ZERO-RATE RELIABILITY OF THE EXPONENTIAL-SERVER TIMING CHANNEL

by

Aaron B. Wagner and Venkat Anantharam

Memorandum No. UCB/ERL M04/5

29 November 2003

ZERO-RATE RELIABILITY OF THE EXPONENTIAL-SERVER TIMING CHANNEL

by

Aaron B. Wagner and Venkat Anantharam

Memorandum No. UCB/ERL M04/5

29 November 2003

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

Zero-Rate Reliability of the Exponential-Server Timing Channel *

Aaron B. Wagner and Venkat Anantharam
Dept. of Electrical Engineering and Computer Sciences
University of California
Berkeley, CA 94720 USA
{awagner,ananth}@eecs.berkeley.edu

November 29, 2003

Abstract

We determine the reliability function of the exponential-server timing channel in the limit as the data rate approaches zero. The limit shows that at low rates, the exponential-server timing channel is strictly more reliable than the Poisson channel without dark current, answering a question Arikan posed in the Information Theory TRANSACTIONS. The proof identifies a distance metric over inputs to timing channels that parallels Euclidean and Hamming distance for conventional channels. We also prove the straight-line bound for the exponential-server timing channel, and a bound on the reliability of timing channels with general service distributions in the limit as the data rate approaches zero.

Index Terms—Timing channel, Poisson channel, error exponent, zero-rate reliability, point process.

1 Introduction

The exponential-server timing channel (ESTC), as introduced by Anantharam and Verdú [1], is one in which the sender chooses the arrival times of identical jobs to a single-server queue with independent, exponentially distributed service times, while the receiver observes the resulting departure times. Timing channels such as the ESTC have been studied in the context of covert communication [2]; by modulating the times at which it performs routine tasks, the sender can transmit information to the receiver in a way that makes it unclear to an observer that communication has occurred. We assume that the service discipline is first-in-first-out, and that the queue is initially empty. Neither of these assumptions is crucial; see Anantharam and

*This research was supported by DARPA Grants F30602-00-2-0538 and N66001-00-C-8062, by Grant N00014-1-0637 from the Office of Naval Research, and by Grant ECS-0123512 and a Graduate Research Fellowship from the National Science Foundation.

Verdú [1] for a discussion of these assumptions and for a proof that the capacity of the channel is μ/e nats per unit time, where $1/\mu$ is the mean service time.

Arikan [3] proves the random-coding and sphere-packing bounds on the reliability function [4] of the ESTC. These are shown in Figure 1, and coincide at rates between $(\mu/4)\log 2$ and capacity.[1] Although the motivation for proving these bounds was to determine the cutoff rate of the channel, which is $\mu/4$ [3], Arikan points out that they provide for an interesting comparison between the ESTC and the Poisson channel without dark current but with a peak power constraint of μ. The latter channel models a direct-detection optical channel; the input is a nonnegative waveform λ_t, which is upper bounded by μ, while the output is a Poisson process with intensity λ_t. The capacity of this channel is also μ/e nats per unit time [5, 6], and its reliability function is known [7, 8] and coincides with the random-coding exponent of the ESTC. Thus the ESTC is at least as reliable as the Poisson channel, and their reliability functions coincide at rates between $(\mu/4)\log 2$ and their common capacity. Arikan [3] poses the problem of determining whether the two reliability functions are identical. We show that they are not by proving that the *zero-rate reliability* of the ESTC, defined as the limit of the reliability function as the rate approaches zero, equals $\mu/2$, which exceeds the zero-rate reliability of the Poisson channel, which is $\mu/4$ [7, 8]. We also provide an improved upper bound on the reliability function of the ESTC at positive rates, and a bound on the zero-rate reliability of timing channels with general service distributions, that is, $\cdot/G/1$ queues. The proofs use point-process techniques that could be of independent interest.

It is known that feedback increases the zero-rate reliability of the Poisson channel from $\mu/4$ to μ [9]. Sundaresan and Verdú [10] show how the ESTC can be emulated on the Poisson channel with feedback by having the feedback encoder modulate the instantaneous intensity to imitate a $\cdot/M/1$ queue. Our results quantify, in the low-rate regime, the penalty associated with using this restricted form of feedback: general feedback provides for a factor of four increase in the reliability of the Poisson channel, whereas using the feedback to emulate a $\cdot/M/1$ queue provides for only a factor of two increase.

The covert communication scenario mentioned earlier provides another motivation, independent of the comparison with the Poisson channel, for finding improved bounds on the reliability function of the ESTC. When communicating covertly, the incentive for the sender to use a code with a short blocklength surpasses the usual motivation of minimizing the coding delay: a short blocklength decreases the likelihood of detection by the observer. Of course, a shorter blocklength generally comes at the expense of a higher error probability. Since the reliability function describes the tradeoff between the blocklength of a code and its error probability, it also captures the tradeoff between the probability of error and the probability of detection by the observer, assuming that the sender uses codes that are optimized for communication. More generally, it would be of interest to study this tradeoff when the codes are optimized with both goals in mind.

We return to the zero-rate reliability of the ESTC. It is instructive to begin a study of this problem by considering the zero-rate reliability of the binary symmetric channel. This channel has a natural distance metric over its inputs, Hamming distance, such that the maximum-likelihood (ML) error probability of a pair of codewords depends only on the Hamming distance between them. Furthermore, for low-rate codes, the error probability of a codeword is

[1]Throughout, log denotes the natural logarithm.

well-approximated by the probability that it is confused with one of its "nearest neighbors." The overall error probability of a low-rate code is then governed by its minimum distance. In turn, the zero-rate reliability of the channel is governed by the largest achievable minimum distance of low-rate codes. This *max-min distance* can be lower bounded via a random-coding argument and upper bounded using the Plotkin bound [4].

The situation is similar for the Gaussian channel, with Euclidean distance replacing Hamming distance [11]. Shannon, Gallager, and Berlekamp [12] extend the approach to a general discrete memoryless channel (DMC) by devising an appropriate distance metric and by replacing the Plotkin bound, which is insufficient in this case, with a more intricate converse.

We show that, despite the memory in the channel, this approach also works for the ESTC. We introduce a distance metric over inputs to timing channels, which parallels Euclidean and Hamming distance for conventional channels, and use it to bound the error probability of a pair of codewords, when used over the ESTC, in terms of the distance between them. Bounds are required here since the exact pairwise error probability under ML decoding is not easily evaluated for this channel. These results are contained in Section 3. In Section 4, we reduce the problem of determining the zero-rate reliability of the ESTC to a combinatorial sphere-packing problem involving our metric: determining the max-min distance. This quantity is lower bounded via a random-coding argument and upper bounded using a variation of the Shannon, Gallager, and Berlekamp converse mentioned above. Our variation uses a inner-product space lemma that can be used to elucidate their proof for discrete memoryless channels. Section 4 also contains a bound on the zero-rate reliability of $\cdot/G/1$ timing channels. The last section contains a proof of the straight-line bound for the ESTC. First we give a mathematical description of the channel.

2 Channel Construction

The blocklength of a block code for the ESTC can be defined in multiple ways, since the transmission and reception time of a codeword can be quite different, and the reception time depends on the channel realization. Two approaches have been used in the past: (*i*) defining the blocklength to be the expected time of the last departure from the queue, averaged over the codebook and the queueing process [1], and (*ii*) restricting the decoder to observe the departure process over a finite interval $[0, T]$, and defining the blocklength to be T [10]. We adopt the latter approach here. Notice that if the decoder is required to decode based on the observation of the departure process over $[0, T]$, then it is pointless for the encoder to submit jobs to the queue after time T. Thus we shall assume that the encoder submits all of its jobs to the queue during $[0, T]$. For $T > 0$, let Ω_T be the set of all counting functions over $[0, T]$, i.e. the set of all right-continuous, nondecreasing, integer-valued functions ω of $[0, T]$ such that $\omega(0) = 0$. When the blocklength is T, we view the input and output of the channel as elements of Ω_T by viewing ω_t as the number of arrivals or departures that have occurred during $(0, T]$. We use the following *window code* definition of Sundaresan and Verdú [10].

Definition 1 ([10]) *An* (n, M, T) encoder *is a mapping f from a set of integers, A, to* Ω_T*, such that* $|A| = M$ *and for all i,* $f_T(i) = n$.[2] *An* (n, M, T) code *is an* (n, M, T) *encoder and a*

[2]Throughout, we write $f_t(i)$ in place of the cumbersome $(f(i))(t)$.

decoder $\varphi : \Omega_T \mapsto A'$, where A' is a superset of A. The data rate of the code is $(1/T)\log M$.

We have followed precedent [1, 3, 10] and required that all codewords deliver the same number of jobs to the queue. Since the decoder only observes the departure process during $[0, T]$, this is no real restriction, because any encoder can be made compliant by adding arrivals at time T to some of the codewords, and this will not change the output law that they induce.

To construct the output law induced by sending an input x over the ESTC with service rate μ, we use Girsanov's Theorem applied to point processes [13, Section VI.2]. Let $\{\mathcal{F}_t\}_{t=0}^T$ be the filtration over Ω_T generated by the coordinate mappings $\omega \mapsto \omega_t$.[3] Throughout, we denote by P_0 the probability measure on $(\Omega_T, \mathcal{F}_T)$ under which the outcomes are distributed as a Poisson process with rate one.[4] The measure is not indexed by T since the measurable space with which it is associated will be clear from the context.

On $(\Omega_T, \mathcal{F}_T)$, define the random variable[5]

$$L = \exp\left[\int_0^T \log[\mu 1(x_t > \omega_{t-})]\, d\omega_t + \int_0^T [1 - \mu 1(x_t > \omega_t)]\, dt\right].$$

By convention, we take $L = 0$ on the event that $\omega_t - \omega_{t-} > 0$ for some t such that $x_t \le \omega_{t-}$. We call an output ω *feasible* for x if $\omega_t \le x_t$ for all $0 \le t \le T$. Evidently, L is positive on the event that ω is feasible for x. Now

$$\int_{\Omega_T} L\, dP_0 = 1$$

[13, Chapter VI, Theorem T4], and if we define P by $dP/dP_0 = L$, then under P, ω is feasible for x a.s. and ω_t has $\mathcal{F}$-intensity $\mu 1(x_t > \omega_{t-})$ [13, Chapter VI, Theorem T3]. So $\{\omega_t\}_{t=0}^T$ is distributed as the output of an initially-empty $\cdot/M/1$ queue with input x, as desired [13, Chapter III, Theorems T7 and T8]. In general, we denote by $P_{x,\mu}$ the probability measure on $(\Omega_T, \mathcal{F}_T)$ under which the outcomes are distributed as the output of a $\cdot/M/1$ queue with input x and service rate μ.

The error probability of a code (f, φ), when used over the ESTC with service rate μ, is defined as

$$P_e(f, \varphi) := \max_i [1 - P_{f(i),\mu}(\varphi^{-1}(i))].$$

3 Error Probability Bounds

Given an encoder f and an output ω, a ML decoder for the ESTC chooses a message i for which ω is feasible and that maximizes

$$\log \frac{dP_{f(i),\mu}}{dP_0} = \int_0^T \log[\mu 1(f_t(i) > \omega_{t-})]\, d\omega_t + \int_0^T [1 - \mu 1(f_t(i) > \omega_t)]\, dt.$$

[3]The σ-field $\mathcal{F}_T$ is the restriction of the σ-field generated by the Skorohod topology on $D[0, 1]$ to Ω_T [14, Theorem 14.5], but this fact is not needed here.

[4]When working with point-process martingales, one typically assumes that the underlying probability space and filtration satisfy the *usual conditions*: $(\Omega_T, \mathcal{F}_T, P_0)$ is complete, $\mathcal{F}$ is right-continuous, and $\mathcal{F}_0$ contains all of the P_0-null sets of $\mathcal{F}_T$. Brémaud [13, pp. 309–310] shows how to modify $\mathcal{F}$ and P_0 to comply with these requirements; in the sequel, we assume this has been done.

[5]Throughout, $\int_a^b$ should be interpreted as $\int_{(a,b]}$. Also, we write indicator functions as $1(\cdot)$ instead of $1_{\{\cdot\}}$.

Now for any message i for which ω is feasible,

$$\int_0^T \log[\mu 1(f_t(i) > \omega_{t-})]\, d\omega_t = \omega_T \log \mu.$$

Therefore, a ML decoder chooses a message in

$$\arg \max_{i:\omega \text{ feasible}} \frac{dP_{f(i),\mu}}{dP_0} = \arg \min_{i:\omega \text{ feasible}} \int_0^T 1(f_t(i) > \omega_t)\, dt.$$

This motivates the following definition.

Definition 2 *Given two elements of Ω_T, u and v, we say that u* leads v by

$$\mathcal{L}(u,v) := \int_0^T 1(u_t > v_t)\, dt.$$

Also let

$$\overline{\mathcal{L}}(u,v) := \max[\mathcal{L}(u,v), \mathcal{L}(v,u)].$$

Define the minimum distance *of an encoder f by*

$$\overline{\mathcal{L}}(f) := \inf_{i \neq j} \overline{\mathcal{L}}(f(i), f(j)).$$

The $\overline{\mathcal{L}}$ function is the distance mentioned in the introduction. It is not a metric on Ω_T because two elements of Ω_T, u and v, that differ only at time T will have $\overline{\mathcal{L}}(u,v) = 0$. It is, however, a pseudo-metric [15], and a metric if we identify elements of Ω_T that differ only at T. Two other properties of $\overline{\mathcal{L}}$ that will be used later are given in the first Lemma, whose straightforward proof is omitted.

