CHAPTER 1
INTRODUCTION & RELEVANT LITERATURE SURVEY
1.1 CODING THEORY: A Brief Introduction & Literature Survey

The last few decades have seen a dramatic change in the way digital data communication and storage systems are achieved around the world. Digital information, its processing, storage and communication have become dominant features of modern day society. Digital information has been proved an asset not only to the educationist but also become a widespread and economical means for facilitating domestic, commercial as well as public service communications. There has been a tremendous dependence of human life on computing and communication devices in the form of cellular communication, satellite communication, CD players, Internet, Modems, Routers, Telephone lines etc.

Though, modern communication devices are highly reliable yet communication in adverse channel could have distortions in the form of dynamic noise, jamming multi access interference, server fading, and dropped calls. In such situations, these devices are proved to be unreliable.

The goal of coding theory is to improve the reliability of communication by devising methods that enable the receiver to decide whether there have been errors during the transmission (error detection), and if there are, to possibly recover the original message (error correction). The discipline of error correcting/detecting codes deals with techniques for encoding to-be-transmitted data and decoding the received data.

The modern coding theory has its roots in pioneering work of Shannon [100] in the field of ‘Communication Theory’. Shannon’s celebrated paper ‘A Mathematical Theory of Communication’ re-established the foundation of ‘engineering of communication’. A communication system, in general, can be represented by the block diagram as shown below:

![Figure 1.1 A Simple Communication System](Image)
In his paper, Shannon gave simple mathematical characterizations and functions of various blocks in the diagram above. Shannon thus, for the first time, formulated the problems of communication in mathematical terms. He masterly succeeded in developing a highly sophisticated and rigorous mathematical theory to lay foundation of a reliable communication system, through unreliable channels. This lead to intensive research in theory and applications of Communication Theory emerging in what is currently called Information Technology, which has profoundly changed the society world over.

Following are the two main reasons of this:

I. The emergence and advancement in communication technology.

II. The use of elegant mathematics in communication theory.

Coding has acquired an important place with its increasing demands due to its applications in computers, space communications, distant control systems, electronic devices and also in medical science.

Studies in coding theory branched out in mainly following two directions:

I. For noiseless channels.

II. For noisy channels.

The study for noiseless channels was made with variable length codes. One of the most significant results in this direction was Kraft’s Inequality [4]. Later on, McMillan [87] extended this result for some uniquely decodable codes. In 1962, Huffman [67] constructed a method for optimal instantaneous codes that virtually answered all questions in search for the most efficient codes for noiseless channels.

Another direction in which Coding Theory developed was the area of error correcting codes.

Highly significant papers of Hamming [60] and Golay [52] explored this area for work initially. The work of Hamming was concerned to the code construction and bounds. It was Hamming [60] who introduced the basic concepts of linear parity check, parity check matrix and a metric. While, the work of Golay [52] was related to almost all perfect codes.
An important class of codes was constructed by Reed [95] and Muller [90]. They also invented the notion of threshold coding. Slepian [108, 109], Reed [95] and Muller [90] are known to set the foundation of algebraic coding theory. Reed [95] and Muller [90] made the use of finite rings, algebra and field in code construction. Slepian [108, 109] introduced the concept of group codes. The important concept of asymptotic construction for block codes was introduced by Elias [40]. Convolution codes were also invented by Elias [40]. Also Elias bound is one of the tightest known asymptotic bound on minimum distance.

Hamming [60] was the first to establish the norms regarding the minimum distance for the performance of a code. He gave well-known Hamming bound on minimum distance of a code. Later on, Plotkin [93], Varshamov [113] and Gilbert [51] also gave some bounds on minimum distance of a code.

Then came major breakthrough in the construction of error correcting codes due to the work of Reed and Solomon [96] and Bose and Choudhary [16, 17] and Hocquenghen [66] independently. They gave a method of constructing binary codes for correcting multiple random errors. These codes are known as BCH codes. Peterson [91] showed these codes to be cyclic. BCH codes were also generalized by Gorenstein and Zierler [57]. They also gave a decoding algorithm for these codes. Decoding algorithms for these codes were also given by Peterson [91], Berlakmap [10] and Massey [85].

Including BCH codes, there is another technique for code construction that is based on the roots of polynomials which have coefficients in a Galois Field. Mattson and Solomon [86] introduced a new technique that is known as ‘associated polynomial approach to coding’. Following this technique, Kasami, Lin and Peterson [74] obtained one of the general classes of codes. These codes include BCH codes as a subclass. Thereafter, the focus of research in coding deviated to design a family of codes that have codes with larger code length and much better performances.

On the one hand Shannon’s theory was an inspiration to coding theory but on the other hand it was a source of frustration also because nobody was able to construct a family of asymptotically good codes. Then Elias [40] came with partial success in obtaining asymptotically good codes by using product codes. The notion of concatenated codes was introduced by Forney [48]. He gave a class of codes which are asymptotically error free in a stronger way than the Elias codes. The first major success in this direction was
achieved by Justesen [72]. He concatenated short random codes with long Reed-Solomon codes. A large class of algebraic codes was constructed by Goppa [54, 55, 56]. For large value of $n$, these codes approach Extended Varshamov-Gilbert bound that was given by Sharma and Dass [102]. Later on, these Goppa codes were generalized by Helgert [61, 62, 63, 64] and Mandelbaum [83]. The codes given by Helgert are known as ‘Alternate codes’ [62, 63]. These codes include many important classes of some well-known codes viz. Chien–Choy generalized BCH codes [23], generalized Srivastava codes due to Helgert [64], BCH codes, Srivastava codes [10] etc. Helgert [62] also gave a decoding method for alternate codes.

Nowadays coding is used in a broad range of communication systems and also in storage mediums on a large scale. Survey articles by Van Lint [81], Wolf [115], Kautz and Levit [75], Dass and Das [28], Berlakmap [13], Assmus and Mattson [5] and others show the importance and the development of the subject in various fields of practical approaches.

