An Alternative Proof of the Rate-Distortion Function of Poisson Processes with a Queueing Distortion Measure

Todd P. Coleman*, Negar Kiyavash†, and Vijay G. Subramanian‡
* ECE Dept & Coordinated Science Laboratory, UIUC. colemant@uiuc.edu
† CS Dept & Information Trust Institute, UIUC. kiyavash@uiuc.edu
‡Hamilton Institute, NUIM. vijay.subramanian@nuim.ie

Abstract

This paper presents a proof of the rate distortion function of a Poisson process with a queuing distortion measure that is in complete analogy with the proofs associated with the rate distortion functions of a Bernoulli source with Hamming distortion measure and a Gaussian source with squared-error distortion measure. Analogous to those problems, the distortion measure that we consider is related to the logarithm of the conditional distribution relating the input to the output of a well-known channel coding problem, specifically the Anantharam and Verdu "Bits through Queues" [1] coding problem. Our proof of the converse utilizes McFadden's point process entropy formulation [2] and involves a number of mutual information inequalities, one of which exploits the maximum-entropy achieving property of the Poisson process. Our test channel uses Burke's theorem [3], [4] to prove achievability.

I. Introduction

In their landmark "Bits through Queues" paper Anantharam & Verdu [1] concretely illustrated the advantages of coding through the *timing* of packets in packet systems. For a memoryless arrival process of rate λ, the First-Come First-Serve (FCFS) $\cdot/M/1$ continuous-time queue with service rate $\mu > \lambda > 0$ has a capacity $C(\lambda)$ given by

$$C(\lambda) = \lambda \log \frac{\mu}{\lambda}, \quad \lambda < \mu \text{ nats/s.} \tag{1}$$

Then the capacity of the FCFS $\cdot/M/1$ continuous-time queue with service rate μ is given by the maximum of $C(\lambda)$ over all possible arrival rates $\lambda < \mu$, namely,

$$C = e^{-1}\mu \text{ nats/s,} \tag{2}$$

where the maximum corresponding to (2) is achieved in (1) at $\lambda = e^{-1}\mu$.

The remarkable aspect of the "Bits through Queues" result [1] is that despite a queueing system having non-linearities, memory, and non-stationarities, there still exists

T. Coleman would like to acknowledge the support of this research through the grant DARPA ITMANET program via US Army RDECOM contract W911NF-07-1-0029.

V. Subramanian would like to acknowledge the support of this research through SFI grant 03/IN3/I346.

a closed-form solution for the FCFS $\cdot/M/1$ queue. This fact, along with the fact that an exponential service timing channel (ESTC) can be seen as an "entropy-increasing operator" [5, Sec. 2],[6, Thm 1] and also the saddle-point property of mutual information for the ESTC [1, Thm 3], makes the ESTC appear to be analogous to the binary symmetric channel (BSC) and additive white Gaussian noise (AWGN) channels.

The latter two channels have canonical rate-distortion problems associated with them, where the distortion measure is related to the logarithm of the conditional distribution relating the input to the output of the channels. A thorough discussion of this and the applicability of this to optimality of uncoded transmissions can be found in [7]. In these two cases (AWGN and BSC), simple mutual information inequalities exploiting maximum-entropy distributions and entropy manipulations lead to a closed-form lower bound - which can be shown to be tight by developing elegant "test-channel" achievability arguments [8, Sec. 10.3].

Given the entropy-maximizing property of the exponential distribution and the pre-eminence of the Poisson process in point-process theory [9] rate distortion functions under various distortion measures have been sought in the literature [10], [11], [12], [13], [14], [15]. A large class of distortion measures, including those that consider a magnitude error criterion and the reproduction of the number of events in a given interval and the times between successive events, lead to non-causal reproductions [10], [11], [12]. Gallager [13] considered transmission of point processes through a causal system such that the output point process always follows the input but the order of messages need not be preserved. Considering the output of such a system to be a reproduction of the input point process and using average delay per message as the distortion measure, Gallager [13] derived a minimum information rate for the Poisson process for a given fidelity of reproduction that included both the timing information and the message ordering. Bedekar [15] refined Gallager's approach by insisting upon in-order delivery of the messages through the system and defining the distortion measure to be the average service-time D of a hypothetical FCFS queue that would result in the output. For a Poisson message arrival stream with rate λ Bedekar [15] showed that the rate-distortion function $R(D)$ is given as follows:

$$R(D) = \begin{cases} -\log(\lambda D) \text{ bits/message} & \text{if } 0 < \lambda D \leq 1; \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

Related work pertaining to minimizing mutual information across special classes of causal systems is the jamming work of Giles and Hajek [16].

