2011

Physical Information Theoretic Bounds on Energy Costs for Error Correction

Natesh Ganesh
University of Massachusetts Amherst, nganesh@engin.umass.edu

PHYSICAL INFORMATION THEORETIC BOUNDS ON ENERGY COSTS FOR ERROR CORRECTION

A Thesis Presented

by

NATESH GANESH

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

MASTER OF SCIENCE IN ELECTRICAL AND COMPUTER ENGINEERING

September 2011

Electrical and Computer Engineering

PHYSICAL INFORMATION THEORETIC BOUNDS ON ENERGY COSTS FOR ERROR CORRECTION

A Thesis Presented

by

NATESH GANESH

Approved as to style and content by:

Neal G. Anderson, Chair

Dennis Goeckel, Member

Eric Polizzi, Member

Christopher V. Hollot, Department Chair
Electrical and Computer Engineering

Matha, Pitha, Guru, Dheivam.

ACKNOWLEDGMENTS

I would like to thank my advisor and mentor Prof. Neal Anderson for all his guidance, patience and support. I would also like to thank my colleague and friend Ilke Ercan for taking the time to help me whenever I have needed it. I am grateful to my family for always being there for me, and my friends Shrey, Vaishu, Nandhini, Mala, Aandhai, Kodi, Caroline, Joel and my partner in crime Prakash and many more for their support and for providing me with some necessary and unnecessary distractions. A special thanks to all members of the UMBDT for always giving me something to look forward to at the end of a day. Joga Bonito.

ABSTRACT

PHYSICAL INFORMATION THEORETIC BOUNDS ON ENERGY COSTS FOR ERROR CORRECTION

SEPTEMBER 2011

NATESH GANESH

B.Tech., NATIONAL INSTITUTE OF TECHNOLOGY, TRICHY, INDIA

M.S.E.C.E., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Neal G. Anderson

With diminishing returns in performance with scaling of traditional transistor devices, there is a growing need to understand and improve potential replacements technologies. Sufficient reliability has not been established in these devices and additional redundancy through use of fault tolerance and error correction codes are necessary. There is a price to pay in terms of energy and area, with this additional redundancy. It is of utmost importance to determine this energy cost and relate it to the increased reliability offered by the use of error correction codes. In this thesis, we have determined the lower bound for energy dissipation associated with error correction using a linear (n,k) block code. The bound obtained is implementation independent and is derived from fundamental considerations and it allows for quantum effects in the channel and decoder. We have also developed information theoretic efficacy measures that can quantify the performance of the error correction and their relationship to the corresponding energy cost.

TABLE OF CONTENTS

LIST OF FIGURES

Figure **Page**

CHAPTER 1

INTRODUCTION

1.1 Motivation and Overview

Nanotechnology requires the manipulation of matter on an atomic and molecular scale, and involves structures which are in the range of 1-100 nm in at least one dimension. Nanotechnology can be used to create devices that can be used to perform computation. Though these devices will not replace complementary metal-oxide-semiconductor(CMOS) transistor devices in the nearby future, it is of utmost importance that research is carried out to develop these devices as well as new architectures and tools to study and maximize their potential.

In computing, nanotechnology could follow two possible paths - Evolutionary and Revolutionary [1]. In computing, the evolutionary path deals with continuing scaling of CMOS devices. There are however many drawbacks to this including fundamental lower limit on transistor sizes, interconnect scaling, disproportional increase in performance with device density, power density, etc. Revolutionary nanoelectronics research involves the study of nanotube transistors, spintronics, magnetic memory devices, quantum devices, molecular devices and optoelectronics. Extensive research in these technologies to improve there reliability and signal to noise ratio (SNR) is necessary if they are to replace CMOS in the future.

One of the major problems facing us with respect to future nanodevices is that of high manufacturing defect rates. These devices which may operate at lower noise margin result in greater soft errors. Low power devices possess low energy barriers between different logic states and are very prone to thermal noise. Further the fun-

damental quantum nature of these devices renders them probabilistic and introduces quantum noise which cannot be eliminated. Thus there is a need to achieve robustness in nanoscale technology in order for them to be used effectively (for computation). This is has been done so through defect tolerance built into the fabric, as well as through the use of error-correction codes/techniques [1], [2].

Addition of redundancy in the form of error-correction codes leads to increase in design complexity, chip area and power consumption. Since power dissipation and energy efficiency have been identified as an important bottleneck in the growth of future technologies, the focus of this study has been to determine fundamental lower bounds on power dissipation associated with performing error correction on physical realizations of signal states. Robustness is achieved through the use of a linear (n,k) single error correcting code. Emphasis has also been laid on identifying and testing potential information theoretic quantities that will provide an insight into the relationship between the "amount" of error correction achieved with the energy required to do so.

Chapter 2 and Chapter 3 present the theoretical background of the concepts used in this thesis. We start Chapter 2 by introducing the concept of error correcting codes, their construction, the parameters involved and the various properties associated with them. We shall also look at the decoding strategy used. This will be followed by an introduction into information theory. Starting from the concept of information entropy, we shall move onto relative entropy and mutual information.

Chapter 3 will introduce principles associated with the physical realization of information bearing systems. Decoherence and the relation between information theoretic and thermodynamic quantities are also discussed. The chapter will also explore the Referential approach to physical information theory and in depth, analyze the implications of this approach, for certain established ideas of relevance to this work [12], [13]. The derivation and use of computational efficacy measures are also touched

upon. In Chapter 4, how information is encoded in a physical quantum system using density matrices and the need for changes to the efficacy measures introduced in Chapter 3 are presented. The changes in the definitions are studied on an example system, and are followed up by extending the usage of these definitions to more complex systems. We have then focused on our system of interest and developed the necessary formulation for performance metrics and energy dissipation. In the last section of this chapter, the results we have obtained are presented, followed by a discussion of the various challenges faced and the interesting questions that have been raised over the course of this work. Chapter 5 is a brief summary of what has been covered in this thesis.

CHAPTER 2

TECHNICAL BACKGROUND I: ERROR CORRECTION CODES AND CLASSICAL INFORMATION THEORY

2.1 Introduction to Error-Correction Codes

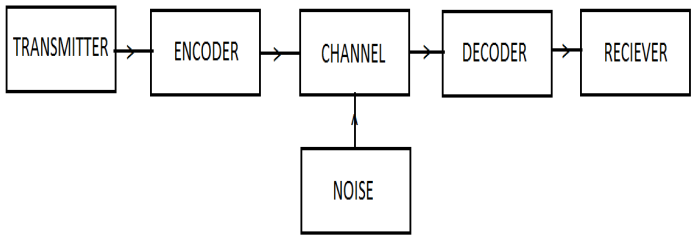

Figure 2.1. Block diagram of a simple communication system with encoding (after [5])

There are two structurally different types of codes which are in use today: Block codes and Convolutional codes. For block codes, the encoder divides the information sequence into message blocks of k-tuple message words $\boldsymbol{u}$ =(u_0,u_1,....,u_k). The encoder then transforms the k-tuple message word $\boldsymbol{u}$ into n-tuple code word $\boldsymbol{v}$. Thus with the binary alphabet, 2^k message words are converted into the code words. This set of 2^k out of the possible 2^n words is called a (n,k) block code. The code rate R is defined

as number of information bits entering the encoder per transmitted symbol and thus R=k/n.

We have $k \leq n$ for error-correction codes, which implies the code rate R is lesser than or equal to 1. The (n-k) bits added to each message to form a code word are called parity bits. These redundant bits provide the code with the capability of combating noise. The choice of the number of parity bits in a (n,k) block code is a major issue while designing the encoder.

2.1.1 Linear Block Codes

For ease of code generation and analysis, we shall focus upon a subclass of codes within the class of block codes called linear block codes. A block code of length n and 2^k code words is called a linear (n,k) code if and only if its 2^k code words form a k-dimensional subspace of the vector space of all the n-tuples over the field $GF(2)$ [4], [5], [6]. This means that a binary block code is linear if and only if the modulo-2 sum of any two code words is also a code word. For any two code words $\boldsymbol{v}_i$ and $\boldsymbol{v}_j$

$$\boldsymbol{v}_i \bigoplus \boldsymbol{v}_j = \boldsymbol{v}_k \tag{2.1}$$

where $\boldsymbol{v}_k$ is also a code word.

Since a (n,k) linear code C is a k-dimensional subspace of the vector space $\mathbf{V}_{\boldsymbol{n}}$ of all the binary n-tuples, it is possible to find k linearly independent code words, $\{\boldsymbol{g}_0, \boldsymbol{g}_1, ..., \boldsymbol{g}_{k-1}\}$, in C such that every code word $\boldsymbol{v}$ in C is a linear combination of these k code words; that is

$$\boldsymbol{v} = u_0\boldsymbol{g}_0 + u_1\boldsymbol{g}_1 + ... + u_{k-1}\boldsymbol{g}_{k-1} \tag{2.2}$$

where u_i=0 or 1 for $0 \leq$ i $\leq k$. We can arrange these k linearly independent code

words as the rows of a $k \times n$ matrix as shown below:

$$\mathbf{G} = \begin{bmatrix} \boldsymbol{g}_0 \\ \boldsymbol{g}_1 \\ \vdots \\ \boldsymbol{g}_{k-1} \end{bmatrix} = \begin{bmatrix} g_{00} & g_{01} & g_{02} & \cdots & g_{0,(n-1)} \\ g_{10} & g_{11} & g_{12} & \cdots & g_{1,(n-1)} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ g_{k-1,0} & g_{k-1,1} & g_{k-1,2} & \cdots & g_{k-1,n-1} \end{bmatrix} \tag{2.3}$$

where $\boldsymbol{g}_i$=(g_{i0},g_{i1},....,$g_{(i,n-1)}$)) for 0≤i≤k. If $\boldsymbol{u}$=(u_0,u_1,...,u_{k-1}) is the message to be encoded, the corresponding code word is given as

$$\boldsymbol{v} = \boldsymbol{u} \cdot \mathbf{G} = (u_0, u_1, ..., u_{k-1}) \cdot \begin{bmatrix} \boldsymbol{g}_0 \\ \boldsymbol{g}_1 \\ \vdots \\ \boldsymbol{g}_{k-1} \end{bmatrix} = u_0\boldsymbol{g}_0 + u_1\boldsymbol{g}_1 + ... + u_{k-1}\boldsymbol{g}_{k-1} \tag{2.4}$$

G is called the generator matrix and its rows span the (n,k) linear code C. There is also another useful matrix associated with every linear code. For any $k \times n$ matrix, there exists a *(n-k)*$\times n$ matrix **H** with (n-k) linearly independent rows such that any vector in the row space of **G** is orthogonal to the rows of **H**. Hence we can describe a linear code C generated by **G** as: A n-tuple $\boldsymbol{v}$ is a code word generated in C by **G** if and only if $\boldsymbol{v} \cdot \mathbf{H}^T = 0$.

2.1.2 Minimum Distance of a Block Code

A very important parameter of a linear block code is its minimum distance, which determines its error-detecting and error-correcting capabilities. Let $\boldsymbol{v}$=(v_0,v_1,....,v_{n-1}) be a binary n-tuple. The *Hamming weight* of $\boldsymbol{v}$, denoted by w($\boldsymbol{v}$) is defined as the number of nonzero components of $\boldsymbol{v}$. The *Hamming distance* between two n-tuples

$\boldsymbol{u}$ and $\boldsymbol{v}$, denoted as d($\boldsymbol{u}$,$\boldsymbol{v}$), is defined as the number of places where they differ. For a block code C, the *minimum distance*, denoted as d_{min} is defined as

$$d_{min} = min\{d(\boldsymbol{u},\boldsymbol{v}) : \boldsymbol{u},\boldsymbol{v} \in C, \boldsymbol{u} \neq \boldsymbol{v}\}. \tag{2.5}$$

Since in a linear block code, the modulo-2 sum of two code words is another code word, the minimum distance of code C can also be defined as the smallest Hamming weight of a nonzero code word. Thus d_{min}=w_{min}= min(w($\boldsymbol{x}$): $\boldsymbol{x}$$\in$$C$, $\boldsymbol{x}$$\neq$0)

2.1.3 Syndrome and Error Detection

Consider an (n,k) linear code with generator matrix **G** and parity-check matrix **H**. Let $\boldsymbol{v}$=($v_0, v_1, ..., v_{n-1}$) be the transmitted code word over a noisy channel and $\boldsymbol{r}$=($r_0, r_1, ..., r_{n-1}$) be the received vector at the output of the channel. Due to the channel noise, $\boldsymbol{r}$ may be different from $\boldsymbol{v}$. The vector sum

$$\boldsymbol{e} = \boldsymbol{r} + \boldsymbol{v}. \tag{2.6}$$

is a n-tuple, where e_i=1 for $r_i \neq v_i$ and e_i=0 if $r_i = v_i$. This n-tuple is called the error pattern or vector, which simply displays the positions where the received vector $\boldsymbol{r}$ differs from the transmitted code word $\boldsymbol{v}$. The 1's in $\boldsymbol{e}$ are the *transmission errors* caused by the channel noise. It is very clear that

$$\boldsymbol{r} = \boldsymbol{v} + \boldsymbol{e}. \tag{2.7}$$

On receiving $\boldsymbol{r}$, the decoder must determine whether $\boldsymbol{r}$ contains any transmission errors and if the presence of errors is detected, the decoder will have to take action to locate and correct them.

When $\boldsymbol{r}$ is received, the decoder computes the *(n-k)*-tuple:

$$\boldsymbol{s} = \boldsymbol{r} \cdot H^T \tag{2.8}$$

which is called the *syndrome* of $\boldsymbol{r}$. Now $\boldsymbol{s}$=0 if and only if $\boldsymbol{r}$ is a code word, and $\boldsymbol{s}\neq 0$ if and only if $\boldsymbol{r}$ is not a codeword. It is possible that the errors in certain error vectors are not detectable i.e. $\mathbf{r}$ contains errors but $\boldsymbol{s} = \boldsymbol{r} \cdot H^T$=0. This happens when the error pattern $\boldsymbol{e}$ is identical to a non-zero code word. In this event, $\boldsymbol{r}$ is the sum of two code words, which is a code word as well, and consequently $\boldsymbol{s} = \boldsymbol{r} \cdot H^T$=0. Error patterns of this kind are called *undetectable error patterns*. When an undetectable error pattern occurs the decoder makes a decoding error.

The syndrome $\boldsymbol{s}$ computed from the received vector depends only upon $\boldsymbol{e}$ and not on transmitted word $\boldsymbol{v}$.

$$\mathbf{s} = \mathbf{r} \cdot H^T = \mathbf{(v+e)}H^T = \boldsymbol{v} \cdot H^T + \boldsymbol{e} \cdot H^T \tag{2.9}$$

However since $\boldsymbol{v} \cdot H^T$=0, the above Eq(2.9) reduces to

$$\boldsymbol{s} = \boldsymbol{e} \cdot H^T \tag{2.10}$$

Any error-correction scheme would be a method of solving the (n-k) linear equations of Eq. (2.8) for the error vector $\boldsymbol{e}$. Once that is found, we can take $\boldsymbol{r+e}$ and get the actual transmitted word. However this is not simple as the linear equations of Eq(2.8) do not have a unique solution, but have 2^k solutions. Thus there are 2^k error patterns that result in the same syndrome, and the true error pattern is just

one of them. The decoder has to hence decide the right solution from 2^k candidates and must do so in a way to reduce the probability of decoding error.

If the channel is a binary symmetric channel or **BSC** with crossover probability less than 0.5, the most probable error pattern is the one that has the smallest number of non-zero digits. This idea is used in the construction of a standard array which is used in syndrome decoding [4].

2.1.4 Hamming Codes

The Hamming codes are a family of single error correcting linear codes. For any integer $m \geq 3$, there exists a Hamming code with the following parameters:

Code length: $n=2^m-1$

Number of information symbols: $k=2^m-m-1$

Number of parity check symbols: $n-k=m$

Error correcting capability: $t=1(d_{min}=3)$

For $m=3$, we have $n=7$ and $k=4$, and thus the Hamming (7,4) code which is capable of correcting all single-bit errors, or detecting all single-bit and double-bit errors. The Hamming distance between the transmitted and received words should not be greater than one for it to be correctable. The parity check matrix for the (7,4) code is given by

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 \end{bmatrix} \tag{2.11}$$

The Hamming(7,4) code is a *perfect code*, i.e. it satisfies the Hamming bound [5]. The weight distribution of a Hamming code of length n=$2^m - 1$ is given by the following polynomial,

$$A(z) = \frac{1}{n+1}\{(1+z)^n + n(1-z)(1-z^2)^{(n-1)/2}\} \tag{2.12}$$

where the number of code vectors of weight i, A_i, is simply the coefficient of z^i in the polynomial, which is called the *weighted enumerator* for the Hamming code. For m=3, n=7 and k=4, we have the weight distribution for the Hamming (7,4) code to be A_0=1,$A_3 = A_4$=7 and A_7=1.

The extended Hamming $(2^m, 2^m - m - 1, 1)$ code is formed by the addition of a parity bit to the $(2^m - 1, 2^m - m - 1, 1)$ code. For example, the Hamming (8,4) code is formed by the addition of a parity bit to the Hamming (7,4) code. The addition of the parity bit allows us to increase the minimum Hamming distance to 4. This enables the correction of a single-bit error as well detection of a double-bit error at the same time and they belong to the class of codes called Single Error Correcting and Double Error Detecting codes or SECDED codes. However the Hamming $(2^m, 2^m - m - 1, 1)$ code is not a perfect code.

2.2 Introduction to Information Theory

2.2.1 Classical Information Theory

Information theory seeks to obtain the fundamental limits on the reliability of compressing and exchanging data. The theory, originally used in the communication field, has since developed and found applications in a wide variety of disciplines. Application of this theory to nanoelectronic circuits is necessary as their intended purpose includes communication, computation and information storage. It can also

be used to develop many important performance metrics which will help in the exploration of future devices. Information theory allows us to connect such performance measures directly with related thermodynamic quantities such as thermodynamic entropy and energy dissipation, and hence provide us with important knowledge on the capabilities of these nanoelectronic circuits.

2.2.2 Entropy and Information

The concept of entropy, originally a thermodynamic construct, has been adapted to other fields of study, including information theory. It is commonly claimed that thermodynamic entropy can be interpreted as an application of the information entropy concept to a highly specific set of physical questions. Entropy is defined as the measure of the disorder of a system, which can be expressed as

A state of high order=low probability

A state of low order=high probability

The information entropy was introduced in 1948 by Claude Shannon through his exploration of the entropy of random variables and random processes in [3]. Information entropy is often eponymously called Shannon entropy or Shannon information. Before delving deeper into the formal definition of information and the formulation of information entropy, let us first see how the information content of an object is determined.

2.2.3 What is the Information Content of an Object?

In order to determine the information content of an object, let us introduce a sender (i.e. us) and a receiver, say a friend who shares some background knowledge with us (e.g. the same language or other sets of prior agreements that make commu-

nication possible), but who does not know the state of our object. The information content of the object is defined as "the size of the set of instructions that our friend requires to be able to reconstruct the object, or better the state of the object.

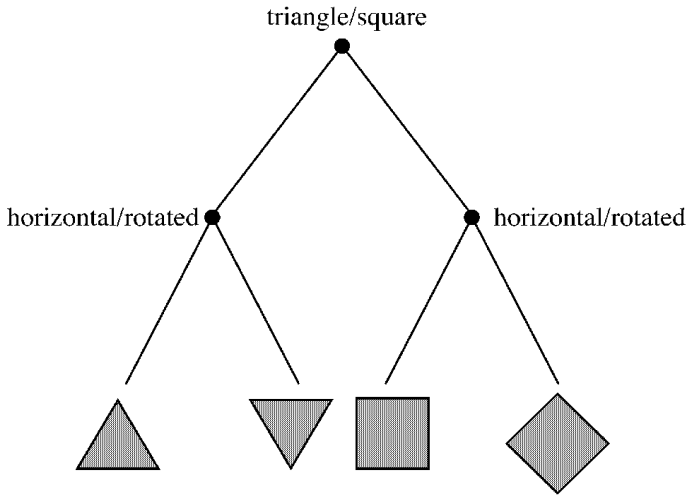

Figure 2.2. Decision tree explaining information content of an object (after [11])

The Figure 2.2.3 displays an example of a decision tree. Two binary choices have to be made to identify the shape (triangle or square) and the orientation (horizontal or rotated). If we send with equal probabilities one of the four objects, two bits of information is transmitted. The information content of an object can easily be obtained by counting the number of equally likely binary choices. In classical information theory, a variable which assumes the values of 0 or 1 equally likely, is called a bit. It can be said that n bits of information can be encoded in a system when instructions in the form of n binary choices needs to be transmitted to identify or regenerate the state of the system.

2.2.4 Shannon Information

The Shannon entropy is a measure of the uncertainty associated with a random variable. For an event X with n outcomes, (x_i, i = 1, 2, 3,..., n) the information

entropy, denoted by H(X) or H($\{p(x_i)\}$), is defined as

$$H(X) = -\sum_{i=1}^{n} p(x_i) log_b p(x_i). \tag{2.13}$$

where p(x_i) is the probability mass function of the outcome x_i, b is the base of the logarithm used. The unit of the information entropy H depends on the value of base b, and is expressed in bits when b= 2 and in nats when b= e. For our purposes we shall use b= 2.

The above definition was evolved for Shannon entropy by imposing three reasonable conditions on the quantitative measure of the "information content of an event". These are

(i) Information is non-negative.

(ii) Least probable events provide the most information.

(iii) Information is additive for independent events.

To understand the relationship between entropy and probability of an event, consider the following experiment. Consider tossing a coin with known (not necessarily equal) probabilities of coming up heads or tails. The entropy of the unknown result of a toss of the coin is maximized if the coin is fair (i.e., heads and tails equally probable). This is the situation of maximum uncertainty, as it is most difficult to predict the outcome of a toss; the result of each toss of the coin delivers a one bit of information. However, if we know the coin is not fair, but comes up heads or tails with probabilities p and q, then there is less uncertainty. Every time, one side is more likely to come up than the other. The reduced uncertainty is reflected in a lower entropy, as on average, each toss of the coin delivers less than a full bit of information. A double-headed coin which never comes up tails is an extreme case in which there is no uncertainty. The entropy is zero and each toss of the coin delivers no information.

The use of the logarithm function allows entropy to follow the first and third condition (assuming entropy vanishes for p(x_i)=0). Suppose there is a set of n mutually exclusive events (a_j, j= 1, 2,..., n) each with equal probability p(a_j)=1/n. The Shannon entropy of this set of events is equal to $log_b n$ units. Consider another set of m mutually exclusive events which are independent from the previous set of events, with the probability of each event given as $1/m$. The Shannon entropy associated with this set is $log_b m$ units. Considering both the sets together, i.e. for the set of mn possible events each with a probability of $1/mn$, the Shannon entropy is $log_b(mn)$=$log_b m$+$log_b n$ units which is the sum of the Shannon entropies of the two independent sets of events.