There are also some books that are good references for coding theory and its development. Some of those are mentioned below:

The books by Peterson and Weldon [92], Abramson [2], Berlakmap [11, 12], Blake [14], Blake and Mullin [15], Baylis [8], Clark, Jr. and Cain [25], Justesen and Hoholdt [71], Jones and Jones [69], Mac Williams and Sloane [82], Mann [84], Morelos-Zaragoza [89], Lin [79] and Lin and Costello [80].

A major part of coding theory is devoted to the study of linear block codes for correcting and detecting random errors. The other important area, in which the growth of coding occurred, is burst error correcting and detecting codes.

In most of the communication systems, it is seen that the occurrence of the errors is more adjacent rather than in a random manner. This led to the study of burst error correcting and detecting codes. Abramson [1] started this era by obtaining results for single error and double-adjacent error correcting codes. In fact, Abramson [1], Regier [97] and Melas [88] set the foundation of burst-error-correcting codes. Generalizing Abramson’s work, the most successful burst-error-correcting codes were first designed by Fire [43]. He depicted the more general concept of clustered errors, which are popularly known as ‘burst errors’. Fire [43] considered two types of burst errors; one is
‘open-loop burst error’ or simply burst error and other is ‘closed-loop burst error’. A survey of work in this area was done by Forney [49].

In communication over a memory less channel, random errors are added to the message while in a certain memory channel; the errors are in the form of bursts. So there is a need to design codes that are capable of correcting and detecting burst errors.

Although the work in the area of burst errors was initiated by Abramson [1] but Burton [20], Regier [97], Campopiano [21], Elspas and Short [42], Gross [58], Fire [43] and Melas [88] gave great contribution in this area. Elspas [41] generalized the work of Abramson [1] over a general Galois field $GF(q)$. The study of burst-error-correcting codes has been nicely treated by Peterson [91] and Peterson and Weldon [92].

In many burst-noise channels, bursts donot occur singly but in the form of bursts of bursts or multiple bursts. Remarkable studies in this direction were made by Stone [110], Wolf [116] and Bridwell and Wolf [19]. Among various types of burst errors, Chien and Tang [24] considered a different type of burst which in literature is known as CT burst. Later, Dass [26] modified the definition of CT burst and termed that modified burst as ‘burst of length $b$ (fixed)’. These types of burst errors occur in the channels that do not produce burst error near the end of the codeword.

In certain systems like lightening and other short term intermittent disturbances which introduce burst errors usually operate in such a way that over a given length, some digits are received correctly while others are corrupted. Such situations demand the development of codes that can detect and correct a burst with weight lying between a given range. Such types of bursts are known as ‘moderate-density open-loop burst errors’ [68]. The study in this direction was made by Sharma and Gupta [104], Dass and Sobha [27] and Jain [68]. Further, Jain [68] studied cyclic codes detecting moderate-density open loop burst error and CT moderate density open-loop burst error.

It has been noticed that when there are large number of messages, the code words are quite long and within a given length bursts repeat themselves. This led to the study of repeated burst error detecting and correcting codes. The concept of repeated bursts was introduced by Dass and Verma [33] and studied linear codes correcting repeated burst errors. Beraradi, Dass and Verma [9] obtained lower and upper bounds on the number of parity-check digits required for a linear code that is capable of detecting 2-repeated
burst errors and also capable of detecting and simultaneously correcting such errors. In general, Dass and Verma [34] obtained results regarding the number of parity-check digits for detecting and simultaneously correcting \( m \)-repeated burst errors. Codes for detecting and simultaneously correcting repeated low-density burst errors of length \( b \) or less with weight \( w \) or less were also studied by Dass and Verma [35]. Later on, Dass and Verma [36] studied codes that are capable of correcting 2-repeated low-density bursts of length \( b \) or less with weight \( w \) or less.

Also there is another kind of repeated burst errors, i.e., repeated burst errors of length \( b \) (fixed) defined by Dass, Garg and Zannetti [31]. Linear codes capable of detecting and simultaneously correcting 2-repeated bursts of length \( b \) (fixed) were studied by Dass and Garg [29]. Dass, Garg and Zannetti [32] obtained lower and upper bound on the number of parity-check digits required for a linear code that is capable of correcting repeated burst errors of length \( b \) (fixed). An upper bound on the number of parity-check digits for a code to detect \( m \)-repeated burst of length \( b \) (fixed) was also derived by them. Dass and Garg [30] also obtained bounds on the number of parity-check digits for a linear code that can detect and simultaneously correct repeated low-density burst of length \( b \) (fixed) with weight \( w \) or less.

In some practical channels like semiconductors and super computers [3], there occurs a different kind of burst error that is called ‘solid burst error’. Solid bursts are extensively studied by many authors [99,103,106]. Recently, the systematic study of linear codes detecting and correcting solid burst errors was done by Das [37]. He also generalized the area of study of multiple solid bursts [106] by defining the concept of repeated solid burst errors [38, 39].

The investigations in this thesis are over various types of repeated burst errors and multiple burst errors with reference to Hamming weight. This chapter is intended to be an introduction to the investigations reported in Chapters II to VII. We shall therefore confine ourselves to only relevant and connected aspects.

1.2 COMMUNICATION SYSTEM

As we have mentioned earlier, a communication system, as shown by a block diagram in figure 1.1, have mainly following five parts:
**Source:** A source is one that produces a message or sequence of messages to be communicated to the receiver. The source output might represent, for example, the output of a set of sensors in a space probe, a sensory input to a biological organism, or a target in a radar system.

**Encoder:** An encoder is one that represents any processing of the source output prior to transmission. The processing might include, for example, any combination of modulation, data reduction and insertion of redundancy to combat the channel noise.

**Channel:** A channel is the medium for transmitting signals from transmitter to receiver. It may be a telephone line, a high frequency radio link, a space communication link, storage medium, or a biological organism. A typical storage medium may be semiconductor memories, magnetic tapes, magnetic disks, CDs, DVDs.

