

Fall Semester 2010

Prof. Dr. Amos Lapidoth
Dr. ir. Jos H. Weber

Master Thesis

On Achieving the Minimum Energy for Sending One Single Bit with Feedback

Shahab Asoodeh

December 2010

Advisor: MSc. Ligong Wang

*To Haji and Soheila;
the bests in the world*

*and the soul of my Grandpa;
who always inspired me
and just left me*

Acknowledgments

I would like firstly to thank Amos Lapidoth for mentoring, support, encouragement, sharing his time with me and passion for research. His office's door was always open to me and my questions; TODA RABA. A big thank you also goes to Ligong for his helpful guidelines during this projects. I want also acknowledge the research freedom that Jos gave me when I was in Delft and also for his favors which enabled me to go easily to ETH for my project. I did major part of this thesis at ETH Zürich.

I owe another thank you to my friends Shagy, Măthias and Márin in Zürich with whom my stay in Zürich was so pleasant. The financial support of IDEA league is warmly acknowledged.

Last but not least, I am indebted, as usual, to my parents, Haji and Soheila and also my brothers Ehsan and Hesam, for their love, encouragement and unlimited selflessness.

I have consulted the greatest wizard and the greatest mathematician of all times and places. I asked them whether I would ever write these lines. The wizard said that I would not. The mathematician said that I would. Now, thanks to Popper I know they were both wrong. For if the former were right, these lines would not have ever been written; for if the latter were right, these lines would have been written long ago.

Delft, December 6, 2010

Shahab Asoodeh

Abstract

This master thesis was aimed to provide methods for coding construction which incurs the minimum energy when sending one single bit of information in the presence of noiseless feedback.

At first, we tried to find a coding scheme for general channels having a free symbol. Motivated by the original coding scheme proposed by Verdú in his paper which has to do with hypothesis testing and Stein lemma, we focused on sequential hypothesis testing (Wald testing) in the presence of feedback. It is proved that sequential hypothesis testing can not be exploited generally to achieve the minimum energy.

Next, we narrowed down the problem and considered only Gaussian channel. This thesis gave an optimal coding scheme -which incurs the minimum possible energy for sending one single bit- based on MMSE estimation. We then showed that in infinite-bandwidth Gaussian channel a single bit can be sent reliably with the same amount of energy that would be required per bit if one were sending a large number of bits.

In next step, we considered the transmission of a single bit over the discrete-time and also continuous-time Poisson channel with noiseless and delayless feedback. Firstly, the least amount of energy for sending reliably a single bit over Poisson channel was calculated and then a coding scheme which achieves it, was proposed. We showed that in the absence of peak-power constraints, sending reliably a single bit over Poisson channel requires, on the average at least, half the energy of a photon. We also extend these results for sending reliably finite number of bits. Finally, we showed that, in the absence of peak-power constraints, the amount of energy required to send reliably one bit is more than the amount of energy per bit when sending large number of bits. For the Poisson channel with peak power constraint, we also proposed two different coding schemes which can be used with bounded energy when error probability is bounded.

Zusammenfassung

Diese Diplomarbeit zielt darauf ab, Methoden zur Kodierungskonstruktion unter Erreichung der Kapazität pro Einheitskosten anzubieten. Zunächst versuchten wir, ein Kodierungsschema für allgemeine Kanäle zu finden, um ein Gratis-Symbol zu haben. Motiviert sind wir durch die ursprüngliche Kodierung, welche in der Arbeit von Verdú vorgeschlagen wurde. Diese hat mit dem Hypothesentest und Stein Lemma zu tun, dabei konzentrierten wir uns unter Präsenz von Feedback auf den sequentiellen Hypothesentest (Wald-Test). Im Anhang B ist bewiesen, dass der sequentielle Test nicht zur Erreichung der Kapazität pro Einheitskosten eingesetzt werden kann. Weiter wechselten wir die Methode auf Gauss-Kanäle und gaben ein optimales Kodierungsschema (im Sinne von Kapazität pro Einheitskosten) an. Dies basiert auf einer MMSE-Schätzung, welche als Sonderfall des Turin-Schemas erscheint. Weiter wurden Poisson-Kanäle in idealen und nicht-idealen Fällen berücksichtigt. Es wurde bewiesen, dass es kein Kodierungsschema gibt, welches die Kapazität pro Einheitskosten in Poisson-Kanälen erzielen kann, wenn die Kapazität pro Einheitskosten unendlich gross ist (dies passiert, wenn entweder der Dunkelstrom null ist oder keine Beschränkung der Spitzenleistung gibt).

Contents

Acknowledgments	V
Abstract	VII
Table of Contents	IX
List of Figures	XI
1 Introduction	1
2 Capacity Per Unit Cost and Feedback	9
2.1 Capacity Per Unit Cost	9
2.2 Communication with Feedback	15
3 Sequential Hypothesis Testing	18
4 Gaussian Channels	21
4.1 Elias Scheme	22
4.2 Gallager Feedback scheme for high SNR regime	24
4.3 Turin's Scheme	27
4.4 MMSE-based Coding Scheme	29
4.5 On-Off Feedback Scheme	34
5 Poisson Channel, Infinite Capacity Per Unit Cost	38
5.1 Introduction	39
5.2 Capacity per Unit Cost for Poisson Channel	43
5.3 Coding scheme in Discrete-Time Poisson Channels	46
5.4 Energy Analysis in Poisson Channels	53
5.5 Further Discussion on Main Result	65

6	Poisson Channel, Finite Capacity Per Unit Cost	69
6.1	One-bit Quantizer	70
6.2	Density-based Stopping Time	74
7	Concluding Remarks and Outlook	79
A	Improved PPM coding scheme with Feedback	83
B	Another Proof for Theorem 5.3	85
C	Publication	86
	Bibliography	91

List of Figures

1.1	Cost versus number of symbols required to transmit one bit of information through AWGN channel	4
5.1	The Markov chain regarding the coding scheme in Poisson channel; each state represents the input sequence X_j , $j = 1, 2, \dots, N$ and also the state S represents the state of stopping before transmitter reaches N , i.e., $j = N$	50
5.2	The spent energy versus number of transmission for different schemes; constant-input signal, linearly and logarithmically increasing-input signal, when the termination step or N varies from $N = 500$ to $N = 1500$. The energy spent are formulated in (5.11), (5.12) and (5.13).	54
5.3	The spent energy versus number of transmission for different schemes; constant-input signal, linearly and logarithmically increasing-input signal, when the termination step or N varies from $N = 2000$ to $N = 15000$. The energy spent are formulated in (5.11), (5.12) and (5.13).	55
5.4	The comparison of spent energy in three previously mentioned coding schemes with decreasing-input scheme; when the termination step varies. $N = 500, 1000, 2000, 3000$. The spent energy in decreasing-input scheme is shown in (5.14).	56
6.1	The energy consumption and error probability using the proposed coding scheme. Note that α should be always smaller than A . . .	73
6.2	The energy consumption and error probability using the second proposed coding scheme.	75

Chapter 1

Introduction

Missing a train is only painful if you run after it!.

—N. N. Taleb

This master thesis looks at one of the important features of communication, namely, energy. In this project the minimum energy for sending one single bit over two well-known channel families; Gaussian and Poisson, is firstly computed and then for both of these two channels, the optimal coding scheme (in the sense of energy) for sending one single bit is proposed.

The explosive growth and widespread application of high rate wireless communication have prompted waves of research and standard development activities. While semiconductor processing speed has been increasing exponentially, doubling almost every two years, processor power consumption also continues to grow by 150% every two years. In contrast, advances in battery technology have not kept pace, with capacity increasing at a modest rate of 10% every two years. This leads to an increasingly large gap between power need and battery capacity. Energy efficiency, therefore, is becoming an increasingly important factor for wireless communication. Hence a practical problem of broad interest is to transmit a message with minimum energy by which the life time of batteries can be prolonged and thus saving money. In the conventional communication -which has been so far considered in information theory- a long codeword is transmitted for each message. Suppose k bits are to be transmitted over the channel, then we encode it into a n -bit codeword and send it. For channels with additive white Gaussian noise, Shannon [1] has demonstrated that for transmitting

k bits with probability of block error ϵ using total energy E and n degrees of freedom ¹, the minimum energy per bit, $E_b = E/k$, in the limit $k \rightarrow \infty$, $k/n \rightarrow 0$ and $\epsilon \rightarrow 0$ tends to the fundamental limit of $(E_b)_{min} = N_0 \ln 2$ where $N_0/2$ is the noise power per degree of freedom. Instead of allowing $k/n \rightarrow 0$, frequently one fixes the rate $R = k/n$, in which case E_b becomes a function of R and sensitive to modulation [2].

Obviously the assumption of infinite number of information bits (k) and transmission symbols (n) is not realistic, because at the decoder part we have to wait for infinite time to get all symbols and also we have to have an infinite-size buffer to memorize all received symbols. To solve the latter problem (long buffer) we consider sending a *finite* number of information bits. Going beyond the asymptotic limit $k \rightarrow \infty$, the minimum E_b under fixed rate R for finite k has been studied numerically [3] and [4] and tightly approximated [5] and [6].

In this project we consider sending one *single* bit over two well-known and most applicable channel families; Gaussian and Poisson when noiseless feedback is present. Gaussian channel is so applicable that can model almost all additive-noisy wireless communication. Poisson channel is too used to model the optical communication using direct detection which is assumed in many optical communication systems ².

The main reasons why the problem of sending one single bit is considered in this project are briefly as follows:

- In some communication we need to transmit only a bit with high accuracy, for example "fire" or "not fire" in military or "emergency state" or "normal state" in e-health communication.
- When sending one single bit, the decoder part does not have to buffer formerly received bits. In this case, after (large and sometimes random) time, decoder learns that the transmitted bit was 1 or 0 without any buffer.

¹Throughout this thesis, by degree of freedom we mean the number of transmissions or number of channel uses for each message being transmitted.

²Direct detection (or incoherent) is a way of decoding in optical communication which basically relies on the collection of enough signal energy at the receiver.

- The idea of sending one single bit will trigger the research on minimum energy analysis to send *finite* number of bits in general channels as it has already been addressed for Gaussian channel in [7]. As first step we analyzed the energy analysis for k bits in Poisson channels in section 5.5. There has been an open-problem since 2007 about "*what is the minimum energy for sending finite number of bits over general channels and how to achieve it?*".

Verdú in [8] generalized the concept of energy with the notion of *cost function*. Cost function, $b[x]$, assigns a non-negative number to each input x in input alphabet. The cost corresponds to energy if $b[x] = x^2$ in Gaussian channel and if $b[x] = x$ in Poisson channel. The capacity-cost function, $C(\beta)$, is defined as the maximum number of bit per symbols that can be transmitted reliably over channel with average cost per symbol not exceeding β . It is quite well known that the capacity-cost function for Gaussian channel having variance σ^2 and cost function as energy, i.e., $b[x] = x^2$, is $C(\beta) = \frac{1}{2}\log(1 + \beta/\sigma^2)$. Clearly $1/C(\beta)$ and $\beta/C(\beta)$ are the number of required symbols and its associated cost (viz. energy) to transmit one bit of information. Figure 1.1 depicts the cost per bit versus number of symbols per bit for a discrete time additive white Gaussian channel. It is obvious in this figure that the cost per bit blows up if more than one bit is to be transmitted per symbol, while the cost per bit is close to its infimum for more than 3 or 4 symbol-per-bit. This figure also shows that the minimum energy per bit is $N_0\ln 2$ as theoretically shown in [9] and [1].

The minimum cost incurred by the transmission of one bit of information through the channel is a fundamental figure of merit showing the most frugal and economical way to proceed the reliable communication. Its reciprocal is called "*capacity per unit cost*". Therefore as a benchmark for energy consumption in different transmission schemes, it is crucial to investigate the maximum number of bits per unit cost that can be transmitted over the channel (i.e., capacity per unit cost) in lieu of conventional channel capacity, especially when the main limitation is on the input cost rather than on the number of degrees of freedom. Hence, the search for most energy-efficient coding scheme (when k is large) is turned into the search for coding scheme which can achieve the capacity per unit cost.

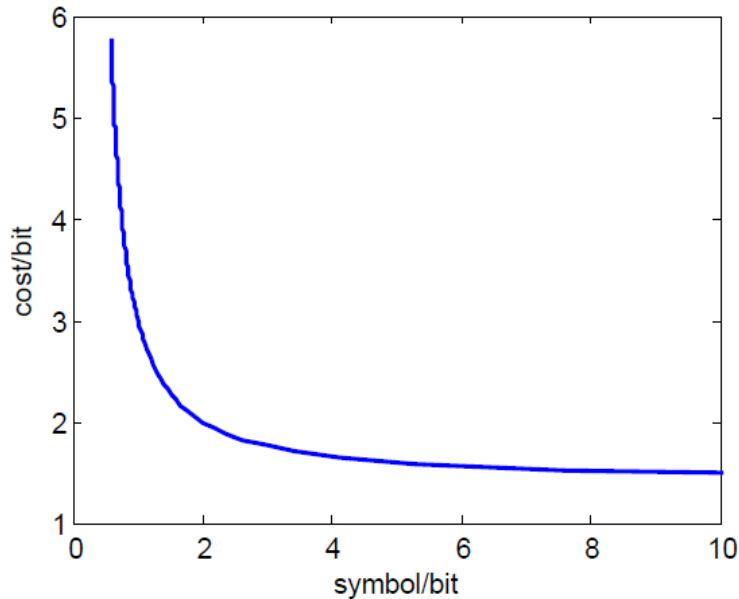


Figure 1.1: Cost versus number of symbols required to transmit one bit of information through AWGN channel

The capacity per unit cost of a general memoryless channel was computed by Verdú in [8] using the capacity-cost function. The capacity has a particularly simple form when one of the input symbols is of zero cost. Since the capacity of a memoryless channel is not increased when causal feedback from the receiver to the transmitter is available, this is also the capacity per unit cost in the presence of feedback. However, feedback sometimes greatly simplifies the coding schemes that are needed to achieve capacity.

This project was firstly intended to find a coding scheme that achieves the capacity per unit cost in the presence of feedback for general channels. Motivated by original coding scheme proposed by Verdú, we started by sequential hypothesis testing [10]. In this technique the blocklength (the size of codeword) is treated as a random variable (i.e., Markov stopping time) so that the resulting scheme would be variable block-length code. The sequential hypothesis testing is based on Wald identity which is stated in chapter 3. In this scheme receiver keeps monitoring the log-likelihood ratio, S_n , in each step and compares it with two constants, $A \geq 0$ and $B \leq 0$, which are considered as boundaries of ambiguity area. The receiver sends the trans-

mitter the signal "ready to decode" by which the transmission is stopped, when either $S_n > A$ or $S_n < B$. The main theorem regarding the performance of coding scheme based on sequential hypothesis testing is shown in chapter 3. We then analyzed the minimum energy per bit in two particular channels; Gaussian and Poisson.

Feedback coding scheme in Gaussian channel has so far undergone extensive researches, in various guises. Elias [11] presented a feedback coding scheme to transmit a normally-distributed source symbol over Gaussian channel whose mean-square estimation error goes down geometrically with n . Schalkwijk and Kailath (SK) [12] proposed another feedback scheme based on stochastic approximation procedure. SK scheme is defined in this thesis in terms of Elias scheme together with PAM modulation. This helps us to better understand the coding scheme which was recently proposed by Gallager for high SNR regime whose needed energy is smaller than 5 and error probability at time i decreases as an i th order exponent [13]. This coding scheme is briefly described in chapter 4. A year before SK scheme was published, Turin proposed a general class of coding scheme which could provably achieve the capacity per unit cost, and consequently, the minimum energy per bit. His scheme was based on continuous-time analysis of sequential hypothesis testing and was proved to achieve the conventional capacity of continuous-time Gaussian channel with unlimited degrees of freedom which equals the capacity per unit cost of equivalent discrete-time channel. Turin in [14] has also conjectured that binary feedback coding scheme (sending 0 or 1 when transmitted bit is 1 and 0 or -1 when transmitted bit is 0 depending on the previously received symbols) can also achieve the capacity per unit cost in Gaussian channel but he could not later prove this conjecture in [15]. His analysis showed that binary feedback coding scheme can achieve $C \ln 2$ where C is the capacity per unit cost. This conjecture has been recently proved to be optimal from capacity per unit cost viewpoint, by Polyanskiy *et al.* [7].

This thesis has also proposed another feedback coding scheme based on Minimum Mean Square Error (MMSE) estimation. This technique has been proved to achieve the minimum energy for sending reliably one single bit. Hence, in Gaussian channel for sending one bit of information, to save energy and money at best case we should use MMSE estimation at the receiver part.

Coding scheme for the Poisson channel has been studied by Wyner in [16] without feedback. In this seminal contribution, he has also computed the reliability function (i.e., the exponential behavior of the probability of error for the best code, as the coding delay is increased with the rate held fixed) of the Poisson channel without feedback. The effect of feedback on capacity of Poisson channel has been further explored by Frey [17] who considered channels with time-varying dark current³ and showed that as long as the dark current is deterministic, capacity remains unchanged by the introduction of feedback. Consequently, feedback can not change the capacity per unit cost of Poisson channel. In this thesis we always assume that dark current is deterministic and time-invariant. Lapidoth [18] proposed a feedback coding scheme and proved that it can achieve the Sphere Packing error exponent which is known to be unachievable without feedback for rates below the critical rate.

To our best knowledge, the capacity per unit cost and/or minimum energy analysis for Poisson channel has not so far been considered. We considered this problem in chapters 5 and 6 depending if the capacity per unit cost is finite or infinite. In chapter 5, we firstly compute the minimum energy required for sending one single bit reliably and then propose the optimal coding scheme which achieves the minimum energy. We then extend this scheme to achieve the minimum energy for sending finite k bits. In chapter 6, in finite capacity per unit cost, we give two different techniques which are easily implementable. In this chapter, we showed that if error probability is bounded, the energy can be saved by using a one-bit quantizer. The optimal coding scheme in this case is still an open-problem.

It is worthwhile to point out, at final, that energy for sending reliably one single bit may be totally different from the energy *per bit* for sending reliably large number of bits. These are two parameters that have been identified equal for a couple of years (because they are the same in Gaussian channel).

In a nutshell, this thesis is organized as follows:

³In Poisson channel, environmental noise and radiation show themselves as dark current. If dark current is zero, the Poisson channel can be modeled as Z-channel and is called *ideal*. This term is explained in details in chapter 5.

- Chapter 2 presents some relevant introductory concepts on feedback coding and also on capacity per unit cost. There, the achievability of capacity per unit cost is shown using Stein lemma.
- Chapter 3 gives a feedback coding scheme based on sequential hypothesis testing and then proves its performance from the energy point of view.
- Chapter 4 reviews some well known feedback coding schemes in Gaussian channel including: Elias scheme and M-PAM version of Schalkwijk-Kailath scheme; both of which can not achieve capacity per unit cost. General Turin's scheme in continuous-time has also been presented. The more recent result on binary feedback coding scheme is given in this chapter. Both these two schemes are shown to be achieving capacity per unit cost. Finally our new coding scheme is proposed and then proved to be optimal.
- Chapter 5 computes the least amount of energy for sending reliably one single bit over Poisson channel (when capacity per unit cost is infinite) and then gives an optimal coding scheme to achieve it. This chapter does an enlightening comparison between the energy for sending reliably one single bit and the energy per bit when sending $k \rightarrow \infty$ bits.
- Chapter 6 extends the techniques of previous chapter for the case when capacity per unit cost is not infinity. In this chapter two different and easily implementable techniques are given. Both are good when the error probability is bounded (say, less than 10^{-5}) and both are variable-length codes. Hence the decoding delay ⁴ is random and its expected value is therefore less than fixed-length codes. Their performances, energy consumption and error probabilities are given in this chapter.
- Chapter 7 presents the conclusions and some further topics of research, besides the proposition of some open questions that arose during this project.

⁴The time that decoder should wait till it can decide.

Chapter 2

Capacity Per Unit Cost and Feedback

A nice adaptation of conditions will make almost any hypothesis agree with the phenomena. This will please the imagination but does not advance our knowledge.

–J. Black

2.1 Capacity Per Unit Cost

A communication channel is characterized by an input alphabet $\mathcal{X}$, output alphabet $\mathcal{Y}$ and also output distribution given input is known or transition laws, i.e., $\Pr(Y = y|X = x)$. The n inputs are instances of the input process represented probabilistically by the random vector $X^n \triangleq (X_1, X_2, \dots, X_n)$ with joint probability function $P_{X^n}(x^n) \triangleq \Pr(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n)$ and also similar representation for output instances, $P_{Y^n}(y^n) \triangleq \Pr(Y_1 = y_1, Y_2 = y_2, \dots, Y_n = y_n)$. In the practical communication input is usually subjected to a constraint which is modeled by a cost function. Cost function can take any non-negative number, $b : \mathcal{X} \rightarrow \mathbb{R}_0^+ = [0, \infty)$ which associates a non-negative number $b[x]$ to each element x of the input alphabet. As we intend to make use of the channel many times in succession, we interpret the cost of sending x^n as an additive cost, i.e., $b[x^n] = \sum_{i=1}^n b[x_i]$. In general, the probability of sending each possible input vector, x^n , need not be equal. Then the average cost of n -dimensional random vector is $E[b[X^n]] = \sum_{i=1}^n E[b[X_i]]$. An n -dimensional random vector X^n is said to be β -admissible if it satisfies $E[b[X^n]] \leq \beta$. The maximum number of bits per symbol that can be transmitted over the channel with average cost

per symbol not exceeding β is the capacity-cost function, $C(\beta)$. The more mathematically formal definition of capacity-cost function is as follows.

Let (n, M, ν, ϵ) define a code in which n is blocklength, M is the number of codewords, each codeword $(x_{i,1}, x_{i,2}, \dots, x_{i,n})$, $i = 1, 2, \dots, M$ is ν -admissible and the average (over the ensemble of the equiprobable messages) probability of decoding the correct message is better than $1 - \epsilon$.

Definition 2.1.[19]: Given $0 \leq \epsilon < 1$ and $\beta > 0$, a non-negative number R is an ϵ -achievable rate with cost per symbol not exceeding β , i.e., $n\beta$ -admissible, if for every $\gamma > 0$ there exists n_0 such that if $n \geq n_0$, then an $(n, M, n\beta, \epsilon)$ code can be found whose rate satisfies $\log M > n(R - \gamma)$. Furthermore, R is said to be achievable if it is ϵ -achievable for all $0 < \epsilon < 1$. Then the maximum achievable rate, with cost per symbol not exceeding β is the channel capacity denoted by $C(\beta)$. Generally the function $C : \mathbb{R}_0^+ \rightarrow \mathbb{R}_0^+$ is referred to as the capacity-cost function.

Gallager in [20] has proved that the channel capacity-cost function can be computed using the following formula:

$$C(\beta) = \sup_{\text{all } n\beta\text{-admissible } X^n} I(X^n; Y^n)$$

where $I(.,.)$ denotes the mutual information and the supremum is taken to be zero if there is no $n\beta$ -admissible X^n .

