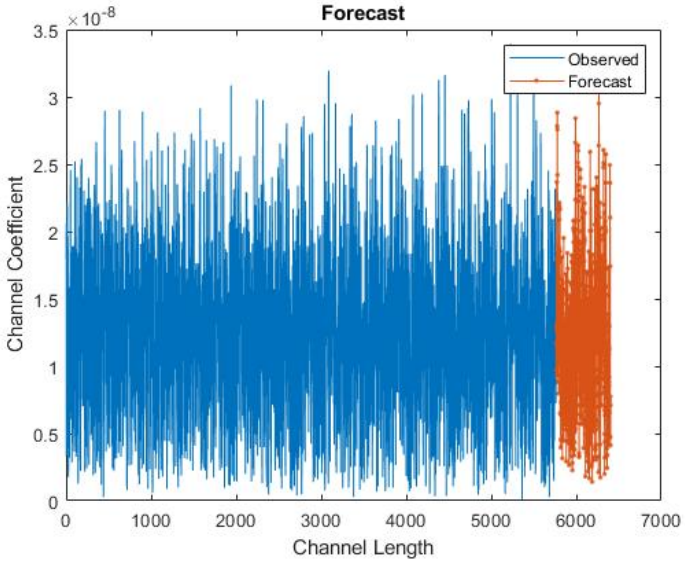

Figure 4.11. Concatenation of actual channel and forecasted channel.

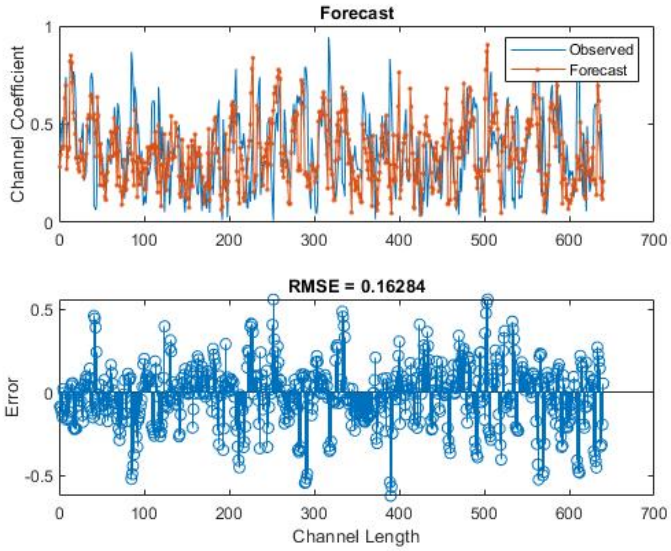

Figure 4.12. Comparison and RMSE of forecasting.

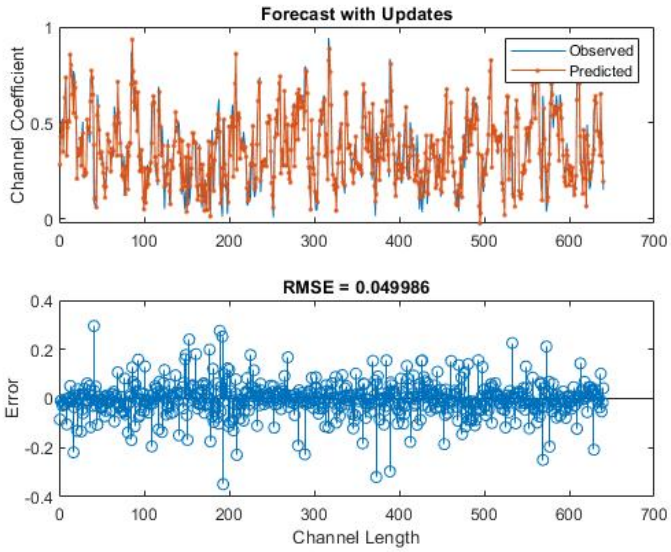

Figure 4.13. Comparison and RMSE of forecasting by using actual observation.

4.3.2 Power Allocation Based on Cooperative Communication

In this dissertation, in the half-duplex single-relay cooperative system, a node power allocation scheme is proposed: when the information transmission rate of the destination node reaches the maximum value, the power of the source node and the relay node is obtained. Under the constraints of the total power of the nodes, after solving the objective function of maximizing the information rate of the destination node, the power values of the source node and the relay node are obtained, that is, the optimal power value. We present the Monte Carlo simulation results in this chapter, and evaluate the impact of the proposed scheme on the system performance by comparing the equal power allocation scheme, the non-cooperative scheme [85] and the power allocation scheme proposed in this dissertation. Among them, in the non-cooperative scheme, there is only one link between node S and node D, and the gain cannot be obtained through the cooperative relay node. The equal power scheme is to set the transmission power of the source node and the relay node to be equal and remain unchanged, which can be regarded as a special case of this scheme.

Table 4.2. Simulation Parameters

Parameter	Value
Minimum distance of S-D link	20m
Variance of noise (σ^2)	1
Maximum transmitted power of source node	20dBm
Maximum transmitted power of relay node	20dBm
Transmission time	1s

In this dissertation, the MATLAB simulation is used to test the performance, and the three links do not affect each other, and the Rayleigh fading link with the mean value of 1 is set. The lower limit of the link distance between the source node and the destination node is set as 20 meters, the noise variance is sigma, the maximum power of the source node and the relay node is both 20dBm, and the transmission time is 1 second. Table 4.2 shows the simulation parameters.

4.3.2.1 Information Transmission Rate of Destination Node vs Power of Source Node

The initial value of the power of the source node is set to 10dBm, which is linearly increased to 20dBm according to the step size of about 1dBm. The influence of the transmit power of the source node on the information rate of the destination node under different power schemes is analyzed as shown in Fig. 4.14.

Fig. 4.14 describes the change of the information rate of the destination node when the power of the source node changes. Comparing several power allocation schemes under the same conditions, it can be concluded that when the transmission power of the source node is low, the information transmission rate of the destination node can be well improved, which shows that the information transmission rate of the destination node increases with the increase as the transmit power increases. Secondly, the simulation results show that comparing the non-cooperative scheme and the equally power distribution algorithm with

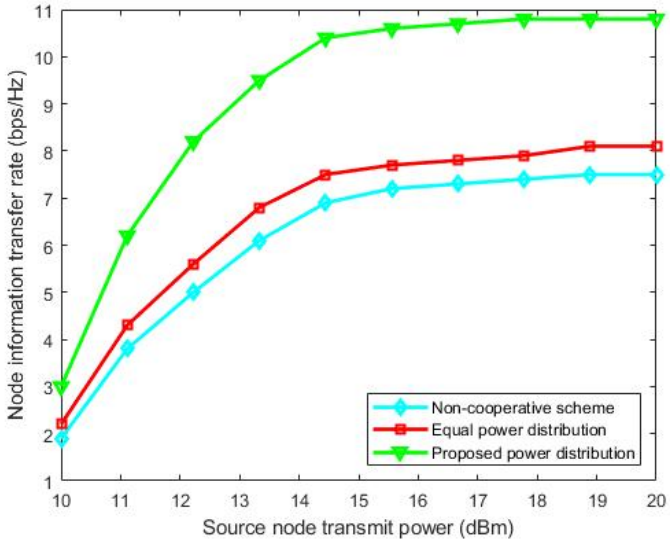

Figure 4.14. Information transmission rate of destination node vs power of source node.

the power distribution algorithm performed in this dissertation, the power distribution algorithm achieved in this dissertation has higher information rate changes of destination nodes, the equal power allocation scheme is second, and the non-cooperative scheme is the worst . The scheme in this dissertation can maximize the information rate of the destination node under the premise of constraining the total transmit power of the node, and effectively reduce the transmit power of the source node. Compared with the other two schemes, this scheme can improve the information transmission rate of the destination node, especially when the transmit power is small.