The channel is usually subject to various types of noise disturbances, which on a telephone line, for example, might take the form of a time-varying frequency response, crosstalk from other lines, thermal noise, impulse noise, impulsive switching noise. A channel subject to noise is called ‘noisy channel’.

“The theory of error correcting codes corrects errors due to noise.”

**Decoder:** A decoder is one that represents the processing of channel output with the objective of producing at the destination an accepted replica of the output.

**Destination:** A destination or receiver is the person or object for whom the message is intended. For example, a computer or a communication system for which message is intended is a receiver.

### 1.3 LINEAR BLOCK CODES

There can be various ways of coding a source. In algebraic coding theory, for logical implementation, the code characters are taken as the elements of a finite field or that of a finite ring. In our study, we are concerned only to Hamming metric. Therefore, in this thesis a linear code will be considered as a subspace of all $n$-tuples over Galois field $GF(q)$.

Code words can be formed from code characters. In the study of error-correcting codes, in most of the situations the words are taken as blocks of same length. We will be restricting to this framework of studies only.
The most commonly used codes are the linear codes, which are linear spaces over some particular finite field. Specifically, a linear code can be defined as follows:

**Linear Code:** A linear code of length $n$ is the subspace of the space of all $n$-tuples over a finite field $GF(q)$.

If we mention a linear code as an $(n,k)$ code then $k$ refers to the dimension of the code. In an $(n,k)$ linear code, there are exactly $q^k$ code words.

A linear code can be best described in terms of its generator matrix and parity-check matrix. These are defined as follows:

**Generator Matrix:** A matrix $G$ is said to be the generator matrix of a linear code if the row space of $G$ is the given code.

**Parity-check Matrix:** A matrix $H$ is said to be the parity-check matrix of a linear code if the code is the null space of the matrix $H$.

If $H$ is a parity-check matrix of a linear code $C$, then an $n$-tuple $\vec{u}$ is a code word if and only if $\vec{u}$ is orthogonal to every row of $H$, i.e.,

$$\vec{u}H^T = 0 \iff \vec{u} \in C.$$ 

(1.3.1)

**Syndrome of an $n$-tuple:** For an $n$-tuple $\vec{u}$, $\vec{u}H^T$ is called the syndrome of the $n$-tuple $\vec{u}$.

In an $(n,k)$ linear code, $k$ represents the information positions and it can have arbitrarily assigned values while the remaining $(n-k)$ positions are called parity check positions and are determined by a set of rules or equations.

### 1.4 WEIGHT AND DISTANCE

In the study of error-correcting codes weight and distance play an important role. It was clearly pointed out by Berlakmap [11] that the notion of weight distinguishes the theory of linear codes from classical linear algebra. In coding theory, there are various weights and distances are used but one and almost universally used weight is Hamming weight and so the Hamming distance [60]. Some weights and distances are given below, regarding which the studies have been made in coding theory:

i. Hamming weight and distance[60],
ii. Lee weight and distance [78],

iii. Sharma-Kaushik weight and distance [105].

In this thesis we are concerned to Hamming weight and Hamming distance only.

**Hamming Weight** [60]: The Hamming weight of an $n$-tuple is the number of its non-zero entries.

More specifically, the Hamming weight of a vector $X = (X_1, X_2, \ldots, X_n)$ denoted by $W(X)$ is the number of non-zero components of $X$. Each component $X_i$ is an element of $GF(q)$. The vector $X$ is also called an $n$-tuple or a code word.

**Hamming Distance** [60]: The Hamming distance between two $n$-tuples is the number of positions in which they differ.

More specifically, the Hamming distance between two vectors $X$ and $Y$, denoted by $d(X, Y)$ is the Hamming weight of the vector $(X - Y)$. It is equal to the number of positions in which they differ, i.e.,

$$d(X, Y) = W(X - Y) = W(Y - X) = \text{the number of positions in which } X \text{ and } Y \text{ differ}.$$

Imposing the ideas of weight and distance over an $n$-vector, we also associate it with a code.

**Minimum Weight of a Code**: The minimum weight of a code is defined as the number of non-zero weights of the words in the code.

**Minimum Distance of a Code**: The minimum distance of a code is defined as the minimum of the distances between all pairs of distinct code words.

For a linear code, minimum distance and minimum weight coincide.

**1.5 ERROR PATTERN AND COSET DECOMPOSITION**

When messages are transmitted over a noisy channel, then an error vector is added to them and therefore, they are not received correctly.

**Error Vector**: If the $n$-vector $\vec{v} = (b_1, b_2, \ldots, b_n)$ is received vector and $\vec{u} = (a_1, a_2, \ldots, a_n)$ is the transmitted vector then the difference $\vec{e} = \vec{u} - \vec{v} = (b_1 - a_1, b_2 - a_2, \ldots, b_n - a_n)$ is called the error vector or an error pattern.
Since

$$\overline{uH} = \overline{0}$$

$$\Rightarrow \overline{vH} = \overline{eH}.$$

Thus the syndrome of a received vector is equal to the syndrome of the error vector added to the transmitted codeword.

**Coset Decomposition:** A coset decomposition of the space of $n$-tuples over $GF(q)$ with respect to its subspace $C$ (the code) is very useful in formulating many important results in coding theory. It is a very simple proposition in algebra that decomposition of the space of $n$-tuple over $GF(q)$ into cosets is complete and unique in the sense that every $n$-tuples over $GF(q)$ is in one and only one coset. All $n$-tuples in a coset over $GF(q)$ have syndrome.

**Weight of a Coset:** The weight of a coset is defined as the minimum of the weights of $n$-tuples over $GF(q)$ in the coset.

If we consider an $(n,k)$ linear code then the number of cosets is $q^{n-k}$ and therefore, there are at most $(q^{n-k} - 1)$ nonzero error patterns that belong to different cosets can be corrected.