2.2.4.1 Relative entropy and Shannon Mutual Information

The relative entropy is a measure of the distance between two distributions. For two probability mass functions $p(x)$ and $q(x)$, the relative entropy or the *Kullback Leibler distance* D($p \parallel q$) is defined as

$$D(p \parallel q) = \sum_{x \in X} p(x) log \frac{p(x)}{q(x)} \tag{2.14}$$

The concept of mutual information is now introduced, which is a measure of the amount of information that one random variable contains about another random variable. Consider two random variables X and Y with a joint probability mass function $p(x,y)$ and marginal mass functions $p(x)$ and $p(y)$. The mutual information I(X;Y) is the relative entropy between the joint distribution and the product of the marginal distributions $p(x)p(y)$, i.e.,

$$\text{I(X;Y)= D}(p(x,y) \| p(x)p(y))$$

$$= \sum_{x \in X} \sum_{y \in Y} p(x,y) log \frac{p(x,y)}{p(x)p(y)} \tag{2.15}$$

Using pre-defined quantities like the joint and conditional entropy from joint and conditional probability distributions, the mutual information I(X;Y) can be expressed as

$$\begin{aligned} \text{I(X;Y)} &= \text{H(X)+H(Y)-H(X,Y)} \\ &= \text{H(X)-H(X|Y)} \\ &= \text{H(Y)-H(Y|X)} \end{aligned}$$

where H(X,Y) is the entropy associated with the joint probability distributions of X and Y. H(X|Y) and H(Y|X) are the conditional entropies [4].

$$H(X,Y) = -\sum_{x \epsilon X} \sum_{y \epsilon Y} p(x,y) log_2 p(x,y) \tag{2.16}$$

$$H(Y|X) = -\sum_{x \epsilon X, y \epsilon Y} p(x,y) log_2 \frac{p(x)}{p(x,y)} \tag{2.17}$$

$$H(Y|X) = -\sum_{x \epsilon X, y \epsilon Y} p(x,y) log_2 \frac{p(y)}{p(x,y)} \tag{2.18}$$

.

Mutual information is symmetric i.e.,I(X;Y)=I(Y;X), which implies there is as much information about X in Y as there is information about Y in X. It is bounded as $0 \leq \text{I(X;Y)} \leq \min(\text{H(X)},\text{H(Y)})$. Equality in the lower bound is achieved when X and Y are independent, i.e. when there is no information in Y about X. For $\text{H(X)} \leq \text{H(Y)}$, equality is achieved in the upper bound when Y uniquely determines X. This implies that all information about X is in Y.

Heuristically I(Y;X) can also be viewed as the "entropy in Y that is attributable to X". This must hence be equal to the total entropy(H(Y)) less the entropy that is

not attributable to X (H(Y|X)). Hence I(Y;X)=H(Y)-H(Y|X).

2.2.5 Binary Symmetric Channel

A binary symmetric channel (or BSC) is a common discrete memoryless communications channel model used in coding theory and information theory. In this model, a transmitter wishes to send a bit (a zero or a one), and the receiver receives a bit. It is assumed that the bit is usually transmitted correctly, but that it will be "flipped" with a small probability (the "crossover probability"). This channel is often used by theorists because it is one of the simplest noisy channels to analyze. Many problems in communication theory can be reduced to a BSC. On the other hand, being able to transmit effectively over the BSC can give rise to solutions for more complicated channels.

A binary symmetric channel with crossover probability e denoted by BSC_e, is a channel with binary input and binary output and probability of error e; that is, if X is the transmitted random variable and Y the received variable, then the channel is characterized by the conditional probabilities

Pr(Y=0|X=0)= 1-e

Pr(Y=1|X=0)= e

Pr(Y=0|X=1)= e

Pr(Y=1|X=1)= 1-e

It is assumed that $0 \leq e \leq 0.5$. If $e \geq 0.5$, then the receiver can swap the output (interpret 1 when it sees 0, and vice versa) and obtain an equivalent channel with crossover probability 1-$e \leq 0.5$. The capacity of a channel is defined as the tightest upped bound on the amount of information that can be reliably transmitted over a channel. The capacity of the binary symmetric channel is 1-H(e)(as shown in figure 2.2.5) where H(e) is the binary entropy function which is equal to

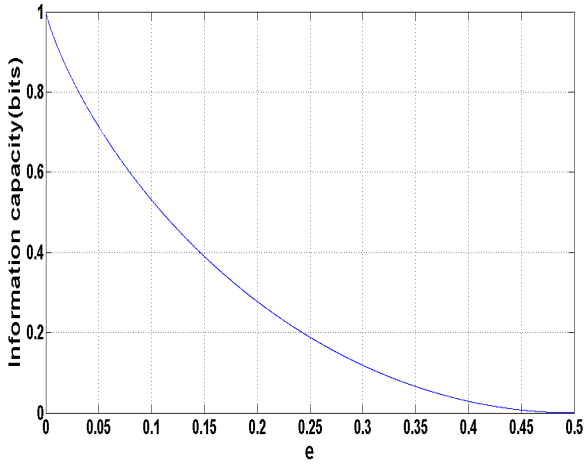

Figure 2.3. Channel capacity vs e

$$H(e) = -e \times log_2(e) - (1-e) \times log_2(1-e) \tag{2.19}$$

CHAPTER 3

TECHNICAL BACKGROUND II: PHYSICAL INFORMATION THEORY-REFERENTIAL APPROACH AND ITS IMPLICATIONS

In the previous chapter, a brief introduction to classical information theory was made. The important quantities such as entropy and mutual information were defined and their properties explored. The quantum equivalent of these classical information theoretic quantities will now be introduced, and the important theorems in physical information theory will be briefly discussed. In this chapter, we shall also develop the Referential approach to information theory and its implications on the lower bounds for entropic and energy costs in information processing physical systems [12], [13], [15]. Noisy computational channels and the need for information theoretic measures to quantify the "how well" a logical operation has been carried out is also discussed [13],[14]. These measures are defined and derived and their relationship to the information loss in a process is stated. In the next chapter, the referential approach is applied to the problem of communicating information using a Hamming code. Required generalizations of the approach were derived and extended to more composite systems.

3.1 Physical Information in Quantum Systems

Information is encoded in the states of classical and quantum systems. In quantum systems,the encoding of information is done by using the quantum state vectors of the system of interest. Yet in most cases, the state vector of a quantum system is either not defined, or not known and only probabilities for various state vectors are

available. In such situations, the density matrix formalism is used. Adaptation of the notion of entropy to the field of quantum mechanics is provided by John von Neumann with the introduction of von Neumann entropy. It is an essential concept in determining the maximum amount of information that can be obtained from a quantum mechanical system. We will now introduce the concept of von Neumann entropy and follow it up with the Holevo information and its significance.

3.1.1 Density Matrix Formalism and Von Neumann Entropy

The density matrix formalism is used in the case where the state vector for a system is not defined or the state vector is not known; only the probabilities of various vectors are known. The density matrix operator is a positive operator with unit trace defined on a complex Hilbert space, which represents the state space of the system.

Consider a quantum system that is known to be in some state from the fixed set $\mid \psi_i\rangle$, where the $\mid \psi_i\rangle$ are normalized but need not be orthogonal. Let p_i indicate the probability that the system is in state $\mid \psi_i\rangle$. The possible states of the system, together with their corresponding probabilities, constitutes an *ensemble of pure states* denoted as $\{p_i,\mid \psi_i\rangle\langle\psi_i \mid\}$. We can associate a density operator

$$\hat{\rho} = \sum_i p_i \mid \psi_i\rangle\langle\psi_i \mid \tag{3.1}$$

with such an ensemble, which acts as the statistical description of the state.
In a more general case, we construct an ensemble of mixed states $\{p_i, \hat{\rho}_i\}$ with the density operator

$$\hat{\rho} = \sum_i p_i \hat{\rho}_i \tag{3.2}$$

with

$$\hat{\rho}_i = \sum_n^N p_n^{(i)} \mid \psi_n^{(i)}\rangle\langle\psi_n^{(i)} \mid \tag{3.3}$$

The von Neumann entropy (or quantum entropy) associated with a density operator $\hat{\rho}$ is

$$S(\hat{\rho}) = -Tr[\hat{\rho} log \hat{\rho}] \tag{3.4}$$

If the density operator $\hat{\rho}$ can be written in the form of a spectral decomposition as

$$\hat{\rho} = \sum_i \lambda_i |\lambda_i\rangle\langle\lambda_i| \tag{3.5}$$

where λ_i and $|\lambda_i\rangle$ are the eigenvalues and eigenvectors respectively. Then $log\hat{\rho}$ is an operator given as

$$\log \hat{\rho} = \sum_i log(\lambda_i) |\lambda_i\rangle\langle\lambda_i| \tag{3.6}$$

S($\hat{\rho}$) maps the density operator into a real number, much as the Shannon entropy H($\{p_i\}$) maps a probability distribution $\{p_i\}$ into a real number. S($\hat{\rho}$) can be most conveniently calculated by solving for the eigenvalues $\{\lambda_i\}$ of $\hat{\rho}$ and applying the result: *the von Neumann entropy of $\hat{\rho}$ is the Shannon entropy of its eigenvalue spectrum.*

$$S(\hat{\rho}) = -\sum_i \lambda_i log_2 \lambda_i \tag{3.7}$$

The von Neumann entropy is non-negative and S($\rho = \hat{|\psi\rangle}\langle\psi|$)=0 for pure state $|\psi\rangle$ as the density operator for any pure state has identically one eigenvalue which is λ=1. Furthermore it is bounded as

$$\sum_i p_i S(\hat{\rho}_i) \leq S(\hat{\rho}) \leq H(\{p_i\}) + \sum_i p_i S(\hat{\rho}_i) \tag{3.8}$$

Equality is achieved in the upper bound when the density operators $\hat{\rho}_i$ have support on orthogonal spaces i.e.,

$$\hat{\rho}_i\hat{\rho}_{i\prime}=\delta_{ii'}\hat{\rho_i^2} \ \forall \ i, \ i'$$

For an ensemble of quantum signal states $\epsilon = \{p_i, \hat{\rho}_i\}$, S($\hat{\rho}$) can be thought of as the "entropy of the average signal state", while the quantity $\sum_i p_i S(\hat{\rho}_i)$ represents the "average entropy of the signal states" and H($\{p_i\}$) is the *preparation entropy*, which is the Shannon entropy of the information source driving the state preparation process. The bounds now say that entropy of the average channel state is never less than the average entropy of the channel state and never greater than the average of the channel state plus the *preparation entropy*. For pure signal states this bound reduces to

$$0 \leq S(\hat{\rho}) \leq H(\{p_i\}). \tag{3.9}$$

The bound can be rewritten for $\hat{\rho} = \sum_i p_i\hat{\rho}_i$ as

$$0 \leq \chi(\epsilon) \leq H(\{p_i\}) \tag{3.10}$$

with

$$\chi(\epsilon) = S(\hat{\rho}) - \sum_i p_i S(\hat{\rho}_i) \tag{3.11}$$

which is called the *Holevo information* or sometimes the *entropy defect* for the ensemble $\epsilon = \{p_i, \hat{\rho}_i\}$. $\chi(\epsilon)$ and I(X;Y) have many similar properties. These include

•$\chi(\epsilon)$ and I(Y:X) have the same bounds. They are both ≥ 0 and upper bounded by the Shannon entropy of the source.

•$\chi(\epsilon)$ is the entropy of the average channel state less the average entropy of the channel state, while I(Y;X) is the entropy of the average output distribution less the average entropy of the output distribution [10].

3.2 Decoherence

Decoherence is a process through which superposition states of a quantum system are reduced to mixtures of orthogonal states in some particular basis. It requires a certain type of interaction of the system with its environment and provides an explanation on why measurements made on quantum system yield classical results. Decoherence coverts the initial pure state of a system S into a mixture of orthogonal eigenstates, increasing the von Neumann entropy from $S(\hat{\rho}^{(S)})$=0 to $S(\hat{\rho}^{(S)'}) = H(\{q_j\})$. The transformation

$$|\psi\rangle\langle\psi| \longrightarrow \sum_j q_j |a_j\rangle\langle a_j|$$

is an example of decoherence [10].

The orthogonal states into which the system is reduced by decoherence are called pointer basis. The probability of obtaining $|a_j\rangle\langle a_j|$ is given by $q_j = |\langle a_j|\psi\rangle|^2$ and depends upon the initial state of the system. An observable $\hat{\Lambda}^S$'s eigenvectors will constitute the pointer basis emerging from decoherence-for evolutions dominated by interactions $\hat{H_I}^{S\varepsilon}$ that commute with $\hat{\Lambda}^S$:

$$\left[\hat{\Lambda}^S, \hat{H_I}^{S\varepsilon}\right] = 0.$$

The following formulation will provide a better understanding of the idea. Let the initial state of the system $|\psi\rangle$ be a superposition of the energy eigenstates

$$|\psi\rangle = \sum_j c_j |E_j\rangle \tag{3.12}$$

with $c_j = \langle R_j|\psi\rangle$. The corresponding density operator is

$$\hat{\rho}^S = \sum_j \sum_{j'} c_j c_{j'}^* |E_j\rangle\langle E_{j'}|. \tag{3.13}$$

This can be written as

$$\hat{\rho}^S = \sum_j q_j |E_j\rangle\langle E_j| + \sum_j \sum_{j' \neq j} c_j c_{j'}^* |E_j\rangle\langle E_{j'}| \tag{3.14}$$

Comparing with the final density operator

$$\hat{\rho}^{S'} = \sum_j q_j |E_j\rangle\langle E_j| \tag{3.15}$$

reveals that the environmental interaction has only left the diagonal terms in the density operator and eliminated all off-diagonal terms. Since the off-diagonal terms are called quantum coherences, any process which removes these terms is classified as decoherence.

From the derivations in [10], we know that decoherence is never complete, but for realistic interactions and large environments, the quantum coherences nearly vanish after an extremely short time, which is referred to as the decoherence time. This is usually orders of magnitude shorter than the time scales associated with the other active dynamical processes and allows us to assume that decoherence is instantaneous and complete. A much deeper analysis is available in [17].

3.3 Landauer's Principle

The principle was first argued by Rolf Landauer and perhaps best restated by Bennett is [8] which states that "any logically irreversible manipulation of informa-

tion, such as the erasure of a bit or the merging of two computation paths, must be accompanied by a corresponding entropy increase in non-information bearing degrees of freedom of the information processing apparatus or its environment". The principle allows us to relate thermodynamical quantities to the amount of information associated with the system. Specifically,

$$\Delta S \geq -k_B \ln(2) \Delta I \tag{3.16}$$

$$\Delta E \geq -k_B T \ln(2) \Delta I \tag{3.17}$$

where k_B is the Boltzmann's constant, T is the absolute temperature of the environment and $-\Delta I$ is the amount of information lost in an operation. The first form is called the "entropic form" and associated a minimum entropy increase (in thermodynamic units) of $k_B \ln(2)$ per bit of information lost in the operation. The second form is called the "energetic form" associated a minimum energy of $k_B T \ln(2)$ per bit of information lost.

The very same inequalities arise in a wide variety of scenarios and definitions for relevant quantities. However, a common feature amongst the various scenarios that lead to these two inequalities assume the loss of information from a physical system as a state transformation that reduces uncertainty in the system state, as quantified by a self-referential information measure defined in the terms of the state of the system undergoing the information loss. In the next chapter, we shall introduce the Referential approach where quantities are described with respect to a referent which remains unchanged during the process of information loss.

3.4 Referential Approach to Physical Information Theory

3.4.1 "Information is always about something else!!"

The referential approach [12], [13], [15] is based on the idea that information is always a measure of correlation between the system and a referent. The approach has many significant advantages, one being the clear divide between the entropic self information of a system and the mutual information of a system with a referent. Since quantum mutual information is defined between two different systems, it cannot be defined between the density operators of the same system at two different time instants. However the referential approach allows the calculation of information loss in the system over time with respect to an unchanged referent. Furthermore, in terms of engineering applications in computing systems, the approach proves to be very beneficial, as the information we manipulate and perform operations upon are usually physical encodings of input information which is present in another location, for example the memory which can act as our referent. Information stored in the memory are not changed until the computation using them is completed and hence are perfect for providing an unchanged referent. Since such memory elements like flip-flops and latches that provide storage capabilities are used in abundance in the intermediate stages of multiple cycle calculations, analyzing such processes using the referential approach can provide crucial insight. Thus the referential approach to physical information theory must be explored in detail to reap its full benefits.

3.4.2 Logical Irreversibility and Information Loss

Before we discuss the framework for studying physical systems in which information is encoded, and the entropy and energy costs associated with them when logical operations are involved, it is important to understand that not all information loss is unnecessary. In a communication channel, we require that the output has all the information about the input i.e., no information be lost in the channel for perfect

communication. Loss of information is undesirable in this case. However even in ideal computation channels, information is necessarily lost while going from input to output that directly implements a logically irreversible operation. Such irreversible operations include AND, OR, NAND, NOR, etc which form the cornerstones of logical operations that are performed in all general purpose computing (these operations have the property that the number of inputs d is greater than the number of outputs r). Thus using the referential approach, the information about some input referent R that is lost going from input X to output Y is

$$-\Delta I = I(R;X) - I(R;Y) = H(R|Y) - H(R|X) \tag{3.18}$$

Thus for a channel that implements a logically irreversible transformation, it follows that $-\Delta I > 0$. Winograd and Cowan in [9], identified this connection and stated that "the destruction of information" as the defining feature of computation.

> We say that computation occurs if H(X|Y) greater than 0 i.e, if the output symbols do not completely specify the input configurations; and we say that communication occurs if H(X|Y)=0, i.e. if the output symbols completely specify the input configurations...It follows...that computation occurs if H(X) is greater than H(Y),i.e. if information is lost going from X and Y.

3.4.3 Framework and Definitions

Input and Output Ensembles

In order to consider the implementation of a d-input r-output logical transformation $\mathcal{L}$ via evolution of the system $\mathcal{S}$, we must define an $\mathcal{L}$-referent $\mathcal{R}_{\mathcal{L}}$ associated with an d-input r-output logical transformation $\mathcal{L}$. The $\mathcal{L}$-referent consists of

• A bipartite quantum system $\mathcal{R}_{\mathcal{L}} = \mathcal{R}_{in}\mathcal{R}_{out}$.

- A set $\{\hat{r}_i^{\mathcal{R}_{in}}\}$ of d distinguishable pure states of $\mathcal{R}_{in}$.
- A set $\{\hat{r}_i^{\mathcal{R}_{out}}\}$ of r distinguishable pure states of $\mathcal{R}_{out}$.
- A set of $\{\hat{r}_i^{\mathcal{R}_{\mathcal{L}}}\}$ of d product states.

$$\hat{r}_i^{\mathcal{R}_{\mathcal{L}}} = \hat{r}_i^{\mathcal{R}_{in}} \otimes \hat{r}_j^{\mathcal{R}_{out}} \forall i \epsilon \{i\}_j = \{i \mid \mathcal{L}(x_i) = y_j\} \tag{3.19}$$

of $\mathcal{R}_{\mathcal{L}}$, where $\mathcal{L}$ is a logical transformation that maps d logical input states $x_i \epsilon \{x_i\}$ into r logical output states $y_j \epsilon \{y_j\}$ via $x_i \rightarrow \mathcal{L}(x_i) = y_j$. The input referent in most applications will be a real physical system which contains a physical instantiation of the logical input that will remain unchanged till the process of computation is complete. These include the cache, latches and flip-flops in the intermediate stages of a multi-staged logical computation. The output referent is a perfect physical instantiation of the logical outputs of a perfect logical transformation. It need not exist and as the name suggests, it provides a reference to which we can compare our actual physical outputs of the logical transformation.

The input ensemble is given by

$$\epsilon_X^{\mathcal{R}_{\mathcal{L}}\mathcal{S}} = \{p_i, \hat{\rho}_i^{\mathcal{R}_{\mathcal{L}}\mathcal{S}}\} \tag{3.20}$$

where p_i is the probability that $\mathcal{R}_{\mathcal{L}}\mathcal{S}$ is initially prepared in the state $\hat{\rho}_i^{\mathcal{R}_{\mathcal{L}}\mathcal{S}} = \hat{r_i}^{\mathcal{R}_{\mathcal{L}}\mathcal{S}} \otimes \hat{\rho_i^{\mathcal{S}}}$ corresponding to the i-th logical input x_i. The density operator describing the statistical state of this ensemble is

$$\hat{\rho}^{\mathcal{R}_{\mathcal{L}}\mathcal{S}} = \sum_{i=1}^{d} p_i \hat{\rho}_i^{\mathcal{R}_{\mathcal{L}}\mathcal{S}} \tag{3.21}$$

In order to obtain the output ensemble, all the members of the input ensemble must be evolved via $\hat{\mathcal{B}}$, a quantum operation (which is a linear, completely positive

map from the set of density operators into itself) to obtain the evolved input ensemble

$$\epsilon_X^{\mathcal{R}_\mathcal{L}\mathcal{S}'} = \{p_i, \hat{\rho}_i^{\mathcal{R}_\mathcal{L}\mathcal{S}'}\} \tag{3.22}$$

where $\hat{\rho}_i^{\mathcal{R}_\mathcal{L}\mathcal{S}'} = \hat{r}_i^{\mathcal{R}_\mathcal{L}} \otimes \mathcal{B}(\hat{\rho}_i^{\mathcal{S}})$. The elements of the output ensemble

$$\epsilon_Y^{\mathcal{R}_\mathcal{L}\mathcal{S}'} = \{q_j, \hat{\rho}_j^{\mathcal{R}_\mathcal{L}\mathcal{S}'}\} \tag{3.23}$$

can then be projected out of the statistical state

$$\hat{\rho}^{\mathcal{R}_\mathcal{L}\mathcal{S}'} = \sum_{i=1}^{d} p_i \hat{\rho}_i^{\mathcal{R}_\mathcal{L}\mathcal{S}'} \tag{3.24}$$

of the evolved input ensemble as

$$\hat{\rho}_j^{\mathcal{R}_\mathcal{L}\mathcal{S}'} = \frac{1}{q_j} \hat{\Pi}_j^{\mathcal{R}_\mathcal{L}\mathcal{S}} \hat{\rho}^{\mathcal{R}_\mathcal{L}\mathcal{S}'} \hat{\Pi}_j^{\mathcal{R}_\mathcal{L}\mathcal{S}} = \sum_{i \in \{i\}_j} p_i^{(j)} \hat{\rho}_i^{\mathcal{R}_\mathcal{L}\mathcal{S}'} \tag{3.25}$$

where $\hat{\Pi}_j^{\mathcal{R}_\mathcal{L}\mathcal{S}}$ is the projector associated with the j-th logical output and is given by

$$\hat{\Pi}_j^{\mathcal{R}_\mathcal{L}\mathcal{S}} = \sum_{i \epsilon \{i\}_j} \hat{\pi}_i^{\mathcal{R}_\mathcal{L}\mathcal{S}} \tag{3.26}$$

on $H^{\mathcal{R}_\mathcal{L}} \bigotimes H^{\mathcal{S}}$ with $\hat{\pi}_i^{\mathcal{R}_\mathcal{L}\mathcal{S}} = \hat{r}_i^{\mathcal{R}_\mathcal{L}} \bigotimes \hat{\pi}_i^{\mathcal{S}}$, where $\hat{\pi}_i^{\mathcal{S}}$ is the identity for the support of $\mathcal{B}(\hat{\rho}_i^{\mathcal{S}_{(in)}})$, and $\Pi_j^{\mathcal{R}_\mathcal{L}\mathcal{S}}$ is the identity for the support subspace associated with the j-th output.

