

Functional Covering of Point Processes

Nirmal V. Shende^{ID} and Aaron B. Wagner^{ID}, *Fellow, IEEE*

Abstract—We introduce a new distortion measure for point processes called **functional-covering distortion**. It is inspired by intensity theory and is related to both the covering of point processes and logarithmic-loss distortion. We obtain the distortion-rate function with feedforward under this distortion measure for a large class of point processes. For Poisson processes, the rate-distortion function is obtained under a general condition called **constrained functional-covering distortion**, of which both covering and functional-covering are special cases. Also for Poisson processes, we characterize the rate-distortion region for a two-encoder CEO problem and show that feedforward does not enlarge this region.

Index Terms—CEO problem, lossy compression, point process, Poisson process, rate-distortion.

I. INTRODUCTION

THE CLASSICAL theory of compression [2] focuses on discrete-time, sequential sources. The theory is thus well-suited to text, audio, speech, genomic data, and the like. Continuous-time signals are typically handled by reducing to discrete-time via projection onto a countable basis. Multi-dimensional extensions enable application to images and video.

Point processes model a distinct data type that appears in diverse domains such as neuroscience [3], [4], [5], [6], [7], [8], communication networks [9], [10], [11], imaging [12], [13], blockchains [14], [15], [16], [17], and photonics [18], [19], [20], [21], [22]. Formally, a point process can be viewed as a random counting measure on some space of interest [23], or if the space is a real line, a random counting function; we shall adopt the latter view. Informally, it may be viewed as simply a random collection of points representing epochs in time or points in space.

Compression of point processes emerges naturally in several of the above domains. Sub-cranial implants need to

communicate the timing of neural firings to a monitoring station over a wireless link that is low-rate because it must traverse the skull [24], [25]. In network flow correlation analysis, one cross-correlates packet timings from different links in the network [11]; this requires communication of the packet timings from one place to another. Compressing point process realizations in 3-D (also known as *point clouds*) arises in computer vision [26], [27], [28], and so on.

Various specialized approaches have been developed for compressing point processes, and in particular for measuring distortion. One natural approach is for the compressed representation to itself be a point-process realization [29], [30], [31], [32], [33], [34]. More relevant to the present paper, Lapidoth et al. [35] introduced a *covering* distortion measure, where the reconstruction of a point process on $[0, T]$ is a subset $\hat{Y}$ of $[0, T]$ that must contain all the points, and the distortion is the Lebesgue measure of the covering set (see also Shen et al. [36]).

If we encode the subset $\hat{Y}$ as an indicator function

$$\hat{Y}_t = \begin{cases} 1 & \text{if } t \in \hat{Y} \\ 0 & \text{otherwise,} \end{cases} \quad (1)$$

then $\hat{Y}_t = 0$ guarantees that no point occurred at time t while $\hat{Y}_t = 1$ indicates that a point may have occurred at t . More generally, $\hat{Y}_t$ could take many values and encode the relative belief that there is a point at t . Inspired by this observation, and the notion of logarithmic-loss distortion [37], [38], we consider the following formulation. 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) and a non-negative reconstruction $\hat{y}_0^T$, we define the *functional-covering distortion* as

$$d(\hat{y}_0^T, y_0^T) \triangleq \int_0^T \hat{y}_t dt - \int_0^T \log(\hat{y}_t) dy_t. \quad (2)$$

This is related to the covering distortion measure in the following sense. If we impose that $\hat{y}_t \in \{0, 1\}$, then (2) reduces to the covering distortion measure. Yet it is natural to consider the distortion in (2) without such a restriction, or with a more general set of allowable values for $\hat{y}_t$. In fact, there are advantages to not restricting $\hat{y}_0^T$ to the set $\{0, 1\}$. Consider a remote source setting where the encoder cannot access the point-process source directly, but instead observes a thinned version in which some of the points in the source point process are deleted randomly. Then, in case of the covering distortion the reconstruction can only be the entire interval $[0, T]$ (i.e., $\hat{y}_t = 1, t \in [0, T]$). On the other hand, under functional-covering distortion the problem has a nontrivial solution.

This idea is related to the *intensity* of a point process. Heuristically, given a random variable M , the intensity of

Manuscript received 15 April 2022; revised 23 January 2023 and 13 May 2023; accepted 17 May 2023. Date of publication 26 May 2023; date of current version 13 June 2023. This work was supported in part by the U.S. National Science Foundation under Grant CCF-2008266, Grant CCF-1956192, Grant CCF-1934985, and Grant CCF-1617673; in part by the U.S. Army Research Office under Grant W911NF-18-1-0426; and in part by a gift from Google. This paper was presented in part at the IEEE International Symposium on Information Theory, Paris, July 2019 [DOI: 10.1109/ISIT.2019.8849723]. (Corresponding author: Aaron B. Wagner.)

Nirma V. Shende was with the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY 14853 USA. He is now with the SSD DSP Architecture, Marvell Technology, Inc., Santa Clara, CA 95054 USA (e-mail: nshende@marvell.com).

Aaron B. Wagner is with the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY 14853 USA (e-mail: wagner@cornell.edu).

This article has supplementary downloadable material available at <https://doi.org/10.1109/JSAIT.2023.3279194>, provided by the authors.

Digital Object Identifier 10.1109/JSAIT.2023.3279194

a point process represented by a counting function Y_0^T is a nonnegative process Γ_0^T such that $P(Y_t - Y_{t-\Delta} = 1 | M, Y_0^{t-\Delta}) \approx \Gamma_t \Delta$ (see Definition 2 for the precise statement). Now consider a discretized version of (2), written in a way that reflects the randomness of the source Y_t :

$$\sum_{\ell=1}^{\lfloor T/\Delta \rfloor} (\hat{y}_{\ell\Delta} \Delta - (Y_{\ell\Delta} - Y_{(\ell-1)\Delta}) \log(\hat{y}_{\ell\Delta})). \quad (3)$$

Fix ℓ , and consider selecting $\hat{y}_{\ell\Delta}$ given the message M and the history of the source $\{Y_{k\Delta}\}_{k=1}^{\ell-1}$ in order to minimize

$$\mathbb{E}[\hat{y}_{\ell\Delta} \Delta - (Y_{\ell\Delta} - Y_{(\ell-1)\Delta}) \log(\hat{y}_{\ell\Delta}) | M, \{Y_{k\Delta}\}_{k=1}^{\ell-1}]. \quad (4)$$

Defining $\Gamma_{\ell\Delta}$ via

$$P(Y_{\ell\Delta} - Y_{(\ell-1)\Delta} = 1 | M, \{Y_{k\Delta}\}_{k=1}^{\ell-1}) = \Gamma_{\ell\Delta} \Delta, \quad (5)$$

and assuming $P(Y_{\ell\Delta} - Y_{(\ell-1)\Delta} > 1 | M, \{Y_{k\Delta}\}_{k=1}^{\ell-1})$ is negligible, then (4) is approximately

$$\hat{y}_{\ell\Delta} \Delta - \Gamma_{\ell\Delta} \Delta \log(\hat{y}_{\ell\Delta}), \quad (6)$$

which is minimized by the choice $\hat{y}_{\ell\Delta} = \Gamma_{\ell\Delta}$. Similarly, if $\hat{y}_{\ell\Delta}$ may depend on M but not on the past of the source process, then the optimal choice of $\hat{y}_{\ell\Delta}$ is $P(Y_{\ell\Delta} - Y_{(\ell-1)\Delta} = 1 | M) / \Delta$. For technical reasons this is not an intensity, however [39, Definition D7, p. 27]. Either way, we see that (2) encourages the decoder to output a process that describes the relative likelihood of a point at each instant of time.

The relation between the functional-covering distortion measure and logarithmic-loss [37], [38] is as follows. Recall that in the context of a memoryless source over a finite alphabet $\mathcal{X}$, the logarithmic-loss distortion measure involves the decoder, given the message $M = m$, outputting a probability distribution $\hat{p}_i(y)$ for each source symbol Y_i in the sequence. The distortion over a block of size n is then

$$\frac{1}{n} \sum_{i=1}^n \log \frac{1}{\hat{p}_i(Y_i)}, \quad (7)$$

the expectation of which is a sum of cross-entropies and is minimized by the choice $\hat{p}_i(y) = P(Y_i = y | M = m)$. With this choice, the expected distortion is $\frac{1}{n} \sum_{i=1}^n H(Y_i | M)$ and the distortion-rate function is simply

$$\max(H(Y) - R, 0). \quad (8)$$

If the source has memory, then the distortion-rate function generalizes to $\max(H(\{Y\}) - R, 0)$, where $H(\{Y\})$ is the entropy rate of the source, if the decoder is permitted to output a conditional distribution $\hat{p}_i(y_i | y_{i-1}, \dots, y_1)$ for each i . Mathematically, this is equivalent to having the decoder output an unconditional distribution $\hat{p}_i(y_i)$ but giving the decoder *feedforward* [40], [41], [42], [43], [44], [45] information, i.e., providing the decoder with $Y_1, \dots, Y_{i-1}$ before it outputs the marginal distribution $\hat{p}_i(y_i)$.

