Comparison of Exact and Approximate Multi-User Detection for GSM

Lili Nie

Informatics and Mathematical Modelling
Technical University of Denmark

Abstract

In today's Group Spécial Mobile (GSM) system, interference is one of the main constraints in increasing cellular capacity. Multi-User Detection (MUD) is a kind of Interference Cancellation (IC) technique, which can be combined with other IC methods, such as antenna diversity, whitening. This dissertation investigates exact and approximate MUD GSM receivers. It is shown that exact MUD solution provides a big Bit Error Rate (BER) gain compared to conventional receivers. However, it has exponential complexity, making it infeasible to implement it on the limited Mobile Station (MS). In this thesis, the approximation to the exact solution is based on Mean Field theory. Two sub-optimum algorithms: Fully Factorized Mean Field (FFMF) receiver and Structured Mean Field (SMF) receiver are implemented and evaluated. FFMF has very low complexity, comparable to that of the Linear Minimum Mean Squared Error (LMMSE) receiver, but much better BER performance for interference dominated scenarios. The SMF receiver gives faster convergence speed. However, for one Cochannel Interference (CCI), its performance is only close to that of FFMF solution in most of the tested CIR range and better than that of FFMF receiver at low CIR values.

Besides, topics such as digital phase modulation, GSM basics, the multi-path fading channel and conventional GSM receivers are also studied.

Keywords: Adjacent Channel Interference (ACI), Cochannel Interference (CCI), Fully Factorized Mean Field (FFMF), Group Spécial Mobile (GSM), Interference Cancellation (IC), Inter-Symbol Interference (ISI), Linear Minimum Mean Squared Error (LMMSE), Maximum a Posteriori (MAP), Multi-User Detection (MUD), Structured Mean Field (SMF).

Preface

This thesis is submitted for the official examination for the degree of Master of Science. The work for this project has been conducted from July 1, 2004 to March 31, 2005. It has been done in Modem System Design (MSD), Nokia Danmark A/S. Supervisors of this project are Associate Professor Ole Winther, Informatics and Mathematical Modelling (IMM), Technical University of Denmark (DTU), and Ph.D. student Lars Puggaard Bøgild Christensen, IMM, DTU and Nokia Danmark.

I would like to express my gratitude to the supervisors, who have guided me for this project and given me good suggestions and advice. Special thanks belongs to Ph.D. Pedro Højen-Sørensen, Nokia, for the interesting discussions about mean field approximation and for taking his time to join the project meetings. I would also like to thank Ph.D. Hong Liu, Nokia, for various inspiring discussions. I am grateful to Ph.D. Niels Mørch, line manager of MSD Copenhagen, Nokia, who gave me the opportunity to do this project in his group. Hong Liu, Lars P.B. Christensen, Ole Winther, Pedro Højen-Sørensen, Stefan Klukowski are gratefully acknowledged for their efforts to review this thesis.

Copenhagen, March 31, 2005

Lili Nie

Contents

List of Figures

List of Abbreviations

8-PSK	Octagonal Phase Shift Keying
ACI	Adjacent Channel Interference
AMP	Amplitude Modulated Pulse
AWGN	Additive White Gaussian Noise
BCJR	Bahl, Cocke, Jelinek, and Raviv
BER	Bit Error Rate
BLUE	Best Linear Unbiased Estimate
BPSK	Binary Phase Shift Keying
BS	Base Station
CCI	Cochannel Interference
CDMA	Code Division Multiple Access
CEPT	Conférence des Administrations Européenne des Postes et Télécommuniations
CIR	Carrier to Interference Ratio
DIR	Dominating to remaining Interference Ratio
DSP	Digital Signal Processing
DPSK	Differential Phase Shift Keying
EDGE	Enhanced Data rates for Global Evolution
EM	Expectation Maximization
ETSI	European Telecommunications Standards Institute
FDD	Frequency Division Duplex
FDMA	Frequency Division Multiple Access
FFMF	Fully Factorized Mean Field
FH	Frequency Hopping
FHMM	Factorial Hidden Markov Model

FIR	Finite-duration Impulse Response
FSK	Frequency Shift Keying
GERAN	GSM EDGE Radio Access Network
GMSK	Gaussian Minimum Shift Keying
GSM	Group Spécial Mobile
GPRS	General Packet Radio Service
HMM	Hiden Markov Model
HT	Hilly Terrain
IC	Interference Cancellation
ISI	Inter-Symbol Interference
JD	Joint Detection
JLS	Joint Least Square
LCL	Linear Conjugate Linear
LMMSE	Linear Minimum Mean Squared Error
LS	Least Square
LoS	Line of Sight
MA	Multiple Access
MAP	Maximum a Posteriori
MF	Mean Field
MIMO	Multi-input Multi-output
ML	Maximum Likelihood
MLSE	Maximum Likelihood Sequence Estimation
MoU	Memorandum of Understanding
MS	Mobile Station
MSK	Minimum Shift Keying
MUD	Multi-User Detection
OQPSK	Offset Quadrature Phase Shift Keying
PDF	Probability Density Function
QPSK	Quadrature Phase Shift Keying
RA	Rural Area
SAIC	Single Antenna Interference Cancellation

SbS Symbol-by-Sybmol

SMF Structured Mean Field

SNR Signal to Noise Ratio

SUD Single User Detection

TDMA Time Division Multiple Access

TS Training Sequence

TU Typical Urban

UMTS Universal Mobile Telecommunications System

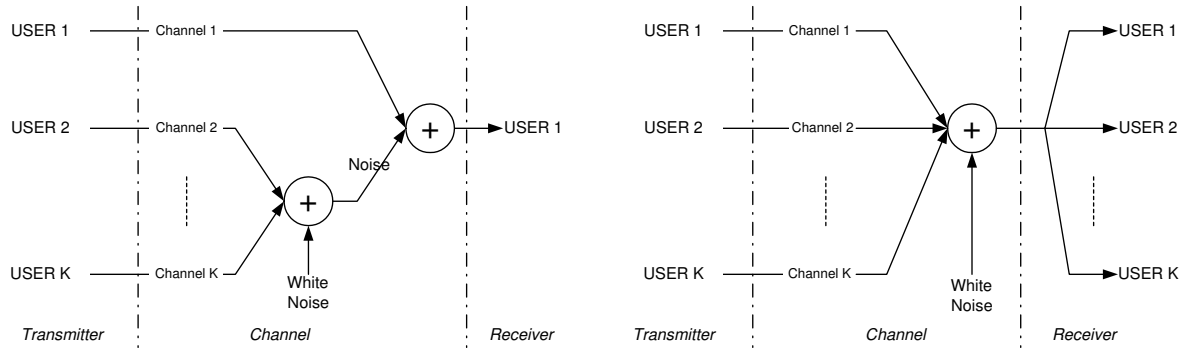
UWB Ultra Wide-band

VA Viterbi Algorithm

Chapter 1

Introduction

Development of mobile communication system has been growing dramatically for the last two decades. Due to the limited and expensive frequency resources and the fast growing market, the design target of GSM system changed from covering larger area in the early days to increasing capacity recently. The traditional methods for increasing capacity, such as power control, cell splitting, sectoring, zone micro cell technique, antenna space diversity and polarization diversity in base stations have already been widely used (p. 54-63, 332-335 in [1]). Capacity can be increased futher by tighten the frequency reuse factor, but interference from other users will also increase accordingly, thus decrease signal condition.



(a) Single User Detection

(b) Multi-User Detection

Figure 1.1: SUD vs. MUD

Conventional GSM receivers extract information only from the desired user while treating noise from channel and interference as white noise. It works well when interference is not so strong and white noise is dominant, but might fail if interference becomes dominant. As illustrated in Figure 1.1 (a), the overall noise includes white noise and contributions from other users, therefore, modeling it as white noise is not completely correct, especially when the power from other users is strong. In this case, it makes sense to investigate the properties of the interference and use these properties for better detection of the desired information. Thus we get several IC techniques, such as whitening filter, antenna diversity in the receiver side (ref. Section 1.2), or MUD.

The objective of this project is to evaluate and design MUD receiver algorithms for interference dominated scenarios.

1.1 Why Multi-User Detection?

MUD aims to provide better detection of the desired user by means of detecting interference and subtracting its contribution. From Figure 1.1 (b), we can see that the difference, compared to conventional receivers, lies in the simutanous detection of the desired user (USER1) as well as interferers (USER 2 to USER K).

For interference dominated networks, the theoretical optimal MUD algorithm (p. 851-852 in [2] and p. 154-209 in [3]), in the sense of BER, outperforms conventional receivers. It seems that the problem is easily solved. Not quite. The optimal MUD algorithm has complexity exponential in number of users and length of channel, making it impossible to implement in today's MS, where size and computation power are limiting factors. Hence suboptimal MUD has become a hot research topic recently. The goal is to make the best tradeoff between performance and complexity.

Further, if the complexity of the MUD algorithm can be reduced to a certain level, it is possible to combined it with other IC algorithms, e.g. whitening filter or antenna diversity.

1.2 Other IC Techniques

There exist alternative IC methods. One relies on antenna diversity, e.g. [4]. It uses antenna array to receive multiple signals and then cancel the interference. It separates signals in the spatial domain and has very impressive BER performance. However, it is in general very complex, expensive and space costing in the limited physical size of MS. Dual antenna solution has received attention lately, as it seems to have an acceptable tradeoff between MS size and BER gain. In stead of using multiple antennas, Single Antenna Interference Cancellation (SAIC) is termed to represent IC solution that uses only one antenna. The benefit of SAIC is that it saves space on the MS, also it only requires a software update on the receiver. It is a comparably cheap yet effective IC solution and becomes very interesting for operators and handsets manufacturers.

Another alternative is the whitening approach, which suppresses interference in a statistial fashion. Whitening receiver is the common SAIC solution now. Whitening filter models interference as colored noise, it finds the statistical properties of interference and tries to convert it to white noise [5]. More about whitening filter is found in Appendix B. MUD IC approach relies on joint channel estimation. Compared to MUD method, whitening filter avoids the estimation of the propagation channel of interference, but instead, the statistics of interference needs to be estimated.

Conventional receivers perform well enough in good signal condition. All IC methods, such as antenna diversity, whitening filter or MUD algorithm, require more computations than conventional receivers, the purpose is to provide better data detection in interference dominated scenario.

If computation cost permits, those IC methods can also be combined with each other. For instance, use whitening filter for dual antenna setup to achieve better BER than SAIC, or apply whitening filter to MUD algorithm, so that the desired user and strong interferers are jointly detected, weak interferers are transformed to white noise, etc.

1.3 Project Scope

In the field of MUD for GSM, there are many interesting topics that can be studied. In order to perform this project within the given time, the following conditions or assumptions are made to constrain the scope of this thesis work.

- Using SAIC solution instead of antenna diversity, since MS size is still recognized as a serious constraint for a multiple antenna solution. Moreover, MUD algorithms derived from SAIC can be easily adjusted to fit into multiple antenna context.
- Consider the coexistence of CCI and ACI (Definition of CCI and ACI will be explained shortly in 2.2.3). ACI can be rejected by receiver filtering, but when ACI is too strong, it is also effective in subtracting it by MUD algorithm.
- The desired user and interferers are synchronized. It is not true for the present GSM networks. But there are many discussions, especially in the US, that synchronizing the Base Station (BS)es in GSM gives reasonable tradeoff between investment and capacity gain. Besides, the synchronization technology is mature and available on the market. So this assumption may come true soon, when GSM operators are willing to pay for synchronizing their networks.
- Multi-path channel with AWGN. This is a realistic channel in cellular wireless communication systems, and it will be used for testing receiver algorithms in this thesis.
- Signal complex Low-Pass equivalence modeling. Thus symbols, channel parameters and noise are complex. Furthermore, the modulated signal is moved from carrier frequency to base band.
- Focus on GMSK modulation for both desired and interference. Enhanced Data rates for Global Evolution (EDGE) utilizes $\frac{3\pi}{8}$ shifted Octagonal Phase Shift Keying (8-PSK), provided a three fold gain compared to GMSK. However, EDGE is not considered in this project.
- Evaluate receiver performance by all three measures, that is the trade-off among BER, complexity and convergence speed, if applicable.

1.4 Thesis Structure

General knowledge relating to this project is introduced in Chapter 2. Thereafter, digital phase modulation and Laurent linear approximation to GMSK modulation are addressed in Chapter 3. After that, multi-path fading channel is discussed in Chapter 4.

Conventional GSM receivers are presented in Chapter 5, followed by the optimum MUD receiver in Chapter 6. Two suboptimal receivers are derived and evaluated in Chapter 7. Summary of this project and suggestions for future research are given in Chapter 8.

1.5 Summary

This chapter first introduces the difference between conventional receivers and MUD algorithms. Interference problem associated with cellular network is the driver for IC technology, because how well receivers can work against interference is one of the constraint of cellular network capacity.

MUD is one of the IC methods, other than MUD, there are also whitening approach or multiple antenna solution, etc. BER performance of optimal MUD solution in theory is attractive, but the complexity is too high, so suboptimal solution are needed. This project focuses on comparison of exact and approximate MUD solutions, the specification of this project in also given in this chapter. In the end, layout of this thesis is introduced.

Chapter 2

GSM Basics

This chapter presents relevant knowledge of GSM. A common digital communication system block diagram is introduced first. Afterwards, an overview of GSM system is given. At the end, the interference scenarios in GSM are discussed, as it leads to our problem why MUD receivers are needed to cancel the interference.

2.1 General Digital Communication System

The basic building blocks of a digital communications system can be illustrated by Figure 2.1 [2, 6]. One of the benefits of a digital communication system is its efficiency in spectrum. That is, it occupies less bandwidth to transmit same amont of information campared to analog system. Thus digital system is of great interest for mobile communication.

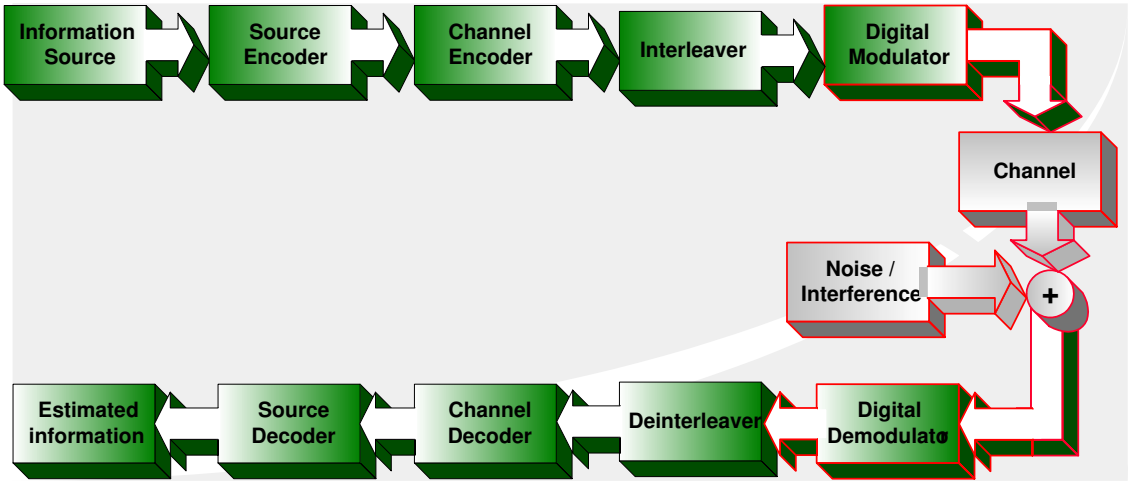


Figure 2.1: Digital Communication System Block Diagram

From top-left corner, information source is the information that needs to be transmitted, e.g. voice, data, image, etc. The redundancy of information source is removed (to its entropy) by source encoder. Channel encoder introduces redundancy again, in such a way that the receiver can use its property to overcome the distortion from the channel. Interleaver shuffles bits so that they can

be modeled as random. The "random" bits are modulated into physical signal, then transmitted through the channel, be it cable, air, etc. For radio channel, e.g. in GSM, due to the channel effects, BER performance is always worse than wired channel. Besides, thermal noise exists in any communication system operating above 0 K. Therefore, the received signal is distorted by the noise in the channel, which brings challenge in detection of the desired information.

From the receiver side, it is the reverse process of the transmitter. Since the modulation type is known by the receiver, demodulator can estimate what has been transmitted from the corrupted signal. Deinterleaver put the bits back to the right order, and it is the opposite process of what has been done by interleaver. Channel decoder corrects errors caused by the channel, and finally after source decoder, it is the estimated desired information.

This project mainly deals with those three blocks marked in the red in Figure 2.1; they are modulator, channel, and demodulator. The emphasis is on the demodulator part, as several detection algorithms are studied later. These three blocks will be discussed in the following three chapters in the sequence they appear. In Chapter 3, digital phase modulation family is presented. GMSK is deployed by GSM, and $\frac{3\pi}{8}$ shifted 8-PSK is implemented for EDGE to increase the data rate to three times of GMSK. After that, channel charateristics and modeling are discussed thoroughly in Chapter 4. In Chapter 5, conventional receivers are introduced.

2.2 Basic Knowledge of GSM

The above section gives a brief introduction to digital communication systems and points out the main issues relevant to the scope of this project. This section gives an overview of the specific digital wireless communication system - GSM, as the work of this thesis investigates the application of MUD algorithms for GSM system. GSM evolution is introduced first, followed by MA schemes, Time Division Multiple Access (TDMA) frame and burst structures. Interference scenarios in GSM are discussed in the end.

2.2.1 GSM evolution

It was back in 1897, Guglielmo Marconi first demonstrated the continuous contact with ships sailing the English channel through radio interface [1]. Historically, development of mobile communication system has been slow in the beginning. The concept of cellular system makes wireless communication technology available for personal communications. It was first promoted by AT&T Bell laboratory in 1947 and widely developed in 1960s and 1970s. Public cellular mobile phone system starts from analog in 1977 and goes digital in 1987.

Beginning with analog cellular telephony, cellular networks were developed, however locally they are not compatible to each other. It means subscribers could not make a call from one network to another. In 1982, GSM was formed at the Conférence des Administrations Européenne des Postes et Télécommuniations (CEPT). The purpose of this study group was to develop a pan-European solution for mobile communication network. The benefits offered by digital network, such as the high quality transmission, encrypted speech and data, cheaper and smaller handsets, motivated GSM group to implement a digital specification from the start. However, the final decision was not made until February 1987. With the demand for a single and permanent organization, the GSM Permanent Nucleus settled its headquarter in Paris in 1986. The responsibility of developing the specification for a pan-European communication network passed to European Telecommunications Standards

Institute (ETSI) in 1989. With the contribution from the co-operative environment and improved resources, the majority of phase 1 GSM 900 specifications was published in 1990. Requested by UK, DCS 1800 (renamed to GSM 1800 in 1997) was implemented in parallel with GSM. While the real launch of GSM in late 1992, it has 1 million subscribers by the end of 1993. It was growing faster than everyone could imagine. Expanding from the initial pan-European network, GSM was going global. By June 1995, Memorandum of Understanding (MoU) was officially founded as an association in Switzerland, and has 239 members from 109 countries now. GSM phase 2 specification was launched in the same year [7].

2.2.2 Multiplexing and MA Scheme

GSM system has two transmission directions, BS to MS is called downlink and MS to BS is called uplink. The multiplexing is implemented by Frequency Division Duplex (FDD), i.e. downlink and uplink occupy two different frequency bands.

In order to assign more than one subscriber, a MA scheme is needed. There are basically three types of MA scheme, namely TDMA, Frequency Division Multiple Access (FDMA), Code Division Multiple Access (CDMA). It describes the way to separate channels, be it time, frequency, or code, see Figure 2.2.

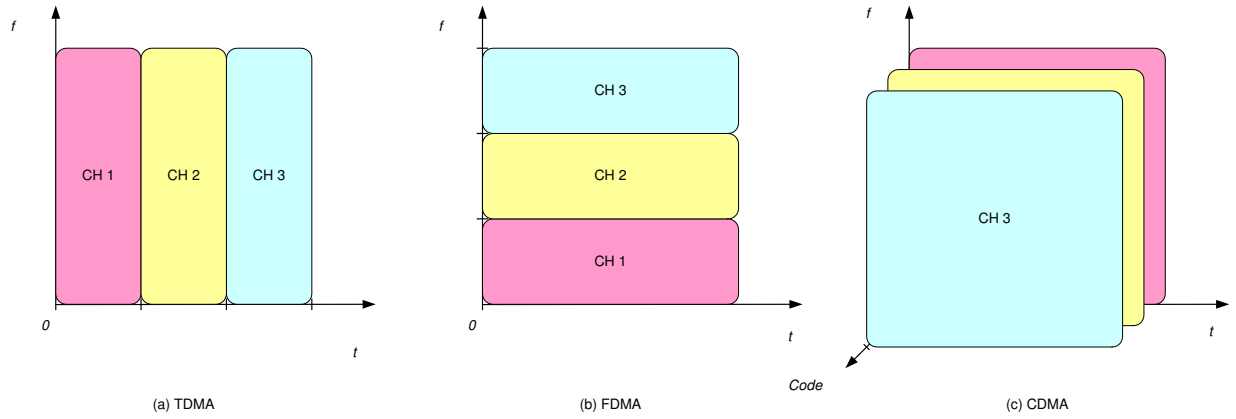


Figure 2.2: MA Schemes

For TDMA, as shown in Figure 2.2 (a), channels are separated in time domain. Time axis is divided into time frames and time slots, each user occupies one time slot in every time frame. Since only one user is allowed to transmit in a particular time, it can use all the bandwidth available during its transmission time slot.

In FDMA, channels are separated on frequency axis. Each user can transmit within the limited bandwidth in unlimited time, see Figure 2.2 (b).

CDMA doesn't have limitation on time or frequency. Several users can transmit all the time in the available bandwidth. Channels are separated by their own codes, see Figure 2.2 (c). By carefully choosing the code, e.g. orthogonal codes, it is possible that the transmitted information is only understandable by the desired receiver. For other receivers, this information is not understandable and normally regarded as noise. Therefore, power control is crucial for CDMA systems, as otherwise it is hard to design efficient receivers.

GSM combines FDMA and TDMA. It means that channels are separated by 200 kHz in frequency domain and each user is allowed to transmit during its time slot in every time frame. Time slot is usually refered to TDMA burst, or just burst.

In GSM system, a multi-frame is 120 ms, which contains 26 frames. One frame consists of 8 bursts, and each burst has duration 577 μs. Figure 2.3 is a graphical illustration of relations between multi-frame, frame, and burst. It also shows the structure of a GSM normal burst. Besides normal burst, there are also random access burst, frequency correction burst, synchronization burst. Details of GSM physical layer is available in [8, 6].

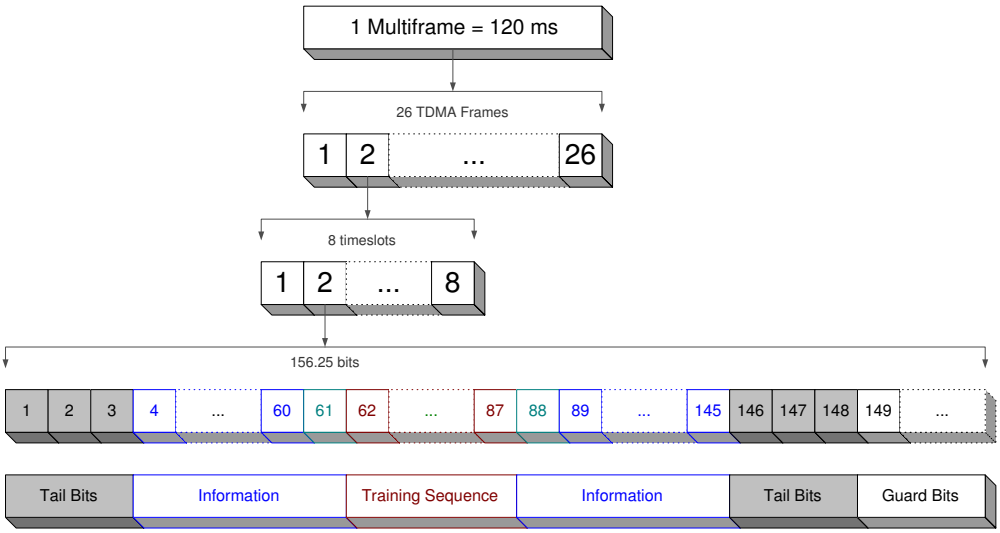


Figure 2.3: GSM TDMA frame and normal burst structure

A normal GSM burst contains $148+8.25 = 156.25$ bits. They are tail bits, information bits (contains two stealing bits), Training Sequence (TS) and guard bits, as shown in Figure 2.3. One bit duration is therefore $T_b = 577\ \mu s/156.25\ bits = 3.69\ \mu s/bit$, the bit rate in GSM is $R_b = 1/T_b = 270.83\ kbps$. Tail bits are located in the beginning and end of a burst as guard period and always set to zero. They are very helpful for initiali<ing demodulators. TS is in the middle of a burst, which is used for synchronization and channel estimation. There are in total 8 TSs in GSM system, which are chosen for their good autocorrelation properties (p.4 and 16 in [8]). Guard period contains no data and has duration of 8.25 T_b. $2(57+1) = 116$ information bits are differentially encoded in GSM, ref. (3.9). Inside information bits part, those two bits located next to each side of TS are stealing flag, which tells decoder the function of burst.

One interesting thing is that the data rate in GSM is 270.83 $kbps$, whereas the channels are separated only 200 kHz apart. The purpose is to efficiently use the limited frequency resource and support as many users as possible. GSM makes a compromise between efficiency and performance, which can be roughly shown as Figure 2.4.

The non-ideal spectrum and compressed channel bandwidth cause spectral overlap, which is illustrated as the blue area in Figure 2.4. This overlap approximately gives capacity gain by $\frac{270.83}{200} - 1 = 35\%$, however, also brings challenge to design receivers to combat ACI. One way to solve this problem is to choose a proper modulation method, so that the power spectrum in the blue area is sufficiently low, so that the overlap can be safely ignored. The selected GMSK modulation scheme provides a good spectral property, that is a narrow main lobe and low side lobes. It assures that most energy is within 200 kHz band, details about modulation are described in Chapter 3. Note

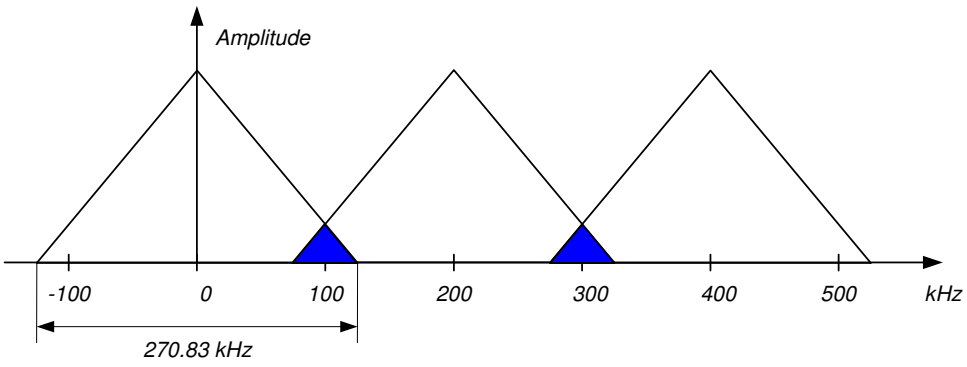


Figure 2.4: GSM channel overlap illustration

that, although the side lobes of GMSK are quite low, when ACI is very strong compared to the desired signal, the small amount of energy spilled into the desired band can also be significant, and its contribution into the desired signal can not be ignored. Interference problem is discussed in the next section.

2.2.3 Interference scenario

Mobile cellular network is made from a number of base stations, each of them covering a certain geographical area, named 'cell'. Neighborhood cells use different radio frequencies in order to minimize interference with each other. However, same frequency band may be reused by other base stations that are far away. Therefore, there is a tradeoff between the interference level and frequency efficiency.

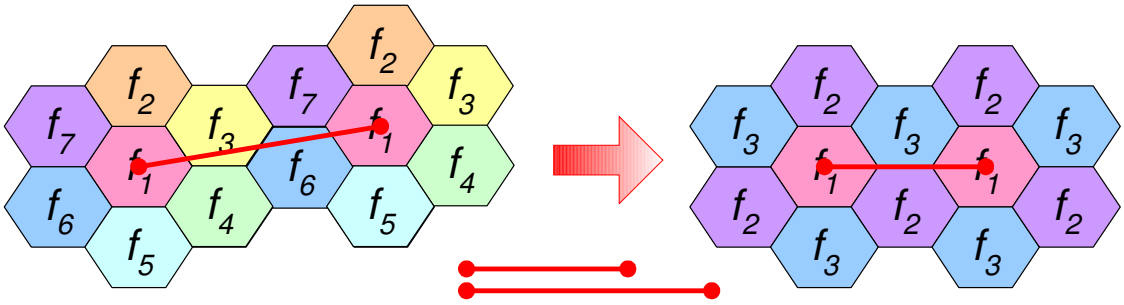


Figure 2.5: Cell layout, freq. reuse factor 7 vs. 3

Figure 2.5 shows two examples of cell layout for GSM system. On the left-hand side, the frequency reuse factor is 7, which means each cell uses one of seven frequencies in this pattern (seven cells inside bold lines). It is a classic layout since the basic pattern, cell f_1 to f_7, can be 'copied' and 'pasted' to make a larger coverage, and the distance between cells using same channel is unchanged. Apparently, it is an easy way to plan a network that covers a large area, however, not the most efficient way. Frequency reuse factor can be tightened as long as the interference from other cells is below a certain limit. As it is said in Chapter 1, how much this factor can be tightened depends on the receiver's ability to resist interference. Cell layout on the right-hand side of Figure 2.5 is an example of frequency reuse factor of 3. The red line represents the distance between cells using

same frequency. Graphically, it shows this distance becomes smaller when tightening the frequency reuse factor, which indicates interference becomes stronger. The power of interference is not only determined by cell layout, but also many other factors. For instance, the propagation path between the interfering transmitter and desired receiver, number of interferers, etc. In frequency domain, interference is classified as ACI and CCI.

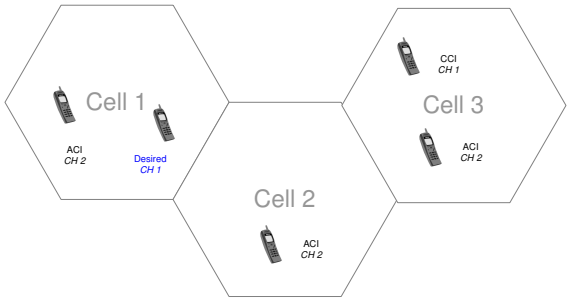


Figure 2.6: Existence of ACI and CCI in the cell layout for frequency reuse factor 3

As stated in the previous section, ACI is caused by the energy from the adjacent channels (up and/or down 200 kHz in GSM) leaking into the desired channel. ACI turns worse under some circumstances, e.g. the power from the adjacent channel is much higher than the desired, channel effects (e.g. Doppler shift, multi-path channel, details refer to Chapter 4), the modulation scheme (details in Chapter 3) itself has tails extending into neighboring bands. As shown in Figure 2.6, ACI exists in both its own cell and other cells.

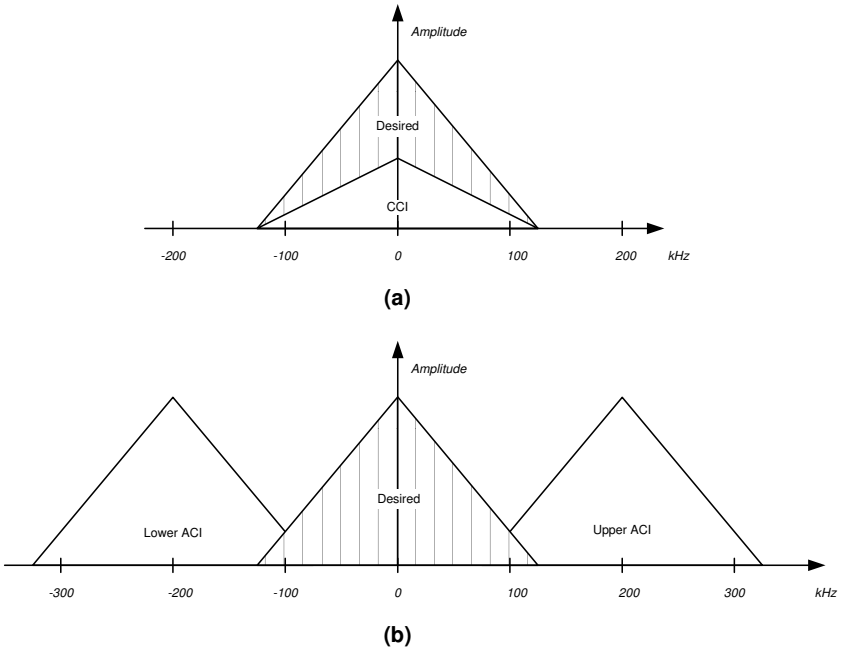


Figure 2.7: Frequency domain drawing of ACI and CCI

Besides ACI, there is also CCI. As shown in Figure 2.6, CCI only comes from other cells that using the same channel (frequency/carrier). In one cell, only one user is allowed to use a specific freqnecy/carrier at a given time. CCI can be quite significant if the geographical distance between

cells is very small [1]. Figure 2.7 is an illustration for ACI and CCI in frequency domain.