Lemma 1

(a) If u and v are elements of Ω_T and $x = \min(u,v)$, then

$$\mathcal{L}(u,x) = \overline{\mathcal{L}}(u,x) = \mathcal{L}(u,v)$$

and

$$\mathcal{L}(v,x) = \overline{\mathcal{L}}(v,x) = \mathcal{L}(v,u).$$

(b) If $i \neq j$ and ω is feasible for $f(i)$ and $f(j)$, then

$$\max[\mathcal{L}(f(i),\omega), \mathcal{L}(f(j),\omega)] \geq \overline{\mathcal{L}}(f(i), f(j)) \geq \overline{\mathcal{L}}(f).$$

Evidently, ML decoding for the ESTC can be described as finding, among those codewords for which the output is feasible, one that leads the output by the smallest amount. Thus $\overline{\mathcal{L}}$ is similar to Euclidean distance over $\mathbb{R}^n$ and Hamming distance over $\{0,1\}^n$ in that each governs ML decoding for the channel with the lowest capacity in its class: ML decoding for the ESTC is governed by $\overline{\mathcal{L}}$ distance and the ESTC has the lowest capacity of all timing channels with fixed service rate [1, 10], ML decoding for the Gaussian channel is governed by Euclidean distance and the Gaussian channel has the lowest capacity of all additive noise channels with fixed SNR [4, Theorem 7.4.3], and ML decoding for the binary symmetric channel is governed by Hamming distance and the binary symmetric channel has the lowest capacity of all binary memoryless channels with the sum of the crossover probabilities of the two inputs held fixed. Unlike the Gaussian and binary symmetric channels, though, for the ESTC the ML error probability is difficult to evaluate even for just two messages, except in special cases. To proceed, we use a suboptimal decoder that approximates ML decoding. Part (b) of Lemma 1 proves that the decoding rule is well defined.

Definition 3 *Given an encoder f and an output ω, the* bounded-distance decoder φ_B *declares that message i was sent if ω is feasible for $f(i)$ and $\mathcal{L}(f(i),\omega) < \overline{\mathcal{L}}(f)$. If no message satisfies this property, then it declares an error (i.e., it chooses an integer outside the message set).*

This decoder is similar to a ML decoder, except that it quits and declares an error if the channel output is farther than $\overline{\mathcal{L}}(f)$ from all codewords for which it is feasible.

Lemma 2 *For any (n, M, T) encoder f, the error probability under bounded-distance decoding satisfies*

$$P_e(f,\varphi_B) \leq \Pr\left[\sum_{i=1}^{n} S_i \geq \overline{\mathcal{L}}(f)\right], \tag{1}$$

where $S_1,\ldots,S_n$ are the i.i.d. service times, each exponentially distributed with mean $1/\mu$. If $\overline{\mathcal{L}}(f)/T > \lambda/\mu$, where $\lambda = n/T$ is the arrival rate *of the code, this yields*

$$P_e(f,\varphi_B) \leq \exp\left[-T\left(\frac{\mu\overline{\mathcal{L}}(f)}{T} - \lambda - \lambda\log\left(\frac{\mu\overline{\mathcal{L}}(f)}{\lambda T}\right)\right)\right].$$

Proof. If message i is sent, then an error occurs only if $\mathcal{L}(f(i),\omega) \geq \overline{\mathcal{L}}(f)$, where ω is the output of the queue. By viewing $S_1,\ldots,S_n$ as the service times of the n jobs, we can couple them to ω such that

$$\mathcal{L}(f(i),\omega) \leq \sum_{i=1}^{n} S_i \quad \text{a.s.}$$

This proves the first conclusion. The second conclusion follows from the first by applying the Chernoff bound. □

Observe that the bounded-distance decoder corrects errors up to the minimum distance of the encoder, compared to half the minimum distance, which one might expect. Even so, the bounded-distance decoder is strictly suboptimal compared to a ML decoder. This result and the next one, however, together imply that there is little loss in using the bounded-distance decoder when the arrival rate is very low. We shall see in the next section that the zero-rate reliability of the ESTC is achieved by codes with vanishing arrival rates. Consequently, the bounded-distance decoder will be sufficient for our purposes.

Lemma 3 *Let f be an (n, M, T) encoder. Then for any decoder φ, the error probability satisfies*

$$P_e(f,\varphi) \geq \frac{1}{2}\exp\left[-\mu\overline{\mathcal{L}}(f)\right],$$

where $1/\mu$ is the mean service time.

Proof. There exists a pair of messages $i \neq j$ such that $\overline{\mathcal{L}}(f(i), f(j)) = \overline{\mathcal{L}}(f)$. Let p_i and p_j be their error probabilities, and define x by $x_t = \min[f_t(i), f_t(j)]$. Let $D_i \in \mathcal{F}_T$ be the decision region for i and let $D_j \in \mathcal{F}_T$ be the decision region for j. Also let

$$Y_{f(i)} = \{\omega \in \Omega_T : \omega \text{ is feasible for } f(i)\},$$

and define $Y_{f(j)}$ and Y_x similarly. Then

$$p_i = \int_{D_i^c \cap Y_{f(i)}} \exp\left[\int_0^T \log[\mu 1(f_t(i) > \omega_{t-})]\, d\omega_t + T - \mu\mathcal{L}(f(i), \omega)\right] dP_0.$$

But on $Y_{f(i)}$,

$$\int_0^T \log[\mu 1(f_t(i) > \omega_{t-})]\, d\omega_t = \omega_T \log\mu,$$

so

$$\begin{aligned}
p_i &= \int_{D_i^c \cap Y_{f(i)}} \exp\left[\omega_T \log\mu + T - \mu\mathcal{L}(f(i), \omega)\right]\, dP_0 \\
&\geq \int_{D_i^c \cap Y_x} \exp\left[\omega_T \log\mu + T - \mu(\mathcal{L}(f(i), x) + \mathcal{L}(x, \omega))\right]\, dP_0 \\
&\geq \exp[-\mu\overline{\mathcal{L}}(f(i), f(j))] \int_{D_i^c \cap Y_x} \exp\left[\omega_T \log\mu + T - \mu\mathcal{L}(x, \omega)\right]\, dP_0,
\end{aligned}$$

where we have used the triangle inequality for $\overline{\mathcal{L}}$ and part (a) of Lemma 1. Similarly,

$$p_j \geq \exp[-\mu\overline{\mathcal{L}}(f(i), f(j))] \int_{D_j^c \cap Y_x} \exp\left[\omega_T \log\mu + T - \mu\mathcal{L}(x, \omega)\right]\, dP_0.$$

But

$$\int_{Y_x} \exp\left[\omega_T \log\mu + T - \mu\mathcal{L}(x, \omega)\right]\, dP_0 = \int_{Y_x} \frac{dP_{x,\mu}}{dP_0}\, dP_0 = 1.$$

So since D_i and D_j are disjoint,

$$\int_{D_i^c \cap Y_x} \exp\left[\omega_T \log\mu + T - \mu\mathcal{L}(x, \omega)\right]\, dP_0 + \int_{D_j^c \cap Y_x} \exp\left[\omega_T \log\mu + T - \mu\mathcal{L}(x, \omega)\right]\, dP_0 \geq 1.$$

Therefore at least one of these integrals must exceed $1/2$. $\square$

4 Zero-Rate Reliability

Let

$$P_e(n, M, T) := \inf\{P_e(f, \varphi) : (f, \varphi) \text{ is an } (n, M, T) \text{ code}\}$$

and

$$P_e(\cdot, M, T) := \inf_n P_e(n, M, T).$$

Then the *reliability function* (or *error exponent*) of the channel at rate $R > 0$ is defined by

$$E(R) := \limsup_{T \to \infty} -\frac{1}{T} \log P_e(\cdot, \lceil \exp(RT) \rceil, T).$$

The *zero-rate reliability* or *zero-rate error exponent* is defined by

$$E(0) := \lim_{R \to 0} E(R) = \sup_{R>0} E(R).$$

We have followed precedent [3, 8, 4, 11, 9, 16] and defined the reliability function using the limit supremum. Our lower bounds on the reliability function, however, continue to hold if the reliability function is defined using the limit infimum instead. For the ESTC, there is little reason to believe that the limit fails to exist, although this has not been proven, even for discrete memoryless channels. For a timing channel with a general service distribution, however, the limit can fail to exist, as the following example shows. Consider any service distribution that places all of its mass on the points of a lattice in $\mathbb{R}$. Then for any $M \geq 2$, the infimum of the error probability over all $(\cdot, M, T)$ codes is $\Pr(S_1 \geq T)$. If one uses our reliability function definition, this gives

$$E(R) = \limsup_{T \to \infty} -\frac{1}{T} \log \Pr(S_1 \geq T)$$

for all R. One can create a lattice distribution, however, for which the limit infimum is strictly smaller.

We return to the ESTC. Arikan [3] proves the following random-coding and sphere-packing bounds [17] on its reliability function.

Theorem 1 ([3]) *Let $E_{sp}(R)$ be defined parametrically for $0 < R < \mu/e$ by*

$$E_{sp}(R) = \frac{\mu}{(1+\rho)^{(1+\rho)/\rho}}[\rho - \log(1+\rho)],$$
$$R = \frac{\mu}{\rho(1+\rho)^{(1+\rho)/\rho}} \log(1+\rho),$$

where ρ ranges over $(0, \infty)$, and let $E_{sp}(0) = \mu$. For $R_c := (\mu/4)\log 2 < R < \mu/e$, let $E_r(R) = E_{sp}(R)$, while for $0 \leq R \leq R_c$, let $E_r(R) = \mu/4 - R$. Then the reliability function of the ESTC with mean service time $1/\mu$ satisfies

$$E_r(R) \leq E(R) \leq E_{sp}(R)$$

for all $0 \leq R < \mu/e$.

We call $E_r(R)$ the *random-coding exponent* and $E_{sp}(R)$ the *sphere-packing exponent* for the channel. These functions are shown in Figure 1. Recall that the reliability function of the Poisson channel without dark current but with a peak power constraint of μ equals $E_r(R)$ [7, 8]. Thus determining whether this channel and the ESTC have identical reliability functions amounts to determining whether $E(R) = E_r(R)$ for all R. Theorem 1 shows that this holds for R in $[R_c, \mu/e]$, so we focus on rates in $[0, R_c)$, and in particular on rate zero, since the random coding bound is rarely tight in the low-rate regime. In particular, Gallager's expurgated bound [4] shows that the random-coding bound lies strictly below the reliability function at sufficiently low rates for any DMC with nonzero capacity. Indeed, it turns out that $E(0) > E_r(0)$ for the ESTC, although a direct application of the expurgated bound fails because the pairwise error probability bound on which it is based is not easily evaluated. One can replace this bound with (1), however, then mimic the derivation of the expurgated bound using results on the $\overline{\mathcal{L}}$-distance between two randomly chosen codewords (e.g. Corollary 1 to follow). Using this approach one can obtain a lower bound on the reliability function at all rates that, in particular, proves that $E(0) \geq \mu/2 > E_r(R)$.

This bound, while valid at all rates, improves upon the random-coding bound only at very low rates, however. Therefore we discard this approach in favor of one that only yields the zero-rate bound, $E(0) \geq \mu/2$, but that better illustrates the vital role of the *max-min distance*.

A Max-Min Distance

Definition 4 *For n in $\mathbb{N}$ and t in $[0,1)$, let*

$$M(n,t) = \sup\left\{M : \textit{ there exists an } (n, M, 1) \textit{ encoder } f \textit{ with } \overline{\mathcal{L}}(f) \geq t\right\}.$$

Define

$$L_0 = \sup\left\{t \in [0,1) : \limsup_{n\to\infty} \frac{1}{n} \log M(n,t) > 0\right\}.$$

One can verify that $M(n,t) \downarrow n+1$ as $t \uparrow 1$ for all n and that

$$M(n,t) < \infty \text{ for all } n \text{ and } t > 0. \tag{2}$$

We call L_0 the *max-min distance* of the channel. Later we show that it satisfies $L_0 = 1/2$. First we link it to the zero-rate reliability.

Proposition 1 *The zero-rate reliability of the ESTC with mean service time $1/\mu$ satisfies $E(0) = \mu L_0$.*

Proof. We first show that $E(0) \leq \mu L_0$. Let $0 < \delta < E(0)$ and let $\left\{\left(f^{(k)}, \varphi^{(k)}\right)\right\}$ be a sequence of codes such that $f^{(k)}$ is (n_k, M_k, T_k) for each k and

$$T_k \uparrow \infty, \tag{3}$$

$$\lim_{k\to\infty} \frac{1}{T_k} \log M_k =: R > 0, \tag{4}$$

and

$$\lim_{k\to\infty} -\frac{1}{T_k} \log P_e\left(f^{(k)}, \varphi^{(k)}\right) = E(R) \geq E(0) - \delta. \tag{5}$$

If $\sup_k n_k/T_k = \infty$, then we can remedy this by choosing $K > 1$ such that $1+K\log K - K > 2E(R)/\mu$, then discarding any points beyond the first $\lceil K\mu T_k \rceil =: N_k$ in each codeword of $f^{(k)}$. If $g^{(k)}$ is the new encoder then a straightforward coupling argument gives

$$\left|P_e\left(g^{(k)}, \varphi^{(k)}\right) - P_e\left(f^{(k)}, \varphi^{(k)}\right)\right| \le \Pr\left[\sum_{i=1}^{N_k} S_i \le T_k\right],$$

where $S_1, \ldots, S_n$ are i.i.d. exponential with mean $1/\mu$. Invoking the Chernoff bound gives

$$\begin{aligned}\left|P_e\left(g^{(k)}, \varphi^{(k)}\right) - P_e\left(f^{(k)}, \varphi^{(k)}\right)\right| &\le \exp[-N_k(1/K - 1 - \log(1/K))] \\ &\le \exp[-2E(R)T_k],\end{aligned}$$

so that (5) above holds with $g^{(k)}$ in place of $f^{(k)}$. Therefore we shall assume that

$$\sup_k \frac{n_k}{T_k} < \infty.$$

By Lemma 3,

$$P_e\left(f^{(k)}, \varphi^{(k)}\right) \ge \frac{1}{2}\exp\left[-\mu\overline{\mathcal{L}}\left(f^{(k)}\right)\right].$$

Thus

$$0 < E(0) - \delta \le \mu \cdot \liminf_{k\to\infty} \frac{\overline{\mathcal{L}}\left(f^{(k)}\right)}{T_k}.$$

Observe that $\overline{\mathcal{L}}\left(f^{(k)}\right)/T_k$ is the minimum distance of the encoder obtained by time-scaling $f^{(k)}$ to have blocklength one. Then since $M_k \to \infty$ by (3) and (4), and

$$\liminf_{k\to\infty} \overline{\mathcal{L}}\left(f^{(k)}\right)/T_k > 0,$$

it must be that $n_k \to \infty$ by (2). This combined with the observation

$$\liminf_{k\to\infty} \frac{\log M_k}{n_k} \ge \inf_k \frac{T_k}{n_k} \cdot \lim_{k\to\infty} \frac{1}{T_k}\log M_k > 0$$

implies

$$\limsup_{k\to\infty} \frac{\overline{\mathcal{L}}\left(f^{(k)}\right)}{T_k} \le L_0.$$

Thus $E(0) - \delta \le \mu L_0$. Since $\delta > 0$ was arbitrary, we conclude $E(0) \le \mu L_0$.

