Robust Decoding for Timing Channels

Rajesh Sundaresan, *Student Member, IEEE,* and Sergio Verdú, *Fellow, IEEE*

Abstract—To transmit information by timing arrivals to a single-server queue, we consider using the exponential server channel's maximum-likelihood decoder. For any server with service times that are stationary and ergodic with mean $1/\mu$ seconds, we show that the rate $e^{-1}\mu$ nats per second (capacity of the exponential server timing channel) is achievable using this decoder. We show that a similar result holds for the timing channel with feedback. We also show that if the server jams communication by adding an arbitrary amount of time to the nominal service time, then the rate $e^{-1}\mu_1\mu_2/(\mu_1 + \mu_2)$ nats per second is achievable with random codes, where the nominal service times are stationary and ergodic with mean $1/\mu_1$ seconds, and the arithmetic mean of the delays added by the server does not exceed $1/\mu_2$ seconds. This is a model of an arbitrarily varying channel where the current delay and the current input can affect future outputs. We also show the counterpart of these results for single-server discrete-time queues.

Index Terms—Arbitrarily varying channel, channels with feedback, mismatched decoder, point-process channel, robust decoding, single-server queue, timing channels.

I. INTRODUCTION

CONSIDER the problem of transmitting information through the epochs at which packets arrive at a single-server queue [1]. All packets are identical and information is contained only in the times of arrival of these packets. The service times cause delays that corrupt the input information. If the service times are independent and exponentially distributed with mean $1/\mu$ seconds, the capacity of this channel is $e^{-1}\mu$ nats per second when the cost of transmission is the expected time for the last packet to exit the system [1]. Furthermore, if the service times are independent and identically distributed (i.i.d.) with mean $1/\mu$ seconds, but are not exponentially distributed, then we can communicate reliably at a rate $e^{-1}\mu$ nats per second [1]. Thus among all servers with i.i.d. service times of mean $1/\mu$ seconds, the exponential server has the least capacity. These results in [1] assume that both the encoder and the decoder know the distribution of the service times.

When the service times are independent and exponentially distributed, the corresponding maximum-likelihood decoder is easy to implement. Given the sequence of times at which packets depart from the queue, the decoder finds the codeword that explains the sequence of departures with the smallest sum of service times. To do this, the decoder needs only additions, sub-

tractions, and comparisons. Since the exponential server has the least capacity, and its maximum-likelihood decoder uses simple functions, we consider using this decoding strategy when the service times are not exponentially distributed. In this case, although the above decoder is suboptimal, its simplicity and general applicability are appealing.

In this paper, we show that we can communicate reliably at a rate $e^{-1}\mu$ nats per second using the above decoding strategy when the distribution of service times, known to the encoder, is stationary and ergodic with mean $1/\mu$ seconds. In other words, the decoder need not know the true distribution of the service times to achieve $e^{-1}\mu$ nats per second.

Consider the following definition of the cost of transmission. Suppose that the decoder has to make decisions based only on departures that occur within a certain time window. If the cost of transmission is the length of the time window of observation, then we show that we can communicate reliably at $e^{-1}\mu$ nats per second. The service times are stationary and ergodic with mean $1/\mu$ seconds. Under this new definition of the cost of transmission, we also show that $e^{-1}\mu$ nats per second is the largest rate achievable on the exponential server channel. We do this by mapping any strategy on the timing channel to an equivalent strategy with complete feedback on the point-process channel [2].

Discrete-time queues were studied in [3] and [4]. The maximum-likelihood decoder for the server with independent and geometrically distributed service times is simple. We argue that using this decoder, the capacity of the geometric server channel is achievable when the distribution of service times is stationary and ergodic with mean $1/\mu$ slots. If the cost of transmission is the length of the observation window, then we show the converse for the geometric server channel by mapping any communication strategy on this timing channel to an equivalent strategy with complete feedback on a binary memoryless channel.

Timing information can be transmitted covertly by transmitting innocuous information in the contents of packets, which may be subject to eavesdropping. Since service times corrupt information encoded in the arrival epochs of packets, we consider the following *jamming strategy* employed by the server to hamper covert communication. Every packet suffers a delay (extra service time) in addition to the nominal service time (which is stationary and ergodic with mean $1/\mu_1$ seconds). If these delays are without limits, then communication in the timing channel can be jammed completely at the expense of information throughput in packet contents. We, therefore, require that the arithmetic mean of these delays be smaller than $1/\mu_2$ seconds. We call the resulting channel the *jammed timing channel*. This channel is similar to the arbitrarily varying channel (AVC) introduced in [5]. An important distinction between the jammed timing channel and the memoryless AVC

([5]–[8] and references therein) is that in the jammed timing channel current input and delay can affect future outputs.

We prove an achievability result in the situation where the jammer does not know the true codebook in use, but knows only a distribution from which the codebook is selected. In particular, the rate $e^{-1}\mu_1\mu_2/(\mu_1 + \mu_2)$ nats per second is achievable with random codes on the jammed timing channel. When the nominal service times are independent and exponentially distributed, we argue that the rate $e^{-1}\mu_1\mu_2/(\mu_1 + \mu_2)$ nats per second is also the largest achievable with random codes, giving us a reduction in capacity by a factor $\mu_2/(\mu_1 + \mu_2)$.

We now briefly survey previous works relevant to our study. The use of the exponential server's maximum-likelihood decoder when the service times are not exponentially distributed is an instance of decoder mismatch. In the context of discrete memoryless channels (DMC), suppose that the communication system operates under a channel with transition probability matrix $W(\cdot|\cdot)$. The decoder performs maximum-likelihood decoding assuming that the DMC is characterized by $V(\cdot|\cdot)$, i.e., for a received sequence y^n, it chooses the codeword x^n that maximizes $V(y^n|x^n)$, where n is the number of uses of the channel. Reference [9] showed that using the mismatched decoder, we can communicate reliably at a rate

$$\sup_{P_X} E\left[\log \frac{V(Y|X)}{Q_Y(Y)}\right] \tag{1}$$

where Q_Y is the marginal distribution of the output under the mismatch channel V and the input distribution P_X. The expectation in (1) is with respect to the joint distribution under the true channel W and the input distribution P_X. This result was extended to discrete channels with memory in [10]. Since these results have not been proved for channels with memory that have continuous inputs and outputs, we first show the achievability of (1) for such channels and then apply this result to the timing channel. The proof, though different from the proofs in [9] and [10], is a simple extension of the proof of [11, Lemma 6.9].

Although rates possibly larger than (1) are achievable with mismatched decoding ([12]–[14] and references therein), achievability of a rate that is analogous to (1) is enough to show the results in this paper.

This paper extends the parallelism found in [1] between the exponential server timing channel and the discrete-time additive white Gaussian noise channel with an input power constraint. Consider the additive noise channel. For n uses of the channel, each codeword is a point in $\mathcal{R}^n$ having power smaller than nP. It is well known that for any stationary, ergodic, zero-mean noise process with variance σ^2, the rate $(1/2)\log[1 + P/\sigma^2]$ nats per channel use is achievable using the minimum Euclidean distance criterion for decoding. A version of this result is the direct part of [15, Theorem 1]. A stronger version of the direct part when $\sigma^2 < P$ is given in [8]. The minimum Euclidean distance criterion for decoding is the maximum-likelihood decoding when the noise is independent and has the Gaussian distribution; the capacity in this case is $(1/2)\log[1 + P/\sigma^2]$ nats per channel use. The timing channel counterparts of this result are Theorems 1 and 2 in Section II. As in [1], the analogy is rooted in the fact that the exponential

distribution and the Gaussian distribution are similar mutual information saddle points [16].

A similar result is known for a convex and compact family Θ of DMC's. For an input distribution P and a DMC W, let $I(P, W)$ denote the mutual information. Let P^* and $W^* \in \Theta$ attain the saddle point of the mutual information functional, i.e.,

$$\max_P \min_{W \in \Theta} I(P, W) = \min_{W \in \Theta} \max_P I(P, W) = I(P^*, W^*).$$

Suppose now that the channel is characterized by $W \in \Theta$. Then $I(P^*, W^*)$ is achievable over the DMC W using a maximum-likelihood decoder for the DMC with stochastic matrix W^* [17] (see also [18, Sec. IV-B–4]).

The jammed timing channel is similar in spirit to the Gaussian arbitrarily varying channel (Gaussian AVC) [8], [19], in which a jammer changes the mean of the Gaussian noise subject to a power constraint. Theorem 3 in Section II is related to results in [19] for random codes in the Gaussian AVC. The capacity of the Gaussian AVC, when the jammer knows the codebook, is known [8]. We do not know if an analogous result holds on the jammed timing channel, when the jammer knows the codebook. In the discrete-time case, however, we can apply the "elimination" technique of [6] to get a nonrandom coding strategy if a certain amount of information can be transmitted by the packet contents. Only a negligible fraction of packet contents need be used.

The rest of the paper is organized as follows. Section II states the basic definitions and results. Section II-A covers the mismatched decoding problems for the continuous-time single-server queue. Section II-B studies the jammed timing channel. Section II-C discusses the signaling channel, or the timing channel with feedback. Section II-D describes the discrete-time single-server queue. Section II-E shows the converses for the exponential and the geometric server channels. Section II-F collects several observations on our results. The proofs are in Section III.

II. DEFINITIONS AND RESULTS

A. Continuous-Time Single-Server Queue

This subsection deals with mismatched decoding for the continuous-time single-server queue without feedback. The definitions of the relevant quantities are as in [1], but written in our notation. Let $\mathcal{R}_+ = [0, \infty)$, $\mathcal{Z}_+ = \{0, 1, \cdots\}$ and $\mathcal{N} = \{1, 2, \cdots\}$. We assume that the following conditions hold.

- The queue is work-conserving, i.e., if a packet departs after service and another one is in the queue, then the server begins to serve the packet in the queue.
- The queue is initially empty, and the server follows a first-in-first-out service discipline.
- The sequence $(S_k : k \in \mathcal{N})$ of service times is a stationary and ergodic process with mean $1/\mu$ seconds.

For each $n \in \mathcal{N}$, the input to the queuing system is a vector $x^n = (x_1, \cdots, x_n)$ of n nonnegative interarrival times, such that the kth arrival occurs at time $\sum_{i=1}^k x_i$, $k = 1, \cdots, n$. The decoder observes $y^n = (y_0, y_1, \cdots, y_n)$, where $y_0 = 0$, and y_k is the time between the $(k-1)$st and the kth departures, $k = 1, \cdots, n$.

For each $n \in \mathcal{N}$, the input alphabet is $\mathcal{R}_+^n$, and the output alphabet is $\mathcal{R}_+^{n+1}$. The σ-algebras associated with the alphabets are the product Borel σ-algebras. Let $E \subset \mathcal{R}_+^{n+1}$ be a Borel set and $x^n \in \mathcal{R}_+^n$. A *transition probability function* [20, p. 315], $P_{Y^n|X^n}$, from the input space to the output space, is a mapping $(x^n, E) \rightarrow P_{Y^n|X^n}(E|x^n)$ having the following measurability properties: a) for each $x^n \in \mathcal{R}_+^n$, the mapping $E \rightarrow P_{Y^n|X^n}(E|x^n)$ is a probability measure on the output space and b) for each Borel set $E \subset \mathcal{R}_+^{n+1}$, the mapping $x^n \rightarrow P_{Y^n|X^n}(E|x^n)$ is measurable with respect to the input space. A *channel* is a sequence (parameterized by n) of transition probability functions from the input space to the output space.

Fix $n \in \mathcal{N}$. Let s_k be the service time of the kth packet, $k = 1, \cdots, n$. The observable y^n can be described as follows. Let w_k be the amount of time for which the server is idle between the $(k-1)$st departure and the kth arrival, i.e.,

$$w_k = \max \left\{ 0, \sum_{i=1}^{k} x_i - \sum_{i=0}^{k-1} y_i \right\}, \qquad k = 1, \cdots, n. \quad (2)$$

Thus if the kth arrival occurs before the $(k-1)$st departure, the idling time w_k is 0. The interdeparture times are then given by

$$y_k = \begin{cases} 0, & k = 0, \\ w_k + s_k, & k = 1, \cdots, n. \end{cases} \quad (3)$$

The stationary and ergodic process $(S_k : k \in \mathcal{N})$, (2) and (3) induce the true channel $(P_{Y^n|X^n} : n \in \mathcal{N})$, which is a sequence of transition probability functions from the input space to the output space.