Also we have $p_i^{(j)} = \frac{p_i}{q_j}$,and as expected

$$q_j = Tr[\hat{\Pi}_j^{\mathcal{R}_\mathcal{L}\mathcal{S}} \hat{\rho}^{\mathcal{R}_\mathcal{L}\mathcal{S}'}] = \sum_{i \in \{i\}_j} p_i \tag{3.27}$$

We then define N output ensembles

$$\epsilon_j^{\mathcal{R}_\mathcal{L}\mathcal{S}'} = \{p_i^{(j)}, \hat{\rho}_i^{\mathcal{R}_\mathcal{L}\mathcal{S}'} \mid i \in \{i\}_j\} \tag{3.28}$$

associated with the r logical outputs, and the j-th reduced density operator is given by

$$\hat{\rho}_j^{\mathcal{S}'} = Tr_{\mathcal{R}_\mathcal{L}}[\hat{\rho}_j^{\mathcal{R}_\mathcal{L}\mathcal{S}'}] = \sum_{i \in \{i\}_j} p_j^{(i)} \mathcal{B}(\hat{\rho}_i^{(in)}) \tag{3.29}$$

$\hat{\rho}_j^{\mathcal{S}'}$ is the physical representation of the j-th output stage or y_j.

The j-th reduced density operator of the system is

$$\hat{\rho}_j^{\mathcal{S}'} = Tr_{\mathcal{R}_\mathcal{L}}[\hat{\rho}_j^{\mathcal{R}_\mathcal{L}\mathcal{S}'}] \tag{3.30}$$

This provides a statistical representation of the logical output-y_j for input distribution $\{p_i\}$ in the device state S alone.

3.4.4 Information Processing and Corresponding Entropic and Energy Cost

Consider a closed composite system consisting of an "information bearing" subsystem RS and environment ε. Let the states of $\mathcal{R}$ and $\mathcal{S}$ be initially correlated and assume that $\mathcal{RS}$ is initially isolated from ε. Initial state of the global system is

$$\hat{\rho} = \hat{\rho}^{\mathcal{RS}} \otimes \hat{\rho}^{\hat{\varepsilon}} \tag{3.31}$$

and the quantum mutual information between $\mathcal{R}$ and $\mathcal{S}$ is

$$S(\hat{\rho}^{\mathcal{R}};\hat{\rho}^{\mathcal{S}}) = S(\hat{\rho}^{\mathcal{R}}) + S(\hat{\rho}^{\mathcal{S}}) - S(\hat{\rho}^{\mathcal{RS}}) \tag{3.32}$$

The initial total entropy is given as

$$S_{tot}(\hat{\rho}) = k_B \ln(2)[S_{tot}(\hat{\rho}^{\mathcal{RS}}) + S_{tot}(\hat{\rho}^{\varepsilon})] \tag{3.33}$$

3.4.4.1 Information Processing

An operation processing information about $\mathcal{R}$ which is encoded in $\mathcal{S}$ is given as an unitary evolution of $\mathcal{RS}\varepsilon$ that involves only interactions between $\mathcal{S}$ and ε.

$$\hat{\rho}' = \hat{\mathcal{U}}\hat{\rho}\hat{\mathcal{U}}^{\dagger} \tag{3.34}$$

for which

$$\hat{\mathcal{U}} = \hat{\mathcal{U}}^{\mathcal{R}} \otimes \hat{\mathcal{U}}^{\mathcal{S}\varepsilon} \tag{3.35}$$

The interactions between $\mathcal{S}$ and ε will generally decrease the correlations between $\mathcal{R}$ and $\mathcal{S}$. Thus information about $\mathcal{R}$ is lost in $\mathcal{S}$ during the operation.
Final quantum mutual information between $\mathcal{R}$ and $\mathcal{S}$ is

$$S(\hat{\rho}^{\mathcal{R}'};\hat{\rho}^{\mathcal{S}'}) = S(\hat{\rho}^{\mathcal{R}'}) + S(\hat{\rho}^{\mathcal{S}'}) - S(\hat{\rho}^{\mathcal{RS}'}) \tag{3.36}$$

and final total entropy is

$$S_{tot}(\hat{\rho}') = k_B \ln(2)[S(\hat{\rho}^{\mathcal{RS}'}) + S(\hat{\rho}^{\varepsilon'})] \tag{3.37}$$

.

3.4.4.2 Information Loss and Change in Entropy

The change in total entropy during the information processing operation

$$\Delta S = S_{tot}(\hat{\rho}') - S_{tot}(\hat{\rho}) \tag{3.38}$$

The change in quantum mutual information is given by

$$\Delta I \equiv S(\hat{\rho}^{\mathcal{R}'}; \hat{\rho}^{\mathcal{S}'}) - S(\hat{\rho}^{\mathcal{R}}; \hat{\rho}^{\mathcal{S}}) \tag{3.39}$$

From Anderson [14], using these definitions we can show that

$$\Delta S \geq -k_B \ln(2) \Delta I \tag{3.40}$$

3.4.4.3 Information Loss and Energy Flow

In order to study the energy costs of operations that discard information, like irreversible logical operations, it is assumed that the environment is initially a thermal bath at temperature T. Thus the initial state of ε is described by the canonical density operator

$$\hat{\rho}^{\varepsilon} = Z^{-1} exp\left(-\frac{\hat{H}_{\varepsilon}}{k_B T}\right) \tag{3.41}$$

where $\hat{H}_{\varepsilon}$ is the Hamiltonian of the bath, T is the bath temperature, and

$$Z = Tr\left[exp\left(-\frac{\hat{H}_{\varepsilon}}{k_B T}\right)\right] \tag{3.42}$$

is the partition function. The expected energy increase in the environment engendered by information loss is

$$\Delta\langle E\rangle^{\varepsilon} = \langle E^{\varepsilon'}\rangle - \langle E^{\varepsilon}\rangle = Tr[\hat{\rho}^{\varepsilon'}\hat{H}_{\varepsilon}] - Tr[\hat{\rho}^{\varepsilon}\hat{H}_{\varepsilon}] \tag{3.43}$$

Consider the quantity

$$\Delta\langle E^{\varepsilon}\rangle - T\Delta S^{\varepsilon} \text{ where } \Delta S^{\varepsilon} = S(\hat{\rho}^{\varepsilon'}) - S(\hat{\rho}^{\varepsilon})$$

Following the derivation in [13], and using

$$\ln \hat{\rho}^{\varepsilon} = -\frac{\hat{H}^{\varepsilon}}{k_B T} - \ln Z \tag{3.44}$$

we get

$$\Delta\langle E^{\varepsilon}\rangle - T\Delta S^{\varepsilon} = k_B T(Tr[\hat{\rho}^{\varepsilon'} \ln \hat{\rho}^{\varepsilon'}] - Tr[\hat{\rho}^{\varepsilon'} \ln \hat{\rho}^{\varepsilon}]) \tag{3.45}$$

which is the relative entropy between initial and final environment states. Since relative entropy is nonnegative for any two density operators, we obtain the inequality

$$\Delta\langle E^{\varepsilon}\rangle \geq T\Delta S^{\varepsilon} \tag{3.46}$$

From the entropic derivation of Landauer's Principle, it is known

$$\Delta S = \Delta S^{\mathcal{RS}} + \Delta S^{\varepsilon} \geq -k_B \ln(2)\Delta I \tag{3.47}$$

$$\Delta S^{\varepsilon} \geq -k_B \ln 2[\Delta I + \Delta S^{\mathcal{RS}}] \tag{3.48}$$

Substituting into Eq. (3.46), we get

$$\Delta\langle E^{\varepsilon}\rangle \geq -k_B T \ln 2[\Delta I + \Delta S^{\mathcal{RS}}] \tag{3.49}$$

Since we know

$$\Delta I + \Delta S^{\mathcal{RS}} = \Delta S^{\mathcal{S}} = S(\hat{\rho}^{\mathcal{S}'}) - S(\hat{\rho}^{\mathcal{S}}) \tag{3.50}$$

this gives

$$\Delta\langle E^{\varepsilon}\rangle \geq -k_B T \ln(2) \Delta S^{\mathcal{S}} \tag{3.51}$$

This inequality implies that there is a minimum environmental energy increase of $k_B T \ln(2)$ associated with every operation that reduces the system entropy $\Delta S^{\mathcal{S}}$ by 1 bit, regardless of how much information is lost. The bound thus accommodates scenarios in which entropy of $\mathcal{S}$ is increased and energy is transferred out of the environment during processes that cause loss of information. This stands in contrast with the traditional form of Landauer's Principle which associates a energy transfer into the environment with loss of information.

3.4.5 Noisy Computation Channels

A d-input, r-output discrete channel with $0< q_{j/i} <1$ for at least one of the outputs y_j, cannot be associated with the implementation of any logical transformation, since direct implementation requires that each x_i map into one and only one output y_j and this requirement is not met if $0< q_{j/i} <1$ for any $q_{j/i}$.

Thus rather than trying to answer the question "what logical transformation $\mathcal{L}$ is implemented by the noisy channel", we should try and answer the question "How well does the noisy computational channel implement the logical transformation $\mathcal{L}$". The information theoretic efficacy measures from [13] capture and quantify this and the relationship of these measures to the information loss is also studied.

3.4.5.1 Representational Faithfulness

For the computational channel to "complete the work" of implementing a logical transformation L, then all device input states "belonging to" the same logical output of $\mathcal{L}$ must evolve into the same device output state $\mathcal{U}_{\mathcal{L}}(\mathcal{S}_i^{(in)}) = \mathcal{S}_j^{(out)} \forall i \in \{i\}_j = \{i \mid \mathcal{L}(x_i) = y_j\}$. This condition requires that the evolved states should contain no information that could help identify the state $\mathcal{S}_i^{(in)} \in \{\mathcal{S}_i^{(in)}\}_j$ from which it is evolved. This implies

$$I(\hat{\rho}_j^{\mathcal{R}_{in}}; \hat{\rho}_j^{\mathcal{S}'}) = \chi(\epsilon_j^{\mathcal{S}'}) = 0. \tag{3.52}$$

From this, the following definition of representational faithfulness can be developed [13],[10].

> **Definition** For a quantum machine that implements a logical transformation $\mathcal{L}$ and input distribution $\{p_i\}$, the *representational faithfulness* is
>
> $$f_{\mathcal{L}} \equiv 1 - \frac{1}{H_{\mathcal{L}}(X/Y)} \sum_{j=1}^{N} q_j \chi(\epsilon_j^{\mathcal{S}'}) \tag{3.53}$$
>
> where q_j and $H_{\mathcal{L}}(X/Y)$ are the j-th output probability and the conditional entropy associated with the logical transformation $\mathcal{L}$ for input distribution $\{p_i\}$ and $\chi(\epsilon_j^{\mathcal{S}'})$ is the Holevo information associated with the ensemble $\epsilon_j^{\mathcal{S}'} = \{p_i^{(j)}, \hat{\rho}_i^{\mathcal{S}'} \mid i \in \{i\}_j\}$ of final reduced device states $\hat{\rho}_i^{\mathcal{S}'}$ representing the logical output states y_j of $\mathcal{L}$.

$f_{\mathcal{L}} H_{\mathcal{L}}(X/Y)$ is the average over all logical outputs, of information about the logical input that is lost in producing the physical representations of the logical outputs. It is bounded as $0 \geq f_{\mathcal{L}} \geq 1$.

3.4.5.2 Computational Fidelity

This efficacy measure is concerned with the distinguishability of the output states independent of their faithfulness. It is related to the amount of information about

the correct logical output-encoded in output referent states- that is reflected in the final physical state of $\mathcal{S}$,i.e by the quantum mutual information

$$I(\hat{\rho}^{\mathcal{R}_{out}};\hat{\rho}^{\mathcal{S}'}) = S(\hat{\rho}^{\mathcal{R}_{out}}) + S(\hat{\rho}^{\mathcal{S}'}) - S(\hat{\rho}^{\mathcal{R}_{out}\mathcal{S}'}) = \chi(\epsilon_Y^{\mathcal{S}'}). \tag{3.54}$$

From [13],[10] we obtain the following definition

Definition For a quantum machine implementing the logical transformation $\mathcal{L}$ and input distribution $\{p_i\}$, the *computational fidelity* is

$$F_{\mathcal{L}} \equiv \frac{1}{H_{\mathcal{L}}(Y)}\chi(\epsilon_Y^{\mathcal{S}'}) \tag{3.55}$$

where $H_{\mathcal{L}}(Y)$ is the entropy associated with the logical transformation $\mathcal{L}$ for input distribution $\{p_i\}$ and $\chi(\epsilon_Y^{\mathcal{S}'})$ is the Holevo information associated with the ensemble $\epsilon_{Y\mathcal{S}')=\{q_j,\hat{\rho}_j^{\mathcal{S}'}\}}$ of final device states representing the logical outputs y_j of $\mathcal{L}$.

$F_{\mathcal{L}}H_{\mathcal{L}}(Y)$ indicates the amount of information about the logical output that is present in the final device state. Computational fidelity is bounded as $0 \geq F_{\mathcal{L}} \geq 1$.

3.4.5.3 Information Loss in Terms of Computational Fidelity and Representational Faithfulness

Using mutual information, the information about the logical input X that is lost as the system $\mathcal{S}$ evolves from its initial to final state to implement the logical transformation is

$$-\Delta I \equiv I(\hat{\rho}^{\mathcal{R}_{in}};\hat{\rho}^{\mathcal{S}}) - I(\hat{\rho}^{\mathcal{R}_{in}};\hat{\rho}^{\mathcal{S}'}) \tag{3.56}$$

If $\mathcal{S}$ initially holds all the information about X, since the x_i are encoded in distinguishable input states of $\mathcal{S}$, then $I(\hat{\rho}^{\mathcal{R}_{in}}; \hat{\rho}^{\mathcal{S}})$=H(X) and information loss is

$$-\Delta I = H(X) - \chi(\epsilon_X^{\mathcal{S}'}) \tag{3.57}$$

where $I(\hat{\rho}^{\mathcal{R}_{in}}; \hat{\rho}^{\mathcal{S}'}) = \chi(\epsilon_X^{\mathcal{S}'})$ and since $\chi(\epsilon_X^{\mathcal{S}'}) \leq$H(X), $-\Delta I \geq 0$. Using the definitions for computational fidelity and representational faithfulness, the information loss can be written as(from [13])

$$-\Delta I = f_{\mathcal{L}} H_{\mathcal{L}}(X/Y) + (1 - F_{\mathcal{L}}) H_{\mathcal{L}}(Y) \tag{3.58}$$

The first term indicates the necessary *desirable* information loss that is required to produce faithful representations of logical output states in channels implementing the logical transformation. The second term accounts for the *undesirable* information loss associated with the indistinguishability of the output states. From the equation 3.58, we can clearly see that if the channel flawlessly implements the logical transformation $\mathcal{L}$ i.e. $F_{\mathcal{L}} = 1, f_{\mathcal{L}} = 1$, then $-\Delta I = H_L(X/Y)$ and if a channel that produce unfaithful ($f_{\mathcal{L}} = 0$) yet perfectly distinguishable outputs ($F_{\mathcal{L}} = 1$), information loss $-\Delta I = 0$ which is what is expected in a perfect communication channel.

3.4.5.4 Lower Bound on Energy Dissipation in Terms of Efficacy Measures

In Eq. (3.58), we have related the information loss in a logical transformation L with the efficacy which indicated "how well the logical transformation L was

achieved". Since the information loss is directly related with the heat dissipation to the environment, substituting Eq. (3.58) in Eq. (3.49), we get

$$\Delta\langle E^{\epsilon}\rangle \geq -k_B T \ln 2[f_{\mathcal{L}} H_{\mathcal{L}}(X/Y) + (1 - F_{\mathcal{L}}) H_{\mathcal{L}}(Y) + \langle \Delta S_i^{\mathcal{S}}\rangle] \tag{3.59}$$

where $\langle \Delta S_i^{\mathcal{S}}\rangle$ is the average change in the von Neumann entropies of the system state during the logical transformation $\mathcal{L}$.

We thus have a very important relation between the lower bound on the physical cost the user must pay to achieve a logical transformation, in terms of the performance metrics fidelity and faithfulness which indicate how well the logical transformation was performed.

In the next chapter, the concepts that we have discussed in this chapter will be used to derive the lower bound on energy dissipation associated with performing error-correction on a noisy system encoded using a linear (n,k) code. We will also explain why generalizations to the efficacy measures discussed in this chapter are needed to study the performance in the error-correction case, support the generalization with an illustrative example and extend it to composite systems.

CHAPTER 4

LOWER BOUNDS ON ENERGY DISSIPATION ASSOCIATED WITH ERROR-CORRECTION USING A (N,K) LINEAR BLOCK CODE

The goal of this thesis is to study and determine the lower bound of energy dissipation associated with performing error-correction using a (n,k) linear block code. The preliminary concepts dealing with linear codes and information theory have been outlined in chapter 1. In the previous chapter, physical information theory was explored in detail and the lower bounds on entropic and energy cost associated with a logical transformation L were discussed using the referential approach to physical information theory. Furthermore, computational efficacy measures were described and their relationship to information loss and minimum energy dissipation stated. In this chapter, we will formulate the input and output ensembles for the system when a linear (n,k) block code is used. The input ensemble system experiences noise and is then subjected to the decoding logical transformation to form the output ensemble as indicated in Figure 4.1. At this point, we will discuss the necessary generalizations to the computational efficacy measures, that is required to account for noise in the input states and derive the corresponding relationship to information loss and dissipation costs. The generalized efficacy measures are applied on an illustrative example system, as well extended for composite systems. The generalization of the efficacy measures, application to an example system and their extension to composite systems represent a significant contribution of this thesis.

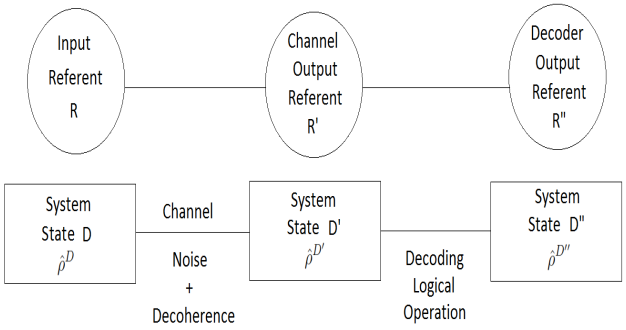

Figure 4.1. Block Diagram indicating the System states as it experiences Noise and is then subject to Decoding, along with the Corresponding Referents

4.1 Formulation

As it is outlined in the previous two chapters, the physical encoding of a logical state x_i in an information bearing system can be regarded as preparation of the system in a quantum state $\hat{\rho}_i$. Consider the n-tuple codeword as being perfectly encoded in the initial state of the system. The input referent $\mathcal{R}$ has 2^k states given by orthogonal density operators $\hat{\rho}_i^{\mathcal{R}}$ with i=0,1,...,2^{k-1}. The input ensemble of the bipartite system $\mathcal{RS}$ is given by $\epsilon_X^{\mathcal{S}} = \{p_i, \hat{\rho}_i^{\mathcal{RS}}\}$ with $\hat{\rho}_i^{\mathcal{RS}} = \hat{\rho}_i^{\mathcal{R}} \otimes \hat{\rho}_i^{\mathcal{S}}$ and $\hat{\rho}_i^{\mathcal{RS}}\hat{\rho}_j^{\mathcal{RS}} = \delta_{ij}\ \forall$ i,j. Since the physical state $\mathcal{S}$ is perfectly correlated to the input referent $\mathcal{R}$

$$\hat{\rho}^{\mathcal{RS}} = \sum_{i=0}^{2^{k-1}} p_i \hat{\rho}_i^{\mathcal{RS}} \tag{4.1}$$

$$I(\hat{\rho}^{\mathcal{R}}; \hat{\rho}^{\mathcal{S}}) = \chi(\epsilon_{\mathcal{R}}^{\mathcal{S}}) = H(\mathcal{R}). \tag{4.2}$$

where $\chi(\epsilon_{\mathcal{R}}^{\mathcal{S}})$ is the Holevo information term which indicates the amount of information the system state $\mathcal{S}$ contains about referent $\mathcal{R}$.

This perfect encoding of the input referent then experiences noise, to form a noisily encoded ensemble i.e the channel output. The noise can be represented using the operator $\hat{\mathcal{N}}$. The input ensemble is evolved to form the noisily encoded ensemble given by $\epsilon_{\mathcal{X}}^{\mathcal{S}'} = \{p_i, \hat{\rho}_i^{\mathcal{R}\mathcal{S}'}\}$ with

$$\hat{\rho}_i^{\mathcal{R}\mathcal{S}'} = \hat{\rho}_i^{\mathcal{R}} \otimes \hat{\mathcal{N}}(\hat{\rho}_i^{S}) \tag{4.3}$$

$$\hat{\rho}_i^{\mathcal{R}\mathcal{S}'} = \hat{\rho}_i^{\mathcal{R}} \otimes \hat{\rho}_i^{\mathcal{S}'} \tag{4.4}$$

$$\hat{\rho}^{\mathcal{R}\mathcal{S}'} = \sum_{i=0}^{2^{k-1}} p_i \hat{\rho}_i^{\mathcal{R}\mathcal{S}'} \tag{4.5}$$

and we have

$$I(\hat{\rho}^{\mathcal{R}}; \hat{\rho}^{\mathcal{S}'}) = \chi(\epsilon_{\mathcal{R}}^{\mathcal{S}'}) \tag{4.6}$$

where $\chi(\epsilon_{\mathcal{R}}^{\mathcal{S}'})$ is the Holevo information term and indicates the amount of information the system state $\mathcal{S}'$ which represents the channel output, contains about the initial referent $\mathcal{R}$.