Now let $0 < \epsilon < L_0$. By the definition of L_0, there exists a sequence of encoders $\{f^{(k)}\}$ such that $f^{(k)}$ is $(n_k, M_k, 1)$ and $\overline{\mathcal{L}}(f^{(k)}) \ge L_0 - \epsilon$ for each k, and

$$\lim_{k\to\infty} \frac{1}{n_k}\log M_k =: R > 0.$$

Let $0 < \lambda < \mu$, and for each k, redefine $f^{(k)}$ to be the $(n_k, M_k, n_k/\lambda)$ encoder obtained by dilating time by a factor of n_k/λ. Consider decoding $f^{(k)}$ using the bounded-distance decoder.

By choosing λ sufficiently small, we may assume that $L_0 - \epsilon > \lambda/\mu$, which implies that $\overline{\mathcal{L}}(f^{(k)})/(n_k/\lambda) > \lambda/\mu$ for each k. Then by Lemma 2,

$$P_e\left(f^{(k)}, \varphi_B\right) \leq \exp\left[-\frac{n_k}{\lambda}\left(\frac{\mu\overline{\mathcal{L}}(f^{(k)})}{n_k/\lambda} - \lambda - \lambda\log\left(\frac{\mu\overline{\mathcal{L}}(f^{(k)})}{n_k}\right)\right)\right]$$
$$\leq \exp\left[-\frac{n_k}{\lambda}\left(\mu(L_0-\epsilon) - \lambda - \lambda\log\left(\frac{\mu(L_0-\epsilon)}{\lambda}\right)\right)\right].$$

Since

$$\lim_{k\to\infty}\frac{1}{n_k/\lambda}\log M_k = \lambda R > 0,$$

we have

$$E(0) \geq \limsup_{k\to\infty} -\frac{1}{n_k/\lambda}\log P_e\left(f^{(k)}, \varphi_B\right)$$
$$\geq \mu(L_0-\epsilon) - \lambda - \lambda\log\left[\frac{\mu(L_0-\epsilon)}{\lambda}\right].$$

Taking $\lambda \to 0$ and $\epsilon \to 0$ establishes that $E(0) \geq \mu L_0$ and completes the proof. $\square$

The problem of determining $E(0)$ for the ESTC is thus reduced to the combinatorial problem of determining L_0. In the next subsection, we lower bound L_0 via a random-coding argument. In the subsequent subsection, we prove a coincident upper bound.

B Direct Result

We define an *n-point Poisson process* over $[0, T]$ to be a Poisson process over this interval conditioned on having exactly n points. We will construct a random $(n, M, 1)$ encoder by choosing the codewords as independent n-point Poisson processes over $[0, 1]$. First we examine the $\overline{\mathcal{L}}$-distance between a pair of n-point Poisson processes, as a prelude to bounding the minimum distance of such an encoder.

If $\left\{F_t^{(n)}\right\}_{t=0}^{1}$ is the empirical distribution function of n i.i.d. random variables that are uniformly distributed over $[0, 1]$, then $nF_t^{(n)}$ is an n-point Poisson process over $[0, 1]$. As such, $\left(nF_t^{(n)} - nt\right)/\sqrt{n}$ converges weakly in $D[0, 1]$ to a Brownian bridge as n tends to infinity, and if $G_t^{(n)}$ is an independent copy of $F_t^{(n)}$, then

$$X_t^{(n)} = \frac{nF_t^{(n)} - nG_t^{(n)}}{\sqrt{2n}}$$

also converges weakly in $D[0, 1]$ to a Brownian bridge [18, Theorem 14.15]. The amount of time that a Brownian bridge over $[0, 1]$ spends strictly positive is uniformly distributed over $[0, 1]$, as is, of course, the amount of time that it spends strictly negative [18, Theorem 13.17]. But the amount of time that $X^{(n)}$ spends strictly positive is precisely the amount by which $nF^{(n)}$ leads $nG^{(n)}$. It follows that $\mathcal{L}\left(nF^{(n)}, nG^{(n)}\right)$ converges weakly to the uniform distribution over $[0, 1]$, as does $\mathcal{L}\left(nG^{(n)}, nF^{(n)}\right)$, and $\overline{\mathcal{L}}\left(nF^{(n)}, nG^{(n)}\right)$ converges weakly to the uniform distribution over $[1/2, 1]$ [18, Lemma 14.10]. From these observations one might guess that $L_0 = 1/2$. This is true, although the proof does not use any of these weak convergence facts explicitly. We do, however, need the following large deviations result.

Lemma 4 *Let $\left\{X_t^{(n)}\right\}_{t=0}^{1}$ and $\left\{Y_t^{(n)}\right\}_{t=0}^{1}$ be independent n-point Poisson processes over* $[0,1]$. *Let*

$$\tau_n = \int_0^1 1\left(X_t^{(n)} = Y_t^{(n)}\right)\,dt.$$

Then for all δ in $[0,1)$,

$$\lim_{n\to\infty} -\frac{1}{n}\log\Pr(\tau_n > \delta) = -\log(1-\delta^2).$$

Proof. Let $\{S_n\}_{n=0}^{2n}$ be a simple random walk on $\mathbb{Z}$, with $S_0 = 0$, and let $A = \{i : S_i = 0\}$. Also let $Z_1,\ldots,Z_{2n}$ be the order statistics of $2n$ i.i.d. uniform $[0,1]$ random variables, independent of S, and write $Z_0 = 0$ and $Z_{2n+1} = 1$. Then

$$\Pr(\tau_n > \delta) = \Pr\left[\sum_{i\in A}(Z_{i+1} - Z_i) > \delta \,\middle|\, S_{2n} = 0\right].$$

Since $\{Z_{i+1} - Z_i\}_{i=0}^{2n}$ are exchangeable and independent of S, this implies

$$\begin{aligned}\Pr(\tau_n > \delta) &= \Pr\left[\sum_{i=0}^{|A|-1}(Z_{i+1} - Z_i) > \delta \,\middle|\, S_{2n} = 0\right]\\ &= \Pr(Z_{|A|} > \delta \mid S_{2n} = 0).\end{aligned}$$

The conditional distribution of $|A|$ can be calculated explicitly,

$$\Pr(|A| = m \mid S_{2n} = 0) = \frac{2^{m-1}(m-1)}{2n-(m-1)}\binom{2n-(m-1)}{n}\binom{2n}{n}^{-1}$$

[19, Section III.4], so

$$\begin{aligned}\Pr(\tau_n > \delta) &= \sum_{m=2}^{n+1}\Pr(Z_m > \delta)\Pr(|A| = m \mid S_{2n} = 0)\\ &= \sum_{m=2}^{n+1}\sum_{i=0}^{m-1}\binom{2n}{i}\delta^i(1-\delta)^{2n-i}\frac{2^{m-1}(m-1)}{2n-(m-1)}\binom{2n-(m-1)}{n}\binom{2n}{n}^{-1}\\ &= \sum_{m=1}^{n}\sum_{i=0}^{m}\binom{2n}{i}\delta^i(1-\delta)^{2n-i}\frac{2^m m}{2n-m}\binom{2n-m}{n}\binom{2n}{n}^{-1}. \qquad (6)\end{aligned}$$

For x and y in $[0,1]$, let

$$f_\delta(x,y) = 2H\left(\frac{y}{2}\right) + y\log\delta + (2-y)\log(1-\delta) + x\log 2 + (2-x)H\left(\frac{1}{2-x}\right) - 2\log 2,$$

where, here and throughout, H is the binary entropy function with natural logarithms. From the combinatorial bound

$$\frac{1}{2n+1}\exp\left[2nH\left(\frac{i}{2n}\right)\right] \leq \binom{2n}{i} \leq \exp\left[2nH\left(\frac{i}{2n}\right)\right], \tag{7}$$

[20, Example 12.1.3], it follows that the summand in (6) is upper bounded by

$$\frac{(2n+1)m}{2n-m}\exp\left[nf_\delta\left(\frac{m}{n},\frac{i}{n}\right)\right].$$

Since $i \leq m$, this is further bounded by

$$\frac{(2n+1)m}{2n-m}\exp\left[n\sup_{0\leq y\leq x\leq 1} f_\delta(x,y)\right].$$

Applying this bound to each term in (6) shows

$$\liminf_{n\to\infty} -\frac{1}{n}\log\Pr(\tau_n > \delta) \geq -\sup_{0\leq y\leq x\leq 1} f_\delta(x,y).$$

Using (7) and the continuity of f_δ, one can show that in fact

$$\lim_{n\to\infty} -\frac{1}{n}\log\Pr(\tau_n > \delta) = -\sup_{0\leq y\leq x\leq 1} f_\delta(x,y).$$

And an elementary calculation shows that

$$\sup_{0\leq y\leq x\leq 1} f_\delta(x,y) = \log\left(1-\delta^2\right).$$

□

Corollary 1 *In the context of the previous lemma, for all* $0 < \delta < 1/2$,

$$\liminf_{n\to\infty} -\frac{1}{n}\log\Pr\left[\overline{\mathcal{L}}\left(X^{(n)},Y^{(n)}\right) < \frac{1}{2}-\delta\right] \geq -\log(1-4\delta^2).$$

Proof. Observe that

$$\left\{\overline{\mathcal{L}}\left(X^{(n)},Y^{(n)}\right) < \frac{1}{2}-\delta\right\} \subset \left\{\int_0^1 1\left(X_t^{(n)} = Y_t^{(n)}\right)\,dt > 2\delta\right\}.$$

Now take probabilities and invoke the previous result. □

Proposition 2 *The max-min distance of the ESTC satisfies* $L_0 \geq 1/2$.

Proof. Let $0 < \delta < 1/2$, and choose γ such that $0 < \gamma < -(1/2)\log(1-4\delta^2)$. Fix n, and consider selecting the codewords of an $(n, \lceil \exp(\gamma n) \rceil, 1)$ encoder $f^{(n)}$ independently as n-point Poisson processes over $[0, 1]$. Then using the notation of Lemma 4 and the union bound,

$$\Pr\left(\overline{\mathcal{L}}\left(f^{(n)}\right) < 1/2 - \delta\right) \leq (\exp(\gamma n) + 1)^2 \Pr\left(\overline{\mathcal{L}}\left(X^{(n)}, Y^{(n)}\right) < 1/2 - \delta\right).$$

Letting n tend to infinity and invoking Corollary 1 shows that $\Pr(\overline{\mathcal{L}}(f^{(n)}) < 1/2 - \delta)$ tends to zero. This proves that for all sufficiently large n, an $(n, \lceil \exp(\gamma n) \rceil, 1)$ encoder exists with a minimum distance of at least $1/2 - \delta$. But δ was arbitrary. □

Before turning to the converse, we note that there is another, entirely different, method of randomly choosing codewords to construct $E(0)$-achieving codes. Let J and K be natural numbers, and let $N_1, N_2, \ldots, N_K$ be i.i.d. random variables uniformly distributed over $\{0, \ldots, J\}$. Construct a codeword with blocklength one by first placing N_1 points at time $1/(K+1)$ and $J - N_1$ points at time $2/(K+1)$. Then place N_2 additional points at time $2/(K+1)$ and $J - N_2$ points at time $3/(K+1)$, etc. Two independent codewords constructed in this fashion will have a $\overline{\mathcal{L}}$-distance that is close to $1/2$ with high probability if K and J are large. In fact, these codewords can replace the Poisson-distributed codewords in the proof of Proposition 2. This construction is inspired by a comment in Section IV of Arikan [3], and is related to the constrained encoders defined later and to the discrete-time models studied by Bedekar and Azizoğlu [21].

C Converse

One approach to showing that the max-min distance is upper bounded by $1/2$ is to bound the minimum distance of an encoder by the average,

$$\overline{\mathcal{L}}(f) \leq \frac{1}{M(M-1)} \sum_{i=1}^{M} \sum_{j \neq i} \overline{\mathcal{L}}(f(i), f(j)),$$

and then show that the average is asymptotically upper bounded by $1/2$. This technique of bounding the minimum distance of an encoder by the average yields a tight upper bound on the max-min distance of the binary-symmetric channel [4], the Gaussian channel [11], and "pairwise reversible" discrete memoryless channels [12], but not the ESTC. Indeed, although we shall see that $L_0 = 1/2$, for arbitrarily large n and M there exists an $(n, M, 1)$ encoder with an average $\overline{\mathcal{L}}$-distance near one: let $0 < \delta < 1$ and consider the $(n, n+1, 1)$ encoder f in which, for each m in $\{1, \ldots, n+1\}$,

$$f_t(m) = \begin{cases} 0 & \text{if } 0 \leq t < \delta \\ m - 1 & \text{if } \delta \leq t < 1 \end{cases}.$$

This encoder has $\overline{\mathcal{L}}(f) = 1 - \delta$, and by replicating all $n+1$ codewords k times, we obtain an $(n, (k+1)(n+1), 1)$ encoder whose minimum distance is zero but whose average distance is

$$\frac{n(k+1)(1-\delta)}{k(n+1)+n},$$

which is close to one if n and k are large and δ is small.

Thus this approach fails for the ESTC, as it does for a general DMC [12]. Shannon, Gallager, and Berlekamp [12] approach the problem for discrete memoryless channels by focusing on a class of encoders they call *ordered*. For the ESTC, the analogue of this class is the following.

Definition 5 *If u and v are elements of Ω_T, then u* leads *v if $\mathcal{L}(u,v) \geq \mathcal{L}(v,u)$. An encoder f is* ordered *if $f(i)$ leads $f(j)$ for all $i \leq j$.*

Shannon, Gallager, and Berlekamp [12] bound the minimum distance of an ordered code using a rather abstruse procedure of manipulating codes that involves transforming the original code into several smaller codes with increasing blocklengths. A simple bound on the size of the largest ordered subcode within a given code, which is reproduced below, then suffices to upper bound the max-min distance of the channel.

