THE STOCHASTIC-CALCULUS APPROACH TO SELECTED TOPICS IN INFORMATION THEORY

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Nirmal Vijay Shende

December 2019

THE STOCHASTIC-CALCULUS APPROACH TO SELECTED TOPICS IN INFORMATION THEORY

Nirmal Vijay Shende, Ph.D.

Cornell University 2019

We study the following three information-theoretic problems using tools derived from stochastic calculus: the multi-receiver Poisson channel, lossy compression of point-processes, and the second-order coding rate in discrete memoryless channels (DMCs) with feedback.

We obtain a general formula for the mutual information involving the point processes that allows for conditioning and the use of auxiliary random variables. We then use this formula to compute necessary and sufficient conditions under which one Poisson channel is less noisy and/or more capable than another, which turn out to be distinct from the conditions under which this ordering holds for the discretized versions of the channels. We also use the general formula to determine the capacity region of various multi-receiver Poisson channel.

We introduce a new distortion measure for point processes called functional-covering distortion. We obtain the distortion-rate function with feedforward under this distortion measure for a large class of point processes. For Poisson processes, stronger results are obtained by constraining the reconstruction. We derive the rate-distortion function for this constrained functional-covering and show that feed-forward does not improve it. Moreover, we characterize the rate-distortion region for a two-encoder CEO problem for Poisson process and show that feedforward does not improve this region. As a corollary, we obtain the rate-distortion region of remote Poisson source. A strong data processing inequality for Poisson processes

under superposition is derived to prove the converse of the CEO problem.

For DMCs, we show that feedback does not improve the second-order coding rate for a class of DMCs which complements the class of channels for which feedback is known to improve the second-order coding rate. We derive an upper bound on the achievable rate with feedback utilizing a novel proof technique for general DMCs.

BIOGRAPHICAL SKETCH

Nirmal Shende was born in Yavatmal, India in 1987. He received the B.Tech. degree in electronics and communication engineering from the Visvesvaraya National Institute of Technology, Nagpur, India, and the M.E. degree in telecommunication engineering from the Indian Institute of Science, Bengaluru, India. He obtained the Ph.D. degree with the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY, USA, where he was a recipient of the Irwin M. and Joan K. Jacobs Fellowship.

To Aai and Baba.

ACKNOWLEDGEMENTS

I would like to express my humblest gratitude to Prof. Wagner. This thesis would not have been possible without his endless patience during (and even after) my initial days as a PhD researcher. His clear, concise, and well-organized approach to problem formulation and solving is something I wish to emulate in my professional career. I thank Prof. Acharya for being on my thesis committee, numerous enjoyable discussions, collaborations, and for inviting me for dinner on many occasions. I also thank Profs. Studer and Tong for being on my committee.

Secondly, I would like to thank my family. Both Aai and Baba have been instrumental in the PhD journey. Nikhil dada has been my inspiration to pursue a PhD and he is the most vocal supporter of this endeavor. Prachi wahini has been taking exceptional care of me during my India trips. My nieces Sai and Niraja have been a delight to interact with. Special thanks go to cousin Sayli and Sunit for hosting me while I searched for jobs after the defense.

I thank my colleagues in the research group Omer, Ibrahim, Sinem, Xiaoqing, Yi, Yuguang. I especially cherish the friendship and many discussions I had with Omer and Ibrahim over all these years. I will also miss the lunches and coffees with other friends from the department including Charles, Ramina, Kolbeinn, and Sam.

I also was fortunate to have made many friends outside the department. Ajay, Pooja, Shreesha, Amandeep, Aravind, Vinay, Adarsha, Gautam, Jay have all contributed to the fun and relaxation I had outside of the work. I also thank my old friends who have helped me one way or another. I have enjoyed many technical and nontechnical discussions with Shashank and Ameya. I am also grateful to Ashwin, Chaitanya, Ketki, and Chinmayee for many encouraging chats and texts.

I wish to thank ECE department staff, particularly Scott Coldren for taking care of all bureaucratic procedures. I am grateful to the National Science Foundation for sponsoring the research via grants.

CONTENTS

LIST OF FIGURES

LIST OF THEOREMS

CHAPTER 1

INTRODUCTION

We consider three problems in information theory. The first two are related, in which we study multi-receiver Poisson channels and lossy compression of point-processes, with emphasis on Poisson processes. Third pertains to the role of feedback in discrete memoryless channels. A common theme in all of these topics is that we have used stochastic calculus to solve these problems. Each technical chapter of the thesis is self-contained and can be read independently. Below we provide a brief introduction to these topics, our approach, and results.

1.1 Multi-Receiver Poisson Channels

The Poisson channel models a direct-detection optical communication system in which the input to the channel X_0^T represents the strength of the optical input signal, and the output of the channel is a Poisson process with rate $aX_0^T + \lambda$, where a accounts for attenuation and λ represents the rate of the dark current. Capacity studies of this channel have been ongoing since it was introduced as a viable model in [44, 48].

Broadly speaking, the channel has been studied using two mathematical approaches. Early work calculated mutual information and related quantities for the channel using stochastic calculus and, in particular, the theory of point process martingales [17, 30]. Most later work followed the approach of Wyner [68] who argued that the encoder and decoder could be restricted to use the channel so that it behaves like a discrete-time, memoryless, binary channel, with no essential loss of performance. One then applies standard techniques for such channels [6, 22, 36, 56].

We espouse the former approach in this paper, both on the general principle that, when the existing tools are insufficient for a new problem, it is preferable to extend the tools rather than to reduce the problem, and for certain pragmatic reasons. The reduction to a discrete-time binary channel is somewhat involved, and it must be reproved for each new variation. Once the appropriate stochastic-calculus-based tools have been developed, on the other hand, they can be directly applied to new problems. Moreover, it is unclear how to extend Wyner's [68] reduction to some setups, such as the wiretap version of the channel considered herein.

Of course, the stochastic calculus approach also has its disadvantages: it requires more sophisticated mathematics, and one cannot apply results from the extensive literature on discrete memoryless channels. One cannot even presume that the capacity is governed by the maximal mutual information, for instance, an oversight in the early work that used this approach. On the other hand, once the necessary tools are developed, coding theorems follow expeditiously.

The goal is to develop those tools that are necessary for various multi-decoder extensions of the Poisson channel. The two-decoder Poisson channel consists of a single transmitter (which inputs process X_0^T) and two receivers with output processes Y_0^T and Z_0^T, where Y_0^T and Z_0^T are Poisson process with rates $a_y X_0^T + \lambda_y$ and $a_z X_0^T + \lambda_z$, respectively. We shall consider both the broadcast channel (either with independent or degraded message sets) and the wiretap channel (where one of the receivers is an eavesdropper).

We derive a general formula for the mutual information over a Poisson channel, which generalizes an existing formula [17, 30] by allowing the use of auxiliary random variables and conditioning. We also obtain a continuous-time Csiszár-sum-like identity for Poisson channels. Using these tools, we obtain necessary and sufficient conditions for which the broadcast channel is less noisy and more capable, and show that these orderings are in fact equivalent. These conditions turn out not to be equivalent, however, to the analogous conditions for the discrete-time binary channel obtained as a reduction of the Poisson channel [33], indicating that some care is required when interpreting results obtained via this reduction. We also rederive the capacity of the more capable broadcast channel with independent message sets (found earlier using the reduction method [33]), extend the secrecy capacity results of the degraded wiretap channel to the more capable wiretap channel, and obtain the capacity of the broadcast channel with degraded message sets.

Although, here we focus on multi-receiver Poisson channels, the tools and techniques developed in this chapter should be useful in analyzing other multi-user Poisson channel models as well. These include Poisson multiple access channel [1, 40] and Poisson interference channel [38].

1.2 Functional Covering of Point Processes

For a realization of a counting (or point) process $y_0^T = (y_t : t \in [0,T])$ (i.e. y_t is integer valued, non-decreasing, and has unit jumps)[1] and a non-negative function $\hat{y}_0^T$, we define the *functional-covering distortion* as

$$d(\hat{y}_0^T, y_0^T) \triangleq \int_0^T \hat{y}_t \, dt - \log(\hat{y}_t) \, dy_t. \tag{1.1}$$

The above distortion measure is inspired by the covering distortion measure in [42, 45], where the reconstruction is a subset of $[0,T]$ containing all the points

[1] Strictly speaking, a point process is a collection of epochs which correspond to jumps of the counting process. However, since both of these viewpoints are equivalent in the sense that one can be obtained from the other, we do not distinguish between these two.

of Y_0^T, and the distortion is the Lesbegue measure of the covering set. If we impose that $\hat{y}_t \in \{0,1\}$, then the above mentioned distortion is indeed the covering distortion measure. Hence it is natural to consider the distortion in (1.1) without such restriction. Moreover, there are advantages of not restricting $\hat{y}_0^T$ to be binary $\{0,1\}$. Consider a remote source setting where the encoder cannot access the point-process source directly, but instead observes a thinned version where some of the points in the source point process are deleted randomly. Then, in case of the covering distortion the reconstruction has to be the whole set $[0,T]$ (i.e. $\hat{y}_t = 1, t \in [0,T]$). However, the distortion measure in (1.1) has a non-trivial solution to this problem.

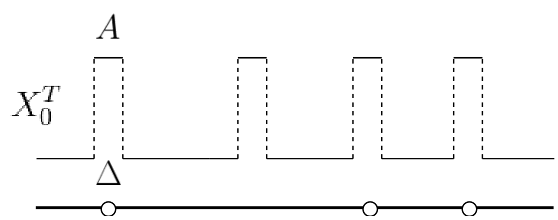

Figure 1.1: A possible reconstruction for a point process shown at the bottom in terms of arrival instants. The reconstruction levels here are binary, 0 and A. The functional-covering distortion in this case is $4A\Delta - 3\log(A)$.

The above distortion measure is also related to the logarithmic-loss function used for discrete memoryless sources (DMS) [12, 13]. If we constrain $\hat{y}_0^T$ to be bounded, then we can use Girsanov theorems [9, Chapter VI, Theorems T2-T4] to define a probability measure on the set of all counting processes using $\hat{y}_0^T$, and the distortion can be defined as the expectation of the negative logarithm of the Radon-Nikodym derivative between this probability measure and an appropriately chosen reference measure, which is equivalent to (1.1). However, we will allow $\hat{y}_0^T$ to be unbounded but integrable $\mathbb{E}[\int_0^T \hat{Y}_t\, dt] < \infty$.

Heuristically, given a random variable M, the intensity of a point process Y_0^T is a non-negative process Γ_0^T such that $P(Y_{t+\Delta}-Y_t = 1|M, Y_0^t) \approx \Gamma_t\Delta$ (see Definition 10 for the precise statement). Looking at (1.1), we expect any optimal $\hat{Y}_0^T$ (in the rate-distortion trade-off sense) to be related to the intensity of Y_0^T. In fact, we will see in the proof of Theorem 10 that an optimal reconstruction $\hat{Y}_0^T$ is the intensity of Y_0^T given the encoder's output.

Rate-distortion problems when the reconstruction is taken as a realization of point process are considered in [4, 11, 35, 54, 64], where typically the distortion is defined in terms of the inter-arrival times of these point processes.

A major contribution of this work is Theorem 7, where we derive the mutual information between point-processes with intensities and arbitrary random variables. This is the most general expression of mutual information involving point process with intensities. This expression subsumes the existing formulae for mutual informations involving doubly stochastic Poisson processes [17,30,59] and queuing processes [63] as special cases. The other theorems proved in this chapter are: We obtain the rate-distortion trade-off with feedforward for the functional-covering distortion measure for point processes which admit intensities (see Theorem 10). For Poisson processes, we obtain the rate-distortion region when the reconstruction function $\hat{y}_0^T$ is constrained to take value in a subset of reals (Theorem 11). The covering distortion in [42] is a special case of this constrained covering distortion, hence the rate-distortion function in [42, Theorem 1] can be obtained as a special case of this theorem. We characterize the rate-distortion region for a two-encoder Poisson CEO problem (see Figure 3.1)) under functional-covering distortion in Theorem 12. To prove the converse of the CEO problem, we derive a strong data processing inequality for Poisson processes under superposition (see

Theorem 8), which complements the strong data processing inequality for Poisson processes under thinning due to Wang in [66]. We also provide a self-contained proof of Wang's theorem in Theorem 9. The solution to the CEO problem gives the rate-distortion trade-off for the remote Poisson sources.

1.3 Second-Order Coding Rate in Discrete Memoryless Channels with Feedback

Consider the canonical communications model consisting of a single encoder sending bits to a single decoder over a discrete memoryless channel (DMC). We assume the alphabets are finite, the channel law is completely known, and that the transmission is fixed rate, i.e., the decoding of the entire message must occur at a prespecified time.

It is well-known that feedback does not improve the capacity of a DMC [55]. Recently, a novel mechanism was introduced by which feedback can improve coding performance for some DMCs, even when the coding is high-rate and fixed-blocklength and the channel is known and memoryless [2, 65]. For channels with multiple capacity-achieving input distributions that give rise to information densities with different variances, which are called *compound-dispersion* channels (see Definition 26), feedback codes exist that asymptotically outperform the best non-feedback codes. In channel-coding terms, the idea is that, with compound-dispersion channels, the encoder can use codewords with symbols drawn from multiple input distributions such that the mean of the rate of information conveyance across the channel is the same under all of these distributions (namely, the Shannon capacity), but the variance is different. The encoder then monitors the progress of transmission via the feedback link and uses a "bold" input distribution when a decoding error is expected and a "timid" input distribution when it is not.

In the context of feedback communication, this shows that timid/bold coding improves the second-order coding rate compared with the best non-feedback code for all compound-dispersion channels. We here show a converse result, namely that feedback does not improve the second-order coding rate of simple (i.e., non-compound) dispersion channels, improving upon [50, Theorem 15]. Thus, timid/bold coding provides a second-order coding rate improvement whenever such an improvement is possible. The converse is obtained by using the code modification technique of Fong and Tan [24] along with a "Berry-Esseen"-type martingale CLT and large deviations results for martingales. In particular, this settles the problem of determining whether feedback improves the second-order coding rate for a given DMC.

For compound-dispersion channels, it is not clear if timid/bold coding is an optimal feedback signaling scheme. To shed some light on this question, we provide the first nontrivial impossibility result for the second-order coding rate of feedback communication over DMCs. The technical challenge in proving such a result is that martingale central limit theorems do not appear to provide useful bounds. Instead, we obtain the result using tools from stochastic calculus, namely, martingale embeddings, change-of-time methods, and finally solving a stochastic differential equation. The bound on the second-order coding rate that we obtain is functionally identical to the second-order coding rate achieved by timid/bold coding, although evaluated at different channel parameters. The two bounds coincide for some channels but not in general.

CHAPTER 2

MULTI-RECEIVER POISSON CHANNELS

2.1 Preliminaries

We will construct a probability space $(\Omega, \mathcal{F}, P)$ on which all stochastic processes considered here are defined. For a finite $T > 0$, let $(\mathcal{F}_t : t \in [0, T])$ be an increasing family of σ-fields with $\mathcal{F}_T \in \mathcal{F}$. Stochastic processes are denoted as $X_0^T = \{X_t, 0 \leq t \leq T\}$. X_{t-} denotes $\lim_{\delta \to 0^+} X_{t-\delta}$ when $t > 0$, and equals X_0 when $t = 0$. The process X_0^T is said to be *adapted* to the history $(\mathcal{F}_t : t \in [0, T])$ if X_t is $\mathcal{F}_t$ measurable for all $t \in [0, T]$. The internal history recorded by the process X_0^T is denoted by $\mathcal{F}_t^X = (\sigma(X_s) : s \in [0, t])$, where $\sigma(A)$ denotes the σ-field generated by A. A process X_0^T is called $(\mathcal{F}_t : t \in [0, T])$-*predictable* if X_0 is $\mathcal{F}_0$ measurable and the mapping $(t, \omega) \to X_t(\omega)$ defined from $(0, T) \times \Omega$ into $\mathbf{R}$ (the set of real numbers) is measurable with respect to the σ-field over $(0, T) \times \Omega$ generated by rectangles of the form

$$(s, t] \times A; \quad 0 < s \leq t \leq T, \quad A \in \mathcal{F}_s. \tag{2.1}$$

Let $\mathcal{N}_0^T$ denote the set of counting realizations (or point-process realizations) on $[0, T]$, i.e., if $N_0^T \in \mathcal{N}_0^T$, then for $t \in [0, T]$, $N_t \in \mathbf{N}$ (the set of non-negative integers), is right continuous, and has unit jumps with $N_0 = 0$.

For two given σ-fields $\mathcal{F}_1$ and $\mathcal{F}_2$, the smallest σ-field containing the union of these two fields is denoted by $\mathcal{F}_1 \vee \mathcal{F}_2$. For two measurable spaces $(\Omega_1, \mathcal{F}_1)$ and $(\Omega_2, \mathcal{F}_2)$, the product space is denoted by $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2)$. We say that $A \rightleftarrows B \rightleftarrows C$ forms a Markov chain under measure P, if A and C are conditionally

This work was presented at the IEEE Int. Symposium on Information Theory (ISIT), Barcelona, July 2016 [57]

independent given B under P. $P \ll Q$ denotes that the probability measure P is absolutely continuous with respect to the measure Q. $\mathbf{1}\{\mathsf{E}\}$ denotes the indicator function for an event E and $\log(x)$ is the natural logarithm of x. Convergence in probability and almost sure (a.s.) convergence are denoted by $\xrightarrow{\text{p}}$ and $\xrightarrow{\text{a.s.}}$, respectively. Throughout this paper we will adopt the convention that $0\log(0) = 0$, $\exp(\log(0)) = 0$, and $0^0 = 1$.

We will use the following form of Jensen's inequality.

Lemma 1 *If $\phi(x)$ is a convex function, then*

$$\mathbb{E}[\phi(X)] \geq \mathbb{E}[\phi(\mathbb{E}[X|A, B])] \geq \mathbb{E}[\phi(\mathbb{E}[X|A])] \geq \phi(\mathbb{E}[X]).$$

We now recall the definition of mutual information for general ensembles and its properties. Let A, B, and C be measurable mappings defined on a given probability space $(\Omega, \mathcal{F}, P)$, taking values in $(\mathcal{A}, \mathfrak{F}^A)$, $(\mathcal{B}, \mathfrak{F}^B)$, and $(\mathcal{C}, \mathfrak{F}^C)$ respectively. Consider partitions of Ω, $\mathfrak{Q}_A = \{\mathtt{A}_i, 1 \leq i \leq N_A\} \subseteq \sigma(A)$ and $\mathfrak{Q}_B = \{\mathtt{B}_j, 1 \leq j \leq N_B\} \subseteq \sigma(B)$. Wyner defined the conditional mutual information $I(A; B|C)$ as [67]

$$I(A; B|C) = \sup_{\mathfrak{Q}_A, \mathfrak{Q}_B} \mathbb{E}\left[\sum_{i,j=1,1}^{N_A, N_B} P(\mathtt{A}_i, \mathtt{B}_j|C) \log\left(\frac{P(\mathtt{A}_i, \mathtt{B}_j|C)}{P(\mathtt{A}_i|C)P(\mathtt{B}_j|C)}\right)\right], \tag{2.2}$$

where the supremum is over all such partitions of Ω. Wyner showed that $I(A; B|C) \geq 0$ with equality if and only if $A \rightleftarrows C \rightleftarrows B$ forms a Markov chain [67, Lemma 3.1], and that (generally referred to as) Kolmogrov's formula holds [67, Lemma 3.2]

$$I(A, C; B) = I(A; B) + I(C; B|A). \tag{2.3}$$

Hence if $I(A; B) < \infty$, then $I(C; B|A) = I(A, C; B) - I(A; B)$. The data processing inequality can be obtained from (3.3) as well: if $A \rightleftarrows C \rightleftarrows B$ forms a Markov chain, then $I(A;B) \leq I(C;B)$.

Denote by $P^{A,B}$, the joint distribution of A and B on the space $(\mathcal{A}\times\mathcal{B}, \mathfrak{F}^A\otimes\mathfrak{F}^B$), i.e.,

$$P^{A,B}(dA \times dB) = P((A^{-1}(dA), B^{-1}(dB)), \quad dA \in \mathfrak{F}^A, dB \in \mathfrak{F}^B.$$

Similarly, P^A and P^B denote the marginal distributions. Gelfand and Yaglom [27] proved that if $P^{A,B} \ll P^A \times P^B$, then the mutual information $I(A;B)$ (defined via (3.2) by taking $\sigma(C)$ to be the trivial σ-field) can be computed as

$$I(A;B) = \mathbb{E}\left[\log\left(\frac{dP^{A,B}}{d(P^A \times P^B)}\right)\right]. \tag{2.4}$$

A sufficient condition for $P^{A,B} \ll P^A \times P^B$ is that $I(A;B) < \infty$ [28, Lemma 5.2.3, p. 92]. We will also require the following result [67, Lemma 2.1]:

Lemma 2 (Wyner's Lemma) *If M is a finite alphabet random variable, then*

$$I(M;U_0^T) = H(M) - \mathbb{E}\left[H(M|U_0^T)\right],$$

where

$$H(M|U_0^T) = -\sum_m P(M=m|U_0^T)\log\left(P(M=m|U_0^T)\right),$$

and $H(M)$ is the entropy of M.

2.2 Doubly-Stochastic Poisson Process

Definition 1 *Let X_0^T be a non-negative process. A counting process N_0^T is called* a doubly-stochastic Poisson process with rate process X_0^T *under measure P if*

- *for an interval* $[s,t] \in [0,T]$

$$P(N_t - N_s = k | X_0^T) = \frac{1}{k!}\left(\int_s^t X_\tau \, d\tau\right)^k \exp\left(-\int_s^t X_\tau \, d\tau\right), \text{ for } k \in \mathbf{N}$$

with convention $0^0 = 1$,

- *conditioned on* X_0^T *the increments in disjoint intervals of* $[0,T]$ *are independent.*

Throughout this paper, the rate process X_0^T will be a bounded càdlàg (right continuous with left limits) process.

Definition 2 *If* N_0^T *is a counting process adapted to the history* $(\mathcal{F}_t : t \in [0,T])$, *then* N_0^T *is said to have* $(P, \mathcal{F}_t : t \in [0,T])$-intensity $\Gamma_0^T = \{\Gamma_t, 0 \leq t \leq T\}$, *where* Γ_0^T *is a non-negative measurable process if*

- Γ_0^T *is* $(\mathcal{F}_t : t \in [0,T])$*-predictable,*
- $\int_0^T \Gamma_t \, dt < \infty$, *P-a.s.,*
- *and for all non-negative* $(\mathcal{F}_t : t \in [0,T])$*-predictable processes* C_0^T:[1]

$$\mathbb{E}\left[\int_0^T C_s \, dN_s\right] = \mathbb{E}\left[\int_0^T C_s \Gamma_s \, ds\right].$$

Definition 3 *Given a doubly-stochastic Poisson process* N_0^T, *a counting process* $\tilde{N}_0^T$ *is called* the time-reversed N_0^T process *if* $\tilde{N}_0 = 0$ *and for* $t \in (0,T]$, $\tilde{N}_t = N_T - N_{(T-t)-}$.

Definition 4 *Fix* $0 \leq t_1 < t_2 \leq T$. *Given a doubly-stochastic Poisson process* N_0^T, $N_{t_1}^{t_2}$ *will denote a point process on* $[0,T]$ *which has no arrival before* t_1, *after*

[1]The limits of the Lebesgue-Stieltjes integral $\int_a^b$ are to be interpreted as $\int_{(a,b]}$.

t_2, and the same arrivals as process N_0^T on the interval $[t_1, t_2]$. Specifically, let $\hat{N}_t$ denote the value of the process $N_{t_1}^{t_2}$ at time t. Then

$$\begin{aligned} \hat{N}_t &= 0, \quad t < t_1, \\ &= N_t - N_{t_1}, \quad t_1 \le t \le t_2, \\ &= N_{t_2} - N_{t_1}, \quad t_2 < t \le T. \end{aligned}$$

Lemma 3 *Suppose N_0^T is a doubly-stochastic Poisson process with rate process X_0^T under measure P and $\tilde{N}_0^T$ is the time-reversed N_0^T process. Then $\tilde{N}_0^T$ is a doubly-stochastic Poisson process with rate process $\tilde{X}_0^T = \left\{\tilde{X}_t = X_{(T-t)-} : t \in [0,T]\right\}$ under measure P.*

Proof: See Section 2.8. □

Lemma 4 *Suppose N_0^T is a doubly-stochastic Poisson process with rate process Λ_0^T under measure P and $A \rightleftarrows \Lambda_0^T \rightleftarrows N_{t_1}^{t_2}$ is a Markov chain. Let $\hat{N}_0^T = \{\hat{N}_t : t \in [0,T]\}$, where $\hat{N}_t$ is the value of $N_{t_1}^{t_2}$ at time $t \in [0,T]$, i.e., the process $\hat{N}_0^T$ has no arrivals prior to t_1 and after t_2 and the same arrivals instants as process N_0^T for $t \in [t_1, t_2]$. Then for $\mathcal{F}_t = \sigma(A) \vee \mathcal{F}_T^{\Lambda} \vee \mathcal{F}_t^{\hat{N}}$, the $(P, \mathcal{F}_t : t \in [0,T])$-intensity of N_0^T is $\hat{\Lambda}_0^T = \left\{\hat{\Lambda}_t = \mathbf{1}\{t_1 \le t \le t_2\}\Lambda_t, t \in [0,T]\right\}$. Also, for $\mathcal{G}_t = \sigma(A) \vee \mathcal{F}_t^{\hat{N}}$, there exists a $(\mathcal{G}_t : t \in [0,T])$-predictable process Π_0^T such that Π_0^T is the $(P, \mathcal{G}_t : t \in [0,T])$-intensity of $\hat{N}_0^T$ and $\Pi_t = \mathbb{E}[\hat{\Lambda}_t | \mathcal{G}_t]$ P-a.s. for each $t \in [0,T]$.*

Proof: See Section 2.8. □

2.3 Channel Model

The two-user Poisson Channel considered here consists of an encoder $\mathscr{E}_x^T$ and two decoders $\mathscr{D}_y^T$ and $\mathscr{D}_z^T$. Let $\mathcal{X}_0^T$ denote the set of all waveforms over $[0,T]$ which are non-negative, right continuous with left limits, and peak power limited by unity. This is the set of inputs to the channel, i.e., $X_0^T = \{X_t, 0 \leq X_t \leq 1, t \in [0,T]\}$. The received signal at the first receiver Y_0^T is a doubly-stochastic Poisson process with rate $a_y X_0^T + \lambda_y$. Here $a_y \geq 0$ accounts for possible attenuation of the signal at the first receiver and $\lambda_y \geq 0$ is the dark current intensity due to background noise and is independent of the input process X_0^T. Similarly the received signal at the second receiver is Z_0^T, where Z_0^T is a doubly-stochastic Poisson process with rate $a_z X_0^T + \lambda_z$ with $a_z, \lambda_z \geq 0$.

Let $(\mathcal{X}_0^T, \mathfrak{F}^X)$ denote the input space, where $\mathfrak{F}^X$ is the σ-field on $\mathcal{X}_0^T$ generated by the open sets of $\mathcal{X}_0^T$ when endowed with the Skorohod topology [7, Chapter 3, Section 12, p. 121]. Similarly, let $(\mathcal{N}_0^T, \mathfrak{F}^Y)$ and $(\mathcal{N}_0^T, \mathfrak{F}^Z)$ be the first and second receiver's output space respectively, where $\mathfrak{F}^Y$ and $\mathfrak{F}^Z$ are the σ-field generated by the open sets of $\mathcal{N}_0^T$ when endowed with the Skorohod topology. Let $P_0^{Y_0^T}$ (respectively $P_0^{Z_0^T}$) be the probability measure on the first receiver's (respectively second receiver's) output space such that point process Y_0^T (respectively Z_0^T) is a unit-rate Poisson process. Then we will take the output space of the channel to be the product space $(\mathcal{N}_0^T \times \mathcal{N}_0^T, \mathfrak{F}^Y \otimes \mathfrak{F}^Z)$ and our reference measure P_0 will be the product measure $P_0 = P_0^{Y_0^T} \times P_0^{Z_0^T}$. Fix $x_0^T \in \mathcal{X}_0^T$, and let $\Xi_{x_0^T}(\cdot)$ denote the transition probability function from the input space $(\mathcal{X}_0^T, \mathfrak{F}^X)$ to the output space $(\mathcal{N}_0^T \times \mathcal{N}_0^T, \mathfrak{F}^Y \otimes \mathfrak{F}^Z)$. The channel is modeled through the following Radon-Nikodym derivative:

$$\frac{d\Xi_{x_0^T}}{dP_0}(y_0^T, z_0^T) = \prod_{u=y,z} p_u(x_0^T, u_0^T), \tag{2.5}$$

where

$$p_u(x_0^T, u_0^T) = \exp\left(\int_0^T \log(a_u x_t + \lambda_u)\, du_t + 1 - (a_u x_t + \lambda_u)\, dt\right), \tag{2.6}$$

where we recall the convention $\exp(\log(0)) = 0$. Then due to Girsanov's theorems [9, Chapter VI, Theorems T2-T4, p. 165-168], the process U_0^T has $(\mathcal{F}_t^U : t \in [0,T])$-intensity $a_u x_0^T + \lambda_u$ under probability measure $\Xi_{x_0^T}$ for $(u, U) \in \{(y, Y), (z, Z)\}$. Note that the above model implies that for given $x_0^T \in \mathcal{X}_0^T$, processes Y_0^T and Z_0^T are independent doubly-stochastic Poisson processes with rate processes $a_y x_0^T + \lambda_y$ and $a_z x_0^T + \lambda_z$ respectively [9, Theorem T4, Chapter II, p. 25].

Let M be a random variable on a measurable space $(\mathcal{M}, \mathfrak{F}^M)$. For the most part of this paper M will represent a message intended for either or both of the users, in which case $\mathcal{M}$ is a finite set and we will take $\mathfrak{F}^M$ to be the power set of $\mathcal{M}$. However, in proving Theorem 3 to follow, we will take the space $(\mathcal{M}, \mathfrak{F}^M)$ to be isomorphic to the input space $(\mathcal{X}_0^T, \mathfrak{F}^X)$. Let $\mu_m(dx_0^T)$ denote the transition probability function from $(\mathcal{M}, \mathfrak{F}^M)$ to the input space $(\mathcal{X}_0^T, \mathfrak{F}^X)$. Let $\nu(dm)$ be a probability measure on $(\mathcal{M}, \mathfrak{F}^M)$. Then these measures induce a joint measure P on $(\Omega, \mathcal{F})$, where

$$\Omega = \mathcal{M} \times \mathcal{X}_0^T \times \mathcal{N}_0^T \times \mathcal{N}_0^T$$
$$\mathcal{F} = \mathfrak{F}^M \otimes \mathfrak{F}^X \otimes \mathfrak{F}^Y \otimes \mathfrak{F}^Z$$
$$P = \nu(dm)\mu_m(dx_0^T) P_0^{Y_0^T}(dy_0^T) P_0^{Z_0^T}(dz_0^T) \prod_{u=y,z} p_u(x_0^T, u_0^T). \tag{2.7}$$

From (2.7), we have $M \rightleftarrows X_0^T \rightleftarrows (Y_0^T, Z_0^T)$ and $Y_0^T \rightleftarrows X_0^T \rightleftarrows Z_0^T$ forming a Markov chain under P. These Markov chain structures will play a triple role in the upcoming analysis. First, the former implies the finiteness of mutual information quantities (and hence absolute continuity of measures) of the form $I(A; U_{t_1}^{t_2})$ for $U \in \{Y, Z\}$, where $A \rightleftarrows X_0^T \rightleftarrows U_{t_1}^{t_2}$ is a Markov chain (see Lemma 5). Second, the former allows us compute the log-likelihood ratio martingales through the intensity of the point process $U_{t_1}^{t_2}$ (see Theorem 1). Finally, the latter coupling is useful for proving impossibility results (cf. Theorem 2 to follow). The capacity regions defined subsequently, however, only depend on the two marginal distributions of Y_0^T and Z_0^T given X_0^T. Thus our capacity results hold for any channels for which Y_0^T and Z_0^T are Poisson processes with rate $a_y X_0^T + \lambda_y$ and $a_z X_0^T + \lambda_z$, respectively.

We will assume that the given filtration $(\mathcal{F}_t : t \in [0,T])$, P, and $\mathcal{F}$ satisfy the "usual conditions" [9, Chapter III, p. 75]: $\mathcal{F}$ is complete with respect to P, $\mathcal{F}_t$ is right continuous, and $\mathcal{F}_0$ contains all the P-null sets of $\mathcal{F}_t$.

In the rest of this paper we will consider mappings A and B from Ω in (2.7) to a component space $\mathcal{N}_0^T$ or $\mathcal{M}$ of Ω: A can be M itself, or A can be a portion of arrival time process Y_0^T or Z_0^T on the interval $[s_1, s_2]$, which we model as a point process on $\mathcal{N}_0^T$ with no arrival prior to s_1 and after s_2. Fix $0 \le t_1 < t_2 \le T$ and consider the process $U_{t_1}^{t_2}$. Denote by $\hat{U}_t$ its value at time $t \in [0,T]$. Let $\hat{U}_0^T = \{\hat{U}_t : t \in [0,T]\}$. Note that $U_{t_1}^{t_2}$ and $\hat{U}_0^T$ are exactly the same process, but we use $\hat{U}_0^T$ for notational convenience. We will use the following condition to verify that the mutual information $I(A; U_{t_1}^{t_2})$ is finite.

Lemma 5 *If A is such that $A \rightleftarrows X_0^T \rightleftarrows U_{t_1}^{t_2}$ forms a Markov chain under measure P, then with $\hat{U}_0^T = \{\hat{U}_t : t \in [0,T]\}$, where $\hat{U}_t$ is the value of $U_{t_1}^{t_2}$ at time $t \in [0,T]$*

$$I(A; U_{t_1}^{t_2}) < \infty,$$

and thus

$$P^{A,\hat{U}_0^T} \ll P^A \times P^{\hat{U}_0^T} \ll P^A \times P_0^{\hat{U}_0^T},$$

where $P_0^{\hat{U}_0^T}$ is the distribution of process $\hat{U}_0^T$ under the measure $P_0^{U_0^T}$.

Proof: See Section 2.8. □

In particular the above lemma implies that if $(A, B) \rightleftarrows X_0^T \rightleftarrows U_{t_1}^{t_2}$ is a Markov chain, then $I(A; U_{t_1}^{t_2})$ and $I(A; U_{t_1}^{t_2}|B)$ are finite. The mutual information expressions considered in the sequel will be of this form. The following theorem provides a way of computing such expressions. It will be applied repeatedly in the later sections.

Theorem 1 (Log Radon-Nikodym derivatives and mutual information expression for

Fix $0 \leq t_1 < t_2 \leq T$*, and let* $(u, U) \in \{(y, Y), (z, Z)\}$*.*

1. *Log Radon-Nikodym derivatives:*

   Let $A \rightleftarrows X_0^T \rightleftarrows U_{t_1}^{t_2}$ *be a Markov chain. Denote by* $\hat{U}_t$ *the value of* $U_{t_1}^{t_2}$ *at time* $t \in [0, T]$*. Let* $\hat{U}_0^T = \{\hat{U}_t : t \in [0, T]\}$*. Let* $\tilde{P}^{A, \hat{U}_0^T} = P^A \times P_0^{\hat{U}_0^T}$*. From Lemma 5,* $P^{A, \hat{U}_0^T} \ll \tilde{P}^{A, \hat{U}_0^T}$*. Then*

$$\log\left(\frac{dP^{A,\hat{U}_0^T}}{d\tilde{P}^{A,\hat{U}_0^T}}\right) = \int_{t_1}^{t_2} \log(a_u \Pi_t + \lambda_u) dU_t + 1 - (a_u \Pi_t + \lambda_u)\, dt, \tag{2.8}$$

   where the above equality is $P^{A,\hat{U}_0^T}$*-a.s., and* Π_0^T *is a* $(\sigma(A) \vee \mathcal{F}_t^{\hat{U}}, t \in [0, T])$*-predictable process satisfying for each* $t \in [t_1, t_2]$*,*

$$\Pi_t = \mathbb{E}[X_t | A, U_{t_1}^t], \quad P^{A,\hat{U}_0^T}\text{-a.s.}$$ [2]

2. *Mutual Information Expressions:*

[2] Here we have abused notation slightly since this random variable will be defined on a larger probability space in the proof.

Suppose that the Markov chain $(A,B) \rightleftarrows X_0^T \rightleftarrows U_{t_1}^{t_2}$ *holds. Then*

$$\begin{aligned} I\left(A;U_{t_1}^{t_2}\middle|B\right) &= \int_{t_1}^{t_2} \mathbb{E}[\phi_u(\mathbb{E}[X_t|U_{t_1}^t,A,B])] - \mathbb{E}[\phi_u(\mathbb{E}[X_t|U_{t_1}^t,B])]\,dt \\ &= \int_{t_1}^{t_2} \mathbb{E}[\phi_u(\mathbb{E}[X_{t-}|U_t^{t_2},A,B])] - \mathbb{E}[\phi_u(\mathbb{E}[X_{t-}|U_t^{t_2},B])]\,dt \\ &= \int_{t_1}^{t_2} \mathbb{E}[\phi_u(\mathbb{E}[X_t|U_t^{t_2},A,B])] - \mathbb{E}[\phi_u(\mathbb{E}[X_t|U_t^{t_2},B])]\,dt, \end{aligned}$$

where for $u \in \{y,z\}$ *we define*

$$\phi_u(x) = (a_u x + \lambda_u)\log(a_u x + \lambda_u),$$

with convention that $0\log(0) = 0$*. Note that* $\phi_u(x)$ *is convex and continuous for* $x \in [0,1]$*.*

If $A = X_0^T$, then the identity (2.8) is true by definition (cf. (2.6)). It is also known when A is independent of X_0^T [9, (5.6), p. 181]. Those two cases suffice to compute the quantities $I(X_0^T;Y_0^T)$ and $I(X_0^T;Z_0^T)$. By allowing for arbitrary A in (2.8), we can compute mutual information expressions involving auxiliary random variables, which are needed for multiterminal problems.

Proof: We will consider the measurable space $(\mathcal{A}\times\mathcal{X}_0^T\times\mathcal{N}_0^T, \mathfrak{F}^A\otimes\mathfrak{F}^X\otimes\mathfrak{F}^{\tilde{U}})$. Here $\mathcal{A}$ is the set on which A takes values and $\mathfrak{F}^A$ is its σ-field. Let $\tilde{P}^{A,X_0^T,\hat{U}_0^T}$ be defined as

$$\tilde{P}^{A,X_0^T,\hat{U}_0^T} = P^{A,X_0^T} \times P_0^{\hat{U}_0^T},$$

i.e., under $\tilde{P}^{A,X_0^T,\hat{U}_0^T}$, $\hat{U}_0^T$ is a Poisson process with deterministic rate μ_0^T, independent of A and X_0^T, where

$$\mu_t = \mathbf{1}\{t_1 \le t < t_2\}.$$

Let $\mathcal{G}_t = \mathcal{F}_t^{\hat{U}} \vee \sigma(A)$. Since under $\tilde{P}^{A,X_0^T,\hat{U}_0^T}$, A is independent of $\hat{U}_0^T$, using Lemma 4 we conclude that the $(\tilde{P}^{A,X_0^T,\hat{U}_0^T}, \mathcal{G}_t : t \in [0,T])$-intensity of $\hat{U}_0^T$ is μ_0^T.

Since $I(A, X_0^T; \hat{U}_0^T) = I(X_0^T; \hat{U}_0^T) < \infty$, we have that $P^{A,X_0^T,\hat{U}_0^T} \ll P^{A,X_0^T} \times P^{\hat{U}_0^T}$ [28, Lemma 5.2.3, p. 92]. Using the fact that $P^{\hat{U}_0^T} \ll P_0^{\hat{U}_0^T}$ we get [31, Chapter 1, Exercise 19, p. 22]

$$P^{A,X_0^T,\hat{U}_0^T} \ll \tilde{P}^{A,X_0^T,\hat{U}_0^T}.$$

Let

$$\mathcal{L} = \frac{dP^{A,X_0^T,\hat{U}_0^T}}{d\tilde{P}^{A,X_0^T,\hat{U}_0^T}}$$

denote the Radon-Nikodym derivative on the space $(\mathcal{A} \times \mathcal{X}_0^T \times \mathcal{N}_0^T, \mathfrak{F}^A \otimes \mathfrak{F}^X \otimes \mathfrak{F}^{\tilde{U}})$. Consider the mapping $(a, x_0^T, \hat{u}_0^T) \to (a, \hat{u}_0^T)$ from $(\mathcal{A} \times \mathcal{X}_0^T \times \mathcal{N}_0^T)$ to $(\mathcal{A} \times \mathcal{N}_0^T)$. Since $\sigma(A, U_0^{\hat{T}}) = \mathcal{G}_T$, $\frac{dP^{A,\hat{U}_0^T}}{d\tilde{P}^{A,\hat{U}_0^T}}$ can be computed as [28, Lemma 5.2.4, p. 96]

$$\frac{dP^{A,\hat{U}_0^T}}{d\tilde{P}^{A,\hat{U}_0^T}} = \mathbb{E}_{\tilde{P}}[\mathcal{L}|\mathcal{G}_T].$$

Here the subscript $\tilde{P}$ indicates that the expectation is taken with respect to $\tilde{P}^{A,X_0^T,\hat{U}_0^T}$. Towards this end define process L_0^T as

$$L_t = \mathbb{E}_{\tilde{P}}[\mathcal{L}|\mathcal{G}_t], \quad t \in [0, T].$$

Then L_0^T is a $(\tilde{P}^{A,X_0^T,\hat{U}_0^T}, \mathcal{G}_t)$ non-negative absolutely-integrable martingale.

By the martingale representation theorem, the process L_0^T can be written as [9, Chapter III, Theorem T17, p. 76] (where we have taken $\sigma(A)$ to be the "germ σ-field"):

$$L_t = 1 + \int_0^t K_s(d\hat{U}_s - \mu_s ds),$$

where K_0^T is a $(\mathcal{G}_t : t \in [0, T])$-predictable process which satisfies $\int_0^T |K_t|\mu_t \, dt < \infty$ $\tilde{P}^{A,X_0^T,\hat{U}_0^T}$-a.s. Applying [43, Lemma 19.5, p. 315], we can write L_t as

$$L_t = \exp\left(\int_0^t \log(\Psi_s) d\hat{U}_s + (1 - \Psi_s)\mu_s \, ds\right) \quad t \in [0, T], \tag{2.9}$$

where Ψ_0^T is a non-negative $(\mathcal{G}_t : t \in [0,T])$-predictable process, and $\Psi_t < \infty$ $\tilde{P}^{A,X_0^T,\hat{U}_0^T}$-a.s. for $t \in [0,T]$. Let

$$\hat{\Psi}_t = \Psi_t \mu_t = \mathbf{1}\{t_1 \le t < t_2\} \Psi_t.$$

Since the candidate intensity $\hat{\Psi}_0^T$ is not known to satisfy $\int_0^T \hat{\Psi}_t \, dt < \infty$, we cannot apply [9, Chapter VI, Theorems T2-T3, p. 166] directly. Instead, we first mimic the proof of [9, Chapter VI, Theorem T3, p. 166] to get following result.

Lemma 6 *For all non-negative $(\mathcal{G}_t : t \in [0,T])$-predictable processes C_0^T*

$$\mathbb{E}\left[\int_0^T C_t \hat{\Psi}_t \, dt\right] = \mathbb{E}\left[\int_0^T C_t \, d\hat{U}_t\right],$$

where the above expectation is with respect to the measure $P^{A,X_0^T,\hat{U}_0^T}$.

Proof: See Section 2.8. □

Taking $C_t = 1$ in the above equality yields

$$\mathbb{E}\left[\int_0^T \hat{\Psi}_t \, dt\right] = \mathbb{E}\left[\int_0^T d\hat{U}_t\right] < \infty.$$

Hence $\int_0^T \hat{\Psi}_t \, dt < \infty$ $P^{A,X_0^T,\hat{U}_0^T}$-a.s. and we conclude that the $(P^{A,X_0^T,\hat{U}_0^T}, \mathcal{G}_t : t \in [0,T])$-intensity of $\hat{U}_0^T$ is $\hat{\Psi}_0^T$.

Moreover due to uniqueness of predictable intensities [9, Theorem T12, Chapter II, p. 31], from Lemma 4, we can take for $t_1 \le t \le t_2$ $P^{A,X_0^T,\hat{U}_0^t}$-a.s.

$$\Psi_t = a_u \Pi_t + \lambda_u, \tag{2.10}$$

where for each $t \in [t_1, t_2]$,

$$\Pi_t = \mathbb{E}[X_t | A, \hat{U}_0^t]. \tag{2.11}$$

Noting that process $\hat{U}_0^T$ has no arrivals prior to t_1 and later than t_2, and the same arrivals as U_0^T between t_1 and t_2, substituting value of Ψ_t from (2.10), (3.7) yields

$$\log\left(\frac{dP^{A,\hat{U}_0^T}}{d\tilde{P}^{A,\hat{U}_0^T}}\right) = \log(L_T)$$
$$= \int_{t_1}^{t_2} \log(a_u\Pi_t + \lambda_u)d\,U_t + 1 - (a_u\Pi_t + \lambda_u)\,dt, \qquad (2.12)$$

where $\Pi_t = \mathbb{E}[X_t|A, U_{t_1}^t]$ $P^{A,\hat{U}_0^T}$-a.s. for each $t \in [t_1, t_2]$. This proves part (1) of the theorem.

Writing (2.12) in terms of Ψ_t, we get

$$\log\left(\frac{dP^{A,\hat{U}_0^T}}{d\tilde{P}^{A,\hat{U}_0^T}}\right) = \int_0^T \log(\Psi_t)d\hat{U}_t + (1-\Psi_t)\mu_t\,dt, \qquad (2.13)$$

and recalling that Ψ_0^T is $(\mathcal{G}_t : t \in [0,T])$-predictable

$$\mathbb{E}\left[\log\left(\frac{dP^{A,\hat{U}_0^T}}{d\tilde{P}^{A,\hat{U}_0^T}}\right)\right] = \mathbb{E}\left[\int_0^T \log(\Psi_t)d\hat{U}_t\right] + \int_0^T (1-\mathbb{E}[\Psi_t])\mu_t\,dt$$
$$= \mathbb{E}\left[\int_0^T \log(\Psi_t)\Psi_t\mu_t\,dt\right] + \int_{t_1}^{t_2} 1-\mathbb{E}[\Psi_t]\,dt$$
$$= \int_{t_1}^{t_2} \mathbb{E}[\Psi_t\log(\Psi_t)] + 1 - \mathbb{E}[\Psi_t]\,dt$$
$$= \int_{t_1}^{t_2} \mathbb{E}\left[\log(a_u\mathbb{E}[X_t|A,U_{t_1}^t] + \lambda_u)(a_u\mathbb{E}[X_t|A,U_{t_1}^t] + \lambda_u)\right] + 1 - (a_u\mathbb{E}[X_t] + \lambda$$
$$= \int_{t_1}^{t_2} \mathbb{E}\left[\phi_u\left(\mathbb{E}[X_t|A,U_{t_1}^t]\right)\right] + 1 - (a_u\mathbb{E}[X_t] + \lambda_u)\,dt. \quad (2.14)$$

Similarly

$$\mathbb{E}\left[\log\left(\frac{dP^{\hat{U}_0^T}}{dP_0^{\hat{U}_0^T}}\right)\right] = \int_{t_1}^{t_2} \mathbb{E}\left[\phi_u\left(\mathbb{E}[X_t|U_{t_1}^t]\right)\right] + 1 - (a_u\mathbb{E}[X_t] + \lambda_u)\,dt. \qquad (2.15)$$

Using (3.4) and Lemma 5 we can compute the mutual information expression

$$I(A; U_{t_1}^{t_2}) = I(A; \hat{U}_0^T)$$

$$
\begin{aligned}
&= \mathbb{E}\left[\log\left(\frac{dP^{A,\hat{U}_0^T}}{d(P^A \times P^{\hat{U}_0^T})}\right)\right] \\
&= \mathbb{E}\left[\log\left(\frac{dP^{A,\hat{U}_0^T}/d\tilde{P}^{A,\hat{U}_0^T}}{d(P^A \times P^{\hat{U}_0^T})/d\tilde{P}^{A,\hat{U}_0^T}}\right)\right] \\
&= \mathbb{E}\left[\log\left(\frac{dP^{A,\hat{U}_0^T}/d\tilde{P}^{A,\hat{U}_0^T}}{dP^{\hat{U}_0^T}/dP_0^{\hat{U}_0^T}}\right)\right] \\
&= \mathbb{E}\left[\log\left(\frac{dP^{A,\hat{U}_0^T}}{d\tilde{P}^{A,\hat{U}_0^T}}\right)\right] - \mathbb{E}\left[\log\left(\frac{dP^{\hat{U}_0^T}}{dP_0^{\hat{U}_0^T}}\right)\right] \\
&= \int_{t_1}^{t_2} \mathbb{E}[\phi_u(\mathbb{E}[X_t|U_{t_1}^t, A])] - \mathbb{E}[\phi_u(\mathbb{E}[X_t|U_{t_1}^t])]\, dt. \qquad (2.16)
\end{aligned}
$$

Now we use Kolmogorov's formula and the fact that all the mutual information expressions are finite due to Lemma 5:

$$
\begin{aligned}
I(A; U_{t_1}^{t_2}|B) =& I(A, B; U_{t_1}^{t_2}) - I(B; U_{t_1}^{t_2}) \\
=& \int_{t_1}^{t_2} \mathbb{E}[\phi_u(\mathbb{E}[X_t|U_{t_1}^t, A, B])] - \mathbb{E}[\phi_u(\mathbb{E}[X_t|U_{t_1}^t])]\, dt \\
&- \int_{t_1}^{t_2} \mathbb{E}[\phi_u(\mathbb{E}[X_t|U_{t_1}^t, B])] - \mathbb{E}[\phi_u(\mathbb{E}[X_t|U_{t_1}^t])]\, dt \\
=& \int_{t_1}^{t_2} \mathbb{E}[\phi_u(\mathbb{E}[X_t|U_{t_1}^t, A, B])] - \mathbb{E}[\phi_u(\mathbb{E}[X_t|U_{t_1}^t, B])]\, dt. \qquad (2.17)
\end{aligned}
$$

Now define a new point process $\tilde{U}_0^T$ as the time-reversed version of the process $\hat{U}_0^T$. From Lemma 3, $\tilde{U}_0^T$ is a doubly-stochastic Poisson process with rate process

$$
\tilde{\Lambda}_0^T = \{(a_u\tilde{X}_t + \lambda_u)\mathbf{1}\{T - t_2 \le t < T - t_1\}, t \in [0, T]\},
$$

where $\tilde{X}_t = X_{(T-t)-}$. Let $\tilde{U}_t$ denote the value of process $\tilde{U}_0^T$. Then

$$
\begin{aligned}
I(A; U_{t_1}^{t_2}|B) &= I(A; \hat{U}_0^T|B) \\
&= I(A; \tilde{U}_0^T|B) \\
&= \int_{T-t_2}^{T-t_1} \mathbb{E}[\phi_u(\mathbb{E}[\tilde{X}_s|\tilde{U}_{T-t_2}^s, A, B])] - \mathbb{E}[\phi_u(\mathbb{E}[\tilde{X}_s|\tilde{U}_{T-t_2}^s, B])]\, ds
\end{aligned}
$$

$$= \int_{T-t_2}^{T-t_1} \mathbb{E}[\phi_u(\mathbb{E}[X_{(T-s)-}|U_{T-s}^{t_2}, A, B])] - \mathbb{E}[\phi_u(\mathbb{E}[X_{(T-s)-}|U_{T-s}^{t_2}, B])]\, ds$$

$$= \int_{t_1}^{t_2} \mathbb{E}[\phi_u(\mathbb{E}[X_{t-}|U_t^{t_2}, A, B])] - \mathbb{E}[\phi_u(\mathbb{E}[X_{t-}|U_t^{t_2}, B])]\, dt. \tag{2.18}$$

Note that since a càdlàg process can have at most countably many jumps over a bounded interval $[t_1, t_2]$ [7, Section 12, Lemma 1, p. 122], we have

$$\int_{t_1}^{t_2} \mathbf{1}\{X_{t-} \neq X_t\} = 0.$$

Taking expectation and using Fubini's theorem

$$\frac{1}{t_2 - t_1} \int_{t_1}^{t_2} P(X_{t-} \neq X_t) = 0.$$

Thus

$$P(X_{S-} = X_S) = 1, \tag{2.19}$$

where we have defined S to be a random variable uniformly distributed over $[t_1, t_2]$ and independent of all other σ-fields. We can then write $I(A; U_{t_1}^{t_2}|B)$ as

$$\begin{aligned} I(A; U_{t_1}^{t_2}|B) &= \int_{t_1}^{t_2} \mathbb{E}[\phi_u(\mathbb{E}[X_{t-}|U_t^{t_2}, A, B])] - \mathbb{E}[\phi_u(\mathbb{E}[X_{t-}|U_t^{t_2}, B])]\, dt \\ &= (t_2 - t_1)\mathbb{E}[\phi_u(\mathbb{E}[X_{S-}|U_S^{t_2}, A, B])] - \mathbb{E}[\phi_u(\mathbb{E}[X_{S-}|U_S^{t_2}, B])] \\ &\overset{(a)}{=} (t_2 - t_1)\mathbb{E}[\phi_u(\mathbb{E}[X_S|U_S^{t_2}, A, B])] - \mathbb{E}[\phi_u(\mathbb{E}[X_S|U_S^{t_2}, B])] \\ &= \int_{t_1}^{t_2} \mathbb{E}[\phi_u(\mathbb{E}[X_t|U_t^{t_2}, A, B])] - \mathbb{E}[\phi_u(\mathbb{E}[X_t|U_t^{t_2}, B])]\, dt, \end{aligned}$$

where for (a) we have used (2.19). This completes the proof of part (2) of the theorem. □

We now derive some properties of $I(A; U_0^T|B)$.