To see the connection to the functional-covering distortion measure, recall that discrete entropy can be expressed in terms of the relative entropy against the uniform distribution, and

$$\log \frac{1}{\hat{p}(Y_i)} \quad (9)$$

can be viewed as, up to an additive constant, the log-likelihood ratio between the uniform distribution and $\hat{p}(Y_i)$. For the case of point processes, the role of the uniform distribution may be played by a unit-rate Poisson process. Heuristically, consider a point process $\{Y_t\}$ in which $\hat{Y}_t \Delta$ describes the probability of a point in the interval $[t, t + \Delta]$ given the strict past $\{Y_s\}_{s < t}$. Then the log-likelihood ratio of a unit-rate Poisson process against the source evaluated on a realization Y_0^T is approximately

$$\begin{aligned} & \log \left(\prod_{\ell=1}^{\lfloor T/\Delta \rfloor} \frac{\left((1 - \hat{Y}_{(\ell-1)\Delta}) \Delta \right)^{\mathbf{1}\{Y_{\ell\Delta} > Y_{(\ell-1)\Delta}\}} e^{-\Delta}}{\left(\hat{Y}_{(\ell-1)\Delta} \Delta \right)^{\mathbf{1}\{Y_{\ell\Delta} > Y_{(\ell-1)\Delta}\}} (1 - \hat{Y}_{(\ell-1)\Delta})} \right) \\ &= \sum_{\ell=1}^{\lfloor T/\Delta \rfloor} -\Delta - \mathbf{1}\{Y_{\ell\Delta} > Y_{(\ell-1)\Delta}\} \log(\hat{Y}_{(\ell-1)\Delta}) \\ & \quad - \mathbf{1}\{Y_{\ell\Delta} = Y_{(\ell-1)\Delta}\} \log(1 - \hat{Y}_{(\ell-1)\Delta}) \end{aligned} \quad (10)$$

$$\begin{aligned} & \approx \sum_{\ell=1}^{\lfloor T/\Delta \rfloor} -\Delta - \mathbf{1}\{Y_{\ell\Delta} > Y_{(\ell-1)\Delta}\} \log(\hat{Y}_{(\ell-1)\Delta}) \\ & \quad + \hat{Y}_{(\ell-1)\Delta} \Delta \mathbf{1}\{Y_{\ell\Delta} = Y_{(\ell-1)\Delta}\} \end{aligned} \quad (11)$$

$$\approx \int_0^T (\hat{Y}_t - 1) dt - \int_0^T \log(\hat{Y}_t) dY_t, \quad (12)$$

which differs from (2) only by the constant term $-T$. This argument can be made rigorous via Girsanov's theorem for point processes [39, Ch. VI, Ths. T2–T4]. Specifically, if the reconstruction $\hat{Y}_0^T$ is assumed to be bounded then it can be used to define a probability measure on the space of point-processes $(\mathcal{N}_0^T, \mathcal{F}^N)$ via the following Radon-Nikodym derivative,

$$\frac{dP_{\hat{Y}_0^T}}{dP_0}(y_0^T) = \exp \left(\int_0^T \log(\hat{y}_t) dy_t - \int_0^T (\hat{y}_t - 1) dt \right), \quad (13)$$

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 $\hat{y}_0^T$ [39, Ch. VI, Ths. T2–T4] and the functional-covering distortion is related to the above Radon-Nikodym derivative as

$$d(\hat{Y}_0^T, y_0^T) = -\log \left(\frac{dP_{\hat{Y}_0^T}}{dP_0}(y_0^T) \right) + T. \quad (14)$$

Thus an alternative view of the functional-covering distortion measure is that the reconstruction $\hat{y}_t$ describes a distribution over the source realizations and (2) is simply the log-loss distortion measure. Note, however, that since $\hat{y}_t \Delta$ describes the probability of a point in $[t, t + \Delta]$ given the history of the source process, either the decoder effectively needs feedforward, as in the above discussion on intensity, or the distribution induced by $\hat{y}_0^T$ needs to be such that the source is conditionally Poisson, i.e., memoryless, given $\hat{y}_0^T$. Regarding the former case, note that the concept of feedforward is somewhat subtle in continuous time; mathematically, it amounts to assuming that $\hat{y}_0^T$ is *predictable* [39, Definition D4, p. 8] with respect to the message and the history of the source process. In the latter case, the functional-covering distortion measure is an affine function of the negative log-likelihood of a Poisson channel with input $\hat{y}_0^T$ and output y_0^T , analogous to the way

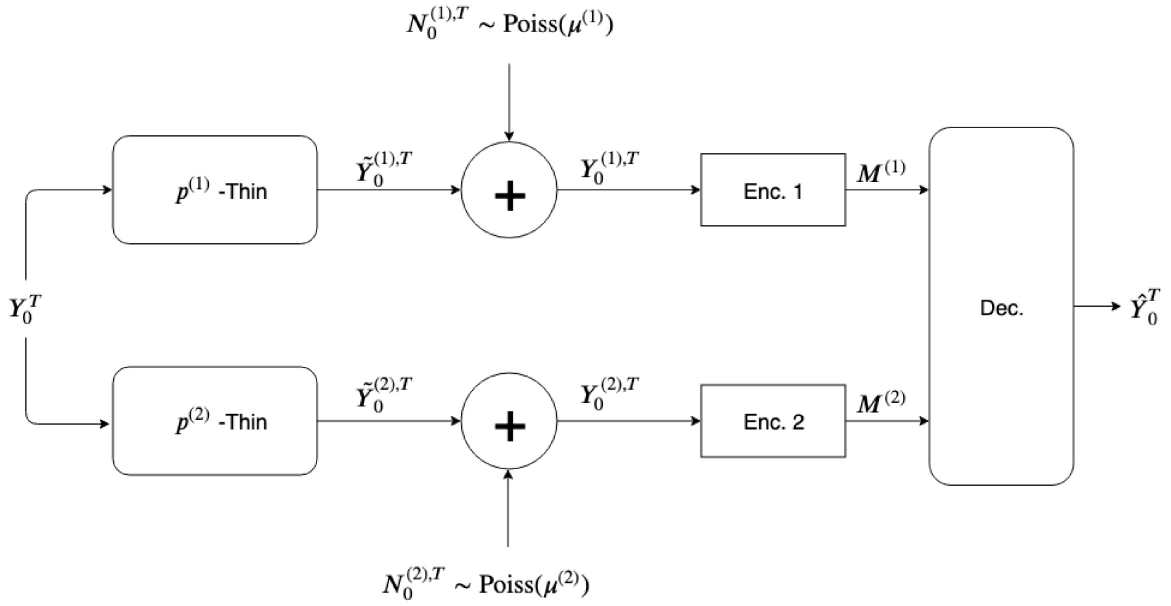


Fig. 1. Poisson CEO Problem.

Hamming distance is an affine function of the negative log-likelihood of a binary symmetric channel and squared error is an affine function of the negative log-likelihood of a Gaussian channel.

We shall consider both the feedforward and memoryless approaches. For any arbitrary source with an intensity, we show in Theorem 4 and Corollary 1 that if the decoder has feedforward information then the rate-distortion function is linear as in (8). For the case of a Poisson process source, in Theorem 5 we determine the rate-distortion tradeoff with the added restriction that for some given set $\mathcal{A}$, we have $\hat{y}_t \in \mathcal{A}$ for all t . We find that this tradeoff is unaffected by the presence of feedforward. Taking $\mathcal{A} = \{0, 1\}$ recovers the result of Lapidot et al. [35, Th. 1] for the covering distortion measure (Corollary 3) and taking $\mathcal{A} = [0, \infty)$ recovers the unconstrained case.

Recall that the covering distortion measure yields a trivial rate-distortion tradeoff in the remote-source setting in which the encoder observes a thinned version of the source. Here we determine the rate-distortion function under the functional-covering distortion measure for a Poisson source that is observed after thinning and superposition with an independent Poisson process. We find that the rate-distortion function is not degenerate. In fact, in Theorem 6 we solve the more general two-encoder CEO problem (see Fig. 1), again finding that the rate-distortion function is unaffected by feedforward. It is notable that our scheme for the CEO problem does not require binning (cf. [46, Sec. 15.4]). A result for the CEO problem without thinning and with the covering distortion measure was earlier obtained by Wang [47].

To prove these results, we establish various technical tools that are useful for characterizing mutual information involving continuous-time point processes. Theorem 1 provides a general formula for computing mutual information in terms of intensities; it subsumes existing formulae for mutual informations involving doubly stochastic Poisson processes [48], [49], [50]

and queueing processes [51] as special cases. We also prove two strong data processing inequalities. Theorem 2 provides a strong data processing inequality for Poisson processes under superposition, which complements the strong data processing inequality for Poisson processes under thinning due to Wang [47]. We also provide a self-contained proof of Wang's theorem in Theorem 3. These results may have independent use.