It is also proved in [19] that $C(\beta)$ is concave and nondecreasing, so we can trade off the number of symbols (or channel use) and the cost it takes to send every bit of information. The minimum cost incurred by the transmission of one bit of information through the channel is a fundamental criterion to determine the best and easiest way (from economical point of view) to communicate reliably. This quantity is so important that its reciprocal is given a name, *capacity per unit cost* which is defined similarly to the conventional capacity, except here we have to divide the logarithm of number of codewords to their cost. It is of great interest for designers (and anyone who wants to save energy and money) to investigate the maximum number of bits per unit cost.

Then, channel capacity per unit cost is formally defined in [8] as following:

Definition 2.2. Given $0 \leq \epsilon < 1$, a nonnegative number¹ $\mathbf{R}$ is an ϵ -achievable rate per unit cost if for every $\gamma > 0$, there exists $\nu_0 \in \mathbb{R}_0^+$ such that if $\nu > \nu_0$, then an (n, M, ν, ϵ) code can be found with $\log M > \nu(\mathbf{R} - \gamma)$. $\mathbf{R}$ is achievable per unit cost if it is ϵ -achievable per unit cost for all $0 \leq \epsilon < 1$ and the capacity per unit cost is the maximum achievable rate per unit cost.

The following theorem, proved by Verdú, links the capacity per unit cost to capacity-cost function introduced before [8].

Theorem 2.1 The capacity per unit cost of a memoryless channel is equal to

$$C = \sup_{\beta > 0} \frac{C(\beta)}{\beta} = \sup_X \frac{I(X; Y)}{E[b[X]]}$$

where $I(X; Y)$ denotes the mutual information and $E[X]$ denotes the expectation with respect to random variable X .

It is obvious that if $b[x] = 1$ for all $x \in \mathcal{X}$, then the channel capacity per unit cost reduces to the conventional Shannon capacity of unconstrained channel. In other words, theorem 2.1 is a generalization of conventional capacity, until no further restriction is imposed on cost function. It turns out that having a free symbol (a symbol whose cost is zero) renders the analysis much easier and leads to the formula which is no longer dependent on capacity-cost function. Suppose there exists an input symbol, x , whose cost is zero, $b[x] = 0$ and let label this symbol as 0 hereafter. Verdú in his seminal paper [8] proved the following theorem which brings significant ease over the theorem 2.1 as it requires the computation of $C(\beta)$ in order to calculate capacity per unit cost. But before the theorem, we need to define the *Kullback-Leibler divergence* for two probability mass function p and q over the same alphabet $\mathcal{X}$, as follows

$$D(p||q) = \sum_{x \in \mathcal{X}} p(x) \log \frac{p(x)}{q(x)} = E_p \left[\log \frac{p(X)}{q(X)} \right].$$

where $E_p[\cdot]$ means the expectation with respect to probability mass function $p(x)$.

¹Information per symbol and information per unit cost are different and therefore written differently here, former by lightface and latter by boldface letters.

Theorem 2.2. *If there is a free input symbol, let's say 0, then the capacity per unit cost of a memoryless channel is equal to*

$$C = \sup_{x \in \mathcal{X}'} \frac{D(P_{Y|X=x} || P_{Y|X=0})}{b[x]}$$

where $P_{Y|X=x}$ denotes the output distribution given the fact that the input was x , $D(P||Q)$ is the Kullback-Leibler divergence between two probability distributions P and Q and also $\mathcal{X}'$ is the input alphabet excluded free symbol, $\mathcal{X}' = \mathcal{X} - \{0\}$.

Since we will need the technique used in the Verdú's proof of achievability, we now proceed with this proof. But before we proceed with proof, an important point should be mentioned. In the presence of free input symbol we can write

$$C = \sup_{\beta > 0} \frac{C(\beta)}{\beta} = \lim_{\beta \downarrow 0} \frac{C(\beta)}{\beta}$$

This can be clarified as follows:

Suppose two different average power constraints β_1 and $\beta_2 > \beta_1$ are given, so we can write

$$\beta_1 = \frac{\beta_1}{\beta_2} \cdot \beta_2 + (1 - \frac{\beta_1}{\beta_2}) \cdot 0$$

hence the concavity of $C(\beta)$ implies

$$\begin{aligned} C(\beta_1) &\geq \frac{\beta_1}{\beta_2} C(\beta_2) + (1 - \frac{\beta_1}{\beta_2}) C(0) \\ &= \frac{\beta_1}{\beta_2} C(\beta_2) \end{aligned}$$

the second line is due to the fact that $C(0) = 0$. The inequality $\frac{C(\beta_1)}{\beta_1} \geq \frac{C(\beta_2)}{\beta_2}$ can thus be obtained which proves that $\frac{C(\beta)}{\beta}$ is monotonically non-increasing function in β , so its supremum happens in β_{min} which is zero.

Now let's proceed with the proof for achievability of theorem 2.2.

Proof. Fix $x_0 \in \mathcal{X}'$ and integers $M, N > 0$. We build a matrix of size $M \times N$ which correspond to codeword $m \in \{1, 2, \dots, M\}$. The columns in this matrix are simply the same and each element in the column is constructed as follows:

$$r_i = \begin{cases} x_0, & i = m \\ 0, & i \neq m \end{cases}$$

Hence we have only one nonzero element in each column and only one non-zero row. The blocklength will therefore be $n = MN$ and consequently $\nu = Nb[x_0]$. In the receiver side, decoder observes a matrix $\{y_{ij}, 1 \leq i \leq M, 1 \leq j \leq N\}$ of size $M \times N$ which he knows it consists of all-zero rows and one all-one row, the task of decoding is just to determine the position of all-one row. This problem translates into binary hypothesis testing where, given the codeword m was transmitted, the elements in m th row, $y_{ij}, i = m$, have distribution $P_{Y|X=x_0}$ and the other elements have distribution $P_{Y|X=0}$. Putting in different way, decoder gets N i.i.d samples for each r_i , hence M binary hypothesis testing can be carried out where the null and alternative hypotheses are defined as follows:

$$\begin{aligned} H_{i0} : r_i &= 0 \\ H_{i1} : r_i &= x_0 \end{aligned}$$

As a hypothesis testing, we have to deal with two different kind of error denoted by α_N and β_N and defined as below

$$\begin{aligned} \alpha_{iN} &= \Pr[\hat{r}_i = x_0 | r_i = 0] \\ \beta_{iN} &= \Pr[\hat{r}_i = 0 | r_i = x_0] \end{aligned}$$

where $\hat{r}_i$ is what decoder guess about r_i , bigger than a threshold x_0 and smaller than threshold 0 is guessed. The threshold is chosen in a way that $\beta_{iN} \leq \epsilon/2$. The message is declared to be m if

$$\hat{r}_i = \begin{cases} x_0 & i = m \\ 0 & i \neq m \end{cases}$$

and error will be declared if either no or more than one non-zero r_i is detected. The proof concludes with the largest values of messages, M , which remains the error probability smaller than fixed ϵ . Since the error probability for the case when message m is sent, P_m , is independent of the value of m , we can assume that message 1 has been sent, as we always do in the

proof of Shannon channel capacity. Hence we can upper bound the error probability P_1 as follows

$$P_1 \leq \beta_{1N} + (M - 1)\alpha_{1N} \quad (2.1)$$

Using Stein lemma [21] which is about error probability of Neyman-Pearson hypothesis testing, we can get for sufficiently large N and for every $\gamma > 0$

$$\alpha_{1N} \leq \exp[-ND(P_{Y|X=x}||P_{Y|X=0}) + N\gamma] \quad (2.2)$$

Hence having inequalities 2.1 and 2.2 at hand, if the following inequality holds for the rate per unit cost (i.e., logarithm of number of messages over their cost)

$$\frac{\log M}{Nb[x_0]} < \frac{D(P_{Y|X=x}||P_{Y|X=0})}{b[x_0]} - \frac{2\gamma}{b[x_0]}$$

we can make sure that $P_1 \leq \epsilon$ for sufficiently large N , and therefore $D(P_{Y|X=x}||P_{Y|X=0})/b[x_0]$ is an achievable rate per unit cost. This concludes the proof of achievability.

□

Remarks

- Although the codebook here consists of codewords whose lengths are of order $O(M \log M)$, Verdú has proved that allowing the codeword's block length to grow logarithmically with M will also lead to achieve the same rate per unit cost.
- This scheme is an orthogonal signaling, i.e., no nonzero component overlap, but Sahai proposed a semi-orthogonal signaling which also achieves the capacity per unit cost [22]. Indeed his idea is the repeated Pulse Position Modulation (PPM) in which previous symbols plus new one are transmitted in each channel use. Although he did not incorporate feedback in his scheme, he claimed that feedback can reduce greatly the complexity of encoder while all distance properties and expected probability of error are held unchanged.

Using this formula, the capacity per unit cost for Gaussian channel will be $1/N_0 \ln 2$ which coincides with result of Golay [9] in 1939 that the minimum energy per bit for Gaussian channel is $N_0 \ln 2$.

In the PPM scheme above, the entire $M \times N$ matrix has to be sent as a block and then a suboptimal yet simple independent Neyman Pearson hypothesis testing are performed on each of the M rows based on all N elements of the rows and finally a decision is made. Feedback can dramatically simplify the complexity of the encoding and decoding. We came up with an improved PPM coding scheme with feedback which is explained in Appendix A.

2.2 Communication with Feedback

In this section we give a brief overview on main definitions in communication schemes with feedback which will later be used often.

Schalkwijk and Kailath (SK) [12] and [23] developed a block coding scheme in AWGN channels with an average energy constraint on the input when a noiseless feedback is present. Interestingly, they showed that the probability of error can be made to decay double exponentially in the block length. Kramer [24] showed with an intermittent feedback scheme that the probability of error can be made to decay N -fold exponentially. More recently, Gallager and Nakiboglu [13] proposed a feedback coding scheme for high SNR regime which yielded N -fold exponential probability of error with finite energy which is proved to be smaller than 5. The extremely good performances of SK and SK-like schemes are directly dependent on having unlimited peak power available. This allows them to correct the errors which happen from strong and rare channel behaviors compared to channels having peak power constraint. Indeed, it has been shown in [25] and [24] that if the channel input has peak power constraint, SK scheme will collapse to single exponentially error decay. Schalkwijk and Barron [26] proposed a sequential² coding scheme for AWGN channels with a peak energy constraint. In their coding scheme, information is revealed to the encoder through a random process incurring variable delay. It is worthy to point out that Kim *et al.*, [27] have recently shown that even if the feedback is

²In sequential coding, the information source reveals the information, which is supposed to be transmitted, to the encoder progressively.

ideal except for additive Gaussian noise, then the error probability decreases only as a single exponential in blocklength, although the exponent increases with increasing signal-to-noise ratio in the feedback channel.

Let us precisely define the main entities which are needed to understand feedback coding strategies.

Definition 2.3. (Stopping Time). *An integer values random variable τ is called stopping time with respect to a random sequence $\{X_i\}_{i=1}^{\infty}$ if, conditioned on the sequence X^n , the event $\{\tau = T\}$ is independent of future $\{X_i\}_{i>n}$, more mathematically, $\{\tau \leq n\} \in \mathcal{F}_n$ where $\mathcal{F}_n$ is the minimum σ -algebra generated by X^n .*

Definition 2.4. (Feedback Block encoder). *Given a message $W \in \mathcal{W}$, a block encoder with feedback is shown as a sequence of encoder functions*

$$\Phi = \{f_n : \mathcal{W} \times \mathcal{Y}^{n-1} \rightarrow \mathcal{X}\}$$

where $\mathcal{Y}^{n-1}$ represents the previous received symbols fed back to transmitter through noiseless feedback link. Encoder sequence is linked to stopping time T defined by the previous outputs and then the encoder sequence will be $\{f_i\}_{i=1}^T$.

Definition 2.5. (Feedback Block Decoder). *a block decoder is shown as a sequence of decoder functions as follows*

$$\Psi = \{g_n : \mathcal{Y}^n \rightarrow \mathcal{W}\}$$

Which is linked to the stopping time T defined by output sequence $\{Y_i\}$. Then decoder message will be $\hat{W} = g_T(Y^T)$.

Any feedback coding scheme is illustrated by triple $C = (\Phi, \Psi, T)$. Scheme C may specify a fixed length or variable length coding scheme depending on the definition of stopping time. Error probability for code C is the probability that the decoded message does not match the message transmitted, averaged over all possible transmitted message. Hence error probability can be written as

$$P_e(C) = \frac{1}{|\mathcal{W}|} \sum_{w \in \mathcal{W}} \Pr(g_T(Y^T) \neq w | w \text{ was sent})$$

Variable length block code can be best understood by an example on erasure channel. Feedback presence can come up a possibility of a block code so called "*resend until non-erasure*".

Example 2.1 (Variable length block code in erasure channel). Consider a message set of size 2^k . Each message is represented as an k -dimensional sequence, $w = \{w_i\}_{i=1}^k$ where $w_i \in \{0, 1\}$. The encoded sequence is obtained by applying following encoder sequence

$$f_n(w, y^{n-1}) = w_j$$

where $w_j = 1 + \sum_{i=1}^{n-1} 1_{\{y_i \neq e\}}$. The decoding time is the first time that the decoder sees k non-erased symbols at the channel output, so

$$T = \min \left\{ n : \sum_{i=1}^n 1_{\{y_i \neq e\}} = k \right\}$$

The decoder is sequence of functions $\{g_n\}$ which output further samples until k non-erased symbols have been observed;

$$g_n = \begin{cases} \text{further sample} & \sum_{i=1}^n 1_{\{y_i \neq e\}} < k \\ \hat{w} = \{y_{\bar{e},1}, y_{\bar{e},2}, \dots, y_{\bar{e},k}\} & \text{otherwise} \end{cases}$$

where $y_{\bar{e},i}$ means the i th non-erased symbol of the channel output. This coding scheme has obviously variable blocklength and zero error probability. Stopping time, T , is the sum of independent geometric random variables, τ_i , representing the required expected time to transmit each bit reliably through erasure channel. Therefore decoding delay is $T = \sum_{i=1}^k \tau_i$ and then it can be concluded that $E[T] = \frac{k}{1-\epsilon}$.

Chapter 3

Sequential Hypothesis Testing

If people do not believe that mathematics is simple, it is only because they do not realize how complicated life is.

–John L. von Neumann

This chapter concerns the attempts that we made to achieve the capacity per unit cost for general channels. The general coding scheme proposed in the original paper of Verdú [8] motivated us to work on sequential hypothesis testing in order to find the general feedback coding scheme which can achieve capacity per unit cost. One drawback of hypothesis testing is its dependence on assumption of N i.i.d samples which means the constant input sequence in the coding context. We know that this is not optimal in Gaussian channel and it is in Poisson channel. Hence this method fails to give the optimal coding scheme for general channel, but however we tried it to see what is the best energy that we can get from this technique. Before we proceed with the main theorem, Wald identity which is the main ingredient of sequential hypothesis testing, is described.

Theorem 3.1 (Wald Identity). *Let $y_1, y_2, \dots$, be i.i.d samples with mean value $\mu = E[y_1]$. Let M be any integer-valued random variable such that $M = n$ is measurable on $y_1, y_2, \dots, y_n$ for all $n = 1, 2, \dots$ and assume that $E[M] < \infty$ then $E\left[\sum_{k=1}^M y_k\right] = \mu E[M]$.*

Proof. Let S_n denote the sum $y_1 + y_2 + \dots + y_n$, then the expected

value of sum of random number of i.i.d samples can be written as follows:

$$E[S_M] = E[E[S_M|M]] = E[ME[y_1]] = \mu E[M]$$

Note that when $E[S_M]$ is conditioned on M , M can be treated as deterministic value.

□

Suppose that we keep sending free symbol for H_0 and a constant non-free symbol from input alphabet for H_1 . Let us put the encoding and decoding algorithm in the purely mathematical way in what follows. Suppose $y_1, y_2, \dots$ are i.i.d observations from probability distribution f which may be f_0 or f_1 . Assume that if f_0 is true distribution, sampling costs nothing but if f_1 is true, sampling costs something fixed. Our preferred action is to observe $y_1, y_2, \dots$ *ad infinitum* when f_0 is true and to stop sampling as soon as possible. Hence the test of H_0 against H_1 is the stopping time T , i.e., if $T < \infty$, H_0 is rejected and if $T = \infty$, H_0 is accepted. We need to seek the stopping time for which $P_0(T < \infty) \triangleq \Pr(T < \infty|H_0)$ is minimum and also $E_1[T] \triangleq E[T|H_1]$ is also minimum. Minimizing both is impossible, because they are inversely related to each other so we need to keep $P_0(T < \infty)$ small and then get $E_1[T]$ minimized. The best candidate for this goal is one-sided Wald sequential test, because it is well-known that among all sequential and non-sequential tests, Wald test minimizes sample size [10].

Getting back to log-likelihood ratio as

$$S_n = \sum_{j=1}^n \log \frac{f_1(y_j)}{f_0(y_j)}$$

we can define the stopping time as follows: $T(b) \triangleq \inf\{n : S_n \geq b\}$. It is clear that in this case, $P_0(T(b) < \infty) \leq e^{-b}$. Hence the interesting case to consider is when $b \rightarrow \infty$. The following theorem shows that letting $b \rightarrow \infty$ results in energy to be infinity. We have set $S_n = z_1 + z_2 + \dots + z_n$ where $z_i = \log \frac{f_1(y_i)}{f_0(y_i)}$.

Theorem 3.2. $E[T(b)] < \infty$ for all $b < \infty$ and

$$\lim_{b \uparrow \infty} \frac{E[T(b)]}{b} = \frac{1}{D(f_1||f_0)}.$$

Proof. Suppose initially that z_i are limited, i.e., there is a constant c for which $\Pr(z_i > c) = 0$. Wald identity [28] implies:

$$D(f_1||f_0)E[\min\{T, n\}] = E[S_{\min\{T, n\}}] \leq b + c$$

letting $n \rightarrow \infty$, we can obviously get

$$D(f_1||f_0)E[T] \leq b + c < \infty$$

hence we can easily get

$$\lim_{b \uparrow \infty} \frac{E[T(b)]}{b} \leq \frac{1}{D(f_1||f_0)} \quad (3.1)$$

The reverse inequality is soon acquired again by Wald identity as follows:

$$\frac{E[T]}{b} = \frac{E[S_T]}{b} \cdot \frac{1}{D(f_1||f_0)} \geq \frac{1}{D(f_1||f_0)}$$

therefore the theorem is proved for the case that all log-likelihood ratio is limited. In general case, let $c > 0$ and define $z_n^c = z_n 1_{\{z_n \leq c\}} + c 1_{\{z_n > c\}}$, where c is sufficiently large. Let $\mu^c = E[z_1^c]$ and also define $T^c(b)$ by analogy with $T(b)$. Then we can get the following expression from what we got in (3.1) knowing that $T^c \geq T$:

$$\lim_{b \uparrow \infty} \frac{E[T(b)]}{b} \leq \frac{1}{\mu^c}$$

and since $\mu^c \rightarrow D(f_1||f_0)$ as $c \rightarrow \infty$; then this expression is generalized of (3.1). Reverse inequality is again immediate by Wald identity.

□

We thought that sequential hypothesis testing can help to achieve a finite energy, but the preceding theorem shows that getting the error probability to decay exponentially to zero causes energy to blow up.

In next chapters we decided to examine the problem of achieving capacity per unit cost (or equivalently achieving the minimum energy per bit) individually in some particular channels. The minimum energy per bit analysis and how to achieve it in Gaussian and Poisson channel will follow in next chapters.

Chapter 4

Gaussian Channels

A nice adaptation of conditions will make almost any hypothesis agree with the phenomena. This will please the imagination but does not advance our knowledge.

–J. Black

This chapter concerns the minimum energy for communication over Gaussian channel. Although the Gaussian channel is largely used in modeling the channels in wireless communication, no easily understandable and implementable feedback coding scheme has ever been proposed for it. We took advantage of Minimum Mean Square Error (MMSE) estimation in our newly proposed coding scheme for sending reliably one single bit which is proved to achieve the minimum energy per bit, $(E_b)_{min}$. We know the availability of the feedback link (even if noiseless) can not increase the channel capacity of the forward link (and so capacity per unit cost), so $(E_b)_{min}$ is exactly the same as what Shannon and Golay suggested, [1] and [9]. Hence, we proposed a coding scheme whose incurred energy for transmitting reliably one single bit in the presence of feedback is $N_0 \ln 2$.

Based on figure 1.1, it is clear that energy per bit escalates rapidly with transmission rate, that is, sending one bit in each symbol (or channel use) requires infinite energy. In other words, the cost per bit is close to its infimum, $N_0 \ln 2$ [9], if we use more symbols per bit. This is why it is always preferred to spread the transmission of one bit over infinite degree of freedom (i.e., infinite number of channel uses). Using a large number of degree of freedom per transmitted bit allows use to spend as little energy per bit as possible.

In first section, Elias scheme [11] is introduced. Although Elias scheme is not energy efficient, it gives a good understanding and intuition of feedback

scheme also it is needed to explain the Gallager's scheme in subsequent section. Gallager *et al.*, [13] modified SK scheme in high SNR and proved that energy in this case can be bounded and always smaller than 5. Third section concerns the Turin's scheme [14], which achieves the capacity per unit cost. Turin originally proved the optimality of his scheme in continuous time which is pointed out briefly in section 4.3. Fourth section gives our new coding scheme which can be thought as a special case of Turin's scheme. Fifth section talks about seemingly new feedback coding scheme given by Polyanskiy *et al.*, [7] which achieves the capacity per unit cost, as well. In fact this the scheme is exactly the same as one proposed by Turin in [15], but he unfortunately failed to prove the optimality.

In what follows and also next chapter, we never consider the bandwidth constraints, that is, we can use the forward channel as many times as we want.

4.1 Elias Scheme

Let $X^n = X_1, X_2, \dots, X_n$ be $n > 1$ successive inputs to a discrete-time additive memoryless Gaussian channel with ideal feedback. Then the channel outputs are $Y^n = Y_1 Y_2, \dots, Y_n$ which satisfy $Y^n = X^n + Z^n$ where Z^n is an n -tuple of statistically independent Gaussian random variables, each with zero mean and variance σ_Z^2 , i.e., $Z_i \sim \mathcal{N}(0, \sigma_Z^2)$. without loss of generality σ_Z^2 is set to be one. The channel input constraint (cost constraint) is modeled as the energy of each codeword, in other words, $b[x] = x^2$. Hence, $\frac{1}{n} \sum_{i=1}^n E[X_i^2] \leq S$ where S is a given total power constraint. Let U_1 be the random source symbol to be transmitted to receiver using codeword X^n . An ideal feedback is present, that is, previous outputs, Y^{i-1} are made known to the transmitter in time to generate X_i and therefore each input is a function of source symbol, U_1 and also previous outputs, $X_i = f(U_1, Y^{i-1})$. Elias further assumed that $U_1 \sim \mathcal{N}(0, \sigma_1^2)$. Elias constructed simple scheme to map U_1 to n -dimensional channel input in which the variance of error decreases so fast.

