

Covering Point Patterns

Amos Lapidoth, Andreas Malär, and Ligong Wang

Abstract—A source generates a point pattern consisting of a finite number of points in an interval. Based on a binary description of the point pattern, a reconstructor must produce a covering set that is guaranteed to contain the pattern. We study the optimal trade-off (as the length of the interval tends to infinity) between the description length and the least average Lebesgue measure of the covering set. The trade-off is established for point patterns that are generated by homogeneous and inhomogeneous Poisson processes. The homogeneous Poisson process is shown to be the most difficult to describe among all point patterns.

We also study a Wyner-Ziv version of this problem, where some of the points in the pattern are revealed to the reconstructor but not to the encoder. We show that this scenario is as good as when they are revealed to both encoder and reconstructor. A connection between this problem and the queueing distortion is established via feedforward. Finally, we establish the aforementioned trade-off when the covering set is allowed to miss some of the points in the pattern at a certain cost.

Index Terms—Poisson process, rate-distortion problem, side information, Wyner-Ziv problem, feedforward.

I. INTRODUCTION

Imagine a controller that receives a request that a computer be on at certain epochs. If the controller could describe these epochs to the computer with infinite precision, then the computer would turn itself on only at these epochs and be in sleep mode at all other times. If the controller cannot describe the epochs at all, then the computer must be on all the time. In this paper we study the trade-off between the bit rate with which the epochs can be described and the percentage of time the computer must be on.

More specifically, we consider a source that generates a “point pattern” consisting of a finite number of points in the interval $[0, T]$. Based on a binary description of the pattern, a reconstructor must produce a “covering set”: a subset of $[0, T]$ containing all the points. There is a trade-off between the description length and the minimal Lebesgue measure of the covering set. This trade-off is formulated as a continuous-time rate-distortion problem in Section II. In this paper we investigate this trade-off in the limit where T tends to infinity.

For point patterns that are generated by a homogeneous Poisson process of intensity λ , we show that, for the reconstructor to produce covering sets of average measure not

exceeding DT , the required description rate in bits per second is $-\lambda \log D$ [1]. This result is closely related to results on the capacity of the ideal peak-limited Poisson channel [2]–[5]. In fact, in the spirit of [6], the two problems may be considered dual, although this duality is not exactly the same as the one formulated in [6]. Also, [6] considers discrete-time channels and sources, whereas our problem is in continuous time.

Rate-distortion problems for Poisson processes under different distortion measures were studied in [7]–[11]. It is interesting that our rate-distortion function, $-\lambda \log D$, is equal to the one in [11], where a queueing distortion measure was considered. This is no coincidence: the connection between our and the queueing distortion measure is established when we introduce *feedforward* to the problem in Section VI. This connection is similar to the connection between the Poisson channel and the queueing channel [12].

One nice feature of our distortion measure is that it can be naturally extended to any measure space on which a Poisson process can be defined, whereas the previously-studied distortion measures [7]–[11] only apply to Poisson processes on the real line.

We also derive a formula for the required description rate when the point pattern is generated by an inhomogeneous Poisson process. More generally, we show that the homogeneous Poisson process is the most difficult to cover, in the sense that any point process that, with high probability, has no more than λT points in $[0, T]$ can be described with $-\lambda \log D$ bits per second. This is true even if an adversary selects an arbitrary point pattern, provided that the encoder and the reconstructor are allowed to use random codes. Here we note that, after the appearance of [1], a stronger result was shown in [13] that *deterministic* codes of the same rate are sufficient to cover arbitrary point patterns.

We then consider a Wyner-Ziv setting [14] of the problem, where some points in the pattern are revealed to the reconstructor but the encoder does not know which ones. This can be viewed as a dual problem to the Poisson channel with noncausal side information [15]. We show that in this setting one can achieve the same minimum rate as when the transmitter *does* know the reconstructor’s side information.

Finally we consider the case where the covering set is allowed to miss some of the points in the pattern, and where every missed point adds a certain cost to the overall distortion. The exact expression for the rate-distortion function is established for the homogeneous Poisson process. This problem was also considered in [13] in a perhaps more natural setting.

A. Notation

We use lower-case letters like x to denote numbers, and upper-case letters like X to denote random variables. Exceptions to this rule include T for time, R for rate, and D

This work was presented in part at the 2011 IEEE International Symposium of Information Theory (ISIT) in Saint Petersburg, Russia.

A. Lapidoth is with ETH Zurich, Switzerland (e-mail: lapidoth@isi.ee.ethz.ch).

A. Malär was with the Signal and Information Processing Laboratory, ETH Zurich, Switzerland. He is now with Malcom AG, Zurich, Switzerland (e-mail: andreas@malcom.ch).

L. Wang is with ETIS, CNRS UMR 8051, ENSEA, Université de Cergy-Pontoise, France (e-mail: ligong.wang@ensea.fr). Part of this work was conducted while L. Wang was with the Signal and Information Processing Laboratory, ETH Zurich, Switzerland.

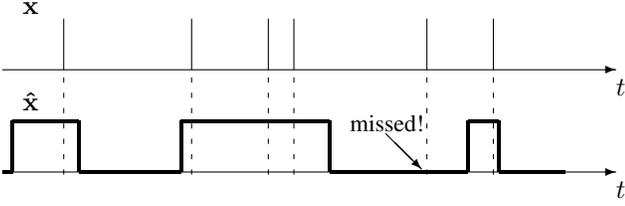


Fig. 1. Illustration of the problem. Here $\hat{\mathbf{x}}$ is not allowed because it does not cover all the points.

for expected distortion, which are all deterministic. Boldface lower-case letters like \mathbf{x} denote vectors, functions from the reals, or point patterns, depending on the context. If \mathbf{x} is a vector, x_i denotes its i th element. If \mathbf{x} is a function, $x(t)$ denotes its value at $t \in \mathbb{R}$. And if \mathbf{x} is a point pattern, $n_{\mathbf{x}}(\cdot)$ denotes its counting function, with $n_{\mathbf{x}}(t_2) - n_{\mathbf{x}}(t_1)$ being the number of points in \mathbf{x} that fall in the interval $(t_1, t_2]$. Boldface upper-case letters like \mathbf{X} denote random vectors, random functions, or random point processes. The random counting function corresponding to a point process \mathbf{X} is denoted $N_{\mathbf{X}}(\cdot)$.

We denote by $\text{Ber}(p)$ the Bernoulli distribution of parameter p , which assigns probability p to the outcome 1 and probability $(1 - p)$ to the outcome 0.

Unless stated otherwise, we use binary logarithm and measure information in bits throughout this paper.

The rest of this paper is arranged as follows: in Sections II and III we present our results for homogeneous and inhomogeneous Poisson processes, respectively; in Section IV we present our results for general point processes and arbitrary point patterns; in Section V we discuss the Wyner-Ziv setting; in Section VI we introduce feedforward and demonstrate a coding scheme based on existing work on the queueing distortion; in Section VII we present our result for the case where missing a point is allowed but at some cost; and in Section VIII we conclude the paper with some remarks.

II. HOMOGENEOUS POISSON PROCESSES

Consider a homogeneous Poisson process \mathbf{X} of intensity λ on the interval $[0, T]$. Its counting function $N_{\mathbf{X}}(\cdot)$ satisfies

$$\Pr [N_{\mathbf{X}}(t + \tau) - N_{\mathbf{X}}(t) = k] = \frac{e^{-\lambda\tau} (\lambda\tau)^k}{k!} \quad (1)$$

for all $\tau \in [0, T]$, $t \in [0, T - \tau]$ and $k \in \{0, 1, \dots\}$.

The encoder maps the realization of the Poisson process to a message in $\{1, \dots, 2^{TR}\}$, where R is the description rate in bits per second. The reconstructor then maps this message to a $\{0, 1\}$ -valued, Lebesgue-measurable, signal $\hat{x}(t)$, $t \in [0, T]$. We wish to minimize the length of the region where $\hat{x}(t) = 1$ while guaranteeing that all points in the original Poisson process lie in this region. See Figure 1 for an illustration.

More formally, we formulate this problem as a continuous-time rate-distortion problem, where the distortion between the point pattern \mathbf{x} and the reproduction signal $\hat{\mathbf{x}}$ is defined as

$$d(\mathbf{x}, \hat{\mathbf{x}}) \triangleq \begin{cases} \frac{\mu_{\mathbb{L}}(\hat{x}^{-1}(1))}{T}, & \text{if all points in } \mathbf{x} \text{ are in } \hat{x}^{-1}(1) \\ \infty, & \text{otherwise,} \end{cases} \quad (2)$$

where $\hat{x}^{-1}(1)$ is the set of all $t \in [0, T]$ such that $\hat{x}(t) = 1$, and where $\mu_{\mathbb{L}}(\cdot)$ denotes the Lebesgue measure.

We say that (R, D) is an achievable rate-distortion pair for \mathbf{X} if, for every $\epsilon > 0$, there exists some $T_0 > 0$ such that, for every $T > T_0$, there exist an encoder $f_T(\cdot)$ and a reconstructor $\phi_T(\cdot)$ of rate $R + \epsilon$ bits per second that, when applied to \mathbf{X} on $[0, T]$, result in

$$\mathbb{E} [d(\mathbf{X}, \phi_T(f_T(\mathbf{X})))] \leq D + \epsilon. \quad (3)$$

Denote by $R(D, \lambda)$ the minimal rate R such that (R, D) is achievable for the Poisson process of intensity λ . Define

$$R_{\text{Pois}}(D, \lambda) \triangleq \begin{cases} -\lambda \log D \text{ bits per second,} & D \in (0, 1) \\ 0, & D \geq 1. \end{cases} \quad (4)$$

Theorem 1 (Homogeneous Poisson). *For all $D, \lambda > 0$,*

$$R(D, \lambda) = R_{\text{Pois}}(D, \lambda). \quad (5)$$

Note: Theorem 1 can be alternatively expressed as

$$D(R, \lambda) = 2^{-R/\lambda}, \quad R, \lambda > 0, \quad (6)$$

where $D(R, \lambda)$ denotes the minimal expected distortion D such that (R, D) is achievable for the Poisson process of intensity λ .