Verdu [14] considered a different approach where the message inter-arrival times were reproduced to a fidelity of D seconds under the constraint that the last arrival is declared no sooner than it occurs. The rate distortion function $R(D)$ that results in [14] for a

Poisson process of rate λ is given by

$$R(D) = \begin{cases} -\lambda \log(\lambda D) \text{ bits/second} & \text{if } 0 < \lambda D \leq 1; \\ 0 & \text{otherwise,} \end{cases} \tag{4}$$

where the optimal reproduction is by a counting process with inter-arrival times given by $\hat{X} = \max(0, X - W) + D$ where $X \sim \exp(\lambda)$ is the random-variable corresponding to the inter-arrival time of the Poisson process and W is a random-variable independent of X such that the marginal distributions are $\exp(1/D)$. The sequence of W random-variables are chosen such that the constraint on the last arrival time mentioned earlier is respected.

Wagner and Anantharam [17] embarked upon understanding the similarities between the BSC, AWGN, and ETSC channels by developing a distance metric for the $\cdot/M/1$ queuing system that is analogous to Hamming/Euclidean distances for BSC/AWGN channels; in that, it is related to the logarithm of the channel's conditional density. Using this distance metric they were able to characterize the zero-rate reliability of the ESTC. This led the authors to conjecture that an analogous rate-distortion problem with an appropriate distortion measure should lead to an elegant set of mutual information inequalities, a closed-form lower bound, and an elegant "test-channel" to illustrate achievability. We accomplish that in what follows using the distance metric developed in [17] per unit time to be the distortion measure.

This paper is organized as follows. In Section II we state a few preliminaries that are helpful in proving our result. The specific distortion measure and identification of the rate-distortion function is then presented in Section III. We then present in Section IV a detailed comparison with existing results in the literature, and finally conclude in Section V.

II. Preliminaries

The entropy on $[0, T]$ of a point process $\mathcal{P}$ with arrival times $\{P_1, P_2, \ldots\}$ is defined [2] as the sum of its *numerical entropy* and its *positional entropy*:

$$h_T(\mathcal{P}) := H(M) + E_M[h(P_1, \ldots, P_m | M = m)],$$

where M is a random variable denoting the number of arrivals in $[0, T]$, $H(\cdot)$ is discrete entropy, $h(\cdot)$ is differential entropy, and given $M = m$, $\{P_1, \ldots, P_m\}$ are the locations (in time) of the arrivals. Also define the counting function $N_{\mathcal{P}}(T)$ at time $T \geq 0$ associated with point process $\mathcal{P}$ as follows

$$N_{\mathcal{P}}(T) := \sup\{n \in \mathbb{N} : P_n \leq T\}.$$

We will restrict our attention to non-explosive point processes whereby $N_{\mathcal{P}}(T)$ is finite (*a.s.*) for all $0 \leq T < +\infty$.

Analogously, we can define the joint entropy on $[0, T]$ of two point processes $\mathcal{P}^a$ and $\mathcal{P}^b$ by

$$h_T(\mathcal{P}^a, \mathcal{P}^b) := H(M_a, M_b) + E_{M_a, M_b}[h(P_1^a, \ldots, P_{m_a}^a, P_1^b, \ldots, P_{m_b}^b | M_a = m_a, M_b = m_b)],$$

where M_a and M_b are the random variables denoting the number of arrivals of processes $\mathcal{P}^a$ and $\mathcal{P}^b$ respectively, and $\{P_1^a, \ldots, P_{m_a}^a\}$ and $\{P_1^b, \ldots, P_{m_b}^b\}$ are the locations (in time) of the arrivals of processes $\mathcal{P}^a$ and $\mathcal{P}^b$ given $M_a = m_a$ and $M_b = m_b$. From this we can define the conditional point process entropy simply as

$$h_T(\mathcal{P}^a | \mathcal{P}^b) := h_T(\mathcal{P}^a, \mathcal{P}^b) - h_T(\mathcal{P}).$$