After the system has been affected by noise, decoding, which is a logical transformation is performed on the system. Thus channel outputs become inputs to the decoding operation. The noisy ensemble is evolved by a operator given by $\hat{\mathcal{U}}$ to give $\epsilon_X^{\mathcal{S}''} = \{p_i, \hat{\rho}_i^{\mathcal{R}\mathcal{S}''}\}$.

$$\hat{\rho}^{\mathcal{R}\mathcal{S}''} = \sum_{i=0}^{2^{k-1}} p_i \hat{\rho}_i^{\mathcal{R}\mathcal{S}''} \tag{4.7}$$

and the quantum mutual information between the system state at the decoder output $\mathcal{S}''$ and referent $\mathcal{R}$ is given by

$$I(\hat{\rho}^{\mathcal{R}}; \hat{\rho}^{\mathcal{S}''}) = \chi(\epsilon_{\mathcal{R}}^{\mathcal{S}''}) \tag{4.8}$$

.

The information loss associated during the process of decoding is

$$-\Delta I_{decoding} = I(\hat{\rho}^{\mathcal{R}}; \hat{\rho}^{\mathcal{S}'}) - I(\hat{\rho}^{\mathcal{R}}; \hat{\rho}^{\mathcal{S}''}) = \chi(\epsilon_{\mathcal{R}}^{\mathcal{S}'}) - \chi(\epsilon_{\mathcal{R}}^{\mathcal{S}''}) \tag{4.9}$$

This can be substituted in the entropic form of Landauer's Principle given by Eq. (3.40) to obtain the lower bound on entropy change during decoding

$$\Delta S^{total} \geq -k_B \ln(2) \Delta I_{decoding} \tag{4.10}$$

$$\Delta S^{total} \geq -k_B \ln(2) [\chi(\epsilon_{\mathcal{R}}^{\mathcal{S}'}) - \chi(\epsilon_{\mathcal{R}}^{\mathcal{S}''})] \tag{4.11}$$

The terms are then substituted into the energy form of Landauer's Principle given by Eq. (3.51)

$$\Delta E^{\varepsilon} \geq -k_B T \ln(2) \Delta S^{\mathcal{S}} \tag{4.12}$$

$$\Delta E^{\varepsilon}_{decoding} \geq k_B T \ln(2) [-\Delta I_{decoding} + \sum_{i=0}^{2^{k-1}} p_i (S(\hat{\rho}_i^{\mathcal{S}'}) - S(\hat{\rho}_i^{\mathcal{S}''})] \tag{4.13}$$

Thus we have established the lower bound on the energy cost of decoding. Since the user will only pay the energy cost associated with decoding, it will be extremely useful to obtain the relationship between the energy dissipation of decoding and the efficacy measures of the logical operation. However the information loss in the above equation cannot be simply expanded using the definitions of computational efficacy explained in the previous chapter. The earlier definitions had assumed that the input states of the logical transformation L were a perfect encoding of the referent and hence contained all the information about the referent. However the inputs to the decoder, which are the outputs of the noisy channel are not so. Since the communication channel can be considered as the computational channel performing the identity operation, the noisy communication channel will output a noisy encoding

of the initial referent states and hence will not be completely correlated to the referent. Thus in order to study the heat dissipation associated with the decoder in terms of its performance measures, there is a need to develop the required tools. Thus we need to generalize the efficacy measures, test it on an example system and extend it to composite systems.

4.2 Development of Generalized Efficacy Measures Required to Study the Decoder Performance

4.2.1 Computational Fidelity Revisited

Recall that the computational fidelity measure is used to quantify the amount of information about the "correct output" of a logical transformation that is in the noisy channel output. For the quantum L-machine and input distribution $\{p_i\}$, the computational fidelity

$$F_{\mathcal{L}} = \frac{\chi(\epsilon_Y^{\mathcal{D}'})}{H_{\mathcal{L}}(Y)} \tag{4.14}$$

where $H_{\mathcal{L}}(Y)$ is the entropy associated with the logical transformation L for input distribution $\{p_i\}$ and $\chi(\varepsilon_Y^{\mathcal{D}'})$ is the Holevo information associated with the ensemble $\epsilon_Y^{\mathcal{D}'} = \{q_j, \hat{\rho}_j^{\mathcal{D}'}\}$ of final device states representing the logical outputs y_j of $\mathcal{L}$.

This definition assumes that the input states of the system are perfectly correlated to the input referent and hence can contain a maximum of $H_{\mathcal{L}}(Y)$ amount of information about the output referent. However if the input states are not perfectly orthogonal to each other, i.e. not perfectly correlated to the input referent, then the earlier definition of fidelity is not valid. In fact, the fidelity is not defined in such a case. With noise and quantum effects being a widespread phenomenon in nanoscale devices, it might be very common that the input states of the system are non-orthogonal and not completely correlated to the input referent. We also have problems when it comes to defining the computational fidelity of an individual stage

in a multistage computation, because here again unless the states of the system at an intermediate stage are perfectly orthogonal, the computational fidelity for the next stage cannot be defined. For example, consider a two staged operation, in which for the first stage 8 inputs are mapped into 4 outputs and in the second stage 4 inputs are mapped into 2 outputs. Since the four output states of the first stage act as inputs to the second, they have to be orthogonal and perfectly correlated to the output referent of the first stage, in order for us to be able to define the fidelity of the second stage. However this need not always happens, as the four output states of the first operation can be non-orthogonal.

Thus we need to generalize to account for the non-orthogonality of the input system states when formulating the computational fidelity measure. In general the maximum amount of information that the input system state $\mathcal{D}$ for a logical operation can have about the output referent Y is given by the Holevo information term $\chi(\epsilon_Y^{\mathcal{D}}) \leq$H(Y), with equality obtained when the states of $\mathcal{D}$ are orthogonal and perfectly correlated to the input referent. Since fidelity is the ratio of the amount of information that the system state $\mathcal{D}'$ has about the output referent Y to the maximum information it could have contained, computational fidelity is defined as

$$F_{\mathcal{L}} = \frac{\chi(\epsilon_Y^{\mathcal{D}'})}{\chi(\epsilon_Y^{\mathcal{D}})} \tag{4.15}$$

As before, fidelity is bounded as $0 \leq F_{\mathcal{L}} \leq 1$ but a fidelity of 1 does not mean that the states of the system are perfectly orthogonal and distinguishable. The new definition indicates the amount of non-orthogonality that has been introduced between input states that produce different logical outputs. Thus a computational fidelity of 1, indicates that the logical transformation did not introduce any more non-orthogonality into the system than what was already present between input states that map into different logical outputs. This new definition, allows us to deal with noisily encoded input states which are very likely in nanoelectronic systems and thus

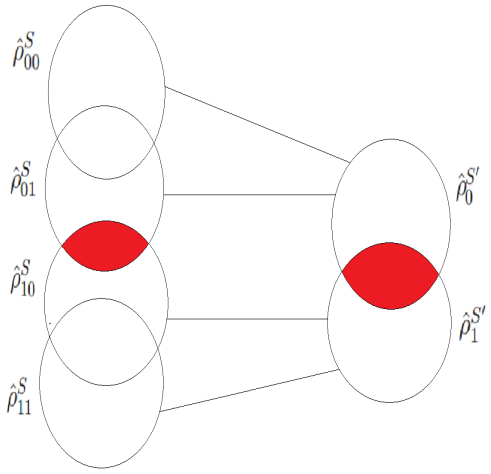

Figure 4.2. Meaning of Computational Fidelity in a 4-input 2-output Logical Operation. The colored region indicates the increase in non-orthogonality between system states that map into different logical output, which is captured by the Computational Fidelity measure.

allow us to deal with a wider range of scenarios, as well define fidelity for the individual stages of a multi-stage logical transformation.

4.2.2 Representational Faithfulness Revisited

Recall that the representational faithfulness is a measure of how well the computational channel "completed the work" of implementing the logical transformation $\mathcal{L}$. For a quantum machine that implements a logical transformation $\mathcal{L}$ and input distribution $\{p_i\}$, the *representational faithfulness* is

$$f_{\mathcal{L}} \equiv 1 - \frac{1}{H_{\mathcal{L}}(X/Y)} \sum_{j=1}^{N} q_j \chi(\epsilon_j^{\mathcal{D}'}) \tag{4.16}$$

where q_j and $H_{\mathcal{L}}(X/Y)$ are the j-th output probability and the conditional entropy associated with the logical transformation $\mathcal{L}$ for input distribution $\{p_i\}$ and $\chi(\epsilon_j^{\mathcal{D}'})$ is the Holevo information associated with the ensemble $\epsilon_j^{\mathcal{D}'} = \{p_i^{(j)}, \hat{\rho}_i^{\mathcal{D}'} \mid i \in \{i\}_j\}$ of final reduced device states $\hat{\rho}_i^{\mathcal{D}'}$ representing the logical output states y_j of $\mathcal{L}$.

For a logical transformation to be faithfully implemented, the evolved states of the device should not contain any information that could help identify the state $\mathcal{D}_i \in \{\mathcal{D}_i\}_j$ from which it is evolved. Quantitatively this is described as

$$I(\hat{\rho}_j^{\mathcal{R}}; \hat{\rho}_j^{\mathcal{D}'}) = 0 \tag{4.17}$$

The maximum value of this term is equal to

$$H_j(X/y_j) = -\sum_{i \in \{i\}_j} p_i^{(j)} log_2 p_i^{(j)}. \tag{4.18}$$

However this is true if and only the referent is perfectly encoded in the initial device states $\mathcal{D}$. If the system state $\mathcal{D}$ is a noisy encoding of the referent $\mathcal{R}$, then the quantity $I(\hat{\rho}_j^{\mathcal{R}}; \hat{\rho}_j^{\mathcal{D}'})$ achieves its maximum value of

$$\chi(\epsilon_j^{\mathcal{D}}) = I(\hat{\rho}_j^{\mathcal{R}}; \hat{\rho}_j^{\mathcal{D}}) \tag{4.19}$$

where $\chi(\epsilon_j^{\mathcal{D}})$ is the Holevo information associated with the ensemble $\epsilon_j^{\mathcal{D}} = \{p_i^{(j)}, \hat{\rho}_i^{\mathcal{D}}\}$. Thus the difference $\chi(\epsilon_j^{\mathcal{D}}) - \chi(\epsilon_j^{\mathcal{D}'})$ can serve as a basis for how well y_j is faithfully represented in $\mathcal{D}$. The output average

$$\sum_j q_j [\chi(\epsilon_j^{\mathcal{D}}) - \chi(\epsilon_j^{\mathcal{D}'})] = \left(1 - \frac{\sum_j q_j \chi(\epsilon_j^{\mathcal{D}'})}{\sum_j q_j \chi(\epsilon_j^{\mathcal{D}})}\right) \sum_j q_j \chi(\epsilon_j^{\mathcal{D}}) \tag{4.20}$$

can hence be used as measure of the representational faithfulness of the L-machine

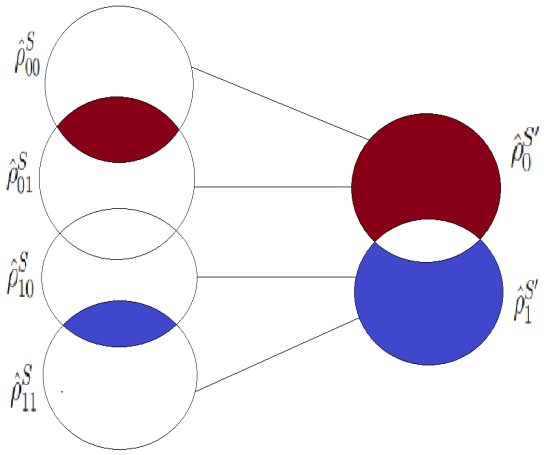

Figure 4.3. Meaning of Representational Faithfulness in a 4-input 2-output Logical Operation. The colored regions indicate the necessary increase in non-orthogonality between system states that map into the same logical output, which is captured by the Representational Faithfulness measure.

as a whole. Thus faithfulness is given as

$$f_{\mathcal{L}} \equiv \left(1 - \frac{\sum_j q_j \chi(\epsilon_j^{\mathcal{D}'})}{\sum_j q_j \chi(\epsilon_j^{\mathcal{D}})}\right). \tag{4.21}$$

As before, representational faithfulness is bounded as $0 \leq f_{\mathcal{L}} \leq 1$. The definition of faithfulness can be interpreted as the amount of necessary non-orthogonality or indistinguishability that is introduced between inputs that map into the same logical output. A faithfulness of 1 indicates that there is no way to distinguish between inputs that map into the same logical output and a faithfulness of zero implies that the inputs are no less indistinguishable than they were before undergoing the logical operation. Like in computational fidelity, the change in definition allowed us to define faithfulness measures for individual stages in a multistage logical computation and also to account for noisily encoded input systems.

4.2.3 Information Loss in Terms of the Efficacy Measures

Let the information loss as device state $\mathcal{D}$, which is a noisy encoding of the input referent $\mathcal{R}$, evolves to device state $\mathcal{D}'$, implementing the logical transformation $\mathcal{L}$ be -$\Delta I_{\mathcal{L}}$, with the output referent being denoted as $\mathcal{R}'$.

$$-\Delta I_{\mathcal{L}} = I(\hat{\rho}^{\mathcal{R}}; \hat{\rho}^{\mathcal{D}}) - I(\hat{\rho}^{\mathcal{R}}; \hat{\rho}^{\mathcal{D}'}) = \chi(\epsilon_{\mathcal{R}}^{\mathcal{D}}) - \chi(\epsilon_{\mathcal{R}}^{\mathcal{D}'}) \tag{4.22}$$

The first term of this equation can be written as

$$\chi(\epsilon_{\mathcal{R}}^{\mathcal{D}}) = \chi(\epsilon_{\mathcal{R}'}^{\mathcal{D}}) + \sum_j q_j \chi(\epsilon_j^{\mathcal{D}}) \tag{4.23}$$

with

$$\chi(\epsilon_{\mathcal{R}'}^{\mathcal{D}}) = S(\hat{\rho}^{\mathcal{D}}) - \sum_j q_j S(\hat{\rho}_j^{\mathcal{D}}) \tag{4.24}$$

$$\chi(\epsilon_j^{\mathcal{D}}) = S(\hat{\rho}_j^{\mathcal{D}}) - \sum_{i \in \{i\}_j} p_i^{(j)} S(\hat{\rho}_i^{\mathcal{D}}) \tag{4.25}$$

Similarly $\chi(\epsilon_{\mathcal{R}}^{\mathcal{D}'})$ can be written as

$$\chi(\epsilon_{\mathcal{R}}^{\mathcal{D}'}) = \chi(\epsilon_{\mathcal{R}'}^{\mathcal{D}'}) + \sum_j q_j \chi(\epsilon_j^{\mathcal{D}'}) \tag{4.26}$$

Substitution into the equation for information loss gives us

$$-\Delta I_{\mathcal{L}} = \chi(\epsilon_{\mathcal{R}'}^{\mathcal{D}}) + \sum_j q_j \chi(\epsilon_j^{\mathcal{D}}) - \chi(\epsilon_{\mathcal{R}'}^{\mathcal{D}'}) - \sum_j q_j \chi(\epsilon_j^{\mathcal{D}'}) \tag{4.27}$$

or

$$-\Delta I_{\mathcal{L}} = \left(1 - \frac{\chi(\epsilon^{\mathcal{D}'}_{\mathcal{R}'})}{\chi(\epsilon^{\mathcal{D}}_{\mathcal{R}'})}\right)\chi(\epsilon^{\mathcal{D}}_{\mathcal{R}'}) + \left[1 - \frac{\sum_j q_j \chi(\epsilon^{\mathcal{D}'}_j)}{\sum_j q_j \chi(\epsilon^{\mathcal{D}}_j)}\right]\sum_j q_j \chi(\epsilon^{\mathcal{D}}_j). \tag{4.28}$$

Using the modified definitions of computational fidelity and representational faithfulness defined in the previous subsections, the information loss can written as

$$-\Delta I_{\mathcal{L}} = (1 - F_{\mathcal{L}})\chi(\epsilon^{\mathcal{D}}_{\mathcal{R}'}) + f_{\mathcal{L}} \sum_j q_j \chi(\epsilon^{\mathcal{D}}_j) \tag{4.29}$$

As indicated in [13] and in the earlier chapter, the information loss is resolved into components related to faithful representation and computational fidelity. The first term accounts for the *undesirable* information loss associated with the amount of indistinguishability introduced between input states that map into different logical output states. The term vanishes for $F_{\mathcal{L}}$=1, that is when there is no more non-orthogonality between input states that map into different logical output states than there was before initially. The second term accounts for the *desirable* information loss that is required for faithful representations of logical outputs states in a logical transformation. It achieves its maximum value of $\sum_j q_j \chi(\epsilon^{\mathcal{D}}_j)$ for $f_{\mathcal{L}}$=1 and vanishes for completely unfaithful representations, i.e $f_{\mathcal{L}}$=0.

Thus Eq. (4.29) relates the information loss associated with the logical transformation when the input states are not perfectly correlated to the input referent, with the efficacy measures. The equation can be further substituted in Eq. (3.49) to obtain the heat dissipation to the environment:

$$\Delta\langle E^{\varepsilon}\rangle \geq -k_B T \ln(2) \left[(1 - F_{\mathcal{L}})\chi(\epsilon^{\mathcal{D}}_{\mathcal{R}'}) + f_{\mathcal{L}} \sum_j q_j \chi(\epsilon^{\mathcal{D}}_j) + \langle \Delta S^{\mathcal{D}}_i \rangle\right] \tag{4.30}$$

.

The above equation would enable us to understand the relation between the performance metrics of a logical transformation when the input states are themselves noisy and the heat dissipation associated with performing the logical transformation on these inputs which are not perfectly correlated to the input referent. This allows us to deal with a much wider range of scenarios that are more likely to occur due to noise and quantum effects present in nanoelectronic systems.

4.2.4 Generalized Computational Fidelity and Representational Faithfulness: An Illustrative Example

In the following section, we will apply the generalized computational fidelity and representational faithfulness formulation to access how well classical binary addition is implemented by a quantum controlled-NOT gate that is subject to decoherence. The example system used is borrowed from [14] and will help us study the effect of input state indistinguishability and environmental interactions on the fidelity and faithfulness of the logical operation. The quantum CNOT gate is used for the very fact that it is not designed to perform binary addition and hence we will be able to determine how well it did.

Setup

The controlled-NOT gate or CNOT gate is a quantum gate that flips the second target qubit, if and only if the first control qubit is 1. The logical operation L is the classical binary addition, which maps logical inputs $x_i = A_i B_i$ into logical outputs $y_j = C_j S_j$ as

$$x_0 = 00 \longrightarrow y_0 = 00$$
$$x_1 = 01 \longrightarrow y_1 = 01$$
$$x_2 = 10 \longrightarrow y_1 = 01$$
$$x_3 = 11 \longrightarrow y_2 = 10$$

where A_i and B_i are the input(summand) bits and C_j and S_j are the carry(C=AB) and sum (S=A$\oplus$B) output bits respectively.

The logical input $x_i = A_i B_i$ is physically encoded into the initial state of $\mathcal{AB}$ as

$$\hat{\rho}_i^X = |x_i^{\mathcal{AB}}\rangle\langle x_i^{\mathcal{AB}}| \tag{4.31}$$

where

$$|x_i^{\mathcal{AB}}\rangle = |\psi_{A_i}^{\mathcal{A}}\rangle \otimes |\psi_{B_i}^{\mathcal{B}}\rangle \tag{4.32}$$

with

$$|\psi_0^{\mathcal{A}}\rangle = cos\theta|0^{\mathcal{A}}\rangle + sin\theta|1^{\mathcal{A}}\rangle$$

$$|\psi_0^{\mathcal{A}}\rangle = sin\theta|0^{\mathcal{A}}\rangle + cos\theta|1^{\mathcal{A}}\rangle$$

and $|\psi_0^{\mathcal{B}}\rangle$ and $|\psi_1^{\mathcal{B}}\rangle$ defined similarly. The parameter θ controls the indistinguishability of the encoding states $\hat{\rho}_i^{\mathcal{X}}$ corresponding to the four logical states x_i. For $\theta = 0$, the states are orthogonal and perfectly distinguishable and for $\theta = \frac{\pi}{4}$, the states are identical and perfectly indistinguishable. The input distribution $\{p_i\}$ is uniform with $p_i = \frac{1}{4}$ for all i. The input density matrices are represented as shown below

$$\hat{\rho}_i^{\mathcal{X}} = \begin{bmatrix} \langle 00|\hat{\rho}_i|00\rangle & \langle 00|\hat{\rho}_i|01\rangle & \langle 00|\hat{\rho}_i|10\rangle & \langle 00|\hat{\rho}_i|11\rangle \\ \langle 01|\hat{\rho}_i|00\rangle & \langle 01|\hat{\rho}_i|01\rangle & \langle 01|\hat{\rho}_i|10\rangle & \langle 01|\hat{\rho}_i|11\rangle \\ \langle 10|\hat{\rho}_i|00\rangle & \langle 10|\hat{\rho}_i|01\rangle & \langle 10|\hat{\rho}_i|10\rangle & \langle 10|\hat{\rho}_i|11\rangle \\ \langle 11|\hat{\rho}_i|00\rangle & \langle 11|\hat{\rho}_i|01\rangle & \langle 11|\hat{\rho}_i|10\rangle & \langle 11|\hat{\rho}_i|11\rangle \end{bmatrix} \tag{4.33}$$

$$\hat{\rho}_0^{\mathcal{X}} = \begin{bmatrix} cos^4\theta & cos^3\theta sin\theta & cos^3\theta sin\theta & sin^2\theta cos^2\theta \\ cos^3\theta sin\theta & sin^2\theta cos^2\theta & sin^2\theta cos^2\theta & sin^3\theta cos\theta \\ cos^3\theta sin\theta & sin^2\theta cos^2\theta & sin^2\theta cos^2\theta & sin^3\theta cos\theta \\ sin^2\theta cos^2\theta & sin^3\theta cos\theta & sin^3\theta cos\theta & sin^4\theta \end{bmatrix} \tag{4.34}$$

$$\hat{\rho}_1^{\mathcal{X}} = \begin{bmatrix} sin^2\theta cos^2\theta & cos^3\theta sin\theta & sin^2\theta cos^2\theta & sin^3\theta cos\theta \\ cos^3\theta sin\theta & cos^4\theta & cos^3\theta sin\theta & cos^2\theta sin^2\theta \\ sin^2\theta cos^2\theta & cos^3\theta sin\theta & sin^2\theta cos^2\theta & sin^3\theta cos\theta \\ sin^3\theta cos\theta & sin^2\theta cos^2\theta & sin^3\theta cos\theta & sin^4\theta \end{bmatrix} \tag{4.35}$$

$$\hat{\rho}_2^{\mathcal{X}} = \begin{bmatrix} sin^2\theta cos^2\theta & sin^3\theta cos\theta & sin^2\theta cos^2\theta & cos^3\theta sin\theta \\ sin^3\theta cos\theta & sin^4\theta & sin^3\theta cos\theta & sin^2\theta cos^2\theta \\ sin^2\theta cos^2\theta & sin^3\theta cos\theta & sin^2\theta cos^2\theta & cos^3\theta sin\theta \\ cos^3\theta sin\theta & sin^2\theta cos^2\theta & cos^3\theta sin\theta & cos^4\theta \end{bmatrix} \tag{4.36}$$

$$\hat{\rho}_3^{\mathcal{X}} = \begin{bmatrix} sin^4\theta & sin^3\theta cos\theta & sin^3\theta cos\theta & sin^2\theta cos^2\theta \\ sin^3\theta cos\theta & sin^2\theta cos^2\theta & sin^2\theta cos^2\theta & cos^3\theta sin\theta \\ sin^3\theta cos\theta & sin^2\theta cos^2\theta & sin^2\theta cos^2\theta & cos^3\theta sin\theta \\ sin^2\theta cos^2\theta & cos^3\theta sin\theta & cos^3\theta sin\theta & cos^4\theta \end{bmatrix} \tag{4.37}$$

