# Neurobiophysical Communication

Aman Chawla

# Contents

# Chapter 1

# Introduction

In this dissertation, we examine the node of Ranvier and electrodynamics near the node. In this context we recall the statement by Paul A. M. Dirac who said, in Lectures on Quantum Mechanics, that, "...one would like to take into account the possibility that Maxwell's equations are not accurately valid. When one goes to distances very close to the charges that are producing the fields, one may have to modify Maxwell's field theory so as to make it into a non-linear electrodynamics." Thus we focus on the quantum theory of near field electrodynamics and apply it to the vicinity of the node of Ranvier. This node is then modelled in MATLAB. Several such nodes combine with internodes to form an axon and several such axons combine to form nerve bundles. Finally, we simulate the overall ephaptic-dynamical properties of a nerve bundle. We begin our modeling at the ion channel level.

Then, in the theory section (Chapters 3 through 6), we will present results on the functional significance of the ephaptic interaction, the performance of the neuronal-Poisson channel and the error exponent of the additive White Gaussian noise channel.

# Chapter 3

# Ephaptic Communication

## 3.1 Information Theoretic Analysis

### 3.1.1 Introduction

The ephaptic interaction between axons is an interesting systemic phenomenon. In ephaptic coupling, action potentials (impulses) on adjacent axons get coupled to each other such that their velocities of propagation become synchronized at zero (primarily, but also other possible stable points) relative separation between the impulses. The functional role of such synchronization has remained somewhat of a mystery. To quote Scott [39]

> . . . it is interesting to consider whether such groups of [synchronized] impulses might play functional roles on the fiber bundles of motor nerves, sensory nerves, . . .

In this chapter we are able to present a very simple information theoretic analysis which shows that if independent stimuli are presented to two adjacent ephaptically interacting axons with differing conduction velocities, then the output at the ends of these axons is precisely the stimulus with the greater entropy or *surprise*.

We rely on the concept of temporal coding of information in a neuron wherein information is encoded in the precise time at which an impulse is received at the output of an axon. Note that in this paper we will use entropy interchangeably with surprise-value or surprise.

### 3.1.2 Prior work

The most recent work on ephaptic coupling is [1] where they state that increasing the applied extracellular field increased the phase locking of spikes to the field in a group of 25 neurons. The analysis in this paper is for 2 neurons, but the principles are easily generalized to any number of ephaptically coupled neurons. The conclusion that ephaptic coupling is a mechanism for selecting and conveying to the downstream destination neurons a random variable with the largest of the input random variable surprise-values is unchanged. Other relevant works include [29, 36, 2].

### 3.1.3 Entropic Analysis of Ephaptic Coupling

Consider an impulse traveling down an axon. The time of arrival of the impulse at the end or terminal of the axon is a random variable which takes on values in a continuous but bounded subset of the real numbers. The range of this set depends on the length of the axon, the conduction velocity and the range of possible impulse onset times. Thus given impulses traveling on two adjacent axons, we view them as carrying information in the precise time at which they will arrive at the ends of their individual axonal paths. If these were two electrically separated (ephaptically non-interacting) axons, we obtain two random variables, call them $X$ and $Y$ at their outputs. Here the value of $X$ is the time of arrival at the first axon's output and the value of $Y$ is that for the second. Thus, $X \in [t_1, t_2]$ and $Y \in [t_3, t_4]$.

Now let us introduce the element of ephaptic interaction between the two axons. As elucidated in Scott [39], Chapter 8, regardless of the underlying model, impulses traveling with differing velocities on adjacent coupled fibers undergo a process of change in velocities where the slower one catches up and the faster one slows down until they travel together at exactly the same velocity, with 0 or other relative spacing between the impulses.

If we analyze this in terms of what is happening to the temporally coded information carried by the two pulses, we immediately see that the information is being modified since the potential time of arrival is changing due to the ephaptic interaction. How exactly is it being modified? Well, that is simple. Since $time = \frac{distance}{speed}$, the time of arrival of the slower pulse is getting earlier, while that of the faster pulse is getting later until they equate at the point $t_{sync} = \frac{d}{w_1 v_1 + w_2 v_2}$ where $d$ is the distance they travel 'together'. Assume that the impulses synchronize fairly rapidly so that the majority of the time that they travel, they travel together and $d$ is approximately the length of the axon.

Suppose $v_1 > v_2$ and $t_{sync} \in [t_1 = \frac{d}{v_1}, t_2 = \frac{d}{v_2}]$. Then $t_{sync}$ can be written as $t_{sync} = \lambda_1 t_1 + (1 - \lambda_1) t_2$ since $[t_1, t_2]$ is a convex set. Here $\lambda_1$ is a constant that depends on $d, v_1, v_2, w_1, w_2 = 1 - w_1$. Furthermore, redundant information is being created since the times of arrival of synchronized pulses on the two coupled axons' ends are the same. So a natural question is: what could be the purpose of synchronization and redundancy creation?

We will come back to this question. For the moment consider two random variables $X$ and $Y$, where $X$ is independent of $Y$. It can be shown [11] that if $Z = X + Y$ then $H(Z) \geqslant H(X)$ and $H(Z) \geqslant H(Y)$ where $H$ stands for the Shannon entropy. Consequently, $H(Z) \geqslant \max\{H(X), H(Y)\}$. The equality holds if $X = f(Z)$ and $Y = g(Z)$ where $f, g$ are two deterministic functions.

Now if $X$ represents the time of arrival of the first pulse and $Y$ represents that of the second pulse were they not ephaptically coupled, then we can create a new random variable $Z$ which will represent the time of arrival of the coupled pulses. Then, following the representation of $t_{sync}$ derived earlier,

$$Z = \lambda_1 X + \lambda_2 Y \tag{3.1}$$

where $\lambda_2 = 1 - \lambda_1$. Define new random variables $X' = \lambda_1 X$ and $Y' = \lambda_2 Y$. Then they are still independent since $\lambda_1$ is a constant. And so

$$Z = X' + Y' \tag{3.2}$$

Note that still, $H(X') = H(X)$ and $H(Y') = H(Y)$. By the result from [11] Exercise 2.14 alluded to above, we can state that

$$H(Z) \geqslant \max\{H(X'), H(Y')\} \tag{3.3}$$

This equation is the central equation of this paper from which we will draw all our conclusions. Here $Z$ represents the synchronized time of arrival random variable.

### 3.1.4 Consequences

The main consequence of equation (3.3) is that at the output of the two axons is delivered a new random variable $Z$ whose surprise or entropy is greater than or equal to the larger of the two surprises that would be conveyed by the ephaptically non-interacting axons. Thus ephaptic coupling acts as an entropy *amplifier* or at the very least, a greater entropy *selector*. It delivers at the destinations copies of a new random variable ($Z$) whose surprise-value is at least as large as the larger of the two surprise-values that were input into the system.

Coming back to the question we asked, it is now clear that synchronization and redundancy creation are serving to magnify (convey to more downstream destinations) the greater of the two surprises that were input into the system. This has relevance to the mechanism by which an organism *pays* more attention to the more information bearing of two stimuli presented to it. Even higher up in the cortical organization, the organism will *pay* more attention to the more information bearing of two choices. While the random variable at the ouptut is neither of the input random variables, its surprise value is characterized as above.

This indicates that it is the surprise value of stimulus presented to a neuron that is fundamental in what *matters* to the neuron, rather than the precise random variable that bears that surprise. Thus ephaptic coupling has illuminated for us the mechanism of all choosing or selecting activity in the mammalian nervous system, where one stimulus (or event or memory or image) is preferred over another. It suggests that we simply choose that which is conveying more information to us and store it up (or further process it) in a representation such as $Z$ above.

Now it is a central result in information theory that if the entropy of a source is less than the capacity of a channel over which it is intended to be communicated, then it can be communicated effectively. Thus if the larger of the two surprise values is lower than the capacity of the downstream channel over which it is to be transmitted, then that channel can accomodate this information. Organisms with larger capacity nervous systems are therefore able to exploit the presence of this entropy selector to choose higher-entropy stimuli and store, internally transport and process them.

# Chapter 4

# Burnashev's Converse for the AWGN channel

Information theory involves the study of channels and their capacities and reliabilities with and without feedback. The most widely studied channel is the Additive White Gaussian Noise (AWGN) channel. Here we consider a continuous time AWGN channel with perfect feedback. This channel is connected to Brownian motion via a stochastic equation [21]. We consider data transmission over this channel. The decoder makes the decisions as to when to stop decoding. This is known as the random transmission time setting introduced first in 1976 [5]. The codes that arise in this setting are variable length codes in continuous time. For these codes, we follow [4] in proving a sequence of results that converge in an upper bound to the expected decoding time for communication over this channel.

The continuous time AWGN channel is discussed in [16]. The same channel with feedback was first considered in [37] and later studied in [38]. Burnashev [5] studied the reliability function of a discrete memoryless channel with perfect feedback. The reliability function of a continuous time infinite bandwidth AWGN channel with noisy feedback was studied in [9]. That work can be made systemic if we consider the noise in the forward and feedback links to be correlated instead of independent. However, this will not be pursued in this thesis due to time limitations. In this chapter we consider the Burnashev converse, but a simplified version [4], and in continuous time for the AWGN channel.

The main assumptions on the channel are peak and average power constraints and right-continuity of sample paths of the input. Another pervading assumption is that the filtration to which the ouptut is adapted satisfies the usual conditions, defined in the text. The Encoder must also satisfy a certain strong condition called *Nack-continuity*, defined in the text. Finally, a technical condition on the output measure (see footnote 1) must hold.

This paper is structured as follows. In section 1 we present a review of the probability concepts and results that will be used in the paper and also introduce the notation. In section 2 we review the channel model being studied. In section 3 we begin proving a sequence of results that lead to the converse over this channel. Finally we conclude in section 4.

## 4.1 Preliminaries and Background

We first need a bit of notation and probability background that will set the stage for studying the continuous time AWGN channel with perfect feedback.

### 4.1.1 Notation and Probability Review

Let $\Omega$ be a set, called the sample space. Let $\mathcal{F}$ be a set of some subsets of $\Omega$, the measurable subsets. $\mathcal{F}$ is called the sigma algebra. Now we define a map $\mathbb{P} : \Omega \to [0, 1]$, called the probability measure which assigns weights or probabilities to elements of $\mathcal{F}$. This triple $(\Omega, \mathcal{F}, \mathbb{P})$ is called a probability triple or a probability space. It is the fundamental object in probability theory. Associated with a probability space, we have random variables, which are measurable maps [8] from the sample space $\Omega$ to the real line $\mathbb{R}$.

For the study of the continuous time AWGN channel we will need to be able to study continuous time stochastic processes. A definition is in order [41].

**Definition** (Continuous Time Stochastic Process) A *continuous time stochastic process* $\{x(t) : t \in \mathbb{T}\}$ is a collection $x(t) = x(t, \omega)$ of random variables defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ indexed by $t \in \mathbb{T}$, where $\mathbb{T}$ is a finite or infinite interval in $\mathbb{R}$, and $\omega \in \Omega$.

The central issue with continuous time stochastic processes as defined above, is that certain events may not be measurable because $\mathbb{T}$ is uncountable. To get around this problem, one puts the probability measure directly on the space of functions that represent the realization of the stochastic process, instead of on the sample space associated with each $x(t)$ .

With this definition of stochastic processes in continuous time, we are now in a position to briefly present martingales in continuous time. This will lead to the all-important *Doob's optional stopping time theorem* for continuous time martingales. Again, a few definitions are in order, following [41].

**Definition** (Filtration) Let $(\Omega, \mathcal{F})$ be a space with a $\sigma-$field. A family $\{\mathcal{F}_t : t \geqslant 0\}$ of sub-$\sigma$-fields of $\mathcal{F}$ is called a *filtration* if $\mathcal{F}_s \subset \mathcal{F}_t$ if $s \leqslant t$.