Elias strategy is started by sending the linear scaling of source symbol U_1 as first channel input, i.e., $X_1 = \frac{\sqrt{S_1}}{\sigma_1} U_1$ where $S_1 = E[X_1^2]$. The source symbol needs to be scaled to meet the second moment constraint. Receiver performs the Minimum Mean Square Error (MMSE) estimate of X_1 and

outputs $E[X_1|Y_1] = \frac{S_1 Y_1}{1+S_1}$ and consequently $E[U_1|Y_1] = \frac{\sigma_1 \sqrt{S_1} Y_1}{1+S_1}$ as the first estimate of source symbol. Transmitter is informed about what has been received through ideal feedback. In second step $U_2 = U_1 - E[U_1|Y_1]$ replaces the source symbol and then second channel input is $X_2 = \frac{\sqrt{S_2}}{\sigma_2} U_2$ in which σ_2^2 is the variance of error in first estimate, $\sigma_2^2 = \frac{\sigma_1^2}{1+S_1}$ and again $S_1 = E[X_1^2]$. Elias proved that the best choice for S_i subject to power constraint is just $S_i = S$ for $1 \leq i \leq n$. In the second step, receiver again calculates the MMSE estimate $E[U_2|Y_2] = \frac{\sigma_2 \sqrt{S_2} Y_2}{1+S_2}$ and then transmitter computes the error in the estimation as $U_3 = U_2 - E[U_2|Y_2]$. Putting two estimates in one expression, we get

$$U_3 = U_1 - E[U_1|Y_1] - E[U_2|Y_2]$$

Thus U_3 is the error of estimating source symbol based on $E[U_1|Y_1] + E[U_2|Y_2]$. This procedure keeps going on and receiver gets always the better estimate of U_1 and also transmitter keeps sending the linearly scaled version of estimation error, generally the formula for this procedure is as follows:

$$\begin{aligned} X_i &= \frac{\sqrt{S_i}}{\sigma_i} U_i; \\ E[U_i|Y_i] &= \frac{\sigma_i \sqrt{S_i} Y_i}{1 + S_i}; \\ U_{i+1} &= U_i - E[U_i|Y_i]; \\ \sigma_{i+1}^2 &= \frac{\sigma_i^2}{1 + S_i}. \end{aligned}$$

Running this iterative procedure for variance of estimation error yields:

$$\sigma_{n+1}^2 = \frac{\sigma_1^2}{\prod_{i=1}^n (1 + S_i)}$$

Elias proved that this total error estimation subject to the constraint $\sum_{i=1}^n S_i = nS$ gets minimized by choosing all S_i equal to each other and hence $S_i = S$. It is worthwhile to point out that although Elias scheme is restricted to the source symbols being Gaussian distributed, it can be efficiently used for the general coding scheme in AWGN channel where the assumption of Gaussian distributed for source symbol is dropped. In this case, the first channel input, X_0 is again linear scaling version of source symbol U_0 (which is not Gaussian distributed). Receiver observes $Y_0 =$

$X_0 + Z_0$ and then feeds it back to the transmitter. Knowing X_0 , transmitter determines Z_0 . From now on the situation resembles the Elias scheme if we consider Z_0 as U_1 in above, then $U_1 \sim \mathcal{N}(0, \sigma_Z^2)$ is the goal to be estimated using the Elias scheme. Let $\hat{U}_1$ be the estimated version of U_1 using Elias scheme. Then source symbol U_0 is obtained by $X_0 = Y_0 - \hat{U}_1$. This scheme yields an almost trivial derivation of the SK scheme [9]. This scheme is briefly explained in the following section.

4.2 Gallager Feedback scheme for high SNR regime

As said in previous section Schalkwijk-Kailath (SK) scheme can be derived using Elias scheme. Suppose the source is a set of M equiprobable symbols denoted by $\{1, 2, \dots, M\}$. The source signal, U_0 is a standard M -PAM modulation of the source symbol. That is, for each symbol m , $1 \leq m \leq M$, m is mapped into a the signal a_m being defined as $a_m = m - (M + 1)/2$ which renders M signals in U_0 symmetric around 0 with unit distance. Then the initial channel input is $X_0 = \frac{\sqrt{S_0}}{\sigma_0^2} U_0$ in which $\sigma_0^2 = (M^2 - 1)/12$ and S_0 is the allocated energy. Thus X_0 is M -PAM encoding with signal separation $d_0 = \frac{\sqrt{S_0}}{\sigma_0}$. The received signal, $Y_0 = X_0 + Z_0$ is known to transmitter, as well, through the noiseless feedback. Determining Z_0 , transmitter can from now on use Elias scheme because Z_0 is assumed to be distributed as $\mathcal{N}(0, 1)$. Then the next unconstrained signal, U_1 is Z_0 and the next transmitted signal is $X_1 = \sqrt{S_1} U_1$ because, $\sigma_1^2 = E[U_1] = 1$. S_1 is again the allocated energy to the second transmission step. Elias scheme is performed for next $n - 1$ steps and then the estimate of Z_0 is combined with Y_0 to get the value of X_0 . To optimize the Elias scheme all $S_i = S_1$ for all $1 \leq i \leq n - 1$ and then the power constraint will reduce to $S_0 + (n - 1)S_1 = nS$. At time $(n - 1)$ the estimated version of Z_0 is

$$E[Z_0|Y^{n-1}] = \sum_{i=1}^{n-1} E[U_i|Y_i]$$

and U_n which is regarded as the error in the estimating Z_0 up to time $(n - 1)$ is equal to $Z_0 - E[Z_0|Y^{n-1}]$ is distributed as $\mathcal{N}(0, \sigma_n^2)$ in which, due to (4.1)

$$\sigma_n^2 = \frac{1}{(1 + S_1)^{n-1}}$$

Hence all the receiver should do is just to have to detect X_0 from $Y_0 -$

$E[Z_0|Y^{n-1}] = X_0 + U_n$. Detecting an M -PAM signal from a the signal plus the independently normally distributed noise goes wrong if and only if the value of noise is bigger than half of the distance of M -PAM signal points, i.e. $|U_n| \geq d_0 = 0.5\sqrt{\frac{12S_0}{M^2-1}}$. Taking the variance of U_n into account, the error probability can be found as¹:

$$P_e \leq 2Q(\gamma_n)$$

where $\gamma_n = 0.5\sqrt{\frac{12S_0(1+S_1)^{n-1}}{M^2-1}}$.

To minimize the error probability, γ_n has to be maximized. To this end, S_0 and S_1 shall be found by solving the following maximization problem

$$\arg \max S_0(1+S_1)^{n-1}$$

$$\text{subject to } S_0 + (n-1)S_1 = nS$$

Taking derivative leads to solutions $S_1 = S - \frac{1}{n}$ and $S_0 = S + 1 - \frac{1}{n}$, hence we have to assume that $nS > 1$ ². The minimum error probability can therefore be

$$P_e \leq 2Q(\psi_n)$$

where $\psi_n = \sqrt{\frac{3(1+S-\frac{1}{n})^n}{M^2-1}}$.

If we want to transmit one bit of information, i.e., $M = 2$, the minimum amount of energy that this scheme incurs, is always bigger than one and error probability is always smaller than $P_e \leq 0.3172$. That is, this scheme is not optimum from energy point of view, or it can not achieve the capacity per unit cost.

As said before, this scheme is a trivial modification of SK scheme but Gallager *et al* [13], explored new strategy which is again based on the use of M -PAM at time zero but is quite different from the SK scheme at time 1 to $n-1$. This new strategy is restricted to the situation in which SNR is high so that the distance between adjacent PAM signal points in X_0 is large

¹In the original paper [13], the authors found the exact value for error probability. To do so, they consider the fact that the smallest and the largest signals have only one neighbor, while the others have two. This causes the error probability to be slightly smaller than $2Q(\gamma_n)$.

²Otherwise S_1 is assumed to be zero which means all energy is used in first step and we do not benefit from feedback.

compared to the noise standard deviation. Then the new strategy leads to the better performance than Elias scheme.

Given that a_m is the sample value of the PAM source signal U_0 , the sample value of X_0 is $a_m d_0$ in which d_0 as usual equals $d_0 = \frac{\sqrt{S_0}}{\sigma_0^2}$. Obviously, the probability distribution of received signal, Y_0 , is $\mathcal{N}(a_m d_0, 1)$, or in other words, it is the density of Z_0 shifted to center $a_m d_0$. Maximum Likelihood detection here acts as a quantizer which takes Y_0 to the nearest signal point. Let $\hat{m}_0$ be the estimate of m at time zero. Error happens if and only if the noise exceeds half of distance of a_m to its neighbors which is d_0 . Thus error probability is $2Q(d_0/2)$. Feedback informs the transmitter about $\hat{m}_0$ by which it can build next signal U_1 without second moment constraint. Indeed, U_1 is the shifted version of original M -PAM signal, $U_1 = \hat{m}_0 - m$ where m is the original source message to be transmitted. In fact, noise at time zero can make the receiver choose a_{m+1} or a_{m-1} depending on the value of noise and due to the assumption of high SNR, the probability of other choices for $\hat{m}_0$ (other than $\hat{m}_0 = m \pm 1$) is so small that it can be ignored. Hence U_1 is the integer-valued error in the first estimate, $a_{\hat{m}_0}$, of U_0 , so U_1 is either ± 1 with probability $2Q(d_0/2)$ or 0 with probability $1 - 2Q(d_0/2)$. Next transmitted signal, X_1 is linear scaling of U_1 as $X_1 = \frac{\sqrt{S_1}}{\sqrt{E[U_1^2]}} U_1$ in which S_1 is the allocated energy for second transmission and $E[U_1^2] = 2Q(d_0/2)$, again linear scaling of U_1 is chosen so that the second moment constraint or power constraint is fulfilled. It can be easily seen that X_1 is not only the shifted version of X_0 but is the scaled up version as well. It means that if S_0 is sufficiently large, X_1 is exponentially increasing with S_0 .

After sending X_1 , the situation is exactly the same as when sending X_0 with this difference that here the distances between adjacent signal points increase exponentially in S_0 . Hence the error probability for another ML estimation will be decreasing as double exponential in S_0 . Running the same procedure for third signal, we get error probability being decreased as third order exponential in S_0 . Gallager proved that ML estimation in each step is simply just a d_i -quantization where d_i is distance between adjacent signal points at time i and then he showed that

$$E[U_i^2] \leq \frac{1.6}{d_i} \exp\left(-\frac{d_i^2}{8}\right).$$

Based on this inequality, Gallager suggested to take $d_i = \sqrt{8} \exp(\frac{d_{i-1}^2}{16})$.

In fact, this choice of d_i causes the expected energy in step i to be as small as possible, because $d_i^2/8 = \exp(d_{i-1}^2/8)$, hence the expected energy in step i is smaller or equal than $12.8/d_{i-1}$, i.e., $E[X_i^2] \leq \frac{12.8}{d_{i-1}}$. The interesting results that are proved in [11] are as follows; for $d_0 \geq 4$

$$\sum_{i=1}^{\infty} E[X_i^2] \leq 5$$

$$\Pr(\hat{m}_i \neq m) \leq \frac{1}{g_{i+1}(2)}$$

where $g_i(x) = \exp(\exp(\dots(\exp(x))\dots))$ with i exponentials. It means that with finite energy, this scheme yields a very small error probability which decays with n order exponentials at time n .

In fact, Gallager made use of the idea of Zigangirov (i.e., using PAM centered on the receiver's current tentative decision) to show the rapid expansion of signal points with small energy. The main difference between this scheme and one proposed by Zigangirov in [29] is the fact that energy in Zigangirov's scheme increases linearly with time.

Gallager again proposed a two-phase feedback coding scheme whose first phase is done using Elias scheme and then at time $n_1 - 1$ the equivalent received signal $Y_0 - E[Z_0|Y_1^{n_1-1}]$ is used in an ML decoder to detect the original PAM signal X_0 in the presence of additive Gaussian noise of variance $\sigma_{n_1}^2$. He assured that if n_1 is chosen large enough to fulfill $d_{n_1-1} \geq 4$, then the situation at time $n_1 - 1$ can be assumed at time 0 of the high SNR scheme.

4.3 Turin's Scheme

After discussing some well-known feedback coding strategies for Gaussian channels (in infinite degree of freedom regime); we come to the main challenge of this thesis which is about coding strategy for sending one bit of information achieving the channel capacity per unit cost. Turin [14] proposed an idea and analyzed the continuous time feedback to show the optimality of his idea. He worked with two signals b_+ and b_- to convey one bit of information (so two waveforms have to be assumed uniformly distributed) through a Gaussian channel whose variance is $N_0/2$. The output of channel in the beginning is therefore either $Y = Sb_- + n$ (hypothesis H_-) or $Y = Sb_+ + n$ (hypothesis H_+) in which S is a parameter identified later.

Receiver looks at the signal Y on the interval $[t_0, t]$ where $t \in [t_0, \infty)$, which is called Y_t , hereafter. Let $S(t)$ denote the log posteriori ratio of received signal up to time $t \geq t_0$, i.e.,

$$S(t) = \ln \frac{\Pr(H_+|Y_t)}{\Pr(H_-|Y_t)}$$

Like other sequential analyses, two parameters S_- and S_+ are considered to bound the ambiguity area for $S(t)$, in other words, decision is made in favor of H_+ (resp. H_-) as soon as $S(t)$ crosses S_+ (resp. S_-). The ambiguity area spans $[t_0, T]$ where T is a stopping time. Due to the continuity of posteriori probability ratio, the procedure stops when $S(t) = S_{\pm}$ (instead of having to write $S(t) < S_-$ and $S(t) > S_+$). Having the previously received signals at hand, transmitter makes next signal as $b_-(S(t), t)$ or $b_+(S(t), t)$. If $p(e|H_-)$ and $p(e|H_+)$ represent the probability of error when H_- and H_+ are true, respectively, then $p(e|H_-) = e^{-S_+}(1 - p(e|H_+))$. This becomes clear as follows: whenever it is decided that H_+ is true, that is, the test stops with $S(t) = S_+$ then from definition of log-likelihood ratio we have $p(H_-|Y_T) = e^{-S_+}p(H_+|Y_T)$ and then averaging it over all observations Y_T which leads to decision H_+ , yields $p(e|H_-) = e^{-S_+}(1 - p(e|H_+))$.

Hence given the required error probability (which is the average of $p(e|H_-)$ and $p(e|H_+)$), constants S_+ and S_- can be calculated using the following expression:

$$S_{\pm} = \ln \frac{1 - p(e|H_{\pm})}{p(e|H_{\mp})}$$

So far every things are pretty straightforward and quite general for all encoding sequences, that is, for any $b_{\pm}(S(t), t)$. Turin proposed to work with $b_{\pm}(S(t), t) = \pm U_{\pm}(S(t))\sigma(t)$, and he called his scheme as "*uncertainty adjustment*" because the amplitude of waveform $\sigma(t)$ is adjusted by receiver's uncertainty. He further proved the optimal value for these quantities, he showed that if there is no peak power constraint so that peak power can be so big (and average power is so small) then the best choices are

$$\begin{aligned} U_+(S(t)) &= \frac{1}{1 + e^{S(t)}} \\ U_-(S(t)) &= \frac{1}{1 + e^{-S(t)}} \\ \sigma(t) &= K, \quad t \geq t_0 \end{aligned}$$

then choosing a reasonable value for K leads this scheme to achieve capacity per unit cost³, $C = \frac{P_{av}}{N_0 \ln 2}$ in which P_{av} is just the average power transmitted over the infinite sequence. Turin gave a more explicit scheme, for illustrative purposes, as $b_{\pm}(S(t), t) = \pm 0.5\sigma(t)[1 \mp \text{sgn}(S(t))]$. Hence if H_+ is true, whenever $S(t) > 0$ the transmitter is turned off and it is turned on when $S(t) < 0$ and conversely for sending H_- . This is how to save the energy in this coding scheme. It is interesting to know that this scheme was dealt with in [15] and claimed to achieve $C \ln 2$ at best, that is, it can not achieve the capacity per unit cost. But recently Polyanskiy *et al.* [7] showed that this scheme (on-off scheme) can indeed achieve the capacity per unit cost. So this is the simplest optimal scheme for AWGN channel. This scheme will be explained more in section 4.5. It is worthy to point out that Turin made error probability so small by choosing very large $S = S_- = -S_+$, in this case error probability for both hypothesis are equal and totally $p_e \leq e^{-S}$.

In next section we propose the new coding scheme which at final turns out to be special-case of Turin scheme with much easier and more interesting analysis.

4.4 MMSE-based Coding Scheme

In this section we propose a new scheme which is easy to implement and also to understand.

Here two signals are $b_- = -\frac{1}{2}$ and $b_+ = +\frac{1}{2}$ and $\sigma(t)$ is assumed to have fixed value of S (which can be regarded as the allocated energy in each step). In the new perspective, $U(S(t))$ is the error of MMSE estimation of transmitted signal based on the received signals (like Elias scheme). Lets analyze this scheme in discrete time and then work with the required energy.

Suppose U_0 is the source message to be transmitted which is equally likely either $b_- = -\frac{1}{2}$ (hypothesis H_0) and $b_+ = +\frac{1}{2}$ (hypothesis H_1). In the first step, $X_1 = U_0 S$ is sent. After the receiver observes the corrupted version of transmitted signal, Y_1 , it calculates the MMSE estimate of U_0

³Note that C in this formula is conventional capacity of continuous-time Gaussian channel. We know that the capacity of continuous-time Gaussian channel in infinite degree of freedom regime is equivalent to the capacity per unit cost of an equivalent discrete time channel [8].

and then feeds it back to transmitter through noiseless feedback link. Let U_1 denote the error of MMSE estimate of U_0 based on Y_1 ,

$$U_1 = U_0 - E[U_0|Y_1]$$

then the second transmitted signal is $X_2 = SU_1$.

The reason behind choosing MMSE can be explained as follows. Suppose second transmission signal denoted as $U_0 - u_1^*$ and u_1^* is chosen in a way so that the total energy is minimized, so putting in another way

$$u_1^* = \arg \min_{u_1} E[(U_0 - u_1)^2]$$

which leads to the MMSE estimation, $u_1^* = E[U_0|Y_1]$. Hence if the receiver is in the right direction in estimating U_0 , next step has smaller amplitude and consequently smaller energy and of course if the estimation direction is wrong, transmitter will send more energy to get the receiver into the right direction. This procedure is done for third step to get

$$\begin{aligned} U_2 &= U_0 - E[U_0|Y_1Y_2], \\ X_3 &= SU_2 \end{aligned}$$

This is because $u_2^* = E[U_0|Y_1Y_2]$.

Running this procedure for n steps, we can get following

$$\begin{aligned} u_n^* &= E[U_0|Y^n] \\ U_n &= U_0 - u_n^* \\ X_{n+1} &= SU_n \end{aligned}$$

Note that all MMSE estimations $u_1^*, u_2^*, \dots, u_n^*$ are known for both receiver and transmitter, hence receiver can add u_{i-1}^* to Y_i to get U_0 corrupted by noise. The receiver has therefore observed n repetitions of source message. This makes the situation for receiver similar to the case as if the simple repetition code was used, in other words the error probability is simply $p_e = Q\left[\sqrt{\frac{S^2n}{2N_0}}\right]$.

Now let's find the expression for MMSE estimation, as follows:

$$\begin{aligned}
u_n^* &= E[U_0|Y^n] \\
&= \frac{1}{2} \Pr(U_0 = +\frac{1}{2}|Y^n) - \frac{1}{2} \Pr(U_0 = -\frac{1}{2}|Y^n) \\
&= -\frac{1}{2} + \Pr(b_+|Y^n)
\end{aligned} \tag{4.1}$$

Hence when H_1 is the true hypothesis, $\Pr(b_+|Y^n)$ tends to one and $u_n^* \rightarrow 1/2 = b_+$ and when H_0 is true, $\Pr(b_+|Y^n)$ tends to zero and $u_n^* \rightarrow -1/2 = b_-$.

Since the hypotheses H_0 and H_1 are assumed to be equally likely, the situation is symmetric with respect to b_- and b_+ , so to compute the expected transmitted energy after n th transmission step, E_n , one can assume that H_1 is true hypothesis. Hence E_n can be written as

$$E_n = \frac{S^2}{4} + S^2 \sum_{i=2}^n E[(U_0 - u_{i-1}^*)^2]$$

Substituting $U_0 = \frac{1}{2}$ and also u_n^* from (4.1) into expression for E_n , one can get

$$\begin{aligned}
E_n &= \frac{S^2}{4} + S^2 \sum_{i=1}^{n-1} E\left[\left[1 - \Pr(b_+|Y^i)\right]^2\right] \\
E_n &= \frac{S^2}{4} + S^2 \sum_{i=1}^{n-1} E\left[\left[\Pr(b_-|Y^i)\right]^2\right]
\end{aligned} \tag{4.2}$$

In order to compute the expectation in (4.2), log-likelihood ratio can be used. Let

$$\lambda_j = \ln \frac{\Pr(b_+|Y^j)}{\Pr(b_-|Y^j)} = \ln \frac{\Pr(Y^j|b_+)}{\Pr(Y^j|b_-)}$$

be the likelihood ratio at time j , then

$$\Pr(b_-|Y^j) = \frac{1}{1 + e^{\lambda_j}} \tag{4.3}$$

Having assumed that H_1 is true, one can write the received signal at step j as $Y_j = S(b_+ - u_{j-1}^*) + n_j$, $j = 2, \dots, n$, and log-likelihood ratio as

$$\begin{aligned}
\lambda_j &= \ln \frac{\exp\left(-\frac{1}{N_0} \sum_{i=1}^j [Y_i - S(b_+ - u_i^*)]^2\right)}{\exp\left(-\frac{1}{N_0} \sum_{i=1}^j [Y_i - S(b_- - u_i^*)]^2\right)} \\
\lambda_j &= \frac{S}{N_0} \sum_{i=1}^j (S + 2n_i)
\end{aligned} \tag{4.4}$$

which shows that λ_j is distributed as $\mathcal{N}(\frac{S^2}{N_0}j, 2\frac{S^2}{N_0}j)$.