In time domain, interference can be categorized into synchronous or asynchronous. Figure 2.8 is an example of the relations between desired burst and interference burst in time domain. Interfering burst can be either ACI or CCI.

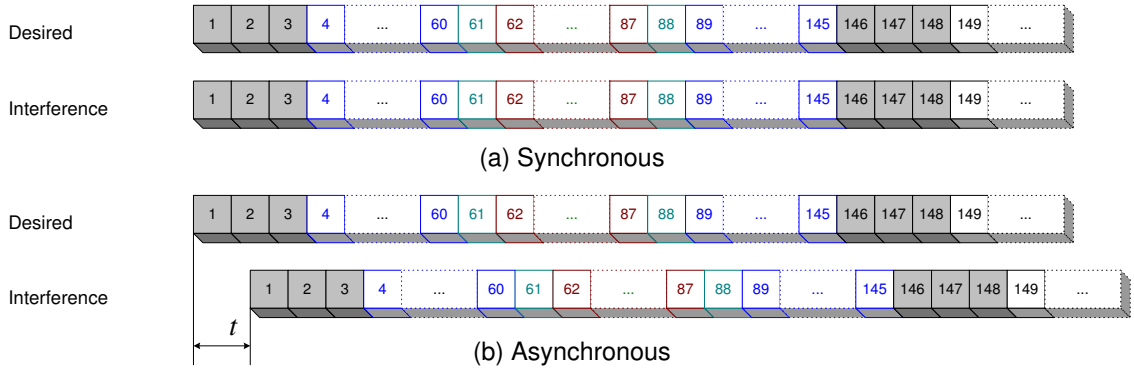


Figure 2.8: Example of Synchronous /Asynchronous Interference for normal burst

In pratice, there is normally more than one interferer. A lot of contribution has been done to model the realistic interference scenarios for testing the receiver performance. The latest GSM EDGE Radio Access Network (GERAN) specification describes five test cases [9]. Those test cases are used to evaluate receiver performance later in this thesis.

Several parameters are used to compare receiver algorithms. Receiver performance is normally measured by the BER. BER is defined as the probability that a bit is in error. If N_b is number of tested bits, and N_e is number of bits that are detected wrong, then BER can be written as:

$$BER = \frac{N_e}{N_b} \tag{2.1}$$

For AWGN channel, BER is normally determined by the Signal to Noise Ratio (SNR). It is the average received signal power for the desired user divided by the noise power in the band of interest. The ratio is normally expressed in decibels (dB). If the signal power is E_s, the double-side Additive White Gaussian Noise (AWGN) power density is $N_0/2$, and baseband signal has double-side bandwidth of $2B_s$, SNR is expressed as:

$$SNR = \frac{E_s}{N_0 B_s} \tag{2.2}$$

For the interference dominated scenario, which means SNR is sufficiently high, the main aspect to influence BER performance is not SNR, but the ratio between the average received signal power from the desired user and the average interference power, it is called Carrier to Interference Ratio (CIR). If the total number of users is K, where the first user is the desired user and rest are interferers. Using superscript (k) to denote different users, then CIR can be written as:

$$CIR = \frac{E_s^{(1)}}{\sum_{k=2}^{K} E_s^{(k)}} \tag{2.3}$$

Note that with the notation of user index, SNR can be rewritten as:

$$SNR = \frac{E_s^{(1)}}{N_0 B_s} \tag{2.4}$$

Normally, receiver performance is not only limited by CIR, but also the distribution among the interferers. Dominating to remaining Interference Ratio (DIR) is used to describe the ratio between the average receive signal power from dominating interference and the sum of other interferers' average powers. If the k'-th user is the dominant interferer, $DIR^{(k')}$ is given by:

$$DIR^{(k')} = \frac{E_s^{(k')}}{\sum_{k=2,\ k\neq k'}^{K} E_s^{(k)}} \tag{2.5}$$

When there is no interference, receiver BER performance is determined by SNR, the higher SNR, the better BER performance (lower BER), this is proved later by simulation results for conventional receivers in Section 5.4. For more complicated test case, i.e. more than one interferer, receiver performance is mainly determined by the power distribution among interferers [10].

2.3 Summary

In this chapter, a general digital communication block diagram and the basic functions of each block are given first. Modulator, propagation channel and demodulator are the most interesting blocks for the scope of this project, and they will be discussed in the following three chapters.

A brief introduction to GSM evolution is also presented and MA schemes are introduced thereafter. GSM deploys a combination of FDMA and TDMA. Channels are separated 200 kHz apart in frequency axis. For each channel, information is transmitted burst by burst. A TDMA burst contains 148 bits. A normal GSM burst has 26 bits TS in the middle of the burst, 58×2 information bits allocated on each side of TS, 3 tail bits in both end of a burst, 8.25 bits are transmitted between two burst as guard period. Other than normal burst, there are also Random Access Burst, Frequency Correction Burst, and Synchronization Burst, they are used to assist the transmission/reception process.

GSM is a cellular digital communication system, ACI and CCI exist due to the cellular layout and the limited channel bandwidth of GSM. As a consequence of increasing capacity, ACI and CCI become severe for conventional GSM receivers. A better detection algorithm is required to handle the increasing interference, IC technique is named for this purpose. Due to the ever increasing demand to increase capacity of GSM network, IC becomes a hot research topic recently. MUD is one of IC solution.

Chapter 3

Digital Phase Modulation

After bits are formed into a burst format, it needs to be carried by a physical signal in order to accomplish transmission via the radio interface. This task is performed by the modulator. The modulator is the last block in the transmitter of a general digital communication system, as shown in Figure 2.1. A modulator converts the signal source (baseband signal) to a bandpass signal centered in a certain carrier frequency. This process is called modulation. Modulation can be performed by changing the carrier's amplitude, frequency, or phase.

A number of issues were considered for choosing the modulation scheme for GSM:

1. Constant envelope: As the transmission medium in GSM is air, the amplitude of the received signal is unpredictable due to propagation loss and fading effects from the channel. Constant envelope modulation makes the receiver design easier as the amplitude does not contain any useful information. Therefore, digital phase modulation was a suitable candidate.

2. Efficiency: Because the frequency resource is quite expensive, in order to make GSM a cheap and attractive system, one of the design targets was to assign as many subscribers as possible, and also assure the quality of service. In terms of physical layer, that is to get a high bit rate per channel and low BER.

3. Spectrum: Limiting the interference between neighboring channels requires a tight spectrum for each channel.

Owing to the above considerations, GSM deploys a member of the digital phase modulation family - GMSK, with $BT = 0.3$, B is the 3 dB bandwidth, T is the symbol duration, which is denoted as T_s in this thesis. In GSM, symbol duration is the same as bit duration, i.e. $T_s = T_b$.

This chapter introduces several digital phase modulation methods, explaining why GMSK was selected among others. These methods are described with the emphasis on bit rate per channel and BER performance. BPSK is presented first, followed by DPSK, QPSK, OQPSK, MSK and GMSK. Afterwards, an amplitude modulation approximation to GMSK is presented, which simplifies receiver design.

3.1 BPSK

The simplest type of digital phase modulation is BPSK. It sets the carrier's phase to either of two different values to represent ±1. Those two phases are separated by π to achieve the best possible BER performance in an AWGN channel.

$$S_{BPSK}(t) = \begin{cases} A\cos(2\pi f_c t + \theta_0) & \text{for } \mathrm{b_n} = 1 \quad (\mathrm{n}-1)\mathrm{T_b} \leq \mathrm{t} \leq \mathrm{nT_b} \\ A\cos(2\pi f_c t + \pi + \theta_0) & \text{for } \mathrm{b_n} = -1 \quad (\mathrm{n}-1)\mathrm{T_b} \leq \mathrm{t} \leq \mathrm{nT_b} \end{cases} \tag{3.1}$$

where n is time index, f_c is the carrier frequency. θ_0 is the initial phase, which is ignored hereafter for simplification. A is amplitude, the energy per bit is $E_b = \frac{1}{2}A^2T_b$.

Defining an analog signal $b(t) = b_n \quad (n-1)T_b \leq t \leq nT_b$, and using a trigonometric property, (3.1) with $\theta_0 = 0$ can be rewritten in amplitude modulation form, that is:

$$S_{BPSK}(t) = Ab(t)\cos(2\pi f_c t) \tag{3.2}$$

For BPSK is symbol duration is the same as bit duration, i.e. $T_s = T_b$. and symbol rate is therefore given as:

$$R_{s,BPSK} = \frac{1}{T_s} \tag{3.3}$$

An example for the binary sequence $[1, -1, -1, 1]$ is shown in Figure 3.1, the red dashed line represents $b(t)$, and the black curve is the modulated signal. Note that, the phase transition is accomplished in a very short instant, it is a kind of discrete digital phase modulation, and it is different from continuous phase modulation, for instance MSK. MSK is introduced shortly in Section 3.5.

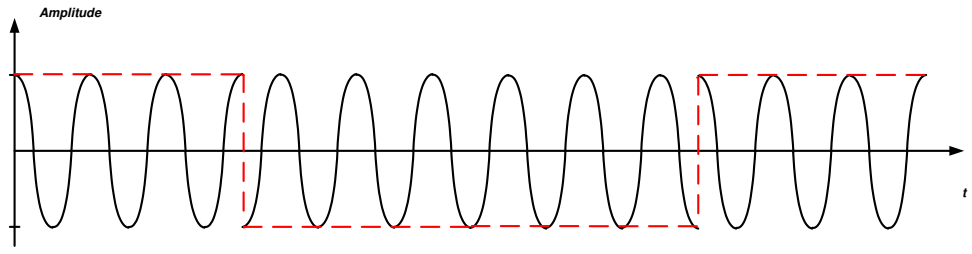


Figure 3.1: BPSK modulation for the sequence [1,-1,-1,1]

From Figure 3.1 we can see that when $b(t)$ changes its sign, which corresponds to a phase shift of π. In this case, the envelope of the carrier to return to zero momentarily. This can also be explained by the constellation diagram. For $\theta_0 = 0$, the constellation diagram can be depicted as in Figure 3.2. The blue dashed line denotes the phase shift of π.

Now, let's analyze BER for BPSK. As mentioned in Section 2.1, received signal is corrupted by noise from the channel. For instance, if binary symbol $-\sqrt{E_b}$ is transmitted and the noise power

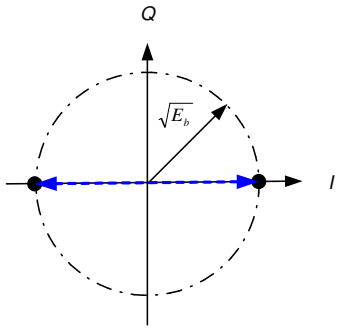


Figure 3.2: BPSK constellation diagram

is $+1.5\sqrt{E_b}$, after demodulation, the received signal is then $+0.5\sqrt{E_b}$. If using hard decision, i.e. $\text{sgn}(0.5\sqrt{E_b}) = +\sqrt{E_b}$, therefore this bit is detected wrong. In this example, BER equals to the probability of noise power $\geq \sqrt{E_b}$, thus is determined by the Probability Density Function (PDF) of noise. PDF of normal Gaussian distribution is given in [2]:

$$P(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-m_x)^2/2\sigma^2} \tag{3.4}$$

where m_x is the mean and σ^2 is the variance of the random variable. It is called white Gaussian noise because it has a flat spectrum density over the entire frequency axis, and is usually described by the double-sided density $N_0/2$ $watt/Hz$.

In AWGN channel, the probability of noise power $\geq \sqrt{E_b}$ is described by error function $Q(x)$. $Q(x)$ is the integral of the normalized Gaussian function from x to ∞. The bigger of x, the smaller of $Q(x)$ and it is defined as:

$$Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} \exp(\frac{-\tau^2}{2}) d\tau \tag{3.5}$$

So for normal Gaussian distribution, the error probability in AWGN channel for BPSK is described by:

$$P_{e,BPSK} = Q\Big(\sqrt{\frac{2E_b}{N_0}}\Big) \tag{3.6}$$

For BPSK, the Euclidean distance (or mean square error) between constellation points is $2\sqrt{E_b}$, as shown in Figure 3.2. For a more general case, an upper bound of the error probability for any constellation in an AWGN channel can be obtained by the union bound (p. 237 in [1] and [11]). If s_i is the i-th signal point in the constellation and d_{ij} is the Euclidean distance between $s_i(t)$ and $s_j(t)$ in the constellation, defined as:

$$d_{ij}^2 = \int_0^{T_b} \big|s_i(t) - s_j(t)\big|^2 dt \tag{3.7}$$

The average error probability for a particular modulation is given as:

$$P\Big(e|s_i(t)\Big) \leq \sum_{j \neq i} Q\left(\frac{d_{ij}}{\sqrt{2N_0}}\right) \tag{3.8}$$

It is the sum of the probability that $s_i(t)$ is detected as $s_j(t)$ for all $j \neq i$. Note that, since $Q(x)$ decreases fast when x increase, BER for any constellation is mainly determined by the nearest constellation points.

3.2 DPSK

BPSK modulation needs a coherent receiver, which is complex and expensive. DPSK is a slight modification to BPSK in order to use non-coherent receiver. The change is that input bits are differentially encoded prior to phase selection. That is:

$$d_n = b_n \oplus b_{n-1} \tag{3.9}$$

where *oplus* denotes exclusive or. For $b_n \in [\pm 1]$,

$$d_n = b_n b_{n-1} \tag{3.10}$$

Note that for differentially encoded bits, one detection error causes two bits to be detected wrong, thus the error probability of DPSK is worse than that of BPSK. In AWGN channel, BER for DPSK is given as (p.243 in [1]):

$$P_{e,DPSK} = \frac{1}{2} \exp\Big(-\frac{E_b}{N_0}\Big) \tag{3.11}$$

3.3 QPSK

BPSK modulation uses two phases separated by π to represent ± 1. From Figure 3.2, we can see that the circle can be more efficiently used by adding more constellation points on the circle. This is normally called M-ary PSK ($M = 2^{n_b}$), in this case, every constellation point represents n_b bits. The phase separation is $2\pi/M$.

QPSK modulation is a kind of M-ary PSK, with $M = 2^2 = 4$. It is implemented by two carrier signals in quadrature:

- $\cos(2\pi f_c t)$: I branch
- $\sin(2\pi f_c t)$: Q branch

Information bits are first split into two groups according to the time they appear: even and odd time index. Assuming the bits are already differentially encoded by (3.9), the input sequence is split as follows:

$$[d_0\ d_1\ d_2\ d_3\ d_4\ d_5\ \cdots] \rightarrow \begin{cases} [d_0\ d_2\ d_4\ \cdots] & \text{for I branch} \\ [d_1\ d_3\ d_5\ \cdots] & \text{for Q branch} \end{cases} \tag{3.12}$$

Like for BPSK, assign two analog signals to denote the digital bits for I and Q branches respectively:

$$\begin{cases} d_I(t) = d_n \quad for\ n\ \in [0, 2, 4, \cdots], \quad nT_b \leq t \leq (n+2)T_b \\ d_Q(t) = d_n \quad for\ n\ \in [1, 3, 5, \cdots], \quad nT_b \leq t \leq (n+2)T_b \end{cases} \tag{3.13}$$

Note that the symbol duration is then doubled to $T_s = 2T_b$.

Feed $d_I(t)$ and $d_Q(t)$ into the modulator's I and Q branch separately, the modulated signal is the sum of these two quadrature terms:

$$S_{QPSK}(t) = \frac{A}{\sqrt{2}} d_I(t) \cos(2\pi f_c t) + \frac{A}{\sqrt{2}} d_Q(t) \sin(2\pi f_c t) \tag{3.14}$$

Comparing (3.14) to (3.2), it is easily seen that, a single branch (I or Q) is the same as BPSK modulation. The sum of these two BPSK signals is the QPSK signal. Applying trigonometric identities, (3.14) can also be written as:

$$S_{QPSK}(t) = A \cos(2\pi f_c t + \theta_i) \qquad\qquad i \in [1, 2, 3, 4] \tag{3.15}$$

where $\theta_i \in [0, \pm\pi/2, \pi]$ for the four possible $d_I(t)$, $d_Q(t)$ combinations. The I and Q branches of the QPSK modulated signal for the sequence $[1, -1, -1, 1]$ are shown in Figure 3.3, again the red dashed line represents $b_I(t)$ and $b_Q(t)$. The final output is the sum of the I and Q branches. The phase of the modulated signal changes every T_s, with one of four possible phases $(0, \pm\pi/2, \pi)$, and each phase represents two bits (for I and Q branches).

Since $T_s = 2T_b$, the bandwidth for QPSK is then half of BPSK modulation. It means that data throughput is doubled within the same bandwidth compared to BPSK. Symbol rate for QPSK is half of $R_{s,BPSK}$ and is given as:

$$R_{s,QPSK} = \frac{1}{T_s} \tag{3.16}$$

The constellation diagram is shown in Figure 3.4. Blue dashed lines denote phase transitions. Note that the envelope of QPSK signals returns to zero momentarily, as it happens for BPSK.

For QPSK, one symbol represents two bits, so $E_s = 2E_b$. As shown in Figure 3.4, the Euclidean distance of the nearest constellation points is $\sqrt{2E_s} = 2\sqrt{E_b}$. Inserting it into (3.8) and the error probability of QPSK in AWGN channel appears:

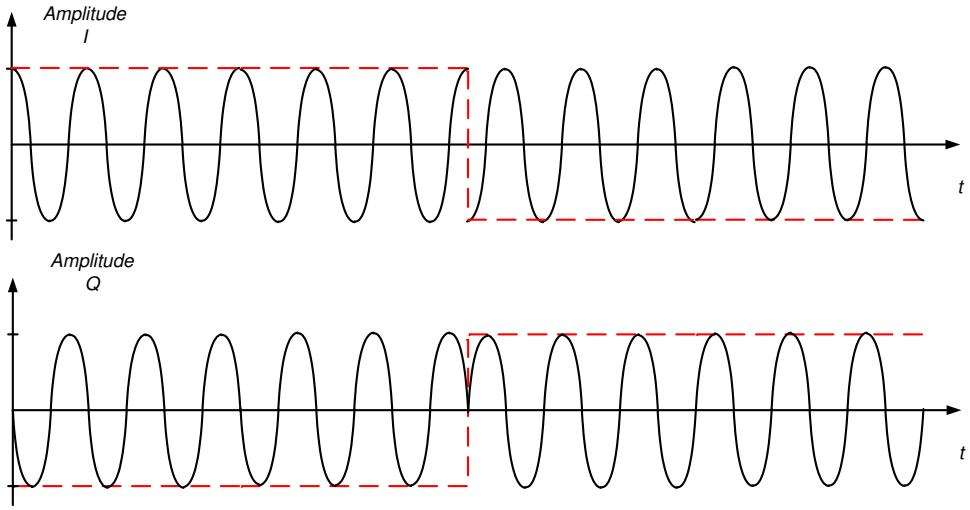


Figure 3.3: QPSK modulation

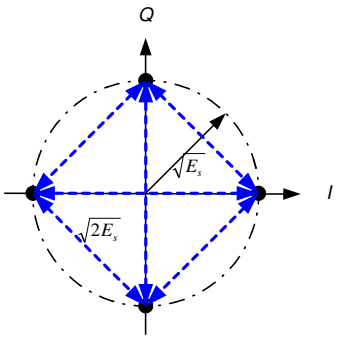


Figure 3.4: QPSK constellation diagram

$$P_{e,QPSK} = Q\left(\sqrt{\frac{2E_b}{N_0}}\right) \tag{3.17}$$

Although QPSK doubles the capacity, the BER performance is the same as BPSK, which is due to the orthogonal property between I and Q branches. Since $\cos(2\pi f_c t)$ and $\sin(2\pi f_c t)$ are orthogonal, they do not interfere with each other, thus the error performance does not get worse.

3.4 OQPSK

QPSK doubles the spectrum efficiency compared to BPSK. However, the envelope returns to zero occasionally due to the phase change of π (as shown in Figure 3.3), which makes it difficult to design an efficient amplifier. OQPSK makes a small modification to QPSK to solve this problem.

OQPSK delays the Q part by one bit duration T_b compared to QPSK. The modulated signal of OQPSK for the sequence $[1, -1, -1, 1]$ is shown in Figure 3.5.

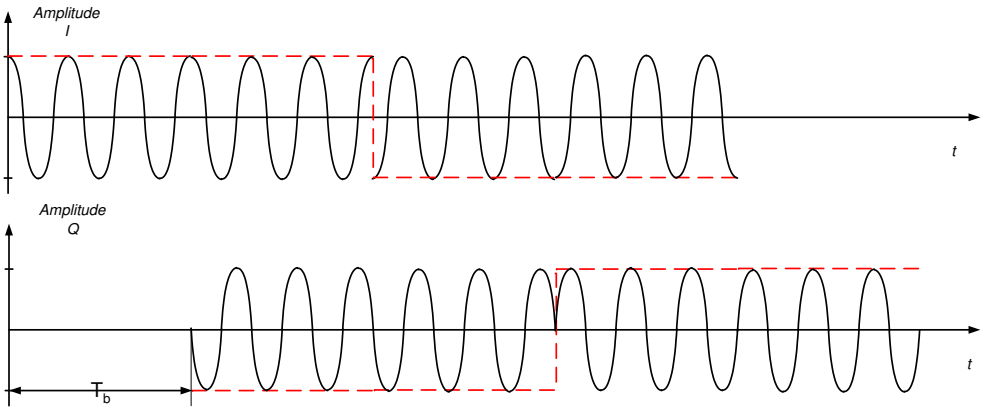


Figure 3.5: OQPSK modulation

The offset of T_b between the I and Q branches helps to avoid a phase shift of π, the maximum phase shift is $\pm\pi/2$ for OQPSK, but $\pm\pi$ for QPSK. Thus the envelope never returns to zero for any phase shift. It can be illustrated by the constellation diagram in Figure 3.6. Blue dashed lines denote phase shifts, and those two lines crossing origin are removed compared to Figure 3.4, since phase phase shifts of $\pm\pi$ do not exist.

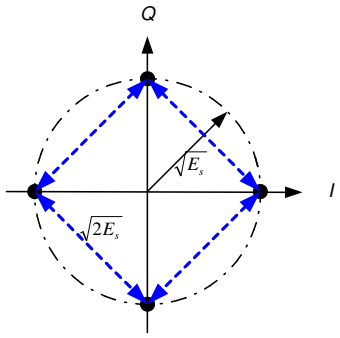


Figure 3.6: OQPSK constellation diagram

The time shift of the Q branch does not influence symbol rate, spectrum and BER, compared to that of QPSK. Symbol rate for OQPSK is:

$$R_{s,OQPSK} = \frac{1}{T_s} \tag{3.18}$$

The error probability in AWGN channel with double-sided noise density $N_0/2$ $watt/Hz$ is:

$$P_{e,OQPSK} = Q\Big(\sqrt{\frac{2E_b}{N_0}}\Big) \tag{3.19}$$

3.5 MSK

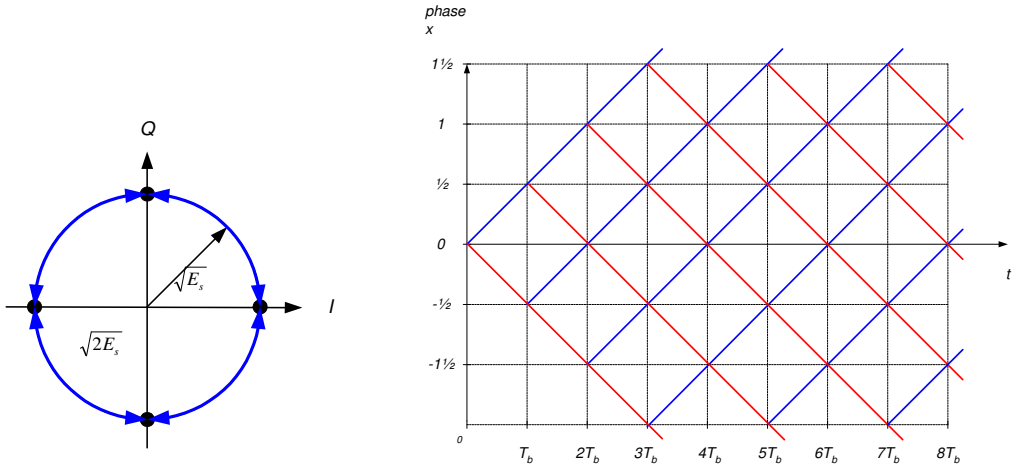
MSK is derived from OQPSK. Instead of using rectangular pulse shapes $d_I(t)$ and $d_Q(t)$ to modulate the I and Q branches, MSK uses a half-cycle sinusoidal pulses. The MSK modulated signal for a binary sequence of length N_{MSK} is defined as:

$$S_{MSK}(t) = \frac{A}{\sqrt{2}} \sum_{n=0}^{N_{MSK}-1} d_I(t) m(t-2nT_b) \cos(2\pi f_c t) + \frac{A}{\sqrt{2}} \sum_{n=0}^{N_{MSK}-1} d_Q(t) m(t-2nT_b-T_b) \sin(2\pi f_c t) \tag{3.20}$$

where $m(t)$ is the half sinusoidal pulse, defined as:

$$m(t) = \begin{cases} \sin\left(\frac{\pi t}{2T_b}\right) & 0 \le t \le 2T_b \\ 0 & other\ t \end{cases} \tag{3.21}$$

For BPSK, QPSK and OQPSK, phase shift happens at the instant when one bit changes sign compared to previous bit. Then phase is unchanged during the time of T_b. As shown in (3.20), the phase shifts linearly by $\pm\pi/2$ during one bit duration T_b for MSK. This kind of modulation methods is also called continuous phase shift keying. The benefit is that the power spectral density has low side lobes and a wider main lobe than QPSK/OQPSK.



(a) *MSK constellation diagram* (b) *phase tree for MSK*

Figure 3.7: MSK constellation diagram and phase tree

The constellation diagram is shown in Figure 3.7 (a), where the phase transition is illustrated with the blue solid line to represent the continuous linear phase shift property. Also observe that, the envelope of MSK is also constant like for OQPSK. Figure 3.7 (b) is an other way to illustrate the linear phase shift property. It is termed phase tree. The blue path corresponds to $d_n = 1$, whereas red path corresponds to $d_n = -1$. From Figure 3.7 (b), we can see that the MSK signal can also be expressed in a more direct form:

$$\begin{aligned} S_{MSK}(t) &= A\cos\left(2\pi f_c t + \sum_{n=-\infty}^{\infty} d_n \theta_{MSK}(t - nT_b)\right) \\ &= A\mathrm{Re}\left\{\exp\left(j\left(2\pi f_c t + \sum_{n=-\infty}^{\infty} d_n \theta_{MSK}(t - nT_b)\right)\right)\right\} \end{aligned} \tag{3.22}$$

where $\theta_{MSK}(t)$ is the phase pulse, which is written as:

$$\theta_{MSK}(t) = \frac{\pi}{2}\frac{t}{T_b} \qquad for \;\; 0 \leq t \leq T_b \tag{3.23}$$

The differentiation of the phase pulse is the frequency pulse, and vice versa. Thus phase modulation can also be described in the frequency domain. The frequency pulse for MSK is a rectangular pulse, which is written as:

$$\varphi_{MSK} = \frac{1}{2T_b} \qquad for \;\; 0 \leq t \leq T_b \tag{3.24}$$

It shows that MSK shifts the carrier frequency by $\pm\frac{1}{2T_b}$ to represent binary input symbols. It is therefore a kind of digital Frequency Shift Keying (FSK). These two frequencies are separated by $\frac{1}{T_b}$, which is the minimum allowed frequency separation for orthogonal carriers, thus it gets its name MSK.

The example for input the sequence $[1, \; -1, \; -1, \; 1]$ is shown in Figure 3.8.

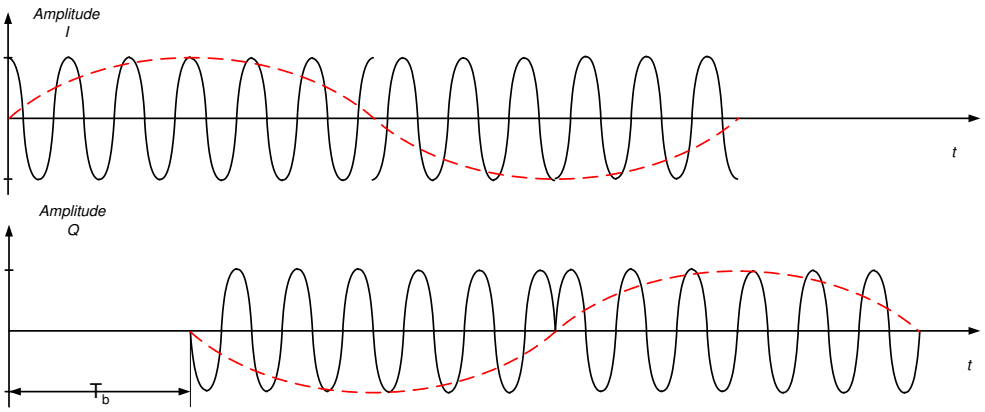


Figure 3.8: MSK modulation

The symbol rate is the same as for QPSK/OQPSK:

$$R_{s,MSK} = \frac{1}{T_s} \tag{3.25}$$

As shown in Figure 3.7 (a), the Euclidean distance between constellation points is the same as in QPSK, thus the error probability is also the same.

$$P_{e,MSK} = Q\left(\sqrt{\frac{2E_b}{N_0}}\right) \tag{3.26}$$

3.6 GMSK

Until here, BPSK, DPSK, QPSK, OQPSK and MSK have been introduced. They all belong to the constant envelope modulation family. Constant envelope modulation eases amplifier design. Because it does not require a linear amplifier as the envelope doesn't contain useful information anyway. It seems that the requirements, listed in the introduction part of this chapter, can be fulfilled by MSK.

There are other modulation schemes that have better spectrum performance than MSK. GMSK is one of them.

GMSK is derived from MSK by feeding the rectangular frequency pulse φ_{MSK} to a Gaussian low pass filter to obtain a Gaussian shaped frequency pulse, hence the name GMSK. The frequency pulse is:

$$\varphi_{GMSK}(t) = \frac{1}{\sqrt{2\pi}\sigma T_b} \exp\left(\frac{-t^2}{2\sigma^2 T_b^2}\right) \tag{3.27}$$

where σ is defined as:

$$\sigma = \frac{\sqrt{\ln 2}}{2\pi B T_b} \tag{3.28}$$

The phase pulse is the integration of the frequency pulse:

$$\theta_{GMSK}(t) = \frac{\pi}{2}\int_{-\infty}^{t} \varphi_{GMSK}(\tau)d\tau = \frac{\pi}{2}\int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}\sigma T_b} \exp\left(\frac{-\tau^2}{2\sigma^2 T_b^2}\right)d\tau \tag{3.29}$$

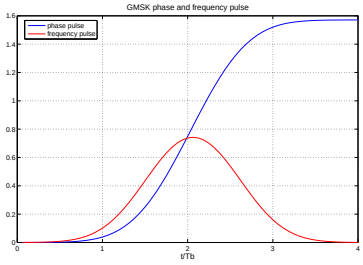


Figure 3.9: GMSK phase and frequency pulse

$\varphi_{GMSK}(t)$ and $\theta_{GMSK}(t)$ are non causal and infinite long. However, they are normally truncated to a duration of a few bit periods. For GMSK with $BT = 0.3$, the truncation of $\varphi_{GMSK}(t)$ to $4T_b$ makes it accumulate 99.99% of $\pi/2$ [12]. The truncated $\varphi_{GMSK}(t)$ and $\theta_{GMSK}(t)$ are illustrated in Figure 3.9. A GMSK modulated signal can be described as (3.30):

$$S_{GMSK}(t) = A\mathrm{Re}\left\{ \exp\left(j\Big(2\pi f_c t + \sum_{n=-\infty}^{\infty} d_n \cdot \theta_{GMSK}(t - nT_b)\Big)\right)\right\} \tag{3.30}$$

where d_n is the differentially encoded bits, given by (3.9). T_b is the symbol duration, $\theta_{GMSK}(t)$ is the phase function. The phase change of GMSK is non-linear, which is different from MSK. However, the power spectral density function has lower side lobes and narrower main lobe compared to MSK.

The symbol rate of GMSK is the same as QPSK/OQPSK and MSK, that is:

$$R_{s,GMSK} = \frac{1}{2T_b} \tag{3.31}$$

BER of GMSK in AWGN channel is given as (p. 264 in [1]):

$$P_{e,GMSK} = Q\Big(\sqrt{\frac{2\zeta E_b}{N_0}}\Big) \tag{3.32}$$

where ζ is a function of BT product. For $BT = 0.3$, GMSK suffers less than 1 dB degradation in error performance compared to MSK [13]. On the other hand, it has a better spectrum than MSK. The spectrum property makes GMSK more attractive than MSK therefore GMSK, with $BT = 0.3$, is chosen as the modulation scheme for GSM.