### 1.6 BOUNDS ON NUMBER OF PARITY-CHECKS AND MINIMUM DISTANCE OF A CODE

In this Section we mention some well-known bound on the number of parity-check digits of a linear code with respect to Hamming weight [60] only.

A code must use minimum numbers of parity-checks for keeping its error-correction/detection capability intact. But it is not always possible to determine the exact number of parity checks. Such situations demand to have bounds on parity checks.

**Hamming Sphere-Packing Bound:** This bound was obtained by Hamming [60].

If an $(n,k)$ linear code over $GF(q)$, a Galois field with $q$ elements, is capable of correcting $t$ or less (Hamming) errors then

$$n - k \geq \log \left[ 1 + \binom{n}{1}(q-1) + \ldots + \binom{n}{t}(q-1)^t \right].$$
A refinement of this bound was obtained by Wax [114].

In 1960, Plotkin [93] obtained some closer bounds and relations over the number of parity checks and minimum distance for a linear/non-linear block codes. There are also some important contributions in this area, by Helgert and Stinaff [65], Johnson [70] and Bambah, Joshi and Luther [6].

**Plotkin Bound:** Plotkin [93] derived a necessary lower bound on number of parity checks by employing the technique of taking average. This bound is stated as below:

If \( n \geq \frac{(qd - 1)}{(q - 1)} \), the number of parity checks required to achieve minimum weight \( d \) in an \( n \)-symbol linear block code is at least

\[
\left\lceil \frac{(qd - 1)}{(q - 1)} \right\rceil - 1 - \log_q d.
\]

**Varshamov-Gilbert Bound:** A general lower bound on the number of code words in a code with given length and minimum distance was given by Gilbert [51] and Varshamov [114] independently. This states that:

A sufficient condition for the existence of an \((n,k)\) linear code over \(GF(q)\) with minimum Hamming distance at least \(d\) is that \(n,k,q\) and \(d\) are such that they satisfy the inequality

\[
q^{n-k} > \sum_{i=0}^{d-2} \binom{n-1}{i} (q-1)^i.
\]

Later Sacks [98] gave a method to obtain above bound by constructing a parity-check matrix.

**Regier Bound** [97]: This bound was given by Regier in 1960, and is stated below:

In order to correct all burst errors of length \(b\) or less, a linear block code must have at least \(2b\) parity-check symbols. In order to correct all bursts of length \(b\) or less and simultaneously detect all bursts of length \(l \geq b\) or less, the code must have at least \((l+b)\) parity-check symbols.

**Extended Varshamov-Gilbert Bound:** This bound was obtained by Sharma and Dass [102]. This bound is an extension of the Varshamov-Gilbert bound for a code that has
no burst of length $b$ or less as a code word. This bound assures the existence of a code that can detect all error patterns which are either bursts of length $b$ or less or have weight $(w-1)$ or less.

There exists an $(n,k)$ linear code with minimum weight $w$ that has no non-zero burst of length $b$ or less as a code word $(w \leq b)$ satisfying the inequality

$$\sum_{i=0}^{w-2} \binom{n}{i}(q-1)^i + \sum_{j=1}^{b-1} \binom{b-1}{j}(q-1)^j \geq q^{n-k}.$$  

**Extended Sphere-Packing Bound:** Sharma and Dass [102] extended the Hamming sphere-packing bound to a code which is capable of correcting all bursts of length $b$ or less.

The number of parity-check symbols in any linear code that corrects all bursts of length $b$ or less with weight $w$ or less $(w \leq b)$ is at least

$$\log_q \left[ q^{n-1}[(q-1)(n-w+1)+1] + (q-1)^2 \sum_{j=w+1}^{b} (n-j+1)[1+(q-1)]^{1-j-w} \right].$$

An $(n,k)$ linear code capable of correcting all combinations of $m$ or fewer errors and all bursts of length $b$ or less must satisfy

$$n-k \geq \log_q \left[ \sum_{i=0}^{m} \binom{n}{i}(q-1)^i + (q-1)^2 \sum_{i=m+1}^{b} (n-i+1)[q^{i-2} - [1+(q-1)]^{i-2,m-2}] \right].$$

### 1.7 CYCLIC CODE

The notion of cyclic codes was introduced by Prange [94]. Cyclic codes are very simply mechanized using general linear finite-state switching circuits. In modern age coding, these linear codes have become very important in error detection and correction. Cyclic codes have a highly mathematical structure and so the decoding can be accomplished through the solution of a set of polynomial equations, the roots of which determine the error locations. These codes possess the minimum weight property and guarantee to detect and correct all errors up to a certain number and of a defined minimum weight. One of the most important classes of multiple random errors correcting codes is BCH codes that are cyclic codes.
Cyclic codes are not only important for random error correction but also important for burst error correction and detection. There are Hamming codes, BCH codes and other multiple random error correcting codes for solving the problem of single and multiple random error correction. But in many channels, errors do not occur independently but in clusters. So there is a need to develop codes for detecting and correcting burst errors.

In order to define a cyclic code, first we give some definitions on polynomial arithmetic.

The algebra of polynomials modulo \( p(x) \) is a field if \( p(x) \) is an irreducible polynomial with coefficients in a field \( F \).

The field of polynomials over \( GF(q) \) mod \( p(x) \), is called the ground field and here \( p(x) \) is an irreducible polynomial of degree \( m \), is called the Galois field of \( p^m \) elements, or \( GF(p^m) \). \( GF(p) \) is called the ground field and \( GF(p^m) \) is called the extension field.

If the ground field contains \( q \) elements then the extension field has \( q^k \) elements.

While defining a cyclic code, an \( n \)-tuple is considered to be an element of the algebra of polynomials \( P_n \) mod \( x^n - 1 \). An \((n,k)\) cyclic code has code words of length \( n \) and \( k \) information digits.