The remainder of the paper is organized as follows. Section II introduces the necessary notation and contains the mutual information identities and inequalities. Section III defines the functional covering distortion measure precisely and contains point-to-point results for general sources. Section IV contains point-to-point results for Poisson sources, and Section V contains results for the CEO problem. Many of the proofs and auxiliary results are contained in the supplementary material.

II. POINT PROCESSES, INTENSITIES, AND MUTUAL INFORMATION

Our treatment follows Brémaud [39], to which the reader is referred for additional background. We 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 $\mathcal{F}_T \subseteq \mathcal{F}$. We will assume that the given filtration $(\mathcal{F}_t : t \in [0, T])$, P , and $\mathcal{F}$ satisfy the "usual conditions" [39, Ch. 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 $(t, \omega) \rightarrow X_t(\omega)$ 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

$$(s, t] \times A; \quad 0 < s \leq t \leq T, \quad A \in \mathcal{F}_s. \quad (15)$$

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 $A \rightleftharpoons B \rightleftharpoons C$ forms a Markov chain under measure P if A and C are conditionally independent given B under P . The notation $P \ll Q$ indicates that the probability measure P is absolutely continuous with respect to the measure Q . The indicator function for the event $\mathbf{E}$ is denoted by $\mathbf{1}\{\mathbf{E}\}$. All logarithms and exponentiations are base e . The superscripts $(x)^+$ and $(x)^-$ denote the positive ($\max(x, 0)$) and the negative part ($-\min(x, 0)$) of x respectively. The ceiling of x is denoted by $\lceil x \rceil$. Throughout this paper we adopt the convention that $0 \log(0) = 0$, $\exp(\log(0)) = 0$, and $0^0 = 1$.

Definition 1: We use $\phi(x)$ to denote $x \log(x)$.

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 nonnegative integers), is right continuous, and has unit increasing jumps with $N_0 = 0$. Let $\mathcal{F}^N$ be the σ -field generated by the open sets of $\mathcal{N}_0^T$ when endowed with the Skorohod topology [52, Sec. 12].

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 $\Gamma_0^T = (\Gamma_t : t \in [0, T])$, where Γ_0^T is a nonnegative 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 nonnegative $(\mathcal{F}_t : t \in [0, T])$ -predictable processes C_0^T ¹:

$$\mathbb{E} \left[\int_0^T C_s dN_s \right] = \mathbb{E} \left[\int_0^T C_s \Gamma_s ds \right]. \quad (16)$$

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 3: A point process Y_0^T is said to be a *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 a Poisson process [39, Th. T4, Ch. II, p. 25] and vice versa [39, Sec. 2, Ch. II, p. 23].

Definition 4: $P_0^{Y_0^T}$ denotes the distribution of a point process Y_0^T (on the space $(\mathcal{N}_0^T, \mathcal{F}^N)$) under which Y_0^T is a Poisson process with unit rate.

The following theorem allows us to express the mutual information involving 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 [50, Th. 1]. See Appendix A in the supplementary material for a review of how mutual information is defined for general ensembles such as point processes.

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

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

$$\mathbb{E} \left[\int_0^T |\phi(\Lambda_t)| dt \right] < \infty, \quad (17)$$

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

$$I(M; Y_0^T) = \mathbb{E} \left[\int_0^T \phi(\Gamma_t) - \phi(\Lambda_t) dt \right]. \quad (18)$$

See Appendix C in the supplementary material for the proof. To understand this result intuitively, let $h(Y_0^T)$ denote the negative relative entropy of the process Y_0^T against the unit-rate Poisson process, and define $h(Y_0^T|M)$ similarly. Then we expect to have $I(M; Y_0^T) = h(Y_0^T) - h(Y_0^T|M)$. But from (13) in the introduction

$$h(Y_0^T) = \mathbb{E} \left[\int_0^T (\Lambda_t - 1) dt - \int_0^T \log(\Lambda_t) dY_t \right] \quad (19)$$

$$= \int_0^T (\mathbb{E}[\Lambda_t] - 1) dt - \int_0^T \mathbb{E}[\phi(\Lambda_t)] dt, \quad (20)$$

where (20) follows because $\mathbb{E}[\int_0^T \log(\Lambda_t) dY_t] = \mathbb{E}[\int_0^T \log(\Lambda_t) \Lambda_t dt] = \mathbb{E}[\int_0^T \phi(\Lambda_t) dt]$. Similarly, we have

$$h(Y_0^T|M) = \int_0^T (\mathbb{E}[\Gamma_t] - 1) dt - \int_0^T \mathbb{E}[\phi(\Gamma_t)] dt, \quad (21)$$

Subtracting (21) from (20) gives (18).

We shall require two strong data processing inequalities, one for superposition and one for thinning. These will be used in the proofs of our coding theorems.

Theorem 2: Let 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,

$$I(M; Y_0^T) = \mathbb{E} \left[\int_0^T \phi(\Gamma_t) - \phi(\lambda) dt \right] \quad (22)$$

$$I(M; Z_0^T) \leq \mathbb{E} \left[\int_0^T \phi(\Gamma_t + \mu) - \phi(\lambda + \mu) dt \right]. \quad (23)$$

Proof: Since $M \rightleftharpoons Y_0^T \rightleftharpoons Z_0^T$ forms a Markov chain, the data processing inequality gives $I(M; Z_0^T) \leq I(M; Y_0^T) < \infty$. Applying Theorem 1 and using the uniqueness of intensities,

$$I(M; Y_0^T) = \mathbb{E} \left[\int_0^T \phi(\Gamma_t) - \phi(\lambda) dt \right] \quad \text{and} \quad (24)$$

$$I(M; Z_0^T) = \mathbb{E} \left[\int_0^T \phi(\hat{\Gamma}_t) - \phi(\hat{\lambda}_t) dt \right], \quad (25)$$

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 10, 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 (25) and applying Jensen's inequality yields

$$I(M; Z_0^T) = \mathbb{E} \left[\int_0^T \phi(\mathbb{E}[\Gamma_t|M, Z_0^{t-}] + \mu) - \phi(\lambda + \mu) dt \right] \quad (26)$$

$$\leq \mathbb{E} \left[\int_0^T \phi(\Gamma_t + \mu) - \phi(\lambda + \mu) dt \right]. \quad (27)$$

Definition 5: 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.

The following theorem was first proven by Wang [47]. We provide a self-contained proof that uses Theorem 1 and Lemma 11.

Theorem 3: 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 be obtained from p -thinning of Y_0^T such that the thinning operation is independent of M . Then

$$I(M; Z_0^T) \leq (1-p)I(M; Y_0^T). \quad (28)$$

Proof: The data processing inequality gives $I(M; Z_0^T) \leq I(M; Y_0^T) < \infty$. Applying Theorem 1,

$$I(M; Y_0^T) = \mathbb{E} \left[\int_0^T \phi(\Gamma_t) - \phi(\lambda) dt \right], \quad (29)$$

and

$$I(M; Z_0^T) = \mathbb{E} \left[\int_0^T \phi(\hat{\Gamma}_t) - \phi(\hat{\lambda}_t) dt \right], \quad (30)$$

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 (respectively Z_0^T). Due to the uniqueness of the intensities and Lemma 11, we can take for each $t \in [0, T]$,

$$\hat{\Gamma}_t = (1-p)\mathbb{E}[\Gamma_t|M, Z_0^{t-}] \quad (31)$$

$$\hat{\lambda}_t = (1-p)\lambda. \quad (32)$$

Noting that $\phi((1-p)x) = (1-p)\phi(x) + x\phi(1-p)$, (30) yields

$$I(M; Z_0^T) = (1-p)\mathbb{E} \left[\int_0^T \phi(\mathbb{E}[\Gamma_t|M, Z_0^{t-}]) - \phi(\lambda) dt \right] + \phi(1-p)\mathbb{E} \left[\int_0^T \Gamma_t - \lambda dt \right] \quad (33)$$

$$= (1-p)\mathbb{E} \left[\int_0^T \phi(\mathbb{E}[\Gamma_t|M, Z_0^{t-}]) - \phi(\lambda) dt \right] \quad (34)$$

$$\leq (1-p)\mathbb{E} \left[\int_0^T \phi(\Gamma_t) - \phi(\lambda) dt \right] \quad (35)$$

$$= (1-p)I(M; Y_0^T), \quad (36)$$

where for (34) we have used the fact that $\mathbb{E}[\int_0^T \Gamma_t dt] = \mathbb{E}[\int_0^T \lambda dt]$, and for (35) we have used Jensen's inequality. ■

III. 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. More general results are obtained for Poisson processes in the next sections.