3.7 Linear Approximation of GMSK Signal

GMSK is selected because of its constant envelope and well restricted spectrum. These properties ensure that it is easy to generate the signal and design efficient amplifiers however detection in the receiver is complicated because of the non-linear property. Pierre Laurent presented in his paper [14] that any constant amplitude binary phase modulation can be exactly expressed as a sum of a finite number of time limited Amplitude Modulated Pulse (AMP), and can be approximately represented by only one main pulse. This contribution greatly simplifies receiver design for binary phase modulation due to the linear property. Details of AMP representation can be found in [12, 15] and Section 3.2.2 and 3.2.3 in [16]. The conclusion for decomposition of GMSK signals is introduced here. A GMSK signal can be approximately expressed by the main pulse $C_0(t)$:

$$S_{GMSK}(t) \approx \sum_{n=-\infty}^{\infty} s_n C_0(t - nT_b) \tag{3.33}$$

where

$$s_n = j^{\sum_{m=-\infty}^{n} d_m} \tag{3.34}$$

Assuming initial bits $b_{-\infty} = 1$, and $j^{-1} = -j,\ j^0 = 1,\ j^1 = j,\ j^2 = -1,\ j^3 = -j\ \ \ldots$, inserting (3.9) and using the fact that d_n and b_n are ± 1, (3.34) can be rewritten as:

$$\begin{aligned} s_n &= \prod_{m=-\infty}^{n} j^{d_m} = \prod_{m=-\infty}^{n} j d_m = \prod_{m=-\infty}^{n} j b_{m-1} b_m \\ &= j^n \Big(b_{-\infty} \cdots \underbrace{b_{m-3} b_{m-2}}_{\mathrm{d_{m-2}}} \underbrace{b_{m-2} b_{m-1}}_{\mathrm{d_{m-1}}} \underbrace{b_{m-1} b_m}_{\mathrm{d_m}} \cdots b_{n-1} b_n \Big) \\ &= j^n b_{-\infty} b_n = j^n b_n \end{aligned} \tag{3.35}$$

(3.33) can now be simplified as:

$$S_{GMSK}(t) \approx \sum_{n=-\infty}^{\infty} j^n b_n \cdot C_0(t - nT_b) \tag{3.36}$$

So the non-linear phase modulated GMSK signal in (3.30) is approximated by (3.36). From (3.30) and (3.36), it can be seen that the current output of GMSK modulated signal is determined by the current input bit as well as several previous bits. This phenomenon is called ISI [17]. The linear approximation of the phase modulated signal greatly simplifies data detection. The $C_0(t)$ pulse is one T_b longer than frequency pulse $\varphi_{GMSK}(t)$, for instance, if $\varphi_{GMSK}(t)$ is truncated to $4T_b$, then $C_0(t)$ has duration $5T_b$.

It has been shown that only using the $C_0(t)$ pulse is a very good approximation, which captures most of the energy of the GMSK signal, the approximation error can normally be ignored. However, for our problem of MUD, when the interference is much stronger than the desired signal, it is shown by simulations that this approximation error degrades MUD receiver BER performance.

3.8 Summary

The constant envelope associated with digital phase modulation is attractive, one of the reasons is that it does not require a linear amplifier. It is much easier and cheaper to design a non-linear amplifier. The efficient non-linear amplifier significantly extends the battery charge cycle of MS compared to its linear counterpart. This chapter introduces several digital phase modulation methods with the emphasis on the analysis of symbol rate, BER performance and spectrum.

GMSK has better spectrum properties than other PSK methods, such as BPSK/DPSK, QPSK/OQPSK and MSK. They are tighter main lobe and lower side lobes. The power spectra of those modulation schemes can be found in Figure 5.22, 5.27, 5.28 and 5.41 in [1]. However, GMSK suffers BER degradation compared to MSK and the pulse shape generates ISI. These make it difficult to detect the modulated signal by the receiver. Another challenge is the non-linearity of GMSK. Laurent AMP representation of digital phase modulation shows that most of the energy from GMSK modulated signal can be captured by using only $C_0(t)$ pulse. AMP representation of GMSK signal makes the receiver design simpler and cheaper, so it is normally adopted by GSM receivers.

Chapter 4

Radio Interface

Chapter 3 demonstrates several digital phase modulation techniques. GSM uses the GMSK modulation scheme, with $BT = 0.3$. After data is modulated to a physical signal, it is transmitted through the channel, as seen in Figure 2.1.

The channel is the physical medium used to send the signal from the transmitter to the receiver [2]. In a mobile communication system, this medium is the air, also called air interface or radio interface. The air interface frees the user from connecting wires, the user can move flexibly. On the other hand, unlike for wired communication, the channel between transmitter and receiver is unpredictable and very complex to model. As mentioned in Section 2.1, BER performance of radio channel is generally worse than wired channel.

Path loss, shadowing and multi-path fading effects cause the transmitted signal to be disturbed or obstructed by the channel. Modeling the wireless channel is normally accomplished by statistical means.

In this chapter, the single path AWGN channel is introduced first, and it is used as the building block to understand the more realistic channel-the multi-path channel with AWGN. The properties and models of the multi-path channel with AWGN are described afterwards. Based on the channel models, the signal model in the multi-path channel with AWGN is given thereafter, which is used in the rest of the thesis to explain receiver algorithms.

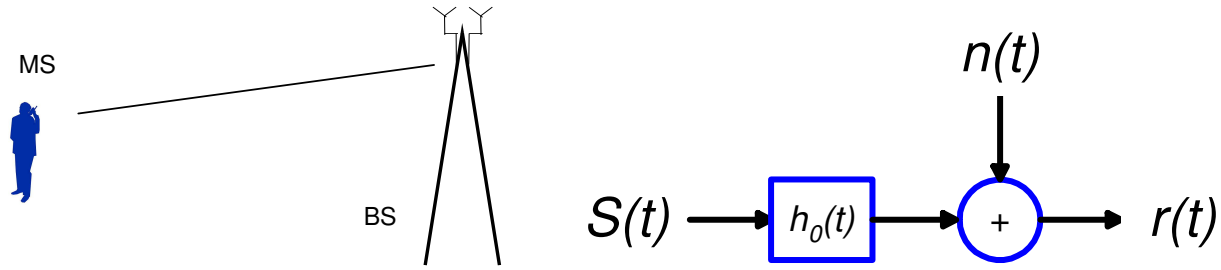
4.1 Single Path AWGN Channel

A radio channel can be as simple as in Figure 4.1 (a), which means there is only one path between transmitter and receiver, with the presence of AWGN. The transmitted signal passes through the radio channel causing the amplitude and phase to be changed and AWGN to be added.

Figure 4.1 (b) is the model of the example in Figure 4.1 (a), the received signal $r(t)$ may be expressed as:

$$r(t) = h_0(t) \cdot S(t) + n(t) = A(t)e^{j\phi(t)} \cdot S(t) + n(t) \qquad h_0(t) \in \mathbb{C} \tag{4.1}$$

where $h_0(t)$ is the transfer function, which is described by amplitude $A(t)$ and phase change $\phi(t)$.



(a) *Example* (b) *Model*

Figure 4.1: A single path AWGN channel

Depending on the channel condition, $A(t)$ and $\phi(t)$ can be time-variant or constant. $S(t)$ is the modulated signal and $n(t)$ is White Gaussian noise with zero mean and variance σ^2.

4.2 Channel Effects

The single path AWGN channel model gives a basic relation between transmitted and received signal, which is described by the transfer function $h_0(t)$. However the practical radio channel is normally much more complex. There are three basic mechanisms of wave propagation: reflection, diffraction and scattering [1]. Reflection is caused by a radio wave encountering a large obstacle compared to its wave length λ. Diffraction happens when the wave travels around the obstacle, also called shadowing. Scattering occurs when the radio wave hits a small object compared to λ. An example of radio wave propagation from BS and MS is shown in Figure 4.2.

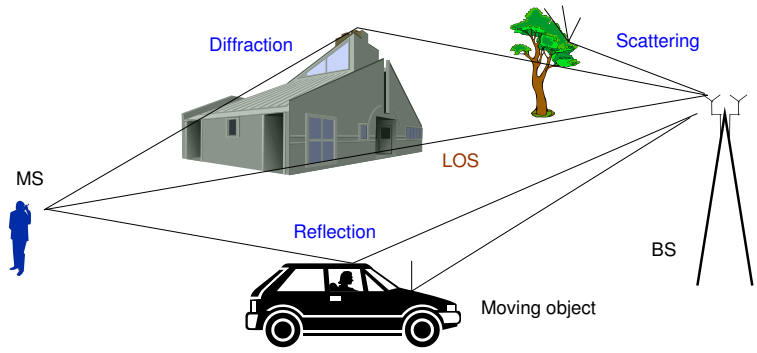


Figure 4.2: An example of multi-path propagation

Generally, there is more than one replica of the transmitted signal arriving at the MS at slightly different times. In other words the signal travels along different paths from transmitter to receiver, hence the name multi-path channel.

Each path is determined by the surrounding objects, which are random. Different paths are considered to be independent. The movement of MS and surrounding objects makes the channel parameters change over time. Since the multi-path channel is non-stationary and unpredictable, it is normally modeled statistically.

The time difference results in a phase difference in the radio waves, therefore the sum of replicas at

the receiver, may be constructive or destructive. The envelope of the modulated signal $S(t)$ is no longer constant due to the multi-path phenomenon. The variation of the amplitude of the received signal is called signal fading [2].

Over a large transmitter-receiver distance (up to several km), the average power of the received signal is degraded, this is called the large-scale path loss model [18].The attenuation depends on the transmission path, and is normally derived empirically based on the measurements. On top of the large scale path loss curve, the amplitude of a radio signal fluctuates rapidly for a short distance or transmission time. The small-scale fading model, or fast fading, is used to describe the channel characteristics over a short period of time or distance [1]. It is normally within several λ. For a 1 GHz signal, λ is 0.3 m.

4.2.1 Rayleigh and Ricean fading

A fast fading can be modeled as a zero mean and complex valued Gaussian process. The envelope is then Rayleigh distributed. It is called a Rayleigh fading channel. The PDF of Rayleigh distribution is given as (p. 173 in [1]):

$$P_{Rayleigh}(r) = \begin{cases} \frac{r}{\sigma^2} e^{\frac{-r^2}{2\sigma^2}} & \text{for } r \geq 0 \\ 0 & \text{for } r < 0 \end{cases} \tag{4.2}$$

where σ^2 is the time average power of the received signal before envelope detection, r is the envelope voltage. A Rayleigh fading channel may undergo 100 fades /sec at speeds of 80 km/h and a frequency of 900 MHz [13].

If the channel has an individual dominant multi-path component, e.g. line-of-sight component, then it cannot be modeled as zero mean anymore, the envelope has a Rice distribution property, and the channel is called a Ricean fading channel. The PDF of a Ricean fading channel is given by (p. 176 in [1]):

$$P_{Ricean}(r) = \begin{cases} \frac{r}{\sigma^2} e^{\frac{-(r^2+A^2)}{2\sigma^2}} I_0(\frac{Ar}{\sigma^2}) & \text{for } A \geq 0,\ r \geq 0 \\ 0 & \text{for } r < 0 \end{cases} \tag{4.3}$$

where A is the dominant signal peak, $I_0(s)$ is the modified Bessel function of the first kind and zero-order, which is given as:

$$I_0(x) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \exp(x \cos\theta) d\theta \tag{4.4}$$

Obviously, if $A \to 0$, it is the same as the Rayleigh distribution.

4.2.2 Delay spread

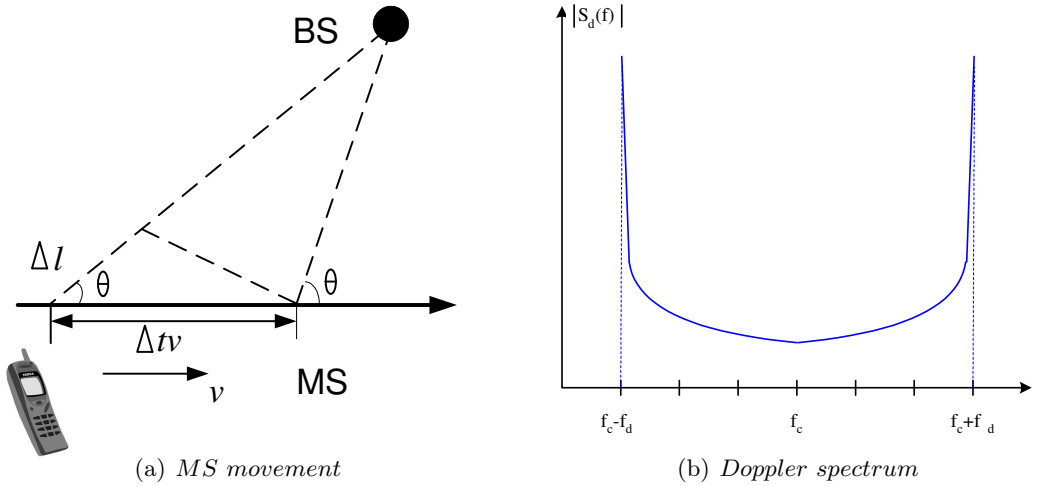
Delay spread is used to describe the channel effect caused by the delay of signal. For now, disregard the large-scale path loss effect here, and normalize the power of the transmitted signal $S(t)$ to 1. The received signal $r(t)$ from two paths with a small time delay T_d is:

$$\begin{aligned} r(t) &= S(t) + S(t - T_d) \\ R(f) &= S(f) + S(f)e^{-j2\pi fT_d} = S(f)[1 + e^{-j2\pi fT_d}] \\ &= 2S(f)\cos \pi fT_d e^{-j\pi fT_d} \end{aligned} \tag{4.5}$$

In the time domain, when $T_d > 0$, delay spread introduces ISI, which is a function of time. From the expression of the spectrum, we can see that $R(f)$ is the product of the original spectrum $S(f)$ and a cosine function, this results in frequency selective fading.

4.2.3 Doppler effect

When MS moves, it causes the distance between MS and BS to change accordingly (unless the MS moves around a circle with BS being in the origin). For a radio wave, the change in distance along the propagation path causes a change in phase of the wave. The phase change is a function of the MS velocity, and makes the carrier frequency change. This frequency shift is named Doppler shift [1, 3].



(a) *MS movement* (b) *Doppler spectrum*

Figure 4.3: Doppler Effect

For instance, MS moves at a constant velocity v, as shown in Figure 4.3 (a). Assume BS is far away, so that the angle θ between the MS moving direction and the propagation path can be considered constant for a short time Δt. During Δt, the distance between MS and BS changes $\Delta l = \Delta t\ v \cos\theta$. Δl leads to a phase change of $\Delta\phi = 2\pi\Delta l/\lambda = 2\pi\Delta t\ v\cos\theta/\lambda$ for the carrier signal.

This phase change results in a frequency to be shifted by (4.6) [1]:

$$f_d = \frac{1}{2\pi}\frac{\Delta\phi}{\Delta t} = \frac{v}{\lambda}\cos\theta \tag{4.6}$$

where λ is the wavelength. Clearly:

$$\begin{cases} f_d > 0 & if\ \theta < 90^o \\ f_d < 0 & if\ 90^o < \theta < 180^o \end{cases}$$

If the carrier frequency is f_c, the Doppler shift expands the spectrum to the range from f_c - f_d to $f_c + f_d$. This range is called the Doppler spread. The Doppler effect introduces an irreducible error floor.

The Doppler spectrum is shown in Figure 4.3 (b), and is given as:

$$S_d(f) = \frac{1}{\sqrt{1 - \left(\frac{f - f_c}{f_d}\right)^2}} \qquad f_c - f_d < f < f_c + f_d \tag{4.7}$$

4.3 Signal Model in Multi-path Channel with AWGN

As discussed before, the mobile radio channel is random and time-variant. On the receiver side, radio wave is sampled at some sample rate. For band-limited signal, the minimum sample rate is twice the highest frequency, which is known as the Nyquist rate. It means the analog signal can be reconstructed from the discrete signal without losing any information (refer to page 72-73 in [2]). As described in Section 3.6, GMSK has infinite bandwidth, but most of the energy is contained in the main lobe. So GSM receiver normally utilizes two times oversampling to recover most of the energy from the signal. Oversampling factor is denoted as N_{sps} in this thesis. After receiver sampling, the multi-path channel can be modeled as a multi-tap, time varying Finite-duration Impulse Response (FIR) filter, which is shown in Figure 4.4.

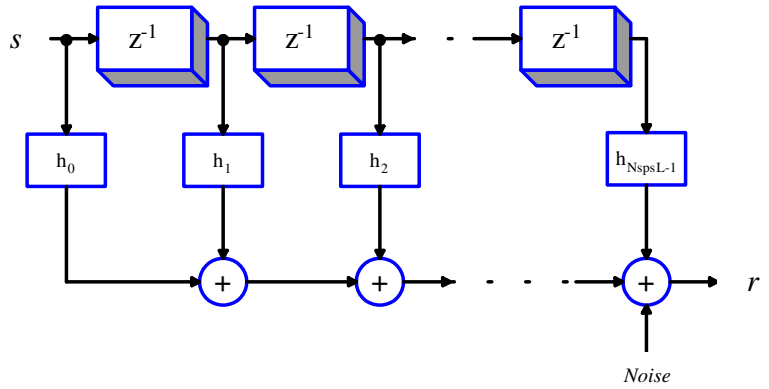


Figure 4.4: FIR filter structure to model multi-path channel with AWGN

It is an expansion of the single path channel model. h denotes the complex filter coefficients. z^{-1} denotes time delay. This delay can be of any value, and is normally quantized to sample duration $\frac{T_s}{N_{sps}}$, as delays less than $\frac{T_s}{N_{sps}}$ can be described by a phase shift.

Each channel parameter h is determined by the ever changing surrounding objects, thus it is usually

modeled as a wide sense stationary random process. If a signal source fulfills the following conditions, it is called wide sense stationary process [2]:

- the mean value is independent of time (a constant)
- the autocorrelation function satisfies:

$$\phi(t_1, t_2) = \phi(t_1 - t_2) \tag{4.8}$$

The statistical characteristics of wide sense stationary process can be applied in order to derive the channel parameters. As just discussed, filter coefficients h can be modeled as Rayleigh distributed or Ricean distributed. If the FIR filter has length L, the signal model can be written as:

$$\mathbf{r}_n = \sum_{l=0}^{L-1} \mathbf{h}_l s_{n-l} + \boldsymbol{\epsilon}_n \tag{4.9}$$

where n denotes time index. $\mathbf{r}$ is the received signal, $\mathbf{h}$ is the overall channel impulse response, including pulse shaping from the modulator. s is the transmitted symbol. ϵ is independent identically distributed (i.i.d.) AWGN with zero mean and variance σ^2. The reason why $\mathbf{r}$, $\mathbf{h}$ and $\boldsymbol{\epsilon}$ are in vector format is receiver oversampling.

$$\mathbf{r}_n = \left[r_{(n-1)N_{sps}+1}, r_{(n-1)N_{sps}+2}, \ldots, r_{nN_{sps}}\right]^T \qquad \mathbf{r}_n \in \mathbb{C}^{N_{sps}\times 1}$$

$$\mathbf{h}_l = \left[h_{lN_{sps}}, h_{lN_{sps}+1}, \ldots, h_{lN_{sps}-1}\right]^T \qquad \mathbf{h}_l \in \mathbb{C}^{N_{sps}\times 1}$$

$$\boldsymbol{\epsilon}_n = \left[\epsilon_{(n-1)N_{sps}+1}, \epsilon_{(n-1)N_{sps}+2}, \ldots, \epsilon_{nN_{sps}}\right]^T \qquad \boldsymbol{\epsilon}_n \in \mathbb{C}^{N_{sps}\times 1}$$

For MUD receiver, the signal model in (4.9) is expanded to cover more users. If there are totally K users, using superscript (k) to denote the $k-th$ user, the observation at time index n can be expressed as:

$$\mathbf{r}_n = \sum_{l=0}^{L-1} \sum_{k=1}^{K} \mathbf{h}_l^{(k)} s_{n-l}^{(k)} + \boldsymbol{\epsilon}_n \tag{4.10}$$

Further, let $\mathbf{r}$ hold all of the received samples, $\mathbf{s}^{(k)}$ for the k-th user's burst and $\mathbf{s}$ for all K bursts. $\mathrm{N_s}$ is the burst length and L is the channel impulse response length:

$$\mathbf{r} = \begin{bmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \\ \vdots \\ \mathbf{r}_{N_s+L-1} \end{bmatrix} \qquad \mathbf{s}^{(k)} = \begin{bmatrix} s_1^{(k)} \\ s_2^{(k)} \\ \vdots \\ s_{N_s}^{(k)} \end{bmatrix} \qquad \mathbf{s} = \begin{bmatrix} \mathbf{s}^{(1)} \\ \mathbf{s}^{(2)} \\ \vdots \\ \mathbf{s}^{(K)} \end{bmatrix}$$

Sometimes it is convenient to use matrix notation instead of sums, thus we also derive it here. Build the channel impulse response matrix $\mathbf{B}$ as:

$$\mathbf{B} = \left[\begin{array}{cccc} \mathbf{h}_0^{(1)}, \mathbf{h}_1^{(1)}, \cdots, \mathbf{h}_{L-1}^{(1)}, & \mathbf{h}_0^{(2)}, \mathbf{h}_1^{(2)}, \cdots, \mathbf{h}_{L-1}^{(2)}, \cdots, & \mathbf{h}_0^{(K)}, \mathbf{h}_1^{(K)}, \cdots, \mathbf{h}_{L-1}^{(K)} \end{array} \right]$$

$$\mathbf{h}_l^{(k)} \in \mathbb{C}^{N_{sps} \times 1} \qquad\qquad l \in [0, 1, \cdots, L-1], \qquad k \in [1, 2, \cdots, K]$$

$\mathbf{B}^{(k)}$ is a sub-matrix of $\mathbf{B}$ for the k-th user:

$$\mathbf{B}^{(k)} = \left[\mathbf{h}_0^{(k)}, \mathbf{h}_1^{(k)}, \cdots, \mathbf{h}_{L-1}^{(k)} \right]$$

Also stack symbols in vector $\mathbf{d}_n$ as follows:

$$\mathbf{d}_n = \left[\begin{array}{cccc} s_n^{(1)}, s_{n-1}^{(1)}, \cdots, s_{n-L+1}^{(1)}, & s_n^{(2)}, s_{n-1}^{(2)}, \cdots, s_{n-L+1}^{(2)}, \cdots, & s_n^{(K)}, s_{n-1}^{(K)}, \cdots, s_{n-L+1}^{(K)} \end{array} \right]^T$$

Similarly, $\mathbf{d}_n^{(k)}$ is a sub-vector of $\mathbf{d}_n$:

$$\mathbf{d}_n^{(k)} = \left[s_n^{(k)}, s_{n-1}^{(k)}, \cdots, s_{n-L+1}^{(k)}, \right]^T$$

The matrix notation of the signal model in (4.10) is then expressed as:

$$\mathbf{r}_n = \mathbf{B}\mathbf{d}_n + \boldsymbol{\epsilon}_n \tag{4.11}$$

(4.11) is identical to (4.10), but in matrix notations. When $K = 1$, single user signal model in (4.9) can be expressed in matrix format as:

$$\mathbf{r}_n = \mathbf{B}^{(1)}\mathbf{d}_n^{(1)} + \boldsymbol{\epsilon}_n \tag{4.12}$$

From (4.10), it is clear that the observation $\mathbf{r}_n$ is not only determined by s_n, but also the previous ones. The presence of ISI introduces correlation between symbols, therefore an equalizer is needed to remove the correlation and detect the symbols. With the signal model derived in this chapter, it is possible to discuss receiver algorithms in Chapter 5, 6 and 7.

4.4 Summary

This chapter discusses radio channel characteristics and modeling. Single path AWGN channel is introduced first. In this example, received signal is only corrupted by white Gaussian noise.

Reflection, diffraction and scattering of wave form cause more than one path between transmitter and receiver. Multi-path channel is a more realistic model for mobile radio transmission. Multi-path channel can be categorized into large scale path loss and small scale fading. Large scale path loss means received signal power decreases as a function of the distance between transmitter and receiver. Fading is due to the multi-path transmission, that is the sum of received signal from different paths may be constructive or destructive. Fading channel can be modeled as Rayleigh distributed or Rice distributed if there is an individual dominant component. Besides, delay spread and Doppler shift are analyzed in both time and frequency domain. Depending on the channel condition, multi-path fading channel may be frequency selective or frequency flat.

In GSM standard, multi-path channels are classified into:

1. Typical Urban (TUx)
2. Rural Area (RAx)
3. Hilly Terrain (HTx)

Their profiles are available in Appendix C. x represents mobile speed. Simulations appearing later in this thesis are normally done in the TU channel. From the receiver's point of view, multi-path channel can be modeled as FIR filter, and signal models for both single user and multi-user in multi-path channel with AWGN are derived based on FIR filter structure. Signal model in matrix format is also given in this chapter.

Chapter 5

Conventional Receivers

In Chapter 4, channel characteristics and channel models are presented, and the signal models for both SUD and MUD are derived. This chapter introduces several conventional data detection methods.

Data detection means detecting from given observations what symbols have been transmitted. Knowledge such as modulation scheme and IQ correlations can be used as side information to assist data detection. Channel estimation is needed in order to perform data detection. Problems of channel estimation and data detection can be solved together or separately. For instance, the Expectation Maximization (EM) technique solves these two problems together [19, 20]. In the E-step, estimate symbols with the estimated channel parameters from M-step. In the M-step, channel parameters are re-estimated given the estimated symbols from E-step. In this thesis, separate data detection methods are investigated, since the algorithms derived here can normally be integrated into the EM algorithm.

The following two chapters focus on data detection and channel parameters are therefore assumed available. A brief introduction of how to estimate channel those is given in Appendix A. For a known channel data detection concerns estimating symbols from received samples. As discussed in Chapter 3 and 4, modulation scheme and channel effects cause ISI. It means the received sample at time n is determined by the symbol transmitted as time n as well as several previous symbols. Thus, an equalizer is needed to compensate ISI. The ideal equalizer would cancel ISI completely. However, due to the existence of noise the equalizer can not achieve perfect performance, there is always some residual ISI and some small tracking error.

This chapter presents three types of conventional GSM receivers:

1. Matched Filter - the optimal linear receiver for an AWGN channel without ISI

2. LMMSE - optimal linear equalizer for an ISI AWGN channel

3. Symbol-by-Sybmol (SbS) MAP - optimal non-linear symbol detector for an AWGN channel with ISI.

5.1 Matched Filter

In AWGN non-ISI channel, the received signal is only corrupted by White Gaussian noise. In this case, Matched Filter can achieve the optimal BER performance. If $\mathbf{g}_t$ is the transmission pulse shape, and $\mathbf{n}_n$ is the white noise sample, that is:

$$\mathbf{g}_t = [h_0^{(1)}, h_1^{(1)}, \ldots, h_{LN_{sps}-1}^{(1)}]^T \qquad \mathbf{n}_n = [\epsilon_n, \epsilon_{n+1}, \ldots, \epsilon_{n+LN_{sps}-1}]^T$$

Note that $h_l^{(1)}$ normally denotes tap of the overall channel impulse response including modulation pulse and channel transfer function in this thesis. But in AWGN non-ISI channel channel, if the transfer function is normalized to 1 and the phase change caused by the channel is disregarded from the discussion in this section, the overall channel impulse response is then the same as the modulation pulse shape. So the received signal $\mathbf{o}_n$ can be expressed as:

$$\mathbf{o}_n = \mathbf{g}_t s_n + \mathbf{n}_n \tag{5.1}$$

Matched filter coefficients $\mathbf{g}_r$ are the time-inverse and complex conjugated version of $\mathbf{g}_r$ (p. 238-244 in [2]), that is:

$$\mathbf{g}_r = [h_{LN_{sps}-1}^{(1)*}, h_{LN_{sps}-2}^{(1)*}, \ldots, h_0^{(1)*}]^T$$

Thus the output from Matched Filter appears:

$$\mathbf{s}_{n,\ MF} = \mathbf{g}_r^T \otimes \mathbf{r}_n^T = s_n^{(1)} \mathbf{g}_r^T \otimes \mathbf{g}_t^T + \mathbf{g}_r^T \otimes \mathbf{n}_n^T \tag{5.2}$$

The convolution $\mathbf{g}_r^T \otimes \mathbf{g}_t^T$ generates the highest power at lag LN_{sps}. Hence, if using hard decision, the estimated symbol $\breve{s}_n^{(1)}$ by Matched Filter can be found from the sign of the LN_{sps}-th element in $\mathbf{s}_{n,\ MF}^{(1)}$.

In GSM, GMSK modulation pulse itself generates ISI. Further, multi-path channel effects introduce more ISI. Hence for GMSK, BER performance of Matched Filter in AWGN non-ISI channel is quite close but not exactly the same as the theoretical BER. And it gets worse for multi-path channel with AWGN since ISI becomes worse. This analysis is proved by the simulation plots in Figure 5.4 and 5.5.

5.2 LMMSE Receiver

LMMSE is a linear data detector that gives better BER than Matched Filter when ISI presents. It can be viewed as a Wiener filter, which minimizes the mean squared estimation error. Mean squared estimation error serves as a cost function. If $\breve{s}_{w,\ n}$ denotes the Wiener filter output, the estimation error $e_{w,\ n}$ is:

$$e_{w,\,n} = s_n - \breve{s}_{w,\,n} \tag{5.3}$$

The mean squared error is then denoted as:

$$E\left[\mid e_{w,\,n}\mid^2\right]_n = E\left[|s_n - \breve{s}_{w,\,n}|^2\right]_n \tag{5.4}$$

$[\cdot]_n$ denotes average with respect to time. Filter coefficients of LMMSE receiver are obtained by minimizing the mean square error:

$$\mathbf{w}_o = \operatorname{argmin} E\left[|s_n - \breve{s}_{w,\,n}|^2\right]_n \tag{5.5}$$

If the Wiener filter $\mathbf{w}_o$ has length L_w, normally $L_w > L$, let $\mathbf{x}_n$ contain received samples, and $\mathbf{v}^{(k)}$ hold channel impulse response for the k-th user:

$$\mathbf{x}_n = \begin{bmatrix} \mathbf{r}_n \\ \mathbf{r}_{n+1} \\ \vdots \\ \mathbf{r}_{n+L_w-1} \end{bmatrix} \qquad \mathbf{v}^{(k)} = \begin{bmatrix} \mathbf{h}_0^{(k)} \\ \vdots \\ \mathbf{h}_{L-1}^{(k)} \\ 0 \\ \vdots \\ 0 \end{bmatrix} \qquad \mathbf{x}_n,\ \mathbf{v}^{(k)} \in \mathbb{C}^{L_w \times N_{sps}}$$

then the output of Wiener filter is:

$$\breve{s}_{w,\,n} = \mathbf{w}_o^T \mathbf{x}_n \tag{5.6}$$

If hard decision is used, the estimated symbol from the LMMSE receiver is:

$$\breve{s}_n = \operatorname{sgn}\Big(\breve{s}_{w,\,n}\Big) \tag{5.7}$$

R_{xx} is the covariance of $\mathbf{x}_n$ and it is given as (Section5.4 in [21]):

$$R_{xx} = E\left[\mathbf{x}_n \mathbf{x}_n^H\right]_n = \sum_{m=0}^{L_w} \sum_{k=1}^{K} \mathbf{v}^{(k)}(m)\mathbf{v}^{(k)}(m)^H + \sigma^2 \mathbf{I} \qquad m \in [0, 1, \ldots, L_w - 1] \tag{5.8}$$

where $\mathbf{v}^{(k)}(m)$ denotes shifting $\mathbf{v}^{(k)}$ up by m and pad zero. $\mathbf{I}$ is identity matrix and has dimension $L_w N_{sps}$ by $L_w N_{sps}$. σ^2 is noise variance.

$\mathbf{p}_{sx}$ is the cross correlation and given as:

$$\mathbf{p}_{sx} = E\left[s_n \mathbf{x}_n\right]_n = \mathbf{v}_n^{(1)} \tag{5.9}$$

LMMSE filter coefficients can be found by Wiener-Hopf equations (p. 203-206 in [3]):

$$\mathbf{w}_o = \mathbf{R}_{xx}^{-1} \mathbf{p}_{sx} \tag{5.10}$$

It is shown in [5, 22], that for any real-valued constellation and complex-value ISI channel impulse response, such as de-rotated GMSK, both $E[\mathbf{x}_n \mathbf{x}_n^H]_n$ and $E[\mathbf{x}_n \mathbf{x}_n^T]_n$ are needed to get sufficient statistics. Because $E[\mathbf{x}_n \mathbf{x}_n^T]_n$ is non-zero, include it can provide a gain in performance. However, if the transmitted symbols are complex, e.g. M-QAM, then $E[\mathbf{x}_n \mathbf{x}_n^T]_n$ can be shown zero and include it will not give a gain.

