AN ABSTRACT OF THE THESIS OF

Zhou Fang for the degree of Master of Science in Electrical and Computer Engineering presented on October 31, 2018.

Title: New Detection Methods for Transmission at Faster-Than-Nyquist Rates Using Quasi-Orthogonal Sequences

Abstract approved: _____

Huaping Liu

Direct sequence spread spectrum (DSSS) was initially used for anti-jamming in military applications and was later developed in other commercial applications. Walsh codes, which are a set of mutually orthogonal codes, are one type of the spreading codes used in DSSS systems. For any spreading codes of length N, these exist exactly N mutually orthogonal spreading codes; applying more than N codes to transmit data simultaneously will introduce interference among the transmitted data. For interference-free detection, the symbol rate is thus bounded by the Nyquist transmission rate with DSSS codes, assuming the traditional correlation receivers. The goal of this thesis is to explore new detection methods for signaling with faster-than-Nyquist transmission rates. First, new sets of quasi-orthogonal codes are designed to allow simultaneous transmission of data with more than N codes of length N each. Thus there exists mutual interference among the transmitted data. Then, a multi-layer perceptron (MLP) detection scheme with three layers is proposed for detecting the signal transmitted at faster-than-Nyquist rates using the quasi-orthogonal sequences. Finally, a long-short-term-memory (LSTM) scheme is proposed for detection in the proposed signaling scheme. The performances of these schemes are analyzed and simulated, and compared with the traditional correlation detection schemes.

New Detection Methods for Transmission at Faster-Than-Nyquist Rates Using Quasi-Orthogonal Sequences

by

Zhou Fang

A THESIS

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Master of Science

Presented October 31, 2018
Commencement June 2019

Master of Science thesis of Zhou Fang presented on October 31, 2018.

APPROVED:

Major Professor, representing Electrical and Computer Engineering

---

Head of the School of Electrical Engineering and Computer Science

---

Dean of the Graduate School

I understand that my thesis will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my thesis to any reader upon request.

Zhou Fang, Author

ACKNOWLEDGEMENTS

First, I would like to thank my entire family, my husband Wei, my father Shuming, and my mother Liyu for their love, patience and support, which have been the powerful source for me to be successful in study and life.

I would also like to thank my major advisor Dr. Huaping Liu. He provided a great opportunity for me to conduct valuable research work and guided me through my graduate study process. His valuable support and guidance made it possible for me to complete this thesis work.

Thanks to my Computer Science minor advisor Dr. Liang Huang, for his help and I appreciate the excellent lectures he gave in the classes that were helpful to my thesis work.

Thanks to Dr. Bechir Hamdaoui, Dr. Jinsub Kim and Graduate Council Representative Dr. Bo Zhao, for serving on my graduate committee and for reviewing the manuscript.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

LIST OF FIGURES

LIST OF FIGURES (Continued)

LIST OF TABLES

Chapter 1: INTRODUCTION

1.1 Background and Motivation

Digital communication can be traced back to 1837. Morse developed electric telegraph that translated English letters into a sequence of dots and dashes. This was the precursor of the variable-length source coding method [27]. In 1924, Nyquist determined the the maximum data rate without inter-symbol interference (ISI) that can be transmitted over a transmission channel. That is now called the Nyquist Criterion (for ISI-free digital signaling). In 1948, Shannon established the relationship between channel capacity in Gaussian noise and gave birth to a new field called information theory. Shannon's results were precursors to the work of Hamming, who provided the work on error detecting and error correction theorem to reduce the influence of channel noise. There are more powerful works in digital communications based on the early works of Nyquist, Shannon and Hamming, etc [35]. The general digital communication structures are consisted of these advanced works in Figure 1.1.

The information source will be encoded by source coding in source encoder. Source

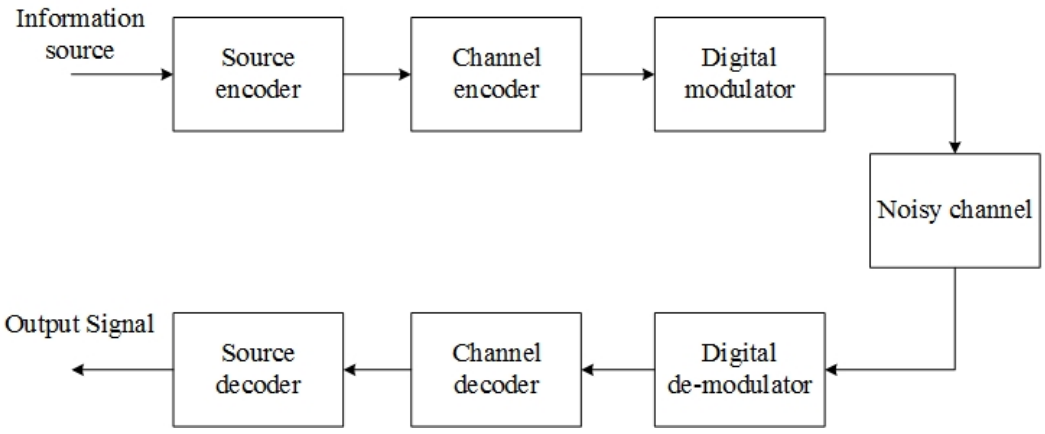

Figure 1.1: General Digital Communication Structure.

coding is used to remove redundant information so that bandwidth will be minimized for

efficient transmission. Famous source coding techniques include Huffman codes, Lemple-Ziv codes, Fano codes, etc. In the channel encoder, information signal is encoded by adding some extra bits to protect information from noise and interference. Common channel codes include block codes, convolutional codes, turbo codes, LDPC codes, etc. In radio communications, the encoded information is carrier-modulated and then transmitted to the receiver via an antenna(s).

Compared to analog communications, digital communications has many advantages, such as better privacy and security to protect information by data compression and error correction coding and high noise tolerance. In analog communications, even a small noise signal can have a large impact on received signal. Digital communications could be cheaper than analog communications due to the usage of advanced digital very-large-scale integration (VLSI) technology.

Spread spectrum, when used for digital communications, uses direct sequence, frequency hopping or a hybrid of both model for multiple access. In such a system, the information signal is spread with a spreading code, and then modulated onto a carrier. A noise-like signal is transmitted. The same spreading code is used in the receiver to recover the original information via correlation and integration. Resistance to interference and security are achieved by exploiting different spreading codes. Interference introduced by other spreading codes is spread over a much wider bandwidth than the original information bandwidth and in the despreading process the desired signal is effectively affected by only the interference in its information bandwidth. Since the transmitted spread information can be detected only by the receiver that has same spreading code, information is protected from interception.

The motivation of the work in this thesis is shown in Figure 1.2. Walsh codes are widely used in spread spectrum communications. With Walsh codes, there is no inter-symbol interference, but the symbol rate is still bounded by the Nyquist transmission rate. To improve channel use efficiency beyond the Nyquist rate via a scheme that uses spreading codes, new spreading codes are needed to allow faster-than-Nyquist (FTN) rate. In this case, there will be mutual interference among the spreading codes. Correlation detection can be used to recover a received signal that is affected by the non-zero cross-correlation among the spreading codes simultaneously transmitted but is not necessarily the best detection scheme, and depending on the symbol rate of transmission, the corresponding BER performance may be unacceptable. The goal of this thesis is to

explore new detection methods that achieve a better performance than the traditional correlation detector.

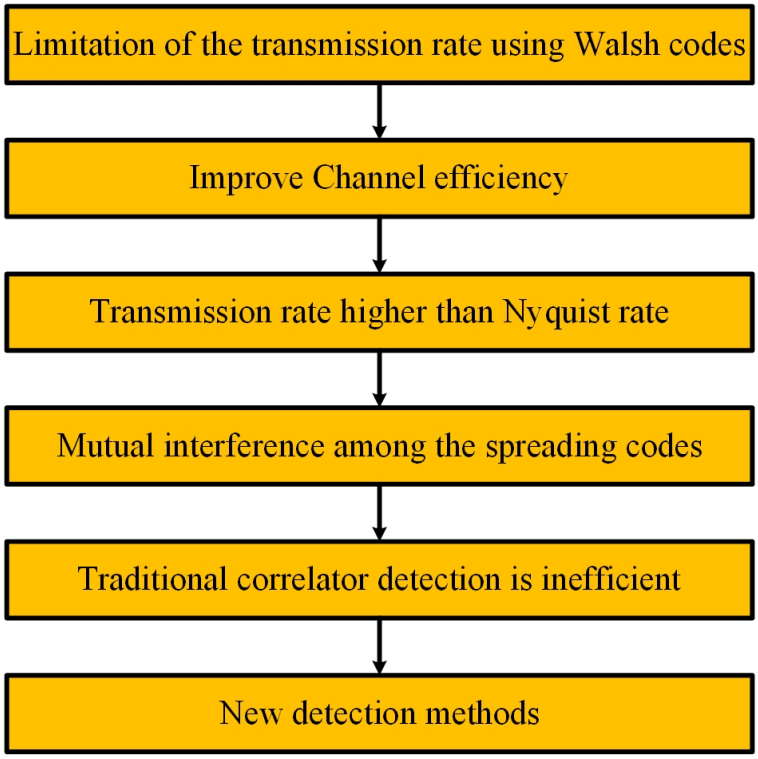

Figure 1.2: Motivation.

1.2 Thesis Outline

In this thesis, new spreading codes, called quasi-orthogonal sequences (QOS), that can provide a large amount of spreading codes of fixed length are proposed. Because the cross-correlation among the QOS is non-zero, a correlation detection method may not perform well. New detection methods are then proposed to detect the transmitted signals that use QOS to transmit at the FTN rate. The first detection scheme is called multi-layer perceptron (MLP) and the second one is called long-short-term memory (LSTM). These detection methods are able to detect the desired user's signal from interference and noise better than the traditional correlation detector.

Chapter 2 gives a review of spread spectrum. The processing gain in spread spectrum is introduced. Different spread spectrum transmissions such as DSSS, frequency hopping spread spectrum (FHSS) and Hybrid DS/FH are described. The classical spreading codes used in spread spectrum, such as Gold codes, small Kasami set, large Kasami set and Walsh codes are also introduced.

In Chapter 3, DSSS transmitter and receiver structures are described. Based on the analysis spread spectrum transmission methods and existing spreading codes, this thesis adopts QOS as the spreading codes to provide a large number of spreading codes. The definition of BER performance for faster-than-Nyquist signals is presented and compared with BER performance for signaling at Nyquist. A three-layer MLP scheme and an LSTM detection method are developed to detect the QOS signals in the presence of interference and noise.

Chapter 4 presents the simulation results for the three-layer MLP and LSTM detection schemes. QOS of lengths 16, 32, 64 and 128 are adopted. Theoretical and simulation results for correlation detection are compared at different signal-to-noise ratio (SNR) values.

Chapter 5 summarizes the spreading codes used, new detection methods provided and simulation results from Chapter 3 and 4. Conclusions are made in terms of BER performance and complexity of the new detection methods.

Chapter 2: BACKGROUND REVIEW

2.1 Introduction

Spread spectrum was traditionally used in military anti-jamming communications because this technique affords protection against high interference with limited signal power [26]. Other applications include satellite-positioning systems (GPS), 3G mobile telecommunications, wireless LAN, Bluetooth, ZigBee and Wi-Fi [34][28][9][19]. Spread spectrum uses wide-band and noise-like signals to transmit information that only can be detected by using the spreading codes used in the transmitter. Spread spectrum signals have a wider frequency band than the information that they are carrying to make information more noise-like.

The power spectral density of signal and interference in transmission are shown in Figure 2.1. User information signal is a narrow band signal before spreading. The use of spreading codes makes user information signals appear wide-band and noise-like. At a receiver, same spread code is used to recover a desired user's information signal from the received signal, while the interference spreads over the whole wideband. In such a system, the desired user's information signal could still be recovered from interference with a high power.

This chapter begins with introduction of spread spectrum. Three commonly used spread spectrum techniques are described in Section 2.2. In spread spectrum, a set of unique spreading codes are used to spread users' signals, which is discussed in Section 2.3. The primary interest of this thesis, new spreading codes and new detection methods that achieve a better performance, will be introduced in Chapter 3.

2.2 Spread Spectrum

There are several types of spread spectrum techniques. Different spreading codes are used to transmit different signals in the time domain. Processing gain is to evaluate the interference-immunity performance. It is defined as the differential between output and

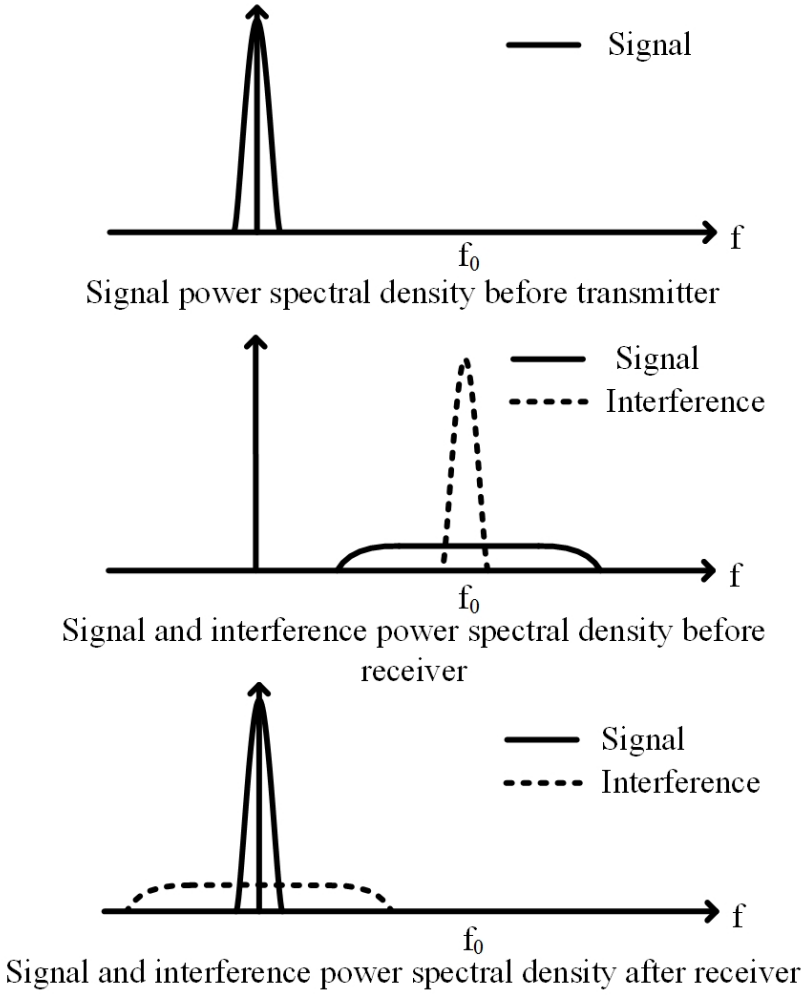

Figure 2.1: Signal Power Density of Signal and Interference in Transmission.

input SNRs [7]. It can be also expressed by the length of the spreading code N as [8].

$$\mathrm{PG} = 10 \log_{10} \frac{\mathrm{SNR}_{\mathrm{out}}}{\mathrm{SNR}_{\mathrm{in}}} = 10 \log_{10} N. \tag{2.1}$$

The processing gain is only related to the length of the spreading code, and will not be affected by the cross-correlation between spreading codes.

In this section, direct sequence spread spectrum, frequency hopping spread spectrum, and hybrid DS/FH spread spectrum (HSS) are briefly described.

2.2.1 Direct Sequence Spread Spectrum

DSSS is the best known and most widely used spread spectrum technology. It was first used in military in 1940s, and now are widely used in ZigBee and Wi-Fi systems [18][7]. DSSS does not require a high speed, fast settling frequency synthesizer, which makes it easier to implement for some cases.