We follow this approach, except that we have replaced the code manipulation procedure and some of the surrounding analysis with a simple result about inner-product spaces, Lemma 5. Evidently this replacement is a natural one, since it can also be made in the original proof for discrete memoryless channels, making the proof somewhat more transparent.

The ESTC presents another obstacle that a DMC does not, however. Due to the memory in the channel, the contribution to $\mathcal{L}(u,v)$ made by a small time interval $[t, t+dt]$, $1(u_t > v_t)\, dt$, depends not only on the arrivals during $[t, t+dt]$ but on all previous arrivals. In fact, the effect of an arrival does not diminish with time at all: an arrival near zero has the same effect on $1(u_t > v_t)$ as one just prior to t. Our solution is again to reduce to a class of encoders that mitigate the problem. These encoders have the property that all codewords "agree" at certain times meaning they have all delivered the same number of jobs to the queue at those times. This limits the encoder's use of the long-term memory of the channel.

Definition 6 *An (n, M, T) encoder f is L-*constrained *if for all i and j,*

$$f_{kT/L}(i) = f_{kT/L}(j)$$

for all k in $\{0, \ldots, L-1\}$.

We shall bound the minimum distance of ordered, constrained encoders using the following lemma about inner-product spaces.

Lemma 5 *Let H be a real inner-product space. Let $x_1, \ldots, x_{2^r}$ be elements of H such that $||x_i|| \leq 1$ for all i. For a nonempty subset A of $\{1, \ldots, 2^r\}$, let*

$$x_A = \frac{1}{|A|} \sum_{i \in A} x_i.$$

Then there exist nonempty, contiguous, and disjoint subsets of $\{1, \ldots, 2^r\}$, A and B, such that $||x_A - x_B|| \leq 2/\sqrt{r}$.

Proof. Suppose $||x_A - x_B|| > 2/\sqrt{r}$ for all nonempty, contiguous, and disjoint A and B. Then by the parallelogram law for inner-product spaces [22]

$$||x_{1,2}||^2 = \frac{1}{2}||x_1||^2 + \frac{1}{2}||x_2||^2 - \frac{1}{4}||x_1 - x_2||^2 < 1 - \frac{1}{r}.$$

Similarly,

$$\max(||x_{3,4}||^2, ||x_{5,6}||^2, \ldots, ||x_{2^r-1,2^r}||^2) < 1 - \frac{1}{r}.$$

Then

$$||x_{1,2,3,4}||^2 = \frac{1}{2}||x_{1,2}||^2 + \frac{1}{2}||x_{3,4}||^2 - \frac{1}{4}||x_{1,2} - x_{3,4}||^2 < 1 - \frac{2}{r}.$$

Continuing,

$$0 \leq ||x_{1,\ldots,2^r}||^2 < 1 - \frac{r}{r} = 0.$$

This contradiction proves the result. □

We shall also use the following fact about L^1 distance, whose straightforward proof is omitted.

Lemma 6 *Let $\mathcal{X}$ and $\mathcal{Y}$ be finite alphabets. Let p_1 and p_2 be probability distributions on $\mathcal{X}$, and let q_1 and q_2 be probability distributions on $\mathcal{Y}$. Let $||\cdot||_1$ denote the usual vector 1-norm, so that*

$$||p_1 - p_2||_1 = \sum_{x \in \mathcal{X}} |p_1(x) - p_2(x)|$$

is the L^1 distance between p_1 and p_2. Then

$$||p_1 q_1 - p_2 q_2||_1 \leq ||p_1 - p_2||_1 + ||q_1 - q_2||_1,$$

where $p_1 q_1$ and $p_2 q_2$ denote product measures on $\mathcal{X} \times \mathcal{Y}$.

Lemma 7 *For any ordered, L-constrained $(n, 2^r, 1)$ encoder f,*

$$\overline{\mathcal{L}}(f) \leq \frac{1}{2} + \frac{2\sqrt{1 + n/L}}{\sqrt{r}}.$$

Proof. We may assume that the message set is $\{1, \ldots, 2^r\}$. For a subset I of messages, a time t in $[0, 1]$, and an integer l in $\{0, \ldots, n\}$, let

$$\gamma_t^I(l) = \frac{1}{|I|} \sum_{i \in I} 1(f_t(i) = l).$$

For fixed I, we view γ^I as describing the empirical distribution of the codewords in I as a function of time. Thus we view it as a trajectory through the probability simplex,

$$\gamma^I : [0, 1] \mapsto \left\{ x \in \mathbb{R}_+^{n+1} \text{ such that } \sum_{i=0}^{n} x_i = 1 \right\}.$$

The set of all such trajectories can be embedded in the real linear space

$$\left\{\alpha : [0,1] \mapsto \mathbb{R}^{n+1} : \int_0^1 ||\alpha_t||_2^2 \, dt < \infty\right\},$$

where $||\cdot||_2$ denotes the usual vector 2-norm. We endow this linear space with the inner product

$$< \alpha, \beta > = \int_0^1 (\alpha_t)^T(\beta_t) \, dt,$$

where ^{T} denotes vector transpose. Observe that for every message i,

$$< \gamma^i, \gamma^i > = \int_0^1 ||\gamma_t^i||_2^2 \, dt = 1.$$

So by Lemma 5, there exist disjoint and contiguous sets of messages, A and B, such that

$$\int_0^1 ||\gamma_t^A - \gamma_t^B||_2^2 \, dt \leq \frac{4}{r}. \tag{8}$$

We may assume without loss of generality that all elements of A are strictly less than all elements of B. Consider averaging the $\overline{\mathcal{L}}$-distance between the messages in A and B. Since $\overline{\mathcal{L}}(f)$ cannot exceed this average and f is ordered, we have

$$\begin{aligned}
\overline{\mathcal{L}}(f) \leq \frac{1}{|A|}\frac{1}{|B|}\sum_{i\in A}\sum_{j\in B}\overline{\mathcal{L}}(f(i), f(j)) &= \frac{1}{|A|}\frac{1}{|B|}\sum_{i\in A}\sum_{j\in B}\mathcal{L}(f(i), f(j)) \\
&= \frac{1}{|A|}\frac{1}{|B|}\sum_{i\in A}\sum_{j\in B}\int_0^1 1(f_t(i) > f_t(j)) \, dt \\
&= \int_0^1 \sum_{l=0}^{n}\sum_{m=0}^{n} \gamma_t^A(l)\gamma_t^B(m)1(l > m) \, dt.
\end{aligned}$$

If γ^A and γ^B were equal, then this last integral would be upper bounded by $1/2$. Since they are "close" by (8), the integral cannot exceed $1/2$ by much. To quantify this, we compare γ^A and γ^B to $\gamma^{A\cup B}$:

$$\begin{aligned}
&\sum_{l=0}^{n}\sum_{m=0}^{n} \gamma_t^A(l)\gamma_t^B(m)1(l > m) \\
&\qquad = \sum_{l=0}^{n}\sum_{m=0}^{n} \left[\gamma_t^{A\cup B}(l)\gamma_t^{A\cup B}(m) + \gamma_t^A(l)\gamma_t^B(m) - \gamma_t^{A\cup B}(l)\gamma_t^{A\cup B}(m)\right] 1(l > m).
\end{aligned}$$

Now

$$\sum_{l=0}^{n}\sum_{m=0}^{n} \gamma_t^{A\cup B}(l)\gamma_t^{A\cup B}(m)1(l > m) \leq 1/2,$$

and

$$\sum_{l=0}^{n}\sum_{m=0}^{n} \left[\gamma_t^A(l)\gamma_t^B(m) - \gamma_t^{A\cup B}(l)\gamma_t^{A\cup B}(m)\right] 1(l > m) \leq ||\gamma_t^A\gamma_t^B - \gamma_t^{A\cup B}\gamma_t^{A\cup B}||_1.$$

Lemma 6 shows that

$$||\gamma_t^A\gamma_t^B - \gamma_t^{A\cup B}\gamma_t^{A\cup B}||_1 \le ||\gamma_t^A - \gamma_t^{A\cup B}||_1 + ||\gamma_t^B - \gamma_t^{A\cup B}||_1,$$

and

$$||\gamma_t^A - \gamma_t^{A\cup B}||_1 + ||\gamma_t^B - \gamma_t^{A\cup B}||_1 = ||\gamma_t^A - \gamma_t^B||_1$$

is easily verified by direct calculation. Using these facts and integrating over t, we obtain

$$\overline{\mathcal{L}}(f) \le 1/2 + \int_0^1 ||\gamma_t^A - \gamma_t^B||_1 \, dt. \tag{9}$$

We proceed by using (8) to bound the integral. This can be accomplished by defining

$$\overline{n}_t = \sup_i f_t(i)$$

and

$$\underline{n}_t = \inf_i f_t(i),$$

so that all codewords have delivered between $\underline{n}_t$ and $\overline{n}_t$ jobs to the queue by time t. Then we apply Schwarz's inequality twice to obtain

$$\begin{aligned}
\int_0^1 ||\gamma_t^A - \gamma_t^B||_1 \, dt &= \int_0^1 \sum_{l=\underline{n}_t}^{\overline{n}_t} |\gamma_t^A(l) - \gamma_t^B(l)| \, dt \\
&\le \int_0^1 \sqrt{1+\overline{n}_t - \underline{n}_t} \sqrt{\sum_{l=0}^{n} [\gamma_t^A(l) - \gamma_t^B(l)]^2} \, dt \\
&\le \sqrt{\int_0^1 1 + \overline{n}_t - \underline{n}_t \, dt} \sqrt{\int_0^1 ||\gamma_t^A - \gamma_t^B||_2^2 \, dt.}
\end{aligned} \tag{10}$$

But $\overline{n}_t \le \overline{n}_{(\lfloor tL \rfloor+1)/L}$ and $\underline{n}_t \ge \underline{n}_{\lfloor tL \rfloor/L}$ for all t in $[0,1)$, so

$$\begin{aligned}
\int_0^1 1 + \overline{n}_t - \underline{n}_t \, dt &\le 1 + \int_0^1 \overline{n}_{(\lfloor tL \rfloor+1)/L} - \underline{n}_{\lfloor tL \rfloor/L} \, dt \\
&= 1 + \frac{1}{L}\sum_{i=0}^{L-1} \overline{n}_{(i+1)/L} - \underline{n}_{i/L}.
\end{aligned}$$

Since the encoder is L-constrained, $\overline{n}_{i/L} = \underline{n}_{i/L}$ for all i. Thus the summation is telescoping, and

$$1 + \frac{1}{L}\sum_{i=0}^{L-1} \overline{n}_{(i+1)/L} - \underline{n}_{i/L} = \frac{L + \overline{n}_1 - \underline{n}_0}{L}.$$

Thus

$$\int_0^1 1 + \overline{n}_t - \underline{n}_t \, dt \le \frac{L+n}{L}.$$

Combining this inequality with (8) and (10) gives

$$\int_0^1 \|\gamma_t^A - \gamma_t^B\|_1 \, dt \le \frac{2\sqrt{1+n/L}}{\sqrt{r}}.$$

Combining this with (9) proves the result. □

Lemma 8 *For any $(n, M, 1)$ encoder f and any L, there is a subset of messages, A, such that f restricted to A is L-constrained and*

$$|A| \ge M\binom{L+n-1}{L-1}^{-1}.$$

Proof. Let

$$B = \left\{ x \in \{0, \dots, n\}^L : \sum_{i=1}^{L} x_i = n \right\}.$$

The cardinality of B can be calculated explicitly,

$$|B| = \binom{L+n-1}{L-1}$$

[19, Section II.5]. For x in B, let

$$A_x = \{i : f_{j/L}(i) - f_{(j-1)/L}(i) = x_j \text{ for all } j \in \{1, \dots, L\}\}.$$

Then f restricted to any A_x is L-constrained, and the largest of these sets must have at least

$$M\binom{L+n-1}{L-1}^{-1}.$$

messages. □

The next lemma is Lemma 4.2 in Shannon, Gallager, and Berlekamp [12]. We include a proof for completeness.

Lemma 9 ([12]) *For any $(n, M, 1)$ encoder f, there exists a subset of messages, B, and a bijection $\pi : \{1, \dots, |B|\} \mapsto B$ such that $|B| \ge \log_2 M$ and $g := f \circ \pi$ is an ordered $(n, |B|, 1)$ encoder.*

Proof. Call the original message set A and initialize $B = \emptyset$. Observe that every encoder has at least one message that leads at least half of the other messages. Find such a message i_1 in A, move it from A to B, set $\pi(1) = i_1$, and delete from A all messages that i_1 does not lead. Then repeat the procedure, setting $\pi(2)$ equal to the chosen message i_2. Continue repeating this procedure until A is empty. It follows by induction that after the kth iteration, there are at least $(M+1)/2^k - 1$ messages remaining in A. Thus there is at least one message remaining after $\lfloor \log_2 M \rfloor$ iterations, so at least $\lfloor \log_2 M \rfloor + 1$ iterations are possible. □

Proposition 3 *The max-min distance of the ESTC satisfies $L_0 \le 1/2$.*

Proof. Let $\{f^{(k)}\}_{k=1}^{\infty}$ be a sequence of encoders such that $f^{(k)}$ is $(n_k, M_k, 1)$, $n_k \uparrow \infty$, and

$$\lim_{k\to\infty} \frac{1}{n_k} \log M_k = R > 0.$$

It suffices to show that

$$\limsup_{k\to\infty} \overline{\mathcal{L}}\left(f^{(k)}\right) \leq 1/2.$$

For this, choose $\alpha > 0$ sufficiently small so that

$$(\alpha+1)H\left(\frac{\alpha}{\alpha+1}\right) < R/2,$$

set $L_k = \lceil \alpha n_k \rceil$, and for each k let $\tilde{f}^{(k)}$ be an L_k-constrained $(n_k, \tilde{M}_k, 1)$ encoder with

$$\tilde{M}_k \geq M_k \binom{L_k + n_k - 1}{L_k - 1}^{-1},$$

obtained by restricting $f^{(k)}$ to a subset of messages as in Lemma 8. Since $\overline{\mathcal{L}}\left(f^{(k)}\right) \leq \overline{\mathcal{L}}\left(\tilde{f}^{(k)}\right)$, it suffices to show that

$$\limsup_{k\to\infty} \overline{\mathcal{L}}\left(\tilde{f}^{(k)}\right) \leq 1/2.$$

Now

$$\frac{1}{n_k} \log \tilde{M}_k \geq \frac{1}{n_k} \log M_k - \frac{1}{n_k} \log \binom{L_k + n_k - 1}{L_k - 1}.$$

Using (6), this gives

$$\frac{1}{n_k} \log \tilde{M}_k \geq \frac{1}{n_k} \log M_k - \frac{L_k + n_k - 1}{n_k} H\left[\frac{L_k - 1}{L_k + n_k - 1}\right].$$

The second term converges to $(\alpha+1)H(\alpha/(\alpha+1))$ as k tends to infinity, thus

$$\liminf_{k\to\infty} \frac{1}{n_k} \log \tilde{M}_k > R/2 > 0.$$

This implies that $\tilde{M}_k \to \infty$. Let $r_k = \lfloor \log_2 \log_2 \tilde{M}_k \rfloor$. By Lemma 9, we can find an $(n_k, 2^{r_k}, 1)$ ordered, L_k-constrained encoder, $g^{(k)}$, such that $\overline{\mathcal{L}}\left(g^{(k)}\right) \geq \overline{\mathcal{L}}\left(\tilde{f}^{(k)}\right)$. Applying Lemma 7 to $g^{(k)}$ shows that

$$\overline{\mathcal{L}}\left(g^{(k)}\right) \leq \frac{1}{2} + \frac{2\sqrt{1 + n_k/L_k}}{\sqrt{r_k}}.$$

Since $r_k \to \infty$ and $n_k/L_k \to 1/\alpha$, the proof is complete. □

Theorem 2 follows immediately from Propositions 1 through 3.