Lemma 7 *If $(A, B) \rightleftarrows X_0^T \rightleftarrows U_0^T$ is a Markov chain, then*

$$\lim_{\delta \to 0^+} \frac{1}{\delta} I\left(A; U_t^{t+\delta} \middle| U_0^t, B\right) = \mathbb{E}[\phi_u(\mathbb{E}[X_t|U_0^t, A, B])] - \mathbb{E}[\phi_u(\mathbb{E}[X_t|U_0^t, B])]$$

and

$$\lim_{\delta\to 0^+}\frac{1}{\delta}I\left(A;U_{t-\delta}^t\middle|U_t^T,B\right)=\mathbb{E}[\phi_u(\mathbb{E}[X_{t-}|U_t^T,A,B])]-\mathbb{E}[\phi_u(\mathbb{E}[X_{t-}|U_t^T,B])].$$

Proof: See Section 2.8. □

Lemma 8 *If A and B are such that $(A,B)\rightleftarrows X_0^T\rightleftarrows U_0^T$ is a Markov chain, then both $\frac{1}{\delta}I\left(A;U_s^{s+\delta}\middle|U_0^s,B\right)$ and $\frac{1}{\delta}I\left(A;U_{s-\delta}^s\middle|U_s^T,B\right)$ are bounded uniformly over s and $\delta>0$.*

Proof: See Section 2.8. □

Combining Lemmas 7 and 8 yields the chain rule for mutual information in continuous time.

Lemma 9 *If $(A,B)\rightleftarrows X_0^T\rightleftarrows U_0^T$ is a Markov chain, then*

$$I(A;U_0^t|B)=\lim_{\delta\to 0^+}\frac{1}{\delta}\int_0^t I\left(A;U_s^{s+\delta}\middle|U_0^s,B\right)\,ds,$$
$$I(A;U_t^T|B)=\lim_{\delta\to 0^+}\frac{1}{\delta}\int_t^T I\left(A;U_{s-\delta}^s\middle|U_s^T,B\right)\,ds.$$

Proof: See Section 2.8. □

We now prove an identity which parallels the Csiszár sum identity [14] for discrete memoryless channels.

Theorem 2 (Csiszár-sum-like-identity for Poisson channel) *With the channel model in (2.7):*

$$\lim_{\epsilon\to 0^+}\int_0^T\frac{1}{\epsilon}I\left(Z_{t-\epsilon}^t;Y_0^t\middle|Z_t^T,M\right)\,dt=\lim_{\epsilon\to 0^+}\int_0^T\frac{1}{\epsilon}I\left(Y_t^{t+\epsilon};Z_t^T\middle|Y_0^t,M\right)\,dt,\qquad(2.20)$$

where we take $U_s^{t_2} = U_0^{t_2}$ *if* $s < 0$, *and* $U_{t_1}^s = U_{t_1}^T$ *if* $s > T$. *This implies*

$$\int_0^T \mathbb{E}[\phi_y(\mathbb{E}[X_t|Y_0^t, M])] - \mathbb{E}[\phi_z(\mathbb{E}[X_t|Z_t^T, M])] \, dt =$$
$$\int_0^T \mathbb{E}[\phi_y(\mathbb{E}[X_t|Y_0^t, Z_t^T, M])] - \mathbb{E}[\phi_z(\mathbb{E}[X_t|Y_0^t, Z_t^T, M])] \, dt. \tag{2.21}$$

Proof: Noting that since $(M, Z_0^T) \rightleftarrows X_0^T \rightleftarrows Y_0^T$ is a Markov chain, the mutual information expressions considered below are finite. Using [67, Lemma 3.3] we get

$$\begin{aligned}\int_0^T I\left(Z_{t-\epsilon}^t; Y_0^t \middle| Z_t^T, M\right) \, dt &= \int_0^T I\left(Z_{t-\epsilon}^t, Z_t^T; Y_0^t \middle| M\right) - I\left(Z_t^T; Y_0^t \middle| M\right) \, dt \\ &= \int_0^T I\left(Z_{t-\epsilon}^T; Y_0^t \middle| M\right) \, dt - \int_0^T I\left(Z_t^T; Y_0^t \middle| M\right) \, dt.\end{aligned} \tag{2.22}$$

Similarly,

$$\begin{aligned}\int_0^T I\left(Y_t^{t+\epsilon}; Z_t^T \middle| Y_0^t, M\right) \, dt &= \int_0^T I\left(Y_0^{t+\epsilon}; Z_t^T \middle| M\right) \, dt - \int_0^T I\left(Y_0^t; Z_t^T \middle| M\right) \, dt \\ &= \int_\epsilon^{T+\epsilon} I\left(Y_0^t; Z_{t-\epsilon}^T \middle| M\right) \, dt - \int_0^T I\left(Y_0^t; Z_t^T \middle| M\right) \, dt.\end{aligned} \tag{2.23}$$

From (2.22) and (2.23),

$$\begin{aligned}&\int_0^T \frac{1}{\epsilon} I\left(Z_{t-\epsilon}^t; Y_0^t \middle| Z_t^T, M\right) \, dt - \int_0^T \frac{1}{\epsilon} I\left(Y_t^{t+\epsilon}; Z_t^T \middle| Y_0^t, M\right) \, dt \\ &\quad = \int_0^T \frac{1}{\epsilon} I\left(Z_{t-\epsilon}^T; Y_0^t \middle| M\right) \, dt - \int_\epsilon^{T+\epsilon} \frac{1}{\epsilon} I\left(Y_0^t; Z_{t-\epsilon}^T \middle| M\right) \, dt \\ &\quad = \int_0^\epsilon \frac{1}{\epsilon} I\left(Y_0^t; Z_{t-\epsilon}^T \middle| M\right) \, dt - \int_T^{T+\epsilon} \frac{1}{\epsilon} I\left(Y_0^t; Z_{t-\epsilon}^T \middle| M\right) \, dt.\end{aligned} \tag{2.24}$$

Taking limits, we will consider both terms separately

$$\begin{aligned}\lim_{\epsilon \to 0^+} \int_0^\epsilon \frac{1}{\epsilon} I\left(Y_0^t; Z_{t-\epsilon}^T \middle| M\right) \, dt &\overset{(a)}{\leq} \lim_{\epsilon \to 0^+} \int_0^\epsilon \frac{1}{\epsilon} I\left(Y_0^t; Z_0^T \middle| M\right) \, dt \\ &\overset{(b)}{\leq} \lim_{\epsilon \to 0^+} \int_0^\epsilon \frac{1}{\epsilon} I\left(Y_0^\epsilon; Z_0^T \middle| M\right) \, dt\end{aligned}$$

$$= \lim_{\epsilon \to 0^+} I\left(Y_0^\epsilon; Z_0^T \middle| M\right)$$
$$\stackrel{(c)}{=} 0, \tag{2.25}$$

where, for (a) and (b) we have used the fact that $I(U_{t_1}^{t_2}; A|B)$ is monotonic in t_1 and t_2 since

$$I\left(A; U_{t_1}^{t_2} \middle| B\right) = \int_{t_1}^{t_2} \mathbb{E}[\phi_u(\mathbb{E}[X_t|U_{t_1}^t, A, B])] - \mathbb{E}[\phi_u(\mathbb{E}[X_t|U_{t_1}^t, B])]\, dt.$$

As the integrand is non-negative due to Jensen's inequality, $I\left(A; U_{t_1}^{t_2} \middle| B\right)$ is non-increasing in t_1 for fixed t_2 and non-decreasing in t_2 for fixed t_1. Also, since the integrand is bounded,

$$\lim_{t_2 \to t_1^+} I\left(A; U_{t_1}^{t_2} \middle| B\right) = 0.$$

This gives (c). Similarly,

$$\lim_{\epsilon \to 0^+} \frac{1}{\epsilon} \int_T^{T+\epsilon} I\left(Y_0^t; Z_{t-\epsilon}^T \middle| M\right)\, dt = 0.$$

This proves part (1). Since $\frac{1}{\epsilon} I\left(Z_{t-\epsilon}^t; Y_0^t \middle| Z_t^T, M\right)$ and $\frac{1}{\epsilon} I\left(Y_t^{t+\epsilon}; Z_t^T \middle| Y_0^t, M\right)$ are bounded over $\epsilon > 0$ from Lemma 8, we use the dominated convergence theorem to swap the integral and limit in (2.20) to get

$$\int_0^T \lim_{\epsilon \to 0^+} \frac{1}{\epsilon} I\left(Z_{t-\epsilon}^t; Y_0^t \middle| Z_t^T, M\right)\, dt = \int_0^T \lim_{\epsilon \to 0^+} \frac{1}{\epsilon} I\left(Y_t^{t+\epsilon}; Z_t^T \middle| Y_0^t, M\right)\, dt. \tag{2.26}$$

Taking $U = Z$, $A = Y_0^t$ and $B = M$ in the left-hand side of (2.26), Lemma 7 gives

$$\int_0^T \lim_{\epsilon \to 0^+} \frac{1}{\epsilon} I\left(Z_{t-\epsilon}^t; Y_0^t \middle| Z_t^T, M\right)\, dt = \int_0^T \mathbb{E}[\phi_z(\mathbb{E}[X_{t-}|Y_0^t, Z_t^T, M])] - \mathbb{E}[\phi_z(\mathbb{E}[X_{t-}|Z_t^T, M])]\, dt.$$

Since X_0^T is a càdlàg process, we can repeat the same argument as in the proof of Theorem 1 to replace X_{t-} in the above integral with X_t. We get

$$\int_0^T \lim_{\epsilon \to 0^+} \frac{1}{\epsilon} I\left(Z_{t-\epsilon}^t; Y_0^t \middle| Z_t^T, M\right)\, dt = \int_0^T \mathbb{E}[\phi_z(\mathbb{E}[X_t|Y_0^t, Z_t^T, M])] - \mathbb{E}[\phi_z(\mathbb{E}[X_t|Z_t^T, M])]\, dt. \tag{2.27}$$

Similarly, taking $U = Y$, $A = Z_t^T$ and $B = M$ in the right hand side of (2.26), Lemma 7 gives

$$\int_0^T \lim_{\epsilon \to 0^+} \frac{1}{\epsilon} I\left(Y_t^{t+\epsilon}; Z_t^T \middle| Y_0^t, M\right) dt = \int_0^T \mathbb{E}[\phi_y(\mathbb{E}[X_t|Y_0^t, Z_t^T, M])] - \mathbb{E}[\phi_y(\mathbb{E}[X_t|Y_0^t, M])]\, dt. \tag{2.28}$$

The second part of the lemma now follows since (2.27) and (2.28) are equal from (2.26). □

2.4 Comparison of Two Receivers

Motivated by the definition for the discrete memoryless channels [36], we define a less noisy receiver and a more capable receiver for the two-user Poisson channel as follows.

Definition 5 (Less Noisy Receiver) *Receiver 1 is said to be less noisy than receiver 2 if $I(M; Y_0^T) \geq I(M; Z_0^T)$ for all possible M in (2.7), where $M \rightleftarrows X_0^T \rightleftarrows (Y_0^T, Z_0^T)$ is a Markov chain.*

Definition 6 (More Capable Receiver) *Receiver 1 is said to be more capable than receiver 2 if $I(X_0^T; Y_0^T) \geq I(X_0^T; Z_0^T)$ for all probability measures on the input space $(\mathcal{X}_0^T, \mathfrak{F}^X)$.*

We shall call a channel with a less noisy receiver to be a less noisy Poisson channel and similarly a channel with a more capable receiver to be a more capable Poisson channel.

Theorem 3 (More capable and less noisy Poisson channel) *In a two-user Poisson channel the following conditions are equivalent:*

(I) $\Phi(x) = \phi_y(x) - \phi_z(x)$ *is a convex function over* $[0, 1]$.

(II) Receiver 1 is less noisy than receiver 2.

(III) Receiver 1 is more capable than receiver 2.

(IV) The channel parameters satisfy

- $a_y \geq a_z$ *and* $a_y^2\lambda_z \geq a_z^2\lambda_y$*; or*
- $0 < a_y < a_z$ *and* $a_y^2(a_z + \lambda_z) \geq a_z^2(a_y + \lambda_y)$.

Proof: To prove (I) implies (II), note that Theorem 1 yields

$$\begin{aligned}
I(M; Y_0^T) - I(M; Z_0^T) &= \int_0^T \mathbb{E}[\phi_y(\mathbb{E}[X_t|Y_0^t, M])] - \mathbb{E}[\phi_y(\mathbb{E}[X_t|Y_0^t])]\, dt \\
&\quad - \int_0^T \mathbb{E}[\phi_z(\mathbb{E}[X_t|Z_t^T, M])] - \mathbb{E}[\phi_z(\mathbb{E}[X_t|Z_t^T])]\, dt \qquad (2.29) \\
&= \int_0^T \mathbb{E}[\phi_y(\mathbb{E}[X_t|Y_0^t, M])] - \mathbb{E}[\phi_z(\mathbb{E}[X_t|M, Z_t^T])]\, dt \\
&\quad - \int_0^T \mathbb{E}[\phi_y(\mathbb{E}[X_t|Y_0^t])] - \mathbb{E}[\phi_z(\mathbb{E}[X_t|Z_t^T])]\, dt \\
&\overset{(a)}{=} \int_0^T \mathbb{E}[\phi_y(\mathbb{E}[X_t|Y_0^t, Z_t^T, M])] - \mathbb{E}[\phi_z(\mathbb{E}[X_t|Y_0^T, Z_t^T, M])]\, dt \\
&\quad - \int_0^T \mathbb{E}[\phi_y(\mathbb{E}[X_t|Y_0^t, Z_t^T])] - \mathbb{E}[\phi_z(\mathbb{E}[X_t|Y_0^t, Z_t^T])]\, dt \\
&= \int_0^T \mathbb{E}[\Phi(\mathbb{E}[X_t|Y_0^t, Z_t^T, M])] - \mathbb{E}[\Phi(\mathbb{E}[X_t|Y_0^t, Z_t^T])]\, dt \qquad (2.30)
\end{aligned}$$

where (a) is due to Theorem 2. Since $\Phi(x)$ is a convex function, Jensen's inequality gives

$$\begin{aligned}
I(M; Y_0^T) - I(M; Z_0^T) &= \int_0^T \mathbb{E}[\Phi(\mathbb{E}[X_t|Y_0^t, Z_t^T, M])] - \mathbb{E}[\Phi(\mathbb{E}[X_t|Y_0^t, Z_t^T])]\, dt \\
&\geq 0. \qquad (2.31)
\end{aligned}$$

Note that (II) implies (III) trivially. We now prove that (III) implies (I). There exists a sequence of input distributions (indexed by n), such that X_0^T is binary and stationary with the following limit [17, 30]

$$\lim_{n\to\infty} \mathbb{E}\left[\phi_u\left(\mathbb{E}[X_t|U_0^t]\right)\right] = \phi_u(\mathbb{E}[X_t]).$$

Thus choosing X_t such that $P(X_t = p) = 1 - P(X_t = q) = \alpha$, $0 \leq \alpha \leq 1$ and taking the limit gives

$$\alpha\phi_y(p) + (1-\alpha)\phi_y(q) - \phi_y(\alpha p + (1-\alpha)q) \geq \alpha\phi_z(p) + (1-\alpha)\phi_z(q) - \phi_z(\alpha p + (1-\alpha)q).$$

Therefore

$$\alpha\Phi(p) + (1-\alpha)\Phi(q) \geq \Phi(\alpha p + (1-\alpha)q).$$

Hence $\Phi(x)$ is a convex function.

The channel parameters for which the channel is less noisy can be obtained by calculating conditions under which the second derivative of $\Phi(x)$ is non-negative for $0 \leq x \leq 1$. □

Note that these channel parameters include the parameters for which the channel is known to be stochastically degraded [41]

$$a_y \geq a_z, \qquad a_y\lambda_z \geq a_z\lambda_y. \tag{2.32}$$

The conditions given in Theorem 3 differ from the conditions under which the discretized Poisson channel is more capable. A discretized Poisson channel is a discrete memoryless channel in which the input is binary and constant over τ-duration intervals, where τ is very small. The output in an interval is taken to be "1" if there are one or more arrivals during this interval and "0" otherwise. Wyner [67] shows that, for the purposes of reliable communication, the Poisson channel is equivalent to its discretized version, so that coding theorems for the former may be inferred from the latter. This equivalence carries over to Poisson broadcast channels [41].

Kim *et al.* [33] determine the range of parameters under which the discretized Poisson broadcast channel is less noisy and more capable. The conditions under which the discretized channel is less noisy match those in Theorem 3. The conditions for the discretized channel to be more capable, however, are strictly weaker: if $a_y = 0.4$, $\lambda_y = 0$, $a_z = \lambda_z = 1$, for example, the discretized channel is more capable [33, Theorem 1], whereas the continuous-time, continuous-space channel considered here is not. To see the reason behind this, consider a sequence of input distributions (indexed by n) as in the proof of Theorem 3, such that X_0^T is binary and stationary with the following limit for $u \in \{y, z\}$ [17, 30]

$$\lim_{n\to\infty} \mathbb{E}\left[\phi_u\left(\mathbb{E}[X_t|U_0^t]\right)\right] = \phi_u(\mathbb{E}[X_t]).$$

Then choosing X_t such that $P(X_t = 1) = P(X_t = 0.9) = 0.5$, and taking the limit gives

$$\lim_{n\to\infty} \frac{1}{T} I(X_0^T; Z_0^T) \approx 6.41 \times 10^{-4} > 5.26 \times 10^{-4} \approx \lim_{n\to\infty} \frac{1}{T} I(X_0^T; Y_0^T).$$

If X_0^T only takes values in $\{0, 1\}$, on the other hand, then this inequality is impossible. Of course, for the purposes of reliable communication, X_0^T need only take values in $\{0, 1\}$, as noted above.

Nair [47] defines one discrete memoryless channel to be *essentially more capable* than another if a condition similar to the usual definition of "more capable" holds under a restricted set of input distributions that dominates all others in certain single-letter mutual information expressions. The statement that one discretized Poisson channel is more capable than another thus translates into something akin to "essentially more capable" when expressed in terms of the underlying continuous Poisson channels. This analogy is not exact, however, in that "essentially more capable" is defined in terms of mutual information expressions while the reduction from the Poisson channel to its discretized version is operational. All of this indicates that some care is required when translating statements between the Poisson channel and its discretized version.

We next apply the results obtained thus far to characterize the capacity (regions) for several multi-receiver communication problems. The first of these is the more-capable Poisson broadcast channel. Our result here is less general than that obtained by Kim *et al.* [33], although our proof is more self contained in that it does not require a discretization argument. We then prove new results on the Poisson broadcast channel with degraded message sets and the Poisson wiretap channel.

2.5 More Capable Poisson Broadcast Channel

We first prove several lemmas. Let $T_n = n\tau$ for some $\tau > 0$. Construct an auxiliary process $V_0^{T_n}$ to be piecewise constant, taking value in the finite alphabet $\mathcal{V} = \{1, \ldots, K_v\}$ as follows. We divide the interval $[0, T_n]$ into n intervals each of equal length τ. The process will be constant on each of these sub-intervals with value given by

$$V_t = \bar{V}_i \text{ for } (i-1)\tau \le t < i\tau, \quad i = 1, 2, \ldots, n \tag{2.33}$$

where $\bar{V}_i$'s are independent and identically distributed random variables with $P(\bar{V}_i = j) = \alpha_j,\ j \in \mathcal{V}$. Let $\mathcal{V}_0^{T_n}$ denote the collection of all such processes. The input waveform $X_0^{T_n}$ is binary and piecewise constant with

$$X_t = \bar{X}_i \text{ for } (i-1)\tau \le t < i\tau, \quad i = 1, 2, \ldots, n \tag{2.34}$$

where

$$P(\bar{X}_i = 1|\bar{V}_i = j) = 1 - P(\bar{X}_i = 0|\bar{V}_i = j) = p_j. \tag{2.35}$$

The following lemma shows that with the above input to the channel, we have essentially decomposed the single channel use into n independent and identical channel uses.

Lemma 10 *Let $U_t^{(i)}$ be the point process corresponding to the arrival time process $U_{(i-1)\tau}^{i\tau}$. The joint distribution of processes $(\bar{V}_i, \bar{X}_i, U_t^{(i)} : t \in [(i-1)\tau, i\tau])$ is independent and identical across the disjoint blocks for $i = 1, \ldots, n$ and $U \in \{Y, Z\}$.*

For fixed $V_0^{T_n} \in \mathcal{V}_0^{T_n}$, let $P^{X_0^{T_n}|V_0^{T_n}}$ denote the probability measure on the input space from the construction in (2.33)-(2.35). Then the probability measure on $(\mathcal{N}_0^{T_n}, \mathfrak{F}^Y)$ for fixed $V_0^{T_n}$ is [31, Lemma 1.41, p. 21]

$$P^{Y_0^{T_n}|V_0^{T_n}}(dy_0^{T_n}) = \int_{\mathcal{X}_0^T} P^{X_0^{T_n}|V_0^{T_n}} p_y(x_0^{T_n}, y_0^{T_n}) P_0(dy_0^{T_n}).$$

Let

$$Q^{V_0^{T_n}, X_0^{T_n}, Y_0^{T_n}} = P^{V_0^{T_n}} \times P^{X_0^{T_n}|V_0^{T_n}} \times P^{Y_0^{T_n}|V_0^{T_n}}. \tag{2.36}$$

Hence under $Q^{V_0^{T_n}, X_0^{T_n}, Y_0^{T_n}}$, the joint distribution of $(V_0^{T_n}, X_0^{T_n})$ and $(V_0^{T_n}, Y_0^{T_n})$ is the same as that under P, and $X_0^{T_n} \rightleftarrows V_0^{T_n} \rightleftarrows Y_0^{T_n}$ forms a Markov chain.

Definition 7 *The following mutual information densities are defined whenever the corresponding Radon-Nikodym derivatives exist and are strictly positive, in which* *case we will say that the mutual information densities exist.*

$$
\begin{aligned}
\mathfrak{i}(X_0^{T_n};Y_0^{T_n}) &= \log\left(\frac{dP^{X_0^{T_n},Y_0^{T_n}}}{d(P^{X_0^{T_n}}\times P^{Y_0^{T_n}})}\right)\\
\mathfrak{i}(X_0^{T_n};Y_0^{T_n}|V_0^{T_n}) &= \log\left(\frac{dP^{V_0^{T_n},X_0^{T_n},Y_0^{T_n}}}{dQ^{V_0^{T_n},X_0^{T_n},Y_0^{T_n}}}\right)\\
\mathfrak{i}(V_0^{T_n};Z_0^{T_n}) &= \log\left(\frac{dP^{V_0^{T_n},Z_0^{T_n}}}{d(P^{V_0^{T_n}}\times P^{Z_0^{T_n}})}\right).
\end{aligned}
$$

Lemma 11 *The mutual information densities in Definition 7 exist, and for all* $\epsilon > 0$ *there exists* $\bar{\tau}$ *and* N *such that if* $n \geq N$ *and* $\tau \leq \bar{\tau}$ *then*

$$
\begin{aligned}
P\left(\left|\frac{1}{T_n}\mathfrak{i}(X_0^{T_n};Y_0^{T_n}) - \left(\mathbb{E}\left[\phi_y(X_0)\right] - \phi_y(\mathbb{E}[X_0])\right)\right| > \epsilon\right) &\leq \epsilon\\
P\left(\left|\frac{1}{T_n}\mathfrak{i}(V_0^{T_n};Z_0^{T_n}) - \left(\mathbb{E}\left[\phi_z\left(\mathbb{E}[X_0|\bar{V}_1]\right)\right] - \phi_z(\mathbb{E}[X_0])\right)\right| > \epsilon\right) &\leq \epsilon\\
P\left(\left|\frac{1}{T_n}\mathfrak{i}(X_0^{T_n};Y_0^{T_n}|V_0^{T_n}) - \left(\mathbb{E}\left[\phi_y(X_0)\right] - \mathbb{E}\left[\phi_y\left(\mathbb{E}[X_0|\bar{V}_1]\right)\right]\right)\right| > \epsilon\right) &\leq \epsilon. \quad (2.37)
\end{aligned}
$$

Proof: See Section 2.8. □

Lemma 12 *If user 1 is more capable than user 2, then*

$$
\int_0^T \mathbb{E}[\phi_z(\mathbb{E}[X_t|M,Y_0^t])]\,dt \geq \int_0^T \mathbb{E}[\phi_z(\mathbb{E}[X_t|M,Z_t^T])]\,dt. \quad (2.38)
$$

Proof: See Section 2.8. □

2.5.1 Encoding and Decoding

An (L_y, L_z, T) *code* for the Poisson broadcast channel consists of a source (equipped with an encoder $\mathscr{E}_x^T$) and two receivers each with a decoder ($\mathscr{D}_y^T$ and $\mathscr{D}_z^T$). The source has two independent messages M_y and M_z for the first and second user, respectively, where M_y and M_z are uniformly distributed on sets $\mathcal{M}_y = \{1, \dots, L_y\}$ and $\mathcal{M}_z = \{1, \dots, L_z\}$, respectively.

Given messages M_y and M_z the encoder selects a waveform in $\mathcal{X}_0^T$

$$\mathscr{E}_x^T : \mathcal{M}_y \times \mathcal{M}_z \to \mathcal{X}_0^T. \tag{2.39}$$

Let $\Delta_{x_0^T}(dx_0^T)$ be the Dirac measure on the input space induced by the given messages m_y, m_z, and the encoder $\mathscr{E}_x^T$. Then the probability space $(\Omega, \mathcal{F}, P)$ is

$$\begin{aligned}
\Omega &= \mathcal{M}_y \times \mathcal{M}_z \times \mathcal{X}_0^T \times \mathcal{N}_0^T \times \mathcal{N}_0^T \\
\mathcal{F} &= 2^{\mathcal{M}_y \times \mathcal{M}_z} \otimes \mathfrak{F}^X \otimes \mathfrak{F}^Y \otimes \mathfrak{F}^Z \\
P &= \nu(m_y, m_z) \Delta_{\mathscr{E}_x^T(m_y, m_z)}(dx_0^T) P_0^Y(dy_0^T) P_0^Z(dz_0^T) \prod_{u=y,z} p_u(x_0^T, u_0^T).
\end{aligned} \tag{2.40}$$

Here $\nu(m_y, m_z)$ is the uniform distribution on $\mathcal{M}_y \times \mathcal{M}_z$, and $2^{\mathcal{M}_y \times \mathcal{M}_z}$ is the power set of $\mathcal{M}_y \times \mathcal{M}_z$.

On observing Y_0^T and Z_0^T, each decoder chooses a message

$$\begin{aligned}
\mathscr{D}_y^T &: \mathcal{N}_0^T \to \mathcal{M}_y \\
\mathscr{D}_z^T &: \mathcal{N}_0^T \to \mathcal{M}_z.
\end{aligned} \tag{2.41}$$

The average probability of error for this code is

$$\mathrm{P}_e = \frac{1}{L_y L_z} \sum_{m_y=1, m_z=1}^{L_y, L_z} P\left\{ \{\mathscr{D}_y^T(Y_0^T) \neq m_y\} \bigcup \{\mathscr{D}_z^T(Z_0^T) \neq m_z\} \middle| M_y = m_y, M_z = m_z \right\}. \tag{2.42}$$

A rate pair (R_y, R_z) is said to be *achievable* if for all $\epsilon > 0$ and sufficiently large T, there exists an (L_y, L_z, T) code such that

$$\begin{aligned}
\frac{\log(L_y)}{T} &\geq R_y - \epsilon \\
\frac{\log(L_z)}{T} &\geq R_z - \epsilon \\
\mathrm{P}_e &\leq \epsilon.
\end{aligned} \tag{2.43}$$

The capacity region (C_y, C_z) is the closure of achievable rate pairs.

Theorem 4 (Capacity of more capable Poisson broadcast channel) *The capacity of the more capable Poisson broadcast channel when receiver 1 is more capable than receiver 2 is given by the convex hull of the union over all* $0 \le \alpha \le \frac{1}{2}$ *and* $0 \le p, q \le 1$ *of rate pairs satisfying*

$$R_y \le C_y = \alpha(p\phi_y(1) + (1-p)\phi_y(0) - \phi_y(p)) + (1-\alpha)(q\phi_y(1) + (1-q)\phi_y(0) - \phi_y(q))$$
$$R_z \le C_z = \alpha\phi_z(p) + (1-\alpha)\phi_z(q) - \phi_z(\alpha p + (1-\alpha)q).$$

Although the proof of the above theorem can be found in [33], we provide an alternate proof using tools derived from stochastic calculus without resorting to the discretization of the continuous-time, continuous-space Poisson channel. Similar proof techniques will be used in proving the capacity theorem of the Poisson broadcast channel with degraded message set to follow. The achievability and converse arguments are provided in next two subsections.

2.5.2 Achievability

We first note that that C_y and C_z are upper bounded by the point-to-point capacity of the single-receiver Poisson channel to the first and second user respectively, which for the channel parameters (a_u, λ_u), $u \in \{y, z\}$ is given by [17, 30, 68]

$$C_u^{\text{pp}} = \max_{0 \le \kappa \le 1} \kappa\phi_u(1) + (1-\kappa)\phi_u(0) - \phi_u(\kappa).$$

Let $\kappa = \alpha p + (1-\alpha)q$, and using the convexity of ϕ_u :

$$
\begin{aligned}
C_y &= \alpha(p\phi_y(1) + (1-p)\phi_y(0) - \phi_y(p)) + (1-\alpha)(q\phi_y(1) + (1-q)\phi_y(0) - \phi_y(q)) \\
&= (\alpha p + (1-\alpha)q)\phi_y(1) + (\alpha(1-p) + (1-\alpha)(1-q)\phi_y(0)) - (\alpha\phi_y(p) + (1-\alpha)\phi_y(q)) \\
&\le (\alpha p + (1-\alpha)q)\phi_y(1) + (\alpha(1-p) + (1-\alpha)(1-q)\phi_y(0)) - \phi_y(\alpha p + (1-\alpha)q) \\
&= \kappa\phi_y(1) + (1-\kappa)\phi_y(0) - \phi_y(\kappa) \\
&\le C_y^{\mathrm{pp}}.
\end{aligned}
$$

Likewise

$$
\begin{aligned}
C_z &= \alpha\phi_z(p) + (1-\alpha)\phi_z(q) - \phi_z(\alpha p + (1-\alpha)q) \\
&\le \alpha p\phi_z(1) + \alpha(1-p)\phi_z(0) + (1-\alpha)q\phi_z(1) + (1-\alpha)(1-q)\phi_z(0) - \phi_z(\alpha p + (1-\alpha)q) \\
&= \kappa\phi_z(1) + (1-\kappa)\phi_z(0) - \phi_z(\kappa) \\
&\le C_z^{\mathrm{pp}}.
\end{aligned}
$$

Thus if α, p, and q are such that either C_y or C_z is zero, then achievability follows from the point-to-point achievability argument in [68]. Hence we consider the cases when both of these quantities are strictly positive. Let $T_n = n\tau$ for some finite $\tau > 0$. Construct an auxiliary process $V_0^{T_n}$ to be a piecewise constant binary-valued process. We divide the interval $[0, T_n]$ into n intervals each of equal length τ. The process will be constant on each of these sub-intervals with value given by

$$V_t = \bar{V}_i \text{ for } (i-1)\tau \le t < i\tau, \quad i = 1, 2, \ldots, n \tag{2.44}$$

where $\bar{V}_i$'s are independent and identically distributed Bernoulli random variables with $P(\bar{V}_i = 1) = \alpha$.

The input waveform $X_0^{T_n}$ is binary and piecewise constant with

$$X_t = \bar{X}_i \text{ for } (i-1)\tau \le t < i\tau, \quad i = 1, 2, \ldots, n \tag{2.45}$$

where

$$P(\bar{X}_i = 1|\bar{V}_i = 1) = 1 - P(\bar{X}_i = 0|\bar{V}_i = 1) = p$$
$$P(\bar{X}_i = 1|\bar{V}_i = 0) = 1 - P(\bar{X}_i = 0|\bar{V}_i = 0) = q. \tag{2.46}$$

An application of Lemma 11 yields:

Lemma 13 *Let* $\tilde{C}_y = \alpha\phi_y(p) + (1-\alpha)\phi_y(q) - \phi_y(\alpha p + (1-\alpha)q)$*. For all* $\epsilon > 0$ *there exist* $\bar{\tau}$ *and* N *such that if* $n \geq N$ *and* $\tau \leq \bar{\tau}$*, then*

$$P\left(\left|\frac{1}{T_n}\mathfrak{i}(X_0^{T_n}; Y_0^{T_n}) - (C_y + \tilde{C}_y)\right| > \epsilon\right) \leq \epsilon$$
$$P\left(\left|\frac{1}{T_n}\mathfrak{i}(V_0^{T_n}; Z_0^{T_n}) - C_z\right| > \epsilon\right) \leq \epsilon$$
$$P\left(\left|\frac{1}{T_n}\mathfrak{i}(X_0^{T_n}; Y_0^{T_n}|V_0^{T_n}) - C_y\right| > \epsilon\right) \leq \epsilon.$$

Proof: See Section 2.8. □

Encoding Operation

We use superposition coding. Fix $\delta > 0$, and let $R_y = C_y - \delta$ and $R_z = C_z - \delta$. We generate $L_z = \exp(T_n R_z)$ many $V_0^{T_n}$ waveforms (indexed by $j = 1, \ldots, L_z$) independently according to (2.44). For each $V_0^{T_n}(j)$, we generate $L_y = \exp(T_n R_y)$ many independent $X_0^{T_n}$ waveforms (indexed by $i = 1, \ldots, L_y$) according to (2.45) and (2.46). To transmit messages (M_y, M_z), encoder sends $X_0^{T_n}(M_y, M_z)$ over the channel.

Decoding Operation

For a received $Z_0^{T_n}$, the second receiver considers only those $V_0^{T_n}$ for which both $\frac{1}{T_n}\log\left(\frac{dP^{V_0^{T_n},Z_0^{T_n}}}{d\tilde{P}^{V_0^{T_n},Z_0^{T_n}}}\right)$ and $\frac{1}{T_n}\log\left(\frac{dP^{Z_0^{T_n}}}{d\tilde{P}^{Z_0^{T_n}}}\right)$ (calculated using Theorem 1) are finite. We note that $\{\Pi_t : t \in [0,T]\}$ as in Theorem 1 is $V_0^{T_n}$, $Z_0^{T_n}$ measurable. It seeks the unique j among all such waveforms such that

$$\frac{1}{T_n}\mathfrak{i}(V_0^{T_n}(j); Z_0^{T_n}) = \frac{1}{T_n}\log\left(\frac{dP^{V_0^{T_n},Z_0^{T_n}}}{d\tilde{P}^{V_0^{T_n},Z_0^{T_n}}}\right) - \frac{1}{T_n}\log\left(\frac{dP^{Z_0^{T_n}}}{dP_0^{Z_0^{T_n}}}\right) \geq C_z - \gamma_z \quad (2.47)$$

for some $\gamma_z > 0$, and outputs $\hat{M}_z = j$. If the decoder does not find any such $V_0^{T_n}$, or if it finds more than one $V_0^{T_n}$ that satisfy (2.47), then the decoder arbitrarily outputs some $\hat{M}_z \in [1, \ldots, L_z]$.

The first receiver decodes both M_y and M_z, and we declare an error if either or both messages are decoded incorrectly. It seeks a unique i and j that satisfy both

$$\frac{1}{T_n}\mathfrak{i}(X_0^{T_n}(i,j); Y_0^{T_n}) \geq C_y + \tilde{C}_y - \gamma_y \quad (2.48)$$

and

$$\frac{1}{T_n}\mathfrak{i}(X_0^{T_n}(i,j); Y_0^{T_n} | V_0^{T_n}(j)) \geq C_y - \gamma_y. \quad (2.49)$$

The decoder considers only those $X_0^{T_n}$ and $V_0^{T_n}$ for which the above random variables are well defined (i.e., they do not evaluate to $\infty - \infty$) and finite.

Without loss of generality assume that $X_0^{T_n}(1,1)$ was transmitted. Let $P_{e,0}^{(z)}$ denote the probability of the error event that the second decoder does not find any $V_0^{T_n}$ that satisfies (2.47). Due to Lemma 13, $\mathbb{E}_{\mathcal{C}}[P_{e,0}^{(z)}]$ can be made arbitrarily small, where $\mathbb{E}_{\mathcal{C}}$ denotes expectation with respect to random code book generation. Let $\mathsf{E}_{e,j}^{(z)}$ denote the error event that for some $j \neq 1$, $V_0^{T_n}(j)$ satisfies (2.47), and let

$P_{e,j}^{(z)}$ denote the corresponding error probability. Then we have for $j \neq 1$

$$\begin{aligned}\mathbb{E}_{\mathcal{C}}[P_{e,j}^{(z)}] &= \int_{V_0^{T_n}, Z_0^{T_n}} \mathbf{1}\{\mathsf{E}_{e,j}^{(z)}\} d(P^{V_0^{T_n}} \times P^{Z_0^{T_n}}) \\ &\leq \exp(-T_n(C_z - \gamma_z)) \int_{V_0^{T_n}, Z_0^{T_n}} \mathbf{1}\{\mathsf{E}_{e,j}^{(z)}\} dP^{V_0^{T_n}, Z_0^{T_n}} \\ &\leq \exp(-T_n(C_z - \gamma_z)).\end{aligned}$$

By the union bound

$$\begin{aligned}\mathbb{E}_{\mathcal{C}}[P_e^{(z)}] &\leq \mathbb{E}_{\mathcal{C}}[P_{e,0}^{(z)}] + \sum_{j=2}^{L_z} \mathbb{E}_{\mathcal{C}}[P_{e,j}^{(z)}] \\ &\leq \mathbb{E}_{\mathcal{C}}[P_{e,0}^{(z)}] + \exp(-T_n(C_z - R_z - \gamma_z)). \end{aligned} \tag{2.50}$$

Thus $\mathbb{E}_{\mathcal{C}}[P_e^{(z)}]$ can be made arbitrarily small.

Similar to the second decoder, the average probability $\mathbb{E}_{\mathcal{C}}[P_{e,0}^{(y)}]$ that the first receiver cannot find any (i, j) that satisfy both (2.48) and (2.49) can be made small due to Lemma 13. Let $\mathsf{E}_{e,(i,j)}^{(y)}$ denote the error event that for some $(i, j) \neq (1, 1)$, (i, j) satisfies both (2.48) and (2.49). First consider $\mathsf{E}_{e,(i,j)}^{(y)}$ for $j \neq 1$. For this case $X_0^{T_n}(i, j)$ and $Y_0^{T_n}$ are independent, and for $j \neq 1$, the corresponding error probability $P_{e,(i,j)}^{(y)}$ is upper bounded by the probability that (i, j) satisfies (2.48).

$$\begin{aligned}\mathbb{E}_{\mathcal{C}}[P_{e,(i,j)}^{(y)}] &\leq \int_{X_0^{T_n}, Y_0^{T_n}} \mathbf{1}\{\mathsf{E}_{e,(i,j)}^{(y)}\} d(P^{X_0^{T_n}} \times P^{Y_0^{T_n}}) \\ &\leq \exp(-T_n(C_y + \tilde{C}_y - \gamma_y)) \int_{X_0^{T_n}, Y_0^{T_n}} \mathbf{1}\{\mathsf{E}_{e,(i,j)}^{(y)}\} dP^{X_0^{T_n}, Y_0^{T_n}} \\ &\leq \exp(-T_n(C_y + \tilde{C}_y - \gamma_y)).\end{aligned}$$

When $j = 1$, and $i \neq 1$, $X_0^{T_n}(i, 1) \rightleftarrows V_0^{T_n}(1) \rightleftarrows Y_0^{T_n}$ is a Markov chain. The average probability that $V_0^{T_n}(1)$ and $X_0^{T_n}(i, 1)$ for $i \neq 1$ satisfies (2.49) is

$$\int_{V_0^{T_n}, X_0^{T_n}, Y_0^{T_n}} \mathbf{1}\{\mathsf{E}_{e,(i,1)}^{(y)}\} dQ^{V_0^{T_n}, X_0^{T_n}, Y_0^{T_n}},$$

where $Q^{V_0^{T_n},X_0^{T_n},Y_0^{T_n}}$ is defined in (2.36). Thus for $i \neq 1$, we can upper bound $\mathbb{E}_{\mathcal{C}}[P_{e,(i,1)}^{(y)}]$ as

$$\begin{aligned}\mathbb{E}_{\mathcal{C}}[P_{e,(i,1)}^{(y)}] &\leq \int_{V_0^{T_n},X_0^{T_n},Y_0^{T_n}} \mathbf{1}\{\mathsf{E}_{e,(i,1)}^{(y)}\} dQ^{V_0^{T_n},X_0^{T_n},Y_0^{T_n}} \\ &\leq \exp(-T_n(C_y - \gamma_y)) \int_{V_0^{T_n},X_0^{T_n},Y_0^{T_n}} \mathbf{1}\{\mathsf{E}_{e,(i,1)}^{(y)}\} dP^{V_0^{T_n},X_0^{T_n},Y_0^{T_n}} \\ &\leq \exp(-T_n(C_y - \gamma_y)).\end{aligned}$$

The average probability of error can be upper bounded using the union bound as

$$\begin{aligned}\mathbb{E}_{\mathcal{C}}[P_e^{(y)}] &\leq \mathbb{E}_{\mathcal{C}}[P_{e,0}^{(y)}] + \sum_{(i,j)\neq(1,1)}^{L_y,L_z} \mathbb{E}_{\mathcal{C}}[P_{e,(i,j)}^{(y)}] \\ &= \mathbb{E}_{\mathcal{C}}[P_{e,0}^{(y)}] + \sum_{i=2}^{L_y} \mathbb{E}_{\mathcal{C}}[P_{e,(i,1)}^{(y)}] + \sum_{i=1,j=2}^{L_y,L_z} \mathbb{E}_{\mathcal{C}}[P_{e,(i,j)}^{(y)}] \\ &= \mathbb{E}_{\mathcal{C}}[P_{e,0}^{(y)}] + (L_y - 1)\mathbb{E}_{\mathcal{C}}[P_{e,(2,1)}^{(y)}] + L_y(L_z - 1)\mathbb{E}_{\mathcal{C}}[P_{e,(1,2)}^{(y)}] \\ &\leq \mathbb{E}_{\mathcal{C}}[P_{e,0}^{(y)}] + \exp(R_y T_n)\exp(-T_n(C_y - \gamma_y)) + \exp((R_y + R_z)T_n)\exp(-T_n(C_y + \tilde{C}_y - \gamma_y)) \\ &= \mathbb{E}_{\mathcal{C}}[P_{e,0}^{(y)}] + \exp(-T_n(C_y - R_y - \gamma_y)) + \exp(-T_n(C_y + \tilde{C}_y - (R_y + R_z) - \gamma_y)),\end{aligned} \tag{2.51}$$

which can be made arbitrarily small since $R_y = C_y - \delta$ and

$$\begin{aligned}R_y + R_z &= \alpha(p\phi_y(1) + (1-p)\phi_y(0) - \phi_y(p)) + (1-\alpha)(q\phi_y(1) + (1-q)\phi_y(0) - \phi_y(q)) \\ &\quad + \alpha\phi_z(p) + (1-\alpha)\phi_z(q) - \phi_z(\alpha p + (1-\alpha)q) - 2\delta \\ &\leq \alpha(p\phi_y(1) + (1-p)\phi_y(0) - \phi_y(p)) + (1-\alpha)(q\phi_y(1) + (1-q)\phi_y(0) - \phi_y(q)) - 2\delta \\ &\quad + \alpha\phi_y(p) + (1-\alpha)\phi_y(q) - \phi_y(\alpha p + (1-\alpha)q) \\ &= C_y + \tilde{C}_y - 2\delta,\end{aligned}$$

where we have used the more capable property of the channel:

$$\alpha\phi_z(p)+(1-\alpha)\phi_z(q)-\phi_z(\alpha p+(1-\alpha)q) \leq \alpha\phi_y(p)+(1-\alpha)\phi_y(q)-\phi_y(\alpha p+(1-\alpha)q).$$

Hence by Markov's inequality, for a given $\epsilon > 0$ there exists N and $\bar{\tau}$ such that for all $n \geq N$, and $\tau \leq \bar{\tau}$, a codebook with $T = n\tau$ satisfying (2.43) can be found.

