CHAPTER 1

INTRODUCTION

1.1 WIRELESS COMMUNICATION AND CODING

“Attention, the Universal! By kingdoms, right wheel!” This prophetic phrase represents the first telegraph message on record. Samuel F.B. Morse sent it over a 16 km line in 1838. Thus a new era was born: the era of electrical communication.

Now, over a century and a half later, communication engineering has advanced to the point that earthbound TV viewers watch astronauts working in space. Telephone, radio, and TV are integral parts of our life. Long-distance circuits span the globe carrying text, data, voice, and images. Computers talk to computers via intercontinental networks. Wireless personal communication devices keep us connected wherever we go. Certainly great strides have been made since the days of Morse.

This thesis describes the research work that has been done in order to improve quality and performance in all aspects of the Viterbi decoder in digital communication systems. In particular, a novel minimized memory and non-polynomial convolutional Viterbi decoder is presented. In addition to this, a low power Survivor Memory Unit (SMU) and Trace back output unit designs that have been developed. These are discussed together with the results from testing them.
It is remarkable that the earliest form of electrical communication, namely telegraphy developed by Samuel Morse in 1837, was a digital communication system. Although Morse was responsible for the development of the first electrical digital communication system, the beginnings of what we now regard as modern digital communications system is from the work of Nyquist in 1928. His studies led him to conclude that for binary data transmission (transmitting 1’s and 0’s) over a noiseless channel of bandwidth $W$ Hertz, the maximum pulse rate is $2W$ pulses per second without any cross symbol interference.

Hartley extended this work in 1928 to non-binary data transmission, while Kolmogorov and Wiener independently in 1939 and 1942, respectively, and solved the problem of optimally estimating a signal in the presence of additive noise. In 1948 Shannon established the mathematical foundation for information transmission and derived fundamental limits for digital communication systems. His work can arguably be considered as the true beginning of the information age.

Another important contribution to the field of digital communication is the work of Kotelnikov in 1947, which provided a coherent analysis and consequently a principle for optimal design of such systems. Wozencraft and Jacobs later extended his work in 1957, leading to the principles used to design the communication systems of today.

The work of Hamming in 1950 on error control coding to combat detrimental effects of channel noise completes the classic contributions to modern digital communication systems.

Today wireless technology is one of the most important systems in the field of communication engineering. While it has been a topic of research since 1960, the past decade has seen a surge of research activities in this area.
This is due to the confluence of many factors. First, there has been an explosive increase in the tether less connectivity driven so far mainly by cellular telephony but expected to be soon eclipsed by wireless data applications. Second, the dramatic progress in VLSI technology has enabled small area low power implementation of various signal processing algorithms and coding and decoding techniques. The channel coder accepts message bits and adds redundancy according to a prescribed rule, thereby producing encoded data at a higher bit rate. The channel decoder exploits the redundancy to decide which message bit was actually transmitted. The combined goal of the channel encoder and decoder is to minimize the effect of channel noise. The Viterbi algorithm is widely used for the elimination of the potential noise in a data stream. Convolutional encoding with Viterbi decoding is a Forward Error Correction (FEC) technique that is particularly suited to a channel in which the transmitted signal is corrupted mainly by Additive White Gaussian Noise (AWGN). Convolutional coding has been used in digital wireless communication systems including mobile communications and deep space communications. An advantage of convolutional coding is that it can be applied to a continuous data stream as well as to blocks of data. IS-95, a wireless digital cellular standard for code division multiple access (CDMA) employs convolutional coding.

The Viterbi decoding algorithm, proposed in 1967 by Viterbi, is a decoding process for convolutional codes in memory-less channels. The algorithm can be applied to a host of problems encountered in the design of digital wireless communication systems. The decoder operates by finding the maximum likelihood decoding sequence. Many researchers have extended his work by finding good convolutional codes, exploring the performance limits of the technique, and varying decoder design parameters to optimize the implementation of the technique in hardware and software (Wicker 1995).
1.1.1 Information Theory

Information theory is the study of how the amount of content in a stream of data may be evaluated, and how fast it may in principle be shipped from place to place by a given communication channel Mackay (2003). The channel may need the data in a specific form and may corrupt it by randomly introducing errors. The subject is thus built on discrete probability theory as its mathematical base. It is somewhat high level in hierarchy, giving bounds and existence proofs without always any explicit means of implementation.