A. Proof of Theorem 1 via Discretization

To prove Theorem 1, we propose a scheme to reduce the original problem to one for a discrete memoryless source. This is reminiscent of Wyner's scheme for reducing the peak-limited Poisson channel to a discrete memoryless channel [4]. We shall show the optimality of this scheme in Lemma 1, and we shall then prove Theorem 1 by computing the best rate that is achievable using this scheme.

Scheme 1. *We divide the time interval $[0, T]$ into T/Δ slots¹ of duration Δ . The encoder first maps the original point pattern \mathbf{x} to a $\{0, 1\}$ -valued vector \mathbf{x}^Δ of (T/Δ) components in the following way: if \mathbf{x} has at least one point in the slot $((i-1)\Delta, i\Delta]$, then we set the i th component of \mathbf{x}^Δ to 1. Otherwise, we set it to zero. The encoder then maps \mathbf{x}^Δ to a message in $\{1, \dots, 2^{TR}\}$.*

Based on the encoder's message, the reconstructor produces a $\{0, 1\}$ -valued length- (T/Δ) vector $\hat{\mathbf{x}}^\Delta$ that meets the distortion criterion

$$\mathbb{E} [d^\Delta(\mathbf{X}^\Delta, \hat{\mathbf{X}}^\Delta)] \leq D + \epsilon, \quad (7)$$

¹If T is not divisible by Δ , we replace it with $T' = \lceil T/\Delta \rceil \Delta$. When Δ tends to zero, the difference between RT and RT' tends to zero. Consequently, we shall ignore this edge effect and assume that T is divisible by Δ .

where the distortion measure $d^\Delta(\cdot, \cdot)$ between vectors is defined in terms of the single-letter distortion function

$$d^\Delta(0, 0) = 0 \quad (8a)$$

$$d^\Delta(0, 1) = 1 \quad (8b)$$

$$d^\Delta(1, 0) = \infty \quad (8c)$$

$$d^\Delta(1, 1) = 1. \quad (8d)$$

It then maps $\hat{\mathbf{x}}^\Delta$ to the piecewise-constant continuous-time signal $\hat{\mathbf{x}}$

$$\hat{x}(t) = \hat{x}_{\lceil \frac{t}{\Delta} \rceil}^\Delta, \quad t \in [0, T]. \quad (9)$$

Scheme 1 reduces the task of designing a code for \mathbf{X} subject to the distortion $d(\cdot, \cdot)$ to the task of designing a code for the vector \mathbf{X}^Δ subject to the distortion $d^\Delta(\cdot, \cdot)$ because

$$d(\mathbf{x}, \hat{\mathbf{x}}) = d^\Delta(\mathbf{x}^\Delta, \hat{\mathbf{x}}^\Delta). \quad (10)$$

When \mathbf{X} is a Poisson process of intensity λ , the components of \mathbf{X}^Δ are independent and identically distributed (IID), and each is $\text{Ber}(1 - e^{-\lambda\Delta})$. Let $R_\Delta(D, \lambda)$ denote the rate-distortion function for \mathbf{X}^Δ and $d^\Delta(\cdot, \cdot)$. If we combine Scheme 1 with an optimal code for \mathbf{X}^Δ subject to $\mathbb{E}[d^\Delta(\mathbf{X}^\Delta, \hat{\mathbf{X}}^\Delta)] \leq D + \epsilon$, we can achieve any rate that is larger than

$$\frac{R_\Delta(D, \lambda) \text{ bits}}{\Delta \text{ seconds}}. \quad (11)$$

The next lemma, which is reminiscent of [5, Theorem 2.1], shows that when we let Δ tend to zero, there is no loss in optimality in using Scheme 1.

Lemma 1. For all $D, \lambda > 0$,

$$R(D, \lambda) = \lim_{\Delta \downarrow 0} \frac{R_\Delta(D, \lambda)}{\Delta}. \quad (12)$$

Proof: Clearly, for all $\Delta > 0$, (11) cannot exceed $R(D, \lambda)$. To prove (12) we only need to show

$$R(D, \lambda) \leq \lim_{\Delta \downarrow 0} \frac{R_\Delta(D, \lambda)}{\Delta}. \quad (13)$$

We prove this in the following way: Given any rate-distortion code with 2^{TR} codewords $\hat{\mathbf{x}}_m$, $m \in \{1, \dots, 2^{TR}\}$ that achieves expected distortion D , we shall construct a new code that can be constructed through Scheme 1, that contains $(2^{TR} + 1)$ codewords, and that achieves an expected distortion that is arbitrarily close to D .

Denote the codewords of our new code by $\hat{\mathbf{w}}_m$, where $m \in \{1, \dots, 2^{TR} + 1\}$. We choose the last codeword to be the constant 1. We next describe our choices of the other codewords. For every $\epsilon > 0$ and every $\hat{\mathbf{x}}_m$, we can approximate the set $\{t: \hat{x}_m(t) = 1\}$ by a set \mathcal{A}_m that is equal to a finite, say N_m , union of open intervals. More specifically,

$$\mu_L(\hat{x}_m^{-1}(1) \triangle \mathcal{A}_m) \leq 2^{-TR}\epsilon, \quad (14)$$

where \triangle denotes the symmetric difference between two sets (see, e.g., [16, Chapter 3, Proposition 15]). Define

$$\mathcal{B} \triangleq \bigcup_{m=1}^{2^{TR}} (\hat{x}_m^{-1}(1) \setminus \mathcal{A}_m), \quad (15)$$

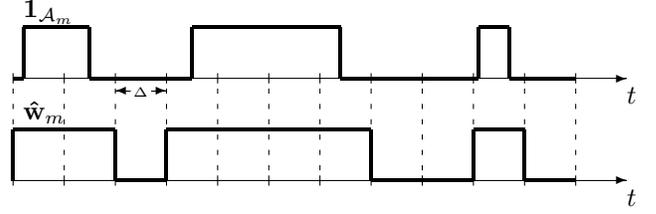


Fig. 2. Constructing $\hat{\mathbf{w}}_m$ from \mathcal{A}_m .

and note that by (14)

$$\mu_L(\mathcal{B}) \leq \epsilon. \quad (16)$$

For each \mathcal{A}_m , $m \in \{1, \dots, 2^{TR}\}$, define

$$\mathcal{T}_m \triangleq \{t \in [0, T]: ((\lceil t/\Delta \rceil - 1)\Delta, \lceil t/\Delta \rceil \Delta) \cap \mathcal{A}_m \neq \emptyset\}. \quad (17)$$

We now construct $\hat{\mathbf{w}}_m$, $m \in \{1, \dots, 2^{TR}\}$ as

$$\hat{\mathbf{w}}_m = \mathbf{1}_{\mathcal{T}_m}, \quad (18)$$

where $\mathbf{1}_S$ denotes the indicator function of the set S . Note that $\mathcal{A}_m \subseteq \mathcal{T}_m = \hat{w}_m^{-1}(1)$. See Figure 2 for an illustration of this construction. Let

$$N \triangleq \max_{m \in \{1, \dots, 2^{TR}\}} N_m.$$

It can be seen that

$$\mu_L(\hat{w}_m^{-1}(1)) - \mu_L(\mathcal{A}_m) \leq 2N\Delta, \quad m \in \{1, \dots, 2^{TR}\}. \quad (19)$$

Our encoder works as follows: if \mathbf{x} contains no point in \mathcal{B} , it maps \mathbf{x} to the same message as the given encoder; otherwise it maps \mathbf{x} to the index $(2^{TR} + 1)$ of the all-one codeword.

To analyze the distortion, first consider the case where \mathbf{x} contains no point in \mathcal{B} . In this case, all points in \mathbf{x} must be covered by the selected codeword $\hat{\mathbf{w}}_m$. By (14) and (19), the difference $d(\mathbf{x}, \hat{\mathbf{w}}_m) - d(\mathbf{x}, \hat{\mathbf{x}}_m)$, if positive, can be made arbitrarily small by choosing small ϵ and Δ (independently of \mathbf{x}). Next consider the case where \mathbf{x} does contain points in \mathcal{B} . By (16), the probability that this happens can be made arbitrarily small by choosing ϵ small, therefore its contribution to the expected distortion can also be made arbitrarily small. We conclude that our code $\{\hat{\mathbf{w}}_m\}$ can achieve a distortion that is arbitrarily close to the distortion achieved by the original code $\{\hat{\mathbf{x}}_m\}$. ■

We next proceed to prove Theorem 1. We derive $R(D, \lambda)$ by computing the right-hand side of (12) using [17, Theorem 9.3.2], which is a generalization of Shannon's formula for the rate-distortion function of a discrete memoryless source [18] to unbounded distortion measures. This gives us

$$R_\Delta(D, \lambda) = \min_{\substack{P_{\hat{X}^\Delta | X^\Delta}: \\ \mathbb{E}[d^\Delta(X^\Delta, \hat{X}^\Delta)] \leq D}} I(X^\Delta; \hat{X}^\Delta). \quad (20)$$

First consider the case where $D \in (0, 1)$. It is clear that, to satisfy $\mathbb{E}[d^\Delta(X^\Delta, \hat{X}^\Delta)] \leq D$, one must choose $\hat{X}^\Delta = 1$ with probability one whenever $X^\Delta = 1$. Assume that Δ is small enough so that

$$1 - e^{-\lambda\Delta} < D. \quad (21)$$

Then the conditional probability for $\hat{X}^\Delta = 1$ when $X^\Delta = 0$ should be chosen such that the probability $\hat{X}^\Delta = 1$ is equal to D . Hence the optimal conditional distribution is

$$P_{\hat{X}^\Delta|X^\Delta}^*(1|0) = 1 - e^{-\lambda\Delta} + De^{-\lambda\Delta}, \quad (22a)$$

$$P_{\hat{X}^\Delta|X^\Delta}^*(1|1) = 1. \quad (22b)$$

We compute the mutual information $I(X^\Delta; \hat{X}^\Delta)$ corresponding to this conditional law as follows

$$R_\Delta(D, \lambda) = I(X^\Delta; \hat{X}^\Delta) \quad (23)$$

$$= H(\hat{X}^\Delta) - H(\hat{X}^\Delta|X^\Delta) \quad (24)$$

$$= H(\hat{X}^\Delta) - e^{-\lambda\Delta} H(\hat{X}^\Delta|X^\Delta = 0) - (1 - e^{-\lambda\Delta}) H(\hat{X}^\Delta|X^\Delta = 1) \quad (25)$$

$$= H_b(D) - e^{-\lambda\Delta} H_b(1 - e^{-\lambda\Delta} + De^{-\lambda\Delta}), \quad (26)$$

where $H_b(\cdot)$ denotes the binary entropy function:

$$H_b(p) = p \log \frac{1}{p} + (1 - p) \log \frac{1}{1 - p}, \quad p \in [0, 1]. \quad (27)$$