4.3.2.2 Information Transmission Rate of Destination Node vs Power of Relay Node

The initial value of the relay node power is set to 10dBm, and it is linearly increased to 20dBm in steps of about 1dBm. Compare the power allocation scheme proposed in this dissertation with the equal power distribution, and study the influence of the relay node power on the information rate of the destination node as shown in the Fig. 4.15 shown.

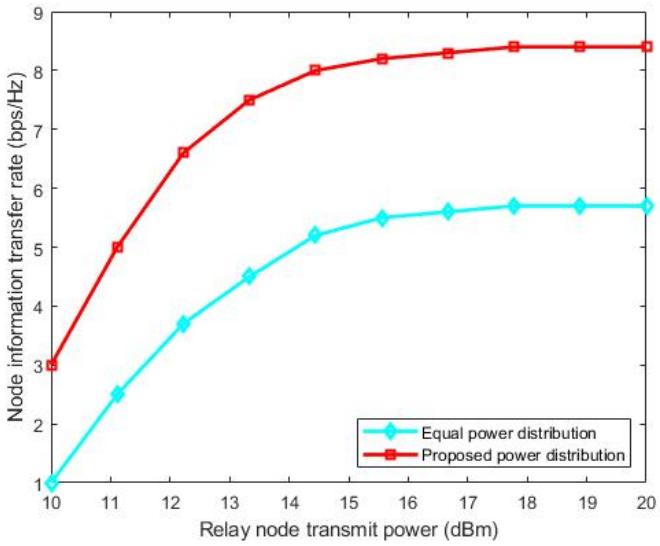

Figure 4.15. Information transmission rate of destination node vs power of relay node.

Fig. 4.15 describes the change of the information rate of the destination node when the power of the relay node changes. First of all, the figure shows that in the process of linear increase of relay node transmit power, the change of destination node information rate shows an upward trend, this is because with the increase of relay node transmit power, the power used for information transmission becomes larger, Therefore, the information transmission rate at the destination node is improved. Secondly, compared with the equal power scheme, the proposed scheme obtains better information rate of the destination node, because cooperative diversity helps to improve the information transmission rate at the destination node. In addition, Fig. 4.15 presents that when the relay node power is low, both power allocation schemes improve the information transmission rate of the destination node. However, the scheme proposed in this dissertation has a higher information transmission rate of the destination node when the transmit power of the relay node changes, reflecting the constraints on the total power of the node. The power allocation scheme proposed in this dissertation can maximize the information rate of the destination node. The

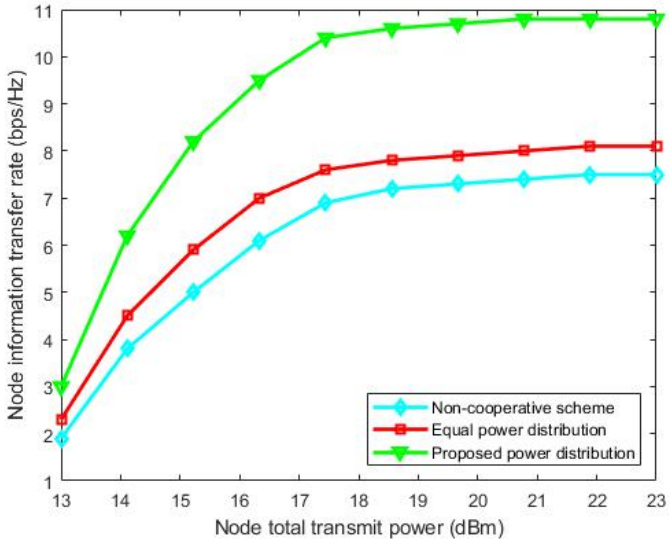

Figure 4.16. Information transmission rate of destination node vs total power of nodes.

scheme not only improves the information transmission rate of the destination node, but also reduces the power consumption of the relay node.

4.3.2.3 Information Transmission Rate of Destination Node vs Total Power of Nodes

The initial value of the total node power is set to 13dBm, which is linearly increased to 23dBm in steps of about 1dBm. The influence of the total node transmit power on the information rate of the destination node in the three schemes is analyzed as shown in Fig. 4.16.

Fig. 4.16 describes the change of the information rate of the corresponding destination node when the total power of the node changes. As can be seen from the figure, the scheme proposed in this dissertation has the following two phenomena compared with the other two: first, when the total power of the node is very small, the information transmission rate of the destination node increases rapidly; When the power is large, the interference limitation exists, so that the information transmission rate of the destination node tends to

a stable value. Secondly, the change of destination node information rate corresponding to the proposed scheme is better than the other two schemes, especially when the total node power is small, the performance of this scheme is better.

It can be seen in Fig. 4.16 that under the same conditions, when the total transmit power of the node is low, the information transmission rate of the destination node can be better improved, which shows that the information transmission rate of the destination node increases with the increase of the total transmit power of the node. The simulation results show that the proposed scheme can maximize the information rate of the destination node under the constraint of the total node power. Compared with the other three schemes, this scheme not only improves the information transmission rate of the destination node, but also reduces the node transmit power.

4.3.2.4 SNR vs Power of Relay Node

The initial value of the relay node power is set to 20dBm, and it is linearly increased to 20dBm in about 1dBm steps. The influence of the relay node transmit power corresponding to the two schemes on the SNR is shown in Figure 4.17.

Fig. 4.17 describes the change of the corresponding SNR when the power of the relay node changes. The following phenomenon exists: when the power of the relay node increases, the SNR increases with the increase of the power of the relay node. Under the same conditions, when the transmit power of the relay node is increased, the signal-to-noise ratio can be better improved, which shows that the signal-to-noise ratio increases with the increase of the transmit power of the relay node. Therefore, the scheme proposed in this dissertation can better improve the system performance.

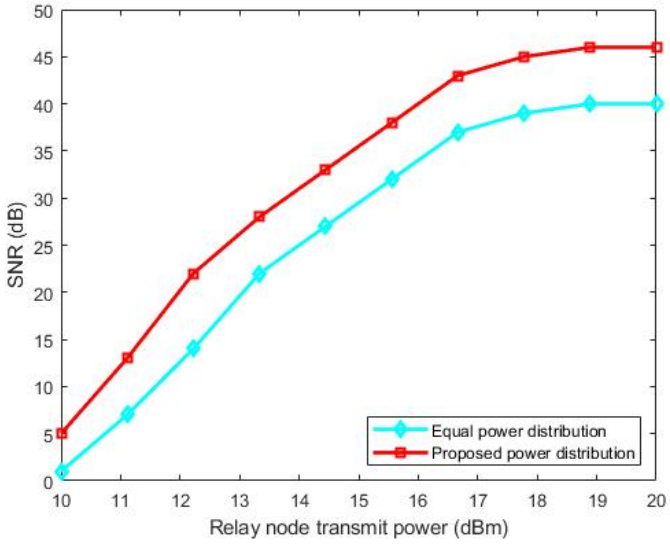

Figure 4.17. SNR vs power of relay node.

4.3.2.5 Information Transmission Rate of Destination Node vs Distance of Relay Node

The initial value of the relay node position is set to 20m, and linearly increases to 110m in steps of 10m. Comparing the power allocation scheme proposed in this dissertation with the equal power allocation scheme, the difference between the information transmission rate of the destination node and the position of the relay node The relationship is shown in Figure 4.18.