**Cyclic Code** [76]: An \((n,k)\) block code is said to be a cyclic code if for every code word

\[
f(x) = a_0 + a_1 x + \ldots + a_{n-1} x^{n-1},
\]

the word

\[
f(x) = a_{n-1} + a_0 x + \ldots + a_{n-2} x^{n-2},
\]

is also a code word. Equation (1.7.2) is equivalent to \( xf(x) \), viz:

\[
xf(x) = a_0 x + a_1 x^2 + \ldots + a_{n-1} x^n = a_{n-1} + a_0 x + \ldots + a_{n-2} x^{n-1} + a_{n-1} (x^n - 1)
\]

\[
\{xf(x)\} = a_{n-1} + a_0 \{x\} + a_1 \{x^2\} + \ldots + a_{n-2} \{x_{n-1}\} = f(x)
\]

where \( \{ \} \) denotes algebraic operation of polynomials modulo \( x^n - 1 \).
It is well known that in an \((n,k)\) cyclic code a burst of length \((n-k)\) or less is always detectable. Two known results in this direction, as given in Peterson and Weldon [92], are given below:

**Theorem 1.7.1:** No code vector of an \((n,k)\) cyclic code is a burst of length \((n-k)\) or less. Therefore, every \((n,k)\) cyclic code can detect any burst of length \((n-k)\) or less.

**Theorem 16.2:** The fraction of bursts of length \(b > n-k\) that can be undetected by any \((n,k)\) cyclic code is

\[
\begin{cases} 
    q^{-(n-k-1)} & \text{if } b = n-k + 1 \\
    \frac{q}{q-1} & \text{if } b > n-k + 1.
\end{cases}
\]

### 1.8 Burst Errors: Some Definitions

In this section, we are stating some definitions of various types of burst errors related to our study. Although, we will repeat these definitions in further respective chapters.

Since the development of various burst error detecting and correcting codes, several variants and modifications of the definition of a burst error came up depending upon the various kinds of channels, in use.

A burst of length \(b\) may be defined as follows, as given by Fire [43]:

**Definition 1.8.1 Burst-Error:** A burst of length \(b\) is a vector whose only non-zero components are among some \(b\) consecutive components, the first and the last component of which is non-zero.

**Example:** \((000120410)\) is a burst of length 5 over \(GF(5)\).

Fire called such errors as ‘open-loop burst errors’

There is also another kind of burst error due to Chien and Tang [24] that has attracted attention of many researchers. In literature, this burst is known as CT burst.

A CT burst may be defined as follows:
**Definition 1.8.2 CT Burst-Error:** A CT burst of length $b$ is a vector whose only non-zero components are confined to some $b$ consecutive positions, the first of which is non-zero.

**Example:** $(0001110000)$ is a CT burst of length 4 over $GF(2)$.

This definition was further modified by Dass [26] and he defined ‘burst of length $b$ (fixed)’ as follows:

**Definition 1.8.3 Burst of Length $b$ (Fixed):** A burst of length $b$ (fixed) is an $n$-tuple whose only non-zero components are confined to $b$ consecutive positions, the first of which is non-zero and the number of its starting positions in an $n$-tuple is among the first $n - b + 1$ components.

**Example:** $(1001000000)$ is a burst of length 7 (fixed) over $GF(2)$.

Amongst several generalizations of bursts, there is a good deal of research work devoted to the study of multiple burst correcting and detecting codes.

A multiple burst may be defined as follows [19]:

**Definition 1.8.4 Multiple Burst Error:** A pattern of $m$ bursts of length $b$ is called an $m$-multiple burst of length $b$.

**Example:** $(01101001110100101011000)$ is a 4-multiple burst of length 4 $GF(2)$.

Looking at the fact that there are many channels that requires systematic study of multiple burst errors, Berardi, Dass and Verma [9] proposed a new kind of burst error which they termed as ‘2-repeated burst error’.

A 2-repeated burst error may be defined as follows:

**Definition 1.8.5 2-Repeated Burst:** A 2-repeated burst of length $b$ is a vector of length $n$ whose only non-zero components are confined to two distinct sets of $b$ consecutive components the first and the last component of each set being non-zero.

**Example:** $(000120400100300)$ is a 2-repeated burst of length 4 over $GF(5)$.

Dass and Verma [33] also gave a generalization of this error by defining ‘$m$-repeated burst error’.
An $m$-repeated burst error may be defined as follows:

**Definition 1.8.6 $m$-Repeated Burst-Error:** An $m$-repeated burst of length $b$ is a vector of length $n$ whose only non-zero components are confined to $m$ distinct sets of $b$ consecutive components, the first and the last component of each set being non-zero.

**Example:** $(0010200241003140000)$ is a 3-repeated burst of length 3 over $GF(5)$.

Later, Dass, Garg and Zanneti [32] gave another kind of repeated burst error, i.e., 2-repeated burst of length $b$ (fixed).

A 2-repeated burst of length $b$ (fixed) may be defined as follows:

**Definition 1.8.7 2-Repeated Burst Error of Length $b$ (Fixed):** A 2-repeated burst of length $b$ (fixed) is an $n$-tuple whose only non-zero components are confined to two distinct sets of $b$ consecutive digits, the first component of each set is non-zero and the number of its starting positions is among the first $n - 2b + 1$ components.

**Example:** $(1000001000)$ is a 2-repeated burst of length up to 4(fixed).

Generalizing the concept of 2-repeated burst of length $b$ (fixed), Dass, Garg and Zannetti [32] defined $m$-repeated burst of length $b$ (fixed).

An $m$-repeated burst of length $b$ (fixed) may be defined as follows:

**Definition 1.8.8 $m$-Repeated Burst Error of Length $b$ (Fixed):** An $m$-repeated burst of length $b$ (fixed) is an $n$-tuple whose only non-zero components are confined to $m$ distinct sets of $b$ consecutive digits, the first component of each set is non-zero and the number of its starting positions is among the first $n - mb + 1$ components.