Using (12) and (26) we have

$$R(D, \lambda) = \left. \frac{\partial (H_b(D) - e^{-\lambda\Delta} H_b(1 - e^{-\lambda\Delta} + De^{-\lambda\Delta}))}{\partial \Delta} \right|_{\Delta=0} \quad (28)$$

$$= \lambda e^{-\lambda\Delta} H_b(1 - e^{-\lambda\Delta} + De^{-\lambda\Delta}) + \lambda e^{-\lambda\Delta} (1 - D) \log \frac{e^{\lambda\Delta} - De^{\lambda\Delta}}{1 - e^{\lambda\Delta} + De^{\lambda\Delta}} \Big|_{\Delta=0} \quad (29)$$

$$= \lambda H_b(D) + \lambda(1 - D) \log \frac{1 - D}{D} \quad (30)$$

$$= -\lambda \log D, \quad D \in (0, 1). \quad (31)$$

The case where $D \geq 1$ is simple: one can use the constant signal $\hat{X}(t) = 1, t \in [0, T]$ as the reconstruction irrespective of the realization of \mathbf{X} , hence

$$R(D, \lambda) = 0, \quad D \geq 1. \quad (32)$$

Combining (31) and (32) yields (5).

B. Alternative Proof of Converse

In the above proof of Theorem 1, the converse is based on Lemma 1, i.e., on the optimality of the discretization approach. We next provide an alternative proof of the converse, which does not rely on this optimality. One advantage of this proof over the previous one is that it can be applied to general measure spaces; see Section II-C.

We shall show that, for any positive ε and δ , for large enough T , any code that achieves expected distortion $D < 1$ must have rate at least

$$R \geq (1 - \delta)(\lambda - \varepsilon) \log \frac{1 - \delta}{D}. \quad (33)$$

To this end, let $K \triangleq N_{\mathbf{X}}(T)$ be the total number of points in \mathbf{X} in the interval $[0, T]$, and let $M \in \{1, \dots, 2^{TR}\}$ be the label of the chosen codeword. Note that, assuming the expected distortion is finite (i.e., no arrival is missed), the distortion $d(\mathbf{x}, \hat{\mathbf{x}})$ is uniquely determined by the realization of

$M = m$: it is the Lebesgue measure of the m th reconstruction signal divided by T , which we denote d_m .

Conditional $K = k$, the k arrivals in \mathbf{x} , when randomly labeled, are IID uniformly on $[0, T]$; see, e.g., [19, Theorem 2.4.6]. It then follows that the probability that the m th reconstruction signal covers all these k points is d_m^k , which implies

$$P_{M|K}(m|k) \leq d_m^k, \quad (34)$$

i.e.,

$$k \log \frac{1}{d_m} \leq \log \frac{1}{P_{M|K}(m|k)}. \quad (35)$$

The inequality is because the m th codeword may cover \mathbf{x} and still not be picked if another codeword of smaller distortion also covers \mathbf{x} . Averaging over the message m we have

$$k \sum_{m=1}^{2^{TR}} P_{M|K}(m|k) \log \frac{1}{d_m} \leq H(M|K = k) \leq TR. \quad (36)$$

Because the function $a \mapsto \log(1/a)$ is convex, we further have by (36) and Jensen's inequality

$$TR \geq k \log \frac{1}{\mathbb{E} \left[d(\mathbf{X}, \hat{\mathbf{X}}) \mid K = k \right]}, \quad (37)$$

for all $k = 0, 1, \dots$. Averaging over k yields

$$TR \geq \sum_{k=0}^{\infty} P_K(k) k \log \frac{1}{\mathbb{E} \left[d(\mathbf{X}, \hat{\mathbf{X}}) \mid K = k \right]}. \quad (38)$$

Let \mathcal{L} be the event $\{K \geq (\lambda - \varepsilon)T\}$. Since K has the Poisson distribution of mean λT , we know that, for large enough T , the probability of \mathcal{L} is at least $1 - \delta$. This combined with the fact that $d(\mathbf{x}, \hat{\mathbf{x}})$ is nonnegative implies

$$\mathbb{E} \left[d(\mathbf{X}, \hat{\mathbf{X}}) \mid \mathcal{L} \right] \leq \frac{\mathbb{E} \left[d(\mathbf{X}, \hat{\mathbf{X}}) \right]}{1 - \delta} = \frac{D}{1 - \delta}. \quad (39)$$

Continuing from (38) we have

$$TR \geq \sum_{k \geq (\lambda - \varepsilon)T} P_K(k) k \log \frac{1}{\mathbb{E} \left[d(\mathbf{X}, \hat{\mathbf{X}}) \mid K = k \right]} \quad (40)$$

$$\geq (\lambda - \varepsilon)T \sum_{k \geq (\lambda - \varepsilon)T} P_K(k) \log \frac{1}{\mathbb{E} \left[d(\mathbf{X}, \hat{\mathbf{X}}) \mid K = k \right]} \quad (41)$$

$$\geq (1 - \delta)(\lambda - \varepsilon)T \mathbb{E} \left[\log \frac{1}{\mathbb{E} \left[d(\mathbf{X}, \hat{\mathbf{X}}) \mid K \right]} \mid \mathcal{L} \right] \quad (42)$$

$$\geq (1 - \delta)(\lambda - \varepsilon)T \log \frac{1}{\mathbb{E} \left[d(\mathbf{X}, \hat{\mathbf{X}}) \mid \mathcal{L} \right]} \quad (43)$$

$$\geq (1 - \delta)(\lambda - \varepsilon)T \log \frac{1 - \delta}{D}, \quad (44)$$

where (42) follows because the probability of \mathcal{L} is at least $1 - \delta$; (43) follows from the convexity of $a \mapsto \log(1/a)$ and Jensen's inequality; and (44) follows from (39). This establishes (33).

C. Fixed-Interval Formulation and General Measure Spaces

We now introduce an alternative formulation for Theorem 1. In this formulation, instead of letting the length of the interval tend to infinity, we keep this interval fixed and let the *intensity* go to infinity: the source \mathbf{X} is a Poisson process of intensity $T \cdot \lambda$ on the interval $[0, 1]$, where we let T tend to infinity. The distortion is now defined as

$$d(\mathbf{x}, \hat{\mathbf{x}}) \triangleq \begin{cases} \mu_L(\hat{x}^{-1}(1)), & \text{if all points in } \mathbf{x} \text{ are in } \hat{x}^{-1}(1) \\ \infty, & \text{otherwise.} \end{cases} \quad (45)$$

A rate-distortion pair (R, D) is said to be achievable if, for large enough T , there exists a codebook of size 2^{TR} with expected distortion arbitrarily close to D , and $R(D, \lambda)$ is the smallest R such that (R, D) is achievable. It is easy to verify that $R(D, \lambda)$ still satisfies (5), and that the proofs in Sections II-A and II-B are both valid (after minor adjustments).

The above formulation allows us to extend Theorem 1 to general measure spaces (instead of an interval on the real line with the Lebesgue measure), as well as to inhomogeneous Poisson processes. We next discuss general measure spaces; inhomogeneous Poisson processes will be discussed in the Section III.

Consider any measurable space on a set \mathcal{S} equipped with a measure $\mu(\cdot)$ that satisfies the following:

- It is non-atomic: any measurable $\mathcal{A} \subseteq \mathcal{S}$ such that $\mu(\mathcal{A}) > 0$ contains a measurable subset $\mathcal{B} \subset \mathcal{A}$ such that $0 < \mu(\mathcal{B}) < \mu(\mathcal{A})$.
- The measure of \mathcal{S} itself is finite: $\mu(\mathcal{S}) < \infty$. Without loss of generality, let $\mu(\mathcal{S}) = 1$.

Then there exist Poisson processes on \mathcal{S} with $\mu(\cdot)$ as their “mean measure”; see, e.g., [20]. Specifically, for any $\lambda T > 0$, there exists a Poisson process \mathbf{X} on \mathcal{S} such that the number of arrivals in \mathbf{X} in any measurable $\mathcal{A} \subseteq \mathcal{S}$ has the Poisson distribution of mean $\lambda T \mu(\mathcal{A})$. We can hence formulate a rate-distortion problem by replacing the interval $[0, 1]$ by \mathcal{S} , and by replacing the Lebesgue measure $\mu_L(\cdot)$ by $\mu(\cdot)$.

It is easy to see that Theorem 1 holds on such measure spaces. First, note that Scheme 1 is applicable because $\mu(\cdot)$ is non-atomic, hence the achievability part of Theorem 1 on \mathcal{S} can be proven exactly as in Section II-A. The optimality of this discretization approach in Lemma 1, however, relies on certain properties of the Lebesgue measure on an interval, and is difficult to extend to \mathcal{S} . But the converse proof in Section II-B is applicable to a homogeneous Poisson process on \mathcal{S} .

The simplest examples for \mathcal{S} are multi-dimensional Euclidean spaces.

III. INHOMOGENEOUS POISSON PROCESSES

In this section we consider the case where \mathbf{X} is an inhomogeneous Poisson process. As in Section II-C, we fix the time interval and let the intensity of the Poisson process tend to infinity. Let \mathbf{X} be a Poisson process on the interval $[0, 1]$ of intensity $T \cdot \lambda(t)$, $t \in [0, 1]$, where $\lambda(\cdot)$ is Lebesgue measurable, nonnegative, and bounded on $[0, 1]$, and where

$T > 0$ is a scaling factor. Hence the counting function $N_{\mathbf{X}}(\cdot)$ satisfies

$$\Pr [N_{\mathbf{X}}(t + \tau) - N_{\mathbf{X}}(t) = k] = \frac{e^{-\alpha} \alpha^k}{k!} \quad (46)$$

for all $\tau \in [0, 1]$, $t \in [0, 1 - \tau]$ and $k \in \{0, 1, \dots\}$, where

$$\alpha = T \int_t^{t+\tau} \lambda(s) ds. \quad (47)$$

We define the rate-distortion function $R(D, \lambda)$ in the same way as in Section II-C.