Definition 1: An (n, M, T, ε)-*code* consists of a codebook of M codewords and a decoder. Each codeword is a vector of n nonnegative interarrival times $(x_1, \cdots, x_n)$. The decoder, after observing the n departures, selects the correct codeword with probability greater than $1 - \varepsilon$, under equiprobable codewords and $P_{Y^n|X^n}$. The nth departure occurs on the average (under equiprobable codewords and $P_{Y^n|X^n}$) no later than T. The rate of the code is $(\log M)/T$. Rate R is *achievable* if, for every $\gamma > 0$, there exists a sequence of $(n, M_n, T_n, \varepsilon_n)$-codes that satisfies $(\log M_n)/T_n > R - \gamma$ for all sufficiently large n, and $\lim_{n \to \infty} \varepsilon_n = 0$.

We now describe the mismatch channel $(Q_{Y^n|X^n} : n \in \mathcal{N})$ according to which the decoder performs maximum-likelihood decoding. Let a synchronizing zeroth packet be sent at $t = 0$ and interpret y_0 as the amount of unfinished work at $t = 0$, including the service time of the zeroth packet (i.e., the time at which the zeroth packet departs from the system). Let the number of packets in the queue at $t = 0$ have the equilibrium distribution that is associated with an $M/M/1$ queue [21, pp. 48–49] having input rate $\lambda' < \mu$ packets per second. The mismatch channel is then the channel induced by the process $(S_k : k \in \mathcal{N})$ that is independent and exponentially distributed with mean $1/\mu$ seconds. It will soon be clear that the decoding strategy does not depend on the parameter λ.

Let $e_\mu(x)$ denote the exponential density function $\mu e^{-\mu x}$, $x \in \mathcal{R}_+$, having mean $1/\mu$. The random variable

Y_0 has the exponential density $e_{\mu-\lambda'}(y_0)$ under $Q_{Y^n|X^n}$ [1], for every $n \in \mathcal{N}$. In contrast, $Y_0 = 0$ under $P_{Y^n|X^n}$, for every $n \in \mathcal{N}$.

Let π denote the Lebesgue measure (the argument will indicate the appropriate space). Fix $n \in \mathcal{N}$. Using (2) and (3), and the density for exponentially distributed service times, $Q_{Y^n|X^n}$ can be written as

$$dQ_{Y^n|X^n}(y^n|x^n) = d\pi(y^n)p(x^n, y^n) \quad (4)$$

for every $x^n \in \mathcal{R}_+^n$, where

$$p(x^n, y^n) \triangleq e_{\mu-\lambda'}(y_0) \prod_{k=1}^{n} e_\mu(y_k - w_k), \qquad \lambda' < \mu. \quad (5)$$

Let the distribution P_{X^n} on the input space be given by

$$dP_{X^n}(x^n) = d\pi(x^n) \prod_{k=1}^{n} e_{\lambda'}(x_k), \qquad \lambda' < \mu. \quad (6)$$

This is the distribution of the first n arrivals induced by the Poisson arrival process with rate λ'. Let Q_{X^n, Y^n} denote the joint distribution under the input distribution P_{X^n} (cf. (6)) and $Q_{Y^n|X^n}$ (cf. (4)). The joint distribution Q_{X^n, Y^n} can then be written unambiguously as

$$dQ_{X^n, Y^n}(x^n, y^n) = dP_{X^n}(x^n)\, d\pi(y^n)p(x^n, y^n) \quad (7)$$

due to Fubini's Theorem [22, Theorem 18.3, p. 238]. Let Q_{Y^n} denote the marginal distribution of $Y^n = (Y_0, \cdots, Y_n)$ under Q_{X^n, Y^n}. Let $P_{X^n} \times Q_{Y^n}$ denote the joint distribution under which the random variables X^n and Y^n are independent, and have marginal distributions P_{X^n} and Q_{Y^n}, respectively.

As a consequence of (7), we have that

$$Q_{X^n, Y^n} \ll P_{X^n} \times Q_{Y^n}$$

[23, Corollary 5.3.1, p. 112], and that $Q_{Y^n} \ll \pi_{Y^n}$. A version of the Radon–Nikodym derivative $dQ_{X^n, Y^n}/d(P_{X^n} \times Q_{Y^n})$ is the function f given by

$$f(x^n, y^n) = \begin{cases} p(x^n, y^n)/m(y^n), & \text{if } m(y^n) > 0 \\ 1, & \text{if } m(y^n) = 0 \end{cases} \quad (8)$$

where

$$m(y^n) \triangleq \int_{\mathcal{R}_+^n} dP_{X^n}(x^n)p(x^n, y^n), \qquad y^n \in \mathcal{R}_+^{n+1}. \quad (9)$$

We can easily verify that

$$dQ_{Y^n}(y^n) = d\pi(y^n)m(y^n). \quad (10)$$

Clearly, the function f (cf. (8)) satisfies

$$\int_{\mathcal{R}_+^n} dP_{X^n}(x^n)f(x^n, y^n) = Ef(X^n, y^n) = 1 \quad (11)$$

for every $y^n \in \mathcal{R}_+^{n+1}$. The output of an $M/M/1$ system with input rate $\lambda' < \mu$ is a Poisson process with rate λ' (see, for e.g., [21, Fact 2.8.2, p. 60]). Consequently, under Q_{X^n, Y^n}, the

random vector $(Y_1, \cdots, Y_n)$ is a vector of independent and exponentially distributed random variables with mean $1/\lambda'$ seconds, i.e.,

$$dQ_{Y_1, \cdots, Y_n}(y_1, \cdots, y_n) = d\pi(y_1, \cdots, y_n) \prod_{k=1}^{n} e_{\lambda'}(y_k). \tag{12}$$

We will use (10), (11) and (12) in the proof of our results.

We now describe the mismatched decoder. The decoder makes a decision based on f (cf. (8)) as follows. Let the codebook be $\{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_M\}$, where $\boldsymbol{x}_i \in \mathcal{R}_+^n$ for $i = 1, \cdots, M$. The decoder $\phi_f : \mathcal{R}_+^{n+1} \to \{0, 1, \cdots, M\}$ maps the observed interdeparture times y^n to

$$\phi_f(y^n) \triangleq \begin{cases} i, & \text{if } \max_{j \neq i} f(\boldsymbol{x}_j, y^n) < f(\boldsymbol{x}_i, y^n) \\ 0, & \text{if no such } i \text{ exists.} \end{cases} \tag{13}$$

We interpret the output 0 as an error in decoding.

From Lemma 2 in Section III, $m(y^n) = 0$ if and only if $y_0 + y_1 = 0$, in which case $\phi_f(y^n) = 0$. But $m(Y^n) = 0$ with zero probability under Q_{Y^n}. When $m(y^n) > 0$, which is the case almost always, the decoder ϕ_f tries to pick the unique codeword that maximizes $p(\cdot, y^n)$ (cf. (5)). This is the same as picking the unique codeword (among the compatible ones) that minimizes the sum of service times, $\sum_{k=1}^{n} (y_k - w_k)$, or, equivalently, maximizes the sum of idling times of the server, $\sum_{k=1}^{n} w_k$. When the decoder cannot find such a unique codeword, it declares 0, an error. The only functions required to make this decision are additions, subtractions, and comparisons. Although these functions are simple, $\sum_{k=1}^{n} w_k$ must be evaluated for every codeword before a decision is made. Since the number of codewords is exponential in time, the number of operations performed to decode is exponential in time.

Theorem 1: Let the queue be work-conserving and initially empty. Let the server follow a first-in-first-out service discipline with stationary and ergodic service times of mean $1/\mu$ seconds. The rate $e^{-1}\mu$ nats per second is then achievable using the decoding rule in (13).

The result [1, Theorem 7] on the achievability of $e^{-1}\mu$ nats per second for i.i.d. service times is a special case of Theorem 1. To transmit reliably at $e^{-1}\mu$ nats per second on such a channel, maximum-likelihood decoding is not required; the decoder ϕ_f is sufficient. This decoder is therefore robust to the distribution of service times. The decoder's robustness, however, does not imply that a single sequence of codes works for all stationary and ergodic distributions of the service times. Furthermore, Theorem 1 does not give rates at which the probability of error goes to zero. The term "robust" should therefore be interpreted with caution. We only argue that, knowing the true channel, a sequence of good codes with decoder ϕ_f and rate close to $e^{-1}\mu$ nats per second can be selected.

Suppose that the codebook is such that for every codeword, the last arrival occurs before $t = T'$. Then the decoder ϕ_f need not observe departures beyond $t = T'$. This is because of the following. Suppose that y^n satisfies $\sum_{k=0}^{n} y_k > T'$. Given any candidate codeword, the server is not idle beyond T', i.e., the quantity that is required to make a decision, $\sum_{k=1}^{n} w_k$, can be evaluated upon observation of departures in the time window

$[0, T']$. Departures in $[0, T']$ therefore constitute a set of sufficient statistics for determining the input codeword. This is not surprising because of the memoryless property of exponential service times.

Now suppose that the decoder observes only the departures that occur in $[0, T]$, where T is known to the encoder. Clearly, it is useless to have arrivals after time T. This motivates the following definition.

Definition 2: An (n, M, T, ε)-*window-code* consists of a codebook of M codewords and a decoder. Each codeword is a vector of n nonnegative interarrival times $(x_1, \cdots, x_n)$. The nth arrival of every codeword occurs before time T. The decoder, after observing departures in $[0, T]$, selects the correct codeword with probability greater than $1 - \varepsilon$, under equiprobable codewords and $P_{Y^n | X^n}$. The rate of the window-code is $(\log M)/T$. Rate R is *achievable with window-codes* if, for every $\gamma > 0$, there exists a sequence of $(n, M_n, T_n, \varepsilon_n)$-window-codes that satisfies

$$(\log M_n)/T_n > R - \gamma$$

for all sufficiently large n, and $\lim_{n \to \infty} \varepsilon_n = 0$.

Thus the term "window-code" refers to a code whose decoder observes only those departures that occur in the window $[0, T]$; the cost of transmission for the window-code is T seconds. In contrast, the code in Definition 1 has a decoder that observes all the n departures; the cost of transmission in that case is the expected time of the nth departure, i.e., the expected time the decoder waits to gather the output data.

Theorem 2: Let the queue be work-conserving and initially empty. Let the server follow a first-in-first-out service discipline with stationary and ergodic service times of mean $1/\mu$ seconds. The rate $e^{-1}\mu$ nats per second is then achievable with window-codes using the decoding rule in (13).

B. Jammed Timing Channel—Random Codes

We now consider the jammed timing channel where the server acts as an adversary. The queue is initially empty, and the server follows a first-in-first-out service discipline. The process $(S_k : k \in \mathcal{N})$ of nominal service times is stationary and ergodic with mean $1/\mu_1$ seconds. Fix $n \in \mathcal{N}$. The server (jammer) includes a delay of z_k seconds to the service time of the kth packet, $k = 1, \cdots, n$. We call $\boldsymbol{z} = (z_1, \cdots, z_n) \in \mathcal{R}_+^n$ the *state sequence*, because it determines the state of the channel. The resulting service time for the kth packet is $S_k + z_k$ seconds, $k = 1, \cdots, n$. If no constraints are imposed on the state sequence, communication in the timing channel can be jammed completely at the expense of information throughput in packet contents. We impose the following constraint. For a code with n packets, we allow only those state sequences that satisfy

$$l(\boldsymbol{z}) \triangleq \frac{1}{n} \sum_{k=1}^{n} z_k \leq \frac{1}{\mu_2}$$

i.e., a total delay of at most n/μ_2 seconds is allowed for all the n packets. Each state sequence $\boldsymbol{z}$ induces a transition probability

function from the input space to the output space, denoted by $W^n(\boldsymbol{z})$. We need communication strategies that perform well for every state sequence $\boldsymbol{z}$ that satisfies $l(\boldsymbol{z}) \leq 1/\mu_2$.

The problem of finding achievable rates for deterministic codes, i.e., when the codebook is known to the jammer, appears to be nontrivial. Instead of fixing a single good (deterministic) codebook, we allow communication strategies with random codes. The encoder chooses the codebook that is used for transmission from a set of possible codebooks. The decoder knows this selection. The jammer, however, is ignorant of the selected codebook. Its partial knowledge is modeled by a distribution on the set of codebooks. Such a code is usually called in the AVC literature, somewhat deceptively, a *random code*.

Given a selected codebook $\boldsymbol{c}$, the decoder is ϕ_f (cf. (8) and (13)). For the codebook $\boldsymbol{c}$, the average probability of error (over equiprobable codewords) is denoted by $P_e(\boldsymbol{c}, \phi_f, W^n(\boldsymbol{z}))$ when the state sequence is $\boldsymbol{z}$.