If $\Omega$ is the space of functions discussed in the preceding paragraph, defined on $\mathbb{T} \subset \mathbb{R}^+$, it comes with a natural filtration $\mathcal{F}_t = \sigma\{x(s) : s \leqslant t\}$, where $\sigma$ stands for the 'smallest $\sigma-$field generated by'.

**Definition** (Martingale) Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a filtration $\mathcal{F}_t \subset \mathcal{F}$ indexed by $t \geqslant 0$, a family $x(t) = x(t, \omega)$ of random variables is called a *martingale relative to* $(\Omega, \mathcal{F}, \mathbb{P})$ if the following are true.

- For almost all $\omega$, $x(t, \omega)$ has left and right limits at every $t$ and is continuous from the right, i.e., $x(t, \omega) = x(t + 0, \omega)$.
- For each $t \geqslant 0$, $x(t)$ is $\mathcal{F}_t$ measurable and integrable.
- For $0 \leqslant s \leqslant t$,

$$E[x(t)|\mathcal{F}_s] = x(s) \text{ a.e.} \tag{4.1}$$

If the process $x(t)$ satisfies

$$E[x(t)|\mathcal{F}_s] \geqslant x(s) \text{ a.e.} \tag{4.2}$$

then it is called a *submartingale* and if the above inequality is reversed it is called a *supermartingale*.

Next we define stopping times in continuous time and their associated $\sigma-$fields.

**Definition** (Stopping Time) A *stopping time* for the filtration $\{\mathcal{F}_t\}$ is a random variable $\tau(\omega) : \Omega \to [0, +\infty]$ that satisfies for every $t \geqslant 0$,

$$E_t = \{\omega : \tau(\omega) \leqslant t\} \in \mathcal{F}_t \tag{4.3}$$

**Definition** (Corresponding $\sigma-$field) Given a stopping time $\tau$, the corresponding $\sigma-$field $\mathcal{F}_t$ is defined as consisting of those sets $A \in \mathcal{F}$ that satisfy for each $t \geqslant 0$,

$$A \cap \{\omega : \tau(\omega) \leqslant t\} \in \mathcal{F}_t \tag{4.4}$$

To study mutual information in the white Gaussian setting we will need another definition and related theorem.

**Definition** (Absolute continuity) Let $\mu$ and $\nu$ be probability measures on a measurable space $(\mathbb{X}, \mathcal{B}(\mathbb{X}))$. We say that $\mu$ is *absolutely continuous* with respect to $\nu$, i.e., $\mu \prec \nu$, if $\mu(A) = 0$ for every $A \in \mathcal{B}(\mathbb{X})$ such that $\nu(A) = 0$.

**Theorem 4.1.1** *(Radon-Nikodym) The* Radon-Nikodym *theorem states that, $\mu \prec \nu \Rightarrow \exists$ a $\nu-$integrable function $\phi(x)$ such that*

$$\mu(A) = \int_A \phi(x) d\nu(x), \forall A \in \mathcal{B}(\mathbb{X}). \tag{4.5}$$

*The function $\phi(x)$ is called the Radon-Nikodym derivative and is written in the form*

$$\phi(x) = \frac{d\mu}{d\nu}(x). \tag{4.6}$$

### 4.1.2 Some more background

Here we state a few useful results from Karatzas and Shreve [28].

Given a stochastic process, the simplest choice of a filtration is that generated by the process itself, i.e.,

$$\mathcal{F}_t^X = \sigma(X_s; 0 \leqslant s \leqslant t),$$

the smallest $\sigma-$field with respect to which $X_s$ is measurable for every $s \in [0, t]$. We interpret $A \in \mathcal{F}_t^X$ to mean that by time $t$, an observer of $X$ knows whether or not $A$ has occured.

Let $\{\mathcal{F}_t; t \geqslant 0\}$ be a filtration. We define

$$\mathcal{F}_{t^-} \triangleq \sigma(\bigcup_{s<t} \mathcal{F}_s)$$

to be the $\sigma-$field of events strictly prior to $t > 0$ and

$$\mathcal{F}_{t^+} \triangleq \bigcap_{\epsilon>0} \mathcal{F}_{t+\epsilon}$$

to be the $\sigma-$field of events immediately after $t \geqslant 0$. We decree

$$\mathcal{F}_{0-} \triangleq \mathcal{F}_0$$

and say that the filtration $\{\mathcal{F}_t\}$ is right- (left-) continuous if $\mathcal{F}_t = \mathcal{F}_{t^+}$ (respectively $\mathcal{F}_t = \mathcal{F}_{t^-}$) holds for every $t \geqslant 0$. Similarly, we call a sample path *RCLL* if it is right-continuous on $[0, \infty)$ with finite left-hand limits on $(0, \infty)$. A stochastic process $X$ is *adapted* to the filtration $\{\mathcal{F}_t\}$ if, for each $t \geqslant 0$, $X_t$ is an $\mathcal{F}_t$-measurable random variable. A useful definition:

**Definition** (Usual Conditions) A filtration $\{\mathcal{F}_t\}$ is said to satisfy the *usual conditions* if it is right-continuous and $\mathcal{F}_0$ contains all the $P-$negligible events in $\mathcal{F}$.

Now we are in a position to state *Doob's optional sampling theorem* for continuous time martingales.

**Theorem 4.1.2** *(Optional Sampling Theorem) Let $\{X_t, \mathcal{F}_t; 0 \leqslant t < \infty\}$ be a right-continuous sub-martingale with a last element $X_\infty$, and let $S \leqslant T$ be two optional times of the filtration $\{\mathcal{F}_t\}$. We have*

$$E(X_T|\mathcal{F}_{s^+}) \geqslant X_s \ a.s. \ \mathbb{P}. \tag{4.7}$$

*If $S$ is a stopping time, then $\mathcal{F}_s$ can replace $\mathcal{F}_{s^+}$ above. In particular, $EX_T \geqslant EX_0$, and for a martingale with a last element we have $EX_T = EX_0$.*

Finally, we need this theorem from [28].

**Theorem 4.1.3** *Let $X = \{X_t, \mathcal{F}_t; 0 \leqslant t < \infty\}$ be a submartingale, and assume the filtration $\{\mathcal{F}_t\}$ satisfies the usual conditions. Then the process $X$ has a right-continuous modification if and only if the function $t \mapsto EX_t$ from $[0, \infty)$ to $\mathbb{R}$ is right-continuous. If this right-continuous modification exists, it can be chosen so as to be RCLL and adapted to $\{\mathcal{F}_t\}$, hence a submartingale with respect to $\{\mathcal{F}_t\}$.*

For futher background material on stochastic processes, the reader is referred to [3] and for stochastic filtering to the development in [26, 43, 23].

## 4.2 The Channel

### 4.2.1 The Channel Model

In this section we define the continuous time AWGN channel with feedback and discuss its properties, following [21]. The channel input is decided by an *encoder*. Thus,

$$X_t = f_t(\xi, Y_0^{t-}) \tag{4.8}$$

where $\xi \in \Xi$ is the particular message being transmitted and $Y_0^{t-}$ represents the causal (ideal) feedback from the output of the channel to the encoder. At any instant $t$, the encoder could choose to input any length signal (from null to a finite duration signal) into the channel, depending on the feedback. This makes the channel input variable in length.

The *decoder* is a map $D(\cdot) : \mathcal{Z} \to \mathcal{W}$ where $\mathcal{Z}$ is the receiver observation space until the decoding time $T$. Since the channel inputs are variable in length, the output space also consists of variable length signals. In addition, the decoding time $T$ is a stopping time with respect to the receiver observation $Y^t$. The decision to stop decoding is made by the decoder based on the output observations so far. This adds another cause of variability in the length of the ouptut signals.

The forward *channel* is an AWGN channel in continuous time with perfect feedback. It adds a white Gaussian noise $\nu(t)$ to the input $x(t)$. The channel can be written as:

$$y(t) = x(t) + \nu(t), \ \ t \in \mathbb{T} \tag{4.9}$$

If we integrate (4.9) we get,

$$Y(t) = \int_0^t x(u)du + B(t), \ \ t \geqslant 0. \tag{4.10}$$

This equation can be considered as a mathematical model of a white Gaussian channel [21] where $B(t)$ is Brownian motion. In [21] the author also specifies that the channel must satisfy the following four conditions.

1. $\{\nu(t) : t \in \mathbb{T}\}$ is a Gaussian process.

2. The message $\xi$ is independent of the noize $\nu(t)$.

3. The channel input $x(t)$ satisfies (4.8).

4. The stochastic equation (4.9) has a unique solution $Y = \{Y(t) : t \in \mathbb{T}\}$.

### 4.2.2 The Power Constraints

In this subsection we introduce the power constraints that will be imposed on the channel. The first is an average power constraint which says that:

$$\frac{1}{T}\int_0^T E[|x(t)|^2]dt < P_{av} \tag{4.11}$$

for $P_{av} < \infty$.

The second is a peak power constraint which says that:

$$0 \leqslant |x(t)| \leqslant P_{peak} \tag{4.12}$$

where $P_{peak} < \infty$.

### 4.2.3 Probabilistic Framework for Analyzing the AWGN channel with Perfect Feedback

There are fundamentally two sources of randomness in a communication problem over a channel with additive noise and perfect feedback: (i) the randomness associated with which message is input into the system and (ii) the randomness associated with which noise stochastic process is input into the system.

Let us denote by $\mathcal{W}$ the sample space associated with the messages and by $\Omega$ the sample space associated with the noise stochastic process $\nu(t)$. Then the overall sample space for the experiment of using the channel is $\mathcal{W} \times \Omega$. The source of randomness is thus the message generated on $(\mathcal{W}, 2^{\mathcal{W}})$ and the noise stochastic process generated on $(\Omega, \mathcal{F})$. The combined space would be $(\mathcal{W} \times \Omega, 2^{\mathcal{W}} \times \mathcal{F}, \mathbb{P})$.

Having defined notation, background and the channel, we now present a converse for the reliability function of this channel in terms of a lower bound on the expected decoding time.

## 4.3 The Converse

### 4.3.1 Binary Hypothesis Testing with Feedback in Continuous Time

#### The first proposition

In this section we will prove the first proposition from [4], for the continuous time AWGN channel with perfect feedback.

**Proposition 4.3.1** *For any binary hypothesis testing scheme for a continuous time channel with feedback and right-continuous sample paths of the output,*

$$D(P_A||P_N) \leqslant C_1 E[T|A] \tag{4.13}$$

*where $T$ is the decision stopping time, $E[T] < \infty$, and $E[T|A]$ denotes the expectation of $T$ conditioned on hypothesis A. A requirement is that the probability distribution on the output signals given either hypothesis $A$ or $N$ must be mutually absolutely continuous and under the $N$ hypothesis, the output process must be continuous in probability.*

*We discuss the first of the two requirements. In the proof of Theorem 6.2.1 in [21], the author concludes that $\mu_{Y|\xi}$ is absolutely continuous with respect to $\mu_B$. This would hold for both $\xi = A$ and $\xi = N$. If we assume[1] that $\mu_B$ is absolutely continuous with respect to $\mu_{Y|\xi=N}$, then by transitivity of absolute continuity, $\mu_{Y|\xi=A} \prec \mu_{Y|\xi=N}$. The condition on the input signal for this to hold can be stated as*

$$\int_0^T E[|x(t)|^2|\xi]dt < \infty, \text{ with probability 1.} \tag{4.14}$$

*which is satisfied if the average power constraint (4.11) is met.*

The second of the two requirements is that under the $N$ hypothesis, the output process must be continuous in probability at every time $t \in \mathbb{T}$. This is the only strong assumption in this paper, going beyond right-continuity [2]. Formally it can be stated as the requirement that for every $\epsilon > 0$,

$$\lim_{s \to t} P_N(\{\omega \in \Omega || Y_s(\omega) - Y_t(\omega)| \geqslant \epsilon\}) = 0. \tag{4.15}$$

This imposes the condition that under the N hypothesis, the encoder must input a process that is essentially continuous in probability. Together with the continuity of Brownian motion, this would guarantee that the output process is also continuous in probability. We call this the *Nack-continuity* assumption on the Encoder.