DSSS transmission structure is shown in Figure 2.2. The noise-like signal is generated from the signal bit and spreading code. The spreading code used in DSSS transmission carries a higher frequency than the signal bit. This helps increase the signal's resistance to interference. If any bits in the spread signal are distorted by noise or interference from other information signals during transmission, the original signal bit may still be recovered via the despreading process. The time interval of the spreading code is the bit duration T_b. Each fragment of the spreading code is called a chip with a duration T_c, which is typically much shorter than the bit duration. Let the length of spreading code be N. The relationship among length of the spreading code, code duration and chip duration is $N = T_b/T_c$. The base-band spread signal is then modulated onto a carrier via a chosen modulation scheme such as binary phase-shift keying (BPSK), quadrature phase-shift keying (QPSK), or quadrature amplitude modulation (QAM), etc. This way, the bandwidth of the DSSS signal is much wider than the bandwidth if the information signal is simply modulated onto carrier. To decode and reconstruct the signal bit, the received signal is carrier-demodulated to reconstitute high speed baseband data stream. This data stream is then multiplied with the spreading code to recover the signal bit. Only the signal bit that is modulated with this spreading code will be reconstructed. Other signal bits and noise that are not modulated with same spreading code will be

interference and noise, and are ignored in a correlation type of receiver.

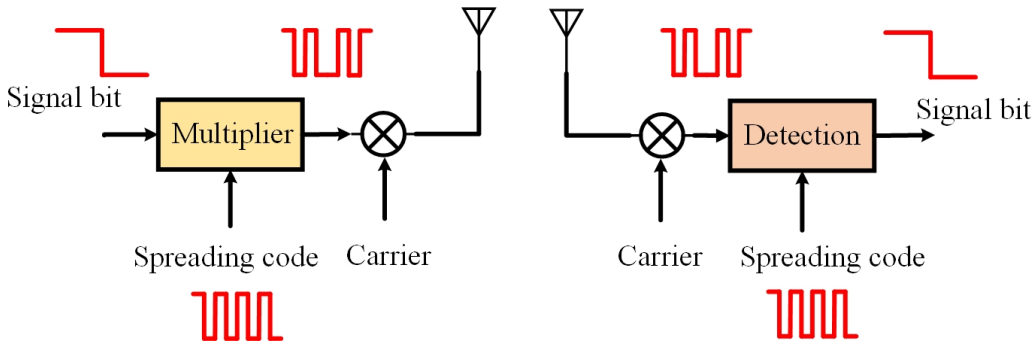

Figure 2.2: DSSS Transmission Dtructure.

2.2.2 Frequency Hopping Spread Spectrum

FHSS is a different approach to DSSS. Rather than modulating the spread signal onto a fixed carrier, the signal bit is unchanged but directly modulate onto a carrier of varying frequencies. This is a method of transmitting the signal bit by rapidly switching the carrier frequency across numerous channels defined by a hopping sequence, which is already known to the transmitter and receiver. It is used in military and Bluetooth [19]. The narrow-band interference resistance comparison between DSSS and FHSS can be found in [23]. There are two types of FHSS according to the hopping rate [21][37]. On the one hand, a system is considered to be fast FHSS (FFHSS) if the hopping rate is greater than the symbol rate, in which case a symbol can be transmitted by several frequencies. On the other hand, a system is considered to be slow FHSS (SFHSS), in which case several symbols may be transmitted at the same frequency. The narrow-band interference of moderate level can be mitigated by DSSS. DSSS may fail when interference is strong. FHSS can work in a strong interference environment, even though the interference is not completely rejected.

The FHSS structure is shown in Figure 2.3. Frequency hopping modulation is similar to frequency shift keying (FSK) except that the set of frequency choices is greatly expanded. Frequency hopping systems often have a large number of frequencies available while FSK often use two frequency choices. A frequency hopping system consists of a

code generator and a frequency synthesizer that can output a frequency responding to the code from the code generator. The signal bit modulates with a frequency hopping carrier to form a FHSS signal. The receiver has the same hopping sequence used in the transmitter, so the frequency synthesizer can generated the same frequency hopping pattern as applied at the transmitter. The signal bit can also be recovered by correlation.

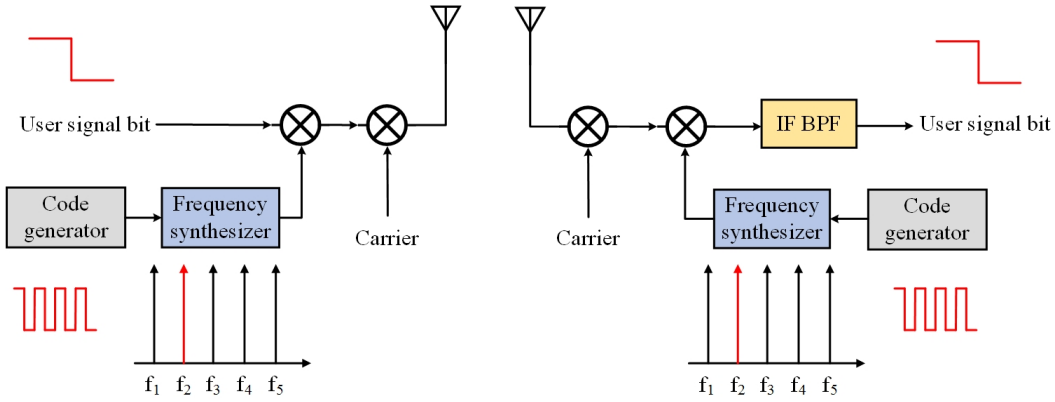

Figure 2.3: FHSS Transmission Structure.

2.2.3 Hybrid DS/FH Spread Spectrum

A hybrid DS/FH spread spectrum may be classified into hybrid DS/SFH spread spectrum and hybrid DS/FFH spread spectrum depending on frequency hopping rate [36][25]. Hybrid DS/FH spread spectrum has some desirable characteristics that DSSS or FHSS does not have. For example, hybrid DS/FH spread spectrum provides a higher level of security than DSSS and FHSS, because it combines DSSS and FHSS. Although the complexity of hybrid DS/FH spread spectrum appears to be higher than DSSS or FHSS, its implementation is not much more complex than DSSS or FHSS, since a shorter spreading code used in hybrid DS/FH spread spectrum can achieve the same or even better performance than DSSS or FHSS with a longer spreading code.

Hybrid DS/FH spread spectrum structure is shown in Figure 2.4. In such a system, a bit is identified not only by DSSS spreading code, but also frequency hopping pattern in FHSS. A bit is first modulated a spreading sequence, and then modulated by a frequency hopping carrier. At the receiver, the spread sequence that carries a desired

bit is generated by de-modulating with its frequency hopping carrier. The bit is finally recovered by de-spreading the received spread sequence.

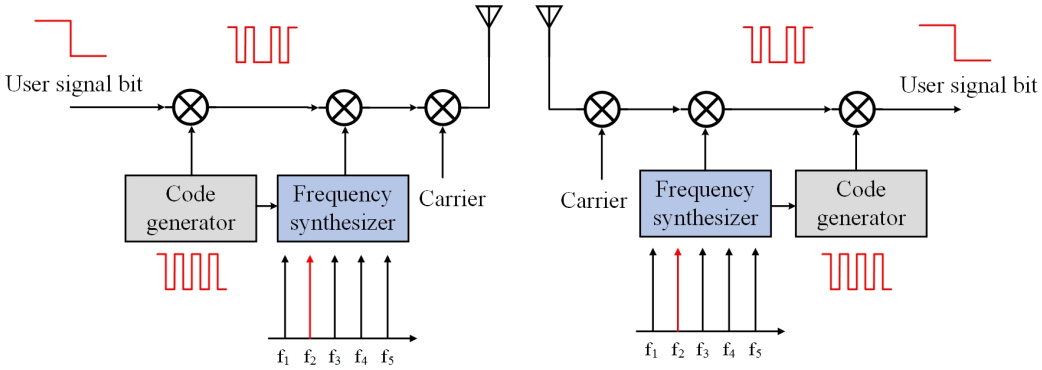

Figure 2.4: Hybrid DS/FH Transmission Structure.

Four types of hybrid DS/FH spread spectrum are shown in Figure 2.5. Hybrid DS/FH spread spectrum can be realized with a combination of different frequency hopping and the number of signal bits per spreading code. Hybrid DS/SFH spread spectrum means more than one signal bit are transmitted on one frequency hop. One or more spreading codes can be modulated with one signal bit. Hybrid DS/FFH spread spectrum means more than one frequency hop are used as carrier for one signal bit. At the same time, one or more spreading codes can be modulated with one signal bit. If two signal bits use the same spreading code to transmit simultaneously, they are not interfered with each other when they use different frequency hops. If two signal bits are transmitted by the same frequency hop carrier, then they can be recovered without interference when they use different spreading codes.

2.3 Spreading Codes

There are two types of spreading codes according to the cross-correlation between non-orthogonal codes and orthogonal codes. Gold codes, small Kasami set and large Kasami set are non-orthogonal codes. Walsh codes are orthogonal codes.

Before introducing spreading codes used in spread spectrum, we describe the M-sequence. M-sequence is a type of pseudo-random binary sequence that is generated by maximal linear feedback shift registers (LFSR). A polynomial function can be associated

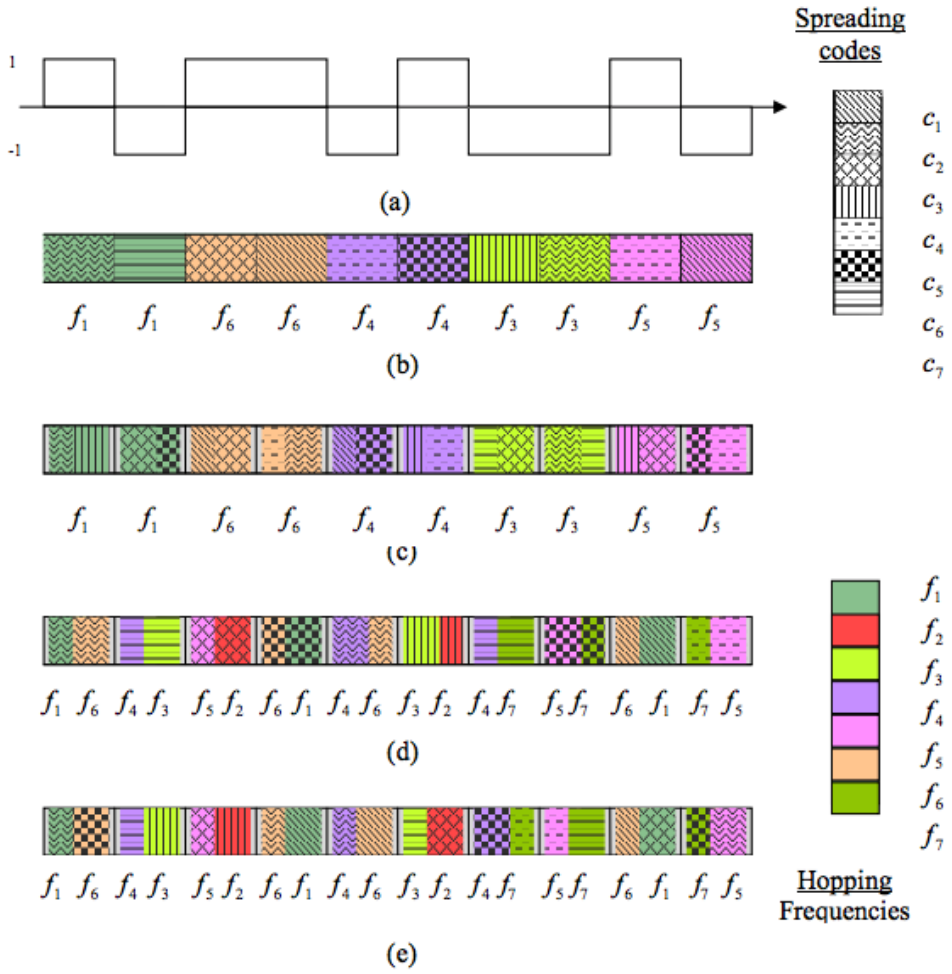

Figure 2.5: Concept of Hybrid DS/SFH and Hybrid DS/FFH Spread Spectrum: (a) User signal sequence (b) Hybrid DS/SFH spread spectrum with 2 bits/hop and 1 spreading code/bit (c) Hybrid DS/SFH spread spectrum with 2 bits/hop and 2 spreading codes/bit (d) Hybrid DS/FFH spread spectrum with 2 hops/bit and 1 spreading code/bit (e) Hybrid DS/FFH spread spectrum with 2 hops/bit and 2 spreading codes/bit [30].

with a LFSR. Its degree is the number of shift register used, and has coefficients that are either 0 or 1. If a coefficient is 0, then there is no feedback; otherwise, there is a *mod* 2 operation for all feedback as input. The output of the LFSR is periodic with a maximum period $N = 2^m - 1$, where m is the degree of the polynomial function. Hence, the length of an M-sequence is N. The LFSR structure of the polynomial function $1+x+x^3$ is shown in Figure 2.6. The state of this polynomial function is shown in Table 2.1. From Table 2.1, the state 111, 011, 101, 010, $\cdots$, 111 repeats every 7 cycles. Thus, the maximum period of the polynomial function $1 + x + x^3$ is 7 and the M-sequence generated from $1+x+x^3$ is 1110100. The M-sequence is critical since Gold codes, small Kasami set and large Kasami set are all obtained based on M-sequence.

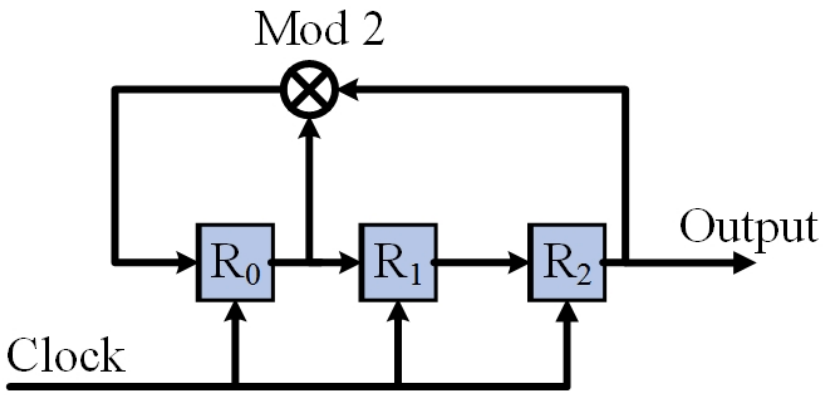

Figure 2.6: Linear Feedback Shift Register Structure for $1 + x + x^3$.

Table 2.1: Register state for $1 + x + x^3$

Clock	**Feedback**	**Register 0**	**Register 1**	**Register 2**
0	0	1	1	1
1	1	0	1	1
2	0	1	0	1
3	0	0	1	0
4	1	0	0	1
5	1	1	0	0
6	1	1	1	0
7	0	1	1	1

In this section, non-orthogonal codes are introduced. The characteristics and generation of Gold codes, small Kasami set and large Kasami set are introduced in Sections 2.3.1 and 2.3.2. Binary orthogonal codes, i.e., Walsh codes, are described in Section 3.2.3.

2.3.1 Gold Codes

Gold codes were invented by Robert Gold in 1967 [10]. One M-sequence and another LFSR M-sequence are combined to generate the set of Gold codes. Gold codes have uniform and bounded cross-correlation that makes it useful in spread spectrum [24]. The cross-correlation function are three valued: -1, $-t(m)$ or $t(m)-2$ for all length $N = 2^m - 1$, where $t(m) = 1 + 2^{\lfloor (m+2)/2 \rfloor}$ [6], and $\lfloor\rfloor$ is the floor function.

Gold codes can be constructed from a preferred pair of M-sequences, m_1 and m_2. Consider an M-sequence m_1. Another M-sequence, m_2, can be constructed by sampling m_1 by every q symbols. The definition of preferred pair includes: (a) m is not divisible by 4; (b) $m_1 = m_2[q]$, where q is odd and either $q = 2^k + 1$ or $q = 2^{2k} - 2^k + 1$; (c) The greatest common divisor $gcd(m, k)$ is 1 when m is odd and $gcd(m, k)$ equals 2 for $m = 2(mod\ 4)$. Finding a preferred pair is a critical step for Gold codes. Table 2.2 shows some preferred pairs for constructing Gold codes. The set of Gold codes is then defined as:

$$S_{gold} = \{m_1, m_2, m_1 \oplus m_2, m_1 \oplus Tm_2, m_1 \oplus T^2 m_2, \cdots, m_1 \oplus T^{N-1} m_2\}, \tag{2.2}$$

where T represents a left cyclic shift and $\oplus$ represents mod 2 operation. For length N, there are a total of $N + 2$ Gold codes.