The quantum CNOT gate maps the initial state $\hat{\rho}^{\mathcal{AB}}$ of the two qubit system $\mathcal{AB}$ into a final state $\hat{\rho}^{\mathcal{AB}'}$ via the unitary transformation

$$\hat{\rho}^{\mathcal{AB}'} = \hat{\mathcal{U}}\hat{\rho}^{\mathcal{AB}}\hat{\mathcal{U}}^{\dagger} \tag{4.38}$$

with

$$\hat{\mathcal{U}} = |00\rangle\langle 00| + |01\rangle\langle 01| + |11\rangle\langle 10| + |10\rangle\langle 11| \tag{4.39}$$

$|01\rangle$ for example indicates the vector $|0^{\mathcal{A}}\rangle \otimes |1^{\mathcal{B}}\rangle$, where $\{|0^{\mathcal{A}}\rangle, |1^{\mathcal{A}}\rangle\}$ and $\{|0^{\mathcal{B}}\rangle, |1^{\mathcal{B}}\rangle\}$

are basis sets spanning the Hilbert spaces H^A and H^B of the control and target qubits respectively. Considering a noisy CNOT operation that, with probability P, decoheres $\hat{\rho}^{\mathcal{AB}'}$ into a mixture $\sum_{k=0}^{3} \hat{\Pi}_k \hat{\rho}^{\mathcal{AB}'} \hat{\Pi}_k$ of orthogonal pointer states $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$, where for example $\hat{\Pi}_1 = |01\rangle\langle 01|$. The output density operator is thus of the form

$$\hat{\rho}^{\mathcal{Y}'} = (1-P)\hat{\rho}^{\mathcal{AB}'} + P \sum_{k=0}^{3} \hat{\Pi}_k \hat{\rho}^{\mathcal{AB}'} \hat{\Pi}_k \tag{4.40}$$

Note that the input and output states are denoted as $\hat{\rho}_i^{\mathcal{X}'}$ and $\hat{\rho}_j^{\mathcal{Y}'}$ respectively to maintain consistency with the notations used in the previous sections, even though the system is $\mathcal{AB}$ in both input and output. The output states are obtained from $\hat{\rho}_i^{\mathcal{X}'} = \hat{\mathcal{U}} \hat{\rho}_i^{\mathcal{X}} \hat{\mathcal{U}}^\dagger$ as shown below

$$\hat{\rho}_j^{\mathcal{Y}'} = \frac{1}{q_j} \sum_{i \epsilon \{i\}_j} p_i \hat{\rho}_i^{\mathcal{X}'} \tag{4.41}$$

where $q_j = \sum_{i \epsilon \{i\}_j} p_i$. The output density operator $\hat{\rho}^{\mathcal{Y}'}$ can be written as

$$\hat{\rho}^{\mathcal{Y}'} = \sum_{j=0}^{2} q_j \hat{\rho}_j^{\mathcal{Y}'} \tag{4.42}$$

Let $a = sin\theta cos\theta$, and the output matrices are given as

$$\hat{\rho}_0^{\mathcal{Y}'} = \begin{bmatrix} cos^4\theta & a(1-P)cos^2\theta & (1-P)a^2 & a(1-P)cos^2\theta \\ a(1-P)cos^2\theta & a^2 & a(1-P)sin^2\theta & (1-P)a^2 \\ (1-P)a^2 & a(1-P)sin^2\theta & sin^4\theta & a(1-P)sin^2\theta \\ a(1-P)cos^2\theta & (1-P)a^2 & a(1-P)sin^2\theta & a^2 \end{bmatrix} \tag{4.43}$$

$$\mathcal{Y}'= \begin{bmatrix} a^2 & \frac{1}{2}(1-P)a & (1-P)a^2 & \frac{1}{2}(1-P)a \\ \frac{1}{2}(1-P)a & \frac{1}{2}\{cos^4\theta + sin^4\theta\} & \frac{1}{2}(1-P)a & (1-P)a^2 \\ (1-P)a^2 & \frac{1}{2}a & a^2 & \frac{1}{2}(1-P)a \\ \frac{1}{2}(1=P)a & (1-P)a^2 & \frac{1}{2}(1-P)a & \frac{1}{2}\{cos^4\theta + sin^4\theta\} \end{bmatrix} \quad (4.44)$$

$\hat{\rho}_1$

$$\hat{\rho}_2^{\mathcal{Y}'} = \begin{bmatrix} sin^4\theta & a(1-P)sin^2\theta & (1-P)a^2 & a(1-P)sin^2\theta \\ a(1-P)sin^2\theta & a^2 & a(1-P)cos^2\theta & (1-P)a^2 \\ (1-P)a^2 & a(1-P)cos^2\theta & cos^4\theta & a(1-P)cos^2\theta \\ a(1-P)sin^2\theta & (1-P)a^2 & a(1-P)cos^2\theta & a^2 \end{bmatrix} \quad (4.45)$$

$$\hat{\rho}^{Y'} = \begin{bmatrix} \frac{1}{4} & \frac{1}{2}a(1-P) & (1-P)a^2 & \frac{1}{2}a(1-P) \\ \frac{1}{2}a(1-P) & \frac{1}{4} & \frac{1}{2}a(1-P) & (1-P)a^2 \\ (1-P)a^2 & \frac{1}{2}a(1-P) & \frac{1}{4} & \frac{1}{2}a(1-P) \\ \frac{1}{2}a(1-P) & (1-P)a^2 & \frac{1}{2}a(1-P) & \frac{1}{4} \end{bmatrix} \quad (4.46)$$

Fidelity of the logical transformation $\mathcal{L}$ is given as

$$F_{\mathcal{L}} = \frac{\chi(\epsilon_Y^{\mathcal{Y}'})}{\chi(\epsilon_Y^{\mathcal{Y}})} \quad (4.47)$$

where

$$\chi(\epsilon_Y^{\mathcal{Y}'}) = S(\hat{\rho}^{\mathcal{Y}'}) - \sum_{j=0}^{2} q_j S(\hat{\rho}_j^{\mathcal{Y}'}) \quad (4.48)$$

$$\chi(\epsilon_Y^{\mathcal{Y}}) = S(\hat{\rho}^{\mathcal{Y}}) - \sum_{j=0}^{2} q_j S(\hat{\rho}_j^{\mathcal{Y}}) \quad (4.49)$$

where $\hat{\rho}^{\mathcal{Y}}$ and $\hat{\rho}_j^{\mathcal{Y}}$ for j=0,1,2 is given as

$$\hat{\rho}_0^{\mathcal{Y}} = \begin{bmatrix} cos^4\theta & acos^2\theta & acos^2\theta & a^2 \\ acos^2\theta & a^2 & a^2 & asin^2\theta \\ acos^2\theta & a^2 & a^2 & asin^2\theta \\ a^2 & asin^2\theta & asin^2\theta & sin^4\theta \end{bmatrix} \tag{4.50}$$

$$\hat{\rho}_1^{\mathcal{Y}} = \begin{bmatrix} a^2 & \frac{1}{2}a & \frac{1}{2}a & a^2 \\ \frac{1}{2}a & \frac{1}{2}\{cos^4\theta + sin^4\theta\} & a^2 & \frac{1}{2}a \\ \frac{1}{2}a & a^2 & \frac{1}{2}\{cos^4\theta + sin^4\theta\} & \frac{1}{2}a \\ a^2 & \frac{1}{2}a & \frac{1}{2}a & a^2 \end{bmatrix} \tag{4.51}$$

$$\hat{\rho}_2^{\mathcal{Y}} = \begin{bmatrix} sin^4\theta & asin^2\theta & asin^2\theta & a^2 \\ asin^2\theta & a^2 & a^2 & acos^2\theta \\ asin^2\theta & a^2 & a^2 & acos^2\theta \\ a^2 & acos^2\theta & acos^2\theta & cos^4\theta \end{bmatrix} \tag{4.52}$$

and

$$\hat{\rho}^{\mathcal{Y}} = \begin{bmatrix} \frac{1}{4} & \frac{1}{2}a & \frac{1}{2}a & a^2 \\ \frac{1}{2}a & \frac{1}{4} & a^2 & \frac{1}{2}a \\ \frac{1}{2}a & a^2 & \frac{1}{4} & \frac{1}{2}a \\ a^2 & \frac{1}{2}a & \frac{1}{2}a & \frac{1}{4} \end{bmatrix} \tag{4.53}$$

With all the required matrices defined clearly, the fidelity $F_{\mathcal{L}}$ can be calculated. Similarly for the representational faithfulness $f_{\mathcal{L}}$ given as

$$f_{\mathcal{L}} = 1 - \frac{\sum_{j=0}^{2} q_j \chi(\epsilon_j^{\mathcal{Y}'})}{\sum_{j=0}^{2} q_j \chi(\epsilon_j^{\mathcal{Y}})}$$

where

$$\chi(\epsilon_j^{\mathcal{Y}'}) = S(\hat{\rho}_j^{\mathcal{Y}'}) - \frac{1}{q_j} \sum_{i\epsilon\{i\}_j} p_i S(\hat{\rho}_i^{\mathcal{X}'}) \tag{4.54}$$

$$\chi(\epsilon_j^{\mathcal{Y}}) = S(\hat{\rho}_j^{\mathcal{Y}}) - \frac{1}{q_j} \sum_{i\epsilon\{i\}_j} p_i S(\hat{\rho}_i^{\mathcal{X}}) \tag{4.55}$$

From the above equations, the faithfulness with which the noisy CNOT gate implements binary addition can be calculated for different values of P and θ. The results of the calculation are presented below.

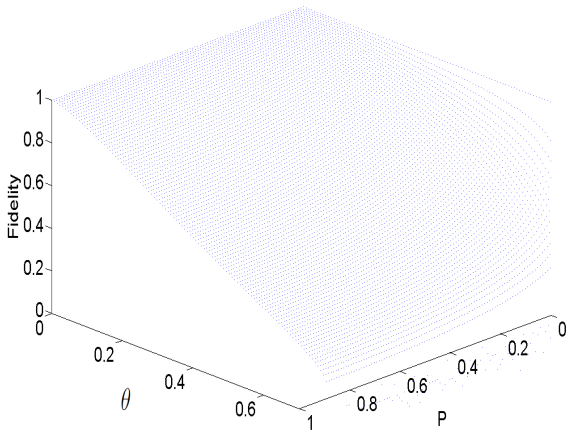

Figure 4.4. Variation in Computational Fidelity of CNOT gate with variation in θ and P

From the Figure 4.4, we can use the generalized definition of computational fidelity to allow for the input states to be noisy, the variation with P and θ is very insightful. It is important to remember that $F_{\mathcal{L}}$ is a measure of the amount of non-orthogonality introduced between system states that map into different logical outputs. For θ=0, that is if the system states are orthogonal initially, then $F_{\mathcal{L}}$=1 for all values of P. Binary addition can be implemented with unit fidelity even though the logical CNOT and binary addition operations are distinctly different because the outputs can be unambiguously inferred. Also for P=0, $F_{\mathcal{L}}$=1 for all values of θ, indicating that when there is no decoherence, the CNOT operation does not lead to

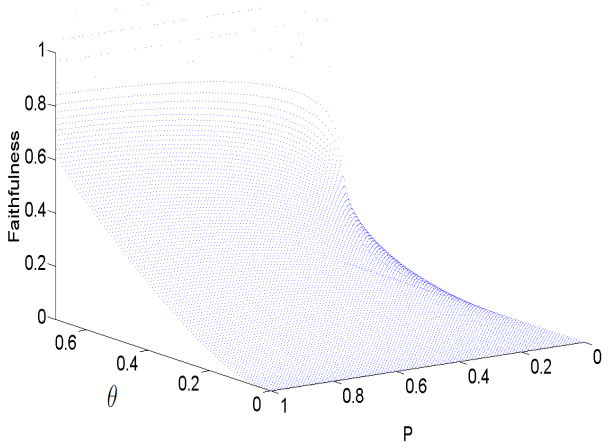

Figure 4.5. Variation in Representational Faithfulness of CNOT gate with variation in θ and P

increase in non-orthogonality between states that map into different logical outputs, than what was already present. We can also see that with increasing P for any fixed θ in the regime $[0,\frac{\pi}{4}]$, $F_{\mathcal{L}}$ decreases, reflecting the reduction in available information that accompanies loss of structure in output state caused by decoherence.

Figure 4.5 indicates the variation of $f_{\mathcal{L}}$ with θ and P. It is important to note that $f_{\mathcal{L}}$ is a measure of the amount of non-orthogonality introduced between system states that map into the same logical output. For θ=0, for all values of P, the CNOT gate evolves into completely distinguishable orthogonal outputs states and the input from which a output arose can be identified, hence $f_{\mathcal{L}}$=0. For any P varying from 0 to 1, for θ in the given regime, there is a increase in $f_{\mathcal{L}}$ with increasing θ as there is increased non-orthogonality in the system states. However for a given θ, with increasing value of the decoherence probability P, there is a decrease in $f_{\mathcal{L}}$ as the decoherence leads to creation of completely distinguishable orthogonal states whereas the logical operation requires non-orthogonality to exist between system states that map into the same logical output.

We were thus able to study the variation of the generalized computational fidelity and representational faithfulness measures of a quantum CNOT gate, subject to decoherence, performing classical binary addition. The results provided insight into the effect of non-orthogonality between input system states and that of decoherence on the performance of a logical operation.

4.2.5 Generalized Efficacy Measures for Two-Stage Logical Computations

One of the biggest advantages of the generalization in the efficacy measures formulation is that it allows us to clearly define the computational fidelity and representational faithfulness of individual stages of a multistage computation and differentiate it from the efficacy measures of the entire operation.

The next task would then be understand the relationship between the efficacy measures of the individual stages and that of the entire logical operation, and develop a clear formulation for it. This formulation would be very important as it would allow us to associate the performance of an entire logical operation with that of the individual stages. Further use of this in Eq. (3.49) would allow us to associate dissipation costs associated with an entire logical operation with the individual steps involved in the achieving it. For the error-correction problem, this formulation would help us associate the information transmitted from end-to-end with the performance metrics of the decoder and hence is very significant.

Since from Eq. (4.29), we know that information loss in a logical transformation is clearly divided into *desirable* and *undesirable* information loss. This concise division in $-\Delta I_{\mathcal{L}}$ is used to determine the required relationship between the various efficacy measures. In order to derive the relationship for a general N-stage computation, we shall start initially with a two stage logical transformation and then proceed from there. Let $\mathcal{L}_1$ and $\mathcal{L}_2$ indicate the individual stages of the two stage logical transformation, which is known as $\mathcal{L}$ from end-to-end. For example, if $\mathcal{L}_1$ was the

AND operation and $\mathcal{L}_2$ was the NOT operation performed on the outputs of the AND operation, then $\mathcal{L}$ is the NAND operation. We hence need to obtain the efficacy measures for $\mathcal{L}_1$ and $\mathcal{L}_2$ and relate it with that of $\mathcal{L}$.

Let the initial input referent to the logical transformation be X with a the probability distribution $\{p_i\}$. The input ensemble of the system states are given as $\epsilon_X^{\mathcal{D}} = \{p_i, \hat{\rho}_i^{\mathcal{D}}\}$. The amount of information about X that is present in the system states is given by the Holevo information $\chi(\epsilon_X^{\mathcal{D}})$ where

$$\chi(\epsilon_X^{\mathcal{D}}) = S(\hat{\rho}^{\mathcal{D}}) - \sum_i p_i S(\hat{\rho}_i^{\mathcal{D}}) \tag{4.56}$$

and

$$\hat{\rho}^{\mathcal{D}} = \sum_i p_i \hat{\rho}_i^{\mathcal{D}} \tag{4.57}$$

.

The system $\mathcal{D}$ undergoes the logical transformation $\mathcal{L}_1$ to produce the evolved input states $\mathcal{D}'$. Let Y be the output referent associated with the logical transformation $\mathcal{L}_1$ and input probability distribution $\{p_i\}$. Let q_j be the probability of the j-th logical output. The fidelity and faithfulness for the logical transformation $\mathcal{L}_1$ is given as

$$F_{\mathcal{L}_1} = \frac{\chi(\epsilon_Y^{\mathcal{D}'})}{\chi(\epsilon_Y^{\mathcal{D}})} \tag{4.58}$$

$$f_{\mathcal{L}_1} = 1 - \frac{\sum_j q_j \chi(\epsilon_j^{\mathcal{D}'})}{\sum_j q_j \chi(\epsilon_j^{\mathcal{D}})} \tag{4.59}$$

The evolved input states $\mathcal{D}'$ which are correlated to Y, now act as the inputs to second logical transformation $\mathcal{L}_2$. The amount of information that $\mathcal{D}'$ contains about Y is given by the Holevo information $\chi(\epsilon_Y^{\mathcal{D}'})$, where

$$\chi(\epsilon_Y^{\mathcal{D}'}) = S(\hat{\rho}^{\mathcal{D}'}) - \sum_j q_j S(\hat{\rho}_j^{\mathcal{D}'}) \tag{4.60}$$

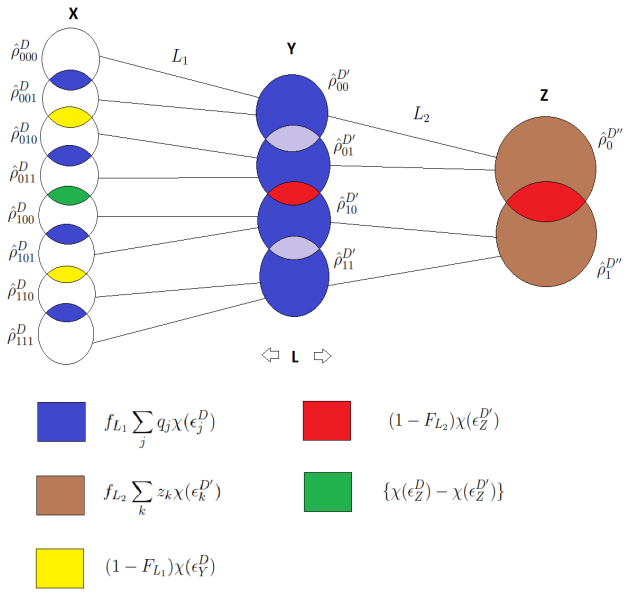

Figure 4.6. System D Undergoing a Two-staged Logical Operation

and

$$S(\hat{\rho}^{\mathcal{D}'}) = \sum_j q_j \hat{\rho}_j^{\mathcal{D}'} \tag{4.61}$$

$$\hat{\rho}_j^{\mathcal{D}'} = \frac{1}{q_j} \sum_{i\epsilon\{i\}_j} p_i \hat{\rho}_i^{\mathcal{D}'} \tag{4.62}$$

The system state evolves from $\mathcal{D}'$ to $\mathcal{D}''$ to implement the logical transformation $\mathcal{L}_2$. Let Z be the output referent of the logical transformation $\mathcal{L}_2$ with input distribution $\{q_j\}$ and z_k, the probability of the k-th logical output. The fidelity and faithfulness associated with $\mathcal{L}_2$ is given as

$$F_{\mathcal{L}_2} = \frac{\chi(\epsilon_Z^{\mathcal{D}''})}{\chi(\epsilon_Z^{\mathcal{D}'})} \tag{4.63}$$

$$f_{\mathcal{L}_2} = 1 - \frac{\sum_k z_k \chi(\epsilon_k^{\mathcal{D}''})}{\sum_k z_k \chi(\epsilon_k^{\mathcal{D}'})}. \tag{4.64}$$

If the entire two stage logical operation L is considered as $\mathcal{L}_1$ followed by $\mathcal{L}_2$, then the input and output referents are given as X and Z. The system state evolved from

$\mathcal{D}$ to $\mathcal{D}''$ to implement the entire logical transformation. The fidelity and faithfulness of L is given as

$$F_{\mathcal{L}} = \frac{\chi(\epsilon_Z^{\mathcal{D}''})}{\chi(\epsilon_Z^{\mathcal{D}})} \tag{4.65}$$

$$f_{\mathcal{L}} = 1 - \frac{\sum_k z_k \chi(\epsilon_k^{\mathcal{D}''})}{\sum_k z_k \chi(\epsilon_k^{\mathcal{D}})} \tag{4.66}$$

In order to relate all of these efficacy measures together, we need to study their relationship to the information lost. If we can identify the *desirable* and *undesirable* information lost in the two stage computation $\mathcal{L}$ and relate it with the information lost in the individual stages $\mathcal{L}_1$ and $\mathcal{L}_2$, the important relationship between the efficacy measures can be obtained.