Theorem 2 *The zero-rate reliability of the ESTC satisfies $E(0) = \mu/2$, where $1/\mu$ is the mean service time.*

D General Service Distributions

The lower bound on $E(0)$ can be generalized to $\cdot/G/1$ queues with little additional effort. An examination shows that the hypothesis that the service distribution is exponential is used only in Lemma 2 to estimate large deviation probabilities involving the service times. Thus generalizing the lower bound is only a matter of replacing the exponential distribution's large deviations rate function with the rate function of an arbitrary service distribution. We give the details of this procedure, along with a simple upper bound on $E(0)$ for $\cdot/G/1$ timing channels, in this subsection.

Consider a $\cdot/G/1$ queue with i.i.d. service times $S_1, S_2, \ldots$. Let $E[S_1] = 1/\mu$, which we permit to be zero or infinity. Let Λ be the logarithmic moment generating function of S_1,

$$\Lambda(\theta) = \log E[\exp(\theta S_1)],$$

and let Λ^* be its Fenchel-Legendre transform [23],

$$\Lambda^*(x) = \sup_{\theta \in \mathbb{R}} [\theta x - \Lambda(\theta)].$$

Lemma 10 *The error probability of an (n, M, T) encoder f and the bounded-distance decoder φ_B used over a $\cdot/G/1$ timing channel satisfies*

$$P_e(f, \varphi_B) \leq \Pr\left[\sum_{i=1}^{n} S_i > \overline{\mathcal{L}}(f)\right].$$

If $\overline{\mathcal{L}}(f)/T > \lambda/\mu$, where $\lambda = n/T$, this yields

$$P_e(f, \varphi_B) \leq \exp\left[-\lambda T \Lambda^*\left(\frac{\overline{\mathcal{L}}(f)}{\lambda T}\right)\right].$$

Proof. The first conclusion can be proven using the coupling argument used in the proof of Lemma 2. The second conclusion follows from the first by applying the Chernoff bound and using the fact that

$$\Lambda^*(x) = \sup_{\theta \geq 0} [\theta x - \Lambda(\theta)] \tag{11}$$

for all $x \geq 1/\mu$ [23, Lemma 2.2.5]. □

Theorem 3 *For a $\cdot/G/1$ timing channel, the zero-rate reliability satisfies*

$$\frac{1}{2} \liminf_{t \to \infty} -\frac{1}{t} \log \Pr(S_1 \geq t) \leq E(0) \leq \limsup_{t \to \infty} -\frac{1}{t} \log \Pr(S_1 \geq t).$$

There exists a channel for which the lower bound holds with equality while the upper bound is strict, and another channel for which the opposite is true.

Proof. We shall first show that

$$\frac{1}{2} \lim_{x \to \infty} \frac{\Lambda^*(x)}{x} \leq E(0). \tag{12}$$

It follows from standard results in convex analysis [24, Theorems 8.5 and 13.3] that

$$\lim_{x\to\infty} \frac{\Lambda^*(x)}{x} = \sup\{\theta \in \mathbb{R} : \Lambda(\theta) < \infty\}. \tag{13}$$

If

$$\lim_{x\to\infty} \frac{\Lambda^*(x)}{x} = 0,$$

then (12) is obvious. If the limit is positive, then by (13), there exists a positive θ such that $\Lambda(\theta) < \infty$, which implies that $\mu > 0$. Let $0 < \epsilon < 1/2$, and choose λ such that $0 < \lambda < \mu(1/2 - \epsilon)$. By Proposition 2, there exists a sequence of encoders $\{f^{(k)}\}$ such that $f^{(k)}$ is $(n_k, M_k, 1)$ and $\overline{\mathcal{L}}(f^{(k)}) \geq 1/2 - \epsilon$ for each k, and

$$\lim_{k\to\infty} \frac{1}{n_k} \log M_k = R > 0.$$

For each k, redefine $f^{(k)}$ to be the $(n_k, M_k, n_k/\lambda)$ encoder obtained by dilating time by a factor of n_k/λ. We decode using the bounded-distance decoder. By Lemma 10,

$$P_e\left(f^{(k)}, \varphi_B\right) \leq \exp\left[-n_k \Lambda^*\left(\frac{\overline{\mathcal{L}}(f^{(k)})}{n_k}\right)\right].$$

By (11), Λ^* is nondecreasing on $[1/\mu, \infty)$, so since $\overline{\mathcal{L}}(f^{(k)})/n_k \geq (1/2 - \epsilon)/\lambda > 1/\mu$,

$$P_e\left(f^{(k)}, \varphi_B\right) \leq \exp\left[-n_k \Lambda^*\left(\frac{1/2 - \epsilon}{\lambda}\right)\right].$$

This yields a lower bound on $E(0)$,

$$E(0) \geq \lambda \Lambda^*\left(\frac{1/2 - \epsilon}{\lambda}\right).$$

Taking λ to zero then ϵ to zero proves (12). Using the formula

$$E[T] = \int_0^\infty \Pr(T \geq t)\, dt,$$

for nonnegative random variables T, we have for all $\theta \geq 0$,

$$E[\exp(\theta S_1)] = 1 + \int_0^\infty \theta \Pr(S_1 \geq t) \exp(\theta t)\, dt.$$

From this and (13) it follows that[6]

$$\liminf_{t\to\infty} -\frac{1}{t} \log \Pr(S_1 \geq t) \leq \lim_{x\to\infty} \frac{\Lambda^*(x)}{x}, \tag{14}$$

[6]In fact, equality holds in (14), although this fact is not needed.

which proves the lower bound, by (12). The upper bound follows by observing that for any (n, M, T) code (f, φ) with $M \geq 2$,

$$P_e(f, \varphi) \geq \Pr(S_1 \geq T), \tag{15}$$

which implies

$$E(0) \leq \limsup_{T \to \infty} -\frac{1}{T} \log \Pr(S_1 \geq T).$$

By Theorem 2, for the ESTC, the lower bound holds with equality while the upper bound is strict. For the lattice example mentioned at the beginning of this section, on the other hand, the opposite is true. □

If the reliability function is defined using the limit infimum instead of the limit supremum, Theorem 3 can be strengthened to

$$\frac{1}{2} \liminf_{T \to \infty} -\frac{1}{T} \log \Pr(S_1 \geq T) \leq E(0) \leq \liminf_{T \to \infty} -\frac{1}{T} \log \Pr(S_1 \geq T),$$

and equality is achieved by the same channels as before. Evidently, Theorem 3 determines $E(0)$ only when both bounds are infinity or zero. Both bounds are infinity, which implies that $E(0) = \infty$, if the service distribution has bounded support or, for example, if it is a one-sided Gaussian distribution. This latter example is somewhat surprising since the zero-error capacity [4] of the Gaussian service distribution is zero, by (15).

A heavy-tailed service distribution is an example for which both bounds are zero, which implies that $E(0) = 0$. A $\cdot/G/1$ timing channel with a heavy-tailed service distribution with finite mean therefore has the unusual property that its reliability function is zero everywhere, even though its capacity is positive [10]. A consequence is that although the $\cdot/M/1$ queue has the lowest capacity of all $\cdot/G/1$ queues with fixed mean [1, 10], it does not have the lowest reliability at any rate below its capacity.

5 Straight-line Bound

In this section we prove that for the ESTC, any line connecting the point $(0, E(0))$ and a point on the sphere-packing exponent upper bounds the reliability function. That is,

$$E(\theta R) \leq \theta E_{sp}(R) + (1 - \theta)\frac{\mu}{2}, \tag{16}$$

for all R and all θ in $[0, 1]$. The best such bound, called the straight-line exponent, is tangent to the sphere-packing exponent and is shown in Figure 1.

It would be obvious that the straight-line exponent upper bounds the reliability function if it was known that the reliability function of the ESTC is a convex function of the rate. But the reliability function is not known to be convex, even for discrete memoryless channels. Nevertheless, it is possible to prove (16) using an approach developed by Shannon, Gallager, and Berlekamp [16] for discrete memoryless channels that is based on list codes. We shall extend their approach to the ESTC here. First we define list codes.

Definition 7 *An* (n, M, L, T) list code *is an* (n, M, T) *encoder* $f : A \mapsto \Omega_T$ *and a decoder* $\varphi : \Omega_T \mapsto 2^{A'}$*, where* A' *is a superset of* A*, such that for all* ω *in* Ω_T*,* $|\varphi(\omega)| \leq L$*. The* discrimination *of the code is* $(1/T)\log(M/L)$*, and the error probability is*

$$P_e(f, \varphi) = \max_i P_{f(i),\mu}(\omega : i \notin \varphi(\omega)).$$

Let

$$P_e(n, M, L, T) := \inf\{P_e(f, \varphi) : (f, \varphi) \text{ is an } (n, M, L, T) \text{ list code}\}$$

and

$$P_e(\cdot, M, L, T) := \inf_n P_e(n, M, L, T).$$

In Appendix A, we show that for all $R_2 > R_1 \geq 0$,

$$\limsup_{T\to\infty} -\frac{1}{T}\log P_e(\cdot, \lceil\exp(R_2T)\rceil, \lceil\exp(R_1T)\rceil, T) \leq E_{sp}(R_2 - R_1). \tag{17}$$

The proof extends a previously published proof for the case $R_1 = 0$ [3] and corrects a minor omission in it.

Consider now an $(n, M, 1, T)$ code. Temporarily, let us make the unjustified assumption that departures during $(0, T_1]$ and $(T_1, T]$ are conditionally independent given the transmitted codeword. Then for any $L \geq 1$, if L codewords are more likely than the transmitted codeword to cause the observed departures during $(0, T_1]$, and if at least one of those L codewords is more likely than the transmitted codeword to cause the observed departures during $(T_1, T]$, then a ML decoding error will occur. Thus, given our assumption, one is lead to conjecture that [16]

$$P_e(\cdot, M, 1, T) \geq P_e(\cdot, M, L, T_1)P_e(\cdot, L+1, 1, T - T_1). \tag{18}$$

This next result shows that, even without the assumption, (18) is essentially valid.

Lemma 11 *For any* $T = T_1 + T_2$ *and any* M*,* L*, and* n*,*

$$P_e(n, M, 1, T) \geq P_e(\cdot, \lceil M/(n+1)\rceil, L, T_1)P_e(\cdot, L+1, 1, T_2).$$

Proof. Let (f, φ) be an $(n, M, 1, T)$ code. Find a subset A of messages such that $|A| = \lceil M/(n+1)\rceil$ and such that for all i and j in A,

$$f_{T_1}(i) = f_{T_1}(j) =: k.$$

Let $\tilde{f}$ denote f restricted to A. Then $(\tilde{f}, \varphi)$ is an $(n, \lceil M/(n+1)\rceil, 1, T)$ code and

$$P_e(f, \varphi) \geq P_e\left(\tilde{f}, \varphi\right),$$

so it suffices to show

$$P_e\left(\tilde{f}, \varphi\right) \geq P_e(\cdot, \lceil M/(n+1)\rceil, L, T_1)P_e(\cdot, L+1, 1, T_2).$$

For i in A let $D_i = \varphi^{-1}(i)$. Then for all i in A,

$$\begin{aligned}
P_{\tilde{f}(i),\mu}(D_i^c) = \int_{\Omega_T} 1(D_i^c) \exp\Bigg[&\int_0^T \log\left[\mu 1\left(\tilde{f}_t(i) > \omega_{t-}\right)\right] d\omega_t + \\
&\int_0^T \left[1 - \mu 1\left(\tilde{f}_t(i) > \omega_t\right)\right] dt\Bigg] dP_0 \\
= \int_{\Omega_T} 1(D_i^c) \exp\Bigg[&\int_0^{T_1} \log\left[\mu 1\left(\tilde{f}_t(i) > \omega_{t-}\right)\right] d\omega_t + \\
&\int_{T_1}^T \log\left[\mu 1\left(\tilde{f}_t(i) > \omega_{t-}\right)\right] d\omega_t + \\
&\int_0^{T_1} \left[1 - \mu 1\left(\tilde{f}_t(i) > \omega_t\right)\right] dt + \\
&\int_{T_1}^T \left[1 - \mu 1\left(\tilde{f}_t(i) > \omega_t\right)\right] dt\Bigg] dP_0.
\end{aligned}$$