Using (4.4) and (4.3), the required expectation for computing energy is written as

$$\begin{aligned}
E[\Pr(b_-|Y^j)^2] &= E\left[\left(\frac{1}{1+e^{\lambda_j}}\right)^2\right] \\
&= E\left[\left(\frac{e^{-\xi_j}}{e^{-\xi_j}+e^{\xi_j}}\right)^2\right]
\end{aligned}$$

In which $\xi_j = \frac{1}{2}\lambda_j$ and therefore is Gaussian distributed $\mathcal{N}(\frac{S^2}{2N_0}j, \frac{S^2}{2N_0}j) = \mathcal{N}(ja, ja)$ where $a = \frac{S^2}{2N_0}$. Then the expression for energy is easily expanded as

$$\begin{aligned}
E_n &= \frac{S^2}{4} + S^2 \sum_{i=1}^{n-1} E[(\Pr(b_-|Y^i))^2] \\
&= \frac{S^2}{4} + S^2 \sum_{i=1}^{n-1} \frac{1}{\sqrt{2\pi ia}} \int_{-\infty}^{+\infty} \frac{\exp\left[-\frac{(\xi_i+ia)^2}{2ia}\right]}{(e^{-\xi_i}+e^{\xi_i})^2} d\xi
\end{aligned}$$

Denoting the integral involved in above expression by I , we can write:

$$\begin{aligned}
I &= \frac{1}{2} \int_{-\infty}^{+\infty} \frac{\exp\left[-\frac{(\xi_i-ia)^2}{2ia}\right] + \exp\left[-\frac{(\xi_i+ia)^2}{2ia}\right]}{(e^{-\xi_i}+e^{\xi_i})^2} d\xi_i \\
&= \frac{e^{-ia/2}}{2} \int_{-\infty}^{+\infty} \frac{\exp\left[-\frac{(\xi_i)^2}{2ia}\right]}{e^{-\xi_i}+e^{\xi_i}} d\xi_i \\
&= e^{-ia/2} \int_0^{+\infty} \frac{\exp\left[-\frac{(\xi_i)^2}{2ia}\right]}{e^{-\xi_i}+e^{\xi_i}} d\xi_i
\end{aligned}$$

and the expression for energy can be simplified when substituting I as

$$\begin{aligned}
E_n &= \frac{S^2}{4} + S^2 \sum_{i=1}^{n-1} \frac{1}{\sqrt{2\pi ia}} \int_{-\infty}^{+\infty} \frac{\exp\left[-\frac{(\xi_i - ia)^2}{2ia}\right]}{(e^{-\xi_i} + e^{\xi_i})^2} d\xi_i \\
&= \frac{S^2}{4} + S^2 \sum_{i=1}^{n-1} \frac{e^{-ia/2}}{\sqrt{2\pi ia}} \int_0^{+\infty} \frac{\exp\left[-\frac{(\xi_i)^2}{2ia}\right]}{e^{-\xi_i} + e^{\xi_i}} d\xi_i \quad (4.5)
\end{aligned}$$

Deriving error probability, $p_e = Q[\sqrt{na}]$, to zero would require na to tend to infinity, but a should be obviously kept as small as possible. Hence n or number of transmission should go to infinity. Keeping a so small, one can ignore the first term in E_n . If n tends to infinity the sum in above expression is approximated by integral, the total energy can then be written as

$$\begin{aligned}
E_\infty &= 2N_0 \int_0^\infty \frac{e^{-t/2}}{\sqrt{2\pi t}} \int_0^{+\infty} \frac{\exp\left[-\frac{\xi^2}{2t}\right]}{e^{-\xi} + e^{\xi}} d\xi dt \\
&= \frac{2N_0}{\sqrt{2\pi}} \int_\xi \frac{1}{e^{-\xi} + e^{\xi}} \int_t \frac{\exp(-\frac{\xi^2}{2t})}{\sqrt{t}} e^{-t/2} dt d\xi
\end{aligned}$$

Second integral represents the Laplace transform of function $f(t) = \frac{\exp(-\frac{\xi^2}{2t})}{\sqrt{t}}$ when the Laplace parameter is $s = \frac{1}{2}$. The Laplace transform of this function can be computed⁴ as follows

$$\mathcal{L}[f(t)] = \sqrt{\frac{\pi}{s}} e^{(-2\sqrt{\xi^2 s/2})}$$

and consequently $\int_t \frac{\exp(-\frac{\xi^2}{2t})}{\sqrt{t}} e^{-t/2} dt = \sqrt{2\pi} e^{-\xi}$. Hence we can compute the total energy as

$$E_\infty = 2N_0 \int_0^\infty \frac{e^{-\xi}}{e^{-\xi} + e^{\xi}} d\xi = N_0 \ln 2$$

which equals $1/C$ where C represents capacity per unit cost in AWGN channel, i.e., $C = \frac{1}{N_0 \ln 2}$.

Although this analysis is totally different from that of Turin which was based on posteriori probabilities ratio, the scheme is pretty similar to Turin's, in other words, this scheme is a special case of Turin's scheme when the encoding function is just the error of receiver's MMSE estimation of the source message up to previous step.

⁴For example, look at: <http://eqworld.ipmnet.ru/en/auxiliary/inttrans/laplace3.pdf>

4.5 On-Off Feedback Scheme

In previous section, a feedback coding scheme was proposed which could achieve the capacity per unit cost for AWGN channel. That scheme could be explained with the aid of the MMSE estimation and its error. As said before, feedback is used not to increase the capacity (and capacity per unit cost) but to decrease the complexity of encoder and decoder. One of the other advantages of feedback is the possibility of zero error communication when $N \rightarrow \infty$. In Turin's scheme in each step a number in $\mathbb{R}$ is sent over forward link, now this question arises that whether there exists an optimal feedback coding scheme which requires only a *binary* input to be sent in each step. Although this idea is by no means new and has previously treated (e.g. look [15]), it has recently been analyzed and proved to achieve the capacity per unit cost [7].

Lets get back to the Turin scheme and the explicit model he proposed. Turin proposed the on-off model for feedback communication which is built by having the encoders act as comparator with zero level, in other words, log-posteriori probability ratio is compared with zero, if positive, H_1 is more probable given all received signals and if negative, H_0 is more probable, hence in next step zero or $b_{\pm}$ is sent depending on whether or not the receiver is in the right direction for decoding. This leads to the following encoder model:

$$f_n(U_0, Y^{n-1}) \triangleq 2U_0d[1 \mp \text{sgn}(S_{n-1})]$$

where like before $U_0 = \pm 1/2$, d determines the energy allocated for each step and

$$\begin{aligned} S_n &= \log \frac{\Pr(U_0 = b_+ | Y^n)}{\Pr(U_0 = b_- | Y^n)} \\ &= \log \frac{\Pr(Y^n | U_0 = b_+)}{\Pr(Y^n | U_0 = b_-)} = \sum_{j=1}^n \log \frac{\Pr(Y_j | Y^{j-1}, U_0 = b_+)}{\Pr(Y_j | Y^{j-1}, U_0 = b_-)} \\ &= \sum_{j=1}^n \log \frac{\Pr(Y_j | f_j(b_+, Y^{j-1}))}{\Pr(Y_j | f_j(b_-, Y^{j-1}))} \end{aligned}$$

is the log-posteriori probability ratio at time n . Letting $W \triangleq 2U_0$, one can write:

$$f_n(U_0, Y^{n-1}) = \begin{cases} Wd & \text{sgn}(S_{n-1}) = -W \\ 0 & \text{otherwise} \end{cases}$$

This situation has symmetry, too, with respect to W , so we can from now on assume that $W = +1$. Hence assuming $W = +1$ and ignoring the alternative in encoding at time n , one can get

$$\begin{aligned} S_n &= S_{n-1} + \log \frac{\Pr(Y_n | f_n(W = +1, Y^{n-1}))}{\Pr(Y_n | f_n(W = -1, Y^{n-1}))} \\ &= S_{n-1} + \frac{1}{2N_0} \left(\frac{1}{2} d^2 + d\nu_n \right) \end{aligned} \quad (4.6)$$

where ν_n in noise component at time n .

It is obvious that Y_n is submartingale drifting to infinity. This means that⁵ $E[S_n | \mathcal{F}_{n-1}]$ is bigger (or equal) than Y^{n-1} . We know that thanks to model, transmitter sends $+d$ if and only if $S_n < 0$, and zero otherwise. Combining these two facts, it can be concluded that the number of non-zero transmitting signals are finite, because otherwise the drift of submartingale would not be positive.

If we have assumed $W = -1$, then we would have had a supermartingale drifting to $-\infty$ and again knowing the fact that the transmitter sends non-zero signal providing that $S_n > 0$ will yield that the number of non-zero transmitted signals is finite. A supermartingale drifting to $-\infty$ when H_0 is true and a submartingale drifting to ∞ when H_1 is true imply that $P_{Y^\infty|W=+1}$ and $P_{Y^\infty|W=-1}$ are mutually singular, in other words, W can be retrieved from Y^∞ with zero error probability.

To compute the energy required in this scheme, we have to compute

$$\begin{aligned} E_\infty &= E[\|\mathbf{X}\|^2] = \sum_{j=1}^{\infty} E[\|X_j^2\|] = \sum_{j=1}^{\infty} d^2 E[1_{\{S_j \leq 0\}}] \\ &= \sum_{j=1}^{\infty} d^2 \Pr(S_j \leq 0) \end{aligned}$$

where $\mathbf{X}$ is the codeword used to convey the message.

⁵ $\mathcal{F}_n$ is the minimum σ -algebra generated by S^{n-1} . Hence conditioning on $\mathcal{F}_n, S^{n-1}$ is no longer random variable.

To be able to find the $E[1_{\{S_j \leq 0\}}]$, we replace $d\nu_n$ in (4.6) with $\sqrt{\frac{N_0}{2}}(W_{nd^2} - W_{(n-1)d^2})$ where W_s is the standard Brownian motion. This is true because we know $\nu_n \sim \mathcal{N}(0, N_0/2)$ and $W_s - W_t \sim \mathcal{N}(0, s - t)$. After substituting, S_n can be written as

$$S_n = \frac{1}{2N_0} \left(\frac{s}{2} + \sqrt{\frac{N_0}{2}} W_s \right)_{s=nd^2}$$

Hence the total number of j for which $S_j < 0$, is equivalent to the number of negative samples we made out of Brownian motion $\frac{s}{2} + \sqrt{\frac{N_0}{2}} W_s$ when the grid size is nd^2 . The interesting case is when $d \rightarrow 0$, in this case the continuity of W_s (for every realization) implies that $\|\mathbf{X}\|^2$ is the total time T that Brownian motion $\frac{s}{2} + \sqrt{\frac{N_0}{2}} W_s$ spends below zero, in other words when $d \rightarrow 0$

$$\lim_{d \rightarrow 0} \sum_{j=1}^{\infty} d^2 1_{\{S_j \leq 0\}} = \int_0^{\infty} 1_{\{\frac{t}{2} + \sqrt{\frac{N_0}{2}} W_t \leq 0\}} dt$$

or equivalently,

$$\begin{aligned} \lim_{d \rightarrow 0} \sum_{j=1}^{\infty} d^2 E[1_{\{S_j \leq 0\}}] &= \int_0^{\infty} \Pr[\frac{t}{2} + \sqrt{\frac{N_0}{2}} W_t \leq 0] dt \\ &= N_0 \end{aligned}$$

So the energy that is spent by this scheme to transmit one bit of information is N_0 which is the optimum, in other words this scheme can achieve the capacity per unit cost, although Turin in [15] tried to use another method to show this scheme is not optimal.

Comparing this scheme with our proposed scheme in previous section shows that this is the simplest feedback encoding one can think of, because here we only send a 0, +1 or -1 but in latter scheme we have to send the real number in each step. However both of these two coding schemes have suffered from the necessity of rate-unlimited feedback link. It is because receiver has to feed back a random variable taking value on an infinite set (in former scheme MMSE estimation $u_n^* \in [-\frac{1}{2}, +\frac{1}{2}]$ has to be fed back and in latter one $Y_n \in \mathbb{R}$ is fed back). To overcome this problem in Gaussian channels no scheme has so far been proposed which achieves capacity per unit cost and also error probability tending to zero. The most straightforward candidate

for solution is *decision feedback scheme* in which receiver only needs to send back one bit, either zero or one which represents "*further sample required*" and "*ready-to-decode*" signals. Transmitter keep sending until it receives ready-to-decode signal. The only result (as far as we know) about energy in Gaussian channel when decision feedback is present, is due to the recent paper, [7] in which it is proved that in order to transmit one bit of information in Gaussian channel in decision feedback scheme error probability is $\epsilon \leq e^{-\frac{E}{N_0}}$, so to get error probability tending to zero we need an infinite energy. However this result is given for a scheme in which codewords are either d or $-d$, so zero is not sent no matter what transmitter observes out of feedback link.

This technique, which has a lot to do with *Stopping Times*, works efficiently in Poisson channels which is explained in details in next chapter.

Chapter 5

Poisson Channel, Infinite Capacity Per Unit Cost

The supreme task of the physicist is to arrive at those universal elementary laws from which the cosmos can be built up by pure deduction. There is no logical path to these laws; only intuition, resting on sympathetic understanding of experience, can reach them.

–A. Einstein

Previous chapter was totally about capacity per unit cost in Gaussian channels and dealt with some coding schemes which could achieve it. We stated that variable length block codes are efficient in Poisson channel, particularly in ideal Poisson channels. Here we have assumed that capacity per unit cost is infinity, hence the minimum energy per bit is zero. In section 5.2 we see when capacity per unit cost is infinity. It is therefore clear that minimum energy per bit is zero. In this chapter we prove that this cannot be achieved. We then calculate the minimum energy for sending one single bit and propose an optimal coding scheme. To our best knowledge this is the first result on energy analysis in Poisson channel when number of bit is finite, and also first proof that minimum energy for sending reliably one bit is totally different from minimum energy per bit when sending large number of bits. We also give an optimal coding scheme for reliably transmitting k bits (finite number of bits).

Poisson channel is a mathematical model that attempts to describe communication via a direct detection optical channel. Direct detection is largely used in many optical communication systems. The channel input represents the squared magnitude of the electric field incident on the photo-detector

and its output is just a counting process describing the arrival time of photons, as registered by detector. In other words, the output has the statistics of Poisson process with instantaneous rate given by the sum of instantaneous input X_t and a non-negative constant term called *dark current* denoted by λ_0 which accounts for background radiation. In ideal Poisson, λ_0 is assumed to be zero.

The capacity of the Poisson channel has been found by Davis [30] who also showed that the availability of causal and ideal feedback can not increase the capacity, like as we know about memoryless channels¹. Hence we again expect the advantages that feedback brings for AWGN channels, also in Poisson channels.

We saw in previous chapter that when ideal feedback is introduced in AWGN channel, the error probability can be made to decay as N -fold exponentials, but of course this drastic improvement can not be obtained in Poisson channels. This is mainly because AWGN channel does not have peak power constraint but Poisson channel does. Since the probability of error is limited by the probability that no counts are registered by the detectors, the error probability is decreasing at best exponentially with coding delay.

In this chapter, we firstly compute the capacity per unit cost for Poisson channels find out when it becomes infinity. Then we calculate and prove the minimum energy for sending one bit and then propose a coding scheme to achieve it. At final we extend the result and obtain an optimal coding scheme for transmitting reliably k bits which unlike conventional information theory does not tend to infinity and is a finite number.

5.1 Introduction

Many of the general Poisson channel's properties are based on the notion of Poisson distribution which can be explained as following:

Let X and Y be a pair of random variables taking values in $\mathbb{R}_0^+ = [0, \infty)$ and the set of non-negative integers $\mathbb{Z}_0^+$, respectively and also let $P_{Y|X}(y|x)$ denote the distribution of output conditioned on $X = x$, then the Poisson distribution implies that:

¹This is not so surprising if we take into account the fact that increments in Poisson process are independent, so the channel can be regarded as memoryless channel.

$$P_{Y|X}(y|x) = \frac{1}{k!} x^k e^{-x}, \quad k = 0, 1, \dots$$

in other words, given $X = x$, random variable Y has above distribution with mean (or rate) equal to x .

In the continuous-time Poisson process the intensity is the process $\{X(t), 0 \leq t \leq T\}$ or equivalently, X_0^T , where $X(t)$ takes value in $[0, \infty)$. The output is a realization of a Poisson point process $\{Y(t)\}$ whose time-varying expectation at any time t is equal to the integral of $X(t)$. Precisely, the output process $Y(t)$ satisfies the following properties:

1. $Y(0) = 0$
2. For all $t_0 = 0 < t_1 < \dots < t_n$, the increments, $Y(t_1) - Y(t_0)$, $Y(t_2) - Y(t_1)$, $\dots$, $Y(t_n) - Y(t_{n-1})$ are independent random variables
3. For $0 \leq t \leq T$, $0 \leq s + t \leq T$ and non-negative integers k , the increments have the Poisson distribution

$$\Pr[Y(t+s) - Y(s) = k] = \frac{\lambda^k e^{-\lambda}}{k!}$$

where

$$\lambda = \int_s^{t+s} X(\tau) d\tau$$

In Poisson channel, input is waveform $X(t)$ and output is just a Poisson counting process $Y(t)$ according to the number of photon registered. For the sake of brevity, let Y_s^t denote the number of photon registered in the duration $[s, t]$. Hence one can then write

$$\Pr[Y_s^t = k] = \frac{\lambda^k e^{-\lambda}}{k!}$$

where

$$\lambda = \int_s^{t+s} X(\tau) d\tau$$

in *ideal* Poisson channel and

$$\lambda = \int_s^{t+s} X(\tau) d\tau + \xi_0 t$$

in *non-ideal* Poisson channel in which ξ_0 is dark current.

There are usually two constraints imposed on Poisson channel which are, *peak power constraint*

$$X(t) \leq A$$

and *average power constraint*

$$\Lambda(t) \triangleq \int_0^t X(\tau) d\tau \leq \sigma A t, \quad 0 \leq \sigma \leq 1$$

Physically, any jump in $Y(\cdot)$ can be viewed as corresponding to photon arrivals to the receiver, so increment $Y(t+s) - Y(s)$ is exactly number of photons arriving in the duration $[s, t+s]$.

Discrete-time Poisson process can be regarded as the samples of the increments of output in continuous-time Poisson process. The precise procedure of discretizing is treated in next sections. However generally the discrete-time Poisson process is defined by the input and output sequences whose values take on $\mathbb{R}_0^+$ and $\mathbb{Z}_0^+$, respectively. If dark current in discrete time fashion is denoted by λ_0 ² then the output counting process has a Poisson distribution of mean $x + \lambda_0$. Thus

$$\Pr[Y = y | X = x] = e^{-(x+\lambda_0)} \frac{(x+\lambda_0)^y}{y!}; \quad y \in \mathbb{Z}_0^+, x \geq 0$$

This channel is often used to model pulse-amplitude modulation optical communication with a direct-detection receiver. Here the input x is proportional to the product of the transmitted light intensity by the pulse duration; λ_0 similarly models the time-by-intensity product of the background radiation and also output Y models the number of photons arriving at the receiver during the pulse duration.

Here a peak power constraint on the transmitter is accounted for by

$$0 \leq X \leq A$$

and the average input constraint by

$$E[X] \leq \sigma A, \quad 0 \leq \sigma \leq 1.$$

²The relation between dark current in discrete-time and continuous-time becomes clear in next section.

The case $\sigma = 1$ corresponds to the absence of an average power constraint, whereas $\sigma \ll 1$ corresponds to a very small peak power constraint.

After some engineering definitions for Poisson point process, now it is turn for mathematics to come to the stage and define more precisely.

Definition 5.1 (Point Process) *Let S be a second countable space equipped with its Borel σ -algebra. Let $\mathfrak{R}$ denote the set of locally finite counting measure on S and $\mathfrak{S}$ denote the smallest σ -algebra on $\mathfrak{R}$ that renders all the point counts. Then a point process on S is a measurable map $\xi : \Omega \rightarrow \mathfrak{R}$ from a probability space $(\Omega, \mathcal{F}, P)$ to the measurable space $(\mathfrak{R}, \mathfrak{S})$.*

A point process is somehow depicted as a random element whose values are "point patterns". It is sufficient for more applied purposes to think of a point pattern as a countable pattern.

Definition 5.2. *Let $(\Omega, \mathcal{F})$ be a measurable space. A history $(\mathcal{F}_t)$ $t \geq 0$ on $(\Omega, \mathcal{F})$ is a family of sub- σ -fields of $\mathcal{F}$ such that for all $0 \leq s \leq t$, $\mathcal{F}_s \subset \mathcal{F}_t$.*

In words, a history is an increasing family of sub- σ -fields of $\mathcal{F}$ indexed by the non-negative real numbers. It is also called a *filtration*. There is another interesting definition for history which is borrowed from [31]. For any random variable X_t , *internal history* is defined as

$$\mathcal{F}_t^X = \sigma(X_s, s \in [0, t]), \quad t \geq 0.$$

any set $\mathcal{F}_t$ such that

$$\mathcal{F}_t \supset \mathcal{F}_t^X \quad t \geq 0$$

is called a *history* of X_t . Then one can also say: X_t is *adapted* to $\mathcal{F}_t$. Now everything is ready to define formally Poisson process.

Definition 5.3. *Let $Y(t)$ be a point process adapted to a history $\mathcal{F}_t$ and let λ_t be a non-negative measurable process (all given on the*

same probability space $(\Omega, \mathcal{F}, P)$). Suppose that λ_t is $\mathcal{F}_0$ -measurable and $\int_0^t \lambda_s ds < \infty$. If for all $0 \leq s \leq t$ and all $k \geq 0$

$$\Pr(Y(t) - Y(s) = k | \mathcal{F}_s) = e^{-\Lambda} \frac{\Lambda^k}{k!}$$

where

$$\Lambda = \int_s^t \lambda_u du$$

then $Y(t)$ is a $\mathcal{F}_t$ stochastic Poisson process with intensity λ_t , or $Y(t)$ can admit the $\mathcal{F}_t$ intensity λ_t through the above law.

If in above definition $\lambda_t = f(t, \psi_t)$ for some appropriately measurable non-negative function f and for some measurable process ψ_t , and if $\mathcal{F}_0$ contains $\mathcal{F}_\infty^\psi$, then $Y(t)$ is Poisson process which is *driven* by ψ_t . This is the case for intensity of Poisson channel when feedback is present. In the presence of feedback, intensity of process is a function of channel output history, in other words, $\psi_t = Y(s)$, $s < t$ and λ_t is built using $\sigma(Y_s, s < t)$. In this case one can say λ_t is $\mathcal{F}_t$ -progressive process.

5.2 Capacity per Unit Cost for Poisson Channel

The capacity per unit cost for Poisson channel can be directly obtained from the formula proposed by Verdú in [8] for channel having one zero input symbol. If we set the cost function to be $b[x] = x$, then the free input symbol is $X = 0$. For a Poisson channel under an average power constraint β and no peak power constraint, i.e., $E[b[X]] = E[X] \leq \beta$, the capacity per unit cost is given by

$$C = \sup_{\beta > 0} \frac{C(\beta)}{\beta} = \lim_{\beta \downarrow 0} \frac{C(\beta)}{\beta}$$

where $C(\beta)$ which is the channel capacity-cost function is defined in chapter 2. The second equality in above equation was clarified in chapter 2.