Theorem 2 (Inhomogeneous Poisson). *The rate-distortion function $R(D, \lambda)$ is given by*

$$R(D, \lambda) = \min_{\substack{\bar{D}: [0, 1] \rightarrow [0, 1] \\ \int_0^1 \bar{D}(t) dt \leq D}} \int_0^1 \lambda(t) \log \frac{1}{\bar{D}(t)} dt. \quad (48)$$

Moreover, if

$$\sup_t \lambda(t) \leq \frac{\bar{\lambda}}{D}, \quad (49)$$

where $\bar{\lambda} \triangleq \int_0^1 \lambda(t) dt$, then (48) simplifies to

$$R(D, \lambda) = \bar{\lambda} \log \frac{1}{D} + \bar{\lambda} \log \bar{\lambda} - \int_0^1 \lambda(t) \log \lambda(t) dt. \quad (50)$$

Note that, unlike most previous works on rate-distortion problems with memory [21], Theorem 2 does not require that the source be stationary or ergodic.

The proof of Theorem 2 is divided into three parts: the achievability part of (48), the converse part of (48), and the proof of (50).

A. Achievability of (48)

As in Section II-A, the coding scheme we propose is based on discretization, but now the length of each interval depends on T and is chosen to be (Δ/T) . The resulting discrete-time source \mathbf{X}^Δ is a sequence of independent but *not* identically distributed binary random variables: X_i^Δ is $\text{Ber}(1 - e^{-\lambda_i \Delta})$ where

$$\lambda_i \triangleq \frac{T}{\Delta} \int_{(i-1)\Delta/T}^{i\Delta/T} \lambda(t) dt. \quad (51)$$

The discrete-time distortion function is again given by (8).

Fix a measurable function $\tilde{D}(t)$, $t \in [0, 1]$. Our discrete-time codebook is generated as follows. We first generate 2^{TR} codewords (covering sets) independently, where in each codeword, the i th symbol is chosen to be 1 with probability

$$D_i \triangleq \frac{T}{\Delta} \int_{(i-1)\Delta/T}^{i\Delta/T} \tilde{D}(t) dt \quad (52)$$

independently of other symbols in the codeword. We then append the all-one sequence to the codebook, so the total size of the codebook is $2^{TR} + 1$.

The encoder maps \mathbf{x}^Δ to the codeword with the smallest number of ones that covers \mathbf{x}^Δ .

We next analyze the performance of this random codebook. We consider three different cases. The first is the case where \mathbf{x}^Δ does *not* satisfy

$$\sum_{i=1}^{T/\Delta} x_i^\Delta \log \frac{1}{D_i} \leq \Delta \sum_{i=1}^{T/\Delta} \lambda_i \log \frac{1}{D_i} + \epsilon T \quad (53)$$

where ϵ is a fixed positive number. We show that the probability that \mathbf{X}^Δ does not satisfy (53) tends to zero as T tends to infinity. Indeed, we have

$$\mathbb{E} \left[\sum_{i=1}^{T/\Delta} X_i^\Delta \log \frac{1}{D_i} \right] = \sum_{i=1}^{T/\Delta} (1 - e^{-\lambda_i \Delta/T}) \log \frac{1}{D_i} \quad (54)$$

$$\leq \Delta \sum_{i=1}^{T/\Delta} \lambda_i \log \frac{1}{D_i} \quad (55)$$

whereas

$$\text{var} \left(\sum_{i=1}^{T/\Delta} X_i \log \frac{1}{D_i} \right) = \sum_{i=1}^{T/\Delta} \text{var}(X_i) \left(\log \frac{1}{D_i} \right)^2 \quad (56)$$

$$\leq \sum_{i=1}^{T/\Delta} \mathbb{E}[X_i^2] \left(\log \frac{1}{D_i} \right)^2 \quad (57)$$

$$= \sum_{i=1}^{T/\Delta} (1 - e^{-\lambda_i \Delta/T}) \left(\log \frac{1}{D_i} \right)^2 \quad (58)$$

$$\leq \Delta \sum_{i=1}^{T/\Delta} \lambda_i \left(\log \frac{1}{D_i} \right)^2. \quad (59)$$

The expressions on the right-hand side of (55) and (59) both tend to infinity linearly with T as T tends to infinity. It then follows by Chebyshev's inequality that the probability that \mathbf{X}^Δ does not satisfy (53) tends to zero as T tends to infinity. Since the distortion for any \mathbf{x}^Δ is at most one, we conclude that the total contribution to the expected distortion from this first case can be neglected.

The second case we consider is where \mathbf{x}^Δ does satisfy (53) but no covering set except the all-one set covers \mathbf{x}^Δ . According to our construction, the probability π that a randomly chosen covering set covers \mathbf{x}^Δ is

$$\pi \triangleq \prod_{j: x_j^\Delta=1} D_j \quad (60)$$

$$= \prod_{i=1}^{T/\Delta} (D_i)^{x_i^\Delta} \quad (61)$$

$$= 2^{-\sum_{i=1}^{T/\Delta} x_i^\Delta \log \frac{1}{D_i}} \quad (62)$$

$$\geq 2^{-\Delta \sum_{i=1}^{T/\Delta} \lambda_i \log \frac{1}{D_i} + \epsilon T} \quad (63)$$

where the last inequality follows because we are now considering the case where (53) does hold. Since the first 2^{TR} covering sets are chosen IID, the probability that none of them covers

\mathbf{x}^Δ is $(1 - \pi)^{2^{TR}}$, which tends to zero as T tends to infinity provided that

$$R > \lim_{T \rightarrow \infty} \frac{\Delta}{T} \sum_{i=1}^{T/\Delta} \lambda_i \log \frac{1}{D_i} + \epsilon \quad (64)$$

$$= \int_0^1 \lambda(t) \log \frac{1}{\bar{D}(t)} dt + \epsilon \quad (65)$$

where the last step follows from, e.g., [22, Theorem 8.8]. Hence, provided that R satisfies (65), the contribution to the expected distortion from this second case can also be neglected.

The last case is where \mathbf{x}^Δ satisfies (53) and where one of the first 2^{TR} randomly-generated codewords covers \mathbf{x}^Δ . As we have shown, when R satisfies (65), the overall expected distortion is dominated by the expected distortion in this last case. Fix any \mathbf{x}^Δ satisfying (53). Recalling that the encoder selects the smallest covering set that covers \mathbf{x}^Δ , and since the minimum is upper-bounded by the average, the expected distortion associated with \mathbf{x}^Δ and the randomly chosen codebook, conditional on the chosen covering set not being the all-one set, is upper-bounded by the expected size of a randomly generated covering set conditional on it covering \mathbf{x}^Δ . This upper bound is easy to compute: conditional on a covering set $\hat{\mathbf{X}}^\Delta$ that was randomly chosen as above covering \mathbf{x}^Δ , its symbols \hat{X}_i^Δ are independently distributed according to

$$P_{\hat{X}_i^\Delta}(1) = \begin{cases} 1, & x_i^\Delta = 1 \\ D_i, & x_i^\Delta = 0. \end{cases} \quad (66)$$

The conditional expectation of the distortion associated with a random covering set conditional on it covering \mathbf{x}^Δ is

$$\frac{\Delta}{T} \sum_{i=1}^{T/\Delta} \mathbb{E}[\hat{X}_i^\Delta] = \frac{\Delta}{T} \left(\sum_{i=1}^{T/\Delta} D_i (1 - x_i^\Delta) + \sum_{i=1}^{T/\Delta} x_i^\Delta \right) \quad (67)$$

$$\leq \frac{\Delta}{T} \left(\sum_{i=1}^{T/\Delta} D_i + \sum_{i=1}^{T/\Delta} x_i^\Delta \right) \quad (68)$$

$$= \int_0^1 \tilde{D}(t) dt + \frac{\Delta}{T} \sum_{i=1}^{T/\Delta} x_i^\Delta. \quad (69)$$

Hence the expected distortion in the last case is upper bounded by

$$\int_0^1 \tilde{D}(t) dt + \frac{\Delta}{T} \sum_{i=1}^{T/\Delta} \mathbb{E}[X_i^\Delta | \mathbf{X}^\Delta \text{ satisfies (53)}]. \quad (70)$$

The second term in (70) vanishes as Δ tends to zero because

$$\sum_{i=1}^{T/\Delta} \mathbb{E}[X_i^\Delta] \leq T \int_0^1 \lambda(t) dt \quad (71)$$

for all $\Delta > 0$, and because the probability that \mathbf{X}^Δ satisfies (53) is bounded away from zero for all fixed ϵ . We have hence proved that, if R satisfies (65), then the expected distortion for our randomly generated codebook can be made arbitrarily close to $\int_0^1 \tilde{D}(t) dt$. It then follows that there must exist a deterministic codebook that achieves this distortion, and the achievability part of (48) is established.

B. Converse of (48)

We next prove the converse part of (48). To this end, first note that the discretization proposed in Section III-A—when combined with an optimal discrete-time code—is optimal as Δ tends to zero. Indeed, one can easily verify that the proof of Lemma 1 is still valid in the inhomogeneous case.