It is well known that for all continuous-time point processes [9] $\mathcal{P}$ with rate at most $\lambda > 0$ (i.e, $E[N_{\mathcal{P}}(T)] \leq \lambda T$ for all $T \in [0, +\infty)$), a Poisson process of rate λ maximizes the entropy [2]. The entropy rate of a Poisson process of rate λ over $(0, T]$ is given by

$$h_T(\mathcal{P}) = T\lambda(1 - \log \lambda).$$

III. The Canonical Rate-Distortion Function in Analogy with "Bits Through Queues"

We now consider the canonical rate-distortion function for any two point processes. Over $[0, T]$, consider two point processes $\mathcal{X}$ and $\hat{\mathcal{X}}$. Denote the arrival times of $\mathcal{X}$ by $\{X_i\}$ and the arrival times of $\hat{\mathcal{X}}$ by $\{\hat{X}_i\}$. We remind the reader that the associated counting functions of $\mathcal{X}$ and $\hat{\mathcal{X}}$ are $N_X(t)$ and $N_{\hat{X}}(t)$, respectively. For any two point processes $\mathcal{X}$ and $\hat{\mathcal{X}}$ such that $N_X(T) = N_{\hat{X}}(T)$, and $N_{\hat{X}}(t) \geq N_X(t),\ \forall t \in [0, T]$ define

$$\mathcal{S} = \mathcal{X} \diamond \hat{\mathcal{X}}$$

as the point process with **inter-arrival times** $\{S_i\}$ given by the induced service times of a FCFS queueing system with $\hat{\mathcal{X}}$ as the input and $\mathcal{X}$ as the output as shown in Figure 1. Note the abuse of notation in that for the process $\mathcal{S}$, the $\{S_i\}$s are the inter-arrival times. Specifically, $S_i = X_i - \max(X_{i-1}, \hat{X}_i)$; see Figure 2 for the validity of this relationship. Note that the counting function for $\mathcal{S}$, $N_S(t)$ is uniquely defined on $[0, \sum_{i=1}^{N_X(T)} S_i]$.

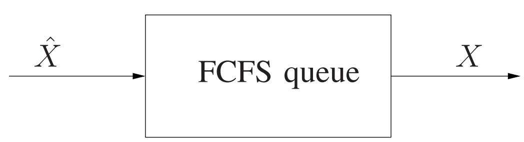

Fig. 1. The Diagram of interpreting $\mathcal{X}$ as the output of a FCFS queueing system with $\hat{\mathcal{X}}$ as the input. The point process $\mathcal{S}$ is defined by treating inter-arrival time S_i as the ith service time.

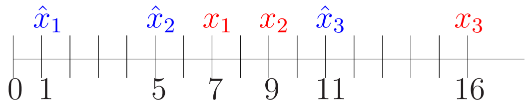

Fig. 2. The arrivals (blue) and departures (red) from a FCFS $\cdot/M/1$ Queue. Note that the service time s_1 for the first packet is given by $x_1 - \hat{x}_1 = 6$. Note that the first packet does not depart from the queue until after the second packet arrives. Thus the service time s_2 for the second packet is given by $s_2 = x_2 - x_1$, because the server starts working on the second packet once the first packet departs. The second packet departs before the third arrival $\hat{x}_3$. Thus the third service time is simply $s_3 = x_3 - \hat{x}_3$. So in general, it follows that $s_i = x_i - \max(\hat{x}_i, x_{i-1})$.

With this definition, we can now define the distortion between any two realizations $\mathbf{x}$ and $\hat{\mathbf{x}}$ of point processes $\mathcal{X}$ and $\hat{\mathcal{X}}$. Define

$$d_T(\mathbf{x}, \hat{\mathbf{x}}) = \begin{cases} \frac{1}{T}\sum_{i=1} s_i & \text{if } N_X(T) = N_{\hat{X}}(T) \text{ and } N_{\hat{X}}(t) \geq N_X(t),\ \forall t \in [0,T] \\ \infty & \text{otherwise} \end{cases} \quad (5)$$

where $s_i = x_i - \max(x_{i-1}, \hat{x}_i)$ defines a realization $\mathbf{s}$ of $\mathcal{S}$.

For the rate-distortion problem we consider $\hat{\mathcal{X}}$ to be a reproduction of $\mathcal{X}$ with average distortion measure $\limsup_{T\to+\infty} E[d_T(\mathcal{X}, \hat{\mathcal{X}})]$. We now state the rate-distortion theorem, show the converse, and prove achievability.