Table 2.2: Preferred Pair for Gold Codes for $m = 3, 5, 6, 7$

m	**Preferred Polynomial 1**	**Preferred Polynomial 2**
3	$1 + x + x^3$	$1 + x^2 + x^3$
5	$1 + x^2 + x^5$	$1 + x^2 + x^3 + x^4 + x^5$
6	$1 + x + x^6$	$1 + x + x^2 + x^5 + x^6$
7	$1 + x^3 + x^7$	$1 + x + x^2 + x^3 + x^7$
9	$1 + x^4 + x^9$	$1 + x^3 + x^4 + x^6 + x^9$

Gold codes can generate $N + 2$ codes by mod 2 operation from two M-sequences of

same length.

2.3.2 Small Kasami Set And Large Kasami Set

Kasami set is proposed by Tadao Kasami in 1966 [16]. There are two types of Kasami set: small Kasami set and large Kasami set. The cross-correlation of the small Kasami set is same as Gold codes. It is a three-valued function: -1, $-t(m)$ or $t(m)-2$, for all lengths $N = 2^m - 1$, where $t(m) = 1 + 2^{\lfloor (m+2)/2 \rfloor}$. For the large Kasami set, the cross-correlations are limited to five values: -1, $-1 \pm 2^{m/2}$ or $\pm 2^{m/2}$ [6].

For the small Kasami set, the generation process is similar to gold codes, but m must be an even number [3]. First, a decimation sequence m_2 is obtained by decimating an M-sequence m_1 by $2^{m/2}+1$. m_2 is also a periodic M-sequence with a period $2^{m/2}-1$. Then the small Kasami set can be generated by performing mod 2 operation between m_1 and cyclically left shift sequence on m_2 as expressed in Eq. (2.3),

$$S_{small} = \{m_1, m_1 \oplus m_2, m_1 \oplus Tm_2, m_1 \oplus T^2 m_2, ..., m_1 \oplus T^{2^{m/2}-2} m_2\} \tag{2.3}$$

where T represents a left cyclic shift and $\oplus$ represents mod 2 operation. A small Kasami code set of length N has $2^{m/2}$ codes.

The large Kasami set contains both the small Kasami set and gold codes of length $N = 2^m - 1$ as its subsets where m is even. M-sequences m_2 and m_3 are obtained from decimating M-sequence m_1 by $2^{m/2}+1$ and $2^{(m+2)/2}+1$. Then the large Kasami set can be generated by performing mod 2 operation between m_1 and the cyclically left-shift sequence on m_2 and m_3 as shown in Eq. (2.4),

$$S_{large} = \{m_1, m_1 \oplus m_2, m_1 \oplus m_3, m_1 \oplus Tm_2, ..., m_1 \oplus T^{2^{m/2}-2} m_2 \oplus T^{2^{m/2}-2} m_3\} \tag{2.4}$$

where T represents a left cyclic shift and $\oplus$ represents mod 2 operation. The number of large Kasami sequences is $N = 2^{3m/2}$, if $m = 0(mod\ 4)$; If $m = 2(mod\ 4)$, the number of large Kasami sequences is $2^{3m/2} + 2^{m/2}$ [6][39].

2.3.3 Walsh Codes

Walsh codes are the most well-known and widely used spreading codes. It is a set of codes that perfectly distinguish different signal bits in CDMA [13]. Walsh codes are orthogonal to each other when they are synchronized in time. The orthogonal characteristics are expressed below

$$\sum_{i=1,j=1,j\neq i}^{2^m} w_i w_j = 0, \tag{2.5}$$

where w_i and w_j are two codes from the Walsh matrix and 2^m is code length.

Walsh codes come from rows of the Walsh matrix that was proposed in 1923 [15]. For a set of Walsh codes of length 2^m, there are 2^m Walsh codes. The dimension of this Walsh matrix is $2^m \times 2^m$. The Walsh matrix is defined recursively as

$$\begin{aligned} W_1 &= [1] \\ W_2 &= \begin{bmatrix} W_1 & W_1 \\ W_1 & -W_1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \\ &\cdots\cdots \\ W_n &= \begin{bmatrix} W_{n-1} & W_{n-1} \\ W_{n-1} & -W_{n-1} \end{bmatrix} \end{aligned} \tag{2.6}$$

However, the number of Walsh codes of length N is $N = 2^m$; thus, only 2^m codes can be used as the spreading codes of length 2^m.

2.4 Conclusion

This chapter briefly introduced several spread spectrum transmission methods including DSSS, FHSS and Hybrid DS/FH. These spread spectrum transmission methods have advantages and disadvantages, depending on the channel environment and implementation complexity. Non-orthogonal spreading codes are also introduced, including Gold codes, small Kasami set and large Kasami set, followed by orthogonal spreading codes, the Walsh codes. Although these spreading codes have good auto-correlation (except Walsh codes) and zero or bounded cross-correlations, but the number of spreading codes is limited. That is impossible to transmit information bits that have symbol rate higher

than Nyquist rate. New spreading codes need to be provided in this situation. Even we have large amount of spreading codes, the cross-correlation between spreading codes has negative influence to detection. Hence, new detection methods also need to be provided.

Chapter 3: NEW SPREADING CODES AND DETECTION METHODS

3.1 Introduction

As discussed in the previous chapter, for orthogonal codes, Walsh codes only have 2^m codes of length 2^m. For existing non-orthogonal codes (Gold codes, small Kasami set and large Kasami set), the number of codes is also limited. The limitation of the number of codes restricts the channel efficiency. Also, traditionally correlation detection is used to recover signal bit but when non-orthogonal codes are used, this traditional detection method will be affected by mutual interference among the spreading codes. In order to improve the channel efficiency, information can be transmitted at a symbol rate that is faster-than-Nyquist rate by a new set of spreading codes. In such case, the number spreading codes for a fixed length is larger than the number of Walsh codes. Thus, such spreading codes are no longer mutually orthogonal. The more non-orthogonal codes are used to transmit the more severe interference will be. New detection methods are desirable to achieve an acceptable performance when transmission rate for transmission at faster-than-Nyquist rates.

A new set of spreading codes named quasi-orthogonal sequences (QOS) will be discussed in Section 3.2, which provides a large amount of codes of a fixed length with acceptable cross-correlation. Moreover, the BER upper bound for faster-than-Nyquist signaling with correlation detection is derived in Section 3.3. Two new detection methods are also proposed: a three-layer MLP detection and the LSTM detection schemes. The new spreading codes and detection methods aim to improve the channel efficiency while keep an acceptable BER performance.

3.2 Quasi-Orthogonal Sequences

A new set of QOS spreading codes must have two properties: low cross-correlation and large number of codes. Yang *et al.* provide a new set of codes named quasi-orthogonal

sequences [38]. QOS sequences do not mean orthogonality, but the cross-correlation between codes is small compared to its auto-correlation. For length 2^m, compared to Walsh codes which provides 2^m codes with zero cross-correlation, QOS provides at least 2×2^m codes with small cross-correlation.

When a single code f_i of length $N = 2^m$ is added into a set of Walsh codes W_m of length N, the cross-correlation between f_i and W_m is defined as $R_{\max}(f_i, W_m)$. Let f_i go through all Walsh codes of length N, we will get a set of $R_{\max}(f_i, W_m), i = 1, \cdots, N$. Let $\theta_{\min}N$ be the minimum achievable cross-correlation between f_i and W_m, given by

$$\theta_{\min}N = \min_{i=1}^{N} R_{\max}(f_i, W_m). \tag{3.1}$$

The cross-correlation introduced by any code f_i to Walsh codes W_m is at least $\theta_{\min}N$. QOS is defined as as [38]:

Definition: Let $W_m = \{w_j, j = 0, 1, 2, \cdots, 2^{m-1}\}$ be the set of Walsh codes of length $N = 2^m$. A family of $F = \{f_i, i = 1, 2, \cdots, M\}$ of M codes of length N is said to be quasi-orthogonal if the followings are satisfied:

a) F contains W_m.

b) $R_{ij} \leq \theta_{\min}N$ for any i and $j (i \neq j)$.

c) For any $f_i \in F$, but $\notin W_m$, $w_j \in W_m$, and any integer L, r, where $L = 2^l, 2 \leq l \leq m$ and $0 \leq r \leq N/L - 1$.

$$\left| \sum_{t=rL}^{rL+L-1} (-1)^{f_i + w_j} \right| \leq \theta_{\min}L. \tag{3.2}$$

Condition *a*) means that F should include W_m; Condition *b*) means the cross-correlation between any two codes from F will be no greater than to $\theta_{\min}N$; Condition *c*) is necessary for practical use that some signal bit may employ a window of length 2^{m-1}, instead of 2^m. If signal bit employs a shorter length, then the cross-correlation is the minimum value according to condition *c*). Every sub-block of length N will have the minimum value of cross-correlation.

For positive $i \leq m$, we define sequences $x_1, x_2, x_3, \cdots, x_m$ of length 2^m as:

$$\begin{aligned} x_1 &= 010101 \cdots 010101 \\ x_2 &= 001100 \cdots 110011 \\ x_3 &= 000111 \cdots 000111 \\ &... \\ x_m &= 000000 \cdots 111111 \end{aligned} \tag{3.3}$$

Then we can construct k special boolean functions $f_1, f_2, f_3, \cdots, f_k$ of length 2^m from $x_1, x_2, x_3, \cdots, x_m$. For example, there are four boolean functions of length 16:

$$\begin{aligned} f_1 &= x_1x_2 \oplus x_2x_3 \oplus x_1x_3 \oplus x_3x_4 \\ f_1 &= x_1x_2 \oplus x_1x_3 \oplus x_1x_4 \oplus x_2x_4 \\ f_1 &= x_1x_2 \oplus x_2x_3 \oplus x_2x_4 \oplus x_3x_4 \\ f_1 &= x_1x_2 \oplus x_1x_4 \oplus x_3x_4 \end{aligned} \tag{3.4}$$

where x_i and x_j are multiplied by the bit $(i, l = 1, \cdots, 4)$ and $\oplus$ is the sign of mod 2.

The set of

$$F = \bigcup_{i=0}^{k} (f_i + W_m) \tag{3.5}$$

is QOS, where f_0 is zero and W_m is Walsh codes. Furthermore, the size of F is $(k+1)2^m$. There are 80 QOS codes of length 16, of which 16 are Walsh codes. The maximum cross-correlation for QOS of length 16 is 0.25. Table 3.1 shows the maximum cross-correlations and the number of codes for QOS of different lengths. As the code length increases, there are more usable QOS codes and with lower cross-correlation. For Walsh code of length 256, there are only 256 number of codes, but there are 5376 QOSs.

3.3 BER With Correlation Detection

As discussed in Section 2.3, if Walsh codes are used for transmission, there is not interference the data symbols carried by the codes, but the symbol rate is bounded by the Nyquist transmission rate. The QOS discussed in Section 3.2 allows us to transmit at faster-than-Nyquist rates, but interference will no longer be zero. Thus the penalty of

Table 3.1: Cross-Correlation and the Number of QOS for Different Lengths

Length(2^m)	**Max Cross-correlation**	**The Number of Codes**
8	0.5	40
16	0.25	80
32	0.25	352
64	0.125	576
128	0.125	2944
256	0.0625	5376

transmission at such rate on the BER performance is of interest.

In this section, BER will be derived for two cases: BPSK signaling at Nyquist rate with correlation detection, the BER upper bound for BPSK signaling at faster-than-Nyquist rate with correlation detection.

3.3.1 BER for Nyquist Signaling

It is well knows that at Nyquist transmission rate, there is no inter-symbol interference and the maximum rate is $2B$, where B is baseband bandwidth. In a BPSK system, the received baseband signal, a vector of $L \times 1$ with each element representing the signal in one chip interval, is expressed as

$$\boldsymbol{r} = \sum_{i=1}^{L} b_i \boldsymbol{c}_i + \boldsymbol{n} \tag{3.6}$$

where $\boldsymbol{c}_i$ is a sequence of length L, $b_i \in \{-1, 1\}$ is a signal bit to be transmitted with energy E_b, and $\boldsymbol{n}$ is the additive white Gaussian noise.

Without loss of generality, let b_1 be the desired bit to detect, which is carried by code $\boldsymbol{c}_1$. Because of the orthogonality among the set of codes, the cross-correlation between between the codes is 0. The baseband signal for the desired bit after despreading is expressed as

$$r_1 = b_1 + n \tag{3.7}$$

where the noise component has a two-sided power spectral density $\sigma^2 = \frac{N_0}{2}$, i.e., $n \sim$

$N(0, \sigma^2)$. This the bit error probability is

$$P_b = P(n > \sqrt{E_b}) = \int_{\sqrt{E_b}}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} dx \tag{3.8}$$

This equation can be simplified using the Q-function as

$$\begin{aligned} P_b &= Q\left(\frac{\sqrt{E_b}}{\sigma}\right) \\ &= Q\left(\sqrt{\frac{2E_b}{N_0}}\right). \end{aligned} \tag{3.9}$$

3.3.2 BER Upper Bound for Signaling at Faster-Than-Nyquist Rates

If QOS are used for transmission at faster-than-Nyquist rate, i.e., ($> 2B$), then interference among the data symbols carried by the sequences are no longer zero. In a BPSK system, the received baseband signal, an $L \times 1$ vector, is expressed as

$$\boldsymbol{r} = \sum_{i=1}^{K} b_i \boldsymbol{c}_i + \boldsymbol{n} \tag{3.10}$$

where $\boldsymbol{c}_i$ is a QOS of length L, $b_i \in \{-1, 1\}$ the i-th bit with energy E_b, K is the total number of QOS ($K > L$) used simultaneously to transmit a set of data bits, and $\boldsymbol{n}$ is the additive white Gaussian noise vector.

Again, let the desired bit be b_1, which is carried by a QOS $\boldsymbol{c}_1$. For the other $K - 1$ QOS, some may have 0 cross-correlation with $\boldsymbol{c}_1$ and let the cross-correlations of the remaining QOS with $\boldsymbol{c}_1$ be $\delta_{1,i}, i = 2, 3, \cdots, k$, where $k < K$ is the number of QOS that have non-zero cross-correlation with $\boldsymbol{c}_1$. Let $n_r = \sum_{i=2}^{k} b_i \delta_{1,i} + n$. The received signal for bit b_1 after despreading is written as

$$\begin{aligned} r_1 &= b_1 + \boldsymbol{c}_1^T \sum_{i=2}^{k} b_i c_i + \boldsymbol{c}_1^T \boldsymbol{n} \\ &= b_1 + \sum_{i=2}^{k} b_i \delta_{1,i} + n \\ &= b_1 + n_1 \end{aligned} \tag{3.11}$$

where $(.)^T$ stands for transpose and for simplicity n_1 is simply written as n.

Random variable $Y_2 = b_2, Y_3 = b_3, \cdots, Y_k = b_k$ are independent identically distributed and $P(Y_i = -1) = P(Y_i = 1) = \frac{1}{2}, i = 2, 3, \cdots, k$. Let random variable $X_2 = b_2\delta_{1,2}, X_3 = b_3\delta_{1,3}, \cdots, X_k = b_k\delta_{1,k}$. $X_2, X_3, \cdots, X_k$ are also independent identically distributed nd $P(X_i = -\delta_{1,k}) = P(X_i = \delta_{1,k}) = \frac{1}{2}, i = 2, 3, \cdots, k$.