If $\omega^{(1)}$ is a counting function in Ω_{T_1} and $\omega^{(2)}$ is a counting function in Ω_{T_2}, we construct the counting function $(\omega^{(1)}, \omega^{(2)})$ in Ω_T by setting

$$\left(\omega^{(1)}, \omega^{(2)}\right)_t = \begin{cases} \omega_t^{(1)} & \text{for } t \text{ in } [0, T_1] \\ \omega_{T_1}^{(1)} + \omega_{t-T_1}^{(2)} & \text{for } t \text{ in } (T_1, T] \end{cases}.$$

This construction defines an isomorphism between $(\Omega_T, \mathcal{F}_T, P_0)$ and the product space

$$(\Omega_{T_1}, \mathcal{F}_{T_1}, P_0) \times (\Omega_{T_2}, \mathcal{F}_{T_2}, P_0),$$

endowed with the product σ-field and measure [22]. Applying this isomorphism to the integral and using Tonelli's Theorem [22] gives

$$\begin{aligned}
P_{\tilde{f}(i),\mu}(D_i^c) = \int_{\Omega_{T_1}} \exp\Bigg[&\int_0^{T_1} \log\left[\mu 1\left(\tilde{f}_t(i) > \omega_{t-}^{(1)}\right)\right] d\omega_t^{(1)} + \\
\int_0^{T_1} \left[1 - \mu 1\left(\tilde{f}_t(i) > \omega_t^{(1)}\right)\right] dt\Bigg] &\int_{\Omega_{T_2}} 1\left(\left(\omega^{(1)}, \omega^{(2)}\right) \in D_i^c\right) \\
\exp\Bigg[&\int_{T_1}^T \log\left[\mu 1\left(\tilde{f}_t(i) > \omega_{T_1}^{(1)} + \omega_{(t-T_1)-}^{(2)}\right)\right] d\omega_t^{(2)} + \\
&\int_{T_1}^T \left[1 - \mu 1\left(\tilde{f}_t(i) > \omega_{T_1}^{(1)} + \omega_{t-T_1}^{(2)}\right)\right] dt\Bigg] dP_0\, dP_0.
\end{aligned} \tag{19}$$

For a moment let us view the mapping

$$i \in A \mapsto \left\{\tilde{f}_t(i) - k\right\}_{t=T_1}^T$$

as an $(n-k, \lceil M/(n+1) \rceil, T_2)$ encoder. For a given $\omega^{(1)}$ in Ω_{T_1}, this encoder has a natural decoder,

$$\varphi_2\left(\omega^{(2)}\right) = \varphi\left(\omega^{(1)}, \omega^{(2)}\right).$$

The integral over Ω_{T_2} in (19) can be viewed as the error probability of message i in this code, when the queue initially contains $k - \omega_{T_1}^{(1)}$ jobs whose departure times are unknown to the transmitter. A straightforward argument involving genies shows that the presence of the initial jobs can only raise the minimum achievable error probability. Thus there must exist at least one i in A such that

$$\int_{\Omega_{T_2}} 1\left((\omega^{(1)}, \omega^{(2)}) \in D_i^c\right) \exp\left[\int_{T_1}^{T} \log\left[\mu 1\left(\tilde{f}_t(i) > \omega_{T_1}^{(1)} + \omega_{(t-T_1)-}^{(2)}\right)\right] d\omega_t^{(2)} + \int_{T_1}^{T}\left[1 - \mu 1\left(\tilde{f}_t(i) > \omega_{T_1}^{(1)} + \omega_{t-T_1}^{(2)}\right)\right] dt\right] dP_0 \geq P_e(n-k, \lceil M/(n+1)\rceil, 1, T_2).$$

Similarly, in any subset of A of size $L+1$ there must exist an i such that

$$\int_{\Omega_{T_2}} 1\left((\omega^{(1)}, \omega^{(2)}) \in D_i^c\right) \exp\left[\int_{T_1}^{T} \log\left[\mu 1\left(\tilde{f}_t(i) > \omega_{T_1}^{(1)} + \omega_{(t-T_1)-}^{(2)}\right)\right] d\omega_t^{(2)} + \int_{T_1}^{T}\left[1 - \mu 1\left(\tilde{f}_t(i) > \omega_{T_1}^{(1)} + \omega_{t-T_1}^{(2)}\right)\right] dt\right] dP_0 \geq P_e(n-k, L+1, 1, T_2). \quad (20)$$

Let $I(\omega^{(1)})$ be the set of i in A such that (20) holds, and note that this set must contain at least $|A| - L$ messages. Then for all i in A,

$$P_{\tilde{f}(i),\mu}(D_i^c) \geq \int_{\Omega_{T_1}} \exp\left[\int_{0}^{T_1} \log\left[\mu 1\left(\tilde{f}_t(i) > \omega_{t-}^{(1)}\right)\right] d\omega_t^{(1)} + \int_{0}^{T_1}\left[1 - \mu 1\left(\tilde{f}_t(i) > \omega_{t}^{(1)}\right)\right] dt\right] 1\left(i \in I\left(\omega^{(1)}\right)\right) P_e(n-k, L+1, 1, T_2)\, dP_0,$$

so

$$P_e\left(\tilde{f}, \varphi\right) \geq P_e(n-k, L+1, 1, T_2) \cdot \max_i \int_{\Omega_{T_1}} \exp\left[\int_{0}^{T_1} \log\left[\mu 1\left(\tilde{f}_t(i) > \omega_{t-}^{(1)}\right)\right] d\omega_t^{(1)} + \int_{0}^{T_1}\left[1 - \mu 1\left(\tilde{f}_t(i) > \omega_{t}^{(1)}\right)\right] dt\right] 1\left(i \in I\left(\omega^{(1)}\right)\right) dP_0. \quad (21)$$

If we now view

$$i \in A \mapsto \left\{\tilde{f}_t(i)\right\}_{t=0}^{T_1}$$

as a $(k, \lceil M/(n+1)\rceil, T_1)$ encoder, then the maximum over i in (21) is the error probability under the list decoder $\omega^{(1)} \mapsto A \backslash I(\omega^{(1)})$. From this we can conclude

$$\max_i \int_{\Omega_{T_1}} \exp\left[\int_{0}^{T_1} \log\left[\mu 1\left(\tilde{f}_t(i) > \omega_{t-}^{(1)}\right)\right] d\omega_t^{(1)} + \int_{0}^{T_1}\left[1 - \mu 1\left(\tilde{f}_t(i) > \omega_{t}^{(1)}\right)\right] dt\right] 1\left(i \in I\left(\omega^{(1)}\right)\right) dP_0 \geq P_e(k, \lceil M/(n+1)\rceil, L, T_1).$$

Thus

$$P_e\left(\tilde{f},\varphi\right) \geq P_e(n-k, L+1, 1, T_2)P_e(k, \lceil M/(n+1)\rceil, L, T_1)$$
$$\geq P_e(\cdot, L+1, 1, T_2)P_e(\cdot, \lceil M/(n+1)\rceil, L, T_1).$$

□

Theorem 4 is the main result of this section.

Theorem 4 *For the ESTC with service rate μ, for any θ in $[0,1]$ and any $R > 0$,*

$$E(\theta R) \leq \theta E_{sp}(R) + (1-\theta)\frac{\mu}{2}.$$

Proof. We may assume that θ is in $(0,1)$. Let $\left\{(f^{(k)}, \varphi^{(k)})\right\}_{k=1}^{\infty}$ be a sequence of codes such that $f^{(k)}$ is (n_k, M_k, T_k) and

$$T_k \uparrow \infty, \tag{22}$$

$$\lim_{k\to\infty}\frac{1}{T_k}\log M_k = \theta R, \tag{23}$$

and

$$\lim_{k\to\infty} -\frac{1}{T_k}\log P_e\left(f^{(k)}, \varphi^{(k)}\right) = E(\theta R). \tag{24}$$

If $\sup_k n_k/T_k = \infty$, then we can use the technique described in the proof of Proposition 1 to modify the code such that $\sup_k n_k/T_k < \infty$, without violating (22) – (24). Therefore we shall assume that $\sup_k n_k/T_k = \alpha < \infty$. Let $\epsilon > 0$. Then by Lemma 11,

$$P_e\left(f^{(k)}, \varphi^{(k)}\right) \geq P_e\left(\cdot, \lceil M_k/(n_k+1)\rceil, \lceil\exp(\epsilon T_k)\rceil, \theta T_k\right) P_e\left(\cdot, \lceil\exp(\epsilon T_k)\rceil + 1, 1, (1-\theta)T_k\right)$$
$$\geq P_e\left(\cdot, \lceil M_k/(1+\alpha T_k)\rceil, \lceil\exp(\epsilon T_k)\rceil, \theta T_k\right) P_e(\cdot, \lceil\exp(\epsilon T_k)\rceil + 1, 1, (1-\theta)T_k).$$

So by (17) and Theorem 2,

$$\begin{aligned} E(\theta R) &\leq \limsup_{k\to\infty} -\frac{1}{T_k}\log P_e\left(\cdot, \lceil M_k/(1+\alpha T_k)\rceil, \lceil\exp(\epsilon T_k)\rceil, \theta T_k\right) \\ &\quad + \limsup_{k\to\infty} -\frac{1}{T_k}\log P_e\left(\cdot, \lceil\exp(\epsilon T_k)\rceil + 1, 1, (1-\theta)T_k\right) \\ &\leq \theta E_{sp}(R - \epsilon/\theta) + (1-\theta)\frac{\mu}{2}. \end{aligned}$$

But ϵ was arbitrary and $E_{sp}(\cdot)$ is continuous. □

A Sphere-Packing Bound with List Decoding

Arikan [3] proves the sphere-packing bound for the ESTC without list decoding,

$$\limsup_{T\to\infty} -\frac{1}{T}\log P_e(\cdot, \lceil\exp(RT)\rceil, T) \leq E_{sp}(R). \tag{25}$$

The proof is a change of measure or "alternate channel" argument [17, Theorem 2.5.3], where the alternate channel in this case is an ESTC with a reduced service rate. That is, the proof bootstraps a converse coding theorem for the ESTC to obtain the exponential error probability bound (25). An examination of the proof reveals that extending this result to list codes (17) requires generalizing the converse coding theorem to list codes (c.f. [17, p. 196]); the remainder of the argument is the same.

The converse coding theorem used by Arikan, Theorem 4 of Anantharam and Verdú [1], must be replaced on other grounds, however, because Arikan uses the window code definition of Sundaresan and Verdú [10], while Anantharam and Verdú use a different block code definition. In Anantharam and Verdú [1], the receiver observes the entire departure process, and the blocklength of a code is governed by the expected time of the last departure, compared to Definition 1 for window codes. This distinction, while minor, prevents one from easily applying coding theorems for one definition to the other. Sundaresan and Verdú [10] prove a coding theorem for window codes, but they only prove that the overall capacity is μ/e nats per unit time. For the sphere-packing bound, one requires the result that the capacity achievable with codes with average departure rate λ does not exceed $\lambda \log \mu/\lambda$ nats per unit time [3].

Thus a converse coding theorem for window codes with a particular average departure rate is required to properly prove (25) and its extension to list codes (17). Even though this converse is all that is needed, we provide a complete proof of (17) because combining the converse with the change of measure argument shortens the overall proof. Note that this will also prove (25). We begin with two Lemmas about the $\cdot/M/1$ queue, both of which are proved in Appendix B.

Lemma 12 *Let $\{D_t\}_{t=0}^T$ be the departure process of an initially empty $\cdot/M/1$ queue with service rate ν and arrival process $\{A_t\}_{t=0}^T$. Then*

$$E\left[\int_0^T \nu 1(A_t > D_t)\, dt\right] = E[D_T]$$

and

$$E\left[\left(D_T - \int_0^T \nu 1(A_t > D_t)\, dt\right)^2\right] = E[D_T].$$

Lemma 13 *Let $\{D_t\}_{t=0}^T$ be the departure process of an initially-empty $\cdot/M/1$ queue with deterministic arrival process $\{A_t\}_{t=0}^T$. Then* $\mathrm{Var}(D_T) \leq E[D_T]$.

Definition 8 *For x in Ω_T, define the* average departure rate *of a $\cdot/M/1$ queue with input x and service rate ν as*

$$\lambda(x,\nu) = \frac{1}{T}\int_{\Omega_T} \omega_T \, dP_{x,\nu}.$$

For an (n, M, T) encoder f, let

$$\lambda(f,\nu) = \frac{1}{M}\sum_{i=1}^{M} \lambda(f(i),\nu)$$

and

$$C(f,\nu) = \lambda(f,\nu) \log \frac{\nu}{\lambda(f,\nu)}$$

with the convention $C(f,0) = 0$.

The next Lemma lower bounds the error probability of a list code subject to some technical conditions. The bound is a composite of the sphere-packing bound for constant-composition codes [17, Theorem 2.5.3] and the Wolfowitz converse [4, Theorem 5.8.5]. For the remainder of this appendix, we shall abbreviate

$$\alpha = \sup_{x \in (0,1]} x|1 + \log x|$$

and

$$\beta = \sup_{x \in (0,1]} x(1 + \log x)^2,$$

both of which are finite.