Definition 6: Given a point process $Y_0^T \in \mathcal{N}_0^T$, and a nonnegative function $\hat{y}_0^T$, the *functional-covering distortion* d is

$$d(\hat{y}_0^T, Y_0^T) \triangleq \int_0^T \hat{y}_t dt - \int_0^T \log(\hat{y}_t) dy_t, \quad (37)$$

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 as the message, constrained via predictability. In particular, we will call $\hat{Y}_0^T$ an *allowable reconstruction with feedforward* if it is nonnegative and $(\sigma(Y_0^T) : t \in [0, T])$ -predictable. Let $\hat{\mathcal{Y}}_{0,FF}^T$ denote the set of all $\hat{y}_0^T$ processes which are allowable reconstructions with feedforward.

Definition 7: A (T, R, D) code with feedforward consists of an encoder f

$$f : \mathcal{N}_0^T \rightarrow \{1, \dots, \dots, \lceil \exp(RT) \rceil\} \quad (38)$$

and a decoder g

$$g : \{1, \dots, \lceil \exp(RT) \rceil\} \times \mathcal{N}_0^T \rightarrow \hat{\mathcal{Y}}_{0,FF}^T \quad (39)$$

satisfying

$$\mathbb{E} \left[\int_0^T \hat{Y}_t dt \right] < \infty \quad (40)$$

and the distortion constraint

$$\mathbb{E} \left[\frac{1}{T} d(\hat{Y}_0^T, Y_0^T) \right] \leq D. \quad (41)$$

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 8: The *minimum achievable distortion with feedforward at rate R and blocklength T* is

$$D_F^*(R, T) \triangleq \inf \{D : \text{there exists a } (T, R, D) \text{ code} \quad (42)$$

$$\text{with feedforward}\}. \quad (43)$$

Definition 9: The *distortion-rate function with feedforward* is

$$D_F(R) \triangleq \limsup_{T \rightarrow \infty} D_F^*(R, T). \quad (44)$$

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.

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

Theorem 4: Let Y_0^T be a point process with $(\mathcal{F}_t^Y : t \in [0, T])$ -intensity Λ_0^T such that

$$\mathbb{E} \left[\int_0^T |\phi(\Lambda_t)| dt \right] < \infty. \quad (45)$$

Let

$$\Xi(Y_0^T) \triangleq \frac{1}{T} \mathbb{E} \left[\int_0^T \Lambda_t - \phi(\Lambda_t) dt \right], \quad (46)$$

and

$$\delta_T \triangleq P(Y_T = 0) < 1. \quad (47)$$

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

$$\Xi(Y_0^T) - R - \frac{1}{T} \leq D_F^*(R, T) \leq \Xi(Y_0^T) - (1 - \delta_T)R + \frac{1}{T}. \quad (48)$$

Remark 1: Per (20), $\Xi(Y_0^T)$ can be interpreted as an entropy rate of the source. Compare (8).

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

$$\frac{1}{T} \mathbb{E}[d(\Lambda_0^T, Y_0^T)] = \frac{1}{T} \mathbb{E}\left[\int_0^T \Lambda_t dt - \int_0^T \log(\Lambda_t) dY_t\right] \quad (49)$$

$$= \frac{1}{T} \mathbb{E}\left[\int_0^T \Lambda_t - \phi(\Lambda_t) dt\right] \quad (50)$$

$$= \Xi(Y_0^T). \quad (51)$$

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 5, 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 $\{2, \dots, J\}$. From Theorem 1, 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 $\mathbb{E}[\int_0^T \Gamma_t dt] = \mathbb{E}[\int_0^T \Lambda_t dt] < \infty$, and from Theorem 1, $\mathbb{E}[\int_0^T \log(\Gamma_t) dY_t] < \infty$. Hence

$$\frac{1}{T} \mathbb{E}[d(\Gamma_0^T, Y_0^T)] = \frac{1}{T} \mathbb{E}\left[\int_0^T \Gamma_t dt - \int_0^T \log(\Gamma_t) dY_t\right] \quad (52)$$

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

$$\begin{aligned} \frac{1}{T} H(M) &= -\frac{1}{T} (\delta_T \log(\delta_T) + (1 - \delta_T) \log(1 - \delta_T)) \\ &\quad + \frac{1 - \delta_T}{T} (\log(J - 1)) \end{aligned} \quad (53)$$

$$\geq \frac{1 - \delta_T}{T} (\log(J - 1)) \quad (54)$$

$$\geq \frac{1 - \delta_T}{T} \log(J / \exp(1)) \quad (55)$$

$$\geq (1 - \delta_T)R - \frac{1}{T}, \quad (56)$$

where for (54), we have used the bound $-\delta_T \log(\delta_T) - (1 - \delta_T) \log(1 - \delta_T) \geq 0$; for (55), we have used the inequality

$J - 1 \geq J / \exp(1)$ when $J \geq 2$; and for (56), we used the fact that $RT \leq \log(J)$.

$H(M)$ also satisfies

$$\frac{1}{T} H(M) = \frac{1}{T} I(M; Y_0^T) \quad (57)$$

$$= \frac{1}{T} \mathbb{E}\left[\int_0^T \log(\Gamma_t) dY_t\right] - \frac{1}{T} \mathbb{E}\left[\int_0^T \phi(\Lambda_t) dt\right], \quad (58)$$

where, for (57) we have used Lemma 2; for (58) we have used Theorem 1. The average distortion can be bounded as follows:

$$\frac{1}{T} \mathbb{E}[d(\Gamma_0^T, Y_0^T)] = \frac{1}{T} \mathbb{E}\left[\int_0^T \Gamma_t dt - \log(\Gamma_t) dY_t\right] \quad (59)$$

$$= \frac{1}{T} \mathbb{E}\left[\int_0^T \Gamma_t dt\right] - \frac{1}{T} \mathbb{E}\left[\int_0^T \log(\Gamma_t) dY_t\right] \quad (60)$$

$$= \frac{1}{T} \mathbb{E}\left[\int_0^T \Lambda_t dt\right] - \frac{1}{T} \mathbb{E}\left[\int_0^T \log(\Gamma_t) dY_t\right] \quad (61)$$

$$= \frac{1}{T} \mathbb{E}\left[\int_0^T \Lambda_t dt\right] - \frac{1}{T} H(M) \quad (62)$$

$$- \frac{1}{T} \mathbb{E}\left[\int_0^T \phi(\Lambda_t) dt\right] \quad (63)$$

$$\leq \frac{1}{T} \mathbb{E}\left[\int_0^T \Lambda_t - \phi(\Lambda_t) dt\right] \quad (64)$$

$$- (1 - \delta_T)R + \frac{1}{T} \quad (65)$$

$$= \Xi(Y_0^T) - (1 - \delta_T)R + \frac{1}{T}, \quad (66)$$

where, for (60), we have used the fact that $\mathbb{E}[\int_0^T \log(\Gamma_t) dY_t] < \infty$ due to Theorem 1; for (61), we used the equality $\mathbb{E}[\int_0^T \Gamma_t dt] = \mathbb{E}[\int_0^T \Lambda_t dt]$; for (63), we used (58); and for (65), we used (56).

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

$$R + \frac{1}{T} \geq \frac{1}{T} \log(J) \geq \frac{1}{T} H(M) = \frac{1}{T} I(M; Y_0^T), \quad (67)$$

where the last equality follows because of Lemma 2.

Since $I(M; Y_0^T) < \infty$, we conclude from Theorem 1 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

$$I(M; Y_0^T) = \mathbb{E}\left[\int_0^T \phi(\Gamma_t) dt\right] - \mathbb{E}\left[\int_0^T \phi(\Lambda_t) dt\right]. \quad (68)$$

Hence from (67)

$$R \geq \frac{1}{T} \mathbb{E}\left[\int_0^T \phi(\Gamma_t) dt\right] - \frac{1}{T} \mathbb{E}\left[\int_0^T \phi(\Lambda_t) dt\right] - \frac{1}{T}. \quad (69)$$

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

$$D \geq \frac{1}{T} \mathbb{E}\left[d(\hat{Y}_0^T, Y_0^T)\right] \quad (70)$$

$$= \frac{1}{T} \mathbb{E} \left[\int_0^T \hat{Y}_t dt - \int_0^T \log(\hat{Y}_t) dY_t \right] \quad (71)$$

$$= \frac{1}{T} \mathbb{E} \left[\int_0^T \hat{Y}_t - \log(\hat{Y}_t) \Gamma_t dt \right], \quad (72)$$

where in the last line we have used Lemma 12.