When discussing information theory it is hard not to mention Claude Elwood Shannon (April 30, 1916 - February 24, 2001), an American electrical engineer and mathematician. He is called “the father of information theory”, and was the founder of practical digital circuit design theory. Shannon published two important research articles in the year 1948. These work focused on the problem of how to best encode the information a sender wants to transmit. In this fundamental work he used tools in probability theory, developed by Norbert Wiener, which were in their nascent stages of being applied to communication theory at that time. Shannon developed information entropy as a measure for the uncertainty in a message while essentially inventing what is now known as the dominant form of “information theory”.

One of the most fundamental results of this theory is Shannon’s source coding theorem, which establishes that on average the number of bits needed to represent a random variable \( X \), which is given by the entropy \( H(X) \) and is defined as Shannon (1948):

\[
H(X) = \sum_{x \in X} p(x) \log p(x) \quad (1.1)
\]
where, $X$ and $p(x)$ represent the possible values of $X$ and their probability, respectively. This equation plays a central role in information theory as measurements of information, choice and uncertainty, reflecting the real life fact that an unusual message contains more information than a normal one and thus may be more difficult for us to understand. Therefore, more bits are required in order to describe it more clearly than a normal message. Mackay (2003) summarizes this theorem as: “$N$ independent identically-distributed (i.i.d.) random variables each with entropy $H(X)$ can be compressed into more than $NH(X)$ bits with negligible risk of information loss, as N tends to infinity; but conversely, if they are compressed into fewer than $NH(X)$ bits it is virtually certain that information will be lost”. When applying the source-coding theorem to communications over a noisy channel, Shannon invented the noisy channel-coding theorem. This states that reliable communication is possible over noisy channels provided that the rate of communication is below a certain threshold called the channel capacity. This is also called the Shannon limit or Shannon capacity (Shannon 1948).

Nowadays, Shannon’s limit becomes the ideal objective for most designers of the communication systems. Coding techniques are essential for a communication system to achieve high performance.

### 1.1.2 Coding Theory

Coding theory is more practical compared with other theories in the information theory domain. It is primarily concerned with finding the methods, called codes, for increasing the efficiency and accuracy of data communication over a noisy channel as close to the theoretical limit that Shannon proved as possible. These codes can be mainly subdivided into source coding (Entropy encoding) and channel coding (Error correction coding) (Huffman and Pless 2003). This thesis is only concerned with channel
coding, as it is widely used to improve the reliability of communication on digital channels by detecting and correcting errors (Huffman and Pless 2003).

Although there are many forms of coding schemes, they all have two basic features in common (Clark and Cain 1981). One is the use of redundancy. Coded digital messages always contain extra or redundant symbols. In fact, these “redundant” symbols are not really redundant as they contain the information to accentuate the uniqueness of each message so that the channel disturbance is unlikely to destroy the message by corrupting enough of the symbols in it. The second feature is noise averaging (Clark and Cain 1981). This is achieved by making the redundant symbols depend on a span of several information symbols. This means the redundant symbols not only make the sent message more distinctive but also contain the information of the transmitted message itself. Therefore, each symbol of the message actually contains less transmitted information and thus causes less damage when it is corrupted by noise.

Two kinds of codes are mainly used in modern communication: block codes and convolutional codes (Huffman and Pless 2003; Clark and Cain 1981). This classification is based on the presence or absence of memory in the encoders for these two codes. An encoder for a block code is memoryless as it maps a \( k \)-symbol input sequence into an \( n \)-symbol code words sequence. Therefore, each \( n \)-symbol output only depends upon a specific input \( k \)-symbol block and the encoder has no “memory” of other previous input symbols. For the block codes, there is no correlation between the encoded output code words. In contrast, the output of encoding a convolutional code is determined by the current input and a span \( \nu \) of the preceding input symbols. The encoder for a certain amount of the time span memorizes each input so that it affects not only the current output but also the next \( \nu \) output code words.
Although, codes can also be classified as linear or nonlinear, almost all the coding schemes used in practical applications are linear codes due to their significantly simplified mathematical representations. For this reason, the codes mentioned in this thesis are all linear unless otherwise specified.

Nowadays, convolutional codes become the ideal objective for most designers of the wireless communication systems.