Suppose $Z_1, Z_2, \cdots, Z_n$ are independent and identically distributed random variables, each with expectation μ and variance σ^2. By invoking the central limit theorem, the mean and variance of $Z_1 + Z_2, \cdots, +Z_n$ are $n\mu$ and $n\sigma^2$, respectively. Hence,

$$X_2 + X_3, \cdots + X_k = \sum_{i=2}^{k} b_i\delta_{1,i} \sim N\left(0, \sum_{i=2}^{k} \delta_{1,i}^2\right). \tag{3.12}$$

Also, $\sum_{i=2}^{k} b_i\delta_{1,i}$ and n are independent. Thus, $n_r = \sum_{i=2}^{k} b_i\delta_{1,i} + n$ can be written as

$$n_r = \sum_{i=2}^{k} b_i\delta_{1,i} + n \sim N\left(0, \sum_{i=2}^{k} \delta_{1,i}^2 + \sigma^2\right). \tag{3.13}$$

However, the central limit theorem is accurate only when the number of random variables is sufficiently large; otherwise, it can only be regarded as an upper bound. The upper bound on the bit error probability is

$$P_b \leq P(n_r > \sqrt{E_b}) = \int_{\sqrt{E_b}}^{\infty} \frac{1}{\sqrt{2\pi(\sum_{i=2}^{k} \delta_{1,i}^2 + \sigma^2)}} e^{-\frac{x^2}{2(\sum_{i=2}^{k} \delta_{1,i}^2 + \sigma^2)}} dx. \tag{3.14}$$

This equation can be simplified by using the Q-function as

$$\begin{aligned} P_b &\leq Q\left(\sqrt{\frac{E_b}{\sum_{i=2}^{k} \delta_{1,i}^2 + \sigma^2}}\right) \\ &\leq Q\left(\sqrt{\frac{E_b}{\sum_{i=2}^{k} \delta_{1,i}^2 + \frac{N_0}{2}}}\right). \end{aligned} \tag{3.15}$$

This equation also shows that interference power for any bit increases as the number of QOS used increases, as expected. Consequently, BER performance gets worse due to accumulation of the non-zero cross-correlation $\sum_{i=2}^{k} \delta_{1,i}^2$. As an extreme case, when the

number of QOS increases to infinity, BER will approach 0.5.

With some careful selection, QOS of lengths 16, 32, 64 and 128 all result in the same BER for the same transmission rate. The BER upper bounds for for different transmission rates (note that $2B$, where B is the baseband bandwidth, is the Nyquist transmission rate) versus SNR are shown in Table 3.2.

Table 3.2: BER Upper Bound for Faster-Than-Nyquist Signal

SNR (dB)	**Transmission Rate (B)**			
	2.25	**2.5**	**2.75**	**3**
0	0.10295	0.12411	0.14253	0.15866
1	0.08320	0.10692	0.12756	0.14553
2	0.06594	0.09179	0.11440	0.13407
3	0.05137	0.07877	0.10306	0.12420
4	0.03948	0.06781	0.09344	0.11584
5	0.03009	0.05875	0.08541	0.10885
6	0.02288	0.05137	0.07878	0.10306
7	0.01746	0.04542	0.07335	0.09835
8	0.01346	0.04069	0.06894	0.09442
9	0.01053	0.03693	0.06539	0.09130
10	0.00842	0.03393	0.06253	0.08878

Table 3.3: The Number of QOS Required for Different Transmission Rates

Length	**Transmission Rate (B)**			
	2.25	**2.5**	**2.75**	**3**
16	18	20	22	24
32	36	40	44	48
64	72	80	88	96
128	144	160	176	192

3.4 Detection Method

The traditional correlation detection scheme includes two steps in the receiver: multiplication and integration. In multiplication step, the received signal is multipled by the despreading code. After multiplication, all the chips in the sequence are integrated to generate the output in the integration step. But the correlation detector might not

work well when QOS are used to transmit at faster-than-Nyquist rates because of the co-channel interference to each bit. Zhang *et al.* propose a detection method that uses the second order moment of auto-correlation to improve the performance of detection [40]. Simulation results indicate BER performance in frequency domain is better than in time domain by adopting a fourth-moment correlation scheme in [41]. Shen *et al.* describe a detection method that is efficient in complex environments; it uses a time-varying signal processing technique to mitigate interference and uses the minimum mean-square estimation to detect the signal [31]. However, these detection methods are not suitable for the signals with QOS, since all these detection methods focus on orthogonal codes.

In this section, the traditional correlation detection will be used as a baseline case, and a three-layer MLP detection scheme and an LSTM detection scheme will be proposed for QOS detection.

3.4.1 Correlation Detection

In a correlation detector for BPSK signals as shown in Figure 3.1, a bit decision of '1' is made if the received power is higher than the threshold 0; otherwise, a bit decision of '-1' is made. A question being answered here is: are there detectors that will work better than the traditional correlation detector for transmission at faster-than-Nyquist rates with QOS? This is studied in the next sections.

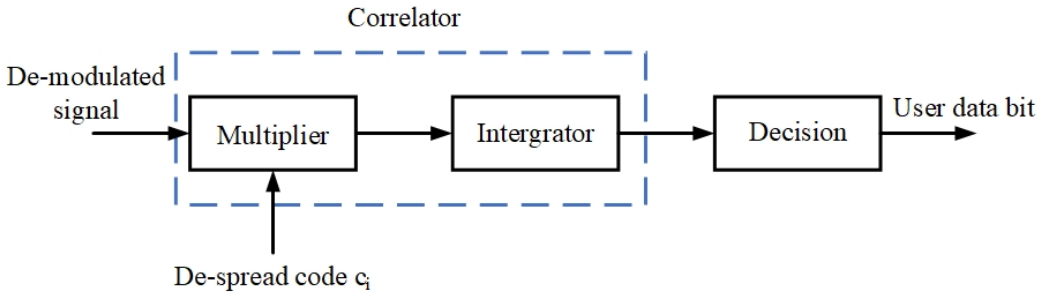

Figure 3.1: Correlation Detection Structure.

3.4.2 Three-Layer MLP Detection

A critical step in the correlation detector is the multiplier. This detection scheme can be regarded as a one-layer detection scheme.

Multi-layer perceptron is a feed-forward neural network that generates a set of outputs from a set of inputs. An MLP structure includes at least three layers. The first layer is the input layer; the last layer is the output layer. Intermediate layers are hidden layers; there is at least one hidden layer in MLP. MLP uses back-propagation for training the network. Back-propagation is a commonly used method by the gradient descent optimization algorithm that calculates the gradients of the loss function. These gradients are also known as errors. Then the weight matrix is adjusted in MLP by these gradients.

MLP is a traditional deep learning method, but has been used as signal detection method in CDMA and channel estimation method in OFDM. Kechriotis *et al.* use neural work in synchronous and asynchronous CDMA and get a better performance than correlation detection [17]. Aazhang *et al.* adopt MLP for signal detection in CDMA systems, and the performance is shown to be much better than traditional correlation detection [1]. Chinthaginjala *et al.* develop a three-layer MLP detection scheme in MC-CDMA, which is shown to have a good performance in canceling multiple access interference compared to maximal ratio combining, equal gain combining, and minimum mean square error [4]. MLP is proposed as channel estimator in OFDM systems are is compared with least-squares (LS) algorithm, radial basis function neural network (RBF), and minimum mean-square error algorithm (MMSE) [32]. It is shown that MLP has a better BER performance and mean-square error than LS algorithm and RBF algorithm, but has a similar performance as MMSE. MLP is a lower complexity, faster convergence, and better performance method than existing schemes to reduce the powder envelope fluctuations in OFDM [14].

A three-layer MLP detection structure is shown in Figure 3.2. The forward process will be introduced first. After the de-spreading multiplier module, instead of integrating all bits in sequence to get a final result, a three-layer MLP structure is applied to separate desired signal from interference and noise. The outputs of MLP is the probabilities of bit ‘1’ and bit ‘-1’.

The de-spread signal will be the input layer for MLP detection. The input size is

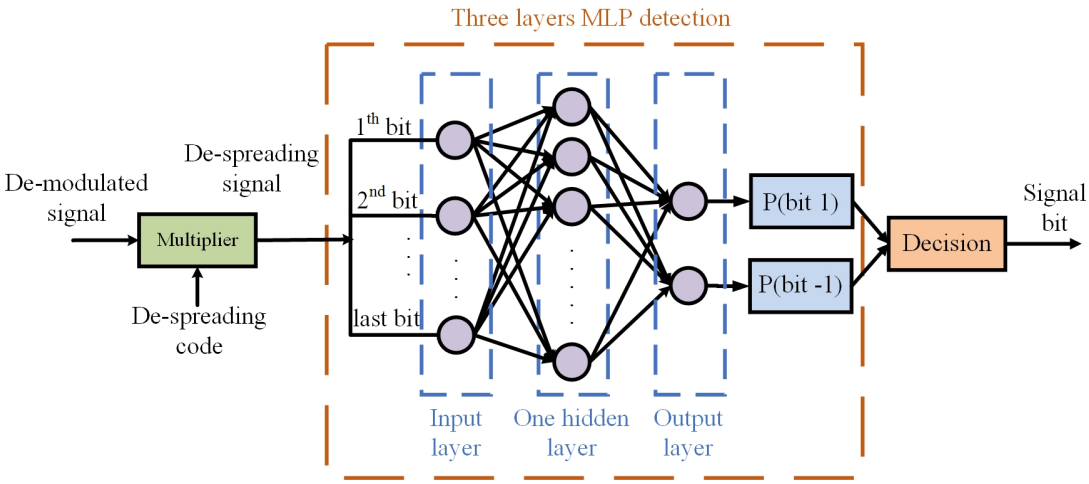

Figure 3.2: Three-Layer MLP Detection Structure in DSSS.

equal to sequence length L. Every bit in the de-spread signal is one node in the input layer. If the de-spread signal length is 16, then there are 16 nodes in the input layer. All sequence bits feed into MLP as input nodes simultaneously. Suppose the input vector is $\boldsymbol{x}$ (dimension is $L \times 1$), the hidden layer can be constructed as

$$h(x) = g(W_1^T x + b_1) \tag{3.16}$$

where W_1 is the weight matrix $(L \times N_h)$ in the hidden layer, b_1 is a bias vector with dimension $L \times 1$ in the hidden layer, and $g(x)$ is a non-linear active function. Relu function $g(x) = \max(0, x)$ is used as the active function. Each column w_i^1 in W_1 represents the weight from input nodes to the i-th hidden neuron.

The output is obtained by

$$f(x) = W_2^T h(x) + b_2 \tag{3.17}$$

where W_2 is the weight matrix $(N_h \times 2)$ in the output layer and b_2 is a bias vector with dimension 2×1 in the output layer. Each column w_j^2 in W_2 represents the weight from the hidden neuron to the j-th output nodes. The output $f(x)$ is the probability for predicted classes.

The goal of back-propagation is to map actual outputs closer to target outputs by

updating the weights and biases in the neural network. The output errors are computed by comparing actual outputs and target outputs in the output layer. Because all the neurons in the network except input nodes have contributions to the actual outputs, the output errors are transmitted backwards from the output layer to previous layers that have influence on output errors. Every neuron in the network will be informed its contributions to the output layer. By this information, the weight matrix and bias will be updated in each layer to reduce the output errors. This is the back-propagation process to reduce output errors.

Cross-entropy loss function is used to measure the performance of a classification model. The output of the cross-entropy loss function is a probability value. The cross-entropy loss function in our classification is

$$E = -y \log f(x) - (1 - y) \log(1 - f(x)) \tag{3.18}$$

where y is the binary indicator ('0' or '1') if the output label is the correct classification for observation and $f(x)$ is the correct prediction probability.

Eq. (3.18) shows that a low prediction probability to the true output class and a high prediction probability to the wrong output class result in a high cross-entropy loss value. For example, if a prediction probability is 0.1 for the true output label, then the loss value will be high. If a loss function value is zero, then it means that the probability is 1 for the true output class, which means that the prediction perfectly fits the actual classification result.

Then gradient descent is used to look for the minimum value of the cross-entropy loss function in back-propagation. In our three-layer MLP, w_1, w_2, b_1 and b_2 need to be optimized to minimize the loss value. The effect of these parameters on the output errors could be defined by partial derivative as $\frac{\delta E}{\delta w_1}$, $\frac{\delta E}{\delta w_2}$, $\frac{\delta E}{\delta b_1}$ and $\frac{\delta E}{\delta b_2}$. By chain rule, we

get

$$\begin{aligned}
\frac{\delta E}{\delta w_1} &= \frac{\delta E}{\delta f(x)} \frac{\delta f(x)}{\delta h(x)} \frac{\delta h(x)}{\delta g(x)} \frac{\delta g(x)}{\delta w_1} \\
\frac{\delta E}{\delta w_2} &= \frac{\delta E}{\delta f(x)} \frac{\delta f(x)}{\delta w_2} \\
\frac{\delta E}{\delta b_1} &= \frac{\delta E}{\delta f(x)} \frac{\delta f(x)}{\delta h(x)} \frac{\delta h(x)}{\delta g(x)} \frac{\delta g(x)}{\delta b_1} \\
\frac{\delta E}{\delta b_2} &= \frac{\delta E}{\delta f(x)} \frac{\delta f(x)}{\delta b_2}.
\end{aligned} \tag{3.19}$$

The weights and biases can be updated by learning rate η. Learning rate is an important hyper-parameter in deep learning that updates the weights and biases with corresponding gradients. The updated weights and biases are

$$\begin{aligned}
w_1^{'} &= w_1 - \eta \frac{\delta E}{\delta w_1} \\
w_2^{'} &= w_2 - \eta \frac{\delta E}{\delta w_2} \\
b_1^{'} &= b_1 - \eta \frac{\delta E}{\delta b_1} \\
b_2^{'} &= b_2 - \eta \frac{\delta E}{\delta b_2}.
\end{aligned} \tag{3.20}$$

Through an applicable learning rate, the cross-entropy loss could converge to the minimum value quickly.

Learning rate and the number of neurons in the hidden layer are two important parameters in MLP and must be carefully considered. If the learning rate is too small, gradient decent may be too slow to converge, known as under-fitting. However, when it is too large, gradient decent may overshoot the minimum value that may fail to converge, known as over-fitting. The number of neurons has a tremendous effect on the output results. If the number of neurons is too small, it causes under-fitting. If there are too many neurons in the hidden layer, it may result in over-fitting. According to [11], the number of hidden neurons should be less than twice the size of the input layer. According to these, the parameters of three-layers MLP detection structure is shown in Table 3.4. The number of neurons in hidden layer is $\frac{3}{2}$ of the size of input layer. Learning rates are chose from $0.0004, 0.0005, 0.0006$.

Table 3.4: Parameters of Three-Layer MLP Detection for Different Code Lengths

Length(2^m)	**Neurons in Hidden layer**	**Learning Rate**
16	24	0.0004, 0.0005, 0.0006
32	48	0.0004, 0.0005, 0.0006
64	96	0.0004, 0.0005, 0.0006
128	192	0.0004, 0.0005, 0.0006

3.4.3 LSTM Detection

Recurrent neural network (RNN) is a neural network designed to predict time series prediction problems. Since the received signal is a time series signal and the desired signal is a single bit, RNN is suitable to predict this many-to-one problem.

The idea behind RNN is to make use of the information among the bits in sequence. In a traditional neural network, like MLP, we assume that all input nodes are independent of one another. That is, previous nodes will not affect the current node. But in detecting the desired signal, this is not a good idea. Every bit in sequence is dependent upon other bits. Since the bits in a sequence should be all negative value or all positive value after de-spreading in a noiseless channel, in a noisy channel, the current bit should have a high probability to be a negative value if most of previous bits are negative and vice verse. The output of each node depends on the previous computations in RNN.

RNN has been a popular detection method used in signal detection. For example, it is used to detect different signal bits in DSSS [33]. The performance matches the maximum likelihood detection. Chuah *et al.* provide a three-layer RNN to implement linear de-correlation detection in DSSS and get a robust performance [5]. CNN and RNN are combined for channel equalization for non-linear distortion and other signal in [20], where it is found that the performance in QPSK outperforms other equalizers by average 2 to 5 dB at low SNR. Luo *et al.* develop a neural network that combines CNN and RNN to predict channel state information [22]. Experimental results show that it achieves a highly accurate prediction and it converges quickly.

A traditional RNN structure for many-to-one model is shown in Figure 3.3. It includes an input layer with nodes $x_1, x_2, ..., x_n$, one hidden layer with cells $h_1, h_2, ..., h_n$ and an output layer y. In this model, $x_0, x_1, ..., x_n$ are inputs, $h_0, h_1, ..., h_n$ are hidden states in time series, W_{hh} is hidden cell to hidden cell weight, W_{xh} is the input layer to

hidden layer weight, W_{hh} and W_{xh} are re-used at every time step, f_w is the tanh function with parameter W_{hh}, and b_1 is the bias vector in the hidden layer.