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

$$\mathbb{E} \left[\int_0^T \log(\hat{Y}_t) \Gamma_t dt \right] \leq \mathbb{E} \left[\int_0^T \phi(\Gamma_t) - \Gamma_t + \hat{Y}_t dt \right] \quad (73)$$

$$= \mathbb{E} \left[\int_0^T \phi(\Gamma_t) dt \right] - \mathbb{E} \left[\int_0^T \Gamma_t dt \right] + \mathbb{E} \left[\int_0^T \hat{Y}_t dt \right]. \quad (74)$$

From (72) and (69), we deduce

$$R + D \geq \frac{1}{T} \mathbb{E} \left[\int_0^T \phi(\Gamma_t) dt \right] - \frac{1}{T} \mathbb{E} \left[\int_0^T \phi(\Lambda_t) dt \right] + \frac{1}{T} \mathbb{E} \left[\int_0^T \hat{Y}_t dt \right] - \frac{1}{T} \mathbb{E} \left[\int_0^T \log(\hat{Y}_t) dY_t \right] - \frac{1}{T} \quad (75)$$

$$\geq \frac{1}{T} \mathbb{E} \left[\int_0^T \Gamma_t dt \right] - \frac{1}{T} \mathbb{E} \left[\int_0^T \phi(\Lambda_t) dt \right] - \frac{1}{T} \quad (76)$$

$$\geq \frac{1}{T} \mathbb{E} \left[\int_0^T \Lambda_t dt \right] - \frac{1}{T} \mathbb{E} \left[\int_0^T \phi(\Lambda_t) dt \right] - \frac{1}{T} \quad (77)$$

$$= \Xi(Y_0^T) - \frac{1}{T}, \quad (78)$$

where, for (76) we have used (74); and for (77) we used the fact that $\mathbb{E}[\int_0^T \Gamma_t dt] = \mathbb{E}[\int_0^T dY_t] = \mathbb{E}[\int_0^T \Lambda_t dt]$. 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

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

Then

$$D_F(R) = \bar{\Xi}(Y) - R. \quad (79)$$

Proof: The corollary follows from the definition $D_F(R) = \limsup_{T \rightarrow \infty} D_F^*(R, T)$ and from the bounds on $D_F^*(R, T)$ in the Theorem 4. ■

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.

IV. 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 $\hat{Y}_0^T$ that are nonnegative 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 values in $\mathcal{A}$, so that for all $t \in [0, T]$, $\hat{Y}_t \in \mathcal{A}$.

Definition 10: A (T, R, D) code consists of an *encoder* f

$$f : \mathcal{N}_0^T \rightarrow \{1, \dots, \lceil \exp(RT) \rceil\} \quad (80)$$

and a *decoder* g

$$g : \{1, \dots, \lceil \exp(RT) \rceil\} \rightarrow \hat{\mathcal{Y}}_0^T \quad (81)$$

satisfying

$$\hat{Y}_t \in \mathcal{A}, \mathbb{E} \left[\int_0^T \hat{Y}_t dt \right] < \infty \quad (82)$$

and the distortion constraint

$$\frac{1}{T} \mathbb{E} \left[d(\hat{Y}_0^T, Y_0^T) \right] \leq D. \quad (83)$$

As before, 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 11: 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 $\lim_{n \rightarrow \infty} T_n = \infty$.

Definition 12: The *rate-distortion region* $\mathfrak{RD}_{\mathcal{A}}^{\mathcal{P}}$ is the set of all achievable rate-distortion vectors (R, D) .

The rate-distortion region $\mathfrak{RD}_{\mathcal{A}}^{\mathcal{P},F}$ for a Poisson source with feedforward is defined analogously, except that a code is defined via Definition 7 with the addendum that we require $\hat{Y}_t \in \mathcal{A}$ for all t .

Theorem 5: The rate-distortion region for the constrained functional-covering of a Poisson process with rate $\lambda > 0$ is given by

$$\mathfrak{RD}_{\mathcal{A}}^{\mathcal{P}} = \mathfrak{RD}_{\mathcal{A}}^{\mathcal{P},F} = \mathfrak{RD}, \quad (84)$$

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

$$R \geq \lambda \sum_{k=1}^4 \beta_k \log \left(\frac{\beta_k}{\alpha_k} \right) \quad (85)$$

$$D \geq \sum_{k=1}^4 \alpha_k \Psi_{\mathcal{A}} \left(\frac{\lambda \beta_k}{\alpha_k} \right), \quad (86)$$

where

$$\Psi_{\mathcal{A}}(u) \triangleq \inf_{v \in \mathcal{A}} v - u \log(v) \quad (87)$$

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

$$R \triangleq \lambda \sum_{k=1}^4 \beta_k \log\left(\frac{\beta_k}{\alpha_k}\right) \quad (88)$$

$$D \triangleq \sum_{k=1}^4 \alpha_k \Psi_{\mathcal{A}}\left(\frac{\lambda \beta_k}{\alpha_k}\right). \quad (89)$$

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 set it equal to 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 $k \in \{1, 2, 3, 4\}$,

$$P(\bar{U} = k | \bar{Y} = 1) = \beta_k \quad (90)$$

$$P(\bar{U} = k | \bar{Y} = 0) = \alpha_k. \quad (91)$$

Define the discretized distortion function

$$\bar{d}(\hat{y}, \bar{y}) \triangleq \hat{y} - \frac{\log(\hat{y})}{\Delta} \mathbf{1}\{\bar{y} = 1\} \quad \hat{y} \in \mathcal{A}, \bar{y} \in \{0, 1\}. \quad (92)$$

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

$$\left| \Psi_{\mathcal{A}}\left(\frac{\lambda \beta_k}{\alpha_k}\right) - \left(v - \frac{\lambda \beta_k}{\alpha_k} \log(v)\right) \right| \leq \frac{\epsilon}{4}, \quad (93)$$

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 $\bar{d}(\hat{Y}, \bar{Y})$ with respect to Δ is given by the following lemma. ■

Lemma 1:

$$\lim_{\Delta \rightarrow 0} \frac{I(\bar{U}; \bar{Y})}{\Delta} = R \quad (94)$$

$$\lim_{\Delta \rightarrow 0} \mathbb{E}[\bar{d}(\hat{Y}, \bar{Y})] \leq D + \frac{\epsilon}{4}. \quad (95)$$

Proof: Please see the supplementary material. ■

Let

$$\kappa \triangleq \max_{k \in \{1, 2, 3, 4\}} \left| \log(\hat{Y}(k)) \right|. \quad (96)$$

Due to [53, Th. 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

$$\bar{f} : (\bar{Y}_j : j \in \{1, \dots, n\}) \rightarrow \{1, \dots, L\} \quad (97)$$

$$\bar{g} : \{1, \dots, L\} \rightarrow (\hat{Y}_j : j \in \{1, \dots, n\}) \quad (98)$$

satisfying

$$\frac{1}{n} \log(L) \leq I(\bar{U}; \bar{Y}) + \bar{\epsilon} \quad (99)$$

$$\mathbb{E} \left[\frac{1}{n} \sum_{j=1}^n \bar{d}(\hat{Y}_j, \bar{Y}_j) \right] \leq \mathbb{E}[\bar{d}(\hat{Y}, \bar{Y})] + \bar{\epsilon}. \quad (100)$$

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

$$\hat{Y}_t \triangleq \sum_{j=1}^n \hat{Y}_j \mathbf{1}\{t \in ((j-1)\Delta, j\Delta]\} \quad t \in [0, T]. \quad (101)$$

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:

$$\frac{1}{T} d(\hat{Y}_0^T; Y_0^T) = \frac{1}{T} \int_0^T \hat{Y}_t dt - \frac{1}{T} \int_0^T \log(\hat{Y}_t) dY_t \quad (102)$$

$$= \frac{1}{n} \sum_{j=1}^n \hat{Y}_j - \frac{1}{T} \sum_{j=1}^n \log(\hat{Y}_j) \bar{\bar{Y}}_j \quad (103)$$

$$= \frac{1}{n} \sum_{j=1}^n \hat{Y}_j - \frac{1}{n\Delta} \sum_{j=1}^n \log(\hat{Y}_j) \bar{Y}_j - \frac{1}{T} \sum_{j=1}^n \log(\hat{Y}_j) (\bar{\bar{Y}}_j - 1) \mathbf{1}\{\bar{\bar{Y}}_j > 1\} \quad (104)$$

$$= \frac{1}{n} \sum_{j=1}^n \bar{d}(\hat{Y}_j, \bar{Y}_j) - \frac{1}{T} \sum_{j=1}^n \log(\hat{Y}_j) (\bar{\bar{Y}}_j - 1) \mathbf{1}\{\bar{\bar{Y}}_j > 1\} \quad (105)$$

$$\leq \frac{1}{n} \sum_{j=1}^n \bar{d}(\hat{Y}_j, \bar{Y}_j) + \frac{\kappa}{T} \sum_{j=1}^n (\bar{\bar{Y}}_j - 1) \mathbf{1}\{\bar{\bar{Y}}_j > 1\} \quad (106)$$

$$\leq \frac{1}{n} \sum_{j=1}^n \bar{d}(\hat{Y}_j, \bar{Y}_j) + \frac{\kappa}{T} \sum_{j=1}^n \bar{\bar{Y}}_j \mathbf{1}\{\bar{\bar{Y}}_j > 1\}, \quad (107)$$

where for (106), we have used the definition of κ in (96), since $\bar{\bar{Y}}_j > 1$ implies $\bar{Y}_j = 1$ which implies $\bar{\bar{Y}}_j > 0$ in order for $\bar{d}(\hat{Y}_j, 1) < \infty$, which occurs a.s. since $\mathbb{E}[\bar{d}(\hat{Y}, \bar{Y})] < \infty$ so long as Δ is sufficiently small.