2.5.3 Converse

Suppose that (R_y, R_z) is achievable. Then there exists a code such that (2.43) holds. For $(u, U) \in \{(y, Y), (z, Z)\}$, let $\tilde{R}_u = \frac{\log(L_u)}{T}$. Then

$$\begin{aligned}
\tilde{R}_u T = \log(L_u) = H(M_u) &= \mathbb{E}[H(M_u | U_0^T)] + I(M_u; U_0^T) \\
&\overset{(a)}{\leq} H\left(M_u | \mathscr{D}_u^T(U_0^T)\right) + I(M_u; U_0^T) \\
&\overset{(b)}{\leq} H(P_e^{(u)}) + P_e^{(u)} \log(L_u) + I(M_u; U_0^T).
\end{aligned}$$

Here $P_e^{(y)}$ and $P_e^{(z)}$ are the average probability of error at the first and second receiver respectively. Since $M_u \rightleftarrows U_0^T \rightleftarrows \mathscr{D}_u^T(U_0^T)$ is a Markov chain, $I(M_u; U_0^T) \geq I(M_u; \mathscr{D}_u^T(U_0^T))$. Then applying Lemma 15 gives (a), and (b) is an application of Fano's inequality. Hence

$$\begin{aligned}
\tilde{R}_u &\leq \frac{1}{T(1 - P_e^{(u)})} \left(I(M_u; U_0^T) + H(P_e^{(u)})\right) \\
&\leq \frac{1}{T(1-\epsilon)} \left(I(M_u; U_0^T) + H(\epsilon)\right). \qquad (2.52)
\end{aligned}$$

Thus

$$\begin{aligned}
R_u \leq \frac{\log(L_u)}{T} + \epsilon &= \tilde{R}_u + \epsilon \\
&\leq \frac{1}{T(1-\epsilon)} \left(I(M_u; U_0^T) + H(\epsilon)\right) + \epsilon. \qquad (2.53)
\end{aligned}$$

Now consider

$$\begin{aligned}
\frac{1}{T} I(M_y; Y_0^T) &\leq \frac{1}{T} I(M_y; M_z Y_0^T) \\
&\overset{(a)}{=} \frac{1}{T} I(M_y; Y_0^T | M_z) \\
&\overset{(b)}{=} \frac{1}{T} I(M_y M_z; Y_0^T) - \frac{1}{T} I(M_z; Y_0^T) \\
&\overset{(c)}{\leq} \frac{1}{T} I(X_0^T; Y_0^T) - \frac{1}{T} I(M_z; Y_0^T) \\
&\overset{(d)}{=} \frac{1}{T} I(X_0^T; Y_0^T | M_z) \qquad (2.54)
\end{aligned}$$

$$\begin{aligned}
&\overset{(e)}{=} \frac{1}{T}\int_0^T \left(\mathbb{E}[\phi_y(X_t)] - \mathbb{E}[\phi_y(\mathbb{E}[X_t|Y_0^t, M_z])]\right)\, dt. \\
&\overset{(f)}{=} \mathbb{E}[\phi_y(X_S)] - \mathbb{E}[\phi_y(\mathbb{E}[X_S|Y_0^S, M_z])]. \qquad (2.55)
\end{aligned}$$

Here, (a) is due to the independence of M_y and M_z,

(b) due to an application of Kolmogrov's formula,

(c) follows since $M_y, M_z \rightleftarrows X_0^T \rightleftarrows Y_0^T$ forms a Markov chain,

(d) follows since $M_z \rightleftarrows X_0^T \rightleftarrows Y_0^T$ forms a Markov chain,

(e) is an application of Theorem 1, and

(f) follows by defining S to be a random variable uniformly distributed on $[0, T]$, and independent of all σ-fields on $(\Omega, \mathcal{F})$.[3]

Similarly,

$$\begin{aligned}
\frac{1}{T} I(M_z; Z_0^T) &\overset{(a)}{=} \frac{1}{T}\int_0^T \mathbb{E}[\phi_z(\mathbb{E}[X_t|Z_t^T, M_z])] - \mathbb{E}[\phi_z(\mathbb{E}[X_t|Z_t^T])]\, dt \\
&\overset{(b)}{\leq} \frac{1}{T}\int_0^T \mathbb{E}[\phi_z(\mathbb{E}[X_t|Z_t^T, M_z])]\, dt - \phi_z\left(\frac{1}{T}\int_0^T \mathbb{E}[X_t]\, dt\right) \\
&\overset{(c)}{\leq} \frac{1}{T}\int_0^T \mathbb{E}[\phi_z(\mathbb{E}[X_t|Y_0^t, M_z])]\, dt - \phi_z\left(\frac{1}{T}\int_0^T \mathbb{E}[X_t]\, dt\right) \\
&\overset{(d)}{=} \mathbb{E}[\phi_z(\mathbb{E}[X_S|Y_0^S, M_z])] - \phi_z(\mathbb{E}[X_S]). \qquad (2.56)
\end{aligned}$$

Here, (a) follows from Theorem 1,

(b) from Jensen's inequality applied to the convex function ϕ_z,

(c) is due to Lemma 12, and

(d) holds since S is the random variable, uniformly distributed on $[0, T]$ and independent of all other variables.

Since the capacity region is convex, to show that the rate-pair (R_y, R_z) is contained in the region in the statement of the theorem, we use a supporting-

[3] S can be defined by extending the probability space $(\Omega, \mathcal{F})$ in (2.40) to $(\Omega \times [0, T], \mathcal{F} \otimes \mathfrak{B}([0, T]))$, where $\mathfrak{B}([0, T])$ is the Borel σ-field on $[0, T]$.

hyperplane argument. It suffices to show that for any $\mu_y, \mu_z \geq 0$,

$$\sup_{R_y, R_z} \mu_y R_y + \mu_z R_z \leq \sup_{\substack{0 \leq \alpha \leq 1/2 \\ 0 \leq p, q, \leq 1}} \mu_y C_y + \mu_z C_z.$$

Note that (2.52), (2.55), and (2.56) imply

$$\mu_y R_y + \mu_z R_z \leq \mu_y \mathbb{E}[\phi_y(X_S)] - \mu_z \phi_z(\mathbb{E}[X_S]) - \mathbb{E}[K_\mu(\mathbb{E}[X_S|Y_0^S, M_z])] + \varepsilon(\epsilon), \tag{2.57}$$

where

$$K_\mu(x) = \mu_y \phi_y(x) - \mu_z \phi_z(x), \tag{2.58}$$

and $\varepsilon(\epsilon) \to 0$ as $\epsilon \to 0$. We now use Fenchel-Eggleston-Carathéodory's theorem [15, Lemma 15.4, Chapter 15, p. 310]. Since $K_\mu(x)$ is a continuous function, there exist $0 \leq \alpha \leq 1/2$, $0 \leq p, q \leq 1$ such that

$$\mathbb{E}[K_\mu(\mathbb{E}[X_S|Y_0^S, M_z])] = \alpha K_\mu(p) + (1-\alpha) K_\mu(q), \tag{2.59}$$

$$\mathbb{E}[X_S] = \mathbb{E}[\mathbb{E}[X_S|Y_0^S, M_z]] = \alpha p + (1-\alpha) q. \tag{2.60}$$

Due to the convexity of $\phi_y(x)$ and $0 \leq X_S \leq 1$ with $\mathbb{E}[X_S] = \alpha p + (1-\alpha)q$,

$$\begin{aligned} \mathbb{E}[\phi_y(X_S)] &\leq \mathbb{E}[X_S]\phi_y(1) + (1 - \mathbb{E}[X_S])\phi_y(0) \\ &= (\alpha p + (1-\alpha)q)\phi_y(1) + (\alpha(1-p) + (1-\alpha)(1-q))\phi_y(0). \end{aligned} \tag{2.61}$$

Equations (2.57)-(2.61) give

$$\begin{aligned} \mu_y R_y + \mu_z R_z \leq & \mu_y((\alpha p + (1-\alpha)q)\phi_y(1) + (\alpha(1-p) + (1-\alpha)(1-q))\phi_y(0)) - \mu_z \phi_z(\alpha p + (1-\alpha)q \\ & - \alpha(\mu_y \phi_y(p) - \mu_z \phi_z(p)) - (1-\alpha)(\mu_y \phi_y(q) - \mu_z \phi_z(q)) + \varepsilon(\epsilon) \\ = & \mu_y \left[(\alpha p + (1-\alpha)q)\phi_y(1) + (\alpha(1-p) + (1-\alpha)(1-q))\phi_y(0) - \alpha \phi_y(p) - (1-\alpha)\phi_y(\right. \\ & + \mu_z \left[\alpha \phi_z(p) + (1-\alpha)\phi_z(q) - \phi_z(\alpha p + (1-\alpha)q)\right] + \varepsilon(\epsilon) \\ = & \mu_y C_y + \mu_z C_z + \varepsilon(\epsilon) \\ \leq & \sup_{\substack{0 \leq \alpha \leq 1/2 \\ 0 \leq p, q, \leq 1}} \mu_y C_y + \mu_z C_z + \varepsilon(\epsilon). \end{aligned} \tag{2.62}$$

Since ϵ is arbitrary, taking $\epsilon \to 0$ we get the converse part of the theorem.

2.6 More Capable Poisson Wiretap Channel

2.6.1 Encoding and Decoding

Here we will consider the first receiver to be the legitimate user and the second receiver to be an eavesdropper. The transmitter (equipped with a stochastic encoder $\mathscr{E}_x^T$) wishes to communicate a message M, which is uniformly distributed on $\mathcal{M} = \{1, \ldots, L\}$, to the legitimate user (equipped with decoder $\mathscr{D}_y^T$). To transmit message $M = m$, the encoder chooses an input waveform $X_0^T \in \mathcal{X}_0^T$. Upon observing Y_0^T, the legitimate decoder chooses a symbol $\hat{M} \in \mathcal{M}$. We will call such an arrangement an (L, T) *code*. The average probability of error at the legitimate receiver is

$$\mathrm{P}_e = \frac{1}{L} \sum_{m=1}^{L} P(\mathscr{D}_y^T(Y_0^T) \neq m | M = m). \tag{2.63}$$

The metric to measure the secrecy will be $\frac{1}{T} I(M; Z_0^T)$.

Definition 8 *A secrecy rate R_s is said to be* achievable *for the Poisson wiretap channel if for all $\epsilon > 0$ and for all sufficiently large T, there exists an (L, T) code such that*

$$\begin{aligned} \frac{\log(L)}{T} &\geq R_s - \epsilon \\ P_e &\leq \epsilon \\ \frac{1}{T} I(M; Z_0^T) &\leq \epsilon. \end{aligned} \tag{2.64}$$

The secrecy capacity is defined to be the supremum of achievable secrecy rate.

Theorem 5 (Capacity of more capable Poisson wiretap channel) *The secrecy capacity of the more capable Poisson wiretap channel is*

$$C_s = \max_{0\leq\alpha\leq 1} \alpha\Phi(1) + (1-\alpha)\Phi(0) - \Phi(\alpha), \tag{2.65}$$

where we recall $\Phi(x) = \phi_y(x) - \phi_z(x)$ *and* $\Phi(x)$ *is a convex function.*

Note that this capacity expression is same as that of the capacity of the degraded Poisson wiretap channel in [39]. Since the achievability argument is identical to that for the degraded Poisson wiretap channel in [39, Section III], we shall only prove the converse here.

2.6.2 Converse

Suppose R_s is achievable. Then there exists an (L, T) code satisfying (2.64). Let $R = \frac{\log(L)}{T}$, then

$$\begin{aligned} RT = \log(L) = H(M) =& \mathbb{E}\left[H(M|Y_0^T)\right] + I(M;Y_0^T) \\ &\overset{(a)}{\leq} H\left(M|\mathscr{D}_y^T(Y_0^T)\right) + I(M;Y_0^T) \\ &\overset{(b)}{\leq} H(\mathrm{P}_e) + \mathrm{P}_e\log(L) + I(M;Y_0^T). \end{aligned}$$

Since $M \rightleftarrows Y_0^T \rightleftarrows \mathscr{D}_y^T(Y_0^T)$ is a Markov chain, $I(M;Y_0^T) \geq I(M;\mathscr{D}_y^T(Y_0^T))$. Then applying Lemma 15 gives (a), and (b) is an application of Fano's inequality. This gives

$$\begin{aligned} R &\leq \frac{1}{T(1-P_e)}(I(M;Y_0^T) + H(\mathrm{P}_e)) \\ &= \frac{1}{T(1-P_e)}(I(M;Y_0^T) - I(M;Z_0^T) + H(\mathrm{P}_e) + I(M;Z_0^T)) \\ &\leq \frac{1}{T(1-\epsilon)}((I(M;Y_0^T) - I(M;Z_0^T) + H(\epsilon)) + \frac{\epsilon}{1-\epsilon}. \end{aligned}$$

Now consider

$$
\begin{aligned}
\frac{1}{T}\left(I(M;Y_0^T)-I(M;Z_0^T)\right) &\overset{(a)}{=}\frac{1}{T}\int_0^T \mathbb{E}[\phi_y(\mathbb{E}[X_t|Y_0^t,M])]-\mathbb{E}[\phi_y(\mathbb{E}[X_t|Y_0^t])]\,dt\\
&\quad-\frac{1}{T}\int_0^T \mathbb{E}[\phi_z(\mathbb{E}[X_t|Z_t^T,M])]-\mathbb{E}[\phi_z(\mathbb{E}[X_t|Z_t^T])]\,dt\\
&=\frac{1}{T}\int_0^T \mathbb{E}[\phi_y(\mathbb{E}[X_t|Y_0^t,M])]-\mathbb{E}[\phi_z(\mathbb{E}[X_t|Z_t^T,M])]\,dt\\
&\quad-\frac{1}{T}\int_0^T \mathbb{E}[\phi_y(\mathbb{E}[X_t|Y_0^t])]-\mathbb{E}[\phi_z(\mathbb{E}[X_t|Z_t^T])]\,dt\\
&\overset{(b)}{=}\frac{1}{T}\int_0^T \mathbb{E}[\phi_y(\mathbb{E}[X_t|Y_0^t,Z_t^T,M])]-\mathbb{E}[\phi_z(\mathbb{E}[X_t|Y_0^t,Z_t^T,M])]\,dt\\
&\quad-\frac{1}{T}\int_0^T \mathbb{E}[\phi_y(\mathbb{E}[X_t|Y_0^t,Z_t^T])]-\mathbb{E}[\phi_z(\mathbb{E}[X_t|Y_0^t,Z_t^T])]\,dt\\
&=\frac{1}{T}\int_0^T \mathbb{E}[\Phi(\mathbb{E}[X_t|Y_0^t,Z_t^T,M])]-\mathbb{E}[\Phi(\mathbb{E}[X_t|Y_0^t,Z_t^T])]\,dt\\
&\overset{(c)}{\leq}\frac{1}{T}\int_0^T \mathbb{E}[\Phi(X_t)]-\Phi(\mathbb{E}[X_t])\,dt\\
&\overset{(d)}{=}\max_{0\leq\alpha\leq 1}\alpha\Phi(1)+(1-\alpha)\Phi(0)-\Phi(\alpha)\\
&=C_s.
\end{aligned}
$$

Here, for (a) we have used Theorem 1,

for (b) we have used Theorem 2,

for (c) we have applied Jensen's inequality to both terms in the integral, and

(d) follows from fixing the mean of the input distribution to α and maximizing over all such distributions and then maximizing over α. Due to the convexity of $\Phi(x)$, the maximizing distribution puts mass on the extreme points $\{0,1\}$, that is, mass $1-\alpha$ on 0 and mass α on 1.

Hence we get,

$$
\begin{aligned}
R_s &\leq \frac{\log(L)}{T}+\epsilon\\
&\leq \frac{C_s}{1-\epsilon}+\frac{H(\epsilon)}{T(1-\epsilon)}+\frac{\epsilon}{1-\epsilon}.
\end{aligned}
$$

Since ϵ is arbitrary, taking $\epsilon\to 0$ we get the converse part of the theorem.

2.7 General Poisson Broadcast Channel with Degraded Message Sets

In this setting the transmitter has a common message $M_o \in \mathcal{M}_0 = \{1, \ldots, L_0\}$ for both of the users and a private message $M_y \in \mathcal{M}_y = \{1, \ldots, L_y\}$ for the first user. Messages M_0 and M_y are assumed to be independent and uniformly distributed on their respective support. The transmitter uses an encoder $\mathscr{E}_x^T$ which maps these messages into an input X_0^T

$$\mathscr{E}_x^T : \mathcal{M}_0 \times \mathcal{M}_y \to \mathcal{X}_0^T.$$

Upon observing Y_0^T, the first receiver estimates both common and private messages using decoder $\mathscr{D}_y^T$

$$\mathscr{D}_y^T : \mathcal{N}_0^T \to \mathcal{M}_0 \times \mathcal{M}_y.$$

Similarly the second receiver employs $\mathscr{D}_z^T$ to decode the common message

$$\mathscr{D}_z^T : \mathcal{N}_0^T \to \mathcal{M}_0.$$

We will call the above setup an (L_0, L_y, T) *code*. The average probability of error of this code is

$$\mathrm{P}_e = \frac{1}{L_0 L_y} \sum_{m_0=1, m_y=1}^{L_0, L_y} P\left\{ \{\mathscr{D}_y^T(Y_0^T) \neq (m_0, m_y)\} \bigcup \{\mathscr{D}_z^T(Z_0^T) \neq m_0\} \middle| M_0 = m_0, M_y = m_y \right\}.$$

The rate pair (R_0, R_y) is said to be *achievable* if for any $\epsilon > 0$ and for any sufficiently large T, there exists an (L_0, L_y, T) code such that

$$\begin{aligned} \frac{\log(L_0)}{T} &\geq R_0 - \epsilon \\ \frac{\log(L_y)}{T} &\geq R_y - \epsilon \\ \mathrm{P}_e &\leq \epsilon. \end{aligned} \tag{2.66}$$

The capacity region is the closure of the achievable rate pairs. Let $P_{e,0}^{(y)}$, $P_{e,y}^{(y)}$ denote the average probability of error in decoding messages M_0 and M_y, respectively, at the first receiver and similarly let $P_{e,0}^{(z)}$ denote the average probability of error at the second receiver. Then for a given code

$$\max(P_{e,0}^{(y)}, P_{e,y}^{(y)}, P_{e,0}^{(z)}) \leq \mathrm{P}_e \leq P_{e,0}^{(y)} + P_{e,y}^{(y)} + P_{e,0}^{(z)}. \tag{2.67}$$

Theorem 6 (Capacity the Poisson broadcast channel with degraded message sets) *The capacity region of the general Poisson broadcast channel with degraded message sets is given by the union over all* $0 \leq \alpha_i, p_i \leq 1$, $i = 1, 2, 3$ *with* $\sum_{i=1}^{3} \alpha_i = 1$ *of rate pairs satisfying:*

$$R_0 \leq C_z$$
$$R_0 + R_y \leq \hat{C}_y + \tilde{C}_y$$
$$R_0 + R_y \leq C_z + \hat{C}_y,$$

where

$$C_z = \sum_{i=1}^{3} \alpha_i \phi_z(p_i) - \phi_z\left(\sum_{i=1}^{3} \alpha_i p_i\right)$$
$$\hat{C}_y = \sum_{i=1}^{3} \alpha_i \left(p_i \phi_y(1) + (1-p_i)\phi_y(0) - \phi_y(p_i)\right)$$
$$\tilde{C}_y = \sum_{i=1}^{3} \alpha_i \phi_y(p_i) - \phi_y\left(\sum_{i=1}^{3} \alpha_i p_i\right).$$

2.7.1 Achievability

We will show the achievability of the formally larger region:

$$R_y \leq \hat{C}_y$$

$$R_0 \leq C_z$$

$$R_0 + R_y \leq \hat{C}_y + \tilde{C}_y. \tag{2.68}$$

The above region turns out to equal the region in the statement of the theorem, which will follow from the converse proven later. To see that the region in (2.68) indeed contains the one given in the theorem, it suffices to show that the rate pair $\bar{R}_0 = C_z > 0$ and $\bar{R}_y = \min((\hat{C}_y + \tilde{C}_y - C_z), \hat{C}_y) > 0$ is in (2.68). This follows since $\bar{R}_0$ and $\bar{R}_y$ satisfy

$$\bar{R}_y \leq \hat{C}_y$$

$$\bar{R}_0 = C_z$$

$$\bar{R}_0 + \bar{R}_y \leq \hat{C}_y + \tilde{C}_y.$$

We use superposition coding and a similar argument as that used in the achievability proof for the more capable Poisson broadcast channel with independent message sets. We divide the interval $[0, T_n]$ into n intervals each of equal length $\tau = T_n/n$. Here we take $V_0^{T_n}$ to be a ternary stochastic process. The process will be constant on each of these sub-interval with value given by

$$V_t = \bar{V}_i \text{ for } (i-1)\tau \leq t < i\tau, \quad i = 1, 2, \ldots, n \tag{2.69}$$

where $\bar{V}_i$ are independent and identically distributed random variables with

$$P(\bar{V}_i = j) = \alpha_j, j \in \{1, 2, 3\}. \tag{2.70}$$

We construct the input processes, X_0^T, as binary and piecewise constant with

$$X_t = \bar{X}_i \text{ for } (i-1)\tau \leq t < i\tau, \quad i = 1, 2, \ldots, n, \tag{2.71}$$

and

$$P(\bar{X}_i = 1 | \bar{V}_i = j) = 1 - P(\bar{X}_i = 0 | \bar{V}_i = j) = p_j, j \in \{1, 2, 3\}. \tag{2.72}$$

Lemma 11 gives that for all $\epsilon > 0$ there exists $\bar{\tau}$ and N such that if $n \geq N$ and $\tau \leq \bar{\tau}$ then

$$P\left(\left|\frac{1}{T_n}\mathfrak{i}(V_0^{T_n}; Z_0^{T_n}) - C_z\right| > \epsilon\right) \leq \epsilon$$
$$P\left(\left|\frac{1}{T_n}\mathfrak{i}(X_0^{T_n}; Y_0^{T_n}) - (C_y + \tilde{C}_y\right)| > \epsilon\right) \leq \epsilon$$
$$P\left(\left|\frac{1}{T_n}\mathfrak{i}(X_0^{T_n}; Y_0^{T_n}|V_0^{T_n}) - C_y\right| > \epsilon\right) \leq \epsilon. \tag{2.73}$$

Encoding and Decoding Operation

Let (R_0, R_y) be strictly positive, satisfying (2.68), and let $\tilde{R}_u = R_u - \delta$, $u \in \{0, y\}$ for some $\delta > 0$. We generate $L_0 = \exp(T_n \tilde{R}_0)$ many $V_0^{T_n}$ waveforms (indexed by $j = 1, \ldots, L_0$) independently according to (2.69) and (2.70). For each $V_0^{T_n}(j)$, we generate $L_y = \exp(T_n \tilde{R}_y)$ many independent $X_0^{T_n}$ waveforms (indexed by $i = 1, \ldots, L_y$) according to (2.71) and (2.72). To transmit messages (M_0, M_y), the encoder sends $X_0^{T_n}(M_0, M_y)$ over the channel.

Both of the receivers consider only those inputs for which the mutual information densities (in Definition 7) evaluate to a finite value (computed using Theorem 1) for given received point process. The first receiver seeks unique i and j that satisfy both

$$\frac{1}{T_n}\mathfrak{i}(X_0^{T_n}(i,j); Y_0^{T_n}) \geq C_y + \tilde{C}_y - \gamma_y \tag{2.74}$$

and

$$\frac{1}{T_n}\mathfrak{i}(X_0^{T_n}(i,j); Y_0^{T_n}|V_0^{T_n}(j)) \geq C_y - \gamma_y. \tag{2.75}$$

The second decoder finds the unique j such that

$$\frac{1}{T_n}\mathfrak{i}(V_0^{T_n}(j); Z_0^{T_n}) \geq C_z - \gamma_z \tag{2.76}$$

for some $\gamma_z > 0$. Without loss of generality assume that $X_0^T(1,1)$ was transmitted over the channel. Using a similar argument as that for the error analysis in the achievability proof of the more capable channel with independent messages we get the following. Since

$$\begin{aligned} \tilde{R}_0 + \tilde{R}_y &= C_y + \tilde{C}_y - 2\delta \\ \tilde{R}_y &= C_y - \delta, \end{aligned} \tag{2.77}$$

the expectation (over random codebook generation) of the average probability of error at the first receiver can be made arbitrarily small. Similarly, as $\tilde{R}_0 = C_z - \delta$, the expectation of the average probability of error at the second receiver can be made arbitrarily low. Hence there exists a sequence of codebooks which achieve the rates in (2.68) with arbitrarily low probability of error.

2.7.2 Converse

For a given sequence of (L_0, L_y, T) codes, using Lemma 15 and Fano's inequality, we get

$$\begin{aligned} R_0 &\le \frac{1}{T(1-\epsilon)}\big(I(M_0; Z_0^T) + H(\epsilon)\big) + \epsilon \\ R_y &\le \frac{1}{T(1-\epsilon)}\big(I(M_y; Y_0^T) + H(\epsilon)\big) + \epsilon \\ R_0 + R_y &\le \frac{1}{T(1-\epsilon)}\big(I(M_0, M_y; Y_0^T) + H(\epsilon)\big) + 2\epsilon, \end{aligned}$$

where we have used the fact that the first user needs to decode both M_0 and M_y, whereas second receiver requires only M_0. We now upper bound the mutual information expressions in the above inequalities.

$$\frac{1}{T} I(M_0; Z_0^T) \overset{(a)}{=} \frac{1}{T}\int_0^T \mathbb{E}[\phi_z(\mathbb{E}[X_t|Z_t^T, M_0])] - \mathbb{E}[\phi_z(\mathbb{E}[X_t|Z_t^T])]\, dt$$

$$
\begin{aligned}
&\overset{(b)}{\leq} \frac{1}{T}\int_0^T \mathbb{E}[\phi_z(\mathbb{E}[X_t|Z_t^T, M_0])]\,dt - \phi_z\left(\frac{1}{T}\int_0^T \mathbb{E}[X_t]\,dt\right) \\
&\overset{(c)}{\leq} \frac{1}{T}\int_0^T \mathbb{E}[\phi_z(\mathbb{E}[X_t|Y_0^t, Z_t^T, M_0])]\,dt - \phi_z\left(\frac{1}{T}\int_0^T \mathbb{E}[X_t]\,dt\right) \\
&\overset{(d)}{=} \mathbb{E}[\phi_z(\mathbb{E}[X_S|Z_S^T, Y_0^S, M_0])] - \phi_z(\mathbb{E}[X_S]). \qquad (2.78)
\end{aligned}
$$

In (a), we have used Theorem 1,

in (b) and (c), we have applied Jensen's inequality to the second and first terms in the integrand, respectively, and

in (d), we have defined S to be a random variable, uniformly distributed on $[0, T]$ and independent of all other random variables and processes. Now consider $\frac{1}{T}I(M_0, M_y; Y_0^T)$.

$$
\begin{aligned}
\frac{1}{T}I(M_0, M_y; Y_0^T) &\overset{(a)}{\leq} \frac{1}{T}I(X_0^T; Y_0^T) \\
&= \frac{1}{T}\int_0^T \mathbb{E}[\phi_y(X_t)] - \mathbb{E}[\phi_y(\mathbb{E}[X_t|Y_0^t])]\,dt \\
&\overset{(b)}{\leq} \frac{1}{T}\int_0^T \mathbb{E}[\phi_y(X_t)]\,dt - \phi_y\left(\frac{1}{T}\int_0^T \mathbb{E}[X_t]\,dt\right) \\
&\overset{(c)}{=} \mathbb{E}[\phi_y(X_S)] - \phi_y\left(\mathbb{E}[X_S]\right). \qquad (2.79)
\end{aligned}
$$

Here (a) is due to the Markov chain $(M_0, M_y) \rightleftarrows X_0^T \rightleftarrows Y_0^T$,

(b) is due Jensen's inequality, and

(c) follows because S is a uniformly distributed on $[0, T]$.

Similar to (2.54), we can show

$$
R_y \leq \frac{1}{T(1-\epsilon)}(I(X_0^T; Y_0^T|M_0) + H(\epsilon)) + \epsilon.
$$

Now consider

$$
\begin{aligned}
\frac{1}{T}I(X_0^T; Y_0^T|M_0) + \frac{1}{T}I(M_0; Z_0^T) \overset{(a)}{=} &\frac{1}{T}\int_0^T \mathbb{E}[\phi_y(X_t)] - \mathbb{E}[\phi_y(\mathbb{E}[X_t|Y_0^t, M_0])]\,dt \\
&+ \frac{1}{T}\int_0^T \mathbb{E}[\phi_z(\mathbb{E}[X_t|Z_t^T, M_0])] - \mathbb{E}[\phi_z(\mathbb{E}[X_t|Z_t^T])]\,dt
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{1}{T}\int_0^T \mathbb{E}[\phi_y(X_t)] - \mathbb{E}[\phi_z(\mathbb{E}[X_t|Z_t^T])]\,dt \\
&\quad + \frac{1}{T}\int_0^T \mathbb{E}[\phi_z(\mathbb{E}[X_t|Z_t^T, M_0])] - \mathbb{E}[\phi_y(\mathbb{E}[X_t|Y_0^t, M_0])]\,dt \\
&\overset{(b)}{\leq} \frac{1}{T}\int_0^T \mathbb{E}[\phi_y(X_t)]\,dt - \phi_z\left(\frac{1}{T}\int_0^T \mathbb{E}[X_t]\,dt\right) \\
&\quad + \frac{1}{T}\int_0^T \mathbb{E}[\phi_z(\mathbb{E}[X_t|Z_t^T, M_0])] - \mathbb{E}[\phi_y(\mathbb{E}[X_t|Y_0^t, M_0])]\,dt \\
&\overset{(c)}{=} \frac{1}{T}\int_0^T \mathbb{E}[\phi_y(X_t)]\,dt - \phi_z\left(\frac{1}{T}\int_0^T \mathbb{E}[X_t]\,dt\right) \\
&\quad + \frac{1}{T}\int_0^T \mathbb{E}[\phi_z(\mathbb{E}[X_t|Y_0^t, Z_t^T, M_0])] - \mathbb{E}[\phi_y(\mathbb{E}[X_t|Y_0^t, Z_t^T, M_0])]\,dt \\
&\overset{(d)}{=} \mathbb{E}[\phi_y(X_S)] - \phi_z\left(\mathbb{E}[X_S]\right) \\
&\quad + \mathbb{E}[\phi_z(\mathbb{E}[X_S|Y_0^S, Z_S^T, M_0])] - \mathbb{E}[\phi_y(\mathbb{E}[X_S|Y_0^S, Z_S^T, M_0])].
\end{aligned} \tag{2.80}
$$

Here, (a) is due to Theorem 1,

(b) is due to Jensen's inequality,

(c) is due to Theorem 2, and

(d) follows because S is uniformly distributed on $[0, T]$ and independent of all other random variables.

Now we use Fenchel-Eggleston-Carathéodory's theorem [15, Lemma 15.4, Chapter 15, p. 310]. Since $\phi_y(x)$ and $\phi_z(x)$ are continuous functions, there exist $0 \leq p_1, p_2, p_3 \leq 1$ and $0 \leq \alpha_1, \alpha_2, \alpha_3 \leq 1$ with $\sum_{i=1}^3 \alpha_i = 1$ such that

$$
\begin{aligned}
\mathbb{E}[\phi_y(\mathbb{E}[X_S|Z_S^T, Y_0^S, M_0])] &= \sum_{i=1}^3 \alpha_i \phi_y(p_i) \\
\mathbb{E}[\phi_z(\mathbb{E}[X_S|Z_S^T, Y_0^S, M_0])] &= \sum_{i=1}^3 \alpha_i \phi_z(p_i) \\
\mathbb{E}[X_S] = \mathbb{E}[\mathbb{E}[X_S|Z_S^T, Y_0^S, M_0]] &= \sum_{i=1}^3 \alpha_i p_i.
\end{aligned} \tag{2.81}
$$

Due to the convexity of ϕ_u,

$$\begin{aligned}\mathbb{E}[\phi_u(X_S)] &\leq \mathbb{E}[X_S]\phi_u(1) + (1-\mathbb{E}[X_S])\phi_u(0)\\ &= \sum_{i=1}^{3}\alpha_i p_i \phi_u(1) + \left(1-\sum_{i=1}^{3}\alpha_i p_i\right)\phi_u(0)\\ &= \sum_{i=1}^{3}\alpha_i\left(p_i\phi_u(1) + (1-p_i)\phi_u(0)\right). \end{aligned} \tag{2.82}$$

Substituting we get the following. From (2.78)

$$\begin{aligned}R_0 &\leq \mathbb{E}[\phi_z(\mathbb{E}[X_S|Z_S^T, Y_0^S, M_0])] - \phi_z(\mathbb{E}[X_S]) + \varepsilon(\epsilon)\\ &= \sum_{i=1}^{3}\alpha_i\phi_z(p_i) - \phi_z\left(\sum_{i=1}^{3}\alpha_i p_i\right) + \varepsilon(\epsilon)\\ &= C_z + \varepsilon(\epsilon)\end{aligned}$$

where $\varepsilon(\epsilon) \to 0$ as $\epsilon \to 0$. From (2.79) we get

$$\begin{aligned}R_0 + R_y &\leq \mathbb{E}[\phi_y(X_S)] - \phi_y\left(\mathbb{E}[X_S]\right) + \varepsilon'(\epsilon)\\ &\leq \sum_{i=1}^{3}\alpha_i\left(p_i\phi_y(1) + (1-p_i)\phi_y(0)\right) - \phi_y\left(\sum_{i=1}^{3}\alpha_i p_i\right) + \varepsilon'(\epsilon)\\ &= \hat{C}_y + \tilde{C}_y + \varepsilon'(\epsilon).\end{aligned}$$

where $\varepsilon'(\epsilon) \to 0$ as $\epsilon \to 0$. Finally (2.80) gives

$$\begin{aligned}R_0 + R_y &\leq \mathbb{E}[\phi_y(X_S)] - \phi_z\left(\mathbb{E}[X_S]\right) + \mathbb{E}[\phi_z(\mathbb{E}[X_S|Y_0^S, Z_S^T, M_0])] - \mathbb{E}[\phi_y(\mathbb{E}[X_S|Y_0^S, Z_S^T, M_0])] + \varepsilon(\epsilon)''\\ &\leq \sum_{i=1}^{3}\alpha_i\left(p_i\phi_y(1) + (1-p_i)\phi_y(0)\right) - \phi_z\left(\sum_{i=1}^{3}\alpha_i p_i\right)\\ &\quad + \sum_{i=1}^{3}\alpha_i\phi_z(p_i) - \sum_{i=1}^{3}\alpha_i\phi_y(p_i) + \varepsilon''(\epsilon)\\ &= \hat{C}_y + C_z + \varepsilon''(\epsilon),\end{aligned}$$

where $\varepsilon''(\epsilon) \to 0$ as $\epsilon \to 0$. As ϵ is arbitrary, taking $\epsilon \to 0$ completes the converse argument.

2.8 Proofs of Lemmas

Proof of Lemma 3: Let $[s,t] \in [0,T]$, and $k \in \mathbf{N}$ then

$$\begin{aligned} P(\tilde{N}_t - \tilde{N}_s = k|\tilde{X}_0^T) &= P(N_{(T-s)-} - N_{(T-t)-} = k|X_0^T) \\ &= \frac{1}{k!}\left(\int_{T-t}^{T-s} X_\tau \, d\tau\right)^k \exp\left(-\int_{T-t}^{T-s} X_\tau \, d\tau\right) \\ &= \frac{1}{k!}\left(\int_t^s \tilde{X}_\tau \, d\tau\right)^k \exp\left(-\int_t^s \tilde{X}_\tau \, d\tau\right), \end{aligned} \tag{2.83}$$

where we have used the fact that since X_0^T is càdlàg, the set $\{t : X_{t-} \neq X_t, t \in [0,T]\}$ is at most countable [7, Section 12, Lemma 1, p. 122]. Since the new process $\tilde{N}_0^T$ is obtained by time reversing the process N_0^T, it has the independent increment property. □

Proof of Lemma 4: For $0 \leq s < t \leq T$

$$\begin{aligned} \mathbb{E}[\hat{N}_t - \hat{N}_s|\mathcal{F}_s] &= \mathbb{E}[\hat{N}_t - \hat{N}_s|A, \Lambda_0^T, \hat{N}_0^s] \\ &\stackrel{(a)}{=} \mathbb{E}[\hat{N}_t - \hat{N}_s|\Lambda_0^T, \hat{N}_0^s] \\ &\stackrel{(b)}{=} \int_s^t \hat{\Lambda}_u \, du \\ &\stackrel{(c)}{=} \mathbb{E}\left[\int_s^t \hat{\Lambda}_u \, du \Big| \mathcal{F}_s\right]. \end{aligned} \tag{2.84}$$

Here, (a) is due to the fact that if $A \rightleftarrows \Lambda_0^T \rightleftarrows (\hat{N}_0^s, \hat{N}_s^T)$ is a Markov chain then so is $A \rightleftarrows (\Lambda_0^T, \hat{N}_0^s) \rightleftarrows \hat{N}_s^T$ [31, Proposition 6.8, p.111], and then using [31, Proposition 6.6, p.111],

(b) is due to Definition 1 and the independent increment property of Poisson processes, and

(c) is due to the fact that Λ_0^T is measurable with respect to $\mathcal{F}_t$ for all $t \in [0,T]$.

Then from (2.84) and [9, Chapter II, Section 2, p. 23-24] we get that for all non-negative $(\mathcal{F}_t : t \in [0,T])$-predictable processes C_0^T

$$\mathbb{E}\left[\int_0^T C_s \, d\hat{N}_s\right] = \mathbb{E}\left[\int_0^T C_s \hat{\Lambda}_s \, ds\right]. \tag{2.85}$$

Also, $\hat{\Lambda}_0^T$ is $\mathcal{F}_0$-measurable and thus $(\mathcal{F}_t : t \in [0,T])$-predictable. Hence the $(P, \mathcal{F}_t : t \in [0,T])$-intensity of $\hat{N}_0^T$ is $\hat{\Lambda}_0^T$.

Let D_0^T be a non-negative $(\mathcal{G}_t : t \in [0,T])$-predictable process. As $\mathcal{G}_t \subseteq \mathcal{F}_t$, it is also $(\mathcal{F}_t : t \in [0,T])$-predictable. Hence

$$\mathbb{E}\left[\int_0^T D_s \, d\hat{N}_s\right] = \mathbb{E}\left[\int_0^T D_s \hat{\Lambda}_s \, ds\right]. \tag{2.86}$$

Let $\Pi_t = \mathbb{E}[\hat{\Lambda}_t | \mathcal{G}_{t-}]$, $t \in [0,T]$. Then the process Π_0^T is $(\mathcal{G}_t : t \in [0,T])$-predictable [18, Chapter 6, Theorem 43, p. 103]. Hence

$$\begin{aligned}
\mathbb{E}\left[\int_0^T D_s \Pi_s \, ds\right] &= \mathbb{E}\left[\int_0^T D_s \mathbb{E}[\hat{\Lambda}_s | \mathcal{G}_{s-}] \, ds\right] \\
&\stackrel{(a)}{=} \mathbb{E}\left[\int_0^T \mathbb{E}[D_s \hat{\Lambda}_s | \mathcal{G}_{s-}] \, ds\right] \\
&= \mathbb{E}\left[\int_0^T D_s \hat{\Lambda}_s \, ds\right] \\
&\stackrel{(b)}{=} \mathbb{E}\left[\int_0^T D_s \, d\hat{N}_s\right].
\end{aligned}$$

Here, (a) is due to the fact that D_s is $\mathcal{G}_{s-}$ measurable [9, Exercise E10, Chapter I, p. 9], and

(b) is due to (3.66).

Hence the $(P, \mathcal{G}_t : t \in [0,T])$-intensity of $\hat{N}_0^T$ is Π_0^T. Since for each $t \in [0,T]$, $\hat{N}_{t-} = \hat{N}_t$ P-a.s., we can take

$$\Pi_t = \mathbb{E}[\hat{\Lambda}_t | \mathcal{G}_{t-}] = \mathbb{E}[\hat{\Lambda}_t | \mathcal{G}_t] \quad P\text{-a.s.}$$

□

Proof of Lemma 5: Using the data processing inequality

$$I(A;\hat{U}_0^T) = I(A;U_{t_1}^{t_2}) \leq I(X_0^T;U_{t_1}^{t_2})$$
$$\leq I(X_0^T;U_0^T) < \infty,$$

where the last inequality is due to [17, 30, 68]. Hence $P^{A,\hat{U}_0^T} \ll P^A \times P^{\hat{U}_0^T}$.

From (2.7) we get that $P^{U_0^T} \ll P_0^{U_0^T}$. Let N be such that $P_0^{\hat{U}_0^T}(\mathsf{N}) = 0$. Then $P_0^{\hat{U}_0^T}(\mathsf{N}) = P_0^{U_0^T}((\hat{U}_0^T)^{-1}\mathsf{N}) = 0$. Hence $P^{U_0^T}((\hat{U}_0^T)^{-1}\mathsf{N}) = P^{\hat{U}_0^T}(\mathsf{N}) = 0$. Thus

$$P^{\hat{U}_0^T} \ll P_0^{\hat{U}_0^T}.$$

This gives $P^A \times P^{\hat{U}_0^T} \ll P^A \times P_0^{\hat{U}_0^T}$ [31, Chapter 1, Exercise 19, p. 22]. □

Proof of Lemma 6: Recall that L_0^T can be written as

$$L_t = \exp\left(\int_0^t \log(\Psi_s) d\hat{U}_s + (1-\Psi_s)\mu_s \, ds\right).$$

We note that for $t \in [0,T]$ L_t satisfies

$$L_t = \begin{cases} L_{t-} & \text{if } \hat{U}_t - \hat{U}_{t-} = 0, \\ \Psi_t L_{t-} & \text{if } \hat{U}_t - \hat{U}_{t-} = 1. \end{cases} \tag{2.87}$$

Let C_0^T be a non-negative $(\mathcal{G}_t : t \in [0,T])$-predictable process. Then

$$\begin{aligned} \mathbb{E}\left[\int_0^T C_t \, d\hat{U}_t\right] &\overset{(a)}{=} \mathbb{E}_{\tilde{P}^{A,\hat{U}_0^T}}\left[L_T \int_0^T C_t \, d\hat{U}_t\right] \\ &\overset{(b)}{=} \mathbb{E}_{\tilde{P}^{A,\hat{U}_0^T}}\left[\int_0^T L_t C_t \, d\hat{U}_t\right] \\ &\overset{(c)}{=} \mathbb{E}_{\tilde{P}^{A,\hat{U}_0^T}}\left[\int_0^T \Psi_t L_{t-} C_t \, d\hat{U}_t\right] \\ &\overset{(d)}{=} \mathbb{E}_{\tilde{P}^{A,\hat{U}_0^T}}\left[\int_0^T \Psi_t L_{t-} C_t \mu_t \, dt\right] \\ &\overset{(e)}{=} \mathbb{E}_{\tilde{P}^{A,\hat{U}_0^T}}\left[\int_0^T \Psi_t L_t C_t \mu_t \, dt\right] \\ &\overset{(f)}{=} \mathbb{E}_{\tilde{P}^{A,\hat{U}_0^T}}\left[\int_0^T \hat{\Psi}_t L_t C_t \, dt\right] \end{aligned}$$

$$\overset{(g)}{=} \mathbb{E}_{\tilde{P}^{A,\hat{U}_0^T}}\left[L_T \int_0^T \hat{\Psi}_t C_t \, dt\right]$$
$$\overset{(h)}{=} \mathbb{E}\left[\int_0^T \hat{\Psi}_t C_t \, dt\right],$$

where, (a) follows since L_T is the Radon-Nikodym derivative $\frac{dP^{A,\hat{U}_0^T}}{d\tilde{P}^{A,\hat{U}_0^T}}$,

(b) follows due to [9, T19 Theorem, Appendix A2, p. 302],

(c) follows due to (3.61),

(d) follows since the $(\tilde{P}^{A,\hat{U}_0^T}, \mathcal{G}_t : t \in [0,T])$-intensity of $\hat{U}_0^T$ is μ_0^T, and L_{t-} being a left-continuous adapted process is $(\mathcal{G}_t : t \in [0,T])$-predictable,

(e) follows since the Lebesgue measure of the set $\{t : t \in [0,T], L_{t-} \neq L_t\}$ is zero due to (3.61),

(f) follows from the definition $\hat{\Psi}_t = \Psi_t \mu_t$,

(g) again follows due to [9, T19 Theorem, Appendix A2, p. 302],

(h) again follows since L_T is the Radon-Nikodym derivative $\frac{dP^{A,\hat{U}_0^T}}{d\tilde{P}^{A,\hat{U}_0^T}}$. □

Proof of Lemma 7 : Let

$$f(t) = \mathbb{E}[\phi_u(\mathbb{E}[X_t | U_0^t, A, B])]. \tag{2.88}$$

We will first show that $f(t)$ is right continuous. Let $\tilde{\delta}_n$ be a non-increasing positive subsequence approaching 0 as $n \to \infty$. Define the following (suppressing the time index t)

$$\mathcal{H}_n = \mathcal{F}^U_{(t+\tilde{\delta}_n)} \vee \sigma(A) \vee \sigma(B) \tag{2.89}$$

$$\mathsf{X}_n = X_{t+\tilde{\delta}_n}. \tag{2.90}$$

Since the sample paths of X_t are right-continuous

$$\lim_{n\to\infty} \mathsf{X}_n \to X_t$$

and $\mathcal{H}_1 \supset \mathcal{H}_2 \supset \dots$, we have the following equalities P-a.s.

$$\begin{aligned}
\lim_{n\to\infty} \mathbb{E}[X_{t+\tilde{\delta}_n}|U_0^{t+\tilde{\delta}_n}, A, B] &\overset{(a)}{=} \lim_{n\to\infty} \mathbb{E}[\mathsf{X}_n|\mathcal{H}_n] \\
&\overset{(b)}{=} \mathbb{E}\left[X_t \middle| \bigcap_n \mathcal{H}_n\right] \\
&= \mathbb{E}\left[X_t \middle| \bigcap_{\epsilon>0} \mathcal{F}^U_{t+\epsilon} \vee \sigma(A) \vee \sigma(B)\right] \\
&\overset{(c)}{=} \mathbb{E}[X_t|\mathcal{F}^U_t \vee \sigma(A) \vee \sigma(B)] \\
&= \mathbb{E}[X_t|U_0^t, A, B]. \qquad (2.91)
\end{aligned}$$

Here, (a) is due to the definition of X_n and $\mathcal{H}_n$,

(b) is due to the backwards analogue of the dominated convergence theorem for conditional expectation [20, Exercise 5.6.2, p. 265] (recall that X_t is bounded), and

(c) is due to the right continuity of the filtration $\mathcal{F}^U_t \vee \sigma(A) \vee \sigma(B)$ [9, Theorem T25, Appendix A2, p. 304].

Since $\phi_u(x)$ is a continuous function and X_t is a bounded random variable

$$\lim_{\tilde{\delta}_n\to 0^+} \mathbb{E}[\phi_u(\mathbb{E}[X_{t+\tilde{\delta}_n}|U_0^{t+\tilde{\delta}_n}, A, B])] = \mathbb{E}[\phi_u(\mathbb{E}[X_t|U_0^t, A, B])],$$

and hence

$$\lim_{\delta\to 0^+} \mathbb{E}[\phi_u(\mathbb{E}[X_{t+\delta}|U_0^{t+\delta}, A, B])] = \mathbb{E}[\phi_u(\mathbb{E}[X_t|U_0^t, A, B])].$$

Similarly,

$$\lim_{\delta\to 0^+} \mathbb{E}[\phi_u(\mathbb{E}[X_{t+\delta}|U_0^{t+\delta}, B])] = \mathbb{E}[\phi_u(\mathbb{E}[X_t|U_0^t, B])].$$

Since $(A,B) \rightleftarrows X_0^T \rightleftarrows (U_0^t, U_t^{t+\delta})$ and $U_0^t \rightleftarrows X_0^T \rightleftarrows U_t^{t+\delta}$ are Markov chains, [31, Proposition 6.8, p. 111] implies $(A, B, U_0^t) \rightleftarrows X_0^T \rightleftarrows U_t^{t+\delta}$ is also a Markov chain. Taking $t_1 = t$, $t_2 = t+\delta$, Theorem 1 yields

$$\begin{aligned}
\lim_{\delta\to 0^+} \frac{1}{\delta} I\left(A; U_t^{t+\delta} \middle| U_0^t, B\right) &= \lim_{\delta\to 0^+} \frac{1}{\delta} \int_t^{t+\delta} \mathbb{E}[\phi_u(\mathbb{E}[X_s|U_t^s, U_0^t, A, B])] - \mathbb{E}[\phi_u(\mathbb{E}[X_s|U_t^s, U_0^t, B])]\, ds \\
&= \mathbb{E}[\phi_u(\mathbb{E}[X_t|U_0^t, A, B])] - \mathbb{E}[\phi_u(\mathbb{E}[X_t|U_0^t, B])], \qquad (2.92)
\end{aligned}$$

where the last equality is due to the fact that if $f(x)$ is right continuous at t, then

$$\lim_{\delta\to 0^+} \frac{1}{\delta}\int_t^{t+\delta} f(s)\,ds = f(t).$$

Let $\tilde{U}_0^T$ to be the time-reversed U_0^T process. Then $\tilde{U}_0^T$ is a doubly-stochastic Poisson process with rate process $\{\tilde{X}_t = X_{(T-t)-}, t \in [0,T]\}$, and

$$\begin{aligned}\lim_{\delta\to 0^+} \frac{1}{\delta} I\left(A; U_{t-\delta}^t \middle| U_t^T, B\right) &= \lim_{\delta\to 0^+} \frac{1}{\delta} I\left(A; \tilde{U}_{T-t}^{T-t+\delta} \middle| \tilde{U}_0^{T-t}, B\right) \\ &= \mathbb{E}[\phi_u(\mathbb{E}[\tilde{X}_{(T-t)-}|\tilde{U}_0^{T-t}, A, B])] - \mathbb{E}[\phi_u(\mathbb{E}[\tilde{X}_{T-t}|\tilde{U}_0^{T-t}, B])] \\ &= \mathbb{E}[\phi_u(\mathbb{E}[X_{t-}|U_t^T, A, B])] - \mathbb{E}[\phi_u(\mathbb{E}[X_{t-}|U_t^T, B])].\end{aligned} \tag{2.93}$$

□

Proof of Lemma 8: We have

$$\begin{aligned}\frac{1}{\delta} I\left(A; U_s^{s+\delta} \middle| U_0^s, B\right) &= \frac{1}{\delta}\int_s^{s+\delta} \mathbb{E}[\phi_u(\mathbb{E}[X_r|U_0^r, A, B])] - \mathbb{E}[\phi_u(\mathbb{E}[X_r|U_0^r, B])]\,dr \\ &\le 2\phi_u^*,\end{aligned} \tag{2.94}$$

where $\phi_u^* = \max_{0\le x\le 1} |\phi_u(x)|$. The second part of the lemma follows similarly. □

Proof of Lemma 9: Consider

$$\begin{aligned}I(A; U_0^t|B) &\overset{(a)}{=} \int_0^t \mathbb{E}[\phi_u(\mathbb{E}[X_s|U_0^s, A, B])] - \mathbb{E}[\phi_u(\mathbb{E}[X_s|U_0^s, B])]\,ds \\ &\overset{(b)}{=} \int_0^t \lim_{\delta\to 0^+} \frac{1}{\delta} I\left(A; U_s^{s+\delta} \middle| U_0^s, B\right)\,ds \\ &\overset{(c)}{=} \lim_{\delta\to 0^+} \frac{1}{\delta}\int_0^t I\left(A; U_s^{s+\delta} \middle| U_0^s, B\right)\,ds.\end{aligned} \tag{2.95}$$

Here, (a) is due to Theorem 1,

(b) is due to Lemma 7, and

(c) is due to Lemma 8 and the dominated convergence theorem.