Let \mathbf{X}^Δ be the discretized source sequence and $\hat{\mathbf{X}}^\Delta$ be the reconstruction sequence. Suppose there exists a code of size 2^{TR} that achieves expected distortion D . Because \mathbf{X}^Δ , the label of the chosen codeword, and $\hat{\mathbf{X}}^\Delta$ form a Markov chain,

$$I(\mathbf{X}^\Delta; \hat{\mathbf{X}}^\Delta) \leq TR. \quad (72)$$

Since the components of \mathbf{X}^Δ are independent, this further implies

$$\sum_{i=1}^{T/\Delta} I(X_i^\Delta; \hat{X}_i^\Delta) \leq TR, \quad (73)$$

i.e.,

$$\frac{1}{T} \sum_{i=1}^{T/\Delta} I(X_i^\Delta; \hat{X}_i^\Delta) \leq R. \quad (74)$$

To establish the converse of (48) we shall show that, for all $\Delta > 0$, the left-hand side of (74) is lower-bounded by the right-hand side of (48) whenever

$$\frac{\Delta}{T} \sum_{i=1}^{T/\Delta} d^\Delta(X_i^\Delta, \hat{X}_i^\Delta) \leq D. \quad (75)$$

Consider any $i \in \{1, \dots, T/\Delta\}$. Recall that $X_i^\Delta \sim \text{Ber}(1 - e^{-\lambda_i \Delta})$, where λ_i is given in (51). Since \hat{X}_i^Δ is $\{0, 1\}$ -valued, it is $\text{Ber}(D_i)$ for some $D_i \in [0, 1]$. Then we must have the following conditional probabilities

$$P_{\hat{X}_i^\Delta | X_i^\Delta}(1|1) = 1 \quad (76a)$$

$$P_{\hat{X}_i^\Delta | X_i^\Delta}(1|0) = 1 - e^{-\lambda_i \Delta} + D_i e^{-\lambda_i \Delta}. \quad (76b)$$

These imply

$$\begin{aligned} I(X_i^\Delta; \hat{X}_i^\Delta) \\ = H_b(D_i) - e^{-\lambda_i \Delta} H_b(1 - e^{-\lambda_i \Delta} + D_i e^{-\lambda_i \Delta}). \end{aligned} \quad (77)$$

The right-hand side of (77) equals zero when $\Delta = 0$. Its first derivative with respect to Δ is

$$\lambda_i e^{-\lambda_i \Delta} \log \frac{1}{1 - e^{-\lambda_i \Delta} + D_i e^{-\lambda_i \Delta}}$$

which, at $\Delta = 0$, equals

$$\lambda_i \log \frac{1}{D_i}.$$

Its second derivative with respect to Δ equals

$$\begin{aligned} \lambda_i^2 e^{-\lambda_i \Delta} \log e \\ \cdot \left(\frac{e^{-\lambda_i \Delta} - D_i e^{-\lambda_i \Delta}}{1 - e^{-\lambda_i \Delta} + D_i e^{-\lambda_i \Delta}} - \ln \frac{1}{1 - e^{-\lambda_i \Delta} + D_i e^{-\lambda_i \Delta}} \right), \end{aligned}$$

which is positive for all $\Delta > 0$. Using Taylor's theorem with Cauchy remainder, from (77) we obtain

$$I(X_i^\Delta; \hat{X}_i^\Delta) \geq \lambda_i \Delta \log \frac{1}{D_i}. \quad (78)$$

Summing over i and dividing by T we obtain

$$\frac{1}{T} \sum_{i=1}^{T/\Delta} I(X_i^\Delta; \hat{X}_i^\Delta) \geq \frac{1}{T} \sum_{i=1}^{T/\Delta} \lambda_i \Delta \log \frac{1}{D_i}. \quad (79)$$

Recalling (51), and noting that, to achieve expected distortion D one must have

$$\frac{\Delta}{T} \sum_{i=1}^{T/\Delta} D_i \leq D, \quad (80)$$

we conclude that the right-hand side of (79) is lower-bounded by the right-hand side of (48). Since this is true for all $\Delta > 0$, and since the discretization approach is optimal as Δ tends to zero, the converse part of (48) is established.

C. Proof of (50)

For every $\tilde{D}: [0, 1] \rightarrow \mathbb{R}_0^+$,

$$\begin{aligned} \int_0^1 \lambda(t) \log \frac{1}{\tilde{D}(t)} dt \\ = \bar{\lambda} \cdot \int_0^1 \frac{\lambda(t)}{\bar{\lambda}} \cdot \log \frac{\lambda(t)/\bar{\lambda}}{\tilde{D}(t)/\bar{D}} dt + \bar{\lambda} \log \frac{1}{\bar{D}} \\ + \bar{\lambda} \log \bar{\lambda} - \int_0^1 \lambda(t) \log \lambda(t) dt \end{aligned} \quad (81)$$

where

$$\bar{D} \triangleq \int_0^1 \tilde{D}(t) dt. \quad (82)$$

Such a \tilde{D} is feasible in (48) if $\tilde{D}(t) \leq 1$, $t \in [0, 1]$, and $\bar{D} \leq D$. Both $\lambda(t)/\bar{\lambda}$ and $\tilde{D}(t)/\bar{D}$ are probability density functions on $[0, 1]$, so the first term on the right-hand side of (81) is a relative entropy, which is nonnegative and equals zero if and only if

$$\frac{\lambda(t)}{\bar{\lambda}} = \frac{\tilde{D}(t)}{\bar{D}} \quad \text{almost everywhere.} \quad (83)$$

The first two terms on the right-hand side of (81) will thus be minimized if (83) holds and $\bar{D} = D$, i.e., if $\tilde{D}(t) = D \cdot \lambda(t)/\bar{\lambda}$ for all $t \in [0, 1]$. This choice satisfies $\bar{D} \leq 1$ if (49) holds, and leads to the right-hand side of (50).

IV. GENERAL PROCESSES AND ARBITRARY PATTERNS

By (48), when $D \in (0, 1)$,

$$R(D, \boldsymbol{\lambda}) \leq \int_0^1 \lambda(t) \log \frac{1}{D} dt \quad (84)$$

$$= \bar{\lambda} \log \frac{1}{D} \quad (85)$$

$$= R(D, \bar{\lambda}). \quad (86)$$

where (86) follows from (4). Thus, among all Poisson processes of the same average intensity, the homogeneous Poisson process is the most difficult to describe: it requires the largest rate to achieve the same distortion. As we next show, the homogeneous Poisson process is the most difficult to describe among *all* point processes, not only among Poisson processes.

In this section we switch back to the original formulation where the source is a point pattern on $[0, T]$ where T tends to infinity.

Consider a point process \mathbf{Y} for which there exists some λ such that

$$\lim_{t \rightarrow \infty} \Pr \left[\frac{N_{\mathbf{Y}}(t)}{t} > \lambda + \delta \right] = 0 \quad \text{for all } \delta > 0. \quad (87)$$

For example, \mathbf{Y} could be an ergodic process whose expected number of points per second is less than or equal to λ .

Theorem 3 (General processes). *The pair $(R_{\text{Pois}}(D, \lambda), D)$ is achievable on any point process satisfying (87).*

Before proving Theorem 3, we state a stronger result. Suppose that a point pattern \mathbf{z} is generated by an adversary with the only constraint that it be in the interval $[0, T]$ and that it contain no more than λT points. The corresponding counting function $n_{\mathbf{z}}(\cdot)$ must hence satisfy

$$n_{\mathbf{z}}(T) \leq \lambda T. \quad (88)$$

Suppose further that the encoder and the reconstructor are allowed to use random codes. That is, they fix a distribution on all (deterministic) codes of a given rate on $[0, T]$, and they use this distribution to generate a code, which is not revealed to the adversary. They then apply it to the point pattern \mathbf{z} chosen by the adversary. We say that (R, D) is achievable with random coding against an adversary subject to (88) if, for every $\epsilon > 0$, there exists some T_0 such that, for every $T > T_0$, there exists a random code on $[0, T]$ of rate $R + \epsilon$ such that the expected distortion between *any* \mathbf{z} respecting (88) and its reconstruction is smaller than $D + \epsilon$.

Theorem 4 (Arbitrary patterns). *The pair $(R_{\text{Pois}}(D, \lambda), D)$ is achievable with random coding against any adversary respecting (88).*

Proof: When $D \geq 1$ the result is obvious because the encoder does not need to describe the pattern: the reconstructor simply produces the all-one function, yielding distortion 1 for any \mathbf{z} .

Next consider $D \in (0, 1)$. We use Scheme 1 of Section II-A to reduce the problem to one of random coding for an arbitrary discrete-time sequence \mathbf{z}^Δ . Here the vector \mathbf{z}^Δ is $\{0, 1\}$ -valued, has (T/Δ) components, and satisfies

$$\sum_{i=1}^{T/\Delta} z_i^\Delta \leq \lambda T. \quad (89)$$

We shall construct a random code of rate (R/Δ) which, when applied to any \mathbf{z}^Δ satisfying (89), yields

$$\mathbb{E} \left[d^\Delta(\mathbf{z}^\Delta, \hat{\mathbf{z}}^\Delta) \right] < D + \epsilon, \quad (90)$$

where the random vector $\hat{\mathbf{z}}^\Delta$ is the result of applying the random encoder and decoder to \mathbf{z}^Δ . Combined with Scheme 1 this random code will yield a random code for the continuous-time point pattern \mathbf{z} that achieves the rate-distortion pair (R, D) .

Our discrete-time random code consists of 2^{TR} $\{0, 1\}$ -valued, length- (T/Δ) random sequences $\hat{\mathbf{z}}_m^\Delta$, $m \in \{1, \dots, 2^{TR}\}$. The first sequence $\hat{\mathbf{z}}_1^\Delta$ is chosen deterministically to be the all-one sequence. The other $2^{TR} - 1$

sequences are drawn independently, with the components of each sequence drawn IID $\text{Ber}(D)$.

To describe a source sequence \mathbf{z}^Δ , the encoder looks for a codeword $\hat{\mathbf{z}}_m^\Delta$, $m \in \{2, \dots, 2^{TR}\}$ such that

$$\hat{z}_{m,i}^\Delta = 1 \text{ whenever } z_i^\Delta = 1. \quad (91)$$

If it finds one or more such codewords, it sends the index of the first one; otherwise it sends the index 1. The reconstructor produces the sequence $\hat{\mathbf{z}}_m^\Delta$, where m is the index it received from the encoder.