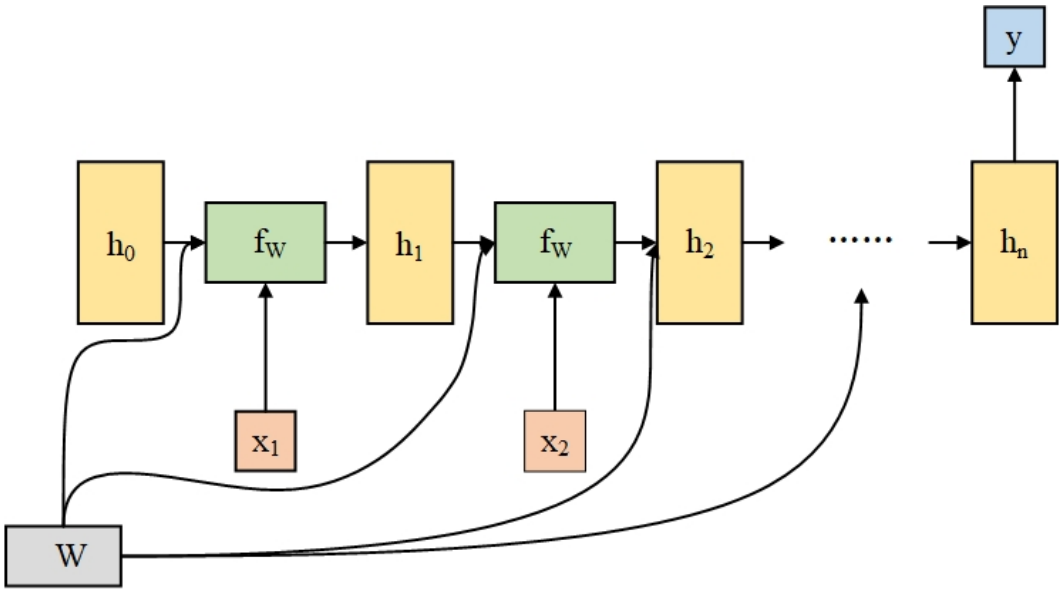

Figure 3.3: RNN Structure for Many Input and One Output.

We can process a sequence with bit x_0 to x_n by applying a recurrence formula at every time step as

$$\begin{aligned} h_t &= f_w(h_{t-1}, x_t) + b_1 \\ &= \tanh(W_{hh}h_{t-1} + W_{xh}x_t) + b_1, t = 1, \cdots, n. \end{aligned} \tag{3.21}$$

In order to simplify Eq. (3.21), $W = [W_{hh}\ W_{xh}]$. Then we get

$$h_t = \tanh(W \begin{bmatrix} h_{t-1} \\ x_t \end{bmatrix}) + b_1. \tag{3.22}$$

In the output layer, W_{hy} is the weight matrix of the hidden layer to the output layer and b_2 is the bias vector. The output y is

$$y = W_{hy}h_n + b_2. \tag{3.23}$$

The performance of RNN is measured by a loss function E. Weights and biases need to be optimized for best loss value. The back-propagation $\frac{\delta E}{\delta W}$ is defined as

$$\frac{\delta E}{\delta W} = \frac{\delta E}{\delta y}\frac{\delta y}{\delta h_n}\frac{\delta h_n}{\delta W} + \frac{\delta E}{\delta y}\frac{\delta y}{\delta h_n}\frac{\delta h_n}{\delta h_{n-1}}\frac{\delta h_{n-1}}{\delta W} + \cdots + \frac{\delta E}{\delta y}\frac{\delta y}{\delta h_n}\frac{\delta h_n}{\delta h_{n-1}}\cdots\frac{\delta h_1}{\delta W}. \tag{3.24}$$

In order to simplify this equation, we focus on descent item $\frac{\delta h_t}{\delta h_{t-1}}$. From Eq. (3.22), we get

$$\begin{aligned}\frac{\delta h_t}{\delta h_{t-1}} &= \frac{\delta h_t}{\delta f_w}\frac{\delta f_w}{\delta h_{t-1}} \\ &= \frac{\delta h_t}{\delta f_w}W \\ &= \tanh^{'} W\end{aligned} \tag{3.25}$$

The back-propagation $\frac{\delta E}{\delta W}$ can be simplified as

$$\frac{\delta E}{\delta W} = \frac{\delta E}{\delta y}\frac{\delta y}{\delta h_n}\frac{\delta h_n}{\delta W} + \frac{\delta E}{\delta y}\frac{\delta y}{\delta h_n}\frac{\delta h_{n-1}}{\delta W}\tanh^{'} W + \cdots + \frac{\delta E}{\delta y}\frac{\delta y}{\delta h_n}\frac{\delta h_1}{\delta W}(\tanh^{'} W)^{n-1} \tag{3.26}$$

For simplicity, assume W is diagonalizable. Then W can be expressed as

$$W = E^{-1}\begin{bmatrix}\lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_k\end{bmatrix}E \tag{3.27}$$

where $\lambda_1, \lambda_2, \cdots, \lambda_k$ are the singular values of W and k depends on the size of W.

Hence, W^{n-1} can be expressed as

$$W^{n-1} = E^{-1}\begin{bmatrix}\lambda_1^{n-1} & & & \\ & \lambda_2^{n-1} & & \\ & & \ddots & \\ & & & \lambda_k^{n-1}\end{bmatrix}E. \tag{3.28}$$

There are two situations that have negative influence on the optimization of W. Suppose the largest singular value is $\lambda_{\max}$. If $\lambda_{\max} > 1$, then during the back-propagation path, the gradient $\frac{\delta E}{\delta W}$ becomes very large and is multiplied by the weight matrix W over

and over again. That is called exploding gradient. Exploding gradient can result in an unstable network that is unable to learn new information from training data. Another situation is when $\lambda_{\max} < 1$, the gradient $\frac{\delta E}{\delta W}$ becomes very small after back-propagation. That is vanishing gradient. Vanishing gradient leads to long time training process and the prediction accuracy will decrease [2][12]. If all singular values are 1, then the training process becomes strange because the gradient does not change in back-propagation.

An efficient method to figure out exploding gradient and vanishing gradient is long short term memory (LSTM) shown in Figure 3.4. A LSTM cell is composed of a forget gate f that makes a decision on what information we are going to erase in the cell state, an input gate i that makes a decision on whether new information to put into the cell state, an update gate g that createa new information, and an output gate o that decides what part of cell state that output to next LSTM cell.

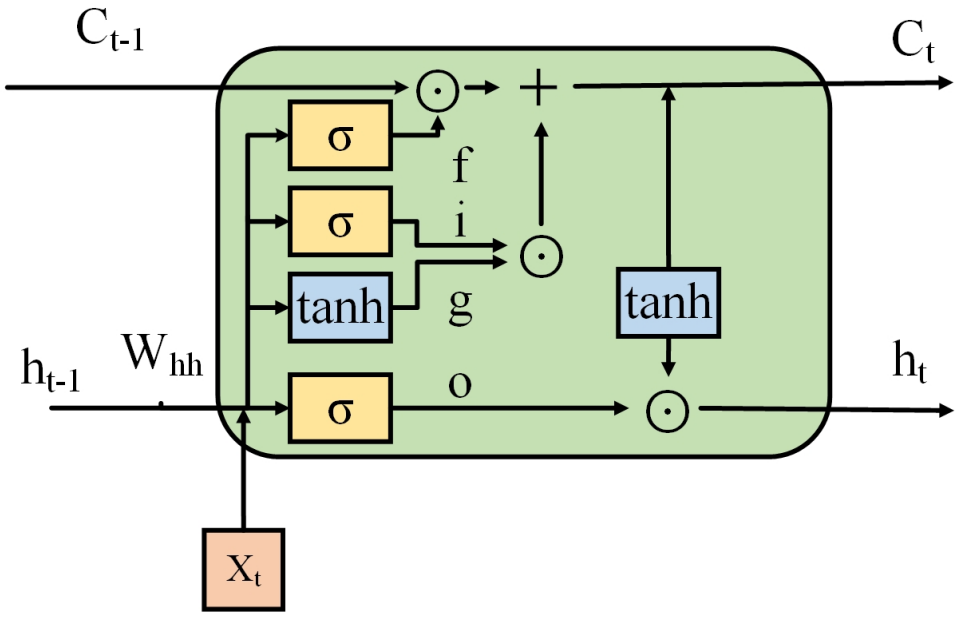

Figure 3.4: LSTM Cell Structure.

Since some bits in the sequence may be affected by interference and noise, which makes it unreliable to support the final prediction, we want to filter them out to maintain the accuracy of prediction. A forget gate f is used to make the selection from previous

cell state as

$$f = \sigma\left(W_f \begin{bmatrix} h_{t-1} \\ x_t \end{bmatrix} + b_f\right). \tag{3.29}$$

In Eq. (3.29), W_f and b_f are the weight matrix and bias vector in the forget gate, respectively. The next step is to decide what new information from the input and previous cell state to store in the current cell state. An input gate i decides whether to update new information and an update gate g creates new information to be updated. In Eq. (3.30), W_i and b_i are the weight matrix and bias in the input gate. In Eq. (3.31), W_g and b_g are the weight matrix and bias vector in the update gate.

$$i = \sigma\left(W_i \begin{bmatrix} h_{t-1} \\ x_t \end{bmatrix} + b_i\right) \tag{3.30}$$

$$g = \tanh\left(W_g \begin{bmatrix} h_{t-1} \\ x_t \end{bmatrix} + b_g\right). \tag{3.31}$$

A result from Hadamard product between i and g is used to update the current cell state C_t from C_{t-1}

$$C_t = f \odot C_{t-1} + i \odot g. \tag{3.32}$$

Finally the output h_t depends on output gate and current cell. The output gate decides what part of the current cell to output. In Eq. (3.33), W_o and b_o are the weight matrix and bias vector in output gate, respectively.

$$\begin{aligned} o &= \sigma\left(W_o \begin{bmatrix} h_{t-1} \\ x_t \end{bmatrix} + b_o\right) \\ h_t &= o \odot \tanh(C_t). \end{aligned} \tag{3.33}$$

A back-propagation to the weight of LSTM is

$$\frac{\delta E}{\delta W} = \frac{\delta E}{\delta h_n}\frac{\delta h_n}{\delta C_n}\frac{\delta C_n}{\delta W} + \frac{\delta E}{\delta h_n}\frac{\delta h_n}{\delta C_n}\frac{\delta C_n}{\delta C_{n-1}}\frac{\delta C_{n-1}}{\delta W} + \cdots + \frac{\delta E}{\delta h_n}\frac{\delta h_n}{\delta C_n}\frac{\delta C_n}{\delta C_{n-1}}\cdots\frac{\delta C_1}{\delta W} \tag{3.34}$$

The gradient $\frac{\delta C_t}{\delta C_{t-1}} = f$; hence, Eq. (3.34) can be written as

$$\frac{\delta E}{\delta W} = \frac{\delta E}{\delta h_n}\frac{\delta h_n}{\delta C_n}\frac{\delta C_n}{\delta W} + \frac{\delta E}{\delta h_n}\frac{\delta h_n}{\delta C_n}\frac{\delta C_{n-1}}{\delta W} f_n + \cdots + \frac{\delta E}{\delta h_n}\frac{\delta h_n}{\delta C_n}\frac{\delta C_1}{\delta W} f_n f_{n-1} \cdots f_1. \quad (3.35)$$

Instead of multiplying a weight matrix at every time step in RNN, a forget gate is multiplied in LSTM. Avoid multiplying with a weight matrix in the gradient has two advantages. First, multiplying with a sigmoid element f is better than a matrix in term of computation; Second, since the forget gate value is different at different time step, the gradient avoids multiplying the same value over and over again. Hence, exploding gradient and vanishing gradient are well controlled by LSTM.

The LSTM detection structure is shown in Figure 3.5. After the de-spreading multiplier module, instead of integrating all bits in sequence to get a final result, every bit in de-spread signal puts into LSTM detection as the input layer. The predicted output y is fed into the decision module to decide the signal bit.

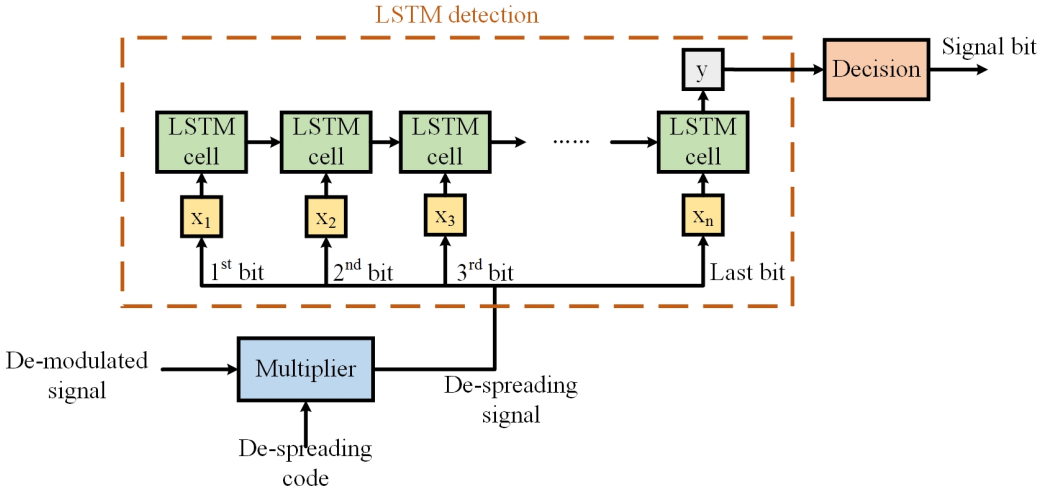

Figure 3.5: LSTM Detection Structure.

The learning rate and the number of neurons in the hidden layer are two important parameters in LSTM and must be carefully considered. Same as MLP detection, a small learning rate and a small hidden layer size will result in under-fitting. A large learning rate and a large hidden layer size will result in over-fitting. Based on these, the

parameters of LSTM detection structure are shown in Table 3.5. The hidden layer size is $\frac{3}{4}$ of the size of the input layer. Learning rates are chosen from $0.0003, 0.0004, 0.0005$.

Table 3.5: Parameters of LSTM Detection for Different Code Lengths.

Length(2^m)	**Neurons in Hidden layer**	**Learning Rate**
16	12	0.0003, 0.0004, 0.0005
32	24	0.0003, 0.0004, 0.0005
64	48	0.0003, 0.0004, 0.0005
128	96	0.0003, 0.0004, 0.0005

3.5 Summary

The use of QOS for transmission at faster-than-Nyquist rates is discussed, followed by the derivation of the BER upper bound with correlation detection. Through carefully selection of QOS, a relatively low BER can be achieved. After brief introduction of correlation detection, a three-layer MLP detection scheme and an LSTM detection scheme, two deep-learning methods that are traditionally used in classification prediction, are developed for detecting the transmitted information when QOS are used for transmission at faster-than-Nyquist rates. The two proposed prediction methods could extract useful information out of strong interference and noise, making them suitable detection methods for the proposed signaling scheme.

Chapter 4: RESULTS

4.1 Introduction

To verify the BER performance of the proposed three-layer MLP and LSTM detection schemes, results for different numbers of spreading codes are obtained in this chapter. There are 60000 pieces of data to evaluate the performance of correlation detection as a reference, while the two new detection methods adopt 60000 pieces of training data to build the model and 6000 pieces of test data to evaluate the performance with cross-entropy loss function and SGD optimizer. Since the number of QOS used to transmit information bits is inversely proportional to the BER performance, the maximum number of spreading codes should be no more than 1.5 times the number of Walsh codes. For length 16, the number of codes chosen are 18, 20, 22 and 24. For length 32, the number of codes chosen are 36, 40, 44 and 48. For length 64, the number of codes chosen are 72, 80, 88 and 96. For length 128, the number of codes chosen are 144, 160. 176 and 192. SNR ranges from 0 to 10 dB with a step size of 1 dB.