**Proof** An important fact to keep in mind is that all the stochastic processes under consideration have right-continuous sample paths. This is because the Brownian process underlying the Gaussian noise process is a continuous process and the encoder is assumed to be such that the channel input process is right-continuous. This inturn guarantees right-continuity of the output.

Since the main issue is the move from a discrete set of time points to a continuum, the key fact that we will need is the following theorem by Striebel [40]. Other references for relevant notation and model set up include [25], [24] and [15].

---

[1] We don't believe that this is a strong assumption since absolute continuity is not an anti-symmetric relation.

[2] Right-continuity only guarantees that the output process is *right*-continuous in probability.

**Theorem 4.3.2** *(Striebel's Theorem) Let $\{\mathcal{X}(T), \mathcal{A}, P\}$ and $\{\mathcal{X}(T), \mathcal{A}, Q\}$ be two stochastic processes such that $P$ is absolutely continuous with respect to $Q$ over $\mathcal{A}$ and the $Q$ process is continuous in probability. Then for any set $D = (t_1, t_2, \cdots)$ dense in $T$, the derivatives $dP^\infty/dQ^\infty$ and $dP/dQ$ coincide a.s. $Q$.*

We move to the continuum case in two steps. We begin by recalling the discrete definition of $V_n$ from [4]:

$$V_n = \ln \frac{P_A(Y_1, \ldots, Y_n)}{P_N(Y_1, \ldots, Y_n)}. \tag{4.16}$$

The motivation for the above equation is a likelihood ratio test. We draw a parallel between the Radon-Nikodym derivative and the likelihood ratio based on [15]. Now let

$$V_D = \ln \frac{P_A(Y_{t_1}, Y_{t_2}, \ldots, Y_{t_\infty})}{P_N(Y_{t_1}, Y_{t_2}, \ldots, Y_{t_\infty})} \tag{4.17}$$

where $D = \{t_1, t_2, \ldots\}$ is dense in $\mathbb{T}$. Then $V_D$ coincides with

$$V_t = \ln \frac{P_A(Y_0^t)}{P_N(Y_0^t)} \tag{4.18}$$

as per Theorem (4.3.2) above. This equation holds with probability 1, as per the theorem. Now we rewrite $V_t$ as

$$V_t = \ln \frac{P_A(Y_{[0,t]})}{P_N(Y_{[0,t]})}. \tag{4.19}$$

We want

$$V_t = \int_0^t U_s ds \tag{4.20}$$

for some $U_s$. Consider,

$$\sum_D \ln \frac{P_A(Y_t|Y_{[0,t)})}{P_N(Y_t|Y_{[0,t)})} = \ln \frac{P_A(Y_{[0,t_\infty]})}{P_N(Y_{[0,t_\infty]})} \tag{4.21}$$

where $\sum_D$ is a countably infinite sum over $D$. This follows by Theorem 8.2 in Chapter 1 of [14]. But the right hand side of the above equation coincides with

$$\ln \frac{P_A(Y_{[0,t]})}{P_N(Y_{[0,t]})} = V_t \tag{4.22}$$

with probability 1, as per Theorem (4.3.2). So we let

$$U_s = \ln \frac{P_A(Y_s|Y_{[0,s)})}{P_N(Y_s|Y_{[0,s)})} \tag{4.23}$$

Then

$$\sum_D U_s = V_t. \tag{4.24}$$

Now consider

$$\begin{aligned} E[U_s|\text{Ack}, y_0^{s-}] &= E[\ln \frac{f_A(y_s|y_0^{s-})}{f_N(y_s|y_0^{s-})}|\text{Ack}, y_0^{s-}] \\ &= \int_{y_s \in \mathbb{R} = \mathcal{Y}} f_A(y_s|y_0^{s-}) \ln \frac{f_A(y_s|y_0^{s-})}{f_N(y_s|y_0^{s-})} dy_s \\ &= \int_{\mathbb{R}} f_A(y_s|A, y_0^{s-}) \ln \frac{f_A(y_s|A, y_0^{s-})}{f_N(y_s|N, y_0^{s-})} dy_s \\ &= \int_{\mathbb{R}} f_A(y_s|x_s^A = f_s(A, y_0^{s-})) \cdot \\ &\quad \ln \frac{f_A(y_s|x_s^A = f_s(A, y_0^{s-}))}{f_N(y_s|x_s^N = f_s(N, y_0^{s-}))} dy_s \end{aligned} \tag{4.25}$$

The channel is given by:

$$y(t) = x(t) + \nu(t). \tag{4.26}$$

and suppose $\nu(t)$ is distributed as $\mathcal{N}(0, 1)$. Then we get:

$$\begin{aligned} y_s^A &= x_s^A + \nu_s \\ y_s^N &= x_s^N + \nu_s \end{aligned} \tag{4.27}$$

where $s \in \mathbb{T}$ is the time subscript. Therefore $y_s^A$ is distributed as $\mathcal{N}(x_s^A, 1)$. Similarly, $y_s^N$ is distributed as $\mathcal{N}(x_s^N, 1)$. Thus,

$$f_A(y_s|x_s^A) = \frac{1}{\sqrt{2\pi}} e^{\frac{-(y_s - x_s^A)^2}{2}} \tag{4.28}$$

and similarly

$$f_N(y_s|x_s^N) = \frac{1}{\sqrt{2\pi}} e^{\frac{-(y_s - x_s^N)^2}{2}} \tag{4.29}$$

and so we get

$$\begin{aligned} E[U_s|\text{Ack}, y_0^{s-}] &= \int_{\mathbb{R}} f_A(y_s|x_s^A = f_s(A, y_0^{s-})) \cdot \\ &\quad \ln \frac{f_A(y_s|x_s^A = f_s(A, y_0^{s-}))}{f_N(y_s|x_s^N = f_s(N, y_0^{s-}))} dy_s \\ &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{\frac{-(y_s - x_s^A)^2}{2}} \ln \frac{e^{\frac{-(y_s - x_s^A)^2}{2}}}{e^{\frac{-(y_s - x_s^N)^2}{2}}} dy_s \end{aligned}$$

If we define $C_1$ for this channel as

$$C_1 = \max_{x,x' \in \mathbb{R}} \int_{\mathbb{R}} f(y|x) \ln \frac{f(y|x)}{f(y|x')} dy \tag{4.30}$$

and if $C_1 < \infty$ then

$$\begin{aligned} E[U_s|\text{Ack}, y_0^{s-}] &= \int_{\mathbb{R}} f_A(y_s|x_s^A = f_s(A, y_0^{s-})) \cdot \\ &\quad \ln \frac{f_A(y_s|x_s^A = f_s(A, y_0^{s-}))}{f_N(y_s|x_s^N = f_s(N, y_0^{s-}))} dy_s \\ &\leqslant C_1 \end{aligned}$$

Here $f_t(\cdot,\cdot)$ is the encoder function at time $t$. Consequently, by applying (4.24), $\{V_t - tC_1\}$ is a supermartingale under hypothesis $A$. We note that the requirement of finitude of $C_1$ imposes a peak power constraint on the channel inputs.

Now we will use Theorem (4.1.2) in its supermartingale form. The main condition to check is that $\{V_t - tC_1, \mathcal{F}_t; 0 \leqslant t < \infty\}$ is a right-continuous supermartingale. Since the output of the channel is right-continuous (by right continuity assumption on channel inputs; see also page 230 in [21]) and $V_t$ is determined by the output[3], the supermartingale under consideration is right-continuous and satisfies the Theorem (4.1.2) requirements. We can therefore apply Theorem (4.1.2) to conclude that $E[V_T - C_1 T|A] \leqslant E[V_0]$. But $V_0 = \ln \frac{f_A(Y_0)}{f_N(Y_0)}$. Now $f_A, f_N$ are positive functions that map $\mathbb{R} \mapsto [0,1]$ and ln is strictly increasing (monotonic) on the positive reals. Moreover, ln evaluates to a negative value on $(0,1)$. Therefore the expectation $E[V_0]$ is non-positive. Thus we conclude,

$$E[V_T - C_1 T|A] \leqslant 0. \tag{4.31}$$

This completes the proof of the first proposition. ∎

**The first lemma**

The first lemma from [4] goes through in the continuous time setting without any changes. This is because the definition of the sets $\mathcal{S}$ and $\bar{\mathcal{S}}$ does not change in the continuous time setting. We state the lemma below for convenience.

[3]As defined in (4.18), $V_t$ is a right continuous process. This is proved in Appendix 1.

**Lemma 4.3.3** *The error probability of a binary hypothesis test performed across a continuous time AWGN channel with feedback and variable-length codes is lower bounded by*

$$P_e \geqslant \frac{\min\{p_A, p_N\}}{4} e^{-C_1 E[T]} \tag{4.32}$$

*where $p_A$ and $p_N$ are the a priori probabilities for the hypotheses.*

### 4.3.2 Expected Decoding Time and Channel Capacity in Continuous Time

Let the output process $Y(t)$ be adapted to the filtration $\{\mathcal{F}_t\}$. Then by definition, for each $t \geqslant 0$, $Y_t$ is an $\mathcal{F}_t-$measurable random variable. Moreover, the filtration specifies all the available information at any given time. To be specific, we let $\mathcal{F}_t$ be the filtration generated by the output process. Thus

$$\mathcal{F}_t = \sigma(Y_s : 0 \leqslant s \leqslant t). \tag{4.33}$$

Similarly, we let $\mathcal{G}_t$ be the filtration generated by the channel input process $\{X(t), t \geqslant 0\}$. Thus

$$\mathcal{G}_t = \sigma(X_s : 0 \leqslant s \leqslant t). \tag{4.34}$$

The purpose of this filtering formulation is to estimate the message based on $\mathcal{F}_t$.
Let

$$\pi_t(\cdot) \equiv \mathbb{P}(w \in \cdot|\mathcal{F}_t) \tag{4.35}$$

be the regular conditional probability distribution of the message $w$ given $\mathcal{F}_t$ [4]. That is, $\pi_t$ is a map from $2^{\mathcal{W}} \times \Omega$ to $[0, 1]$ such that [43]

1. For any $\omega \in \Omega$, $\pi_t(\cdot, \omega)$ is a probability measure on $\mathcal{W}$.

2. For any $A \in 2^{\mathcal{W}}$, $\pi_t(A, \cdot)$ is a $\mathcal{F}_t$-measurable random variable, and

3. For any $A \in 2^{\mathcal{W}}$, we have

$$\pi_t(A, \omega) = \mathbb{P}(w \in A|\mathcal{F}_t)(\omega), \text{ a.s. } \omega. \tag{4.36}$$

[4] We assume that this regular conditional probability exists since the spaces involved are standard [18].