Let $\boldsymbol{C}$ be a random variable taking values in the family of all codebooks that have M codewords, and such that the nth arrival in each codeword occurs before T. The parameters (n, M, T) of the random variable $\boldsymbol{C}$ will be clear from the context. The following definition is an extension of window-codes and achievability with window-codes (Definition 2) for the jammed timing channel.

Definition 3: An (n, M, T, ε)-*random window-code* consists of a probability distribution for $\boldsymbol{C}$, and a decoder ϕ that depends on the codebook realization. Each realization $\boldsymbol{c}$ is a set of M codewords. Each codeword is a vector of n nonnegative interarrival times. The nth arrival of every codeword occurs before time T. The decoder, knowing the codebook realization $\boldsymbol{c}$, makes a decision after observing departures in $[0, T]$. The average probability of error satisfies $E[P_e(\boldsymbol{C}, \phi, W^n(\boldsymbol{z}))] \leq \varepsilon$ for every $\boldsymbol{z}$ with $l(\boldsymbol{z}) \leq 1/\mu_2$. Rate R is *achievable with random window-codes* if, for every $\gamma > 0$, there exists a sequence of $(n, M_n, T_n, \varepsilon_n)$-random window-codes that satisfies $(\log M_n)/T_n > R - \gamma$ for all sufficiently large n, and $\lim_{n \to \infty} \varepsilon_n = 0$.

Theorem 3: On the jammed timing channel, the rate $e^{-1}\mu_1\mu_2/(\mu_1 + \mu_2)$ nats per second is achievable with random window-codes using the decoding rule in (13).

C. Signaling Channel

In a telephone signaling channel [1], the encoder knows exactly when the packet is removed from service, i.e., complete feedback is available. The encoder can make use of this information to avoid queuing altogether, and the resulting channel is a simple additive noise channel.

On this channel, the rate $e^{-1}\mu$ nats per second is clearly achievable in the presence of complete feedback. Indeed, the encoder can ignore feedback completely, and use a code suggested by Theorem 1. Making use of feedback, however, leads to another decoder that achieves $e^{-1}\mu$ nats per second. It will be clear from the following definition that feedback is used only to avoid queuing.

The sequence of service times is stationary and ergodic with mean $1/\mu$ seconds. An (n, M, T, ε)-*feedback code* consists of

codebook of M codewords and a decoder. Each message is an n-vector $(x_1, \cdots, x_n)$ of positive real numbers. The first arrival occurs at $t = x_1$. The kth component, x_k, is the amount of time the encoder will wait after the $(k-1)$st departure, before sending the kth arrival, $k = 2, \cdots, n$. The last packet exits on the average before T. The encoder thus makes use of feedback to avoid queuing and to control completely the idling times of the server. Feedback however is not used to choose the waiting times x_k. The rate of the feedback code is $(\log M)/T$ nats per second. The decoder, after observing the interdeparture times $y^n = (y_1, \cdots, y_n)$, makes the correct decision with probability larger than $1 - \varepsilon$ when averaged over equiprobable codewords.

Rate R is *achievable with feedback* if, for every $\gamma > 0$, there exists a sequence of $(n, M_n, T_n, \varepsilon_n)$-feedback codes that satisfies $(\log M_n)/T_n > R - \gamma$ for all sufficiently large n, and $\lim_{n \to \infty} \varepsilon_n = 0$.

The decoder chooses the codeword that explains the received sequence of departures with the minimum sum of service times. For a candidate codeword, $\boldsymbol{x} = (x_1, \cdots, x_n)$, let

$$d(\boldsymbol{x}, y^n) \triangleq \sum_{k=1}^n d'(y_k - x_k)$$

where

$$d'(r) \triangleq \begin{cases} r, & \text{if } r \geq 0 \\ +\infty, & \text{if } r < 0. \end{cases}$$

Given the codebook $\{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_M\}$, the decoder φ_d maps y^n to

$$\varphi_d(y^n) \triangleq \begin{cases} i, & \text{if } d(\boldsymbol{x}_i, y^n) < \min_{j \neq i} d(\boldsymbol{x}_j, y^n) \\ 0, & \text{if no such } i \text{ exists} \end{cases} \quad (14)$$

where an output 0 is interpreted as an error in decoding.

Theorem 4: Let the queue be work-conserving and initially empty. Let the server follow a first-in-first-out service discipline with stationary and ergodic service times of mean $1/\mu$ seconds. The rate $e^{-1}\mu$ nats per second is then achievable with feedback using the decoding rule in (14).

We remark that Theorem 4 is not implied by previous results on mismatched decoding. In particular, we cannot apply [12, Theorem 2] because it precludes input distributions with infinite differential entropy. For the input distribution that we choose, differential entropy does not exist. This input distribution is the one that attains the mutual information saddle point [16, Theorem 1].

D. Discrete-Time Single-Server

In this subsection, we describe the discrete-time single-server queuing system [3], [4], and state the counterparts of Theorems 1, 2, and 3. The proofs are omitted because they are analogous to the continuous-time case.

In the discrete-time model, arrivals and departures occur only at integer-valued epochs, called slots. At most one packet arrives and at most one packet departs in each slot. Unserved packets are stored in a queue. The queue is work-conserving, and the server follows a first-in-first-out service discipline. The sequence $(S_k : k \in \mathcal{N})$ of nominal service times is an $\mathcal{N}$-valued,

stationary and ergodic process with mean $1/\mu_1$ slots, $0 < \mu_1 < 1$. Each packet requires at least one slot of service.

For each $n \in \mathcal{N}$, the input is a vector of n interarrival times, $x^n = (x_1, \cdots, x_n)$. The decoder observes $y^n = (y_0, y_1, \cdots, y_n)$, where $y_0 = 1$, and y_k is the time (in slots) between the $(k-1)$st and the kth departures, $k = 1, \cdots, n$. We set $y_0 = 1$ because $x_1 \geq 1$ and $s_1 \geq 1$, i.e., the first slot does not give any information about the transmitted message. The input alphabet is $\mathcal{N}^n$ and the output alphabet is $\mathcal{N}^{n+1}$. The σ-algebras associated with the alphabets are the collection of all the corresponding subsets.

Fix $n \in \mathcal{N}$. Given the sequence of service times $(s_1, \cdots, s_n) \in \mathcal{N}^n$, the interdeparture times in $y^n = (y_0, y_1, \cdots, y_n)$ are

$$
y_k = \begin{cases} 1, & k = 0 \\ s_k + w_k, & k = 1, \cdots, n \end{cases} \tag{15}
$$

where

$$
w_k = \max\left\{ 0, \sum_{j=1}^{k} x_j - \sum_{j=0}^{k-1} y_j \right\}
$$

is the server's idling time before serving the kth packet. The stationary and ergodic process $(S_k: k \in \mathcal{N})$ and (15) induce the true channel $(P_{Y^n|X^n}: n \in \mathcal{N})$.

The definitions of (n, M, T, ε)-code, achievability, (n, M, T, ε)-window-code and achievability with window-codes are analogous to those in Definitions 1 and 2.

Fix $n \in \mathcal{N}$. For the jammed timing channel (discrete time), the state sequence $\boldsymbol{z} = (z_1, \cdots, z_n) \in \mathcal{Z}_+^n$ satisfies the constraint

$$
l(\boldsymbol{z}) \triangleq \left(\sum_{k=1}^{n} z_k \right) \bigg/ n \leq 1/\mu_2, \qquad \mu_2 > 0.
$$

As in the continuous-time case, each $\boldsymbol{z}$ induces a transition probability function $W^n(\boldsymbol{z})$ from the input space to the output space.

The definitions of (n, M, T, ε)-random window-code and achievability with random window-codes are analogous to those in Definition 3.

We now describe the mismatch channel $(Q_{Y^n|X^n}: n \in \mathcal{N})$ based on which the decoder performs maximum-likelihood decoding. We say that a random variable X has the $\mathrm{Geo}^+(\lambda)$ distribution, $0 < \lambda < 1$, if

$$
P\{X = x\} \triangleq g_\lambda(x) \triangleq \lambda(1 - \lambda)^{x-1}, \qquad x \in \mathcal{N}.
$$

Let a synchronizing zeroth packet be sent at $t = 0$ and interpret y_0 as the amount of unfinished work at $t = 0$, including the service time of the zeroth packet. Let the number of packets in the queue at $t = 0$ have the equilibrium distribution that is associated with the queue having $\mathrm{Geo}^+(\lambda')$-distributed arrivals, $0 < \lambda' < \mu_1 < 1$, and $\mathrm{Geo}^+(\mu_1)$-distributed service times. This queuing system is the discrete-time counterpart of the $M/M/1$ system. The mismatch channel is then the channel induced by the process $(S_k: k \in \mathcal{N})$ of independent and $\mathrm{Geo}^+(\mu_1)$-distributed service times. Fix $n \in \mathcal{N}$. Using (15), we

see that the mismatch transition probability function $Q_{Y^n|X^n}$ is the probability mass function (pmf) on $\mathcal{N}^{n+1}$ given by

$$
Q_{Y^n|X^n}(y^n|x^n) = g_{\mu-\lambda'}(y_0) \prod_{k=1}^{n} g_\mu(y_k - w_k)
$$

for every $x^n \in \mathcal{N}^n$. Let the pmf on the input alphabet $\mathcal{N}^n$ be

$$
P_{X^n}(x^n) = \prod_{k=1}^{n} g_{\lambda'}(x_k), \quad 0 < \lambda' < \mu_1 < 1.
$$

For $(x^n, y^n) \in \mathcal{N}^n \times \mathcal{N}^{n+1}$, let

$$
f(x^n, y^n) \triangleq Q_{Y^n|X^n}(y^n|x^n)/m(y^n) \tag{16}
$$

where

$$
m(y^n) \triangleq \sum_{x^n \in \mathcal{N}^n} P_{X^n}(x^n) \cdot Q_{Y^n|X^n}(y^n|x^n), \quad y^n \in \mathcal{N}^{n+1}.
$$

The function f satisfies $Ef(X^n, y^n) = 1$ for every $y^n \in \mathcal{N}^{n+1}$; the expectation is with respect to P_{X^n}. Given a codebook with M codewords, the decoder is the function $\phi_f: \mathcal{N}^{n+1} \to \{0, 1, \cdots, M\}$ defined in (13) with f as in (16).

Theorem 5: Let the discrete-time queue be work-conserving and initially empty. Let the server follow a first-in-first-out service discipline with stationary and ergodic nominal service times of mean $1/\mu_1$ slots. The following statements then hold.

a) The rate $\log\left[1 + \mu_1(1 - \mu_1)^{(1-\mu_1)/\mu_1}\right]$ nats per slot is achievable using the (discrete-time) decoding rule in (13).

b) The rate $\log\left[1 + \mu_1(1 - \mu_1)^{(1-\mu_1)/\mu_1}\right]$ nats per slot is achievable with window-codes using the decoding rule in (13).

c) On the jammed timing channel, the rate $\log\left[1 + \mu(1 - \mu)^{(1-\mu)/\mu}\right]$ nats per slot, where $\mu = \mu_1\mu_2/(\mu_1 + \mu_2)$, is achievable with random window-codes using the decoding rule in (13).

On the discrete-time jammed timing channel, if each packet carries a nonzero amount of information, we can apply the *elimination* technique of [6] to get a nonrandom communication strategy that has rate $\log\left[1 + \mu(1 - \mu)^{(1-\mu)/\mu}\right]$ nats per slot. The first step in this technique is *random code reduction* [11, Lemma 6.8], which we now describe.

Given a codebook $\boldsymbol{c}$ and the decoder ϕ_f, let $P_{e,i}(\boldsymbol{c}, \phi_f, W^n(\boldsymbol{z}))$ be the probability of error when the state sequence is $\boldsymbol{z}$ and the transmitted message is i, where $1 \leq i \leq M$. In the rest of this subsection, let the definition of (n, M, T, ε)-random window-code be analogous to that in Definition 3 with the condition on the average probability of error replaced by

$$
\max_{1 \leq i \leq M_n} E[P_{e,i}(\boldsymbol{C}, \phi_f, W^n(\boldsymbol{z}))] \leq \varepsilon
$$

i.e., a condition on the maximum probability of error.