BER performance comparison between theoretical and simulation result of correlation detection using QOS are discussed Section 4.2. Three-layer MLP detection and LSTM detection results are discussed at Section 4.3 and Section 4.4, respectively.

4.2 Theoretical and Simulation Result of Correlation Detection

QOS of lengths 16, 32, 64 and 128 will have the same theoretical BER by code selection that is shown in Table 3.2. For correlation detection, each simulation result is the averaged BER among all spreading codes.

Theoretical and simulation results for QOS of lengths 16, 32, 64 and 128 are shown in Figure 4.1, 4.2, 4.3, 4.4, respectively. The maximum cross-correlation values for these QOS lengths are 0.25, 0.25, 0.125 and 0.125, respectively. For different number of spreading codes when SNR ranges from 0 dB to 10 dB, the simulation and theoretical results match well, validating the theoretical BER upper bound for faster-than-Nyquist

signaling with BPSK.

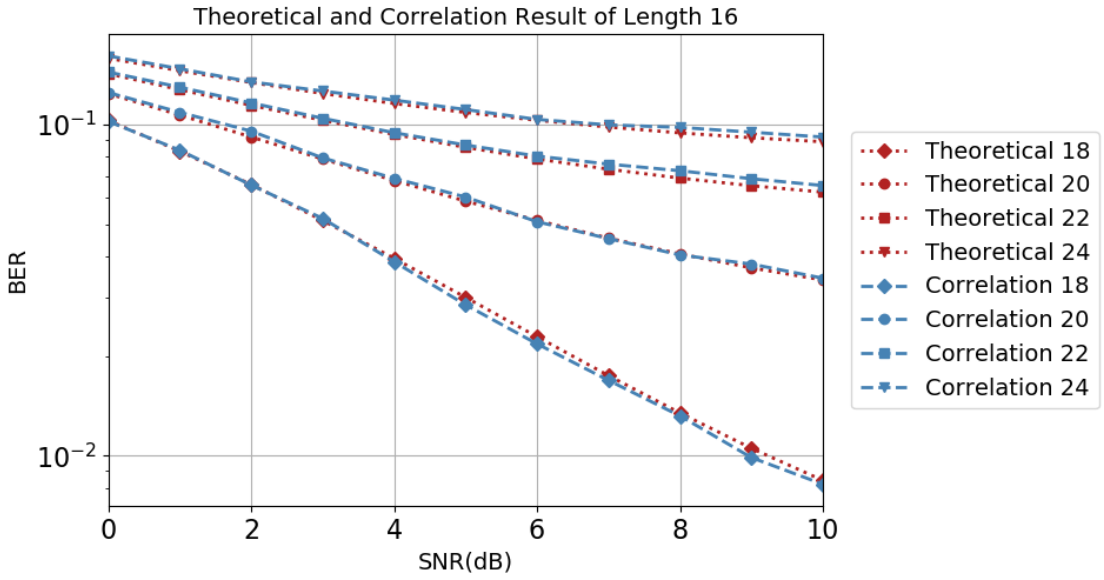

Figure 4.1: Theoretical and Simulation Results of Length 16.

4.3 Three-Layer MLP Detection Results

Three-layer MLP detection is more complex than correlation detection due to intermediate layers and back-propagation. Instead of making decision by comparing integration result with a threshold, the three-layer MLP detection scheme builds a model with 60000 pieces of training data going through 50 epoches, and then the prediction is made by the trained model for 6000 pieces of test data to generate the predicted signal bits. It turns out that three-layer MLP detection has a better BER performance than correlation detection.

For correlation detection and three-layer MLP detection of length 16, each simulation result is the averaged BER among the spreading codes shown in Figure 4.5. The BER improvement by three-layer MLP detection of code length 16 is shown in Table 4.1. For the number of codes 18, 20, 22 and 24, BER improvements increase from 7.792% to 85.772%, 11.559% to 69.285%, 13.420% to 54.307% and 9.075% to 39.397%, respectively,

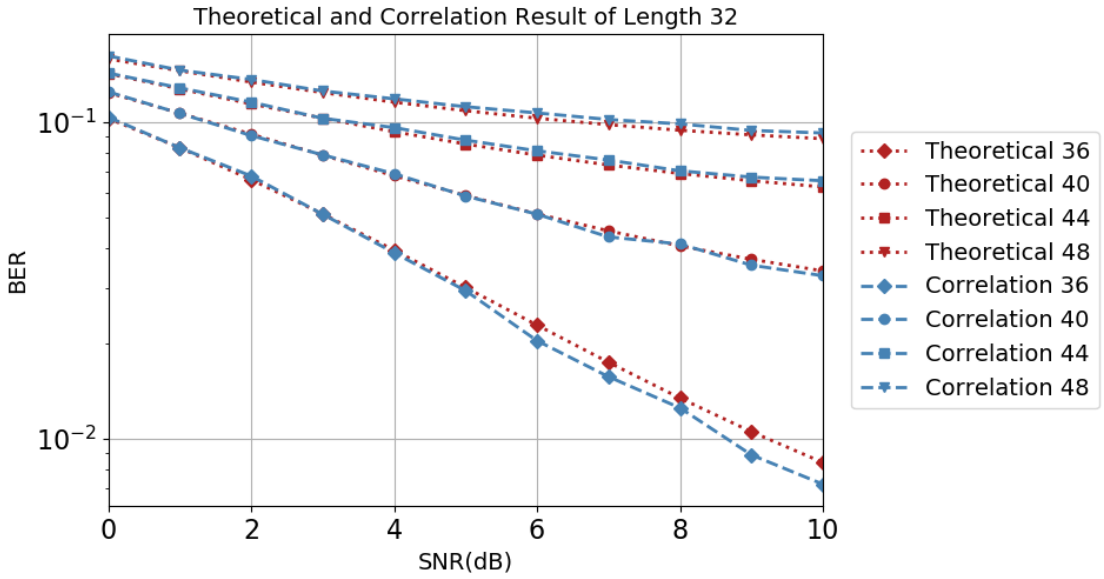

Figure 4.2: Theoretical and Simulation Results of Length 32.

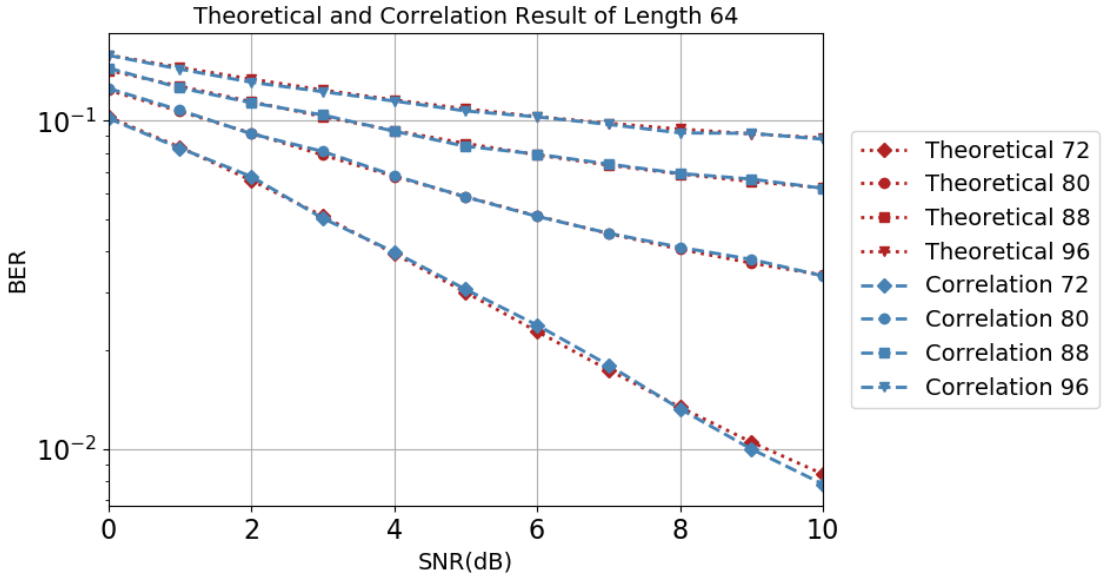

Figure 4.3: Theoretical and Simulation Results of Length 64.

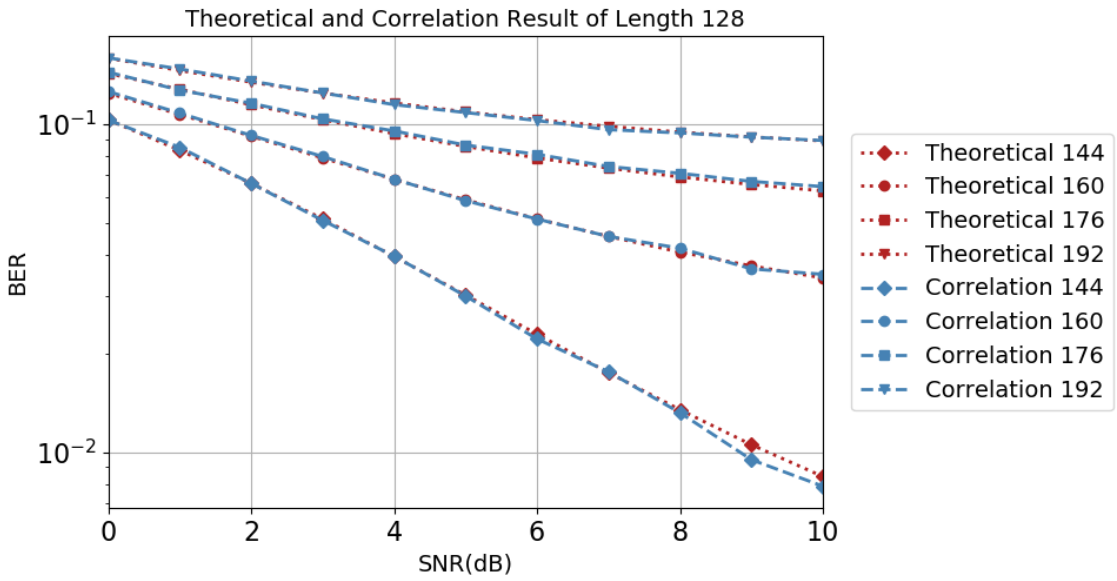

Figure 4.4: Theoretical and Simulation Results of Length 128.

when SNR ranges from 0 dB to 10 dB. The minimum BER improvement is 7.792%, achieved at the number of codes 18 and SNR 0 dB. The maximum BER improvement is 85.772%, achieved at the number of codes 18 and SNR 10 dB. In three-layer MLP detection of length 16, improvement decreases as SNR decreases and as the number of codes used increases.

Table 4.1: Improvement of Three-Layer MLP Detection for Length 16

Improvement	**The number of codes**			
	18	**20**	**22**	**24**
Minimum(%)	7.792	11.559	13.420	9.075
Maximum(%)	85.772	69.285	53.307	39.397

Because simulation for three-layer MLP is time-consuming, the BER performance for length 32, 64 and 128 represents the BER with one particular code, instead of average BER over all codes. The comparison between correlation detection and three-layer MLP detection of length 32 is shown in Figure 4.6. The BER improvement by one spreading code of code length 32 is shown in Table 4.2. For the number of codes 36, 40, 44 and

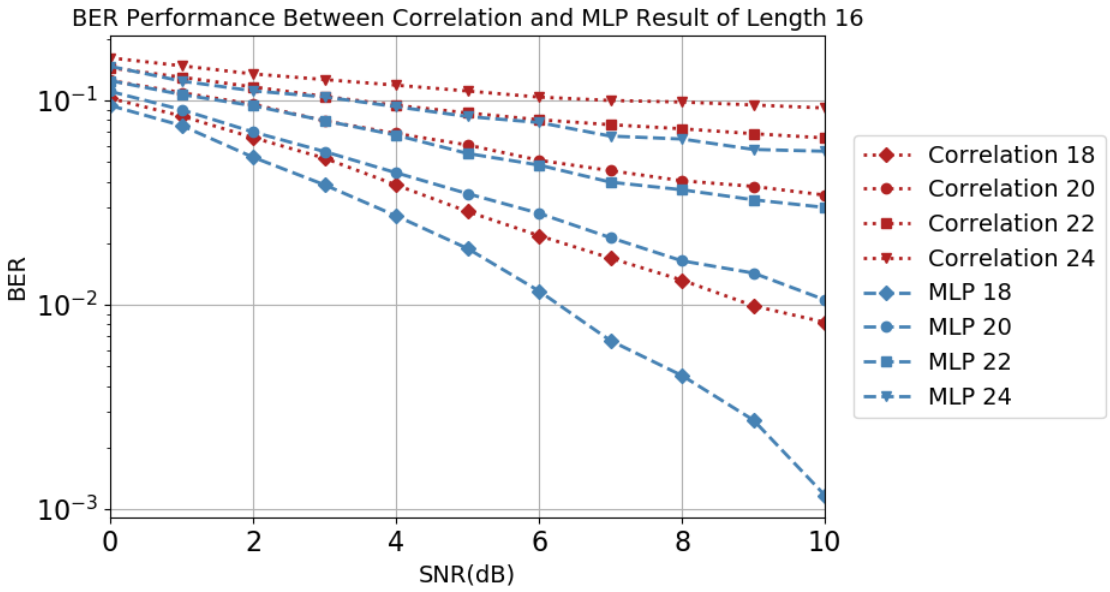

Figure 4.5: BER Performance between Correlation and MLP Results of Length 16.

48, BER improvements increase from 8.826% to 51.624%, 10.756% to 61.337%, 12.242% to 50.795% and 9.176% to 40.235%, respectively, when SNR ranges from 0 dB to 10 dB. The minimum BER improvement is 8.826%, achieved at the number of codes 36 and SNR 1 dB. The maximum BER improvement is 61.337%, achieved at the number of codes 40 and SNR 10 dB.

Table 4.2: Improvement of Three-Layer MLP Detection for Length 32 (One code)

Improvement	**The number of codes**			
	36	**40**	**44**	**48**
Minimum(%)	8.826	10.756	12.242	9.176
Maximum(%)	51.624	61.337	50.795	40.235

The comparison between correlation detection and three-layer MLP detection of length 64 is shown in Figure 4.7. The BER improvement of code length 64 is shown in Table 4.3. For the number of codes 72, 80, 88 and 96, BER improvements increase from 2.827% to 66.800%, 7.206% to 65.422%, 11.425% to 49.780% and 3.504% to 31.635%, respectively, when SNR ranges from 0 dB to 10 dB. The minimum BER improvement

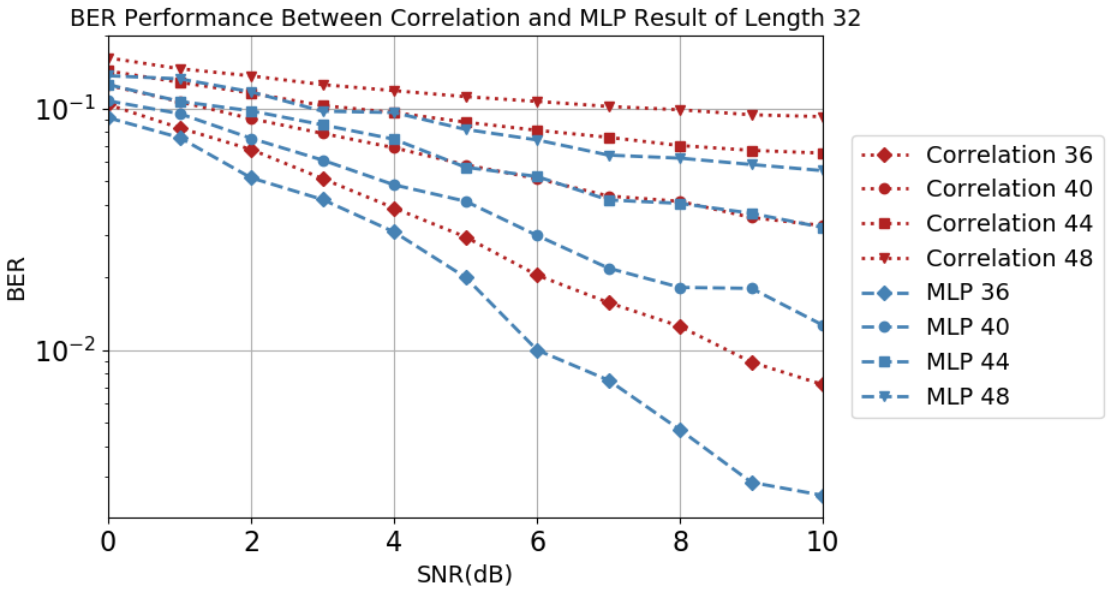

Figure 4.6: BER Performance between Correlation and MLP Results of Length 32.

is 2.827%, achieved at the number of codes 72 and SNR 0 dB. The maximum BER improvement is 66.800%, achieved at the number of codes 72 and SNR 9 dB

Table 4.3: Improvement of Three-Layer MLP Detection for Length 64 (One code)

Improvement	**The number of codes**			
	72	**80**	**88**	**96**
Minimum(%)	2.827	7.206	11.425	3.504
Maximum(%)	66.800	65.422	49.780	31.635

Correlation detection and three-layer MLP detection of length 128 are shown in Figure 4.8. The BER improvement of code length 128 is shown in Table 4.4. For the number of codes 144, 160, 176 and 192, BER improvements increase from 1.236% to 65.189%, 13.601% to 60.741%, 7.714% to 47.742% and 7.440% to 32.252%, respectively, when SNR ranges from 0 dB to 10 dB. The minimum BER improvement is 1.236%, achieved at the number of codes 144 and SNR 0 dB. The maximum BER improvement is 65.189%, achieved at the number of codes 144 and SNR 10 dB.