Figure 4.18 depicts the change in the information rate of the destination node when the position of the relay node changes. As can be seen from the figure, compared with the equal power allocation scheme, the proposed scheme in this dissertation has the following two phenomena: first, when the relay node is located closer, the information rate of the destination node decreases rapidly with its increase; When the position of the successor node is very large, the existence of interference limitation makes the information transmission rate of the destination node tend to a stable value. The whole curve shows a downward trend, indicating that the information rate of the destination node decreases with the increase of

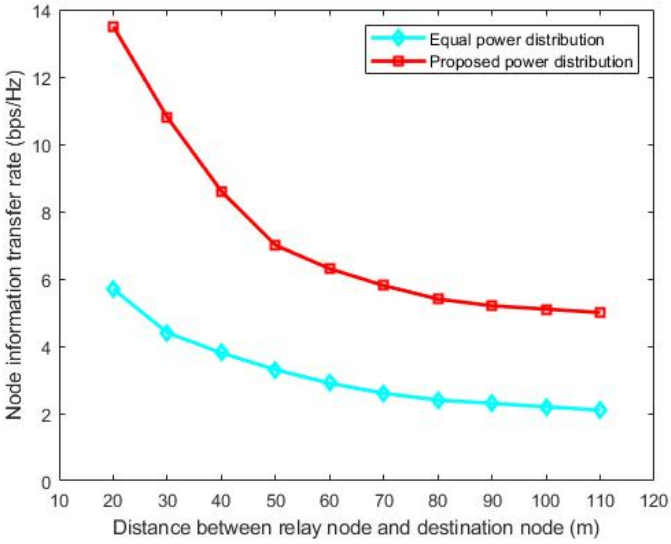

Figure 4.18. Information transmission rate of destination node vs distance of relay node.

the distance of the relay node. System performance can be improved by reducing the distance between relay nodes. Secondly, the proposed scheme is better than the equal power allocation scheme in response to the change of the information rate of the destination node.

CHAPTER 5

CONCLUSIONS AND FUTURE RESEARCH

5.1 Conclusions

In the first part of this dissertation, we have proposed the channel forecasting in 5G Wireless Communication Based on Granger Causality and Transfer Entropy, and the power allocation based on Inverse Water-filling algorithm. Our work consists of four main parts: Firstly, we performed the Granger causality test on two independent, random and real 5G channels of the simulation, ensuring the correlation between the two channels [14]. Secondly, we used the transfer entropy method to forecast the above two 5G channels and obtain two forecasted 5G channels. Third, we calculated the RMSE of the original channel and the forecasted channel to ensure the high accuracy of the forecasting [14]. At the same time, we computed the CRLB of the forecasted 5G channel and showed that the variance of the forecasted parameters is close to the CRLB. Finally, for the two forecasted 5G channels, we performed power allocation and comparison based on the Equal Gain algorithm and the Inverse Water-filling algorithm. Simulations in Chapter 2 validate these theoretical results [14].

In the second part, the corresponding channel model is obtained based on 28GHz indoor and through-wall wireless channel measurement. The main work content can be divided into the following chapters.

Chapter 3.1 introduces the background and current status of 5G Massive MIMO today, and introduces the basic knowledge required for wireless channel modeling. Chapter 3.2 illustrates the concept of MIMO channel characteristics and modeling methods, which includes the MIMO channel characteristics and some popular statistical and analyt-

ical channel models. The measurement and collection of experimental data scenarios of this dissertation is discussed in chapter 3.3. In this dissertation, we used 8 x 8 URA massive MIMO antennas at TX side and RX side, the TX signal is with central frequency of 28GHz and 800MHz bandwidth. The two scenarios are indoor OLS and through-wall respectively. Chapter 3.4 is the statistical channel modeling and parameter analysis based on received measurement signal. The GSCM method is used. Simulations and performance are proposed in chapter 3.5.

For the last part, the 5G channel is forecasted by using LSTM network. Based on that, cooperative communication is proposed to improve the performance of wireless communication. Chapter 4.1 introduces the background and methodologies of RNN and LSTM networks demonstrates their possibility for 5G channel forecasting. A proposed power allocation algorithm based on cooperative communication is illustrated in chapter 4.2. Chapter 4.3 is the simulation and performance analysis of above two sections. In the chapter, the optimal LSTM model parameters are obtained to achieve the RMSE about 0.05 of testing data set. In addition, the comparisons of proposed cooperative communication power allocation scheme with non-cooperative system and equal power allocation scheme discussed.

5.2 Future Research

5.2.1 5G Ultra-dense Networking and Resource Allocation

In higher frequency bands such as millimeter waves, there are more abundant spectrum resources. But as the frequency increases, the signal penetration will be worse, resulting in more severe attenuation. In future work, we should focus on wireless communication (millimeter wave) in higher frequency bands and obtain better wireless communication quality through more optimized channel information extraction methods and modeling methods. Also, the scenario measured in this dissertation are indoor scenes and indoor

through-wall scenes. The main propagation environment is only indoor objects such as desks and computers. It does not consider factors such as moving people and receiving users moving. In practice, users are basically moving, and the blockage of the human body also affects high frequency communications. Moreover, outdoor scenes also have some important influencing factors for high-frequency signal propagation such as foliage, rain attenuation, etc. Therefore, in the future, 5G ultra-dense networking and resource allocation will be a meaningful research direction.

5.2.1.1 Ultra-dense Networking Architecture

In the development process of the wireless communication architecture, the core network and the access network are developing in the direction of simplicity, flatness, and intelligence. 5G ultra-dense networking simplifies the association between the core network and the access network by separating the control plane and the data plane, making network service functions more intelligent and 5G network functions more open. There are two main types of ultra-dense networking architectures: macro-micro and micro-micro base station deployment scenarios [91].

1. Macro-micro base station deployment scenarios

   The 5G ultra-dense networking realizes the separation of the user plane and the control plane through the way that the micro base station is responsible for the capacity and the macro base station is responsible for the coverage. According to the characteristics of flexible deployment of micro base stations, the collaborative management of resources between micro base stations is adopted to meet the needs of the development of access network services. At the same time, as a centralized control module in the access network, the macro base station can coordinate interference between micro base stations and allocate system resources. The macro base station is mainly responsible for maintaining the control plane of the terminal equipment

to carry the wireless resource control connection service. The terminal equipment selects the specific bearer base station according to the service requirements. The macro base station carries the delay-sensitive, low bandwidth demand voice, remote driving and other services, and the micro base station is responsible Video transmission and download services that are not sensitive to time delay and require large bandwidth. Generally speaking, the macro base station is responsible for the control plane, and the micro base station is mainly responsible for improving the system capacity.

The location of the micro base station is fixed, always within the coverage of the macro base station, and only provides a user plane connection for the terminal. The macro base station is always connected with the micro base stations within range to control wireless resources. At this time, it can simplify the operation of the user equipment to switch at the micro base station, reducing the pressure on the core network. The macro base station can maintain radio resource control services and some low-rate services, and improve the user experience of terminal equipment during frequent handovers. The macro base station can also use the access resource optimization algorithm to optimize the terminal user's micro base station selection strategy, inter-micro base station interference and load balancing and other issues, reasonably allocate micro base station channels and computing resources, and improve terminal user experience.

