Chapter 1

Introduction

As science and technology advance, reliable digital data transmission and storage systems are becoming more and more necessary. We use computers and other communication devices to develop these systems. In 1948, Shannon [70] established in his paper that error occurring over noisy communication channels can be corrected if information rate is below the channel capacity. It can be achieved through proper coding of the message. Although his work lacks the method and technique by which it can be achieved, he has only given probabilistic results when it is possible. In 1950, Hamming [36], an American Mathematician, was the first to give a method for constructing a class of codes and coined the first kind of error-correcting codes, known as the Hamming codes, at Bell Laboratory. These codes are used in the computer storage system. Then some important classes of codes like Golay codes and BCH codes were discovered by Golay [35] in 1949, and by Hocquenghem [37] in 1959, Bose and Chaudhuri [48] in 1960 independently. The Reed-Solomon code, another potent error-correcting code developed by Reed and Solomon [73] in 1960, is a subfamily of the BCH codes. In 1970, Hsiao [39] developed a new group of codes known as *odd-weight-column-code*. This code provides an improvement over the Hamming codes. It reduces errors on deployments, improves reliability, and increases the speed of the code. As a result, *odd-weight-column-code* became suitable for applications for IBM and the computer industry worldwide [16, 50].

In 1959, Abramson [2] developed a class of codes that can correct two or fewer consecutive errors. Fire [29] generalised this and developed a class of codes that

can correct any b or less consecutive errors (called burst errors). In 1960, Gilbert [32, 33] brought a thrilling elegance to low-density binary burst-correcting codes. The work of Gilbert led to the improvement of low-density burst detection and correction codes. In [52], Neumann expanded the work of Gilbert strengthening Gilbert's outcomes at the burst correction capability of the codes and presented comparable results on burst detecting capability. Neumann's work was basically an extension of Huffman codes [40]. His results showed that the burst-correcting codes were more efficient than the formerly constructed codes. A study on low-density cyclic burst errors was also presented concurrently by Wyner [81]. A lithographic stage for semiconductors is an example where this type of error occurs. In [54], a category of linear block codes that can correct all burst errors occurring in one-dimensional, two-dimensional, and three-dimensional communication channels was studied. Further, Chen [17] defined a category of binary linear codes for protecting computer memory defects. These codes can mask three or fewer defects and correct multiple random errors.

Cyclic codes are another class of codes that are used for detecting and correcting errors in storage systems. Cyclic code was introduced by Prange [59] in the year 1957. Later, Peterson [56] developed more results and laid the outline for much of the present-day theory. This is a type of linear block code in which the cyclic shift of any codeword results in another codeword. This algebraic property proved to be an excellent choice for designing the codes. An (n, k) cyclic code can be defined by a generator polynomial of degree $n-k$. Cyclic code has the ability to combat single and multiple bursts along with random errors. Cyclic code is one of the most widely used codes developed in recent years. Even Hamming code is a cyclic code. Many well-known codes, like BCH code [48], Reed-Muller code [56], Reed-Solomon code [73], Goppa code [73], fall into the category of cyclic codes.

In coding theory, mathematicians have derived many upper and lower bounds on the parameters of the codes. One goal of the work is to find a suitable one that can provide fewer parity check digits in order to have better information rate. The Hamming bound was developed by Hamming [36] in 1950. It is also known as Sphere Packing Bound. Hamming bound is famous for giving the maximum number

of codewords in a code of length n with minimum distance d. In 1957-58, Varshamov [77] and Sacks [62] gave a famous bound on existence of a linear code, known as Gilbert-Varshamov bound. Campopiano bound was given by Campopiano [15] in 1962. The bound gives us the minimum number of parity-check digits required for a linear code that corrects all single burst errors. A bound (known as Plotkin bound) was derived by Plotkin [58] in 1960. This was a better bound than Hamming on the minimum distance of a linear code and gave a more accurate result when the minimum distance is very close to the length of the code. Huffman and Pless [41] derived the upper bound for nonlinear code. Elias [26] has inhanced Plotkin bound further to derive a bound that gives a higher range of minimum distances. Griesmer bound is another upper bound (developed by Griesmer [34]) on the length of a linear code with respect to the dimension and minimum distance of the code. Unlike other bounds, this bound is valid only for linear codes. After Griesmer derived his bound, Solomon and Stiffler [68], and Belov [10] studied and obtained simplex codes that attained Griesmer bound. In 1964, Singleton [67] derived a simple upper bound on the number of codewords known as Singleton bound. The codes satisfying Singleton bound with equality are called the Maximum Distance Separable (MDS) codes. Studies on the class of MDS codes had been carried out independently by Assmus et al. [8], Forney [30], Kassami and Lin [44], and many others.