Theorem 3.1: The rate-distortion function for the source a Poisson process with rate λ with distortion measure (5) is given by

$$R(D) = \begin{cases} -\lambda \log(D) \text{ bits/second} & D \in (0,1); \\ 0 & D \geq 1. \end{cases}$$

The proof is presented in two parts: the converse in Section III-A and achievability in Section III-B. The converse consists of lower bounding the mutual information between the input and output by the service-time process. Thereafter, the entropy maximizing property of the Poisson process is used to derive the expression for $R(D)$. Then the achievability result follows quite easily considering an $M/M/1$ queue in steady-state and invoking Burke's theorem [3], [4] to show that every step in the proof of the converse is satisfied with equality in this setting.

A. Converse

Fix a distortion value $D \in (0,1)$. Consider any joint distribution on $\mathcal{X}$ and $\hat{\mathcal{X}}$ satisfying the distortion constraint, where $\mathcal{X}$ is a Poisson process of rate λ on $[0,T]$. Note that because the finite distortion constraint is met, with probability one, $N_X(T) = N_{\hat{X}}(T)$. Define

$$\mu \triangleq \frac{\lambda}{D}. \quad (6)$$

Note given $0 < \lambda < \mu < \infty$, for any valid $\mathcal{X}$ and $\hat{\mathcal{X}}$ defined on $[0, T]$ we have

$$
\begin{aligned}
I_T(\mathcal{X}; \hat{\mathcal{X}}) &= h_T(\mathcal{X}) - h_T(\mathcal{X}|\hat{\mathcal{X}}) \\
&= \lambda T(1 - \log \lambda) - h_T(\mathcal{X}|\hat{\mathcal{X}}) && (7)\\
&= \lambda T(1 - \log \lambda) - h_T(\mathcal{X} \diamond \hat{\mathcal{X}}|\hat{\mathcal{X}}) && (8)\\
&\geq \lambda T(1 - \log \lambda) - h_T(\mathcal{X} \diamond \hat{\mathcal{X}}) = \lambda T(1 - \log \lambda) - h_T(\mathcal{S}) \\
&\geq \lambda T(1 - \log \lambda) - \mu \left(\frac{\lambda T}{\mu}\right)(1 - \log \mu) && (9)\\
&= \lambda T \log \left(\frac{\mu}{\lambda}\right) = -\lambda T \log(D) \\
\Rightarrow \frac{1}{T} I_T(\mathcal{X}; \hat{\mathcal{X}}) &\geq -\lambda \log(D)
\end{aligned}
$$

where (7) follows because $\mathcal{X}$ is a Poisson process with rate λ, (8) follows by referring to Appendix I, and (9) follows because of the argument below.

Note the following properties for $\mathcal{S}$, namely,

1) $\mathcal{S} = \mathcal{X} \diamond \hat{\mathcal{X}}$ is a point process with an average of λT spikes; and
2) $\frac{1}{T} E\left[\sum_i S_i\right] \leq D = \frac{\lambda}{\mu}$, and thus $E\left[\sum_i S_i\right] \leq \frac{\lambda T}{\mu}$.

Thus $h_T(\mathcal{S})$ is upper-bounded by the entropy of the maximum-entropy point process among all point processes of duration $\frac{\lambda T}{\mu}$ and rate at most μ, which is a Poisson process on $[0, \frac{\lambda T}{\mu}]$ of rate μ.

B. Achievability

Generate codewords $\hat{\mathcal{X}}$ according to a Poisson process with rate (λ) on $[0, T]$. For any $\hat{\mathcal{X}}$, consider the output of an ESTC with rate $\mu > \lambda$ in steady-state with input process $\hat{\mathcal{X}}$, and denote the output process as $\mathcal{X}$. Note that by Burke's theorem [3], [4] the departure process $\mathcal{X}$ is also a Poisson process with rate λ and thus, by defining $D = \frac{\lambda}{\mu}$, we have that