2. Macro-micro base station deployment scenarios

   The micro-micro base station deployment method of 5G ultra-dense networking is to use dynamic clustering to divide 5G micro base stations into multiple virtual macro cells, and share the signals, channels, computing power and other resources of multiple base stations in the same cell. In this way, network functions virtualization (NFV) macro base stations are constructed, and the micro base stations in the same cell im-

plement functions similar to the control plane transmission of the macro base station on these resources. The NFV macro base station carries the functions of the control plane and is responsible for coordinating and managing the channel and computing power resources in the cell, while the micro base station has the same function and is responsible for transmitting user data. The ultra-dense networking of micro-micro base station architecture can flexibly construct NFV macro base stations according to business changes, optimize the system's interference management between micro base stations, wireless resource allocation and other issues, and improve network capacity while adapting to task requirements for end users. Provide a better business experience.

5.2.1.2 The Key Technology of Ultra-dense Networking

Ultra-dense networking reduces the path loss between base stations and user terminals through dense deployment of micro base stations, and increases network capacity. At the same time, the dense distribution of base stations also increases the interference to the system. While improving network capacity, effective interference management and resource allocation have become the research focus of ultra-dense networking. The key technologies to increase system capacity and enhance system performance in ultra-dense networking include non-orthogonal multiple access (NOMA) technology and unmanned aerial vehicle (UAV) technology.

1. NOMA technology in ultra-dense networking

   From 1G to 4G, wireless communication systems have adopted Frequency Division Multiple Access (FDMA), Time Division Multiple Access (TDMA), Code Division Multiple Access (CDMA) and Orthogonal Frequency Division Multiple Access (OFDMA) are key technologies. These multiple access schemes all belong to Orthogonal Multiple Access (OMA) technology, and wireless resources in the time

domain, frequency domain, and code domain are orthogonally allocated to multiple users according to the user's combination. In a communication system using the OMA technology, a relatively simple and low-cost receiver is used [92]. When the complexity of the signal carried by the user is low, it can be easily separated at the receiving end. The number of users that OMA technology can serve is also limited by the number of orthogonal resources available in the system. In the process of channel transmission, the orthogonality of the signal is also affected by time or frequency.

As an alternative to OMA, NOMA technology is mainly characterized by supporting more users to access the communication system with the help of non-orthogonal resource allocation. In addition, complex inter-user interference cancellation technology is used to solve the problem of exponential increase in receiver complexity. According to the difference of multiplexing domain, NOMA technology can be divided into two types: power domain NOMA and coding domain NOMA. In the power domain, power is allocated to users according to the channel quality of different users, and multiple users share the same time-frequency resources at the same time. On the receiver side, the power domain NOMA uses the difference of user power and uses SIC technology to process user signals and eliminate interference [93]. The coding domain NOMA is similar to CDMA or multi-carrier CDMA, and it is more inclined to use non-orthogonal sequences with low density sequences or low correlation, and the implementation is more complicated.

2. UAV technology in ultra-dense networking

   UAVs have the characteristics of multiple functions and strong maneuverability. According to different application methods and purposes, low-altitude UAVs are widely used in the public domain. In the process of wireless communication, drones can assist the ground communication network, by installing communication equipment as a flight base station or flight buffer, to provide ground users with enhanced or

emergency wireless communication when the traffic demand is large and the system capacity is insufficient. business. In addition, UAVs can also be used as air nodes for various applications such as task scheduling, offloading, and energy transmission. Most of the existing work is mainly about drones assisting ground network communications, realizing millimeter wave communications, beamforming, emergency communications, and wireless charging [94].

As an alternative or supplementary technology to the ground 5G communication network, UAVs are flying in the air, with a high probability of signal transmission through the line-of-sight link, which is conducive to improving the reliability of long-distance signal transmission. In addition, the flying height of the UAV can be adjusted to change the coverage of the UAV and the communication quality of ground users. Compared with fixed ground communication infrastructure equipment, UAVs can dynamically plan flight trajectories and adapt to real-time business needs. In addition, as an aerial base station, UAVs have no site lease costs, and require less ground infrastructure and power supply from the power grid, and are easy to manage and maintain. Utilizing a large number of UAVs can form an ultra-dense UAV network to provide targeted instant messaging services to ground users. The ultra-dense UAV network can be flexibly and quickly deployed to areas with insufficient network capacity or interrupted communication services according to business needs, so as to achieve rapid and effective auxiliary communication and emergency communication [95].

5.2.2 Millimeter Wave Communication Based on NOMA Technology

In recent years, the development of new spectrum resources to meet the ever-increasing speed requirements of the 5G era has become a hot research direction. Millimeter wave is considered to be a promising technology in the fifth-generation mobile communication net-

work due to its large bandwidth in the high-frequency region [10]. The severe propagation path loss and low penetration capability of millimeter wave signals require the help of new multiple access technologies, especially when it is necessary to support large-scale connections in dense networks [11]. NOMA is regarded as an important technology to realize the ultra-high capacity and large-scale access requirements of 5G networks [11]. Therefore, one of the best ways to further improve the spectrum efficiency of the upcoming 5G era is to introduce NOMA into millimeter wave communications [11] [96] [97] [98] [99].

Faced with the explosive development of future network requirements, the 5G communication process must focus on mobile communication methods that can further explore the ability to improve frequency efficiency. In the next decade of wireless communication, NOMA can be regarded as an important technology to meet the ultra-high capacity and large connection requirements of 5G systems [13] [100] [101]. Driven by the unprecedented growth of new network-based smart devices and innovative applications, emerging complex service requirements have accelerated the development of new multiple-access technologies for 5G networks. NOMA technology can be mainly divided into two categories, namely code domain NOMA and power domain NOMA [102]. Compared with OMA, NOMA can further enhance user links and support more system functions.

The key idea of NOMA is to use the power domain for multiple access, while previous generations of mobile networks have always relied on the time/frequency/code domain. Take the traditional OFDMA used by 3GPP-LTE as an example. The main problem of this orthogonal multiple access technology is that when certain bandwidth resources (such as sub-carrier channels) are allocated to users with poor channel conditions, its spectrum efficiency is very low. In addition, by using NOMA, each user can access all sub-carrier channels. Therefore, users with poor channel conditions can still access the bandwidth resources allocated to users with poor channel conditions, thereby greatly improving channel quality and spectrum efficiency. In addition, compared with regular opportunistic user

scheduling that only serves users under strong channel conditions, NOMA has achieved a good balance between system throughput and user fairness. In other words, NOMA can provide services to users with different channel conditions in time, which provides the possibility to meet the 5G requirements of ultra-low latency and ultra-high connectivity [103]. With the continuous research of 5G and the continuous enhancement of chip processing technology, the processing capability of the receiver has been greatly improved. Therefore, inevitably, NOMA has become a popular area for industry and academia. The main research in academia is the power domain NOMA [103] [104].

NOMA proposes a new power domain dimension, which uses the principle of superposition coding to send information. So that the same frequency resources, such as frequency, time, and code domain, can be shared by users with different channel conditions. The main principle of power multiplexing is to allocate different powers to different users according to the differences in the channel conditions of different users. In order to ensure the equality of users, more power should be allocated to users with weak channel conditions. Therefore, while ensuring the quality of service for weak users, the overall quality of the system can be improved. At the receiving port, the serial interference cancellation (SIC) method can be used to eliminate the interference between each user [105] [106]. The main concept of SIC is to gradually eliminate interference between different users according to different user channel conditions. For example, the user with the strongest signal can be detected first. Then, the strongest user encodes and modulates its signal again, and then removes its signal from the composite signal. The second strongest signal follows the same process, in fact it has become the strongest signal. When only one signal is detected, the weakest user decodes its information without any interference.

By performing high SNR analysis, it is easy to illustrate the performance gain of NOMA compared to conventional OMA. Under the implementation of OMA, the received data rates of the two users are $1/2log_2(1+\rho|h_A|)^2$ and $1/2log_2(1+\rho|h_B|)^2$ respectively.