Fix the time $t$. Then, let

$$p_{\max} = \max_{w \in \mathcal{W}} \pi_t(\{w\}) \tag{4.37}$$

be the maximum a posteriori probability. Once this maximization is performed and the corresponding maximizing $w$ is chosen, the conditional (conditioned on $\mathcal{F}_t$) probability of error is given by

$$P_e(\mathcal{F}_t) = 1 - p_{\max}. \tag{4.38}$$

The expectation of (4.38) with respect to the probability measure $\mathbb{P}$ is the unconditional probability of error

$$P_e = E[P_e(\mathcal{F}_t)]. \tag{4.39}$$

For any fixed $\delta > 0$, we can define a stopping time $\tau$ as the first time that the error probability goes below $\delta$, if this happens before $T$, and as $T$, otherwise. Thus,

$$\tau = \inf\{t : (P_e(\mathcal{F}_t) \leqslant \delta) \text{ or } (t = T)\}. \tag{4.40}$$

If $P_e(\mathcal{F}_\tau)$ exceeds $\delta$, then we are certain that $\tau = T$, and $P_e(\mathcal{F}_t) > \delta$ for all $0 \leqslant t \leqslant T$, so the event $P_e(\mathcal{F}_\tau) > \delta$ is included in the event $P_e(\mathcal{F}_T) > \delta$. (We have inclusion instead of equality since $P_e(\mathcal{F}_\tau) \leqslant \delta$ does not exclude $P_e(\mathcal{F}_T) > \delta$. [4]) Thus,

$$\Pr\{P_e(\mathcal{F}_\tau) > \delta\} \leqslant \Pr\{P_e(\mathcal{F}_T) > \delta\} \leqslant \frac{P_e}{\delta} \tag{4.41}$$

where the second inequality is an application of Markov's inequality.

Given a particular realization $y_{[0,t]}$ we will denote the entropy of the regular conditional probability (4.35) (see also [18], section 7.2) $\pi_t(\cdot)$ as $\mathcal{H}(W|\mathcal{F}_t)$. Then, by point (2) in the definition of regular conditional probability, $\mathcal{H}(W|\mathcal{F}_t)$ is a random variable and $E[\mathcal{H}(W|\mathcal{F}_t)] \triangleq H(W|Y_{[0,t]})$. Here the expectation is with respect to the probability measure on $\mathcal{F}_t$.

If $P_e(\mathcal{F}_\tau) \leqslant \delta \leqslant \frac{1}{2}$, then from Fano's inequality it follows that

$$\mathcal{H}(W|\mathcal{F}_\tau) \leqslant h_b(\delta) + \delta \ln M. \tag{4.42}$$

The expected value of $\mathcal{H}(W|\mathcal{F}_\tau)$ can be bounded as follows:

$$\begin{aligned}
E[\mathcal{H}(W|\mathcal{F}_\tau)] &= E[\mathcal{H}(W|\mathcal{F}_\tau)|P_e(\mathcal{F}_\tau) \leqslant \delta]\Pr\{P_e(\mathcal{F}_\tau) \leqslant \delta\} \\
&+ E[\mathcal{H}(W|\mathcal{F}_\tau)|P_e(\mathcal{F}_\tau) > \delta]\Pr\{P_e(\mathcal{F}_\tau) > \delta\} \\
&\leqslant (h_b(\delta) + \delta \ln M)\Pr\{P_e(\mathcal{F}_\tau) \leqslant \delta\} \\
&+ (\ln M)\Pr\{P_e(Y_\tau) > \delta)\} \\
&\leqslant h_b(\delta) + (\delta + \frac{P_e}{\delta})\ln M.
\end{aligned}$$

Just as $Y_t$ is an $\mathcal{F}_t$-measurable random variable, we define $Y_{t-}$ to be a $\mathcal{F}_{t-}$-measurable random variable, both on the same space $(\mathcal{W} \times \Omega, 2^{\mathcal{W}} \times \mathcal{F}, \mathbb{P})$. We note that $\mathcal{F}_{t-} \subset \mathcal{F}_t$. Thus $Y_0^{t-}$ carries less "information" than $Y_0^t$. Similarly we define $X_{t-}$ to be a $\mathcal{G}_{t-}$-measurable random variable. Along the same lines as $\mathcal{H}(W|\mathcal{F}_t)$, we can also define $\mathcal{H}(W|\mathcal{F}_{t-})$ with respect to the corresponding regular conditional probability $\pi_{t-}(\cdot)$. We define $E[\mathcal{H}(W|\mathcal{F}_{t-})] \triangleq H(W|Y_{[0,t-)})$. Now we will prove the following lemma:

**Lemma 4.3.4** *For any* $0 < \delta \leqslant \frac{1}{2}$,

$$E[\tau] \geqslant \frac{1}{C}(\ln M(1 - \delta - \frac{P_e}{\delta}) - h_b(\delta)). \tag{4.43}$$

**Proof** Consider

$$\begin{aligned}
E[\mathcal{H}(W|\mathcal{F}_{t-}) - \mathcal{H}(W|\mathcal{F}_t)|\mathcal{F}_{t-}] &= E[\mathcal{H}(W|\mathcal{F}_{t-})|\mathcal{F}_{t-}] \\
&- E[\mathcal{H}(W|\mathcal{F}_t)|\mathcal{F}_{t-}] \\
&\triangleq H(W|Y_{[0,t-)}, \mathcal{F}_{t-}) \\
&- H(W|Y_{[0,t]}, \mathcal{F}_{t-}) \\
&\triangleq I(W; Y_t|\mathcal{F}_{t-}) \\
&\overset{(a)}{\leqslant} I(X_t; Y_t|\mathcal{F}_{t-}) \\
&\overset{(b)}{\leqslant} C
\end{aligned}$$

where the conditional mutual information is normally defined (as the divergence between two conditional probability measures) (see [21], equation (1.6.6)) and (a) follows from the data processing inequality and the fact that given $X_t$, the random variable $Y_t$ is independent of the message $W$ when the feedback $Y_0^{t-}$ is known. (b) follows since $C$ is obtained by maximizing the mutual information. *Note:* The finitude of $C$ is required for this inequality to be useful. This inturn imposes an average power constraint on the channel input.

The last inequality says that $\{\mathcal{H}(W|\mathcal{F}_{t-}) + tC\}$ is a submartingale. This follows from Lemma 2.68.2 in [8] (ascribed to Doob). We note that $t- = t = t+$ since $t$ is a continuous function of $t$.

Now we will apply Doob's optional sampling theorem (4.1.2) to this submartingale. We need to ensure that the submartingale is right-continuous. As shown in Appendix 2, Theorem (4.1.3) implies that this submartingale has a right continuous modification which is a submartingale with respect to the same filtration. Therefore, in what follows we replace the original submartingale $\{\mathcal{H}(W|\mathcal{F}_{t-}) + tC\}$ by the right-continuous modification $\{\mathcal{H}(W|\mathcal{F}_t) + tC\}$.

Clearly, $E[\tau] \leqslant E[T] < \infty$. Then, by the theorem (4.1.2), we conclude that at time $\tau$, the expected value of the modified submartingale must be above the initial value $\ln M$. Hence [4],

$$\begin{aligned} \ln M = \mathcal{H}(W|\mathcal{F}_0) &\leqslant E[\mathcal{H}(W|\mathcal{F}_\tau) + \tau C] \\ &= E[\mathcal{H}(W|\mathcal{F}_\tau)] + E[\tau]C \\ &\leqslant h_b(\delta) + (\delta + \frac{P_e}{\delta}) \ln M + E[\tau]C. \end{aligned} \tag{4.44}$$

Solving for $E[\tau]$ yields

$$E[\tau] \geqslant \frac{1}{C}(\ln M(1 - \delta - \frac{P_e}{\delta}) - h_b(\delta)). \tag{4.45}$$

This concludes the proof of the lemma. ∎

### 4.3.3 Burnashev's Lower Bound in Continuous Time

The goal for this section is to combine the Lemmas (4.3.3) and (4.3.4) in order to obtain a bound on the overall expected decoding time.

Proposition 2 in [4] goes through without any change from the discrete to the continuous setting. The only change is the replacement of conditional probabilities by regular conditional probabilities. We have:

**Proposition 4.3.5** *If $C_1 < \infty$*

$$\lambda\pi(w|\mathcal{F}_{t-}) \leqslant \pi(w|\mathcal{F}_t) \leqslant \frac{\pi(w|\mathcal{F}_{t-})}{\lambda}. \tag{4.46}$$

where $\lambda = \inf_{x,y} f(y|x)$. $\lambda > 0$ since there is a peak power constraint (4.12). This then gives a minimum transition probability such that $f(y|x) \geqslant \lambda$ for all $x, y$.

Now we are in a position to state the main result of this paper.

**Theorem 4.3.6** *The expected decoding time $T$ of a variable-length block code for a peak and average power constrained continuous time AWGN channel used with feedback satisfying the assumptions of right-continuity of sample paths and Nack-continuity of the Encoder is lower-bounded by*

$$E[T] \geqslant \left(1 - \delta - \frac{P_e}{\delta}\right)\frac{\ln M}{C} + \frac{-\ln P_e}{C_1} - \frac{h_b(\delta)}{C} + \frac{\ln(\lambda\delta) - \ln 4}{C_1} \tag{4.47}$$

*where $M$ is the cardinality of the message set, $P_e$ the error probability, $\lambda = \min_{x\in\mathbb{R}, y\in\mathbb{R}} f(y|x)$, and $\delta$ is any number satisfying $0 < \delta \leqslant \frac{1}{2}$.*

**Proof** The following initial portion of the proof closely follows section 5 of [4] since the arguments are nearly identical. The objective is to lower-bound the probability of error of a decoder that decides at time $T$. The key idea is that a binary hypothesis decision such as deciding whether or not $W$ lies in some set $\mathcal{G}$ can be made at least as reliably as a decision on the value of $W$ itself.

Given a set $\mathcal{G}$ of messages, consider deciding between $W \in \mathcal{G}$ and $W \notin \mathcal{G}$ in the following way: given access to the original decoder's estimate $\hat{W}$, declare that $W \in \mathcal{G}$ if $\hat{W} \in \mathcal{G}$, and declare $W \notin \mathcal{G}$ otherwise. This binary decision is always correct when the original decoder's estimate $\hat{W}$ is correct. Hence, the probability of error of this (not necessarily optimal) binary decision rule cannot exceed the probability of error of the original decoder, for any set $\mathcal{G}$. Thus, the error probability of the optimal decoder deciding at time $T$ whether or not $W \in \mathcal{G}$ is a lower bound to the error probability of any decoder that decodes $W$ itself at time $T$. This fact is true even if the set $\mathcal{G}$ is chosen at a particualr stopping time $\tau$ and the error probabilities we are calculating are conditioned on the observation $Y^\tau$ or the sub-$\sigma$-field $\mathcal{F}_\tau$.

For every realization of $Y^\tau$, the message set can be divided into two parts, $\mathcal{G}(\mathcal{F}_\tau)$ and its complement $\mathcal{W}\backslash\mathcal{G}(\mathcal{F}_\tau)$, in such a way that both parts have an *a posteriori* probability greater than $\lambda\delta$. The rest of this paragraph describes how this is possible. From the definition of $\tau$, at time $\tau-$ the *a posteriori* probability of every message is smaller than $1 - \delta$. This implies that the sum of the *a posteriori* probabilities of any set of $M-1$ messages is greater than $\delta$ at time $\tau-$, and by Proposition (4.3.5), greater than $\lambda\delta$ at time $\tau$. In particular, $P_e(\mathcal{F}_\tau) \geqslant \lambda\delta$. We separately consider the cases $P_e(\mathcal{F}_\tau) \leqslant \delta$ and $P_e(\mathcal{F}_\tau) > \delta$. In the first case, $P_e(\mathcal{F}_\tau) \leqslant \delta$, let $\mathcal{G}(\mathcal{F}_\tau)$ be the set consisting of only the message with the highest *a posteriori* probability at time $\tau$. The *a posteriori* probability of $\mathcal{G}(\mathcal{F}_\tau)$ then satisfies $\Pr\{\mathcal{G}(\mathcal{F}_\tau)\} \geqslant 1-\delta \geqslant 1/2 \geqslant \lambda\delta$. As argued above, its complement (the remaining $M-1$ messages) also has *a posteriori* probability greater than $\lambda\delta$, thus for this $\mathcal{G}(\mathcal{F}_\tau), \pi\{\mathcal{G}(\mathcal{F}_\tau)|\mathcal{F}_\tau\} \in [\lambda\delta, 1-\lambda\delta]$. In the second case, namely when $P_e(\mathcal{F}_\tau) > \delta$, the *a posteriori* probability of each message is smaller than $1-\delta$. In this case, the set $\mathcal{G}(\mathcal{F}_\tau)$ may be formed by starting with the empty set and adding messages in arbitrary order until the threshold $\delta/2$ is exceeded. This ensures that the *a priori* probability of $\mathcal{G}(\mathcal{F}_\tau)$ is greater than $\lambda\delta$. Notice that the threshold will be exceeded by at most $1-\delta$, thus the complement set has an *a posteriori* probability of at least $\delta/2 > \lambda\delta$. Thus, $\pi\{\mathcal{G}(\mathcal{F}_\tau)|\mathcal{F}_\tau\} \in [\lambda\delta, 1-\lambda\delta]$.