From these figures, the three-layer MLP detection has better performance than cor-

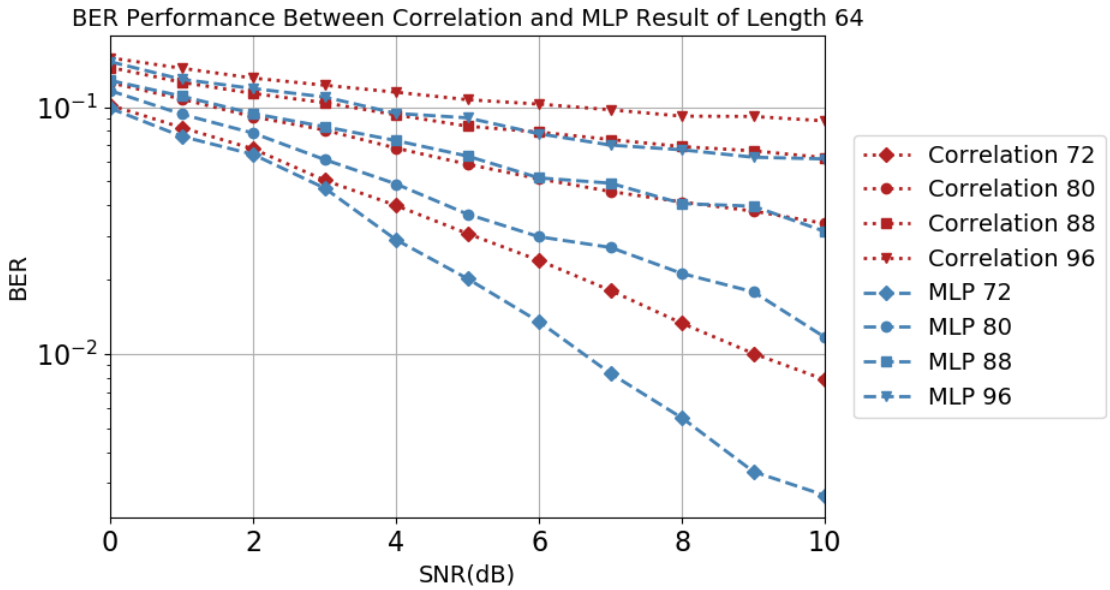

Figure 4.7: BER Performance between Correlation and MLP Results of Length 64.

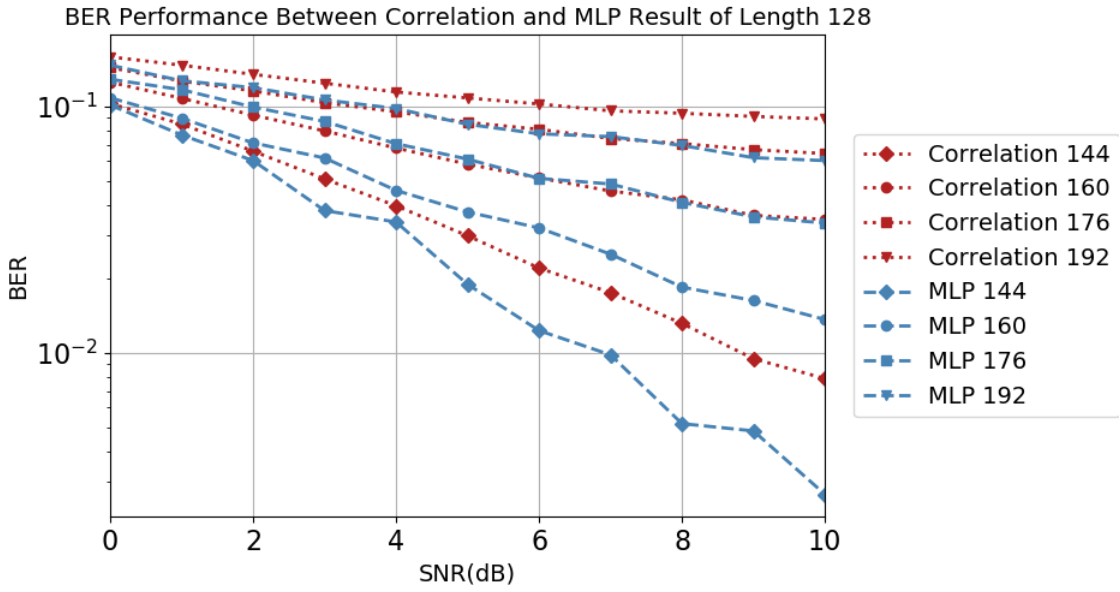

Figure 4.8: BER Performance between Correlation and MLP Results of Length 128.

Table 4.4: Improvement of Three-Layer MLP Detection for Length 128 (One code)

Improvement	**The number of codes**			
	144	**160**	**176**	**192**
Minimum(%)	1.236	13.601	7.714	7.440
Maximum(%)	65.189	60.741	47.742	32.252

relation detection in noisy environments. With the same number of codes used, BER performance improves as SNR increases. At the same SNR level, BER performance improves when fewer QOS are used, which is expected since fewer QOS means less interference. A higher SNR and lower interference would contribute to a better BER performance.

4.4 LSTM Detection Results

LSTM detection builds the model with 60000 pieces of training data. The trained model outputs the predicted signal bits after going through 6000 pieces of testing data. Finally, the BER can be calculated with the predicted and actual transmitted information bits.

The averaged BER performance from all QOS for correlation and LSTM detection of length 16 is shown in Figure 4.9. The BER improvement by LSTM detection of code length 16 is shown in Table 4.5. For the number of codes 18, 20, 22 and 24, BER improvements increase from 7.709% to 87.846%, 12.390% to 70.255%, 13.420% to 54.307% and 9.513% to 38.650%, respectively, when SNR ranges from 0 dB to 10 dB. The minimum BER improvement is 7.709%, achieved at the number of codes 18 and SNR 0 dB. The maximum BER improvement is 87.846%, achieved at the number of codes 18 and SNR 10 dB.

Table 4.5: Improvement of LSTM Detection for Length 16

Improvement	**The number of codes**			
	18	**20**	**22**	**24**
Minimum(%)	7.709	12.390	13.420	9.513
Maximum(%)	87.846	70.255	54.307	38.650

Because the simulation for LSTM detection is time-consuming, the BER performances for lengths 32, 64 and 128 is obtained on one code, instead of the average over

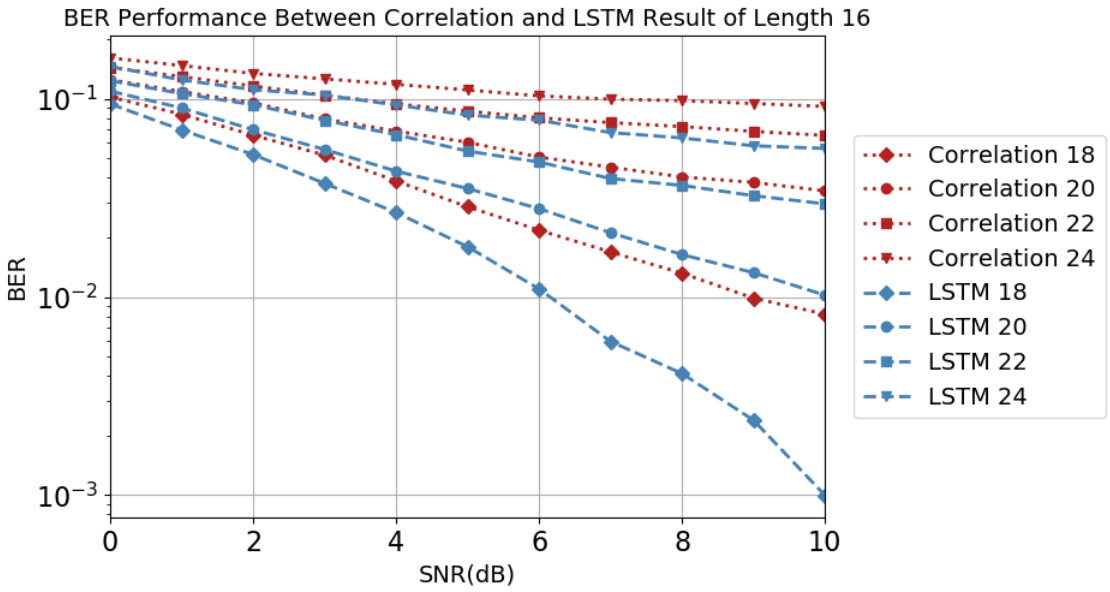

Figure 4.9: BER Performance between Correlation and LSTM Results of Length 16.

all codes. The comparison between correlation and LSTM detections of length 32 is shown in Figure 4.10. The BER improvement by LSTM detection of code length 32 is shown in Table 4.6. For the number of codes 36, 40, 44 and 48, BER improvements increase from 9.426% to 68.718%, 13.565% to 64.388%, 14.574% to 51.315% and 11.111% to 40.783%, respectively, when SNR ranges from 0 dB to 10 dB. The minimum BER improvement is 9.426%, achieved at the number of codes 36 and SNR 1 dB. The maximum BER improvement is 68.718%, achieved at the number of codes 36 and SNR 9 dB.

Table 4.6: Improvement of LSTM Detection for Length 32 (One code)

Improvement	**The number of codes**			
	36	**40**	**44**	**48**
Minimum(%)	9.426	13.565	14.574	11.111
Maximum(%)	68.718	64.388	51.315	40.783

The comparison between correlation and LSTM detection of length 64 is shown in Figure 4.11. The BER improvement by LSTM detection of code length 64 is shown in Table 4.7. For the number of codes 72, 80, 88 and 96, BER improvements increase from

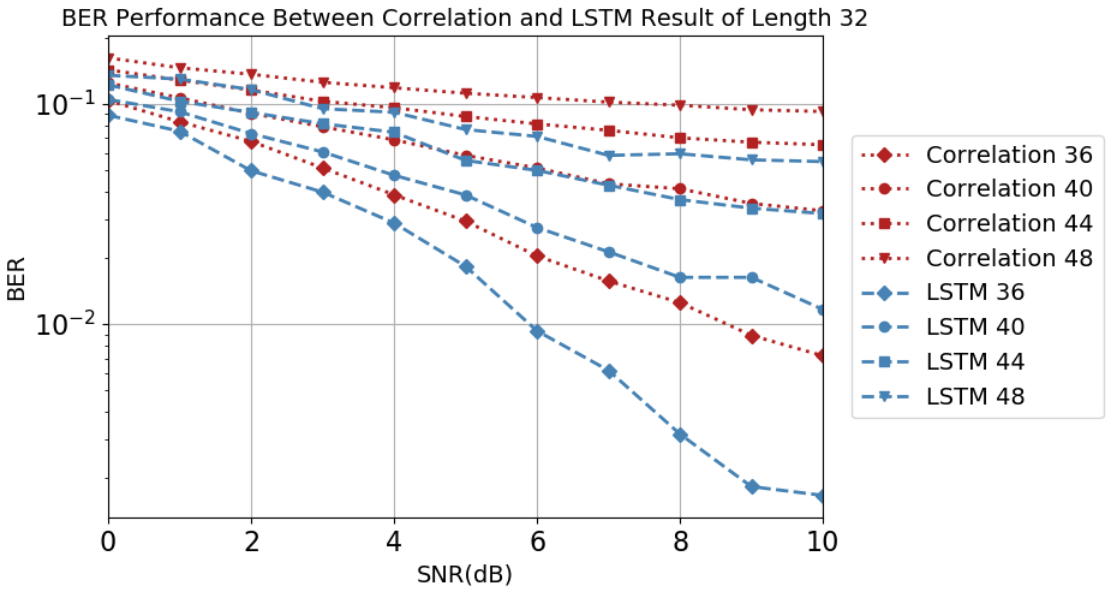

Figure 4.10: BER Performance between Correlation and LSTM Results of Length 32.

4.790% to 81.399%, 9.194% to 74.815%, 12.353% to 52.180% and 6.030% to 35.639% respectively when SNR ranges from 0 dB to 10 dB. The minimum BER improvement is 4.790%, achieved at the number of codes 72 and SNR 0 dB. The maximum BER improvement is 81.399%, achieved at the number of codes 72 and SNR 10 dB.

Table 4.7: Improvement of LSTM Detection for Length 64 (One code)

Improvement	**The number of codes**			
	72	**80**	**88**	**96**
Minimum(%)	4.790	9.194	12.353	6.030
Maximum(%)	81.399	74.815	52.180	35.639

The comparison between correlation and LSTM detection of length 128 is shown in Figure 4.12. The BER improvement by LSTM detection of code length 128 is shown in Table 4.8. For the number of codes 144, 160, 176 and 192, BER improvements increase from 7.882% to 86.962%, 16.651% to 68.409%, 9.552% to 49.117% and 8.805% to 34.677%, respectively, when SNR ranges from 0 dB to 10 dB. The minimum BER improvement is 7.882%, achieved at the number of codes 144 and SNR 0 dB. The max-

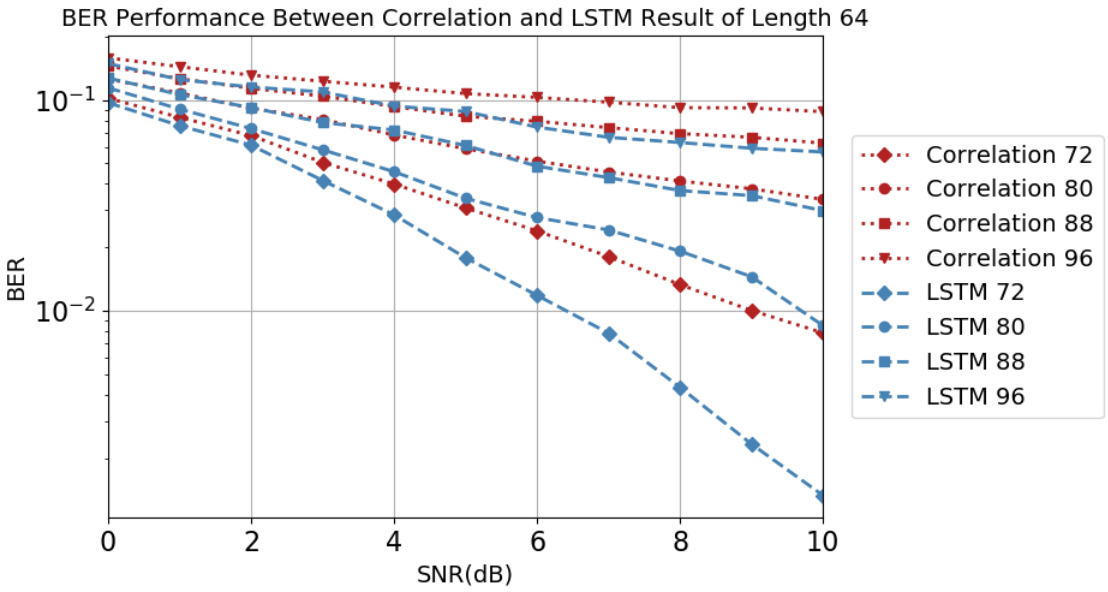

Figure 4.11: BER Performance between Correlation and LSTM Results of Length 64.

imum BER improvement is 86.962%, achieved at the number of codes 144 and SNR 10 dB.