We can easily modify the proof to show the following extension to Theorem 5 c). On the jammed timing channel, for every $\gamma > 0$, there exists a sequence of $(n, M_n, T_n, \varepsilon_n)$-random window-codes that satisfies

i) $(\log M_n)/T_n > \log\left[1 + \mu(1 - \mu)^{(1-\mu)/\mu}\right] - \gamma$

for all sufficiently large n;

ii) $\max\limits_{\boldsymbol{z}:\, l(\boldsymbol{z})\leq 1/\mu_2}\ \max\limits_{1\leq i\leq M_n} E[P_{e,i}(\boldsymbol{C},\phi_f,W^n(\boldsymbol{z}))]\leq\varepsilon_n$

for every $n\in\mathcal{N}$, and

iii) $\lim_{n\to\infty}\varepsilon_n = 0$.

Now suppose that the error probabilities need not vanish. Fix $n\in\mathcal{N}$. On the discrete-time channel, the cardinality of $\{W^n(\boldsymbol{z}):l(\boldsymbol{z})\leq 1/\mu_2\}$ is upper-bounded by $(1+n/\mu_2)^n$. We can therefore apply random code reduction [11, Lemma 6.8] to get the following. Given $\varepsilon>0$ and $\gamma>0$, for all sufficiently large n, we can find a set of n^2 codebooks $\{\boldsymbol{c}_j:j=1,\cdots,n^2\}$, where each codebook has parameters (n,M_n,T_n), the set of codebooks satisfies

$$\max_{\boldsymbol{z}:\, l(\boldsymbol{z})\leq 1/\mu_2}\ \max_{1\leq i\leq M_n}\frac{1}{n^2}\sum_{j=1}^{n^2} P_{e,i}(\boldsymbol{c}_j,\phi_f,W^n(\boldsymbol{z}))<\varepsilon\quad (17)$$

and

$$\frac{\log M_n}{T_n} > \log\left[1+\mu(1-\mu)^{(1-\mu)/\mu}\right]-\gamma.$$

If each packet carries C_0 nats of information, $C_0>0$, then we can employ the elimination technique of [6] as follows. Fix $n\in\mathcal{N}$. Let the set of equiprobable messages be $\{1,\cdots,n^2\}\times\{1,\cdots,M_n\}$. Given a message (j,i) from this set, choose the codebook $\boldsymbol{c}_j$. Transmit i on the jammed timing channel using codebook $\boldsymbol{c}_j$. Convey the codebook index j to the receiver using the first $2\log n$ nats (of the nC_0 nats) of packet contents. We thus use only a negligible fraction, $(2\log n)/(nC_0)$, of the packet contents. The *average* probability of error over equiprobable codewords is smaller than ε for this (nonrandom) communication strategy because of (17).

E. Converses

In this subsection we state converse results for the continuous-time (resp., discrete-time) queue with independent and exponentially (resp., geometrically) distributed service times. Converses to Theorems 1, 4, and 5 a) were shown in [1] and [3].

Theorem 6: For the continuous-time system, let the queue be work-conserving and initially empty. Furthermore, let the nominal service times be independent and exponentially distributed with mean $1/\mu_1$ seconds.

a) The largest rate achievable with window-codes is $e^{-1}\mu_1$ nats per second.

b) On the jammed timing channel, the largest rate achievable with random window-codes is $e^{-1}\mu_1\mu_2/(\mu_1+\mu_2)$ nats per second.

Similarly, for the discrete-time system, let the nominal service times be independent and $\mathrm{Geo}^+(\mu_1)$-distributed with mean $1/\mu_1$ slots.

c) The largest rate achievable with window-codes is $\log\left[1+\mu_1(1-\mu_1)^{(1-\mu_1)/\mu_1}\right]$ nats per slot.

d) On the jammed timing channel, the largest rate achievable with random window-codes is $\log\left[1+\mu(1-\mu)^{(1-\mu)/\mu}\right]$ nats per slot, where $\mu=\mu_1\mu_2/(\mu_1+\mu_2)$.

Proofs of a) and b) are in Section III. The key idea in the proof of a) is to map any window-code on the timing channel to an equivalent strategy with complete feedback on the point-process channel. We now prove c). We omit the proof of d) because it is analogous to the proof of b).

Proof of c): The service times are independent and have the $\mathrm{Geo}^+(\mu_1)$ distribution. The key idea here is to map any window-code on the timing channel to a strategy with complete feedback on a binary memoryless channel.

Fix $n\in\mathcal{N}$. Suppose that the codebook is designed to transmit M_n messages. Each message maps to a codeword $(x_1,\cdots,x_n)\in\mathcal{N}^n$ of interarrival times that satisfies

$$\sum_{k=1}^{n} x_k < T_n.$$

The decoder observes departures until slot T_n. Let $1\{\cdot\}$ denote the indicator function of an event. Let the output be the binary-valued vector $(v_1,\cdots,v_{T_n})$ given by

$$v_k = 1\{\text{Departure in slot } k\},\qquad k=1,\cdots,T_n.\quad (18)$$

Clearly, $v_1=0$ because $x_1\geq 1$ and each packet requires at least one slot of service.

Fix a codeword $(x_1,\cdots,x_n)$. Let $(A_k:1\leq k<T_n)$ be the cumulative arrival process;

$$A_k = \sum_{i=1}^{k} 1\{x_1+\cdots+x_i\leq k\},\qquad 1\leq k<T_n$$

denotes the number of arrivals in the first k slots. Analogously, the cumulative departure process is $(D_k:1\leq k<T_n)$, where $D_k=\sum_{i=1}^{k} v_i$ is the number of departures in the first k slots, $1\leq k<T_n$. The number of packets that remain in the system at the end of the kth slot is A_k-D_k, $1\leq k<T_n$.

If $A_k-D_k=0$, the queue is empty at the end of the kth slot, and hence no packet exits in the $(k+1)$st slot. If $A_k-D_k>0$, a packet is served in the $(k+1)$st slot. Using the memoryless property of geometric service times, this packet departs in the $(k+1)$st slot with probability μ_1, or stays in the system with probability $1-\mu_1$, independent of the past.

The timing channel therefore behaves like a binary memoryless Z-channel, W, with $W(1|1)=\mu_1$ and $W(1|0)=0$. The inputs to the Z-channel are $u_1=0$, and $u_{k+1}=1\{A_k-D_k>0\}$, $k=1,\cdots,T_n-1$. The output sequence is $(v_1,\cdots,v_{T_n})$ given by (18).

Any window-code on the timing channel is therefore equivalent to the above strategy with complete feedback on the memoryless Z-channel. Complete feedback is necessary because the $(k+1)$st input, u_{k+1}, depends on the past departures (outputs) through D_k.

The capacity of the timing channel (for window-codes) is therefore upper-bounded by the capacity of the memoryless Z-channel with complete feedback. Feedback does not increase the capacity of the memoryless Z-channel; the upper bound is, therefore,

$$\max_{0\leq p\leq 1}\left[h(p\mu_1)-ph(\mu_1)\right]=\log\left[1+\mu_1(1-\mu_1)^{(1-\mu_1)/\mu_1}\right].$$

This completes the proof. $\qquad\square$

We can, in fact, say more about the converse. For window-codes with rate above

$$\log\left[1 + \mu_1 (1 - \mu_1)^{(1-\mu_1)/\mu_1}\right]$$

nats per slot, the probability of error goes to 1. This follows from the strong converse for DMC's with feedback [11 p. 2.5.16(c)].

F. Discussion

Theorems 1 and 2 show that the exponential server channel's maximum-likelihood decoder ϕ_f (cf. (8) and (13)) is a robust decoder. Suppose that the service times are stationary and ergodic with mean $1/\mu$ seconds. When the cost of transmission is the expected departure time of the last packet, the rate $e^{-1}\mu$ nats per second is *achievable* using the decoder ϕ_f (Theorem 1). A *window-code* is one where the decoder makes a decision based on departures in a certain time window, and all arrivals fall within this time window. The decoder ϕ_f does not have to look beyond the time window to make a decision. The rate $e^{-1}\mu$ nats per second is *achievable with window-codes* using the decoder ϕ_f (Theorem 2). Furthermore, when the service times are independent and exponentially distributed, this rate is the largest achievable with window-codes (Theorem 6 a)]. We prove this result by mapping any window-code on the timing channel to an equivalent strategy with complete feedback on the point-process channel. Using feedback on the timing channel to avoid queuing, the rate $e^{-1}\mu$ nats per second is *achievable with feedback* using the decoder φ_d (Theorem 4). The mutual information saddle-point inequality plays a crucial role in the proof of our achievability results under mismatch.

On the jammed timing channel, the jammer (server) includes an arbitrary amount of delay to the service time of each packet. This is done to diminish the transmission capabilities of the timing channel. The total delay for all the n packets cannot exceed n/μ_2 seconds. The nominal service times are stationary and ergodic with mean $1/\mu_1$ seconds. Let $\mu = \mu_1\mu_2/(\mu_1 + \mu_2)$. The rate $e^{-1}\mu$ nats per second is *achievable with random window-codes* using the decoder ϕ_f (Theorem 3). Furthermore, when the service times are independent and exponentially distributed, $e^{-1}\mu$ nats per second is the largest rate achievable with random window-codes (Theorem 6 b)).

Analogous results hold for the discrete-time single-server queuing system (Theorems 5 and 6 c), d)). Furthermore, if each packet carries a nonzero amount of information, there is a nonrandom communication strategy to transmit information reliably on the jammed timing channel (cf. discussion following Theorem 5). This strategy uses only a negligible fraction of packet contents. We do not know if a similar result holds for the continuous-time system. Suppose that the jammer is now aware of the codebook in use, and there is no side channel available. We do not know the (deterministic-code) capacity of such a jammed timing channel.

We conclude this section with the following observation. We can map any window-code on the timing channel to an equivalent strategy with complete feedback on the point-process channel (binary memoryless Z-channel in the discrete-time case). Theorem 2 (resp., Theorem 5 b)) therefore gives us an alternative capacity achieving strategy with complete feedback

on the point-process channel (resp., Z-channel). Furthermore, it is well known that the capacities of the point-process channel and the discrete memoryless channel do not increase with feedback. This fact gives a simple explanation of why the capacity of the exponential server (resp., geometric server) channel does not increase with feedback.

III. PROOFS

In this section we prove Theorems 1–4 and 6 a) and b). We begin with a simple extension of [11, Lemma 6.9]. We provide the proof because of some minor variations in the statement of the lemma and its applicability to standard measurable spaces. A *standard measurable space* is a measurable space $(A, \mathcal{A})$ that is isomorphic to $(F, \mathcal{F})$, where F is a Borel subset of $[0, 1]$ and $\mathcal{F}$ is the Borel σ-algebra on F.

Let $(A, \mathcal{A})$ and $(B, \mathcal{B})$ be two standard measurable spaces. Let $P_{Y|X}$ be a transition probability function from $(A, \mathcal{A})$ to $(B, \mathcal{B})$. Let P_X be a probability measure on $(A, \mathcal{A})$. P_X and $P_{Y|X}$ induce a joint probability measure $P_{X,Y}$ [22, P. 18.25 (b), p. 247], and a marginal probability measure P_Y [22, P. 18.25 (d), p. 247], on the appropriate spaces.

Let $\boldsymbol{c} = \{x_1, \cdots, x_M\}$ be a codebook of M codewords, $x_i \in A$ for $i = 1, \cdots, M$. Let $g: A \to \mathcal{R}_+$ be a measurable function that represents the constraint on the inputs. For a fixed $\Gamma \in \mathcal{R}_+$, we require that $g(x) \leq \Gamma$ for every $x \in \boldsymbol{c}$. Fix a set $H \in \mathcal{B}$. Let $f: A \times B \to \mathcal{R}_+$ be a measurable function. Let $\phi_{f,H}: B \to \{0, 1, \cdots, M\}$ denote the mapping

$$\phi_{f,H}(y) \triangleq \begin{cases} i, & \text{if } y \in H \text{ and } \max_{j \neq i} f(x_j, y) < f(x_i, y) \\ 0, & \text{if } y \notin H \text{ or if no such } x_i \text{ exists in } \boldsymbol{c}. \end{cases}$$

In other words, when $y \in H$, the decoder looks for a unique codeword that maximizes $f(\cdot, y)$. Given a codebook $\boldsymbol{c}$, the encoder and the decoder are thus fixed. An error occurs whenever the decoder declares a symbol that is different from the transmitted symbol. Let $P_{e,i}(\boldsymbol{c}, \phi_{f,H}, P_{Y|X})$ denote the probability of error when the codebook is $\boldsymbol{c}$ and the transmitted message is i.