1.2 CONVOLUTIONAL CODES

Convolutional coding technique was first introduced by Elias in (1955). A binary convolutional encoder is normally represented by the values of three parameters: $n, m$ and $k$. The values of $n$ and $m$ indicate that each $n$-bit input yields a $m$-bit output so that the code has a rate of $R$, where

$$R = \frac{n}{m}$$ (1.2)

A convolutional encoder is considered as a finite-state machine. The parameter $k$ is called constraint length which equals the shift register stages of the encoder. The principle of “overlap” is extensively used in convolutional codes. In a rate 1/2 convolutional code, for example, each input is “overlapped” with several previous inputs to produce each pair of encoded symbols. It is not able to divide coded sequence into blocks as each coded symbol pair is interlocked with its neighbors. This provides convolutional codes with great distance and error correction features. The Voyager program uses a convolutional code with a constraint length $k$ of 7 and a rate $R$ of 1/2. Longer constraint lengths produce more powerful codes, but the complexity of decoding operations increases exponentially with constraint length, limiting these more powerful codes to deep space missions. Mars Pathfinder, Mars Exploration Rover and the Cassini probe to Saturn use a $k$ of 15 and a rate of 1/6; this code performs approximately 2 dB better than the simpler $k = 7$ code at an additional cost of 256 times in the decoding complexity.
1.3 DECODING CONVOLUTIONAL CODES

Convolutional coding techniques show significant error protection features over block codes. The decoding techniques for convolutional codes, thus have become the subject of much research interest. There are three basic methods for decoding convolutional codes: maximum-likelihood decoding of Viterbi algorithm (Viterbi 1967), sequential decoding (Wozencraft and Jacobs 1957), and syndrome decoding Clark and Cain (1981).

The Viterbi decoding algorithm was first introduced by Viterbi (1971), Viterbi (1967). By estimating the most-likely received sequence, the Viterbi algorithm achieves its optimum performance in BER. Since the most-likely received sequence is a relative measurement, then to achieve the most optimum performance, all possible code words need to be compared. This requires extensive hardware computation and storage. Much research work is therefore concerned with minimizing the computation complexity of the Viterbi decoding while increasing its performance.

Viterbi decoding has the advantage that it has a fixed decoding time. It is well suited to hardware decoder implementation. But its computational requirements grow exponentially as a function of the constraint length, so it is usually limited in practice to the constraint length (K) of 9 or less. Advanced Wireless Technologies achieved a speed of 2 Mbps for a constraint length of 9 (Fettweis and Meyr 1991). Convolutional coding with Viterbi decoding process has been the predominant Forward Error Correction (FEC) technique used in space communications, particularly in geostationary satellite communication systems, such as VSAT (very small aperture terminal).

Shortly after Elias discovered convolutional codes, Wozencraft and Jacobs (1957) devised a decoding technique which is called “sequential
decoding”. It is a trial-and-error search decoding technique that provides performance that can meet or exceed that of Viterbi decoders. The Fano algorithm (Fano 1963) and stack sequential decoding algorithm (Jelinek 1969) are two major sequential decoding techniques. The main difference between the Viterbi algorithm and a sequential decoding algorithm is that the Viterbi algorithm is a “one direction” decoding process which computes all paths of code words, whereas the sequential decoding only retains a minimum number of code words for a path and needs to “go back” if the path is not correct (Fano 1963; Jelinek 1969). Basically, a sequential decoder performs the search in a sequential manner always operating only on a single path. Each time the decoder moves forward a “trial” decision is made. If an incorrect decision is made, subsequent extensions of the path will be wrong. When the decoder recognizes this situation, it searches back and tries alternate paths until it finally decodes successfully. The drawback thus is the substantial amount of computation required to try alternative paths and recover the correct one. A major research topic, therefore, is to find the optimum parameters to allow quick recognition of an incorrect decision and quick recovery of the correct paths in order to minimize the computation problem (Clark and Cain 1981).

Syndrome decoding sacrifices BER performance in exchange for simplified computation. One widely used syndrome decoding technique is table look-up feedback decoding (Clark and Cain 1981). Instead of estimating the correct data, a syndrome decoder seeks the errors in the received sequence. It calculates the syndrome, which only contains the information of the error patterns in the received data and uses it against the pre-computed syndrome in a look-up table. Thus the corresponding error pattern can be identified and the correct data can then be recovered. A syndrome decoder using a look-up table requires simple hardware implementation. However, the drawback is the BER performance degradation. Another major type of
syndrome decoding technique is called “threshold decoding”. It was discovered by Massey (1963) and can achieve relatively higher BER performance than table look-up feedback decoding while still requiring a simple implementation.