Lemma 14 *Let (f, φ) be an (n, M, L, T) list code with discrimination $\Delta > 0$. Suppose that $0 \leq \rho < \nu \leq \mu$ and $0 < \epsilon < \Delta$ are such that*

$$\rho \log \frac{\nu}{\rho} \leq \Delta - \epsilon$$

and at least one of the following holds:

$$\text{(i) } \rho = \lambda(f, \nu) \text{ and } \sup_i |\lambda(f(i), \nu) - \rho| \leq \frac{\epsilon \rho}{4 \alpha \nu}, \text{ or}$$
$$\text{(ii) } \sup_i \lambda(f(i), \nu) \leq \rho \leq \frac{\nu}{e}.$$

Then the error probability of (f, φ) when used over the ESTC with service rate μ satisfies

$$P_e(f, \varphi) \geq \exp\left[-\left(\sup_i \lambda(f(i), \nu) \cdot D(\nu||\mu) + \delta\right) T\right] \cdot \left[1 - \frac{4\nu[D(\nu||\mu)^2 + (1 - \mu/\nu)^2]}{\delta^2 T} - \frac{16\nu(\beta + 1)}{\epsilon^2 T} - \exp\left(-\frac{\epsilon T}{4}\right)\right]$$

for all $\delta > 0$, where

$$D(\nu||\mu) = \log \frac{\nu}{\mu} + \frac{\mu}{\nu} - 1.$$

Proof. If $\rho = 0$ then, under both (*i*) and (*ii*), $\sup_i \lambda(f(i), \nu) = 0$. This implies $\sup_i \lambda(f(i), \mu) = 0$, which implies $P_e(f, \varphi) = 1$ since $\Delta > 0$. Thus the conclusion holds in this case, so suppose that $\rho > 0$. On $(\Omega_T, \mathcal{F}_T)$, define the random variable

$$L = \exp(\omega_T \log \rho + (1 - \rho)T),$$

so that under Q, defined by $dQ/dP_0 = L$, $\{\omega_t\}_{t=0}^T$ is distributed as a Poisson process with rate ρ. Let

$$A_i = \left\{\log \frac{dP_{f(i),\mu}}{dP_{f(i),\nu}} \geq -(\lambda(f(i), \nu) D(\nu||\mu) + \delta)T\right\}$$

and

$$B_i = \left\{ \log \frac{dP_{f(i),\nu}}{dQ} \leq \left(\Delta - \frac{\epsilon}{4}\right) T \right\}.$$

Let $D_i = \{\omega \in \Omega_T : i \in \varphi(\omega)\}$, so that

$$P_e(f, \varphi) = \max_i P_{f(i),\mu}(D_i^c).$$

We will lower bound the average error probability,

$$\frac{1}{M} \sum_{i=1}^{M} P_{f(i),\mu}(D_i^c),$$

which of course will also lower bound $P_e(f, \varphi)$. Now

$$\begin{aligned}
P_{f(i),\mu}(D_i^c) &\geq P_{f(i),\mu}(D_i^c, A_i, B_i) \\
&= \int_{D_i^c \cap A_i \cap B_i} \frac{dP_{f(i),\mu}}{dP_{f(i),\nu}} \, dP_{f(i),\nu} \\
&\geq \int_{D_i^c \cap A_i \cap B_i} \exp\left[-\left(\lambda(f(i),\nu) D(\nu \| \mu) + \delta\right) T\right] dP_{f(i),\nu} \\
&= \exp\left[-\left(\lambda(f(i),\nu) D(\nu \| \mu) + \delta\right) T\right] \left(P_{f(i),\nu}(A_i, B_i) - P_{f(i),\nu}(D_i, A_i, B_i)\right).
\end{aligned}$$

Thus

$$\begin{aligned}
\frac{1}{M} \sum_{i=1}^{M} P_{f(i),\mu}(D_i^c) &\geq \exp\left[-\left(\sup_i \lambda(f(i),\nu) \cdot D(\nu \| \mu) + \delta\right) T\right] \cdot \\
&\qquad \frac{1}{M} \sum_{i=1}^{M} [P_{f(i),\nu}(A_i, B_i) - P_{f(i),\nu}(D_i, A_i, B_i)] \\
&\geq \exp\left[-\left(\sup_i \lambda(f(i),\nu) \cdot D(\nu \| \mu) + \delta\right) T\right] \cdot \\
&\qquad \frac{1}{M} \sum_{i=1}^{M} [1 - P_{f(i),\nu}(A_i^c) - P_{f(i),\nu}(B_i^c) - P_{f(i),\nu}(D_i, A_i, B_i)]. \quad (26)
\end{aligned}$$

Next we bound the summation. Addressing the terms in reverse order,

$$\begin{aligned}
P_{f(i),\nu}(D_i, A_i, B_i) &= \int_{D_i \cap A_i \cap B_i} \frac{dP_{f(i),\nu}}{dQ} \, dQ \\
&\leq \int_{D_i \cap A_i \cap B_i} \exp\left[\left(\Delta - \frac{\epsilon}{4}\right) T\right] \, dQ \\
&\leq \exp\left[\left(\Delta - \frac{\epsilon}{4}\right) T\right] \, Q(D_i).
\end{aligned}$$

Then since $\sum_i 1(D_i) \leq L$,

$$\frac{1}{M}\sum_{i=1}^{M} P_{f(i),\nu}(D_i, A_i, B_i) \leq \exp\left[\left(\Delta - \frac{\epsilon}{4}\right)T\right]\frac{1}{M}\int_{\Omega_T}\sum_{i=1}^{M} 1(D_i)\, dQ$$
$$\leq \exp\left[\left(\Delta - \frac{\epsilon}{4}\right)T\right]\frac{L}{M}$$
$$= \exp\left(-\frac{\epsilon T}{4}\right). \tag{27}$$

Next,

$$\log\frac{dP_{f(i),\nu}}{dQ} = \int_0^T \log\frac{\nu 1(f_t(i) > \omega_{t-})}{\rho}\, d\omega_t + \int_0^T \left[\rho - \nu 1(f_t(i) > \omega_t)\right]\, dt.$$

But

$$\int_0^T \log\frac{\nu 1(f_t(i) > \omega_{t-})}{\rho}\, d\omega_t = \omega_T \log\frac{\nu}{\rho} \quad P_{f(i),\nu} - \text{a.s.},$$

so

$$P_{f(i),\nu}(B_i^c) = P_{f(i),\nu}\left[\omega_T \log\frac{\nu}{\rho} + \rho T - \int_0^T \nu 1(f_t(i) > \omega_t)\, dt > \left(\Delta - \frac{\epsilon}{4}\right)T\right].$$

Since $\rho\log(\nu/\rho) + \epsilon \leq \Delta$ by assumption, this implies

$$P_{f(i),\nu}(B_i^c) \leq P_{f(i),\nu}\left[\omega_T \log\frac{\nu}{\rho} + \rho T - \int_0^T \nu 1(f_t(i) > \omega_t)\, dt > \left(\rho\log\frac{\nu}{\rho} + \frac{3\epsilon}{4}\right)T\right]$$
$$= P_{f(i),\nu}\left[(\omega_T - \lambda(f(i),\nu)T + \lambda(f(i),\nu)T - \rho T)\left(\log\frac{\nu}{\rho} - 1\right)\right.$$
$$\left. - \left(\int_0^T \nu 1(f_t(i) > \omega_t)\, dt - \omega_T\right) > \frac{3\epsilon T}{4}\right].$$

If (i) holds, then

$$(\lambda(f(i),\nu)T - \rho T)\left(\log\frac{\nu}{\rho} - 1\right) \leq |\lambda(f(i),\nu)T - \rho T|\left|\log\frac{\nu}{\rho} - 1\right|$$
$$\leq \frac{\epsilon\rho T}{4\alpha\nu}\left|1 + \log\frac{\rho}{\nu}\right|.$$

Thus by the definition of α, we can conclude that

$$(\lambda(f(i),\nu)T - \rho T)\left(\log\frac{\nu}{\rho} - 1\right) \leq \frac{\epsilon T}{4}. \tag{28}$$

If (*ii*) holds, then the left-hand side of (28) is nonpositive. So either way, (28) holds and

$$\begin{aligned}
P_{f(i),\nu}(B_i^c) &\le P_{f(i),\nu}\left[|\omega_T - \lambda(f(i),\nu)T| \cdot \left|\log\frac{\nu}{\rho} - 1\right| + \right. \\
&\qquad \left. \left|\int_0^T \nu 1(f_t(i) > \omega_t)\,dt - \omega_T\right| > \frac{\epsilon T}{2}\right] \\
&\le P_{f(i),\nu}\left[|\omega_T - \lambda(f(i),\nu)T| \cdot \left|\log\frac{\nu}{\rho} - 1\right| > \frac{\epsilon T}{4}\right] + \\
&\qquad P_{f(i),\nu}\left[\left|\int_0^T \nu 1(f_t(i) > \omega_t)\,dt - \omega_T\right| > \frac{\epsilon T}{4}\right].
\end{aligned}$$

Using Chebyshev's inequality and Lemmas 12 and 13, this gives

$$P_{f(i),\nu}(B_i^c) \le \frac{16\lambda(f(i),\nu)\left(\log\frac{\nu}{\rho} - 1\right)^2}{\epsilon^2 T} + \frac{16\lambda(f(i),\nu)}{\epsilon^2 T}.$$

Thus

$$\begin{aligned}
\frac{1}{M}\sum_{i=1}^{M} P_{f(i),\nu}(B_i^c) &\le \frac{16\rho\left(\left(\log\frac{\nu}{\rho} - 1\right)^2 + 1\right)}{\epsilon^2 T} \\
&\le \frac{16\nu(\beta+1)}{\epsilon^2 T}.
\end{aligned} \tag{29}$$

Using Chebyshev's inequality and Lemmas 12 and 13 in a similar way, one can show that

$$\frac{1}{M}\sum_{i=1}^{M} P_{f(i),\nu}(A_i^c) \le \frac{4\nu[D(\nu\|\mu)^2 + (1-\mu/\nu)^2]}{\delta^2 T}. \tag{30}$$

Substituting (27), (29), and (30) into (26) completes the proof. □

Proof of (17) (c.f. [3, Proposition 3]). It suffices to show that every sequence of list codes $\{(f^{(k)}, \varphi^{(k)})\}_{k=1}^{\infty}$ such that $(f^{(k)}, \varphi^{(k)})$ is (n_k, M_k, L_k, T_k) with $T_k \to \infty$ and

$$\lim_{k\to\infty} \frac{1}{T_k}\log\frac{M_k}{L_k} = \Delta > 0$$

satisfies

$$\liminf_{k\to\infty} -\frac{1}{T_k}\log P_e\left(f^{(k)}, \varphi^{(k)}\right) \le E_{sp}(\Delta). \tag{31}$$

Let F be the set of all nondecreasing functions g mapping $[0,\mu]$ into itself such that $g(0) = 0$ and

$$\sup_{|u-v|\le t} |g(u) - g(v)| \le t$$

for all $t \ge 0$. To see that the average departure rate $\lambda(x,\cdot)$ of any input x to the ESTC must be an element of F, observe that $\lambda(x,\mu)T/\mu$ is the expected amount of time during $[0,T]$ that the rate-μ server is busy when servicing x, by Lemma 12. Then if $\nu \le \mu$, a simple coupling argument shows that

$$\frac{\lambda(x,\nu)T}{\nu} \ge \frac{\lambda(x,\mu)T}{\mu}, \tag{32}$$

which implies

$$0 \le \lambda(x,\mu) - \lambda(x,\nu) \le (\mu - \nu)\frac{\lambda(x,\mu)}{\mu} \le \mu - \nu.$$

Now let K be a natural number, and to each g in F associate the K-dimensional vector of integers y defined by

$$y_i = \left\lfloor \frac{g(i\mu/K)K}{\mu} \right\rfloor$$

for i in $\{1,\dots,K\}$. One can verify that the number of possible vectors is 2^K, and that two functions g and h that map to the same vector must satisfy

$$\sup_{v\in[0,\mu]} |g(v) - h(v)| \le \frac{\mu}{K}.$$

Choose $K = \lceil \mu\sqrt{T_k} \rceil$. Let $A_y^{(k)}$ be the set of messages i such that $\lambda\left(f^{(k)}(i),\cdot\right)$ is mapped to the vector y, and observe that $\left\{A_y^{(k)}\right\}$ is a partition of the message set of $f^{(k)}$. Let $\tilde{f}^{(k)}$ be an $(n_k, \tilde{M}_k, T_k)$ encoder obtained by restricting $f^{(k)}$ to an $A_y^{(k)}$ containing the maximum number of messages. Then $P_e\left(\tilde{f}^{(k)}, \varphi^{(k)}\right) \le P_e\left(f^{(k)}, \varphi^{(k)}\right)$ so it suffices to show

$$\liminf_{k\to\infty} -\frac{1}{T_k}\log P_e\left(\tilde{f}^{(k)}, \varphi^{(k)}\right) \le E_{sp}(\Delta). \tag{33}$$

Note that

$$\lim_{k\to\infty} \frac{1}{T_k}\log\frac{\tilde{M}_k}{L_k} = \Delta,$$

and by omitting finitely many codes in the sequence, if necessary, we may assume that $\tilde{M}_k > L_k$ for all k.

First suppose that $\liminf_k \lambda\left(\tilde{f}^{(k)}, \mu\right) = 0$. Fix any $0 < \rho \le \mu/e$ such that $\rho\log(\mu/\rho) < \Delta$. Then

$$\sup_i \lambda\left(\tilde{f}^{(k)}(i), \mu\right) \le \lambda\left(\tilde{f}^{(k)}, \mu\right) + \frac{1}{\sqrt{T_k}} \le \rho$$

for infinitely many k. For these k, applying Lemma 14 with $\nu = \mu$ gives

$$P_e\left(\tilde{f}^{(k)}, \varphi^{(k)}\right) \ge \exp(-\delta T_k)\cdot\left[1 - \frac{16\mu(\beta+1)}{(\Delta - \rho\log(\mu/\rho))^2 T_k} - \exp\left(-\frac{(\Delta - \rho\log(\mu/\rho))T_k}{4}\right)\right]$$

for all $\delta > 0$. This implies

$$\liminf_{k\to\infty} -\frac{1}{T_k}\log P_e\left(\tilde{f}^{(k)}, \varphi^{(k)}\right) \le \delta.$$

Taking δ to zero proves (33) in this case since E_{sp} is nonnegative.