Lemma 1: Let $(A, \mathcal{A})$ and $(B, \mathcal{B})$ be standard alphabet spaces. Let $\Gamma \in \mathcal{R}_+$ and $\delta \in (0, 1)$. Let P_X be a probability measure on $(A, \mathcal{A})$ that satisfies

$$P_X\{g(X) \leq \Gamma\} > 1 - \delta. \tag{19}$$

Let $f: A \times B \to \mathcal{R}_+$ be a measurable function that satisfies

$$Ef(X, y) = 1 \tag{20}$$

for every $y \in B$. Let $H \in \mathcal{B}$. There exists a random variable $\boldsymbol{C}$ that takes values in the set of codebooks with M codewords of block length 1, such that for any realization $\boldsymbol{c}$ of this random variable, $g(x) \leq \Gamma$ for every $x \in \boldsymbol{c}$. Furthermore, for every $\beta > 0$, every $i \in \{1, \cdots, M\}$

$$EP_{e,i}(\boldsymbol{C}, \phi_{f,H}, P_{Y|X}) \leq \frac{P_{X,Y}\{f(X, Y) \leq \beta\}}{1-\delta} + \frac{M}{\beta(1-\delta)^2} + \frac{P_Y\{Y \notin H\}}{1-\delta}.$$

Proof: Let

$$A_\Gamma = \{x \in A : g(x) \le \Gamma\}$$

and

$$\mathcal{A}_\Gamma = \{G \cap A_\Gamma : G \in \mathcal{A}\}.$$

From (19), $P_X\{A_\Gamma\} > 1 - \delta$. Let $P_X^{A_\Gamma}$ be the restriction of P_X to $(A_\Gamma, \mathcal{A}_\Gamma)$. Let P_X' be the probability measure on $(A_\Gamma, \mathcal{A}_\Gamma)$ given by

$$dP_X'(x) = \frac{dP_X^{A_\Gamma}(x)}{P_X\{A_\Gamma\}}, \qquad x \in A_\Gamma. \tag{21}$$

Fix $M \in \mathcal{N}$. Each codebook is an M-tuple of elements from A_Γ. Let the codebook random variable $\boldsymbol{C} = (X_1, \cdots, X_M)$ have the product distribution

$$dP_{\boldsymbol{C}}(x_1, \cdots, x_M) = dP_X'(x_1)dP_X'(x_2)\cdots dP_X'(x_M) \tag{22}$$

on $(A_\Gamma^M, \mathcal{A}_\Gamma^M)$, where $\mathcal{A}_\Gamma^M$ is the product Borel σ-algebra on A_Γ^M. Each codeword is therefore drawn independently from A_Γ according to P_X'. Clearly, for any realization $\boldsymbol{c}$ of the codebook random variable, $g(x) \le \Gamma$ for every $x \in \boldsymbol{c}$.

Consider an auxiliary threshold decoder that declares i as the transmitted message only if, for the received $y \in H$, x_i is the only codeword that satisfies $f(x_i, y) > \beta$. Otherwise, this auxiliary decoder declares 0, an error. Whenever this auxiliary decoder declares a nonzero symbol, it agrees with the output of the decoder $\phi_{f,H}$. The auxiliary decoder therefore performs worse than $\phi_{f,H}$.

The error probability given that message i is transmitted, averaged over the random selection of the codebook, is the same for all $i = 1, \cdots, M$. We therefore assume without loss of generality that the message $i = 1$ is transmitted. For a fixed codebook $(x_1, \cdots, x_M)$, we are interested in the probability of error when x_1 is transmitted. The error event depends on the entire codebook. Let $E \in \mathcal{B}$. It can be verified that the mapping defined by $((x_1, \cdots, x_M), E) \to P_{Y|X}(E|x_1)$ is a transition probability function from $(A_\Gamma^M, \mathcal{A}_\Gamma^M)$ to $(B, \mathcal{B})$. It represents the probability of an event E given that the codebook is $(x_1, \cdots, x_M)$ and x_1 is transmitted.

For any set $F \subset A_\Gamma^M \times B$ that is measurable relative to the product σ-algebra on $A_\Gamma^M \times B$, let

$$\Pr\{F\} \triangleq \int_{A_\Gamma^M} dP_{\boldsymbol{C}}(x_1, \cdots, x_M)$$
$$\cdot \int_B dP_{Y|X}(y|x_1)1\{(x_1, \cdots, x_M, y) \in F\} \tag{23}$$

i.e., probability of the event F given that message $i = 1$ is transmitted, averaged over the random selection of codebooks.

If the auxiliary decoder makes an error, then one of the following events should have occurred:

   i) $f(x_1, y) \le \beta$;
   ii) $f(x_j, y) > \beta$, for some $j \ne 1$; or
   iii) $y \notin H$.

It is therefore sufficient to upper-bound the probabilities of these events.

Using (19), (21)–(23), we can upper-bound the probability of event i) by

$$\Pr\{f(X_1, Y) \le \beta\} \le P_{X,Y}\{f(X, Y) \le \beta\}/(1 - \delta).$$

To upper-bound the probability of event ii), observe that

$$\Pr\{f(X_j, Y) > \beta, \text{ for some } j \ne 1\}$$
$$\overset{\text{a)}}{\le} M \cdot \Pr\{f(X_2, Y) > \beta\}$$
$$\overset{\text{b)}}{=} M \int_{A_\Gamma \times A_\Gamma} dP_X'(x_1)dP_X'(x_2)$$
$$\cdot \int_B dP_{Y|X}(y|x_1)1\{f(x_2, y) > \beta\}$$
$$\overset{\text{c)}}{\le} \frac{M}{(1-\delta)^2} \int_{A \times A} dP_X(x_1)dP_X(x_2)$$
$$\cdot \int_B dP_{Y|X}(y|x_1)1\{f(x_2, y) > \beta\}$$
$$\overset{\text{d)}}{=} \frac{M}{(1-\delta)^2} \int_A dP_X(x_2) \int_B dP_Y(y)1\{f(x_2, y) > \beta\}$$
$$\overset{\text{e)}}{\le} \frac{M}{(1-\delta)^2} \int_A dP_X(x_2) \int_B dP_Y(y)\frac{f(x_2, y)}{\beta}$$
$$\overset{\text{f)}}{=} \frac{M}{\beta(1-\delta)^2} \int_B dP_Y(y)$$
$$= \frac{M}{\beta(1-\delta)^2}.$$

In the above chain of inequalities, a) follows from the union bound for probabilities, b) from (22) and (23), c) from (19) and (21), d) from Fubini's theorem and [22, P. 18.25 (d), p. 247], e) from the fact $1\{f(x_2, y) > \beta\} \le f(x_2, y)/\beta$, and f) from Fubini's theorem and (20).

Under maximum-likelihood decoding,

$$f = dP_{X,Y}/d(P_X \times P_Y);$$

step f) would then be unnecessary, and the last equality would follow immediately after e). Under mismatched decoding, (20) is sufficient to obtain the last equality.

The probability of event iii) is upper-bounded by

$$\Pr\{Y \notin H\} \le P_{X,Y}\{Y \notin H\}/(1 - \delta)$$
$$= P_Y\{Y \notin H\}/(1 - \delta).$$

This completes the proof of the lemma. $\qquad\square$

Fix $n \in \mathcal{N}$. We apply Lemma 1 to the timing channel with $A = \mathcal{R}_+^n$ and $B = \mathcal{R}_+^{n+1}$. Let β^n play the role of β. Let $0 < \lambda'' < \lambda' < \mu$. Fix $\delta \in (0, 1)$. Let P_{X^n} be as defined in (6). $P_{Y^n|X^n}$ is the transition probability function from the input space to the output space that is induced by a stationary and ergodic process of service times, and (2) and (3). Let P_{X^n, Y^n} denote the joint distribution under P_{X^n} and $P_{Y^n|X^n}$. Let P_{Y^n} denote the marginal of Y^n under P_{X^n, Y^n}.

Fix $M \in \mathcal{N}$. A codebook is an M-tuple of elements from $\mathcal{R}_+^n$. Let the function g that denotes the input constraint in Lemma 1 be

$$g(x^n) \triangleq \left(\sum_{k=1}^n x_k\right)/n$$

where $x^n = (x_1, \cdots, x_n)$. We require that $g(x^n) \leq 1/\lambda''$ for every x^n in the codebook. By the weak law of large numbers for i.i.d. random variables, we have that for every $\delta \in (0, 1)$

$$P_{X^n}\{g(X^n) \leq 1/\lambda''\} \geq 1 - \delta \qquad (24)$$

for all sufficiently large n.

The function f in (8) satisfies $E[f(X^n, y^n)] = 1$ for every $y^n \in \mathcal{R}_+^{n+1}$. We deal with two decoders. The decision set H for the decoder $\phi_{f, H}$ will be either

$$\left\{ y^n \in \mathcal{R}_+^{n+1} \colon \sum_{k=0}^{n} y_k \leq n/\lambda'' \right\}$$

or $\mathcal{R}_+^{n+1}$. When $H = \mathcal{R}_+^{n+1}$, the entire output set, we denote the corresponding decoder by ϕ_f after omitting the subscript H.

Corollary 1: Fix $\delta \in (0, 1)$ and $M \in \mathcal{N}$. Fix $n \in \mathcal{N}$ so that $P_{X^n}\{g(X^n) \leq 1/\lambda''\} \geq 1 - \delta$. There exists a random variable $\boldsymbol{C}$ that takes values in the set of codebooks with M codewords, such that for any realization $\boldsymbol{c}$ of this random variable, $g(x^n) \leq 1/\lambda''$ for every $x^n \in \boldsymbol{c}$. Furthermore, for every $\beta > 0$

a) $E \dfrac{1}{M} \displaystyle\sum_{i=1}^{M} P_{e, i}(\boldsymbol{C}, \phi_f, P_{Y^n|X^n})$

$$\leq \frac{P_{X^n, Y^n}\{f(X^n, Y^n) \leq \beta^n\}}{1 - \delta} + \frac{M}{\beta^n(1 - \delta)^2},$$

and with $H = \{y^n \in \mathcal{R}_+^{n+1} \colon \sum_{k=0}^{n} y_k \leq n/\lambda''\}$

b) $E \dfrac{1}{M} \displaystyle\sum_{i=1}^{M} P_{e, i}(\boldsymbol{C}, \phi_{f, H}, P_{Y^n|X^n})$

$$\leq \frac{P_{X^n, Y^n}\{f(X^n, Y^n) \leq \beta^n\}}{1 - \delta} + \frac{M}{\beta^n(1 - \delta)^2}$$

$$+ \frac{P_{Y^n}\left\{ \sum_{k=0}^{n} Y_k > n/\lambda'' \right\}}{1 - \delta}.$$

Proof: The corollary follows from Lemma 1 after averaging the probability of error over the M equiprobable codewords. $\square$

Since the function f depends on the quantity m (cf. (9)), we need the following lemma to evaluate

$$P_{X^n, Y^n}\{f(X^n, Y^n) \leq \beta^n\}.$$

Lemma 2: Let $0 < \lambda' < \mu$. For $y^n \in \mathcal{R}_+^{n+1}$, the function m satisfies $m(y^n) = 0$ if and only if $y_0 + y_1 = 0$.

Proof: Fix $y^n \in \mathcal{R}_+^{n+1}$ so that $y_0 + y_1 > 0$ and $\sum_{k=0}^{n} y_k < \infty$. Let

$$S(y^n) \triangleq \left\{ x^n \in \mathcal{R}_+^n \colon \sum_{k=1}^{n} x_k \leq y_0 + y_1 \right\}.$$

Fix $x^n \in S(y^n)$. Then $0 \leq w_1 = \max\{0, x_1 - y_0\} \leq y_1$ and $w_k = 0$, $k = 2, \cdots, n$. These two conditions and (5) imply that $p(x^n, y^n) \geq u(y^n)$, where

$$u(y^n) \triangleq (\mu - \lambda')\mu^n \exp\left\{ -\mu \sum_{k=0}^{n} y_k \right\}.$$

Observe that $u(y^n) > 0$. Moreover, for any $x^n \in S(y^n)$

$$\prod_{k=1}^{n} e_{\lambda'}(x_k) \geq v(y^n) > 0$$

where

$$v(y^n) \triangleq (\lambda')^n \exp\{-\lambda'(y_0 + y_1)\}.$$

Furthermore,

$$\int_{S(y^n)} d\pi(x^n) = (y_0 + y_1)^n / n! > 0.$$

After substitution of these quantities in (9), we get

$$m(y^n) \geq \int_{S(y^n)} d\pi(x^n) \prod_{k=1}^{n} e_{\lambda'}(x_k) p(x^n, y^n)$$

$$\geq v(y^n) u(y^n) \int_{S(y^n)} d\pi(x^n)$$

$$> 0.$$

Conversely, if $y_0 + y_1 = 0$, then either $x_1 = 0$ or $p(x^n, y^n) = 0$, and thus

$$m(y^n) = \int_{\mathcal{R}_+^n} dP_{X^n}(x^n) p(x^n, y^n) 1\{x_1 = 0\} = 0.$$

This completes the proof of the lemma. $\square$

We now show that under a mild condition on the process of service times the quantity $P_{X^n, Y^n}\{f(X^n, Y^n) \leq \beta^n\}$ goes to zero as $n \to \infty$ if β is chosen judiciously.