From (5.10), it is seen that LMMSE receiver only captures $E[\mathbf{x}_n \mathbf{x}_n^H]_n$. $E[\mathbf{x}_n \mathbf{x}_n^T]_n$ can be obtained by a modification to LMMSE receiver. That is to split received signal into real and imaginary parts and term it IQ LMMSE receiver. The implementation of IQ LMMSE is quite similar as a normal LMMSE receiver, but with IQ-split notations for vectors and matrices mentioned earlier. Let subscript I denote real part and Q denote imaginary one. Modify vectors and matrices as follows:

$$\mathbf{x}_{n,\ IQ} = \begin{bmatrix} \mathbf{x}_{n,\ I} \\ \mathbf{x}_{n,\ Q} \end{bmatrix} \quad \mathbf{w}_{o,\ IQ} = \begin{bmatrix} \mathbf{w}_{o,\ I} \\ \mathbf{w}_{o,\ Q} \end{bmatrix} \quad \mathbf{p}_{sx,\ IQ} = \begin{bmatrix} \mathbf{p}_{sx,\ I} \\ \mathbf{p}_{sx,\ Q} \end{bmatrix} \quad \mathbf{R}_{xx,\ IQ} = \begin{bmatrix} \mathbf{R}_{x_I x_I} & \mathbf{R}_{x_I x_Q} \\ \mathbf{R}_{x_Q x_I} & \mathbf{R}_{x_Q x_Q} \end{bmatrix}$$

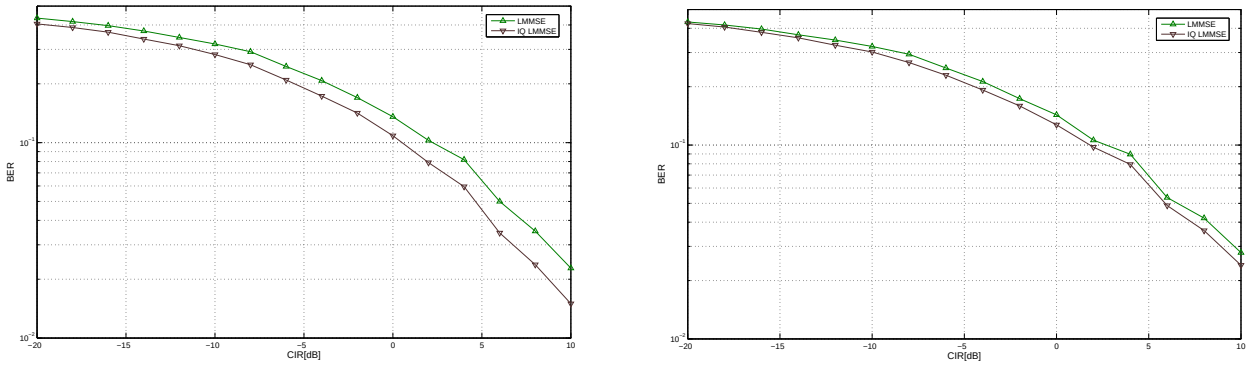
IQ LMMSE filter coefficients are obtained by:

$$\mathbf{w}_{o,\ IQ} = \mathbf{R}_{xx,\ IQ}^{-1} \mathbf{p}_{sx,\ IQ} \tag{5.11}$$

The output for IQ LMMSE is then:

$$\breve{s}_{w,\ IQ,\ n} = \mathbf{w}_{o,\ IQ}^T \mathbf{x}_{n,\ IQ} \tag{5.12}$$

Simulation for LMMSE and IQ LMMSE receivers is shown in Figure 5.1. Note that, IQ LMMSE equalizer requires to calculate the inversion of matrix $\mathbf{R}_{xx,\ IQ}$, which is twice as big as $\mathbf{R}_{xx}$, therefore the complexity of IQ LMMSE receiver is at least doubled compared to LMMSE receiver. Moreover, IQ splitting is a non-linear process, so IQ LMMSE is not a linear receiver anymore. It can also be understood as the Linear Conjugate Linear (LCL) technique. It has two filters, one for received signal and another one for conjugated received signal.



(a) *Perfect channel parameter* (b) *JLS channel estimation with length* 5

Figure 5.1: Simulation of LMMSE/IQ LMMSE receivers, filter length = 11, SNR = 20 dB, GMSK modulation, 1 CCI, TU3 channel

5.3 Symbol-by-Symbol MAP Equalizer for SUD

In section 5.2, the optimum linear equalizer - LMMSE is discussed. For GMSK modulated signals, BER performance can be improved by IQ splitting. However, this is a non-linear process and approximately doubles complexity compared to LMMSE.

BER in an AWGN ISI channel can be further improved by another type of non-linear signal processing, which models the signal coming from a first order Hidden Markov source [15]. SbS MAP equalizer is based on this model, besides there are also MLSE and ML equalizer. So the difference and similarity among SbS MAP, ML and MLSE are analyzed first. Then the derivation of SbS MAP equalizer is presented. Optimum MUD equalizer is developed from SbS MAP for SUD, which is discussed in Chapter 6.

5.3.1 Comparison of MLSE, Symbol-by-Symbol MAP and ML

First, we look into the difference between SbS MAP and MLSE. MLSE is an optimal solution in terms of minimizing the probability of word error for convolutional codes, invented by Viterbi in 1967 and usually called the Viterbi Algorithm (VA) [23]. MLSE minimizes sequence error by maximizing likelihood of observation $\mathbf{r}$ for unknown sequence $\mathbf{s}$. The sequence estimated by MLSE is given as:

$$\breve{\mathbf{s}} = \underset{\mathbf{s}\in[-1,1]^{N_s}}{\operatorname{argmax}} P(\mathbf{r} \mid \mathbf{s}) \tag{5.13}$$

However, MLSE does not necessarily minimize symbol error probability. SbS MAP minimizes the symbol error probability given received samples, and can be implemented by the algorithm of Bahl, Cocke, Jelinek, and Raviv (BCJR) [24]. If hard decision is applied, the estimated symbol by SbS MAP equalizer is given as:

$$\breve{s}_n = \underset{s_n\in[-1,1]}{\operatorname{argmax}} P(s_n \mid \mathbf{r}) \tag{5.14}$$

The main difference between MLSE and SbS MAP is that MLSE gives optimum hard sequence estimation whereas SbS MAP is optimum for Symbol-by-Sybmol soft decisions. It means BER of SbS MAP is better than MLSE, and sequence error of SbS MAP may be worse than MLSE.

Now, let's find out the similarity between SbS MAP and ML. If Bayes' rule (ref. p.245 in [2]) is applied to (5.14):

$$P(s_n \mid \mathbf{r}) = \frac{P(\mathbf{r} \mid s_n)P(s_n)}{P(\mathbf{r})} \tag{5.15}$$

The ML equalizer is given as:

$$\breve{s}_n = \underset{\mathbf{s}_n \in [-1,1]}{\operatorname{argmax}} P(\mathbf{r} \mid s_n) \tag{5.16}$$

In (5.15), $P(\mathbf{r})$ is a constant for a given $\mathbf{r}$. Compare (5.14) and (5.16), (5.15) shows that SbS MAP and ML are the same for uniformly distributed $P(s_n)$, sometimes it is written as ML/MAP equalizer.

To summarize the comparison among MLSE, SbS MAP and ML equalizers, the following conclusions are given:

1. MLSE is the optimal hard decision sequence estimator, which is implemented by VA algorithm.
2. SbS MAP is the optimal soft decision symbol estimator and can be implemented by BCJR algorithm.
3. SbS MAP and ML are the same for equal probable input.
4. $BER_{SbS\ MAP} < BER_{MLSE}$, and $BER_{SbS\ MAP} = BER_{ML}$ for uniformly distributed symbol.

MLSE is of great interest when decoding channel codes because these are convolutional codes. SbS The SbS MAP equalizer for SUD has received increasing interest since Turbo codes were introduced in 1993. This project does not cover the channel decoder part, symbols are assumed independent and uniformly distributed, hence the derivation of SbS MAP for SUD is discussed next.

5.3.2 Symbol-by-Symbol MAP equalizer

In (5.14), the problem of maximizing the probability of symbols given the received signal in an AWGN ISI channel can be modeled as a signal coming from a first order Hidden Markov source [15]. In 1974, Bahl, Cocke, Jelinek and Raviv proposed the SbS MAP algorithm for minimizing symbol error probability [24]. Recall single user signal model in (4.9), for a channel with memory, the observation $\mathbf{r}_n$ depends on s_n as well as the previous $L-1$ symbols. Therefore, the first-order Markov state, at time n, can be defined from the previous $L-1$ symbols, say state vector $\mathbf{a}_n$ is formed as:

$$\mathbf{a}_n = \left[\hat{s}_{n-1}, \hat{s}_{n-2}, \cdots, \hat{s}_{n-L+1}\right]^T \tag{5.17}$$

where $\hat{s}_n$ is the hypothesized value of s_n from constellation Ω; for GMSK we have $\Omega = [\pm 1]$. The subscript n is the time index. There are $Q =\mid \Omega \mid^{L-1}$ different combinations of hypothesized values, numbered from 0 to $Q-1$. So (5.17) can be rewritten as:

$$\mathbf{a}_n(p) = \left[\hat{s}_{n-1}(p), \hat{s}_{n-2}(p), \cdots, \hat{s}_{n-L+1}(p)\right]^T \qquad p \in [0, 1, \cdots, Q-1] \tag{5.18}$$

$\hat{s}_{n-1}(p)\ \cdots\ \hat{s}_{n-L+1}(p)$ are the hypothesized symbol values for state $\mathbf{a}_n\ (p)$. An example for $L = 3$, $Q =\mid \Omega \mid^{L-1}= 4$ is given in Table 5.1.

$\mathbf{a}_n\ (p)$	$[\hat{s}_{n-1}\ (p)\ ,\ \hat{s}_{n-2}\ (p)\ ]$
$\mathbf{a}_n\ (0)$	$[-1, -1]$
$\mathbf{a}_n\ (1)$	$[-1, +1]$
$\mathbf{a}_n\ (2)$	$[+1, -1]$
$\mathbf{a}_n\ (3)$	$[+1, +1]$

Table 5.1: Example of state vector for $L = 3$

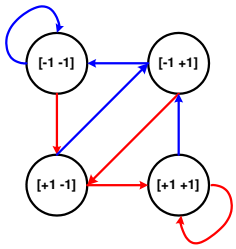


Figure 5.2: State transition diagram for the example in Table 5.1

Similar to (5.18), state $\mathbf{a}_{n+1}(q)$ is defined as:

$$\mathbf{a}_{n+1}(q) = \left[\hat{s}_n(q), \hat{s}_{n-1}(q), \cdots, \hat{s}_{n-L+2}(q)\right]^T \qquad q \in [0, 1, \cdots, Q-1] \tag{5.19}$$

Figure 5.2 is a diagram of possible transitions between states, blue arrow for input -1, and red for $+1$. With the state vectors defined in Table 5.1, a trellis diagram for input $[+1, +1, -1, +1, -1]$ is shown in Figure 5.3. A black dot denotes the state vector, lines between the states denote transitions. Comparing (5.18) and (5.19), we can see that $\hat{s}_n(q)$ determines the transition path in the trellis diagram for a specific $\mathbf{a}_n(p)$. Since $\hat{s}_n(q) \in \Omega$, so there are $\mid \Omega \mid$ possible transition paths from $\mathbf{a}_n(p)$ to $\mathbf{a}_{n+1}(q)$. This is also shown in Figure 5.3 that there are $\mid \Omega \mid$ possible parent states for a specific state $\mathbf{a}_{n+1}(q)$.

We now, define a vector $\boldsymbol{\xi}_n^{(p,\ q)}$, which holds all the hypothesized symbols associated with the transition $\mathbf{a}_n\ (p) \rightarrow \mathbf{a}_{n+1}\ (q)$, $\boldsymbol{\xi}_n^{(p,\ q)}$ has dimension L. Let $(p,\ q)$ stand for the transition $\mathbf{a}_n\ (p) \rightarrow \mathbf{a}_{n+1}\ (q)$.

$$\boldsymbol{\xi}_n^{(p,\ q)} = \left[\hat{s}_n(q), \hat{s}_{n-1}(p), \cdots, \hat{s}_{n-L+1}(p)\right]^T \tag{5.20}$$

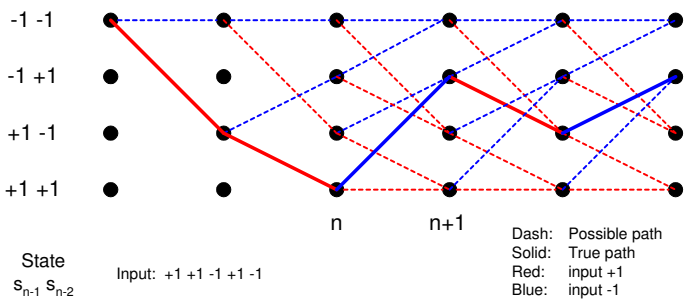


Figure 5.3: 4-states trellis

Let $\mathcal{A}$ hold the set of transitions $(p,\ q)$ where $\hat{s}_n(q)$ takes on the specific hypothesized value $\widetilde{s}_n$. From the signal model in (4.9), the posterior probability given the observation $\mathbf{r}$ is [25]:

$$\begin{aligned}
P(\widetilde{s}_n \mid \mathbf{r}) &= \sum_{(p,\ q)\in\mathcal{A}} P\Big(\mathbf{a}_n(p),\ \mathbf{a}_{n+1}(q) \mid \mathbf{r}\Big) \\
&= \frac{1}{P(\mathbf{r})} \sum_{(p,\ q)\in\mathcal{A}} P\Big(\mathbf{a}_n(p),\ \mathbf{a}_{n+1}(q),\ \mathbf{r}\Big) \\
&= \frac{1}{P(\mathbf{r})} \sum_{(p,\ q)\in\mathcal{A}} P\Big(\mathbf{a}_n(p),\ \mathbf{a}_{n+1}(q),\ \mathbf{r}_1^{n-1},\ \mathbf{r}_n,\ \mathbf{r}_{n+1}^{N_s+L-1}\Big) \\
&= \frac{1}{P(\mathbf{r})} \sum_{(p,\ q)\in\mathcal{A}} P\Big(\mathbf{a}_n(p),\ \mathbf{a}_{n+1}(q),\ \mathbf{r}_1^{n-1},\ \mathbf{r}_n\Big) P\Big(\mathbf{r}_{n+1}^{N_s+L-1} \mid \mathbf{a}_n(p),\ \mathbf{a}_{n+1}(q),\ \mathbf{r}_1^{n-1},\ \mathbf{r}_n\Big) \\
&= \frac{1}{P(\mathbf{r})} \sum_{(p,\ q)\in\mathcal{A}} P\Big(\mathbf{a}_n(p),\ \mathbf{r}_1^{n-1}\Big) P\Big(\mathbf{a}_{n+1}(q),\ \mathbf{r}_n \mid \mathbf{a}_n(p),\ \mathbf{r}_1^{n-1}\Big) \cdot \\
&\qquad\qquad P\Big(\mathbf{r}_{n+1}^{N_s+L-1} \mid \mathbf{a}_n(p),\ \mathbf{a}_{n+1}(q),\ \mathbf{r}_1^{n-1},\ \mathbf{r}_n\Big)
\end{aligned} \tag{5.21}$$

where

$$\mathbf{r}_{n_1}^{n_2} = \begin{bmatrix} \mathbf{r}_{n_1} \\ \vdots \\ \mathbf{r}_{n_2} \end{bmatrix} \qquad n_1, n_2 \in [1, 2, \cdots, N_s + L - 1] \tag{5.22}$$

Using the property of a first order Hidden Markov source [25, 15], knowledge of $\mathbf{a}_{n+1}$ accounts for the dependence of its parent $\mathbf{a}_n$ and observation $\mathbf{r}_1^n$. Similarly, knowledge of $\mathbf{a}_n$ also accounts for the dependence of $\mathbf{r}_1^{n-1}$, (5.21) is reduced to:

$$P\Big(\widetilde{s}_n \mid \mathbf{r}\Big) = \frac{1}{P(\mathbf{r})} \sum_{(p,\ q)\in\mathcal{A}} P\Big(\mathbf{a}_n(p),\ \mathbf{r}_1^{n-1}\Big) P\Big(\mathbf{a}_{n+1}(q),\ \mathbf{r}_n \mid \mathbf{a}_n(p)\Big) P\Big(\mathbf{r}_{n+1}^{N_s+L-1} \mid \mathbf{a}_{n+1}(q)\Big) \tag{5.23}$$

We assign three new variables for the three probability terms in (5.23). Forward variable $\alpha_n(p)$, which accounts for the information from the past to time n,

$$\alpha_n(p) = P\Big(\mathbf{a}_n(p),\ \mathbf{r}_1^{n-1}\Big) \tag{5.24}$$

Transition variable $\gamma_n(p,\ q)$ that determines the probability for a transition $(p,\ q)$, that is:

$$\gamma_n(p,\ q) = P\Big(\mathbf{a}_{n+1}(q),\ \mathbf{r}_n \mid \mathbf{a}_n(p)\Big) \tag{5.25}$$

Backward variable $\beta_{n+1}(q)$, which accounts for information from the future back to time $n+1$:

$$\beta_{n+1}(q) = P\Big(\mathbf{r}_{n+1}^{N_s+L-1} \mid \mathbf{a}_{n+1}(q)\Big) \tag{5.26}$$

(5.23) can be rewritten as:

$$P\big(\widetilde{s}_n \mid \mathbf{r}\big) = \frac{1}{P(\mathbf{r})} \sum_{(p,\ q)\in\mathcal{A}} \alpha_n(p)\gamma_n(p,\ q)\beta_{n+1}(q) \tag{5.27}$$

The forward and backward variables can be calculated recursively. For $\alpha_{n+1}(q)$:

$$\begin{aligned}
\alpha_{n+1}(q) &= P\Big(\mathbf{a}_{n+1}(q),\ \mathbf{r}_1^n\Big) = P\Big(\mathbf{a}_{n+1}(q),\ \mathbf{r}_1^{n-1},\ \mathbf{r}_n\Big) \\
&= \sum_{p\in Q} P\Big(\mathbf{a}_{n+1}(q),\ \mathbf{a}_n(p),\ \mathbf{r}_1^{n-1},\ \mathbf{r}_n\Big) \\
&= \sum_{p\in Q} P\Big(\mathbf{a}_n(p),\ \mathbf{r}_1^{n-1}\Big) P\Big(\mathbf{a}_{n+1}(q),\ \mathbf{r}_n \mid \mathbf{a}_n(p),\ \mathbf{r}_1^{n-1}\Big) \\
&= \sum_{p\in Q} P\Big(\mathbf{a}_n(p),\ \mathbf{r}_1^{n-1}\Big) P\Big(\mathbf{a}_{n+1}(q),\ \mathbf{r}_n \mid \mathbf{a}_n(p)\Big) \\
&= \sum_{p\in Q} \alpha_n(p)\gamma_n(p,\ q)
\end{aligned} \tag{5.28}$$

Similarly, $\beta_n(p)$ can also be calculated recursively, that is:

$$\begin{aligned}
\beta_n(p) &= P\Big(\mathbf{r}_n^{N_s+L-1} \mid \mathbf{a}_n(p)\Big) = P\Big(\mathbf{r}_n,\ \mathbf{r}_{n+1}^{N_{sps}+L-1} \mid \mathbf{a}_n(p)\Big) \\
&= \sum_{q\in Q} P\Big(\mathbf{r}_n,\ \mathbf{r}_{n+1}^{N_s+L-1},\ \mathbf{a}_{n+1}(q) \mid \mathbf{a}_n(p)\Big) \\
&= \sum_{q\in Q} P\Big(\mathbf{r}_{n+1}^{N_s+L-1} \mid \mathbf{a}_{n+1}(q),\ \mathbf{a}_n(p),\ \mathbf{r}_n\Big) P\Big(\mathbf{a}_{n+1}(q),\ \mathbf{r}_n \mid \mathbf{a}_n(p)\Big) \\
&= \sum_{q\in Q} P\Big(\mathbf{r}_{n+1}^{N_{sps}+L-1} \mid \mathbf{a}_{n+1}(q)\Big) P\Big(\mathbf{a}_{n+1}(q),\ \mathbf{r}_n \mid \mathbf{a}_n(p)\Big) \\
&= \sum_{q\in Q} \beta_{n+1}(q)\gamma_n(p,\ q)
\end{aligned} \tag{5.29}$$

Since α and β can be calculated recursively, the question left to solve is to derive $\gamma_n(p,\ q)$. Recall (5.25):

$$\begin{aligned}\gamma_n(p,\ q) &= P\Big(\mathbf{a}_{n+1}(q),\ \mathbf{r}_n \mid \mathbf{a}_n(p)\Big) \\ &= P\Big(\mathbf{r}_n \mid \mathbf{a}_{n+1}(q),\ \mathbf{a}_n(p)\Big) P\Big(\mathbf{a}_{n+1}(q) \mid \mathbf{a}_n(p)\Big)\end{aligned} \tag{5.30}$$

The first term in (5.30) is the likelihood of $\mathbf{r}_n$ for the transition $(p,\ q)$. For White Gaussian noise, with zero mean and variance σ^2, likelihood of observations $\mathbf{r}_n$ given $\mathbf{a}_{n+1}(q)$ and $\mathbf{a}_n(p)$ is:

$$P\Big(\mathbf{r}_n \mid \mathbf{a}_{n+1}(q),\ \mathbf{a}_n(p)\Big) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^{N_{sps}} \exp\frac{-1}{2\sigma^2}\Big(\mathbf{r}_n - \widetilde{\mathbf{r}}_n^{(p,\ q)}\Big)^H\Big(\mathbf{r}_n - \widetilde{\mathbf{r}}_n^{(p,\ q)}\Big) \tag{5.31}$$

where $\widetilde{\mathbf{r}}_n^{(p,\ q)}$ denotes the output for $\boldsymbol{\xi}_n^{(p,\ q)}$, by the matrix notation signal model in (4.11), $\widetilde{\mathbf{r}}_n^{(p,\ q)}$ can be expressed as:

$$\widetilde{\mathbf{r}}_n^{(p,\ q)} = \mathbf{B}\boldsymbol{\xi}_n^{(p,\ q)} \tag{5.32}$$

Further using definitions of $\mathbf{a}_{n+1}(q)$ and $\mathbf{a}_n(p)$ in (5.18) (5.19), the second term in (5.30) is:

$$\begin{aligned}P\Big(\mathbf{a}_{n+1}(q) \mid \mathbf{a}_n(p)\Big) &= \frac{P\Big(\mathbf{a}_{n+1}(q),\ \mathbf{a}_n(p)\Big)}{P\big(\mathbf{a}_n(p)\big)} = \frac{P\Big(\mathbf{a}_n(p),\ \hat{s}_n(q)\Big)}{P\big(\mathbf{a}_n(p)\big)} \\ &= \frac{P\Big(\mathbf{a}_n(p) \mid \hat{s}_n(q)\Big)P\big(\hat{s}_n(q)\big)}{P\big(\mathbf{a}_n(p)\big)}\end{aligned} \tag{5.33}$$

Note that, from the definition of state vector $\mathbf{a}_n(p)$ in (5.18), $\mathbf{a}_n(p)$ is independent of $\hat{s}_n(q)$, thus $P\Big(\mathbf{a}_n(p) \mid \hat{s}_n(q)\Big) = P\big(\mathbf{a}_n(p)\big)$, (5.33) is then reduced to:

$$P\Big(\mathbf{a}_{n+1}(q) \mid \mathbf{a}_n(p)\Big) = \frac{P\big(\mathbf{a}_n(p)\big)P\big(\hat{s}_n(q)\big)}{P\big(\mathbf{a}_n(p)\big)} = P\big(\hat{s}_n(q)\big) \tag{5.34}$$

Accordingly, (5.30) is reduced to:

$$\gamma_n(p,\ q) = P\Big(\mathbf{r}_n \mid \mathbf{a}_{n+1}(q),\ \mathbf{a}_n(p)\Big)P\big(\hat{s}_n(q)\big) \tag{5.35}$$

With (5.35), rewriting (5.27):

$$\begin{aligned} P\big(\widetilde{s}_n \mid \mathbf{r}\big) &= \frac{1}{P(\mathbf{r})} \sum_{(p,\ q)\in\mathcal{A}} \alpha_n(p)\gamma(p,\ q)\beta_{n+1}(q) \\ &= \frac{1}{P(\mathbf{r})} \sum_{(p,\ q)\in\mathcal{A}} \alpha_n(p)\beta_{n+1}(q) P\Big(\mathbf{r}_n \mid \mathbf{a}_{n+1}(q),\ \mathbf{a}_n(p)\Big) P\big(\hat{s}_n(q)\big) \\ &= \frac{P(\widetilde{s}_n)}{P(\mathbf{r})} \sum_{(p,\ q)\in\mathcal{A}} \alpha_n(p)\beta_{n+1}(q) P\Big(\mathbf{r}_n \mid \mathbf{a}_{n+1}(q),\ \mathbf{a}_n(p)\Big) \end{aligned} \tag{5.36}$$

The last equal is true because the summation is over all $(p,\ q) \in \mathcal{A}$ and $\mathcal{A}$ denotes all the transitions that correspond to $\hat{s}_n(q) = \widetilde{s}_n$. $P(\widetilde{s}_n)$ is a prior of symbol. For equal probable symbol, e.g. like in GSM, $P(\widetilde{s}_n)$ doesn't influence the maximization procedure. Moreover, $P(\mathbf{r})$ doesn't affect this procedure neither since it is a constant for a given $\mathbf{r}$. So $P(\widetilde{s}_n)$ and $P(\mathbf{r})$ are normally discarded to make it simple.

Further using (5.31) and (5.34) to (5.30), we have:

$$\gamma_n(p,\ q) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^{N_{sps}} \exp\frac{-1}{2\sigma^2}\big(\mathbf{r}_n - \widetilde{\mathbf{r}}_n^{(p,\ q)}\big)^H\big(\mathbf{r}_n - \widetilde{\mathbf{r}}_n^{(p,\ q)}\big) P\big(\hat{s}_n(q)\big) \tag{5.37}$$

The complexity of SbS MAP equalizer is $\mathcal{O}(|\Omega|^L)$. From the likelihood term given in (5.31), we can see that the SbS MAP algorithm requires lots of exponential operations, it can be conveniently implemented in the logarithm domain to avoid exponential calculations. It can be further simplified under the assumption that the sum of the noise and interference is Gaussian, the approximation given in (5.38) can be applied, and the algorithm is then normally called Max-log-MAP equalizer. However the total operations remain the same, i.e. $\mathcal{O}(|\Omega|^L)$.

$$\ln(\mathrm{e}^{\alpha_1} + \mathrm{e}^{\alpha_2} + \ldots + \mathrm{e}^{\alpha_n}) \ \approx\ \max(\alpha_1, \alpha_2, \ldots, \alpha_n) \tag{5.38}$$

The approximation error is:

$$0 \ \le\ \ln(\mathrm{e}^{\alpha_1} + \mathrm{e}^{\alpha_2} + \ldots + \mathrm{e}^{\alpha_n}) \ - \max(\alpha_1, \alpha_2, \ldots, \alpha_n) \ \le\ \ln(n) \tag{5.39}$$

Max-log-MAP slightly suffers BER degradation compared to SbS MAP, i.e. $BER_{Max-log-MAP} > BER_{SbS\ MAP}$. If only using hard decision and only the forward path, Max-log-MAP gives the same BER as MLSE. That is $BER_{Max-log-MAP} = BER_{MLSE} > BER_{SbS\ MAP}$.

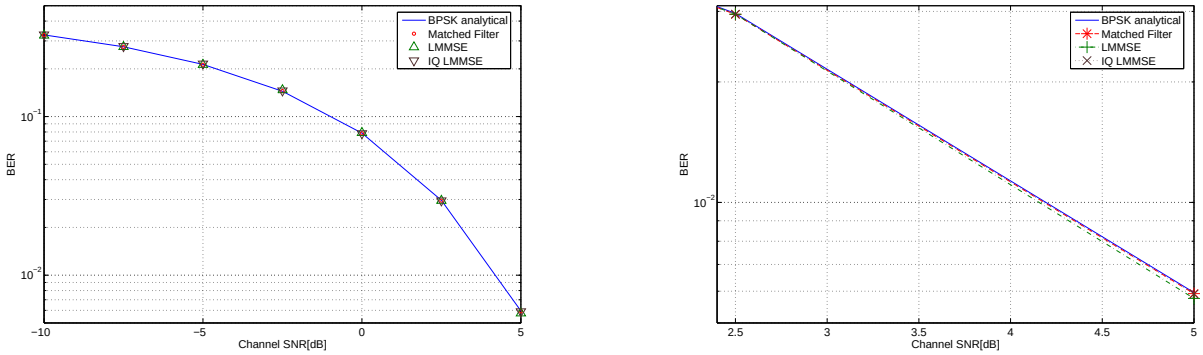
5.4 Simulation

Monte-Carlo simulations of these three GSM receivers are demonstrated in this section. Performance is evaluated by BER. Simulation is completed when 1000 errors are found or at least 1000 desired bursts are tested. The simulated burst is the GSM normal burst. Errors are counted for information

bits only. The tested conventional receivers are Matched Filter, LMMSE equalizer, IQ LMMSE equalizer and Log-Max-MAP equalizer.

Channel parameters are assumed known by the receiver and four test scenarios are chosen:

1. AWGN single path channel. GMSK modulation has the inherent ISI. In order to show Matched Filter is the optimal receiver for non ISI scenario, non ISI pulse is used instead of GMSK pulse in this test case. Since there is no coupling between bits, there is no need to show the performance of the SbS MAP equalizer for SUD in this test case. Simulations are plotted in Figure 5.4.
2. Static Multi-path channel with AWGN, no interference. Channel parameters are unchanged for the entire test. Channel taps are from instantaneous samples of TUx channel, i.e. $[-0.4113 + 0.9837i, 0.1177 - 0.4396i, 0.2541 - 0.0495i, -0.0348 + 0.2310i, 0, 0.1329 - 0.1217i]$, see Figure 5.5 (a).
3. TU50 channel, no interference. It is similar as the above test case, but the channel impulse response is changing from time to time. These test results are shown in Figure 5.5 (b).
4. TU3 channel. One CCI (GMSK) is included. BER performance is presented in Figure 5.6.



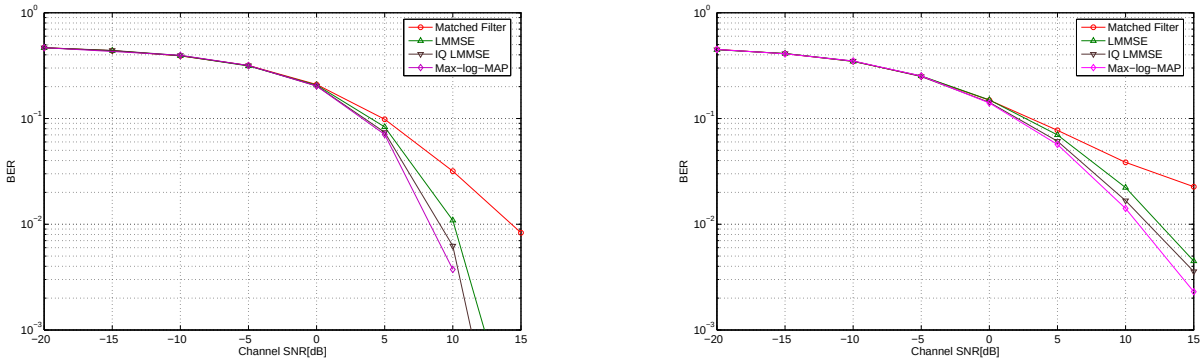
(a) SNR -10 dB *to* 5 dB (b) SNR -2.5 dB *to* 5 dB

Figure 5.4: Simulation of conventional receivers in AWGN single path channel, non ISI pulse, LMMSE and IQ LMMSE filter length=11, N_{sps}=1

The Matched Filter is optimal for non ISI type of modulation in AWGN single path channel. Hence a non ISI modulation pulse is chosen instead of GMSK in the first test case. The LMMSE and IQ LMMSE solutions are reduced to Matched Filter in this test case. Therefore all tested receivers achieve BPSK analytical BER performance, as seen in Figure 5.4.

For multi-path channel with AWGN, all receivers give worse BER than in AWGN single path channel. It is due to the fading effect from the channel. In this case, Matched Filter is not optimal anymore. Both LMMSE and IQ LMMSE receivers outperform the Matched Filter and the SbS MAP equalizer provides the best performance as shown in Figure 5.5. BER curves in Figure 5.5 (a) are tested in static channel, whereas BER curves in Figure 5.5 (a) are the averaged BER performance in time-varying multi-path fading channel with AWGN. That is why BER curves in Figure 5.5 (a) is better than those in Figure 5.5 (b).

