Covering Point Patterns

Amos Lapidoth
ETH Zurich
8092 Zurich, Switzerland
amos.lapidoth@ethz.ch

Andreas Malär
ETH Zurich
8092 Zurich, Switzerland
amalaer@ee.ethz.ch

Ligong Wang
ETH Zurich
8092 Zurich, Switzerland
wang@isi.ee.ethz.ch

***Abstract*—An encoder observes a point pattern—a finite number of points in the interval $[0, T]$—which is to be described to a reconstructor using bits. Based on these bits, the reconstructor wishes to select a subset of $[0, T]$ that contains all the points in the pattern. It is shown that, if the point pattern is produced by a homogeneous Poisson process of intensity λ, and if the reconstructor is restricted to select a subset of average Lebesgue measure not exceeding DT, then, as T tends to infinity, the minimum number of bits per second needed by the encoder is $-\lambda \log D$. It is also shown that, as T tends to infinity, any point pattern on $[0, T]$ containing no more than λT points can be successfully described using $-\lambda \log D$ bits per second in this sense. Finally, a Wyner-Ziv version of this problem is considered where some of the points in the pattern are known to the reconstructor.**

I. Introduction

An encoder observes a point pattern—a finite number of points in the interval $[0, T]$—which is to be described to a reconstructor using bits. Based on these bits, the reconstructor wishes to produce a covering-set—a subset of $[0, T]$ containing all the points—of least Lebesgue measure. There is a trade-off between the number of bits used and the Lebesgue measure of the covering-set. This trade-off can be formulated as a continuous-time rate-distortion problem (Section III). In this paper we investigate this trade-off in the limit where $T \to \infty$.

When the point pattern is produced by a homogeneous Poisson process, this problem is closely related to that of transmitting information through an ideal peak-limited Poisson channel [1], [2], [3], [4]. In fact, the two problems can be considered dual in the sense of [5]. However, the duality results of [5] only apply to discrete memoryless channels and sources, so they cannot be directly used to solve our problem. Instead, we shall use a technique that is similar to Wyner's [3], [4] to find the desired rate-distortion function. We shall show that, if the point pattern is the outcome of a homogeneous Poisson process of intensity λ, and if the reconstructor is restricted to select covering-sets of average measure not exceeding DT, then the minimum number of bits per second needed by the encoder to describe the pattern is $-\lambda \log D$.

Previous works [6], [7] have studied rate-distortion functions of the Poisson process with different distortion measures. It is interesting to notice that our rate-distortion function, $-\lambda \log D$, is equal to the one in [7], where a queueing distortion measure is considered. This is no coincidence, since the Poisson channel is closely related to the queueing channel introduced in [8].

We also show that the 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 even true if an adversary selects the point pattern provided that the pattern contains no more than λ points per second and that the encoder and the reconstructor are allowed to use random codes.

Finally, we consider a Wyner-Ziv setting [9] of the problem where some points in the pattern are known to the reconstructor but the encoder does not know which ones they are. This can be viewed as a dual problem to the Poisson channel with noncausal side-information [10]. We show that in this setting one can achieve the same minimum rate as when the transmitter *does* know the reconstructor's side-information.

The rest of this paper is arranged as follows: in Section II we introduce some notation; in Section III we present the result for the Poisson process; in Section IV we present the results for general point processes and arbitrary point patterns; and in Section V we present the results for the Wyner-Ziv setting.

II. Notation

We use a lower-case letter like x to denote a number, and an upper-case letter like X to denote a random variable. We use a boldface lower-case letter like $\mathbf{x}$ to denote a vector, a function of reals, or a point pattern, and it will be clear from the context which one we mean. If $\mathbf{x}$ is a vector, x_i denotes its ith element. If $\mathbf{x}$ is a function, $x(t)$ denotes its value at $t \in \mathbb{R}$. If $\mathbf{x}$ is a point pattern, we use $n_{\mathbf{x}}(\cdot)$ to denote its counting function, so $n_{\mathbf{x}}(t_2) - n_{\mathbf{x}}(t_1)$ is the number of points in $\mathbf{x}$ that fall in the interval $(t_1, t_2]$. We use a bold-face upper-case letter like $\mathbf{X}$ to denote a random vector, a random function, or a random point process. The random counting function corresponding to a point process $\mathbf{X}$ is denoted by $N_{\mathbf{X}}(\cdot)$.