Some noteworthy reference books of error-correcting codes are: Peterson [55], Abramson [4], Ash [7], Berlekamp [11], Van Lint [75], Van Lint [76], Vermani [78], Pless [57], Jones et al. [43], Feng et al. [28], Todd [72], Neubauer et al. [51], Klove [45], Bose [13], Niederreiter et al. [53], You [82], Bierbrauer [12], Dougherty [25], Tomlinson et al. [73], Howe et al. [38], Kwong et al. [46], Ball [9], Shang et al. [64], etc. These books are written in chronological order as per the year of publication.

1.1 Terminologies and preliminaries

Any subset C of a vector space V^n of n-tuples over $GF(q)$ is called a `code` over $GF(q)$. If C is a subspace, then it is called a `linear code`. An (n, k) linear code is a subspace of V^n whose dimension is k and n is called the length of the code.

The elements of C are called `codewords` or `codevectors`. The `Hamming distance` between two vectors is the number of components that vary. The `Hamming weight` of a vector is the number of nonzero components. The `minimum weight` of a code is the minimum of all weights of nonzero codewords. The `minimum distance` of a code is the minimum of all distances between any two codewords. In linear code, the minimum weight and the minimum distance coincide. A `generator matrix` of an (n, k) linear code is a matrix of order $k \times n$ whose rows are the basis elements of the code. A `parity check matrix` of an (n, k) linear code is an $(n-k) \times n$ matrix whose rows form a basis for the dual of the code. Let $w \in V^n$ be a vector, and H be the parity check matrix of a linear code. Then wH^T is called the `syndrome` of the vector w.

Standard array: The standard array of a linear code is an array consisting of all the n-tuples in the space of vector space. It is arranged by placing all the codewords of the linear code as the first row with the zero vector at the left. The next row is constructed by placing any non-zero codevector with smaller weight below the zero vector, then adding the new vector to each codeword and placing it just below the codewords. The process continued until all the n-tuples were present in the array. Each row is a `coset` of the code and the first element (smallest weight vector) in each row of the array is known as `coset leader`.

Encoding and decoding: Let G and H be generator and parity check matrices of an (n, k) linear code C. If $x = (x_1, x_2, \ldots, x_k)$ be the message, then x is *encoded* as $v = xG$ by the code C (clearly $v \in C$). Suppose v is sent through communication channel and $w = (w_1, w_2, \ldots, w_n)$ is received. Then decoding is done as follows:

Case i: If $wH^T = 0$, then there is no error in transmission. This means w is *decoded* as $v = w$ and accordingly we can get the message x.

Case ii: If $wH^T \neq 0$, then there is error in transmission. In this case, we locate the syndrome wH^T in standard array with corresponding coset leader e and w is *decoded* as $v = w - e$ and accordingly we can get the message x.

Weight distribution: The (Hamming) `weight distribution` of a code provides insight into the error detecting/correcting properties of a code. This is very helpful

in determining the number of errors that are detectable/correctable by the code. Usually, the weights of the errors are not uniformly distributed. Smallest weight error vector has more chance than the higher one. If the weight distribution of the code or the error pattern is known, probability of incorrect output can be estimated. If $W(i)$ is the number of codewords of weight i in a code C, the list $W(i)$ for $0 \leq i \leq n$ is called the weight distribution or weight spectrum of C. The weight distribution of a code was first given by Hamming [36] in the 1950s. In 1963, MacWilliam [49] gave the fascinating relation of weight distributions between a linear code and its dual code. For more details on weight distribution, one can see the book: Huffman and Pless [41], or Peterson and Weldon [56]. Some recent works in this topic can be found in [20, 21, 60, 65].