The proof of the second part of the lemma follows similarly. □

Proof of Lemma 11: The existence of $\mathfrak{i}(X_0^{T_n}; Y_0^{T_n})$ and $\mathfrak{i}(V_0^{T_n}; Z_0^{T_n})$ is due to Lemma 5. The existence of $\mathfrak{i}(X_0^{T_n}; Y_0^{T_n}|V_0^{T_n})$ is discussed in a later part of this proof. We will use the measure $\tilde{P}$ as defined in Theorem 1. Using Theorem 1 we have $P^{V_0^{T_n}, X_0^{T_n}, Z_0^{T_n}}$-a.s.

$$\frac{1}{T_n} \log\left(\frac{dP^{V_0^{T_n}, Z_0^{T_n}}}{d\tilde{P}^{V_0^{T_n}, Z_0^{T_n}}}\right) = \frac{1}{T_n}\int_0^{T_n} \log(a_z\Pi_t + \lambda_z) dZ_t + 1 - (a_z\Pi_t + \lambda_z)\, dt$$
$$= \frac{1}{n\tau}\sum_{i=1}^{n}\int_{(i-1)\tau}^{i\tau} \log(a_z\Pi_t + \lambda_z)\, dZ_t^{(i)} + 1 - (a_z\Pi_t + \lambda_z)\, dt, \tag{2.96}$$

where $\left\{Z_t^{(i)}, t \in [(i-1)\tau, i\tau]\right\}$ is the point process corresponding to $Z_{(i-1)\tau}^{i\tau}$, and for $t \in [0, T]$,

$$\Pi_t = E[X_t|Z_0^t, V_0^{T_n}], \quad P^{V_0^{T_n}, X_0^{T_n}, Z_0^{T_n}}\text{-a.s.}$$

Let

$$\Psi_i^{(1)} = \frac{1}{\tau}\int_{(i-1)\tau}^{i\tau} \log(a_z\Pi_t + \lambda_z)\, dZ_t^{(i)}, \tag{2.97}$$

then $\Psi_i^{(1)}$, for $i = 1, 2, \ldots, n$ are independent and identically distributed with

$$\begin{aligned}
\mathbb{E}[|\Psi_1^{(1)}|] &= \frac{1}{\tau}\mathbb{E}\left[\left|\int_0^{\tau} \log(a_z\Pi_t + \lambda_z)\, dZ_t\right|\right] \\
&\le \frac{1}{\tau}\mathbb{E}\left[\int_0^{\tau} |\log(a_z\Pi_t + \lambda_z)|\, dZ_t\right] \\
&= \frac{1}{\tau}\int_0^{\tau} \mathbb{E}\left[|\phi_z(\Pi_t)|\right] dt \\
&\le \phi_z^* < \infty,
\end{aligned} \tag{2.98}$$

where $\phi_z^* = \max_{0\le x\le 1} \phi_z(x)$, and we have used the fact that the $(P, \sigma(\bar{V}_1) \vee \mathcal{F}_t^Z : t \in [0, T])$-intensity of $Z_0^{T_n}$ is $a_z\Pi_t + \lambda_z$ (Lemma 4). Thus by the strong law of large numbers [31, Theorem 4.23, p.73]

$$\frac{1}{n}\sum_{i=1}^{n}\Psi_i^{(1)} \to \mathbb{E}[\Psi_1^{(1)}] = \frac{1}{\tau}\int_0^{\tau} \mathbb{E}\left[\phi_z\left(\mathbb{E}[X_t|Z_0^t, \bar{V}_1]\right)\right] dt \tag{2.99}$$

almost surely. Now let

$$\Psi_i^{(2)} = \frac{1}{\tau}\int_{(i-1)\tau}^{i\tau} a_z\Pi_t + \lambda_z \, dt,$$

for which the law of large numbers gives

$$\frac{1}{n}\sum_{i=1}^{n}\Psi_i^{(2)} \xrightarrow{\text{a.s.}} \mathbb{E}[\Psi_1^{(2)}] = \frac{1}{\tau}\int_0^{\tau} a_z\mathbb{E}[X_t] + \lambda_z \, dt. \tag{2.100}$$

Thus

$$\frac{1}{T_n}\log\left(\frac{dP^{V_0^{T_n},Z_0^{T_n}}}{d\tilde{P}^{V_0^{T_n},Z_0^{T_n}}}\right) \xrightarrow{\text{a.s.}} \frac{1}{\tau}\int_0^{\tau}\mathbb{E}\left[\phi_z\left(\mathbb{E}[X_t|Z_0^t,\bar{V}_1]\right)\right] + 1 - (a_z\mathbb{E}[X_t]+\lambda_z)\, dt. \tag{2.101}$$

Similarly $P^{V_0^{T_n},X_0^{T_n},Z_0^{T_n}}$-a.s.

$$\frac{1}{T_n}\log\left(\frac{dP^{Z_0^{T_n}}}{dP_0^{Z_0^{T_n}}}\right) \xrightarrow{\text{a.s.}} \frac{1}{\tau}\int_0^{\tau}\mathbb{E}\left[\phi_z\left(\mathbb{E}[X_t|Z_0^t]\right)\right] + 1 - (a_z\mathbb{E}[X_t]+\lambda_z)\, dt. \tag{2.102}$$

This gives $P^{V_0^{T_n},X_0^{T_n},Z_0^{T_n}}$-a.s.

$$\begin{aligned}
\frac{1}{T_n}\mathfrak{i}(V_0^{T_n};Z_0^{T_n}) &= \log\frac{dP^{V_0^{T_n},Z_0^{T_n}}}{d(P^{V_0^{T_n}}\times P^{Z_0^{T_n}})} \\
&= \frac{1}{T_n}\log\left(\frac{dP^{V_0^{T_n},Z_0^{T_n}}}{d\tilde{P}^{V_0^{T_n},Z_0^{T_n}}}\right) - \frac{1}{T_n}\log\left(\frac{dP^{Z_0^{T_n}}}{dP_0{}^{Z_0^{T_n}}}\right) \\
&\xrightarrow{\text{a.s.}} \frac{1}{\tau}\int_0^{\tau}\mathbb{E}\left[\phi_z\left(\mathbb{E}[X_t|Z_0^t,\bar{V}_1]\right)\right] - \mathbb{E}\left[\phi_z\left(\mathbb{E}[X_t|Z_0^t]\right)\right]\, dt \\
&= \frac{1}{\tau}I(\bar{V}_1;Z_0^{\tau})
\end{aligned} \tag{2.103}$$

as $n\to\infty$, and we have used Theorem 1. From Lemma 7

$$\lim_{\tau\to 0^+}\frac{1}{\tau}I(\bar{V}_1;Z_0^{\tau}) = \mathbb{E}\left[\phi_z\left(\mathbb{E}[X_0|\bar{V}_1]\right)\right] - \phi_z\left(\mathbb{E}[X_0]\right). \tag{2.104}$$

Thus given any $\epsilon > 0$, we can choose $\bar{\tau}$ such that

$$\left|\frac{1}{\tau^*}I(\bar{V}_1;Z_0^{\tau^*}) - \left(\mathbb{E}\left[\phi_z\left(\mathbb{E}[X_0|\bar{V}_1]\right)\right] - \phi_z\left(\mathbb{E}[X_0]\right)\right)\right| \le \frac{\epsilon}{2}, \tag{2.105}$$

and then choosing N large enough we can ensure that

$$P\left(\left|\frac{1}{T_N}\mathfrak{i}(V_0^{T_N};Z_0^{T_N}) - \left(\mathbb{E}\left[\phi_z\left(\mathbb{E}[X_0|\bar{V}_1]\right)\right] - \phi_z\left(\mathbb{E}[X_0]\right)\right)\right| > \epsilon\right) \leq \epsilon. \tag{2.106}$$

Note that $\mathcal{V}_0^{T_n}$ and $\mathcal{X}_0^{T_n}$ here are effectively finite alphabets. For the space $(\mathcal{N}_0^{T_n}, \mathfrak{F}^Y)$, the σ-field $\mathfrak{F}^Y$ is the restriction of the σ-field generated by the Skorohod topology on $D[0,1]$ to $\mathcal{N}_0^{T_n}$. This makes $(\mathcal{N}_0^{T_n}, \mathfrak{F}^Y)$ a standard space [7, Theorem 12.2, p. 128] and [28, Section 1.5, p. 12]. Consider

$$\begin{aligned} I(V_0^{T_N}, X_0^{T_N}; Y_0^{T_n}) &= I(X_0^{T_n}; Y_0^{T_N}) + I(V_0^{T_n}; Y_0^{T_N}|X_0^{T_N}) \\ &= I(X_0^{T_n}; Y_0^{T_N}) < \infty. \end{aligned} \tag{2.107}$$

This gives $P^{V_0^{T_n},X_0^{T_n},Y_0^{T_n}} \ll P^{V_0^{T_n},X_0^{T_n}} \times P^{Y_0^{T_n}}$. Thus from [28, Corollary 5.5.3, p. 125], $\mathfrak{i}(X_0^{T_n};Y_0^{T_n}|V_0^{T_n})$ exists and $P^{V_0^{T_n},X_0^{T_n},Y_0^{T_n}}$-a.s. satisfies

$$\frac{1}{T_n}\mathfrak{i}(X_0^{T_n};Y_0^{T_n}|V_0^{T_n}) = \frac{1}{T_n}\mathfrak{i}(V_0^{T_n},X_0^{T_n};Y_0^{T_n}) - \frac{1}{T_n}\mathfrak{i}(V_0^{T_n};Y_0^{T_n}). \tag{2.108}$$

Here, we have used the fact that since $\frac{1}{T_n}\mathbb{E}[|\mathfrak{i}(V_0^{T_n};Y_0^{T_n})|] < \infty$, $\frac{1}{T_n}\mathfrak{i}(V_0^{T_n};Y_0^{T_n})$ is $P^{V_0^{T_n},X_0^{T_n},Y_0^{T_n}}$-a.s. finite. Also $P^{X_0^{T_n},Y_0^{T_n}} \ll P^{X_0^{T_n}} \times P^{Y_0^{T_n}}$ (since $I(X_0^{T_n};Y_0^{T_n}) < \infty$), and $V_0^{T_n} \rightleftarrows X_0^{T_n} \rightleftarrows Y_0^{T_n}$ being a Markov chain, [28, Corollary 5.5.4, p.126] yields

$$\mathfrak{i}(V_0^{T_n},X_0^{T_n};Y_0^{T_n}) = \mathfrak{i}(X_0^{T_n};Y_0^{T_n}), \quad P^{V_0^{T_n},X_0^{T_n},Y_0^{T_n}}\text{-a.s.}$$

Since $P^{X_0^{T_n},Y_0^{T_n}}$-a.s.

$$\begin{aligned} \mathfrak{i}(X_0^{T_n};Y_0^{T_n}) &= \log\left(\frac{dP^{X_0^{T_n},Y_0^{T_n}}}{d\left(P^{X_0^{T_n}} \times P^{Y_0^{T_n}}\right)}\right) \\ &= \log\left(\frac{dP^{X_0^{T_n},Y_0^{T_n}}}{d\tilde{P}^{X_0^{T_n},Y_0^{T_n}}}\right) - \log\left(\frac{dP^{Y_0^{T_n}}}{dP_0^{Y_0^{T_n}}}\right), \end{aligned}$$

we have from Theorem 1, $P^{V_0^{T_n},X_0^{T_n},Y_0^{T_n}}$-a.s.

$$\begin{aligned} \frac{1}{T_n}\log\left(\frac{dP^{X_0^{T_n},Y_0^{T_n}}}{d\tilde{P}^{X_0^{T_n},Y_0^{T_n}}}\right) &= \frac{1}{T_n}\int_0^{T_n} \log(a_yX_t+\lambda_y)dY_t + 1 - (a_yX_t+\lambda_y)\,dt \\ &\xrightarrow{a.s.} \frac{1}{\tau}\int_0^{\tau} \mathbb{E}\left[\phi_y\left(X_t\right)\right] + 1 - \left(a_y\mathbb{E}[X_t]+\lambda_y\right)dt, \end{aligned}$$

where the a.s. convergence can be shown by using an argument similar to that used for the second user. Similarly for the second term, $P^{V_0^{T_n},X_0^{T_n},Y_0^{T_n}}$-a.s.,

$$\frac{1}{T_n}\log\left(\frac{dP^{Y_0^{T_n}}}{dP_0^{Y_0^{T_n}}}\right) = \frac{1}{T_n}\int_0^{T_n}\log(a_y\Pi_t' + \lambda_y)dY_t + 1 - (a_y\Pi_t' + \lambda_y)\,dt$$
$$\xrightarrow{a.s.} \frac{1}{\tau}\int_0^{\tau}\mathbb{E}\left[\phi_y\left(\mathbb{E}[X_t|Y_0^t]\right)\right] + 1 - (a_y\mathbb{E}[X_t] + \lambda_y)\,dt,$$

where $\Pi_t' = \mathbb{E}[X_t|Y_0^t]$ $P^{V_0^{T_n},X_0^{T_n},Y_0^{T_n}}$-a.s. Hence we have

$$\frac{1}{T_n}\mathfrak{i}(X_0^{T_n};Y_0^{T_n}) \xrightarrow{\text{a.s.}} \frac{1}{\tau}\int_0^{\tau}\mathbb{E}\left[\phi_y(X_t)\right] - \mathbb{E}\left[\phi_y\left(\mathbb{E}[X_t|Y_0^t]\right)\right]\,dt$$
$$= \frac{1}{\tau}I(X_0^{\tau};Y_0^{\tau}) = \frac{1}{\tau}I(X_0;Y_0^{\tau}),$$

where we have used the fact that X_0^{τ} is constant over the interval $[0,\tau)$ and Theorem 1. From Lemma 7

$$\lim_{\tau\to 0^+}\frac{1}{\tau}I(X_0;Y_0^{\tau}) = \mathbb{E}\left[\phi_y(X_0)\right] - \phi_y\left(\mathbb{E}[X_0]\right).$$

Also, similar to the second receiver, we can show that for a given $\epsilon > 0$ there exists N and $\bar{\tau}$ such that $n \geq N$ and $\tau \leq \bar{\tau}$ implies that

$$P\left(\left|\frac{1}{T_n}\mathfrak{i}(V_0^{T_n};Y_0^{T_n}) - \left(\mathbb{E}\left[\phi_y\left(\mathbb{E}[X_0|\bar{V}_1]\right)\right] - \phi_y(\mathbb{E}[X_0])\right)\right| > \epsilon\right) \leq \epsilon \tag{2.109}$$

Since $P^{V_0^{T_n},X_0^{T_n},Y_0^{T_n}}$-a.s.

$$\frac{1}{T_n}\mathfrak{i}(X_0^{T_n};Y_0^{T_n}|V_0^{T_n}) = \frac{1}{T_n}\mathfrak{i}(X_0^{T_n};Y_0^{T_n}) - \frac{1}{T_n}\mathfrak{i}(V_0^{T_n};Y_0^{T_n}).$$

Thus for given $\epsilon > 0$ there exists N and $\bar{\tau}$ such that $n \geq N$ and $\tau \leq \bar{\tau}$ implies that

$$P\left(\left|\frac{1}{T_n}\mathfrak{i}(X_0^{T_n};Y_0^{T_n}|V_0^{T_n}) - \left(\mathbb{E}\left[\phi_y(X_0)\right] - \mathbb{E}\left[\phi_y\left(\mathbb{E}[X_0|\bar{V}_1]\right)\right]\right)\right| > \epsilon\right) \leq \epsilon. \tag{2.110}$$

□

Proof of Lemma 12: Note that

$$\begin{aligned}
&\int_0^T \mathbb{E}[\phi_z(\mathbb{E}[X_t|M, Y_0^t])] - \mathbb{E}[\phi_z(\mathbb{E}[X_t|M, Z_t^T])]\, dt \\
\stackrel{(a)}{=} &\int_0^T \mathbb{E}[\phi_y(\mathbb{E}[X_t|M, Y_0^t])] - \mathbb{E}[\phi_z(\mathbb{E}[X_t|M, Z_t^T])]\, dt \\
&- \int_0^T \mathbb{E}[\phi_y(\mathbb{E}[X_t|M, Y_0^t])] - \mathbb{E}[\phi_z(\mathbb{E}[X_t|M, Y_0^t])]\, dt \\
\stackrel{(b)}{=} &\int_0^T \mathbb{E}[\phi_y(\mathbb{E}[X_t|M, Y_0^t, Z_t^T])] - \mathbb{E}[\phi_z(\mathbb{E}[X_t|M, Y_0^t, Z_t^T])]\, dt \\
&- \int_0^T \mathbb{E}[\phi_y(\mathbb{E}[X_t|M, Y_0^t])] - \mathbb{E}[\phi_z(\mathbb{E}[X_t|M, Y_0^t])]\, dt \\
\stackrel{(c)}{=} &\int_0^T \mathbb{E}[\Phi(\mathbb{E}[X_t|M, Y_0^t, Z_t^T])] - \mathbb{E}[\Phi(\mathbb{E}[X_t|M, Y_0^t])]\, dt \\
\stackrel{(d)}{\geq} &\ 0.
\end{aligned}$$

In (a) we have added and subtracted a term,

(b) is due to Theorem 2,

(c) is due to the definition of $\Phi(x)$, and

(d) is due to convexity of $\Phi(x)$ and Jensen's inequality. □

Proof of Lemma 13: In this case we have

$$\begin{aligned}
\mathbb{E}\left[\phi_y(X_0)\right] - \phi_y\left(\mathbb{E}[X_0]\right) &= (\alpha p + (1-\alpha)q)\phi_y(1) + (\alpha(1-p) + (1-\alpha)(1-q))\phi_y(0) - \phi_y(\alpha p + (1 - \\
&= C_y + \tilde{C}_y.
\end{aligned}$$

And

$$\begin{aligned}
\mathbb{E}\left[\phi_z\left(\mathbb{E}[X_0|\bar{V}_1]\right)\right] - \phi_z\left(\mathbb{E}[X_0]\right) &= \alpha\phi_z(p) + (1-\alpha)\phi_z(q) - \phi_z(\alpha p + (1-\alpha)q) \\
&= C_z.
\end{aligned}$$

Similarly,

$$\begin{aligned}
\mathbb{E}\left[\phi_y(X_0)\right] - \mathbb{E}\left[\phi_y\left(\mathbb{E}[X_0|\bar{V}_1]\right)\right] &= (\alpha p + (1-\alpha)q)\phi_y(1) + (\alpha(1-p) + (1-\alpha)(1-q))\phi_y(0) \\
&\quad - \alpha\phi_y(p) - (1-\alpha)\phi_y(q) \\
&= C_y.
\end{aligned}$$

Now applying Lemma 11 proves the statement of the lemma. □

CHAPTER 3

FUNCTIONAL COVERING OF POINT PROCESSES

3.1 Preliminaries

We will consider a probability space $(\Omega, \mathcal{F}, P)$ on which all stochastic processes considered here are defined. For a finite $T > 0$, let $(\mathcal{F}_t : t \in [0,T])$ be an increasing family of σ-fields with $\mathcal{F}_T \in \mathcal{F}$. We will assume that the given filtration $(\mathcal{F}_t : t \in [0,T])$, P, and $\mathcal{F}$ satisfy the "usual conditions" [9, Chapter III, p. 75]: $\mathcal{F}$ is complete with respect to P, $\mathcal{F}_t$ is right continuous, and $\mathcal{F}_0$ contains all the P-null sets of $\mathcal{F}_t$. Stochastic processes are denoted as $\hat{Y}_0^T = \{\hat{Y}_t : 0 \leq t \leq T\}$. The process X_0^T is said to be *adapted* to the history $(\mathcal{F}_t : t \in [0,T])$ if X_t is $\mathcal{F}_t$ measurable for all $t \in [0,T]$. The internal history recorded by the process X_0^T is denoted by $\mathcal{F}_t^X = (\sigma(X_s) : s \in [0,t])$, where $\sigma(A)$ denotes the σ-field generated by A.

A process X_0^T is called $(\mathcal{F}_t : t \in [0,T])$-*predictable* if X_0 is $\mathcal{F}_0$ measurable and the mapping $(t,\omega) \to X_t(\omega)$ defined from $(0,T) \times \Omega$ into $\mathbb{R}$ (the set of real numbers) is measurable with respect to the σ-field over $(0,T) \times \Omega$ generated by rectangles of the form

$$(s,t] \times A; \quad 0 < s \leq t \leq T, \quad A \in \mathcal{F}_s. \tag{3.1}$$

For two measurable spaces $(\Omega_1, \mathcal{F}_1)$ and $(\Omega_2, \mathcal{F}_2)$, the product space is denoted by $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2)$. We say that $A \rightleftarrows B \rightleftarrows C$ forms a Markov chain under measure P, if A and C are conditionally independent given B under P. $P \ll Q$ denotes that the probability measure P is absolutely continuous with respect to the measure Q. $\mathbf{1}\{\mathsf{E}\}$ denotes the indicator function for an event E. $\log(x)$ is the natural

This work was presented at the IEEE Int. Symposium on Information Theory (ISIT), Paris, July 2019 [58]

logarithm of x. $(x)^+$ and $(x)^-$ denote the positive ($\max(x,0)$) and the negative part ($-\min(x,0)$) of x respectively. $\lceil x \rceil$ denotes the ceiling of x. Throughout this paper we will adopt the convention that $0\log(0)=0$, $\exp(\log(0))=0$, and $0^0=1$.

Definition 9 $\phi(x)=x\log(x)$ *with convention that* $0\log(0)=0$.

We note that $\phi(x)$ is convex.

We will use the following form of Jensen's inequality [34, Theorem 7.9, p. 149] and [34, Theorem 8.20, p. 177].

Lemma 14 *If* $f(x)$ *is a convex function and* $\mathbb{E}|X|<\infty$ *then* $\mathbb{E}[f(X)]$ *exists and for any two* σ*-fields* A *and* B

$$\mathbb{E}[f(X)] \geq \mathbb{E}[f(\mathbb{E}[X|A,B])] \geq \mathbb{E}[f(\mathbb{E}[X|A])] \geq f(\mathbb{E}[X]).$$

We now recall the definition of mutual information for general ensembles and its properties. Let A, B, and C be measurable mappings defined on a given probability space $(\Omega,\mathcal{F},P)$, taking values in $(\mathcal{A},\mathfrak{F}^A)$, $(\mathcal{B},\mathfrak{F}^B)$, and $(\mathcal{C},\mathfrak{F}^C)$ respectively. Consider partitions of Ω, $\mathfrak{Q}_A=\{\mathtt{A}_i, 1\leq i\leq N_A\}\subseteq\sigma(A)$ and $\mathfrak{Q}_B=\{\mathtt{B}_j, 1\leq j\leq N_B\}\subseteq\sigma(B)$. Wyner defined the conditional mutual information $I(A;B|C)$ as [67]

$$I(A;B|C)=\sup_{\mathfrak{Q}_A,\mathfrak{Q}_B}\mathbb{E}\left[\sum_{i,j=1,1}^{N_A,N_B}P(\mathtt{A}_i,\mathtt{B}_j|C)\log\left(\frac{P(\mathtt{A}_i,\mathtt{B}_j|C)}{P(\mathtt{A}_i|C)P(\mathtt{B}_j|C)}\right)\right], \tag{3.2}$$

where the supremum is over all such partitions of Ω. Wyner showed that $I(A;B|C)\geq 0$ with equality if and only if $A \rightleftarrows C \rightleftarrows B$ forms a Markov chain [67, Lemma 3.1], and that (generally referred to as) Kolmogrov's formula holds [67, Lemma 3.2]

$$I(A,C;B)=I(A;B)+I(C;B|A). \tag{3.3}$$

Hence if $I(A;B) < \infty$, then $I(C;B|A) = I(A,C;B) - I(A;B)$. The data processing inequality can be obtained from (3.3) as well: if $A \rightleftarrows C \rightleftarrows B$ forms a Markov chain, then $I(A;B) \leq I(C;B)$.

Denote by $P^{A,B}$, the joint distribution of A and B on the space $(\mathcal{A}\times\mathcal{B}, \mathfrak{F}^A \otimes \mathfrak{F}^B)$, i.e.,

$$P^{A,B}(dA \times dB) = P((A^{-1}(dA), B^{-1}(dB)), \quad dA \in \mathfrak{F}^A, dB \in \mathfrak{F}^B.$$

Similarly, P^A and P^B denote the marginal distributions. Gelfand and Yaglom [27] proved that if $P^{A,B} \ll P^A \times P^B$, then the mutual information $I(A;B)$ (defined via (3.2) by taking $\sigma(C)$ to be the trivial σ-field) can be computed as

$$I(A;B) = \mathbb{E}\left[\log\left(\frac{dP^{A,B}}{d(P^A \times P^B)}\right)\right]. \tag{3.4}$$

A sufficient condition for $P^{A,B} \ll P^A \times P^B$ is that $I(A;B) < \infty$ [28, Lemma 5.2.3, p. 92]. We will also require the following result [67, Lemma 2.1]:

Lemma 15 (Wyner's Lemma) *If M is a finite alphabet random variable, then*

$$I(M;U_0^T) = H(M) - \mathbb{E}\left[H(M|U_0^T)\right],$$

where

$$H(M|U_0^T) = -\sum_m P(M=m|U_0^T)\log\left(P(M=m|U_0^T)\right),$$

and $H(M)$ is the entropy of M.

3.2 Point Processes, Intensities, and Mutual Information

Let $\mathcal{N}_0^T$ denote the set of counting realizations (or point-process realizations) on $[0,T]$, i.e., if $N_0^T \in \mathcal{N}_0^T$, then for $t \in [0,T]$, $N_t \in \mathbf{N}$ (the set of non-negative integers), is right continuous, and has unit increasing jumps with $N_0 = 0$. Let $\mathfrak{F}^N$ be the restriction of the σ-field generated by the Skorohod topology on $D[0,1]$ to $\mathcal{N}_0^T$.

Definition 10 *If N_0^T is a counting process adapted to the history $(\mathcal{F}_t : t \in [0,T])$, then N_0^T is said to have $(P, \mathcal{F}_t : t \in [0,T])$-*intensity *$\Gamma_0^T = (\Gamma_t : t \in [0,T])$, where Γ_0^T is a non-negative measurable process if*

- *Γ_0^T is $(\mathcal{F}_t : t \in [0,T])$-predictable,*
- *$\int_0^T \Gamma_t \, dt < \infty$, P-a.s.,*
- *and for all non-negative $(\mathcal{F}_t : t \in [0,T])$-predictable processes C_0^T:*[1]

$$\mathbb{E}\left[\int_0^T C_s \, dN_s\right] = \mathbb{E}\left[\int_0^T C_s \Gamma_s \, ds\right].$$

When it is clear from the context, we will drop the probability measure P from the notation and say N_0^T has $(\mathcal{F}_t : t \in [0,T])$-intensity Γ_0^T.

Definition 11 *A point process N_0^T is said to be* Poisson process with rate λ *if its $(\mathcal{F}_t^Y : t \in [0,T])$-intensity is $(\lambda : t \in [0,T])$.*

The above definition can be shown to imply the usual definition of Poisson process [9, Theorem T4, Chapter II, p. 25] and vice versa [9, Section 2, Chapter II, p. 23].

Definition 12 *$P_0^{Y_0^T}$ denotes the distribution of a point process Y_0^T (on the space $(\mathcal{N}_0^T, \mathfrak{F}^N)$) under which Y_0^T is a Poisson process with unit rate.*

[1]The limits of the Lebesgue-Stieltjes integral $\int_a^b$ are to be interpreted as $\int_{(a,b]}$.

A point processes with stochastic intensity and a Poisson process with unit rate are linked via the following result.

Lemma 16 *Let $P^{Y_0^T}$ be the distribution of a point process Y_0^T such that $P^{Y_0^T} \ll P_0^{Y_0^T}$, then there exists a non-negative predictable process Λ_0^T such that*

$$\frac{dP^{Y_0^T}}{dP_0^{Y_0^T}} = \exp\left(\int_0^T \log(\Lambda_t)\, dY_t - \Lambda_t + 1\, dt\right).$$

Moreover, the $(P^{Y_0^T}, (\mathcal{F}_t^Y : t \in [0,T]))$-intensity of Y_0^T is Λ_0^T. Conversely, if $(P^{Y_0^T}, \mathcal{F}_t^Y : t \in [0,T])$-intensity of Y_0^T is Γ_0^T and $\mathbb{E}_{P^{Y_0^T}}[\int_0^T |\phi(\Gamma_t)|\, dt] < \infty$, then $P^{Y_0^T} \ll P_0^{Y_0^T}$, and the corresponding Radon-Nikodym derivative is given by the above expression, where

$$\mathbb{E}_{P^{Y_0^T}}\left[\int_0^T |\Gamma_t - \Lambda_t|\, dt\right] = 0, \quad \mathbb{E}_{P^{Y_0^T}}\left[\int_0^T \mathbf{1}\{\Gamma_t \neq \Lambda_t\}\, dY_t\right] = 0.$$

In the latter case,

$$\mathbb{E}_{P^{Y_0^T}}\left[\log\left(\frac{dP^{Y_0^T}}{dP_0^{Y_0^T}}\right)\right] = \mathbb{E}_{P^{Y_0^T}}\left[\int_0^T \phi(\Gamma_t) - \Gamma_t + 1\, dt\right].$$

Proof: Please see Section 3.6. □

The following theorem allows us to express the mutual information involving a point processes with intensity and other random variables in terms of the intensity functions. The proof of the theorem is similar to the proof of Theorem 1 in [59].

Theorem 7 (Mutual information for point processes) *Let Y_0^T be a point process with $(\mathcal{F}_t^Y : t \in [0,T])$-intensity Λ_0^T such that $\mathbb{E}[\int_0^T |\phi(\Lambda_t)|\, dt] < \infty$, and let M be a measurable mapping on the given probability space satisfying $I(M; Y_0^T) < \infty$. Then there exists a process Γ_0^T such that Γ_0^T is the $(\mathcal{G}_t = \sigma(M, Y_0^t) : t \in [0,T])$ intensity of Y_0^T and*

$$I(M; Y_0^T) = \mathbb{E}\left[\int_0^T \phi(\Gamma_t) - \phi(\Lambda_t)\, dt\right].$$

Proof: Let P^{M,Y_0^T} denote the joint distribution of M and Y_0^T, and P^M and $P^{Y_0^T}$ denote their marginals, respectively. Since $I(M;Y_0^T) < \infty$, we get that $P^{M,Y_0^T} \ll P^M \times P^{Y_0^T}$ [28, Lemma 5.2.3, p. 92]. Lemma 16 says that $P^{Y_0^T} \ll P_0^{Y_0^T}$, which together with [31, Chapter 1, Exercise 19, p. 22] gives $P^{M,Y_0^T} \ll P^M \times P^{Y_0^T} \ll P^M \times P_0^{Y_0^T}$.

Let $\tilde{P}^{M,Y_0^T} \triangleq P^M \times P_0^{Y_0^T}$ and

$$\mathcal{L} = \frac{dP^{M,Y_0^T}}{d\tilde{P}^{M,Y_0^T}} \tag{3.5}$$

denote the Radon-Nikodym derivative. Since under $\tilde{P}^{M,Y_0^T}$, M and Y_0^T are independent, we note that the $(\tilde{P}^{M,Y_0^T}, (\mathcal{G}_t : t \in [0,T]))$-intensity of Y_0^T is 1 [9, E5 Exercise, Chapter II, p. 28]. Define process L_0^T as

$$L_t = \mathbb{E}_{\tilde{P}}[\mathcal{L}|\mathcal{G}_t], \quad t \in [0,T], \tag{3.6}$$

where $\mathbb{E}_{\tilde{P}}$ denotes that the conditional expectation is taken with respect to the measure $\tilde{P}^{M,Y_0^T}$. Then L_0^T is a $(\tilde{P}^{M,Y_0^T}, \mathcal{G}_t)$ non-negative absolutely-integrable martingale.

By the martingale representation theorem, the process L_0^T can be written as [9, Chapter III, Theorem T17, p. 76] (where we have taken $\sigma(M)$ to be the "germ σ-field"):

$$L_t = 1 + \int_0^t K_s(d\mathcal{Y}_s - ds),$$

where K_0^T is a $(\mathcal{G}_t : t \in [0,T])$-predictable process which satisfies $\int_0^T |K_t|\, dt < \infty$ $\tilde{P}^{M,Y_0^T}$-a.s. Applying [43, Lemma 19.5, p. 315], we can write L_0^T as

$$L_t = \exp\left(\int_0^t \log(\Gamma_s)\, dY_s + (1-\Gamma_s)\, ds\right), \quad t \in [0,T], \tag{3.7}$$

where Γ_0^T is a non-negative $(\mathcal{G}_t : t \in [0,T])$-predictable process, and $\Gamma_t < \infty$ $\tilde{P}^{M,Y_0^T}$-a.s. for $t \in [0,T]$.

Now we can mimic the proof of [9, Chapter VI, Theorems T3, p. 166] to deduce:

Lemma 17 *For all non-negative* $(\mathcal{G}_t : t \in [0,T])$*-predictable processes* C_0^T

$$\mathbb{E}\left[\int_0^T C_t \Gamma_t \, dt\right] = \mathbb{E}\left[\int_0^T C_t \, dY_t\right],$$

where the expectation is taken with respect to measure P.

Proof: Please see the Appendix. □

Taking $C_t = 1$ in the above equality yields

$$\left[\int_0^T \Gamma_t \, dt\right] = \mathbb{E}\left[\int_0^T dY_t\right] = \mathbb{E}\left[\int_0^T \Lambda_t \, dt\right] < \infty. \tag{3.8}$$

Hence $\int_0^T \Gamma_t \, dt < \infty$ P-a.s. and we conclude that the $(P^{M,Y_0^T}, \mathcal{G}_t : t \in [0,T])$-intensity of Y_0^T is Γ_0^T.

Now we will prove

Lemma 18

$$\mathbb{E}\left[\int_0^T \log(\Gamma_t) \, dY_t\right] = \mathbb{E}\left[\int_0^T \phi(\Gamma_t) \, dt\right]. \tag{3.9}$$

Proof: Please see the Appendix. □

Since $\mathbb{E}\left[\log\left(\frac{dP^{M,N_0^T}}{d\tilde{P}^{M,N_0^T}}\right)\right]$ is well defined, (3.5), (3.6), and (3.7) yields

$$\begin{aligned}
\mathbb{E}\left[\log\left(\frac{dP^{M,N_0^T}}{d\tilde{P}^{M,N_0^T}}\right)\right] &= \mathbb{E}\left[\log(L_T)\right] \\
&= \mathbb{E}\left[\int_0^T \log(\Gamma_t) \, dY_t + (1 - \Gamma_t) \, dt\right] \\
&= \mathbb{E}\left[\int_0^T \phi(\Gamma_t) \, dt\right] + \mathbb{E}\left[\int_0^T (1 - \Lambda_t) \, dt\right],
\end{aligned} \tag{3.10}$$

where in the last line we have used Lemma 18 and $\mathbb{E}\left[\int_0^T \Gamma_t\,dt\right] = \mathbb{E}\left[\int_0^T \Lambda_t\,dt\right] < \infty$ from (3.8). Also,

$$\begin{aligned}\mathbb{E}\left[\log\left(\frac{d(P^M \times P^{Y_0^T})}{d\tilde{P}^{M,Y_0^T}}\right)\right] &= \mathbb{E}\left[\log\left(\frac{dP^{Y_0^T}}{dP_0^{Y_0^T}}\right)\right] \\ &\stackrel{(a)}{=} \mathbb{E}\left[\int_0^T \phi(\Lambda_t) + 1 - \Lambda_t\,dt\right] \qquad (3.11) \\ &< \infty,\end{aligned}$$

where we have used Lemma 16 for (a). Using the above inequality and the fact that $\mathbb{E}\left[\log\left(\frac{dP^{M,N_0^T}}{d\tilde{P}^{M,N_0^T}}\right)\right]$ is well defined, we can express the mutual information as

$$\begin{aligned}I(M;Y_0^T) &= \mathbb{E}\left[\log\left(\frac{dP^{M,Y_0^T}}{d(P^M \times P^{Y_0^T})}\right)\right] \\ &= \mathbb{E}\left[\log\left(\frac{dP^{M,N_0^T}}{d\tilde{P}^{M,N_0^T}}\right)\right] - \mathbb{E}\left[\log\left(\frac{d(P^M \times P^{N_0^T})}{d\tilde{P}^{M,N_0^T}}\right)\right], \qquad (3.12)\end{aligned}$$

Now we can compute the mutual information from (3.10), (3.11), and (3.12),

$$\begin{aligned}I(M;Y_0^T) &= \mathbb{E}\left[\int_0^T \phi(\Gamma_t)\,dt\right] + \mathbb{E}\left[\int_0^T (1-\Lambda_t)\,dt\right] - \mathbb{E}\left[\int_0^T \phi(\Lambda_t)\,dt\right] - \mathbb{E}\left[\int_0^T 1 - \Lambda_t\,dt\right] \\ &= \mathbb{E}\left[\int_0^T \phi(\Gamma_t)\,dt\right] - \mathbb{E}\left[\int_0^T \phi(\Lambda_t)\,dt\right] \\ &= \mathbb{E}\left[\int_0^T \phi(\Gamma_t) - \phi(\Lambda_t)\,dt\right].\end{aligned}$$

□

In order to prove the strong data processing inequalities, we now derive some ancillary results regarding the intensity of a point process. Combining [9, T8 Theorem, Chapter II, p. 27] and [9, T9 Theorem, Chapter II, p. 28], we can conclude the following result.

Lemma 19 *Let Γ_0^T be a $(\mathcal{F}_t : t \in [0,T])$-predictable non-negative process satisfying $\int_0^T \Gamma_t\,dt < \infty$ a.s. Let Y_0^T be a point process adapted to $(\mathcal{F}_t : t \in [0,T])$. Then Γ_0^T*

is the $(\mathcal{F}_t : t \in [0,T])$*-intensity of* Y_0^T *if and only if*

$$M_t \triangleq Y_t - \int_0^t \Gamma_s \, ds \quad t \in [0,T]$$

is a $(\mathcal{F}_t : t \in [0,T])$*-local martingale* [2].

Since we impose a stricter condition of finite expectation $\mathbb{E}[\int_0^T \Gamma_t \, dt] < \infty$, the local martingale condition in the above statement can be replaced by the martingale condition.

Lemma 20 *Let* Γ_0^T *be a* $(\mathcal{F}_t : t \in [0,T])$*-predictable non-negative process satisfying* $\mathbb{E}[\int_0^T \Gamma_t \, dt] < \infty$*. Let* Y_0^T *be a point process adapted to* $(\mathcal{F}_t : t \in [0,T])$*. Then* Γ_0^T *is the* $(\mathcal{F}_t : t \in [0,T])$*-intensity of* Y_0^T *if and only if*

$$M_t \triangleq Y_t - \int_0^t \Gamma_s \, ds \quad t \in [0,T]$$

is a $(\mathcal{F}_t : t \in [0,T])$*-martingale.*

Proof: Please see the Appendix. □

Lemma 21 *If a point process* N_0^T *has* $(\mathcal{F}_t : t \in [0,T])$*-intensity* Λ_0^T*, and* $(\mathcal{G}_t : t \in [0,T])$ *is another history for* N_0^T *such that* $\mathcal{G}_t \subseteq \mathcal{F}_t$ *for each* $t \in [0,T]$*, then there exists a process* Π_0^T *such that* Π_0^T *is the* $(\mathcal{G}_t : t \in [0,T])$*-intensity of* N_0^T*, and for each* $t \in [0,T]$*,* $\Pi_t = \mathbb{E}[\Lambda_t | \mathcal{G}_{t-}]$ P*-a.s.*

Proof: Please see the Appendix. □

[2] A process Y_0^T is called a *local martingale* with respect to a filtration $(\mathcal{F}_t : t \geq 0)$ if Y_t is $\mathcal{F}_t$-measurable for each $t \in [0,T]$ and there exists an increasing sequence of stopping times T_n, such that $T_n \to \infty$ and the stopped and shifted processes $(Y_{\min\{t,T_n\}} - Y_0 : t \in [0,T])$ are $(\mathcal{F}_t : t \in [0,T])$-martingales for each n.

Lemma 22 *Let* Y_0^T *be a point process with* $(\mathcal{G}_t \triangleq \sigma(M, Y_0^t) : t \in [0,T])$*-intensity* Γ_0^T *for some* M*. Let* Z_0^T *be obtained adding an independent (of both* M *and* Y_0^T*) point process* N_0^T *with* $(\mathcal{F}_t^N : t \in [0,T])$*-intensity* Π_0^T *to* Y_0^T*. Then* Z_0^T *has a* $(\mathcal{F}_t \triangleq \sigma(M, Z_0^t) : t \in [0,T])$*-intensity* Θ_0^T *which satisfies* $\Theta_t = \mathbb{E}[(\Gamma_t + \Pi_t)|\mathcal{F}_{t-}]$ P*-a.s. for each* $t \in [0,T]$*.*

Proof: Please see the Appendix. □

Theorem 8 (Strong data processing inequality for Poisson processes under Superposi
Suppose Y_0^T *be a Poisson process with rate* λ*,* M *be such that* $I(M; Y_0^T) < \infty$*, and* Γ_0^T *be the* $(\sigma(M; Y_0^t) : t \in [0,T])$*-intensity of* Y_0^T*. Suppose* Z_0^T *is obtained by adding an independent (of* Y_0^T *and* M*) Poisson process with rate* μ *to* Y_0^T*. Then,*

$$I(M; Y_0^T) = \mathbb{E}\left[\int_0^T \phi(\Gamma_t) - \phi(\lambda)\, dt\right],$$
$$I(M; Z_0^T) \leq \mathbb{E}\left[\int_0^T \phi(\Gamma_t + \mu) - \phi(\lambda + \mu)\, dt\right].$$

Proof: Since $M \leftrightarrows Y_0^T \leftrightarrows Z_0^T$ forms a Markov chain, the data processing inequality gives $I(M; Z_0^T) \leq I(M; Y_0^T) < \infty$. Applying Theorem 7 and using the uniqueness of intensities,

$$I(M; Y_0^T) = \mathbb{E}\left[\int_0^T \phi(\Gamma_t) - \phi(\lambda)\, dt\right], \quad \text{and}$$
$$I(M; Z_0^T) = \mathbb{E}\left[\int_0^T \phi(\hat{\Gamma}_t) - \phi(\hat{\lambda}_t)\, dt\right], \tag{3.13}$$

where $\hat{\Gamma}_0^T$ and $\hat{\lambda}_0^T$ are the $(\sigma(M; Z_0^t) : t \in [0,T])$ and $(\mathcal{F}_t^Z : t \in [0,T])$-intensities of Z_0^T. Due to the uniqueness of the intensities and Lemma 22, we get for each $t \in [0,T]$, $\hat{\Gamma}_t = \mathbb{E}[\Gamma_t | M, Z_0^{t-}] + \mu$, and $\hat{\lambda}_t = \lambda + \mu$. Substituting this in (3.13) and applying Jensen's inequality yields

$$\begin{aligned} I(M; Z_0^T) &= \mathbb{E}\left[\int_0^T \phi(\mathbb{E}[\Gamma_t | M, Z_0^{t-}] + \mu) - \phi(\lambda + \mu)\, dt\right], \\ &\leq \mathbb{E}\left[\int_0^T \phi(\Gamma_t + \mu) - \phi(\lambda + \mu)\, dt\right]. \end{aligned}$$

□

Definition 13 *A point process Z_0^T is said to be obtained from p-thinning of a point process Y_0^T, if each point in Y_0^T is deleted with probability p, independent of all other points and deletions.*

Lemma 23 *Suppose that Y_0^T is a point process with $\mathcal{G}_t \triangleq \sigma(M, Y_0^t)$-intensity Γ_0^T such that $\mathbb{E}[\int_0^T \Gamma_t\, dt] < \infty$ and Z_0^T is obtained from p-thinning Y_0^T. Then the $(\mathcal{F}_t \triangleq \sigma(M, Z_0^t) : t \in [0,T])$-intensity of Z_0^T is given by Θ_0^T, where P-a.s. $\Theta_t = (1-p)\mathbb{E}[\Gamma_t | \mathcal{F}_{t-}], t \in [0,T]$.*

Proof: Please see the Appendix. □

The following theorem was first proven by Wang in [66] using a property of certain "contraction coefficient" used in strong data processing inequalities [51]. Here, we provide a self-contained proof which relies on Theorem 7.

Theorem 9 (Strong data Processing inequality for Poisson process under thinning) *Let Y_0^T be a Poisson process with rate λ, and M be such that $I(M; Y_0^T) < \infty$. Let Z_0^T obtained from p-thinning of Y_0^T such that the thinning operation is independent of M. Then*

$$I(M; Z_0^T) \leq (1-p) I(M; Y_0^T).$$

Proof: The data processing inequality gives $I(M; Z_0^T) \leq I(M; Y_0^T) < \infty$. Applying Theorem 7,

$$I(M; Y_0^T) = \mathbb{E}\left[\int_0^T \phi(\Gamma_t) - \phi(\lambda)\, dt\right], \tag{3.14}$$

and

$$I(M; Z_0^T) = \mathbb{E}\left[\int_0^T \phi(\hat{\Gamma}_t) - \phi(\hat{\lambda}_t)\, dt\right], \tag{3.15}$$

where Γ_0^T and λ_0^T (respectively $\hat{\Gamma}_0^T$ and $\hat{\lambda}_0^T$) are the $(\sigma(M; Y_0^t) : t \in [0,T])$ and $(\sigma(Y_0^t) : t \in [0,T])$-intensities (respectively $(\sigma(M; Z_0^t) : t \in [0,T])$ and $(\sigma(Z_0^t) : t \in [0,T])$-intensities) of Y_0^T (Z_0^T). Due to the uniqueness of the intensities and Lemma 23, we can take for each $t \in [0,T]$,

$$\hat{\Gamma}_t = (1-p)\mathbb{E}[\Gamma_t | M, Z_0^{t-}], \quad \hat{\lambda}_t = (1-p)\lambda.$$

Noting that $\phi((1-p)x) = (1-p)\phi(x) + x\phi(1-p)$, (3.15) yields

$$\begin{aligned} I(M; Z_0^T) =& (1-p)\mathbb{E}\left[\int_0^T \phi(\mathbb{E}[\Gamma_t | M, Z_0^{t-}]) - \phi(\lambda)\, dt\right] \\ & + \phi(1-p)\mathbb{E}\left[\int_0^T \Gamma_t - \lambda\, dt\right] \\ \overset{(a)}{=}& (1-p)\mathbb{E}\left[\int_0^T \phi(\mathbb{E}[\Gamma_t | M, Z_0^{t-}]) - \phi(\lambda)\, dt\right] \\ \overset{(b)}{\leq}& (1-p)\mathbb{E}\left[\int_0^T \phi(\Gamma_t) - \phi(\lambda)\, dt\right] \\ =& (1-p) I(M; Y_0^T), \end{aligned}$$

where for (a) we have used the fact that $\mathbb{E}\left[\int_0^T \Gamma_t\, dt\right] = \mathbb{E}\left[\int_0^T 1\, dY_t\right] = \mathbb{E}\left[\int_0^T \lambda\, dt\right]$, for (b) we have used Jensen's inequality. □

We will require the following result [21, Theorem 2.11, p. 106].

Lemma 24 *Suppose that Y_0^T is a Poisson process with rate λ and Z_0^T is obtained from p-thinning of Y_0^T. Let*

$$\hat{Z}_t = Y_t - Z_t \quad t \in [0,T].$$

Then $\hat{Z}_0^T$ and Z_0^T are independent Poisson processes with rates $p\lambda$ and $(1-p)\lambda$ respectively.

The following lemma will be used repeatedly in the converse proofs of the rate-distortion function.

Lemma 25 *Let a point process Y_0^T have an $(\mathcal{F}_t : t \in [0,T])$-intensity Γ_0^T such that $\mathbb{E}\left[\int_0^T \phi(\Gamma_t)\,dt\right] < \infty$. Let $\hat{Y}_0^T$ be an non-negative $(\mathcal{F}_t : t \in [0,T])$-predictable process satisfying $\mathbb{E}\left[\int_0^T \hat{Y}_t\,dt\right] < \infty$. Then*

$$\mathbb{E}\left[\int_0^T \log(\hat{Y}_t)\,dY_t\right] = \mathbb{E}\left[\int_0^T \log(\hat{Y}_t)\Gamma_t\,dt\right].$$

Proof: Please see the Appendix. □

3.3 Functional Covering of Point Processes

In this section, we will consider general point processes and obtain the rate-distortion function under the functional-covering distortion when feedforward is present. Stronger results are obtained for Poisson processes in the next sections.

Definition 14 *Given a point process $y_0^T \in \mathcal{N}_0^T$, and a non-negative function $\hat{y}_0^T$, the* functional-covering distortion *d is*

$$d(\hat{y}_0^T, y_0^T) \triangleq \left(\int_0^T \hat{y}_t\,dt - \log(\hat{y}_t)\,dy_t\right),$$

whenever the expression on the right is well defined.

We will allow the reconstruction function $\hat{Y}_0^T$ to depend on Y_0^T as well along with the message via predictability. In particular, we will call $\hat{Y}_0^T$ as *allowable* *reconstruction with feedforward* if it is non-negative and $(\sigma(Y_0^t) : t \in [0,T])$-predictable. Let $\hat{\mathcal{Y}}_{0,\mathrm{FF}}^T$ denote the set of all $\hat{y}_0^T$ processes which are allowable reconstructions with feedforward.

Definition 15 *A* (T, R, D) code with feedforward *consists of an* encoder f

$$f : \mathcal{N}_0^T \to \{1, \ldots, \ldots, \lceil \exp(RT) \rceil\}$$

and a decoder g

$$g : \{1, \ldots, \lceil \exp(RT) \rceil\} \times \mathcal{N}_0^T \to \hat{\mathcal{Y}}_{0,FF}^T$$

satisfying

$$\mathbb{E}\left[\int_0^T \hat{Y}_t \, dt\right] < \infty$$

and the distortion constraint

$$\mathbb{E}\left[\frac{1}{T} d(\hat{Y}_0^T, Y_0^T)\right] \leq D.$$

We will call the encoder's output $M = f(Y_0^T)$ the *message* and the decoder's output $\hat{Y}_0^T$ the reconstruction.