Lemma 3: Let the process $(S_k \colon k \in \mathcal{N})$ of service times (not necessarily stationary or ergodic) satisfy for every $\alpha > 0$, the condition

$$\lim_{n \to \infty} P_{S^n}\left\{ \frac{1}{n} \sum_{k=1}^{n} S_k > \frac{1}{\mu} + \alpha \right\} = 0. \qquad (25)$$

Then for every $\gamma > 0$

$$\lim_{n \to \infty} P_{X^n, Y^n}\left\{ \frac{1}{n} \log f(X^n, Y^n) \leq \log \frac{\mu}{\lambda'} - \gamma \right\} = 0. \qquad (26)$$

Proof: Observe that $Y_0 = 0$ under P_{X^n, Y^n} for every $n \in \mathcal{N}$. Let

$$T \triangleq \left\{ (x^n, y^n) \in \mathcal{R}_+^n \times \mathcal{R}_+^{n+1} \colon y_0 = 0 \text{ and} \right.$$

$$\left. \sum_{k=1}^{j} (y_k - x_k) \geq 0, \text{ for every } j = 1, \cdots, n \right\}$$

i.e., the set of all pairs (x^n, y^n) such that y^n is a possible sequence of interdeparture times (with $y_0 = 0$) when the se-

quence of interarrival times is x^n. From (2) and (3), we have $P_{X^n, Y^n}\{T\} = 1$. We therefore have from (8) that

$$
P_{X^n, Y^n}\left\{\frac{1}{n}\log f(X^n, Y^n) \le \log\frac{\mu}{\lambda'} - \gamma\right\}
$$
$$
\le P_{X^n, Y^n}\{m(Y^n) = 0; \ (X^n, Y^n) \in T\}
$$
$$
+ P_{X^n, Y^n}\left\{\frac{1}{n}\log\frac{p(X^n, Y^n)}{m(Y^n)} \le \log\frac{\mu}{\lambda'} - \gamma;\right.
$$
$$
\left. m(Y^n) > 0; (X^n, Y^n) \in T\right\}. \quad (27)
$$

From Lemma 2, (2), and (3), we have that

$$
\{m(y^n) = 0; \ (x^n, y^n) \in T\}
$$
$$
= \{y_0 + y_1 = 0; (x^n, y^n) \in T\} \subset \{x_1 = 0\}
$$

and, therefore,

$$
P_{X^n, Y^n}\{m(Y^n) = 0; (X^n, Y^n) \in T\} \le P_{X^n}\{X_1 = 0\} = 0.
$$

We now upper-bound the term in (27). Let $(x^n, y^n) \in T$ and $m(y^n) > 0$. From (10) and (12) we can write

$$
m(y^n) = \prod_{k=1}^n e_{\lambda'}(y_k) \cdot g_{Y_0|Y_1, \cdots, Y_n}(y_0|y_1, \cdots, y_n)
$$

where the function g is the conditional density of Y_0 given $(Y_1, \cdots, Y_n)$ under Q_{Y^n}. Using (5), we get

$$
\frac{1}{n}\log\frac{p(x^n, y^n)}{m(y^n)} = \log\frac{\mu}{\lambda'} + \frac{\lambda'}{n}\sum_{k=1}^n y_k - \frac{\mu}{n}\sum_{k=1}^n(y_k - w_k)
$$
$$
- \frac{1}{n}i_{Y_0; Y_1, \cdots, Y_n}(y_0; y_1, \cdots, y_n) \quad (28)
$$

where $(1/n)i_{Y_0; Y_1, \cdots, Y_n}$ [1, Lemma 3] is the normalized information density of the relevant quantities, under Q_{Y^n}. Observe that

$$
\sum_{k=1}^n y_k \ge \sum_{k=1}^n x_k
$$

and that

$$
\sum_{k=1}^n (y_k - w_k) = \sum_{k=1}^n s_k
$$

the sum of service times [3] (cf. (3)). Using these facts and (28), we get that

$$
\left\{\frac{1}{n}\log\frac{p(x^n, y^n)}{m(y^n)} \le \log\frac{\mu}{\lambda'} - \gamma; m(y^n) > 0; (x^n, y^n) \in T\right\}
$$
$$
\subset \left\{\frac{\lambda'}{n}\sum_{k=1}^n x_k - \frac{\mu}{n}\sum_{k=1}^n s_k\right.
$$
$$
\left. - \frac{1}{n}i_{Y_0; Y_1, \cdots, Y_n}(y_0; y_1, \cdots, y_n) \le -\gamma; (x^n, y^n) \in T\right\}
$$
$$
\subset \left\{\frac{1}{n}i_{Y_0; Y_1, \cdots, Y_n}(y_0; y_1, \cdots, y_n) \ge \frac{\gamma}{3}\right\}
$$

$$
\cup\left\{\frac{1}{n}\sum_{k=1}^n x_k < \frac{1}{\lambda'} - \frac{\gamma}{3\lambda'}\right\}
$$
$$
\cup\left\{\frac{1}{n}\sum_{k=1}^n s_k > \frac{1}{\mu} + \frac{\gamma}{3\mu}; \ (x^n, y^n) \in T\right\}.
$$

Using the above inclusions, the term in (27) is upper-bounded by

$$
P_{Y^n}\left\{\frac{1}{n}i_{Y_0; Y_1, \cdots, Y_n} \ge \frac{\gamma}{3}\right\} + P_{X^n}\left\{\frac{1}{n}\sum_{k=1}^n X_k < \frac{1}{\lambda'} - \frac{\gamma}{3\lambda'}\right\}
$$
$$
+ P_{S^n}\left\{\frac{1}{n}\sum_{k=1}^n S_k > \frac{1}{\mu} + \frac{\gamma}{3\mu}\right\}
$$
$$
(29)
$$

where the last term follows after a change of variables. As $P_{X^n, Y^n}\{Y_0 = 0\} = 1$ for every $n \in \mathcal{N}$, the first term in (29) goes to 0 by [1, Lemma 3]. In fact, from the proof of [1, Lemma 3], this term equals 0 for all sufficiently large n. The second term in (29) goes to 0 by the weak law of large numbers for i.i.d. random variables. The third term in (29) goes to 0 by the assumption of the lemma. □

We are now ready to prove Theorems 2 and 3.

Proof of Theorem 2: Let the assumptions preceding Corollary 1 hold. Fix arbitrary $\gamma > 0$. Let $\log\beta = \log(\mu/\lambda') - \gamma$. We now apply Corollary 1 a). Since the process $(S_k: k \in \mathcal{N})$ is stationary and ergodic, it satisfies (25) by the ergodic theorem (see for example, [21, Theorem 7.3.3, pp. 236–237]), and, therefore, (26) holds by Lemma 3. Choosing M_n so that

$$
\log(\mu/\lambda') - 2\gamma < (\log M_n)/n < \log(\mu/\lambda') - 3\gamma/2
$$

ensures that $\lim_{n\to\infty} M_n/(\beta^n(1-\delta)^2) = 0$. We can, therefore, find a sequence $(\varepsilon_n : n \in \mathcal{N})$, such that

$$
E\frac{1}{M_n}\sum_{i=1}^{M_n} P_{e, i}(\boldsymbol{C}, \phi_f, P_{Y^n|X^n}) \le \varepsilon_n
$$

for every $n \in \mathcal{N}$, and $\lim_{n\to\infty}\varepsilon_n = 0$.

Consequently, for each $n \in \mathcal{N}$, there is a codebook realization $\boldsymbol{c}$ that satisfies

$$
\frac{1}{M_n}\sum_{i=1}^{M_n} P_{e, i}(\boldsymbol{c}, \phi_f, P_{Y^n|X^n}) \le \varepsilon_n.
$$

Furthermore, every codeword $x^n \in \boldsymbol{c}$ satisfies $\sum_{k=1}^n x_k \le n/\lambda''$. The decoder ϕ_f observes only those departures that occur in the time window $[0, n/\lambda'']$. We therefore have a sequence of $(n, M_n, n/\lambda'', \varepsilon_n)$-window-codes that satisfies

$$
\frac{\log M_n}{n/\lambda''} \ge \lambda''\log\frac{\mu}{\lambda'} - 2\gamma\lambda''
$$

for every $n \in \mathcal{N}$, and $\lim_{n\to\infty}\varepsilon_n = 0$. Setting $\lambda' = e^{-1}\mu$ and $\lambda'' = e^{-1}\mu/(1+\gamma)$, we get

$$
\lambda''(\log M_n)/n \ge e^{-1}\mu - 3\gamma e^{-1}\mu/(1+\gamma)
$$

for every $n \in \mathcal{N}$. This means that the rate $e^{-1}\mu$ nats per second is achievable with window-codes (because γ was arbitrary). $\square$

Proof of Theorem 3: Fix arbitrary $\gamma > 0$. The process $(S_k : k \in \mathcal{N})$ of nominal service times is stationary and ergodic. Consider $n \in \mathcal{N}$ and $\boldsymbol{z} \in \mathcal{R}_+^n$ satisfying $l(\boldsymbol{z}) \leq 1/\mu_2$. Let $W^n(\boldsymbol{z})$ be the corresponding transition probability function. Set $1/\mu = 1/\mu_1 + 1/\mu_2$. Since $l(\boldsymbol{z}) \leq 1/\mu_2$, for every $\alpha > 0$, we can write

$$\left\{ \frac{1}{n} \sum_{k=1}^{n} (s_k + z_k) > \frac{1}{\mu} + \alpha \right\} \subset \left\{ \frac{1}{n} \sum_{k=1}^{n} s_k > \frac{1}{\mu_1} + \alpha \right\}. \tag{30}$$

With $W^n(\boldsymbol{z})$ in place of $P_{Y^n|X^n}$ in the proof of Lemma 3, and by using (30), we obtain

$$P_{X^n, Y^n} \left\{ \frac{1}{n} \log f(X^n, Y^n) \leq \log \frac{\mu}{\lambda'} - \gamma \right\}$$
$$\leq P_{Y^n} \left\{ \frac{1}{n} i_{Y_0; Y_1, \cdots, Y_n} \geq \frac{\gamma}{3} \right\}$$
$$+ P_{X^n} \left\{ \frac{1}{n} \sum_{k=1}^{n} X_k < \frac{1}{\lambda'} - \frac{\gamma}{3\lambda'} \right\}$$
$$+ P_{S^n} \left\{ \frac{1}{n} \sum_{k=1}^{n} S_k > \frac{1}{\mu_1} + \frac{\gamma}{3\mu} \right\}. \tag{31}$$

Observe that $P_{X^n, Y^n}\{\cdot\}$ and $P_{Y^n}\{\cdot\}$ depend in general on the state sequence $\boldsymbol{z}$. For sufficiently large n, however, the term $P_{Y^n}\{(1/n)i_{Y_0; Y_1, \cdots, Y_n} \geq \gamma/3\}$ does not depend on $\boldsymbol{z}$ because of the following. Indeed, for every $\boldsymbol{z}$, $P_{Y^n}\{Y_0 > 0\} = 0$. In this case, we can extend the proof of [1, Lemma 1] to get

$$P_{Y^n}\{(1/n)i_{Y_0; Y_1, \cdots, Y_n} \geq \gamma/3\} = 0$$

for every $n > n_0(\gamma, \lambda', \mu)$. The quantity $n_0(\gamma, \lambda', \mu)$, which can be taken as $-(3/\gamma) \log(1 - \lambda'/\mu)$, is independent of $\boldsymbol{z}$. Furthermore, for a fixed n, the two remaining terms in (31) do not depend on $\boldsymbol{z}$ if $l(\boldsymbol{z}) \leq 1/\mu_2$; they go to 0 as $n \to \infty$.