Probability of decoding error: The decoder will produce incorrect output if and only if the syndrome produced corresponds to a correctable error. There is always a possibility of decoding error that exceed the code's guaranteed detection/correction capability. To know the rate at which decoding error occurs, we need to study `Probability of decoding error` of the code. Gallager [31] first introduced the probability of decoding error in 1962. He derived the probability of decoding error over a binary symmetric channel with the maximum likelihood of decoding. For more details on probability of decoding error, one may refer Peterson and Weldon [56], or Sweeney [69]. Some recent works on this direction can be found in [5, 61, 71].

1.2 Error Patterns

1.2.1 Burst errors

Burst errors are one of the most common error patterns present in communication channels. Abramson [2, 3] studied this error pattern in 1959. He developed codes correcting single and double adjacent errors. Fire [29] generalised them in the same year and named them as *open-loop burst errors* (or simply *burst errors*). They are defined as follows.

Definition 1.1. *A burst of length b is a vector whose only nonzero components are among some b consecutive components, the first and the last of which are nonzero.*

Some examples of bursts of length 3 in a vector of length 9 over $GF(3)$ are 101000000, 001010000, 000001210, 000000201.

We call those bursts as end-around (EA) bursts (refer Wainberg et al. [80]) when the b consecutive components are considered cyclically. A proper definition of an EA-burst is given by Villalba et al. [79]. They called such bursts as all-around (AA) bursts.

Definition 1.2. *An n-tuple* $v = (v_0, v_1, \ldots, v_{n-1})$ *is called an all-around (AA) burst of length b (where* $b < n/2$*) if given the consecutive (cyclically) co-ordinates* $v_i, v_{i+1}, \ldots, v_{i+b-1}$*, the co-ordinates* v_i *and* v_{i+b-1} *should be nonzero while* $v_j = 0$ *for* $j \notin \{i+1, i+2, \ldots, i+b-1\}$ *and* $i+b-1 < i$ *when taken modulo n.*

For example, the vectors 1100000001 and 1100000010 are AA-bursts of length 3 and 4 respectively.

1.2.2 Cyclic burst errors

Cyclic bursts are known to be considered first by Abramson [2, 3] in 1960. He proved the existence of an optimum one-burst-correcting code that is Hamming code, a two-burst-correcting code, and a three-burst correcting code. Later, in 1986, Abdel et al. [1] proved the existence of an infinite number of optimum burst correcting cyclic codes. Cyclic burst basically consists of burst and AA-burst.

Definition 1.3. *A cyclic burst of length b is a vector whose only nonzero cyclic bursts are confined to some b consecutive components, the first and the last of which are nonzero.*

For example, $10100000, 00000121, 10000021, 12000001$ are some cyclic bursts of length 3 in a vector of length 8 over $GF(3)$.

1.2.3 CT-burst errors

Alexander, Gryb and Nast [6] studied burst errors in a binary bit of length N that begins with an error bit and ends with $N-1$ bits, whether or not they are error bits. This type of error occurs on telephone lines. In 1965, Chien and Tang [18] gave proper definition of such bursts, which are known as CT-bursts.

Definition 1.4. *A CT-burst of length b is a vector whose only nonzero components are confined to some b consecutive positions, first of which is nonzero.*

For example, $100000000, 000100200, 000102000$ are some CT-bursts of length 4 in a vector of length 9.

1.2.4 Periodic random/periodical burst errors

Errors due to noisy channels in the early days are classified into two broad categories: random and burst errors. In 1994, it has been observed by Lange [47] that there are channels like power lines, data channels in close distance to electronically controlled power supply units or inverters, the car electric, compact discs, and CD-ROM where groups of consecutive errors (random or burst errors) repeat periodically. They are called as `periodic random errors` or `periodical burst errors` accordingly. Such types of errors are also found in lithographic stages for semiconductor fabrication, as observed by Schmitz et al. [63].

Definition 1.5. *An s-periodic random error of length b is an n-tuple whose nonzero components are confined to distinct sets of b consecutive positions, the sets are separated by s zeros and the b components can be filled by any field element.*

For example, 4-periodic random errors of length 3 in a vector of length 16 are 001 0000 110 0000 10, 0 001 0000 110 0000 1, 00 001 0000 110 0000, etc.