Consider the desirable information loss in the second stage of the computation given by $f_{\mathcal{L}_2} \sum_k z_k \chi(\epsilon_k^{\mathcal{D}})$. Adding and subtracting the term $\sum_k z_k \chi(\epsilon_k^{\mathcal{D}})$ we get

$$f_{\mathcal{L}_2} \sum_k z_k \chi(\epsilon_k^{\mathcal{D}}) = \sum_k z_k \chi(\epsilon_k^{\mathcal{D}'}) - \sum_k z_k \chi(\epsilon_k^{\mathcal{D}''}) + \sum_k z_k \chi(\epsilon_k^{\mathcal{D}}) - \sum_k z_k \chi(\epsilon_k^{\mathcal{D}}) \tag{4.67}$$

Since

$$\sum_k z_k \chi(\epsilon_k^{\mathcal{D}}) - \sum_k z_k \chi(\epsilon_k^{\mathcal{D}''}) = f_{\mathcal{L}} \sum_k z_k \chi(\epsilon_k^{\mathcal{D}}) \tag{4.68}$$

Thus substituting we get

$$f_{\mathcal{L}_2} \sum_k z_k \chi(\epsilon_k^{\mathcal{D}}) = f_L \sum_k z_k \chi(\epsilon_k^{\mathcal{D}}) + \sum_k z_k [\chi(\epsilon_k^{\mathcal{D}'}) - \chi(\epsilon_k^{\mathcal{D}})] \tag{4.69}$$

Since we know

$$\chi(\epsilon_k^{\mathcal{D}'}) = S(\hat{\rho}_k^{\mathcal{D}'}) - \frac{1}{z_k} \sum_{j \epsilon \{j\}_k} q_j S(\hat{\rho}_j^{\mathcal{D}'}) \tag{4.70}$$

$$\sum_k z_k \chi(\epsilon_k^{\mathcal{D}'}) = \sum_k z_k S(\hat{\rho}_k^{\mathcal{D}'}) - \sum_j q_j S(\hat{\rho}_j^{\mathcal{D}'}) \tag{4.71}$$

Adding and subtracting $S(\hat{\rho}^{\mathcal{D}'})$ to the above equation, we get

$$\sum_k z_k \chi(\epsilon_k^{\mathcal{D}'}) = \chi(\epsilon_X^{\mathcal{D}'}) - \chi(\epsilon_Z^{\mathcal{D}'}) \tag{4.72}$$

Similarly we can write

$$\sum_k z_k \chi(\epsilon_k^{\mathcal{D}}) = \chi(\epsilon_X^{\mathcal{D}}) - \chi(\epsilon_Z^{\mathcal{D}}) \tag{4.73}$$

The second term in Eq. (4.69) can be written as

$$\sum_k z_k [\chi(\epsilon_k^{\mathcal{D}'}) - \chi(\epsilon_k^{\mathcal{D}})] = \{\chi(\epsilon_X^{\mathcal{D}'}) - \chi(\epsilon_Z^{\mathcal{D}'})\} - \{\chi(\epsilon_X^{\mathcal{D}}) - \chi(\epsilon_Z^{\mathcal{D}})\} \tag{4.74}$$

Since $\chi(\epsilon_X^{\mathcal{D}}) - \chi(\epsilon_X^{\mathcal{D}'})$ is the information about X that is lost as the system evolves from state $\mathcal{D}$ to $\mathcal{D}'$ implementing $\mathcal{L}_1$, it can be written as

$$\chi(\epsilon_X^{\mathcal{D}}) - \chi(\epsilon_X^{\mathcal{D}'}) = f_{\mathcal{L}_1} \sum_j q_j \chi(\epsilon_j^{\mathcal{D}}) + (1 - F_{\mathcal{L}_1}) \chi(\epsilon_Y^{\mathcal{D}}) \tag{4.75}$$

Substituting into the previous equation we get

$$\sum_k z_k [\chi(\epsilon_k^{\mathcal{D}'}) - \chi(\epsilon_k^{\mathcal{D}})] = \{\chi(\epsilon_Z^{\mathcal{D}}) - \chi(\epsilon_Z^{\mathcal{D}'})\} - f_{\mathcal{L}_1} \sum_j q_j \chi(\epsilon_j^{\mathcal{D}}) - (1 - F_{\mathcal{L}_1}) \chi(\epsilon_Y^{\mathcal{D}}) \tag{4.76}$$

Therefore we can write

$$f_{\mathcal{L}_2} \sum_k z_k \chi(\epsilon_k^{\mathcal{D}'}) = f_{\mathcal{L}} \sum_k z_k \chi(\epsilon_k^{\mathcal{D}}) + \{\chi(\epsilon_Z^{\mathcal{D}}) - \chi(\epsilon_Z^{\mathcal{D}'})\} - f_{\mathcal{L}_1} \sum_j q_j \chi(\epsilon_j^{\mathcal{D}}) - (1 - F_{\mathcal{L}_1})\chi(\epsilon_Y^{\mathcal{D}}) \tag{4.77}$$

Rearranging the terms we get

$$f_{\mathcal{L}} \sum_k z_k \chi(\epsilon_k^{\mathcal{D}}) = f_{\mathcal{L}_2} \sum_k z_k \chi(\epsilon_k^{\mathcal{D}'}) + f_{\mathcal{L}_1} \sum_j q_j \chi(\epsilon_j^{\mathcal{D}}) + (1 - F_{\mathcal{L}_1})\chi(\epsilon_Y^{\mathcal{D}}) - \{\chi(\epsilon_Z^{\mathcal{D}}) - \chi(\epsilon_Z^{\mathcal{D}'})\} \tag{4.78}$$

$$f_{\mathcal{L}} = \frac{f_{\mathcal{L}_2} \sum_k z_k \chi(\epsilon_k^{\mathcal{D}'}) + f_{\mathcal{L}_1} \sum_j q_j \chi(\epsilon_j^{\mathcal{D}}) + (1 - F_{\mathcal{L}_1})\chi(\epsilon_Y^{\mathcal{D}}) - \{\chi(\epsilon_Z^{\mathcal{D}}) - \chi(\epsilon_Z^{D'})\}}{\sum_k z_k \chi(\epsilon_k^{\mathcal{D}})} \tag{4.79}$$

The above equation thus relates the representational faithfulness of the entire two staged logical transformation with the efficacy measures of the individual stages. The first term $f_{\mathcal{L}_1} \sum_j q_j \chi(\epsilon_j^{\mathcal{D}})$ indicates the amount of desirable information loss that occurs about referent X, between system states that map into the same logical output when operation $\mathcal{L}_1$ is performed. Similarly $f_{\mathcal{L}_2} \sum_k z_k \chi(\epsilon_k^{\mathcal{D}'})$ indicates the amount of necessary information loss that occurs about referent Y, between system states that map into the same logical output when $\mathcal{L}_2$ is implemented. There is some information loss about referent Y which is undesirable for the implementation of $\mathcal{L}_1$ but is desirable when the entire two-staged logical transformation L is considered. This is indicated by $(1 - F_{\mathcal{L}_1})\chi(\epsilon_Y^{\mathcal{D}}) - \{\chi(\epsilon_Z^{\mathcal{D}}) - \chi(\epsilon_Z^{\mathcal{D}'})\}$.

Using a similar approach of studying the information lost, we can derive the relationship between the computational fidelity of L with the efficacy measures of the individual stages. Consider the term $(1 - F_{\mathcal{L}})\chi(\epsilon_Z^{\mathcal{D}})$. This can be written as

$$(1 - F_{\mathcal{L}})\chi(\epsilon_Z^{\mathcal{D}}) = \chi(\epsilon_Z^{\mathcal{D}}) - \chi(\epsilon_Z^{\mathcal{D}''}) \tag{4.80}$$

Adding and subtracting $\chi(\epsilon_Z^{\mathcal{D}'})$ we get

$$(1 - F_{\mathcal{L}})\chi(\epsilon_Z^{\mathcal{D}}) = \chi(\epsilon_Z^{\mathcal{D}}) - \chi(\epsilon_Z^{\mathcal{D}''}) + \chi(\epsilon_Z^{\mathcal{D}'}) - \chi(\epsilon_Z^{\mathcal{D}'}) \tag{4.81}$$

Since

$$\chi(\epsilon_Z^{\mathcal{D}'}) - \chi(\epsilon_Z^{\mathcal{D}''}) = (1 - F_{\mathcal{L}_2})\chi(\epsilon_Z^{\mathcal{D}'}) \tag{4.82}$$

We can write

$$(1 - F_{\mathcal{L}})\chi(\epsilon_Z^{\mathcal{D}}) = (1 - F_{\mathcal{L}_2})\chi(\epsilon_Z^{\mathcal{D}'}) + \{\chi(\epsilon_Z^{\mathcal{D}}) - \chi(\epsilon_Z^{\mathcal{D}'})\} \tag{4.83}$$

The above equation thus relates the computational fidelity $F_{\mathcal{L}}$ of the two stage logical computation L with the efficacy measures of the individual stages. These equations can be extended for N-stage logical computations and will provide powerful tools in characterizing the performances of individual stages in a multistage computation. Furthermore it will also help associate the energy dissipation associated with each stage to the performance metrics of that stage.

4.2.6 Generalized Efficacy Measures for N-Stage Logical Computations

Consider the N-stage logical transformation shown below

$$X_1 \xrightarrow{L_1} X_2 \xrightarrow{L_2} X_3 \xrightarrow{L_3} \cdots \xrightarrow{L_{n-1}} X_n \xrightarrow{L_n} X_{n+1}$$

$$\mathcal{D}_1 \longrightarrow \mathcal{D}_2 \longrightarrow \mathcal{D}_3 \longrightarrow \cdots \longrightarrow \mathcal{D}_n \longrightarrow \mathcal{D}_{n+1}$$

X_1 is the initial input referent and the X_{i+1} is the output referent of the $\mathcal{L}_i$-th logical transformation. $\mathcal{D}_1$ indicates the initial system state and it evolves to $\mathcal{D}_2$

while implementing $\mathcal{L}_1$, which then evolves to $\mathcal{D}_3$ when implementing $\mathcal{L}_2$ and so on and so forth. The entire end-to-end N-stage logical transformation is called $\mathcal{L}$ and is implemented by evolving the system from $\mathcal{D}_1$ to $\mathcal{D}_{n+1}$.

$$X_1 \xrightarrow{L} X_{n+1}$$

$$\mathcal{D}_1 \longrightarrow \mathcal{D}_{n+1}$$

Following the same procedure as in the previous subsection for the two-stage computation, the relationship between the fidelity and faithfulness of the N-stage computation in terms of the efficacy measures of the individual stages can be derived to be

$$f_{\mathcal{L}} = \frac{\left\{ \sum_{i=1}^{n} f_{\mathcal{L}_i} \sum_a p_a^i \chi(\epsilon_a^{\mathcal{D}_i}) + \sum_{i=1}^{n-1} (1 - F_{\mathcal{L}_i}) \chi(\epsilon_{X_{i+1}}^{\mathcal{D}_i}) - [\chi(\epsilon_{X_{n+1}}^{\mathcal{D}_1}) - \chi(\epsilon_{X_{n+1}}^{\mathcal{D}_n})] \right\}}{\sum_a p_a^{n+1} \chi(\epsilon_a^{\mathcal{D}_1})} \tag{4.84}$$

and furthermore

$$(1 - F_{\mathcal{L}}) \chi(\epsilon_{X_{n+1}}^{\mathcal{D}_1}) = (1 - F_{L_n}) \chi(\epsilon_{X_{n+1}}^{\mathcal{D}_n}) + [\chi(\epsilon_{X_{n+1}}^{\mathcal{D}_1}) - \chi(\epsilon_{X_{n+1}}^{\mathcal{D}_n})] \tag{4.85}$$

where $F_{\mathcal{L}_i}$ and f_{L_i} is the computational fidelity and representational faithfulness of the $\mathcal{L}_i$-th logical transformation and p_a^i is the probability associated with the a-th logical state in the referent X_i.

Over the last section we have generalized the computational fidelity and representational faithfulness measures, applied it to an example system and extended it for more complex systems. We now possess the tools required to characterize the performance the decoding operation and relate it to the amount of information obtained at the decoder output as well the lower bound on energy dissipation.

4.3 Application to the Case of Decoding the Noisy Channel Output

The above changes to the definition of computational efficacy measures are now applied to our system of interest implementing the decoding logical operation. The noise, modeled after a bit-flip channel followed by decoherence is introduced by the environment. There is no necessary heat dissipation to the environment associated with the introduction of this noise into the system the user does not have to pay in terms of energy cost. However when the system undergoes the decoding logical transformation, there is a loss of information and a reduction in the system entropy, which results in a heat dissipation cost that the user must pay. Determining a lower bound on this energy cost, as well as studying the relationship of the computational efficacy measures for the decoding process to this cost is one of the key goals of this thesis.

Since we know

$$-\Delta I_{decoding} = I(\hat{\rho}^{\mathcal{R}}; \hat{\rho}^{\mathcal{D}'}) - I(\hat{\rho}^{\mathcal{R}}; \hat{\rho}^{\mathcal{D}''}) = \chi(\epsilon_{\mathcal{R}}^{\mathcal{D}'}) - \chi(\epsilon_{\mathcal{R}}^{\mathcal{D}''}) \tag{4.86}$$

and

$$\langle E^{\varepsilon}_{decoding} \rangle \geq -k_B T \ln(2) \{ \Delta I_{decoding} + \sum_{i=0}^{2^{k-1}} p_i \left[S(\hat{\rho}_i^{\mathcal{D}'}) - S(\hat{\rho}_i^{\mathcal{D}''}) \right] \} \tag{4.87}$$

Defining $F_{decoding}$ and $f_{decoding}$ as the computational fidelity and representational faithfulness of the decoding operation, we can substitute for $-\Delta I_{decoding}$ and $\langle E^{\varepsilon}_{decoding} \rangle$ in terms of $F_{decoding}$ and $f_{decoding}$. We thus have a lower bound on the average energy dissipation to the environment, in terms of the fidelity and faithfulness of the decoding process. This will allow us to study the trade offs between "how well" the decoding

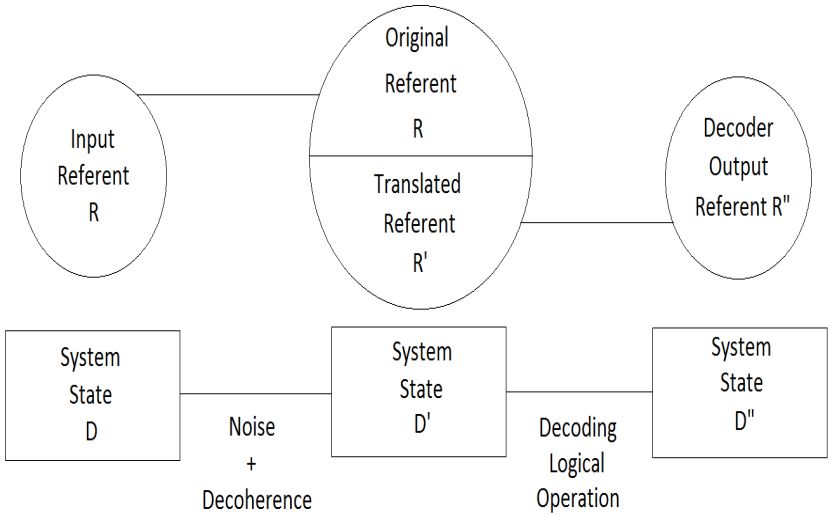

Figure 4.7. Block diagram of the System through the Communication Channel and Decoding Operation

is achieved and the minimum energy cost the user must pay to do so. Regions of diminishing returns for the decoder can also be determined.

In the previous sections, we have derived the equations for information loss and the lower bounds on energy cost for general quantum systems undergoing logical transformations. We shall develop the formulation for a system encoded as the linear codeword, which then experiences channel noise. The noisily encoded system undergoes the logical transformation of decoding to form the decoded output. The information loss and the energy dissipation associated with decoding is expressed in terms of fidelity and faithfulness of the decoding operation.

The input referent of the codewords is given as $\mathcal{R}$ and contains 2^k orthogonal states given by $\hat{\rho}_i^{\mathcal{R}}$ and specified by a probability p_i(let X be the random variable associated with this probability distribution). The input ensemble of the system is given as $\epsilon_X^{\mathcal{D}} = \{p_i, \hat{\rho}_i^{\mathcal{D}}\}$. The amount of information that the input ensemble contains about the referent $\mathcal{R}$ is given by the Holevo information term.

$$\chi(\epsilon_{\mathcal{R}}^{\mathcal{D}}) = S(\hat{\rho}^{\mathcal{D}}) - \sum_{i=0}^{2^{k-1}} p_i S(\hat{\rho}_i^{\mathcal{D}}) \tag{4.88}$$

The system experiences noise given by the quantum operator $\hat{\mathcal{U}}_1$, and is followed by decoherence to form the evolved system state $\mathcal{D}'$. The system ensemble is given as $\epsilon_{\mathcal{R}}^{\mathcal{D}'} = \{p_i, \hat{\rho}_i^{\mathcal{D}'}\}$. The amount of information about referent $\mathcal{R}$ contained in system state $\mathcal{D}'$ is given as

$$\chi(\epsilon_{\mathcal{R}}^{\mathcal{D}'}) = S(\hat{\rho}^{\mathcal{D}'}) - \sum_{i=0}^{2^{k-1}} p_i S(\hat{\rho}_i^{\mathcal{D}'}) \tag{4.89}$$

Since a communication channel is performing the identity operation and the output referent should be the same as the input referent $\mathcal{R}$. However $\mathcal{R}$ cannot be used as the input referent to the decoding logical transformation. Though it might provide necessary results, it is not intuitive as decoding is usually a many-to-one irreversible mapping while using R as the input referent to decoding would imply a one-to-one mapping with the decoder output referent. Studying decoding as a one-to-one mapping implies unnecessary information loss and questions the very need to perform it. Hence to overcome this, a **Referent Translation** is suggested, in which we replace the channel output referent R with another referent $\mathcal{R}'$. After decoherence, the system decoheres into one of the possible 2^n orthogonal pointer states $|j\rangle\langle j|$ with a certain probability. Let us define each of the system states $\hat{\rho}_i^{\mathcal{D}'}$ as

$$\hat{\rho}_i^{\mathcal{D}'} = \frac{1}{\sum_{j=0}^{2^n-1} q_{j/i}} \sum_{j=0}^{2^n-1} q_{j/i} |j\rangle\langle j| \tag{4.90}$$

Since we know that $\sum_{j=0}^{2^n-1} q_{j/i} = 1$,

$$\hat{\rho}_i^{\mathcal{D}'} = \sum_{j=0}^{2^n-1} q_{j/i} |j\rangle\langle j| \tag{4.91}$$

Thus we can write

$$\hat{\rho}^{\mathcal{D}'} = \sum_{i=0}^{2^k-1} p_i \hat{\rho}_i^{\mathcal{D}'} \tag{4.92}$$

$$\hat{\rho}^{\mathcal{D}'} = \sum_{j=0}^{2^n-1} q_j |j\rangle\langle j| \tag{4.93}$$

where

$$q_j = \sum_{i=0}^{2^k-1} p_i q_{j/i} \tag{4.94}$$

q_j will depend upon the value of θ and e and will indicate the probability with which we obtain the pointer state $\hat{\rho}_j^{D'} = |j\rangle\langle j|$.

We shall now define the new referent $\mathcal{R}'$, specified by 2^n orthogonal states $\hat{\rho}_j^{\mathcal{R}'}$ and corresponding probability q_j(let Y be the random variable associated with this distribution). The referent $\mathcal{R}'$ is used to obtain the output referent $\mathcal{R}''$ of the decoding logical operation. The probability distribution associated with the 2^k logical outputs is given by $\{z_m\}$ (and let Z be the random variable associated with this distribution) where

$$z_m = \sum_{j \epsilon \{j\}_m} q_j \tag{4.95}$$

The decoding transformation is implemented by evolving the system from the state $\mathcal{D}'$ to $\mathcal{D}''$ through the quantum operation $\hat{\mathcal{U}}_2$. The output ensemble is given as $\epsilon_Z^{\mathcal{D}''} = \{z_m, \hat{\rho}_m^{\mathcal{D}''}\}$ where

$$\hat{\rho}_m^{\mathcal{D}''} = \frac{1}{z_m} \sum_{j \epsilon \{j\}_m} q_j \hat{\rho}_j^{\mathcal{D}''} \tag{4.96}$$

Under these definitions, we can write

$$\hat{\rho}_i^{\mathcal{D}''} = \sum_{m=0}^{2^k-1} \sum_{j \epsilon \{j\}_m} q_{j/i} \hat{\rho}_j^{\mathcal{D}''} \tag{4.97}$$

The information loss about the referent $\mathcal{R}$ during the logical transformation is given by

$$-\Delta I_{decoding} = \chi(\epsilon_X^{\mathcal{D}'}) - \chi(\epsilon_X^{\mathcal{D}''}) \tag{4.98}$$

Writing $\chi(\epsilon_X^{\mathcal{D}'})$ and $\chi(\epsilon_X^{\mathcal{D}''})$ as

$$\chi(\epsilon_X^{\mathcal{D}'}) = \chi(\epsilon_Z^{\mathcal{D}'}) + \sum_{m=0}^{2^k-1} z_m \chi(\epsilon_m^{\mathcal{D}'}) \tag{4.99}$$

where

$$\chi(\epsilon_Z^{\mathcal{D}'}) = S(\hat{\rho}^{\mathcal{D}'}) - \sum_m z_m S(\hat{\rho}_j^{\mathcal{D}'_m}) \tag{4.100}$$

and since

$$\hat{\rho}_m^{\mathcal{D}'} = \frac{1}{z_m} \sum_{i \epsilon \{i\}_m} p_i \hat{\rho}_i^{\mathcal{D}'} \tag{4.101}$$

$$\chi(\epsilon_m^{\mathcal{D}'}) = S(\hat{\rho}_m^{\mathcal{D}'}) - \frac{1}{z_m} \sum_{i \epsilon \{i\}_m} p_i S(\hat{\rho}_j^{\mathcal{D}'}) \tag{4.102}$$

Add and subtract $\frac{1}{z_m} \sum_{j \epsilon \{j\}_m} q_j S(\hat{\rho}_j^{\mathcal{D}'})$ to the above equation. Rearranging terms we have

$$\chi(\epsilon_m^{\mathcal{D}'}) = S(\hat{\rho}_m^{\mathcal{D}'}) - \frac{1}{z_m} \sum_{j \epsilon \{j\}_m} q_j S(\hat{\rho}_j^{\mathcal{D}'}) - \frac{1}{z_m} \sum_{i \epsilon \{i\}_m} p_i S(\hat{\rho}_j^{\mathcal{D}'} + \frac{1}{z_m} \sum_{j \epsilon \{j\}_m} q_j S(\hat{\rho}_j^{\mathcal{D}'}) \tag{4.103}$$

Similarly

$$\chi(\epsilon_X^{\mathcal{D}''}) = \chi(\epsilon_Z^{\mathcal{D}''}) + \sum_{m=0}^{2^k-1} z_m \chi(\epsilon_m^{\mathcal{D}''}) \tag{4.104}$$

with

$$\chi(\epsilon_Z^{\mathcal{D}''}) = S(\hat{\rho}^{\mathcal{D}''}) - \sum_m z_m S(\hat{\rho}_m^{\mathcal{D}''}) \tag{4.105}$$

$$\chi(\epsilon_m^{D''}) = S(\hat{\rho}_m^{D''}) - \frac{1}{z_m} \sum_{i \epsilon \{i\}_m} p_i S(\hat{\rho}_i^{D''}) \tag{4.106}$$