Definition 16 *The* minimum achievable distortion with feedforward at rate R and blocklength T *is*

$$D_F^*(R, T) \triangleq \inf\{D : \textit{There exists a } (T, R, D) \textit{ code with feedforward}\}.$$

Definition 17 *The* distortion-rate function with feedforward *is*

$$D_F(R) \triangleq \limsup_{T \to \infty} D_F^*(R, T).$$

The minimum achievable rate at distortion D and blocklength T with feedforward $R_F^*(D,T)$ and the rate-distortion function with feedforward $R_F(D)$ can be defined similarly.

$D_F^*(R,T)$ can be characterized via the following theorem for certain point processes.

Theorem 10 (Functional-Covering Distortion with Feedforward) *Let Y_0^T be a point process with $(\mathcal{F}_t^Y : t \in [0,T])$-intensity Λ_0^T such that $\mathbb{E}[\int_0^T |\phi(\Lambda_t)|\, dt] < \infty$. Let*

$$\Xi(Y_0^T) \triangleq \frac{1}{T}\mathbb{E}\left[\int_0^T \Lambda_t - \phi(\Lambda_t)\, dt\right],$$

and

$$\delta_T \triangleq P(Y_T = 0) < 1.$$

Then $D_F^*(R,T)$ satisfies

$$\Xi(Y_0^T) - R - \frac{1}{T} \leq D_F^*(R,T) \leq \Xi(Y_0^T) - (1-\delta_T)R + \frac{1}{T}.$$

Proof:

Achievability:

Recall that since Λ_0^T is the $(\mathcal{F}_t^Y : t \in [0,T])$-intensity of Y_0^T, it is $(\mathcal{F}_t^Y : t \in [0,T])$-predictable, and $\mathbb{E}[\int_0^T |\phi(\Lambda_t)|\, dt] < \infty$ implies $\mathbb{E}[\int_0^T \Lambda_t\, dt] < \infty$. If the decoder outputs Λ_0^T, this leads to distortion

$$\begin{aligned}\frac{1}{T}\mathbb{E}[d(\Lambda_0^T, Y_0^T)] &= \frac{1}{T}\mathbb{E}\left[\int_0^T \Lambda_t\, dt - \log(\Lambda_t)\, dY_t\right] \\ &= \frac{1}{T}\mathbb{E}\left[\int_0^T \Lambda_t - \phi(\Lambda_t)\, dt\right] \\ &= \Xi(Y_0^T).\end{aligned}$$

Thus $D_F^*(0,T) \leq \Xi(Y_0^T)$, and the upper bound in the statement of the theorem holds at $R = 0$.

Now consider the case $R > 0$. Fix $T > 0$ and let $J = \lceil \exp(RT) \rceil$. If $Y_T = 0$, then the encoder sends index $M = 1$. Otherwise, let Θ denote the first arrival instant of the observed point process Y_0^T. From Lemma 16, we have that $P^{Y_0^T} \ll P_0^{Y_0^T}$. Since under $P_0^{Y_0^T}$, Y_0^T is a Poisson process with unit rate, it holds that $P_0^{Y_0^T}(\Theta = t, Y_T > 0) = 0$ for any fixed $t \in [0,T]$. This gives us $P(\Theta = t, Y_T > 0) = 0$ for $t \in [0,T]$. Thus conditioned on the event $Y_T > 0$, Θ has a continuous distribution function F_Θ. The encoder computes $F_\Theta(\Theta)$ which is uniformly distributed over $[0,1]$, which the encoder suitably quantizes to obtain an M which is uniform in $\{2, \ldots, J\}$. From Theorem 7, there exists a $(\sigma(M, Y_0^t) : t \in [0,T])$-predictable process Γ_0^T which is the $(\sigma(M, Y_0^t) : t \in [0,T])$-intensity of Y_0^T. We note that $\mathbb{E}\left[\int_0^T \Gamma_t \, dt\right] = \mathbb{E}\left[\int_0^T \Lambda_t \, dt\right] < \infty$, and from Theorem 7 $\mathbb{E}\left[\int_0^T \log(\Gamma_t) \, dY_t\right] < \infty$. Hence

$$\frac{1}{T}\mathbb{E}[d(\Lambda_0^T, Y_0^T)] = \frac{1}{T}\mathbb{E}\left[\int_0^T \Gamma_t \, dt - \log(\Gamma_t) \, dY_t\right]$$

is well defined. The decoder outputs Γ_0^T as its reconstruction. Then we have,

$$\begin{aligned}
\frac{1}{T}H(M) &= -\frac{1}{T}\left(\delta_T \log(\delta_T) + (1-\delta_T)\log(1-\delta_T)\right) + \frac{1-\delta_T}{T}(\log(J-1)) \\
&\stackrel{(a)}{\geq} \frac{1-\delta_T}{T}(\log(J-1)) \\
&\stackrel{(b)}{\geq} \frac{1-\delta_T}{T}\log(J/\exp(1)) \\
&\stackrel{(c)}{\geq} (1-\delta_T)R - \frac{1}{T},
\end{aligned} \tag{3.16}$$

where for (a), we have used the bound $-\delta_T \log(\delta_T) - (1-\delta_T)\log(1-\delta_T) \geq 0$,

for (b), we have used the inequality $J - 1 \geq J/\exp(1)$ when $J \geq 2$,

for (c), we used the fact that $RT \leq \log(J)$.

$H(M)$ also satisfies

$$
\begin{aligned}
\frac{1}{T}H(M) &\stackrel{(a)}{=} \frac{1}{T}I(M;Y_0^T) \\
&\stackrel{(b)}{=} \frac{1}{T}\mathbb{E}\left[\int_0^T \log(\Gamma_t)\,dY_t\right] - \frac{1}{T}\mathbb{E}\left[\int_0^T \phi(\Lambda_t)\,dt\right],
\end{aligned} \tag{3.17}
$$

where, for (a) we have used Lemma 15,

for (b) we have used Theorem 7.

The average distortion can be bounded as follows:

$$
\begin{aligned}
\frac{1}{T}\mathbb{E}[d(\Lambda_0^T, Y_0^T)] &= \frac{1}{T}\mathbb{E}\left[\int_0^T \Gamma_t\,dt - \log(\Gamma_t)\,dY_t\right] \\
&\stackrel{(a)}{=} \frac{1}{T}\mathbb{E}\left[\int_0^T \Gamma_t\,dt\right] - \frac{1}{T}\mathbb{E}\left[\int_0^T \log(\Gamma_t)\,dY_t\right] \\
&\stackrel{(b)}{=} \frac{1}{T}\mathbb{E}\left[\int_0^T \Lambda_t\,dt\right] - \frac{1}{T}\mathbb{E}\left[\int_0^T \log(\Gamma_t)\,dY_t\right] \\
&\stackrel{(c)}{=} \frac{1}{T}\mathbb{E}\left[\int_0^T \Lambda_t\,dt\right] - \frac{1}{T}H(M) - \frac{1}{T}\mathbb{E}\left[\int_0^T \phi(\Lambda_t)\,dt\right] \\
&\stackrel{(d)}{\leq} \frac{1}{T}\mathbb{E}\left[\int_0^T \Lambda_t - \phi(\Lambda_t)\,dt\right] - (1-\delta_T)R + \frac{1}{T} \\
&= \Xi(Y_0^T) - (1-\delta_T)R + \frac{1}{T},
\end{aligned}
$$

where, for (a), we have used the fact that $\mathbb{E}\left[\int_0^T \log(\Gamma_t)\,dY_t\right] < \infty$ due to Theorem 7,

for (b), we used the equality $\mathbb{E}\left[\int_0^T \Gamma_t\,dt\right] = \mathbb{E}\left[\int_0^T \Lambda_t\,dt\right]$,

for (c), we used (3.17),

for (d), we used (3.16).

Thus we have shown the existence of a (T, R, D) code with feedforward such that $D = \Xi(Y_0^T) - (1-\delta_T)R + \frac{1}{T}$. This gives the upper bound on $D_F^*(R, T)$.

Converse:

For the given (T, R, D) code with feedforward, let $J = \lceil \exp(RT) \rceil$. Then $J \leq \exp(RT) + 1 \leq \exp(RT + 1)$. Thus we have

$$R + \frac{1}{T} \geq \frac{1}{T} \log(J) \geq \frac{1}{T} H(M) \overset{(a)}{=} \frac{1}{T} I(M; Y_0^T), \tag{3.18}$$

where (a) follows because of Lemma 15.

Since $I(M; Y_0^T) < \infty$, we conclude from Theorem 7 that there exists a process Γ_0^T such that Γ_0^T is the $(\mathcal{F}_t = \sigma(M, Y_0^t) : t \in [0, T])$ intensity of Y_0^T and

$$I(M; Y_0^T) = \mathbb{E}\left[\int_0^T \phi(\Gamma_t)\, dt\right] - \mathbb{E}\left[\int_0^T \phi(\Lambda_t)\, dt\right].$$

Hence from (3.18)

$$R \geq \frac{1}{T}\mathbb{E}\left[\int_0^T \phi(\Gamma_t)\, dt\right] - \frac{1}{T}\mathbb{E}\left[\int_0^T \phi(\Lambda_t)\, dt\right] - \frac{1}{T}. \tag{3.19}$$

Let $\hat{Y}_0^T$ denote the decoder's output. The distortion constraint D satisfies

$$\begin{aligned} D \geq \frac{1}{T}\mathbb{E}\left[d(\hat{Y}_0^T, Y_0^T)\right] &= \frac{1}{T}\mathbb{E}\left[\int_0^T \hat{Y}_t\, dt - \log(\hat{Y}_t)\, dY_t\right] \\ &= \frac{1}{T}\mathbb{E}\left[\int_0^T \hat{Y}_t - \log(\hat{Y}_t)\Gamma_t\, dt\right] \end{aligned} \tag{3.20}$$

where in the last line we have used Lemma 25.

Using the inequality $u \log(v) \leq \phi(u) - u + v$, and noting that the individual terms have finite expectations,

$$\begin{aligned} \mathbb{E}\left[\int_0^T \log(\hat{Y}_t)\Gamma_t\, dt\right] &\leq \mathbb{E}\left[\int_0^T \phi(\Gamma_t) - \Gamma_t + \hat{Y}_t\, dt\right] \\ &= \mathbb{E}\left[\int_0^T \phi(\Gamma_t)\, dt\right] - \mathbb{E}\left[\int_0^T \Gamma_t\, dt\right] + \mathbb{E}\left[\int_0^T \hat{Y}_t\, dt\right]. \end{aligned} \tag{3.21}$$

From (3.20) and (3.19), we deduce

$$R + D \geq \frac{1}{T}\mathbb{E}\left[\int_0^T \phi(\Gamma_t)\, dt\right] - \frac{1}{T}\mathbb{E}\left[\int_0^T \phi(\Lambda_t)\, dt\right] + \frac{1}{T}\mathbb{E}\left[\int_0^T \hat{Y}_t\, dt\right] - \frac{1}{T}\mathbb{E}\left[\int_0^T \log(\hat{Y}_t)\, dY_t\right] - \frac{1}{T}$$

$$\overset{(a)}{\geq} \frac{1}{T}\mathbb{E}\left[\int_0^T \Gamma_t\, dt\right] - \frac{1}{T}\mathbb{E}\left[\int_0^T \phi(\Lambda_t)\, dt\right] - \frac{1}{T}$$

$$\overset{(b)}{\geq} \frac{1}{T}\mathbb{E}\left[\int_0^T \Lambda_t\, dt\right] - \frac{1}{T}\mathbb{E}\left[\int_0^T \phi(\Lambda_t)\, dt\right] - \frac{1}{T}$$

$$= \Xi(Y_0^T) - \frac{1}{T},$$

where, for (a) we have used (3.21),

for (b) we use the fact that $\mathbb{E}\left[\int_0^T \Gamma_t\, dt\right] = \mathbb{E}\left[\int_0^T dY_t\right] = \mathbb{E}\left[\int_0^T \Lambda_t\, dt\right]$.

Hence we have shown that for any (T, R, D) code with feedforward, $D \geq \Xi(Y_0^T) - R - 1/T$. This gives us the lower bound on $D_F^*(R, T)$ □

Corollary 1 *Let Y_0^T be a point process with $(\mathcal{F}_t^Y : t \in [0, T])$-intensity Λ_0^T such that*

- $\mathbb{E}[\int_0^T |\phi(\Lambda_t)|\, dt] < \infty$,
- $\bar{\Xi}(Y) \triangleq \limsup_{T\to\infty} \frac{1}{T}\mathbb{E}\left[\int_0^T \Lambda_t - \phi(\Lambda_t)\, dt\right]$ *is finite.*
- $\lim_{T\to\infty} P(Y_T = 0) = 0$.

Then

$$D_F(R) = \bar{\Xi}(Y) - R.$$

Proof: The corollary follows from the definition $D_F(R) = \limsup_{T\to\infty} D_F^*(R, T)$ and from the bounds on $D_F^*(R, T)$ in the Theorem 10. □

Remark 1 *The above distortion-rate function is reminiscent of logarithmic loss distortion-rate function for a DMS. Specifically, for a DMS Y on alphabet $\mathcal{Y}$ let* *reconstruction be a probability distribution function Q on $\mathcal{Y}$. The logarithmic loss distortion is defined as $d_{LL}(y, Q) \triangleq -\log(Q(y))$ and the distortion-rate function is then given by $D(R) = (H(Y) - R)^+$ [12].*

If the reconstruction $\hat{y}_0^T$ is assumed to be bounded then it can be used define a probability measure on the space of point-processes $(\mathcal{N}_0^T, \mathfrak{F}^N)$ via following Radon-Nikodym derivative.

$$\frac{dP_{\hat{y}_0^T}}{dP_0}(y_0^T) = \exp\left(\int_0^T \log(\hat{y}_t)\, dy_t - (\hat{y}_t - 1)\, dt\right),$$

where P_0 is the measure under which Y_0^T is a Poisson process with unit rate. Then the intensity of Y_0^T under this measure is $\hat{y}_0^T$ [9, Chapter VI, Theorems T2-T4] and the functional-covering distortion is related to the above Radon-Nikodym derivative as

$$d(\hat{y}_0^T, y_0^T) = -\log\left(\frac{dP_{\hat{y}_0^T}}{dP_0}(y_0^T)\right) + T.$$

$\Diamond$

Applying the above corollary to a Poisson process with rate $\lambda > 0$, we get that $D_F(R) = \lambda - \lambda \log(\lambda) - R$. As we will see in the next section, this distortion-rate function can be achieved without feedforward.

3.4 Constrained Functional-Covering of Poisson Processes

In this and the next section we focus on Poisson processes. Let $\hat{\mathcal{Y}}_0^T$ denote the set of all functions $\hat{y}_0^T$ which are non-negative and left-continuous with right-limits. We assume that we are given a set $\mathcal{A} \in \mathbb{R}_+$ with at least one positive element. We will constrain the reconstruction function $\hat{Y}_0^T$ to take value in $\mathcal{A}$, so that for all $t \in [0, T]$, $\hat{Y}_t \in \mathcal{A}$.

Definition 18 *A* (T, R, D) *code consists of an* encoder f

$$f : \mathcal{N}_0^T \to \{1, \ldots, \lceil \exp(RT) \rceil\}$$

and a decoder g

$$g : \{1, \ldots, \lceil \exp(RT) \rceil\} \to \hat{\mathcal{Y}}_0^T$$

satisfying

$$\hat{Y}_t \in \mathcal{A},\ \mathbb{E}\left[\int_0^T \hat{Y}_t \, dt\right] < \infty$$

and the distortion constraint

$$\frac{1}{T}\mathbb{E}\left[d(\hat{Y}_0^T, Y_0^T)\right] \le D.$$

We will call the encoder's output $M = f(Y_0^T)$ the *message* and the decoder's output $\hat{Y}_0^T = g(M)$ the reconstruction.

Definition 19 *A rate-distortion vector* (R, D) *is said to be* achievable *if for any* $\epsilon > 0$*, there exists a sequence of* $(T_n, R+\epsilon, D+\epsilon)$ *codes such that* $\lim_{n\to\infty} T_n = \infty$.

Definition 20 *The* rate-distortion region $\mathfrak{RD}_{\mathcal{A}}^{\mathcal{P}}$ *is the intersection of all achievable rate-distortion vectors* (R, D).

When feedforward is present, recall that $\hat{Y}_0^T$ is an allowable reconstruction with feedforward if it is non-negative and $(\sigma(Y_0^t) : t \in [0, T])$-predictable, and that $\hat{\mathcal{Y}}_{0,\mathrm{FF}}^T$ denotes the set of all $\hat{y}_0^T$ processes which are allowable reconstructions with feedforward.

Definition 21 *A* (T, R, D) code with feedforward *consists of an* encoder f

$$f : \mathcal{N}_0^T \to \{1, \ldots, \ldots, \lceil \exp(RT) \rceil\}$$

and a decoder g

$$g : \{1, \ldots, \lceil \exp(RT) \rceil\} \times \mathcal{N}_0^T \to \hat{\mathcal{Y}}_{0,FF}^T$$

satisfying

$$\hat{Y}_t \in \mathcal{A}, \, \mathbb{E}\left[\int_0^T \hat{Y}_t \, dt\right] < \infty$$

and the distortion constraint

$$\mathbb{E}\left[\frac{1}{T} d(\hat{Y}_0^T, Y_0^T)\right] \leq D.$$

The *rate-distortion region* $\mathfrak{RD}_{\mathcal{A}}^{\mathcal{P},\mathrm{F}}$ with feedforward is defined similarly.

Theorem 11 (Constrained functional-covering of Poisson processes) *The rate-distortion region for the constrained functional-covering of a Poisson process with rate* $\lambda > 0$ *is given by*

$$\mathfrak{RD}_{\mathcal{A}}^{\mathcal{P}} = \mathfrak{RD}_{\mathcal{A}}^{\mathcal{P},F} = \mathfrak{RD},$$

where $\mathfrak{RD}$ *is the convex hull of union of sets of rate-distortion vectors* (R, D) *such that*

$$R \geq \lambda \sum_{k=1}^{4} \beta_k \log\left(\frac{\beta_k}{\alpha_k}\right),$$
$$D \geq \sum_{k=1}^{4} \alpha_k \Psi_{\mathcal{A}}\left(\frac{\lambda \beta_k}{\alpha_k}\right),$$

where

$$\Psi_{\mathcal{A}}(u) \triangleq \inf_{v \in \mathcal{A}} v - u \log(v)$$

with the convention that $0\Psi(0/0) = 0$*, and* $[\alpha_k]_{k=1}^4$ *and* $[\beta_k]_{k=1}^4$ *are probability vectors over* $\{1, 2, 3, 4\}$ *satisfying* $\alpha_k = 0 \Rightarrow \beta_k = 0$.

Proof:

Achievability Let

$$R \triangleq \lambda \sum_{k=1}^{4} \beta_k \log\left(\frac{\beta_k}{\alpha_k}\right),$$
$$D \triangleq \sum_{k=1}^{4} \alpha_k \Psi_{\mathcal{A}}\left(\frac{\lambda\beta_k}{\alpha_k}\right).$$

We will show achievability using a $(T, R+\epsilon, D+\epsilon)$ code without feedforward. We will use discretization and results from the rate-distortion theory for discrete memoryless sources (DMS). Define a binary-valued discrete-time process $(\bar{Y}_j : j \in \{1,\dots,n\})$ as follows. If there are one or more arrivals in the interval $((j-1)\Delta, j\Delta]$ of the process Y_0^T, then set $\bar{Y}_j$ to 1, otherwise it equals zero. Since Y_0^T is a Poisson process with rate λ, the components of $(\bar{Y}_j : j \in \{1,\dots,n\})$ are independent and identically distributed with $P(\bar{Y}=1) = 1-\exp(-\lambda\Delta)$. Consider the following "test"-channel for $k \in \{1,2,3,4\}$

$$P(\bar{U}=k|\bar{Y}=1) = \beta_k,$$
$$P(\bar{U}=k|\bar{Y}=0) = \alpha_k. \tag{3.22}$$

Define the discretized distortion function

$$\bar{d}(\hat{y},\bar{y}) \triangleq \hat{y} - \frac{\log(\hat{y})}{\Delta}\mathbf{1}\{\bar{y}=1\} \quad \hat{y}\in\mathcal{A}, \bar{y}\in\{0,1\}. \tag{3.23}$$

The reconstruction $\hat{\bar{Y}}(k)$ is taken as a $v \in \mathcal{A}$ satisfying

$$\left|\Psi_{\mathcal{A}}\left(\frac{\lambda\beta_k}{\alpha_k}\right) - \left(v - \frac{\lambda\beta_k}{\alpha_k}\log(v)\right)\right| \le \frac{\epsilon}{4}, \tag{3.24}$$

where such a v exists due to the definition of $\Psi_{\mathcal{A}}$. We recall that if $\alpha_k = 0$ then $\beta_k = 0$, and hence $P(\bar{U}=k) = 0$ for such a k. The scaling of the mutual information $I(\bar{U};\bar{Y})$ and the distortion function $\bar{d}(\hat{\bar{Y}},\bar{Y})$ with respect to Δ is given by the following lemma.

Lemma 26

$$\lim_{\Delta \to 0} \frac{I(\bar{U}; \bar{Y})}{\Delta} = R$$

$$\lim_{\Delta \to 0} \mathbb{E}[\bar{d}(\hat{\bar{Y}}, \bar{Y})] \leq D + \frac{\epsilon}{4}$$

Proof:

Please see the Appendix. □

Let

$$\kappa \triangleq \max_{\substack{k \in \{1,2,3\} \\ \hat{\bar{Y}}(k) > 0}} \left| \log \left(\hat{\bar{Y}}(k) \right) \right|. \tag{3.25}$$

Due to [25, Theorem 9.3.2, p. 455], for a given $\Delta > 0$, $\bar{\epsilon} > 0$, and all sufficiently large n, there exists an encoder $\bar{f}$ and a decoder $\bar{g}$ such that

$$\bar{f} : (\bar{Y}_j : j \in \{1, \dots, n\}) \to \{1, \dots, L\}$$

$$\bar{g} : \{1, \dots, L\} \to (\hat{\bar{Y}}_j : j \in \{1, \dots, n\})$$

satisfying

$$\frac{1}{n} \log(L) \leq I(\bar{U}; \bar{Y}) + \bar{\epsilon},$$

$$\mathbb{E}\left[\frac{1}{n} \sum_{j=1}^{n} \bar{d}(\hat{\bar{Y}}_j, \bar{Y}_j) \right] \leq \mathbb{E}[\bar{d}(\hat{\bar{Y}}, \bar{Y})] + \bar{\epsilon}. \tag{3.26}$$

Given the above setup, the encoder f upon observing Y_0^T obtains the binary valued discrete time process $(\bar{Y}_j : j \in \{1, \dots, n\})$, and sends $M = \bar{f}(\bar{Y}_j : j \in \{1, \dots, n\})$ to the decoder. The decoder outputs the reconstruction $\hat{Y}_0^T$ as

$$\hat{Y}_t \triangleq \sum_{j=1}^{n} \hat{\bar{Y}}_j \mathbf{1}\left\{ t \in ((j-1)\Delta, j\Delta] \right\} \quad t \in [0, T].$$

Let $\bar{\bar{Y}}_j$ denote the actual number of arrivals of Y_0^T in an interval $((j-1)\Delta, j\Delta]$. Then $\bar{d}$ is related to the original distortion function via the above reconstruction as follows:

$$\begin{aligned}
\frac{1}{T} d(\hat{Y}_0^T; Y_0^T) &= \frac{1}{T}\int_0^T \hat{Y}_t\, dt - \frac{1}{T}\int_0^T \log(\hat{Y}_t)\, dY_t \\
&= \frac{1}{n}\sum_{j=1}^n \hat{\bar{Y}}_j - \frac{1}{T}\sum_{j=1}^n \log(\hat{\bar{Y}}_j)\bar{\bar{Y}}_j \\
&= \frac{1}{n}\sum_{j=1}^n \hat{\bar{Y}}_j - \frac{1}{n\Delta}\sum_{j=1}^n \log(\hat{\bar{Y}}_j)\bar{Y}_j - \frac{1}{T}\sum_{j=1}^n \log(\hat{\bar{Y}}_j)(\bar{\bar{Y}}_j - 1)\mathbf{1}\{\bar{\bar{Y}}_j > 1\}. \\
&= \frac{1}{n}\sum_{j=1}^n \bar{d}(\hat{\bar{Y}}_j, \bar{Y}_j) - \frac{1}{T}\sum_{j=1}^n \log(\hat{\bar{Y}}_j)(\bar{\bar{Y}}_j - 1)\mathbf{1}\{\bar{\bar{Y}}_j > 1\} \\
&\overset{(a)}{\leq} \frac{1}{n}\sum_{j=1}^n \bar{d}(\hat{\bar{Y}}_j, \bar{Y}_j) + \frac{\kappa}{T}\sum_{j=1}^n (\bar{\bar{Y}}_j - 1)\mathbf{1}\{\bar{\bar{Y}}_j > 1\} \\
&\leq \frac{1}{n}\sum_{j=1}^n \bar{d}(\hat{\bar{Y}}_j, \bar{Y}_j) + \frac{\kappa}{T}\sum_{j=1}^n \bar{\bar{Y}}_j\mathbf{1}\{\bar{\bar{Y}}_j > 1\},
\end{aligned}$$

where for (a), we have used the definition of κ in (3.25), since $\bar{\bar{Y}}_j > 1$ implies $\bar{Y}_j = 1$ which implies $\hat{\bar{Y}}_j > 0$ in order for $\bar{d}(\hat{\bar{Y}}_j, 1) < \infty$, which occurs a.s. since $\mathbb{E}[\bar{d}(\hat{\bar{Y}}, \bar{Y})] < \infty$ so long as Δ is sufficiently small.

Hence taking the expectation, we get

$$\begin{aligned}
\mathbb{E}\left[\frac{1}{T} d(\hat{Y}_0^T, Y_0^T)\right] &\leq \mathbb{E}\left[\frac{1}{n}\sum_{j=1}^n \bar{d}(\hat{\bar{Y}}_j, \bar{Y}_j)\right] + \kappa\mathbb{E}\left[\frac{1}{T}\sum_{j=1}^n \bar{\bar{Y}}_j\mathbf{1}\{\bar{\bar{Y}}_j > 1\}\right] \\
&\overset{(a)}{\leq} \mathbb{E}[\bar{d}(\hat{\bar{Y}}, \bar{Y})] + \kappa\mathbb{E}\left[\frac{1}{T}\sum_{j=1}^n \bar{\bar{Y}}_j\mathbf{1}\{\bar{\bar{Y}}_j > 1\}\right] + \bar{\epsilon} \\
&\overset{(b)}{=} \mathbb{E}[\bar{d}(\hat{\bar{Y}}, \bar{Y})] + \kappa(\lambda - \lambda\exp(-\lambda\Delta)) + \bar{\epsilon} \\
&\overset{(c)}{\leq} \mathbb{E}[\bar{d}(\hat{\bar{Y}}, \bar{Y})] + \kappa\lambda^2\Delta + \bar{\epsilon}, \qquad (3.27)
\end{aligned}$$

where, for (a), we have used (3.26),

for (b) we note that $\mathbb{E}[\bar{\bar{Y}}_j\mathbf{1}\{\bar{\bar{Y}}_j > 1\}] = \lambda\Delta - \lambda\Delta\exp(-\lambda\Delta)$,

for (c), we have used the inequality $1 - u \leq \exp(-u)$.

Moreover using (3.26),

$$\frac{1}{T}\log(L) = \frac{1}{n\Delta}\log(L) \leq \frac{I(\bar{U};\bar{Y})}{\Delta} + \frac{\bar{\epsilon}}{\Delta}. \tag{3.28}$$

Now given the rate-distortion vector (R, D) and $\epsilon > 0$, first choose $\Delta < 1$ sufficiently small so that

$$\frac{I(\bar{U};\bar{Y})}{\Delta} \leq R + \frac{\epsilon}{4}$$
$$\mathbb{E}[\bar{d}(\hat{\bar{Y}}, \bar{Y})] \leq D + \frac{\epsilon}{2},$$
$$\kappa\lambda^2\Delta \leq \epsilon/2$$

Then let $\bar{\epsilon} = \Delta\epsilon/4$, and choose a sufficiently large n so that (3.26) is satisfied. From (3.27) and (3.28) we conclude that a sequence of $(T_n, R+\epsilon, D+\epsilon)$ code exists with $T_n = n\Delta$ and $T_n \to \infty$ as $n \to \infty$.

Converse We will prove the converse when feedforward is present. For the given $(T, R+\epsilon, D+\epsilon)$ code with feedforward, let M denote the encoder's output. Since $I(M; Y_0^T) < \infty$, we conclude from Theorem 7 that there exists a process Γ_0^T such that Γ_0^T is the $(\mathcal{F}_t = \sigma(M, Y_0^t) : t \in [0, T])$ intensity of Y_0^T and

$$I(M; Y_0^T) = \mathbb{E}\left[\int_0^T \phi(\Gamma_t)\, dt\right] - T\phi(\lambda). \tag{3.29}$$

We also have

$$\frac{1}{T}I(M; Y_0^T) = \frac{1}{T}H(M) \leq \frac{1}{T}\log\left(\lceil\exp((R+\epsilon)T)\rceil\right) \leq R + \epsilon + \frac{1}{T}.$$

This gives

$$R \geq \frac{1}{T}\mathbb{E}\left[\int_0^T \phi(\Gamma_t)\, dt\right] - \phi(\lambda) - \epsilon - \frac{1}{T}.$$

Let $\hat{Y}_0^T$ denote the decoder's output. The distortion constraint D satisfies

$$\begin{aligned}
D &\geq \frac{1}{T}\mathbb{E}\left[d(\hat{Y}_0^T, Y_0^T)\right] - \epsilon \\
&= \frac{1}{T}\mathbb{E}\left[\int_0^T \hat{Y}_t\, dt - \log(\hat{Y}_t)\, dY_t\right] - \epsilon \\
&\overset{(a)}{=} \frac{1}{T}\mathbb{E}\left[\int_0^T \hat{Y}_t - \Gamma_t \log(\hat{Y}_t)\ dt\right] - \epsilon \\
&\geq \frac{1}{T}\mathbb{E}\left[\int_0^T \inf_{v\in\mathcal{A}} v - \Gamma_t \log(v)\, dt\right] - \epsilon \\
&\overset{(b)}{=} \frac{1}{T}\mathbb{E}\left[\int_0^T \Psi_{\mathcal{A}}(\Gamma_t)\, dt\right] - \epsilon, \qquad (3.30)
\end{aligned}$$

where, for (a) we have used Lemma 25,

for (b), we have used the definition of $\Psi_{\mathcal{A}}$.

Defining S to be uniformly distributed on $[0, T]$, and independent of all other random variables we have

$$R \geq \mathbb{E}\left[\phi(\Gamma_S)\right] - \phi(\lambda) - \epsilon - \frac{1}{T} \qquad (3.31)$$

$$D \geq \mathbb{E}\left[\Psi_{\mathcal{A}}(\Gamma_S)\right] - \epsilon, \qquad (3.32)$$

Now we use Carathéodory's theorem [53, Theorem 17.1]. There exist non-negative $[\eta_k]_{k=1}^4$ and $[\alpha_k]_{k=1}^4$, such that $\sum_{k=1}^4 \alpha_k = 1$ and

$$\mathbb{E}\left[\phi(\Gamma_S)\right] = \sum_{k=1}^4 \alpha_k \phi(\eta_k), \qquad (3.33)$$

$$\mathbb{E}\left[\Psi_{\mathcal{A}}(\Gamma_S)\right] = \sum_{k=1}^4 \alpha_k \Psi_{\mathcal{A}}(\eta_k), \qquad (3.34)$$

$$\mathbb{E}\left[\Gamma_S\right] = \sum_{k=1}^4 \alpha_k \eta_k = \lambda, \qquad (3.35)$$

where in the last line we have used the fact that since Γ_0^T is the $(\sigma(M, Y_0^T) : t \in [0, T])$-intensity of Y_0^T, $\mathbb{E}\left[\int_0^T \Gamma_t\, dt\right] = \mathbb{E}[Y_T] = T\lambda$. Now define

$$\beta_k \triangleq \frac{\alpha_k \eta_k}{\lambda}.$$

We note that $\beta_k = 0$ if $\alpha_k = 0$, and $\sum_{k=1}^{4} \beta_k = 1$. Substituting the above definitions in (3.31)-(3.32), we get

$$\begin{aligned} R &\geq \left(\sum_{k=1}^{4} \alpha_k \eta_k \log(\eta_k) - \lambda \log(\lambda) \right) - \epsilon - \frac{1}{T} \\ &= \lambda \left(\sum_{k=1}^{4} \beta_k \log\left(\frac{\beta_k \lambda}{\alpha_k} \right) \mathbf{1}\{\alpha_k > 0\} - \log(\lambda) \right) - \epsilon - \frac{1}{T} \\ &= \lambda \sum_{k=1}^{4} \beta_k \log\left(\frac{\beta_k}{\alpha_k} \right) - \epsilon - \frac{1}{T}. \end{aligned} \tag{3.36}$$

Likewise,

$$D \geq \sum_{k=1}^{4} \alpha_k \Psi_{\mathcal{A}} \left(\frac{\lambda \beta_k}{\alpha_k} \right) - \epsilon.$$

Since ϵ is arbitrary and T can be made arbitrarily large, we obtain the rate-distortion region in the statement of the theorem. □

If we do not put any restriction on $\mathcal{A}$, i.e. $\mathcal{A}$ is the set of non-negative reals, then we get the functional-covering distortion.

Corollary 2 (Functional Covering of Poisson Processes) *The rate-distortion function for the functional-covering distortion is given by* $R_{FC}(D) = (\lambda - \lambda \log(\lambda) - D)^+$.

Proof: For the functional-covering distortion, $\mathcal{A}$ *is the set of non-negative reals. Hence*

$$\Psi_{\mathcal{A}}(u) = \inf_{v \geq 0} v - u \log(v) = u - u \log(u).$$

For any achievable (R, D) *we have*

$$R \geq \lambda \sum_{k=1}^{4} \beta_k \log\left(\frac{\beta_k}{\alpha_k} \right), \tag{3.37}$$

and

$$\begin{aligned} D &\geq \sum_{k=1}^{4} \alpha_k \Psi_{\mathcal{A}}\left(\frac{\lambda\beta_k}{\alpha_k}\right) \\ &= \sum_{k=1}^{4} \alpha_k \left(\frac{\lambda\beta_k}{\alpha_k} - \frac{\lambda\beta_k}{\alpha_k}\log\left(\frac{\lambda\beta_k}{\alpha_k}\right)\right) \\ &= \lambda - \lambda\log(\lambda) - \lambda\sum_{k=1}^{4}\beta_k\log\left(\frac{\beta_k}{\alpha_k}\right). \end{aligned}$$

Hence

$$R + D \geq \lambda - \lambda\log(\lambda),$$

and this is achieved by $[\alpha_k]_{k=1}^4$ *and* $[\beta_k]_{k=1}^4$ *that yield equality in (3.37).* □

If take $\mathcal{A} = \{0, 1\}$, then we recover the covering distortion in [42].

Corollary 3 (Covering Distortion [42]) *The rate-distortion function for the covering distortion is given by* $R_C(D) = (-\lambda\log(D))^+$.

Proof: For the covering distortion, $\mathcal{A} = \{0, 1\}$. *Hence*

$$\Psi_{\mathcal{A}}(u) = \inf_{v\in\{0,1\}} v - u\log(v) = \mathbf{1}\{u > 0\}.$$

Suppose (R, D) *is in* $\mathfrak{RD}$. *Then*

$$\begin{aligned} D &\geq \sum_{k=1}^{4} \alpha_k \Psi_{\mathcal{A}}\left(\frac{\lambda\beta_k}{\alpha_k}\right) \\ &= \sum_{k=1}^{4} \alpha_k \mathbf{1}\{\beta_k > 0\} \\ &= \sum_{k\in\mathcal{B}} \alpha_k, \end{aligned}$$

where we have defined $\mathcal{B} = \{k : \beta_k > 0\}$*. Similarly,*

$$\begin{aligned} R &\geq \lambda \sum_{k=1}^{4} \beta_k \log\left(\frac{\beta_k}{\alpha_k}\right) \\ &= \lambda \sum_{k \in \mathcal{B}} \beta_k \log\left(\frac{\beta_k}{\alpha_k}\right) \\ &\overset{(a)}{\geq} \lambda \left(\sum_{k \in \mathcal{B}} \beta_k\right) \log\left(\frac{\sum_{k \in \mathcal{B}} \beta_k}{\sum_{k \in \mathcal{B}} \alpha_k}\right) \\ &= \lambda \log\left(\frac{1}{\sum_{k \in \mathcal{B}} \alpha_k}\right) \\ &\geq \left(-\lambda \log(D)\right)^{+}, \end{aligned}$$

where, (a) is due to the log-sum inequality, which can be achieved by setting $\alpha_1 = \min(1, D)$*,* $\alpha_2 = 1 - \alpha_1$*,* $\beta_1 = 1$*,* $\beta_2 = 0$*.*

□

Remark 2 *As in for the general case in Theorem 10 (see Remark 1), the reconstruction* $\hat{y}_0^T$ *(assuming it is bounded) can be used to define a probability measure on the input space* $(\mathcal{N}_0^T, \mathfrak{F}^N)$ *via*

$$\frac{dP_{\hat{y}_0^T}}{dP_0}(y_0^T) = \exp\left(\int_0^T \log(\hat{y}_t)\, dy_t - (\hat{y}_t - 1)\, dt\right),$$

where P_0 *is the measure under which* Y_0^T *is a Poisson process with unit rate. Moreover, in absence of feedforward,* $\hat{y}_0^T$ *is deterministic (it depends only on the encoder's output). Thus the input point-process* Y_0^T *is a non-homogeneous Poisson processes with rate* $\hat{y}_0^T$ *under* $P_{\hat{y}_0^T}$*. As in the general case, the functional-covering distortion is related to the above Radon-Nikodym derivative via*

$$d(\hat{y}_0^T, y_0^T) = -\log\left(\frac{dP_{\hat{y}_0^T}}{dP_0}(y_0^T)\right) + T$$

◊

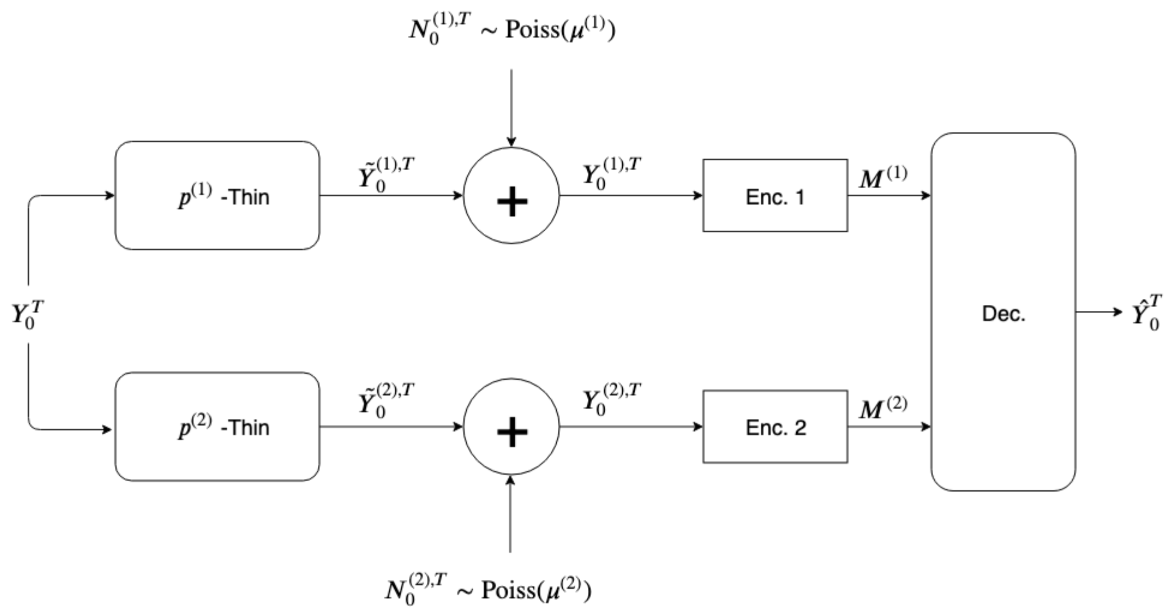

Figure 3.1: Poisson CEO Problem.

3.5 A Poisson CEO Problem

We now consider a distributed lossy compression problem of a Poisson process as shown in Figure 3.1. Our goal is to compress Y_0^T, which is a Poisson process with rate $\lambda > 0$. Each of the two encoders observes a degraded version of Y_0^T, denoted by $Y_0^{(i),T}$, $i \in \{1, 2\}$. Y_0^T is first $p^{(i)}$-thinned to obtain $\tilde{Y}_0^{(i),T}$, and then an independent Poisson process $N_0^{(i),T}$ with rate $\mu^{(i)}$ is added to $\tilde{Y}_0^{(i),T}$ to get $Y_0^{(i),T}$.

Recall that $\hat{\mathcal{Y}}_0^T$ is the set of all non-negative functions $\hat{y}_0^T$ which are left-continuous with right-limits, and

$$d(\hat{y}_0^T, y_0^T) = \int_0^T \hat{y}_t \, dt - \log(\hat{y}_t) \, dy_t.$$

Definition 22 *A* $(T, R^{(1)}, R^{(2)}, D)$ *code for the Poisson CEO problem consists of* encoders $f^{(1)}$ *and* $f^{(2)}$

$$f^{(1)} : \mathcal{N}_0^T \to \{1, \ldots, \lceil \exp(R^{(1)}T) \rceil\}$$
$$f^{(2)} : \mathcal{N}_0^T \to \{1, \ldots, \lceil \exp(R^{(2)}T) \rceil\}$$

and a decoder g

$$g : \{1, \ldots, \lceil \exp(R^{(1)}T) \rceil\} \times \{1, \ldots, \lceil \exp(R^{(2)}T) \rceil\} \to \hat{\mathcal{Y}}_0^T$$

satisfying

$$\mathbb{E}\left[\int_0^T \hat{Y}_t \, dt\right] < \infty$$

and the distortion constraint

$$\frac{1}{T}\mathbb{E}\left[d(\hat{Y}_0^T, Y_0^T)\right] \le D.$$

Definition 23 *A rate-distortion vector* $(R^{(1)}, R^{(2)}, D)$ *is said to be* achievable *for the Poisson CEO problem if for any* $\epsilon > 0$*, there exists a sequence of* T*'s with* $T \to \infty$ *such that* $(T, R^{(1)} + \epsilon, R^{(2)} + \epsilon, D + \epsilon)$ *code exists.*

Definition 24 *The* rate-distortion region *for the Poisson CEO problem* $\mathfrak{RD}^{\mathcal{P}}$ *is the intersection of all achievable rate-distortion vectors* $(R^{(1)}, R^{(2)}, D)$*.*

In presence of feedforward, recall that $\hat{Y}_0^T$ is an allowable reconstruction with feedforward if it is non-negative and $(\sigma(Y_0^t) : t \in [0, T])$-predictable, and that $\hat{\mathcal{Y}}_{0,\text{FF}}^T$ denotes the set of all $\hat{y}_0^T$ processes which are allowable reconstructions with feedforward.

Definition 25 *A* $(T, R^{(1)}, R^{(2)}, D)$ *code with feedforward for the Poisson CEO problem consists of* encoders $f^{(1)}$ *and* $f^{(2)}$

$$f^{(1)} : \mathcal{N}_0^T \to \{1, \ldots, \lceil \exp(R^{(1)}T) \rceil\}$$
$$f^{(2)} : \mathcal{N}_0^T \to \{1, \ldots, \lceil \exp(R^{(2)}T) \rceil\}$$

and a decoder g

$$g : \{1, \ldots, \lceil \exp(R^{(1)}T) \rceil\} \times \{1, \ldots, \lceil \exp(R^{(2)}T) \rceil\} \times \mathcal{N}_0^T \to \hat{\mathcal{Y}}_{0,FF}^T$$

satisfying

$$\mathbb{E}\left[\int_0^T \hat{Y}_t \, dt\right] < \infty$$

and the distortion constraint

$$\frac{1}{T}\mathbb{E}\left[d(\hat{Y}_0^T, Y_0^T)\right] \leq D.$$

The rate-distortion region for the Poisson CEO problem with feedforward, denoted by $\mathfrak{RD}_F^{\mathcal{P}}$, is defined analogously.

Theorem 12 (Poisson CEO problem under functional-covering) *The rate-distortion region for the Poisson CEO problem is given by*

$$\mathfrak{RD}^{\mathcal{P}} = \mathfrak{RD}_F^{\mathcal{P}} = \mathfrak{RD},$$

where $\mathfrak{RD}$ *is the convex hull of union of sets of rate-distortion vectors* $(R^{(1)}, R^{(2)}, D)$ *such that*

$$R^{(1)} \geq \left((1 - p^{(1)})\lambda + \mu^{(1)}\right) \sum_{k=1}^{4} \beta_k^{(1)} \log\left(\frac{\beta_k^{(1)}}{\alpha_k^{(1)}}\right),$$
$$R^{(2)} \geq \left((1 - p^{(2)})\lambda + \mu^{(2)}\right) \sum_{k=1}^{4} \beta_k^{(2)} \log\left(\frac{\beta_k^{(2)}}{\alpha_k^{(2)}}\right),$$
$$D \geq \lambda - \phi(\lambda) - \lambda\left(\sum_{k=1}^{4} \gamma_k^{(1)} \log\left(\frac{\gamma_k^{(1)}}{\alpha_k^{(1)}}\right) + \sum_{k=1}^{4} \gamma_k^{(2)} \log\left(\frac{\gamma_k^{(2)}}{\alpha_k^{(2)}}\right)\right)$$

for some probability vectors $[\alpha_k^{(i)}]_{k=1}^4$, $[\beta_k^{(i)}]_{k=1}^4$, *and* $[\gamma_k^{(i)}]_{k=1}^4$, *where for* $k \in \{1,2,3\}$ *and* $i \in \{1,2\}$

$$\left.\begin{aligned} &\gamma_k^{(i)} = p^{(i)}\alpha_k^{(i)} + (1-p^{(i)})\beta_k^{(i)} \\ &\alpha_k^{(i)} = 0 \Rightarrow \beta_k^{(i)} = 0 \end{aligned}\right\} \quad \text{if } p^{(i)} < 1,$$

$$\alpha_k^{(i)} = \beta_k^{(i)} = \gamma_k^{(i)} \qquad \text{if } p^{(i)} = 1.$$

Proof:

Achievability:

Let

$$R^{(1)} \triangleq \left((1-p^{(1)})\lambda + \mu^{(1)}\right) \sum_{k=1}^{4} \beta_k^{(1)} \log\left(\frac{\beta_k^{(1)}}{\alpha_k^{(1)}}\right)$$
$$R^{(2)} \triangleq \left((1-p^{(2)})\lambda + \mu^{(2)}\right) \sum_{k=1}^{4} \beta_k^{(2)} \log\left(\frac{\beta_k^{(2)}}{\alpha_k^{(2)}}\right)$$
$$D \triangleq \lambda - \phi(\lambda) - \lambda \cdot \left(\sum_{k=1}^{4} \gamma_k^{(1)} \log\left(\frac{\gamma_k^{(1)}}{\alpha_k^{(1)}}\right) + \gamma_k^{(2)} \log\left(\frac{\gamma_k^{(2)}}{\alpha_k^{(2)}}\right)\right).$$

We will show achievability using a $(T, R^{(1)}+\epsilon, R^{(2)}+\epsilon, D+\epsilon)$ code without feed-forward. We will use discretization and results from the rate-distortion theory for discrete memoryless sources (DMS).

First consider the case when for each $i \in \{1,2\}$, at least one of the following conditions is satisfied

C.1 $\beta_k^{(i)} > 0$ for all k,

C.2 $p^{(i)} > 0$.

Fix $\Delta > 0$, and let $T \triangleq n\Delta$ for an integer n. For each $i \in \{1, 2\}$, define a binary valued discrete time process $(\bar{Y}_j^{(i)} : j \in \{1, \ldots, n\})$ as follows. If there are one or more arrivals in the interval $((j-1)\Delta, j\Delta]$ of the process $Y_0^{(i),T}$, then set $\bar{Y}_j^{(i)}$ to 1, otherwise it equals zero. Since $Y_0^{(i),T}$ is a Poisson process with rate $\lambda^{(i)} \triangleq (1 - p^{(i)})\lambda + \mu^{(i)}$, the components of $(\bar{Y}_j^{(i)} : j \in \{1, \ldots, n\})$ are independent and identically distributed with $P(\bar{Y}^{(i)} = 1) = 1 - \exp(-\lambda^{(i)}\Delta)$. Similarly, if $(\bar{Y}_j : j \in \{1, \ldots, n\})$ denotes the discretized process Y_0^T, then we have

$$P\left(\bar{Y}_j^{(i)} : j \in \{1, \ldots, n\} \middle| \bar{Y}_j : j \in \{1, \ldots, n\}\right) = \prod_{j=1}^{n} P(\bar{Y}_j^{(i)} | \bar{Y}_j)$$

due the the memoryless property of Poisson processes and independent thinning. Consider the following "test"-channel for $k \in \{1, 2, 3\}$

$$\begin{aligned} P(\bar{U}^{(i)} = k | \bar{Y}^{(i)} = 1) &= \beta_k^{(i)}, \\ P(\bar{U}^{(i)} = k | \bar{Y}^{(i)} = 0) &= \alpha_k^{(i)}. \end{aligned} \tag{3.38}$$

Define the discretized distortion function

$$\bar{d}(\hat{\bar{y}}, \bar{y}) \triangleq \hat{\bar{y}} - \frac{\log(\hat{\bar{y}})}{\Delta} \mathbf{1}\{\bar{y} = 1\} \quad \hat{\bar{y}} \geq 0, \bar{y} \in \{0, 1\}. \tag{3.39}$$

The reconstruction $\hat{\bar{Y}}$ is taken as

$$\hat{\bar{Y}}(\bar{U}^{(1)}, \bar{U}^{(2)}) = \lambda \hat{\bar{Y}}^{(1)}(\bar{U}^{(1)}) \hat{\bar{Y}}^{(2)}(\bar{U}^{(2)}),$$

where

$$\hat{\bar{Y}}^{(i)}(k) = \begin{cases} \frac{\gamma_k^{(i)}}{\alpha_k^{(i)}} & \text{if } \alpha_k^{(i)} > 0, \\ 1 & \text{otherwise.} \end{cases}$$

We note that since $\gamma_k^{(i)} = p^{(i)}\alpha_k^{(i)} + (1 - p^{(i)})\beta_k^{(i)}$, and at least one of C.1-C.2 is satisfied, $\hat{\bar{Y}}^{(i)}(k) > 0$, and hence $\hat{\bar{Y}} > 0$. Thus the distortion function $\bar{d}(\hat{\bar{Y}}, \bar{Y})$ in (3.39) is bounded. Let

$$\kappa \triangleq \max_{k_1, k_2} \left| \log\left(\hat{\bar{Y}}(k_1, k_2)\right) \right|. \tag{3.40}$$

Due to the Berger-Tung inner bound [26, Theorem 12.1, p. 295], for a given $\Delta > 0$, $\bar{\epsilon} > 0$, and all sufficiently large n, there exists encoders $\bar{f}^{(1)}$ and $\bar{f}^{(2)}$, and a decoder $\bar{g}$ such that for $i \in \{1, 2\}$

$$\bar{f}^{(i)} : (\bar{Y}_j^{(i)} : j \in \{1, \dots, n\}) \to \{1, \dots, L^{(i)}\}$$
$$\bar{g} : \{1, \dots, L^{(1)}\} \times \{1, \dots, L^{(2)}\} \to (\hat{\bar{Y}}_j : j \in \{1, \dots, n\})$$

satisfying

$$\frac{1}{n} \log(L^{(i)}) \leq I(\bar{U}^{(i)}; \bar{Y}^{(i)}) + \bar{\epsilon}, \tag{3.41}$$
$$\mathbb{E}\left[\frac{1}{n} \sum_{j=1}^{n} \bar{d}(\hat{\bar{Y}}_j, \bar{Y}_j)\right] \leq \mathbb{E}[\bar{d}(\hat{\bar{Y}}, \bar{Y})] + \bar{\epsilon}. \tag{3.42}$$

It is noteworthy that the Berger-Tung inner bound has a conditioning term in the mutual-information expression, which in general is a stronger bound than that presented here. However, in our setting we can drop this conditioning as explained in Remark 3 to follow.