We use $\mathrm{Ber}(p)$ to denote the Bernoulli distribution of parameter p, namely, the distribution that has probability p on the outcome 1 and probability $(1-p)$ on the outcome 0.

III. Covering a Poisson Process

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

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

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

The encoder maps the realization of the Poisson process to a message in $\{1, \ldots, 2^{TR}\}$. 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 total 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.

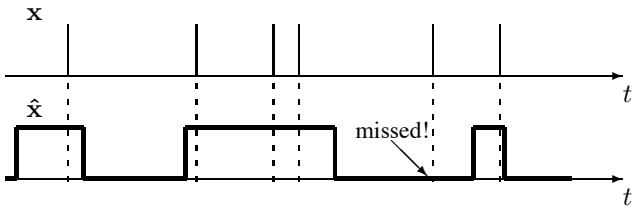

Fig. 1. Illustration of the problem.

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

$$d(\mathbf{x},\hat{\mathbf{x}}) \triangleq \begin{cases} \frac{\mu\left(\hat{x}^{-1}(1)\right)}{T}, & \text{if all points in } \mathbf{x} \text{ are in } \hat{x}^{-1}(1) \\ \infty, & \text{otherwise} \end{cases} \tag{1}$$

where $\mu(\cdot)$ denotes the Lebesgue measure.

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

$$\mathsf{E}\left[d\big(\mathbf{X}, \phi_T\left(f_T(\mathbf{X})\right)\big)\right] \le D + \epsilon.$$

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

$$R_{\text{Pois}}(D,\lambda) \triangleq \begin{cases} -\lambda \log D \text{ bits per second}, & D \in (0,1) \\ 0, & D \ge 1. \end{cases} \tag{2}$$

Theorem 1: For all $D, \lambda > 0$,

$$R(D,\lambda) = R_{\text{Pois}}(D,\lambda). \tag{3}$$

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 [3]. 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 slots of Δ seconds long. The encoder first maps the original point pattern $\mathbf{x}$ to a $\{0,1\}$-valued vector $\mathbf{x}'$ of length $\frac{T}{\Delta}$[1] in the following way: if $\mathbf{x}$ has at least one point in the time-slot $((i-1)\Delta, i\Delta]$, choose $x'_i = 1$; otherwise choose $x'_i = 0$. The encoder then maps $\mathbf{x}'$ to a message in $\{1,\ldots,2^{TR}\}$.

Based on the encoder's message, the reconstructor produces a $\{0,1\}$-valued length-$\frac{T}{\Delta}$ vector $\hat{\mathbf{x}}'$ to meet the distortion criterion

$$\mathsf{E}\left[d'(\mathbf{X}',\hat{\mathbf{X}}')\right] \le D + \epsilon,$$

where the distortion measure $d'(\cdot,\cdot)$ is given by

$$\begin{aligned} d'(0,0) &= 0 \\ d'(0,1) &= 1 \\ d'(1,0) &= \infty \\ d'(1,1) &= 1. \end{aligned}$$

It then maps $\hat{\mathbf{x}}'$ to a continuous-time signal $\hat{\mathbf{x}}$ through

$$\hat{x}(t) = \hat{x}'_{\lceil \frac{t}{\Delta} \rceil}, \quad t \in [0,T].$$

Scheme 1 reduces the task of designing a code for $\mathbf{X}$ subject to distortion $d(\cdot,\cdot)$ to the task of designing a code for the vector $\mathbf{X}'$ subject to the distortion $d'(\cdot,\cdot)$. The way we define $d'(\cdot,\cdot)$ yields the simple relation

$$d(\mathbf{x},\hat{\mathbf{x}}) = d'(\mathbf{x}',\hat{\mathbf{x}}'). \tag{4}$$

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

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

The next lemma, which is reminiscent of [4, 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 $D, \lambda > 0$,

$$R(D,\lambda) = \lim_{\Delta \downarrow 0} \frac{R_\Delta(D,\lambda)}{\Delta}. \tag{5}$$