Table 4.8: Improvement of LSTM Detection for Length 128 (One code)

Improvement	**The number of codes**			
	144	**160**	**176**	**192**
Minimum(%)	7.882	16.651	9.552	8.805
Maximum(%)	86.962	68.409	49.117	34.677

From these figures, LSTM detection has better performance than correlation detection. At the same transmission rate, BER performance improves as SNR increases. At the same SNR level, BER performance improves when fewer QOS are used for simultaneous transmission. In most case, LSTM detection has better improvement than three-layer MLP detection.

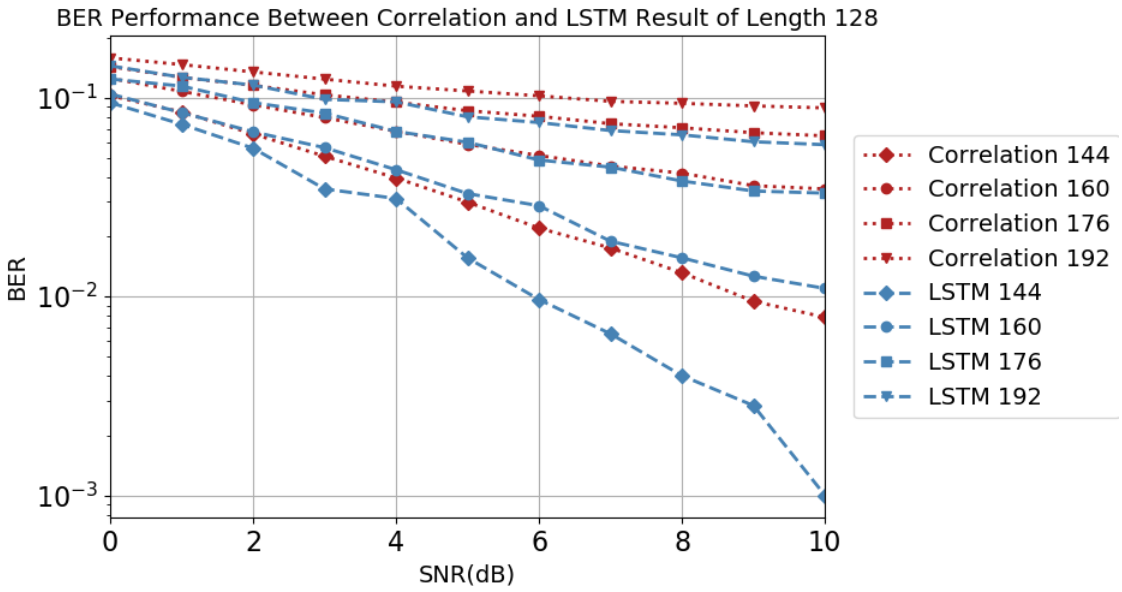

Figure 4.12: BER Performance between Correlation and LSTM Results of Length 128.

4.5 Conclusion

Simulated BER performances and the theoretical BER upper bounds of transmission using QOS with BPSK modulation at faster-than-Nyquist rates are presented in this chapter. Compared with the traditional correlation detection scheme, the three-layer MLP boosts up the performance from 1.236% to 85.772% for QOS lengths of 16, 32, 64 and 128 when SNR ranges from 0 dB to 10 dB. The LSTM detection schemes increases the traditional correlation detection scheme from 4.790% to 87.846% for QOS lengths of 16, 32, 64 and 128 when SNR ranges from 0 dB to 10 dB. It is also observed that LSTM detection has a better performance than three-layer MLP detection in some cases with fewer hidden layer neurons.

Chapter 5: CONCLUSION AND FUTURE WORK

5.1 Conclusion

A new set of quasi-orthogonal sequences that provide a large amount of codes has been presented in this thesis. For length 2^m, compared to Walsh codes which provide 2^m sequences with zero cross-correlation, QOS provides at least 2×2^m codes with small cross-correlation. The non-zero cross-correlation with QOS negatively affects the BER performance of correlation detection.

In order to improve the BER performance, two new detection methods, a three-layer MLP scheme and an LSTM scheme are proposed. These schemes exploit more useful information existed in interference among the simultaneously transmitted sequences which make them less affected by the mutual interference among the codes.

The three-layer MLP detection is a feed-forward artificial neural network that takes a sequential signal bits as its input, and outputs the predicted bits. It implements back-propagation to train the model during which the weights and biases are updated according to the calculated gradients from a loss function. Throughout the training process, the predicted result gets closer to the desired value as weights and biases update. After the despreading multiplier module, the correlation detection generates the final results via accumulation over all chips, whereas the three-layer MLP detection method analyzes further the impact of interference or noise on the desired signals before outputting the final result.

LSTM detection, one type of RNN, faces exploding gradient and vanishing gradient problems. The gradient problem is solved by utilizing input gate, update gate, forget gate and output gate, which stabilizes the whole system. The correlation detection fails to recover the signals mingled with some erroneous bits resulting from interference and noise. After despreading the received signal, the correlation detection generates the final result via accumulation, whereas the LSTM scheme extracts the information between adjacent bits since the model takes every bit as its input in series.

By utilizing the new quasi-orthogonal sequences and the new detection methods

proposed, more more information bits can be transmitted over the same channel with tolerable interference. The two new detection methods have much better performance than correlation detection. With the three-layer MLP detection scheme, the BER performance improves up to 85.772% compared with correlation detection for QOS lengths 16, 32, 64 and 128 when SNR ranges from 0 dB to 10 dB. With the LSTM detection scheme, the BER performance improves up to 87.846% compared with correlation detection for QOS lengths 16, 32, 64 and 128 when SNR ranges from 0 dB to 10 dB.

5.2 Future Work

This thesis expands the study of improving the transmission rate over the Nyquist transmission rate with QOS.

Follow-up work could collect signals from the real world at the receiver, and then test new detection models. In practice the received signal might be affected by multipath fading. A more sophisticated transmission structure is needed to evaluate the performance and compensate for multipath fading. There are several possible changes including enabling transmitters to send signals with different power levels to alleviate multipath fading effects, adopting a compensation module in the receiver to help mitigate fading effects, and adding more layers in the detection model to make it more robust to multipath fading.

Although the three-layer MLP and LSTM detection schemes perform better than correlation detection, they can be further improved. One such improvement is to increase the hidden size or to add the number of hidden layers, to allow it to better exploit the information in interference and noise. The weight and bias updates can also be optimized, such as Adagrad, Adam, RMSprop, Adamax, etc. Other machine learning or deep learning methods can also be applied for such signal detection, such as K-nearest neighbors(KNN), random forest, support vector machine (SVM), etc.

Conventionally, binary sequences are widely used. Walsh codes, Gold codes, small Kasami set, large Kasami set and QOS are all binary codes. Binary codes are easy to implement and to detect because there are only two power levels. M-ary codes, which have shorten lengths compared with binary sequences for the sate transmission rate, could be applied. Nonetheless, it would be more complex to implement the transmitter and receiver structure if M-ary codes are adopted.

Bibliography

[1] B. Aazhang, B.-P. Paris, and G. Orsak, "Neural networks for multi-user detection in CDMA communications," *IEEE Trans. Commun.*, vol. 40, pp. 1212–1222, Jul. 1992.

[2] Y. Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," *IEEE Trans. Neural Networks*, vol. 5, pp. 157–166, 1994.

[3] A. Chandra and S. Chattopadhyay, "Small set orthogonal Kasami codes for CDMA system," in *Proc. IEEE Int. Conf. Computers and Devices for Commun.*, Dec. 2009, Kolkata, India, pp. 1–4.

[4] R. Chinthaginjala and K. Bagadi, "Receiver design using artificial neural network for signal detection in multi carrier–CDMA system," *Int. J. Intelligent Engineering and Systems*, vol. 10, pp. 66–74, 2017.

[5] T. C. Chuah, B. S. Sharif, and O. R. Hinton, "Robust CDMA multiuser detection using a neural network approach," *IEEE Trans. Neural Networks*, vol. 13, pp.1532–1539, 2002.

[6] E. H. Dinan and B. Jabbari, "Spreading codes for direct sequence CDMA and wideband CDMA cellular networks," *IEEE Commun. Mag.*, vol. 36, pp. 48–54, 1998.

[7] R. C. Dixon, *Spread Spectrum Systems,* 2^{nd} *Edition.* Hoboken, New Jersey: John Wiley and Sons, 1984.

[8] J. Fakatselis and M. Belkerdid, "Processing gain for direct sequence spread spectrum communication systems and prism®," *Harris Semiconductor Application Note*, No. AN9633, August 1996.

[9] S. Glisic, Z. Nikolic, N. Milosevic, and A. Pouttu, "Advanced frequency hopping modulation for spread spectrum WLAN," *IEEE J.Select. Areas Commun.*, vol. 18, pp. 16–29, 2000.

[10] R. Gold, "Optimal binary sequences for spread spectrum multiplexing," *IEEE Trans.Information Theory*, vol. 13, pp. 619–621, 1967.

[11] J. Heaton, *Introduction to Neural Networks for Java,* 2^{nd} *Edition.* Heaton Research, Inc., 2008.

[12] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, pp. 1735–1780, 1997.

[13] K. J. Horadam, *Hadamard matrices and their applications.* Princeton, New Jersey: Princeton university press, 2012.

[14] Y. Jabrane, V. P. G. Jiménez, A. Armada, B. A. Said, and A. A. Ouahman, "Reduction of power envelope fluctuations in OFDM signals by using neural networks," *IEEE Commun. Lett.*, vol. 14, pp. 599–601, 2010.

[15] P. P. Kanjilal, *Adaptive prediction and predictive control.* London: Peter Peregrinus Ltd, 1995.

[16] T. Kasami, "Weight distribution formula for some class of cyclic codes," *Coordinated Science Laboratory Report no. R-285*, 1966.

[17] G. I. Kechriotis and E. S. Manolakos, "Hopfield neural network implementation of the optimal CDMA multiuser detector," *IEEE Trans. Neural Networks*, vol. 7, pp. 131–141, 1996.

[18] P. Kinney, "Zigbee technology: Wireless control that simply works," in *Proc. Commun. Design Conf.*, vol. 2, Oct. 2003, San Jose, CA, pp. 1–7.

[19] J. S. Lee, Y. W. Su, and C. C. Shen, "A comparative study of wireless protocols: Bluetooth, UWB, ZigBee, and Wi-Fi," in *Proc.* 33^{rd} *Annual Conf. IEEE Industrial Elect. Soc.,* Mar. 2007, Taipei, Taiwan, PP. 46–51.

[20] Y. Li, M. Chen, Y. Yang, M. T.Zhou, and C. Wang, "Convolutional recurrent neural network-based channel equalization: An experimental study," in *Proc.* 23^{rd} *Asia-Pacific Conf. Commun.*, Dec. 2017, Perth, WA, pp. 1–6.

[21] B. Lojko, "A contribution to the design of a frequency synthesizer for fast frequency-hopped spread-spectrum systems," in *Proc. 2007* 17^{th} *Int. Conf. Radioelektronika*, Apr. 2007, Brno, Czech Republic, pp. 1–5.

[22] C. Luo, J. Ji, Q. Wang, X. Chen, and P. Li, "Channel state information prediction for 5G wireless communications: A deep learning approach," *IEEE Trans. Network Science and Engineering*, pp. 1–11, 2018.

[23] E. McCune, "DSSS vs. FHSS narrowband interference performance issues," *RF Signal Processing Mag.*, pp. 90104, 2000.

[24] M. B. Mollah and M. R. Islam, "Comparative analysis of gold codes with PN codes using correlation property in CDMA technology," in *Proc. 2012 Int. Computer Commun. and Informatics (ICCCI)*, Jan. 2012, Coimbatore, India, pp. 1–6.

[25] M. M. Olama, X. Ma, T. P. Kuruganti, S. F. Smith, and S. M. Djouadi, "Hybrid DS/FFH spread-spectrum: A robust, secure transmission technique for communication in harsh environments," in *Proc. 2011 IEEE MILCOM*, Nov. 2011, Baltimore, MD, pp. 2136–2141.

[26] R. Pickholtz, D. Schilling, and L. Milstein, "Theory of spread-spectrum communications–a tutorial," *IEEE Trans. Commun.*, vol. 30, pp. 855–884, 1982.

[27] J. G. Proakis and D. K. Manolakis, *Digital Signal Processing,* 4^{th} *Edition.* Upper Saddle River, NJ: Prentice-Hall, Inc., 2006.

[28] I. F. Progri, "A MC-CDMA indoor geolocation system," in *Proc.* 16^{th} *Int. Symp. Personal, Indoor and Mobile Radio Commun.*, Sep. 2005, Berlin, Germany, vol. 4, pp.2535–2542.

[29] H. D. Schotten and H. H. Mahram, "Analysis of a CDMA downlink with non-orthogonal spreading sequences for fading channels," in *Proc.* 51^{st} *Veh. Technol. Conf.*, May 2000, Tokyo, Japan, vol. 3, pp. 1782–1786.

[30] Y. Sharma, "Performance study of hybrid spread spectrum techniques," Master's Thesis, University of Tennessee, Knoxville, 2005.

[31] H. Shen and A. P. Suppappola, "Jamming interference suppression and multiuser detection for DSSS systems in time-frequency selective channels," in *Proc.* 4^{th} *IEEE Workshop Signal Processing Advances in Wireless Commun.*, Jun. 2003, Rome, Italy, pp. 294–298.

[32] N. Taşpinar and M. N. Seyman, "Back propagation neural network approach for channel estimation in OFDM system," in *Proc. Wireless Commun., Networking and Inf. Security (WCNIS)*, Jun. 2010, Beijing, China, pp. 265–268.

[33] W. Teich, M. Seidl, and M. Nold, "Multiuser detection for DS-CDMA communication systems based on recurrent neural network structures," in *Proc.* 5^{th} *Int. Symp. Spread Spectrum Techniques and Applications*, Sep. 1998, Sun City, South Africa, pp. 863–867.

[34] D. J. R. V. Nee and A. J. R. M. Coenen, "New fast GPS code-acquisition technique using FFT," *Electronics Lett.*, vol. 27, pp. 158–160, 1991.

[35] A. J. Viterbi, "Wireless digital communication: A view based on three lessons learned," *IEEE Commun. Mag.*, vol. 29, pp. 33–36, 1991.

[36] T. Vlachos and E. Geraniotis, "Performance study of hybrid spread-spectrum random-access communications," *IEEE Trans. Commun.*, vol. 39, pp. 975–985, 1991.

[37] P. Yadav and U. Neelakantan, "Performance analysis of FHSS transceiver model in MATLAB," *Int. Research J. Engineering and Technol.*, vol. 2, pp. 349–351, 2015.

[38] K. Yang, Y. K. Kim, and P. V. Kumar, "Quasi-orthogonal sequences for CDMA systems," *IEEE Trans. Inf.Theory*, vol. 46, pp. 982–993, 2000.

[39] X. Zeng, J. Q. Liu, and L. Hu, "Generalized Kasami sequences: the large set," *IEEE Trans. Inf. Theory*, vol. 53, pp. 2587–2598, 2007.

[40] Z. Zhang, "A detecting algorithm of DSSS signal based on auto-correlation estimation," in *Proc.* 2^{nd} *Advanced Information Technology, Electronic and Automation Control Conf.*, Chongqing, China, Mar. 2017, pp. 137–141.

[41] Z. Zhao, G.Xing, and J. Wu, "The detection methods based on the fourth-order moment slices and the accumulation for DSSS/QPSK signal," in *Proc.* 8^{th} *Int. Conf. Signal Processing*, Beijing, China, Apr. 2006.