**Example:** $(000011001000100000)$ is a 3-repeated burst of length at most 3(fixed).

Dass and Garg also defined a 2-repeated burst of length $b$ (fixed) in terms of its weight as ‘2-repeated low-density burst of length $b$ (fixed)’ [30].

A 2-repeated low-density burst of length $b$ (fixed) may be defined as follows:

**Definition 1.8.13 2-Repeated Low-Density Burst Error:** A 2-repeated low-density burst of length $b$ (fixed) with weight $w$ or less is an $n$-tuple whose only non-zero components are confined to two distinct sets of $b$ consecutive components the first component of each set is non-zero where each set can have at most $w$ non-zero
components \((w \leq b)\), and the number of its starting positions is among the first \(n - 2b + 1\) components.

**Example:** \((00111000000100000)\) is a 2-repeated low-density burst of length up to 6 (fixed) with weight 4 or less.

Dass and Garg [30] defined an \(m\)-repeated low-density burst as follows:

**Definition 1.8.14 \(m\)-Repeated Low-Density Burst Error:** An \(m\)-repeated low-density burst of length \(b\) (fixed) with weight \(w\) or less is an \(n\)-tuple whose only non-zero components are confined to \(m\) distinct sets of \(b\) consecutive components the first component of each set is non-zero where each set can have at most \(w\) non-zero components \((w \leq b)\), and the number of its starting positions is among the first \(n - mb + 1\) components.

**Example:** \((1000110000000100000)\) is a 3-repeated low-density burst of length up to 3 (fixed) with weight 2 or less.

In certain systems due to some particular type of disturbances, it is natural that burst errors occur in such a way that that over a given length some digits are received correctly and others are not. This type of situation demands the study of codes that can correct and/or detect this particular type of burst error. For this purpose Jain [68] defined ‘moderate-density-open loop bursts’.

A moderate-density-open loop burst may be defined as follows:

**Definition 1.8.9 Moderate-Density Open-Loop Burst:** A moderate-density open-loop burst of length \(b\) is a vector whose only non-zero components are confined to some \(b\) consecutive components, the first and the last component of which is non-zero, with weight lying between \(w_1\) and \(w_2\) \((w_1 \leq w_2 \leq b)\).

**Example:** \((00110010100)\) is a moderate-density open-loop burst of length 7 with weight lying between \(w_1 = 2\) and \(w_2 = 7\).

Jain also defied a CT moderate-density burst, as follows [68]:

**Definition 1.8.10 CT Moderate-Density Burst:** A CT moderate-density burst of length \(b\) is a vector whose only non-zero components are confined to some \(b\) consecutive
positions, the first of which is non-zero, with weight lying between \( w_1 \) and \( w_2 \) \( (w_1 \leq w_2 \leq b) \).

**Example:** \((11000000)\) is a CT moderate-density burst of length 4 with weight lying between \( w_1 = 1 \) and \( w_2 = 4 \).

In some practical channels, the burst occur in such a way that all the digits within a burst are corrupted. Such types of bursts are termed as ‘solid burst errors’.

A solid burst may be defined as follows [37]:

**Definition 1.8.11 Solid Burst error:** A solid burst of length \( b \) is a vector whose all the \( b \) consecutive components are non-zero and rest are zero.

**Example:** \((000111110)\) is a solid burst of length 6.

Recently, as a generalization of solid burst error, Dass [40, 41] defined ‘repeated solid burst error’.

A 2-repeated solid burst error may be defined as follows [38]:

**Definition 1.8.12 2-Repeated Solid Burst Error:** A 2-repeated solid burst of length \( b \) is a vector of length \( n \) whose only non-zero components are confined consecutively to two distinct sets of \( b \) consecutive components.

**Example:** \((1110011100)\) is a 2-repeated solid burst of length 3.

An \( m \)-repeated solid burst error may be defined as follows [39]:

**Definition 1.8.13 m-Repeated Solid Burst Error:** An \( m \)-repeated solid burst of length \( b \) is a vector of length \( n \) whose only non-zero components are confined consecutively to \( m \) distinct sets of \( b \) consecutive components.

**Example:** \((111100111101111100111100)\) is a 4-repeated solid burst of length 4.

### 1.9 APPLICATIONS OF CODING THEORY

Coding theory is an important study which attempts to minimize data loss due to errors introduced in transmission from noise, interference or other forces. With a wide range of theoretical and practical applications from digital data transmission to modern medical research, coding theory has helped enable much of the growth in the 20th century. There
are many applications of coding theory in the modern word. A brief review of some applications of coding theory is given below:

**Communication System** [47, 73, 111]: In satellite communication the channel noise can be regarded as additive white Gaussian and therefore, mainly self-orthogonal convolutional codes are used. Specifically a $\frac{1}{2}$ rate convolutional with Viterbi decoding which produces large error correction capabilities is used. In some particular cases, high-rate BCH codes are also used.

In teletext digitized characters and figures are transmitted. These digitized characters and figures overlapped with TV signals. In such type of communication the most common error that occurs is burst error that is due to impulsive noise. In this communication system TV set itself performs the decoding. Therefore, a small decoder is needed and hence a different set of cyclic codes can be used that can be decoded by a relatively simple circuit.

For a space communication system, the main requirements are high reliability and efficiency. Therefore, usually Trellis Coded Modulation (TCM) and Reed-Solomon Codes are considered in these systems.

A Global System for Mobile Communication, i.e., GSM is the digital cellular radio system. When a message is sent through a GSM system, various types of signals are needed to transmit that message over a physical channel. In a GSM system, generally turbo codes are used so that the better efficiency, amplitude and accurate data transmission using channel coding of the speech data, channel coding of the signaling channels and channel coding of the data channels can be provided. In a Code Division Multiple Access (CDMA) system convolutional codes are used because these codes provide a good example of application in error protection in both forward CDMA channel and reverse CDMA channel.