Hence taking expectations, we get

$$\mathbb{E} \left[\frac{1}{T} d(\hat{Y}_0^T; Y_0^T) \right] \leq \mathbb{E} \left[\frac{1}{n} \sum_{j=1}^n \bar{d}(\hat{Y}_j, \bar{Y}_j) \right] \quad (108)$$

$$+ \kappa \mathbb{E} \left[\frac{1}{T} \sum_{j=1}^n \bar{\bar{Y}}_j \mathbf{1}\{\bar{\bar{Y}}_j > 1\} \right] \quad (109)$$

$$\leq \mathbb{E}[\bar{d}(\hat{Y}, \bar{Y})] \quad (110)$$

$$+ \kappa \mathbb{E} \left[\frac{1}{T} \sum_{j=1}^n \bar{Y}_j \mathbf{1} \{ \bar{Y}_j > 1 \} \right] + \bar{\epsilon} \quad (111)$$

$$= \mathbb{E} \left[\bar{d}(\hat{\bar{Y}}, \bar{Y}) \right] + \kappa(\lambda - \lambda \exp(-\lambda \Delta)) + \bar{\epsilon} \quad (112)$$

$$\leq \mathbb{E} \left[\bar{d}(\hat{\bar{Y}}, \bar{Y}) \right] + \kappa \lambda^2 \Delta + \bar{\epsilon}, \quad (113)$$

where, for (111), we have used (100); for (112) we note that $\mathbb{E}[\bar{Y}_j \mathbf{1} \{ \bar{Y}_j > 1 \}] = \lambda \Delta - \lambda \Delta \exp(-\lambda \Delta)$; and for (113), we have used the inequality $1 - u \leq \exp(-u)$.

Moreover using (100),

$$\frac{1}{T} \log(L) = \frac{1}{n\Delta} \log(L) \leq \frac{I(\bar{U}; \bar{Y})}{\Delta} + \frac{\bar{\epsilon}}{\Delta}. \quad (114)$$

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

$$\frac{I(\bar{U}; \bar{Y})}{\Delta} \leq R + \frac{\epsilon}{4} \quad (115)$$

$$\mathbb{E} \left[\bar{d}(\hat{\bar{Y}}, \bar{Y}) \right] \leq D + \frac{\epsilon}{2} \quad (116)$$

$$\kappa \lambda^2 \Delta \leq \frac{\epsilon}{2}. \quad (117)$$

Then let $\bar{\epsilon} = \Delta \epsilon / 4$, and choose a sufficiently large n so that (100) is satisfied. From (113) and (114) we conclude that a sequence of $(T_n, R + \epsilon, D + \epsilon)$ code exists with $T_n = n\Delta$ and $T_n \rightarrow \infty$ as $n \rightarrow \infty$.

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 1 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

$$I(M; Y_0^T) = \mathbb{E} \left[\int_0^T \phi(\Gamma_t) dt \right] - T\phi(\lambda). \quad (118)$$

We also have

$$\frac{1}{T} I(M; Y_0^T) = \frac{1}{T} H(M) \leq \frac{1}{T} \log(\lceil \exp((R + \epsilon)T) \rceil) \quad (119)$$

$$\leq R + \epsilon + \frac{1}{T}. \quad (120)$$

This gives

$$R \geq \frac{1}{T} \mathbb{E} \left[\int_0^T \phi(\Gamma_t) dt \right] - \phi(\lambda) - \epsilon - \frac{1}{T}. \quad (121)$$

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

$$D \geq \frac{1}{T} \mathbb{E} \left[d(\hat{Y}_0^T, Y_0^T) \right] - \epsilon \quad (122)$$

$$= \frac{1}{T} \mathbb{E} \left[\int_0^T \hat{Y}_t dt - \int_0^T \log(\hat{Y}_t) dY_t \right] - \epsilon \quad (123)$$

$$= \frac{1}{T} \mathbb{E} \left[\int_0^T \hat{Y}_t - \Gamma_t \log(\hat{Y}_t) dt \right] - \epsilon \quad (124)$$

$$\geq \frac{1}{T} \mathbb{E} \left[\int_0^T \inf_{v \in \mathcal{A}} v - \Gamma_t \log(v) dt \right] - \epsilon \quad (125)$$

$$= \frac{1}{T} \mathbb{E} \left[\int_0^T \Psi_{\mathcal{A}}(\Gamma_t) dt \right] - \epsilon, \quad (126)$$

where, for (124) we have used Lemma 12; and for (126), 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

$$R \geq \mathbb{E}[\phi(\Gamma_S)] - \phi(\lambda) - \epsilon - \frac{1}{T} \quad (127)$$

$$D \geq \mathbb{E}[\Psi_{\mathcal{A}}(\Gamma_S)] - \epsilon. \quad (128)$$

Now we use Carathéodory's theorem [54, Th. 17.1]. There exist nonnegative $[\eta_k]_{k=1}^4$ and $[\alpha_k]_{k=1}^4$, such that $\sum_{k=1}^4 \alpha_k = 1$ and

$$\mathbb{E}[\phi(\Gamma_S)] = \sum_{k=1}^4 \alpha_k \phi(\eta_k) \quad (129)$$

$$\mathbb{E}[\Psi_{\mathcal{A}}(\Gamma_S)] = \sum_{k=1}^4 \alpha_k \Psi_{\mathcal{A}}(\eta_k) \quad (130)$$

$$\mathbb{E}[\Gamma_S] = \sum_{k=1}^4 \alpha_k \eta_k = \lambda, \quad (131)$$

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 , $\mathbb{E}[\int_0^T \Gamma_t dt] = \mathbb{E}[Y_T] = T\lambda$. Now define

$$\beta_k \triangleq \frac{\alpha_k \eta_k}{\lambda}. \quad (132)$$

We note that $\beta_k = 0$ if $\alpha_k = 0$, and $\sum_{k=1}^4 \beta_k = 1$. Substituting the above definitions in (127) and (128), we obtain

$$R \geq \left(\sum_{k=1}^4 \alpha_k \eta_k \log(\eta_k) - \lambda \log(\lambda) \right) - \epsilon - \frac{1}{T} \quad (133)$$

$$= \lambda \left(\sum_{k=1}^4 \beta_k \log\left(\frac{\beta_k \lambda}{\alpha_k}\right) \mathbf{1}_{\{\alpha_k > 0\}} - \log(\lambda) \right) - \epsilon - \frac{1}{T} \quad (134)$$

$$= \lambda \sum_{k=1}^4 \beta_k \log\left(\frac{\beta_k}{\alpha_k}\right) - \epsilon - \frac{1}{T}. \quad (135)$$

Likewise,

$$D \geq \sum_{k=1}^4 \alpha_k \Psi_{\mathcal{A}}\left(\frac{\lambda \beta_k}{\alpha_k}\right) - \epsilon. \quad (136)$$

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 place any restrictions on $\mathcal{A}$, i.e., if $\mathcal{A}$ is the set of all nonnegative reals, then we obtain the functional-covering distortion.

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

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

$$\Psi_{\mathcal{A}}(u) = \inf_{v \geq 0} v - u \log(v) = u - u \log(u). \quad (137)$$

For any achievable (R, D) we have

$$R \geq \lambda \sum_{k=1}^4 \beta_k \log\left(\frac{\beta_k}{\alpha_k}\right), \quad (138)$$

and

$$D \geq \sum_{k=1}^4 \alpha_k \Psi_{\mathcal{A}}\left(\frac{\lambda \beta_k}{\alpha_k}\right) \quad (139)$$

$$= \sum_{k=1}^4 \alpha_k \left(\frac{\lambda \beta_k}{\alpha_k} - \frac{\lambda \beta_k}{\alpha_k} \log\left(\frac{\lambda \beta_k}{\alpha_k}\right) \right) \quad (140)$$

$$= \lambda - \lambda \log(\lambda) - \lambda \sum_{k=1}^4 \beta_k \log\left(\frac{\beta_k}{\alpha_k}\right). \quad (141)$$

Hence

$$R + D \geq \lambda - \lambda \log(\lambda), \quad (142)$$

and this is achieved by $[\alpha_k]_{k=1}^4$ and $[\beta_k]_{k=1}^4$ that yield equality in (138). ■

If take $\mathcal{A} = \{0, 1\}$, then we recover the covering distortion in [35, Th. 1].