In Figure 5.6, the x axis is changed from SNR to CIR because it is more interesting to evaluate the receiver performance for different CIR when interference is dominant. Thus SNR is set to a



(a) *Static* (b) $TU50$

Figure 5.5: Simulation of conventional receivers, GMSK, LMMSE and IQ LMMSE filter length=11, N_{sps}=2

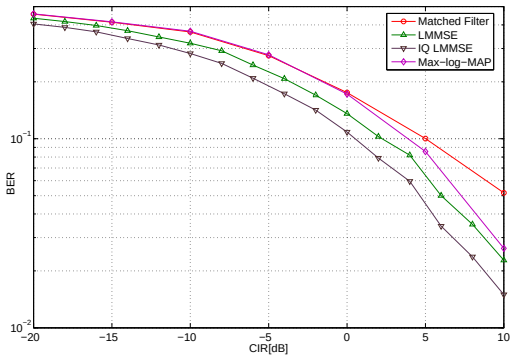


Figure 5.6: Simulation of conventional receivers in TU3 channel, 1 GMSK CCI, SNR = 20 dB, LMMSE and IQ LMMSE filter length=11, N_{sps}=1

high value in this case. As seen in Figure 5.5 for single user case, at SNR = 15 dB, all tested receivers (except Matched Filter) give BER below 10^{-3} in static channel and between 10^{-2} to 10^{-3} in the time-variant fading channel. However, interference degrades BER performance. As shown in Figure 5.6, same receives give BER above 10^{-2} at SNR = 20 dB and CIR = 10 dB. CIR = 10 dB means interfering user power is 10 dB lower than the desired one. In this test case, the receiver performance at SNR = 20 dB is worse than that of single user test case at SNR = 15 dB. It shows that BER performance is quite strongly influenced by the interferer. The reason why LMMSE and IQ LMMSE receivers give better BER is that the statistic for interfering user is included in the implementation of these receivers, referring to (5.8).

5.5 Summary

This chapter discusses about conventional receivers. The Matched Filter, the LMMSE / IQ LMMSE and the SbS MAP equalizers are introduced. The Matched Filter is the optimum receiver for the AWGN channel without ISI. The LMMSE receiver is the optimal linear receiver for ISI channel with AWGN and the performance can be enhanced by IQ-splitting for GMSK modulation. However, IQ LMMSE is at least twice as complex as LMMSE and it is not a linear process.

Based on the first order Hiden Markov Model (HMM), there are several optimal non-linear equalizers. MLSE gives optimum sequence hard decision whereas SbS MAP equalizer provides optimum symbol-by-symbol soft estimation. The SbS MAP equalizer and the ML receiver are the same for equal probable symbol. The SbS MAP equalizer provides better BER than the MLSE solution, and the MLSE receiver gives less sequence error compared to the SbS MAP equalizer. Since data detection algorithms for this project is to detect the interleaved bits, the bits therefore can be assumed to be independent. So, it is more interesting to investigate on the SbS MAP receiver compared to the MLSE solution for the scope of this project.

The derivation of the SbS MAP equalizer is also given in this chapter. Depending on the channel length, the SbS MAP equalizer is normally more complex than linear receivers, it has complexity $\mathcal{O}(|\Omega|^L)$. The SbS MAP solution can be implemented in the logarithm domain to eliminate exponential operations. Further, Max-log-MAP is an approximation to the SbS MAP equalizer however the complexity is the same as the SbS MAP equalizer, i.e.$\mathcal{O}(|\Omega|^L)$. If using only the forward recursion and hard decision, Max-Log-MAP is the same as MLSE, which provides a little worse BER than the SbS MAP equalizer.

Conventional receivers discussed in this chapter are evaluated in four scenarios in the end of this chapter. Max-log-MAP is used instead of the SbS MAP equalizer, since its BER performance is quite close to that of the SbS MAP equalizer. Simulations show that all receivers offer optimal performance in single path AWGN channel with non-ISI modulation pulse. GMSK modulation itself introduces ISI. In the multi-path channel with AWGN, Max-log-MAP provides the best BER and the Matched Filter gives the worst for GMSK. The IQ LMMSE receiver is better than the LMMSE solution for the GMSK modulation on the cost of approximately doubled complexity compared to the LMMSE solution. One CCI with GMSK modulation is included in the last test case, the performance of all conventional receivers is not satisfactory in this test case. Even the power of CCI is 10 dB lower than the desired, its influence on BER is still significant. Thus better receivers are required for interference limited scenario.

Chapter 6

Optimal MUD Receiver in AWGN ISI Channel

From Chapter 5, we know that the optimal symbol detector for the ISI channel with AWGN is the SbS MAP equalizer. However, when interference is present, the performance of the SbS MAP equalizer for SUD is not satisfactory. To improve performance, interference can be taken into account when defining the MAP state vector. Therefore, it is capable to jointly detect both the desired user and interferers. In this way, the optimal MUD receiver is derived, it is also known as exact MUD solution. This chapter discusses optimal MUD algorithm. The derivation is presented first and followed by simulation results for 1CCI in TU3 channel. This plot will appear again as a reference for suboptimal receivers, which will be addressed in Chapter 7.

6.1 Joint MAP Equalizer

Joint MAP equalizer is to maximize the joint probability of estimated symbols for all user given observations $\mathbf{r}$.

$$\breve{\mathbf{s}}_n = \underset{\mathbf{s}_n \in [-1,1]^K}{\operatorname{argmax}} P\big(\widetilde{\mathbf{s}}_n \mid \mathbf{r}\big) \qquad \text{where } \mathbf{s}_n = [s_n^{(1)}, s_n^{(2)}, \cdots, s_n^{(K)}]^T \tag{6.1}$$

$\breve{\mathbf{s}}_n$ denotes estimated symbols. In the scope of this project, only the 1st user is desired, so the ultimate goal is to calculate the posterior for the desired user, this can be achieved by marginalizing over undesired users:

$$P\big(\breve{s}_n^{(1)} \mid \mathbf{r}\big) = \sum_{k=2}^{K} P\big(\breve{s}_n^{(k)} \mid \mathbf{r}\big) \qquad \text{where } \breve{s}_n^{(k)} \in \breve{\mathbf{s}}_n \tag{6.2}$$

Note that Joint MAP is jointly optimal for all users, whereas SbS MAP for SUD is individual optimal for the desired user. The implementation of Joint MAP can be obtained by expanding state vector of SbS MAP for SUD to include more users. Thus it is possible to detect interference. In this chapter, notations are the same as for the SbS MAP equalizer for SUD in Section 5.3.2. The reason

is that SbS MAP for SUD can be viewed as a special case of Joint MAP, where the number of users is $K = 1$. The Joint MAP equalizer state vector $\mathbf{a}_n$ may be expressed as:

$$\mathbf{a}_n = \left[\hat{s}_{n-1}^{(1)}, \hat{s}_{n-2}^{(1)}, \cdots, \hat{s}_{n-L+1}^{(1)}, \hat{s}_{n-1}^{(2)}, \hat{s}_{n-2}^{(2)}, \cdots, \hat{s}_{n-L+1}^{(2)}, \cdots, \hat{s}_{n-1}^{(K)}, \hat{s}_{n-2}^{(K)}, \cdots, \hat{s}_{n-L+1}^{(K)}\right]^T \tag{6.3}$$

As in SbS MAP equalizer for SUD,where $\hat{s}_n^{(k)}$ is the hypothesized value of $s_n^{(k)}$ from constellation Ω. There are $Q =\mid \Omega \mid^{K(L-1)}$ different combinations of hypothesized values, numbered also from 0 to $Q-1$. Rewriting (6.3),

$$\begin{aligned}\mathbf{a}_n(p) = \Big[&\hat{s}_{n-1}^{(1)}(p), \hat{s}_{n-2}^{(1)}(p), \cdots, \hat{s}_{n-L+1}^{(1)}(p), \hat{s}_{n-1}^{(2)}(p), \hat{s}_{n-2}^{(2)}(p), \cdots, \hat{s}_{n-L+1}^{(2)}(p), \cdots, \\ &\hat{s}_{n-1}^{(K)}(p), \hat{s}_{n-2}^{(K)}(p), \cdots, \hat{s}_{n-L+1}^{(K)}(p)\Big]^T\end{aligned} \tag{6.4}$$

$\hat{s}_{n-1}^{(k)}(p)\ \cdots\ \hat{s}_{n-L+1}^{(k)}(p)$ is the hypothesized symbol values for state $\mathbf{a}_n\ (p)$.

For each user, $\mathbf{a}_n^{(k)}$ is a sub-vector of $\mathbf{a}_n$ and $\mathbf{a}_n^{(k)}(p)$ is a sub-vector of $\mathbf{a}_n(p)$:

$$\mathbf{a}_n^{(k)} = \left[\hat{s}_{n-1}^{(k)}, \hat{s}_{n-2}^{(k)}, \cdots, \hat{s}_{n-L+1}^{(k)}\right]^T \tag{6.5}$$

$$\mathbf{a}_n^{(k)}(p) = \left[\hat{s}_{n-1}^{(k)}(p), \hat{s}_{n-2}^{(k)}(p), \cdots, \hat{s}_{n-L+1}^{(k)}(p)\right]^T \tag{6.6}$$

Let $\boldsymbol{\xi}_n^{(p,\ q)}$ denotes the symbol vector that corresponds to a transition from $\mathbf{a}_n(p)$ to $\mathbf{a}_{n+1}(q)$:

$$\begin{aligned}\boldsymbol{\xi}_n^{(p,\ q)} = \Big[&\hat{s}_n^{(1)}(q), \hat{s}_{n-1}^{(1)}(p), \cdots, \hat{s}_{n-L+1}^{(1)}(p), \hat{s}_n^{(2)}(q), \hat{s}_{n-1}^{(2)}(p), \cdots, \hat{s}_{n-L+1}^{(2)}(p), \cdots, \\ &\hat{s}_n^{(K)}(q), \hat{s}_{n-1}^{(K)}(p), \cdots, \hat{s}_{n-L+1}^{(K)}(p)\Big]^T\end{aligned} \tag{6.7}$$

Except for the definition of the state vectors, the derivation of the Joint MAP equalizer is exactly the same as for the SbS MAP equalizer for SUD. Most of the conclusions in 5.3.2 can be reused here. Definitions of $\alpha_n(p)$, $\gamma_n(p,\ q)$, $\beta_{n+1}(q)$ are given in (5.24) (5.25) (5.26).

Same as for SbS MAP for SUD, let $\mathcal{A}$ denote the set of transitions $(p,\ q)$, where vector $\hat{\mathbf{s}}_n(q)$ takes on the specific hypothesized values $\widetilde{\mathbf{s}}_n$. Recall (5.36) for single user case, the posterior for Joint MAP equalizer appears:

$$\begin{aligned}P\big(\widetilde{\mathbf{s}}_n \mid \mathbf{r}\big) &= \frac{1}{P(\mathbf{r})} \sum_{(p,\ q)\in\mathcal{A}} \alpha_n(p)\gamma_n(p,\ q)\beta_{n+1}(q) \\ &= \frac{P(\widetilde{\mathbf{s}}_n)}{P(\mathbf{r})} \sum_{(p,\ q)\in\mathcal{A}} \alpha_n(p)\beta_{n+1}(q) P\Big(\mathbf{r}_n \mid \mathbf{a}_{n+1}(q),\ \mathbf{a}_n(p)\Big)\end{aligned} \tag{6.8}$$

Same as for single user case, $\alpha_n(p)$ and $\beta_{n+1}(q)$ can be calculated from forward and backward recursions as shown in (5.28) and (5.29). Observing (6.8), for uniformly distributed symbol $s_n^{(k)}$, joint probability $P(\widetilde{\mathbf{s}}_n)$ is also uniformly distributed. As for the SbS MAP equalizer for SUD, $P(\widetilde{\mathbf{s}}_n)$ does not influence the maximization procedure. Further $P(\mathbf{r})$ in (6.8) is a constant for given $\mathbf{r}$, it is therefore does not change the maximization steps. Finally, the desired user's posterior is found from (6.2).

Joint MAP can also be implemented in the logarithm domain like the SbS MAP equalizer for SUD. If (5.38) is applied, we obtain the Max-Log Joint MAP equalizer. The complexity of Joint MAP or Max-log Joint MAP is $\mathcal{O}(|\Omega|^{KL})$, which is exponential in K and L.

The complexity is so high, so what is the benefit of Joint MAP? In order to answer this question, it is interesting to investigate the worst case for MUD. To analyze the worst case, we first assume there are only two users and set SNR to a very high value, i.e. $SNR \to \infty$, so that influence from White Gaussian noise can be ignored in the discussion here.

In this case the BER curve for SUD always goes downwards for increasing CIR. There is no doubt that the worst case for SUD is the poor CIR range, this can also be seen in the BER curves in Figure 5.6. This is because the signal condition for the detected user is better in high CIR range, or the receiver is more certain about the detected symbols in good signal conditions. However, the BER curve looks quite different for MUD receivers in this test case.

Let's first look into two extreme cases: either the desired user's power is large corresponding to a high CIR; or the interferer's power is dominant, which means low CIR. The MUD receiver jointly detects these two users. In these two extreme scenarios, the MUD receiver is either very certain about the desired user or about the interference, therefore any strong user can be well detected and subtracted. And when the strong user is subtracted, what if left for the weak user is only white noise. Thus, in these two extreme cases, BER performance can be considered as only depending on the power ratio between the weak signal and the white noise. This is similar to the single user case, but the definition of "SNR" is the power ratio between the weak user and the white Gaussian noise. The interesting thing is that the low CIR range, which is the worst case for single user detection, is actually a good scenario for joint detection.

When the power of these two users is comparable to each other, meaning the uncertainty for both users is equal, that is the worst case for the MUD receiver. It corresponds to $CIR \to 0\ dB$ at $SNR \to \infty$, and worst case will move towards higher CIR values for lower SNR. The reason is that white noise brings uncertainty into the system.

The worst case for more interferers not only depends on CIR value, but also the power distribution and the type of interference, etc [3].

6.2 Simulation

Because the complexity of Joint MAP grows exponentially in K, 1 CCI is chosen in order to finish the simulations in reasonable time.

The above analysis of the worst case can be well illustrated by the simulation as seen in Figure 6.1. This simulation is performed with the TU3 channel, desired user is GMSK modulated, and 1 CCI with GMSK modulation is considered. SNR is 20 dB. Worst case in this simulation appears around $CIR = 15\ dB$. That is interferer power is comparable with the white Gaussian noise. Note that, for $CIR < -5\ dB$, a degradation of BER performance is seen, this is due to the Laurent linear

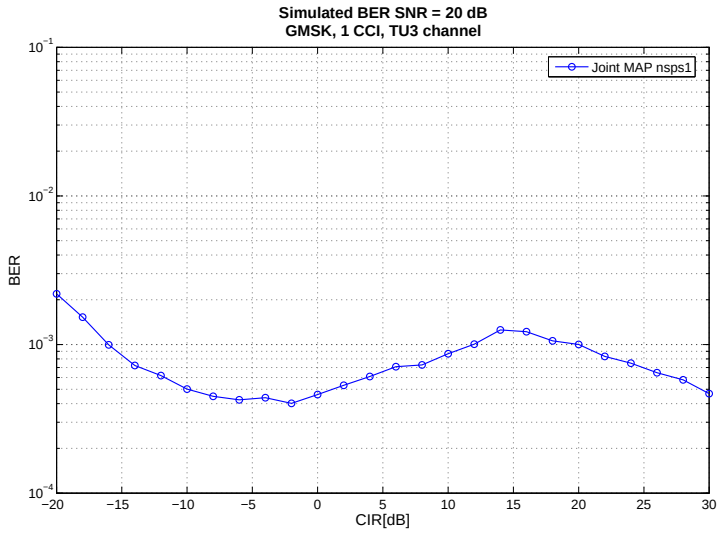


Figure 6.1: Simulation of Max-log Joint MAP for N_{sps}=1

approximation utilized by the receiver. At low CIR, interference power is much stronger than desired user. Although the approximation error for using $C_0(t)$ pulse is quite small compared to the true GMSK pulse, this error from strong interference is significant for the weak desired user.

Comparing to simulation of the same test case in Figure 5.6, the Joint MAP receiver outperforms any other single user receiver in terms of BER. However, as shown in the previous section, the complexity of the exact MUD solution is way to high for implementing it in a MS. So suboptimal solutions are needed for practical applications. The main concern of suboptimal methods is the performance-complexity tradeoff. It is not easy to find the best compromise. Approximations to the exact solution will be discussed in Chapter 7.

6.3 Summary

The focus of this chapter is derivation of the Joint MAP equalizer and performance analysis of the optimum MUD receiver. The Joint MAP algorithm is obtained by expanding the state vector of SbS MAP equalizer for SUD. So the SbS MAP equalizer for SUD can be seen as a special case of Joint MAP, that is $K = 1$. In GSM, only one user is desired, posterior distribution of the desired user can be found by marginalizing over other users. For two co-channel users, the worst case for the MUD receiver is analyzed. It occurs in the $CIR \rightarrow 0\ dB$ when SNR is sufficiently high. Low CIR range, which is the usual worst case for single user detection is actually a good case for MUD.

BER performance and complexity of Joint MAP are also discussed. The advantage of joint detection is big BER gain for interference dominated scenario, and the drawback is the huge complexity.

Simulation results of the exact solution for 1 CCI in the TU3 channel are also given in this chapter. Since the computation time is very long, it is not recommended to test more scenarios. The simulated BER curve well reflects the theoretical analysis: the worst case occurs somewhere above $CIR = 0\ dB$ for 1 CCI, i.e., one interfering user, $K = 2$. The reason BER curve moves up at very low CIR is the approximation error of using only the $C_0(t)$ pulse to represent GMSK signal. The big gain of BER performance compared to SbS MAP for SUD is achieved on the cost of complexity.

Chapter 7

Suboptimal MUD Receivers

The complexity of the exact method grows exponentially with the length of channel and the number of users to detect, i.e. $\mathcal{O}(|\Omega|^{LK})$. Besides, it requires back up memory to store the transition probabilities and forward/backward paths, which can be of huge size. It is therefore infeasible for practical applications. There are a great number of contributions aiming to reduce the computation cost, e.g. [26, 27, 28, 29, 30]. For instance, decision feedback, reducing the number of states, a pre-filter for shortening the channel impulse response are discussed. Mean field theory is one of the simplest approximations to the exact solution, it has a long history in statistical physics [31]. The applications of Mean Field methods to data detection for CDMA system are presented in [32, 33, 34]. In this chapter, two suboptimal receivers are proposed, which are based on Mean Field theory:

1. The Fully Factorized Mean Field (FFMF) approximation. It is an extension to T. Fabricius and O. Winther's work for multi-stage MUD CDMA receivers in single path AWGN channel [34]. In this project, multi-path fading channel with AWGN is considered for the derivation and implementation of FFMF receiver. The forward and backward recursions of the exact solution are avoided by fully factorizing the symbol distribution over both user and time index.

2. The Structured Mean Field (SMF) approximation. It is believed to be a novel approach to approximate the exact solution within telecommunication field. Compared to FFMF approximation, symbol distribution is factorized over only user axis. The first order HMM structure is kept in time axis.

As seen in Figure 7.1 (a), the Joint MAP equalizer models the MUD problem as K first order Hidden Markov Models. Each chain can be viewed as a SbS MAP equalizer for SUD. From the property of first order HMM model, $\mathbf{a}_{n+1}^{(k)}$ depends on $\mathbf{a}_n^{(k)}$ and $\mathbf{a}_n^{(k)}$ depends on $\mathbf{a}_{n-1}^{(k)}$, etc. In order to solve one node in the chain, the entire burst needs to be taken into account, this is the coupling in the time axis. Moreover, the output is generated from K chains, i.e. $\mathbf{r}_n$ comes from $\mathbf{a}_n^{(1)}\ \mathbf{a}_n^{(2)}\ \ldots \mathbf{a}_n^{(K)}$. To solve one chain from the observation $\mathbf{r}$, all K chains need to be considered, this is the coupling between users. As explained in the previous chapter, the Joint MAP equalizer's state vector covers the dependence between users and it utilizes forward-backward recursions to deal with the time dependency. That is also why the Joint MAP equalizer works so well but the computation cost of the exact solution is exponential in both channel length and number of users.

Like in other engineering problems, there is always a tradeoff between performance and cost. Suboptimal solutions aim at reducing the complexity without losing too much performance, that is try to find the best compromise. The basic mechanism of Mean Field approximations is to uncouple these

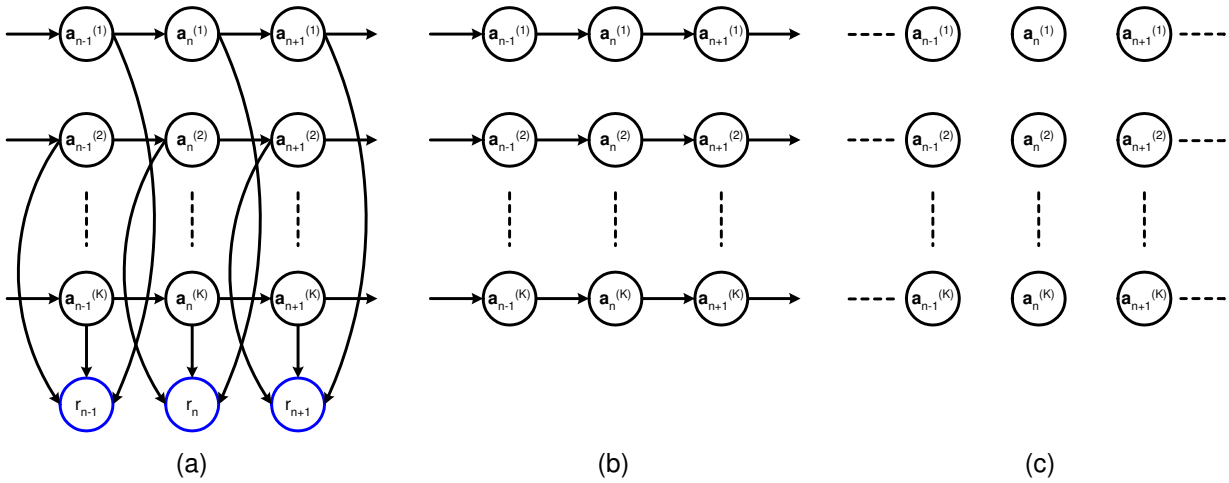


Figure 7.1: Graphical representation of the first order HMM (a) Joint MAP (b) SMF (c) FFMF

K chains in a certain manner, so that the computations fir solving each node in the chain can be simplified. It utilizes an iterative structure. In each iteration, only a part of the chains is updated. The rest is first replaced by its posterior mean from another iteration, and then contributions calculated from the posterior mean is subtracted from the observation. By doing so, what is left from the subtracted observation can be seen as "purely" generated from those nodes that need to be updated from the current iteration. So, those currently updated nodes are "isolated" from the rest of the chain and can be solved independently. In general, Mean Field algorithms can be summarized into the following four steps:

1. Initialize posterior mean of nodes by prior information.
2. Partly update posterior means while fixing the remaining nodes in the chains.
3. Iteratively update other parts of HMM chains as in step 2 until all nodes are updated once.
4. Use the latest posterior means and repeat step 2 and 3 until posterior converge.

As in other iterative methods, Mean Field approximation has the inherent convergence problem, which is not an issue for the exact solution. For every sequence update, certain cost function is reduced because it is a coordinate descent step. The coordinate descent search is to set the partial differentiation of the cost function to zero. But it can be slow. Therefore, convergence will be considered together with BER performance and complexity to evaluate these two Mean Field receivers.

FFMF and SMF receivers are illustrated in Figure 7.1 (b) and (c). The derivations and simulations of these two receivers are discussed in the following two sections respectively.

7.1 Fully Factorized Mean Field Approximation

One way to uncouple the HMM chains in Figure 7.1 (a) is to completely disconnect all nodes, as shown in Figure 7.1 (c). The benefit of doing so is that forward-backward recursions can be avoided

when estimating a single node in the HMM chain. Shortly, it will show that it significantly reduces the computation cost to solve each node separately. To uncouple all the nodes in the HMM chain is equivalent to Fully Factorizing the distribution of symbols, hence the name FFMF.

In this section, a signal model for FFMF receiver is given first, then the derivation of the FFMF receiver is explained. Besides, the influence of a control parameter and receiver oversampling rate are also investigated. Usage scenarios of the FFMF receiver are addressed in the end of this section.

7.1.1 Signal Model for FFMF

From the signal model given in (4.11), we know that the energy from symbol $s_n^{(k)}$ is spread out to LN_{sps} consecutive samples due to ISI. As shown in Figure 7.1 (c), the FFMF algorithm will not have the forward-backward recursions to recover the energy from a single symbol anymore. It is therefore necessary to include a window of observations in order to estimate one symbol. Naturally, at least LN_{sps} samples, which are generated from the symbol, should be taken into account. So the signal model in (4.11) is slightly modified in this section. Observation of LN_{sps} samples involves $K(2L-1)$ symbols. The new model is given as:

$$\mathbf{y}_n = \mathbf{H}\mathbf{b}_n + \mathbf{z}_n \tag{7.1}$$

$\mathbf{y}_n$ is a window of received samples, which captures all energy from $\mathbf{s}_n$. $\mathbf{z}_n$ is the corresponding white Gaussian noise samples.

$$\mathbf{y}_n = \begin{bmatrix} \mathbf{r}_n \\ \mathbf{r}_{n+1} \\ \vdots \\ \mathbf{r}_{n+L-1} \end{bmatrix} \qquad \mathbf{z}_n = \begin{bmatrix} \boldsymbol{\epsilon}_n \\ \boldsymbol{\epsilon}_{n+1} \\ \vdots \\ \boldsymbol{\epsilon}_{n+L-1} \end{bmatrix} \qquad \mathbf{y}_n,\ \mathbf{z}_n \in \mathbb{C}^{N_{sps}L}$$

Symbols are stacked in $\mathbf{b}_n$ as below, $\mathbf{b}_n$ has dimension $K(2L-1)$.

$$\mathbf{b}_n = \left[s_{n+L-1}^{(1)}, s_{n+L-1}^{(2)}, \cdots, s_{n+L-1}^{(K)}, \cdots, s_n^{(1)}, s_n^{(2)}, \cdots, s_n^{(K)}, \cdots, s_{n-L+1}^{(1)}, s_{n-L+1}^{(2)}, \cdots, s_{n-L+1}^{(K)} \right]^T$$

If $\mathbf{A}$ holds the channel impulse response, then $\mathbf{H}$ is built by shifting $\mathbf{A}$, that is:

$$\boxed{\mathbf{A}} = \left[\ \mathbf{h}_0^{(1)}, \mathbf{h}_0^{(2)}, \cdots, \mathbf{h}_0^{(K)}, \quad \mathbf{h}_1^{(1)}, \mathbf{h}_1^{(2)}, \cdots, \mathbf{h}_1^{(K)}, \quad \cdots \quad \mathbf{h}_{L-1}^{(1)}, \mathbf{h}_{L-1}^{(2)}, \cdots, \mathbf{h}_{L-1}^{(K)} \ \right]$$

$$\mathbf{H} = \begin{bmatrix} \mathbf{0} & \cdots & \mathbf{0} & \boxed{\mathbf{A}} & & \\ \mathbf{0} & \cdots & \boxed{\mathbf{A}} & & \mathbf{0} & \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \boxed{\mathbf{A}} & & \mathbf{0} & \cdots & & \mathbf{0} \end{bmatrix} \qquad \begin{aligned} \mathbf{H} &\in \mathbb{C}^{N_{sps}L \times K(2L-1)} \\ \mathbf{0} &\in 0^{N_{sps} \times K} \end{aligned}$$

Besides, the definition of $\mathbf{M}$ is also given here, which is used to explain the derivation of the FFMF receiver later. $\mathbf{M}$ is the covariance matrix of $\mathbf{H}$, where the diagonal is set to zero.

$$\mathbf{M} = \mathbf{H}^H\mathbf{H} - \mathrm{diag}(\mathbf{H}^\mathbf{H}\mathbf{H}) \qquad \mathbf{M} \in \mathbb{C}^{K(2L-1)\times K(2L-1)} \tag{7.2}$$

Until here, we made a small change to the signal model in (4.11). The new signal model (7.1) makes the expressions for derivation of the FFMF algorithm more clear and direct.