Proof: See Appendix. ∎

Proof of Theorem 1: We derive $R(D,\lambda)$ by computing the right-hand side of (5). To compute $R_\Delta(D,\lambda)$ we apply Shannon's formula of the rate-distortion function for a discrete memoryless source [11]:

$$R_\Delta(D,\lambda) = \min_{P_{\hat{Z}|Z}:\mathsf{E}\left[d_\Delta(Z,\hat{Z})\right] \le D} I(Z;\hat{Z}).^2 \tag{6}$$

When $D \in (0,1)$, the conditional distribution $P_{\hat{Z}|Z}$ which achieves the minimum on the right-hand side of (6) is

$$\begin{aligned} P^*_{\hat{Z}|Z}(1|0) &= De^{\lambda\Delta} - e^{\lambda\Delta} + 1, \\ P^*_{\hat{Z}|Z}(1|1) &= 1. \end{aligned}$$

[1] When T is not divisible by Δ, we consider $\mathbf{x}$ as a pattern on $[0,T']$ where $T' = \lceil \frac{T}{\Delta} \rceil \Delta$. When we let Δ tend to zero, the difference between T and T' also tends to zero. Henceforth we ignore this technicality and assume T is divisible by Δ.

[2] Strictly speaking, since our distortion measure is unbounded, we need to modify Shannon's proof of this formula in order to use it for our problem. This can be done by letting the reconstructor produce the all-one sequence, which yields bounded distortion for any source sequence, whenever no codeword can be found that is jointly typical with the source sequence.

Computing the mutual information $I(Z;\hat{Z})$ under this $P^*_{\hat{Z}|Z}$ yields

$$R_\Delta(D,\lambda) = H_\mathrm{b}(D) - e^{-\lambda\Delta} H_\mathrm{b}(De^{\lambda\Delta} - e^{\lambda\Delta} + 1), \quad D \in (0,1), \tag{7}$$

where $H_\mathrm{b}(\cdot)$ denotes the binary entropy function.

When $D \geq 1$, it is optimal to choose $\hat{Z} = 1$ (deterministically), yielding

$$R_\Delta(D,\lambda) = 0, \quad D \geq 1. \tag{8}$$

Combining (5), (7) and (8) and computing the limit as Δ tends to zero yields (3). ∎

IV. Covering General Point Processes and Arbitrary Point Patterns

We next consider a general point process $\mathbf{Y}$. We assume that there exists some λ such that

$$\lim_{t\to\infty} \Pr\left[\frac{N_\mathbf{Y}(t)}{t} > \lambda + \delta\right] = 0 \quad \text{for all } \delta > 0. \tag{9}$$

Condition (9) is satisfied, for example, when $\mathbf{Y}$ is an ergodic process whose expected number of points per second is less than or equal to λ.

Since the Poisson process is memoryless, one naturally expects it to be the most difficult to describe. This is indeed the case, as the next theorem shows.

Theorem 2: The pair $(R_\mathrm{Pois}(D,\lambda), D)$ is achievable on any point process satisfying (9).

Before proving Theorem 2, we state a stronger result. Consider a point pattern $\mathbf{z}$ chosen by an adversary on the interval $[0,T]$ which contains no more than λT points. The corresponding counting function $n_\mathbf{z}(\cdot)$ must then satisfy

$$n_\mathbf{z}(T) \leq \lambda T. \tag{10}$$

The encoder and the reconstructor are allowed to use random codes. Namely, they fix a distribution on all (deterministic) codes of a certain rate on $[0,T]$. According to this distribution, they randomly pick 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 (10) 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* $\mathbf{z}$ satisfying (10) and its reconstruction is smaller than $D + \epsilon$.