Choose λ', λ'', β and the sequence $(M_n : n \in \mathcal{N})$ as in the proof of Theorem 2. From Corollary 1 a), we can find a sequence $(\varepsilon_n : n \in \mathcal{N})$, such that for all sufficiently large n

$$E \frac{1}{M_n} \sum_{i=1}^{M_n} P_{e,i}(\boldsymbol{C}, \phi_f, W^n(\boldsymbol{z})) \leq \varepsilon_n$$

for every $\boldsymbol{z}$ with $l(\boldsymbol{z}) \leq 1/\mu_2$, and $\lim_{n\to\infty} \varepsilon_n = 0$. We have therefore obtained a sequence of $(n, M_n, n/\lambda'', \varepsilon_n)$-random window-codes. The sequence also satisfies

$$\lambda'' \frac{\log M_n}{n} \geq e^{-1}\mu - 3e^{-1}\mu \frac{\gamma}{(1+\gamma)}$$

for every $n \in \mathcal{N}$. $\square$

Proving Theorem 1 requires more work. We need to find an upper bound on the expected time of departure of the nth packet. We first prove the following lemma.

Lemma 4: Let the single-server queue be work-conserving and initially empty. Let the server follow a first-in-first-out service discipline with stationary and ergodic service times of mean $1/\mu$ seconds. Let the queue be driven by a Poisson process of rate $\lambda' < \mu$. Then

$$(1/n) \sum_{k=1}^{n} Y_k \to 1/\lambda'$$

in probability.

Proof: Let $(X_n : n \in \mathcal{N})$ be the process of independent and exponentially distributed interarrival times, and $(S_n : n \in \mathcal{N})$ the stationary and ergodic process of service times. Let $P\{\cdot\}$ denote the probability of an event with respect to the joint process. Let R_n be the waiting time of the nth packet. Observe that

$$\sum_{k=1}^{n} Y_k = \sum_{k=1}^{n} X_k + R_n + S_n. \tag{32}$$

Since $R_n, S_n \geq 0$, we have that for any $\gamma > 0$

$$P\left\{ \frac{1}{n} \sum_{k=1}^{n} Y_k < \frac{1}{\lambda'} - \gamma \right\} \leq P\left\{ \frac{1}{n} \sum_{k=1}^{n} X_k < \frac{1}{\lambda'} - \gamma \right\} \to 0$$

as $n \to \infty$. On the other hand, using (32) and the union bound for probabilities, we get

$$P\left\{ \frac{1}{n} \sum_{k=1}^{n} Y_k > \frac{1}{\lambda'} + \gamma \right\} \leq P\left\{ \frac{R_n}{n} > \frac{\gamma}{3} \right\} + P\left\{ \frac{S_n}{n} > \frac{\gamma}{3} \right\}$$
$$+ P\left\{ \frac{1}{n} \sum_{k=1}^{n} X_k > \frac{1}{\lambda'} + \frac{\gamma}{3} \right\}. \tag{33}$$

The last two terms on the right side of (33) go to 0 as $n \to \infty$. We now upper-bound the first term in (33).

It can be shown that $((X_n, S_n) : n \in \mathcal{N})$ is a stationary and ergodic process. R_n converges in distribution to a finite random variable ϕ that satisfies $P\{\phi < \infty\} = 1$ because $\lambda' < \mu$ [21, Theorem 7.4.5, p. 241]. Observe that

$$\lim_{n\to\infty} P\{R_n \leq C\} = P\{\phi \leq C\}$$

if C is a point of continuity of the cumulative distribution function (cdf) of ϕ. A cdf can have only a countable number of discontinuities. Fix arbitrary $\varepsilon > 0$. Choose $C \in (0, \infty)$ large enough so that $P\{\phi > C\} \leq \varepsilon$ and C is a point of continuity of the cdf of ϕ. Then, for all sufficiently large n

$$P\{R_n > n\gamma/3\} \leq P\{R_n > C\} \leq P\{\phi > C\} + \varepsilon \leq 2\varepsilon.$$

This proves the lemma. $\square$

Proof of Theorem 1: Let the assumptions preceding Corollary 1 hold. Fix $\gamma > 0$. Let

$$\lambda' = e^{-1}\mu, \ \lambda'' = e^{-1}\mu/(1+\gamma), \ \lambda = e^{-1}\mu/(1+2\gamma)$$

and $\log \beta = \log(\mu/\lambda') - \gamma$. Let

$$H = \left\{ y^n \in \mathcal{R}_+^{n+1} : \sum_{k=0}^{n} y_k \leq n/\lambda'' \right\}.$$

Consider the decoder $\phi_{f,H}$. Clearly, this decoder cannot outperform ϕ_f. The process $(S_k \colon k \in \mathcal{N})$ is stationary and ergodic with mean $1/\mu$ seconds; it therefore satisfies (25), and thus (26) holds by Lemma 3. Choosing M_n so that

$$\log(\mu/\lambda') - 2\gamma < (\log M_n)/n < \log(\mu/\lambda') - 3\gamma/2$$

ensures that

$$\lim_{n \to \infty} M_n/(\beta^n (1-\delta)^2) = 0.$$

Since $P_{Y^n}\{Y_0 = 0\} = 1$ for every $n \in \mathcal{N}$ and $1/\lambda'' = 1/\lambda' + \gamma e/\mu$, we can apply Lemma 4 to get

$$\lim_{n \to \infty} P_{Y^n}\{Y^n \notin H\} = 0.$$

From Corollary 1 b), we can find a sequence $(\varepsilon_n : n \in \mathcal{N})$, such that

$$E \frac{1}{M_n} \sum_{i=1}^{M_n} P_{e,i}(\boldsymbol{C}, \phi_{f,H}, P_{Y^n|X^n}) \leq \frac{\varepsilon_n}{2}$$

for every $n \in \mathcal{N}$, and $\lim_{n \to \infty} \varepsilon_n = 0$.

Fix $n \in \mathcal{N}$. We can find a codebook $\boldsymbol{c}(n)$ such that

$$\frac{1}{M_n} \sum_{i=1}^{M_n} P_{e,i}(\boldsymbol{c}(n), \phi_{f,H}, P_{Y^n|X^n}) \leq \frac{\varepsilon_n}{2}.$$

Furthermore, every codeword $x^n \in \boldsymbol{c}(n)$ satisfies $\sum_{k=1}^{n} x_k \leq n/\lambda''$. By removing the $M_n - \lfloor M_n/2 \rfloor$ worst codewords, relabeling the remaining codewords from 1 through $\lfloor M_n/2 \rfloor$, and denoting the resulting codebook as $\boldsymbol{c}'(n)$, we get

$$P_{e,i}(\boldsymbol{c}'(n), \phi_{f,H}, P_{Y^n|X^n}) \leq \varepsilon_n$$

for $i = 1, \cdots, \lfloor M_n/2 \rfloor$. In particular, this implies that

$$P_{Y^n|X^n}\left\{ \frac{1}{n} \sum_{k=0}^{n} Y_k > \frac{1}{\lambda''} \Big| x^n \right\} \leq \varepsilon_n \qquad (34)$$

for every $x^n \in \boldsymbol{c}'(n)$. The next lemma shows that the expected time of the nth departure given each codeword is smaller than n/λ for all sufficiently large n.

Lemma 5: Fix $0 < \lambda < \lambda'' < \lambda' < \mu$. Let $(\varepsilon_n : n \in \mathcal{N})$ satisfy $\lim_{n \to \infty} \varepsilon_n = 0$. Let $(\boldsymbol{c}'(n) : n \in \mathcal{N})$ be a sequence of codebooks that satisfies for each $n \in \mathcal{N}$, the condition (34) and the condition $\sum_{k=1}^{n} x_k \leq n/\lambda''$ for every $x^n \in \boldsymbol{c}'(n)$. Then for all sufficiently large n

$$E\left[(1/n) \sum_{k=0}^{n} Y_k | x^n \right] \leq 1/\lambda$$

for every $x^n \in \boldsymbol{c}'(n)$.

Proof: Fix a $\nu > 0$ so that $1/\lambda'' + \nu < 1/\lambda$. Let

$$F_n \triangleq \left\{ y^n \in \mathcal{R}_+^{n+1} : \sum_{k=0}^{n} y_k > n/\lambda'' \right\}.$$

Let F_n^c denote the complement of the set F_n, and 1_{F_n} the indicator function of F_n. From (34), $P_{Y^n|X^n}\{F_n|x^n\} \leq \varepsilon_n$ for every $x^n \in \boldsymbol{c}'(n)$.

Let S_k be the service time of the kth packet. For a $C > 0$, let

$$G_k \triangleq \{s_k \in \mathcal{R}_+ : s_k > C\}.$$

For every $x^n \in \boldsymbol{c}'(n)$, every $k = 1, \cdots, n$, we have

$$E[S_k 1_{F_n}|x^n] = E[S_k 1_{F_n} 1_{G_k}|x^n] + E[S_k 1_{F_n} 1_{G_k^c}|x^n]$$
$$\leq E[S_k 1_{G_k}|x^n] + C P_{Y^n|X^n}\{F_n|x^n\}. \quad (35)$$

Observe that $E[S_k 1_{G_k}|x^n]$ is independent of x^n and of k because the process $(S_k : k \in \mathcal{N})$ is stationary and independent of the arrivals. Furthermore,

$$E[S] = E[S 1_G] + E[S 1_{G^c}] = 1/\mu < \infty.$$

Using the monotone convergence theorem, we can choose a C large enough so that $E[S 1_G] < \nu/3$. Pick n large enough so that $\max\{1/\lambda'', C\} \cdot \varepsilon_n < \nu/3$. Using (35), we therefore have that for all sufficiently large n

$$E[S_k 1_{F_n}|x^n] \leq \nu/3 + C \varepsilon_n \leq 2\nu/3.$$

Since $\sum_{k=1}^{n} x_k \leq n/\lambda''$ for every $x^n \in \boldsymbol{c}'(n)$, we get

$$E\left[\left(\frac{1}{n} \sum_{k=1}^{n} (x_k + S_k) \right) \cdot 1_{F_n}|x^n \right] \leq \varepsilon_n/\lambda'' + 2\nu/3 \leq \nu.$$

Since we can assume

$$\sum_{k=0}^{n} Y_k \leq \sum_{k=1}^{n} (x_k + S_k)$$

under $P_{Y^n|X^n}\{\cdot|x^n\}$, it follows that

$$E\left[\left((1/n) \sum_{k=0}^{n} Y_k \right) \cdot 1_{F_n}|x^n \right] \leq \nu.$$

Therefore, for every $x^n \in \boldsymbol{c}'(n)$

$$E\left[\frac{1}{n} \sum_{k=0}^{n} Y_k|x^n \right] = E\left[\left(\frac{1}{n} \sum_{k=0}^{n} Y_k \right) \cdot 1_{F_n}|x^n \right]$$
$$+ E\left[\left(\frac{1}{n} \sum_{k=0}^{n} Y_k \right) \cdot 1_{F_n^c}|x^n \right]$$
$$\leq \nu + 1/\lambda''$$
$$< 1/\lambda.$$

This completes the proof of the lemma. $\qquad \square$

Continuing with the proof of Theorem 1, for all sufficiently large n, the expected time of the nth departure is not greater than n/λ (cf. Lemma 5). We therefore have a sequence of $(n, \lfloor M_n/2 \rfloor, n/\lambda, \varepsilon_n)$-codes that satisfies

$$\lambda (\log \lfloor M_n/2 \rfloor)/n > e^{-1}\mu - 5\gamma e^{-1}\mu/(1+2\gamma)$$

for all sufficiently large n, and $\lim_{n \to \infty} \varepsilon_n = 0$. This proves that the rate $e^{-1}\mu$ nats/s is achievable. $\qquad \square$

Proof of Theorem 4: Fix $0 < \lambda < \lambda' < \mu$. Consider an $n \in \mathcal{N}$. We apply Lemma 1 with $A = B = \mathcal{R}_+^n$. Observe that $Y_k = X_k + S_k$, $k = 1, \cdots, n$. Let $P_{Y^n|X^n}$ denote the transition probability function from the input space to the output space. Choose M_n as in the proof of Theorem 2. Let

$g(x^n) \triangleq (1/n) \sum_{k=1}^{n} x_k$. We require that $g(x^n) \leq 1/\lambda - 1/\mu$ for every codeword x^n. Since the mean service time is $1/\mu$ seconds, it follows that the expected time of the nth departure is not greater than n/λ.