If $b = 1$, such errors are simply called *s-periodic errors*. s-periodic error detecting and correcting linear codes and the Hamming weight distribution of the error pattern are studied in [20, 22, 74].

Definition 1.6. *An s-periodical burst error of length b is an n-tuple whose nonzero components are confined to distinct sets of b consecutive positions such that the sets are separated by s positions and first component of each set is nonzero.*

For example, 4-periodical burst errors of length 3 in a vector of length 16 are 100 0000 110 0000 10, 0 101 0000 110 0000 1, 00 101 0000 110 0000, etc.

Note that the sets which are separated by s positions in an s-periodical burst error of length b are nothing but CT-burst errors of length b.

1.2.5 Low-density periodic random/periodical burst errors

Wyner [81] in 1961 introduced the concept of low-density burst errors for low intensity bursts. In such a burst, very few components within the burst are normally disturbed. Motivated by this, for low intensity periodical disturbances, we have considered `low-density periodic random errors` and `low-density periodical burst errors`. They are defined as follows:

Definition 1.7. *A low-density s-periodic random error of length b with weight w ($w \leq b$) is an n-tuple whose non-zero components are confined to distinct sets of b consecutive positions such that the sets are separated by s positions and the number of nonzero components within the b consecutive positions can be at most w.*

For example, 5-periodic random errors of length 4 with weight 2 over $GF(2)$ in a vector of length 21 are 0101 00000 0101 00000 100, 0 0110 00000 1001 00000 10, 00 0101 00000 0101 00000 1, etc.

Definition 1.8. *A low-density s-periodical burst error of length b with weight w ($w \leq b$) is an n-tuple whose non-zero components are confined to distinct sets of b consecutive positions such that the sets are separated by s positions, the first position of each set is nonzero, and the number of nonzero components within the b consecutive positions can be at most w.*

For example, 5-periodical burst errors of length 4 with weight 3 over $GF(2)$ in a vector of length 21 are 1010 00000 1101 00000 100, 0 1010 00000 1101 00000 10,

00 1010 00000 1101 00000 1, etc.

1.2.6 Burst -weight and -distance

Hamming distance [56] has been widely studied and found applications in many coding problems among the various studied standard distances in coding theory. The choice of a distance for a communication channel is vital. One type of distance is better suited than another because different channels produce different types of error patterns, and error patterns and distances are interlinked. In this regard, Wainberg and Wolf [80] in 1972 introduced the concept of burst-b weight and distance in order to correct multiple bursts and multiple erasures. Then Villalba et al. [79] extended the study on this distance and derived extended Reiger-Singleton bound for a linear code having minimum burst-b distance. They also presented a class of Maximum Distance Separable (MDS) codes with respect to burst-b distance. To define burst-b distance, burst-b weight is required to be defined first.

Definition 1.9. *The burst-b weight of a vector v is the minimum number of CT-bursts of length b that cover (cyclically) the nonzero coordinates of a vector v. We denote the burst-b weight of a vector v as $w_b(v)$.*

For instance, $w_2(100101101) = 3$, $w_4(100101101) = 2$.

Definition 1.10. *The burst-b distance d_b between vectors u and v is defined by $d_b(u, v) = w_b(u - v)$.*

Note that the above two definitions were originally defined for binary case, but valid for q-ary case also. Also, burst-b weight and distance coincide with the Hamming weight and distance in the case of $b = 1$. An (n, k) linear code equipped with the minimum burst-b distance d_b is written as (n, k, d_b).

1.3 Some previous results

Some of the important results on which this thesis is based are mentioned below.

Result 1.11. ***(Hamming bound [36]):*** *The maximum number of codewords in a q-ary (n, k) block code with minimum distance at least d is given by*

$$\frac{q^n}{\sum_{i=1}^{\lfloor \frac{d-1}{2} \rfloor} \binom{n}{i} (q-1)^i}.$$

Result 1.12. ***(Plotkin's bound [58]):*** *The minimum weight of a codeword in a q-ary (n, k) linear code is at most as large as the average weight $nq^{k-1}(q-1)/(q^k-1)$.*