Now the capacity per unit cost for Poisson channel can be calculated using the formula given in theorem 2.2 as follows:

Consider a discrete-time Poisson channel with input $X \in \mathbb{R}_0^+$, output $Y \in \mathbb{Z}_0^+$ and channel law $W(y|x) = \Pr(Y = y|X = x) \triangleq \mathcal{P}_{(x+\lambda_0)}$. The input alphabet includes a free symbol which is $X = 0$, hence the capacity per unit cost is

$$\begin{aligned}
 C &= \sup_{x>0} \frac{D(W(\cdot|x)||W(\cdot|0))}{b[x]} \\
 &= \sup_{x>0} \frac{D(\mathcal{P}_{(x+\lambda_0)}||\mathcal{P}_{(\lambda_0)})}{b[x]} \\
 &= \sup_{x>0} \frac{1}{x} \sum_{y=0}^{\infty} \mathcal{P}_{(x+\lambda_0)}(y) \log \left(\frac{\mathcal{P}_{(x+\lambda_0)}}{\mathcal{P}_{(\lambda_0)}} \right) \\
 &= \sup_{x>0} \frac{1}{x} \sum_{y=0}^{\infty} \mathcal{P}_{(x+\lambda_0)}(y) \log \left(\frac{e^{-(x+\lambda_0)}(x+\lambda_0)^y}{e^{-(\lambda_0)}\lambda_0^y} \right) \\
 &= \sup_{x>0} \frac{1}{x} \sum_{y=0}^{\infty} \mathcal{P}_{(x+\lambda_0)}(y) \log \left(e^{-x} \left(1 + \frac{x}{\lambda_0} \right)^y \right) \\
 &= \sup_{x>0} \frac{1}{x} \sum_{y=0}^{\infty} \mathcal{P}_{(x+\lambda_0)}(y) \left(-x + y \log \left(1 + \frac{x}{\lambda_0} \right) \right) \\
 &= \sup_{x>0} \frac{1}{x} \left(\underbrace{-x \sum_{y=0}^{\infty} \mathcal{P}_{(x+\lambda_0)}(y)}_{=1} + \log \left(1 + \frac{x}{\lambda_0} \right) \underbrace{\sum_{y=0}^{\infty} y \mathcal{P}_{(x+\lambda_0)}(y)}_{=(x+\lambda_0)} \right) \\
 &= \sup_{x>0} \left(-1 + \left(1 + \frac{\lambda_0}{x} \right) \log \left(1 + \frac{x}{\lambda_0} \right) \right) \tag{5.1} \\
 &= \infty \quad \forall \lambda_0 \geq 0
 \end{aligned}$$

We can see that the capacity per unit cost of Poisson channel under no peak power constraint is infinity.

Computing capacity per unit cost for continuous-time is a bit more involving. If the channel input in continuous-time fashion is denoted by $\{\lambda(t), t \in [0, T]\}$, we can consider the input alphabet to be the set of waveforms $\{x(t), t \in [0, T_0]\}$ for fix T_0 and then cost function is $b[x(t)] = \int_0^{T_0} x(t)dt$. Hence there is a free symbol; $\{x(t) = 0, t \in [0, T_0]\}$. Without any restriction on the variations of these waveforms, it is clear that if we let $T = nT_0$ and also let codeword $(\{x_1(t)\}, \dots, \{x_n(t)\})$ cor-

respond with the continuous-time input $\lambda(t) = \sum_{i=1}^n x_i(t - iT_0)$, then the model is equivalent to the original one [8].

The relative entropy between $\mathcal{P}_{(x(t)+\lambda_0)}$ and $\mathcal{P}_{(\lambda_0)}$ for $t \in [0, T_0]$ can be computed using the sample function density of a Poisson point process with rate $\{\lambda(t), t \in [0, T_0]\}$ with respect to the probability measured generated by a unit rate Poisson process evaluated at the unordered realization $(t_1, t_2, \dots, t_K)$ [32] as follows:

$$\mathcal{P}_{(\lambda(t))} = \exp\left(\int_0^{T_0} [1 - \lambda(t)] dt\right) \prod_{i=1}^K \lambda(t_i)$$

then

$$\begin{aligned} D(\mathcal{P}_{x(t)+\lambda_0} || \mathcal{P}_{\lambda_0}) &= E \left[\log \frac{\exp\left(-\int_0^{T_0} [x(t) + \lambda_0] dt\right) \prod_{i=1}^K [x(t_i) + \lambda_0]}{\exp\left(-\int_0^{T_0} \lambda_0 dt\right) \prod_{i=1}^K \lambda_0} \right] \\ &= -\int_0^{T_0} x(t) \log e dt + E \left[\sum_{i=1}^K \log \left(1 + \frac{x(t_i)}{\lambda_0} \right) \right] \\ &= -\int_0^{T_0} x(t) \log e dt \\ &\quad + \int_0^{T_0} [x(t) + \lambda_0] \log \left(1 + \frac{x(t)}{\lambda_0} \right) dt \end{aligned}$$

where that the expectation in above expression is taken with respect to the measure generated by the Poisson process with rate $x(t) + \lambda_0$, $t \in [0, T_0]$. Due to the theorem 2.2, the capacity per unit cost is

$$\begin{aligned} C &= \sup_{t \in [0, T_0]} -\log e + \frac{\int_0^{T_0} [x(t) + \lambda_0] \log \left(1 + \frac{x(t)}{\lambda_0} \right) dt}{\int_0^{T_0} x(t) dt} \\ &= \sup_{\alpha \geq 0} -\log e + \left(1 + \frac{\lambda_0}{\alpha} \right) \log \left(1 + \frac{\alpha}{\lambda_0} \right) \end{aligned}$$

Hence again if no peak power constraint is imposed on Poisson channel, the capacity per unit cost is infinity.

Two other important results can be inferred from expression (5.1) as follows:

- Even if Poisson channel is subjected to peak power constraint, $X < A$ (or equivalently $X(t) < A$ in continuous-time fashion),

the capacity per unit cost for *ideal* Poisson channel will be ∞ . This is because setting $\lambda_0 = 0$ causes the argument in (5.1) to become infinity. The capacity per unit cost being ∞ implies that the minimum energy per bit is zero. This arises the challenging problem of finding the coding scheme for ideal Poisson channel whose energy for transmission one bit of information is zero. This will be treated in the following sections.

- In order for *non-ideal* Poisson channel under peak power constraint to achieve the capacity per unit cost, input should take the maximum allowable value, i.e., the peak power constraint. This is because the function $\left(1 + \frac{\lambda_0}{x}\right) \log \left(1 + \frac{x}{\lambda_0}\right)$ is monotonically increasing in x .

5.3 Coding scheme in Discrete-Time Poisson Channels

In chapter 2 capacity per unit cost was introduced as the reciprocal of the minimum energy (or generally cost) per bit. In previous section we showed that the capacity per unit cost of ideal Poisson is infinity, so we expect that the minimum energy incurred by the transmission of one bit of information on ideal Poisson channel would be zero. To ease the analysis of the different schemes from energy point of view and also the error probability we need to know how to discretize the continuous-time Poisson channel and turn it into discrete-time. In this section we firstly explain how to convert the continuous-time Poisson channel into the discrete-time and then we talk about some pretty straightforward schemes in discrete time fashion and show that their performance are asymptotically the same. The main result generally in continuous-time fashion will be given finally in next chapter.

Let input process $X(t)$, $0 \leq t \leq T$ be given. Furthermore let $\nu(t)$ be a counting process which is used to count the number of photon registered in the receiver (or photon detector). For any function $g(t)$, $0 \leq t \leq \infty$, suppose g_a^b denotes $\{g(t) : a \leq t \leq b\}$. Therefore ν_0^T is the number of photons registered during the whole duration of input signal. The following gives a formal definition for coding scheme

in Poisson channel.

Definition 5.4. A code scheme for Poisson channel for transmission of one bit of information with parameters (T, σ, P_e) is defined as

- A set of two waveforms $X_0(t)$ and $X_1(t)$, $0 \leq t \leq T$ which satisfy the peak power constraint $\Pr(X_i(t) > A) = 0$, $0 \leq t \leq T$, $i = 0, 1$ and average power constraint as

$$\frac{1}{T} \int_0^T X_i(t) dt \leq \sigma A, \quad 0 \leq \sigma \leq 1$$

- A decoder mapping $D : \nu_0^T \rightarrow \{0, 1\}$

Then the overall error probability is

$$P_e = \frac{1}{2} \sum_{m=0}^1 \Pr(D(\nu_0^T) \neq m | X_m(t))$$

A code as defined above can be used in a communication system in the usual way to transmit H_1 or H_0 , thus when $X_i(t)$ is sent the waveform $\nu(t)$, $0 \leq t \leq T$ is received and is decoded as $D(\nu_0^T)$. Definition 5.4 does not incorporate feedback. In the presence of feedback, the coding scheme is a bit more technical (refer to [18]).

Definition 5.5. A feedback coding scheme for sending one single bit of information consists of two waveforms (as before). Conditional on H_1 the channel output is a Poisson process, $Y(t)$, which admits the $\mathcal{F}_t$ intensity $X_1(t) + \lambda_0$. That is conditional on H_1 , $Y(t)$ is a point process adapted to some history $\mathcal{F}_t$ and $X_1(t)$ is a non-negative $\mathcal{F}_t$ -progressive process.

Wyner in [16] and [33] proved the equivalence of above coding scheme to the following discrete-time coding scheme:

Let $\{X_i(t)\}_{i=0}^1$ be the code with parameters (T, σ, P_e) as defined above. Let $\epsilon > 0$ be arbitrary, then another code $\{\hat{X}_i(t)\}_{i=0}^1$ with parameters $(T, \sigma, \hat{P}_e)$ exists for which $\hat{P}_e \leq P_e + \epsilon$ and such that the

new code has the following properties. For some sufficiently large N and $\Delta = T/N$,

- (a) The new decoder mapping $\hat{D} : \{\nu(n\Delta)\}_{n=1}^N \rightarrow \{0, 1\}$
- (b) For both of waveforms, $\hat{X}_i(t)$ is constant on the interval $[(n-1)\Delta, n\Delta]$, $1 \leq n \leq N$; further $\hat{X}_i(t)$ takes only the values 0 or A .

In other words, channel output jump process, $\nu(t)$ can be approximated by its uniformly spaced samples $\nu(n\Delta)$ or equivalently by its increments $\nu(n\Delta) - \nu((n-1)\Delta)$.

Here we take a similar discretization procedure; picking a sufficiently small Δ and approximating the continuous-time input by its energy in the interval $[(n-1)\Delta, n\Delta]$. Precisely, given the input waveforms $\{X_i(t)\}_{i=0}^1$, the sequence $\{X_{ij}\}$, $i = 0, 1$, $j = 1, 2, \dots, N$ are built as

$$X_{ij} \triangleq \frac{1}{\Delta} \int_{(j-1)\Delta}^{j\Delta} X_i(t) dt \quad (5.2)$$

in other words, X_{ij} is just the average energy of signal $X_i(t)$ in the interval $[(j-1)\Delta, j\Delta]$. Using the mean value theorem we can make sure that letting $\Delta \rightarrow 0$ implies $\{X_{ij}\}_{j=1}^N \rightarrow X_i(t)$. Hence the procedure degrades the original performance in negligible way when Δ is sufficiently small.

The number of photon in each interval is therefore $\nu_{(j-1)\Delta}^{j\Delta} = \nu(j\Delta) - \nu((j-1)\Delta)$. For sufficiently small value of Δ , $\nu_{(j-1)\Delta}^{j\Delta}$ is either 0 or 1 and the probability of getting more photon in each interval is so small that one can ignore.

The definition of sequence in (5.2) enables us to modify the probability distribution as follows. Let $W(\nu|k) = \Pr(\nu_{(j-1)\Delta}^{j\Delta} = \nu | X_j = k)$ then

$$\begin{aligned} W(0|k) &= e^{-(k+\lambda_0)\Delta} \\ W(1|k) &= 1 - e^{-(k+\lambda_0)\Delta} \end{aligned}$$

The decoding algorithm can be a simple hypothesis testing which tests null hypothesis H_0 against H_1 . The log-likelihood ratio can be written

as

$$\begin{aligned}
S_n &= \ln \frac{\Pr(\nu_0^{n\Delta} | H_1)}{\Pr(\nu_0^{n\Delta} | H_0)} \\
&= \sum_{j=1}^n \ln \frac{\Pr(\nu_{(j-1)\Delta}^{j\Delta} | H_1, \nu_0^{(j-1)\Delta})}{\Pr(\nu_{(j-1)\Delta}^{j\Delta} | H_0, \nu_0^{(j-1)\Delta})} \quad (5.3)
\end{aligned}$$

When feedback is present, transmitter can know the previous outputs, that is, transmitter at time j knows $\nu_0^{(j-1)\Delta}$. Input sequence will then be $X_{ij} = f(H_i, \nu_0^{(j-1)\Delta})$ or equivalently $X_{ij} = f(H_i, S_{j-1})$. Getting back to log-likelihood ratio in (5.3)

$$\begin{aligned}
S_n &= \sum_{j=1}^n \ln \frac{\Pr(\nu_{(j-1)\Delta}^{j\Delta} | H_1, \nu_0^{(j-1)\Delta})}{\Pr(\nu_{(j-1)\Delta}^{j\Delta} | H_0, \nu_0^{(j-1)\Delta})} \\
&= \sum_{j=1}^n \ln \frac{\Pr(\nu_{(j-1)\Delta}^{j\Delta} | X_{1j})}{\Pr(\nu_{(j-1)\Delta}^{j\Delta} | X_{0j})} \quad (5.4)
\end{aligned}$$

The log-likelihood ratio can be further simplified by taking this fact into account that $\nu_{(j-1)\Delta}^{j\Delta}$ is either 0 or 1; if $\nu_{(j-1)\Delta}^{j\Delta} = 0$ then the $\Pr(\nu_{(j-1)\Delta}^{j\Delta} = 0 | X_{1j}) = e^{-\Delta(\lambda_0 + X_{1j})}$ and if $\nu_{(j-1)\Delta}^{j\Delta} = 1$ then the $\Pr(\nu_{(j-1)\Delta}^{j\Delta} = 1 | X_{1j}) = 1 - e^{-\Delta(\lambda_0 + X_{1j})}$. The same holds for X_{0j} , hence one can get this ratio

$$\frac{\Pr(\nu_{(j-1)\Delta}^{j\Delta} | X_{1j})}{\Pr(\nu_{(j-1)\Delta}^{j\Delta} | X_{0j})} = \begin{cases} \frac{e^{-\Delta(\lambda_0 + X_{1j})}}{e^{-\Delta(\lambda_0 + X_{0j})}} & \nu_{(j-1)\Delta}^{j\Delta} = 0 \\ \frac{1 - e^{-\Delta(\lambda_0 + X_{1j})}}{1 - e^{-\Delta(\lambda_0 + X_{0j})}} & \nu_{(j-1)\Delta}^{j\Delta} = 1 \end{cases}$$

Putting this two cases in one compact form and following (5.4), the log-likelihood ratio for ideal Poisson channel can be written as

$$S_n = \sum_{j=1}^n \ln \frac{\nu_{(j-1)\Delta}^{j\Delta} - e^{-\Delta X_{1j}}}{\nu_{(j-1)\Delta}^{j\Delta} - e^{-\Delta X_{0j}}} \quad (5.5)$$

Note that in this case the discrete-time Poisson channel turns into Z-channel. Hence motivated by feedback coding scheme in Z-channel, one can think of the same scheme for discrete-time Poisson channel, as follows:

- To transmit H_0 , we keep sending zero³ no matter what has been registered in receiver, i.e., $X_{0j} = 0, \forall j$. Note that zero is always received without any error in ideal Poisson channel.
- To transmit H_1 , we send $X_{1j} = f(H_1, \nu_0^{(j-1)\Delta})$ until we get one count, then transmission stops.

Above coding scheme is a variable-length code which causes no error as long as the number of transmission is unlimited. However in order to analyze the performance of this system, the decoder is forced to stop after N steps no matter whether or not any count is registered. The main task of transmitter is just to enable the receiver to distinguish H_0 against H_1 based on getting or non getting photon detected. Hence transmitter has to play with transmitted symbols to get a count registered. As soon as a count appears, receiver makes sure that H_1 is true and then informs the transmitter to stop sending.

The Markov chain corresponding to this coding scheme is shown in figure 5.1.

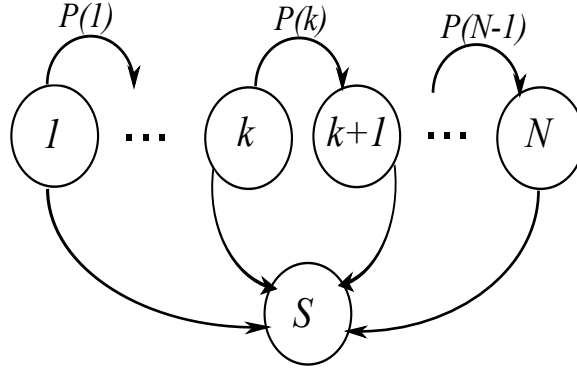


Figure 5.1: The Markov chain regarding the coding scheme in Poisson channel; each state represents the input sequence $X_j, j = 1, 2, \dots, N$ and also the state S represents the state of stopping before transmitter reaches N , i.e., $j = N$.

The Markov chain shown in figure 5.1 outlines the transmission through a Poisson channel. Each transmission step is represented in this chain as a state which is numbered $1, 2, \dots, N$. The stopping state S represents that a count has been registered and then the transmission is

³We will show later that sending always zero for H_0 is optimal which coincides with intuition.

stopped. The probability of going from state k to state $k + 1$ is denoted by $P(k)$ and is equal to $P(k) = e^{-\Delta X_{1k}}$ which is the probability of getting no count in step k when X_{1k} was sent and consequently the probability of moving from state k to state S is $1 - e^{-\Delta X_{1k}}$. The time to stop is defined as the number of transitions the system takes to reach the stopping state. Let $T(j)$ be the expected time to stop of the system when starting from state j , then

$$\begin{aligned} T(j) &= \begin{cases} (1 - P(j)) + P(j)(1 + T(j + 1)) & j < N \\ 1 & j = N \end{cases} \\ &= \begin{cases} 1 + P(j)T(j + 1) & j < N \\ 1 & j = N \end{cases} \end{aligned} \quad (5.6)$$

$$= 1 + \sum_{u=j}^{N-1} \prod_{v=j}^u P(v) + T(N) \prod_{u=j}^{N-1} P(u) \quad (5.7)$$

where $P(j)$ is the probability of passing from state j to state $(j + 1)$ in corresponding Markov chain, $T(N) = 1$ by the definition of state N and (5.7) is obtained by recursively evaluating (5.6). As the system always starts in state 1, we are more interested in $T(1)$. Taking the limit as N grows to infinity and considering the fact that $P(j) < 1$, we have

$$\begin{aligned} \lim_{N \uparrow \infty} T(1) &= 1 + \sum_{u=1}^{\infty} \prod_{v=1}^u P(v) \\ &= 1 + \sum_{u=1}^{\infty} \prod_{v=1}^u e^{-\Delta X_{1v}} \end{aligned} \quad (5.8)$$

If $P(j)$ is constant for each step, that is, the transmitted symbol in each step is the same and equal to $X_{1j} = x$, then $T(1)$ will be

$$\lim_{N \uparrow \infty} T(1) = \frac{1}{1 - e^{-\Delta x}}$$

which coincides with the intuition; because in this case time to stop is simply geometric random variable so its expected random variable is $\frac{1}{p_c}$ in which p_c means the probability of getting a count which equals $1 - e^{-x}$.

Let E denote the energy spent by this scheme in general case when all transmitted symbols are different, then

$$E = \Delta X_{11} + \Delta \sum_{u=1}^{\infty} X_{1(u+1)} \prod_{v=1}^u e^{-\Delta X_{1v}} \quad (5.9)$$

and also the error probability will be

$$P_e = \prod_{j=1}^N e^{-\Delta X_{1j}} \quad (5.10)$$

Note that the error occurs if and only if after N step of transmission, receiver does not get any count. Obviously letting N tend to infinity causes the error probability goes to zero and renders the system zero-error communication.

The first choice for input sequence function which intuitively comes to mind is the increasing function, for example linear or logarithmically. In former case $X_{1j} = jX_{11}$ and in the latter $X_{1j} = X_{11} \log(j)$, $j > 1$ then energy spent by linearly increasing input sequence is

$$E = \Delta X_{11} + \Delta X_{11} \sum_{u=1}^{\infty} (u+1) e^{-\Delta X_{11} \frac{u(u+1)}{2}} \quad (5.11)$$

and by logarithmically increasing input sequence is

$$E = \Delta X_{11} + \Delta X_{11} \sum_{u=1}^{\infty} \log(u+1) e^{-\Delta X_{11} \log(2u!)} \quad (5.12)$$

In both schemes error probability tends to zero when N is sufficiently large. The simplest scheme is to keep sending constant input, X , in each step for which $T(1)$ is

$$T(1) = \frac{1}{1 - e^{-\Delta X}}$$

and then the energy spent is

$$E = \frac{\Delta X}{1 - e^{-\Delta X}} \quad (5.13)$$

Apparently the spent energy in this case tends to one if the input in each step is so small ($\Delta X \rightarrow 0$) which does not coincide with what we theoretically expected. The simulations reveal that the energy

spent by two other previously explained schemes; linearly and logarithmically increasing input signal, are also convergent to one. As said before, the theoretically optimal energy for sending one bit of information in ideal Poisson channel is zero, so we expected to get the energy tending to zero. Later we show that although *theoretically* optimal energy is zero, we can not get the energy smaller than one, thus the simplest scheme is just to send constant signal until we get one count registered.

Figures 5.2 and 5.3 show the spent energy for three abovementioned schemes versus number of transmission as termination step, N , varies. As seen in these simulations as N gets larger, the energy tends to one. Note that in all simulations we have chosen X_{11} being small, i.e., $\Delta X_{11} \rightarrow 0$.

We also thought of decreasing signal as input. That is, transmitter sends decreasing signal as input, while keeping the error probability still tending to zero with N , until a count appears in receiver. In this scheme $X_{1j} = \frac{X_{11}}{j}$ and energy spent is

$$E = \Delta X_{11} + \Delta X_{11} \sum_{u=1}^{\infty} \frac{1}{u+1} e^{-\Delta X_{11} \sum_{j=1}^u \frac{1}{j}} \quad (5.14)$$

The energy for this scheme has also compared to that of other schemes in figure 5.4. The simulations in this figure show that the energy in this case also tending to one. This prompted us to think of the reason behind this observation and the question "can we somehow come up with a coding scheme whose incurred energy is smaller than one?" This question is dealt with and proved to have negative answer in next section.