Let P_X be the mixture of a point mass and an exponential distribution given by

$$P_X\{X = 0\} = \lambda'/\mu$$
$$P_X\{X > x\} = \left(1 - \frac{\lambda'}{\mu}\right) e^{-\lambda' x}, \qquad x \geq 0.$$

Note that P_X is the input distribution that attains the mutual information saddle point [1, Theorem 3], [16, Theorem 1]. Let P_{X^n} be the distribution under which $X^n = (X_1, \cdots, X_n)$ is a vector of i.i.d. random variables with distribution P_X. Observe that if the service times are independent and exponentially distributed with mean $1/\mu$ secomds, then the outputs are independent and exponentially distributed with mean $1/\lambda'$ seconds. Let

$$f(x^n, y^n) \triangleq \prod_{k=1}^{n} \frac{e_\mu(y_k - x_k)}{e_{\lambda'}(y_k)}. \tag{36}$$

This function satisfies $Ef(X^n, y^n) = 1$ for every $y^n \in \mathcal{R}_+^n$. Let $H = \mathcal{R}_+^n$. Observe that the decoder ϕ_f (cf. (13)) with f as in (36) is the same as the decoder φ_d (cf. (14)).

Let P_{X^n, Y^n} be the joint distribution under P_{X^n} and $P_{Y^n|X^n}$. We only need to consider $(x^n, y^n) \in \mathcal{R}_+^n \times \mathcal{R}_+^n$ that satisfy $0 \leq x_k \leq y_k$, $k = 1, \cdots, n$. For such an (x^n, y^n),

$$\frac{1}{n} \log f(x^n, y^n) = \log \frac{\mu}{\lambda'} + \frac{\mu}{n} \sum_{k=1}^{n} x_k - \frac{(\mu - \lambda')}{n} \sum_{k=1}^{n} y_k.$$

From this and the stationarity and ergodicity of $((X_n, Y_n) : n \geq 1)$, we get

$$\lim_{n \to \infty} P_{X^n, Y^n} \left\{ \frac{1}{n} \log f(X^n, Y^n) \leq \log \frac{\mu}{\lambda'} - \gamma \right\} = 0.$$

The rest of the proof is similar to that of Theorem 2.   □

Proof of Theorem 6 a): Fix $n \in \mathcal{N}$. There are M_n messages. Each message corresponds to a sequence of interarrival times; the nth arrival occurs before time T_n. This sequence maps to a (right-continuous with left limits) point process of arrivals $(A_t : t \in [0, T_n])$. A_t is the number of arrivals in $[0, t]$, $t \in [0, T_n]$. Analogously, the observed departures form a (right-continuous with left limits) point process $(D_t : t \in [0, T_n])$, where D_t is the number of departures in $[0, t]$, $t \in [0, T_n]$. Let

$$\mathcal{F}_t^A \triangleq \sigma(A_s : s \in [0, t])$$
$$\mathcal{F}_t^D \triangleq \sigma(D_s : s \in [0, t])$$
$$\mathcal{F}_t^{A, D} \triangleq \sigma\{\mathcal{F}_t^A, \mathcal{F}_t^D\}.$$

For $t = 0$, let $\nu(0, A, D) \triangleq 0$. For $t > 0$, fix an increasing sequence of rational numbers $(r_n : n \in \mathcal{N})$ such that $r_n \uparrow t$, and let

$$\nu(t, A, D) \triangleq \lim_{r_n \uparrow t} (A_{r_n} - D_{r_n}).$$

The quantity $\nu(t, A, D)$ represents the number of packets that remain in the system at time $t-$, i.e., just prior to t. Clearly $\nu(t, A, D)$ is $\mathcal{F}_t^{A, D}$-measurable for every $t \in [0, T_n]$ and $(\nu(t, A, D) : t \in [0, T_n])$ is a left-continuous process. Let

$$\lambda(t, A, D) \triangleq \mu 1\{\nu(t, A, D) > 0\}, \ t \in [0, T_n].$$

Observe that $\lambda(t, A, D)$ is $\mathcal{F}_t^{A, D}$-measurable for every $t \in [0, T_n]$ and that $(\lambda(t, A, D) : t \in [0, T_n])$ is a left-continuous process; it is therefore a *predictable* process [24, Definition 3, p. 173].

Fix arbitrary $t \in [0, T_n]$. If $\lambda(t, A, D) = 0$, there is no packet in the system at time $t-$, and therefore no packet can depart at time t; the intensity of the point process of departures is 0 at time t. If $\lambda(t, A, D) = \mu$, there is at least one packet in the system at time $t-$. Due to the memoryless property of exponential service times, the residual time for the next departure is exponentially distributed with mean $1/\mu$ seconds, independent of the past. In other words, when $\lambda(t, A, D) = \mu$, the intensity of the point process of departures also takes the value μ at time t. The process of departures is therefore a point process with intensity $(\lambda(t, A, D) : t \in [0, T_n])$.

Any window-code on the timing channel is therefore equivalent to the above strategy with complete feedback on the point-process channel with maximum intensity μ and no background intensity. Complete information about the past departures is necessary to determine the intensity of the departures at time t. The capacity of the timing channel (for window-codes) is therefore upper-bounded by the capacity of the point-process channel with complete feedback. From the corollary to [25, Theorem 19.10, pp. 318–320] and the proofs of converse in [2] and [26], this upper bound is $e^{-1}\mu$ nats per second.   □

Proof of Theorem 6 b): The process $(S_k : k \in \mathcal{N})$ of nominal service times is a sequence of independent and exponentially distributed random variables with mean $1/\mu_1$ seconds. Let $\mu_2 \in (0, \infty)$. Let $\mu \triangleq \mu_1 \mu_2/(\mu_1 + \mu_2)$. Suppose that there were a sequence of $(n, M_n, T_n, \varepsilon_n)$-random window-codes that satisfies for some $\alpha > 0$, $(\log M_n)/T_n > e^{-1}\mu + \alpha$ for all sufficiently large n, and $\lim_{n \to \infty} \varepsilon_n = 0$. Then, for some ϕ and $\boldsymbol{C}$

$$\sup_{\boldsymbol{z}:\, l(\boldsymbol{z}) \leq 1/\mu_2} E[P_e(\boldsymbol{C}, \phi, W^n(\boldsymbol{z}))] \leq \varepsilon_n. \tag{37}$$

Choose μ_2' so that $1/\mu_2' < 1/\mu_2$ and $e^{-1}\mu > e^{-1}\mu' - \alpha/2$, where

$$\mu' \triangleq \mu_1 \mu_2'/(\mu_1 + \mu_2').$$

Let P_Z be the distribution given by

$$P_X\{Z = 0\} = \mu'/\mu_1$$
$$P_X\{Z > z\} = \left(1 - \frac{\mu'}{\mu_1}\right) e^{-\mu' z}, \qquad z \geq 0.$$

Note that Z has mean $1/\mu_2' = 1/\mu' - 1/\mu_1$ seconds. Furthermore, if S is independent of Z and exponentially distributed with mean $1/\mu_1$ seconds, then $S + Z$ is exponentially distributed with mean $1/\mu'$ seconds.

Let $\boldsymbol{Z} = (Z_1, \cdots, Z_n)$ be a vector of i.i.d. random variables with common distribution P_Z, independent of the codebook distribution and the nominal service times. We then have

$$E[P_e(\boldsymbol{C}, \phi, W^n(\boldsymbol{Z}))] \leq \sup_{\boldsymbol{z}:\, l(\boldsymbol{z}) \leq 1/\mu_2} E[P_e(\boldsymbol{C}, \phi, W^n(\boldsymbol{z}))]$$
$$+ P_{\boldsymbol{Z}}\{l(\boldsymbol{Z}) > 1/\mu_2\}$$
$$\overset{a)}{\leq} \varepsilon_n + P_{\boldsymbol{Z}}\{l(\boldsymbol{Z}) > 1/\mu_2\}$$

where a) follows from (37). Let

$$\delta_n \overset{\triangle}{=} \varepsilon_n + P_{\boldsymbol{Z}}\{l(\boldsymbol{Z}) > 1/\mu_2\}.$$

From the weak law of large numbers, we get $\lim_{n\to\infty} \delta_n = 0$ because $1/\mu_2' < 1/\mu_2$. Observe that $E[P_e(\boldsymbol{C}, \phi, W^n(\boldsymbol{Z}))]$ is also the expected probability of error (expectation over the codebook distribution) for the exponential server channel with mean service time $1/\mu'$ seconds. We can therefore find for this channel a sequence of (n, M_n, T_n, δ_n)-window-codes with

$$(\log M_n)/T_n \geq e^{-1}\mu' + \alpha/2$$

for all sufficiently large n, and $\lim_{n\to\infty} \delta_n = 0$. Since $e^{-1}\mu'$ nats per seconds is the largest rate achievable with window-codes, we reach a contradiction. $\qquad\square$

REFERENCES

[1] V. Anantharam and S. Verdú, "Bits through queues," *IEEE Trans. Inform. Theory*, vol. 42, pp. 4–18, Jan. 1996.

[2] Y. M. Kabanov, "The capacity of a channel of the Poisson type," *Theory Prob. Appl.*, vol. 23, pp. 143–147, 1978.

[3] A. Bedekar and M. Azizoğlu, "The information-theoretic capacity of discrete-time queues," *IEEE Trans. Inform. Theory*, vol. 44, pp. 446–461, Mar. 1998.

[4] J. A. Thomas, "On the Shannon capacity of discrete-time queues," in *Proc. 1997 IEEE Int. Symp. Information Theory*, Ulm, Germany, July 1997, p. 333.

[5] D. Blackwell *et al.*, "The capacities of certain channel classes under random coding," *Ann. Math. Statist.*, vol. 31, pp. 558–567, 1960.

[6] R. Ahlswede, "Elimination of correlation in random codes for arbitrarily varying channels," Z. *Wahrscheinlichkeitsth. Verw. Gebiete*, vol. 44, pp. 159–175, 1978.

[7] I. Csiszár and P. Narayan, "The capacity of arbitrarily varying channels revisited: Positivity, constraints," *IEEE Trans. Inform. Theory*, vol. 34, pp. 181–193, Mar. 1988.

[8] ——, "Capacity of the Gaussian arbitrarily varying channel," *IEEE Trans. Inform. Theory*, vol. 36, pp. 18–26, Jan. 1991.

[9] T. R. M. Fischer, "Some remarks on the role of inaccuracy in Shannon's theory of information transmission," in *Trans. 8th Prague Conf. Information Theory*, Prague, Czechoslovakia, 1978, pp. 211–226.

[10] G. Kaplan and S. Shamai (Shitz), "Information rates and error exponents of compound channels with application to antipodal signaling in a fading environment," *AEU*, vol. 47, no. 4, pp. 228–239, 1993.

[11] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, 2nd ed. Budapest, Hungary: Academic, 1986.

[12] N. Merhav, G. Kaplan, A. Lapidoth, and S. Shamai (Shitz), "On information rates for mismatched decoders," *IEEE Trans. Inform. Theory*, vol. 40, pp. 1953–1967, Nov. 1994.

[13] I. Csiszár and P. Narayan, "Channel capacity for a given decoding metric," *IEEE Trans. Inform. Theory*, vol. 41, pp. 35–43, Jan. 1995.

[14] A. Lapidoth, "On mismatched decoding," *IEEE Trans. Inform. Theory*, vol. 42, pp. 1439–1452, Sept. 1996.

[15] ——, "Nearest neighbor decoding for additive non-Gaussian noise channels," *IEEE Trans. Inform. Theory*, vol. IT-42, pp. 1520–1529, Sept. 1996.

[16] S. Verdú, "The exponential distribution in information theory" (in Russian), *Probl. Pered. Inform.*, vol. 32, no. 1, pp. 100–111, Jan.–Mar. 1996.

[17] I. Stiglitz, "Coding for a class of unknown channels," *IEEE Trans. Inform. Theory*, vol. IT-12, pp. 189–195, Apr. 1966.

[18] A. Lapidoth and P. Narayan, "Reliable communication under channel uncertainty," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2148–2177, Oct. 1998.

[19] B. Hughes and P. Narayan, "Gaussian arbitrarily varying channels," *IEEE Trans. Inform. Theory*, vol. IT-33, pp. 267–284, Mar. 1987.

[20] K. L. Chung, *A Course in Probability Theory*, 2nd ed. New York, NY: Academic, 1974.

[21] J. Walrand, *An Introduction to Queueing Networks*. Englewood Cliffs, NJ: Prentice-Hall, 1988.

[22] P. Billingsley, *Probability and Measure*, 2nd ed. New York, NY: Wiley, 1986.

[23] R. M. Gray, *Entropy and Information Theory*. New York, NY: Springer-Verlag, 1990.

[24] R. S. Liptser and A. N. Shiryayev, *Statistics of Random Processes*. New York, NY: Springer, 1977, vol. 1.

[25] ——, *Statistics of Random Processes*. New York, NY: Springer, 1978, vol. 2.

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