For any realization of $Y^\tau$ we have the binary hypothesis testing problem, running from $\tau$ until $T$, deciding whether or not $W \in \mathcal{G}(Y^\tau)$. Notice that the *a priori* probabilities of the two hypotheses of this binary hypothesis testing problem are the *a posteriori* probabilities of $\mathcal{G}(\mathcal{F}_\tau)$ and $\mathcal{W}\backslash\mathcal{G}(\mathcal{F}_\tau)$ at time $\tau$ each of which is shown to be greater than $\lambda\delta$ in the preceding paragraph. We apply Lemma (4.3.3) with $A = \mathcal{G}(\mathcal{F}_\tau)$ and $N = \mathcal{W}\backslash\mathcal{G}(\mathcal{F}_\tau)$ to lower-bound the probability of error of the binary decision made at time $T$ and, as argued above, we use the result to lower-bound the probability that $\hat{W} \neq W$. Initially, everything is conditioned on the channel output up to time $\tau$; thus

$$\pi\{\hat{W}(\mathcal{F}_T) \neq W|\mathcal{F}_\tau\} \geqslant \frac{\lambda\delta}{4}e^{-C_1 E[T-\tau|\mathcal{F}_\tau]}. \tag{4.48}$$

Taking the expectation of the above expression over all realizations of $Y^\tau$ yields the unconditional probability of error

$$\begin{aligned} P_e &= E[\pi\{\hat{W}(\mathcal{F}_T) \neq W|\mathcal{F}_\tau\}] \\ &\geqslant E[\frac{\lambda\delta}{4}e^{-C_1 E[T-\tau|\mathcal{F}_\tau]}]. \end{aligned} \tag{4.49}$$

Using the convexity of $e^{-x}$ and Jensen's inequality, we obtain

$$P_e \geqslant \frac{\lambda\delta}{4}e^{-C_1 E[T-\tau]}. \tag{4.50}$$

Solving for $E[T-\tau]$ yields

$$E[T-\tau] \geqslant \frac{-\ln P_e - \ln 4 + \ln(\lambda\delta)}{C_1}. \tag{4.51}$$

Combining Lemma (4.3.4) and (4.51) yields the result of the theorem. ∎

## 4.4 Conclusion

In this chapter we derived a lower bound (4.47) on the expected decoding time of any communication scheme over a power constrained continuous time Additive White Gaussian Noise channel with perfect feedback satisfying the assumptions of right continuous sample paths and *Nack-continuity* of the Encoder as well as a technical condition on the output measure. By the use of Striebel's theorem, a lemma of Doob and regular conditional probability, and related results for continuous time submartingales, we were able to extend the discrete time results of [4] to the continuous time setting, obtaining exact counterparts to the discrete results. This is an important first step. The next step, being presently undertaken, is to study achievability schemes over this communication channel and determine an upper bound on the expected decoding time. The togetherness of the bounds will be crucial in determining the *reliability gap*.

## Appendix 1: Right Continuity of $V_t$

We have (4.18):

$$V_t = \ln \frac{P_A(Y_0^t)}{P_N(Y_0^t)}$$

We must show

$$\lim_{s \downarrow t} V_s = V_t \tag{4.52}$$

Consider the Left Hand Side,

$$\begin{aligned}
\lim_{s\downarrow t} V_s &= \lim_{s\downarrow t} \ln \frac{P_A(Y_0^s)}{P_N(Y_0^s)} \qquad (4.53) \\
&\stackrel{(a)}{=} \ln \frac{\lim_{s\downarrow t} P_A(Y_0^s)}{\lim_{s\downarrow t} P_N(Y_0^s)} \\
&\stackrel{(b)}{=} \ln \frac{P_A(\bigcap_{s>t}\{Y_0^s = y_0^s\})}{P_N(\bigcap_{s>t}\{Y_0^s = y_0^s\})} \\
&\stackrel{(c)}{=} \ln \frac{P_A(\{Y_0^{t+} = y_0^{t+}\})}{P_N(\{Y_0^{t+} = y_0^{t+}\})} \\
&\stackrel{(d)}{=} \ln \frac{P_A(\{Y_0^{t} = y_0^{t}\})}{P_N(\{Y_0^{t} = y_0^{t}\})} \\
&= V_t
\end{aligned}$$

where (a) follows because ln is a continuous function on the positive reals, (b) follows since we are considering the limit of a decreasing sequence of events, (c) follows by definition of $\mathcal{F}_{t+}$ and (d) follows by right continuity of the output sample paths.

# Appendix 2: Application of Theorem (4.1.3)

We must first show that $\{\mathcal{H}(W|\mathcal{F}_{t-}) + tC\}$ is a right continuous process. This follows by right continuity of the regular conditional probability $\pi_{t-}(\cdot)$, a proof that follows along similar lines to Appendix 1 and is therefore omitted. Also needed are continuity of ln, and the Dominated Convergence Theorem.

Once this is established, Theorem (4.1.3) can be used. We primarily need to show that the function $l : t \mapsto E[\mathcal{H}(W|\mathcal{F}_{t-})] = H(W|Y_{[0,t)})$ is right continuous. Since $\mathcal{H}(W|\mathcal{F}_{s-})$ is bounded by $\ln M$, this follows by the Bounded Convergence Theorem. Thus the conclusions of Theorem (4.1.3) follow. In particular, a right continuous modification exists which is a submartingale with respect to the same filtration. This modification is nothing but $\{\mathcal{H}(W|\mathcal{F}_t) + tC\}$.

# Chapter 5

# The noisy feedback Poisson noise channel

## 5.1 Introduction

Neuronal information theory is a relatively new field of study. It involves looking at neuronal channels as information and communication systems and treating their capacities and reliabilities just as we would treat the capacity and reliability of a non-biological channel. A step in this direction was taken in [22], where the authors studied the (Shannon) capacity of a single neuron.

We consider a neuronal forward channel used in conjunction with a feedback neuronal channel. The goal is to determine a bound on the reliability function of this communication system. We parallel the work done in [34] and [10], specializing to the neuronal case. If the noise in the forward and feedback Poisson channel are considered to be correlated, then this becomes a systemic information theoretic study. However due to time and space limitations, this will not be pursued in this thesis.

There exists prior work on the Poission channel (which is a model for the neuronal channel, as we discuss in the next section) with noiseless feedback [30]. However, it appears that there has been no work on the reliability function of neuronal channels with neuronal or noisy feedback, which we address here.

This paper is structured as follows. In Section 2, we describe the neuronal communication model and the communication system. In Section 3, we present an outline of the code used for communication. In Section 4, we analyse the total probability of error of each stage of the code. In Section 5, we bound the expected decoding time of the messages transmitted and, finally, in Section 6, we compute a bound on the reliability function of the communication system. We then conclude in

Section 7.

## 5.2 The Neuronal Model and the Communication System

### 5.2.1 The Neuronal Model

Neuronal communication takes place in one of two possible modes: a mode called rate coding [27] and a mode called temporal coding [13]. In rate coding, which we will exclusively focus on in this paper, a time window is set, and the number of spikes in the interval (the rate) is determined [22]. Let $R$ denote the rate and $\Delta$ denote the interval.

For the rate coding model, the authors in [22] establish that in a certain parameter regime ($\kappa = 1$), the neuronal channel behaves as a Poisson channel where the channel input is a parameter $\theta$ and the channel output is the rate $R$. This channel is shown to be governed by the following probability distribution:

$$p(r|\theta; 1, \Delta) = \left(\frac{\Delta}{\theta}\right)^r \frac{e^{-\Delta/\theta}}{r!} \tag{5.1}$$

where $r$ is a value taken on by the stochastic variable $R$.

Note that although $\theta$ is the channel input, it is nevertheless just a parameter. In fact, $\theta = E[T]$ where $T$ is the inter-spike interval (ISI) and $E[\cdot]$ denotes expected value.

We immediately seek to establish a comparison with the Poisson (direct detection photon) channel from [32]. That channel is governed by the transition probability density function:

$$\Pr\left\{v(t+\tau) - v(t) = j\right\} = \frac{e^{-\Lambda}\Lambda^j}{j!}, \text{ where } j = 0, 1, 2, \ldots \tag{5.2}$$

where

$$\Lambda = \int_t^{t+\tau} \lambda(t')dt', \tag{5.3}$$

$$0 \leqslant \tau, t < \infty \tag{5.4}$$

and

$$0 \leqslant \lambda(t) \leqslant A \tag{5.5}$$

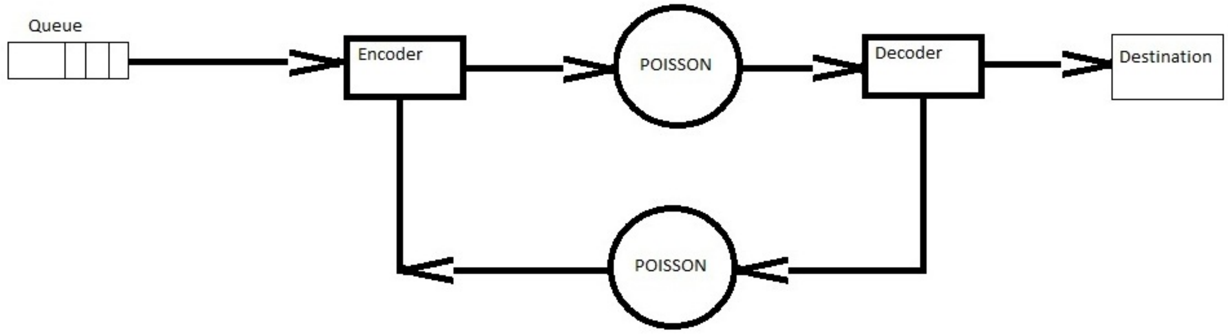


Figure 5.1: The communication system under consideration. The forward Poisson channel is assisted by the presence of a feedback Poisson channel. This situation also models a neuronal feedback system.

is the channel input wave form, defined on $0 \leqslant t < T$ where $A$ is the peak power. Here $\tau$ is the width of the region of integration for $\Lambda$[1]. Also, we have an average power constraint on $\lambda(t)$:

$$\frac{1}{T}\int_0^T \lambda(t)dt \leqslant \sigma A. \tag{5.6}$$

There is a correspondence between $\Lambda$ in (5.2) and $\frac{\Delta}{\theta}$ in (5.1). We note that the larger $\Lambda$ is, the larger is the mean value of the size of the jump in $v(t)$, which equals the number of spikes in the interval of width $\tau$, and the smaller is the expected ISI, $\Delta$ being held constant. This makes intuitive sense: the more the number of spikes in a window, the shorter the average inter spike spacing. Having noted this correspondence between the two channels (neuronal and Poisson), we will now exclusively focus on the direct detection photon (Poisson) channel. Reliability results obtained for the latter channel will thereby give an idea of how well or poorly the neuronal channel can perform in terms of reliability. The reliability is related to the rate at which the probability of error goes to zero with block length or decoding time and is therefore an important performance metric for any channel.