We next analyze the expected distortion of this random code for a fixed \mathbf{z}^Δ satisfying (89). Define

$$\kappa \triangleq \frac{1}{T} \sum_{i=1}^{T/\Delta} z_i^\Delta, \quad (92)$$

and note that by (89) $\kappa \leq \lambda$. Let \mathcal{E} be the event that the encoder cannot find $\hat{\mathbf{z}}_m^\Delta$, $m \in \{2, \dots, 2^{TR}\}$ satisfying (91). If \mathcal{E} occurs, the encoder produces the index 1, and the resulting distortion is 1. The probability that a randomly drawn codeword $\hat{\mathbf{z}}_m^\Delta$ satisfies (91) is

$$D^{\kappa T} \geq D^{\lambda T} = 2^{(\lambda \log D)T}. \quad (93)$$

Because the codewords $\hat{\mathbf{z}}_m^\Delta$, $m \in \{2, \dots, 2^{TR}\}$ are chosen independently, if $R > -\lambda \log D$, then $\Pr[\mathcal{E}] \rightarrow 0$ as $T \rightarrow \infty$. Hence, for large enough T , the contribution to the expected distortion from the event \mathcal{E} can be ignored.

We next analyze the expected distortion conditional on \mathcal{E}^c . The reproduction $\hat{\mathbf{z}}^\Delta$ has the following distribution: at positions where \mathbf{z}^Δ is 1, $\hat{\mathbf{z}}^\Delta$ must also be 1; at other positions the components of $\hat{\mathbf{z}}^\Delta$ are IID $\text{Ber}(D)$. (These components were not “looked at” in the process of generating the index.) Thus, the expected value of $\sum_{i=1}^{T/\Delta} \hat{z}_i^\Delta$ is $\kappa T + D(\frac{T}{\Delta} - \kappa T)$, and

$$\mathbb{E} \left[d^\Delta(\mathbf{z}^\Delta, \hat{\mathbf{z}}^\Delta) \mid \mathcal{E}^c \right] = D + (1 - D)\kappa\Delta, \quad (94)$$

which tends to D as Δ tend to zero. We have thus shown that, for small enough Δ , we can achieve the pair $(R/\Delta, D)$ on \mathbf{z}^Δ using random coding whenever $R > -\lambda \log D$. Consequently, if $R > -\lambda \log D$ then we can also use random coding to achieve (R, D) on the continuous-time point pattern \mathbf{z} . ■

We next use Theorem 4 to prove Theorem 3.

Proof of Theorem 3: It follows from Theorem 4 that, on any point process satisfying (87), the pair $(R_{\text{Pois}}(D, \lambda + \delta), D)$ is achievable with *random coding*. Further, since there is no adversary, the existence of a good random code guarantees the existence of a good deterministic code. Hence $(R_{\text{Pois}}(D, \lambda + \delta), D)$ is also achievable on this process with deterministic coding. Theorem 3 now follows when we let δ tend to zero, because $R_{\text{Pois}}(D, \cdot)$ is a continuous function. ■

V. SOME POINTS ARE REVEALED TO THE RECONSTRUCTOR

In this section we consider a Wyner-Ziv setting for our problem. We first consider the case where \mathbf{X} is a homogeneous Poisson process of intensity λ . (Later we consider an arbitrary point pattern.) Assume that the points in \mathbf{X} are revealed to the reconstructor independently, each with probability p .

Also assume that the encoder does not know which points are revealed to the reconstructor. The encoder maps \mathbf{X} to a message in $\{1, \dots, 2^{TR}\}$, and the reconstructor produces a Lebesgue-measurable, $\{0, 1\}$ -valued signal $\hat{\mathbf{X}}$ on $[0, T]$ based on this message and the positions of the points that it knows. The achievability of a rate-distortion pair is defined in the same way as in Section II. Denote the smallest rate R for which (R, D) is achievable by $R_{\text{WZ}}(D, \lambda, p)$.

Obviously, $R_{\text{WZ}}(D, \lambda, p)$ is lower-bounded by the smallest achievable rate when the transmitter *does* know which points are revealed to the reconstructor. The latter rate is given by $R_{\text{Pois}}(D, (1-p)\lambda)$, where $R_{\text{Pois}}(\cdot, \cdot)$ is given by (4). Indeed, when the encoder knows which points are revealed to the reconstructor, it is optimal for it to describe only the remaining points, which themselves form a Poisson process of intensity $(1-p)\lambda$. The reconstructor then selects a set based on this description to cover the points unknown to it and adds to this set the points it knows. Thus,

$$R_{\text{WZ}}(D, \lambda, p) \geq R_{\text{Pois}}(D, (1-p)\lambda). \quad (95)$$

The next theorem shows that (95) holds with equality.

Theorem 5 (Wyner-Ziv for Poisson). *Knowing the points at the reconstructor only is as good as knowing them also at the encoder:*

$$R_{\text{WZ}}(D, \lambda, p) = R_{\text{Pois}}(D, (1-p)\lambda). \quad (96)$$

To prove Theorem 5, it remains to show that the pair $(R_{\text{Pois}}(D, (1-p)\lambda), D)$ is achievable. We shall prove a stronger result concerning arbitrarily varying sources.

Consider an arbitrary point pattern \mathbf{z} on $[0, T]$ chosen by an adversary. The adversary is allowed to put at most λT points in \mathbf{z} . Also, it must reveal all but at most νT points to the reconstructor, without telling the encoder which points it has revealed. The encoder and the reconstructor are allowed to use random codes, where the encoder is a random mapping from \mathbf{z} to a message in $\{1, \dots, 2^{TR}\}$, and where the reconstructor is a random mapping from this message, together with the point pattern that it knows, to a $\{0, 1\}$ -valued, Lebesgue-measurable signal $\hat{\mathbf{z}}$. The distortion $d(\mathbf{z}, \hat{\mathbf{z}})$ is defined as in (2).

Theorem 6 (Wyner-Ziv for arbitrary patterns). *Against an adversary who puts at most λT points on $[0, T]$ and reveals all but at most νT points to the reconstructor, the rate-distortion pair $(R_{\text{Pois}}(D, \nu), D)$ is achievable with random coding.*

Proof: The case $D \geq 1$ is trivial, so we shall only consider the case where $D \in (0, 1)$. The encoder and the reconstructor first use Scheme 1 as in Section II to reduce the point pattern \mathbf{z} to a $\{0, 1\}$ -valued vector \mathbf{z}^Δ of length (T/Δ) . Define

$$\kappa \triangleq \frac{1}{T} \sum_{i=1}^{T/\Delta} z_i^\Delta, \quad (97)$$

and note that, by assumption, $\kappa \leq \lambda$. If $\kappa \leq \nu$, then we can ignore the reconstructor's side information and use the random code of Theorem 4, so we henceforth assume $\kappa > \nu$.

Denote by \mathbf{s} the point pattern revealed to the reconstructor and by \mathbf{z}^Δ the vector obtained from \mathbf{s} through the discretization

in time of Scheme 1. Since there are at most νT points that are unknown to the reconstructor,

$$\sum_{i=1}^{T/\Delta} s_i^\Delta \geq (\kappa - \nu)T. \quad (98)$$

The encoder first conveys the value of κT to the receiver. Since κT is an integer between 0 and λT , the number of bits per second needed to describe it tends to zero as T tends to infinity, and the overhead in describing κT is negligible.

Next, the encoder and the reconstructor randomly generate $2^{T(R+\tilde{R})}$ independent codewords

$$\hat{\mathbf{z}}_{m,\ell}^\Delta, \quad m \in \{1, \dots, 2^{TR}\}, \ell \in \{1, \dots, 2^{T\tilde{R}}\},$$

where each codeword is generated IID $\text{Ber}(D)$.

To describe \mathbf{z}^Δ , the encoder looks for a codeword $\hat{\mathbf{z}}_{m,\ell}^\Delta$ such that

$$\hat{z}_{m,\ell,i}^\Delta = 1 \text{ whenever } z_i^\Delta = 1. \quad (99)$$

If it finds one or more such codewords, it sends the index m of the first one; otherwise it tells the reconstructor to produce the all-one sequence.

When the reconstructor receives the index m , it looks for an index $\tilde{\ell} \in \{1, \dots, 2^{T\tilde{R}}\}$ such that

$$\hat{z}_{m,\tilde{\ell},i}^\Delta = 1 \text{ whenever } s_i^\Delta = 1. \quad (100)$$

If there is only one such codeword, it produces it as the reconstruction; if there are more than one such codewords, it produces the all-one sequence.

To analyze the expected distortion for \mathbf{z}^Δ over this random code, first consider the event that the encoder cannot find a codeword satisfying (99). The probability that a randomly generated codeword satisfies (99) is $D^{\kappa T}$, so the probability of this event tends to zero as T tends to infinity provided that

$$R + \tilde{R} > -\kappa \log D. \quad (101)$$

Next consider the event that the reconstructor finds more than one $\tilde{\ell}$ satisfying (100). The probability that a randomly generated codeword satisfies (100) is $D^{\sum_{i=1}^{T/\Delta} s_i^\Delta}$. Consequently, by (98) the probability of this event tends to zero as T tends to infinity provided that

$$\tilde{R} < -(\kappa - \nu) \log D. \quad (102)$$

Finally, if the encoder finds a codeword satisfying (99) and the reconstructor finds only one codeword satisfying (100), then the two codewords must be the same.² Following the same calculations as in the proof of Theorem 4, the expected distortion in this case tends to D as Δ tends to zero.

Combining (101) and (102), we can make the expected distortion arbitrarily close to D as $T \rightarrow \infty$ if

$$R > -\nu \log D. \quad (103)$$

Proof of Theorem 5: The claim follows from (95), Theorem 6, and the Law of Large Numbers. ■

²In the general Wyner-Ziv problem there may be an error event in which there is a codeword that is jointly typical with the state sequence but not with the source sequence. This error cannot occur under our coding scheme.

Theorems 5 and 6 are Wyner-Ziv versions of Theorems 1 (on homogeneous Poisson sources) and 4 (on arbitrary sources), respectively. In both cases, knowing some points only at the reconstructor is as good as knowing them also at the encoder. This is also true when the source is an inhomogeneous Poisson process, as in Theorem 2. The proof is a combination of the proofs of Theorems 2 and 5 and is omitted.

