On the Reliability Exponent of the Exponential Timing Channel

Erdal Arikan, *Senior Member, IEEE*

Abstract—We determine the reliability exponent $E(R)$ of the Anantharam–Verdú exponential server timing channel with service rate μ for all rates R between a critical rate $R_c = (\mu/4) \log 2$ and the channel capacity $C = e^{-1}\mu$. For rates between 0 and R_c, we provide a random-coding lower bound $E_r(R)$ and a sphere-packing upper bound $E_{sp}(R)$ on $E(R)$. We also determine that the cutoff rate R_0 for this channel equals $\mu/4$, thus answering a question posed by Sundaresan and Verdú. An interesting aspect of our results is that the lower bound $E_r(R)$ for the reliability exponent of the timing channel coincides with Wyner's reliability exponent for the photon-counting channel with no dark current and with peak power constraint μ. Whether the reliability exponents of the two channels are actually equal everywhere remains open. This shows that the exponential server timing channel is at least as reliable as this type of a photon-counting channel for all rates.

Index Terms—Cutoff rate, photon-counting channel, point process channel, Poisson channel, reliability exponent, sphere-packing exponent, timing channel.

I. INTRODUCTION

NANTHARAM and Verdú [1] introduced the exponential server timing channel and determined that its capacity equals $e^{-1}\mu$ nats per second. More recently, Sundaresan and Verdú [10] considered sequential decoding for this channel and found that its cutoff rate is at least $e^{-1}\mu/2$, half the capacity. They posed the determination of the exact value of the cutoff rate as an open problem. In this paper, we show that the cutoff rate equals $\mu/4$. The main contribution of this paper, however, is the more comprehensive result about the reliability exponent of the channel. We begin by giving a precise statement of the problem.

The exponential server timing channel consists of a $\cdot/M/1$ queue with a first-in first-out service discipline. Service times are independent and exponentially distributed with mean $1/\mu$. Information transmission begins at time 0 with the queue containing a possibly nonzero amount u_0 of unfinished work. The model allows u_0 to be chosen from a probability distribution $r(u_0)$, but u_0, as a random variable, must be independent of the message transmitted. The information to be transmitted is encoded into the arrival process, represented by the interarrival times $x_1, x_2, \ldots$ of customers to the queue. The receiver makes its decisions about the transmitted messages based on its observation of the interdeparture times $y_1, y_2, \ldots$ of customers from the queue, and also its noiseless observation $y_0 = u_0$ of the initial unfinished work.

Given the interarrival times $x^n = (x_1, \ldots, x_n)$ for the first n customers, and the initial load u_0, the conditional probability density of observing y_0 and $y^n = (y_1, \ldots, y_n)$ is specified as

$$p(y^n, y_0 | x^n, u_0) = \delta(y_0 - u_0) \prod_{i=1}^{n} e_\mu(y_i - w_i) \quad (1)$$

where $\delta(\cdot)$ is the delta function, $e_\mu(s) = \mu \exp(-\mu s)$, $s \geq 0$, is the exponential density with mean $1/\mu$, and

$$w_i = \max \left\{ 0, \sum_{j=1}^{i} x_j - \sum_{j=0}^{i-1} y_j \right\} \quad (2)$$

is the server's idling time before the arrival of the ith customer. The quantity $y_i - w_i$ is the service time for customer i and is exponential with parameter μ.

There is more than one reasonable definition of a block code for the timing channel, leading to different capacity results, as explained in [1]. In this paper, we consider only the type of block codes termed *window code* in [9]. By a block code with parameters (n, M, T) we shall mean a collection of M vectors of length n, $x^n(m) = (x_1(m), \ldots, x_n(m))$, $1 \leq m \leq M$. The components of codewords are nonnegative numbers that correspond to customer interarrival times. Transmitting codeword $x^n(m)$ means sending n customers to the queue with the arrival time of the kth customer given by $\sum_{i=1}^{k} x_i(m)$, $1 \leq k \leq n$. The parameter T denotes the deadline for sending the nth customer, thus, $\sum_{i=1}^{n} x_i(m) \leq T$ for every m. The rate of such a code is defined as $R = (\log M)/T$ nats per second. (Throughout, we use natural logarithms.)