The 1/2 here is because the two users divide the bandwidth resources, ρ represents the transmission SNR, and h_A and h_B are user A and user B's channel gain, respectively. We assume that $|h_A|^2 < |h_B|^2$, under high SNR, ρ tends to ∞, the total rate of OMA can be calculated as $1/2log_2(1+\rho|h_A|)^2+1/2log_2(1+\rho|h_B|)^2$. In the case of NOMA, the received data rates of the two users are $log_2(1+\frac{\rho\alpha_A|h_A|^2}{1+\rho\alpha_B|h_A|^2}$ and $log_2(1+\rho\alpha_B|h_B|^2)$ respectively, α_A and α_B are power allocation factors. Therefore, the high SNR estimate for the total NOMA rate is $log_2(\rho\alpha_B|h_B|^2)$. This is much greater than in the case of OMA, especially when the channel gain of user B is much greater than that of user A. In other words, the main reason for the performance improvement of the NOMA technology after use is that the factor 1/2 outside the logarithm of the OMA rate caused by the allocation of bandwidth resources between users has a greater impact than the internal factor. The logarithm of the NOMA rate, used for power allocation.

Downlink and uplink NOMA can be regarded as a special case of multiple access channels (MACs) and broadcast channels (BCs). Therefore, the NOMA rate area is limited by the capacity area of the corresponding MACs and BCs. Compared with the existing information theory work focusing on maximizing system throughput, the main function of NOMA is to achieve a balanced trade-off between system throughput and user fairness. Take the two-user downlink situation as an example again. If the system throughput is the only goal, all power will be allocated to users with stronger channel conditions, which will result in the maximum throughput, but will not be able to serve user A at all. The function of NOMA is to generate greater throughput than OMA, and to ensure fair service to users. This feature is especially important for 5G, because it is foreseeable that 5G will support the functions of the Internet of Things (IoT) to connect trillions of devices. With OMA, connecting thousands of IoT devices (such as vehicles in an in-vehicle ad hoc network for smart transportation) requires thousands of bandwidth channels; however, NOMA can use these devices in a single channel. An important phenomenon in NOMA networks is that

some users with poor channel conditions will experience low data rates. The reason is that these users cannot completely delete the messages of their partners from their observations, which means that they will be subject to strong co-channel interference, so their data rate will be small. In the context of the Internet of Things, because many Internet of Things devices require only a small data rate to provide services, this problem can be avoided. Question [107].

NOMA has following advantages:

1. High bandwidth efficiency: Since NOMA can enable various users to use each resource block (for example, frequency/time) [104], NOMA has high bandwidth efficiency, thereby improving the throughput of the system.
2. Fairness: One of the main functions of NOMA is to provide more power to weak users. By doing so, NOMA can ensure the fairness of users in terms of performance [108].
3. Ultra-high connectivity: It is expected that billions of devices will be connected to 5G systems in the 5G Internet of Things [109]. The existence of NOMA effectively solves this difficult task by using its non-orthogonal features, providing a promising design choice. In contrast to traditional OMA (which requires the same time/frequency resource blocks as the number of devices), NOMA can provide users with a way to reduce the demand for resource blocks.
4. Compatibility: From a theoretical point of view, for any current OMA technology (such as CDMA / FDMA / TDMA / OFDMA), NOMA can be combined again as an ”add-on” method because it utilizes the power domain. Due to the gradual maturity of the SIC method in theory and reality, NOMA can be used in combination with the current MA method.

Therefore, channel modeling for combining millimeter wave and NOMA technology, the user grouping and power allocation scheme under the millimeter wave NOMA and Massive MIMO - NOMA system will be one of the focus of the next work.

REFERENCES

[1] I. ITU-R, "Vision–framework and overall objectives of the future development of imt for 2020 and beyond," *Recommendation ITU*, pp. 2083–0, 2015.

[2] C. E. Shannon, "A mathematical theory of communication," *ACM SIGMOBILE mobile computing and communications review*, vol. 5, no. 1, pp. 3–55, 2001.

[3] C. W. Granger, "Testing for causality: a personal viewpoint," *Journal of Economic Dynamics and control*, vol. 2, pp. 329–352, 1980.

[4] L. Barnett, A. B. Barrett, and A. K. Seth, "Granger causality and transfer entropy are equivalent for gaussian variables," *Physical review letters*, vol. 103, no. 23, p. 238701, 2009.

[5] T. Schreiber, "Measuring information transfer," *Physical review letters*, vol. 85, no. 2, p. 461, 2000.

[6] N. A. Ahmed and D. Gokhale, "Entropy expressions and their estimators for multivariate distributions," *IEEE Transactions on Information Theory*, vol. 35, no. 3, pp. 688–692, 1989.

[7] H. Qiu, S. Xu, F. Han, H. Liu, and B. Caffo, "Robust estimation of transition matrices in high dimensional heavy-tailed vector autoregressive processes," in *Proceedings of the... International Conference on Machine Learning. International Conference on Machine Learning*, vol. 37. NIH Public Access, 2015, p. 1843.

[8] M. K. Samimi and T. S. Rappaport, "3-d statistical channel model for millimeter-wave outdoor mobile broadband communications," in *2015 IEEE International Conference on Communications (ICC)*. IEEE, 2015, pp. 2430–2436.

[9] T. Bai and R. W. Heath, "Coverage and rate analysis for millimeter-wave cellular networks," *IEEE Transactions on Wireless Communications*, vol. 14, no. 2, pp. 1100–1114, 2014.

[10] Z. Pi and F. Khan, "An introduction to millimeter-wave mobile broadband systems," *IEEE communications magazine*, vol. 49, no. 6, pp. 101–107, 2011.

[11] Z. Ding, P. Fan, and H. V. Poor, "Random beamforming in millimeter-wave noma networks," *IEEE access*, vol. 5, pp. 7667–7681, 2017.

[12] F. Al-Ogaili and R. M. Shubair, "Millimeter-wave mobile communications for 5g: Challenges and opportunities," in *2016 IEEE International Symposium on Antennas and Propagation (APSURSI)*. IEEE, 2016, pp. 1003–1004.

[13] C. Chen, W.-D. Zhong, H. Yang, and P. Du, "On the performance of mimo-noma-based visible light communication systems," *IEEE Photonics Technology Letters*, vol. 30, no. 4, pp. 307–310, 2017.

[14] Z. Chen and Q. Liang, "Power allocation in 5g wireless communication," *IEEE Access*, vol. 7, pp. 60 785–60 792, 2019.

[15] F. Han, H. Lu, and H. Liu, "A direct estimation of high dimensional stationary vector autoregressions," *The Journal of Machine Learning Research*, vol. 16, no. 1, pp. 3115–3150, 2015.

[16] A. Shojaie and G. Michailidis, "Discovering graphical granger causality using the truncating lasso penalty," *Bioinformatics*, vol. 26, no. 18, pp. i517–i523, 2010.

[17] K. Zhou, H. Zha, and L. Song, "Learning social infectivity in sparse low-rank networks using multi-dimensional hawkes processes," in *Artificial Intelligence and Statistics*, 2013, pp. 641–649.

[18] E. C. Hall, G. Raskutti, and R. Willett, "Inference of high-dimensional autoregressive generalized linear models," *arXiv preprint arXiv:1605.02693*, 2016.

[19] G. Aulogiaris and K. Zografos, "A maximum entropy characterization of symmetric kotz type and burr multivariate distributions," *Test*, vol. 13, no. 1, pp. 65–83, 2004.