### 5.2.2 The Communication System Based on the Neuronal Model

The communication system that we will consider is illustrated in Figure 1. It consists of a message queue holding classical $M$-ary messages, an encoder, a decoder and the two Poisson channels: one forward and one feedback.

[1]Note that we have set the dark current to 0 throughout this paper.

1. Channel 1 is a forward Poisson channel. In this channel the input is one of $M$ possible waveforms $\lambda_m(t)$ where $1 \leqslant m \leqslant M$ and the output is an independent increments process $v(t)$ determined by (5.2).

2. Channel 2 is a feedback Poisson channel. In this channel the input is one of $M$ possible waveforms $\gamma_m(t)$, $1 \leqslant m \leqslant M$, and the output is an independent increments process $w(t)$ determined by (5.2) with $\lambda(t)$ and $v(t)$ replaced by $\gamma(t)$ and $w(t)$ respectively.

The above model is understood to be an approximation to a neuronal feedback system. For example, the independent increments process assumption, may, in reality, not hold.

## 5.3 A Possible Coding Strategy over the Neuronal Channel

The coding strategy over the neuronal feedback system will consist of two phases each, for the encoder and the decoder. At the outset we specify the following notation: $T_1$ is the duration of signals used by the encoder in Phase 1 to transmit messages over the forward channel. $T_2$ is the duration of signals used by the decoder in Phase 1 to transmit messages over the feedback channel. $T_3$ is the duration of signals used by the encoder in Phase 2 to transmit messages over the forward channel, and finally, $T$ is the duration of signals used by the decoder in Phase 2 to transmit messages over the feedback channel[2]. Let us now begin by describing the encoder's phase one transmission scheme.

### 5.3.1 Phase 1 Encoder

The encoder receives an $M$-ary message from the queue. It sends this message over the forward Poisson channel using one of $M$ possible waveforms $\lambda_m(t)$, each of duration $T_1$ seconds. The waveforms are selected from the Wyner code construction [42].

### 5.3.2 Phase 1 Decoder

The decoder receives the channel output waveform $v(t)$ and performs maximum likelihood detection on it. This detection scheme is discussed in [42]. The probability of error obtained in the absence

[2] $T$ is used instead of $T_4$ in order to distinguish the anytime code that will be used in this stage.

of dark current satisfies [42]:

$$-\frac{1}{T_1}\ln P_e \geqslant Aq - Aq^{(1+\rho)} - \rho R. \tag{5.7}$$

for $R \geqslant 0$, $R = \frac{\ln M}{T_1}$ is the operating data rate in nats per second, $q \in [0,\sigma]$ and $\rho \in [0,1]$ and $T_1$ sufficiently large.

The decoder must tell the encoder which waveform it received. It does this by sending the decoded message back over the feedback Poisson channel using an identification code [7].

### 5.3.3 Phase 2 Encoder

The encoder receives the feedback waveform $w(t)$ and performs identification oriented decoding to take advantage of the fact that it knows which message to look for or "identify" [7]. The maximum probability of error obtained is bounded by

$$P_{e2} \leqslant \delta \max\{1/2, \ln(2/\delta)\} = \delta \ln(2/\delta) \tag{5.8}$$

where $1/2 > \delta > 0$ for the case of no dark current. Now $\epsilon < \delta$ and $\delta = \frac{2}{e^{\epsilon A T_2}}$. Thus we get

$$P_{e2} \leqslant 2\epsilon A T_2 e^{-\epsilon A T_2} \tag{5.9}$$

where $1/2 > \delta > \epsilon > 0$. Thus,

$$-\frac{\ln P_{e2}}{T_2} \geqslant -\frac{\ln(2\epsilon A T_2)}{T_2} + \epsilon A \tag{5.10}$$

The first term vanishes with large $T_2$. So the exponent is asymptotically lower bounded by $\epsilon A$ where $A$ is the peak power of the signal. This is intuitive: the greater the peak power, the larger the reliability function.

The encoder uses this information to send a confirm/deny signal to the decoder. It does this using a signal set of size 2, again from the Wyner [42] construction, consisting of two signals $\lambda_1(t)$ and $\lambda_2(t)$, each of duration $T_3$ seconds.

### 5.3.4 Phase 2 Decoder

The decoder receives this binary message and performs binary maximum likelikhood decoding, obtaining an error probability given by

$$P_{e3} = e^{-T_3 A q}. \tag{5.11}$$

This is because in the case of $M = 2$ messages, an error occurs *iff* there are 0 arrivals in the non-zero region of the transmitted waveform.

Depending on whether it decodes to a confirm or deny, the decoder keeps or discards the message received in Phase 1. It must now inform the encoder of what it did. It does this by transmitting one bit of state information using an anytime code [35].

The probability of error when the encoder decodes this anytime bit is given by (see Appendix):

$$P_{e4} \leqslant K' e^{-dTP_{av}} \tag{5.12}$$

where $d$ is the delay of the anytime bit being decoded, $T$ is the duration of each signal on the anytime tree and $P_{av}$ is the average power of each of the anytime signals. Note that for this stage of the communication process, we assume that the peak power constraint on the feedback channel is relaxed and only an average power constraint exists. This is necessary for the anytime code construction.

## 5.4 Total Probability of Error of the Coding Strategy

An error in transmission can occur in one of the following ways [10].

1. The encoder sends the wrong message. This happens when the anytime bit is decoded incorrectly. The corresponding probability of error is $P_{e4} \leqslant K' e^{-dTP_{av}}$.

2. The encoder sends a confirm when it should have sent a deny. This happens due to identification error, which happens with probability $P_{e2} \leqslant 2\epsilon A_2 T_2 e^{-\epsilon A_2 T_2}$. $A_2$ is the amplitude of the identification signals.

3. The encoder sends a deny, but it is decoded as a confirm. This happens with probability $P_{e3} = e^{-T_3 A_3 q_3}$. $A_3$ is the amplitude of the binary detection scheme signals. Here $q_3 = \frac{1}{2}$.

The total error probability $P$ is bounded by the Union Bound.

$$P \leqslant P_{e4} + P_{e2} + P_{e3}. \tag{5.13}$$

By choosing parameters wisely, we can equate all three exponents to the third one. For the parameters of the scheme, we obtain

$$T_2 = \frac{A_3 T_3 q_3}{\epsilon A_2} \tag{5.14}$$

and

$$d = \frac{T_3 A_3 q_3}{T P_{av}} \tag{5.15}$$

This yields

$$-\ln P \approx -\ln(K' + 2\epsilon A_2 T_2 + 1) + T_3 A_3 q_3 \tag{5.16}$$

Ignoring the sublinear (in $T_2$) term, we can write,

$$-\ln P \approx T_3 A_3 q_3. \tag{5.17}$$

## 5.5 Expected Decoding Time of the Coding Strategy

The following are the cases in which a retransmission of a message is required [10]:

1. The message sent by the encoder is decoded incorrectly. This happens with probability $P_e$. This can be made small by increasing $T_1$.

2. A confirm sent by the encoder is decoded as a deny. This happens with probability $P_{e3}$. This can be made small by increasing $T_3$.

The probability that a message will be retransmitted is $\leqslant P_e + P_{e3}$; therefore, a transmitted message is accepted with probability $\geqslant 1 - P_e - P_{e3}$. The possible message acceptance times are

$$\begin{aligned} &(T_1 + T_2 + T_3), \\ &(d+1)(T_1 + T_2 + T_3), \\ &2(d+1)(T_1 + T_2 + T_3), \ldots \end{aligned} \tag{5.18}$$

See Figure 2-1 in [10] for a timeline.

Thus, the expected duration for a message to be successfully transmitted $\bar{\tau}$ is upper bounded by

$$\begin{aligned} \bar{\tau} &\leqslant (T_1 + T_2 + T_3)\left(1 + \frac{(P_e + P_{e3})}{(1 - P_{e3} - P_e)}(d+1)\right) \\ &\approx T_1 + T_3\left(1 + \frac{A_3 q_3}{A_2 \epsilon}\right) \end{aligned} \tag{5.19}$$

## 5.6 The Reliability Function of the Coding Strategy: A Lower Bound

The reliability function is obtained from the expected decoding time (6.6) and total probability of error (6.5). It is given by $-\log P$ divided by expected decoding time $\bar{\tau}$.

We begin with the bound on $\bar{\tau}$ (6.6).

$$\bar{\tau} \leqslant T_1 + T_3\left(1 + \frac{A_3 q_3}{A_2 \epsilon}\right) \tag{5.20}$$

Next we will use the following facts:

$$T_1 = \frac{\log_2 M}{R_1} \tag{5.21}$$

and

$$T_3 \approx \frac{-\ln P}{A_3 q_3} \tag{5.22}$$

Plugging (5.21) and (5.22) into (5.20) we obtain

$$\begin{aligned} \bar{\tau} &\lessapprox \frac{\log_2 M}{R_1} + \left(-\frac{\ln P}{A_3 q_3}\right)\left(1 + \frac{A_3 q_3}{\epsilon A_2}\right) \\ 1 &\lessapprox \frac{\bar{R}}{R_1} + \frac{E_{\text{feedback}}(\bar{R})}{A_3 q_3}\left(1 + \frac{A_3 q_3}{\epsilon A_2}\right) \end{aligned} \tag{5.23}$$

where $\bar{R} = \frac{\log_2 M}{\bar{\tau}}$ is the average rate of communication of the feedback scheme and $E_{\text{feedback}}(\bar{R}) = \frac{-\ln P}{\bar{\tau}}$ is the reliability function of the scheme at the rate point $\bar{R}$. By rearranging we obtain,

$$E_{\text{feedback}}(\bar{R}) \gtrapprox \left(\frac{1}{\epsilon A_2} + \frac{2}{A_3}\right)^{-1}\left(1 - \frac{\bar{R}}{R_1}\right) \tag{5.24}$$

where $R_1 \leqslant C_{\text{forward}}$. Here $\epsilon A_2 \approx P_{ID}$ has an interpretation as the average power used by each of the identification signals and $A_3 \approx \frac{2sC_{\text{forward}}}{q^*(1-q^*)}$ where $C_{\text{forward}}$ is the capacity of the forward Poisson channel, $s \to \infty$ and $q^* = \min\{\sigma_{\text{forward}}, \frac{1}{2}\}$, $\sigma_{\text{forward}}$ representing the average power constraint on the forward Poisson channel.

The expression (5.24) is the reliability function of the Poisson communication system which models the corresponding neuronal communication system. Since the given scheme achieves this reliability function, this becomes a bar for what the neuronal system can achieve given its particular capacity and average power constraint.

## 5.7 Conclusion

We have modelled a neuronal communication channel as a direct detection photon (Poisson) channel, following [22]. Using two such channels, one as forward and the second as the feedback channel, we examined a communication scheme over this pair of channels. We determined a lower bound on the reliability function of this scheme. The scheme may or may not be optimal, but in either case, this exponent provides a fiducial point for what neuronal feedback systems can achieve in terms of the reliability function. Future studies will involve more realistic neuronal channel models and find ways of computing the reliability functions for those models. The no dark current assumption can also be relaxed.