The above equation is rewritten as before

$$\chi(\epsilon_m^{D''}) = S(\hat{\rho}_m^{D''}) - \frac{1}{z_m} \sum_{j\epsilon\{j\}_m} q_j S(\hat{\rho}_j^{D''}) - \frac{1}{z_m} \sum_{i\epsilon\{i\}_m} p_i S(\hat{\rho}_j^{D''} + \frac{1}{z_m} \sum_{j\epsilon\{j\}_m} q_j S(\hat{\rho}_j^{D''}) \quad (4.107)$$

If the computational fidelity $F_{decoding}$ and representational faithfulness $f_{decoding}$ of the decoding operation which mapped logical inputs from $\mathcal{R}'$ to $\mathcal{R}''$ is defined as below,

$$F_{decoding} = \frac{\chi(\epsilon_Z^{\mathcal{D}''})}{\chi(\epsilon_Z^{\mathcal{D}'})} \quad (4.108)$$

$$f_{decoding} = 1 - \frac{\sum_{m=0}^{2^k-1} z_m \chi'(\epsilon_m^{\mathcal{D}''})}{\sum_{m=0}^{2^k-1} z_m \chi'(\epsilon_m^{\mathcal{D}'})} \quad (4.109)$$

where we define

$$\chi'(\epsilon_m^{\mathcal{D}'}) = S(\hat{\rho}_m^{\mathcal{D}'}) - \frac{1}{z_m} \sum_{j\epsilon\{j\}_m} q_j S(\hat{\rho}_j^{\mathcal{D}'}) \quad (4.110)$$

$$\chi'(\epsilon_m^{\mathcal{D}''}) = S(\hat{\rho}_m^{\mathcal{D}''}) - \frac{1}{z_m} \sum_{j\epsilon\{j\}_m} q_j S(\hat{\rho}_j^{\mathcal{D}''}) \quad (4.111)$$

Substituting into the equation for information loss, we get

$$\begin{aligned} -\Delta I_{decoding} &= (1 - F_{decoding})\chi(\epsilon_Z^{\mathcal{D}'}) + f_{decoding} \sum_{m=0}^{2^k-1} \chi'(\epsilon_m^{\mathcal{D}'}) \\ &+ \sum_{j=0}^{2^n-1} q_j \left[S(\hat{\rho}_j^{\mathcal{D}'}) - S(\hat{\rho}_j^{\mathcal{D}'}) \right] - \sum_{i=0}^{2^k-1} p_i \left[S(\hat{\rho}_i^{\mathcal{D}'}) - S(\hat{\rho}_i^{\mathcal{D}''}) \right] \quad (4.112) \end{aligned}$$

The left hand side, information loss is loss of information in the system states about the initial referent $\mathcal{R}$ at the channel input during the decoding operation. A closer look at the information loss equation reveals that the first two terms $(1 - F_{decoding})\chi(\epsilon_Z^{\mathcal{D}'})$

and $f_{decoding}\sum_{m=0}^{2^k-1}\chi'(\epsilon_m^{\mathcal{D}'})$ indicate information loss about the referent $\mathcal{R}'$ as the system undergoes the decoding process, expressed in terms of its efficacy measures. whereas the term $\sum_{i=0}^{2^k-1} p_i\left[S(\hat{\rho}_i^{\mathcal{D}'}) - S(\hat{\rho}_i^{\mathcal{D}''})\right] - \sum_{j=0}^{2^n-1} q_j\left[S(\hat{\rho}_j^{\mathcal{D}'}) - S(\hat{\rho}_j^{\mathcal{D}'})\right]$ indicates the entropy of the decoder input states that the decoder treats as information of the input states (because of referent translation) and hence must be subtracted while calculating the information loss about referent $\mathcal{R}$.

We can substitute Eq. (4.112) into the equation for entropy change

$$\Delta S^D_{decoding} = k_B \ln(2)\left[-\Delta I_{decoding} + \sum_{i=0}^{2^k-1} p_i\{S(\hat{\rho}_i^{\mathcal{D}'}) - S(\hat{\rho}_i^{\mathcal{D}''})\}\right] \tag{4.113}$$

We get

$$\begin{aligned}\Delta S^D_{decoding} &= k_B \ln(2)\left[(1-F_{decoding})\chi(\epsilon_Z^{\mathcal{D}'}) + f_{decoding}\sum_{m=0}^{2^k-1} z_m\chi'(\epsilon_m^{\mathcal{D}'})\right. \\ &\quad \left. + \sum_{i=0}^{2^n-1} q_j\{S(\hat{\rho}_j^{\mathcal{D}'}) - S(\hat{\rho}_j^{\mathcal{D}''})\}\right] \end{aligned} \tag{4.114}$$

where $\hat{\rho}_j^{\mathcal{D}'}$ and $\hat{\rho}_j^{\mathcal{D}''}$ are already defined. The lower bound on energy dissipation is thus obtained as

$$\begin{aligned}\langle \Delta E^{\varepsilon}_{decoding}\rangle &\geq k_B T \ln(2)\left[(1-F_{decoding})\chi(\epsilon_Z^{\mathcal{D}'}) + f_{decoding}\sum_{m=0}^{2^k-1} z_m\chi'(\epsilon_m^{\mathcal{D}'})\right. \\ &\quad \left. + \sum_{j=0}^{2^n-1} q_j\{S(\hat{\rho}_j^{\mathcal{D}'}) - S(\hat{\rho}_j^{\mathcal{D}''})\}\right] \end{aligned} \tag{4.115}$$

4.3.1 Central Result

The central result of the work done in this thesis is the following theorem

Theorem- A decoding operation performed on a noisily encoded system $\mathcal{D}'$ with computational fidelity $F_{decoding}$ and representational faithfulness $f_{decoding}$, to obtain $\chi(\epsilon_{\mathcal{R}}^{\mathcal{D}''})$ bits of information about the initial referent $\mathcal{R}$, will lead to a minimum average energy dissipation to the environment of

$$\begin{aligned}\langle \Delta E^{\varepsilon}_{decoding} \rangle \;\geq\; & k_B T \ln(2) \Bigg[(1 - F_{decoding}) \chi(\epsilon_Z^{\mathcal{D}'}) + f_{decoding} \sum_{m=0}^{2^k-1} z_m \chi'(\epsilon_m^{\mathcal{D}'}) \\ & + \sum_{j=0}^{2^n-1} q_j \{ S(\hat{\rho}_j^{\mathcal{D}'}) - S(\hat{\rho}_j^{\mathcal{D}''}) \} \Bigg] \end{aligned} \tag{4.116}$$

$\mathcal{D}''$ refers to the state of the system after the decoding operation is performed on it. $\chi(\epsilon_{\mathcal{R}}^{\mathcal{D}''})$ is the Holevo information associated with the system state $\mathcal{D}''$ and indicates the maximum amount of information that the system can possess about the initial referent $\mathcal{R}$. $F_{decoding}$ and $f_{decoding}$ is the computational fidelity and representational faithfulness of the decoding operation and depend purely upon the decoder's characteristics. The first two terms on the right hand side indicate the information lost about the translated referent $\mathcal{R}'$ with probability distribution $\{q_j\}$, as the system evolves from state $\mathcal{D}'$ to $\mathcal{D}''$. $\chi(\epsilon_Z^{\mathcal{D}'})$ indicates the amount of information that D' will contain about the decoder output referent if the decoding is performed ideally, and $\sum_{m=0}^{2^k-1} z_m \chi'(\epsilon_m^{\mathcal{D}'})$ indicates the maximum amount of information that D' has to lose for the decoding operation to be performed ideally. Both of these terms, $\chi(\epsilon_Z^{\mathcal{D}'})$ and $\chi'(\epsilon_m^{\mathcal{D}'})$ depend only upon the channel and input state properties and probability distribution. The last term indicates the average change in self entropy of the system during the logical operation.

We thus have a lower bound on the amount of energy dissipated on performing the decoding operation on the noisy channel output in terms of decoder's performance metrics. This will enable us to study how well the decoder performs the

error-correction operation, the amount of information we save by this process and a lower bound on the energy dissipated in accomplishing this task.

4.3.2 Illustrative Example

In the previous sections, we have derived the equations for information loss and the lower bounds on energy cost for general quantum systems. We shall now study the results of the simulations for a pure state system. The input n-tuple codeword is physically encoded into the system $\mathcal{D}$. If the codeword x_i is a string of binary digits 0's and 1's, they are encoded in the states $|\psi_0\rangle$ and $|\psi_1\rangle$ respectively. Both $|\psi_0\rangle$ and $|\psi_1\rangle$ are as given in the CNOT gate example

$$|\psi_0\rangle = cos\theta|0\rangle + sin\theta|1\rangle$$

$$|\psi_1\rangle = sin\theta|0\rangle + cos\theta|1\rangle$$

Where $|0\rangle$ and $|1\rangle$ are orthogonal basis vectors and θ is a measure of indistinguishability between the two states. Thus a n-tuple codeword is thus formed by the tensor product of n such states. For example, in a Hamming (7,4) code, the 7-digit 0000111 is encoded into the system as $|\psi_0 \otimes \psi_0 \otimes \psi_0 \otimes \psi_0 \otimes \psi_1 \otimes \psi_1 \otimes \psi_1\rangle$.

The input referent of the codewords is given as $\mathcal{R}$ and contains 2^k orthogonal states given by $\hat{\rho}_i^{\mathcal{R}}$ and specified by a probability p_i.The input ensemble of the system is given as $\epsilon_X^{\mathcal{D}} = \{p_i, \hat{\rho}_i^{\mathcal{D}}\}$ with each of the $\hat{\rho}_i^{\mathcal{D}}$ given as

$$\hat{\rho}_i^{\mathcal{D}} = |\psi_{c_1} \otimes \psi_{c_2} \otimes \psi_{c_3} \otimes \psi_{c_4} \otimes ... \otimes \psi_{c_{n-1}} \otimes \psi_{c_n}\rangle\langle\psi_{c_1} \otimes \psi_{c_2} \otimes \psi_{c_3} \otimes \psi_{c_4} \otimes ... \otimes \psi_{c_{n-1}} \otimes \psi_{c_n}| \tag{4.117}$$

where for l=1 to n, c_l=0 or 1. Depending on the value of θ, the $\hat{\rho}_i^{\mathcal{D}}$ may or may not be orthogonal.

If $\hat{\mathcal{U}}$ is the operator associated with the bit-flip noise, with the probability of a bit-flip error being given by the parameter e, that is probability of flipping the state $|0\rangle$ to the state $|1\rangle$ and vice versa. Then for a single bit flip, $\hat{\mathcal{U}}$ is given as

$$\hat{\mathcal{U}} = \sqrt{1-e}\{|0\rangle\langle 0| + |1\rangle\langle 1|\} + \sqrt{e}\{|0\rangle\langle 1| + |1\rangle\langle 0|\} \tag{4.118}$$

Thus the operator for channel noise associated with a n-tuple codeword $\hat{\mathcal{U}}_1$ is given by the tensor product of n single error bit-flip operators $\hat{\mathcal{U}}$.

$$\hat{\mathcal{U}}_1 = \hat{\mathcal{U}} \otimes \hat{\mathcal{U}} \otimes \hat{\mathcal{U}} \otimes \hat{\mathcal{U}} \otimes .... \otimes \hat{\mathcal{U}} \tag{4.119}$$

The system experiences noise given by the operator $\hat{\mathcal{U}}_1$ and is then followed by decoherence on the evolved state to form the system state $\mathcal{D}'$. The system ensemble is given as $\epsilon_{\mathcal{R}}^{\mathcal{D}'} = \{p_i, \hat{\rho}_i^{\mathcal{D}'}\}$. The decoding operation is implemented by evolving the system from the state $\mathcal{D}'$ to $\mathcal{D}''$ through the quantum operation $\hat{\mathcal{U}}_2$. The output ensemble after performing referent translation is given as $\epsilon_Z^{\mathcal{D}''} = \{z_m, \hat{\rho}_m^{\mathcal{D}''}\}$.

The performance metrics of the decoder is solely technology-dependent. However for this thesis, the following error scheme is followed and the performance metrics, information at the decoder outputs and the lower bound on the energy dissipated are calculated and the results are plotted and discussed in the next section. We assume that the errors in the decoder is such that every one of the 2^n noisy channel outputs evolve into the correct codeword state with a probability (1-f), where f is a parameter which ranges in the interval [0,1] and maps into every wrong codeword with a probability $\frac{f}{2^{k-1}}$ each. The performance parameters will depend only upon the value of f and will be independent of the values of e and θ, even though $\chi(\epsilon_{\mathcal{R}}^{\mathcal{D}''})$

will depend upon all three quantities. For j$\epsilon\{j\}_m$, the decoding error scheme can be viewed as evolving any state $\hat{\rho}_j^{\mathcal{D}'}$ as

$$\hat{\rho}_j^{\mathcal{D}'} = |j\rangle\langle j| \longrightarrow (1-f)|m\rangle\langle m| + \sum_{m':j\notin\{j\}_{m'}} \frac{f}{15}|m'\rangle\langle m'| \tag{4.120}$$

As derived in the previous section the information lost in the decoding operation is given as

$$-\Delta I_{decoding} = \chi(\epsilon_{\mathcal{R}}^{\mathcal{D}'}) - \chi(\epsilon_{\mathcal{R}}^{\mathcal{D}''}) \tag{4.121}$$

which can be expressed in terms of the computational fidelity $F_{decoding}$ and representational faithfulness $f_{decoding}$ of the decoding operation as

$$\begin{aligned} -\Delta I_{decoding} &= (1-F_{decoding})\chi(\epsilon_Z^{\mathcal{D}'}) + f_{decoding}\sum_{k=0}^{2^k-1} z_k\chi(\epsilon_k^{\mathcal{D}'}) \\ &+ \sum_{j=0}^{2^n-1} q_j\left[S(\hat{\rho}_j^{\mathcal{D}'}) - S(\hat{\rho}_j^{\mathcal{D}''})\right] - \sum_{i=0}^{2^k-1} p_i\left[S(\hat{\rho}_i^{\mathcal{D}'}) - S(\hat{\rho}_i^{\mathcal{D}''})\right] \end{aligned} \tag{4.122}$$

where

$$F_{decoding} = \frac{\chi(\epsilon_Z^{\mathcal{D}''})}{\chi(\epsilon_Z^{\mathcal{D}'})} \tag{4.123}$$

$$f_{decoding} = 1 - \frac{\sum_{k=0}^{2^k-1} z_k\chi(\epsilon_k^{\mathcal{D}''})}{\sum_{k=0}^{2^k-1} z_k\chi(\epsilon_k^{\mathcal{D}'})} \tag{4.124}$$

For the decoder error pattern that we have chosen for this thesis, calculation of faithfulness produces a value of 1 always. When the fidelity equal to one as well, that is the decoding operation was performed perfectly, we can write $\hat{\rho}_m^{\mathcal{D}''} = |m\rangle\langle m|$ and

an associated probability of z_m, where $|m\rangle\langle m|$ are the 2^k orthogonal output states. Thus

$$\hat{\rho}_i^{\mathcal{D}''} = \sum_{m=0}^{2^k-1} z_{m/i}|m\rangle\langle m| \tag{4.125}$$

The information loss $-\Delta I_{decoding}$ reduces to the Shannon information loss as shown below

$$\chi(\epsilon_R^{\mathcal{D}'}) = S(\hat{\rho}^{\mathcal{D}'}) - \sum_{i=0}^{2^k-1} p_i S(\hat{\rho}_i^{\mathcal{D}'}) \tag{4.126}$$

Since $\hat{\rho}^{\mathcal{D}'} = \sum_{j=0}^{2^n-1} q_j|j\rangle\langle j|$ and $\hat{\rho}_i^{\mathcal{D}'} = \sum_{j=0}^{2^n-1} q_{j/i}|j\rangle\langle j|$,

$$S(\hat{\rho}^{\mathcal{D}'}) = -\sum_{j=0}^{2^n-1} q_j log_2 q_j = H(Y) \tag{4.127}$$

$$S(\hat{\rho}_i^{\mathcal{D}'}) = -\sum_{j=0}^{2^n-1} q_{j/i} log_2 q_{j/i} = H(Y|X) \tag{4.128}$$

Thus we have $\chi(\epsilon_R^{\mathcal{D}'})$=I(Y;X). Similarly we can show that $\chi(\epsilon_R^{\mathcal{D}''})$=I(Z;X). Substituting into the equation for information loss,

$$-\Delta I_{decoding} = I(Y;X) - I(Z;X) \tag{4.129}$$

Since this equation is obtained when the $f_{decoding}$=1 and $F_{decoding}$=1, we also have

$$\begin{aligned} -\Delta I_{decoding} &= (I(Y;Y) - I(Y;Z)) - \sum_{i=0}^{2^k-1} p_i \left[H(Y|x_i) - H(Z|x_i)\right] \\ &\quad - \sum_{j=0}^{2^n-1} q_j \left[H(Y|y_j) - H(Z|y_j)\right] \end{aligned} \tag{4.130}$$

where I(Y;Y)-I(Y;Z) indicates the amount of information loss about the channel out-

puts that occurs during the decoding operation. This is equal to H(Y—Z)=$\sum_{m=0}^{2^k-1} z_m H(Y|z_m)$. Furthermore since $H(Y|y_j)$=0 always and when the decoding operation is executed perfectly $H(Z|y_j)$=0, information lost reduces to

$$-\Delta I_{decoding} = H(Y|Z) - \sum_{i=0}^{2^k-1} p_i \left[H(Y|x_i) - H(Z|x_i)\right] \tag{4.131}$$

Thus substituting Eq. (4.129) into Eq. (3.40) and reducing it, we get

$$\Delta S^D = k_B \ln(2) \left[(I(Y;Y) - I(Z;Y)) + \sum_{j=0}^{2^n-1} q_j \{H(Y|y_j) - H(Z|y_j)\}\right] \tag{4.132}$$

$$\Delta S^D = k_B \ln(2) H(Y|Z) \tag{4.133}$$

and

$$\langle E^{\varepsilon}_{decoding} \rangle \geq k_B T \ln(2) H(Y|Z) \tag{4.134}$$

The values of q_j and z_m depend upon the values of bit flip error (e), angle θ and decoder error (f). However if θ=0 and the input states are completely orthogonal to each other, then the values of q_j are equivalent to those obtained from a classical binary symmetric channel. Eq. (4.134) provides us with the lower bound on energy dissipation for performing the decoding operation in the special case where the initial input states are orthogonal to each other and the decoding operation has been performed perfectly. Since even in post-CMOS nanoelectronic systems, decoding operations are performed by extremely robust CMOS based circuits, this equation will provide a good idea of the least energy that is required to achieve error-correction. The information obtained at the decoder output and lower bound on the energy dissipation for various values of the channel bit flip error (e) , angle (θ) between the input

states and decoder error (f) for the Hamming (7,4) case are calculated and plotted in the next section.

4.3.2.1 Results and Discussion

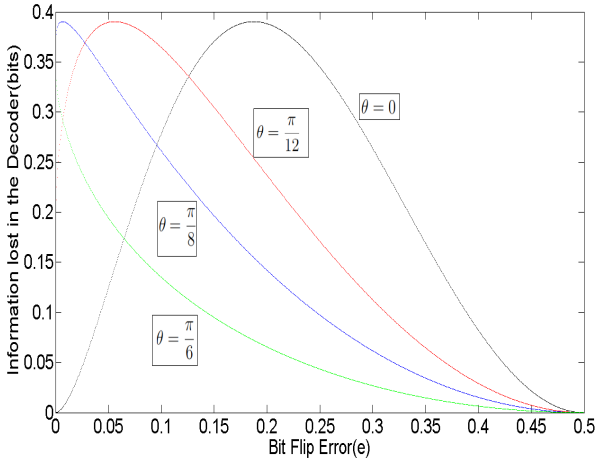

Figure 4.8. Information Loss in the Decoder (bits) vs Channel Bit Flip Error (e) for different θ in Hamming(7,4) case

In Figure 4.8, the information lost about the referent R as the system undergoes the decoding operation is plotted as the channel bit flip error (e) varies from 0 to 0.5. The same is plotted for different values of the angle between the states θ to study the effect of non-orthogonality on the states. The decoding operation is assumed to be perfect, that is $F_{decoding}$ and $f_{decoding}$ are both equal to one, which is possible as they depend only upon the parameter f. We can see that in the figure, for θ=0, there is no information loss at e=0. The information loss then increases with increase in e, reaches a maximum and then decreases to zero. The decrease does not imply that the decoder is performing better and loses lesser amount of information. With increasing e, there is greater information loss in the channel and not enough information about the initial referent R left to be lost in the decoder, culminating to the value of zero for e= 0.5 when all the information is lost in the channel itself. As the value of

θ is increased and the input states are made non-orthogonal, we see that there is information loss even when the bit flip error (e)=0, which increases with increasing e, reaches a maximum and then decreases as before as in the case of $\theta = \frac{\pi}{12}$ and $\theta = \frac{\pi}{8}$. However further increases in θ results in maximum information loss in the decoder for e= 0 and increase in e only results in decrease in information loss(as in the case of $\theta = \frac{\pi}{6}$).

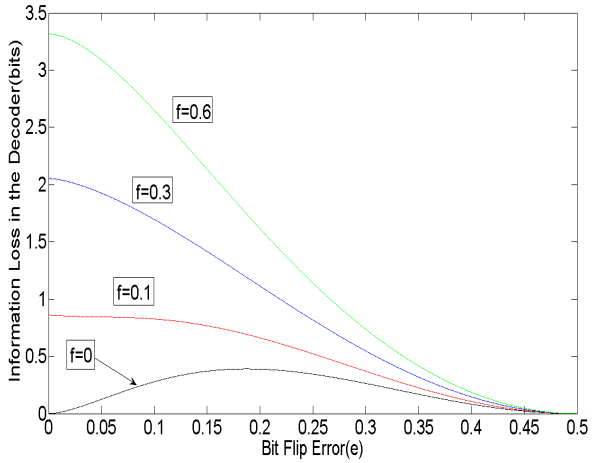

Figure 4.9. Information Loss in the Decoder (bits) vs Channel Bit Flip Error (e) for different values of Decoder Error (f) and θ=0 in Hamming(7,4) case

In Figure 4.9, we have a graph of the information about the referent R lost in the decoder as the bit flip error(e) varies between 0 and 0.5. The Hamming (7,4) code is used in this case and curves for various values of the decoder error probability(f) are calculated. We can see that for f=0, that is if the decoding operation is perfectly carried out, then as indicated in Figure 4.8, information lost in the decoder increases, reaches a maximum and then decreases. As the value of f increases, there is a similar behavior in the manner information loss changes with e, except that there is a non-zero information loss in the decoder at e=0. This information loss is due to the erroneous functioning of the decoder even when there is no information loss in the channel.For further increase in the value of f, information loss achieves a maximum

at e=0 and then continues to decrease with increasing values of e as indicated for f=0.1, f=0.3 and f=0.6.