**Computer System** [45]: In traditional dynamic random access memory (DRAM), Hamming codes are used for error correction. In most of the memory devices information is stored in two dimensions, therefore, errors usually occur in the form of two-dimensional bursts.

In a magnetic storage device, the burst errors occur due to defects of the device or due to dust. In such devices, a more powerful family of codes, i.e., Reed-Solomon codes, is
used to protect data. In these devices, along with Reed-Solomon codes, Fire codes and interleaved Reed-Solomon codes are also used.

In optical disk systems there occur both types of errors random errors and burst errors. So the error rate of device is relatively very high and hence multiple coded Reed-Solomon codes are used for this purpose.

**Genetic Research (Medical science)** [107, 7, 59]: One of the applications of coding theory in medical science is its use in the study of evolution and genetic mutations.

Genetic information in the form of DNA is taken as input, transmitted via the process of replication and amino acid proteins are obtained as an output. In this process, errors are introduced due to fluctuations in heat, radioactivity and other factors. It was suggested by some scientist that some type of error correcting code can be employed in this process to ensure the survival of the species.

Convolutional coding allows for immediate past and future information to be used in the encoding/decoding process.

In a recent study, Srinivas, Jain, saurav and Sikdar [107] showed that the study of repeated burst error correcting/detecting codes plays a significant role in the study where the changes in the neuronal network properties during epileptiform activity in vitro in planar two-dimensional neuronal networks cultured on a multi-electrode array using the in vitro model of stroke-induced epilepsy have been explored.

**1.10 MOTIVATION**

Coding for communication is essential. The major issue is reliability of communication over unreliable/noisy channels. This is achieved through redundancy. Main motivation of the research undertaken is to consider minimum redundancy driven efficiency considerations for a variety of errors mostly found but not studied in the existing literature. This arises by considering repeated bursts of different kinds, an idea that not only extends the idea of burst-error corrections but also generalizes the idea random errors, where in place of errors in single positions these are bursts randomly spread over the code words. In these considerations weight has played key role. However, mainly only two types of weights given by Hamming [60] and Lee [78] have been used in coding theory. Hamming metric [60] and codes studied on that were binary. But in case of Lee [78], the code characters are taken from the ring \( \mathbb{Z}_q \) of integers modulo \( q \). Later
search for all possible distances was accomplished by Sharma and Kaushik [105], over $GF(q)$. This depends on the partition of $GF(q)$ into subsets with specific structure. However, we in this thesis will confine to Hamming case, that when considered over $GF(q)$, also has weight sensitivity.

We, closely following the possible patterns of errors, consider only those errors that can actually occur in the system. Considering in a position possibility of all possible substitution errors, when these are limited to few around brings unnecessary errors for correction. This study of bursts in terms of weight was initiated by Sharma and Dass [101]. Although some results for repeated bursts regarding their weights in terms of specific density were obtained by some authors [30, 35] but there was a need of systematic and close study of various types of repeated bursts that may be helpful in developing more efficient and reliable codes for this digital era where no technology can work without proper encoding and decoding of messages.

This fact led us to study various types of multiple bursts with their different characteristics and needs by examining their capabilities with one of the most important criteria of weight consideration. This study of us may be proved very helpful in the development of various multiple burst error correcting and detecting codes that are highly required for the purpose of reliable communication in different types of practical channels.

### 1.11 OBJECTIVES OF THE STUDY

Major objective of the study reported in this thesis can perhaps be seen in two different ways:

(a) From application point of view to get combinatorial results with weight consideration for efficient communication when repeated burst errors occur, which largely occur when messages are of longer lengths. In fact, by considering bursts of all four different types, the study is very broad in its coverage. Our study provides repeated burst correction/detection criteria for different variety of channels that in practice can be used with currently ever widening phenomena of digital communication.

(b) From theoretical point of view, a major perspective also follows from our study. Reflecting on random and burst errors, as these are, it can be easily seen that burst errors constitute a subclass of random errors. Our study of multiple bursts in message vectors
presents a mathematical formulation of a much broader and general class of errors unifying random errors and multiple burst errors. So much so that random error correcting codes can clearly be viewed as a subclass of multiple burst error correcting codes, with every burst of length one in particular. In fact, our study of multiple burst errors is thus a generalization of random errors and burst errors in one stroke.

Specifically speaking of efficiency, if a repeated burst is considered as a single burst or as a bunch of random errors, we will have to use more parity-checks and that will reduce the efficiency of the code.

Areas of problems mathematically handled, in short, can be put as follows:

- Developing the mathematical formulation for weight calculation of various types of repeated burst errors.
- Introduction of moderate-density repeated burst error and study of cyclic codes detecting repeated burst errors and moderate-density repeated burst errors.
- Study of linear codes those are capable of correcting/detecting repeated solid burst errors.

### 1.12 PROBLEMS WORKED OUT IN THIS THESIS

The thesis contains six chapters and is organized as follows:

**CHAPTER 1**

*Introduction & Relevant Literature Survey*

This chapter is introductory in nature. In which a brief history of Coding Theory and background material relevant for investigations undertaken in later chapters as also some applications of coding theory are given. The chapter also includes, a brief overview of the work reported in the later five chapters.

**CHAPTER 2**

*Results on Weights of Repeated Bursts of Length $b$*

Codes being the mathematical tools for reliable communication over noisy channels, with the help of imbued redundancy, one major problem are to keep the redundancy minimum and improve efficiency. In the early studies, the case being binary, weight
consideration in defining errors did not arise. These were studied with interest over the number of code vectors of a code. Weight distributions, differently of course, have been deeply studied by McWilliams and Sloane [82] in the case of random errors. They studied codes with different weights. But efficiency requires close and critical study of constraints on the errors. In the case of burst errors beside length, the weight provides an important consideration. This is an important area of study for burst errors also, that we have undertaken for multiple burst errors. Our study in this chapter is concerned to codes with weight constraints.