7.1.2 Derivation of FFMF receiver

By using Bayes' rule (5.15), it has been explained in 5.3.1 that the ML equalizer is the same as the SbS MAP equalizer for uniformly distributed symbol. By the signal model in (7.1), the estimated sequence of symbols for observation sequence $\mathbf{y}_n$ is given as:

$$\breve{\mathbf{b}}_n = \mathrm{argmax}\, P(\mathbf{y}_n|\widetilde{\mathbf{b}}_n) \qquad \text{where } \widetilde{\mathbf{b}}_n \in [-1,+1]^{K(2L-1)} \tag{7.3}$$

Note that (7.3)is jointly optimal for vector $\mathbf{b}_n$ and it is an approximation to the exact solution. The reason that the output is a vector is the used sliding window. As explained when introducing the signal model for FFMF receiver, sliding window is needed in order to avoid forward-backward recursions. The log Likelihood of the received signal $\mathbf{y}_n$ given $\mathbf{b}_n$ is:

$$\ln P(\mathbf{y}_n|\mathbf{b}_n) \propto \frac{-1}{2\sigma^2}\big(\mathbf{y}_n - \mathbf{H}\mathbf{b}_n\big)^H\big(\mathbf{y}_n - \mathbf{H}\mathbf{b}_n\big) \tag{7.4}$$

The estimated symbol is obtained when the log likelihood $\ln P(\mathbf{y}_n|\mathbf{b}_n)$ is maximized, i.e. the right hand-side of (7.4) is maximized. For the derivation of the FFMF equations, the distribution of symbols is most interesting. If we multiply out the right-hand side of (7.4), we will find that the log likelihood expression can be divided into two parts: one depends on $\mathbf{b}_n$, and one independent of $\mathbf{b}_n$. Furthermore, replace σ^2 with T. T is the noise power, which is called temperature in physics. If symbol energy is normalized to 1, then $1/T$ is the same as SNR. Later, we will see that T works as a control parameter for adjusting receiver performance. (7.4) can be rewritten as:

$$\ln P(\mathbf{y}_n|\mathbf{b}_n) \propto \frac{-1}{T}E(\mathbf{b}_n) - C_T \tag{7.5}$$

where:

$$E(\mathbf{b}_n) = \frac{1}{2}\Big(\mathbf{b}_n^H\mathbf{M}\mathbf{b}_n - 2\mathrm{Re}\{\mathbf{y}_n^H\mathbf{H}\mathbf{b}_n\}\Big) \tag{7.6}$$

$$C_T = \frac{1}{2T}\Big(\sum_{i=1}^{K(2L-1)}\mathbf{H}^\mathbf{H}\mathbf{H}(i,i) + \mathbf{y}_n^H\mathbf{y}_n\Big) \tag{7.7}$$

(7.7) shows that C_T is not relevant to $\mathbf{b}_n$, i.e. C_T is a constant for fixed T. This implies that C_T can be excluded from the maximization procedure of the ML receiver, at fixed T. So, (7.3) can be modified to:

$$\breve{\mathbf{b}}_n = \operatorname{argmax} P(\mathbf{y}_n|\widetilde{\mathbf{b}}_n) = \operatorname{argmin} E(\mathbf{b}_n) \qquad \widetilde{\mathbf{b}}_n \in [-1,+1]^{K(2L-1)} \tag{7.8}$$

The exact solution finds the true posterior distribution of $\mathbf{b}_n$. For a BPSK modulated signal, let $\mathrm{q_T}(\mathbf{b_n})$ be the ergodic equilibrium distribution of $\mathbf{b}_n$ at temperature T. The exact solution makes the Kullback-Leibler ($\mathcal{KL}$) divergence between the true distribution and $\mathrm{q_T}(\mathbf{b_n})$ zero. The $\mathcal{KL}$ divergence can be written as:

$$\begin{aligned}
\mathcal{KL}\Big[q_T(\mathbf{b}_n), \frac{e^{-\frac{1}{T}E(\mathbf{b}_n)}}{C_T}\Big] &= \sum_{[-1,1]^{K(2L-1)}} q_T(\mathbf{b}_n) \log \frac{q_T(\mathbf{b}_n) C_T}{e^{-\frac{1}{T}E(\mathbf{b}_n)}} \\
&= \sum_{[-1,1]^{K(2L-1)}} q_T(\mathbf{b}_n) \Big(\log q_T(\mathbf{b}_n) + \log C_T + \frac{1}{T} E(\mathbf{b}_n) \Big) \\
&= \sum_{[-1,1]^{K(2L-1)}} q_T(\mathbf{b}_n) \log q_T(\mathbf{b}_n) + \log C_T \sum_{[-1,1]^{K(2L-1)}} q_T(\mathbf{b}_n) \\
&+ \frac{1}{T} \sum_{[-1,1]^{k(2L-1)}} q_T(\mathbf{b}_n) E(\mathbf{b}_n) \\
&= \big\langle \log q_T(\mathbf{b}_n) \big\rangle_T + \frac{1}{T} \big\langle E(\mathbf{b}_n) \big\rangle_T + \log C_T
\end{aligned} \tag{7.9}$$

$\langle \cdot \rangle_T$ denotes average over the ergodic equilibrium distribution $\mathrm{q_T}(\mathbf{b_n})$ at temperature T. The last term in (7.9) is constant, and the first two terms depend on $\mathrm{q_T}(\mathbf{b_n})$, which is called free energy in physics.

$$\mathcal{F}_T = \big\langle \log q_T(\mathbf{b}_n) \big\rangle_T + \frac{1}{T} \big\langle E(\mathbf{b}_n) \big\rangle_T \tag{7.10}$$

Minimizing the $\mathcal{KL}$ divergence is the same as minimizing $\mathcal{F}_T$. Hence, $\mathcal{F}_T$ can be seen as the cost function. ML/MAP gives the posterior distribution $q_T(\mathbf{b}_n)$ that minimize the $\mathcal{KL}$ distance to 0, that is:

$$q_T(\mathbf{b}_n) = \frac{e^{-\frac{1}{T}E(\mathbf{b}_n)}}{C_T} \tag{7.11}$$

The exact solution guarantees to find the global minimum of cost function $\mathcal{F}_T$, but the computation cost is high. To reduce the complexity, we should find an approximation to $\mathrm{q_T}(\boldsymbol{b}_\mathrm{n})$ that makes $\mathcal{F}_T$ simpler to derive. One choice to achieve this goal is to constrain $\mathrm{q_T}(\boldsymbol{b}_\mathrm{n})$ by fully factorizing both over user and time, as shown in Figure 7.1 (c). This factorization is an approximation to the true equilibrium distribution, which is the posterior distribution from ML/MAP. By doing so, the $\mathcal{KL}$

distance cannot be minimized to zero, which means that it can only find a local minimum of cost function instead of the global one. For binary symbol, the constrained $\mathrm{q_T}(\boldsymbol{b}_\mathrm{n})$ can be written as:

$$q_T(\mathbf{b}_n) \approx \prod_{l=-L+1}^{L-1} \prod_{k=1}^{K} q_T(s_{n-l}^{(k)}) = \prod_{l=-L+1}^{L-1} \prod_{k=1}^{K} \Big[\frac{1+m_{n-l,T}^{(k)}}{2}\Big]^{\frac{1+s_{n-l}^{(k)}}{2}} \Big[\frac{1-m_{n-l,T}^{(k)}}{2}\Big]^{\frac{1-s_{n-l}^{(k)}}{2}} \tag{7.12}$$

where $\mathrm{m_{n-l,\ T}^{(k)}} = \Big\langle \mathrm{s_{n-l}^{(k)}} \Big\rangle_\mathrm{T}$ is the average with respect to $\mathrm{q_T}(\mathbf{b}_\mathrm{n})$, or the mean of symbols. With the constrained distribution (7.12), free energy in (7.10) can be written as:

$$\begin{aligned}
\mathcal{F}_T &= \Big\langle \frac{1}{2T}\Big(\mathbf{b}_n^T\mathbf{M}\mathbf{b}_n - 2\mathrm{Re}\{\mathbf{y}_n^H\mathbf{H}\mathbf{b}_n\}\Big)\Big\rangle_T \\
&+ \Big\langle \sum_{l=-L+1}^{L-1}\sum_{k=1}^{K}\Big(\frac{1+s_{n-l}^{(k)}}{2}\log\frac{1+m_{n-l}^{(k)}}{2} + \frac{1-s_{n-l}^{(k)}}{2}\log\frac{1-m_{n-l}^{(k)}}{2}\Big)\Big\rangle_T \\
&= \frac{1}{2T}\Big(\big\langle\mathbf{b}_n^T\big\rangle_T\mathbf{M}\big\langle\mathbf{b}_n\big\rangle_T - 2\mathrm{Re}\{\mathbf{y}_n^H\mathbf{H}\big\langle\mathbf{b}_n\big\rangle_T\}\Big) \\
&+ \Big\langle \sum_{l=-L+1}^{L-1}\sum_{k=1}^{K}\Big(\frac{1+s_{n-l}^{(k)}}{2}\log\frac{1+m_{n-l}^{(k)}}{2} + \frac{1-s_{n-l}^{(k)}}{2}\log\frac{1-m_{n-l}^{(k)}}{2}\Big)\Big\rangle_T
\end{aligned} \tag{7.13}$$

The last equal is true because the diagonal of $\mathbf{M}$ is set to zero in (7.2) and the symbols in $\mathbf{b}_n$ are assumed uncorrelated. For binary symbol:

$$\Big\langle s_n^{(k)} s_m^{(k')}\Big\rangle = \begin{cases} 1 & \text{if } n=m \text{ and } k=k' \\ \Big\langle s_n^{(k)}\Big\rangle\Big\langle s_n^{(k')}\Big\rangle & \text{if } m\neq n \text{ or } k\neq k' \end{cases} \tag{7.14}$$

Proof:

For binary symbol, the average of the product of same symbol is always 1, i.e. $\Big\langle s_n^{(k)} s_n^{(k)}\Big\rangle = 1$.

For the 2nd case in (7.14), if the probability of $s_n^{(k)}$ *being* $+1$ *is* p_1*, then being* -1 *has probability* $1-p_1$*, and if* $s_m^{(k')}$ *being* $+1$ *is* p_2*, then being* -1 *is* $1-p_2$*.*

$$\begin{aligned}
\Big\langle s_n^{(k)} s_m^{(k')}\Big\rangle &= p_1p_2 - p_1(1-p_2) - (1-p_1)p_2 + (1-p_1)(1-p_2) \\
&= p_1p_2 - p_1 + p_1p_2 - p_2 + p_1p_2 + 1 - p_1 - p_2 + p_1p_2 \\
&= 4p_1p_2 - 2p_1 - 2p_2 + 1 \\
&= (2p_1-1)(2p_2-1) \\
&= \Big(p_1 - (1-p_1)\Big)\Big(p_2 - (1-p_2)\Big) \\
&= \Big\langle s_n^{(k)}\Big\rangle\Big\langle s_n^{(k')}\Big\rangle
\end{aligned} \tag{7.15}$$

Now, replace $\left\langle s_{n-l}^{(k)} \right\rangle_T$ by $m_{n-l,T}^{(k)}$ in (7.13), the free energy is reduced to:

$$\begin{aligned}\mathcal{F}_T = &\frac{1}{2T}\Big(\mathbf{m}_n^T \mathbf{M} \mathbf{m}_n - 2\mathrm{Re}\{\mathbf{y}_n^H \mathbf{H} \mathbf{m}_n\}\Big) \\ &+ \sum_{l=-L+1}^{L-1} \sum_{k=1}^{K} \Big(\frac{1+m_{n-l,T}^{(k)}}{2} \log \frac{1+m_{n-l,T}^{(k)}}{2} + \frac{1-m_{n-l,T}^{(k)}}{2} \log \frac{1-m_{n-l,T}^{(k)}}{2} \Big)\end{aligned} \tag{7.16}$$

where

$$\mathbf{m}_n = \Big[m_{n+L-1}^{(1)}, m_{n+L-1}^{(2)}, \cdots, m_{n+L-1}^{(K)}, \cdots, m_n^{(1)}, m_n^{(2)}, \cdots, m_n^{(K)}, \cdots, m_{n-L+1}^{(1)}, m_{n-L+1}^{(2)}, \cdots, m_{n-L+1}^{(K)} \Big]^T$$

A local minimum of $\mathcal{F}_T$ can be found by setting the partial differentiation of $\mathcal{F}_T$ with respect to $m_{n,T}^{(k)}$ to zero, i.e.

$$\frac{\partial \mathcal{F}_T}{\partial m_{n,\,T}^{(k)}} = 0 \qquad k \in [1, 2, \cdots, K] \tag{7.17}$$

This leads to a hyperbolic tangent decision function. The fix point equation for the i-th element in $\mathbf{m}_n$ is:

$$m_{i,\,T} = \tanh\Big[\frac{1}{T} \mathrm{Re}\big\{ \big(\mathbf{y}_n^H \mathbf{H} - \mathbf{M}\mathbf{m}_n \big)_i \big\} \Big] \tag{7.18}$$

$(\cdot)_{\mathrm{i}}$ denotes the i-th element in a vector.

Proof:

Since $m_{n,\,T}^{(k)}$ is the posterior mean of symbol, $m_{n,\,T}^{(k)} \in \mathbb{R}$. Say $m_{n,\,T}^{(k)}$ is the i-th element in $\mathbf{m}_n$ and denoted as $m_{i,\,T}$. Insert(7.16) to (7.17):

$$\begin{aligned}
\frac{\partial \mathcal{F}_T}{\partial m_{n,\,T}^{(k)}} &= \frac{\partial \mathcal{F}_T}{\partial m_{i,\,T}} = \frac{1}{2T}\frac{\partial}{\partial m_{i,\,T}}\Big(\mathbf{m}_n^T\mathbf{M}\mathbf{m}_n - 2\mathrm{Re}\{\mathbf{y}_n^H\mathbf{H}\mathbf{m}_n\}\Big) \\
&\quad + \sum_{l=-L+1}^{L-1}\sum_{k=1}^{K}\frac{\partial}{\partial m_{i,\,T}}\Big(\frac{1+m_{n-l,T}^{(k)}}{2}\log\frac{1+m_{n-l,T}^{(k)}}{2} + \frac{1-m_{n-l,T}^{(k)}}{2}\log\frac{1-m_{n-l,T}^{(k)}}{2}\Big) \\
&= \frac{1}{2T}\left(\Big(\mathbf{M}\mathbf{m}_n\Big)_i + \Big(\mathbf{m}_n^T\mathbf{M}\Big)_i - 2\mathrm{Re}\Big\{\big(\mathbf{y}_n^H\mathbf{H}\big)_i\Big\}\right) \\
&\quad + \left(\frac{1}{2}\log\frac{1+m_{i,\,T}}{2} + \frac{1+m_{i,\,T}}{2}\frac{2}{1+m_{i,\,T}}\frac{1}{2} + \frac{-1}{2}\log\frac{1-m_{i,\,T}}{2} + \frac{1-m_{i,\,T}}{2}\frac{2}{1-m_{i,\,T}}\frac{-1}{2}\right) \\
&= \frac{1}{2T}\left(2\mathrm{Re}\Big\{\Big(\mathbf{M}\mathbf{m}_n\Big)_i\Big\} - 2\mathrm{Re}\Big\{\Big(\mathbf{y}_n^H\mathbf{H}\Big)_i\Big\}\right) + \left(\frac{1}{2}\log\frac{1+m_{i,\,T}}{2} + \frac{1}{2} - \frac{1}{2}\log\frac{1-m_{i,\,T}}{2} - \frac{1}{2}\right) \\
&= \frac{1}{T}\mathrm{Re}\Big\{\Big(\mathbf{M}\mathbf{m}_n - \mathbf{y}_n^H\mathbf{H}\Big)_i\Big\} + \left(\frac{1}{2}\log\frac{1+m_{i,\,T}}{1-m_{i,T}}\right) \\
&= \frac{1}{T}\mathrm{Re}\Big\{\Big(\mathbf{M}\mathbf{m}_n - \mathbf{y}_n^H\mathbf{H}\Big)_i\Big\} + \tanh^{-1}[m_{i,\,T}] = 0
\end{aligned} \tag{7.19}$$

From the last equal in (7.19), we get:

$$\tanh^{-1}[m_{i,\,T}] = \frac{1}{T}\mathrm{Re}\Big\{\Big(\mathbf{y}_n^H\mathbf{H} - \mathbf{M}\mathbf{m}_n\Big)_i\Big\} \tag{7.20}$$

Take tanh *on both sides of (7.19), (7.18) is then proved.*

With (7.18), it is possible to update the posterior mean iteratively until it converges. If using hard decision, the estimated symbol from FFMF receiver is:

$$\breve{s}_n^{(k)} = \mathrm{sgn}\big(m_n^{(k)}\big) \tag{7.21}$$

To implement the FFMF receiver by (7.18), we have two choices, either parallel updating K elements in $\mathbf{m}_n$, or updating only one $(\mathbf{m}_n)_i$ in each iteration. As discussed in the beginning of this chapter, serial update is a coordinate descent step. It fulfills (7.17) and each serial update reduces the free energy but it can be slow. Parallel update has the risk of increasing $\mathcal{F}_T$ in each iteration [34]. So here we only consider serial update. Mapping the FFMF receiver to those four steps described in the introduction part of this chapter, the procedure of FFMF is:

1. First initialize $m_n^{(k)}$. The information part of a burst should be set to zero for binary uniformly distributed symbols. If the TS is known, the TS part should be set accordingly. Tail bits are set to -1, and guard bits are set to $+1$.

2. Update posterior means by (7.18). This is shown below. Other symbols in $\mathbf{m}_n$, which do not need to be updated in this iteration, are obtained by the feed back from other iterations.

$$\Big[\overbrace{m_{n+L-1}^{(1)}, m_{n+L-1}^{(2)}, \cdots, m_{n+L-1}^{(K)}, \cdots}^{\text{decision feedback}}, \underbrace{m_n^{(1)}, m_n^{(2)}, \cdots, m_n^{(K)}}_{\text{update in current iteration}}, \overbrace{\cdots, m_{n-L+1}^{(1)}, m_{n-L+1}^{(2)}, \cdots, m_{n-L+1}^{(K)}}^{\text{decision feedback}}\Big]^T$$

3. Update other symbols from the bursts with newly updated posterior means, so that posterior means for all K bursts are updated once.

4. Repeat step 2 and 3 until all $m_n^{(k)}$ converge.

The posterior mean from FFMF can be used as soft information for other part of the receiver, and estimated symbol can be obtained by hard decision as in (7.21).

7.1.3 Complexity of FFMF receiver

If N_x denotes number of real multiplications, N_+ denotes number of real additions and $N_{\tanh}$ denotes number of tanh operations, computation cost of FFMF receiver is $\mathcal{O}\big(N_{ITE}K(N_x + N_+ + N_{\tanh})\big)$, where

$$N_x = \underbrace{2N_{sps}L}_{(\mathrm{Re}\{\mathbf{y}_n^H\mathbf{H}\})_i} + \underbrace{K(2L-1)}_{(\mathrm{Re}\{\mathbf{M}\mathbf{m}_n\})_i} \tag{7.22}$$

$$N_+ = \underbrace{2N_{sps}L-1}_{(\mathrm{Re}\{\mathbf{y}_n^H\mathbf{H}\})_i} + \underbrace{K(2L-1)-1}_{(\mathrm{Re}\{\mathbf{M}\mathbf{m}_n\})_i} + \underbrace{1}_{(\mathrm{Re}\{\mathbf{y}_n^H\mathbf{H}-\mathbf{M}\mathbf{m}_n\})_i} = 2N_{sps}L + K(2L-1) - 1 \tag{7.23}$$

$N_{\tanh} = 1$. In practice, tanh operation is implemented by a lookup table, so it does not require much computation cost.

Nicely the exponential part of the exact solution $\mathcal{O}(|\Omega|^{KL})$ is gone. This can be quite significant for reducing the complexity, especially for large K or L. The simulation time to estimate one burst is something similar to the LMMSE receiver, which is much faster than exact methods.

Convergence speed, i.e. N_{ITE}, can be analyzed from the simulations. Figure 7.2 is a performance of the FFMF receiver in a TU3 channel, GMSK modulation and with 1 CCI GMSK, SNR = 20 dB. Simulation shows that FFMF always converges in all test cases, and the convergence speed is around $N_{ITE} = 10$.

7.1.4 Investigation on control parameter T

The last aspect of the suboptimal solution is BER. The BER performance of the FFMF receiver is not satisfactory. In general a BER degradation is acceptable for suboptimal methods since complexity is also reduced. The question is how much degradation can be tolerated. Observing the BER curve

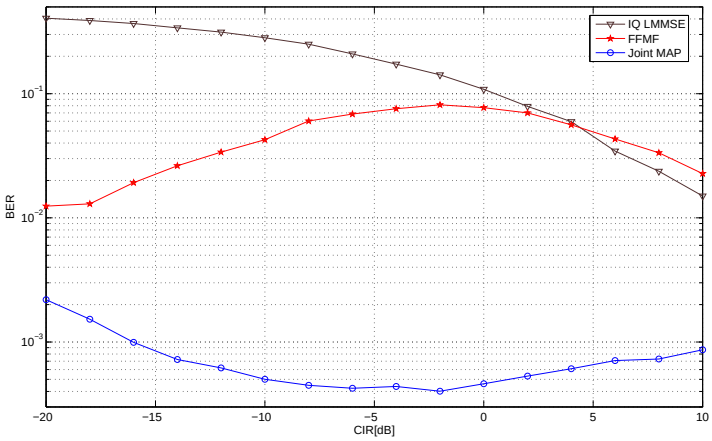


Figure 7.2: Simulation of FFMF receiver in TU3 channel, 1 GMSK CCI, SNR=20 dB

of the FFMF receiver in Figure 7.2, it is noticed that there is a large gain at low CIR compared to the IQ LMMSE receiver. The BER peak appears around $CIR = 0\ dB$, but for high CIR values, it is worse than the IQ LMMSE receiver. It can be explained by the expected local minima problem, which is caused by the approximation error.

As indicated earlier, T can be used as a control parameter for adjusting receiver performance. This is possible because the approximation error of FFMF can be "compensated" by increasing noise power. That is to model the error as extra noise added into the system. Increasing noise power corresponds to increasing T, i.e. to decreasing SNR compared to its true value. In [34], it is proved that there is an optimum T_o that gives global minima of $\mathcal{F}_T$. This optimum T_o can be found from starting the search with a large T higher than the crucial T_c. When $T > T_c$, there is only one minimum of the objective function $\mathcal{F}_T$. T_c is found from Eigenvalues of M [34]. By setting the initial T higher than T_c and taking sufficient small steps to decrease T to T_o, the global minimum of cost function $\mathcal{F}_T$ can be found. This method to find the global minima of $\mathcal{F}_T$ is also known as annealing. Annealing ensures to find the global minimum, however, the challenge is choosing the step size. If the step is infinitely small, the global minimum is guaranteed, but the complexity will grow infinitely large, thus the advantage of FFMF is destroyed.

The theoretical analysis of the existence of T_o in [34] provides a guidance for studying the FFMF receiver implemented in this project. It is worth to try other T values in order to improve BER. But, the concern is that annealing may increase the complexity of FFMF too much, so that it will not be an attractive simple solution anymore. Therefore another method is chosen in this project. Large amount of simulations are performed in order to find out how T can be used to improve the FFMF receiver. Since these experiments are carried out for GSM, the conclusions from these tests are only valid for GSM.

The test setup in Figure 7.2 with $CIR = 10\ dB$ is chosen for the investigation. Since at $CIR = 10\ dB$, there is a big BER loss for the FFMF receiver compared to the IQ LMMSE receiver. These experiments are summarized in Figure 7.3. In Figure 7.3 (a) BER performance at $1/T = 100$ to $1/T = 1$ is tested. The red dot is simulated at $1/T = 100$, which is the true SNR. It gives BER = 4%. Figure 7.3 (b) is zoomed in from (a). $1/T = 5.1$ (red star) provides the min BER = 1.5%, which is defined as optimal $1/T$. If we define the optimal $1/T$ range as the values where BER is at most 10% higher than the minimum BER, this range is then form 4 to 6.7 (pink diamond). A large

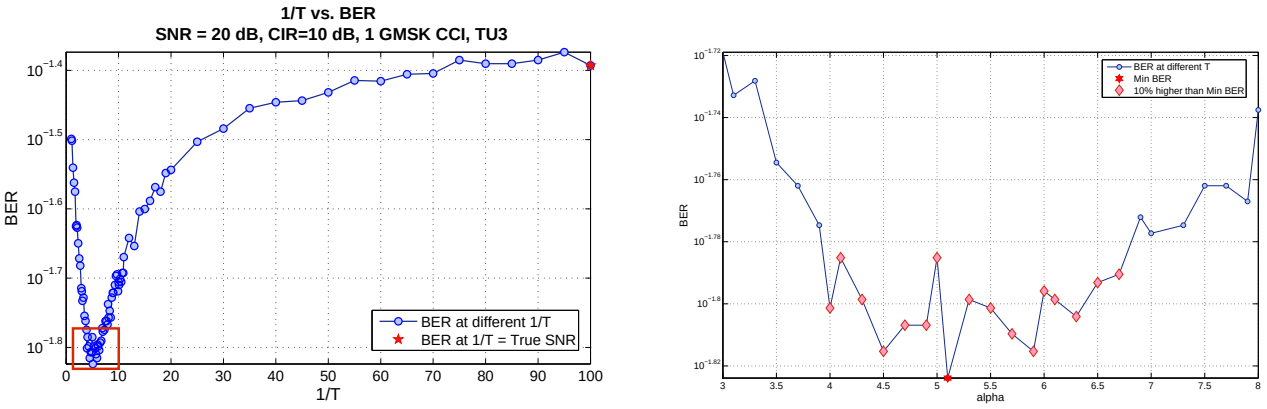
BER gain is achieved by increasing T from its default(=SNR) value. BER then approaches that of IQ LMMSE for high CIR. High BER gain is achieved by increasing T!

The simulations in Figure 7.3 further indicate the existence of an optimal T described in [34]. Similar experiments were conducted for other channel types for the same interference setup at $CIR = 10\ dB$, $SNR = 20\ dB$. The test results are collected in Figure 7.4 (a). The x axis is the $1/T$ value, dots are the points that have been tested and lines show optimum $1/T$ ranges. The experiments show that the optimal $1/T$ range is almost unchanged for different channel types!

Furthermore, optimum $1/T$ values for other CIR values are also found by the same procedure, the results are shown in Figure 7.4 (b). It indicates a direct relation between optimum $1/T$ and CIR.

According to the above simulations, the way T influences FFMF receiver can be concluded as follows:

1. There exists a unique optimum $1/T$, which is normally lower than the true SNR. This optimum $1/T$ value yields minimum BER. If BER can be 10% higher than minimum BER, there is a wide range of optimum $1/T$ values, as shown in Figure 7.3.



(a) *Range from* $1/T = 0.1$ *to* $1/T = 100$ (b) *Finer search in the range* $1/T = 3$ *to* $1/T = 8$

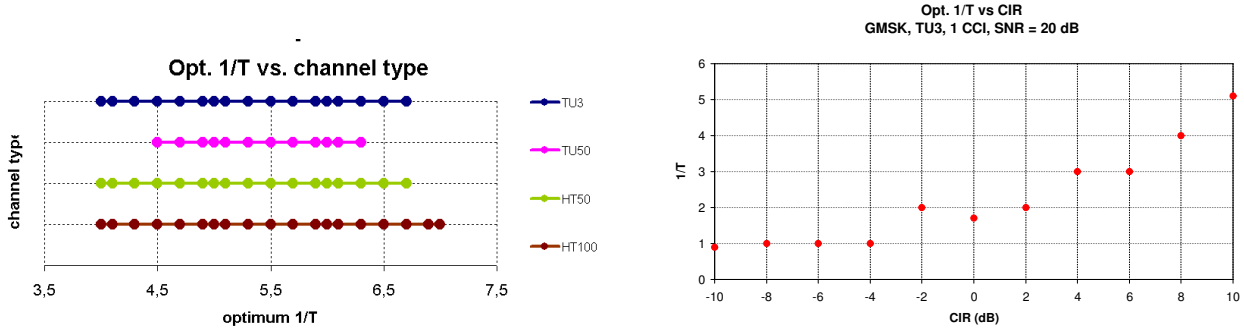
Figure 7.3: Investigation on the influence of control parameter T

2. The optimum $1/T$ value is mainly determined by CIR and true SNR, the channel type has no significant impact. This is shown in Figure 7.4.

These two conclusions are quite useful, as it means that a simple prediction function can be implemented for finding the optimum $1/T$ in (7.18), that is:

$$\text{Optimum } T = f_p(CIR, SNR) \tag{7.24}$$

This prediction function basically only requires little extra computation, so that the FFMF receiver including prediction function is still very simple. By finding the optimum $1/T$, the advantage of low complexity is kept and BER performance is improved. However, CIR and SNR are unknown in practical applications. CIR and SNR are normally estimated by the receiver anyway for other modules e.g. channel estimation. Therefore, they can be borrowed here to predict the optimum $1/T$.



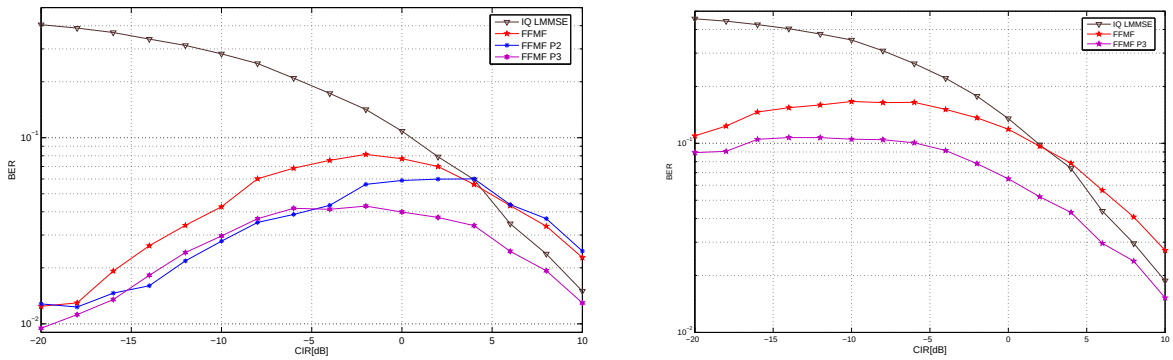
(a) *Optimal* $1/T$ *range vs channel type*

(b) *Optimal* $1/T$ *vs CIR*

Figure 7.4: Relations between optimal 1/T and channel type or CIR value

Another advantage of f_p is that it does not heavily depend on the accuracy of estimated SNR and CIR. This is because the optimum $1/T$ range is rather wide as shown in Figure 7.3 (b) and Figure 7.4 (a). So the output of f_d only needs to fall into the optimum $1/T$ range. This corresponds to relaxing the precision of CIR and SNR estimations.

Simulations of FFMF for two different prediction functions are given in Figure 7.5 (a). Simulations for 2 CCI and 1 ACI with power distribution defined by GERAN DTS3 are collected in Appendix B. Both prediction functions provide a BER gain with almost no extra computation cost. $P3$ will be used for the next test case. That is 2 CCI (GMSK), with CCI#1 $= 0\ dB$ and CCI#2 $= -10\ dB$. Results are plotted in Figure 7.5 (b).



(a) 1 *CCI*

(b) 2 *CCI*

Figure 7.5: Simulation of FFMF receiver in TU3 channel, GMSK

7.1.5 Investigation on receiver oversampling rate for FFMF receiver

It has been discussed earlier that a GSM receiver normally utilizes two times oversampling to capture more energy from desired user, i.e. $N_{sps} = 2$. This also works for the FFMF receiver especially when there is ACI interference, as shown by the simulation plot in Figure 7.6 and B.2.

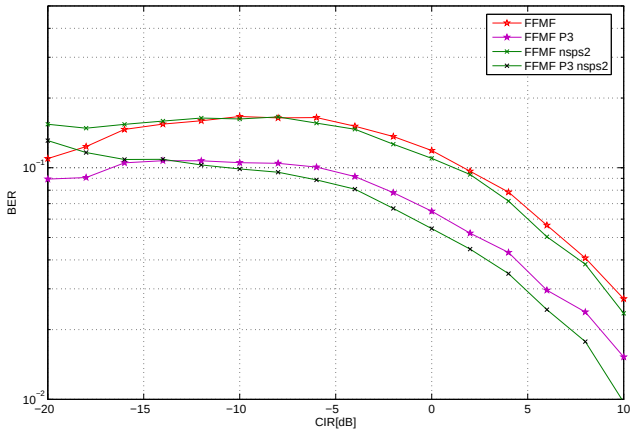


Figure 7.6: Impact of N_{sps} for FFMF receiver: TU3 channel, GMSK, CCI#1 0 dB, CCI#2 -10 dB

7.1.6 FFMF receiver usage scenarios

The computation cost of FFMF receiver is very low which is similar to that of the LMMSE receiver. Therefore, it can be used with other IC techniques e.g. multiple antennas or whitening filter. Simulations with combined FFMF receiver and whitening filter are shown in Figure 7.7 (IQPW4 means IQ-splitting Pre-Whiting Filter with filter length = 4. d2 means detecting two strong users and whitening the weak one).

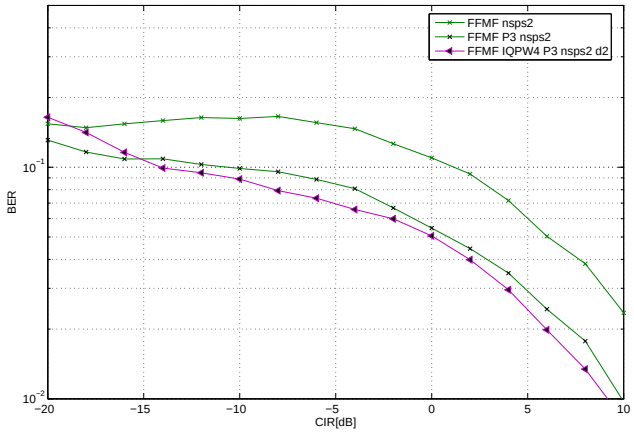


Figure 7.7: FFMF receiver with IQPW4 filter: TU3 channel, GMSK, CCI#1 0 dB, CCI#2 -10 dB

Another usage scenario is single user FFMF receiver for modulation types that have bigger constellation size, e.g. EDGE. Complexity of FFMF is $\mathcal{O}(KN_{ITE})$, for lager $|\Omega|$, the computation cost for the FFMF receiver increases accordingly. But unlike the exact solution, $|\Omega|$ does not significantly influence the computation cost for FFMF receiver.

7.2 Structured Mean Field Approximation

As explained previously, Figure 7.1 (a) shows that the bits are coupled between users through observations, thus the complexity of the exact solution is $\mathcal{O}(|\Omega|^{KL})$. As seen in Figure 7.1 (c), the FFMF solution factorizes the equilibrium distribution of bits over both user index and time index, whereas the SMF receiver only factorizes this distribution over the user index, as demonstrated in Figure 7.1 (b). For each chain, bits are still coupled in the time domain and modeled as a first order Markov process [35]. The complexity for each chain is $\mathcal{O}(|\Omega|^{L})$ and the overall complexity of SMF is reduced to $\mathcal{O}(N_{ITE}K|\Omega|^{L})$ compared to that of the exact solution, where N_{ITE} is the number of iterations.

As seen in the FFMF receiver, it experiences the local minima problem. This problem is compensated by increasing the control parameter T. That is to model the approximation error as white noise. The SMF receiver keeps more structure of the exact solution and it avoids more local minima which occurs to approximation methods, as seen in the FFMF receiver. Thus it provides faster convergence speed.

7.2.1 Derivation of SMF receiver

For the FFMF receiver, symbols are assumed independent, then each node $\mathbf{a}_n^{(k)}$ in Figure 7.1 (c) corresponds to the symbol s_n^k. But for the SMF solution, symbols are still coupled in the time axis. Thus it is better to keep $\mathbf{a}_n$ to describe the distribution of nodes in Figure 7.1 (b). First, the concept of factorizing over user axis is explained. Recall the definitions of state vectors in (6.3) and (6.4). And let $\mathbf{a}$ hold the state vectors of all K bursts and $\mathbf{a}^{(k)}$ contain all N_s state vectors of user k for the entire burst, that is:

$$\mathbf{a} = [\mathbf{a}_1^T, \mathbf{a}_2^T, \ldots, \mathbf{a}_{N_s}^T]^T \tag{7.25}$$

$$\mathbf{a}^{(k)} = [\mathbf{a}_1^{(k)T}, \mathbf{a}_2^{(k)T}, \ldots, \mathbf{a}_{N_s}^{(k)T}]^T \tag{7.26}$$

With this notation, the general structured mean field approximation is given as (p.30 eq.3.9 in [36]):

$$q\Big(\mathbf{a}^{(k)}\Big) \propto \exp\Big(\big\langle \ln P \big\rangle_{q(\mathbf{a} \setminus \mathbf{a}^{(k)})}\Big) \tag{7.27}$$

where $q\Big(\mathbf{a}^{(k)}\Big)$ is the posterior distribution of the k-th chain. $q(\mathbf{a} \setminus \mathbf{a}^{(k)})$ is the posterior distribution of $\mathbf{a}$ except $\mathbf{a}^{(k)}$. P is the true distribution which needs to be approximated. $\langle\cdot\rangle$ denotes the average.