$$
\begin{aligned}
\lim_{T\to\infty} \frac{1}{T} I_T(\mathcal{X}; \hat{\mathcal{X}}) &= \lim_{T\to\infty} \frac{1}{T}[h_T(\mathcal{X}) - h_T(\mathcal{X}|\hat{\mathcal{X}})] \\
&= \lim_{T\to\infty} \frac{1}{T}[\lambda T(1 - \log \lambda) - h_T(\mathcal{X} \diamond \hat{\mathcal{X}}|\hat{\mathcal{X}})] \\
&= \lambda(1 - \log \lambda) - \lim_{T\to\infty} \frac{1}{T} h_T(\mathcal{S}|\hat{\mathcal{X}}) \\
&= \lambda(1 - \log \lambda) - \lim_{T\to\infty} \frac{1}{T} h_T(\mathcal{S}) \\
&= \lambda(1 - \log \lambda) - \lim_{T\to\infty} \frac{1}{T} \mu \left(\frac{\lambda T}{\mu}\right)(1 - \log \mu) \\
&= \lambda(1 - \log \lambda) - \lambda(1 - \log \mu) = \lambda \log \left(\frac{\mu}{\lambda}\right) = -\lambda \log D.
\end{aligned}
$$

Note that this immediately suggests the following coding scheme. For a chosen $D \in (0,1)$ and $T \in (0,+\infty)$, given a realization $\mathbf{x}$ of Poisson process $\mathcal{X}$ with rate λ, generate $2^{TR(D)}$ independent realizations of a rate λ Poisson process $\hat{\mathcal{X}}$ denoting the i^{th} realization by $\hat{\mathbf{x}}(i)$ for $i = 1, 2, \ldots, 2^{TR(D)}$. Now choose $\hat{\mathbf{x}}(i^*)$ as the reproduction of $\mathbf{x}$ where $i^* \in \arg\min_{i=1,2,\ldots,2^{TR(D)}} d_T(\mathbf{x}, \hat{\mathbf{x}}(i))$.

IV. Discussion

The rate distortion function that we have arrived at is exactly the same as in Bedekar [15]. In fact the distortion measures are related and the achievability of the minimum rate using Poisson code-words passed through an $M/M/1$ queue in steady-state is also similar. However, a key distinction is that while our code-words are constructed such that $\hat{\mathcal{X}}$ leads $\mathcal{X}$, i.e., $\hat{X}_i \leq X_i$ for every i, the code-words in [15] are such that the reproductions always follow $\mathcal{X}$. Thus, as an analogy, if one were to construct code-words following the approach outlined in Wyner [18], [19], i.e., sample every D seconds and count the number of arrivals in $[nD, (n+1)D]$ for $n \in \mathbb{N} \cup \{0\}$, then our code-words would assign all the arrivals in $[nD, (n+1)D]$ to nD while the code-words of [15] would assign all of them to $(n+1)D$.

Interpreted appropriately, the rate-distortion function for the problem considered [14] is equivalent in structure to the one we developed, however [14] focusses more on constructing the code-words by approximating the inter-arrival times and relies on the extremal properties of the exponential distribution rather than on queueing. Note that all the code-words in [14] are such that the arrival points (except for the last arrival) of the reproductions can either lead or follow the arrival points of $\mathcal{X}$. There are also other subtle differences in the constraints and distortion measures in our approach *vis-a-vis* [14], namely,

1) Distortion measure: [14] uses the following distortion measure for the reproductions given the source realization

$$\hat{x}_i - \hat{x}_{i-1} 1_{\{i>1\}} \leq x_i - x_{i-1} 1_{\{i>1\}} + D, \ \forall\, i \in \{1, \ldots, m\},$$

where D is a length of time while we consider $\frac{1}{T} E[\sum_{i=1}^{m} S_i] \leq D$ where D is a dimension-less number;

2) Timing constraint: [14] imposes the constraints that the reproductions produce the same number of arrivals as the source realization with the last arrival of the reproduction being after the last arrival of the source realization, i.e.,

$$N_x(T) = N_{\hat{x}}(T) \text{ and } x_{N_x(T)} \leq \hat{x}_{N_{\hat{x}}(T)},$$

whereas we insist that the reproductions lead the original while producing the same number of arrivals, i.e.,

$$\forall\, t \in [0,T] \ N_x(t) \leq N_{\hat{x}}(t) \text{ and } N_x(T) = N_{\hat{x}}(T);$$

3) Nature of Reproductions: A final but important distinction is that our reproductions are always Poisson processes whereas those in [14] are not even renewal processes, although they are, in a sense, close to being renewal processes.

V. Conclusions

In conclusion we would like to emphasize that our main motivation for this work was to develop a mutual information lower-bounding technique using maximum-entropy arguments, and show achievability with an appropriate test-channel - to be in complete analogy to the methodology developed for the AWGN and BSC rate-distortion problems. Also note that the exact same line of analysis follows for the discrete-time queueing case [20], [6]. We plan to consider extensions to networks in the future.