Result 1.13. ***(Varshamov-Gilbert-Sacks bound [56]):*** *The sufficient condition for the existence of a q-ary (n, k) linear code with minimum distance at least $d \geq 3$ is as follows:*

$$q^{n-k} > \sum_{i=0}^{2\lfloor \frac{d-1}{2} \rfloor - 1} \binom{n}{i} (q-1)^i.$$

Result 1.14. ***(Fire bound [29]):*** *The necessary number of parity-check symbols in a q-ary (n, k) linear code that can correct all bursts of length b or less is at least*

$$b - 1 + \log_q \big[(q-1)(n-b+1) + 1\big].$$

Result 1.15. ***(Campopiano bound[15]):*** *There shall always exist a q-ary (n, k) linear code that corrects all bursts of length b or less ($b < n/2$) provided that*

$$n - k > 2(b-1) + \log_q \Big[(q-1)(n-2b+1) + 1\Big].$$

Result 1.16. ***(Singleton bound [67]):*** *For a q-ary (n, k) linear code, the minimum distance d of the code is at most $n - k + 1$, i.e., $d \leq n - k + 1$.*

Result 1.17. *[69] The decoding error probability of a code of length n correcting up to t errors on a binary symmetric channel with cross-over probability p is given by*

$$P_{de} = 1 - \sum_{i=0}^{t} W(i) p^i (1-p)^{n-i},$$

where $W(i)$ is number of codewords of weight i.

Result 1.18. *[80] An (n, k, d_b) code can correct up to $(d_b - 1)/2$ bursts of length up to b each.*

1.4 Plan of the thesis

This thesis studies mainly on the existence of linear codes that correct periodic random/periodical burst errors with or without weight constraint, along with weight distribution of the error patterns and error decoding probability of the codes. We also present a study on the minimum burst-b distance of linear codes with periodical burst-detection and -correction capabilities of codes. In this thesis, weight (distance) is taken in the Hamming sense, except the last chapter. The contents of this thesis are divided into six chapters.

Chapter 1

This chapter is essentially the introductory chapter, discussing the development of the concept of error-correcting codes. Pre-requisites, basic definitions, and some previous results are discussed in this chapter.

Chapter 2

In this chapter, we study necessary and sufficient conditions for the existence of linear codes correcting periodic random errors. We also obtain the Hamming weight distribution of such error pattern and derive an upper bound on the total Hamming weight of all codewords of such error correcting codes. Examples are also provided.

Chapter 3

Chapter 3 extends the study of Chapter 2 and derives the conditions for the existence of linear codes correcting periodic random errors but with (Hamming) weight constraint (i.e., low-density periodic random errors). For this, we first find the weight distribution of low-density periodic random errors and then derive necessary and sufficient conditions for the codes which are followed by examples. Then we present Plotkin's type of bound for the set of all such errors over q-ary n-tuples. Finally, the probability of decoding error of the codes over a memoryless binary symmetric

channel is derived.

Chapter 4

In this chapter, we study linear codes correcting periodical burst errors and derive necessary and sufficient conditions for the existence of such codes. This chapter also gives the (Hamming) weight distribution of periodical burst errors. This is followed by Plotkin's type of bound for the set of periodical burst errors. We conclude this chapter by providing the probability of decoding error of periodical burst error-correcting codes.

Chapter 5

Chapter 5 is also an extension of Chapter 4 where we give necessary and sufficient conditions for the existence of linear codes correcting periodical burst errors having (Hamming) weight constraint (i.e., low-density periodical burst errors). This chapter also gives the weight distribution for low-density periodical burst errors, Plotkin's type of bound for the set of low-density periodical burst errors, and error decoding probability of the codes. In addition to these, we present weight distribution and Plotkins type bound for the set of errors (beyond the correctable periodical bursts) that are detected by the low-density periodical burst correcting codes.

Chapter 6

In Chapter 6, we study bounds on the minimum burst-b distance of any linear code. The periodical burst -detection and -correction capabilities of linear codes with burst-b distance are investigated. Then the same investigation is done for the MDS code C_b, given by Villalba et al. (2016), and its dual code $C_b^{\perp}$. Finally, we provide a decoding procedure for the code C_b in the case of periodical burst errors.

After the last chapter, a separate section titled "Scope for Further Research" is added for future direction of work. At the end, we provide the bibliography of the thesis.