Next suppose that $\liminf_k \lambda\left(\tilde{f}^{(k)}, \mu\right) > 0$. Let $0 < \epsilon < \Delta$. Because we are proving an asymptotic assertion we can focus on those k for which

$$\Delta_k := \frac{1}{T_k} \log \frac{\tilde{M}_k}{L_k} > \epsilon,$$

since this holds eventually. For these k let

$$\nu_k = \sup\left\{0 < \nu \leq \mu : C\left(\tilde{f}^{(k)}, \nu\right) \leq \Delta_k - \epsilon\right\}.$$

Observe that $C\left(\tilde{f}^{(k)}, \nu\right)$ is continuous[7] in ν so for each k

$$C\left(\tilde{f}^{(k)}, \nu_k\right) \leq \Delta_k - \epsilon,$$

with equality if $\nu_k < \mu$. It follows that $\liminf_k \nu_k > 0$. It also follows that if $\liminf_k \lambda\left(\tilde{f}^{(k)}, \nu_k\right) = 0$, there is a subsequence of codes along which $\lambda\left(\tilde{f}^{(k)}, \nu_k\right) \longrightarrow 0$, and along this subsequence, we have $C\left(\tilde{f}^{(k)}, \nu_k\right) \longrightarrow 0$ and thus $\nu_k = \mu$ eventually. Since this would imply that $\liminf_k \lambda\left(\tilde{f}^{(k)}, \mu\right) = 0$, we must have $\liminf_k \lambda\left(\tilde{f}^{(k)}, \nu_k\right) > 0$. This implies that for all sufficiently large k,

$$\sup_i \left|\lambda\left(\tilde{f}^{(k)}(i), \nu_k\right) - \lambda\left(\tilde{f}^{(k)}, \nu_k\right)\right| \leq \frac{1}{\sqrt{T_k}} \leq \frac{\epsilon\lambda\left(\tilde{f}^{(k)}, \nu_k\right)}{4\alpha\nu_k}.$$

By applying Lemma 14 with $\nu = \nu_k$ and $\rho = \lambda\left(\tilde{f}^{(k)}, \nu_k\right)$ we have, for all sufficiently large k,

$$\begin{aligned} P_e\left(\tilde{f}^{(k)}, \varphi^{(k)}\right) \geq \exp&\left[-\left(\left(\lambda\left(\tilde{f}^{(k)}, \nu_k\right) + \frac{1}{\sqrt{T_k}}\right) D(\nu_k||\mu) + \delta\right) T_k\right] \cdot \\ &\left[1 - \frac{4\nu_k[D(\nu_k||\mu)^2 + (1-\mu/\nu_k)^2]}{\delta^2 T_k} - \frac{16\nu_k(\beta+1)}{\epsilon^2 T_k} - \exp\left(-\frac{\epsilon T_k}{4}\right)\right]. \end{aligned}$$

for all δ. This gives

$$\begin{aligned} \liminf_{k\to\infty} -\frac{1}{T_k} \log P_e\left(\tilde{f}^{(k)}, \varphi^{(k)}\right) &\leq \liminf_{k\to\infty} \left[\left(\lambda\left(\tilde{f}^{(k)}, \nu_k\right) + \frac{1}{\sqrt{T_k}}\right) D(\nu_k||\mu) + \delta\right] \\ &= \liminf_{k\to\infty} \left[\lambda\left(\tilde{f}^{(k)}, \nu_k\right) D(\nu_k||\mu) + \delta\right]. \end{aligned}$$

We can now take δ to zero. Arikan [3] shows that for any $\lambda < \nu \leq \mu$,

$$\lambda D(\nu||\mu) \leq E_{sp}\left(\lambda \log \frac{\nu}{\lambda}\right).$$

[7]It is a consequence of (32) that $C\left(\tilde{f}^{(k)}, \nu\right)$ is also nondecreasing in ν, although this fact is not needed.

If $\nu_k < \mu$, then $C\left(\tilde{f}^{(k)}, \nu_k\right) = \Delta_k - \epsilon$, in which case

$$\lambda\left(\tilde{f}^{(k)}, \nu_k\right) D(\nu_k || \mu) \leq E_{sp}(\Delta_k - \epsilon).$$

This inequality also holds when $\nu_k = \mu$ since in this case $D(\nu_k || \mu) = 0$. Thus

$$\begin{aligned}\liminf_{k\to\infty} -\frac{1}{T_k} \log P_e\left(\tilde{f}^{(k)}, \varphi^{(k)}\right) &\leq \liminf_{k\to\infty} E_{sp}(\Delta_k - \epsilon) \\ &= E_{sp}(\Delta - \epsilon).\end{aligned}$$

since E_{sp} is continuous. Letting ϵ tend to zero and again invoking the continuity of E_{sp} establishes (33) in this case and completes the proof. □

B Proof of Lemmas 12 and 13

Proof of Lemma 12. If $\{X_t\}_{t=0}^T$ is a rate-ν Poisson process, we can construct $\{D_t\}_{t=0}^T$ to satisfy $D_0 = 0$ and

$$D_t = \int_0^t 1(A_s > D_{s-})\, dX_s \tag{34}$$

for all t in $(0, T]$. Then

$$E\left[\left(D_T - \int_0^T \nu 1(A_t > D_t)\, dt\right)^2\right] = E\left[\left(\int_0^T 1(A_t > D_{t-})\,(dX_t - \nu dt)\right)^2\right]. \tag{35}$$

By the martingale calculus [25, Chapter 6, Proposition 4.1],

$$E\left[\left(\int_0^T 1(A_t > D_{t-})\,(dX_t - \nu dt)\right)^2\right] = E\left[\int_0^T \nu 1(A_t > D_t)\, dt\right] \tag{36}$$

and

$$E\left[\int_0^T 1(A_t > D_{t-})\,(dX_t - \nu dt)\right] = 0. \tag{37}$$

The conclusion follows by combining these four equations. □

Proof of Lemma 13. As in the previous proof, we construct $\{D_t\}_{t=0}^T$ to satisfy $D_0 = 0$, and

$$D_t = \int_0^t 1(A_s > D_{s-})\, dX_s$$

for all t in $(0, T]$, where X_t is a Poisson process with rate ν. Let $\overline{D}_t = E[D_t]$ and $V_t = \mathrm{Var}[D_t]$. Then if $0 \leq s < t \leq T$,

$$\begin{aligned}V_t &= E\left[(D_t - \overline{D}_t)^2\right] \\ &= E\left[(D_t - D_s + D_s - \overline{D}_s + \overline{D}_s - \overline{D}_t)^2\right].\end{aligned}$$

Expanding this,

$$
\begin{aligned}
V_t - V_s &= E\left[(D_t - D_s)^2\right] - (\overline{D}_t - \overline{D}_s)^2 + 2E\left[(D_t - D_s)(D_s - \overline{D}_s)\right] \\
&\leq E\left[(D_t - D_s)^2\right] + 2E\left[(D_t - D_s)(D_s - \overline{D}_s)\right] \\
&= E\left[\left(\int_s^t 1(A_u > D_{u-})\, dX_u\right)^2\right] + \\
&\quad 2E\left[\int_s^t 1(A_u > D_{u-})\, dX_u (D_s - \overline{D}_s)\right].
\end{aligned}
$$

We shall show that the second expectation is nonpositive. For any integer-valued random variable $\Delta \leq A_s$ a.s., let $Y_s^\Delta = 0$ and

$$
Y_v^\Delta = \int_s^v 1(A_u > Y_{u-}^\Delta + \Delta)\, dX_u
$$

for v in $(s, t]$. Then

$$
E\left[\int_s^t 1(A_u > D_{u-})\, dX_u (D_s - \overline{D}_s)\right] = E\left[Y_t^{D_s}(D_s - \overline{D}_s)\right].
$$

Let $\tilde{D}_s$ be an independent copy of D_s. Then since $Y_t^{\cdot}$ is nonincreasing a.s. as a function of its superscript,

$$
\left[Y_t^{D_s} - Y_t^{\tilde{D}_s}\right]\left[(D_s - \overline{D}_s) - \left(\tilde{D}_s - \overline{D}_s\right)\right] \leq 0 \quad \text{a.s.}
$$

By taking expectations, multiplying out the factors, and combining like terms we obtain

$$
E\left[Y_t^{D_s}(D_s - \overline{D}_s)\right] \leq E\left[Y_t^{\tilde{D}_s}(D_s - \overline{D}_s)\right].
$$

But $Y_t^{\tilde{D}_s}$ and D_s are independent, so the right side is zero. This proves

$$
\begin{aligned}
V_t - V_s &\leq E\left[\left(\int_s^t 1(A_u > D_{u-})\, dX_u\right)^2\right] \\
&\leq E\left[\left(\int_s^t 1(A_s > D_s \text{ or } A_t > A_s)\, dX_u\right)^2\right] \\
&\leq \Pr(A_s > D_s \text{ or } A_t > A_s)\left[(t-s)\nu + ((t-s)\nu)^2\right]. \qquad (38)
\end{aligned}
$$

One could now bound V_T by proving that V is the indefinite integral of its derivative, and then using (38) to bound the derivative. Instead, we use a direct approximation argument. For any n,

$$
V_T = \sum_{k=0}^{n-1} V_{(k+1)T/n} - V_{kT/n}.
$$

If we apply (38) to each summand and then let $n \to \infty$, the dominated convergence theorem implies

$$
V_T \leq E\left[\int_0^T \nu 1(A_t > D_t)\, dt\right],
$$

which implies $V_T \leq E[D_T]$ by Lemma 12. □

References

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

[2] 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, Sept. 2002.

[3] E. Arikan, "On the reliability exponent of the exponential timing channel," *IEEE Trans. Inform. Theory*, vol. 48, no. 6, pp. 1681–1689, Jun. 2002.

[4] R. G. Gallager, *Information Theory and Reliable Communication.* New York: John Wiley & Sons, 1968.

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

[6] M. H. A. Davis, "Capacity and cutoff rate for Poisson-type channels," *IEEE Trans. Inform. Theory*, vol. 26, no. 6, pp. 710–715, Nov. 1980.

[7] A. D. Wyner, "Capacity and error exponent for the direct detection photon channel—Part I," *IEEE Trans. Inform. Theory*, vol. 34, no. 6, pp. 1449–1461, Nov. 1988.

[8] ——, "Capacity and error exponent for the direct detection photon channel—Part II," *IEEE Trans. Inform. Theory*, vol. 34, no. 6, pp. 1462–1471, Nov. 1988.

[9] A. Lapidoth, "On the reliability function of the ideal Poisson channel with noiseless feedback," *IEEE Trans. Inform. Theory*, vol. 39, no. 2, pp. 491–503, Mar. 1993.

[10] R. Sundaresan and S. Verdú, "Robust decoding for timing channels," *IEEE Trans. Inform. Theory*, vol. 46, no. 2, pp. 405–419, Mar. 2000.

[11] C. E. Shannon, "Probability of error for optimal codes in a Gaussian channel," *Bell Syst. Tech. J.*, vol. XXXVIII, no. 3, pp. 611–656, May 1959.

[12] C. E. Shannon, R. G. Gallager, and E. R. Berlekamp, "Lower bounds to error probability for coding on discrete memoryless channels. II," *Inform. Contr.*, vol. 10, pp. 522–552, 1967.

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

[14] P. Billingsley, *Convergence of Probability Measures.* New York: John Wiley & Sons, 1968.

[15] H. L. Royden, *Real Analysis*, 3rd ed. Englewood Cliffs: Prentice Hall, 1988.

[16] C. E. Shannon, R. G. Gallager, and E. R. Berlekamp, "Lower bounds to error probability for coding on discrete memoryless channels. I," *Inform. Contr.*, vol. 10, pp. 65–103, 1967.

[17] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Budapest: Akadémiai Kiadó, 1981.

[18] O. Kallenberg, *Foundations of Modern Probability*, 2nd ed. New York: Springer-Verlag, 2002.

[19] W. Feller, *An Introduction to Probability Theory and Its Applications*, 2nd ed. New York: John Wiley & Sons, 1957, vol. I.

[20] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: John Wiley & Sons, 1991.

[21] A. S. Bedekar and M. Azizoğlu, "The information-theoretic capacity of discrete-time queues," *IEEE Trans. Inform. Theory*, vol. 44, no. 2, pp. 446–461, Mar. 1998.

[22] G. B. Folland, *Real Analysis: Modern Techniques and Their Applications*, 2nd ed. New York: John Wiley & Sons, 1999.

[23] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*. Boston: Jones and Bartlett, 1993.

[24] R. T. Rockafellar, *Convex Analysis*. Princeton University Press, 1970.

[25] E. Wong and B. Hajek, *Stochastic Processes in Engineering Systems*. New York: Springer-Verlag, 1985.

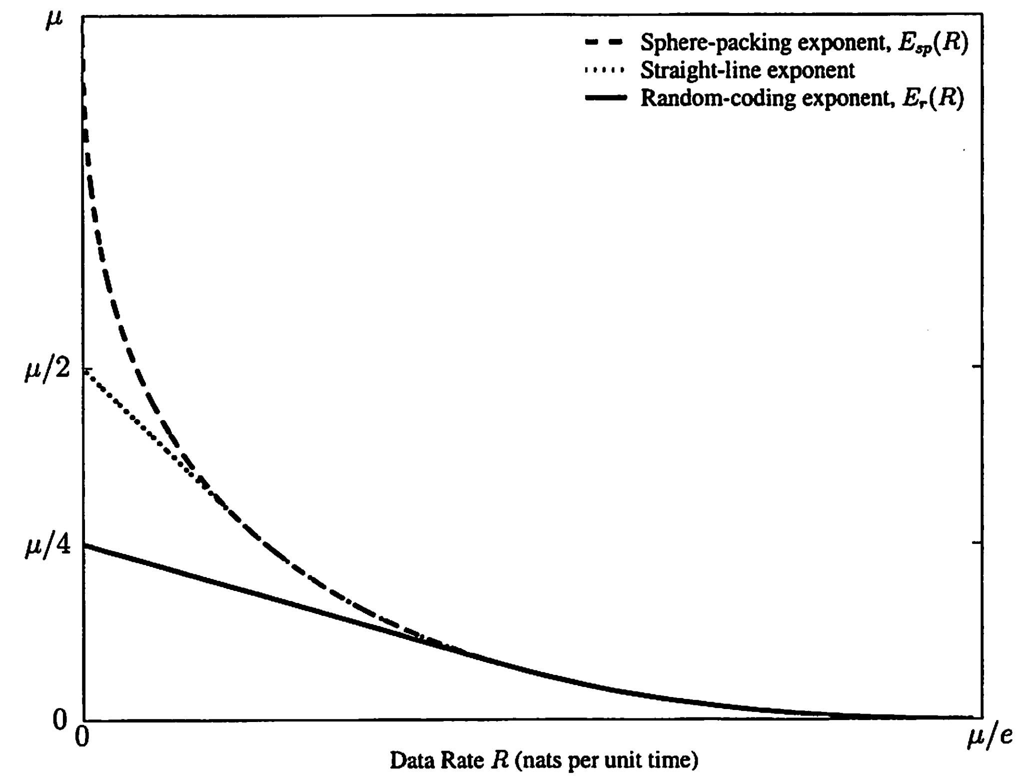

Figure 1: Sphere-packing, straight-line, and random-coding exponents.