[20] M. Bicego, D. Gonzalez-Jimenez, E. Grosso, and J. A. Castro, "Generalized gaussian distributions for sequential data classification," in *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*. IEEE, 2008, pp. 1–4.

[21] K. Zografos and S. Nadarajah, "Expressions for rényi and shannon entropies for multivariate distributions," *Statistics & Probability Letters*, vol. 71, no. 1, pp. 71–84, 2005.

[22] M. Skolnik, "Introduction to radar systems," *IEEE Aerospace and Electronic Systems Magazine*, vol. 16, no. 10, pp. 19–19, 2001.

[23] J. M. Mendel, *Lessons in estimation theory for signal processing, communications, and control*. Pearson Education, 1995.

[24] R. D. Palmer, "Fundamentals of radar signal processing," *Bulletin of the American Meteorological Society*, vol. 89, no. 7, p. 1037, 2008.

[25] M. Zhong, F. Lotte, M. Girolami, and A. Lécuyer, "Classifying eeg for brain computer interfaces using gaussian processes," *Pattern Recognition Letters*, vol. 29, no. 3, pp. 354–359, 2008.

[26] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive mimo for next generation wireless systems," *IEEE communications magazine*, vol. 52, no. 2, pp. 186–195, 2014.

[27] J. Hoydis, C. Hoek, T. Wild, and S. ten Brink, "Channel measurements for large antenna arrays," in *2012 International Symposium on Wireless Communication Systems (ISWCS)*. IEEE, 2012, pp. 811–815.

[28] S. Sangodoyin, V. Kristem, C. U. Bas, M. Käske, J. Lee, C. Schneider, G. Sommerkorn, J. Zhang, R. Thomä, and A. F. Molisch, "Cluster-based analysis of 3d

mimo channel measurement in an urban environment," in *MILCOM 2015-2015 IEEE Military Communications Conference*. IEEE, 2015, pp. 744–749.

[29] M. K. Samimi and T. S. Rappaport, "3-d millimeter-wave statistical channel model for 5g wireless system design," *IEEE Transactions on Microwave Theory and Techniques*, vol. 64, no. 7, pp. 2207–2225, 2016.

[30] C.-K. Wen, S. Jin, K.-K. Wong, J.-C. Chen, and P. Ting, "Channel estimation for massive mimo using gaussian-mixture bayesian learning," *IEEE Transactions on Wireless Communications*, vol. 14, no. 3, pp. 1356–1368, 2014.

[31] A. O. Martinez, E. De Carvalho, and J. Ø. Nielsen, "Massive mimo properties based on measured channels: Channel hardening, user decorrelation and channel sparsity," in *2016 50th Asilomar Conference on Signals, Systems and Computers*. IEEE, 2016, pp. 1804–1808.

[32] Y. Ghasempour, C. R. Da Silva, C. Cordeiro, and E. W. Knightly, "Ieee 802.11 ay: Next-generation 60 ghz communication for 100 gb/s wi-fi," *IEEE Communications Magazine*, vol. 55, no. 12, pp. 186–192, 2017.

[33] R. Kataoka, K. Nishimori, N. Tran, and T. Imai, "Basic performance of massive mimo in indoor scenario at 20-ghz band," in *2015 International Symposium on Antennas and Propagation (ISAP)*. IEEE, 2015, pp. 1–4.

[34] E. Bjornson, L. Van der Perre, S. Buzzi, and E. G. Larsson, "Massive mimo in sub-6 ghz and mmwave: Physical, practical, and use-case differences," *IEEE Wireless Communications*, vol. 26, no. 2, pp. 100–108, 2019.

[35] F. Miramirkhani and M. Uysal, "Channel modeling and characterization for visible light communications," *IEEE Photonics Journal*, vol. 7, no. 6, pp. 1–16, 2015.

[36] M. Series, "Guidelines for evaluation of radio interface technologies for imt-advanced," *Report ITU*, vol. 638, pp. 1–72, 2009.

[37] B. Mondal, T. A. Thomas, E. Visotsky, F. W. Vook, A. Ghosh, Y.-H. Nam, Y. Li, J. Zhang, M. Zhang, Q. Luo, *et al.*, "3d channel model in 3gpp," *IEEE Communications Magazine*, vol. 53, no. 3, pp. 16–23, 2015.

[38] A. Maltsev, R. Maslennikov, A. Sevastyanov, A. Lomayev, and A. Khoryaev, "Statistical channel model for 60 ghz wlan systems in conference room environment," in *Proceedings of the Fourth European Conference on Antennas and Propagation*. IEEE, 2010, pp. 1–5.

[39] G. Calcev, D. Chizhik, B. Goransson, S. Howard, H. Huang, A. Kogiantis, A. F. Molisch, A. L. Moustakas, D. Reed, and H. Xu, "A wideband spatial channel model for system-wide simulations," *IEEE Transactions on Vehicular Technology*, vol. 56, no. 2, pp. 389–403, 2007.

[40] J. F. Monserrat, G. Mange, V. Braun, H. Tullberg, G. Zimmermann, and Ö. Bulakci, "Metis research advances towards the 5g mobile and wireless system definition," *EURASIP Journal on Wireless Communications and Networking*, vol. 2015, no. 1, pp. 1–16, 2015.

[41] N. Khan and C. Oestges, "Impact of transmit antenna beamwidth for fixed relay links using ray-tracing and winner ii channel models," in *Proceedings of the 5th European Conference on Antennas and Propagation (EUCAP)*. IEEE, 2011, pp. 2938–2941.

[42] A. Algans, K. I. Pedersen, and P. E. Mogensen, "Experimental analysis of the joint statistical properties of azimuth spread, delay spread, and shadow fading," *IEEE Journal on selected areas in communications*, vol. 20, no. 3, pp. 523–531, 2002.

[43] L. Liu, C. Oestges, J. Poutanen, K. Haneda, P. Vainikainen, F. Quitin, F. Tufvesson, and P. De Doncker, "The cost 2100 mimo channel model," *IEEE Wireless Communications*, vol. 19, no. 6, pp. 92–99, 2012.

[44] R. Verdone and A. Zanella, *Pervasive mobile and ambient wireless communications: COST action 2100*. Springer Science & Business Media, 2012.

[45] V. Nurmela, A. Karttunen, A. Roivainen, L. Raschkowski, V. Hovinen, J. Y. EB, N. Omaki, K. Kusume, A. Hekkala, R. Weiler, *et al.*, "Deliverable d1. 4 metis channel models," *Proc. Mobile Wireless Commun. Enablers Inf. Soc.(METIS)*, p. 1, 2015.

[46] N. A. Haq and M. Sarvagya, "Analysis on channel parameters and signal processing methods at mm-wave for 5g networks," in *2018 Second International Conference on Advances in Electronics, Computers and Communications (ICAECC)*. IEEE, 2018, pp. 1–6.

[47] C. Masouros, M. Sellathurai, and T. Ratnarajah, "Large-scale mimo transmitters in fixed physical spaces: The effect of transmit correlation and mutual coupling," *IEEE Transactions on Communications*, vol. 61, no. 7, pp. 2794–2804, 2013.

[48] W. Weichselberger, M. Herdin, H. Ozcelik, and E. Bonek, "A stochastic mimo channel model with joint correlation of both link ends," *IEEE Transactions on wireless Communications*, vol. 5, no. 1, pp. 90–100, 2006.

[49] A. M. Sayeed, "Deconstructing multiantenna fading channels," *IEEE Transactions on Signal processing*, vol. 50, no. 10, pp. 2563–2579, 2002.