1.4 VITERBI DECODER APPLICATION CONSIDERATIONS

There are many applications using channel encoding. For example, a typical music CD uses a Reed-Solomon code to correct for scratches and dust. In this application the transmission channel is the CD read out system. Mobile phones also use powerful coding techniques to correct the fading and noise of high frequency radio transmission. From data modems telephone transmission to NASA space programs, all of them employ powerful channel coding to combat noise.

The aim of channel coding is to find codes, which transmit quickly, contain many valid code words and can correct or at least detect many errors. These aims are mutually exclusive however, due to the redundancy and channel capacity correlation illustrated in Shannon’s theorems. Therefore, different codes are optimal for different applications. The required properties of a code mainly depend on the probability of errors occur during transmission. Therefore, to examine the source and properties of errors in a target, implementation is essential for a coding application.

In a typical CD, the impairment is mainly dust or scratches and errors are mainly bursty (Koji Okamoto and Yamamoto 2003). Thus codes are used in an interleaved manner. For the channel with high continuous error probability, convolutional codes are widely used. Deep space communications are limited by the thermal noise of the receiver, which is more of a continuous nature than a bursty nature. Concatenated RS/Viterbi-decoded convolutional coding were used on the Mars Pathfinder, Galileo, Mars Exploration Rover
and Cassini missions to provide optimum BER performance. The concatenated RS Convolutional codes are also extensively implemented in standard satellite digital video broadcasting (DVB) systems. Rapid fading troubles mobile phones. The high frequencies used can cause rapid fading of the signal even if the receiver is moved a few inches. Again convolutional codes are used to combat fading although it normally requires shorter constraint lengths.

For correcting continuous errors, block codes can also be used. The narrowband modems are limited by the noise present in the telephone network and is also modeled better as a continuous disturbance. Block codes are used instead of convolutional codes however, as it requires simpler implementations (Clark and Cain 1981).

The most widely used technique for correcting errors in wireless systems is Viterbi decoded convolutional codes. In different forms, it is used in everything from V.3xseries modems, GSM, the voice channels of 3G and satellite DVB. As the market expands, more and more features, such as watching TV, receiving DVB etc., are being put into handheld devices. The leading standard for mobile TV, DVB-H (Digital Video Broadcasting - Handhelds), has emerged from Nokia and been standardized by the European standards group ETSI, as EN 302,304, etc. This requires powerful error-correction codes to be implemented. Viterbi decoder implementations are complex and dissipate a large amount of power. With the proliferation of battery-powered mobile phones, power dissipation, along with speed and size of the memory, is a major concern in the decoder design. The requirement for lower power dissipation and smaller complexity has encouraged researchers to implement various power reduction techniques to decoder designs in order to improve their power efficiency.
1.5 OBJECTIVES OF THIS WORK

The performance of the Viterbi decoders mostly depends on the selection of code rate and constraint length. For a constraint length 3, the encoder/decoder has 4 possible states and for a constraint length 7, it has 64 states. The performance of the Viterbi decoder has been improved by varying the code rate, constraint length, minimum Hamming distance, code gain and the above combinations. The throughput of the Viterbi decoder has also been increased by implementing parallel and pipelining architecture in the add-compare-select (ACS) unit and the trace back (TB) unit (Fettweis and Meyr 1991). The bit error rate (BER) is an important parameter while studying the performance of the Viterbi decoder for the wireless communication systems (Trung 1992). The probability of error has been reduced by increasing Hamming distance of the convolutional coder by using non-polynomial code generator algorithms.

Recently, power dissipation has also become an important concern, especially in battery-powered applications, such as cellular phones, pagers and laptop computers. Power dissipation can be classified into two categories, static power dissipation and dynamic power dissipation. Typically, static power dissipation is due to various leakage currents, while dynamic power dissipation is a result of charging and discharging the parasitic capacitance of transistors and wires. Since the dynamic power dissipation accounts for about 80 to 90 percent of overall power dissipation in CMOS circuits, a number of techniques have been proposed by various researchers in this field recently. These techniques can be applied at different levels of digital design, such as the algorithmic level, the architectural level, the gate level and, the circuit level (Oh and Hwang 1996).