Theorem 3: The pair $(R_\mathrm{Pois}(D,\lambda), D)$ is achievable with random coding against an adversary subject to (10).

Proof: First note that when $D \geq 1$, the encoder does not need to describe the pattern: the reconstructor simply produces the all-one function, yielding distortion 1 for any $\mathbf{z}$. Hence the pair $(0, D)$ is achievable with random coding.

Next consider $D \in (0,1)$. We use Scheme 1 as in Section III to reduce the original problem to one of random coding for an arbitrary discrete-time sequence $\mathbf{z}'$. Here $\mathbf{z}'$ is $\{0,1\}$-valued, has length $\frac{T}{\Delta}$, and satisfies

$$\sum_{i=1}^{T/\Delta} z'_i \leq \lambda T. \tag{11}$$

We shall construct a random code of rate $\frac{R}{\Delta}$ which, when applied to any $\mathbf{z}'$ satisfying (11), yields

$$\mathsf{E}\Big[d'(\mathbf{z}', \hat{\mathbf{Z}}')\Big] < D + \epsilon,$$

where the random vector $\hat{\mathbf{Z}}'$ is the result of applying the random encoder and decoder to $\mathbf{z}'$. Combined with Scheme 1 this random code will yield a random code on 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-$\frac{T}{\Delta}$ random sequences $\hat{\mathbf{Z}}'_m$, $m \in \{1,\ldots,2^{TR}\}$. The first sequence $\hat{\mathbf{Z}}'_1$ is chosen deterministically to be the all-one sequence. The other $2^{TR}-1$ sequences are drawn independently, with each sequence drawn IID Ber(D).

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

$$\hat{z}'_{m,i} = 1 \text{ whenever } z'_i = 1. \tag{12}$$

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

We next analyze the expected distortion of this random code for a fixed $\mathbf{z}'$ satisfying (11). Define

$$\mu \triangleq \frac{\sum_{i=1}^{T/\Delta} z'_i}{T},$$

and note that by (11) $\mu \leq \lambda$. Denote by $\mathcal{E}$ the event that the encoder cannot find $\hat{\mathbf{z}}'_m$, $m \in \{2,\ldots,2^{TR}\}$ satisfying (12). If $\mathcal{E}$ occurs, the encoder sends 1 and the resulting distortion is equal to 1.

The probability that a randomly drawn codeword $\hat{\mathbf{Z}}'_m$ satisfies (12) is

$$D^{\mu T} \geq D^{\lambda T} = 2^{(\lambda \log D) T}.$$

Because the codewords $\hat{\mathbf{Z}}'_m$, $m \in \{2,\ldots,2^{TR}\}$ are chosen independently, if we choose $R > -\lambda \log D$, then $\Pr[\mathcal{E}] \to 0$ as $T \to \infty$. 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 $\mathcal{E}^\mathrm{c}$. The reproduction $\hat{\mathbf{Z}}'$ has the following distribution: at positions where $\mathbf{z}'$ takes the value 1, $\hat{\mathbf{Z}}'$ must also be 1; at other positions the elements of $\hat{\mathbf{Z}}'$ have the IID Ber(D) distribution. Thus the expected value of $\sum_{i=1}^{T/\Delta} \hat{Z}'_i$ is $\mu T + D(\frac{T}{\Delta} - \mu T)$, and

$$\mathsf{E}\Big[\, d'(\mathbf{z}', \hat{\mathbf{Z}}') \,\Big|\, \mathcal{E}^\mathrm{c} \Big] = D + (1-D)\mu\Delta.$$

When we let Δ tend to zero, this value tends to D. We have thus shown that, for small enough Δ, we can achieve the pair $(R/\Delta, D)$ on $\mathbf{z}'$ with random coding whenever $R > -\lambda \log D$, and therefore we can also achieve (R,D) on the continuous-time point pattern $\mathbf{z}$ with random coding if $R > -\lambda \log D$. ∎

We next use Theorem 3 to prove Theorem 2.