Given the above setup, each encoder $f^{(i)}$ upon observing $Y_0^{(i),T}$ obtains the binary valued discrete time process $(\bar{Y}_j^{(i)} : j \in \{1, \dots, n\})$, and sends $M^{(i)} = \bar{f}^{(i)}(\bar{Y}_j^{(i)} : j \in \{1, \dots, n\})$ to the decoder. The decoder outputs the reconstruction $\hat{Y}_0^T$ as

$$\hat{Y}_t \triangleq \sum_{j=1}^{n} \hat{\bar{Y}}_j \mathbf{1}\left\{t \in ((j-1)\Delta, j\Delta]\right\} \quad t \in [0, T].$$

Let $\bar{\bar{Y}}_j$ denote the actual number of arrivals of Y_0^T in an interval $((j-1)\Delta, j\Delta]$. Then $\bar{d}$ is related to the original distortion function via the above reconstruction as follows:

$$\begin{aligned}
\frac{1}{T} d(\hat{Y}_0^T; Y_0^T) &= \frac{1}{T} \int_0^T \hat{Y}_t \, dt - \frac{1}{T} \int_0^T \log(\hat{Y}_t) \, dY_t \\
&= \frac{1}{n} \sum_{j=1}^{n} \hat{\bar{Y}}_j - \frac{1}{T} \sum_{j=1}^{n} \log(\hat{\bar{Y}}_j) \bar{\bar{Y}}_j
\end{aligned}$$

$$= \frac{1}{n}\sum_{j=1}^{n}\hat{\bar{Y}}_j - \frac{1}{n\Delta}\sum_{j=1}^{n}\log(\hat{\bar{Y}}_j)\bar{Y}_j - \frac{1}{T}\sum_{j=1}^{n}\log(\hat{\bar{Y}}_j)(\bar{\bar{Y}}_j - 1)\mathbf{1}\{\bar{\bar{Y}}_j > 1\}.$$

$$= \frac{1}{n}\sum_{j=1}^{n}\bar{d}(\hat{\bar{Y}}_j, \bar{Y}_j) - \frac{1}{T}\sum_{j=1}^{n}\log(\hat{\bar{Y}}_j)(\bar{\bar{Y}}_j - 1)\mathbf{1}\{\bar{\bar{Y}}_j > 1\}$$

$$\overset{(a)}{\leq} \frac{1}{n}\sum_{j=1}^{n}\bar{d}(\hat{\bar{Y}}_j, \bar{Y}_j) + \frac{\kappa}{T}\sum_{j=1}^{n}(\bar{\bar{Y}}_j - 1)\mathbf{1}\{\bar{\bar{Y}}_j > 1\}$$

$$\leq \frac{1}{n}\sum_{j=1}^{n}\bar{d}(\hat{\bar{Y}}_j, \bar{Y}_j) + \frac{\kappa}{T}\sum_{j=1}^{n}\bar{\bar{Y}}_j\mathbf{1}\{\bar{\bar{Y}}_j > 1\},$$

where for (a), we have used the definition of κ in (3.40).

Hence taking the expectation, we get

$$\begin{aligned}
\mathbb{E}\left[\frac{1}{T}d(\hat{Y}_0^T; Y_0^T)\right] &\leq \mathbb{E}\left[\frac{1}{n}\sum_{j=1}^{n}\bar{d}(\hat{\bar{Y}}_j, \bar{Y}_j)\right] + \kappa\mathbb{E}\left[\frac{1}{T}\sum_{j=1}^{n}\bar{\bar{Y}}_j\mathbf{1}\{\bar{\bar{Y}}_j > 1\}\right] \\
&\overset{(a)}{\leq} \mathbb{E}[\bar{d}(\hat{\bar{Y}}, \bar{Y})] + \kappa\mathbb{E}\left[\frac{1}{T}\sum_{j=1}^{n}\bar{\bar{Y}}_j\mathbf{1}\{\bar{\bar{Y}}_j > 1\}\right] + \bar{\epsilon} \\
&\overset{(b)}{=} \mathbb{E}[\bar{d}(\hat{\bar{Y}}, \bar{Y})] + \kappa(\lambda - \lambda\exp(-\lambda\Delta)) + \bar{\epsilon} \\
&\overset{(c)}{\leq} \mathbb{E}[\bar{d}(\hat{\bar{Y}}, \bar{Y})] + \kappa\lambda^2\Delta + \bar{\epsilon}, \qquad (3.43)
\end{aligned}$$

where, for (a), we have used (3.42),

for (b) we note that $\mathbb{E}[\bar{\bar{Y}}_j\mathbf{1}\{\bar{\bar{Y}}_j > 1\}] = \lambda\Delta - \lambda\Delta\exp(-\lambda\Delta)$,

for (c), we have used the inequality $1 - u \leq \exp(-u)$.

Moreover using (3.41), for $i \in \{1, 2\}$

$$\frac{1}{T}\log(L^{(i)}) = \frac{1}{n\Delta}\log(L^{(i)}) \leq \frac{I(\bar{U}^{(i)}; \bar{Y}^{(i)})}{\Delta} + \frac{\bar{\epsilon}}{\Delta}. \qquad (3.44)$$

The scaling of the mutual information $I(\bar{U}^{(i)}; \bar{Y}^{(i)})$ and the distortion function $\bar{d}(\hat{\bar{Y}}, \bar{Y})$ with respect to Δ is given by the following lemma.

Lemma 27 *For* $i \in \{1, 2\}$

$$\lim_{\Delta\to 0}\frac{I(\bar{U}^{(i)}; \bar{Y}^{(i)})}{\Delta} = R^{(i)},$$

$$\lim_{\Delta\to 0}\mathbb{E}[\bar{d}(\hat{\bar{Y}}, \bar{Y})] = D.$$

Proof: Please see the Appendix. □

Now given the rate-distortion vector $(R^{(1)}, R^{(2)}, D)$ and $\epsilon > 0$, first choose Δ sufficiently small so that

$$\frac{I(\bar{U}^{(i)}; \bar{Y}^{(i)})}{\Delta} \leq R^{(i)} + \frac{\epsilon}{4},$$
$$\mathbb{E}[\bar{d}(\hat{\bar{Y}}, \bar{Y})] \leq D + \frac{\epsilon}{4},$$
$$\kappa\lambda^2\Delta \leq \epsilon/4.$$

Then let $\bar{\epsilon} = \Delta\epsilon/4$, and choose a sufficiently large n so that (3.41) and (3.42) are satisfied. From (3.43) and (3.44) we conclude that a sequence of $(T_n, R^{(1)}+\epsilon, R^{(2)}+\epsilon, D+\epsilon)$ code exists with $T_n = n\Delta$ when at least one of the conditions C.1 or C.2 is satisfied.

Now consider the case when $p^{(i)} = 0$ some $i \in \{1, 2\}$, and for that i, $\beta_k^{(i)} = 0$ for some k's. Say $p^{(1)} = 0$ and $p^{(2)} > 0$. This gives us $\gamma_k^{(1)} = \beta_k^{(1)}$ for $k \in \{1, 2, 3\}$. Then we need to show that the rate-distortion vector

$$R^{(1)} = \left(\lambda + \mu^{(1)}\right) \sum_{k=1}^{4} \beta_k^{(1)} \log\left(\frac{\beta_k^{(1)}}{\alpha_k^{(1)}}\right)$$
$$R^{(2)} = \left((1 - p^{(2)})\lambda + \mu^{(2)}\right) \sum_{k=1}^{4} \beta_k^{(2)} \log\left(\frac{\beta_k^{(2)}}{\alpha_k^{(2)}}\right)$$
$$D = \lambda - \phi(\lambda) - \lambda\left(\sum_{k=1}^{4} \beta_k^{(1)} \log\left(\frac{\beta_k^{(1)}}{\alpha_k^{(1)}}\right) + \sum_{k=1}^{4} \gamma_k^{(2)} \log\left(\frac{\gamma_k^{(2)}}{\alpha_k^{(2)}}\right)\right) \tag{3.45}$$

is achievable. Let $[\hat{\beta}_k^{(1)}]_{k=1}^4 = [1/4, 1/4, 1/4]$ and $[\hat{\alpha}_k^{(1)}]_{k=1}^4 = [1/4, 1/4, 1/4-\nu, 1/4+\nu]$ for some $\nu \in [0, 1/3)$. Then the term

$$\sum_{k=1}^{4} \hat{\beta}_k^{(1)} \log\left(\frac{\hat{\beta}_k^{(1)}}{\hat{\alpha}_k^{(1)}}\right)$$

is continuous in ν and goes from zero to unbounded as ν is increased from zero to $1/4$, hence there exists some $\hat{\nu} \in [0, 1/4)$ such that with $[\hat{\alpha}_k^{(1)}]_{k=1}^4 = [1/4, 1/4, 1/4 - \hat{\nu}, 1/4 + \hat{\nu}]$

$$\sum_{k=1}^{4} \hat{\beta}_k^{(1)} \log\left(\frac{\hat{\beta}_k^{(1)}}{\hat{\alpha}_k^{(1)}}\right) = \sum_{k=1}^{4} \beta_k^{(1)} \log\left(\frac{\beta_k^{(1)}}{\alpha_k^{(1)}}\right). \tag{3.46}$$

We note that this $[\hat{\beta}_k^{(1)}]_{k=1}^4$ satisfies condition C.1. Hence the rate-distortion vector in (3.45) is achievable by using $[\hat{\alpha}_k^{(1)}]_{k=1}^4$ that satisfies (3.46). The case when $p^{(2)} = 0$ or both $p^{(1)} = p^{(2)} = 0$ can be handled similarly.

Converse:

We will prove the converse when feedforward is present. For the given $(T, R^{(1)} + \epsilon, R^{(2)} + \epsilon, D + \epsilon)$ code with feedforward, let $M^{(1)}$ and $M^{(2)}$ denote the first and second encoder's output respectively. We essentially repeat the steps in the converse proof of Theorem 10 to show that

$$\frac{1}{T} I(M^{(1)}, M^{(2)}; Y_0^T) + D \geq \lambda - \phi(\lambda) - \epsilon.$$

Since $I(M^{(1)}, M^{(2)}; Y_0^T) < \infty$, we conclude from Theorem 7 that there exists a process Γ_0^T such that Γ_0^T is the $(\mathcal{F}_t = \sigma(M^{(1)}, M^{(2)}, Y_0^t) : t \in [0, T])$ intensity of Y_0^T and

$$I(M^{(1)}, M^{(2)}; Y_0^T) = \mathbb{E}\left[\int_0^T \phi(\Gamma_t)\, dt\right] - T\phi(\lambda), \tag{3.47}$$

Let $\hat{Y}_0^T$ denote the decoder's output. The distortion constraint D satisfies

$$\begin{aligned} D \geq \frac{1}{T}\mathbb{E}\left[d(\hat{Y}_0^T, Y_0^T)\right] - \epsilon &= \frac{1}{T}\mathbb{E}\left[\int_0^T \hat{Y}_t\, dt - \log(\hat{Y}_t)\, dY_t\right] - \epsilon \\ &= \frac{1}{T}\mathbb{E}\left[\int_0^T \hat{Y}_t - \log(\hat{Y}_t)\Gamma_t\, dt\right] - \epsilon, \end{aligned} \tag{3.48}$$

where for the last equality we have used Lemma 25.

Once again using the inequality $u\log(v) \le \phi(u) - u + v, \quad 0 \le u, v < \infty$, and noting that the individual terms have finite expectations,

$$\begin{aligned}\mathbb{E}\left[\int_0^T \log(\hat{Y}_t)\Gamma_t\,dt\right] &\le \mathbb{E}\left[\int_0^T \phi(\Gamma_t) - \Gamma_t + \hat{Y}_t\,dt\right]\\ &= \mathbb{E}\left[\int_0^T \phi(\Gamma_t)\,dt\right] - \mathbb{E}\left[\int_0^T \Gamma_t\,dt\right] + \mathbb{E}\left[\int_0^T \hat{Y}_t\,dt\right]. \quad (3.49)\end{aligned}$$

From (3.47) and (3.48), we deduce

$$\begin{aligned}\frac{1}{T}I(M^{(1)}, M^{(2)}; Y_0^T) + D &\ge \frac{1}{T}\mathbb{E}\left[\int_0^T \phi(\Gamma_t)\,dt\right] - \phi(\lambda)\\ &\quad + \frac{1}{T}\mathbb{E}\left[\int_0^T \hat{Y}_t\,dt\right] - \frac{1}{T}\mathbb{E}\left[\int_0^T \log(\hat{Y}_t)\,dY_t\right] - \epsilon\\ &\overset{(a)}{\ge} \frac{1}{T}\mathbb{E}\left[\int_0^T \Gamma_t\,dt\right] - \phi(\lambda) - \epsilon\\ &\overset{(b)}{=} \lambda - \phi(\lambda) - \epsilon, \quad (3.50)\end{aligned}$$

where, for (a) we have used (3.48) and (3.49),
for (b) we use the fact that $\mathbb{E}\left[\int_0^T \Gamma_t\,dt\right] = \mathbb{E}\left[\int_0^T\,dY_t\right] = \lambda T$.

We can upper bound the term $I(M^{(1)}, M^{(2)}; Y_0^T)$ as

$$\begin{aligned}I(M^{(1)}, M^{(2)}; Y_0^T) &\overset{(a)}{=} H(M^{(1)}, M^{(2)}) - \mathbb{E}\left[H(M^{(1)}, M^{(2)}|Y_0^T)\right]\\ &\overset{(b)}{=} H(M^{(1)}, M^{(2)}) - \mathbb{E}\left[H(M^{(1)}|Y_0^T)\right] - \mathbb{E}\left[H(M^{(2)}|Y_0^T)\right]\\ &\le H(M^{(1)}) + H(M^{(2)}) - \mathbb{E}\left[H(M^{(1)}|Y_0^T)\right] - \mathbb{E}\left[H(M^{(2)}|Y_0^T)\right]\\ &= I(M^{(1)}; Y_0^T) + I(M^{(2)}; Y_0^T), \quad (3.51)\end{aligned}$$

where, for (a) we have used Lemma 15,
for (b), we used the Markov chain $M^{(1)} \leftrightarrows Y_0^T \leftrightarrows M^{(2)}$.
Combining (3.50) and (3.51) we get

$$D \ge \lambda - \phi(\lambda) - \frac{1}{T}I(M^{(1)}; Y_0^T) - \frac{1}{T}I(M^{(2)}; Y_0^T) - \epsilon. \quad (3.52)$$

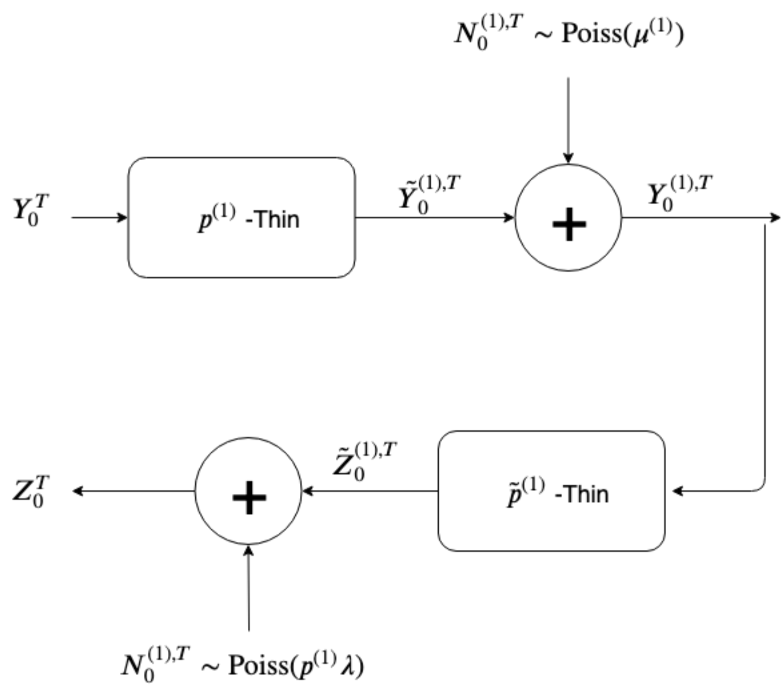

Figure 3.2: Thinning and superposition operations defined in the proof for the first encoder. Note that the joint distribution of $(Y_0^{(1),T}, \tilde{Y}_0^{(1),T}, Y_0^T)$ is same as that of $(Y_0^{(1),T}, \tilde{Z}_0^{(1),T}, Z_0^T)$.

For $i \in \{1,2\}$, using Lemma 15

$$\frac{1}{T} I(M^{(i)}; Y_0^{(i),T}) = \frac{1}{T} H(M^{(i)}) \leq \frac{1}{T} \log\left(\lceil \exp((R^{(i)} + \epsilon)T) \rceil\right) \leq R^{(i)} + \epsilon + \frac{1}{T}. \tag{3.53}$$

We will first consider the case when $p^{(i)} < 1$ for $i \in \{1,2\}$. We shall proceed by defining certain auxiliary processes (see Figure 3.2). Let $\tilde{Z}_0^{(i),T}$ be obtained from $\tilde{p}^{(i)}$-thinning of $Y_0^{(i),T}$, where

$$\tilde{p}^{(i)} = \frac{\mu^{(i)}}{((1 - p^{(i)})\lambda + \mu^{(i)})}.$$

Then using Lemma 24 we can write

$$Y_t^{(i)} = \tilde{Z}_t^{(i)} + \hat{Z}_t^{(i)} \quad t \in [0,T],$$

where $\tilde{Z}_0^{(i),T}$ and $\hat{Z}_0^{(i),T}$ are independent Poisson processes with rates $(1-p^{(i)})\lambda$ and $\mu^{(i)}$ respectively. Whereas, by definition

$$Y_t^{(i)} = \tilde{Y}_t^{(i)} + N_t^{(i)} \quad t \in [0,T],$$

where $\tilde{Y}_t^{(i)}$ and $N_t^{(i)}$ are independent Poisson processes with rates $(1-p^{(i)})\lambda$ and $\mu^{(i)}$ respectively. Hence we conclude that the joint distribution of $(Y_0^{(i),T}, \tilde{Z}_0^{(i),T})$ is identical to the joint distribution of $(Y_0^{(i),T}, \tilde{Y}_0^{(i),T})$. Let $Z_0^{(i),T}$ be obtained by adding an independent Poisson process $\hat{N}_0^{(i),T}$ with rate $p^{(i)}\lambda$ to $\tilde{Z}_0^{(i),T}$,

$$Z_t^{(i)} = \tilde{Z}_t^{(i)} + \hat{N}_t^{(i)} \quad t \in [0,T].$$

Also using Lemma 24 we have

$$Y_t = \tilde{Y}_t^{(i)} + \tilde{\tilde{Y}}_t^{(i)} \quad t \in [0,T],$$

where $\tilde{Y}_0^{(i),T}$ and $\tilde{\tilde{Y}}_0^{(i),T}$ are independent Poisson processes with rates $(1-p^{(i)})\lambda$ and $p^{(i)}\lambda$. Hence the joint distribution of $(Z_0^{(i),T}, \tilde{Z}_0^{(i),T})$ and $(Y_0^T, \tilde{Y}_0^{(i),T})$ are identical. Moreover, $M^{(i)} \rightleftarrows Y_0^{(i),T} \rightleftarrows \tilde{Y}_0^{(i),T} \rightleftarrows Y_0^T$ forms a Markov chain and $M^{(i)} \rightleftarrows Y_0^{(i),T} \rightleftarrows \tilde{Z}_0^{(i),T} \rightleftarrows Z_0^{(i),T}$ forms a Markov chain. This allows us to write

$$\begin{aligned} I\left(M^{(i)}; \tilde{Z}_0^{(i),T}\right) &= I\left(M^{(i)}; \tilde{Y}_0^{(i),T}\right), \\ I\left(M^{(i)}; Z_0^{(i),T}\right) &= I(M^{(i)}; Y_0^T). \end{aligned} \tag{3.54}$$

Since $\tilde{Z}_0^{(i),T}$ is a $\frac{\mu^{(i)}}{((1-p^{(i)})\lambda+\mu^{(i)})}$-thinning of $Y_0^{(i),T}$, Theorem 9 gives

$$I(M^{(i)}; \tilde{Z}_0^{(i),T}) \leq \left(1 - \frac{\mu^{(i)}}{(1-p^{(i)})\lambda + \mu^{(i)}}\right) I(M^{(i)}; Y_0^{(i),T}). \tag{3.55}$$

Also $Z_0^{(i),T}$ is obtained by adding an independent Poisson process with rate $p^{(i)}\lambda$ to $\tilde{Z}_0^{(i),T}$, Theorem 8 yields

$$\begin{aligned} I(M^{(i)}; \tilde{Z}_0^{(i),T}) &= \mathbb{E}\left[\int_0^T \phi(\tilde{\Gamma}_t^{(i)}) - \phi((1-p^{(i)})\lambda)\, dt\right], \\ I(M^{(i)}; Z_0^{(i),T}) &\leq \mathbb{E}\left[\int_0^T \phi(\tilde{\Gamma}_t^{(i)} + p^{(i)}\lambda) - \phi(\lambda)\, dt\right], \end{aligned} \tag{3.56}$$

where, $\tilde{\Gamma}_0^{(i),T}$ is the $(\sigma(M^{(i)}, \tilde{Z}_0^{(i),T}) : t \in [0,T])$-intensity of $\tilde{Z}_0^{(i),T}$. Then we can further lower bound D in (3.52) as

$$\begin{aligned}
D &\geq \lambda - \phi(\lambda) - \frac{1}{T} I(M^{(1)}; Y_0^T) - \frac{1}{T} I(M^{(2)}; Y_0^T) - \epsilon \\
&\overset{(a)}{=} \lambda - \phi(\lambda) - \frac{1}{T} I(M^{(1)}; Z_0^{(1),T}) - \frac{1}{T} I(M^{(2)}; Z_0^{(2),T}) - \epsilon \\
&\overset{(b)}{\geq} \lambda - \phi(\lambda) - \frac{1}{T}\left(\mathbb{E}\left[\int_0^T \phi(\tilde{\Gamma}_t^{(1)} + p^{(1)}\lambda) - \phi(\lambda)\, dt\right]\right) - \frac{1}{T}\left(\mathbb{E}\left[\int_0^T \phi(\tilde{\Gamma}_t^{(2)} + p^{(2)}\lambda) - \phi(\lambda)\, dt\right]\right) \\
&\overset{(c)}{=} \lambda + \phi(\lambda) - \mathbb{E}\left[\phi(\tilde{\Gamma}_{S_1}^{(1)} + p^{(1)}\lambda)\right] - \mathbb{E}\left[\phi(\tilde{\Gamma}_{S_2}^{(2)} + p^{(2)}\lambda)\right] - \epsilon,
\end{aligned}$$

where for (a), we have used (3.54),

for (b), we have used (3.56),

for (c), we define S_1 and S_2 to be uniformly distributed on $[0,T]$, independent of all other random variables and independent of each other as well.

For each $i \in \{1,2\}$, $R^{(i)}$ in (3.53) can be lower bounded as

$$\begin{aligned}
R^{(i)} &\geq \frac{1}{T} I(M^{(i)}; Y_0^{(i),T}) - \epsilon - \frac{1}{T} \\
&\overset{(a)}{\geq} \frac{(1-p^{(i)})\lambda + \mu^{(i)}}{(1-p^{(i)})\lambda} \frac{1}{T} I(M^{(i)}; \tilde{Z}_0^{(i),T}) - \epsilon - \frac{1}{T} \\
&\overset{(b)}{=} \frac{(1-p^{(i)})\lambda + \mu^{(i)}}{(1-p^{(i)})\lambda} \frac{1}{T}\mathbb{E}\left[\int_0^T \phi(\tilde{\Gamma}_t^{(i)}) - \phi((1-p^{(i)})\lambda)\, dt\right] - \epsilon - \frac{1}{T} \\
&\overset{(c)}{=} \frac{(1-p^{(i)})\lambda + \mu^{(i)}}{(1-p^{(i)})\lambda}\mathbb{E}\left[\phi(\tilde{\Gamma}_{S_i}^{(i)}) - \phi((1-p^{(i)})\lambda)\right] - \epsilon - \frac{1}{T},
\end{aligned}$$

where for (a), we have used (3.55),

for (b), we have used (3.56),

for (c), recall that S_1 and S_2 are uniformly distributed on $[0,T]$, independent of all other random variables and independent of each other.

Now we use Carathéodory's theorem [53, Theorem 17.1]. For each $i \in \{1,2\}$, there exist non-negative $[\eta_k^{(i)}]_{k=1}^4$ and $[\alpha_k^{(i)}]_{k=1}^4$, such that $\sum_{k=1}^4 \alpha_k^{(i)} = 1$ and

$$\mathbb{E}\left[\phi(\tilde{\Gamma}_{S_i}^{(i)})\right] = \sum_{k=1}^4 \alpha_k^{(i)} \phi(\eta_k^{(i)}),$$

$$\mathbb{E}\left[\phi(\tilde{\Gamma}^{(i)}_{S_i} + p^{(i)}\lambda)\right] = \sum_{k=1}^{4} \alpha_k^{(i)} \phi(\eta_k^{(i)} + p^{(i)}\lambda),$$
$$\mathbb{E}\left[\tilde{\Gamma}^{(i)}_{S_i}\right] = \sum_{k=1}^{4} \alpha_k^{(i)} \eta_k^{(i)} = (1 - p^{(i)})\lambda,$$

where in the last line we have used the fact that since $\tilde{\Gamma}_0^{(i),T}$ is the $(\sigma(M^{(i)}, \tilde{Z}_0^{(i),T}) : t \in [0, T])$-intensity of $\tilde{Z}_0^{(i),T}$, $\mathbb{E}\left[\int_0^T \tilde{\Gamma}_t^{(i)}\, dt\right] = \mathbb{E}[\tilde{Z}_T^{(i)}] = T(1 - p^{(i)})\lambda$. Hence we have

$$R^{(i)} \geq \frac{(1 - p^{(i)})\lambda + \mu^{(i)}}{(1 - p^{(i)})\lambda} \left(\sum_{k=1}^{4} \alpha_k^{(i)} \phi(\eta_k^{(i)}) - \phi((1 - p^{(i)})\lambda) \right) - \epsilon - \frac{1}{T}, \tag{3.57}$$
$$D \geq \lambda + \phi(\lambda) - \sum_{k=1}^{4} \alpha_k^{(1)} \phi(\eta_k^{(1)} + p^{(1)}\lambda) - \sum_{k=1}^{4} \alpha_k^{(2)} \phi(\eta_k^{(2)} + p^{(2)}\lambda) - \epsilon. \tag{3.58}$$

Now define

$$\beta_k^{(i)} \triangleq \frac{\alpha_k^{(i)} \eta_k^{(i)}}{(1 - p^{(i)})\lambda}, \quad \gamma_k^{(i)} \triangleq p^{(i)} \alpha_k^{(i)} + (1 - p^{(i)}) \beta_k^{(i)}.$$

We note that $\beta_k^{(i)} = 0$ if $\alpha_k^{(i)} = 0$, and $\sum_{k=1}^{4} \beta_k^{(i)} = 1$. Substituting the above definitions in (3.57)

$$\begin{aligned}
R^{(i)} &\geq \frac{(1 - p^{(i)})\lambda + \mu^{(i)}}{(1 - p^{(i)})\lambda} \left(\sum_{k=1}^{4} \alpha_k^{(i)} \eta_k^{(i)} \log(\eta_k^{(i)}) - \phi((1 - p^{(i)})\lambda) \right) - \epsilon - \frac{1}{T} \\
&= ((1 - p^{(i)})\lambda + \mu^{(i)}) \left(\sum_{k=1}^{4} \beta_k^{(i)} \log\left(\frac{\beta_k^{(i)} (1 - p^{(i)})\lambda}{\alpha_k^{(i)}} \right) \mathbf{1}\{\alpha_k^{(i)} > 0\} - \log((1 - p^{(i)})\lambda) \right) - \epsilon - \frac{1}{T} \\
&= ((1 - p^{(i)})\lambda + \mu^{(i)}) \sum_{k=1}^{4} \beta_k^{(i)} \log\left(\frac{\beta_k^{(i)}}{\alpha_k^{(i)}} \right) - \epsilon - \frac{1}{T}.
\end{aligned} \tag{3.59}$$

Likewise,

$$\begin{aligned}
\sum_{k=1}^{4} \alpha_k^{(i)} \phi(\eta_k^{(i)} + p^{(i)}\lambda) &= \sum_{k=1}^{4} \alpha_k^{(i)} \phi\left(\frac{\beta_k^{(i)} (1 - p^{(i)})\lambda}{\alpha_k^{(i)}} + p^{(i)}\lambda \right) \mathbf{1}\{\alpha_k^{(i)} > 0\} \\
&= \sum_{k=1}^{4} \alpha_k^{(i)} \phi\left(\frac{\gamma_k^{(i)}}{\alpha_k^{(i)}} \lambda \right) \mathbf{1}\{\alpha_k^{(i)} > 0\} \\
&= \lambda \sum_{k=1}^{4} \gamma_k^{(i)} \log\left(\frac{\gamma_k^{(i)}}{\alpha_k^{(i)}} \right) + \phi(\lambda).
\end{aligned}$$

Substituting the above in (3.58), we get

$$D \geq \lambda - \phi(\lambda) - \lambda \sum_{k=1}^{4} \gamma_k^{(1)} \log\left(\frac{\gamma_k^{(1)}}{\alpha_k^{(1)}}\right) - \lambda \sum_{k=1}^{4} \gamma_k^{(2)} \log\left(\frac{\gamma_k^{(2)}}{\alpha_k^{(2)}}\right) - \epsilon. \tag{3.60}$$

If either $p^{(i)}$, say $p^{(1)}$, equals 1, then $M^{(1)}$ and Y_0^T are independent so that $I(M^{(1)}; Y_0^T) = 0$, and we can repeat the above steps to show

$$R^{(2)} \geq ((1-p^{(2)})\lambda + \mu^{(2)}) \sum_{k=1}^{4} \beta_k^{(2)} \log\left(\frac{\beta^{(2)}}{\alpha_k^{(2)}}\right) - \epsilon - \frac{1}{T},$$
$$D \geq \lambda - \phi(\lambda) - \lambda \sum_{k=1}^{4} \gamma_k^{(2)} \log\left(\frac{\gamma_k^{(2)}}{\alpha_k^{(2)}}\right) - \epsilon,$$

which is the region in (3.59)-(3.60) with $\alpha_k^{(1)} = \beta_k^{(1)} = \gamma_k^{(1)}$ for $k \in \{1, 2, 3\}$.

Since ϵ is arbitrary, taking $\epsilon \to 0$ and $T \to \infty$ gives us the rate region in the statement of the theorem.

□

Remark 3 *As noted before, there is no sum-rate constraint in the rate-distortion region of the above theorem. This occurs due to the sparsity of points in a Poisson process. After discretizing a Poisson process with rate* λ*, the expected number of ones in the resulting binary process is roughly* λT*, and the rest* $T/\Delta - \lambda T$ *are zeroes. When such sparse binary process is sent via two independent parallel channels as in (3.38), the resulting output processes are almost independent. Specifically, it can be verified that with* $U^{(1)}$ *and* $U^{(2)}$ *as defined in (3.38),* $\lim_{\Delta \to 0} \frac{I(U^{(1)}; U^{(2)})}{\Delta} = 0$*. This implies that the encoders do not need to bin their messages in the achievability argument.*

Corollary 4 (Poisson CEO Problem without Thinning) *If* $p^{(1)} = p^{(2)} = 0$*,* *then the rate-distortion region in Theorem 12 takes a simple form*

$$\frac{\lambda}{\lambda+\mu^{(1)}}R^{(1)}+\frac{\lambda}{\lambda+\mu^{(2)}}R^{(2)}+D\geq\lambda-\phi(\lambda).$$

Corollary 5 (Remote Poisson Source) *Consider a scenario where an encoder wishes to compress a Poisson process with rate $\lambda > 0$, but observes a degraded version of it, where the points are first erased with independent probability $1-p$ and then an independent Poisson process with rate μ is added to it. Then the rate-distortion region (R, D) is the convex hull of the union of all rate-distortion vectors satisfying*

$$R\geq((1-p)\lambda+\mu)\sum_{k=1}^{4}\beta_k\log\left(\frac{\beta_k}{\alpha_k}\right),$$

$$D\geq\lambda-\phi(\lambda)-\lambda\cdot\sum_{k=1}^{4}\gamma_k\log\left(\frac{\gamma_k}{\alpha_k}\right),$$

for some probability vectors $[\alpha_k]_{k=1}^4$, $[\beta_k]_{k=1}^4$, and $[\gamma_k]_{k=1}^4$, where for $k\in\{1,2,3,4\}$

$$\gamma_k=p\alpha_k+(1-p)\beta_k,\quad \alpha_k=0\Rightarrow\beta_k=0.$$

3.6 Proof of Lemmas

Proof of Lemma 16: The first part of the lemma is due to [9, T12 Theorem, Chapter VI, p. 187]. To prove the second part we note that $\mathbb{E}_{P^{Y_0^T}}[\int_0^T|\phi(\Gamma_t)|\,dt]<\infty$ implies $\mathbb{E}_{P^{Y_0^T}}[\int_0^T\Gamma_t\,dt]<\infty$, which in turn gives

$$\int_0^T(1-\sqrt{\Gamma_t})^2\leq\int_0^T(\Gamma_t+1)<\infty,$$

$P^{Y_0^T}$-a.s. Thus applying [43, Theorem 19.7, p. 343], we conclude that $P^{Y_0^T}\ll P_0^{Y_0^T}$. Hence, from the first part of the lemma

$$\frac{dP^{Y_0^T}}{dP_0^{Y_0^T}}=\exp\left(\int_0^T\log(\Lambda_t)\,dY_t-\Lambda_t+1\,dt\right),$$

where the uniqueness of intensity [9, T12 Theorem, Chapter II, p. 31] gives us

$$\mathbb{E}_{P^{Y_0^T}}\left[\int_0^T |\Gamma_t - \Lambda_t|\, dt\right] = 0, \quad \mathbb{E}_{P^{Y_0^T}}\left[\int_0^T \mathbf{1}\{\Gamma_t \neq \Lambda_t\}\, dY_t\right] = 0.$$

Since

$$\mathbb{E}_{P^{Y_0^T}}\left[\int_0^T |\phi(\Gamma_t)|\, dt\right] < \infty,$$

we have

$$\mathbb{E}_{P^{Y_0^T}}\left[\int_0^T (\log(\Gamma_t))^+\, dY_t\right] = \mathbb{E}_{P^{Y_0^T}}\left[\int_0^T (\log(\Gamma_t))^+ \Gamma_t\, dt\right] = \mathbb{E}_{P^{Y_0^T}}\left[\int_0^T (\phi(\Gamma_t))^+\, dt\right] < \infty,$$

and

$$\mathbb{E}_{P^{Y_0^T}}\left[\int_0^T (\log(\Gamma_t))^-\, dY_t\right] = \mathbb{E}_{P^{Y_0^T}}\left[\int_0^T (\log(\Gamma_t))^- \Gamma_t\, dt\right] = \mathbb{E}_{P^{Y_0^T}}\left[\int_0^T (\phi(\Gamma_t))^-\, dt\right] < \infty.$$

Hence

$$\mathbb{E}_{P^{Y_0^T}}\left[\int_0^T \log(\Gamma_t)\, dY_t\right] = \mathbb{E}_{P^{Y_0^T}}\left[\int_0^T \phi(\Gamma_t)\, dt\right] < \infty.$$

Finally,

$$\begin{aligned}
\mathbb{E}_{P^{Y_0^T}}\left[\log\left(\frac{dP^{Y_0^T}}{dP_0^{Y_0^T}}\right)\right] &= \mathbb{E}_{P^{Y_0^T}}\left[\int_0^T \log(\Lambda_t)\, dY_t + \Lambda_t - 1\, dt\right] \\
&\overset{(\mathbf{a})}{=} \mathbb{E}_{P^{Y_0^T}}\left[\int_0^T \log(\Gamma_t)\, dY_t + \Gamma_t - 1\, dt\right] \\
&= \mathbb{E}_{P^{Y_0^T}}\left[\int_0^T \log(\Gamma_t)\, dY_t\right] - \mathbb{E}\left[\int_0^T \Gamma_t - 1\, dt\right] \\
&= \mathbb{E}_{P^{Y_0^T}}\left[\int_0^T \phi(\Gamma_t)\, dt\right] - \mathbb{E}\left[\int_0^T \Gamma_t - 1\, dt\right] \\
&= \mathbb{E}_{P^{Y_0^T}}\left[\int_0^T \phi(\Gamma_t) - \Gamma_t + 1\, dt\right].
\end{aligned}$$

Here, for (a) we have used the uniqueness of the intensity and in the rest of equalities, we have used the finiteness of the expectations $\left[\int_0^T \phi(\Gamma_t)\, dt\right]$, $\mathbb{E}\left[\int_0^T \Gamma_t\, dt\right]$.

□

Proof of Lemma 17: Recall that L_0^T can be written as

$$L_t = \exp\left(\int_0^t \log(\Gamma_s)\, dY_s + (1-\Gamma_s)\, ds\right), \quad t \in [0,T].$$

We note that for $t \in [0,T]$ L_t satisfies

$$L_t = \begin{cases} L_{t-} & \text{if } Y_t - Y_{t-} = 0, \\ \Gamma_t L_{t-} & \text{if } Y_t - Y_{t-} = 1. \end{cases} \tag{3.61}$$

Let C_0^T be a non-negative $(\mathcal{G}_t : t \in [0,T])$-predictable processes. Then

$$\begin{aligned}
\mathbb{E}\left[\int_0^T C_t\, dY_t\right] &\overset{(a)}{=} \mathbb{E}_{\tilde{P}^{M,Y_0^T}}\left[L_T \int_0^T C_t\, dY_t\right] \\
&\overset{(b)}{=} \mathbb{E}_{\tilde{P}^{M,Y_0^T}}\left[\int_0^T L_t C_t\, dY_t\right] \\
&\overset{(c)}{=} \mathbb{E}_{\tilde{P}^{M,Y_0^T}}\left[\int_0^T \Gamma_t L_{t-} C_t\, dY_t\right] \\
&\overset{(d)}{=} \mathbb{E}_{\tilde{P}^{M,Y_0^T}}\left[\int_0^T \Gamma_t L_{t-} C_t\, dt\right] \\
&\overset{(e)}{=} \mathbb{E}_{\tilde{P}^{M,Y_0^T}}\left[\int_0^T \Gamma_t L_t C_t\, dt\right] \\
&\overset{(f)}{=} \mathbb{E}_{\tilde{P}^{M,Y_0^T}}\left[L_T \int_0^T \Gamma_t C_t\, dt\right] \\
&\overset{(g)}{=} \mathbb{E}\left[\int_0^T \Gamma_t C_t\, dt\right],
\end{aligned}$$

where, (a) follows since L_T is the Radon-Nikodym derivative $\frac{dP^{M,Y_0^T}}{d\tilde{P}^{M,Y_0^T}}$,

(b) follows due to [9, T19 Theorem, Appendix A2, p. 302],

(c) follows due to (3.61),

(d) follows since the $(\tilde{P}^{M,Y_0^T}, \mathcal{G}_t : t \in [0,T])$-intensity of Y_0^T is 1, and L_{t-} being a left-continuous adapted process is $(\mathcal{G}_t : t \in [0,T])$-predictable,

(e) follows since the Lebesgue measure of the set $\{t : t \in [0,T], L_{t-} \neq L_t\}$ is zero due to (3.61),

(f) again follows again due to [9, T19 Theorem, Appendix A2, p. 302],

(g) again follows since L_T is the Radon-Nikodym derivative $\frac{dP^{M,Y_0^T}}{d\tilde{P}^{M,Y_0^T}}$. □

Proof of Lemma 18: We will first show that

$$\mathbb{E}\left[\int_0^T (\log(\Gamma_t))^- \, dY_t\right] = \mathbb{E}\left[\int_0^T (\log(\Gamma_t))^- \, \Gamma_t \, dt\right] < \infty.$$

Define $\Gamma_t^{1+} \triangleq \max(\Gamma_t, 1)$ and $\Gamma_t^{1-} \triangleq \min(\Gamma_t, 1)$. We note that $\Gamma_t \le \Gamma_t^{1+} \le \Gamma_t + 1$ and $\Gamma_t = \Gamma_t^{1+}\Gamma_t^{1-}$. Define the process μ_0^T as

$$\mu_t \triangleq \frac{\Gamma_t^{1+}}{\Gamma_t}\mathbf{1}\{\Gamma_t > 0\}, \quad t \in [0, T].$$

Then μ_0^T is a non-negative $(\mathcal{G}_t : t \in [0, T])$-predictable process and

$$\int_0^T \mu_t \Gamma_t \, dt = \int_0^T \Gamma_t^{1+}\mathbf{1}\{\Gamma_t > 0\} \, dt \le \int_0^T (\Gamma_t + 1) \, dt < \infty$$

P-a.s. since $\mathbb{E}[\int_0^T \Gamma_t \, dt] < \infty$. Hence the process $\hat{L}_0^T$ defined as

$$\hat{L}_t \triangleq \exp\left(\int_0^T \log(\mu_t) \, dY_t + (1 - \mu_t)\Gamma_t \, dt\right), \quad t \in [0, T]$$

is a $(P, G_t : t \in [0, T])$ non-negative super-martingale [9, T2 Theorem, Chapter VI, p. 165]. Hence the following chain of inequalities hold

$$\begin{aligned} \mathbb{E}\left[\log(\hat{L}_T)\right] &\overset{(a)}{\le} \log\left(\mathbb{E}[\hat{L}_T]\right) \\ &\overset{(b)}{\le} \log\left(\mathbb{E}[\hat{L}_0]\right) \\ &= 0. \end{aligned} \tag{3.62}$$

Here, for (a) we have used the fact that since L_0^T is a super-martingale, L_T is integrable, and then Jenesn's inequality,

for (b), we have used the fact that $\hat{L}_0^T$ is a super-martingale, hence $\mathbb{E}[\hat{L}_T] \le \mathbb{E}[\hat{L}_0]$.

Let τ_k denote the kth arrival instant of the process Y_0^T, i.e.,

$$\tau_k = \inf\{t \in [0, T] : Y_t = k\},$$

where the infimum of the null set is taken as ∞. Then if $\tau_k \le T$, $\Gamma_{\tau_k} > 0$ P-a.s. [9, T12 Theorem, Chapter II, p. 31]. Hence for $\tau_k \le T$,

$$\log(\mu_{\tau_k}) = \log(\Gamma_{\tau_k}^{1+}) - \log(\Gamma_{\tau_k}) = -\log(\Gamma_{\tau_k}^{1-}) = \left(\log(\Gamma_{\tau_k})\right)^- \quad P-\text{a.s.},$$

Thus we can write

$$\log(\hat{L}_T) = \int_0^T (\log(\Gamma_t))^- \, dY_t + \int_0^T (\Gamma_t - \Gamma_t^{1+}) \mathbf{1}\{\Gamma_t > 0\} \, dt.$$

Using (3.62) we get

$$\mathbb{E}\left[\int_0^T (\log(\Gamma_t))^- \, dY_t + \int_0^T (\Gamma_t - \Gamma_t^{1+}) \mathbf{1}\{\Gamma_t > 0\} \, dt \right] = \mathbb{E}[\log(\hat{L_T})] \leq 0.$$

We note that $\int_0^T (\log(\Gamma_t))^- \, dY_t$ is a non-negative random variable, and

$$\left| \mathbb{E}\left[\int_0^T (\Gamma_t - \Gamma_t^{1+}) \mathbf{1}\{\Gamma_t > 0\} \, dt \right] \right| \leq \mathbb{E}\left[\int_0^T (\Gamma_t + \Gamma_t^{1+}) \, dt \right] \leq \mathbb{E}\left[\int_0^T (2\Gamma_t + 1) \, dt \right] < \infty.$$

Hence we can split the expectation to get

$$\mathbb{E}\left[\int_0^T (\log(\Gamma_t))^- \, dY_t \right] + \mathbb{E}\left[\int_0^T (\Gamma_t - \Gamma_t^{1+}) \mathbf{1}\{\Gamma_t > 0\} \, dt \right] \leq 0,$$

which gives

$$\mathbb{E}\left[\int_0^T (\log(\Gamma_t))^- \, dY_t \right] \leq -\mathbb{E}\left[\int_0^T (\Gamma_t - \Gamma_t^{1+}) \mathbf{1}\{\Gamma_t > 0\} \, dt \right] < \infty. \tag{3.63}$$

Hence

$$\begin{aligned} \mathbb{E}\left[\int_0^T \log(\Gamma_t) \, dY_t \right] &= \mathbb{E}\left[\int_0^T (\log(\Gamma_t))^+ \, dY_t \right] - \mathbb{E}\left[\int_0^T (\log(\Gamma_t))^- \, dY_t \right] \\ &= \mathbb{E}\left[\int_0^T (\log(\Gamma_t))^+ \Gamma_t \, dt \right] - \mathbb{E}\left[\int_0^T (\log(\Gamma_t))^- \, \Gamma_t \, dt \right] \\ &= \mathbb{E}\left[\int_0^T \phi(\Gamma_t) \, dt \right]. \end{aligned} \tag{3.64}$$

□

Proof of Lemma 20: Suppose that Γ_0^T is the $(\mathcal{F}_t : t \in [0,T])$-intensity of Y_0^T. Then applying [9, T8 Theorem, Chapter II, p. 27] with $X_s = 1$ proves M_0^T is a $(\mathcal{F}_t : t \in [0,T])$-martingale. Now suppose that M_0^T is a $(\mathcal{F}_t : t \in [0,T])$-martingale. Consider a simple $(\mathcal{F}_t : t \in [0,T])$-predictable process C_0^T of the form

$$C_t = \mathbf{1}\{\mathcal{E}\}\mathbf{1}\{u < t \leq v \leq T\} \quad \mathcal{E} \in \mathcal{F}_u.$$

Then

$$\begin{aligned}
\mathbb{E}\left[\int_0^T C_s \, d\mathcal{Y}_s\right] &= \mathbb{E}\left[\mathbf{1}\{\mathcal{E}\}(Y_v - Y_u)\right] \\
&= \mathbb{E}\left[\mathbf{1}\{\mathcal{E}\}\mathbb{E}[(Y_v - Y_u)|\mathcal{F}_u]\right] \\
&\overset{(a)}{=} \mathbb{E}\left[\mathbf{1}\{\mathcal{E}\}\mathbb{E}\left[\int_u^v \Gamma_s \, ds \middle| \mathcal{F}_u\right]\right] \\
&= \mathbb{E}\left[\int_0^T C_s \Gamma_s \, ds\right], \qquad (3.65)
\end{aligned}$$

where for (a) we have used the martingale property of M_0^T. Thus for all bounded $(\mathcal{F}_t : t \in [0,T])$-predictable processes C_0^T, (3.65) holds (see the discussion after [9, T6 Theorem, Chapter I, p. 10]). Then by applying the monotone convergence theorem, we can show that (3.65) holds for all non-negative $(\mathcal{F}_t : t \in [0,T])$-predictable processes as well, so that Γ_0^T is the $(\mathcal{F}_t : t \in [0,T])$-intensity of Y_0^T. □

Proof of Lemma 21: There exists a $(\mathcal{G}_t : t \in [0,T])$-predictable process Π_0^T such that P-a.s. $\Pi_t = \mathbb{E}[\Lambda_t|\mathcal{G}_{t-}]$, $t \in [0,T]$ [18, Chapter 6, Theorem 43, p. 103]. We will show that Π_0^T is the $(\mathcal{G}_t : t \in [0,T])$-intensity of N_0^T. Let D_0^T be a non-negative $(\mathcal{G}_t : t \in [0,T])$-predictable process. As $\mathcal{G}_t \subseteq \mathcal{F}_t$, it is also $(\mathcal{F}_t : t \in [0,T])$-predictable. Thus

$$\mathbb{E}\left[\int_0^T D_s \, dN_s\right] = \mathbb{E}\left[\int_0^T D_s \Lambda_s \, ds\right]. \qquad (3.66)$$

Hence

$$\begin{aligned}
\mathbb{E}\left[\int_0^T D_s \Pi_s \, ds\right] &= \mathbb{E}\left[\int_0^T D_s \mathbb{E}[\Lambda_s|\mathcal{G}_{s-}] \, ds\right] \\
&\overset{(a)}{=} \mathbb{E}\left[\int_0^T \mathbb{E}[D_s \Lambda_s|\mathcal{G}_{s-}] \, ds\right] \\
&= \mathbb{E}\left[\int_0^T D_s \Lambda_s \, ds\right] \\
&\overset{(b)}{=} \mathbb{E}\left[\int_0^T D_s \, dN_s\right].
\end{aligned}$$

Here, (a) is due to the fact that D_s is $\mathcal{G}_{s-}$ measurable [9, Exercise E10, Chapter I, p. 9], and
(b) is due to (3.66).