VI. FEEDFORWARD

In this section we consider the problem of covering a homogeneous Poisson process of intensity λ in the presence of instantaneous feedforward from the source to the reconstructor. In this setting, the encoder again maps the realization of the Poisson process to a message, but the reconstructor is different from the one in Section II: The value $\hat{x}(t)$ of the reconstructed signal at time $t \in [0, T]$ is allowed to depend not only on the message but also on the past of the source realization $x(\tau)$, $\tau \in [0, t)$. More formally, given the message from the encoder, $\hat{\mathbf{X}}$ is predictable with respect to the natural filtration associated with \mathbf{X} conditional on the message. The distortion is again given by (2).

It is shown in [23] that feedforward does not improve the rate-distortion function for a discrete memoryless source. This is also true in our setting. Indeed, one can show that in the presence of feedforward the discretization approach in Section II is still optimal in the limit where Δ tends to zero, therefore, by [23], the rate-distortion function remains the same. Our attention in this section is on how feedforward can simplify the reconstruction process. As we shall show, with high probability the covering set needs to be the union of at most $(\lambda + \epsilon)T$ intervals, with ϵ being an arbitrarily small positive number. In contrast, in the coding scheme without feedforward, which we proposed in Section II-A, the total number of disjoint intervals that constitute the covering set tends to infinity as Δ tends to zero.

Our feedforward coding scheme is based on a rate-distortion code for the Poisson process with the *queueing distortion* [11], which we describe below. Consider the high-probability event that \mathbf{X} contains no more than $(\lambda + \epsilon)T$ points. Conditional on this event, [11] guarantees the existence of a codebook of size $2^{T(R_{\text{Pois}}(D, \lambda) + \delta)}$ for some small δ , which consists of point patterns $\tilde{\mathbf{x}}$ each containing $(\lambda + \epsilon)T$ points, such that there exists a mapping from \mathbf{x} to $\tilde{\mathbf{x}}$ satisfying the following two conditions:

- 1) For every $i \in \{1, \dots, (\lambda + \epsilon)T\}$, with probability one the i th arrival in \mathbf{X} occurs after the i th arrival in $\tilde{\mathbf{X}}$.
- 2) Consider \mathbf{X} as the output of feeding $\tilde{\mathbf{X}}$ into a single-server queue, then the expected total service time is at most DT .

After the encoder sends the reconstructor the index for $\tilde{\mathbf{x}}$ according to such a codebook, with the help of instantaneous feedforward, the reconstructor can use the following scheme to produce the covering set $\hat{\mathbf{x}}$.

- At the beginning the reconstruction signal is off: $\hat{x}(0) = 0$ with probability one.

- If an arrival in $\tilde{\mathbf{x}}$ occurs and $\hat{\mathbf{x}}$ is off, then the reconstructor immediately switches on $\hat{\mathbf{x}}$.
- At every t where there is an arrival in the source process \mathbf{x} , the reconstructor compares the total number of points in $\tilde{\mathbf{x}}$ in $[0, t]$ with the total number of points in \mathbf{x} in $[0, t]$ (including the arrival at time t). If the former is larger than the latter, it keeps $\hat{\mathbf{x}}$ on; if the two numbers are equal, it switches $\hat{\mathbf{x}}$ off.

To analyze the performance of this scheme, we first note that 1) guarantees that $\hat{\mathbf{x}}$ be on whenever there is an arrival in the source \mathbf{x} . We further note that the total time during which $\hat{\mathbf{x}}$ is on is exactly the service time when \mathbf{x} is considered as the output of feeding $\tilde{\mathbf{x}}$ into a single-server queue. Therefore, 2) guarantees that the expected total time that $\hat{X}(t)$ is on is at most DT , and our desired distortion is achieved.

The above construction is very similar to the coding scheme in [12] for the Poisson channel with feedback, which is based on codes for the queueing channel.

VII. FINITE COST OF MISSING

So far in this paper we have assumed that the cost of missing an arrival in the point pattern is infinite. In this section we consider the problem where this cost is finite. We concentrate on the case where the point pattern is generated by a homogeneous Poisson process of intensity λ on the interval $[0, T]$. This problem is formulated more naturally in the context where the cost of being on and the cost of missing are considered as two separate criteria, as in [13]. Here, however, we consider a single distortion function that summarizes both costs.

Let the cost of missing one arrival be γ . Specifically, if the total number of points in \mathbf{x} that do not lie in $\hat{x}^{-1}(1)$ is ℓ , then the distortion is given by

$$d(\mathbf{x}, \hat{\mathbf{x}}) = \frac{\mu_{\mathbf{L}}(\hat{x}^{-1}(1)) + \ell\gamma}{T}. \quad (104)$$

The rate-distortion function $R(D, \lambda, \gamma)$ is then defined as in Section II.

Let $D_b(p \| q)$ denote the relative entropy between two Bernoulli distributions of parameters p and q , respectively:

$$D_b(p \| q) \triangleq p \log \frac{p}{q} + (1-p) \log \frac{1-p}{1-q}, \quad p, q \in [0, 1]. \quad (105)$$

The following theorem characterizes $R(D, \lambda, \gamma)$ for the distortion function (104).

Theorem 7 (Finite cost of missing). *The rate-distortion function $R(D, \lambda, \gamma)$ for a homogeneous Poisson process of intensity λ and distortion function (104) is given by*

$$R(D, \lambda, \gamma) = \begin{cases} \lambda \min_q D_b(q \| r(q)), & D < \min\{1, \lambda\gamma\} \\ 0, & D \geq \min\{1, \lambda\gamma\}, \end{cases} \quad (106)$$

where the minimum is over

$$q \in \left(\frac{\lambda\gamma - D}{\lambda\gamma}, 1 \right] \quad (107)$$

and $r(q)$ is given by

$$r(q) = D - \lambda\gamma(1 - q). \quad (108)$$

The intuition behind Theorem 7 is as follows. Let r be the expected per-second Lebesgue measure of the covering set, and let q be the probability that an arrival in \mathbf{X} is covered by the chosen covering set. The expected number of missed arrivals is then $\lambda T(1 - q)$, which contributes $\lambda\gamma(1 - q)$ to the overall distortion. The distortion coming from the Lebesgue measure of the covering set is r , so the overall expected distortion is $\lambda\gamma(1 - q) + r$, hence (108). Next consider the rate required for such a covering set to be found. Heuristically, the covering set can be thought of as drawn with its value at every $t \in [0, T]$ IID according to the Bernoulli distribution of parameter r . We wish to cover proportion q of the approximately λT arrivals in \mathbf{X} . By Sanov's theorem [24], [25], this probability is approximately $2^{-\lambda T D_b(q \| r)}$, hence the claimed expression for $R(D, \lambda, \gamma)$. The formal proof is given below.

Proof of Theorem 7: The case where $D \geq \min\{1, \lambda\gamma\}$ is simple: when $D \geq 1$ the reconstructor can produce the all-one signal, achieving expected distortion 1, and when $D \geq \lambda\gamma$ the reconstructor can produce the all-zero signal, achieving expected distortion $\lambda\gamma$. For the rest of the proof we thus assume $D < \min\{1, \lambda\gamma\}$.

We again apply Scheme 1 to reduce the problem to a discrete-time problem. The discrete-time distortion function (8) is now replaced by

$$d^\Delta(0, 0) = 0 \quad (109a)$$

$$d^\Delta(0, 1) = 1 \quad (109b)$$

$$d^\Delta(1, 0) = \frac{\lambda\Delta}{1 - e^{-\lambda\Delta}} \cdot \frac{\gamma}{\Delta} \quad (109c)$$

$$d^\Delta(1, 1) = 1. \quad (109d)$$

Here, $d^\Delta(1, 0)$ requires some explanation. Given $x^\Delta = 1$, there could be one or more arrivals in the corresponding length- Δ interval, hence the cost of missing in this interval could be γ , 2γ , etc. The expected number of arrivals in this interval, conditional on there being at least one arrival, is $\lambda\Delta/(1 - e^{-\lambda\Delta})$. Hence the right-hand side of (109c) is the *expected* distortion in the corresponding interval in the continuous-time case. It is *not* equal to the actual distortion in the continuous-time case. Since we are interested in the overall expected distortion, (109c) suffices. We assume that the discretization is fine enough so that

$$\Delta < \gamma, \quad (110)$$

which implies

$$\frac{\lambda\Delta}{1 - e^{-\lambda\Delta}} \cdot \frac{\gamma}{\Delta} > 1, \quad (111)$$

i.e.,

$$\lambda\gamma - 1 + e^{-\lambda\Delta} > 0. \quad (112)$$

It is easy to verify that Lemma 1 continues to hold in the following sense: if $R_\Delta(D, \lambda, \gamma)$ is the rate-distortion function of the discrete-time problem that results from applying Scheme 1 and the above distortion function, then

$$R(D, \lambda, \gamma) = \lim_{\Delta \downarrow 0} \frac{R_\Delta(D, \lambda, \gamma)}{\Delta}. \quad (113)$$

We next evaluate $R_\Delta(D, \lambda, \gamma)$. To this end, consider the following conditional distribution

$$P_{\hat{X}^\Delta | X^\Delta}(0|0) = 1 - p \quad (114a)$$

$$P_{\hat{X}^\Delta | X^\Delta}(1|0) = p \quad (114b)$$

$$P_{\hat{X}^\Delta | X^\Delta}(0|1) = 1 - q \quad (114c)$$

$$P_{\hat{X}^\Delta | X^\Delta}(1|1) = q. \quad (114d)$$

To meet the distortion criterion, we need

$$D \geq \mathbb{E} \left[d^\Delta(X^\Delta, \hat{X}^\Delta) \right] \quad (115)$$

$$\begin{aligned} &= P_X(0)P_{\hat{X}|X}(0|0)d^\Delta(0, 0) + P_X(0)P_{\hat{X}|X}(1|0)d^\Delta(0, 1) \\ &\quad + P_X(1)P_{\hat{X}|X}(0|1)d^\Delta(1, 0) \\ &\quad + P_X(1)P_{\hat{X}|X}(1|1)d^\Delta(1, 1) \end{aligned} \quad (116)$$

$$\begin{aligned} &= e^{-\lambda\Delta} \cdot (1 - p) \cdot 0 + e^{-\lambda\Delta} \cdot p \cdot 1 \\ &\quad + (1 - e^{-\lambda\Delta}) \cdot (1 - q) \cdot \frac{\lambda\Delta}{1 - e^{-\lambda\Delta}} \cdot \frac{\gamma}{\Delta} \\ &\quad + (1 - e^{-\lambda\Delta}) \cdot q \cdot 1 \end{aligned} \quad (117)$$

$$= \lambda\gamma + e^{-\lambda\Delta}p - (\lambda\gamma - 1 + e^{-\lambda\Delta})q. \quad (118)$$