Proof of Theorem 2: It follows from Theorem 3 that, on any point process satisfying (9), the pair $(R_\mathrm{Pois}(D, \lambda+\delta), D)$

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 2 now follows when we let δ tend to zero, since $R_{\text{Pois}}(D, \cdot)$ is a continuous function. ∎

V. Some Points are Known 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 each point in $\mathbf{X}$ is known to the reconstructor independently with probability p. Also assume that the encoder does not know which points are known to the reconstructor. The encoder maps $\mathbf{X}$ to a message in $\{1, \ldots, 2^{TR}\}$, 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 he knows. The achievability of a rate-distortion pair is defined in the same way as in Section III. 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 known 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 (2). Indeed, when the encoder knows which points are known to the reconstructor, it is optimal for it to describe only the remaining points, which themselves form a homogeneous 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,

$$R_{\text{WZ}}(D, \lambda, p) \geq R_{\text{Pois}}(D, (1-p)\lambda). \tag{13}$$

The next theorem shows that (13) holds with equality.

Theorem 4: Knowing the points at the reconstructor only is as good as knowing them also at the encoder:

$$R_{\text{WZ}}(D, \lambda, p) = R_{\text{Pois}}(D, (1-p)\lambda). \tag{14}$$

To prove Theorem 4, it remains to show that the pair $(R_{\text{Pois}}(D, (1-p)\lambda), D)$ is achievable. We shall show this as a consequence of 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 $\{1, \ldots, 2^{TR}\}$, 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 (1).

Theorem 5: 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 III to reduce the point pattern $\mathbf{z}$ to a $\{0,1\}$-valued vector $\mathbf{z}'$ of length $\frac{T}{\Delta}$. Define

$$\mu \triangleq \frac{\sum_{i=1}^{T/\Delta} z'_i}{T},$$

and note that, by assumption, $\mu \leq \lambda$. If $\mu \leq \nu$, then we can ignore the reconstructor's side-information and use the random code of Theorem 3. Henceforth we assume $\mu > \nu$.

Denote by $\mathbf{s}$ the point pattern known to the reconstructor and by $\mathbf{s}'$ 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,

$$\sum_{i=1}^{T/\Delta} s'_i \geq (\mu - \nu) T. \tag{15}$$

The encoder conveys the value of μT to the receiver using bits. Since μT 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.

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

$$\hat{\mathbf{z}}'_{m,l}, \quad m \in \{1, \ldots, 2^{TR}\},\ l \in \{1, \ldots, 2^{T\tilde{R}}\},$$

where each codeword is generated IID Ber(D).

To describe $\mathbf{z}'$, the encoder looks for a codeword $\hat{\mathbf{z}}'_{m,l}$ such that

$$\hat{z}'_{m,l,i} = 1 \text{ whenever } z'_i = 1. \tag{16}$$

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 $\tilde{l} \in \{1, \ldots, 2^{T\tilde{R}}\}$ such that

$$\hat{z}'_{m,\tilde{l},i} = 1 \text{ whenever } s'_i = 1. \tag{17}$$

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

To analyze the expected distortion for $\mathbf{z}'$ over this random code, first consider the event that the encoder cannot find a codeword satisfying (16). Note that the probability that a randomly generated codeword satisfies (16) is $D^{\mu T}$, so the probability of this event tends to zero as T tends to infinity provided that

$$R + \tilde{R} > -\mu \log D. \tag{18}$$

Next consider the event that the reconstructor finds more than one $\tilde{l}$ satisfying (17). The probability that a randomly generated codeword satisfies (17) is $D^{\sum_{i=1}^{T/\Delta} s'_i}$. Consequently, by (15) the probability of this event tends to zero as T tends to infinity provided

$$\tilde{R} < -(\mu - \nu) \log D. \tag{19}$$

Finally, if the encoder finds a codeword satisfying (16) and the reconstructor finds only one codeword satisfying (17), then the two codewords must be the same. Following the same calculations as in the proof of Theorem 3, the expected distortion in this case tends to D as Δ tends to zero.