The exact solution, as described in Section 6.1, is to expand the state vector to cover all K users. It corresponds to merge K chains in Figure 7.1 (a) to only one chain, but state vector $\mathbf{a}_n$ is K times as big as $\mathbf{a}_n^{(k)}$. Applying first order HMM properties, i.e. observation $\mathbf{r}_n$ is independent of all other observations and $\mathbf{a}_n$ is independent of $\mathbf{a}_{n-2}$, $\mathbf{a}_{n-3}$, $\ldots$, $\mathbf{a}_1$ given $\mathbf{a}_{n-1}$[35]. So the true joint probability distribution of the sequence of states and observations for one first order HMM chain can be expressed by (p.247 in [35]):

$$\ln q(\mathbf{r},\ \mathbf{a}) = \ln\Big(P(\mathbf{a}_1)P(\mathbf{r}_1 \mid \mathbf{a}_1)\prod_{n=2}^{Ns} P(\mathbf{a}_n \mid \mathbf{a}_{n-1})P(\mathbf{r}_n \mid \mathbf{a}_n)\Big) \tag{7.28}$$

$q(\mathbf{r},\ \mathbf{a})$ is the true distribution that needs to be approximated by the SMF receiver. $P(\mathbf{a}_n \mid \mathbf{a}_{n-1})$, $P(\mathbf{r}_n \mid \mathbf{a}_n)$ on the right hand side of (7.28) are solved by the SbS MAP equalizer for SUD as introduced in 5.3.2, which utilizes forward-backward algorithm.

Now, start the derivation of the SMF approximation to the true distribution given in (7.28). The factorization over users can be expressed as:

$$q(\mathbf{r},\ \mathbf{a}) = q\Big(\mathbf{r},\ \mathbf{a}^{(1)},\ \mathbf{a}^{(2)},\cdots,\ \mathbf{a}^{(K)}\Big) \approx \prod_{k=1}^{K} q\Big(\mathbf{r},\ \mathbf{a}^{(k)}\Big) \tag{7.29}$$

where $q\Big(\mathbf{r},\ \mathbf{a}^{(k)}\Big)$ is the joint probability distribution of observations and nodes in the k-th chain. With this approximation, $\mathbf{a}^{(k)}$ is independent of $\mathbf{a}^{(k')}$ and $q\Big(\mathbf{r},\ \mathbf{a}^{(k)}\Big)$ is independent of $q\Big(\mathbf{r},\ \mathbf{a}^{(k')}\Big)$ given $k \neq k'$. As just discussed, each chain is modeled as a first order Markov process, thus by inserting (7.28) to (7.27), $q\Big(\mathbf{r},\ \mathbf{a}^{(k)}\Big)$ appears:

$$\begin{aligned}
q\Big(\mathbf{r},\ \mathbf{a}^{(k)}\Big) &\propto \exp\Big(\big\langle \ln q(\mathbf{r},\ \mathbf{a})\big\rangle_{q(\mathbf{r},\ \mathbf{a}\ \setminus\ \mathbf{a}^{(k)})}\Big) \\
&= \exp\left(\left\langle \ln\Big(P(\mathbf{a}_1)P(\mathbf{r}_1 \mid \mathbf{a}_1)\prod_{n=2}^{Ns} P(\mathbf{a}_n \mid \mathbf{a}_{n-1})P(\mathbf{r}_n \mid \mathbf{a}_n)\Big)\right\rangle_{q(\mathbf{r},\ \mathbf{a}\ \setminus\ \mathbf{a}^{(k)})}\right) \\
&= \exp\Big(\Big\langle \ln P(\mathbf{a}_1) + \sum_{n=2}^{Ns}\ln P(\mathbf{a}_n \mid \mathbf{a}_{n-1}) + \sum_{n=1}^{Ns}\ln P(\mathbf{r}_n \mid \mathbf{a}_n)\Big\rangle_{q(\mathbf{r},\ \mathbf{a}\ \setminus\ \mathbf{a}^{(k)})}\Big) \\
&= \exp\Big(\Big\langle \ln P(\mathbf{a}_1)\Big\rangle_{q(\mathbf{r},\ \mathbf{a}\ \setminus\ \mathbf{a}^{(k)})} + \sum_{n=2}^{Ns}\Big\langle \ln P(\mathbf{a}_n \mid \mathbf{a}_{n-1})\Big\rangle_{q(\mathbf{r},\ \mathbf{a}\ \setminus\ \mathbf{a}^{(k)})} \\
&\qquad + \sum_{n=1}^{Ns}\Big\langle \ln P(\mathbf{r}_n \mid \mathbf{a}_n)\Big\rangle_{q(\mathbf{r},\ \mathbf{a}\ \setminus\ \mathbf{a}^{(k)})}\Big)
\end{aligned} \tag{7.30}$$

where $q_T\big(\mathbf{r},\ \mathbf{a}\ \setminus\ \mathbf{a}^{(k)}\big)$ denotes the joint probability for the sequence of observations and all states except the nodes of the k-th chain. The terms inside the average that are independent of $q_T\big(\mathbf{r},\ \mathbf{a}\setminus\mathbf{a}^{(k)}\big)$ can be moved out of the average operation. Those terms that depends on $q_T\big(\mathbf{r},\ \mathbf{a}\ \setminus\ \mathbf{a}^{(k)}\big)$ give some constant after average, so that they can be disregarded. For instance, on the second last line in (7.30):

$$\begin{aligned}
\Big\langle \ln P(\mathbf{a}_1)\Big\rangle_{q(\mathbf{r},\ \mathbf{a}\ \setminus\ \mathbf{a}^{(k)})} &= \Big\langle \ln\big(P(\mathbf{a}_1^{(1)})P(\mathbf{a}_1^{(2)})\ldots P(\mathbf{a}_1^{(k)})\ldots P(\mathbf{a}_1^{(K)})\big)\Big\rangle_{q(\mathbf{r},\ \mathbf{a}\ \setminus\ \mathbf{a}^{(k)})} \\
&= \ln P(\mathbf{a}_1^{(k)})\Big\langle \ln\big(P(\mathbf{a}_1^{(1)})P(\mathbf{a}_1^{(2)})\ldots P(\mathbf{a}_1^{(K)})\big) + \Big\rangle_{q(\mathbf{r},\ \mathbf{a}\ \setminus\ \mathbf{a}^{(k)})} \\
&= \ln P(\mathbf{a}_1^{(k)}) + \text{constant} \\
&\propto \ln P(\mathbf{a}_1^{(k)})
\end{aligned} \tag{7.31}$$

Similar technique can be applied to $\Big\langle \ln P(\mathbf{a}_n \mid \mathbf{a}_{n-1}) \Big\rangle_{q(\mathbf{r},\ \mathbf{a}\ \setminus\ \mathbf{a}^{(k)})}$, so that it can be written as:

$$\Big\langle \ln P(\mathbf{a}_n \mid \mathbf{a}_{n-1}) \Big\rangle_{q(\mathbf{r},\ \mathbf{a}\ \setminus\ \mathbf{a}^{(k)})} \qquad \propto \qquad P(\mathbf{a}_n^{(k)} \mid \mathbf{a}_{n-1}^{(k)}) \tag{7.32}$$

Insert (7.31) and (7.32) to (7.30), $\ln q\Big(\mathbf{r},\ \mathbf{a}^{(k)}\Big)$ is then reduced to:

$$\ln q\Big(\mathbf{r},\ \mathbf{a}^{(k)}\Big) \propto \ln P(\mathbf{a}_1^{(k)}) + \sum_{n=2}^{Ns} \ln P(\mathbf{a}_n^{(k)} \mid \mathbf{a}_{n-1}^{(k)}) + \sum_{n=1}^{Ns} \Big\langle \ln P(\mathbf{r}_n \mid \mathbf{a}_n) \Big\rangle_{q(\mathbf{r},\ \mathbf{a}\ \setminus\ \mathbf{a}^{(k)})} \tag{7.33}$$

Until here, (7.33) shows that $\ln q\Big(\mathbf{r},\ \mathbf{a}^{(k)}\Big)$ consists of one part that is independent of $q(\mathbf{r},\ \mathbf{a}\ \setminus\ \mathbf{a}^{(k)})$ and another part that depends on $q(\mathbf{r},\ \mathbf{a}\ \setminus\ \mathbf{a}^{(k)})$. In order to clearly see the similarity between (7.33) and (7.28), modify (7.28) to a closer form as in (7.33), that is:

$$\begin{aligned} \ln q(\mathbf{r},\ \mathbf{a}) &= \ln P(\mathbf{a}_1)\Big(\prod_{n=2}^{Ns} P(\mathbf{a}_n \mid \mathbf{a}_{n-1})\Big)\Big(\prod_{n=1}^{N_s} P(\mathbf{r}_n \mid \mathbf{a}_n)\Big) \\ &= \ln P(\mathbf{a}_1) + \sum_{n=2}^{Ns} P(\mathbf{a}_n \mid \mathbf{a}_{n-1}) + \sum_{n=1}^{N_s} P(\mathbf{r}_n \mid \mathbf{a}_n) \end{aligned} \tag{7.34}$$

Comparing (7.33) and (7.34), when k is fixed, the differences between the exact solution and the SMF approximation are the length of the state vector and the averaged log likelihood expression, i.e. $\Big\langle \ln P\big(\mathbf{r}_n \mid \mathbf{a}_n\big) \Big\rangle_{q_T\left(\mathbf{r},\ \mathbf{a}\ \setminus\ \mathbf{a}^{(k)}\right)}$.

Log likelihood expression can be found from (5.31) and (5.32) and skip the notation for transition $(p,\ q)$ here. $\boldsymbol{\xi}_n$ denotes the more general hypothesized symbol vector, that is:

$$\boldsymbol{\xi}_n = \Big[\hat{s}_n^{(1)}, \underbrace{\hat{s}_{n-1}^{(1)}, \cdots, \hat{s}_{n-L+1}^{(1)}}_{\mathbf{a}_n^{(1)}}, \hat{s}_n^{(2)}, \underbrace{\hat{s}_{n-1}^{(2)}, \cdots, \hat{s}_{n-L+1}^{(2)}}_{\mathbf{a}_n^{(2)}}, \cdots, \hat{s}_n^{(K)}, \underbrace{\hat{s}_{n-1}^{(K)}, \cdots, \hat{s}_{n-L+1}^{(K)}}_{\mathbf{a}_n^{(K)}}\Big]^T$$

The log likelihood expression is shown as:

$$\ln P\Big(\mathbf{r}_n \mid \mathbf{a}_n\Big) \ \propto\ \frac{-1}{2\sigma^2}\Big(\mathbf{r}_n - \mathbf{B}\boldsymbol{\xi}_n\Big)^H\Big(\mathbf{r}_n - \mathbf{B}\boldsymbol{\xi}_n\Big) \tag{7.35}$$

Multiply out right hand-side of (7.35), it consists of one part $E\Big(\boldsymbol{\xi}_n\Big)$ that depends on $\boldsymbol{\xi}_n$ and a constant Z_T that is independent of $\boldsymbol{\xi}_n$. Replace σ^2 with T like for the FFMF receiver, which will be used as a tuning parameter to compensate approximation error. Further, from now on, subscript T is added to the posterior joint distribution $q(\cdot)$ to represent that it is the distribution at T.

$$E(\boldsymbol{\xi}_n) = \frac{1}{2}\boldsymbol{\xi}_n^{(k)H}\mathbf{D}\boldsymbol{\xi}_n - 2\mathrm{Re}\{\mathbf{r}_n^H\mathbf{B}\boldsymbol{\xi}_n\}\Big) \tag{7.36}$$

where:

$$\mathbf{D} = \mathbf{B}^H\mathbf{B} - \mathrm{diag}(\mathbf{B}^H\mathbf{B}) \qquad\qquad \mathbf{D} \in \mathbb{C}^{KL\times KL}$$

Note that the objective of setting diagonal of D to zero is similar as for the FFMF receiver, that is to avoid self coupling between symbols in (7.36). The constant Z_T that is independent of ξ_n appears:

$$Z_T = \frac{1}{2T}\Big(\sum_{i=1}^{KL}\mathbf{B}^H\mathbf{B}(i,i) + \mathbf{r}_n^H\mathbf{r}_n\Big) \tag{7.37}$$

Insert (7.36) (7.37) to (7.35), we get:

$$\ln P\Big(\mathbf{r}_n \mid \mathbf{a}_n\Big) \;\propto\; \frac{-1}{T}E\Big(\boldsymbol{\xi}_n\Big) - Z_T \tag{7.38}$$

Next, let's investigate how to derive the averaged log likelihood expression. Let $\mu_n^{(k)} = \big\langle s_n^{(k)}\big\rangle_T$ be the averaged symbol with respect to $q_T\big(\mathbf{r},\ \mathbf{a}\ \setminus\ \mathbf{a}^{(k)}\big)$. With this notation, let vector $\bar{\boldsymbol{\xi}}_n^{(k)}$ represent the averaged $\boldsymbol{\xi}_n$ over $q_T\big(\mathbf{r},\ \mathbf{a}\ \setminus\ \mathbf{a}^{(k)}\big)$, e.g. if $k=2$, $\bar{\boldsymbol{\xi}}_n^{(2)}$ appears:

$$\bar{\boldsymbol{\xi}}_n^{(2)} = \Big[\overbrace{\mu_n^{(1)},\mu_{n-1}^{(1)},\cdots,\mu_{n-L+1}^{(1)}}^{\text{averaged symbol}},\underbrace{\hat{s}_n^{(2)},\hat{s}_{n-1}^{(2)},\cdots,\hat{s}_{n-L+1}^{(2)}}_{\text{hypothesized symbol}},\overbrace{\cdots,\mu_n^{(K)},\mu_{n-1}^{(K)},\cdots,\mu_{n-L+1}^{(K)}}^{\text{averaged symbol}}\Big]^T$$

From (7.38), the averaged term $\Big\langle \ln P\big(\mathbf{r}_n \mid \mathbf{a}_n\big)\Big\rangle_{q_T(\mathbf{r},\ \mathbf{a}\ \setminus\ \mathbf{a}^{(k)})}$ in (7.33) is found by:

$$\Big\langle \ln P\big(\mathbf{r}_n \mid \mathbf{a}_n\big)\Big\rangle_{q_T(\mathbf{r},\ \mathbf{a}\ \setminus\ \mathbf{a}^{(k)})} \;\propto\; \frac{-1}{2T}E\Big(\bar{\boldsymbol{\xi}}_n^{(k)}\Big) - Z_T \tag{7.39}$$

Inserting (7.39) to (7.33), $q_T\Big(\mathbf{r},\ \mathbf{a}^{(k)}\Big)$ is found by a modified SbS MAP equalizer for SUD. The modification is to use averaged $\bar{\boldsymbol{\xi}}_n^{(k)}$ instead of $\boldsymbol{\xi}_n$ to calculate the averaged likelihood. Again, use $k=2$ as an example, that is:

$$\bar{\boldsymbol{\xi}}_n^{(2)} = \Big[\overbrace{\mu_n^{(1)},\mu_{n-1}^{(1)},\cdots,\mu_{n-L+1}^{(1)}}^{\text{posterior mean}},\hat{s}_n^{(2)},\underbrace{\hat{s}_{n-1}^{(2)},\cdots,\hat{s}_{n-L+1}^{(2)}}_{\text{MAP state vector}},\overbrace{\cdots,\mu_n^{(K)},\mu_{n-1}^{(K)},\cdots,\mu_{n-L+1}^{(K)}}^{\text{posterior mean}}\Big]^T$$

Finally, iteratively update all users until their posterior converge to some criteria, e.g. a sufficiently small difference between new and old posterior. After all the posterior means are converged, if using hard decision, the estimated symbol from SMF receiver is:

$$\breve{s}_n^{(k)} = \operatorname{sgn}\big(\mu_N^{(k)}\big) \tag{7.40}$$

To summarize how SMF receiver works, let's get back to those four steps given in the beginning of this chapter.

1. Initialize $\mu_n^{(k)}$ by the prior. Same as what has been discussed for the first step as in the FFMF receiver.

2. Use the averaged log likelihood expression given in (7.39) in the SbS MAP equalizer to update the current user. If user 2 is the currently updated user, the SMF updates posterior means as follows:

$$\Big[\overbrace{\mu_1^{(1)},\mu_2^{(1)},\cdots,\mu_{Ns}^{(1)}}^{\text{previous updated}},\underbrace{\mu_1^{(2)},\mu_2^{(2)},\cdots,\mu_{Ns}^{(2)}}_{\text{current update}},\overbrace{\mu_1^{(3)},\mu_2^{(3)},\cdots,\mu_{Ns}^{(3)},\cdots,\mu_1^{(K)},\mu_2^{(K)},\cdots,\mu_{Ns}^{(K)}}^{\text{previous updated}}\Big]^T$$

3. Update other user's burst as in step 2 but with newly updated posterior means from previous iteration. Change user index until all K bursts are updated once.

4. Repeat step 2 and 3 until the stop criteria are met.

In the end, posterior mean can be used as soft bit, e.g. input for channel decoder. Estimated symbol can be found by hard decision (7.40).

In $\bar{\boldsymbol{\xi}}_n^{(k)}$, only the k-th user part changes, which means the number of states is reduced from $|\Omega|^{K(L-1)}$ to $|\Omega|^{L-1}$, thus the complexity is changed from $\mathcal{O}(|\Omega|^{KL})$ to $\mathcal{O}(N_{ITE}K|\Omega|^L)$

7.2.2 Simulation

The performance of FFMF and SMF receivers are compared in Figure 7.8. The test case is the same as for Figure 7.2, that is 1 CCI (GMSK), in the TU3 channel. It shows that the SMF receiver performs quite similar to the FFMF receiver without prediction function in most of the tested CIR range and better at low CIR. But it is worse than FFMF receiver with prediction function at higher CIR values. Simulation shows that the SMF receiver can converge within 4 iterations whereas the FFMF solution normally requires 10 iterations. The objective of the SMF receiver is to keep more structure of the exact solution. Thus most of the local minima of the objective function can be rejected. Simulations also show that the SMF receiver gives much faster convergence speed than the FFMF solution. But, it does not necessarily mean that the SMF approximation should give better BER. Because, it is also an approximation to the exact solution and it is difficult to predict which approximation is better.

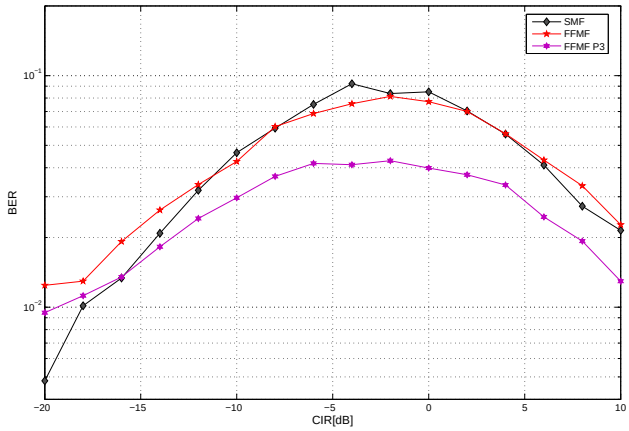
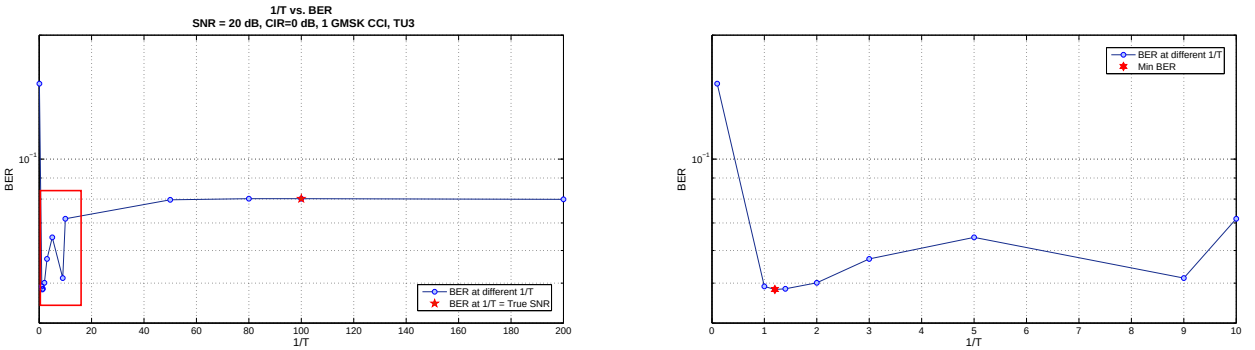


Figure 7.8: Comparison of FFMF and SMF receivers in TU3 channel, 1 GMSK CCI,

7.2.3 Investigation on control parameter T

Like what has been done in 7.1.4, same experiment is conducted for the SMF receiver. The results are shown in Figure 7.9. Similar conclusion as the FFMF receiver can be drawn. T influences BER performance and increasing T gives better BER. However, unlike the FFMF receiver, the optimal T range is quite narrow, that is from 1 to 2. As analyzed in 7.1.4, there is usually CIR and SNR estimation error from the other module in the receiver. So the condition to implement such a prediction function as in FFMF is a comparably wide optimal T range. Therefore, SMF receiver does not meet this condition.



(a) *Range from* $1/T = 0.1$ *to* $1/T = 200$ (b) *Finer search in the range* $1/T = 0.1$ *to* $1/T = 10$

Figure 7.9: Investigation on the influence of control parameter T

7.3 Summary

Two suboptimal receivers are derived in this chapter. They are based on the Mean Field theory. In general, the mean field approximation finds the local minima of the objective (Free energy) by a coordinate descent step, i.e. $\frac{\partial \mathcal{F}_T}{\partial m^{(k)}_{n,\,T}}$. But it can be slow. The goal is to avoid local minima and find

the global minimum of the cost function.

The FFMF method is the first approximation to the exact solution introduced in this chapter. It fully factorizes posterior distribution so that forward-backward recursions are avoided. Due to ISI, a window of received samples are needed to estimate one symbol. The FFMF receiver is very simple so the computation cost is low. Simulation time per burst is something similar to the LMMSE receiver and much shorter than the exact solution. The complexity of FFMF receiver is $\mathcal{O}\big(N_x + N_+ + N_{tanh}\big)$. Its performance can be improved by increasing the control parameter T, where T is the noise power. Increasing T corresponds to model the approximation as white noise. Thus the approximation error can be compensated by increasing T. Simulations show that the optimum T can be predicted from CIR and SNR and it is independent of channel types. Two prediction functions are implemented and both give a BER gain. Two times oversampling provides BER gains for the FFMF receiver, especially when ACI is present. The FFMF receiver can also be used for higher order modulation, but it is more complex.

The SMF receiver is derived by only factorizing posterior distribution over the user index. The first order HMM properties of the exact solution are kept. The complexity of the SMF solution is $\mathcal{O}(N_{ITE}K|\Omega|^L)$. It provides faster convergence speed, however the BER performance is close to that of the FFMF solution in most of the tested CIR range and better than that of the FFMF receiver at low CIR value. T can be used as a control parameter to improve the BER performance. However, simulation shows that the optimum T range is narrow. So implement a prediction function to estimate the optimum T is not practical for the SMF solution. Because there is also some estimation error for SNR and CIR.

Simulations and investigations on the control parameter T show that FFMF receiver offer a very good tradeoff between complexity and performance.

Chapter 8

Conclusion and Future Work

The motivation of this project is the ever increasing demand to add more subscribers to the GSM system. Since it is a cellular network, this need can be met by tightening the frequency reuse factor when planning the network. However, this will increase interference from other users, thus degrade signal conditions. Conventional GSM receivers model noise as AWGN, which may fail in poor signal conditions because noise is not white anymore. Therefore, advanced receivers that provide better data detection in interference dominated networks are in need. MUD is one IC technique. This project has investigated optimal and suboptimal MUD GSM receivers.

8.1 Project Summary

Background knowledge such as common digital communication system blocks, GSM basics, CCI and ACI were given in Chapter 2. Chapter 3 discussed digital phase modulation. Compared to other type of modulation methods, constant envelope, narrow main lobe and low side lobes in the spectrum are the reasons why GMSK was selected for GSM. The physical properties of wave-forms cause radio channel to be unpredictable. The mobile radio channel is normally modeled in a statistical fashion and the signal model in multi-path channel with AWGN was given in Chapter 4.

Conventional receivers such as Matched Filter, LMMSE and ML/MAP were studied in Chapter 5. Matched Filter is only optimal in non-ISI channels, LMMSE is the optimal linear receiver for multi-path channels with AWGN. The SbS MAP equalizer is the optimal non-linear symbol detector for AWGN ISI channel. By redefining the state vector to include more users, the Joint MAP equalizer is derived. In Chapter 6, the exact MUD solution is verified. BER performance is very impressive, however on the cost of huge complexity $\mathcal{O}(|\Omega|^{KL})$. It is nowhere realistic to be implemented in the MS. Approximations which have reasonable computation cost with the compromise of BER are needed.

Most work of this project was presented in Chapter 7, two suboptimal receivers were discussed. The FFMF receiver reduces complexity to $\mathcal{O}\big(KN_{ITE}(N_x + N_+ + N_{tanh})\big)$, where N_{ITE} is normally around 10. Besides numerous tests were performed to investigate the way control parameter T influences BER. The conclusion is that BER performance can be enhanced by adjusting the control parameter T. There is an optimal T range, which can be predicted by CIR and SNR. CIR and SNR are normally estimated for other purposes in the receiver, so they can be reused in the FFMF receiver. This means no extra cost is required for adding prediction function. Because the optimal T range is not narrow, the prediction function does not highly rely on the accuracy of SNR and CIR

estimation. Two prediction functions were implemented and they both give high BER gains. Two times oversampling is suggested for the FFMF receiver. Simulation shows a gain compared to no over sampling, especially for ACI. Another interesting point is that the complexity of the FFMF receiver is not sensitive to the constellation size. It gives rise to the application of the FFMF approximation for other modulation types that have bigger constellation size, e.g. EDGE. The computation cost will be higher than the smaller constellation, but will not be the same situation as the exact solution, i.e. $\mathcal{O}(|\Omega|^{KL})$.

The SMF receiver was also derived and analyzed in Chapter 7. It is an approximation to the exact solution, which keeps the first order HMM model for each chain. The complexity of the SMF receiver is $\mathcal{O}(KN_{ITE}|\Omega|^{L})$, where N_{ITE} is normally around 4. It gives faster convergence speed but provides very close BER performance compared to that of the FFMF receiver without prediction function. Depending on L and K, the SMF approximation may require more computations than FFMF, which indicates the FFMF solution is an attractive approximation in terms of complexity/performance tradeoff.

Joint channel estimation is needed for MUD algorithm, a brief description of JLS channel estimation with known TSs is available in Appendix A. Simulations showed that the degradation of BER due to channel estimation errors was not significant. MUD receivers with an estimated channels still outperform conventional receivers. Whitening filter is the common SAIC solution now. The basic concept of whitening approach is included in Appendix B, the performance of whitening solution and other receivers is compared in the simulation plots in Appendix B.

8.2 Further Discussion

To finish the project in the given time, the scope of this thesis work is focus on topics defined in Section 1.3. However, the following issues should also be considered for practical application of MUD solution in the present GSM network.

1. Asynchronous networks: There are many discussions about synchronizing BSs in GSM network, as it shows big advantages of increased capacity. But networks are not synchronized yet. That is, interference from other cells can occur at any time during a desired burst, as shown in Figure 2.8. Therefore, estimating the timing of interferers is important for MUD receivers or other IC algorithms. Since a TS exists in any GSM burst. Timing estimation of each interfering burst can be accomplish by correlating the received burst with all 8 TSs, this will require a complexity of $\mathcal{O}(8 * 148 * 26)$. A more clever method is needed.

2. Joint channel estimation with unknown TS: In GSM, only the desired TS is known by the receiver. Channel estimation for interferers relies on known TS. Blind channel estimation or semi-blind channel estimation is worth to be investigated. Since computation cost of JLS channel estimation is not high, one way to solve this problem can be to try out all 8 possible TSs and calculate the residual error for the TS part. Then selecting the one that gives minimum residual error.

3. Determine dominant interferers: As shown in the previous chapter, it is a good idea to combine the MUD receiver with a whitening filter, i.e. to detect desired and strong interferers, and whiten the weak interferers. Therefore, a method to identify dominant interferers is interesting for this usage.

4. Interference of other modulation types, e.g. EDGE: In this thesis, only GMSK modulated interferers are considered. It is also interesting to see how MUD receivers perform in GMSK and 8-PSK mixed scenarios.

In GSM, users are separated from the uncorrelated propagation channel, but in CDMA systems, spreading codes can also be used for MUD. Therefore, MUD algorithms for GSM should work better for CDMA or WCDMA. Therefore, the MUD receiver for CDMA or WCDMA is also an interesting topic for investigation in the future work.

Bibliography

[1] Theodore S. Pappaport. *Wireless Communications: Principles and Practive*. Prentice Hall, 1996.

[2] Proakis J G. *Digital Communications*. McGraw-Hill, 3rd edition, 1995.

[3] Sergio Verdu. *Multiuser Detection*. Cambridge University Press, 1998.

[4] Gregory E. Bottomley and Karim Jamal. Adaptive Arrays and MLSE Equalization. *IEEE*, 1995.

[5] Lars P.B. Christensen. Minimum Symbol Error Rate Detection in Single-Input Multiple-Output Channels with Markov Noise. *SPAWC*, 2005. Technical University of Denmark.

[6] Michel Mouly and Marie-Bernadette Pautet. *The GSM System for Mobile Communications*. Cell & SYS, 1992.

[7] History of GSM from GSM WORLD. http://www.gsmworld.com/about/history/index.shtml.

[8] Peter A. Murphy. The GSM Physical Layer. GSM radio channel overview, July 23 1996.

[9] *3GPP TS 45.005*, January 2005. Annex L.

[10] Timo Halonen, Javier Romero, and Juan Melero, editors. *GSM, GPRS and EDGE Performance*. John Wiley & Sons, 2003.

[11] Ziemer R.E. and Peterson R.L. *Introduction to Digital Communications*. Macmillan Publishing Company, 1992.

[12] Richard Booth. A Note on the Application of the Laurent Decomposition to the GSM GMSK Signaling Waveform. Tropian, Inc.

[13] Irving Kalet. *Mordern Digital Modulation Techniques*, chapter VII. CEI-Europe, 2000.

[14] Pierre A. Laurent. Exact and Approximate Construction of Digital Phase Modulations by Superposition of Amplitude Modulated Pulses (AMP). *IEEE Transactions on Communications*, COM-34(2):150–160, February 1986.

[15] Lawrence R. Rabiner. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. *Proceedings of the IEEE*, 77(2):257–286, February 1989.

[16] Thomas Arildsen and Jesper Blauendahl. Advanced Channel Estimation and Multiuser Detection in GSM. Master's thesis, Aalborg University, Institute of Electronic Systems, Department of Communication Technology, 2004.

[17] Jørn Justesen and Knud J. Larsen. Notes for 34230 Telecommunication. COM center, DTU.

[18] Harri Holma and Antti Toskala. *WCDMA FOR UMTS - Radio Access For Third Generation Mobil Communications*. John Wiley & Sons, 2000.

[19] Alexander Kocian and Bernard H. Fleury. EM - Based Joint Data Detection and Channel Estimation of DS-CDMA Signals. *IEEE Transactions on Communications*, 51(10):1709–1720, October 2003.

[20] T.K. Moon. The expectation-maximization algorithm. *IEEE Signal Processing Magazine*, 13(6):47–60, November 1996.

[21] Lars P.B. Christensen. Signal Processing for UWB System. Master's thesis, Informatics and Mathematical Modelling, DTU, 2003.

[22] Wolfgang H. Gerstacker, Robert Schober, and Alexander Lampe. Receivers With Widely Linear Processing for Frequency-Selective Channels. *IEEE Transactions on Communications*, 51(9):1512–1523, September 2003.

[23] Andrew.J. Viterbi. Error bounds for convolutional codes and an asymtotically optimum decoding algorithm. *IEEE Transactions on Information Theory*, IT-13(2):260–269, April 1967.

[24] L.R. BAHL, J. Cocke, F. Jelinek, and J. Raviv. Optimal Decoding of Linear Codes for Minimizing Symbol Error Rate. *IEEE Transations on Information Theory*, pages 284–287, March 1974.

[25] John R. Barry. The BCJR Algorithm for Optimal Equalization. School of Electrical and Computer Engineering, Georgia institute of Technology, March 2nd 2000.

[26] Fathi Raouafi, Adodo Dingninou, and Claude Berrou. Saving memory in turbo-decoders using the Max-Log-MAP algorithm. *IEE, Savoy Place, London WC2R OBI, UK*, pages 14/1–14/4, 1999. The Institudtion of Electrical Engineers.

[27] E. Baccarelli, A. Fasano, S. Gall, and A. Zucchi. A Novel Reduced-Complexity MAP equalizer using sofe-statistics for Decision-Feedback ISI cancellation. INFO-COM Dpt., University of Rome "La Sapienza", Via Eudossiana 18, 00184 Rome, Italy; Telcordia Technologies, formerly Bellcore, 445 South Street, Morristown, NJ 07960, USA.