Combined with (112), this implies that, to meet the distortion constraint, q must satisfy

$$q \geq \frac{\lambda\gamma - D + e^{-\lambda\Delta}p}{\lambda\gamma - 1 + e^{-\lambda\Delta}}. \quad (119)$$

Noting that we are considering the case $D < \{1, \lambda\gamma\}$ and that $p \leq 1$, we further obtain that q must satisfy

$$q \geq \frac{(\lambda\gamma - D + e^{-\lambda\Delta})p}{\lambda\gamma - 1 + e^{-\lambda\Delta}} \quad (120)$$

$$\geq p. \quad (121)$$

But $I(X^\Delta; \hat{X}^\Delta)$ for the conditional law (114) is monotonically decreasing in p for p in the interval $[0, q]$. So, for a fixed q , minimizing $I(X^\Delta; \hat{X}^\Delta)$ is achieved by maximizing p , and we should choose p to achieve (118) with equality:

$$p = e^{\lambda\Delta} ((\lambda\gamma - 1 + e^{-\lambda\Delta})q - \lambda\gamma + D), \quad (122)$$

where q must satisfy

$$q \geq \frac{\lambda\gamma - D}{\lambda\gamma - 1 + e^{-\lambda\Delta}}. \quad (123)$$

With $p = p(q, \Delta)$ chosen as in (122), the mutual information $I(X^\Delta; \hat{X}^\Delta)$ can be expressed as a function of q and Δ (and, of course, D , λ , and γ):

$$\begin{aligned} &I(X^\Delta; \hat{X}^\Delta) \\ &= H(\hat{X}^\Delta) - H(\hat{X}^\Delta | X^\Delta) \end{aligned} \quad (124)$$

$$\begin{aligned} &= H_b(e^{-\lambda\Delta}p(q, \Delta) + (1 - e^{-\lambda\Delta})q) \\ &\quad - e^{-\lambda\Delta}H_b(p(q, \Delta)) - (1 - e^{-\lambda\Delta})H_b(q) \end{aligned} \quad (125)$$

$$\triangleq f(q, \Delta). \quad (126)$$

Fixing any

$$q > \frac{\lambda\gamma - D}{\lambda\gamma}, \quad (127)$$

which satisfies (123) for small enough Δ , we have

$$\lim_{\Delta \downarrow 0} \frac{f(q, \Delta)}{\Delta} = \left. \frac{\partial f(q, \Delta)}{\partial \Delta} \right|_{\Delta=0} \quad (128)$$

$$= \log \frac{1-p(q, 0)}{p(q, 0)} \cdot \left. \frac{\partial (e^{-\lambda \Delta} p(q, \Delta) + (1-e^{-\lambda \Delta})q)}{\partial \Delta} \right|_{\Delta=0} + \lambda H_b(p(q, 0)) - \log \frac{1-p(q, 0)}{p(q, 0)} \cdot \left. \frac{\partial p(q, \Delta)}{\partial \Delta} \right|_{\Delta=0} - \lambda H_b(q) \quad (129)$$

$$= \log \frac{1-p(q, 0)}{p(q, 0)} \cdot \left(-\lambda p(q, 0) + \left. \frac{\partial p(q, \Delta)}{\partial \Delta} \right|_{\Delta=0} + \lambda q \right) + \lambda H_b(p(q, 0)) - \log \frac{1-p(q, 0)}{p(q, 0)} \cdot \left. \frac{\partial p(q, \Delta)}{\partial \Delta} \right|_{\Delta=0} - \lambda H_b(q) \quad (130)$$

$$= \lambda(q - p(q, 0)) \log \frac{1-p(q, 0)}{p(q, 0)} + \lambda H_b(p(q, 0)) - \lambda H_b(q) \quad (131)$$

$$= \lambda q \log \frac{1}{p(q, 0)} + \lambda(1-q) \log \frac{1}{1-p(q, 0)} - \lambda H_b(q) \quad (132)$$

$$= \lambda D_b(q \| p(q, 0)). \quad (133)$$

The proof is complete when we obtain from (122)

$$p(q, 0) = \lambda \gamma q - \lambda \gamma + D, \quad (134)$$

which is the same as $r(q)$ given in (108). ■

VIII. CONCLUDING REMARKS

This paper introduced a new quantization problem for point processes: to describe the points so as to allow the reconstructor to generate a covering set that is small in its Lebesgue measure and that is yet guaranteed to contain all the points. The asymptotic tension between the description length and the size of the covering set was quantified for Poisson processes in various settings. The problem in its various settings provides new insight into the relationship between point processes, Poisson channels, and queues. It is related, but not equivalent, to the continuous-alphabet version [26] of the problem of source coding with lists [27].

Unlike the queueing-motivated quantization problems, our problem does not require a notion of order for the points: it easily generalizes to Poisson random fields or Poisson processes on general nonatomic measure spaces. Moreover—unlike some of the other formulations that deal with the quantization of the inter-arrival times—it is of a purely continuous-time nature.

REFERENCES

- [1] A. Lapidoth, A. Malär, and L. Wang, “Covering point patterns,” in *Proc. IEEE Int. Symp. Inform. Theory*, Saint Petersburg, Russia, July 31–August 5 2011, pp. 149–153.
- [2] Y. Kabanov, “The capacity of a channel of the Poisson type,” *Theory of Probability and Its Appl.*, vol. 23, pp. 143–147, 1978.
- [3] M. H. A. Davis, “Capacity and cutoff rate for Poisson-type channels,” *IEEE Trans. Inform. Theory*, vol. 26, pp. 710–715, Nov. 1980.
- [4] A. D. Wyner, “Capacity and error exponent for the direct detection photon channel — part I,” *IEEE Trans. Inform. Theory*, vol. 34, no. 6, pp. 1449–1461, Nov. 1988.
- [5] —, “Capacity and error exponent for the direct detection photon channel — part II,” *IEEE Trans. Inform. Theory*, vol. 34, pp. 1462–1471, Nov. 1988.
- [6] T. M. Cover and M. Chiang, “Duality between channel capacity and rate distortion with two-sided state information,” *IEEE Trans. Inform. Theory*, vol. 48, no. 6, pp. 1629–1638, June 2002.
- [7] I. Rubin, “Information rates and data-compression schemes for Poisson processes,” *IEEE Trans. Inform. Theory*, vol. 20, no. 2, pp. 200–210, Mar. 1974.
- [8] R. G. Gallager, “Basic limits on protocol information in data communication networks,” *IEEE Trans. Inform. Theory*, vol. 22, no. 4, pp. 385–398, July 1976.
- [9] S. Verdú, “The exponential distribution in information theory,” *Problems of Information Transmission*, vol. 31, no. 1, pp. 86–95, 1996.
- [10] A. S. Bedekar, “On the information about message arrival times required for in-order decoding,” in *Proc. IEEE Int. Symp. Inform. Theory*, Washington DC, June 24–29, 2001.
- [11] T. P. Coleman, N. Kiyavash, and V. G. Subramanian, “The rate-distortion function of a Poisson process with a queueing distortion measure,” in *Proceedings Data Compression Conference 2008*, Cliff Lodge, Snowbird, Utah, USA, Mar. 2008.
- [12] V. Anantharam and S. Verdú, “Bits through queues,” *IEEE Trans. Inform. Theory*, vol. 42, no. 1, pp. 4–18, Jan. 1996.
- [13] A. Mazumdar and L. Wang, “Covering arbitrary point patterns,” in *Proc. 50th Allerton Conf. Comm., Contr. and Comp.*, Allerton H., Monticello, IL, October 1–5, 2012.
- [14] A. D. Wyner and J. Ziv, “The rate-distortion function for source coding with side information at the decoder,” *IEEE Trans. Inform. Theory*, vol. 22, no. 1, pp. 1–10, Jan. 1976.
- [15] S. I. Bross, A. Lapidoth, and L. Wang, “The Poisson channel with side information,” in *Proc. 47th Allerton Conf. Comm., Contr. and Comp.*, Allerton H., Monticello, IL, September 30–October 2, 2009.
- [16] H. L. Royden, *Real Analysis*, 3rd ed. Macmillan Publishing Company, 1988.
- [17] R. G. Gallager, *Information Theory and Reliable Communication*. John Wiley & Sons, 1968.
- [18] C. E. Shannon, “A mathematical theory of communication,” *Bell System Techn. J.*, vol. 27, pp. 379–423 and 623–656, July and Oct. 1948.
- [19] J. R. Norris, *Markov Chains*. Cambridge University Press, 1997.
- [20] J. Kingman, *Poisson Processes*. Oxford University Press, 1992.
- [21] T. Berger and J. D. Gibson, “Lossy source coding,” *IEEE Trans. Inform. Theory*, vol. 44, no. 6, pp. 2693–2723, Oct. 1998.
- [22] W. Rudin, *Real and Complex Analysis*, 2nd ed. McGraw-Hill, 1974.
- [23] T. Weissman and N. Merhav, “On competitive prediction and its relation to rate-distortion theory,” *IEEE Trans. Inform. Theory*, vol. 49, no. 12, pp. 3185–3194, Dec. 2003.
- [24] I. N. Sanov, “On the probability of large deviations of random variables,” *Mat. Sbornik*, vol. 42, pp. 11–44, 1957.
- [25] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. John Wiley & Sons, 1991.
- [26] C. Bunte and A. Lapidoth, “Rényi entropy and quantization for densities,” in *Proc. Inform. Theory Workshop (ITW)*, Hobart, Australia, Nov. 2–5, 2014.
- [27] —, “Encoding tasks and Rényi entropy,” *IEEE Trans. Inform. Theory*, vol. 60, no. 9, pp. 5065–5076, 2014.