Acknowledgements

The authors would like to thank Aaron Wagner for useful comments on an earlier draft of this manuscript, and also Anand Bedekar for providing us with a copy of [15].

References

[1] V. Anantharam and S. Verdu, "Bits through queues," *IEEE Transactions on Information Theory*, vol. 42, no. 1, pp. 4–18, 1996.

[2] J. McFadden, "The entropy of a point process," *SIAM Journal of Applied Mathematics*, vol. 13, pp. 988–994, 65.

[3] P. Bremaud, *Point Processes and Queues: Martingale Dynamics*. New York: Springer-Verlag, 1981.

[4] R. Gallager, *Discrete Stochastic Processes*. Boston, MA: Kluwer, 1996.

[5] B. Prabhakar and N. Bambos, "The entropy and delay of traffic processes in ATM networks," in *Proceedings of the Conference on Information Science and Systems (CISS)*, Baltimore, Maryland, 1995, pp. 448–453.

[6] B. Prabhakar and R. Gallager, "Entropy and the timing capacity of discrete queues," *IEEE Transactions on Information Theory*, vol. 49, no. 2, pp. 357–370, February 2003.

[7] M. Gastpar, B. Rimoldi, and M. Vetterli, "To code, or not to code: Lossy source-channel communication revisited," *IEEE Transactions on Information Theory*, vol. 49, no. 5, pp. 1147–1158, May 2003.

[8] T. M. Cover and J. Thomas, *Elements of Information Theory*, 2nd ed. New York: Wiley, 1991.

[9] D. Daley and D. Vere-Jones, *An Introduction to the theory of point processes*. New York: Springer-Verlag, 1988.

[10] I. Rubin, "Information rates for poisson sequences," *IEEE Transactions on Information Theory*, vol. 19, no. 3, pp. 283–294, 1973.

[11] ——, "Information rates and data-compression schemes for poisson processes," *IEEE Transactions on Information Theory*, vol. 20, no. 2, pp. 200–210, 1974.

[12] H. Sato and T. Kawabata, "Information rates for poisson point processes," *Transactions of IEICE*, vol. E70, no. 9, pp. 817–822, Sept. 1987.

[13] R. G. .Gallager, "Basic limits on protocol information in data communication networks," *IEEE Transactions on Information Theory*, vol. 22, no. 4, pp. 385–398, 1976.

[14] S. Verdu, "The exponential distribution in information theory," *Problems of Information Transmission*, vol. 32, no. 1, pp. 86–95, Jan-Mar 1996.

[15] A. S. Bedekar, "On the information about message arrival times required for in-order decoding," in *Proceedings of the Internation Symposium on Information Theory (ISIT)*, Washington, D.C., 2201, p. 227.

[16] J. Giles and B. Hajek, "An information-theoretic and game-theoretic study of timing channels," *IEEE Trans. Inform. Theory*, vol. 48, no. 9, pp. 2455–2477, 2002.

[17] A. B. Wagner and V. Anantharam, "Zero-rate reliability of the exponential-server timing channel," *IEEE Transactions on Information Theory*, vol. 51, no. 2, pp. 447–465, February 2005.

[18] A. D. Wyner, "Capacity and error exponent for the direct detection photon channel - Part I," *IEEE Transactions on Information Theory*, vol. 34, no. 6, pp. 1449–1461, 1988.
[19] ——, "Capacity and error exponent for the direct detection photon channel - Part II," *IEEE Transactions on Information Theory*, vol. 34, no. 6, pp. 1462–1471, 1988.
[20] A. S. Bedekar and M. Azizoglu, "The information-theoretic capacity of discrete-time queues," *IEEE Transactions on Information Theory*, vol. 44, no. 2, pp. 446–461, 1998.
[21] L. Kleinrock, *Queueing Systems. Vol 1: Theory*. New York, NY: Wiley, 1975.