[50] T. S. Rappaport, S. Y. Seidel, and K. Takamizawa, "Statistical channel impulse response models for factory and open plan building radio communicate system design," *IEEE transactions on communications*, vol. 39, no. 5, pp. 794–807, 1991.

[51] T. S. Rappaport, W. Huang, and M. J. Feuerstein, "Performance of decision feedback equalizers in simulated urban and indoor radio channels," *IEICE Transactions on Communications*, vol. 76, no. 2, pp. 78–89, 1993.

[52] T. S. Rappaport, S. Y. Seidel, and R. Singh, "900-mhz multipath propagation measurements for us digital cellular radiotelephone," *IEEE Transactions on Vehicular Technology*, vol. 39, no. 2, pp. 132–139, 1990.

[53] S. Y. Seidel, T. S. Rappaport, S. Jain, M. L. Lord, and R. Singh, "Path loss, scattering and multipath delay statistics in four european cities for digital cellular and

microcellular radiotelephone," *IEEE Transactions on Vehicular Technology*, vol. 40, no. 4, pp. 721–730, 1991.

[54] T. S. Rappaport *et al.*, *Wireless communications: principles and practice*. prentice hall PTR New Jersey, 1996, vol. 2.

[55] J. E. Nuckols, "Implementation of geometrically based single-bounce models for simulation of angle-of-arrival of multipath delay components in the wireless channel simulation tools, smrcim and sircim," Ph.D. dissertation, Virginia Tech, 1999.

[56] M. Ding, P. Wang, D. López-Pérez, G. Mao, and Z. Lin, "Performance impact of los and nlos transmissions in dense cellular networks," *IEEE Transactions on Wireless Communications*, vol. 15, no. 3, pp. 2365–2380, 2015.

[57] A. Ghosh *et al.*, "5g channel model for bands up to 100 ghz," *5GCM white paper*, 2015.

[58] T. S. Rappaport, E. Ben-Dor, J. N. Murdock, and Y. Qiao, "38 ghz and 60 ghz angle-dependent propagation for cellular & peer-to-peer wireless communications," in *2012 IEEE international conference on communications (ICC)*. IEEE, 2012, pp. 4568–4573.

[59] S. Nie, G. R. MacCartney, S. Sun, and T. S. Rappaport, "28 ghz and 73 ghz signal outage study for millimeter wave cellular and backhaul communications," in *2014 IEEE International Conference on Communications (ICC)*. IEEE, 2014, pp. 4856–4861.

[60] M. Lei, J. Zhang, T. Lei, and D. Du, "28-ghz indoor channel measurements and analysis of propagation characteristics," in *2014 IEEE 25th Annual International Symposium on Personal, Indoor, and Mobile Radio Communication (PIMRC)*. IEEE, 2014, pp. 208–212.

[61] A. H. Jafari, D. López-Pérez, M. Ding, and J. Zhang, "Performance analysis of dense small cell networks with practical antenna heights under rician fading," *IEEE Access*, vol. 6, pp. 9960–9974, 2017.

[62] J. Meredith, "Study on channel model for frequency spectrum above 6 ghz," *3GPP TR 38.900, Jun, Tech. Rep.*, 2016.

[63] A. I. Sulyman, A. Alwarafy, G. R. MacCartney, T. S. Rappaport, and A. Alsanie, "Directional radio propagation path loss models for millimeter-wave wireless networks in the 28-, 60-, and 73-ghz bands," *IEEE Transactions on Wireless Communications*, vol. 15, no. 10, pp. 6939–6947, 2016.

[64] J.-P. Kermoal, L. Schumacher, K. I. Pedersen, P. E. Mogensen, and F. Frederiksen, "A stochastic mimo radio channel model with experimental validation," *IEEE Journal on selected areas in Communications*, vol. 20, no. 6, pp. 1211–1226, 2002.

[65] S. Sun, T. S. Rappaport, T. A. Thomas, and A. Ghosh, "A preliminary 3d mm wave indoor office channel model," in *2015 International Conference on Computing, Networking and Communications (ICNC)*. IEEE, 2015, pp. 26–31.

[66] G. R. Maccartney, T. S. Rappaport, S. Sun, and S. Deng, "Indoor office wideband millimeter-wave propagation measurements and channel models at 28 and 73 ghz for ultra-dense 5g wireless networks," *IEEE access*, vol. 3, pp. 2388–2424, 2015.

[67] S. Sun, T. S. Rappaport, T. A. Thomas, A. Ghosh, H. C. Nguyen, I. Z. Kovács, I. Rodriguez, O. Koymen, and A. Partyka, "Investigation of prediction accuracy, sensitivity, and parameter stability of large-scale propagation path loss models for 5g wireless communications," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 5, pp. 2843–2860, 2016.

[68] A. M. Hunter, J. G. Andrews, and S. Weber, "Transmission capacity of ad hoc networks with spatial diversity," *IEEE Transactions on Wireless Communications*, vol. 7, no. 12, pp. 5058–5071, 2008.

[69] J. G. Andrews, T. Bai, M. N. Kulkarni, A. Alkhateeb, A. K. Gupta, and R. W. Heath, "Modeling and analyzing millimeter wave cellular systems," *IEEE Transactions on Communications*, vol. 65, no. 1, pp. 403–430, 2016.

[70] U. Schilcher, S. Toumpis, M. Haenggi, A. Crismani, G. Brandner, and C. Bettstetter, "Interference functionals in poisson networks," *IEEE Transactions on Information Theory*, vol. 62, no. 1, pp. 370–383, 2015.

[71] A. E. Ampoma, G. Wen, Y. Huang, K. O. Gyasi, P. I. Tebe, and K. Ntiamoah-Sarpong, "Spatial correlation models of large-scale antenna topologies using maximum power of offset distribution and its application," *IEEE Access*, vol. 6, pp. 36 295–36 304, 2018.

[72] S. Wu, C.-X. Wang, M. M. Alwakeel, Y. He, *et al.*, "A non-stationary 3-d wideband twin-cluster model for 5g massive mimo channels," *IEEE journal on selected areas in communications*, vol. 32, no. 6, pp. 1207–1218, 2014.

[73] J. Li, B. Ai, R. He, M. Yang, and Z. Zhong, "On modeling of dense multipath component for indoor massive mimo channels," *IEEE Antennas and Wireless Propagation Letters*, vol. 18, no. 3, pp. 526–530, 2019.

[74] R. Janaswamy, "Angle of arrival statistics for a 3-d spheroid model," *IEEE Transactions on vehicular technology*, vol. 51, no. 5, pp. 1242–1247, 2002.

[75] J. Cao and J. Wang, "Global asymptotic and robust stability of recurrent neural networks with time delays," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 52, no. 2, pp. 417–426, 2005.

[76] T.-c. Fu, "A review on time series data mining," *Engineering Applications of Artificial Intelligence*, vol. 24, no. 1, pp. 164–181, 2011.

[77] X. Zhao, S. Wang, S. Li, and J. Li, "Passive image-splicing detection by a 2-d noncausal markov model," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 25, no. 2, pp. 185–199, 2014.

[78] S. Solari and M. Á. Losada, "Simulation of non-stationary wind speed and direction time series," *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 149, pp. 48–58, 2016.

[79] B. Li, S. Li, A. Nallanathan, and C. Zhao, "Deep sensing for future spectrum and location awareness 5g communications," *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 7, pp. 1331–1344, 2015.

[80] Y. Chen, B. Yang, J. Dong, and A. Abraham, "Time-series forecasting using flexible neural tree model," *Information sciences*, vol. 174, no. 3-4, pp. 219–235, 2005.