Combining (18) and (19), we can make the expected distortion arbitrarily close to D as $T \to \infty$ if

$$R > -\nu \log D.$$

■

Proof of Theorem 4: The claim follows from (13), Theorem 5, and the Law of Large Numbers. ■

APPENDIX

In this appendix we prove Lemma 1. Given any rate-distortion code with 2^{TR} codewords $\hat{\mathbf{x}}_m$, $m \in \{1, \ldots, 2^{TR}\}$ 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 $\hat{\mathbf{w}}_m$, $m \in \{1, \ldots, 2^{TR}+1\}$. We choose the last codeword to be the constant 1. We next describe our choices for the other codewords. For every $\epsilon > 0$ and every $\hat{\mathbf{x}}_m$, we can approximate the set $\{t\colon \hat{x}_m(t) = 1\}$ by a set $\mathcal{A}_m$ that is equal to a finite, say N_m, union of open intervals. More specifically,

$$\mu\left(\hat{x}_m^{-1}(1) \triangle \mathcal{A}_m\right) \le 2^{-TR}\epsilon, \tag{20}$$

where $\triangle$ denotes the symmetric difference between two sets (see, e.g., [12, Chapter 3, Proposition 15]). Define

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

and note that by (20)

$$\mu(\mathcal{B}) \le \epsilon. \tag{21}$$

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

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

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

$$\hat{\mathbf{w}}_m = \mathbf{1}_{\mathcal{T}_m},$$

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

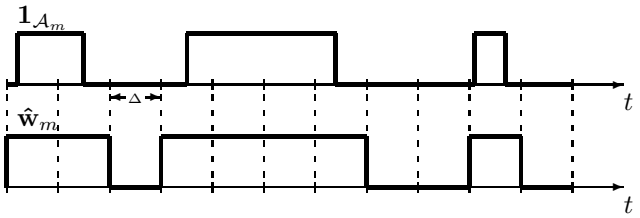

Fig. 2. Constructing $\hat{\mathbf{w}}_m$ from $\mathcal{A}_m$.

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

It can be seen that

$$\mu\left(\hat{w}_m^{-1}(1)\right) - \mu(\mathcal{A}_m) \le 2N\Delta, \quad m \in \{1, \ldots, 2^{TR}\}. \tag{22}$$

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 (20) and (22), 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. Next consider the case where $\mathbf{x}$ does contain points in $\mathcal{B}$. By (21), 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 $\{\hat{\mathbf{x}}_m\}$. This concludes the proof of Lemma 1.

REFERENCES

[1] Y. Kabanov, "The capacity of a channel of the Poisson type," *Theory of Probability and Its Appl.*, vol. 23, pp. 143–147, 1978.

[2] M. H. A. Davis, "Capacity and cutoff rate for Poisson-type channels," *IEEE Trans. Inform. Theory*, vol. 26, pp. 710–715, Nov. 1980.

[3] 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.

[4] ——, "Capacity and error exponent for the direct detection photon channel — part II," *IEEE Trans. Inform. Theory*, vol. 34, pp. 1462–1471, Nov. 1988.

[5] 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.

[6] I. Rubin, "Information rates and data-compression schemes for Poisson processes," *IEEE Trans. Inform. Theory*, vol. 20, no. 2, pp. 200–210, Mar. 1974.

[7] 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.

[8] V. Anantharam and S. Verdú, "Bits through queues," *IEEE Trans. Inform. Theory*, vol. 42, no. 1, pp. 4–18, Jan. 1996.

[9] 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.

[10] S. Bross, A. Lapidoth, and L. Wang, "The Poisson channel with side information," in *Proceedings Forty-Seventh Allerton Conf. Comm., Contr. and Comp.*, Allerton House, Monticello, Illinois, September 30–October 2, 2009.

[11] C. E. Shannon, "A mathematical theory of communication," *Bell System Techn. J.*, vol. 27, pp. 379–423 and 623–656, July and Oct. 1948.

[12] H. L. Royden, *Real Analysis*, 3rd ed. Macmillan Publishing Company, 1988.