A decoder for such a code consists of a collection of disjoint decoding regions D_m, $1 \leq m \leq M$, such that the decoder decides that message m was sent when the observed channel output (y^n, y_0) falls in D_m. The probability of decoding error when message m is transmitted is given by

$$P_{e,m} = \int_{D_m^c} p[y^n | x^n(m), y_0] r(y_0) \, dy^n \, dy_0. \quad (3)$$

The average probability of error for the decoder is defined by

$$P_e = \frac{1}{M} \sum_{m=1}^{M} P_{e,m}. \quad (4)$$

In [9], it is observed that a maximum-likelihood (ML) decoder for an (n, M, T) code on an exponential timing channel may,
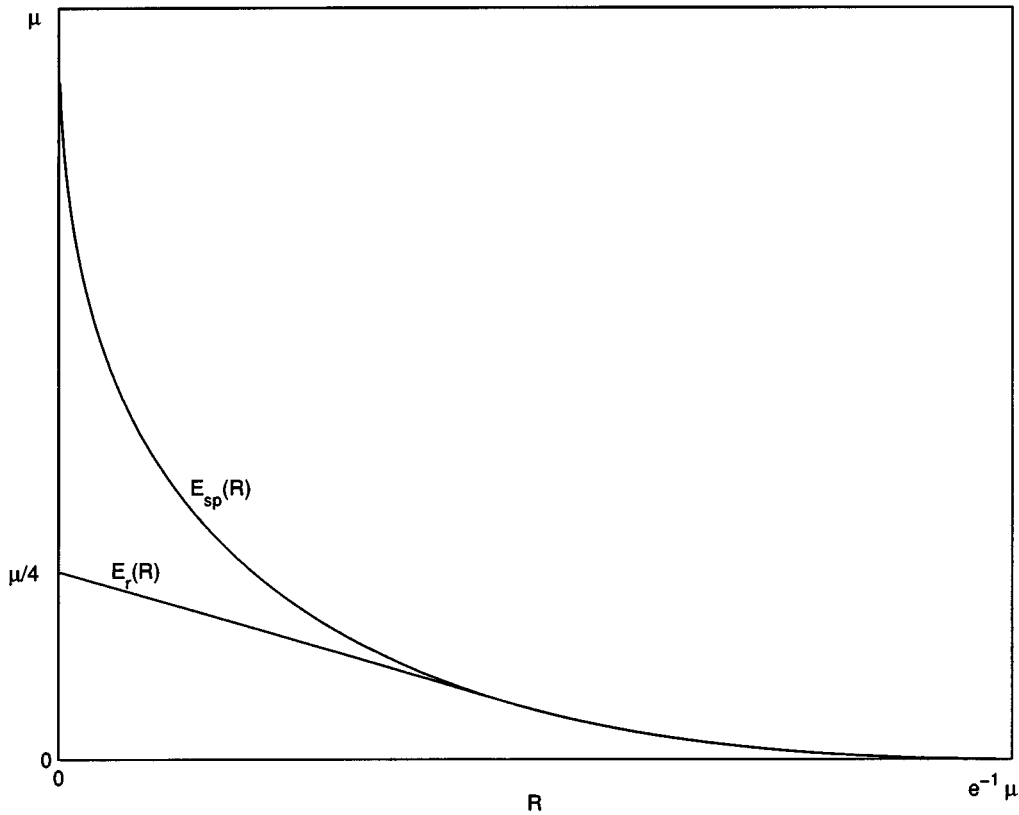
Fig. 1.   Random-coding and sphere-packing exponents.

without loss of optimality, restrict its observation of the channel output to the time window $[0, T]$.

A rate R nats per second is said to be achievable on the $\cdot/M/1$ channel with service rate μ if for any given $\gamma > 0$ and $\epsilon > 0$ there exists an (n, M, T) code with rate greater than $R - \gamma$ and probability of ML decoding error less than ϵ. The supremum of achievable rates is called the capacity of the channel and is given by $C = \mu e^{-1}$ [1], [9].

While the capacity results tell us the highest achievable rates, they do not say much about how fast the probability of ML decoding error can be made to approach zero. To achieve smaller error probabilities at a fixed rate R, one needs to use longer codes as measured by the parameter T. This leads to the following definition. Given a $\cdot/M/1$ channel, its reliability exponent at rate $R > 0$ is defined as

$$E(R) = \limsup_{T \to \infty} -\frac{\log P_e(R, T)}{T} \qquad (5)$$

where $P_e(R, T)$ is the minimum value of the ML decoding error P_e over all (n, M, T) codes for the given channel such that $M = \lceil e^{TR} \rceil$ and $n \geq 1$ is arbitrary. The zero-rate exponent is defined as

$$E(0) = \sup_{R > 0} E(R). \qquad (6)$$

The main result of this paper is the following.