Hence the $(\mathcal{G}_t : t \in [0,T])$-intensity of N_0^T is Π_0^T. □

Proof of Lemma 22: We first note that since Y_0^T and N_0^T are independent, trajectories of Z_0^T are a.s. in $\mathcal{N}_0^T$. The $(\tilde{\mathcal{F}}_t \triangleq \sigma(M, Y_0^t, N_0^t) : t \in [0,T])$-intensities of Y_0^T and N_0^T are Γ_0^T and Π_0^T respectively [9, E5 Exercise, Chapter II, p. 28]. Then for a non-negative $(\tilde{\mathcal{F}}_t : t \in [0,T])$-predictable process C_0^T:

$$\begin{aligned}
\mathbb{E}\left[\int_0^T C_s \, d\mathbb{Z}_s\right] &= \mathbb{E}\left[\int_0^T C_s \, d\mathcal{Y}_s\right] + \mathbb{E}\left[\int_0^T C_s \, dN_s\right] \\
&= \mathbb{E}\left[\int_0^T C_s \Gamma_s \, ds\right] + \mathbb{E}\left[\int_0^T C_s \Pi_s \, ds\right] \\
&= \mathbb{E}\left[\int_0^T C_s(\Gamma_s + \Pi_s) \, ds\right].
\end{aligned}$$

Hence the $(\tilde{\mathcal{F}}_t : t \in [0,T])$-intensity of Z_0^T is $(\Gamma_t + \Pi_t : t \in [0,T])$. Since $\mathcal{F}_t \subseteq \tilde{\mathcal{F}}_t$ the statement of the lemma follows from an application of Lemma 21. □

Proof of Lemma 23: Let $\mathcal{H}_t \triangleq \sigma(M, Y_0^t, Z_0^t)$. We note that the $(\mathcal{H}_t : t \in [0,T])$-intensity of Y_0^T is Γ_0^T. Now we will compute the $(\mathcal{H}_t : t \in [0,T])$-intensity of Z_0^T. Let $(\chi_i : i \in \{1, \dots\})$ denote the sequence of independent and identically distributed Bernoulli random variables which indicate if a particular point in point process $\mathcal{Y}_0^T$ is erased or not. In particular, if $\chi_j = 1$, then the jth point in Y_0^T is retained, so that $\mathbb{E}[\chi_j] = 1 - p$. Then for $0 \le u < v \le T$

$$Z_v - Z_u = \sum_{k=Y_u+1}^{Y_v} \chi_k = \sum_{k=1}^{\infty} \chi_k \mathbf{1}\{Y_u < k \le Y_v\}.$$

Using the monotone convergence theorem for the conditional expectation,

$$\mathbb{E}\left[(Z_v - Z_u)|\mathcal{H}_u\right] = \sum_{k=1}^{\infty} \mathbb{E}[\chi_k \mathbf{1}\{Y_u < k \le Y_v\}|\mathcal{H}_u]$$

$$
\begin{aligned}
&\stackrel{(a)}{=} \sum_{k=1}^{\infty} \mathbb{E}[\chi_k|\mathcal{H}_u]\mathbb{E}[\mathbf{1}\{Y_u < k \le Y_v\}|\mathcal{H}_u] \\
&\stackrel{(b)}{=} (1-p)\mathbb{E}[(Y_v - Y_u)|\mathcal{H}_u] \\
&\stackrel{(c)}{=} (1-p)\mathbb{E}\left[\int_u^v \Gamma_s \, ds \middle| \mathcal{H}_u\right],
\end{aligned}
$$

where, for (a) we have used the fact that given $\mathcal{H}_u$, χ_k is independent of Y_0^T, for (b), we use note that $\mathbb{E}[\chi_k|\mathcal{H}_u] = \chi_k \mathbf{1}\{k \le Y_u\} + (1-p)\mathbf{1}\{k > Y_u\}$, for (c), we have used the martingale property of M_0^T.

Then

$$
\tilde{M}_t \triangleq Z_t - \int_0^t (1-p)\Gamma_s \, ds \quad t \in [0,T].
$$

is a $(\mathcal{H}_t : t \in [0,T])$-martingale. Hence from Lemma 20, the $(\mathcal{H}_t : t \in [0,T])$-intensity of Z_0^T is $((1-p)\Gamma_t : t \in [0,T])$. An application of Lemma 21 proves the statement of the lemma. □

Proof of Lemma 25: We will require the following inequality

$$
u\log(v) \le \phi(u) - u + v, \quad 0 \le u, v < \infty. \tag{3.67}
$$

The inequality can be verified to be true if either or both u, v are zero. If $u, v > 0$, the inequality follows from $\log(u/v) \ge (1 - v/u)$.

Defining $X_t^{1+} \triangleq \max(1, \hat{Y}_t)$, we note that $X_t^{1+} \le \hat{Y}_t + 1$. Consider

$$
\begin{aligned}
\mathbb{E}\left[\int_0^T (\log(\hat{Y}_t))^+ \, dY_t\right] &= \mathbb{E}\left[\int_0^T \log(X_t^{1+}) \, dY_t\right] \\
&\stackrel{(a)}{=} \mathbb{E}\left[\int_0^T \log(X_t^{1+})\Gamma_t \, dt\right] \\
&\stackrel{(b)}{\le} \mathbb{E}\left[\int_0^T \phi(\Gamma_t) - \Gamma_t + X_t^{1+} \, dt\right] \\
&\stackrel{(c)}{=} \mathbb{E}\left[\int_0^T \phi(\Gamma_t) \, dt\right] - \mathbb{E}\left[\int_0^T \Gamma_t \, dt\right] + \mathbb{E}\left[\int_0^T X_t^{1+} \, dt\right] \\
&< \infty,
\end{aligned} \tag{3.68}
$$

where, for (a), we have used the facts that $(X_t^{1+} : t \in [0,T])$ is $(\mathcal{F}_t : t \in [0,T])$-predictable, $\log(X_t^{1+})$ is non-negative, and Γ_0^T is the $(\mathcal{F}_t : t \in [0,T])$-intensity of Y_0^T,

for (b), we note that $\hat{Y}_t^{1+}$ and Γ_t are P-a.s finite, and then use the inequality in (3.67),

for (c), we have used the facts that $\mathbb{E}\left[\int_0^T \phi(\Gamma_t)\, dt\right] < \infty$ (via Theorem 7), $\mathbb{E}\left[\int_0^T \Gamma_t\, dt\right] < \infty$, and $\mathbb{E}\left[\int_0^T X_t^{1+}\, dt\right] \leq \mathbb{E}\left[\int_0^T \hat{Y}_t + 1\, dt\right] < \infty$.

Hence we can write

$$\begin{aligned}\mathbb{E}\left[\int_0^T \log(\hat{Y}_t)\, dY_t\right] &= \mathbb{E}\left[\int_0^T (\log(\hat{Y}_t))^+\, dY_t\right] - \mathbb{E}\left[\int_0^T (\log(\hat{Y}_t))^-\, dY_t\right] \\ &= \mathbb{E}\left[\int_0^T (\log(\hat{Y}_t))^+\Gamma_t\, dt\right] - \mathbb{E}\left[\int_0^T (\log(\hat{Y}_t))^-\Gamma_t\, dt\right] \\ &= \mathbb{E}\left[\int_0^T \log(\hat{Y}_t)\Gamma_t\, dt\right]. \end{aligned} \tag{3.69}$$

□

Proof of Lemma 26: The first part of the lemma is easy to verify using L'Hôpital's rule . For the second part

$$\begin{aligned}\lim_{\Delta\to 0} \mathbb{E}[\bar{d}(\hat{\bar{Y}}, \bar{Y})] &= \lim_{\Delta\to 0} \sum_{k=1}^{4} \hat{\bar{Y}}(k)\exp(-\lambda\Delta)\alpha_k + \left(\hat{\bar{Y}}(k) - \frac{\log(\hat{\bar{Y}}(k))}{\Delta}\right)(1-\exp(-\lambda\Delta)\beta_k \\ &= \sum_{k=1}^{4} \hat{\bar{Y}}(k)\alpha_k - \lambda\log(\hat{\bar{Y}}(k))\beta_k \\ &= \sum_{k=1}^{4} \alpha_k\left(\hat{\bar{Y}}(k) - \frac{\lambda\beta_k}{\alpha_k}\log(\hat{\bar{Y}}(k))\right)\mathbf{1}\{\alpha_k > 0\} \\ &\overset{(a)}{\leq} \sum_{k=1}^{4} \alpha_k\left(\Psi_{\mathcal{A}}\left(\frac{\lambda\beta_k}{\alpha_k}\right) + \frac{\epsilon}{4}\right) \\ &= D + \frac{\epsilon}{4},\end{aligned}$$

where for (a), we have used the definition in (3.24).

□

Proof of Lemma 27: The first limit can be evaluated using L'Hôpital's rule. To compute the second limit, consider

$$\begin{aligned}
\lim_{\Delta\to 0} P(\bar{U}^{(i)} = k|\bar{Y} = 1) &= \lim_{\Delta\to 0} \sum_{l=0}^{1} P(\bar{U}^{(i)} = k, \bar{Y}^{(i)} = l|\bar{Y} = 1) \\
&= \lim_{\Delta\to 0} \sum_{l=0}^{1} P(\bar{Y}^{(i)} = l|\bar{Y} = 1)(\bar{U}^{(i)} = k|\bar{Y}^{(i)} = l) \\
&= p^{(i)}\alpha_k^{(i)} + (1 - p^{(i)})\beta_k^{(i)} \\
&= \gamma_k^{(i)}.
\end{aligned}$$

Then we have

$$\lim_{\Delta\to 0} P(\bar{U}_1^{(1)} = k_1, \bar{U}_1^{(2)} = k_2|\bar{Y} = 1)) = \lim_{\Delta\to 0} P(\bar{U}_1^{(1)} = k_1|\bar{Y} = 1)P(\bar{U}_1^{(2)} = k_2|\bar{Y} = 1) = \gamma_{k_1}^{(1)}\gamma_{k_2}^{(2)}.$$

Recalling that $\alpha_k^{(i)} = 0$ implies $\beta_k^{(i)} = \gamma_k^{(i)} = 0$, we have

$$\begin{aligned}
\lim_{\Delta\to 0} \frac{\mathbb{E}[\log(\hat{\hat{Y}})\mathbf{1}\{\bar{Y} = 1\}]}{\Delta} &= \lim_{\Delta\to 0} \frac{P(\bar{Y} = 1)}{\Delta} \lim_{\Delta\to 0} \mathbb{E}[\log(\hat{\hat{Y}})|\bar{Y} = 1] \\
&= \lambda \sum_{k_1,k_2} \lim_{\Delta\to 0} P(\bar{U} = k_1, \bar{U}_2 = k_2|\bar{Y} = 1)) \log\left(\hat{\hat{Y}}(k_1, k_2)\right) \\
&= \lambda \sum_{k_1,k_2} \gamma_{k_1}^{(1)}\gamma_{k_2}^{(2)} \log\left(\lambda\frac{\gamma_{k_1}^{(1)}\gamma_{k_2}^{(2)}}{\alpha_{k_1}^{(1)}\alpha_{k_2}^{(2)}}\right) \mathbf{1}\{\gamma_{k_1}^{(1)}\gamma_{k_2}^{(2)} > 0\} \\
&= \lambda \sum_{k_1=1}^{4} \gamma_{k_1}^{(1)} \log\left(\frac{\gamma_{k_1}^{(1)}}{\alpha_{k_1}^{(1)}}\right) + \lambda \sum_{k_2=1}^{4} \gamma_{k_2}^{(1)} \log\left(\frac{\gamma_{k_2}^{(2)}}{\alpha_{k_2}^{(2)}}\right) + \phi(\lambda).
\end{aligned}$$

Now to compute $\lim_{\Delta\to 0} \mathbb{E}[\hat{\hat{Y}}]$, we first calculate

$$\begin{aligned}
\lim_{\Delta\to 0} P(\bar{U}_1^{(1)} = k_1, \bar{U}_1^{(2)} = k_2) &= \lim_{\Delta\to 0} P(\bar{U}_1^{(1)} = k_1, \bar{U}_1^{(2)} = k_2|\bar{Y} = 0)P(\bar{Y} = 0) \\
&\quad + \lim_{\Delta\to 0} P(\bar{U}_1^{(1)} = k_1, \bar{U}_1^{(2)} = k_2|\bar{Y} = 1)P(\bar{Y} = 1) \\
&= \lim_{\Delta\to 0} P(\bar{U}_1^{(1)} = k_1, \bar{U}_1^{(2)} = k_2|\bar{Y} = 0) \\
&= \lim_{\Delta\to 0} P(\bar{U}_1^{(1)} = k_1|\bar{Y} = 0)P(\bar{U}_1^{(2)} = k_2|\bar{Y} = 0) \\
&= \alpha_{k_1}^{(1)}\alpha_{k_2}^{(2)}.
\end{aligned}$$

This gives

$$\begin{aligned}\lim_{\Delta\to 0}\mathbb{E}[\hat{\bar{Y}}] &= \sum_{k_1,k_2}\lim_{\Delta\to 0}P(\bar{U}=k_1,\bar{U}_2=k_2)\hat{\bar{Y}}(k_1,k_2)\\ &= \lambda\sum_{k_1,k_2}\alpha^{(1)}_{k_1}\alpha^{(2)}_{k_2}\frac{\gamma^{(1)}_{k_1}\gamma^{(2)}_{k_2}}{\alpha^{(1)}_{k_1}\alpha^{(2)}_{k_2}}\mathbf{1}\{\alpha^{(1)}_{k_1}\alpha^{(2)}_{k_2}>0\}\\ &= \lambda.\end{aligned}$$

Thus

$$\begin{aligned}\lim_{\Delta\to 0}\mathbb{E}[\bar{d}(\hat{\bar{Y}},\bar{Y})] &= \lambda-\phi(\lambda)-\lambda\left(\sum_{k_1=1}^{4}\gamma^{(1)}_{k_1}\log\left(\frac{\gamma^{(1)}_{k_1}}{\alpha^{(1)}_{k_1}}\right)+\sum_{k_2=1}^{4}\gamma^{(2)}_{k_2}\log\left(\frac{\gamma^{(2)}_{k_2}}{\alpha^{(2)}_{k_2}}\right)\right)\\ &= D.\end{aligned}$$

□

CHAPTER 4

(CONVERSE TO) SECOND-ORDER CODING RATE IN DISCRETE MEMORYLESS CHANNELS WITH FEEDBACK

4.1 Preliminaries

Notation

$\mathbb{R}, \mathbb{R}^+$, $\mathbb{R}^-$ and $\mathbb{R}_+$ denote the set of real, positive real, negative real and non-negative real numbers, respectively. $\mathbb{Z}^+$ denotes the set of positive integers. We assume the input alphabet, $\mathcal{X}$, and the output alphabet, $\mathcal{Y}$, are both finite. For a finite set $\mathcal{X}$, $\mathcal{P}(\mathcal{X})$ denotes the set of all probability measures on $\mathcal{X}$. Similarly, for two finite sets $\mathcal{X}$ and $\mathcal{Y}$, $\mathcal{P}(\mathcal{Y}|\mathcal{X})$ denotes the set of all stochastic matrices from $\mathcal{X}$ to $\mathcal{Y}$. Given any $P \in \mathcal{P}(\mathcal{X})$, $\mathcal{S}(P) \triangleq \{x \in \mathcal{X} : P(x) > 0\}$. $\Phi(\cdot)$ and $\phi(\cdot)$ denote the distribution and the density of the standard Gaussian random variable, respectively. $\mathbf{1}\{\cdot\}$ denotes the standard indicator function. For a random variable Z, $\|Z\|_\infty$ denotes its essential supremum (that is, the infimum of those numbers z such that $P(Z \leq z) = 1$. Boldface letters will denote vectors (e.g., $\mathbf{y}^k = [y_1, \ldots, y_k]$) and continuous-time process (e.g., $N = (N_t : t \geq 0)$). We follow the notation of the book of Csiszár-Körner [16] for standard information theoretic quantities. See [32] for standard definitions and notations used in stochastic calculus. Unless otherwise stated, all logarithms and exponentiations are base e.

This work was presented at the IEEE Int. Symposium on Information Theory (ISIT), Vail, June 2018 [60]

Definitions

Given a DMC $W \in \mathcal{P}(\mathcal{Y}|\mathcal{X})$, C denotes the capacity of the channel, and

$$\Pi_W^* \triangleq \{Q \in \mathcal{P}(\mathcal{X}) \colon I(Q; W) = C(W)\} \tag{4.1}$$

denotes the set of capacity-achieving input distributions. There exists a distribution q^* over $\mathcal{Y}$ such that for any $P \in \Pi_W^*$,

$$q^*(y) \triangleq \sum_{x \in \mathcal{X}} P(x) W(y|x). \tag{4.2}$$

Moreover, q^* can be assumed to satisfy $q^*(y) > 0$ for all $y \in \mathcal{Y}$ [25, Corollaries 1 and 2 to Theorem 4.5.1].[1] Define

$$\begin{aligned} \mathfrak{i}^*(X,Y) &\triangleq \log \frac{W(Y|X)}{q^*(Y)}, \\ \nu_x &\triangleq \mathrm{Var}[\mathfrak{i}^*(X,Y)|X = x], \\ V_{\min} &\triangleq \min_{P \in \Pi_W^*} \sum_{x \in \mathcal{X}} P(x) \nu_x, \\ V_{\max} &\triangleq \max_{P \in \Pi_W^*} \sum_{x \in \mathcal{X}} P(x) \nu_x, \\ \nu_{\min} &\triangleq \min_{x \in \mathcal{X}} \nu_x, \\ \nu_{\max} &\triangleq \max_{x \in \mathcal{X}} \nu_x, \\ \mathfrak{i}_{\max} &\triangleq \max_{x \in \mathcal{X}, y \in \mathcal{Y}} |\mathfrak{i}^*(x,y)| \end{aligned}$$

Let $V_{\min}$ and $V_{\max}$ denote V_ε for an arbitrary $\varepsilon \in (0, \frac{1}{2})$ and $\varepsilon \in [\frac{1}{2}, 1)$, respectively, for notational convenience.

Definition 26 *We will call a DMC with*[2] $V_{\min} > 0$ simple-dispersion *if* $V_{\min} = V_{\max}$. *Otherwise, it is called* compound-dispersion.

[1] We can assume without loss of generality that W does not contain an all-zero column.

[2] Note that if $V_{\min} > 0$, then the capacity of the channel is positive.

Remark 4 *The set of compound-dispersion DMCs is not empty. As an example, consider*[3] $p \in (0,1)$ *such that*

$$h(p) + (1-p)\log 2 = h(q), \tag{4.3}$$

for some $q \in (0, 1/2)$*, where* $h(\cdot)$ *denotes the* binary entropy function*, i.e., for any* $r \in [0,1]$*,* $h(r) \triangleq -r\log r - (1-r)\log(1-r)$*. Define* $\mathcal{X} \triangleq \{0,1,2,3,4,5\}$*,* $\mathcal{Y} \triangleq \{0,1,2\}$ *and* $W \in \mathcal{P}(\mathcal{Y}|\mathcal{X})$ *as*

$$W(y|x) \triangleq \begin{bmatrix} p & 0.5(1-p) & 0.5(1-p) \\ 0.5(1-p) & p & 0.5(1-p) \\ 0.5(1-p) & 0.5(1-p) & p \\ q & 1-q & 0 \\ 0 & q & 1-q \\ 1-q & q & 0 \end{bmatrix}. \tag{4.4}$$

One can numerically verify that if $p = 0.8$*, then* $q \approx 0.337$ *satisfies* (4.3) *and the channel defined in* (4.4) *has* $V_{\min} \approx 0.1023$*, which is attained by the uniform input distribution over the set of input symbols* $\{3,4,5\}$*, and* $V_{\max} \approx 0.6919$*, which is attained by the uniform input distribution over the set of input symbols* $\{0,1,2\}$*. Note that for this channel* $\nu_{\min} = V_{\min}$ *and* $\nu_{\max} = V_{\max}$*.* $\diamondsuit$

An (n, R) code with ideal feedback for a DMC consists of an encoder f, which at the kth time instant $(1 \le k \le n)$ chooses an input $x_k = f(m, y_1 \ldots, y_{k-1}) \in \mathcal{X}$, where $m \in \{1, \ldots, \lceil \exp(nR) \rceil\}$ denotes the message to be transmitted, and a decoder g, which maps outputs $(y_1, \ldots, y_n)$ to $\hat{m} \in \{1, \ldots, \lceil \exp(nR) \rceil\}$. Given $\varepsilon \in (0,1)$, define

$$M^*_{\mathrm{fb}}(n, \varepsilon) \triangleq \max\left\{\lceil \exp(nR) \rceil \in \mathbb{R}_+ : \bar{\mathrm{P}}_{\mathrm{e,fb}}(n, R) \le \varepsilon\right\}, \tag{4.5}$$

[3]One can verify that any $p \in [0.8, 1)$ satisfies the following.

where $\bar{\mathrm{P}}_{\mathrm{e}}(n, R)$ denotes the minimum average error probability attainable by any (n, R) code with feedback. Similarly,

$$M^*(n, \varepsilon) \triangleq \max \left\{ \lceil \exp(nR) \rceil \in \mathbb{R}_+ : \bar{\mathrm{P}}_{\mathrm{e,fb}}(n, R) \le \varepsilon \right\}, \tag{4.6}$$

where $\bar{\mathrm{P}}_{\mathrm{e,fb}}(n, R)$ denotes the minimum average error probability attainable by any (n, R) code (without feedback).

Definition 27 *The* second-order coding rate *of a DMC $W \in \mathcal{P}(\mathcal{Y}|\mathcal{X})$ at the average error probability ε is defined as*

$$\liminf_{n\to\infty} \frac{\log M^*(n, \varepsilon) - nC}{\sqrt{n}}. \tag{4.7}$$

The second-order coding rate with feedback is defined analogously.

For a finite set $\mathcal{X}$, $\mathcal{P}(\mathcal{X})$ denotes the set of all probability measures on $\mathcal{X}$. Similarly, for two finite sets $\mathcal{X}$ and $\mathcal{Y}$, $\mathcal{P}(\mathcal{Y}|\mathcal{X})$ denotes the set of all stochastic matrices from $\mathcal{X}$ to $\mathcal{Y}$ $\Phi(\cdot)$ denotes the distribution of the standard Gaussian random variable. $\mathbf{1}\{\cdot\}$ denotes the standard indicator function. Boldface letters such as N will denote continuous-time random process $(N_t : t \ge 0)$.

Channel Model

Let $\mathcal{X}$ be a finite set denoting the set of inputs and $\mathcal{Y}$ be a finite set denoting the set of outputs, then the discrete memoryless channel $W \in \mathcal{P}(\mathcal{Y}|\mathcal{X})$, and let C denote its capacity.

An (n,R) code with complete feedback for this channel consists of an encoder f, which at the kth time instant $(1 \leq k \leq n)$ chooses an input $X_k = f(M, Y_1 \ldots, Y_{k-1})$, where $M \in \{1, \ldots, \lceil \exp(nR) \rceil\}$ denotes the message to be transmitted, and a decoder g, which maps outputs $(Y_1, \ldots, Y_n)$ to $\hat{M} \in \{1, \ldots, \lceil \exp(nR) \rceil\}$. Given any n and $\varepsilon \in (0,1)$,

$$M^*(n,\varepsilon) \triangleq \max\left\{\lceil \exp nR \rceil \in \mathbb{R}_+ \colon \bar{\mathrm{P}}_e(n,R) \leq \varepsilon\right\}, \tag{4.8}$$

where $\bar{\mathrm{P}}_e(n,R)$ denotes the minimum average error probability attainable by any (n,R) code with ideal feedback.

Let $Q^* \in \mathcal{P}(\mathcal{Y})$ denote the distribution on $\mathcal{Y}$ corresponding to any capacity achieving input distribution.

Before we state our results, we recall the following two results. The first result is due to Strassen [61]. For any $W \in \mathcal{P}(\mathcal{Y}|\mathcal{X})$ and $\varepsilon \in (0,1)$,

$$\lim_{n\to\infty} \frac{\log M^*(n,\varepsilon) - nC}{\sqrt{n}} = \sqrt{V_\varepsilon}\Phi^{-1}(\varepsilon). \tag{4.9}$$

That is, the second-order coding rate without feedback is $\sqrt{V_\varepsilon}\Phi^{-1}(\varepsilon)$.

The second result is due to [60, 65]. Consider any $W \in \mathcal{P}(\mathcal{Y}|\mathcal{X})$ with $0 < V_{\min}$ and let $\beta \triangleq \sqrt{\frac{V_{\min}}{V_{\max}}}$.

$$\liminf_{n\to\infty} \frac{\log M^*_{\mathrm{fb}}(n,\varepsilon) - nC}{\sqrt{n}} \geq \begin{cases} \sqrt{V_{\min}}\Phi^{-1}\left(\frac{1}{2\beta}\varepsilon(1+\beta)\right), & \varepsilon \in \left(0, \frac{\beta}{1+\beta}\right], \\ \sqrt{V_{\max}}\Phi^{-1}\left(\frac{1}{2}[\varepsilon(1+\beta) + (1-\beta)]\right), & \varepsilon \in \left(\frac{\beta}{1+\beta}, 1\right). \end{cases} \tag{4.10}$$

4.2 Converse for simple-dispersion channels

As the following theorem shows, feedback does not improve the second-order coding rate for simple-dispersion channels.

Theorem 13 (Impossibility for simple-dispersion channels) *For any* $W \in \mathcal{P}(\mathcal{Y}|\mathcal{X})$ *with* $0 < V_{min} = V_{max}$ *(i.e., simple-dispersion* W*) and* $\varepsilon \in (0,1)$*,*

$$\limsup_{n\to\infty} \frac{\log M^*_{\mathrm{fb}}(n,\varepsilon) - nC}{\sqrt{n}} \le \sqrt{V_{\mathrm{min}}}\Phi^{-1}(\varepsilon).$$

Proof:

The proof of Theorem 13 uses a method of making feedback codes "constant-composition," which is inspired by Fong and Tan's work on parallel Gaussian channels [24]. Fong and Tan have also noted that their techniques can be applied to DMCs to obtain something like Theorem 13 [23].

Definition 28 *The* type *of a sequence* $\mathbf{x}^n$ *is the distribution* $P_{\mathbf{x}^n}$ *on* $\mathcal{X}$ *defined as*

$$P_{\mathbf{x}^n}(a) \triangleq \frac{1}{n}\sum_{k=1}^{n} \mathbf{1}\{x_k = a\}.$$

Definition 29 *For a sequence* $\mathbf{x}^n \in \mathcal{X}^n$*,*

$$\phi_W(\mathbf{x}^n) \triangleq \inf_{P\in\Pi^*_W} d_{TV}(P, P_{\mathbf{x}^n}),$$

where $d_{TV}(P,Q)$ *denotes the total variation distance between distributions* P *and* Q*.*

Definition 30 *Let $\mathcal{T}^n$ denote the set of all probability distributions on $\mathcal{X}$ that are types of some length-n sequence, and define*

$$\mathcal{T}_\gamma^n \triangleq \left\{ T \in \mathcal{T}^n, \inf_{P \in \Pi_W^*} d_{TV}(P,T) > \gamma \right\},$$
$$\mathcal{T}_\gamma^{c,n} \triangleq \left\{ T \in \mathcal{T}^n, \inf_{P \in \Pi_W^*} d_{TV}(P,T) \leq \gamma \right\}.$$

Let $\mathbf{f}(m, \mathbf{y}^i) \triangleq [f(m, \mathbf{y}^0), f(m, \mathbf{y}^1), \ldots, f(m, \mathbf{y}^i)] \in \mathcal{X}^{i+1}$ with the convention that both $\mathbf{y}^0$ and $\mathbf{f}(m, \mathbf{y}^i)$ for $i \leq -1$ are empty strings.

Definition 31 *If Q is a probability distribution on $\mathcal{X}$ and $A \subset \mathcal{X}$ is such that $Q(A) > 0$ then $Q|_A$ is the probability measure*

$$Q_A(x) = \begin{cases} \frac{Q(x)}{Q(A)} & \text{if } x \in A \\ 0 & \text{otherwise.} \end{cases} \tag{4.11}$$

Definition 32 *Given a controller $F : (\mathcal{X} \times \mathcal{Y})^* \mapsto \mathcal{P}(\mathcal{X})$, the $(*, \gamma)$-modified controller $\tilde{F}$ is defined as follows. For $k < n$ and $x^k \in \mathcal{X}^k$, let*

$$\mathcal{X}_{x^k} = \{x : (x^k, x) \text{ is a prefix of some } x^n \in \mathcal{T}_\gamma^{c,n}\}. \tag{4.12}$$

Fix some $x_0 \in \mathcal{X}$ arbitrarily. Let $\tilde{F}(x^k, y^k)$ be a point-mass on x_0 if either $k \geq n$ or $k < n$ but $F(x^k, y^k)(\mathcal{X}_{x^k}) = 0$ (note that the latter includes the case in which $\mathcal{X}_{x^k}$ is empty). Otherwise, let

$$\tilde{F}(x^k, y^k) = F(x^k, y^k)|_{\mathcal{X}_{x^k}}. \tag{4.13}$$

Definition 33 *Given a controller $F : (\mathcal{X} \times \mathcal{Y})^* \mapsto \mathcal{P}(\mathcal{X})$, the (T, γ)-modified controller is defined as in the previous Def but with $\mathcal{T}_\gamma^n$ in place of $\mathcal{T}_\gamma^{c,n}$.*

Lemma 34 in the Appendix states for any $\rho_n > 0$

$$\log M^*(n,\varepsilon) \leq \sup_F \inf_q \left(\log \rho_n - \log \left(1 - \varepsilon - F \circ W \left(\log \frac{\prod_{k=1}^n W(Y_k|X_k)}{q(\mathbf{Y}^n)} \geq \log \rho_n \right) \right) \right), \tag{4.14}$$

where F is a controller: $F : (\mathcal{X} \times \mathcal{Y})^* \to \mathcal{P}(\mathcal{X})$. We shall use (4.52) in Lemma 34 instead of (4.53)-(4.54). Let P denote the distribution $F \circ W$. We will take

$$q(\mathbf{y}^n) = \frac{1}{2} \prod_{k=1}^n q^*(y_k) + \frac{1}{2|\mathcal{T}_\gamma^n|} \sum_{T \in \mathcal{T}_\gamma^n} \prod_{k=1}^n q_T(y_k), \tag{4.15}$$

where

$$q_T(y) \triangleq \sum_{x \in \mathcal{X}} T(x) W(y|x).$$

Let $K_W \triangleq \max\left(2|\mathcal{X}|\nu_{\max}, \frac{8|\mathcal{X}|\nu_{\max}}{V_{\min}}\right)$, and χ_W denote the constant in [8, Corollary to Theorem 2] which depends only upon $\mathbf{i}_{\max}$. Fix $0 < \gamma \leq \frac{V_{\min}}{4|\mathcal{X}|\nu_{\max}}$, and define

$$\begin{aligned}
\delta_n &\triangleq \chi_W \left(\frac{\log n}{\sqrt{n}(V_{\min} - \gamma K_W)^{3/2}} + \sqrt{\gamma K_W} \right), \\
r_n &\triangleq \begin{cases} \sqrt{V_{\min} - \gamma K_W}\,\Phi^{-1}(\varepsilon + 3\delta_n) + \frac{\log 2}{\sqrt{n}} & \varepsilon \in \left(0, \frac{1}{2} - 3\delta_n\right], \\ \sqrt{V_{\min} + \gamma K_W}\,\Phi^{-1}(\varepsilon + 3\delta_n) + \frac{\log 2}{\sqrt{n}} & \varepsilon \in \left(\frac{1}{2} - 3\delta_n, 1\right), \end{cases} \qquad (4.16) \\
\rho_n &\triangleq \exp(nC + \sqrt{n} r_n).
\end{aligned}$$

We now analyze the probability term in (4.14).

$$\begin{aligned}
P\left(\log \frac{\prod_{k=1}^n W(Y_k|X_k)}{q(\mathbf{Y}^n)} \geq \log \rho_n \right) &= P\left(\log \frac{\prod_{k=1}^n W(Y_k|X_k)}{q(\mathbf{Y}^n)} \geq \log \rho_n \bigcap \phi_W(\mathbf{X}^n) \leq \gamma \right) \\
&\quad + P\left(\log \frac{\prod_{k=1}^n W(Y_k|X_k)}{q(\mathbf{Y}^n)} \geq \log \rho_n \bigcap \phi_W(\mathbf{X}^n) > \gamma \right) \\
&= P\left(\log \frac{\prod_{k=1}^n W(Y_k|X_k)}{q(\mathbf{Y}^n)} \geq \log \rho_n \bigcap \phi_W(\mathbf{X}^n) \leq \gamma \right) \\
&\quad + \sum_{T \in \mathcal{T}_\gamma^n} P\left(\log \frac{\prod_{k=1}^n W(Y_k|X_k)}{q(\mathbf{Y}^n)} \geq \log \rho_n \bigcap P_{\mathbf{X}^n} = T \right).
\end{aligned} \tag{4.17}$$

We will now apply the code modification technique of Fong and Tan [24]. Let P_* (respec. P_T) denote the distribution induced by the $(*, \gamma)$-modified (respec. (T, γ)-modified) code.

Lemma 28 *For an event $\mathcal{E} \in \sigma(\mathbf{X}^n, \mathbf{Y}^n)$*

$$P\left(\mathcal{E} \bigcap \phi_W(\mathbf{X}^n) \le \gamma\right) \le P_*(\mathcal{E}),$$
$$P\left(\mathcal{E} \bigcap P_{\mathbf{X}^n} = T\right) \le P_T(\mathcal{E}).$$

Proof: For any (x^n, y^n) such that $\phi_W(x^n) \le \gamma$,

$$P_*((x^n, y^n)) = \prod_{k=1}^{n} \tilde{F}(x_k|x^{k-1}, y^{k-1}) W(y_k|x_k) \tag{4.18}$$
$$= \prod_{k=1}^{n} \frac{F(x_k|x^{k-1}, y^{k-1}) W(y_k|x_k)}{F(\mathcal{X}_{x^{k-1}}|x^{k-1}, y^{k-1})} \tag{4.19}$$
$$\ge \prod_{k=1}^{n} F(x_k|x^{k-1}, y^{k-1}) W(y_k|x_k) \tag{4.20}$$
$$= P(x^n, y^n). \tag{4.21}$$

The proof of the second part is similar. □

Application of the above lemma to (4.17) yields

$$P\left(\log \frac{\prod_{k=1}^{n} W(Y_k|X_k)}{q(\mathbf{Y}^n)} \ge \log \rho_n\right) \le P_*\left(\log \frac{\prod_{k=1}^{n} W(Y_k|X_k)}{q(\mathbf{Y}^n)} \ge \log \rho_n\right) + \sum_{T \in \mathcal{T}_\gamma^n} P_T\left(\log \frac{\prod_{k=1}^{n} W(Y_k|X_k)}{q(\mathbf{Y}^n)} \ge \log \rho_n\right). \tag{4.22}$$

We will now upper bound the first term in the right hand side of the above equation. Let $\mathcal{F}_k = \sigma(M, Y_1, \ldots, Y_k)$, and

$$Z_k \triangleq \mathfrak{i}^*(X_k, Y_k) - \mathbb{E}_*[\mathfrak{i}^*(X_k, Y_k)|\mathcal{F}_{k-1}], \tag{4.23}$$
$$S_k \triangleq \sum_{j=1}^{k} Z_j.$$

$$\begin{aligned} P_*\left(\log \frac{\prod_{k=1}^n W(Y_k|X_k)}{q(\mathbf{Y}^n)} \geq \log \rho_n\right) &\overset{(a)}{\leq} P_*\left(\log \frac{\prod_{k=1}^n W(Y_k|X_k)}{1/2\prod_{k=1}^n q^*(Y_k)} \geq \log \rho_n\right) \\ &= P_*\left(\sum_{k=1}^n \left(\log \frac{W(Y_k|X_k)}{q^*(Y_k)} - C\right) \geq \sqrt{n} r_n - \log 2\right) \\ &\overset{(b)}{\leq} P_*\left(\sum_{k=1}^n (\mathfrak{i}^*(X_k, Y_k) - \mathbb{E}_*[\mathfrak{i}^*(X_k, Y_k)|\mathcal{F}_{k-1}]) \geq \sqrt{n} r_n - \log 2\right) \\ &= P_*\left(\sum_{k=1}^n Z_k \geq \sqrt{n} r_n - \log 2\right), \end{aligned} \tag{4.24}$$

where in (a), we have used the definition of $q(\mathbf{Y}^n)$ in (4.15), and in (b), we have used the fact that $\mathbb{E}_*[\mathfrak{i}^*(X_k, Y_k)|X_k] = \sum_{y \in \mathcal{Y}} W(y|X_k) \log \frac{W(y|X_k)}{Q^*(Y_k)} \leq C$ [25, Theorem 4.5.1].

Lemma 29 *Let $\mathcal{G}_k = \sigma(S_1, \ldots, S_k)$ for $1 \leq k \leq n$, with $\mathcal{G}_0$ being the trivial σ-algebra. Then with $K_W = \max\left(2|\mathcal{X}|\nu_{max}, \frac{8|\mathcal{X}|\nu_{max}}{V_{\min}}\right)$,*

$$V_{\min} - \gamma K_W \leq \frac{1}{n}\sum_{k=1}^n \mathbb{E}_*[Z_k^2|\mathcal{G}_{k-1}] \leq V_{\min} + \gamma K_W, \quad P_*\text{-a.s.}$$
$$\left\| \frac{\sum_{k=1}^n \mathbb{E}_*[Z_k^2|\mathcal{G}_{k-1}]}{\sum_{k=1}^n \mathbb{E}_*[Z_k^2]} - 1 \right\|_\infty \leq \gamma K_W, \quad P_*\text{-a.s.}$$

Proof: The following chain of equalities holds P_*-a.s.,

$$\begin{aligned} \frac{1}{n}\sum_{k=1}^n \mathbb{E}_*[Z_k^2|\mathcal{F}_{k-1}] &= \frac{1}{n}\sum_{k=1}^n \mathbb{E}_*[Z_k^2|X_k] \\ &= \frac{1}{n}\sum_{k=1}^n \mathrm{Var}[\mathfrak{i}(X_k, Y_k)|X_k] \end{aligned}$$

$$= \frac{1}{n}\sum_{k=1}^{n}\sum_{x\in\mathcal{X}} \mathbf{1}\{X_k = x\}\nu_x$$
$$= \sum_{x\in\mathcal{X}} P_{\mathbf{X}^n}(x)\nu_x.$$

Since $\phi_W(\mathbf{X}^n) \le \gamma$, there exists a $\tilde{P} \in \Pi_W^*$ such that $d_{\mathrm{TV}}(\tilde{P}, P_{\mathbf{X}^n}) \le 2\gamma$. Thus we have for each $x \in \mathcal{X}$

$$|\tilde{P}(x) - P_{\mathbf{X}^n}(x)| \le d_{\mathrm{TV}}(\tilde{P}, P_{\mathbf{X}^n}) \le 2\gamma.$$

Thus

$$\begin{aligned}\frac{1}{n}\sum_{k=1}^{n}\mathbb{E}_*[Z_k^2|\mathcal{F}_{k-1}] &= \sum_{x\in\mathcal{X}} P_{\mathbf{X}^n}(x)\nu_x \\ &\le \sum_{x\in\mathcal{X}} \left(\tilde{P}(x) + 2\gamma\right)\nu_x \\ &= \sum_{x\in\mathcal{X}} \tilde{P}(x)\nu_x + 2\gamma\sum_{x\in\mathcal{X}}\nu_x \\ &\le V_{\min} + 2\gamma|\mathcal{X}|\nu_{\max},\end{aligned}$$

where the last step follows since for any $\tilde{P} \in \Pi_W^*$, $\sum_{x\in\mathcal{X}} \tilde{P}(x)\nu_x = V_{\min}$.

Similarly

$$\frac{1}{n}\sum_{k=1}^{n}\mathbb{E}_*[Z_k^2|\mathcal{F}_{k-1}] \ge V_{\min} - 2\gamma|\mathcal{X}|\nu_{\max}.$$

Since $\mathcal{G}_{k-1} \subseteq \mathcal{F}_{k-1}$, taking the conditional expectation with respect to $\mathcal{G}_{k-1}$, we get,

$$V_{\min} - 2\gamma|\mathcal{X}|\nu_{\max} \le \frac{1}{n}\sum_{k=1}^{n}\mathbb{E}_*[Z_k^2|\mathcal{G}_{k-1}] \le V_{\min} + 2\gamma|\mathcal{X}|\nu_{\max}$$

To prove the second part, we note P_*-a.s.,

$$\begin{aligned}\left|\frac{\sum_{k=1}^{n}\mathbb{E}_*[Z_k^2|\mathcal{G}_{k-1}]}{\sum_{k=1}^{n}\mathbb{E}_*[Z_k^2]} - 1\right| &\le \left|\frac{V_{\min} + 2\gamma|\mathcal{X}|\nu_{\max}}{V_{\min} - 2\gamma|\mathcal{X}|\nu_{\max}} - 1\right| \\ &= \frac{4\gamma|\mathcal{X}|\nu_{\max}}{V_{\min} - 2\gamma|\mathcal{X}|\nu_{\max}} \\ &\le \frac{8\gamma|\mathcal{X}|\nu_{\max}}{V_{\min}},\end{aligned}$$

provided $\gamma \leq \frac{V_{\min}}{4|\mathcal{X}|\nu_{\max}}$.

The statement of the lemma now follows by setting $K_W = \max\left(2|\mathcal{X}|\nu_{\max}, \frac{8|\mathcal{X}|\nu_{\max}}{V_{\min}}\right)$.

□

Continuing the chain of expressions in (4.24).

$$
\begin{aligned}
P_*\left(\log \frac{\prod_{k=1}^n W(Y_k|X_k)}{q(\mathbf{Y}^n)} \geq \log \rho_n\right) &= P_*\left(\sum_{k=1}^n Z_k \geq \sqrt{n} r_n - \log 2\right), \\
&\stackrel{(a)}{\leq} P_*\left(\frac{1}{\sqrt{\sum_{k=1}^n \mathbb{E}[Z_k^2]}} \sum_{k=1}^n Z_k \geq \Phi^{-1}(\varepsilon + 3\delta_n)\right) \\
&\stackrel{(b)}{\leq} 1 - \varepsilon - 3\delta_n + \chi_W\left(\frac{n \log n}{(\sum_{k=1}^n \mathbb{E}[Z_k^2])^{3/2}} + \left\| \frac{\sum_{k=1}^n \mathbb{E}[Z_k^2|\mathcal{G}_{k-1}]}{\sum_{k=1}^n \mathbb{E}[Z_k^2]} - \right.\right. \\
&\stackrel{(c)}{\leq} 1 - \varepsilon - 3\delta_n \\
&\qquad + \chi_W\left(\frac{\log n}{\sqrt{n}(V_{\min} - \gamma K_W)^{3/2}} + \sqrt{\gamma K_W}\right) \\
&= 1 - \varepsilon - 2\delta_n, \qquad (4.25)
\end{aligned}
$$

where, for (a) we have used $n(V_{\min} - \gamma K_W) \leq \sum_{k=1}^n \mathbb{E}[Z_k^2] \leq n(V_{\min} + \gamma K_W)$ from Lemma 29,

for (b), we have used the martingale central limit theorem [8, Corollary to Theorem 2], and taking the constant as χ_W (which only depends upon $\mathbf{i}_{\max}$ since $|Z_k| \leq 2\mathbf{i}_{\max}$),

for (c), we have used Lemma 29.

Moving to the second term in (4.22), and noting that $q(\mathbf{Y}^n) \geq \frac{1}{2|\mathcal{T}_\gamma^n|} \prod_{k=1}^n q_T(Y_k)$, we get

$$
\begin{aligned}
\sum_{T \in \mathcal{T}_\gamma^n} P_T\left(\log \frac{\prod_{k=1}^n W(Y_k|X_k)}{q(\mathbf{Y}^n)} \geq \log \rho_n\right) &\leq \sum_{T \in \mathcal{T}_\gamma^n} P_T\left(\log \frac{\prod_{k=1}^n W(Y_k|X_k)}{\frac{1}{2|\mathcal{T}_\gamma^n|} \prod_{k=1}^n q_T(Y_k)} \geq \log \rho_n\right) \\
&= \sum_{T \in \mathcal{T}_\gamma^n} P_T\left(\sum_{k=1}^n \log \frac{W(Y_k|X_k)}{q_T(Y_k)} \geq \log \rho_n - \log 2|\mathcal{T}_\gamma^n|\right).
\end{aligned}
$$

Consider

$$\begin{aligned}
\sum_{k=1}^{n} \mathbb{E}_T \left[\log \frac{W(Y_k|X_k)}{q_T(Y_k)} \middle| \mathcal{F}_{k-1} \right] &= \sum_{x\in\mathcal{X}} \sum_{k=1}^{n} \mathbb{E}_T \left[\log \frac{W(Y_k|X_k)}{q_T(Y_k)} \middle| X_k = x \right] \mathbf{1}\{X_k = x\} \\
&= \sum_{x\in\mathcal{X}} \sum_{k=1}^{n} \sum_{y\in\mathcal{Y}} W(y|x) \log \frac{W(y|x)}{q_T(y)} \mathbf{1}\{X_k = x\} \\
&= n \sum_{x\in\mathcal{X}} T(x) \sum_{y\in\mathcal{Y}} W(y|x) \log \frac{W(y|x)}{q_T(y)} \\
&= nI(T;W).
\end{aligned}$$

Recall that for any $P \in \Pi_W^*$ and $T \in \mathcal{T}_\gamma^n$, $d_{\mathrm{TV}}(P,T) > \gamma > 0$, hence $I(T;W) < C$. Let $K_T \triangleq C - I(T;W) > 0$, and $\tilde{i}_{\max,T} \triangleq \max_{x,y} \left| \log \frac{W(y|x)}{q_T(y)} \right|$. Since

$$\sum_{k=1}^{n} \mathbb{E}_T \left[\log \frac{W(Y_k|X_k)}{q_T(Y_k)} \right] = nI(T;W) < \infty,$$

we conclude that $\tilde{i}_{\max,T} < \infty$ P_T-a.s.

We now show that $\tilde{i}_{\max,T} \leq 2\log n$ P_T-a.s., for all sufficiently large n. Let $W_{\min} \triangleq \min_{x,y:W(y|x)>0} W(y|x)$ and $q_{T,\min} \triangleq \min_{q_T(y)>0} q_T(y)$. Then

$$q_{T,\min} \triangleq \min_{q_T(y)>0} \sum_x T(x)W(y|x) \geq \min_{x,y:W(y|x)>0} W(y|x) \min_{x:T(x)>0} T(x) = \frac{W_{\min}}{n},$$

where the last equality follows since T is a type of a sequence. Thus

$$\begin{aligned}
\tilde{i}_{\max,T} &= \max_{x,y:W(y|x)q_Y(y)>0} \left| \log \frac{W(y|x)}{q_T(y)} \right| \\
&\leq \max_{x,y:W(y|x)q_Y(y)>0} |\log W(y|x)| + \max_{y:q_T(y)>0} |q_T(y)| \\
&\leq |\log W_{\min}| + \left| \log \frac{W_{\min}}{n} \right| \\
&= \log \frac{n}{W_{\min}^2} \\
&\leq 2\log n
\end{aligned}$$

for all sufficiently large n.

Defining $\tilde{Z}_k \triangleq \log \frac{W(Y_k|X_k)}{q_T(Y_k)} - E_T \left[\log \frac{W(Y_k|X_k)}{q_T(Y_k)} \middle| \mathcal{F}_{k-1}\right]$, we have

$$\begin{aligned}
&\sum_{T\in\mathcal{T}_\gamma^n} P_T\left(\sum_{k=1}^n \log \frac{W(Y_k|X_k)}{q_T(Y_k)} \geq \log \rho_n - \log 2|\mathcal{T}_\gamma^n|\right) \\
&= \sum_{T\in\mathcal{T}_\gamma^n} P_T\left(\sum_{k=1}^n \left(\log \frac{W(Y_k|X_k)}{q_T(Y_k)} - E_T\left[\log \frac{W(Y_k|X_k)}{q_T(Y_k)} \middle| F_{k-1}\right]\right) \geq nK_T + \sqrt{n}r_n - \log 2|\mathcal{T}_\gamma^n|\right) \\
&= \sum_{T\in\mathcal{T}_\gamma^n} P_T\left(\sum_{k=1}^n \tilde{Z}_k \geq nK_T + \sqrt{n}r_n - \log 2|\mathcal{T}_\gamma^n|\right) \\
&\overset{(a)}{\leq} \sum_{T\in\mathcal{T}_\gamma^n} P_T\left(\sum_{k=1}^n \tilde{Z}_k \geq nK_T + \sqrt{n}r_n - |\mathcal{X}|\log 2(n+1)\right) \\
&\overset{(b)}{\leq} \sum_{T\in\mathcal{T}_\gamma^n} P_T\left(\sum_{k=1}^n \tilde{Z}_k \geq \frac{nK_T}{2}\right) \\
&\overset{(c)}{\leq} \sum_{T\in\mathcal{T}_\gamma^n} \exp\left(-\frac{nK_T^2}{16\log n}\right) \\
&\overset{(d)}{\leq} \sum_{T\in\mathcal{T}_\gamma^n} \exp\left(-\frac{nK}{\log n}\right) \\
&= |\mathcal{T}_\gamma^n| \exp\left(-\frac{nK}{\log n}\right) \\
&\leq (n+1)^{|\mathcal{X}|} \exp\left(-\frac{nK}{\log n}\right) \\
&\overset{(e)}{\leq} \delta_n, \qquad (4.26)
\end{aligned}$$

where, (a) follows since $|\mathcal{T}_\gamma^n| \leq |\mathcal{T}^n| \leq (n+1)^{|\mathcal{X}|}$,

(b) follows since $\sqrt{n}r_n - |\mathcal{X}|\log 2(n+1) \geq -\frac{nK_T}{2}$ for all sufficiently large n,

(c) follows from Azuma's inequality [5, (3.3), p. 61], and noting that $|\tilde{Z}_k| \leq 2\tilde{i}_{\max,T} \leq 4\log n$,

(d) follows from defining $K \triangleq \min_{T\in\mathcal{T}_\gamma^n} \frac{K_T^2}{16}$,

(e) holds for all sufficiently large n.