The memory management technique mainly deals with the storage of survivor sequences from which the decoded information sequence is
retrieved. The survivor sequences are usually stored in RAM blocks and then traced back. The former architectures in the Viterbi decoder mainly used two memory management techniques, namely Register Exchange Method and Trace-Back Method (Feygin and Gulak 1993).

1.6 PROBLEM FORMULATION

In this thesis, an attempt has been made to study the performance of various important parameters of Viterbi decoders such as speed, BER, low power and memory management. To decode error-dependently means, the decoder should run in an adaptive manner. There are some existing adaptive decoding methods for convolutional codes. Most of the adaptability is achieved by approximating the calculation of the likelihood measurement. In this work, a new non-polynomial convolutional algorithm is proposed which can detect the sequence and has no error prior to the decoding. Thus the Viterbi decoding operation can reduce the probability of error. This non-polynomial convolutional algorithm has been implemented on a FPGA and demonstrates a significant BER reduction at high noise level. The power consumption of the Viterbi decoder is dominated by the consumptions of the PMU and SMU, which average 36.8% and 62.5% respectively. Therefore, a low power design of a Viterbi decoder should target reducing the power dissipation in the PMU and SMU. In the Viterbi decoder, the survivor path storage unit and trace back output unit consists of different function units. For a low power design we proposed a clock blocking and toggle filtering for survivor path and trace back output units of Viterbi decoder.

In a Viterbi decoder, the Survivor Memory Unit (SMU) is a vital part of the design. So far, classical implementations of the SMU employ the register exchange or the trace back approaches. In the conventional trace back implementation, read-write RAM architecture is generally adopted. However, it suffers from large memory requirement and speed penalty. In this research,
a new approach to implement the trace back algorithm targeted at low memory applications is proposed. This research proposed a new architecture for efficient and minimized memory management in Viterbi decoder based on Zig-Zag algorithm. The implementation result shows that the memory size has been reduced to 56.09%. In order to improve the speed and reduce the latency of the Viterbi decoder, we have proposed new architecture incorporate the ACS, TB and their associated circuits have been operated in parallel and deep pipelined manner to achieve higher throughput rate. Simulations on both software and hardware test the new designs. The results provide a clear view of the improvement of the modifications and enable a novel methodology for significantly reducing complexity of convolutional codes.

1.7 ORGANIZATION OF THESIS

The thesis opens with the basic concepts of Information theory and coding theory. The first chapter deals with introduction of convolutional codes, different decoding techniques of convolutional codes and Viterbi decoder applications. The objectives of thesis and problem statement have also been presented in this chapter.

The chapter 2 deals with the literature survey for Viterbi decoder in four different performance measures. In this chapter, present status of Viterbi decoder in the following four aspects like low Bit Error Rate (BER), low power consumption, less memory utilization and high throughput rate.

In chapter 3 describes the novel non-polynomial based Viterbi decoder. The convolutional code structure is introduced with a detailed discussion of the distance properties and BER performance of convolutional codes. The principles of the Viterbi decoding algorithm and the decoding process are represented using non polynomial approach. Basic concepts related to the Viterbi algorithm, such as hard/soft-decision decoding, etc., are
also introduced in this chapter. The non-polynomial Viterbi decoder has been analyzed with different parameters namely, $d_{\text{free}}$ distance, probability of error, code gain and SNR.

In chapter 4 gives the Viterbi decoder design for low power dissipation. A brief description of low-power design techniques investigated in this chapter. This Chapter proposes a novel clock gating and toggle filtering low-power design for Viterbi decoders. It also discusses the power dissipation results of conventional and proposed Viterbi decoders.

Chapter 5 briefly deals an idea about Zig-Zag algorithm for memory management in Viterbi decoder. The architecture of Viterbi Decoder based on Zig-Zag algorithm with main processing units, Branch metric Unit (BMU), Add-Compare-Select (ACS) unit and Survivor memory unit has been presented along with conventional memory management techniques. All the design units of Viterbi decoder have been carefully implemented in FPGA efficiency. Implementation results of the FPGA prototype are also discussed.

Chapter 6 explores the novel design of high speed deep pipelined Viterbi decoder architecture. A brief description of different high speed techniques investigated in this chapter. To meet the high throughput requirement of the modern wireless communication systems, the fully parallel and deep pipelined architecture are presented in this chapter.

Chapter 7 summarizes the contribution of thesis by the comparison of all above proposed methods with existing one. Further research in this area of interest has also been provided in this chapter.