5.4 Energy Analysis in Poisson Channels

In previous section we saw that the energy spent by four different schemes is always tending to one when error probability decays exponentially to zero with N (apart from the last scheme whose error probability is $\Theta(\frac{1}{N})$). In all coding scheme mentioned before, keeping N so large results in energy close to one. This observation cast doubt

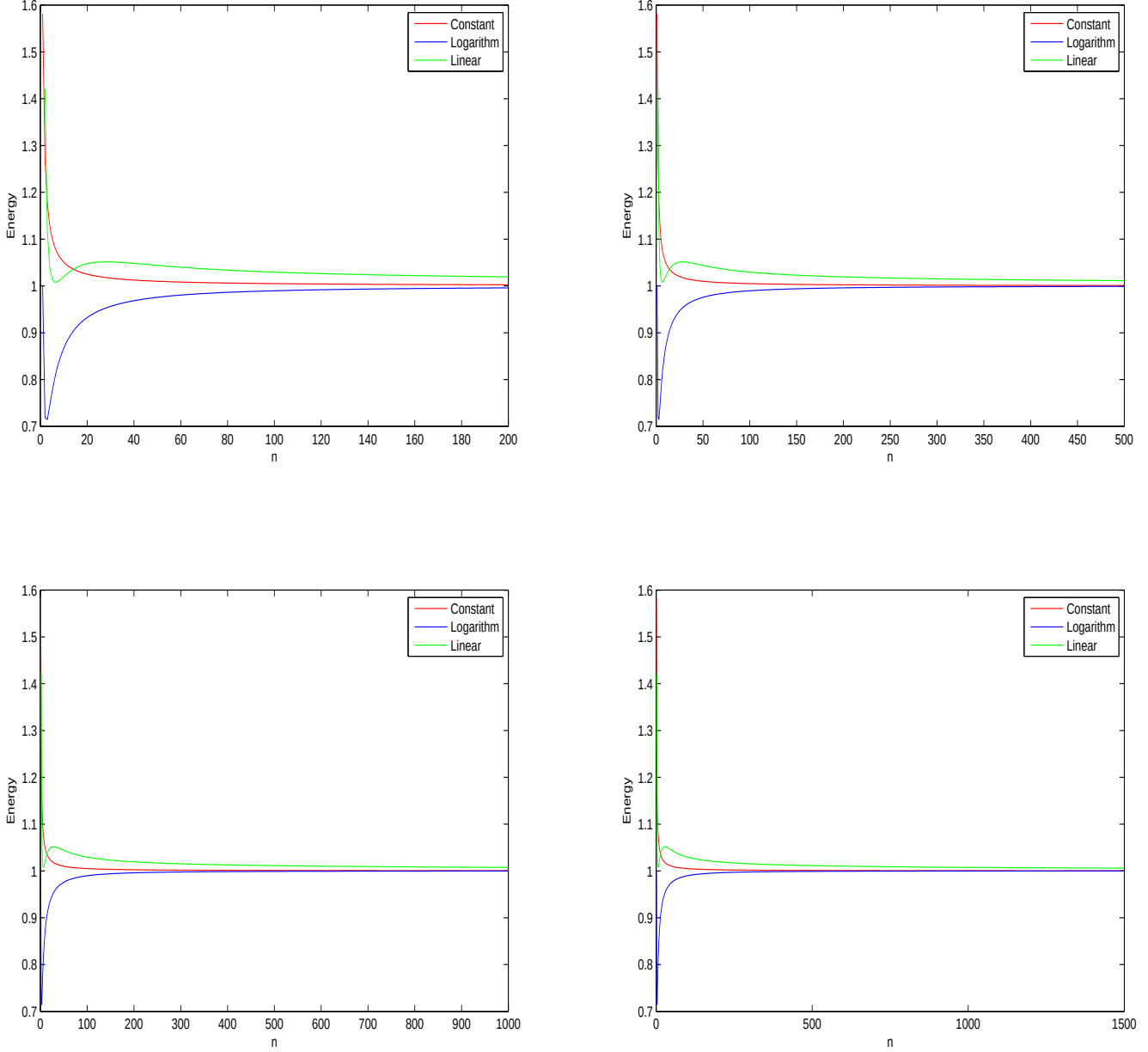


Figure 5.2: The spent energy versus number of transmission for different schemes; constant-input signal, linearly and logarithmically increasing-input signal, when the termination step or N varies from $N = 500$ to $N = 1500$. The energy spent are formulated in (5.11), (5.12) and (5.13).

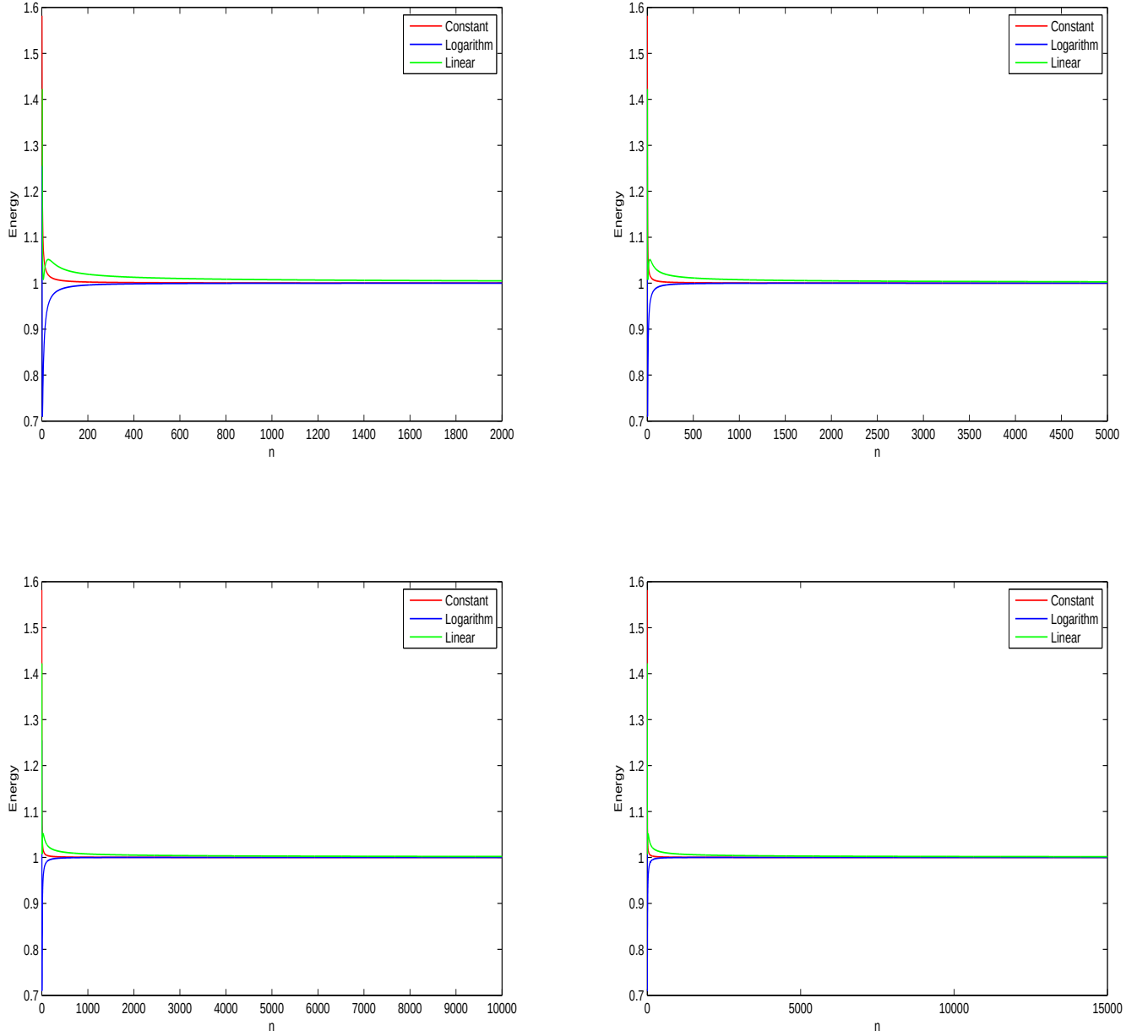


Figure 5.3: The spent energy versus number of transmission for different schemes; constant-input signal, linearly and logarithmically increasing-input signal, when the termination step or N varies from $N = 2000$ to $N = 15000$. The energy spent are formulated in (5.11), (5.12) and (5.13).

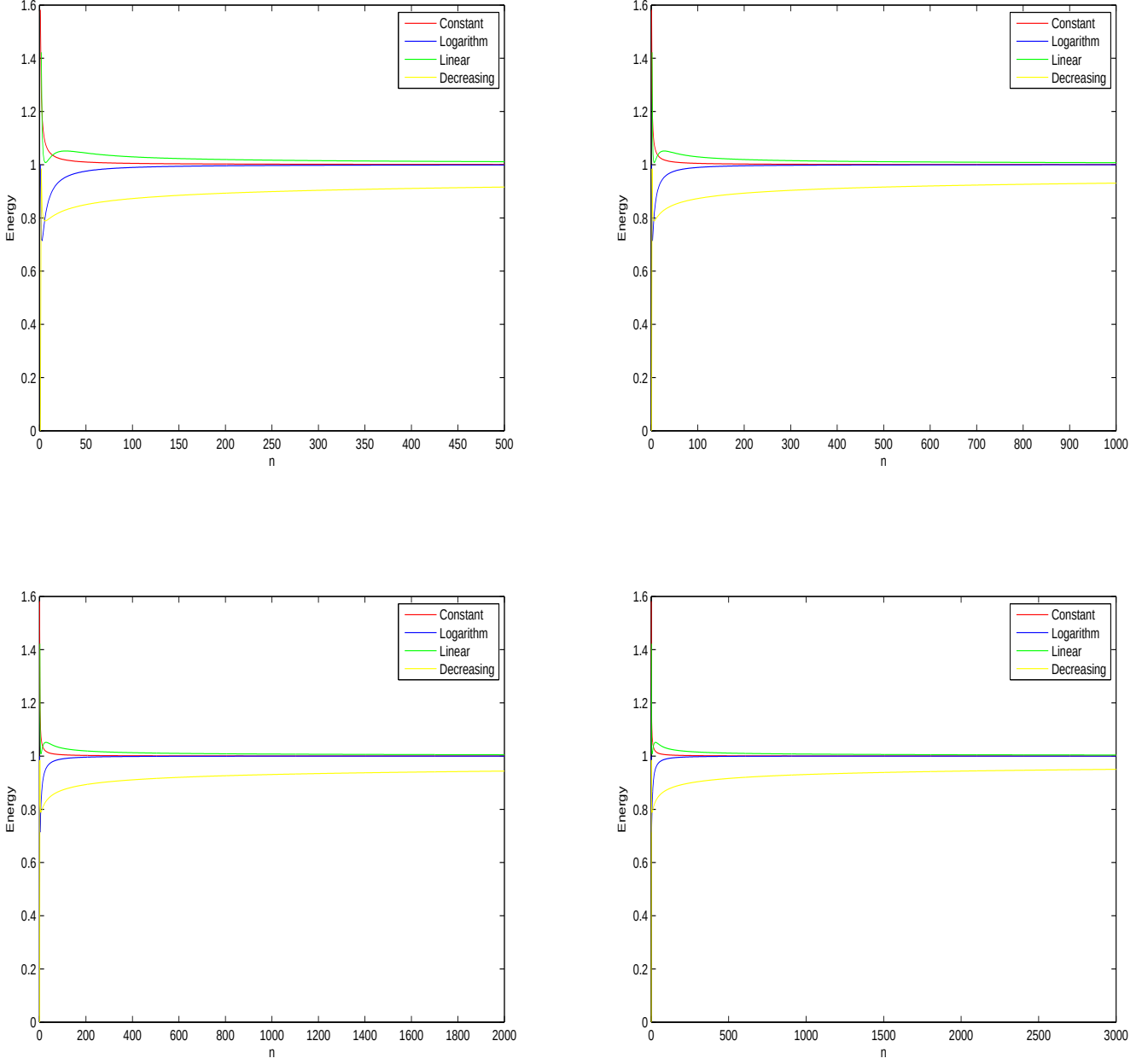


Figure 5.4: The comparison of spent energy in three previously mentioned coding schemes with decreasing-input scheme; when the termination step varies. $N = 500, 1000, 2000, 3000$. The spent energy in decreasing-input scheme is shown in (5.14).

on the original attempt made to find a coding scheme for ideal Poisson channel whose incurred energy is zero and also error probability is arbitrarily small. We take the continuous-time picture of Poisson channel to deal with the *availability* of this optimal coding scheme.

Suppose the input signals $X_1(t)$ and $X_0(t)$ are made in transmitter to convey H_1 and H_0 , respectively, when feedback is present. By this, we mean all $X_{ij} = f(H_i, \nu_0^{j-1})$, $i = 0, 1$, $j = 1, 2, \dots, N$ have been determined which together with sufficiently small Δ can build the continuous-time signals $X_0(t)$ and $X_1(t)$.

But before the main theorem, the following theorem has to be stated. We know that sending zero incurs no cost and also zero can be received correctly without any error. Hence intuitively in the optimum coding scheme, H_0 is conveyed to receiver by always sending zero. The following theorem is to prove that observation.

Theorem 5.1. *Sending always zero to convey H_0 is optimum in ideal Poisson channel.*

Proof: We want to show that there is no loss in optimality in assuming that to send H_0 , a zero waveform is transmitted. Indeed, given any general scheme consider the guess the decoder makes when faced with no counts, let's call that as FALSE and let TRUE be its complement. Consider now a scheme with the same encoding rule for TRUE, with the same decoding rule, but where we send the all-zero input to convey FALSE. The new scheme uses less or equal energy and has the same $\Pr(\text{error}|\text{TRUE})$ and has $\Pr(\text{error}|\text{FALSE}) = 0$. So this coding scheme is better than any other scheme that may be thought of. Since the name we give to hypotheses is immaterial, we can assume that FALSE corresponds to H_0 .

□

This theorem says the total energy for sending one bit of information is minimized if we keep sending zero for H_0 and sending $X_1(t)$ for H_1 .

In light of this intuitive fact, the optimal scheme should consist of two signals $X_1(t)$ and $X_2(t) = 0$. In this case the log-likelihood ratio stated in (5.5) turns into the following (supposing $X_1(t) = X(t)$):

$$S_n = \sum_{j=1}^n \ln \frac{\nu_{(j-1)\Delta}^{j\Delta} - e^{-\Delta X_j}}{\nu_{(j-1)\Delta}^{j\Delta} - 1}$$

hence as soon as a count is registered in receiver, decoder stops and ask the transmitter to stop transmission. So to transmit H_0 we need infinite transmission steps and to convey H_1 only finite many. This idea is also used in chapter 3 for general channel as an attempt to find the optimal scheme. The following theorem computes the energy spent in the optimal scheme. But before we get to the theorem, let's remind that when sending H_0 and H_1 , the total energy will be $\frac{1}{2}(\mathcal{E}_0 + \mathcal{E}_1)$ where $\mathcal{E}_0$ and $\mathcal{E}_1$ are the required energy for transmitting H_0 and H_1 respectively.

Theorem 5.2. *The energy required for transmission of one bit of information in ideal Poisson channel is always bigger than 1/2; and the minimum energy is achieved by $X(t) = k$ and $X_0(t) = 0$.*

Proof. To prove the theorem we need to find the energy spent in the optimal coding scheme, and since it was proved that setting $X_0(t) = 0$ is optimal, in the following we compute the energy required to convey H_0 and H_1 using signals $X(t)$ and constant zero signal $X_0(t)$. Since in this case $X_0(t)$ does not play any role in energy, the following stopping time has to be introduced in a bid to stop the transmission

$$T \triangleq \inf\{n : \nu_{(n-1)\Delta}^{n\Delta} = 1\}$$

then the spent energy becomes

$$\Lambda(T) = \int_0^T X(t) dt.$$

Upon registration of a count, the transmitter stops sending (or transmit zero afterwards).

Note that the error can happen if the horizon is finite, that is, if we force the decoder not to run to infinity and instead make decision at

most at T_{max} , irrespective of photon arrival. Then the error probability is equivalent to the probability that no count is registered up to T_{max} and is equal to $e^{-\Lambda(T_{max})}$, so letting $T_{max} \rightarrow \infty$ leads error probability tending to zero exponentially. If $f_T(t)$ denotes the probability density function of stopping time, then

$$f_T(t) = -\frac{d}{dt} \Pr(T > t).$$

Then the expected energy can obviously be written as

$$E[\Lambda(T)] = \int_0^{T_{max}} \Lambda(t) f_T(t) dt = \int_0^\infty \Lambda(t) f_T(t) dt \quad (5.15)$$

$$= [\Lambda(t) \Pr(T > t)]_\infty^0 + \int_0^\infty \Lambda'(t) \Pr(T > t) dt \quad (5.16)$$

$$= -\Lambda(\infty) \Pr(T = \infty) + \int_0^\infty \Lambda'(t) \Pr(T > t) dt \quad (5.17)$$

$$= \int_0^\infty \Lambda'(t) e^{-\Lambda(t)} dt \quad (5.18)$$

$$= 1 \quad (5.19)$$

where (5.15) is because the transmitter sends zero after one count is registered, (5.16) is just an integration by parts, (5.17) is the evaluation of (5.16) taking into account $\Lambda(0) = 0$, (5.18) uses the observation that $\Pr(T = \infty) = e^{-\Lambda(\infty)}$ which decays exponentially to zero so $\Lambda(\infty) \Pr(T = \infty)$ decays to zero as well and also $\Pr(T > t) = e^{-\Lambda(t)}$ and finally (5.19) becomes clear if we do the following change of variable; $\Lambda(t) = s$ and then the integral will be $\int_0^\infty e^{-s} ds$ which is one. Thus the energy spent by optimal scheme is one. Note that this amount of energy can be achieved by simply sending constant signal as $X(t)$ or equivalently $X_j = k, \forall j$ so the simplest coding scheme is the best one.

□

This theorem shows that the optimal coding scheme incurs one unit of energy which contradicts our theoretical expectation. Getting an infinity capacity per unit cost, we expected to get the energy close to zero but it turns out that energy in ideal Poisson channel can never reach zero and always bigger than one. Another conclusion of this

theorem is that sending constant signal for both H_1 and H_0 is the best scheme and achieves the minimum energy.

The following theorem states the general conclusion for any function of stopping time T in ideal Poisson channel whose input is constant.

Theorem 5.3. *Let T be the stopping time in Poisson channel whose input is constantly k and $f(\cdot)$ be an integrable function, then*

$$E[f(T)] = k \int_0^\infty e^{-kv} f(v) dv$$

Proof.⁴ Unlike before, we use the continuous-time and discrete-time definition of stopping time, intermittently. Firstly note that $\nu_{j\Delta}^{(j+1)\Delta}$ is independent from $\nu_0^{j\Delta}$ and consequently from $\sigma(\nu_0^{j\Delta})$ and also let us approximate the probability of getting count in interval $[j\Delta, (j+1)\Delta]$ as

$$\Pr(\nu_{j\Delta}^{(j+1)\Delta} = 1 | \sigma(\nu_0^{j\Delta})) = 1 - e^{-k\Delta} \approx k\Delta$$

for any infinitesimal Δ and also let $\mathcal{H}_{j\Delta}$ be the filtration generated by the process $1_{\{T > j\Delta\}}$ or equivalently in continuous-time $\mathcal{H}_t$ be the filtration generated by the process $1_{\{T > t\}}$. In fact we have assumed that $s\Delta = r$ and $j\Delta = t$. Then it is easy to realize

$$E[1_{\{T \in [j\Delta, (j+1)\Delta]\}} | \mathcal{H}_{j\Delta}] = k\Delta 1_{\{T > j\Delta\}} \quad (5.20)$$

and also

$$\begin{aligned} E[1_{\{T \in [j\Delta, (j+1)\Delta]\}} | \mathcal{H}_{s\Delta}]_{s < j} &= E[E[1_{\{T \in [j\Delta, (j+1)\Delta]\}} | \mathcal{H}_{j\Delta}] | \mathcal{H}_{s\Delta}]_{s < j} \\ &= k\Delta E[1_{\{T > j\Delta\}} | \mathcal{H}_{s\Delta}]_{s < j} \end{aligned} \quad (5.21)$$

or equivalently in continuous-time:

$$\begin{aligned} E[1_{\{T \in [t, t+\Delta]\}} | \mathcal{H}_r]_{r < t} &= E[E[1_{\{T \in [t, t+\Delta]\}} | \mathcal{H}_t] | \mathcal{H}_r]_{r < t} \\ &= k\Delta E[1_{\{T > t\}} | \mathcal{H}_r]_{r < t} \end{aligned}$$

In what follows, $E[1_{\{T > t\}} | \mathcal{H}_r]_{r < t}$ is computed. Suppose the interval $[r, t]$ is divided into M sub-intervals each of size δ , that is, $\delta = \frac{t-r}{M}$. If

⁴The more straightforward proof is given in Appendix B. @ courtesy of A. Nikeghbali

it is assumed that $T > r$ then

$$E[1_{\{T>t\}}|\mathcal{H}_r]_{r<t} = E\left[\prod_{u=0}^{M-1} 1_{\{T \notin [r+u\delta, r+(u+1)\delta]\}}|\mathcal{H}_r\right] \quad (5.22)$$

$$= E\left[E\left[\prod_{u=0}^{M-1} 1_{\{T \notin [r+u\delta, r+(u+1)\delta]\}}|\mathcal{H}_{t-\delta}\right]|\mathcal{H}_r\right] \quad (5.23)$$

$$= E\left[\prod_{u=0}^{M-2} 1_{\{T \notin [r+u\delta, r+(u+1)\delta]\}} E[1_{\{T \notin [t-\delta, t]\}}|\mathcal{H}_{t-\delta}]|\mathcal{H}_r\right] \\ = (1 - k\delta) E\left[\prod_{u=0}^{M-2} 1_{\{T \notin [r+u\delta, r+(u+1)\delta]\}}|\mathcal{H}_r\right] \quad (5.24)$$

where (5.23) is just a simple chain rule⁵ as we did before in (5.21), next line is because we know that conditioning on $\mathcal{H}_{t-\delta}$, the events $1_{\{T \notin [r+u\delta, r+(u+1)\delta]\}}$ for $u = 0, 1, 2, \dots, M-2$ are no longer random and become deterministic, so they can be taken out of expectation and finally (5.24) is written with the help of (5.20). Hence

$$E[1_{\{T>t\}}|\mathcal{H}_r]_{r<t} = 1_{\{T>r\}} (1 - k\delta) E\left[\prod_{u=0}^{M-2} 1_{\{T \notin [r+u\delta, r+(u+1)\delta]\}}|\mathcal{H}_r\right]$$

The expectation in above expression is the same as that of in (5.22) with this difference that in latter case product runs from $u = 0$ to $u = M-2$. This procedure of conditioning can be continued to get

$$E[1_{\{T>t\}}|\mathcal{H}_r]_{r<t} = 1_{\{T>r\}} (1 - k\delta)^M \\ = 1_{\{T>r\}} \left(1 - k \frac{t-r}{M}\right)^M$$

hence we can deduce the following limit when $M \rightarrow \infty$

$$E[1_{\{T>t\}}|\mathcal{H}_r]_{r<t} \rightarrow 1_{\{T>r\}} e^{-k(t-r)} \quad (5.25)$$

So plugging (5.25) into (5.21) yields

$$E[1_{\{T \in [j\Delta, (j+1)\Delta]\}}|\mathcal{H}_{s\Delta}]_{s<j} = 1_{\{T>s\Delta\}} e^{-k(j-s)\Delta} k\Delta \quad (5.26)$$

⁵Chain rule can generally be expressed as $E[X] = E_Z[E_X[X|Z]]$ in which E_X means the expectation taken with respect to random variable X .