## Appendix: The Anytime Reliability Function of the Poisson Channel

Wyner [42] provides a code construction which is optimal for achieving the exact error exponent of the Poisson channel. It turns out that for the parameter choice of $k = 1$, this code construction yields Pulse Position Modulation. Thus PPM is a special case of the Wyner code. The probability of decoding error for PPM over the Poisson channel is easily computed to be:

$$P_e \leqslant e^{-TP_{av}} \tag{5.25}$$

where $P_{av}$ is the average power of each of the PPM signals and $T$ is the signalling duration. This therefore yields the exponent $P_{av}$.

Now we look at the anytime tree (see also [10]), with PPM signals, each of duration $T$ seconds on each branch. Since, once the bits on the tree differ, the remaining bit sequences have disjoint supports, and the Poisson channel has disjoint intervals being independent, the signal duration of relevance is $(d+1)T$ (see also [35]). Since the derivation in [35] involves only the Union bound and the fact of disjoint supports between the true waveform and any of the false waveforms, we have that [35]:

$$\Pr(\hat{B}_i \neq B_i \text{ at delay d } | B_1^{i-1} \text{ is known}) \leqslant e^{-(d+1)TP_{av}+d\ln 2} \tag{5.26}$$

For large $T$, we ignore the second $d\ln 2$ term in the exponent in what follows. We obtain,

$$\begin{aligned}
\Pr(\hat{B}_i \neq B_i \text{ at delay d}) &\leqslant \sum_{j=0}^{i-1} \Pr(\hat{B}_{i-j} \neq B_{i-j} \\
&\quad \text{at delay d+j } |B_1^{i-1-j} \text{ known}) \\
&\leqslant \sum_{j=0}^{i-1} e^{-(d+1+j)TP_{av}} \\
&< (\sum_{j=0}^{\infty} e^{-(j+1)TP_{av}})e^{-dTP_{av}} \\
&= \left[\frac{e^{-TP_{av}}}{1-e^{-TP_{av}}}\right] e^{-dTP_{av}}
\end{aligned} \tag{5.27}$$

The term in parentheses on the last line is a constant for a given signal set; call it $K'$. Then the anytime reliability function of a Poisson channel under PPM or Wyner ($k = 1$) coding and maximum likelihood decoding is given precisely by $E_a(R) = P_{av}$. Notably, it depends neither on bit position $i$ nor on delay $d$. This completes the derivation of (5.12).

# Chapter 6

# Quantum Feedback Reliability Function

*To be added:* A cloner on the receiver side which can make low fidelity clones of the received state, use one for decoding and one for feedback transmission. Also, entanglement shared between sender and receiver. How these two modifications impact the reliability function.

## 6.1 Introduction

Quantum communication theory is a new field which is developing rapidly. A major inspiration for quantum communication and information theory is classical information theory. A popular means for getting advances in the quantum domain is to consider the results in the classical domain and try to port them to the quantum setting. In this chapter we will attempt to do a similar thing.

We consider a quantum pure-state forward channel used in conjunction with a feedback quantum pure-state channel. The feedback channel is aided by a classical one-way infinite bandwidth additive white gaussian noise channel. The goal is to determine the reliability function of this communication system. We parallel the work done in [34] and [10]. Note that if we consider shared entanglement between the transmitter and receiver, this would become a systemic study.

In section 2 we describe the communication system. In section 3 we setup the code used for communication. In section 4 we analyse the total probability of error of each stage of the code. In section 5 we bound the expected decoding time of the messages transmitted and finally in section 6 we compute the reliability function of the scheme. We then conclude in section 7.

Prior work on the quantum reliability function includes the important work of Jane Liu [31] and

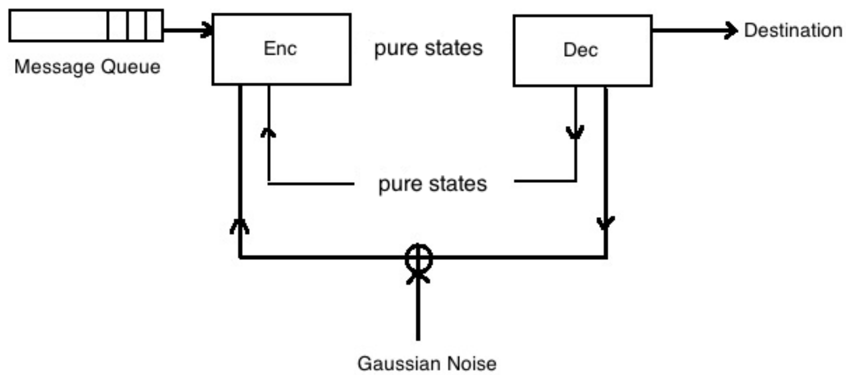


Figure 6.1: The communication system under consideration. The classical Gaussian noise channel allows for synchronization between Encoder and Decoder.

the recent works of M. V. Burnashev [6] and A. S. Holevo [20]. None of these authors consider feedback while computing the reliability function of the communication system.

## 6.2 The Communication System

The communication system consists of a message queue holding classical binary messages, an Encoder, a Decoder and three channels.

1. Channel 1 is a forward pure state quantum channel. In this channel the input is one of two pure states of some quantum system and the output is the same pure state.

2. Channel 2 is a feedback pure state quantum channel. In this channel also the input is one of two pure states of a different quantum system and the output is the same pure state.

3. Channel 3 is a feedback infinite bandwidth additive white gaussian noise channel. In this channel the input is a signal of duration of one channel use of the above two channels, carrying

anytime bit information and the output is also a signal of duration of one channel use though corrupted by Gaussian noise.

## 6.3 The Code

### 6.3.1 Phase 1 Encoder

The Encoder receives a binary message from the queue. It sends this message over the forward quantum channel using one of two pure states.

### 6.3.2 Phase 1 Decoder

The decoder receives these pure states and performs a binary decision test between these states. The average probability of error obtained is given by [19]

$$P_e = \frac{1}{2}[1 - [1 - 4\chi_0\chi_1| < \psi_1|\psi_0 > |^2]^{\frac{1}{2}}] \tag{6.1}$$

The Decoder must tell the Encoder which state it received. It does this by sending the same state as it receives back over the quantum pure-state feedback channel.

### 6.3.3 Phase 2 Encoder

The Encoder receives these pure states and performs a binary decision test between them. The average probability of error obtained is again given by

$$P_e = \frac{1}{2}[1 - [1 - 4\chi_0\chi_1| < \psi_1|\psi_0 > |^2]^{\frac{1}{2}}] \tag{6.2}$$

The Encoder uses this information to send a Confirm/Deny signal to the Decoder. It does this using once again another set of pure states $|\psi_2>$ and $|\psi_3>$.

### 6.3.4 Phase 2 Decoder

The decoder receives this binary message and performs a binary pure state decision test on it, obtaining an average probability of error given by

$$P_{e2} = \frac{1}{2}[1 - [1 - 4\chi_2\chi_3| < \psi_2|\psi_3 > |^2]^{\frac{1}{2}}] \tag{6.3}$$

Depending on whether it decodes to a Confirm or Deny, the Decoder keeps or discards the message received in Phase 1. It must now inform the Encoder of what it did. It does this by transmitting one bit of state information using an anytime code. This is done over an auxiliary feedback channel which is an infinite bandwidth AWGN channel.

The probability of error when the Encoder decodes this anytime bit is given by

$$P_{e3} \leqslant K2^{-WE_{orth}(R^{'})T} \tag{6.4}$$

where $E_{orth}$ is the standard orthogonal signalling error exponent [17], $T$ is the signalling duration corresponding to single channel use on the quantum cahnnels and $R^{'} = \frac{1}{T}$ is the consequent anytime code rate. $W$ is the number of messages transmitted before any one of them could possibly be retransmitted. The code has a tree structure as depicted in [10].

## 6.4 Total Probability of Error

An error in transmission can occur in one of the following ways [10].

1. The Encoder sends the wrong message. This happens when the anytime bit is decoded incorrectly. The corresponding probability of error is $P_{e3} \leqslant K2^{-WE_{orth}(R^{'})T}$.

2. The Encoder sends a Confirm when it should have sent a Deny. This happens with probability $P_e = \frac{1}{2}[1 - [1 - 4\chi_0\chi_1| < \psi_1|\psi_0 > |^2]^{\frac{1}{2}}]$.

3. The Encoder sends a Deny but it is decoded as a Confirm. This happens with probability $P_{e2} = \frac{1}{2}[1 - [1 - 4\chi_2\chi_3| < \psi_2|\psi_3 > |^2]^{\frac{1}{2}}]$.

The total error probability $P$ is bounded by the Union Bound.

$$\begin{aligned} P &\leqslant P_e + P_{e2} + P_{e3} \qquad (6.5) \\ &= \frac{1}{2}[1 - [1 - 4\chi_0\chi_1| < \psi_1|\psi_0 > |^2]^{\frac{1}{2}}] + \frac{1}{2}[1 - [1 - 4\chi_2\chi_3| < \psi_2|\psi_3 > |^2]^{\frac{1}{2}}] + K2^{-WE_{orth}(R^{'})T} \end{aligned}$$

## 6.5 Expected Decoding Time

The following are the cases in which a retransmission of a message is required:

1. The message sent by the Encoder is decoded incorrectly. This happens with probability $P_e$.

2. A Confirm sent by the Encoder is decoded as a Deny. This happens with probability $P_{e2}$.

The probability that a message will be retransmitted is $\leqslant P_e + P_{e2}$. Therefore a transmitted message is accepted with probability $\geqslant 1 - P_e - P_{e2}$. The possible message acceptance times are $(T+T+T), (W+1)(T+T+T), 2(W+1)(T+T+T)$, i.e. $3T, 3T(W+1), 6T(W+1), \ldots$ etc. Thus the expected duration for a message to be successfully transmitted $\bar{\tau}$ is upper bounded by

$$3T\left(1+\frac{(P_e+P_{e2})}{(1-P_{e2}-P_e)}(W+1)\right) \tag{6.6}$$

## 6.6 The Reliability Function: A Lower Bound

The reliability function is obtained from the expected decoding time (6.6) and total probability of error (6.5). It is given by $-\log P$ divided by expected decoding time.

We first compute the base 2 exponents of the probabilities of error $P_e$ and $P_{e2}$. Consider

$$P_e = \frac{1}{2}[1-[1-4\chi_0\chi_1|<\psi_1|\psi_0>|^2]^{\frac{1}{2}}]$$

. Assume that the prior probabilities on the two states are $\frac{1}{2}$ each. Then

$$P_e = \frac{1}{2}[1-[1-|<\psi_1|\psi_0>|^2]^{\frac{1}{2}}]$$

. Thus,

$$-\log_2 P_e = 1-\log_2(1-\sqrt{1-|<\psi_1|\psi_0>|^2})$$

. Using the fact

$$\log_2(1+x) = \frac{1}{\ln 2}[-x-\frac{x^2}{2}-\frac{x^3}{3}-\frac{x^4}{4}-\cdots]$$

, for $x \in \{-1,1\}$ we get

$$-\log_2 P_e = 1+\frac{1}{\ln 2}[\sqrt{1-|<\psi_1|\psi_0>|^2}+\frac{(1-|<\psi_1|\psi_0>|^2)}{2}+\frac{(1-|<\psi_1|\psi_0>|^2)^{\frac{3}{2}}}{3}+\cdots] \tag{6.7}$$

Here $| < \psi_1|\psi_0 > |^2 \in \{0, 1\}$. Therefore $x \in \{0, 1\}$. Thus we can ignore $x^2$ and higher powers thereby obtaining the expression

$$-\log_2 P_e = 1 + \frac{1}{\ln 2}\sqrt{1 - | < \psi_1|\psi_0 > |^2} + \text{sublinear terms} \tag{6.8}$$

. Thus

$$P_e \approx 2^{-(1+\frac{1}{\ln 2}\sqrt{1-|<\psi_1|\psi_0>|^2})}$$

and

$$P_{e2} \approx 2^{-(1+\frac{1}{\ln 2}\sqrt{1-|<\psi_2|\psi_3>|^2})} \tag{6.9}$$

and we have from (6.4)

$$P_{e3} \leqslant K2^{-WE_{orth}(\frac{1}{T})T}$$

.