Corollary 3 (Covering Distortion [35]): 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

$$\Psi_{\mathcal{A}}(u) = \inf_{v \in \{0, 1\}} v - u \log(v) = \mathbf{1}\{u > 0\}. \quad (143)$$

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

$$D \geq \sum_{k=1}^4 \alpha_k \Psi_{\mathcal{A}}\left(\frac{\lambda \beta_k}{\alpha_k}\right) \quad (144)$$

$$= \sum_{k=1}^4 \alpha_k \mathbf{1}\{\beta_k > 0\} \quad (145)$$

$$= \sum_{k \in \mathcal{B}} \alpha_k, \quad (146)$$

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

$$R \geq \lambda \sum_{k=1}^4 \beta_k \log\left(\frac{\beta_k}{\alpha_k}\right) \quad (147)$$

$$= \lambda \sum_{k \in \mathcal{B}} \beta_k \log\left(\frac{\beta_k}{\alpha_k}\right) \quad (148)$$

$$\geq \lambda \left(\sum_{k \in \mathcal{B}} \beta_k \right) \log\left(\frac{\sum_{k \in \mathcal{B}} \beta_k}{\sum_{k \in \mathcal{B}} \alpha_k}\right) \quad (149)$$

$$= \lambda \log\left(\frac{1}{\sum_{k \in \mathcal{B}} \alpha_k}\right) \quad (150)$$

$$\geq (-\lambda \log(D))^+, \quad (151)$$

where (149) is due to the log-sum inequality, which can be achieved by setting $\alpha_1 = \min(1, D)$, $\alpha_2 = 1 - \alpha_1$, $\beta_1 = 1$, $\beta_2 = 0$. ■

V. THE POISSON CEO PROBLEM

We now consider the distributed problem shown in Fig. 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 $Y_0^{(i),T}$, $i \in \{1, 2\}$. Here 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 obtain $Y_0^{(i),T}$.

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

$$d(\hat{y}_0^T, y_0^T) = \int_0^T \hat{y}_t dt - \int_0^T \log(\hat{y}_t) dy_t. \quad (152)$$

Definition 13: A $(T, R^{(1)}, R^{(2)}, D)$ code for the Poisson CEO problem consists of encoders $f^{(1)}$ and $f^{(2)}$,

$$f^{(1)} : \mathcal{N}_0^T \rightarrow \left\{1, \dots, \left\lceil \exp(R^{(1)}T) \right\rceil\right\} \quad (153)$$

$$f^{(2)} : \mathcal{N}_0^T \rightarrow \left\{1, \dots, \left\lceil \exp(R^{(2)}T) \right\rceil\right\}, \quad (154)$$

and a decoder g ,

$$g : \left\{1, \dots, \left\lceil \exp(R^{(1)}T) \right\rceil\right\} \times \left\{1, \dots, \left\lceil \exp(R^{(2)}T) \right\rceil\right\} \rightarrow \hat{\mathcal{Y}}_0^T, \quad (155)$$

satisfying

$$\mathbb{E} \left[\int_0^T \hat{Y}_t dt \right] < \infty, \quad (156)$$

and the distortion constraint

$$\frac{1}{T} \mathbb{E} \left[d(\hat{Y}_0^T, Y_0^T) \right] \leq D. \quad (157)$$

Definition 14: 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 $(T_n, R^{(1)} + \epsilon, R^{(2)} + \epsilon, D + \epsilon)$ codes $T_n \rightarrow \infty$.

Definition 15: The *rate-distortion region* for the Poisson CEO problem $\mathfrak{RD}^{\mathcal{P}}$ is the intersection of all achievable rate-distortion vectors $(R^{(1)}, R^{(2)}, D)$.

The rate-distortion region for the Poisson CEO problem with feedforward, denoted by $\mathfrak{RD}_F^{\mathcal{P}}$, is defined analogously. Note that our formulation differs from that of Wang [47] in that it allows for thinning and we use the functional-covering distortion measure in place of the covering distortion measure.

Theorem 6: The rate-distortion region for the Poisson CEO problem is given by

$$\mathfrak{RD}^{\mathcal{P}} = \mathfrak{RD}_F^{\mathcal{P}} = \mathfrak{RD}, \quad (158)$$

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

$$R^{(1)} \geq \left((1 - p^{(1)})\lambda + \mu^{(1)} \right) \sum_{k=1}^4 \beta_k^{(1)} \log\left(\frac{\beta_k^{(1)}}{\alpha_k^{(1)}}\right) \quad (159)$$

$$R^{(2)} \geq \left((1 - p^{(2)})\lambda + \mu^{(2)} \right) \sum_{k=1}^4 \beta_k^{(2)} \log\left(\frac{\beta_k^{(2)}}{\alpha_k^{(2)}}\right) \quad (160)$$

$$D \geq \lambda - \phi(\lambda) \quad (161)$$

$$- \lambda \left(\sum_{k=1}^4 \gamma_k^{(1)} \log\left(\frac{\gamma_k^{(1)}}{\alpha_k^{(1)}}\right) + \sum_{k=1}^4 \gamma_k^{(2)} \log\left(\frac{\gamma_k^{(2)}}{\alpha_k^{(2)}}\right) \right) \quad (162)$$

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 $k \in \{1, 2, 3, 4\}$ and $i \in \{1, 2\}$

$$\gamma_k^{(i)} = p^{(i)} \alpha_k^{(i)} + (1 - p^{(i)}) \beta_k^{(i)} \quad \left. \begin{array}{l} \alpha_k^{(i)} = 0 \Rightarrow \beta_k^{(i)} = 0 \end{array} \right\} \quad \text{if } p^{(i)} < 1 \quad (163)$$

$$\alpha_k^{(i)} = \beta_k^{(i)} = \gamma_k^{(i)} \quad \text{if } p^{(i)} = 1. \quad (164)$$

Proof: Please see the supplementary material. ■

Remark 2: Note that 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 remaining $T/\Delta - \lambda T$ bits are zeroes. When such a sparse binary process is sent via two independent parallel channels as in (179)-(180), the resulting output processes are almost independent. 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 6 takes a simple form

$$\frac{\lambda}{\lambda + \mu^{(1)}} R^{(1)} + \frac{\lambda}{\lambda + \mu^{(2)}} R^{(2)} + D \geq \lambda - \phi(\lambda). \quad (165)$$

The above result should be compared with that of Wang [47, (37)] for the same problem but with the covering distortion measure.

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

$$R \geq ((1 - p)\lambda + \mu) \sum_{k=1}^4 \beta_k \log\left(\frac{\beta_k}{\alpha_k}\right) \quad (166)$$

$$D \geq \lambda - \phi(\lambda) - \lambda \cdot \sum_{k=1}^4 \gamma_k \log\left(\frac{\gamma_k}{\alpha_k}\right), \quad (167)$$

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

$$\gamma_k = p\alpha_k + (1 - p)\beta_k, \quad \alpha_k = 0 \Rightarrow \beta_k = 0. \quad (168)$$

ACKNOWLEDGMENT

The authors wish to thank the anonymous reviewers for their careful reading of the paper and their helpful comments.