The goal is to compute the expectation of $f(T)$ which can be written as follows

$$\begin{aligned} E[f(T)] &= E \left[\sum_{j=0}^{\infty} 1_{\{j\Delta \leq T \leq (j+1)\Delta\}} f(T) \right] \\ &\approx \sum_{j=0}^{\infty} E [1_{\{j\Delta \leq T \leq (j+1)\Delta\}}] f(j\Delta) \end{aligned} \quad (5.27)$$

where the second line is true because we have assumed that $\Delta \rightarrow 0$. The expectation in (5.27) is the same as what we obtained before in (5.26) when $s = 0$, so we get

$$\begin{aligned} E[f(T)] &\approx \sum_{j=0}^{\infty} e^{-kj\Delta} k\Delta f(j\Delta) \\ &\approx \int_0^{\infty} k e^{-kv} f(v) dv \end{aligned}$$

which concludes the proof. □

After looking into the minimal energy required in ideal Poisson channel, it is worthy to have a look at minimal energy in non-ideal Poisson channel. The result of Theorem 5.2 can be adopted for non-ideal Poisson channel by compressing the time. In this case the same procedure explained for theorem 5.2 will be done in interval $[0, \Delta]$ instead of $[0, T_{max}]$. Hence we need to sub-divide the interval $[0, \Delta]$ into smaller intervals and keep sending the input signal signals during these intervals.

Theorem 5.4. *The minimal possible energy for transmitting one bit of information in non-ideal Poisson channel is $1/2$.*

Proof. This theorem is just a generalized version of theorem 5.2 which was only about the ideal Poisson channel. Here we want to show that the energy can not be smaller than one also for non-ideal Poisson channel.

In this case, we do the sequential procedure stated in proof of theorem 5.2 but just for interval $[0, \Delta]$; in fact we sub-divide this interval to smaller sub-intervals each of size δ , so using the same notion as section 5.3 the input sequence is $X_j\delta$, $j = 1, 2, \dots, N$ where $N = \lceil \frac{\Delta}{\delta} \rceil$. Although in this case the probability of getting count when sending zero is not zero, it is so small that we can ignore. This becomes more clear when we compare $W(1|1)$ with $W(1|0)$ as follows:

$$W(1|0) = 1 - e^{-\lambda_0\delta} \approx \lambda_0\delta$$

and the probability of getting count for H_1 in step j is

$$W(1|1) = 1 - e^{-(X_j + \lambda_0)\delta}$$

In other words, compared to $W(1|1)$ (for large input) we can ignore $W(0|1)$, hence we can still assume the channel is modeled as Z channel, so the definition of stopping time given in the proof of theorem 5.2 is still valid with this difference that here $T_{max} = \Delta$. Also here again we can keep sending zero for H_0 and sending $X_j\delta$ for H_1 as the optimum scheme (based on theorem 5.1).

Let $\Lambda_1(T)$ and $\Lambda_0(T)$ denote the energy required for sending H_1 and H_0 respectively, so

$$\begin{aligned} \Lambda_1(T) &= \delta \sum_{j=1}^{\lceil \frac{T}{\delta} \rceil} (X_j + \lambda) \approx \int_0^T X(t)dt + \lambda(T + \delta) \\ \Lambda_0(T) &= \delta \sum_{j=1}^{\lceil \frac{T}{\delta} \rceil} \lambda \approx \lambda(T + \delta) \end{aligned}$$

recalling $T \leq T_{max} = \Delta$ and also $\Delta \rightarrow 0$, we can deduce the following

$$\begin{aligned} \Lambda_1(T) &\approx \int_0^T X(t)dt \\ \Lambda_0(T) &\approx 0 \end{aligned}$$

Therefore $\Lambda_0(T)$ does not play any role in total energy. Again the expected energy is defined as

$$E[\Lambda(T)] = \int_0^T \Lambda(t) f_T(t)dt$$

where as before $f_T(t) = -\frac{d}{dt} \Pr(T > t)$.

Here again we encounter error when H_1 is sent and no count is registered until $t = \Delta$, so

$$\Pr(\nu_0^\Delta = 0 | H_1) = e^{-\Lambda_1(\Delta)}$$

hence in order to have the zero error probability $\Lambda_1(\Delta)$ should be infinity, in other words, bigger input signal has to be sent for H_1 compared to theorem 5.2.

Taking integration by parts and also recalling (5.15), we get

$$\begin{aligned} E[\Lambda(T)] &= -\Lambda(\Delta) \Pr(T = \Delta) + \int_0^\Delta \Lambda'(t) \Pr(T > t) dt \\ &= -\Lambda(\Delta) e^{-\Lambda(\Delta)} + \int_0^\Delta \Lambda'(t) e^{-\Lambda(t)} dt \\ &= \int_0^\Delta \Lambda'(t) e^{-\Lambda(t)} dt \\ &= 1 \end{aligned}$$

which completes the proof.

□

Based on theorem 5.4 the minimal energy also for non-ideal Poisson channel is one. Hence provably no coding scheme can achieve the capacity per unit cost in both ideal and non-ideal Poisson channel. In other words, in both ideal and non-ideal Poisson channel, one is the minimum energy that we can get which is not the theoretically minimum energy derived by capacity per unit cost (in the absence of peak power constraint).

This section concerned the availability of a coding scheme which can achieve the capacity per unit cost. Based on theorems 5.2 and 5.4 no capacity-per-unit-cost-achieving scheme can be found for ideal and non-ideal Poisson channel. On the hand, we proved that the minimum energy for sending one single bit is not what is given by capacity per unit cost. That is, minimum energy for sending reliably one single bit differs from minimum energy per bit if one were sending large number of bits.

5.5 Further Discussion on Main Result

In previous section two theorems (5.2. and 5.4) were proved which concern energy efficiency of ideal Poisson channel and also non-ideal Poisson channel without any peak power constraint. In this section, we want to deal with the problem of sending k bits in the same channel as previous section. It is known that if the number of information bits goes to infinity, the availability of noiseless feedback does not lower the minimum energy per bit. This section addresses the behavior of the minimum energy per bit for Poisson channel as a function of number of information bits. Obviously in this non-asymptotic regime, noiseless feedback leads to significantly better energy efficiency. This problem for Gaussian channel has been treated theoretically in [7] and tightly approximated in [5] and [6]. But to our best knowledge nobody has ever worked on minimum energy per bit (when k is finite) in Poisson channel in both ideal and non-ideal regimes.

In the last section the minimal energy was proved in infinite-horizon regime. It can also be proved in finite horizon. The optimal coding scheme is to send all-zero input for H_0 and to send waveform $X_1(t)$, $0 \leq t \leq T$ (which is constant for ideal Poisson channel) for H_1 till one count is registered. Conditioned on H_1 the channel output $Y(t)$ admits the $\mathcal{F}_t$ intensity $X_1(t) + \lambda_0$ [refer to Definition 5.5]. That is, conditional on H_1 , $Y(t)$ is a point process⁶ adapted to some history $\mathcal{F}_t$ and $X_1(t)$ is a non-negative $\mathcal{F}_t$ -progressive process. In this case, since the input waveform is random, the energy should be formulated using expectation; namely

$$\mathcal{E}_1 = E \left[\int_0^T X_1(t) dt \right],$$

the expected energy spent to convey H_1 and similarly for H_0

$$\mathcal{E}_0 = E \left[\int_0^T X_0(t) dt \right] = 0.$$

It is known that for all non-negative $\mathcal{F}_t$ -predictable process $C(t)$ [31]

⁶A point process is a measurable mapping from some probability space to the space of point measure; a measure that is locally finite and takes only integer values on some space. For more detailed definition refer to Definition 5.1

$$E \left[\int_0^\infty C(t) dY(t) \right] = E \left[\int_0^\infty C(t) (X_1(t) + \lambda_0) dt \right]$$

We need to argue that there is no loss in optimality in restricting ourselves to a detector that bases its decision on the presence of counts in the interval $[0, T]$. Since this enlarges the set of outcomes yielding the guess H_1 at decoder, this can not increase $\Pr(\text{error}|H_1)$. To send H_0 we still send all-zero input waveform which results in no counts ($\lambda_0 = 0$), hence this does not change $\Pr(\text{error}|H_0)$. Let T_1 denote the random time at which the first count is registered, therefore the following is true

$$\Pr(\text{correct}|H_1) = \Pr(T_1 \leq T|H_1) = E[Y(T_1 \wedge T)|H_1]$$

where $T_1 \wedge T \triangleq \min\{T_1, T\}$. If we define the stochastic process $C(t) = 1_{\{t \leq T_1 \wedge T\}}$, then we can simplify the above calculation as follows

$$\begin{aligned} \Pr(\text{correct}|H_1) &= E \left[\int_0^\infty C(t) dY(t) | H_1 \right] = E \left[\int_0^\infty C(t) X_1(t) dt \right] \\ &= E \left[\int_0^{T_1 \wedge T} X_1(t) dt \right] = \mathcal{E}_1 \end{aligned}$$

Hence to have error probability tending to zero, the expected energy required to transmit H_1 should tend to one. This is the result that was proved in theorem 5.2. Clearly if T is set to be Δ which is sufficiently small value and we transmit high power for H_1 , the result holds for non-ideal Poisson channel without peak power constraint.

Now we want to analyze the energy spent to send $k = \log_2 \mathcal{M}$. The following theorem addresses this problem.

Theorem 5.5. *The minimum energy to convey $k = \log_2 \mathcal{M}$ bits is $\frac{\mathcal{M}-1}{\mathcal{M}}$.*

Proof.

We start from the converse part-that no coding scheme exists yielding less energy. To prove the converse we take advantage of "genie-aided"

argument. Although this proof assumes zero dark current, it can easily be modified for non-ideal Poisson channel when high input power is allowed. Again thanks to theorem 5.1, there is no loss in optimality to consider decoder guessing the transmitted message was zero if and only if no counts were registered. Suppose there is a genie that if a count is registered, tells the decoder which message was sent. With the aid of the genie, decoder may err as long as the message is not zero and we get no count. Previous analysis showed that if the energy for each non-zero message approaches one, the reliable communication can be made. Averaging over the equally likely messages yields the total energy $\frac{\mathcal{M}-1}{\mathcal{M}}$.

The direct part is divided into two parts, in first part we propose a coding scheme for ideal Poisson channel and in second part we modify it to make it suitable for non-ideal Poisson channel.

The simple coding scheme in the presence of dark current is based on *time division* technique. The transmission interval $[0, T]$ is divided into $(\mathcal{M}-1)$ sub-intervals. To send the non-zero message $m \in \{1, \dots, \mathcal{M}-1\}$, we transmit the waveform X_m with high power in the sub-interval

$$[(m-1)T/(\mathcal{M}-1), mT/(\mathcal{M}-1)].$$

We keep sending waveform till we get a count or the time is over, i.e. we reach the end of sub-interval. If no counts were registered, the decoder declares that zero has been transmitted and if a count is registered the decoder declares that the transmitted message was the one corresponding to the sub-interval in which the count was registered. Error probability analysis for this coding scheme is exactly the same as the one considered in theorem 5.2. Hence if the energy for each of these waveforms is one, error probability goes to zero.

In the presence of dark current we use the same coding scheme except that in each sub-interval we choose a very short period $[0, \Delta]$ as our transmission time. We choose sufficiently high power and small transmission time to assure the probability of spurious count is negligible

□

This section addresses a challenging non-asymptotic analysis of energy analysis as a function of number of bits when noiseless feedback

is available. We generalized the theorem 5.4 for sending reliably finite number of bits, obtained the minimum energy and gave the optimum coding scheme which achieves it.

The analyses in [7], [5] and [14] can show that in infinite-degree-of-freedom-Gaussian channel the minimum energy for sending one single bit is exactly the same as minimum energy per bit if large number of bits are to be transmitted. This is what we expected for Poisson channel, but this chapter proved that when sending one single bit in ideal Poisson channel we need $1/2$ energy of photon whereas its minimum energy per bit when sending large number of bits is proved to be zero. Therefore the situation changes drastically in Poisson channel. The logical extension of this chapter is the energy analysis for sending one single bit in non-ideal Poisson channel ($\lambda_0 > 0$) when peak power constraint is also imposed. What is interesting to know is *"what is the least amount of information for sending a single bit over Poisson channel in finite capacity per unit cost regime?"*.

Chapter 6

Poisson Channel, Finite Capacity Per Unit Cost

As if there could be true stories: things happen in one way, and we retell them in the opposite way.
 –J. P. Sartre

In previous chapter where we analyzed the energy in ideal and non-ideal Poisson channel, only average power was assumed to be constrained. That is, in a particular time slot we can transmit a very big power. It was proved that for ideal Poisson channel, we can keep sending a constant waveform $X_1(t) = k$ for arbitrarily small k till a count is registred in output as the optimal coding scheme. The smaller k , the bigger stopping time T . If $\lambda_0 > 0$ and there is no peak power constraint, we choose $k = A \gg 1$ in which A is peak power and $\Delta \ll 1$ and use the above scheme with $T = \Delta$. We make sure that Δ is small enough for the probability of spurious count in the interval $[0, \Delta]$ to be very small ¹. We also have to choose A large enough so that the probability that a mean- $\frac{1}{A}$ exponential exceeds Δ is also very small. It was proved that this coding scheme is optimal in the sense of incurred energy for reliably transmitting one single bit. The straightforward question which may arise is *nt?*. It was shown in section 5.2 that if $\lambda_0 \neq 0$ and bounded input is always to be transmitted, the capacity per unit cost is finite.

¹The probability of spurious count in this interval is $1 - e^{-\Delta\lambda_0}$.

Although we tried many different techniques and we got some interesting results, this problem has not yet been solved. In section 6.1, we will talk about a technique in which a one-bit quantizer is set after channel to output binary data, zero or one, using two different approaches. Although we could not theoretically prove the (in)stability of the energy consumption in this technique, computer simulation was used to show that energy performance in this technique is instable and monotonically increasing when error probability decreases. Section 6.2 concerns another method which imposes a stopping criterion on quantizer's output based on the number of ones and zeros. We could mathematically prove that this method yields an infinite energy for small error probability. At finall, it should be pointed out that both of these two techniques can be perfectly used when error probability is bounded and is not necessarily tending to zero.

6.1 One-bit Quantizer

In previous chapter we emphasized that most energy-efficient coding scheme in both ideal Poisson channel (even with peak power constraint) and non-ideal one without peak power constraint looks like the same with this difference that in former $T \rightarrow \infty$ whereas in latter $T = \Delta$, a sufficiently small number. This technique fails when there is a peak power constraint imposed on non-ideal Poisson channel.

As the first attempt, we tried to process the output of non-ideal channel in such a way to make it behave like ideal channel. To do so, we consider a one-bit quantizer² after channel. Let y_i and z_i are the output of Poisson channel and quantizer, respectively; then we define

$$z_i \triangleq \begin{cases} 1 & \ln \frac{\Pr(y_i|H_1)}{\Pr(y_i|H_0)} > \alpha \\ 0 & \ln \frac{\Pr(y_i|H_1)}{\Pr(y_i|H_0)} < -\alpha \end{cases}$$

or equivalently:

²The effect of one-bit quantizer on capacity per unit cost for Gaussian channel is studied in [34].

$$z_i \triangleq \begin{cases} 1 & y_i > \frac{A+\alpha}{\gamma} \\ 0 & y_i < \frac{A-\alpha}{\gamma} \end{cases}$$

where $\gamma \triangleq \ln(1 + \frac{A}{\lambda_0})$.

By doing this, we can have a channel whose output is binary and with high probability (depending on how large α is) we receive zero in H_0 and receive one in H_1 . Suppose that the similar scheme as ideal Poisson channel is applied for this modified channel; that is, looking at quantizer's output we keep sending input zero or A for H_0 and H_1 till a one appears. Formally stopping time here is defined as

$$T \triangleq \inf \{n : z_n = 1\}$$

Although T is still geometrically distributed (as before) and we can easily calculate $E[T]$, the result is a complicated expression. Letting the probability of event $\{z_i = 1\}$ conditioned on H_0 be denoted by p_e , which is error probability, we can write

$$\begin{aligned} p_e &= \Pr(y_i > \frac{A+\alpha}{\gamma} | H_0) \\ &= e^{-\lambda_0} \sum_{i=\beta}^{\infty} \frac{\lambda_0^i}{i!} \end{aligned}$$

where $\beta \triangleq \lceil \frac{A+\alpha}{\gamma} \rceil$. Then, we know that expected value of T conditioned on H_1 plays key role in energy consumption. If p_1 denotes the probability of getting one when H_1 is true, we can write

$$\begin{aligned} p_1 &= \Pr(y_i > \frac{A+\alpha}{\gamma} | H_1) \\ &= e^{-(A+\lambda_0)} \sum_{i=\beta}^{\infty} \frac{(A+\lambda_0)^i}{i!} \end{aligned}$$

then $E[T | H_1] = \frac{1}{p_1}$.

Obviously to get error probability tend to zero, α should be tending to infinity which yields $E[T | H_1]$ (and consequently energy) to blow up.

To get a good understanding of this scheme we resorted to simulation. During the simulation it turned out that if $\alpha > A$ then energy will increase rapidly but error probability decrease so smoothly hence we decided to keep α equal or smaller than A .

Let peak power constraint and α grow in integer order from 1 to 10, but bear in mind that in our simulation α is set always smaller than or equal peak power constraint, that is, if $A = 1$ then α is either 0 or 1. Figure 6.1 shows the energy and error probability for this scheme. Based on this figure we make sure that this scheme does not yield bounded energy with arbitrarily small error probability. Although error probability is mostly small (except when $\alpha = 0$), energy is increasing as peak power constraint increases, this is against what was proved in previous chapter. We know that if $A \rightarrow \infty$ then energy should tend to 1, but this behavior is not viewed in this simulation, hence we can make sure this scheme is not optimal. Note that in all following simulations, λ_0 is assumed to be 2. Obviously its value does not play important role (as long as it is not zero) on the energy consumption.

Briefly, there are two interesting remarks about this figure which are following:

- When peak power constraint is fixed, increasing and decreasing α cause error probability to decrease and increase, respectively and cause energy to increase and decrease respectively. For example, if $A = 10$ error probability is 2.37×10^{-4} and energy is 11.83 when $\alpha = 5$ and 1.36×10^{-6} and 18.57 when $\alpha = 10$.
- When α is fixed, energy is rather big when $\alpha = A$ then when A increases energy firstly decreases and when it achieves its minimum it starts again to increase. More interestingly, error probability monotonically decreases as A increases. For example if $\alpha = 4$, energy is minimum when $A = 8$ and after $A = 8$ energy increases while error probability keeps decreasing.

Another scheme which is provably more efficient from energy and also error probability point of view is as follows: instead of working with the

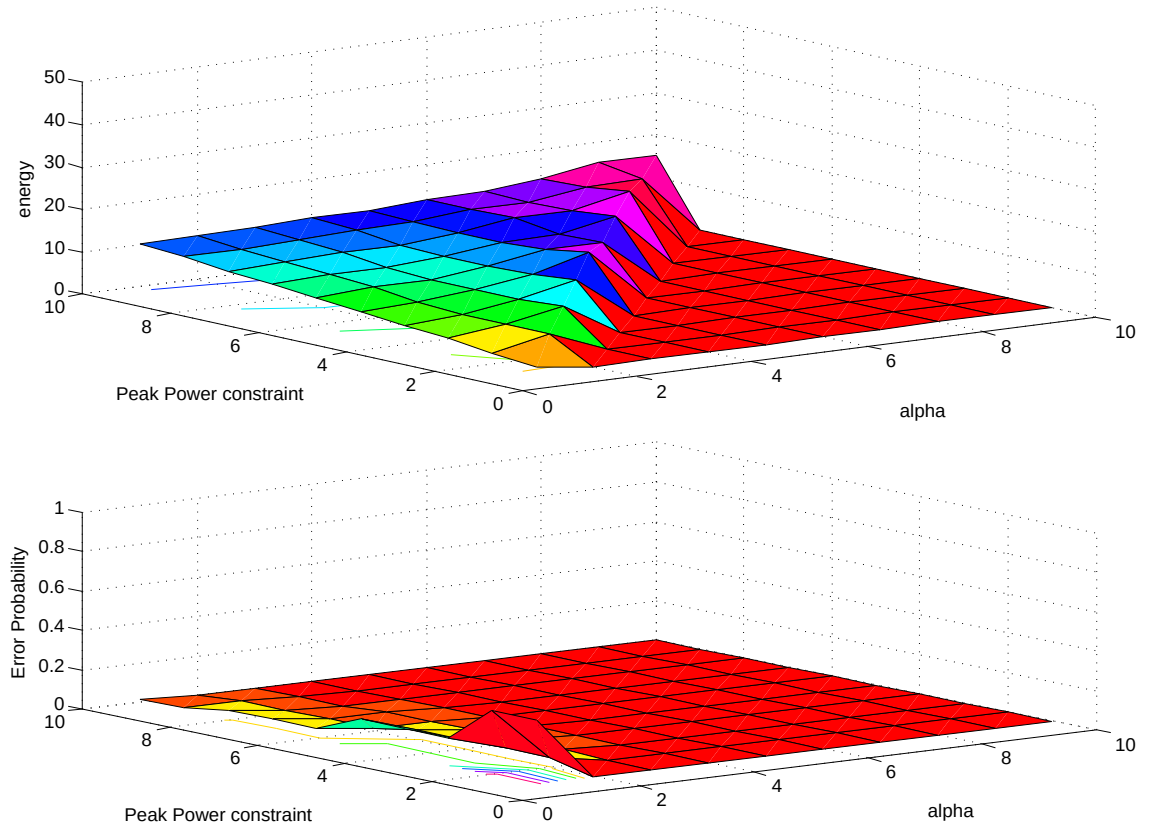


Figure 6.1: The energy consumption and error probability using the proposed coding scheme. Note that α should be always smaller than A .

likelihood ratio of only y_i at time i , the quantizer at time step i compares log likelihood ratio of all past history, $S(i) = \sum_{j=1}^i \ln \frac{\Pr(y_j|H_1)}{\Pr(y_j|H_0)}$, with a pre-defined threshold like α ; more precisely

$$z_i \triangleq \begin{cases} 1 & \sum_{j=1}^i \ln \frac{\Pr(y_j|H_1)}{\Pr(y_j|H_0)} > \alpha \\ 0 & \sum_{j=1}^i \ln \frac{\Pr(y_j|H_1)}{\Pr(y_j|H_0)} < -\alpha \end{cases}$$

or equivalently:

$$z_i \triangleq \begin{cases} 1 & \sum_{j=1}^i y_j > \frac{iA+\alpha}{\gamma} \\ 0 & \sum_{j=1}^i y_j < \frac{iA-\alpha}{\gamma} \end{cases}$$

Figure 6.2 shows the energy consumption and error probability incurred by the new coding scheme. Based on this figure it is interesting to point out that when α is sufficiently large and kept fixed (e.g., $\alpha > 6$), in most of cases minimum energy happens when $A = \alpha + 1$. For instance, if we set $\alpha = 8$, minimum energy belongs to $A = 9$, in this case (for $\alpha = 8$) error probability is always smaller than 10^{-6} .