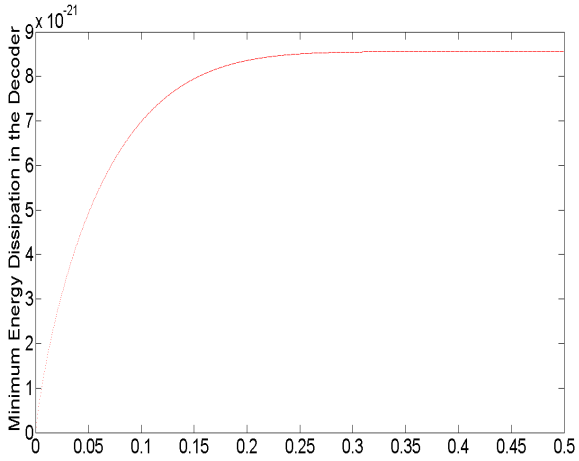

Figure 4.10. Minimum Energy Dissipation in the Decoder (Joules) vs Channel Bit Flip Error (e) for $\theta = 0$ in Hamming (7,4) case

While information about the referent R decreases with increasing e, the energy cost of decoding is a continuous increasing function as indicated in Figure 4.10. Calculation of the energy dissipation in the decoder is independent of the amount of decoder error present, but is dependent of the value of θ for the error scheme used in this thesis. This can be attributed to the simple and highly symmetrical nature of the Hamming (7,4) code used and that of the error scheme. This does not imply that other error schemes will yield similar results and hence should be tested in detail. As the value of θ increases, there is greater information loss at lower values of e, owing to the non-orthogonality and this results in the lower bound achieving the value of $k_B T \ln(2) \times n$ very quickly. Since there is clearly greater information loss in the channel at higher values of e, it is interesting to study whether the energy cost we pay in decoding actually pays dividends in the amount of information we get. For this purpose, we define a new term called **Information saved** which is the difference in the amount of information that is present in the decoder output when error-correction

techniques like a linear code is used and the amount of information the channel output contains about the input when no encoding is used, through a channel with a bit flip error probability of e, assuming the same amount of information is transmitted on the input side. When a (n,k) linear block code is used

$$\chi_{saved} = n \times \chi_{coded}(\epsilon_{R_0}^{D''}) - k \times \chi_{not-coded}(\epsilon_{R_1}^{D'}) \tag{4.135}$$

where $\chi_{coded}(\epsilon_{R_0}^{D''})$ indicates the amount of information the system state D'', after the decoding operation has been performed, about the initial referent R_0 which contains k bits of information. $\chi_{not-coded}(\epsilon_{R_1}^{D'})$ is the amount of information that the system state D', after the system experiences bit flip noise and decoherence about the initial referent R_1 which contains n bits of information.

On a per bit of information transmitted at the input basis,

$$\chi_{saved} = \frac{n \times \chi_{coded}(\epsilon_{R_0}^{D''}) - k \times \chi_{not-coded}(\epsilon_{R_1}^{D'})}{nk} \tag{4.136}$$

From the Figure 4.11, we see that for $f = 0$, that when the decoder is functioning ideally, with increasing e the information saved initially increases and then decreases to zero for $e = 0.5$. As the decoder is functioning without any error $\chi_{saved} \geq 0$ for all values of e. However when f is non-zero, there is a region of bit flip error e when the $\chi_{saved} < 0$, which indicates that there is a range of values of e for which using a linear (n,k) code is disadvantageous as there is dissipation of heat at the decoder without providing adequate amount of information at the decoder output. In this range, it might be more advantageous in terms of information obtained and heat dissipation to not code at all. We also study the effect of the variation in the angle between the input states θ in Figure 4.12. As non-orthogonality is introduced into the input states, there is an expected decrease in the amount of information saved by using

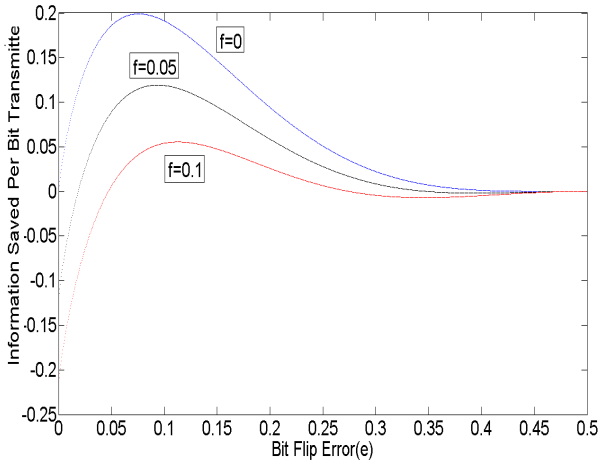

Figure 4.11. Information Saved per Bit Transmitted vs Channel Bit Flip Error (e) for $\theta = 0$ and different values of Decoder Error (f)

linear code. This is seen in figure, where we have plotted the amount of information saved as e varies from 0 to 0.5 for $\theta = 0$, $\frac{\pi}{12}$ and $\frac{\pi}{8}$.

In Figure 4.13, we have studied the characteristics of information saved against the computational fidelity of the decoder(we have studied for only $F_{decoding}$, since for the chosen error scheme in the decoder, faithfulness is always equal to one). This graph is a very useful as it helps us analyze the relationship between how much information is obtained to how well the decoder performed its operation. The graph is plotted for various values of e= 0, 0.1, 0.2 and 0.3 and f is varied in the interval [0,1]. As f increases, the fidelity of the operation decreases and as expected, with decrease in the decoder fidelity, there is a decrease in the amount of information saved. It is clear from the figure that more information about the input referent is saved when the decoder is performing the logical operation more correctly.

Since our focus is to study the amount of heat dissipation associated with achieving a certain amount of error correction, we will study the energy lower bound for corresponding information saved. The two quantities are plotted in Figure 4.14, for varying values of the decoder error (f) as the bit-flip error (e) varies between 0 and 0.5

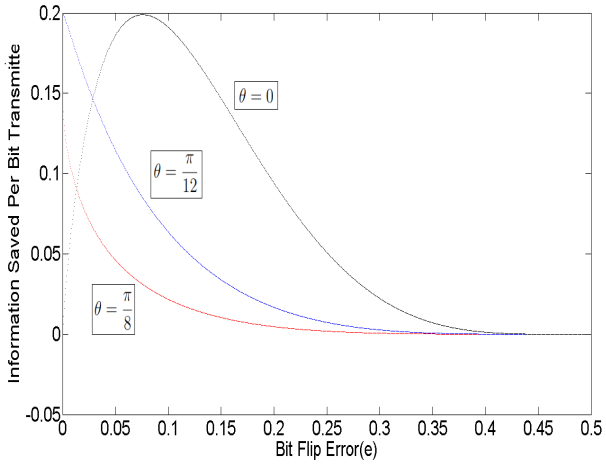

Figure 4.12. Information Saved per Bit Transmitted vs Channel Bit Flip Error (e) for Decoder Error (f)=0 and different values of θ

and θ=0. We can see that in all the cases, the amount of information saved increases reaches a maximum and then decreases but we continue to pay increasing amounts of energy. We see that after information loss reaches a maximum and starts decreasing, we are paying more amounts of energy for diminishing returns in the amount of information saved. This allows us to determine one region of diminishing returns, where use of the (n,k) linear code is not beneficial anymore. If e_{max} is the maximum error probability at which the linear code is to be used, then

$$\left(\frac{\partial\langle E^{\varepsilon}_{decoding}\rangle}{\partial\chi_{saved}}\right)_{e=e_{max}} = 0 \tag{4.137}$$

Thus using the linear (n,k) code for $e \geq e_{max}$ is disadvantageous to the user. Another region of diminishing return is identified from varying the decoder error(f). From Figure 4.14, we see that for non-zero f, there is a range of bit flip error (e) where we have $\chi_{saved} < 0$ but $\langle E^{\varepsilon}_{decoding}\rangle \geq 0$ indicating that we are dissipating more heat to obtain lower amounts of information compared to when not using coding at all. The

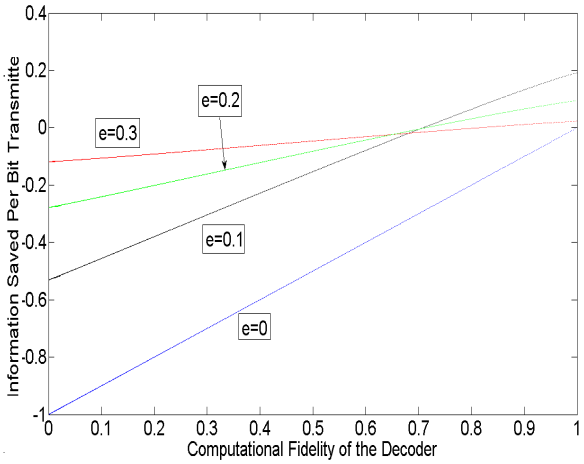

Figure 4.13. Information Saved per Bit Transmitted vs Decoder fidelity for θ=0 and different values of Decoder Error (f)

same can be plotted for other linear (n,k) codes and similar regions of diminishing returns can be identified. If e_0 indicates the maximum value of bit flip error (e) for which χ_{saved}¡0, for a given value of f and θ, we can calculate e_0 by solving for $\chi_{saved} = 0$. The range of e between e_0 and e_{max} indicates the best region to operate the decoder of a (n,k) linear code for a given values of f and θ, in terms of information saved and associated energy dissipation.

In Figure 4.15 above, we have $\langle E^{\varepsilon}_{decoding} \rangle$ vs χ_{saved} for both Hamming (7,4) and Hamming (8,4) code for $\theta = 0$ and decoder error (f)=0. Calculations for the Hamming (8,4) case were made assuming the same system state conditions as the Hamming (7,4) case, discussed in the previous section. From the graph we can see that the Hamming (8,4) code saves more amount of information than the Hamming (7,4) in many regions but also requires dissipation of greater amount of heat. There is also a range of e for which $\chi_{saved} < 0$ for the Hamming (8,4) code and represents a disadvantage when compared to the Hamming (7,4) which always has $\chi_{saved} \geq 0$.

Thus in this section we have analyzed the information loss and lower bound on the decoder dissipation in the decoder as a function of indistinguishability of states

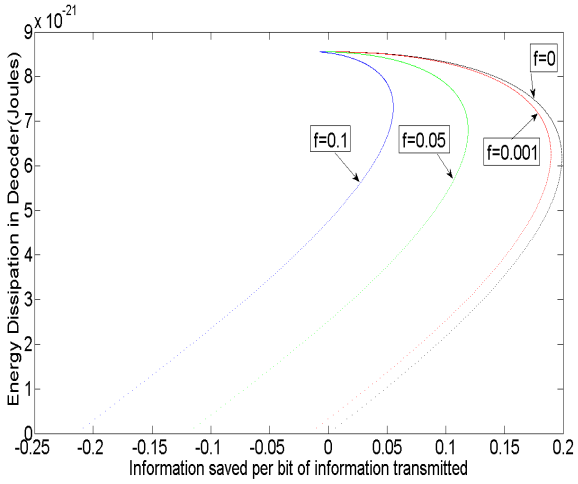

Figure 4.14. Minimum Energy Dissipation in Decoder (Joules) vs Information Saved per Bit of information Transmitted for θ=0 and Decoder Error (f)=0

θ, bit flip error (e) and decoder error (f). We then identified the need to define the new term information saved to understand how beneficiary the error correction technique is and related it to the corresponding energy cost and discovered regions of diminishing returns for the Hamming (7,4) and Hamming (8,4) codes in a special case.

4.4 Obstacles to Obtaining a Tight Lower Bound for Any (n,k) Linear Block Code

Before we proceed to the next chapter to summarize the work that has been carried out , we shall end this chapter discussing work that formed a major part of the thesis. The goal of this thesis is to study and develop a lower bound for energy costs associated with decoding when a (n,k) t-error correcting linear block code was used. The bound required is to be tight, yet independent of the code and hence the coding structure used, and to depend purely on the value of n, k, t and the BSC crossover probability e.

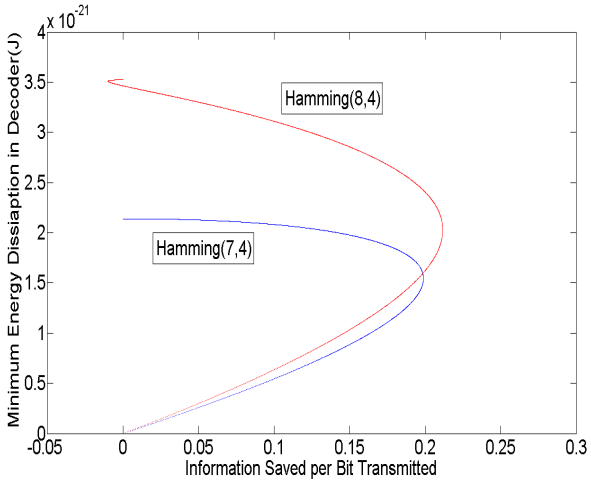

Figure 4.15. Minimum Energy Dissipation in Decoder (Joules) vs Information Saved per Bit of information Transmitted for θ=0 and Decoder Error (f)=0 in Hamming (7,4) and Hamming (8,4) case

In order to calculate the bounds, we required the weight distribution of a Code C which is a (n,k,d_{min}) code with minimum Hamming distance of d_{min} and error correcting capability of $t \leq \frac{d_{min}-1}{2}$. Linear codes are distance invariant, i.e. all the codewords have the same weight distribution, every codeword would see A_i codewords at a distance i. This is not necessarily true for non-linear codes and we require average weight distributions for them. Thus based on the values of n, k and d_{min} and the properties of linear codes, we would require a formula or method to determine the weight distribution of any linear code with these parameters. This would enable us calculate a tight lower bound quite accurately. However such a formula for the weight distributions is not available yet. This is because, in all error correction codes we would like to separate the codewords as much as we possibly can in order to prevent one codeword from being mapped into the other. This translates into what is called the coding theory problem which requires maximizing the distance between the codewords for given values of n, k and d_{min}. It is useful to think of this geometrically. If a binary vector of length n gives the coordinates of a vertex of a

unit cube in n dimensions.Then an (n,k,d_{min}) is a subset of these vertices. Thus the coding theory problem, in a geometrical language is to choose as many vertices of the cube as possible while keeping them a certain distance apart. This is in fact a packing problem, for if the code has a minimum Hamming distance of d_{min}. then the Euclidean distance between the codewords is $\geq \sqrt{d_{min}}$. Thus finding a (n,k,d_{min}) code means finding 2^k non-overlapping spheres of diameter $\sqrt{d_{min}}$ with centers at the vertices of the cube. The analogous problem of placing 2^k points on the surface of a unit sphere in n dimensions is unsolved [5].

What is available in literature are bounds like the Hamming bound, Singleton bound and Plotkin bound that provide bounds on the number of codewords of particular weights that can exist for any (n,k,d_{min}) linear code above a particular Hamming distance d_{min}. Bounds on the cost of decoding developed using these, as well as properties of linear codes are extremely inaccurate and can lead to predicting dissipation costs lower than what is actually possible. This is evident from Eq. (4.116), as for all the terms that need to calculated to determine the lower bound on energy dissipation, it is necessary to know the code structure used and not just the values of n, k, t and e. We only obtain weak bounds using just the values of n, k and t. Hence after exhausting possiblities of determining a bound for any (n,k,d_{min}) linear codes based on the values of n, k and d_{min}, the focus of this work was shifted to more specific Hamming codes.

CHAPTER 5

SUMMARY AND CONCLUSION

With shrinking CMOS device size not being a viable option to obtaining better computational units in terms of power and reliability, there is a need for new computational paradigms to be explored. Fault tolerance and error correction in these new paradigms to provide high accuracy in the computation is of utmost importance, and the energy cost the user must pay in order to do so must be thoroughly studied. In this thesis, error correction was implemented using a linear (n,k) code, and we have investigated the relation between the information lost during the decoding operation performed on noisy channel outputs and the corresponding heat dissipation involved, based on the work of Anderson in [12] and [13]. Based on these results, the thesis had been prepared as follows.

In Chapter 1, we have asserted our motivation and the significance of our work for the emerging technologies. In Chapter 2, we have made a brief introduction into the basic idea of error correction codes and some important parameters involved in them. Following this is introduction to the fundamental ideas of classical information theory.

In Chapter 3, we have introduced physical information theory, following which decoherence and Landauer's Principle has also been discussed. The referent approach to the physical information theory and the changes in ideas to entropic and energy forms of Landauer's principle are stated, with the derivations from [13] shown. Also included is the requirement and definition of computational efficacy measures introduced in [13] and their relationship to the information loss.

Chapter 4 contains a discussion of our system of interest and the formulation for information loss during the decoding operation. We then discover the need to generalize the computational efficacy measures to provide us with the necessary tools to characterize the decoder performance. We have followed it up with the formulation and verification of these generalized definitions for the performance measures on an example system. The results of the variation in fidelity and faithfulness with different parameters were presented and effects of system indistinguishability and environmental interactions on these efficacy measures were understood. Furthermore the relationship between computational fidelity and representational faithfulness of an entire multi-stage computation with that of the individual stages has also been derived. With the necessary mathematical tools at our disposal, we return to the problem of relating decoder performance to the energy dissipation. The new method of Referent Translation method is used to calculate the information loss in the decoder and the corresponding energy lower bounds in terms of the decoder's performance metrics. A formulation for general quantum system with a noisy decoder is first derived, and is then followed up by focusing on a more special case. The results of the thesis were then presented and analyzed. We started with the lower bounds that we have calculated for the pure state quantum system encoded using Hamming (7,4) code based on the formulation in the previous chapter. Variation in information lost, information saved and the associated energy dissipation in the decoder with variation in bit flip error (e), angle θ and decoder error (f) were calculated and studied. Regions of diminishing returns for the linear codes were identified and a comparison between Hamming (7,4) and Hamming (8,4) was carried out.

The approaches discussed in this thesis employ key fundamental concepts of quantum mechanics, physical information theory and error-correction codes relevant to and significant for the analysis and design of nanoscale devices for many years into the future. The goal of this thesis has been to provide a tight lower bound on the

energy cost the user has to pay to perform error-correction. It ties together two of the most important problems that most nanoelectronic devices are sure to face in heat dissipation and the presence of noise and defects. In the process, we developed computational measures to study the efficacy of systems performing computation for a wider range of scenarios like noisy inputs for example, which is very likely when dealing with such systems as well as the Referent Translation method which allows us to employ these efficacy measures. The noise could include both inherent quantum noise, as well as external thermal noise. They can also be used to quantify the performance of individual stages in a multi-stage computation. These efficacy measures provides us with a powerful tool to study and relate how well a system implements a logical operation and the power dissipated in doing so. This would enable us to determine regions of diminishing returns in terms of performance and energy dissipated and provide a better idea on whether a particular nanoelectronic proposal is feasible or not.

In conclusion, the contributions of this thesis are:

• Generalization of the computational efficacy measures- computational fidelity and representational faithfulness.

• Extension of the generalized efficacy measures to first, a two-staged logical computation and then a N-staged logical computation. Derivation of the relationship between the efficacy of the entire multi-stage computation with that of the individual stages.

• Introduction of the Referent Translation approach to allow definition of the required efficacy measures for the decoding problem.

• Determination of the lower bound on energy dissipation in the decoder, in terms of its efficacy measures for a Hamming (7,4) code..

• Determination of areas of diminishing performance returns in terms of information saved and energy dissipated in the decoder for the noisy Hamming (7,4) decoder.

BIBLIOGRAPHY

[1] I.R. Bahar, C. Lau, D. Hammerstrom, D. Marculescu, J. Harlow, A. Orailoglu, W.H. Joyner Jr., M. Pedram, "Architectures for silicon nanoelectronics and beyond," *IEEE Computer Society*, Vol. 40, Issue: 1, pp. 25-33, Jan 2007.

[2] J.R. Heath, P.J. Kuekes, G.S. Snider, R.S. Williams, "A Defect Tolerant Computer Architecture: Opportunities for Nanotechnology," *Science*, New Series, Vol.280, No.5370, pp. 1716-1721, June 1998.

[3] C.E. Shannon, "A Mathematical Theory of Communication," *The Bell System Technical Journal*, volume 27, pp.379-423, 623-656, July 1948.

[4] T.M. Cover and J.A. Thomas, "Elements of Information Theory," A Wiley-InterScience Publication, John Wiley and Sons, Inc., New York, 1991.

[5] F.J. Macwilliams and N.J.A. Sloane, "The Theory of Error-Correcting Codes," Volume 16, North Holland, New York, 1977.

[6] S. Lin and D.J. Costello, "Error Control Coding," 2nd edition, 2005, Pearson Prentice Hall, New York, 1977.

[7] C.J. Isham, "Lectures on Quantum Theory," Imperial College Press, New York, 1995.

[8] C. Bennett, "Notes on Landauer's Principle, reversible computation and Maxwell's Demon," *Studies in History and Philosophy of Modern Physics*, volume 34, pp. 501-510, 2003.

[9] S. Winograd and J.D. Cowan, "Reliable Computation in the Presence of Noise," MIT Press, Cambridge, 1963.

[10] N.G. Anderson, "Physical Information theory," ECE-697 Lecture Notes, University of Massachusetts, Amherst, Spring 2010.

[11] M.B. Plenio and V. Vitelli,"The physics of forgetting: Landauer's principle and information theory," *Contemporary Physics*, volume 42, no.1, pp. 25-60, 2001.

[12] N.G. Anderson, "Information erasure in quantum systems," *Physics Letters A*, Volume 372, Issue 34, pp. 5552-5555, 18 August 2008.

[13] N.G. Anderson, "On the physical implementation of logical transformations: Generalized L-machines," *Theoretical Computer Science*, 411, pp. 4179-4199, 2010.

[14] N.G. Anderson, "An Information-Theoretic Measure for the Computational fidelity of Physical Processes," *Proceedings of the 2008 IEEE International Symposium of Information Theory*, IEEE, pp. 2356-2360, 2008.

[15] N.G. Anderson,"Information Processing Artifacts," In preparation.

[16] M.A. Nielsen and I.L. Chuang, "Quantum Computation and Quantum Information," Cambridge University Press, New York, 2000.

[17] W.H. Zurek, "Environment Induced Superselection Rules," *Phys. Rev. D* Volume 26, Issue 8, pp. 1862-1880, 1982.