[28] Javan Erfanian, Subbarayan Pasupathy, and Glenn Gulak. Reduced Complexity Symbol Detectors with Parallel Structures for ISI Channels. *IEEE Transactions on Communications*, 42(2/3/4):1661–1671, April 1994.

[29] Wolfgang Koch and Alfred Baier. Optimum and Sub-optimum Detection of Coded Data Disturbed by Time-varying InterSymbol Interference. *IEEE*, pages 1679–1684, 1990. Philips Kommunikations Industrie AG, Nurnberg Federal Republic of Germany.

[30] Abdulrauf Hafeez and Wayne E. Stark. Decision Feedback Sequence Estimation for Unwhitened ISI channels with applications to Multiuser Detection. *IEEE Journal on Selected Areas in communications*, 16(9):1785–1975, December 1998.

[31] Manfred Opper and David Saad, editors. *Advanced Mean Field Methods Theory and Practice*. Neural Information Processing, 2001.

[32] Thomas Fabricius and Ole Winther. Correcting the Bias of Subtractive Interference Cancellation in CDMA: Advanced Mean Field Theory. Submitted to IEEE Transactions on Information Theory, 2003.

[33] Thomas Fabricius and Ole Winther. Analysis of Mean Field Annealing in Subtractive Interference Cancellation. Submitted to IEEE Transactions on Communications, 2002.

[34] Thomas Fabricius and Ole Winther. Improved Multistage Detector by Mean-Field Annealing in Multi-User CDMA. *Proceedings of IEEE Intelligent Signal Processing, Applications, and Communication Systems*, 2002.

[35] Zoubin Ghahramani and Michael I. Jordan. Factorial Hidden Markov Models. *Machine Learning*, pages 245–275, 1997. Kluwer Academic Publishers, Boson. Manufactured in The Netherlands.

[36] Pedro A.d.F.R Højen-Sørensen. *Approximating methods for intractable probabilistic models: Applications in Neuroscience*. PhD thesis, Technical University of Denmark, DK-2800 Kongens Lyngby, Denmark, April 2001.

[37] Michel Goosens, Frank Mittlebach, and Alexander Samarin. *The LaTeX Companion*. Addison-Wesley Publishing Company, second edition, May 1994.

[38] Simon Haykin. *Adaptive Filter Theory*. Prentice Hall, 3rd edition, 1996.

[39] Markku Pukkila. Channel Estimation Modeling. Technical report, Nokia Research Center, December 2000. S-72.333 Postgraduate Course in Radiocommunications.

[40] S.M.Kay, editor. *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice Hall, 1998. 595p.

[41] Markku Pukkila. Channel Estimation of Multiple Co-channel signals in GSM. Master's thesis, Helsinki University of Technology, 1997.

[42] *3GPP TS 0.5.05 v8. 15.0*, April 2003. Annex C.

Appendix A

Channel Estimation for MUD

Detection algorithms need knowledge of the channel impulse response, which can be provided by a separate channel estimator. There are two types of channel estimator: blind estimator and non-blind estimator. The blind channel estimator estimates the channel impulse response without any known symbol while the non-blind one uses known information to estimate channel parameters. For a given number of operations, the blind estimator is normally worse than the non-blind estimator. Moreover, as explained in 2.2.2, GSM system always transmits a TS in the middle of a TDMA burst. Therefore the channel estimation is possible by exploiting the known TS and the corresponding received samples. The channel impulse response usually changes burst by burst due to the multi-path fading channel, it is therefore estimated for each burst. There are various channel estimation algorithms, which are similar like the data detection algorithms. For detection algorithms, the unknown information is the symbols and the channel is considered known. For the channel estimator, the channel impulse response is unknown and needs to be estimated from the known data.

The Least Square (LS) channel estimation for SUD is introduced first, then the Joint Least Square (JLS) channel estimator for one desired user and one CCI is discussed. Finally simulation results are presented.

A.1 Least Square (LS) Channel Estimation For SUD

The conventional GSM receivers in Chapter 5 only need channel parameters for the desired user, this can be obtained by a separate channel estimator for the single user. For the AWGN channel, it is shown that the optimum estimator is linear. The concept of the LS [38] method is the same as for the LMMSE receiver introduced in Section 5.2, but the unknown parameter is the channel coefficients. Channel parameters are estimated from the received samples that correspond to the TS. The TS starts from the 62-th symbol in a burst and contains 26 known symbols.

Skip over sampling factor N_{sps} in the discussion here. Let $\mathbf{y}_{ce}$ hold the received samples that are only associated with TS part of the burst and $\mathbf{M}_{LS}$ hold TS. $\mathbf{h}_{LS}$ is the channel impulse response vector that needs to be estimated. $\boldsymbol{\epsilon}_{ce}$ contains the white noise samples corresponding to $\mathbf{y}_{ce}$. The signal model for LS estimation can be expressed as:

$$\mathbf{y}_{ce} = \mathbf{M}_{LS}\mathbf{h}_{LS} + \boldsymbol{\epsilon}_{ce} \tag{A.1}$$

Let $L + P = 26$, where L is the channel impulse response length, P is the number of received samples that are purely generated from the TS and 26 is the length of TS. With the signal model and definitions of r_n, $h_l^{(k)}$, $s_n^{(k)}$ and ϵ_n given in Section 4.3, $\mathbf{y}_{ce}$, $\boldsymbol{\epsilon}_{ce}$, $\mathbf{h}_{LS}$ and $\mathbf{M}_{LS}$ appear:

$$\mathbf{y}_{ce} = \begin{bmatrix} r_{62+L-1} \\ r_{62+L} \\ \vdots \\ r_{62+P+L-2} \end{bmatrix} \quad \boldsymbol{\epsilon}_{ce} = \begin{bmatrix} \epsilon_{62+L-1} \\ \epsilon_{62+L} \\ \vdots \\ \epsilon_{62+P+L-2} \end{bmatrix} \quad \mathbf{h}_{LS} = \begin{bmatrix} h_0^{(1)} \\ h_1^{(1)} \\ \vdots \\ h_{L-1}^{(1)} \end{bmatrix} \quad \mathbf{y}_{ce},\ \boldsymbol{\epsilon}_{ce} \in \mathbb{C}^P, \quad \mathbf{h}_{LS} \in \mathbb{C}^L$$

$$\mathbf{M}_{LS} = \begin{bmatrix} s_{62+L-1}^{(1)} & \cdots & s_{63}^{(1)} & s_{62}^{(1)} \\ s_{62+L}^{(1)} & \cdots & s_{64}^{(1)} & s_{63}^{(1)} \\ \vdots & \ddots & \vdots & \vdots \\ s_{62+L+P-2}^{(1)} & \cdots & s_{62+P}^{(1)} & s_{62+P-1}^{(1)} \end{bmatrix} \qquad \mathbf{M}_{LS} \in \mathbb{R}^{P \times L}$$

The LS rule is to minimize the square error, that is:

$$\hat{\mathbf{h}}_{LS} = \text{argmin} \left[\mid \mathbf{y}_{ce} - \mathbf{M}_{LS}\mathbf{h}_{LS} \mid^2 \right]_n \tag{A.2}$$

where $\hat{\mathbf{h}}_{LS}$ denotes the estimated channel coefficients from the LS estimator. In the AWGN channel, the estimated channel parameters from the LS estimator are found by [39, 40]:

$$\hat{\mathbf{h}}_{LS} = \left(\mathbf{M}_{LS}^H \mathbf{M}_{LS} \right)^{-1} \mathbf{M}_{LS}^H \mathbf{y}_{ce} \tag{A.3}$$

(A.3) is also the Best Linear Unbiased Estimate (BLUE) for the channel coefficients [39, 40]. In GSM, the 16-bit autocorrelation of any TS has the property that the autocorrelation is large at lag 0, and close to zero for other lags. It means that $\mathbf{M}_{LS}^H \mathbf{M}_{LS}$ is a diagonal matrix when $P = 16$ [39, 41]. In this case, (A.3) is reduced to:

$$\hat{\mathbf{h}}_{LS} = \frac{1}{P} \mathbf{M}_{LS}^H \mathbf{y}_{ce} \tag{A.4}$$

The comparison for the performance of the IQ LMMSE receiver in the TU50 channel with LS channel estimation and known channel parameters is seen in Figure A.1(a). LS5 means the LS channel estimation with $L = 5$.

A.2 JLS Channel Estimation for Two Co-Channel Users

The LS channel estimator only supports single user. Joint channel estimation is in need in order to perform MUD. The JLS channel estimator is derived from the LS method discussed in the above

section. Two co-channel synchronized users are considered in this section. The channel impulse response vector $\mathbf{h}_{LS}$ and TS matrix $\mathbf{M}_{LS}$ are expanded to $\mathbf{h}_{JLS}$ and $\mathbf{M}_{JLS}$ to cover two users. $\mathbf{y}_{ce}$ and $\boldsymbol{\epsilon}_{ce}$ are unchanged. The signal model for the JLS estimator is:

$$\mathbf{y}_{ce} = \mathbf{M}_{JLS}\mathbf{h}_{JLS} + \boldsymbol{\epsilon}_{ce} \tag{A.5}$$

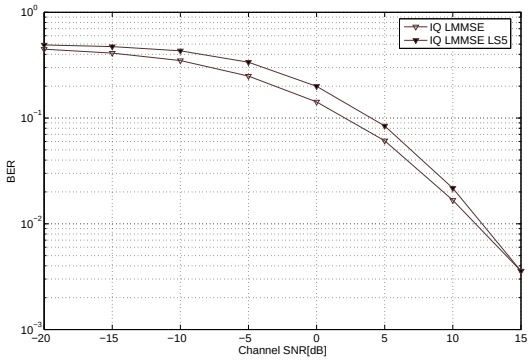
where:

$$\mathbf{h}_{JLS} = \left[h_0^{(1)}, h_1^{(1)}, \cdots, h_{L-1}^{(1)}, h_0^{(2)}, h_1^{(2)}, \cdots h_{L-1}^{(2)}\right]^T \qquad \mathbf{h}_{LS} \in \mathbb{C}^{2L}$$

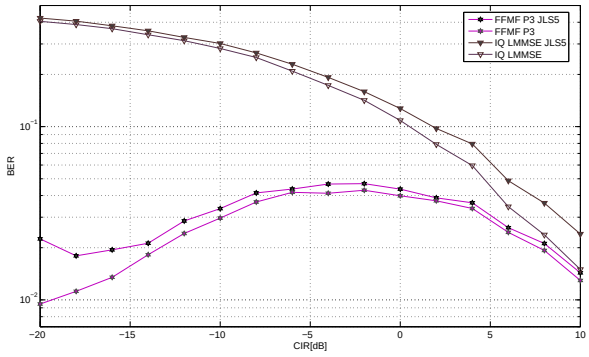
$$\mathbf{M}_{JLS} = \begin{bmatrix} s^{(1)}_{62+L-1} & \cdots & s^{(1)}_{63} & s^{(1)}_{62} & s^{(2)}_{62+L-1} & \cdots & s^{(2)}_{63} & s^{(2)}_{62} \\ s^{(1)}_{62+L} & \cdots & s^{(1)}_{64} & s^{(1)}_{63} & s^{(2)}_{62+L} & \cdots & s^{(2)}_{64} & s^{(2)}_{63} \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ s^{(1)}_{62+L+P-2} & \cdots & s^{(1)}_{62+P} & s^{(1)}_{62+P-1} & s^{(2)}_{62+L+P-2} & \cdots & s^{(2)}_{62+P} & s^{(2)}_{62+P-1} \end{bmatrix} \qquad \mathbf{M}_{JLS} \in \mathbb{R}^{P\times 2L}$$

Like for the LS channel estimation, the estimated channel coefficients in the AWGN channel are found by [39]:

$$\hat{\mathbf{h}}_{JLS} = \text{argmin}\left[\mid \mathbf{y}_{ce} - \mathbf{M}_{JLS}\mathbf{h}_{JLS} \mid^2\right]_n = \left(\mathbf{M}_{JLS}^H\mathbf{M}_{JLS}\right)^{-1}\mathbf{M}_{JLS}^H\mathbf{y}_{ce} \tag{A.6}$$



(a) *LS*, *GMSK*, *TU*50



(b) *JLS*, 1 $CCI(GMSK)$, $SNR = 20$ *dB*, *TU*3

Figure A.1: Simulations of Channel Estimation, IQ LMMSE filter length=11

For the noisy channel estimation, SNR degrades due to the estimation error. Thus the performance of receivers with the channel estimation is worse than that with the perfect channel parameters. Simulations of the JLS for the IQ LMMSE solution and the FFMF receiver with the prediction function are presented in Figure A.1 (b). JLS5 denotes the JLS channel estimator with $L = 5$. Both the LS estimator and the JLS solution produce slightly worse BER than that of the known channel impulse responses. It is because the estimation error. However as shown by the BER curves

in Figure A.1 (a) and (b), this estimation error is not big therefore the receiver BER performance degradation is not significant.

The LS channel estimator and the JLS channel estimator need the knowledge of known TSs. But in GSM, only the desired user's TS is known, the TS information for the interferers needs to be estimated. This can be achieved by correlating the received burst with all 8 TSs. For synchronized case, it requires the computation cost of $\mathcal{O}(8 * 26)$ per interferer, where 8 is the number of TSs and 26 is the length of one TS. For asynchronized case, the complexity is $\mathcal{O}(8 * 148 * 26)$ for each interferer, where 148 is the burst length.

The JLS channel estimation for more synchronized co-channel users is similar to the one just introduced. The difference is the bigger TS matrix $\mathbf{M}_{JLS}$ and larger channel impulse response vector $\mathbf{h}_{JLS}$. The performance of the JLS estimator depends on the cross correlation properties between the TSs. The JLS channel estimation for asynchronized co-channel users or adjacent channel users is more complex.

Appendix B

Whitening Approach

This project focused on SAIC MUD solution, therefore the whitening approach is more relevant to the scope of this project compared to the multiple antenna solution. This appendix gives an overview of the whitening IC approach. This overview is based on Lars Christensen's publication [5]. As indicated in Section 1.2, the whitening solution is to model the interference as the colored noise and convert it to the white noise. It is a single user detection method and the challenge is how to estimate the statistical properties of the interfering signals.

B.1 Whitening Filter Overview

Section 5.3 demonstrated that the optimal symbol detector in the multi-path channel with memoryless noise is the BCJR algorithm [24]. It is shown in [5], the co-channel interference can be well approximated by the Gauss-Markov noise model. The whitening solution is an extension to the BCJR algorithm, which includes the Markov noise model. Let N_m denote the memory of the noise. Recall that L is the channel impulse response length and therefore the channel has memory $L-1$. For the memoryless noise, the SbS MAP equalizer introduced in Section 5.3 has $|\Omega|^{L-1}$ number of states. The Gauss-Markov process introduces N_m memory to the noise. Adding up these two processes corresponds to a total memory of $L-1+N_m$ symbols, therefore there are $|\Omega|^{L-1+Nm}$ number of states for the overall Markov model.

When the Gauss-Markov noise is Gaussian distributed and has zero-mean, it is shown in [5] that only the covariance matrix of the noise is mandatory in order to solve the overall Markov model. The Gaussian distributed Gauss-Markov noise with zero mean is a quite good assumption for the communication system [5]. If the noise vector $\boldsymbol{\epsilon}_{n_1}^{n_2}$ is stacked in the same way as for the received samples in (5.22), the covariance matrix of the noise is shown as:

$$\boldsymbol{\Sigma} = E\left[\boldsymbol{\epsilon}_{n-N_m}^{n}\left(\boldsymbol{\epsilon}_{n-N_m}^{n}\right)^H\right]_n \tag{B.1}$$

Further define $\breve{\Sigma}$ in below:

$$\breve{\boldsymbol{\Sigma}} = E\left[\boldsymbol{\epsilon}_{n-N_m}^{n-1}\left(\boldsymbol{\epsilon}_{n-N_m}^{n-1}\right)^H\right]_n \tag{B.2}$$

It is proved in [5] that the second order statistics used for whitening the noise is found by:

$$\mathbf{W} = \mathbf{\Sigma}^{-1} - \begin{bmatrix} \breve{\mathbf{\Sigma}}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \tag{B.3}$$

Recall the definition of $\widetilde{\mathbf{r}}_n$ in (5.32) and skip the notation for the transition $(p,\ q)$ here. Also stack $\widetilde{\mathbf{r}}_{n_1}^{n_2}$ as in (5.22), the colored Gaussian noise $\widetilde{\boldsymbol{\epsilon}}_{n-1-N_m}^{n-1}$ associated with a state transition is shown as [5]:

$$\widetilde{\boldsymbol{\epsilon}}_{n-1-N_m}^{n-1} = \mathbf{r}_{n-1-N_m}^{n-1} - \widetilde{\mathbf{r}}_{n-1-N_m}^{n-1} \tag{B.4}$$

Then the process to convert the colored noise to the white noise can be expressed as [5]:

$$\big(\widetilde{\boldsymbol{\epsilon}}_{n-1-N_m}^{n-1}\big)^H \mathbf{W} \widetilde{\boldsymbol{\epsilon}}_{n-1-N_m}^{n-1} \tag{B.5}$$

After converting the colored noise to the white noise, the merged Markov model is solved by the SbS MAP equalizer given in Section 5.3 with the longer state vector and whitened noise given in (B.5).

Further it is proved in [5] that the second order statistics of the colored noise, i.e. $\mathbf{W}$, can be expressed as:

$$\mathbf{W} = \mathbf{F}\mathbf{F}^H \tag{B.6}$$

With (B.6), it is possible to implement a pre-whitening filter, that is (B.5) can be modified to:

$$\big(\widetilde{\boldsymbol{\epsilon}}_{n-1-N_m}^{n-1}\big)^H \mathbf{W} \widetilde{\boldsymbol{\epsilon}}_{n-1-N_m}^{n-1} = \big(\widetilde{\boldsymbol{\epsilon}}_{n-1-N_m}^{n-1}\big)^H \mathbf{F}\mathbf{F}^H \widetilde{\boldsymbol{\epsilon}}_{n-1-N_m}^{n-1} = \big(\mathbf{F}^H \widetilde{\boldsymbol{\epsilon}}_{n-1-N_m}^{n-1}\big)^H \big(\mathbf{F}^H \widetilde{\boldsymbol{\epsilon}}_{n-1-N_m}^{n-1}\big) \tag{B.7}$$

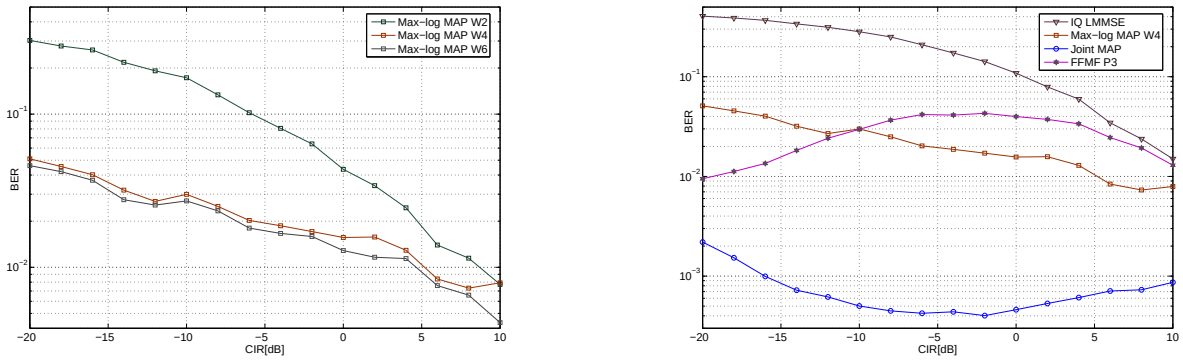
Inserting the definition of $\widetilde{\boldsymbol{\epsilon}}_{n-1-N_m}^{n-1}$ given by (B.4) into (B.7), we get:

$$\begin{aligned} \big(\widetilde{\boldsymbol{\epsilon}}_{n-1-N_m}^{n-1}\big)^H \mathbf{W} \widetilde{\boldsymbol{\epsilon}}_{n-1-N_m}^{n-1} &= \big(\mathbf{F}^H(\mathbf{r}_{n-1-N_m}^{n-1} - \widetilde{\mathbf{r}}_{n-1-N_m}^{n-1})\big)^H \big(\mathbf{F}^H(\mathbf{r}_{n-1-N_m}^{n-1} - \widetilde{\mathbf{r}}_{n-1-N_m}^{n-1})\big) \\ &= \big(\mathbf{F}^H \mathbf{r}_{n-1-N_m}^{n-1} - \mathbf{F}\widetilde{\mathbf{r}}_{n-1-N_m}^{n-1}\big)^H \big(\mathbf{F}^H \mathbf{r}_{n-1-N_m}^{n-1} - \mathbf{F}^H \widetilde{\mathbf{r}}_{n-1-N_m}^{n-1}\big) \end{aligned} \tag{B.8}$$

(B.8) shows $\mathbf{F}^H$ can be served as a pre-whitening filter [5]. From the above derivation, it is seen that using (B.5) or (B.8) to convert the noise to the white noise are identical. The complexity of the whitening receiver is $\mathcal{O}(|\Omega|^{L+N_m})$, which is exponential in both L and N_m.

B.2 Simulation

Figure B.1 (a) shows the BER performance of the whitening receiver for different N_m values. The simulation is performed in the TU3 channel with SNR = 20 dB and one CCI (GMSK) interferer. Increasing N_m from 2 to 4 gives a big BER gain. And increasing it further to 6 provides only a small gain. So choosing $N_m = 4$ is a better tradeoff between the complexity and the performance. For the same test case, other receivers' BER performance is plotted in Figure B.1 (b). Simulation results show that the whitening filer gives better BER than the IQ LMMSE receiver for all the tested CIR values. The whitening approach also shows lower BER compared to the FFMF receiver with the prediction function at higher CIR range and worse BER than the FFMF receiver at low CIR values. This is due to the FFMF receiver is a MUD IC method and has the advantage at low CIR for the single co-channel interferer. The complexity of the whitening filter is higher than that of the IQ LMMSE and the FFMF receivers. The joint MAP equalizer achieves the best BER performance but has the highest complexity.



(a) *different Whitening Filter length*

(b) *comparison with other receivers*

Figure B.1: Simulations of Whitening Filter Approach, SNR = 20 dB, GMSK modulation, 1 CCI, TU3, IQ LMMSE filter length = 11

As indicated in the beginning of this appendix, the Gauss-Markov model is a good approximation to co-channel interferers. So the whitening solution derived with this model works well with co-channel interferers, as seen in Figure B.1 (b). It is known that the energy from the ACI can be suppressed by the receiver filtering. But when the ACI is strong, its influence on BER becomes significant accordingly, especially for the whitening solution. In the next test case two CCI are added, with the relative averaged power from the second one being 10 dB lower than that of the first one. Moreover, one ACI is also included with the averaged power being 3 dB higher than that of the first CCI. ACIU stands for the upper adjacent channel interferer.

Simulation results are collected in Figure B.2. For $N_{sps} = 1$, the FFMF receiver is always better than the whitening filter. However, as discussed earlier, the receiver normally utilizes two times over sampling to overcome the non Nyquist pulse. For $N_{sps} = 2$, the FFMF solution provides better BER than the whitening filter with $N_m = 4$ at $CIR < 0$ dB, and produces slightly worse BER compared to the whitening approach at $CIR > 0$ dB. The reason why the comparison between the FFMF receiver and the whitening receiver looks different in Figure B.1 (b) and Figure B.2 is the presence of ACI. As just stated, the current whitening model only well fits the co-channel interferers whereas MUD solution is suitable for both CCI and ACI scenarios.

Finally, the performance of the FFMF receiver with the prediction function and the whitening

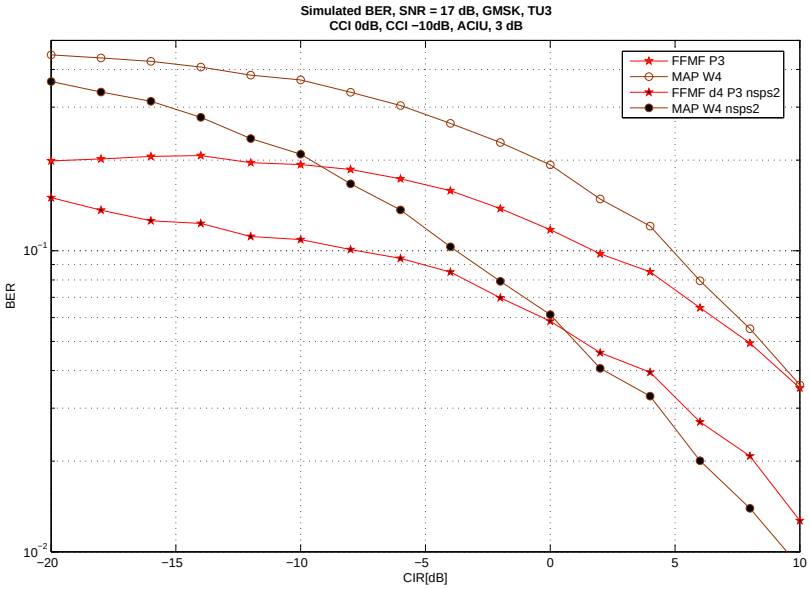


Figure B.2: Simulation of 3 interferers, SNR = 17 dB

solution with $N_m = 4$ is compared in Figure B.3. The power distribution among interferers and white noise is specified in GERAN DTS3 [9]. The FFMF receiver gives some 2 dB gain compared to the whitening receiver at 10% BER. The cross of these two BER curves occurs around $CIR = 0$ dB like in the previous test case. The performance of the FFMF solution at high CIR range might be improved by better tuning the control parameter T. Besides, whitening filter can be combined with suboptimal solutions, e.g. the FFMF receiver, as shown in Figure 7.7.

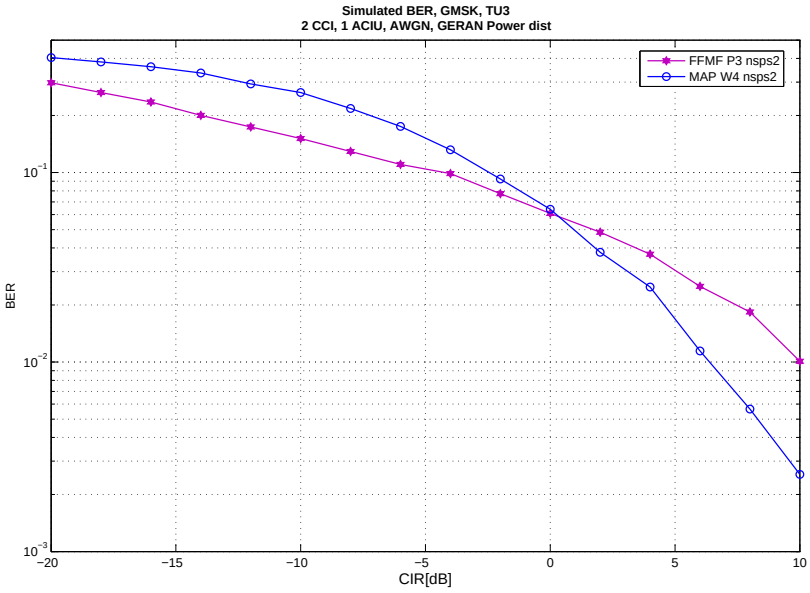


Figure B.3: Simulation of 3 interferers, power distribution is defined in GERAN DTS3

Appendix C

GSM Channel Profile

This appendix illustrates some typical cases of GSM channel profiles [42].

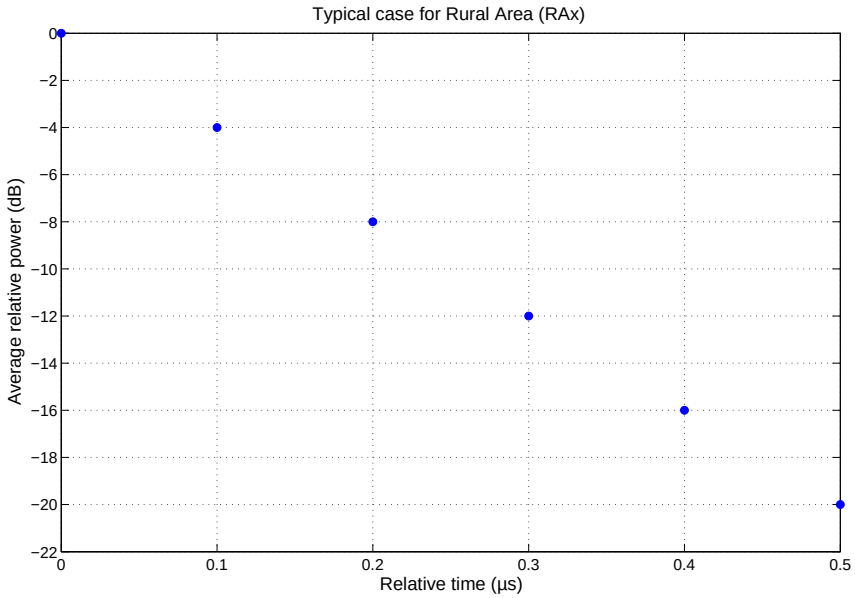


Figure C.1: Typical case for RAx (6 tap setting)

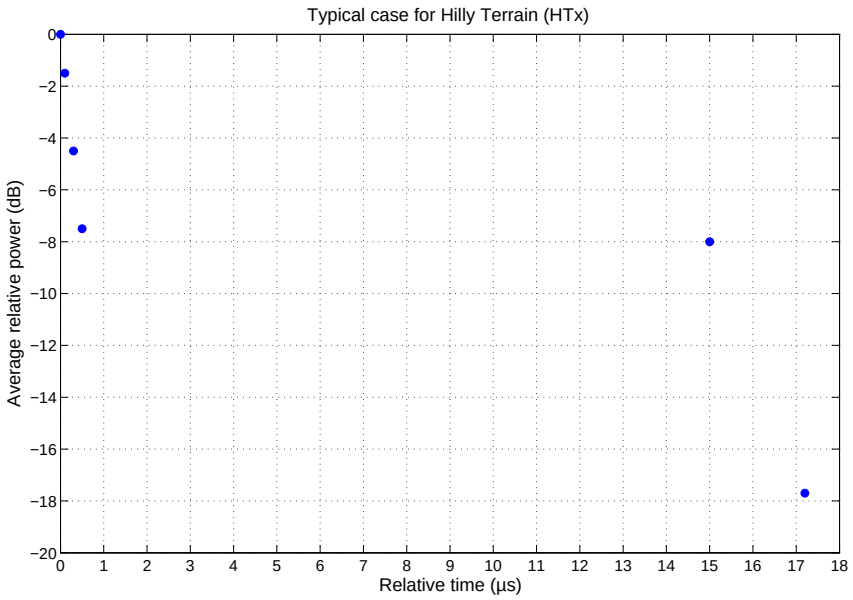


Figure C.2: Typical case for HTx (6 tap setting)

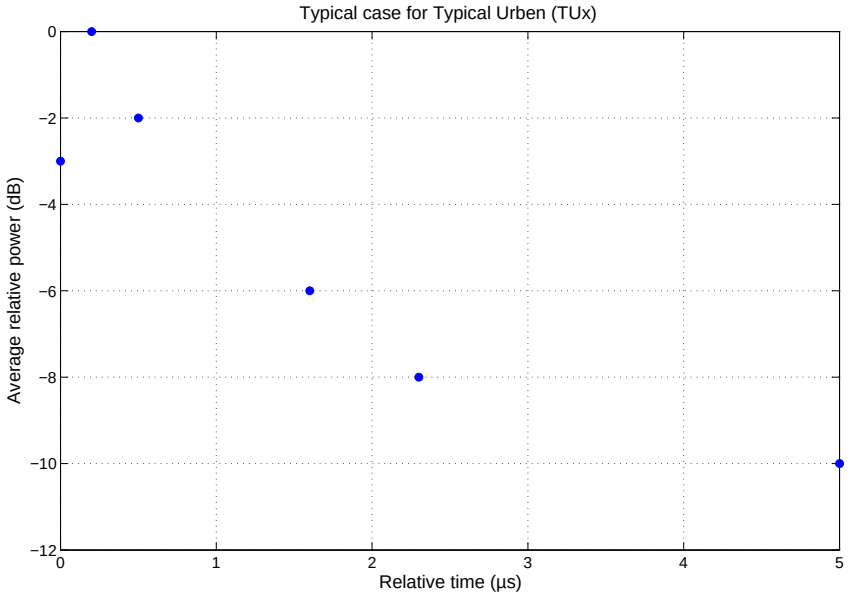


Figure C.3: Typical case for TUx (6 tap setting)

Appendix D

CD Rom

This CD accompanies the thesis "Exact and Approximate Multi-user detection in GSM".

It includes:

- Code used for this project.
- References for the thesis, if possible.
- Simulation data presented in the thesis.
- Thesis including figures and Matlab plots in both "*.fig" and "*.eps" formats. Matlab plots are generated by Matlab 7.