With the wide and varied applications, coding is no longer for binary case as the channels are also of different types. Therefore nature of errors, in particular burst errors, which usually depends upon the type of channel used during the process of transmission. It has been observed that in very busy communication channels, bursts repeat themselves.

In Chapter 2, we study repeated bursts of a given length or less with and without weight constraints.

First we have obtained results on counting the numbers of 2-repeated bursts with and without weight consideration. Then we derived an expression for the total weight of all 2-repeated bursts in a vector of length $n$ is found in this chapter. In our next result, we have given a recurrence relation between total weights of 2-repeated bursts of different lengths. In coding theory, minimum weight being of some importance, we have obtained an expression for the minimum weight of a vector having 2-repeated bursts in the space of all $n$-tuples. We have obtained an expression for the total weight of 2-repeated bursts of length $b$ and weight $w$ or less in the space of all $n$-tuples. A recurrence relation in this very general case is obtained. An upper bound on the minimum weight of a 2-repeated burst of length $b$ with weight $w$ or less is established.

As a generalization of 2-repeated bursts, we have considered $m$-repeated bursts and analogous results are obtained for $m$-repeated bursts in both cases with weight constraints and without weight constraints.

My two published research papers related to this chapter are:


CHAPTER 3

On Weights of Repeated Bursts of Length $b$ (fixed) and Repeated Low-Density Burst of Length $b$ (fixed)

When repeated bursts are introduced by impulse noise, as it happens in the case of large number of practical channels, the length of a repeated burst depends upon the duration of the impulse noise. Also, the number of errors, i.e., the density of errors within a repeated burst clearly depends upon the impulse noise. Depending upon the behavior of the channel, either due to impulse noise or otherwise, the density of errors in the repeated bursts introduced by the channel is either too small or too large. Employing the usual burst-correcting codes is not quite appropriate for this purpose of efficient transmission over such a channel. In such cases we need to consider low-density repeated bursts instead of simple repeated bursts.

In Chapter 3, our study is concerned with repeated burst error of length $b$ (fixed). We start with obtaining results for simple repeated bursts of length $b$ (fixed) and then we find results for repeated low-density bursts of length $b$ (fixed). First we have obtained results for 2-repeated burst errors of length $b$ (fixed). We have obtained results on counting the number of 2-repeated burst errors of length $b$ (fixed) with and without weight constraints. An expression for total weight of 2-repeated bursts of length $b$ (fixed) is derived. An upper bound on the minimum weight element of 2-repeated bursts of length $b$ (fixed) is also obtained. Results regarding total weight and average weight of 2-repeated bursts of length $b$ (fixed) with weight $w$ are obtained. A recurrence relation between weights of different 2-repeated bursts is also obtained. There is a result on average weight of 2-repeated bursts of length $b$ (fixed).
As a generalization of 2-repetated bursts of length $b$ (fixed), we have considered $m$-repetated bursts of length $b$ (fixed) and corresponding results are obtained for $m$-repetated bursts in the both cases with weight constraints and without weight constraints.

Further in an analogous manner, we have obtained results for total numbers of 2-repetated low-density bursts of length $b$ (fixed), total weight of 2-repetated low-density bursts of length $b$ (fixed) and minimum weight of 2-repetated low-density bursts of length $b$ (fixed). After that we have obtained corresponding results for $m$-repetated low-density bursts of length $b$ (fixed).

My research papers related to this chapter are:


(ii) BhuDev Sharma, Barkha Rohtagi and A.K Aggarwal, “Some results on weights of $m$-repeated bursts,” (Communicated)

**CHAPTER 4**

**Moderate-Density Repeated Burst Error Detecting Cyclic Codes**

In this chapter we study cyclic codes detecting repeated burst errors and have considered a different kind of repeated burst, i.e., moderate density repeated burst. Then cyclic codes detecting moderate-density 2-repeated burst errors have been studied. Further, we obtained results for cyclic codes detecting $m$-repeated burst errors and moderate-density $m$-repeated burst errors. We then make a comparative study of results for moderate-density 2-repeated burst error of length $b$ and moderate-density 2-repeated burst errors of length $b$ (fixed). This study shows that there is a difference with regular interval between the values obtained for moderate-density 2-repeated burst error of length $b$ and moderate-density 2-repeated burst errors of length $b$ (fixed), when we apply similar weight constraints on them.

In this study, it is interesting to see that while obtaining results for moderate-density $m$-repeated burst errors, there arise two special cases:

(i) For $m = 1$ and neglecting the consideration of starting positions, the result reduces to the case of moderate-density open-loop burst error in [68].

(ii) For $b = 1$, the result reduces to the case of random errors.
This study may help us in developing codes which detect/correct those errors that are repeated bursts of a given length with weight lying between some pre assigned values. The development of such codes economizes in the parity-check digits required, suitably reducing the redundancy of the code, i.e., suitably increasing the efficiency of transmission.

My research papers related to this chapter:


**CHAPTER 5**

**Codes Correcting and Detecting Repeated Solid Burst Errors**

In situations, where channel introduces repeated burst errors in such a way that each digit in the error vector is nonzero, it is uneconomical to use simple repeated burst error correcting /detecting codes. Such types of repeated burst errors are termed as *repeated solid burst errors*.

In this chapter, we have obtained results for linear codes correcting repeated solid burst errors. An example is given that shows that the syndromes of all repeated solid bursts of length 3 or less are nonzero and distinct. Cyclic codes detecting repeated solid bursts are also studied in this chapter.

My research papers related to this chapter:

(i) Barkha Rohtagi and BhuDev Sharma, “Bounds on codes detecting and correcting 2-repeated solid burst errors,” (Accepted with minor revision in *Journal of Applied Mathematics and Informatics*).
CHAPTER 6

Conclusion and Future Scope

The work reported in Chapters 2 to 5, contains a variety of results on different kinds of multiple-bursts correcting codes. With use of coding through digital operations having widened, and the nature of errors being multiple bursts of one or the other kind, our study should be found helpful in closer examination of efficiency criteria.

*****