From (4.22), (4.25), and (4.26), we get

$$P\left(\log \frac{\prod_{k=1}^n W(Y_k|X_k)}{q(\mathbf{Y}^n)} \geq \log \rho_n\right) \leq 1 - \varepsilon - \delta_n.$$

Plugging the above inequality in (4.14),

$$\log M^*(n,\varepsilon) \le \log \rho_n - \log \delta_n,$$

i.e.,

$$\frac{\log M^*(n,\varepsilon) - nC}{\sqrt{n}} \le r_n - \frac{\log \delta_n}{\sqrt{n}}. \tag{4.27}$$

Using the definition of r_n in (4.16) and taking the limit

$$\limsup_{n\to\infty} \frac{\log M^*(n,\varepsilon) - nC}{\sqrt{n}} \le \begin{cases} \sqrt{V_{\min} - \gamma K_W}\,\Phi^{-1}\left(\varepsilon + \chi_W \sqrt{\gamma K_W}\right) & \varepsilon \in \left(0, \frac{1}{2} - \chi_W\sqrt{\gamma K_W}\right], \\ \sqrt{V_{\min} + \gamma K_W}\,\Phi^{-1}\left(\varepsilon + \chi_W \sqrt{\gamma K_W}\right) & \varepsilon \in \left(\frac{1}{2} - \chi_W\sqrt{\gamma K_W}, 1\right). \end{cases}$$

Now taking $\gamma \to 0$ gives

$$\limsup_{n\to\infty} \frac{\log M^*(n,\varepsilon) - nC}{\sqrt{n}} \le \sqrt{V_{\min}}\,\Phi^{-1}(\varepsilon),$$

proving the theorem. □ □

4.3 Converse for compound-dispersion channels

If the channel is compound dispersion, then feedback improves the second-order coding rate, and 4.10 provides a lower bound on the size of the improvement. The next theorem provides a comparable upper bound.

Theorem 14 (Impossibility for compound-dispersion channels) *Consider any $W \in \mathcal{P}(\mathcal{Y}|\mathcal{X})$ with $0 < \nu_{min}$ and let $\lambda \triangleq \sqrt{\frac{\nu_{min}}{\nu_{max}}}$. Then*

$$\limsup_{n\to\infty} \frac{\log M^*_{\mathrm{fb}}(n,\varepsilon) - nC}{\sqrt{n}} \le \begin{cases} \sqrt{\nu_{min}}\,\Phi^{-1}\left(\frac{1}{2\lambda}\varepsilon(1+\lambda)\right), & \varepsilon \in \left(0, \frac{\lambda}{1+\lambda}\right], \\ \sqrt{\nu_{max}}\,\Phi^{-1}\left(\frac{1}{2}[\varepsilon(1+\lambda) + (1-\lambda)]\right), & \varepsilon \in \left(\frac{\lambda}{1+\lambda}, 1\right). \end{cases} \tag{4.28}$$

Proof:

We begin with a few definitions from stochastic calculus. Throughout we assume that the filtration under consideration is right-continuous and complete (via e.g. [31, Lemma 7.8, p. 124]).

Definition 34 *A process N is called a* local martingale *with respect to a filtration $(\mathcal{F}_t : t \geq 0)$ if N_t is $\mathcal{F}_t$-measurable for each t and there exists an increasing sequence of stopping times T_n, such that $T_n \to \infty$ and the stopped and shifted processes $N^{T_n} \triangleq (N_{\min\{t,T_n\}} - N_0 : t \geq 0)$ are $(\mathcal{F}_t : t \geq 0)$-martingales for each n.*

Definition 35 *The* quadratic variation *of a continuous local martingale N is a unique process $[N]$ such that $N \cdot N - [N]$ is a local martingale. The existence and uniqueness of such process is guaranteed by [31, Theorem 17.5, p. 332].*

Definition 36 *A stochastic process is said to be $\mathcal{F}_t$-*predictable *if it is measurable with respect to the σ-algebra generated by all left-continuous $\mathcal{F}_t$-adapted processes.*

By taking $q(\mathbf{y}^n) = \prod_{i=1}^n q^*(y_i)$ in (4.52) in Lemma 34 in the Appendix, we get

$$\log M^*_{\text{fb}}(n, \varepsilon) \leq \sup_F \left(\log \rho_n - \log \left(1 - \varepsilon - P \left(\sum_{k=1}^n \mathfrak{i}^*(X_k, Y_k) \geq \log \rho_n \right) \right)^+ \right), \tag{4.29}$$

where F is some controller: $F : (\mathcal{X} \times \mathcal{Y})^* \to \mathcal{P}(\mathcal{X})$, and P denotes the distribution $F \circ W$. As in the previous section, we use (4.52) over (4.53)-(4.54) in Lemma 34 because it yields a finite-n result ((4.51) to follow). Fix an arbitrary $\kappa > 0$, let

$K_W \triangleq 16\mathfrak{i}_{\max}^2 \nu_{\max}/\nu_{\min}$, and define

$$\delta_n \triangleq \frac{K_W}{\kappa^2 \sqrt{n}}, \tag{4.30}$$

$$r_n \triangleq \begin{cases} \sqrt{\nu_{\min}}\Phi^{-1}\left(\frac{(1+\lambda)}{2\lambda}\left(\varepsilon + 2\delta_n\right)\right) + \kappa, & 0 < \varepsilon \leq \frac{\lambda}{1+\lambda} - 2\delta_n \\ \sqrt{\nu_{\max}}\Phi^{-1}\left(\frac{(\varepsilon+2\delta_n)(1+\lambda)+(1-\lambda)}{2}\right) + \kappa, & \frac{\lambda}{1+\lambda} - 2\delta_n < \varepsilon < 1. \end{cases} \tag{4.31}$$

$$\rho_n \triangleq \exp(nC + \sqrt{n} r_n). \tag{4.32}$$

The proof will consist of the following steps:

1. We will define a martingale sequence $(S_k, 1 \leq k \leq n)$ such that $P\left(\sum_{k=1}^n \mathfrak{i}^*(X_k, Y_k) \geq \log \rho_n\right) \leq P(S_n \geq r_n)$.
2. We will embed the martingale sequence $(S_k, 1 \leq k \leq n)$ in a Brownian motion $\mathbf{B}$ such that $S_k = B_{T_k}, 1 \leq k \leq n$, where $(T_k, 1 \leq k \leq n)$ are stopping times.
3. We will construct a process $\psi_t \in [\sqrt{\nu_{\min}}, \sqrt{\nu_{\max}}]$ and a Brownian motion $\mathbf{W}$ such that $\int_0^1 \psi_s \, dW_s \approx B_{T_n}$.
4. Applying a theorem from stochastic calculus, we will "mimic" the above Itô process by a solution of a SDE $\hat{\boldsymbol{\xi}}$.
5. Using a result on the optimal control of diffusion processes, we will upper bound the probability $P\left(\hat{\xi}_1 \geq 0\right)$ which will yield an upper bound on $P\left(\int_0^1 \psi_s \, dW_s \geq r_n\right)$.

Define

$$\begin{aligned} \mathcal{F}_k &\triangleq \sigma(M, Y_1, \ldots, Y_k), \\ Z_k &\triangleq \frac{1}{\sqrt{n}}\left(\mathfrak{i}^*(X_k, Y_k) - \mathbb{E}[\mathfrak{i}^*(X_k, Y_k) | \mathcal{F}_{k-1}]\right) \\ S_k &\triangleq \sum_{j=1}^k Z_j, \end{aligned}$$

$$G_k \triangleq \sigma(S_1, \dots, S_k)$$

We note that

$$|Z_k| \leq \frac{2}{\sqrt{n}} \mathbf{i}_{\max}. \tag{4.33}$$

Lemma 30 *The sequence $(S_k, 1 \leq k \leq n)$ is a martingale with respect to the filtration $(\mathcal{G}_k, 1 \leq k \leq n)$ such that*

$$\mathbb{E}[Z_k^2 | \mathcal{F}_{k-1}] \in \left[\frac{\nu_{min}}{n}, \frac{\nu_{max}}{n}\right],$$

and

$$P\left(\sum_{k=1}^{n} \mathbf{i}^*(X_k, Y_k) \geq \log \rho_n\right) \leq P(S_n \geq r_n).$$

Proof:

Since $\mathcal{G}_k \subseteq \mathcal{F}_k$ and

$$\mathbb{E}[Z_k | \mathcal{F}_{k-1}] = 0,$$

taking the conditional expectation with respect to $\mathcal{G}_{k-1}$, we get

$$\mathbb{E}[Z_k | \mathcal{G}_{k-1}] = 0.$$

Thus the sequence $(S_k, 1 \leq k \leq n)$ is a martingale with respect to the filtration $(\mathcal{G}_k, 1 \leq k \leq n)$. Moreover

$$\mathbb{E}[Z_k^2 | \mathcal{F}_{k-1}] = \frac{1}{n} \sum_{x \in \mathcal{X}} \mathbf{1}\{X_k = x\} \nu_x \in \left[\frac{\nu_{\min}}{n}, \frac{\nu_{\max}}{n}\right]. \tag{4.34}$$

Once again taking the conditional expectation with respect to $\mathcal{G}_{k-1}$, we get

$$\mathbb{E}[Z_k^2 | \mathcal{G}_{k-1}] \in \left[\frac{\nu_{\min}}{n}, \frac{\nu_{\max}}{n}\right]. \tag{4.35}$$

Now consider

$$P\left(\sum_{k=1}^{n}\mathfrak{i}^*(X_k,Y_k)\geq\log\rho_n\right)=P\left(\frac{1}{\sqrt{n}}\sum_{k=1}^{n}(\mathfrak{i}^*(X_k,Y_k)-C)\geq r_n\right)$$
$$\leq P\left(\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\left(\mathfrak{i}^*(X_k,Y_k)-\mathbb{E}[\mathfrak{i}^*(X_k,Y_k)|\mathcal{F}_{k-1}]\right)\geq r_n\right)$$
$$=P(S_n\geq r_n), \tag{4.36}$$

where in the middle step we have used the fact that [25, Theorem 4.5.1]

$$\mathbb{E}[\mathfrak{i}^*(X_k,Y_k)|\mathcal{F}_{k-1}]=\mathbb{E}[\mathfrak{i}^*(X_k,Y_k)|X_k]=\sum_{y\in\mathcal{Y}}W(y|X_k)\log\frac{W(y|X_k)}{Q^*(Y_k)}\leq C.$$

□

Lemma 31 *There exists a Brownian motion* $\mathbf{B}$, *and a sequence of non-decreasing stopping times* $T_1,\ldots,T_n$ *such that*

$$S_k=B_{T_k}\ \ a.s.\qquad k\in\{1,\ldots,n\},$$

and if $\tilde{\mathcal{G}}_k=\sigma(S_1,T_1\ldots,S_k,T_k)$, *and* $\tau_k=T_k-T_{k-1}$, *then*

$$E[\tau_k|\tilde{\mathcal{G}}_{k-1}]=\mathbb{E}[Z_k^2|\mathcal{G}_{k-1}], \tag{4.37}$$
$$E[\tau_k^2|\tilde{\mathcal{G}}_{k-1}]\leq 4\mathbb{E}[Z_k^4|\mathcal{G}_{k-1}]. \tag{4.38}$$

Proof: The lemma is a straightforward application of [31, Theorem 14.16, p. 279] to the martingale sequence $(S_k,1\leq k\leq n)$. □

Lemma 32 *There exists a filtration* $\mathcal{H}_t$, *an* $\mathcal{H}_t$*-predictable process* $\boldsymbol{\psi}$, *an* $\mathcal{H}_t$ *Brownian motion* $\mathbf{W}$, *and an* $\mathcal{H}_t$*-stopping time* T_n^* *such that*

1. $\sqrt{\nu}_{min}\leq\psi_t\leq\sqrt{\nu}_{max}$ *a.s.*

2. $\int_0^{T_n^*} \psi_t \, dW_t = B_{T_n} = S_n$.

3. $\mathbb{E}[(T_n^* - 1)^2] \le \frac{K_W^{(1)}}{n}$, *where* $K_W^{(1)} \triangleq 64\mathfrak{i}_{max}^4/\nu_{min}$.

Proof: Define $\boldsymbol{\psi}$ as

$$\psi_t = \begin{cases} \sqrt{nE[\tau_1|\tilde{\mathcal{G}}_0]} & 0 \le t \le \frac{\tau_1}{n\mathbb{E}[\tau_1|\tilde{\mathcal{G}}_0]} \\ \sqrt{nE[\tau_2|\tilde{\mathcal{G}}_1]} & \frac{\tau_1}{n\mathbb{E}[\tau_1|\tilde{\mathcal{G}}_0]} < t \le \frac{\tau_1}{n\mathbb{E}[\tau_1|\tilde{\mathcal{G}}_0]} + \frac{\tau_2}{n\mathbb{E}[\tau_2|\tilde{\mathcal{G}}_1]} \\ \vdots & \vdots \\ \sqrt{nE[\tau_n|\tilde{\mathcal{G}}_{n-1}]} & \sum_{j=1}^{n-1} \frac{\tau_j}{n\mathbb{E}[\tau_j|\tilde{\mathcal{G}}_{j-1}]} < t \le \sum_{j=1}^{n} \frac{\tau_j}{n\mathbb{E}[\tau_j|\tilde{\mathcal{G}}_{j-1}]} \\ \sqrt{\nu}_{\min} & t > \sum_{j=1}^{n} \frac{\tau_j}{n\mathbb{E}[\tau_j|\tilde{\mathcal{G}}_{j-1}]} \end{cases} \tag{4.39}$$

Then, from the above definition, (4.37), and (4.35), it is clear that $\sqrt{\nu}_{\min} \le \psi_t \le \sqrt{\nu}_{\max}$ a.s.

We now employ the change-of-time method (see [3]). To illustrate the reason behind it, consider the stochastic integral

$$\tilde{\xi}_t = \int_0^t \tilde{\psi}_s \, d\tilde{W}_s,$$

with $\tilde{\mathbf{W}}$ being a Brownian Motion and $\tilde{\psi}_s \in [\nu_{\min}, \nu_{\max}]$. Let $\tilde{A}_t \triangleq [\tilde{\xi}]_t = \int_0^t \psi_s^2 \, ds$ [31, Theorem 17.24, p. 344]. Moreover, $\tilde{\xi}_t = \tilde{B}_{\tilde{A}_t}$ for some Brownian motion $\tilde{\mathbf{B}}$ [31, Theorem 18.4, p. 352]. Let $\tilde{T} \triangleq \tilde{A}_1$, then

$$B_{\tilde{T}} = B_{\tilde{A}_1} = \tilde{\xi}_1 = \int_0^1 \tilde{\psi}_s \, d\tilde{W}_s.$$

Hence, if by choosing $\tilde{\xi}$ properly, we could ensure that $\tilde{T} = T_n$ and $\tilde{\mathbf{B}} = \mathbf{B}$, then we would have proven a stronger version of the lemma (with $T_n^* = 1$). However, proving this stronger result appears to be difficult, and hence we allow T_n^* to be random. We continue with the proof of the lemma.

Let $A_t \triangleq \int_0^t \psi_s^2\, ds$, we note that $\mathcal{A}$ is continuous and strictly increasing, and we define the following time changed process $\mathbf{N} \triangleq \mathbf{B} \circ \mathbf{A}$, i.e.,

$$N_t = B_{A_t} = B_{\int_0^t \psi_s^2\, ds},$$

and

$$\mathcal{H}_t \triangleq \sigma(B_{A_s}, 0 \leq s \leq t).$$

Let

$$T_k^* = \sum_{j=1}^{k} \frac{\tau_j}{n\mathbb{E}[\tau_j|\tilde{\mathcal{G}}_{j-1}]}, \quad 1 \leq k \leq n,$$

then it follows that (see Figure 4.1)

$$A_{T_k^*} = \int_0^{T_k^*} \psi_t^2\, dt = \sum_{j=1}^{k} \tau_j = T_k, \quad 1 \leq k \leq n.$$

Hence, $T_n^* = A_{T_n}^{-1}$, where $A_t^{-1}(\omega)$ is the inverse of $A_t(\omega)$ for each ω in the given sample space. We can write

$$T_n = \inf\{t > 0; A_t^{-1} > T_n^*\}$$

Noting that A_t^{-1} is continuous and T_n is a $\sigma(B_s, 0 \leq s \leq t)$-stopping time, applying [31, Proposition 7.9, p. 124], we conclude that $A_{T_n}^{-1} = T_n^*$ is an $\mathcal{H}_t$-stopping time (The role of process X_t in [31, Proposition 7.9, p. 124] is played by A_t^{-1} here).

Now applying [31, Theorem 17.24, p. 344] we get that N is a continuous local martingale with respect to the filtration $\mathcal{H}_t$ with quadratic variation

$$[N] = [\mathbf{B}] \circ \mathbf{A} = \mathbf{A}, \tag{4.40}$$

since $[B]_t = t$ [31, Theorem 18.3, p. 352]. Now we follow the proof of [32, Theorem 4.2, p. 170]. Define $\mathbf{W}$ as

$$W_t = \int_0^t \frac{1}{\psi_s}\, dN_s.$$

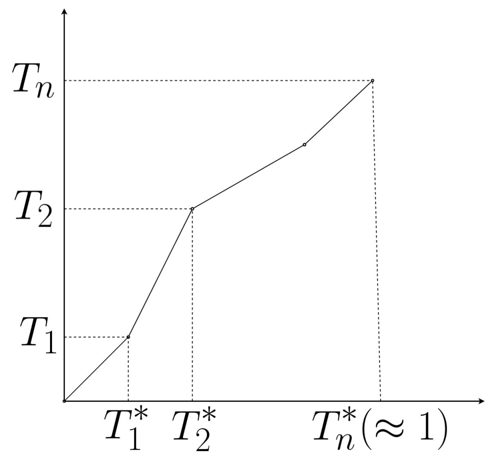

Figure 4.1: Plot of A_t vs t for a fixed ω in the sample space.

Then $\mathbf{W}$ is a continuous local martingale with quadratic variation [31, Lemma 17.10, p.335]

$$[W]_t = \int_0^t \frac{1}{\psi_s^2}\, d[N]_s = \int_0^t \frac{1}{\psi_s^2} \psi_s^2\, ds = t,$$

where we have used [31, Proposition 17.14, p. 338] for the middle equality. Hence $\mathbf{W}$ is a standard Brownian motion with respect to the filtration $\mathcal{H}_t$ [31, Theorem 18.3, p. 352].

Noting that there exists a partition $0 = t_0 < t_1, \ldots, < t_l = t$ such that $\boldsymbol{\psi}$ is constant on $(t_k, t_{k+1}]$ for $0 \leq k \leq l-1$, we can write

$$\int_0^t \psi_s\, dW_s = \sum_{k=0}^{l-1} \psi_{t_k}(W_{t_{k+1}} - W_{t_k}) = \sum_{k=0}^{l-1} \psi_{t_k} \frac{1}{\psi_{t_k}} (N_{t_{k+1}} - N_{t_k}) = N_t.$$

Thus

$$\int_0^{T_n^*} \psi_s\, dW_s = N_{T_n^*} = B_{A_{T_n^*}} = B_{T_n} = S_n. \tag{4.41}$$

Moreover by construction $\boldsymbol{\psi}$ is left-continuous and hence predictable.

Now we bound $\mathbb{E}[(T_n^* - 1)^2]$:

$$\begin{aligned}
\mathbb{E}[(T_n^* - 1)^2] &= \mathbb{E}\left[\left(\sum_{j=1}^{n} \frac{\tau_j}{n\mathbb{E}[\tau_j|\tilde{\mathcal{G}}_{j-1}]} - 1\right)^2\right] \\
&= \mathbb{E}\left[\left(\sum_{j=1}^{n} \frac{\tau_j - \mathbb{E}[\tau_j|\tilde{\mathcal{G}}_{j-1}]}{n\mathbb{E}[\tau_j|\tilde{\mathcal{G}}_{j-1}]}\right)^2\right] \\
&\overset{(a)}{\leq} \frac{1}{\nu_{\min}^2}\mathbb{E}\left[\left(\sum_{j=1}^{n} \tau_j - \mathbb{E}[\tau_j|\tilde{\mathcal{G}}_{j-1}]\right)^2\right] \\
&\overset{(b)}{=} \frac{1}{\nu_{\min}^2}\mathbb{E}\left[\sum_{j=1}^{n} \left(\tau_j - \mathbb{E}[\tau_j|\tilde{\mathcal{G}}_{j-1}]\right)^2\right] \\
&\overset{(c)}{\leq} \frac{1}{\nu_{\min}^2}\mathbb{E}\left[\sum_{j=1}^{n} \mathbb{E}[\tau_j^2|\tilde{\mathcal{G}}_{j-1}]\right] \\
&\overset{(d)}{\leq} \frac{4}{\nu_{\min}^2}\mathbb{E}\left[\sum_{j=1}^{n} \mathbb{E}[Z_j^4|\mathcal{G}_{j-1}]\right] \\
&\overset{(e)}{\leq} \frac{4}{\nu_{\min}^2}\mathbb{E}\left[\sum_{j=1}^{n} \frac{16\mathfrak{i}_{max}^4}{n^2}\right] \\
&= \frac{64\mathfrak{i}_{\max}^4}{n\nu_{\min}^2} \\
&\overset{(f)}{=} \frac{K_W^{(1)}}{n}.
\end{aligned}$$

Here, (a) follows from (4.35) and (4.37),

(b) follows from noting that the sequence $(\tau_j - \mathbb{E}[\tau_j|\tilde{\mathcal{G}}_{j-1}], 1 \leq j \leq n)$ is a martingale difference sequence with respect to the filtration $(\tilde{\mathcal{G}}_j, 1 \leq j \leq n)$, making $(\sum_{j=1}^{k} \tau_j - \mathbb{E}[\tau_j|\tilde{\mathcal{G}}_{j-1}], 1 \leq k \leq n)$ a martingale and the orthogonal increment property of martingales [19, Theorem 5.4.6],

(c) follows from $\mathbb{E}[(\tau_j - \mathbb{E}[\tau_j|\tilde{\mathcal{G}}_{j-1}])^2|\mathcal{G}_{j-1}] = \mathbb{E}[\tau_j^2|\tilde{\mathcal{G}}_{j-1}] - \left(\mathbb{E}[\tau_j|\tilde{\mathcal{G}}_{j-1}]\right)^2$,

(d) follows from (4.38),

(e) follows since $|Z_j| \leq \frac{2}{\sqrt{n}}\mathfrak{i}_{\max}$. a.s. from (4.33),

(f) follows from defining $K_W^{(1)} \triangleq 64\mathfrak{i}_{\max}^4/\nu_{\min}^2$. □

Now define

$$\xi_t \triangleq -(r_n - \kappa) + \int_0^t \psi_s \, dW_s. \tag{4.42}$$

We have the following lemma.

Lemma 33

$$P\left(\int_0^{T_n^*} \psi_s \, dW_s \geq r_n\right) \leq P(\xi_1 \geq 0) + \delta_n.$$

Proof:

$$P\left(\int_0^{T_n^*} \psi_s \, dW_s \geq r_n\right) = P\left(\int_0^1 \psi_s \, dW_s + \theta_n \geq r_n\right),$$

where we have defined θ_n as

$$\theta_n \triangleq \int_0^\infty \mathbf{1}\{1 < s \leq T_n^*\}\psi_s \, dW_s - \int_0^\infty \mathbf{1}\{T_n^* \leq s < 1\}\psi_s \, dW_s.$$

The second moment of θ_n can be bounded as

$$\begin{aligned}
\mathbb{E}[\theta_n^2] &\overset{(a)}{\leq} 2\mathbb{E}\left[\left(\int_0^\infty \mathbf{1}\{1 < s \leq T_n^*\}\psi_s \, dW_s\right)^2\right] + 2\mathbb{E}\left[\left(\int_0^\infty \mathbf{1}\{T_n^* \leq s < 1\}\psi_s \, dW_s\right)^2\right] \\
&\overset{(b)}{=} 2\mathbb{E}\left[\int_0^\infty \mathbf{1}\{1 < s \leq T_n^*\}\psi_s^2 \, ds\right] + 2\mathbb{E}\left[\int_0^\infty \mathbf{1}\{T_n^* \leq s < 1\}\psi_s^2 \, ds\right] \\
&= 2\mathbb{E}\left[\mathbf{1}\{1 < T_n^*\} \int_1^{T_n^*} \psi_s^2 \, ds\right] + 2\mathbb{E}\left[\mathbf{1}\{T_n^* < 1\} \int_{T_n^*}^1 \psi_s^2 \, ds\right] \\
&\leq 2\nu_{\max}\mathbb{E}[|T_n^* - 1|] \\
&\leq 2\nu_{\max}\sqrt{\mathbb{E}[(T_n^* - 1)^2]} \\
&\overset{(c)}{\leq} \frac{K_W}{\sqrt{n}}.
\end{aligned}$$

Here, for (a) we have used the inequality $(a - b)^2 \leq 2a^2 + 2b^2$,

for (b) we have used [32, Problem 2.18, p. 144] [32, Problem 2.18, p. 144],

for (c) we have used Lemma 1, and recalling $K_W = 16\mathfrak{i}_{\max}^2 \nu_{\min}/\nu_{\max} = 2\nu_{\max}\sqrt{K_W^{(1)}}$.

Thus

$$
\begin{aligned}
P\left(\int_0^{T_n^*} \psi_s \, dW_s \ge r_n\right) &= P\left(\int_0^1 \psi_s \, dW_s + \theta_n \ge r_n\right) \\
&= P\left(\int_0^1 \psi_s \, dW_s + \theta_n \ge r_n \bigcap |\theta_n| \le \kappa\right) \\
&\quad + P\left(\int_0^1 \psi_s \, dW_s + \theta_n \ge r_n \bigcap |\theta_n| > \kappa\right) \\
&\le P\left(\int_0^1 \psi_s \, dW_s \ge r_n - \kappa \bigcap |\theta_n| \le \kappa\right) \\
&\quad + P\left(\int_0^1 \psi_s \, dW_s + \theta_n \ge r_n \bigcap |\theta_n| > \kappa\right) \\
&\le P\left(\int_0^1 \psi_s \, dW_s \ge r_n - \kappa\right) \\
&\quad + P\left(|\theta_n| > \kappa\right) \\
&\le P\left(\int_0^1 \psi_s \, dW_s \ge r_n - \kappa\right) + \frac{E[\theta_n^2]}{\kappa^2} \\
&\le P\left(\int_0^1 \psi_s \, dW_s \ge r_n - \kappa\right) + \frac{K_W}{\kappa^2 \sqrt{n}} \\
&= P\left(\int_0^1 \psi_s \, dW_s \ge r_n - \kappa\right) + \delta_n \\
&= P\left(\xi_1 \ge 0\right) + \delta_n.
\end{aligned}
$$

□

Now we apply [10, Corollary 3.7] (see also [29]). There exists a probability space with a measure $\hat{P}$ that supports a process $\boldsymbol{\xi}$ and a Brownian motion $\hat{\mathbf{W}}$ such that

$$
\hat{\xi}_t = -(r_n - a) + \int_0^t \hat{\psi}_s(\hat{\xi}_s) \, d\hat{W}_s, \tag{4.43}
$$

$$
P\left(\xi_t \ge a\right) = \hat{P}\left(\hat{\xi}_t \ge a\right), \qquad a \in \mathbb{R},\, t \ge 0, \tag{4.44}
$$

and $\hat{\psi}_t(\cdot)$ satisfies

$$
\hat{\psi}_t^2(u) = \mathbb{E}[\psi_t^2 | \xi_t = u] \qquad P\text{-a.s., } t \in \mathcal{N}^c,
$$

where $\mathcal{N}$ is a Lebsegue-null set. In particular, we can take $\hat{\psi}_t(u) = \sqrt{\mathbb{E}[\psi_t^2 | \xi_t = u]}$ [62, Section 5.3].

Note that $\boldsymbol{\xi}$ in (4.42) is an Itô process, where, in general the drift coefficient $\boldsymbol{\psi}$ itself can be a stochastic process. The process $\hat{\boldsymbol{\psi}}$, on the other hand has deterministic process $\hat{\boldsymbol{\psi}}$ as a drift coefficient and same one-dimensional law as that of $\boldsymbol{\psi}$ for each t.

Since $\hat{\psi}_t \in [\sqrt{\nu_{\min}}, \sqrt{\nu_{\max}}]$, unique solution in probability law holds for (4.43) [62, Exercise 7.3.3] (see also the discussion after [10, Corollary 3.13]). Thus the setup in (4.43) is *admissible* as defined by McNamara in [46]. McNamara [46, Remark 8] shows that if the goal is to maximize $\hat{P}\left(\bar{\xi}_1 \geq 0\right)$ where

$$\bar{\xi}_t = -(r_n - \kappa) + \int_0^t \bar{\psi}_s(\bar{\xi}_s)\, d\hat{W}_s,$$

by choosing the optimal diffusion coefficient $\bar{\psi}_s(\cdot)$, then such optimal diffusion control is given by

$$\bar{\psi}^{\text{opt}}(u) \triangleq \sqrt{\nu_{\min}}\mathbf{1}\{u > 0\} + \sqrt{\nu_{\max}}\mathbf{1}\{u \leq 0\}. \tag{4.45}$$

Let the corresponding SDE be

$$\bar{\xi}_t^{\text{opt}} \triangleq -(r_n - \kappa) + \int_0^t \bar{\psi}^{\text{opt}}(\bar{\xi}_s^{\text{opt}})\, d\hat{W}_s. \tag{4.46}$$

Thus

$$\hat{P}\left(\hat{\xi}_1 \geq 0\right) \leq \hat{P}\left(\bar{\xi}_1^{\text{opt}} \geq 0\right). \tag{4.47}$$

Using the distribution function of the solution to (4.45) and (4.46) (see [37]), we get

$$\hat{P}\left(\bar{\xi}_1^{\text{opt}} \geq 0\right) = 1 - \frac{2\lambda}{1+\lambda}\Phi\left(\frac{r_n - \kappa}{\sqrt{\nu_{\min}}}\right), \tag{4.48}$$

when $r_n - \kappa \leq 0$, and

$$\hat{P}\left(\bar{\xi}_1^{\text{opt}} \geq 0\right) = \frac{2}{1+\lambda} - \frac{2}{1+\lambda}\Phi\left(\frac{r_n - \kappa}{\sqrt{\nu_{\max}}}\right), \tag{4.49}$$

when $r_n - \kappa > 0$. For our choice of r_n in (4.31), we get

$$\hat{P}\left(\bar{\xi}_1^{\text{opt}} \geq 0\right) = 1 - \varepsilon - 2\delta_n. \tag{4.50}$$

Summarizing the chain of inequalities so far, we get

$$\begin{aligned}
P\left(\sum_{k=1}^{n} \mathfrak{i}^*(X_k, Y_k) \geq \log \rho_n\right) &\leq P(S_n \geq r_n) \\
&= P\left(\int_0^{T_n^*} \psi_s \, dW_s \geq r_n\right) \\
&\leq P\left(\xi_1 \geq 0\right) + \delta_n \\
&= \hat{P}\left(\hat{\xi}_1 \geq 0\right) + \delta_n \\
&\leq \hat{P}\left(\bar{\xi}_1^{\text{opt}} \geq 0\right) + \delta_n \\
&= 1 - \varepsilon - \delta_n.
\end{aligned}$$

Thus from (4.29)

$$\log M^*(n, \varepsilon) \leq nC + \sqrt{n} r_n - \log \frac{K_W}{\kappa^2 \sqrt{n}}, \tag{4.51}$$

and hence

$$\frac{\log M^*(n, \varepsilon) - nC}{\sqrt{n}} \leq r_n - \frac{\log \delta_n}{\sqrt{n}}.$$

From the definition of r_n in (4.31), and taking $n \to \infty$

$$\limsup_{n \to \infty} \frac{1}{\sqrt{n}} \left(\log M_{\text{fb}}^*(n, \varepsilon) - nC\right) \leq \begin{cases} \sqrt{\nu_{\min}} \Phi^{-1}\left(\frac{1}{2\lambda} \varepsilon (1 + \lambda)\right) + \kappa, & \varepsilon \in (0, \frac{\lambda}{1+\lambda}], \\ \sqrt{\nu_{\max}} \Phi^{-1}\left(\frac{1}{2}[\varepsilon(1 + \lambda) + (1 - \lambda)]\right) + \kappa, & \varepsilon \in (\frac{\lambda}{1+\lambda}, 1). \end{cases}$$

Since κ is arbitrary, taking $\kappa \to 0$ we prove the theorem. □

□

4.4 Converse Lemma

The next result is used repeatedly in the paper as a starting point in proving converses. A similar inequality to (4.52) can be found in [24, (42)].

Lemma 34 (Converse) *For any* $\rho > 0$ *and* $\varepsilon > 0$

$$\log M^*_{\text{fb}}(n,\varepsilon) \le \sup_F \inf_{q\in\mathcal{P}(\mathcal{Y}^n)} \left(\log\rho - \log\left[\left(1-\varepsilon-(F\circ W)\left(\sum_{k=1}^n \log\frac{W(Y_k|X_k)}{q(Y_k|\mathbf{Y}^{k-1})} > \log\rho\right)\right)^+\right]\right). \tag{4.52}$$

In particular, if for some α *and* ε*,*

$$\liminf_{n\to\infty}\inf_F \sup_{q\in\mathcal{P}(\mathcal{Y}^n)} (F\circ W)\left(\sum_{k=1}^n \log\frac{W(Y_k|X_k)}{q(Y_k|\mathbf{Y}^{k-1})} \le nC + \alpha\sqrt{n}\right) > \varepsilon, \tag{4.53}$$

then

$$\limsup_{n\to\infty} \frac{\log M^*_{\text{fb}}(n,\varepsilon) - nC}{\sqrt{n}} \le \alpha. \tag{4.54}$$

Proof: Consider an (n,R) feedback code (f,g) with average error probability at most ε. We will denote this code by $\mathfrak{C}$ and its average error probability by $\varepsilon_{\mathfrak{C}}$. Define

$$M^*_{\text{fb},\mathfrak{C}}(n) \triangleq \lceil \exp(nR) \rceil.$$

Then

$$M^*_{\text{fb}}(n,\varepsilon) = \sup_{\mathfrak{C}:\varepsilon_{\mathfrak{C}}\le\varepsilon} M^*_{\text{fb},\mathfrak{C}}(n).$$

The code $\mathfrak{C}$ induces a controller F via

$$F(x_k|\mathbf{x}^{k-1},\mathbf{y}^{k-1}) \triangleq \frac{1}{M^*_{\text{fb},\mathfrak{C}}(n)} \sum_{m=1}^{M^*_{\text{fb},\mathfrak{C}}(n)} \mathbf{1}\{f(m,\mathbf{y}^{k-1}) = x_k\},$$

which, in fact, does not depend on $\mathbf{x}^{k-1}$. Now consider the problem of hypothesis testing where a random variable U taking values in $\mathcal{U}$ can have probability measure P or Q. Upon observing U, the goal is to declare either $U \sim P$ (hypothesis H_1) or $U \sim Q$ (hypothesis H_2). Let $\beta_\alpha(P,Q)$ denote the minimum attainable error probability under Q when the error probability under P does not exceed $1-\alpha$. Then the Neyman-Pearson lemma [52, Proposition II.D.1, p. 33] guarantees that there exists a (possibly randomized) test $T : \mathcal{U} \to \{0,1\}$ (where 0 corresponds to the test selecting Q) such that

$$\sum_{u\in\mathcal{U}} P(u)T(1|u) \geq \alpha, \sum_{u\in\mathcal{U}} Q(u)T(1|u) = \beta_\alpha(P,Q).$$

Then for any $\rho > 0$

$$\begin{aligned}\alpha - \rho\beta_\alpha(P,Q) &\leq \sum_{u\in\mathcal{U}} T(1|u)(P(u) - \rho Q(u)) \\ &\leq \sum_{u\in\mathcal{U}} T(1|u)(P(u) - \rho Q(u))\mathbf{1}\{P(u) > \rho Q(u)\} \\ &= P\left(\frac{P(u)}{Q(u)} > \rho, T = 1\right) - \rho Q\left(\frac{P(u)}{Q(u)} > \rho, T = 1\right) \\ &\leq P\left(\frac{P(u)}{Q(u)} > \rho\right). \end{aligned} \tag{4.55}$$

Fix a $q \in \mathcal{P}(\mathcal{Y}^n)$. Applying [49, Theorem 26] (with $Q_{Y|X} = q$, $\varepsilon' = 1 - 1/M^*_{\mathrm{fb},\mathfrak{C}}(n)$; the assertion there is without feedback but one can verify that it applies to the feedback case as well), we get

$$\beta_{1-\varepsilon_\mathfrak{C}}(F \circ W, F \circ q) \leq \frac{1}{M^*_{\mathrm{fb},\mathfrak{C}}(n)}.$$

Moreover, from (4.55)

$$\alpha \leq (F \circ W)\left(\frac{d(F\circ W)}{d(F\circ q)} > \rho\right) + \rho\beta_\alpha(F \circ W, F \circ q),$$

i.e.,

$$\beta_{1-\varepsilon_\mathfrak{C}}(F \circ W, F \circ q) \geq \frac{1}{\rho}\left(1 - \varepsilon_\mathfrak{C} - (F \circ W)\left(\frac{d(F\circ W)}{d(F\circ q)} > \rho\right)\right)^+.$$

Thus

$$\log M^*_{\mathrm{fb},\mathfrak{C}}(n) \le \log \rho - \log \left[\left(1 - \varepsilon_{\mathfrak{C}} - (F \circ W) \left(\frac{d(F \circ W)}{d(F \circ q)} > \rho \right) \right)^+ \right].$$

Using the fact that $\varepsilon_{\mathfrak{C}} \le \varepsilon$ and that q was arbitrary, we obtain

$$\log M^*_{\mathrm{fb},\mathfrak{C}}(n) \le \inf_{q \in \mathcal{P}(\mathcal{Y}^n)} \log \rho - \log \left[\left(1 - \varepsilon - (F \circ W) \left(\frac{d(F \circ W)}{d(F \circ q)} > \rho \right) \right)^+ \right].$$

Taking the supremum over all controllers F and noting that

$$\frac{d(F \circ W)}{d(F \circ q)} = \prod_{k=1}^{n} \frac{W(y_k | x_k)}{q(y_k | \mathbf{y}^{k-1})},$$

we get

$$\log M^*_{\mathrm{fb}}(n) \le \sup_F \inf_{q \in \mathcal{P}(\mathcal{Y}^n)} \left(\log \rho - \log \left[\left(1 - \varepsilon - (F \circ W) \left(\sum_{k=1}^{n} \log \frac{W(Y_k | X_k)}{q(Y_k | \mathbf{Y}^{k-1})} > \log \rho \right) \right)^+ \right] \right).$$

This establishes (4.52). (4.54) follows directly from (4.52) and (4.53). □

BIBLIOGRAPHY

[1] Ain-ul-Aisha, L. Lai, Y. Liang, and S. Shamai Shitz. On the sum-rate capacity of Poisson miso multiple access channels. *IEEE Transactions on Information Theory*, 63(10):6457–6473, Oct 2017.

[2] Y. Altuğ and A. B. Wagner. Feedback can improve the second-order coding performance in discrete memoryless channels. In *IEEE International Symposium on Information Theory*, pages 2361–2365, June 2014.

[3] O.E. Barndorff-Nielsen and A. Shiryaev. *Change of Time and Change of Measure*, volume 13. World Scientific Publishing Co. Pte. Ltd., 2010.

[4] A. S. Bedekar. On the information about message arrival times required for in-order decoding. In *IEEE Int. Symp. Inf. Theory*, page 227, June 2001.

[5] B. Bercu, B. Delyon, and E. Rio. *Concentration Inequalities for Sums and Martingales*. Springer International Publishing, 2015.

[6] P. Bergmans. Random coding theorem for broadcast channels with degraded components. *IEEE Transactions on Information Theory*, 19(2):197–207, Mar 1973.

[7] P. Billingsley. *Convergence of Probability Measures*. Wiley Series in Probability and Statistics, 2 edition, 1999.

[8] E. Bolthausen. Exact convergence rates in some martingale central limit theorems. *Ann. Probab.*, 10(3):672–688, 1982.

[9] P. Brémaud. *Point Procceses and Queues: Martingale Dynamics*. Springer-Verlag, 1981.

[10] Gerard Brunick and Steven Shreve. Mimicking an Itô process by a solution of a stochastic differential equation. *The Annals of Applied Probability*, 23(4):1584–1628, Aug 2013.

[11] T. P. Coleman, N. Kiyavash, and V. G. Subramanian. The rate-distortion function of a Poisson process with a queueing distortion measure. In *Data Compress. Conf*, pages 63–72, Mar 2008.

[12] Thomas A Courtade and Tsachy Weissman. Multiterminal source coding under logarithmic loss. *IEEE Trans. Info. Theory*, 60(1):740–761, 2014.

[13] Thomas A Courtade and Richard D Wesel. Multiterminal source coding with an entropy-based distortion measure. In *IEEE Int. Symp. Info. Theory*, pages 2040–2044, Jul 2011.

[14] I. Csiszár and J. Körner. Broadcast channels with confidential messages. *IEEE Transactions on Information Theory*, 24(3):339–348, May 1978.

[15] I. Csiszár and J. Körner. *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Cambridge University Press, 2 edition, 2011.

[16] I. Csiszár and J. Körner. *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Cambridge University Press, 2011.

[17] M. Davis. Capacity and cutoff rate for Poisson-type channels. *IEEE Transactions on Information Theory*, 26(6):710–715, Nov 1980.

[18] C. Dellacherie and P. A. Meyer. *Probabilities and Potential B: Theory of Martingales*, volume 72 of *North-Holland Mathematics Studies*. North-Holland, 1982.

[19] R. Durrett. *Probability: Theory and Examples*. Cambridge University Press, fourth edition, 2010.

[20] R. Durrett. *Probabilty Theory and Examples*. Cambridge University Press, fourth edition, 2010.

[21] R. Durrett. *Essentials of Stochastic Processes*. Springer Texts in Statistics. Springer International Publishing, 2016.

[22] A. El Gamal. The capacity of a class of broadcast channels. *IEEE Transactions on Information Theory*, 25(2):166–169, Mar 1979.

[23] S. L. Fong and V. Y. F. Tan. private communication, 2017.

[24] S. L. Fong and V. Y. F. Tan. A tight upper bound on the second-order coding rate of the parallel Gaussian channel with feedback. *IEEE Trans. on Info. Theory*, 63(10):6474–6486, Oct 2017.

[25] Robert G. Gallager. *Information Theory and Reliable Communication*. John Wiley & Sons, Inc., New York, NY, USA, 1968.

[26] Abbas El Gamal and Young-Han Kim. *Network Information Theory*. Cambridge University Press, 2011.

[27] I. M. Gel'fand and A. M. Yaglom. Computation of the amount of information about a stochastic function contained in another such function. *Uspekhi Mat. Nauk*, 12(1):3–52, 1957.

[28] R. M. Gray. *Entropy and Information Theory*. Springer-Verlag, 1990.

[29] I. Gyöngy. Mimicking the one-dimensional marginal distributions of processes having an Ito differential. *Probability Theory and Related Fields*, 71(4):501–516, Dec 1986.

[30] Y. Kabanov. The capacity of a channel of the Poisson type. *Theory of Probabilty and Applications*, 23:143–147, 1978.

[31] O. Kallenberg. *Foundations of Modern Probability*. Springer-Verlag, New York, 2 edition, 2002.

[32] I. Karatzas and S. Shreve. *Brownian Motion and Stochastic Calculus*. Springer-Verlag New York Inc., 1988.

[33] H. Kim, B. Nachman, and A. El Gamal. Superposition coding is almost always optimal for the Poisson broadcast channel. *IEEE Transactions on Information Theory*, 62(4):1782–1794, April 2016.

[34] A. Klenke. *Probability Theory: A Comprehensive Course*. Springer London, second edition, 2013.

[35] Gunther Koliander, Dominic Schuhmacher, and Franz Hlawatsch. Rate-distortion theory of finite point processes. *IEEE Trans. Info. Theory*, 64(8):5832–5861, 2018.

[36] J. Körner and K. Marton. Comparison of two noisy channels. *Topics in information theory*, (16), 1977.

[37] G. L. Kulinich. On the limit behaviour of the solution of a stochastic diffusion equation. *Theory of Probability and its Applications*, 12(3):497–799, Jul. 1967.

[38] L. Lai, Y. Liang, and S. Shamai Shitz. On the capacity bounds for Poisson interference channels. *IEEE Transactions on Information Theory*, 61(1):223–238, Jan 2015.

[39] A. Laourine and A. B. Wagner. The degraded Poisson wiretap channel. *IEEE Transactions on Information Theory*, 58(12):7073–7085, Dec 2012.

[40] A. Lapidoth and S. Shamai. The Poisson multiple-access channel. *IEEE Transactions on Information Theory*, 44(2):488–501, March 1998.

[41] A. Lapidoth, I.E. Telatar, and R. Urbanke. On wide-band broadcast channels. *IEEE Transactions on Information Theory*, 49(12):3250–3258, Dec 2003.

[42] Amos Lapidoth, Andreas Malar, and Ligong Wang. Covering point patterns. *IEEE Trans. Info. Theory*, 61(9):4521–4533, 2015.

[43] Robert S. Liptser and Albert N. Shiryaev. *Statistics of Random Processes II*. Springer-Verlag Berlin Heidelberg, 2 edition, 2001.

[44] J.E. Mazo and J. Salz. On optical data communication via direct detection of light pulses. *The Bell System Technical Journal*, 55(3):347–369, Mar 1976.

[45] Arya Mazumdar and Ligong Wang. Covering arbitrary point patterns. In *Allerton Conf. on Comm., Cont., and Comp.*, pages 2075–2080, Oct 2012.

[46] J. M. McNamara. Optimal control of the diffusion coefficient of a simple diffusion process. *Mathematics of Operations Research*, 8(3):373–380, 1983.

[47] C. Nair. Capacity regions of two new classes of two-receiver broadcast channels. *IEEE Transactions on Information Theory*, 56(9):4207–4214, Sept 2010.

[48] S.D. Personick. Receiver design for digital fiber optic communication systems, II. *The Bell System Technical Journal*, 52(6):875–886, July 1973.

[49] Y. Polyanskiy, H. V. Poor, and S. Verdú. Channel coding rate in the finite blocklength regime. *IEEE Trans. on Info. Theory*, 56(5):2307–2359, May 2010.

[50] Y. Polyanskiy, H. V. Poor, and S. Verdú. Feedback in the non-asymptotic regime. *IEEE Trans. on Info. Theory*, 57(8):4903–4925, Aug 2011.

[51] Yury Polyanskiy and Yihong Wu. Strong data-processing inequalities for channels and bayesian networks. In *Convexity and Concentration*, pages 211–249, New York, NY, 2017. Springer New York.

[52] H. Vincent Poor. *An Introduction to Signal Detection and Estimation*. Springer-Verlag, 2 edition, 1994.

[53] R. Rockafellar. *Convex Analysis*. Princeton University Press, 1997.

[54] Izhak Rubin. Information rates and data-compression schemes for Poisson processes. *IEEE Trans. Info. Theory*, 20(2):200–210, 1974.

[55] C. E. Shannon. The zero error capacity of a noisy channel. *IRE Trans. Inf. Theory*, 2:S8–S19, 1956.

[56] C.E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27(3):379–423, July 1948.

[57] N. V. Shende and A. B. Wagner. The stochastic-calculus approach to multi-receiver Poisson channels. In *2016 IEEE International Symposium on Information Theory*, pages 360–364, July 2016.

[58] N. V. Shende and A. B. Wagner. Functional covering of point processes. In *2019 IEEE International Symposium on Information Theory*, pages 2039–2043, July 2019.

[59] N. V. Shende and A. B. Wagner. The stochastic-calculus approach to multi-receiver Poisson channels. *IEEE Transactions on Information Theory*, 65(8):5007–5027, Aug 2019.

[60] N. V. Shende, A. B. Wagner, and Y. Altuğ. When does feedback improve the second-order coding rate in discrete memoryless channels? In *2018 IEEE International Symposium on Information Theory*, pages 1485–1489, June 2018.

[61] V. Strassen. Asymptotische Abschatzungen in Shannons Informationstheorie. In *Trans. Third Prague Conf Information Theory*, pages 689–723, Prague, 1962.

[62] D.W. Stroock and S.R.S. Varadhan. *Multidimensional Diffusion Processes*. Springer Berlin Heidelberg, 2007.

[63] R. Sundaresan and S. Verdu. Capacity of queues via point-process channels. *IEEE Transactions on Information Theory*, 52(6):2697–2709, June 2006.

[64] S. Verdú. The exponential distribution in information theory. *Probl. Inf. Transm.*, 32(1):86–95, 1996.

[65] A. B. Wagner, N. V. Shende, and Y. Altuğ. A new method for employing feedback to improve coding performance, 2019.

[66] L. Wang. A strong data processing inequality for thinning Poisson processes and some applications. In *IEEE International Symposium on Information Theory*, pages 3180–3184, June 2017.

[67] A.D. Wyner. A definition of conditional mutual information for arbitrary ensembles. *Information and Control*, 38(1):51 – 59, 1978.

[68] A.D. Wyner. Capacity and error exponent for the direct detection photon channel- Part I and II. *IEEE Transactions on Information Theory*, 34(6):1449–1461, Nov 1988.