6.2 Density-based Stopping Time

In previous section a one-bit quantizer was used as an attempt to roughly approximate a non-ideal Poisson channel as an ideal channel by looking at $z(i)$ as the channel output. Simulation showed that having arbitrarily small error probability causes infinite energy even when peak power constraint is big. This section concerns another technique to deal with energy in non-ideal Poisson channel. In the following, we use a similar notion as one explained in Chapter 3, theorem 3.2. The main idea behind this technique is the fact that the number of ones or zeros in the sequence of output's quantizer, z , is an important measure for checking whether H_1 is true or H_0 . Hence we define a stopping time based on number of zero and one in z . Let s_n and r_n denote the number of one and zero in quantizer's output, in other words, $s_n = z_1 + z_2 + \dots + z_n$ and $r_n = n - s_n$. We know that if

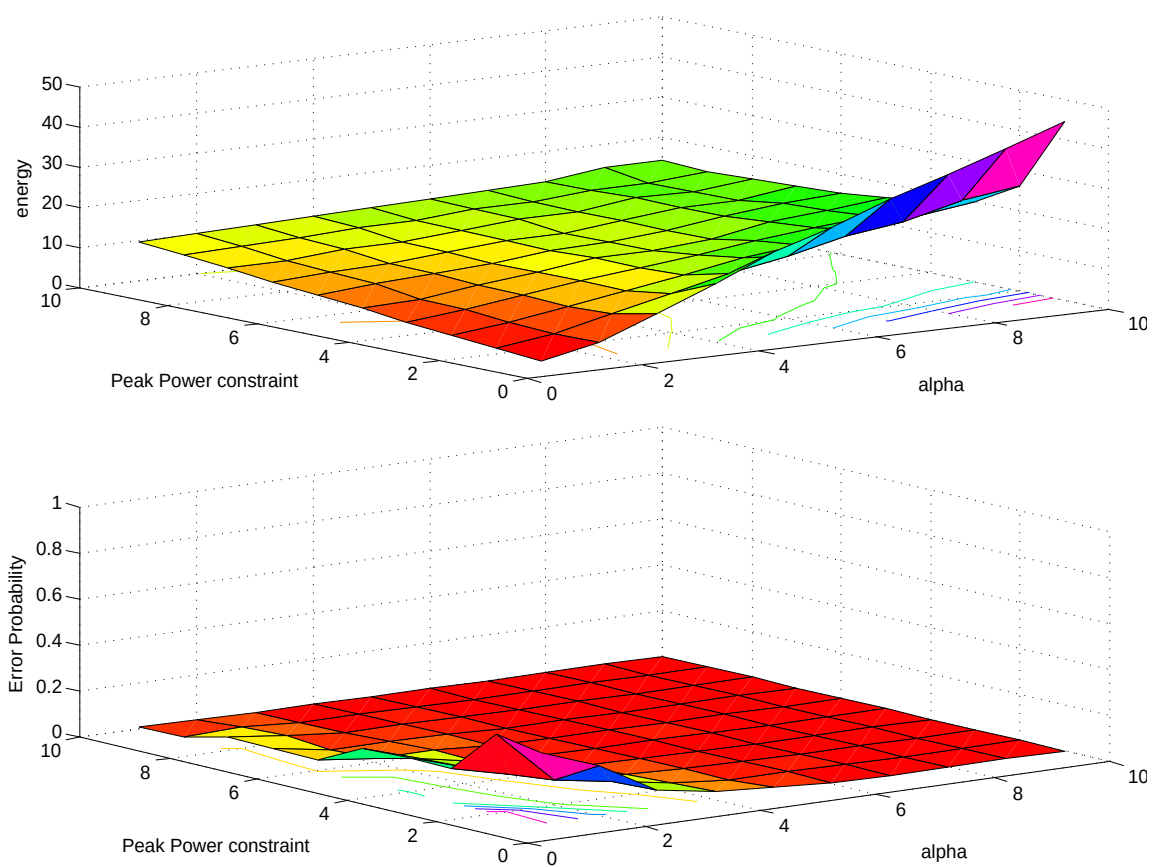


Figure 6.2: The energy consumption and error probability using the second proposed coding scheme.

more one in quantizer's output is received we can guess that H_1 is true and if the number of zeros is more we can declare H_0 . To stop the procedure, the following stopping time is defined:

$$T(c) \triangleq \inf\{n : s_n r_n \geq cn\}$$

for any $c > 0$. Stopping time is written as $T(c)$ to emphasize that T is a function of c . After the procedure is stopped, decoder declares H_0 if the number of zeros is more than ones and vice versa. Let p and $q = 1 - p$ denote the probability of getting one and zero, respectively, in quantizer's output, i.e.,

$$p \triangleq \Pr(z(i) = 1) = \Pr(y(i) > \alpha)$$

Note that in order for this technique to make sense α has to be around $A + \lambda_0$. Obviously to get error probability tend to zero, c should be sufficiently large. The following theorem clearly proves that this technique fails too to achieve finite energy in this problem.

Theorem 6.1: For T defined above

$$\lim_{c \rightarrow \infty} \frac{E[T(c)]}{c} = \frac{1}{pq}$$

To prove this theorem, the following lemma is required which is similar to Wald's identity (proved in Appendix B). Wald's identity showed that for a given i.i.d samples, X_i , with mean μ , $S_n = X_1 + \dots + X_n$ and event $\{\tau = n\}$ which is measurable on sequence $X^n = X_1, X_2, \dots, X_n$, $E[S_\tau] = \mu E[\tau]$. Hence in case $\mu = 0$, Wald's identity gives $E[S_\tau] = 0$ and is not informative about the value of $E[\tau]$. The following lemma is a related fact dealing with that case.

Lemma 6.2. (Wald's identity for second moment) Let $X_1, X_2, \dots$ be i.i.d with $E[X_1] = 0$ and $E[X_1^2] < \infty$. Suppose τ is a stopping time with $E[\tau] < \infty$, then $E[S_\tau^2] = E[X_1^2]E[\tau]$.

Proof of Lemma: Write $E[S_\tau^2] = T_1 + T_2$ where $T_1 \triangleq 2E[\sum_{1 \leq i < j \leq \tau} X_i X_j]$ and $T_2 \triangleq E[\sum_{1 \leq j \leq \tau} X_j^2]$. For terms of T_1 , for each i, j with $1 \leq i < j$,

we have

$$E[X_i X_j 1_{j \leq \tau}] = E[X_i (1 - 1_{\tau \leq j-1}) X_j] = E[X_i (1 - 1_{\tau \leq j-1})] E[X_j] = 0$$

The second quality is true because X_i and $1 - 1_{\tau \leq j-1}$ are both functions of $X_1, \dots, X_{j-1}$, hence so is their product, thus it is independent of X_j . It follows that $T_1 = 0$. Similarly for T_2 , for each j , $E[1_{\tau \geq j} X_j^2] = E[(1 - 1_{\tau \leq j-1}) X_j^2] = \Pr(\tau \geq j) E[X_1^2]$. Since provably $\sum_{j=1}^{\infty} \Pr(\tau \geq j) = E[\tau]$ we can conclude

$$E[S_\tau^2] = T_1 + T_2 = T_2 = E[\tau] E[X_1^2]$$

□

Now we can start the proof the theorem 6.1 which indicates that expected value of stopping time increases as c increases.

Proof of Theorem: To prove the theorem, we start from the following algebraic identity

$$n^{-1} s_n r_n = (q - p)(s_n - np) + npq - (s_n - np)^2 n^{-1}$$

which is immediately verified. Then due to the definition of stopping time, T , and the above algebraic identity together with obvious lower bound $T \geq 4c$,

$$\begin{aligned} c &\leq (q - p)(s_T - pT) + qpT < c + 1 + (s_T - pT)^2 T^{-1} \\ &\leq c + 1 + (s_T - pT)^2 / 4c \end{aligned}$$

If w_i denotes $z_i - p$, then $\tilde{s}_n \triangleq w_1 + w_2 + \dots + w_n = s_n - np$. Hence since $E[w_1] = 0$ and consequently $E[\tilde{s}_T] = 0$, we can use lemma 6.2 to get

$$E[\tilde{s}_T^2] = E[(s_T - pT)^2] = E[T] E[w_1^2] = pqE[T]$$

Therefore

$$c \leq pqE[T] < (c + 1) + pqE[T]/4c$$

and consequently

$$\frac{c}{pq} \leq pqE[T] < \frac{4c(c+1)}{pq(4c-1)}$$

which can clearly give the theorem 6.1.

□

theorem 6.1, similarly to theorem 3.2, has shown that the energy (which is linearly proportional to $E[T]$) will blow up if we force error probability to tend to zero. At the end of this chapter we have to, once again, underscore that capacity per unit cost analysis of non-ideal Poisson channel when peak power constraint is imposed has not been solved in this project. We will absolutely continue looking at this problem in future, although we could not get any promising results so far during this project.

Chapter 7

Concluding Remarks and Outlook

Gravitation is responsible for an apple falling on the ground but it is not responsible for people falling in love.

–Albert Einstein

The work presented here centers around the minimum energy for sending one single bit when noiseless feedback is present. Verdú proposed a general coding scheme for sending $k \rightarrow \infty$ bits whose energy per bit was proved to be minimum [8]. We started in this project to simplify his coding scheme by introducing the noiseless feedback. We dropped the unrealistic assumption of $k \rightarrow \infty$ and look for optimal coding scheme for sending *finite* number of bits. We therefore started working on optimal coding scheme for sending one single bit.

We started by sequential hypothesis testing to come up with the optimal coding scheme, but as proved in chapter 3, this technique fails to do so. That is, in the coding scheme based on sequential hypothesis testing driving error probability to zero would yield total energy to be infinity.

Then two classes of channels were considered: Gaussian and Poisson. A coding scheme for sending one single bit based on MMSE estimation was proposed for Gaussian channel and shown to be achieving

the capacity per unit cost and consequently minimum energy per bit. This coding scheme is very easy to implement and also to understand and although turned out to be special case of Turin's scheme, its optimality analysis is really enlightening and interesting for information-theoretists. It was also proved that minimum energy for sending reliably one single bit is exactly the same as minimum energy per bit when sending large number of bits (which was proposed in [8]).

In Poisson channel, two different cases were considered: finite and infinite capacity per unit cost. In ideal Poisson channel the capacity per unit cost is infinity whether peak-power constraint is imposed or not. Also, the capacity per unit cost for non-ideal Poisson channel is, too, infinity if the peak-power constraint is not imposed. Hence in both cases, the minimum energy to transmit one bit of information is expected to be zero. However, it is proved that in both cases, capacity per unit cost can not be achieved; in other words, the minimum energy for reliably sending one single bit is different from minimum energy per bit if one were sending large number of bits. We then computed the minimum energy for sending one bit and proposed an optimal coding scheme. We proved that in this case, sending reliably one single bit requires at least, on the average, half the energy of a photon. We then computed the minimum energy for sending finite number of bits, k and proposed an optimal coding scheme for it.

Comparison of the results for these two channels is as follows. In infinite-degree-of-freedom (infinite bandwidth) Gaussian channel, a single bit can be sent reliably with the same amount of energy that would be required per bit if one were sending a large number of bits whereas in Poisson channel the energy required for sending one single bit is totally different from the energy per bit when sending large number of bits.

The poisson channel whose capacity per unit cost is finite is dealt with in chapter 6. We proposed two different feedback coding schemes which perform good when error probability is bounded and not necessarily tending to zero. The first scheme is based on one-bit quantizer

after the channel to make the channel behave as ideal Poisson channel (Z-channel) and the second one is based on density-based stopping time. In this chapter, the least amount of energy for sending one single bit and the optimal feedback coding scheme are still open-problem.

Briefly the conclusions are listed as follows:

- The energy efficiency analysis of general feedback coding scheme based on sequential hypothesis testing
- An optimal coding scheme in Gaussian channel based on MMSE estimation
- Showing that in Gaussian channel, minimum energy for sending one single bit equals minimum energy per bit when sending $k \rightarrow \infty$ bits
- Calculation of minimum energy for sending one bit over Poisson channel whose capacity per unit cost is infinity
- An optimal coding scheme for above Poisson channel
- Showing that in above Poisson channel, minimum energy for sending one single bit totally differs minimum energy per bit when sending $k \rightarrow \infty$ bits
- An optimal coding scheme for sending *finite* number of bits
- Two feedback coding schemes which performs good when error probability is bounded, one based on one-bit quantizer and one based on time-varying stopping time.

Future work

As an future work, the feasibility of problem of distinguishing of two hypotheses H_0 and H_1 given Poisson channel output (with $\lambda_0 > 0$ and peak power constraint) should be analyzed. That is, can two hypotheses H_1 and H_0 be distinguishable over non-ideal Poisson channel with

peak power constraint?. We at this moment are really doubtful and pessimistic about that. The main question which is highly related to our problem is, how can one distinguish between two sequences of random variable $\mathbf{X}$ and $\mathbf{X} + \mathbf{a}$ with limited sequence $\mathbf{a}$ ($\mathbf{a} = \{a_1, a_2, \dots\}$). This is because, if $N_\lambda = \mathcal{P}_{(\lambda_0)}$ and $Y_0 = \mathcal{P}_{(X)}$, then output of non-ideal Poisson channel is a version of $Y_\lambda = N_\lambda + Y_0$. We believe that to make two sequences of random variable distinguishable, $\sum_n a_n^2 = \infty$, this is based on our intuition, without any mathematical proof. This needs to be proved or disproved in future work as an start for solving the problem for non-ideal Poisson channel with peak power constraint.

Appendix A

Improved PPM coding scheme with Feedback

In the PPM scheme proposed by Verdu in [8], a big matrix (which plays the same role as codeword) shall be sent to have the receiver perform M suboptimal hypothesis testing. In this appendix we want to make use of feedback to reduce the complexity of this scheme.

In the new PPM scheme in the presence of feedback, one column is sent each time and then transmitter decides what to do in the next step based on what feedback fed back. An $M \times 1$ is sent in first time and all these M received values, $\tilde{r}_i, i = 1, 2, \dots, M$ go through M *Wald sequential* hypothesis tests of which one of following three possible output appear

$$\hat{r}_i = \begin{cases} 0 & \tilde{r}_i < a \\ x_0 & \tilde{r}_i > b \\ \text{further sample} & a \leq \tilde{r}_i \leq b \end{cases}$$

If either 0 or x_0 is detected for the i th component in block, transmitter does not have to retransmit it again in next step but if $\tilde{r}_i$ lays in the ambiguity area, between a and b , then transmitter is informed by feedback about the position of undecided component in block and required to transmit only these component instead of sending whole $M \times 1$ block. Hence the next block to be transmitted is of size $(M - \xi) \times 1$ where ξ represents number of decided components in

previous step. This scheme should go on until receiver observes either 0 or x_0 . Since it is known that the Wald sequential hypothesis test gives the minimum expected blocklength compared to any other (sequential or non-sequential) hypothesis test having the same first and second error probability, we can make sure that this scheme is optimum from blocklength when comparing to Verdu's scheme and other hypothesis tests. It is well-known that in many cases Wald sequential test yields 50% improvement, on average, in blocklength over the regular Neyman Pearson hypothesis testing (c.f. for example [10], table 1). In this scheme a and b have to be chosen in a way that error probabilities satisfy the application need. There are some well-known approximations for first and second error probability and also for expected blocklength in [28] in terms of a and b . A big restriction of this technique which hinders its use for our problem is i.i.d assumption, all samples (for sending 0 or x_0) shall be independent and identically distributed, but most of feedback coding strategies deal with sending something new in next step which is made by feedback output. However, i.i.d assumption does not seem to be a good match for many standard data sets. Recently Gastpar and *et. al* proposed a technique for information estimate by which the i.i.d assumption can be dropped [35].

Appendix B

Another Proof for Theorem 5.3

The more straightforward proof for theorem 5.3 is given in this appendix which was proposed by Prof. A. Nikeghbali.

As explained in (5.27) one can write

$$E[f(T)] \approx \sum_{j=0}^{\infty} E[1_{\{j\Delta \leq T \leq (j+1)\Delta\}}] f(j\Delta) \quad (\text{B.1})$$

now we have to compute the expectation $E[1_{\{j\Delta \leq T \leq (j+1)\Delta\}}]$. It is easy to check out $1_{\{j\Delta \leq T \leq (j+1)\Delta\}} = 1_{\{j\Delta \leq T\}} - 1_{\{T \geq (j+1)\Delta\}}$, hence:

$$\begin{aligned} E[1_{\{j\Delta \leq T \leq (j+1)\Delta\}}] &= E[1_{\{j\Delta \leq T\}} - 1_{\{T \geq (j+1)\Delta\}}] \\ &= E[1_{\{j\Delta \leq T\}}] - E[1_{\{T \geq (j+1)\Delta\}}] \\ &= e^{-kj\Delta} - e^{-k(j+1)\Delta} \\ &= e^{-kj\Delta}(1 - e^{-k\Delta}) \\ &\approx k\Delta e^{-kj\Delta} \end{aligned} \quad (\text{B.2})$$

Plugging (B.2) into (B.1) yields:

$$\begin{aligned} E[f(T)] &\approx \sum_{j=0}^{\infty} k\Delta e^{-kj\Delta} f(j\Delta) \\ &\approx \int_0^{\infty} k e^{-kv} f(v) dv \end{aligned}$$

which is the same as what proved in the original proof of Theorem 5.3.

Appendix C

Publication

Bibliography

- [1] C. E. Shannon. Communication in the presence of noise. *Proc. IRE*, 37:10–21, Jan. 1949.
- [2] S. Verdú. Spectral efficiency in the wideband regime. *IEEE Transaction on Information Theory*, 48, No. 6:1319–143, Jun. 2002.
- [3] D. Buckingham and M. C. Valenti. The information-outage probability of finite length codes over awgn channels. *Proc. Conf. on Info. and Sys. (CISS)*, pages 687–695, Princeton, March 2008.
- [4] D. Divsalar S. Dolinar and F. Pollara. Code performance as a function of block size. Technical report, JPL TDA Progress report, 42-133, Jep Propulsion Laboratory, Pasadena, CA,, 1998.
- [5] Y. Polyanskiy, V. Poor, and S. Verdú. Channel coding rate in the finite blocklength regime. *IEEE Trans. Inform. Theory*, 56, no.5:2307–2359, May 2009.
- [6] Y. Polyanskiy, V. Poor, and S. Verdú. Dispersion of gaussian channels. *Proc. 2009 IEEE Int.Symp. on Information Theory (ISIT)*, pages 2204–2208, Seoul, Korea, July 2010.
- [7] Y. Polyanskiy, V. Poor, and S. Verdú. Minimum energy to send k bits with and without feedback. *Proc. 2010 IEEE Int.Symp. on Information Theory (ISIT)*, pages 221–225, TX, USA, 2010.
- [8] S. Verdú. On channel capacity per unit cost. *IEEE Transaction on Information Theory*, 36, No. 5:1019–1029, 1990.
- [9] M. J. E. Golay. Note on the theoretical efficiency of information in terms of variations. *IRE*, 37:1031, Sep. 1949.

- [10] A. Wald. Sequential tests of statistical hypotheses. *Ann. Math. Stat.*, 16, No. 2:117–186, June 1945.
- [11] P. Elias. Channel capacity without coding. Technical report, Quarterly progress report, MIT research Laboratory of Electronics, Oct. 1956. also in *Lectures on Communication Systems Theory*, E. Baghdady, Ed., New York: McGraw Hill, 1961.
- [12] J.P.M. Schalkwijk and T. Kailath. A coding scheme for additive noise channels with feedback-part i: No bandwidth constraint. *IEEE Transaction on Information Theory*, IT 2, No. 2:172–182, April 1966.
- [13] R. G. Gallager and B. Nakiboğlu. Variations on a theme by schalkwijk and kailath. *IEEE Transaction on Information Theory*, 56, No. 1:6–17, 2010.
- [14] G. L. Turin. Signal design for sequential detection systems with feedback. *IEEE Transaction on Information Theory*, 11, No. 3:401–408, july 1965.
- [15] G. L. Turin. More on uncertainty feedback: the bandlimited case. *IEEE Transaction on Information Theory*, 14, No. 2:1019–1029, March 1968.
- [16] A. D. Wyner. Capacity and error exponent for the direct detection photon channel, part i. *IEEE Transaction on Information Theory*, 34, No. 6:1449–1461, Nov. 1988.
- [17] M. R. Frey. Information capacity of poisson channel. *IEEE Transaction on Information Theory*, 37, No. 3:244–256, March. 1991.
- [18] A. Lapidoth. On the reliability function of the ideal poisson channel with noiseless feedback. *IEEE Transaction on Information Theory*, 39, No. 2:491–503, March 1993.
- [19] R. J. McEliece. *The Theory of Information and Coding: A Mathematical framework for Communication*. MA, Addison-Wesley, 1977.
- [20] R. G. Gallager. *Information Theory and Reliable Communication*. John Wiley, 1968.

- [21] R. E. Blahut. *Principles and Practice of Information Theory*. MA: Addison-Wisely, 1987.
- [22] A. Sahai. Anytime coding on the infinite bandwidth awgn channel: A sequential semi-orthogonal optimal code. Accessible: <http://arxiv.org/abs/cs/0610151>, 2006.
- [23] J.P.M. Schalkwijk. Coding for additive noise channels with feedback-part ii:band-limited channels. *IEEE Transaction on Information Theory*, IT 2, No. 2:183–191, April 1966.
- [24] A. Kramer. Improving communication reliability by use of an intermittent feedback channel. *IEEE Transaction on Information Theory*, 15, No. 1:52–60, Jan. 1969.
- [25] A. D. Wyner. On the schalkwijk-kailath coding scheme with a peak energy constraint. *IEEE Transaction on Information Theory*, 14, No. 1:129–134, Jan. 1968.
- [26] J. P. M. Schalkwijk and M. E. Barron. Sequential signaling under a peak power constraint. *IEEE Transaction on Information Theory*, 17, No. 3:278–282, March 1971.
- [27] Y. H. Kim, A. Lapidoth, and T. Weissman. The gaussian channel with noisy feedback. *Proc. 2007 IEEE Int.Symp. on Information Theory (ISIT)*, pages 1416–1420, June 2007.
- [28] D. Siegmund. *Sequential Analysis Tests and Confidence Intervals*. Springer-Verlag, 1985.
- [29] K. S. Zigangirov. Upper bound for the error probability for channels with feedback. *Problemy Peredaci Informatsii*, 6, No. 2:87–92, 1972.
- [30] M. H. A. Davis. Capacity and cutoff rate for poisson-type channels. *IEEE Transaction on Information Theory*, IT26:710–715, Nov. 1980.
- [31] P. Brémaud. *Point Process and Queues: Martingale Dynamics*. Springer-Verlag, New York, 1981.

- [32] E. L. Lehmann and G. Casella. *Theory of Point Estimation*. Springer, 1998.
- [33] A. D. Wyner. Capacity and error exponent for the direct detection photon channel, part ii. *IEEE Transaction on Information Theory*, 34, No. 6:1461–1471, Nov. 1988.
- [34] T. Koch and A. Lapidoth. Increased capacity per unit cost by oversampling. *To be published in Proc. 26th Convention of Electrical and Electronics Engineers in Israel*, November 17-20, 2010.
- [35] M. Gastpar, C. Gill, R. Patrick, A. G. Huth, and F. E. Theunissen. Anthropic correction of information estimates and its application to neural coding. *IEEE Transaction on Information Theory*, 56, NO. 2:890–900, Feb. 2010.