[81] C. Di, B. Zhang, Q. Liang, S. Li, and Y. Guo, "Learning automata-based access class barring scheme for massive random access in machine-to-machine communications," *IEEE Internet of Things Journal*, vol. 6, no. 4, pp. 6007–6017, 2018.

[82] W.-C. Hong, Y. Dong, L.-Y. Chen, and S.-Y. Wei, "Seasonal support vector regression with chaotic genetic algorithm in electric load forecasting," in *2012 Sixth International Conference on Genetic and Evolutionary Computing*. IEEE, 2012, pp. 124–127.

[83] B. Li, W. Guo, H. Zhang, C. Zhao, S. Li, and A. Nallanathan, "Spectrum detection and link quality assessment for heterogeneous shared access networks," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 2, pp. 1431–1445, 2018.

[84] H. Guo, S. Li, B. Li, Y. Ma, and X. Ren, "A new learning automata-based pruning method to train deep neural networks," *IEEE Internet of Things Journal*, vol. 5, no. 5, pp. 3263–3269, 2017.

[85] C. Di, Q. Liang, F. Li, S. Li, and F. Luo, "An efficient parameter-free learning automaton scheme," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 11, pp. 4849–4863, 2020.

[86] V. K. Tumuluru, P. Wang, and D. Niyato, "A neural network based spectrum prediction scheme for cognitive radio," in *2010 IEEE international conference on communications*. IEEE, 2010, pp. 1–5.

[87] Z. Yan, G. Li, Y. TIan, J. Wu, S. Li, M. Chen, and H. V. Poor, "Dehib: Deep hidden backdoor attack on semi-supervised learning via adversarial perturbation," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 12, 2021, pp. 10 585–10 593.

[88] Y. Zhao, S. Li, and F. Jin, "Identification of influential nodes in social networks with community structure based on label propagation," *Neurocomputing*, vol. 210, pp. 34–44, 2016.

[89] R. Hsieh, A. Seneviratne, H. Soliman, and K. El-Malki, "Performance analysis on hierarchical mobile ipv6 with fast-handoff over end-to-end tcp," in *Global Telecommunications Conference, 2002. GLOBECOM'02. IEEE*, vol. 3. IEEE, 2002, pp. 2488–2492.

[90] C. Di, F. Li, and S. Li, "Sensor deployment for wireless sensor networks: A conjugate learning automata-based energy-efficient approach," *IEEE Wireless Communications*, vol. 27, no. 5, pp. 80–87, 2020.

[91] X. Ge, S. Tu, G. Mao, C.-X. Wang, and T. Han, "5g ultra-dense cellular networks," *IEEE Wireless Communications*, vol. 23, no. 1, pp. 72–79, 2016.

[92] N. Nouri, J. Abouei, M. Jaseemuddin, and A. Anpalagan, "Joint access and resource allocation in ultradense mmwave noma networks with mobile edge computing," *IEEE Internet of Things Journal*, vol. 7, no. 2, pp. 1531–1547, 2019.

[93] X. Liu, X. Chen, Y. Chen, and Z. Li, "Deep learning based dynamic uplink power control for noma ultra-dense network system," in *International Conference on Blockchain and Trustworthy Systems*. Springer, 2019, pp. 774–786.

[94] W. Feng, J. Tang, N. Zhao, Y. Fu, X. Zhang, K. Cumanan, and K.-K. Wong, "Noma-based uav-aided networks for emergency communications," *China Communications*, vol. 17, no. 11, pp. 54–66, 2020.

[95] N. Zhao, X. Pang, Z. Li, Y. Chen, F. Li, Z. Ding, and M.-S. Alouini, "Joint trajectory and precoding optimization for uav-assisted noma networks," *IEEE Transactions on Communications*, vol. 67, no. 5, pp. 3723–3735, 2019.

[96] L. Zhu, Z. Xiao, X.-G. Xia, and D. O. Wu, "Millimeter-wave communications with non-orthogonal multiple access for b5g/6g," *IEEE Access*, vol. 7, pp. 116 123–116 132, 2019.

[97] Z. Wei, L. Zhao, J. Guo, D. W. K. Ng, and J. Yuan, "Multi-beam noma for hybrid mmwave systems," *IEEE Transactions on Communications*, vol. 67, no. 2, pp. 1705–1719, 2018.

[98] A. S. Marcano and H. L. Christiansen, "Performance of non-orthogonal multiple access (noma) in mmwave wireless communications for 5g networks," in *2017 International Conference on Computing, Networking and Communications (ICNC)*. IEEE, 2017, pp. 969–974.

[99] D. Zhang, Z. Zhou, C. Xu, Y. Zhang, J. Rodriguez, and T. Sato, "Capacity analysis of noma with mmwave massive mimo systems," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 7, pp. 1606–1618, 2017.

[100] A. Al Khansa, Y. Yin, G. Gui, and H. Sari, "Power-domain noma or noma-2000?" in *2019 25th Asia-Pacific Conference on Communications (APCC)*. IEEE, 2019, pp. 336–341.

[101] E. Khorov, A. Kureev, and I. Levitsky, "Noma testbed on wi-fi," in *2018 IEEE 29th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*. IEEE, 2018, pp. 1153–1154.

[102] Y. Liu, Z. Qin, M. Elkashlan, Z. Ding, A. Nallanathan, and L. Hanzo, "Non-orthogonal multiple access for 5g and beyond," *Proceedings of the IEEE*, vol. 105, no. 12, pp. 2347–2381, 2017.

[103] A. Benjebbour, Y. Saito, Y. Kishiyama, A. Li, A. Harada, and T. Nakamura, "Concept and practical considerations of non-orthogonal multiple access (noma) for future radio access," in *2013 International Symposium on Intelligent Signal Processing and Communication Systems*. IEEE, 2013, pp. 770–774.

[104] Y. Saito, Y. Kishiyama, A. Benjebbour, T. Nakamura, A. Li, and K. Higuchi, "Non-orthogonal multiple access (noma) for cellular future radio access," in *2013 IEEE 77th vehicular technology conference (VTC Spring)*. IEEE, 2013, pp. 1–5.

[105] J. G. Andrews, "Interference cancellation for cellular systems: a contemporary overview," *IEEE Wireless Communications*, vol. 12, no. 2, pp. 19–29, 2005.

[106] P. Patel and J. Holtzman, "Analysis of a simple successive interference cancellation scheme in a ds/cdma system," *IEEE journal on selected areas in communications*, vol. 12, no. 5, pp. 796–807, 1994.

[107] Z. Ding, Y. Liu, J. Choi, Q. Sun, M. Elkashlan, I. Chih-Lin, and H. V. Poor, "Application of non-orthogonal multiple access in lte and 5g networks," *IEEE Communications Magazine*, vol. 55, no. 2, pp. 185–191, 2017.

[108] S. Timotheou and I. Krikidis, "Fairness for non-orthogonal multiple access in 5g systems," *IEEE Signal Processing Letters*, vol. 22, no. 10, pp. 1647–1651, 2015.

[109] M. Shirvanimoghaddam, M. Dohler, and S. J. Johnson, "Massive non-orthogonal multiple access for cellular iot: Potentials and limitations," *IEEE Communications Magazine*, vol. 55, no. 9, pp. 55–61, 2017.

BIOGRAPHICAL STATEMENT

Zhangliang Chen received the B.S. degree from Tianjin University, Tianjin, China, in 2016. He is currently a ph.D. student with the Department of Electrical Engineering, University of Texas at Arlington, Arlington, TX, USA. His current research interests include wireless communications, massive MIMO, radar sensor networks, wireless sensor networks, and signal processing.