APPENDIX I
PROOF OF (8)

First note that $\mathcal{S} = \mathcal{X} \diamond \hat{\mathcal{X}}$ is uniquely defined given $\mathcal{X}$ and $\hat{\mathcal{X}}$. Also, from the queueing interpretation, it is clear that given $\hat{\mathcal{X}}$, $\mathcal{S}$ uniquely defines $\mathcal{S}$. However, the point process entropy contains both a discrete numerical entropy component as well as a positional entropy, where the latter relies upon a differential entropy term - so just illustrating this bijection does not suffice. We now focus on the differential entropy term pertaining to the positional entropy and show there is no *amplitude scaling*. To do this, we consider conditioning upon $\hat{\mathcal{X}} = \hat{\mathbf{x}}$ and investigate the Jacobian of the mapping between $\mathcal{X}$ and $\mathcal{S}$. Since $N_X(T) = N_{\hat{X}}(T)$ with probability one, it follows that conditioned upon $\hat{\mathcal{X}} = \hat{\mathbf{x}}$, we know exactly how many arrival times there are for $\mathcal{X}$ up until time T, i.e., $N_x(T)$; to be concise, we define $m \triangleq N_x(T)$. Now it suffices to show that

$$\begin{aligned} &h(X_1, \ldots, X_m | N_{\hat{X}}(T) = m, \hat{X}_1 = \hat{x}_1, \ldots, \hat{X}_m = \hat{x}_m) \\ &= h(S_1, \ldots, S_m | N_{\hat{X}}(T) = m, \hat{X}_1 = \hat{x}_1, \ldots, \hat{X}_m = \hat{x}_m). \end{aligned}$$

But note that the service process $\{S_i\}$, given $\{\hat{X}_i\}$, is related to the departure process $\{X_i\}$ by the simple (recursive) relationship [4], [21]

$$S_i = X_i - \max(X_{i-1}, \hat{x}_i).$$

Also see Figure 2. Considering specific realizations note that the mapping from $(x_1, \ldots, x_m)$ to $(s_1, \ldots, s_m)$ is differentiable except on a set $\mathcal{A} \in \Re_+^m$ with Lebesgue measure zero (pertaining to the points of discontinuity of the functions $\max(x_i, s_i)$). We can consider the Jacobian at any point $(x_1, \ldots, x_m)$ not in $\mathcal{A}$, and note that

$$|J(x_1, \ldots, x_m)| = \begin{vmatrix} \frac{\partial s_1}{\partial x_1} & \frac{\partial s_1}{\partial x_2} & \cdots & \frac{\partial s_1}{\partial x_m} \\ \frac{\partial s_2}{\partial x_1} & \frac{\partial s_2}{\partial x_2} & \cdots & \frac{\partial s_2}{\partial x_m} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial s_m}{\partial x_1} & \frac{\partial s_m}{\partial x_2} & \cdots & \frac{\partial s_m}{\partial x_m} \end{vmatrix} = \begin{vmatrix} 1 & 0 & \ldots & 0 & 0 \\ \frac{-1\pm 1}{2} & 1 & \ldots & 0 & \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & \frac{-1\pm 1}{2} & 1 \end{vmatrix} = 1.$$

The $\frac{-1\pm 1}{2}$ simply denotes that if $x_{i-1} > \hat{x}_i$, then $\max(x_{i-1}, \hat{x}_i) = x_{i-1}$ and thus the corresponding Jacobian entry in the $(i, i-1)$ point of the matrix is equal to -1; otherwise, if $x_{i-1} < \hat{x}_i$ then $\max(x_{i-1}, \hat{x}_i) = \hat{x}_i$, and thus the corresponding Jacobian entry in the

$(i, i-1)$ point of the matrix is equal to 0. No matter what the specific values in the $(i, i-1)$ locations of the Jacobian are, note that the determinant is always 1.

Thus it follows that

$$\begin{aligned} & h(X_1, \ldots, X_m | N_{\hat{X}}(T) = m, \hat{X}_1 = \hat{x}_1, \ldots, \hat{X}_m = \hat{x}_m) \\ = \; & h(S_1, \ldots, S_m | N_{\hat{X}}(T) = m, \hat{X}_1 = \hat{x}_1, \ldots, \hat{X}_m = \hat{x}_m). \end{aligned}$$

Consequently,

$$\begin{aligned} & E_M[h(X_1, \ldots, X_m | N_{\hat{X}}(T) = M, \hat{X}_1 = \hat{x}_1, \ldots, \hat{X}_M = \hat{x}_M)] \\ = \; & E_M[h(S_1, \ldots, S_m | N_{\hat{X}}(T) = M, \hat{X}_1 = \hat{x}_1, \ldots, \hat{X}_M = \hat{x}_M)]. \end{aligned}$$

Thus (8) is proved.