We can choose parameters such that all the three above exponents are equal to the first one. This implies that the overlap between $|\psi_0 >$ and $|\psi_1 >$ must be the same as that between $|\psi_2 >$ and $|\psi_3 >$ and also that

$$W = \frac{1 + \frac{1}{\ln 2}\sqrt{1 - | < \psi_1|\psi_0 > |^2}}{E_{orth}(\frac{1}{T})T}$$

.

Thus for the overall probability of error $P$ we have

$$-\log_2 P = 1 + \frac{1}{\ln 2}[\sqrt{1 - | < \psi_1|\psi_0 > |^2}] \tag{6.10}$$

We note in (6.6) that $P_e$ and $P_{e2}$ can be made arbitrarily small. Thus we get

$$\begin{aligned} \bar{\tau} &\leqslant 3T \\ &< \frac{3}{R} \end{aligned} \tag{6.11}$$

where $R$ is the nominal rate in bits per second of communication over a pure-state quantum channel. This gives us an expression for the reliability function:

$$E(\bar{R}) > R\left(\frac{1 + \frac{1}{\ln 2}[\sqrt{1 - | < \psi_1|\psi_0 > |^2}]}{3}\right) \tag{6.12}$$

The bound is tightened when $R$ is maximized. Here we note that classically viewed, the quantum pure-state channel is just a Binary Symmetric Channel with parameter $P_e$. Thus $C_{BSC} = 1 - h_b(P_e)$

bits per channel use is the classical capacity of the quantum channel alone, where $h_b(\cdot)$ is the binary entropy function [12]. We get,

$$E(\bar{R}) > C_{BSC}\left(\frac{1+\frac{1}{\ln 2}[\sqrt{1-|<\psi_1|\psi_0>|^2}]}{3}\right) \tag{6.13}$$

## 6.7 Conclusion

We have analysed a quantum communication channel used with a feedback quantum communication channel and an auxiliary feedback classical infinite bandwidth additive white gaussian noise channel and derived a lower bound on the reliability function of this system when used for communicating classical bits. We have restricted ourselves to a pure state quantum channel. In future we hope to consider a mixed state channel as well as attempt the case of multiple possibilities for each message instead of just a binary scheme. The latter is an open problem since there is no closed form expression known for the probability of error for multiple hypothesis testing quantum problems when there is an arbitrary assignment of costs [19]. We also hope to replace the two feedback channels with a single quantum channel.

*Note:* The above work was done independently (not while affiliated with any university or company).

# Chapter 7

# Conclusion and Future Work

Ephaptic communication is a systemic study; entanglement is a systemic concept; Gaussian communication with pre-shared randomness and correlations between sender and receiver may be systemic; similarly also Poisson communication with pre-shared correlations. Gaussian communication with correlations between the forward and feedback paths (in noise) is definitely systemic. So is Poisson communication with correlations between the forward and feedback path noise. We can study the reliability and the channel capacity in this context.

Systemic information is thus an information theoretic study of the system as a whole rather than of individual parts. This is a sort of global or spread-out information [1]. If the system is broken down into its parts, then this information (theory) becomes meaningless.

It is not a step between point to point and network information theory. Rather it is a concept that is orthogonal to the network concept which is simply a generalization of the point-to-point concept. One could also view the point to point and network scenarios as an unsystemic limit of systemic information theory. There is no systemic information in a network unless we are talking of a quantum network with entangled nodes, or a classical network with correlated noises in the links, or a neuronal network with massive ephaptic cross-talk. Systemic information theory is the study of system-wide information theoretic effects and properties.

We can also study information exchange between synchronizing systems.

We need some general mathematical properties of systemic information such as systemic capacity or a systemic index which measures the systemic-ness of the system. It would ideally quantify how

[1] It hints at the presence of a field of information.

much the systemic-ness adds to the capacity. For example it could be the difference between the capacity/reliability when the system is systemic versus when it is not, when there are correlations versus when there are not, when there is ephaptic communication versus when there is not, when there is quantum entanglement versus when there is not, when there is forward and feedback quantum correlation versus when there is not.

A useful quantity would be the systemic information *gain* defined as the difference between the systemic capacity and the non-systemic capacity. Similarly, the systemic reliability *gain* would be the difference between the systemic reliability and the non-systemic reliability.

There could also be a tripartite or multipartite 'mutual' information as a generalization of mutual information which is always bipartite. This could be related to systemic information. It could even be a tool for "network" information. The right way to study networks.

Define the tripartite systemic information as follows:

$$\begin{aligned} \mathcal{S}(X;Y;Z) &= I(X;Y) - I(X;Y|Z) \\ &= I(X;Z) - I(X;Z|Y) \\ &= I(Y;Z) - I(Y;Z|X) \end{aligned} \tag{7.1}$$

Similarly define the 4-partite systemic information as follows:

$$\mathcal{S}(X;Y;Z;W) = \mathcal{S}(X;Y;Z) - \mathcal{S}(X;Y;Z|W) \tag{7.2}$$

The definitions are certainly symmetric. We must look for an operational meaning of systemic information. Is it the maximum rate at which you can 'spread' (or diffuse) information between three (or four or more) parties? It might be an upper bound to the capacity of any network containing three (or four or more) parties.

It turns out this concept was called 'interaction information' and has been known since the 1950s [33]. However no information theoretic operational meaning has yet been imparted to it, which could make it useful for communication systems understanding, particularly distributed communication systems. Future work will focus on finding this operational meaning.

# Bibliography

[1] Costas Anastassiou, Rodrigo Perin, Henry Markram, and Christof Koch. Ephaptic coupling of cortical neurons. *Nature neuroscience*, 14(2):217–223, 2011.

[2] Bruno Averbeck, Peter Latham, and Alexandre Pouget. Neural correlations, population coding and computation. *Nature reviews. Neuroscience*, 7(5):358–366, 2006.

[3] R. F. Bass. *Stochastic processes*. Cambridge University Press, 2011.

[4] P. Berlin, B. Nakiboglu, B. Rimoldi, and E. Telatar. A simple converse of burnashev's reliability function. *IEEE Transactions on Information Theory*, 55(7), 2009.

[5] M. V. Burnashev. Data transmission over a discrete channel with feedback. random transmission time. *Problemy Peredachi Informatsii*, 12(4), 1976.

[6] M. V. Burnashev. On reliability function of quantum communication channel. *Arxiv quant-ph/9703013*, 1997.

[7] Marat V. Burnashev. On identification capacity of infinite alphabets or continuous-time channels. *IEEE Tran. Inf. Theory*, IT-46(7), 2000.

[8] Vincenzo Capasso and David Bakstein. *An Introduction to Continuous-Time Stochastic Processes*. Birkhauser, 2004.

[9] A. Chawla. Reliability of a gaussian channel in the presence of gaussian feedback. *M. S. Thesis at M. I. T.*, 2006.

[10] A. Chawla. *Reliability of a Gaussian Channel in the Presence of Gaussian Feedback*. Massachusetts Institute of Technology, Master's Thesis, 2006.

[11] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley-Interscience, 2006.

[12] Thomas Cover and Joy Thomas. *Elements of Information Theory*. Wiley-Interscience, 2005.

[13] P. Dayan and L. F. Abbott. *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems*. The MIT Press, Cambridge, Massachusetts, 2001.

[14] J. L. Doob. *Stochastic Processes*. John Wiley And Sons, 1953.

[15] Paul David Feigin. Maximum likelihood estimation for continuous-time stochastic processes. *Advances in Applied Probability*, 8(4):712–736, December 1976.

[16] Robert G. Gallager. *Information Theory and Reliable Communication*. John Wiley And Sons, 1968.

[17] Robert G. Gallager. *Information Theory and Reliable Communication*. Wiley, 1968.

[18] R. M. Gray. *Entropy and Information Theory*. SpringerLink : Bücher. Springer, 2011.

[19] Carl W. Helstrom. *Quantum Detection and Estimation Theory*. Academic Press, 1976.

[20] A. S. Holevo. Reliability function of general classical quantum channel. *IEEE Transactions on Information Theory*, 2000.

[21] Shunsuke Ihara. *Information theory for continuous systems*. World Scientific, 1993.

[22] Shiro Ikeda and Jonathan H. Manton. Capacity of a single spiking neuron channel. *Neural Computation*, 21, 2009.

[23] Andrew H Jazwinski. *Stochastic processes and filtering theory*. Courier Dover Publications, 2007.

[24] Thomas Kailath. A general likelihood-ratio formula for random signals in gaussian noise. *IEEE Transactions on Information Theory*, 15(3):350–361, May 1969.

[25] Thomas Kailath and Moshe Zakai. Absolute continuity and Radon-Nikodym derivatives for certain measures relative to Wiener measure. *The Annals of Mathematical Statistics*, 42(1):130–140, 1971.

[26] Gopinath Kallianpur. *Stochastic filtering theory*. Springer, 1980.

[27] E. Kandel, J. Schwartz, and T. M. Jessel. *Principles of Neural Science*. Elsevier, New York, 1991.

[28] I. Karatzas and S. E. Shreve. *Brownian motion and stochastic calculus*. Springer-Verlag, 1991.

[29] Christoph Kayser, Marcelo Montemurro, Nikos Logothetis, and Stefano Panzeri. Spike-phase coding boosts and stabilizes information carried by spatial and temporal spike patterns. *Neuron*, 61(4):597–608, 2009.

[30] Amos Lapidoth. On the reliability function of the ideal poisson channel with noiseless feedback. *IEEE Transactions on Information Theory*, IT-39(2), March 1993.

[31] Jane Liu. Reliability of quantum-mechanical communication systems. *IEEE Transacations on Information Theory*, 1970.

[32] R. J. McEliece. Practical codes for photon communication. *IEEE Transactions on Information Theory*, IT-27(4), July 1981.

[33] W. J. McGill. Multivariate information transmission. *Psychometrika*, 19:97–116.

[34] A. Sahai and T. Simcek. On the variable-delay reliability function of discrete memoryless channels with access to noisy feedback. *ITW 2004*.

[35] Anant Sahai. Anytime coding on the infinite bandwidth awgn channel: a sequential semi-orthogonal code. *Conference on Information Sciences and Systems, The John Hopkins University*, 2005.

[36] E Salinas and T Sejnowski. Correlated neuronal activity and the flow of neural information. *Nature reviews. Neuroscience*, 2(8):539–550, 2001.

[37] J. P. M. Schalkwijk and T. Kailath. A coding scheme for additive noise channels with feedback – i: No bandwidth constraint. *IEEE Transactions on Information Theory*, April 1966.

[38] J. P. M. Schalkwijk and T. Kailath. A coding scheme for additve noise channels with feedback – part ii: Band-limited signals. *IEEE Transactions on Information Theory*, April 1966.

[39] A. C. Scott. *Neuroscience: A Mathematical Primer*. Springer, 2002.

[40] Charlotte T. Striebel. Densities for stochastic processes. *The Annals of Mathematical Statistics*, 30(2):559–567, June 1959.

[41] S. R. S. Varadhan. *Stochastic processes*. American Mathematical Society, 2007.

[42] Aaron D. Wyner. Capacity and error exponent for the direct detection photon channel - Part I. *IEEE Tran. Inf. Theory*, IT-34(6), 1988.

[43] Jie Xiong. *An introduction to stochastic filtering theory*. Oxford University Press, 2008.