REFERENCES

- [1] N. V. Shende and A. B. Wagner, "Functional covering of point processes," in *Proc. IEEE Int. Symp. Inf. Theory*, 2019, pp. 2039–2043, doi: [10.1109/ISIT.2019.8849723](https://doi.org/10.1109/ISIT.2019.8849723).
- [2] T. Berger, *Rate Distortion Theory: A Mathematical Basis for Data Compression*. Englewood Cliffs, NJ, USA: Prentice Hall, 1971.
- [3] D. H. Johnson, "Point process models of single-neuron discharges," *J. Comput. Neurosci.*, vol. 3, no. 4, pp. 275–299, 1996.
- [4] J. H. Goldwyn, J. T. Rubinstein, and E. Shea-Brown, "A point process framework for modeling electrical stimulation of the auditory nerve," *J. Neurophysiol.*, vol. 108, no. 5, pp. 1430–1452, 2012.
- [5] F. Farkhooi, M. F. Strube-Bloss, and M. P. Nawrot, "Serial correlation in neural spike trains: Experimental evidence, stochastic modeling, and single neuron variability," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 79, no. 2, 2009, Art. no. 21905.
- [6] S. V. Sarma et al., "Using point process models to compare neural spiking activity in the subthalamic nucleus of Parkinson's patients and a healthy primate," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 6, pp. 1297–1305, Jun. 2010.
- [7] E. N. Brown, R. E. Kass, and P. P. Mitra, "Multiple neural spike train data analysis: State-of-the-art and future challenges," *Nat. Neurosci.*, vol. 7, no. 5, pp. 456–461, 2004.
- [8] F. Rieke, D. Warland, R. De Ruyter Van Steveninck, and W. Bialek, *Spikes: Exploring the Neural Code*. Cambridge, MA, USA: MIT Press, 1997.
- [9] J. Giles and B. Hajek, "An information-theoretic and game-theoretic study of timing channels," *IEEE Trans. Inf. Theory*, vol. 48, no. 9, pp. 2455–2477, Sep. 2002.
- [10] M. Shahzad and A. X. Liu, "Accurate and efficient per-flow latency measurement without probing and time stamping," *IEEE/ACM Trans. Netw.*, vol. 24, no. 6, pp. 3477–3492, Dec. 2016.
- [11] Y. Zhu, X. Fu, B. Graham, R. Bettati, and W. Zhao, "On flow correlation attacks and countermeasures in mix networks," in *Proc. 4th Privacy Enhanc. Technol. Workshop (PET)*, 2004, pp. 207–225.
- [12] C. B. A. Börs, "A marked point process model for vehicle detection in aerial LiDAR point clouds," *ISPRS Ann. Photogrammetry, Remote Sens. Spatial Inf. Sci.*, vols. 1–3, Jul. 2012, pp. 93–98.
- [13] Y. Yu, J. Li, H. Guan, C. Wang, and M. Cheng, "A marked point process for automated tree detection from mobile laser scanning point cloud data," in *Proc. Int. Conf. Comp. Vis. Remote Sens.*, 2012, pp. 140–145.
- [14] S. Nakamoto, "A peer-to-peer electronic cash system." 2008. [Online]. Available: bitcoin.org/bitcoin
- [15] Y. Lewenberg, Y. Bachrach, Y. Sompolinsky, A. Zohar, and J. S. Rosenschein, "Bitcoin mining pools: A cooperative game theoretic analysis," in *Proc. Int. Conf. Auton. Agents Multiagent Sys.*, 2015, pp. 919–927.
- [16] Y. Kawase and S. Kasahara, "Transaction-confirmation time for bitcoin: A queueing analytical approach to blockchain mechanism," in *Proc. Int. Conf. Queueing Theory Netw. App.*, 2017, pp. 75–88.
- [17] C. Decker and R. Wattenhofer, "Information propagation in the bitcoin network," in *Proc. IEEE P2P*, 2013, pp. 1–10.
- [18] A. Laourine and A. B. Wagner, "Secrecy capacity of the degraded Poisson wiretap channel," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2010, pp. 2553–2557.
- [19] A. D. Wyner, "Capacity and error exponent for the direct detection photon channel part I," *IEEE Trans. Inf. Theory*, vol. 34, no. 6, pp. 1449–1461, Nov. 1988.
- [20] A. D. Wyner, "Capacity and error exponent for the direct detection photon channel part II," *IEEE Trans. Inf. Theory*, vol. 34, no. 6, pp. 1462–1471, Nov. 1988.
- [21] A. Lapidoth, "On the reliability function of the ideal Poisson channel with noiseless feedback," *IEEE Trans. Inf. Theory*, vol. 39, no. 2, pp. 491–503, Mar. 1993.
- [22] N. Shende and A. B. Wagner, "The stochastic-calculus approach to multiple-decoder Poisson channels," *IEEE Trans. Inf. Theory*, vol. 65, no. 8, pp. 5007–5027, Aug. 2019.
- [23] F. Baccelli and P. Brémaud, *Palm Probabilities and Stationary Queues*. Berlin, Germany: Springer-Verlag, 1987.
- [24] C. Sutardja and J. M. Rabaey, "Isolator-less near-field RFID reader for sub-cranial powering/data link of millimeter-sized implants," *IEEE J. Solid-State Circuits*, vol. 53, no. 7, pp. 2032–2042, Jul. 2018.
- [25] A. K. Skrivervik, A. J. M. Montes, I. V. Trivino, M. Bosiljevac, M. Veljovic, and Z. Sipus, "Antenna design for a cranial implant," in *Proc. Int. Workshop Antenna Technol. (iWAT)*, 2020, pp. 1–4.
- [26] R. L. de Queiroz and P. A. Chou, "Compression of 3D point clouds using a region-adaptive hierarchical transform," *IEEE Trans. Image Process.*, vol. 25, pp. 3947–3956, 2016.
- [27] T. Golla and R. Klein, "Real-time point cloud compression," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, 2015, pp. 5087–5092.
- [28] C. Tu, E. Takeuchi, C. Miyajima, and K. Takeda, "Compressing continuous point cloud data using image compression methods," in *Proc. IEEE 19th Int. Conf. Intell. Transp. Syst. (ITSC)*, 2016, pp. 1712–1719.

- [29] R. Gallager, "Basic limits on protocol information in data communication networks," *IEEE Trans. Inf. Theory*, vol. IT-22, no. 4, pp. 385–398, Jul. 1976.
- [30] A. S. Bedekar, "On the information about message arrival times required for in-order decoding," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2001, p. 227.
- [31] S. Verdú, "The exponential distribution in information theory," *Problemy Peredachi Informatsii*, vol. 32, no. 1, pp. 86–95, 1996.
- [32] T. P. Coleman, N. Kiyavash, and V. G. Subramanian, "The rate-distortion function of a Poisson process with a queueing distortion measure," in *Proc. Data Compress. Conf.*, Mar. 2008, pp. 63–72.
- [33] I. Rubin, "Information rates and data-compression schemes for Poisson processes," *IEEE Trans. Inf. Theory*, vol. IT-20, no. 2, pp. 200–210, Mar. 1974.
- [34] G. Koliander, D. Schuhmacher, and F. Hlawatsch, "Rate-distortion theory of finite point processes," *IEEE Trans. Inf. Theory*, vol. 64, no. 8, pp. 5832–5861, Aug. 2018.
- [35] A. Lapidoth, A. Malar, and L. Wang, "Covering point patterns," *IEEE Trans. Inf. Theory*, vol. 61, no. 9, pp. 4521–4533, Sep. 2015.
- [36] H.-A. Shen, S. M. Moser, and J.-P. Pfister, "Rate-distortion problems of the Poisson process: A group-theoretic approach," in *Proc. IEEE Inf. Theory Workshop (ITW)*, 2021, pp. 1–6.
- [37] T. A. Courtade and R. D. Wesel, "Multiterminal source coding with an entropy-based distortion measure," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2011, pp. 2040–2044.
- [38] T. A. Courtade and T. Weissman, "Multiterminal source coding under logarithmic loss," *IEEE Trans. Inf. Theory*, vol. 60, no. 1, pp. 740–761, Jan. 2014.
- [39] P. Brémaud, *Point Processes and Queues: Martingale Dynamics*. New York, NY, USA: Springer-Verlag, 1981.
- [40] H.-J. Ebeid and T. P. Coleman, "Source coding with feedforward using the posterior matching scheme," in *Proc. IEEE Int. Symp. Inf. Theory*, 2010, pp. 96–100.
- [41] N. Merhav and T. Weissman, "Coding for the feedback Gel'fand-Pinsker channel and the feedforward Wyner-Ziv source," *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 4207–4211, Sep. 2006.
- [42] S. S. Pradhan, "On the role of feedforward in Gaussian sources: Point-to-point source coding and multiple description source coding," *IEEE Trans. Inf. Theory*, vol. 53, no. 1, pp. 331–349, Jan. 2007.
- [43] R. Venkataramanan and S. S. Pradhan, "Source coding with feedforward: Rate-distortion theorems and error exponents for a general source," *IEEE Trans. Inf. Theory*, vol. 53, no. 6, pp. 2154–2179, Jun. 2007.
- [44] R. Venkataramanan and S. S. Pradhan, "Achievable rates for multiple descriptions with feed-forward," *IEEE Trans. Inf. Theory*, vol. 57, no. 4, pp. 2270–2277, Apr. 2011.
- [45] I. Naiss and H. H. Permuter, "Computable bounds for rate distortion with feed forward for stationary and ergodic sources," *IEEE Trans. Inf. Theory*, vol. 59, no. 2, pp. 760–781, Feb. 2013.
- [46] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Hoboken, NJ, USA: Wiley, 2006.
- [47] L. Wang, "A strong data processing inequality for thinning Poisson processes and some applications," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2017, pp. 3180–3184.
- [48] Y. Kabanov, "The capacity of a channel of the Poisson type," *Theory Probab. Appl.*, vol. 23, no. 1, pp. 143–147, 1978.
- [49] M. Davis, "Capacity and cutoff rate for Poisson-type channels," *IEEE Trans. Inf. Theory*, vol. IT-26, no. 6, pp. 710–715, Nov. 1980.
- [50] N. V. Shende and A. B. Wagner, "The stochastic-calculus approach to multi-receiver Poisson channels," *IEEE Trans. Inf. Theory*, vol. 65, no. 8, pp. 5007–5027, Aug. 2019.
- [51] R. Sundaresan and S. Verdú, "Capacity of queues via point-process channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 6, pp. 2697–2709, Jun. 2006.
- [52] P. Billingsley, *Convergence of Probability Measures* (Wiley Series in Probability and Statistics), 2nd ed. Somerset, NJ, USA: Wiley, 1999.
- [53] R. G. Gallager, *Information Theory and Reliable Communication*. New York, NY, USA: Wiley, 1968.
- [54] R. Rockafellar, *Convex Analysis*. Princeton, NJ, USA: Princeton Univ. Press, 1997.