Proposition 1: The reliability exponent of the exponential timing channel with service rate μ is bounded, for every $0 \leq R \leq C = e^{-1}\mu$, by

$$E_r(R) \leq E(R) \leq E_{sp}(R) \qquad (7)$$

where the functions $E_r(R)$ and $E_{sp}(R)$ are defined as follows.

Random-Coding Exponent: For $R \geq C$, $E_r(R) = 0$. For $R_c \leq R < C$, where $R_c \triangleq (\mu/4) \log 2$, $E_r(R)$ is given by the pair of parametric equations

$$E_r(R) = \frac{\mu}{(1+\rho)^{(1+\rho)/\rho}} \left[\rho - \log(1+\rho)\right] \qquad (8)$$

$$R = \frac{\mu}{(1+\rho)^{(1+\rho)/\rho}} \log(1+\rho)^{1/\rho} \qquad (9)$$

where ρ ranges in the interval $(0, 1]$. For $0 \leq R \leq R_c$, $E_r(R) = \mu/4 - R$.

Sphere-Packing Exponent: For $R \geq C$, $E_{sp}(R) = 0$. For $0 < R < C$, $E_{sp}(R)$ is given by

$$E_{sp}(R) = \frac{\mu}{(1+\rho)^{(1+\rho)/\rho}} \left[\rho - \log(1+\rho)\right] \qquad (10)$$

$$R = \frac{\mu}{(1+\rho)^{(1+\rho)/\rho}} \log(1+\rho)^{1/\rho} \qquad (11)$$

where ρ ranges in the interval $0 < \rho < \infty$. Finally, $E_{sp}(0) = \mu$.

We observe that, for rates $R_c \leq R \leq C$, $E_r(R) = E_{sp}(R)$ and $E(R)$ is completely determined. For rates $0 \leq R < R_c$, $E(R)$ remains unknown. A sketch of the functions $E_r(R)$ and $E_{sp}(R)$ is given in Fig. 1.

One of our motivations for studying the reliability exponent of this channel was to determine its *cutoff rate*, which is defined as the largest rate at which sequential decoding can be used within finite expected computation per decoded digit. As shown in [2], the cutoff rate parameter coincides with the R_0 parameter which is the vertical axis intercept of the supporting line of slope -1 to the curve $E(R)$, and for the channel here is given by $R_0 = E_r(0) = \mu/4$. This answers a question raised in [10] about the exact value of the cutoff rate of sequential decoding on this channel.

We finally wish to discuss our results with regard to the analogy, first mentioned in [1], between the exponential timing channel (ETC) and the direct-detection photon channel (also known as the Poisson channel, point process channel, etc.). The input to the photon channel is a waveform $\lambda(t)$ which modulates the power of a transmitted optical field. A photon detector at the receiver produces counts that are modeled as a Poisson process with rate $\lambda_0 + \lambda(t)$, where λ_0 corresponds to inherent background noise (dark current). In general, a photon channel is characterized by its dark-current intensity λ_0 and constraints on peak and average values of $\lambda(t)$. We shall refer to the special case of a photon channel with zero dark current $\lambda_0 = 0$ and only a peak power constraint $0 \leq \lambda(t) \leq \mu$ as an *ideal* photon channel (IPC).

The capacity of an IPC is given by $e^{-1}\mu$ [6], [4], which coincides with the capacity of an ETC with service rate μ. It is interesting that the IPC capacity is achieved by using waveforms with average power $e^{-1}\mu$, analogous to the fact that the capacity-achieving arrival rate for the ETC is $\lambda = e^{-1}\mu$. This analogy carries further to the reliability exponents of the two channels, as we discuss next.

In a seminal work [11], [12], Wyner determined the reliability exponent of the photon channel for all rates between zero and channel capacity. Wyner's result is fully general, covering arbitrary λ_0 and arbitrary peak and average power constraints. It is interesting to note that, as Wyner noted, the photon channel is one of a select few nontrivial channels, such as the infinite-bandwidth additive Gaussian noise channel, whose reliability exponent is known exactly for all rates below capacity.

Wyner's exponent for the special case of an IPC coincides with the random-coding exponent $E_r(R)$ of an ETC, as given by (8), (9). Thus, the reliability exponents for the two channels are identical for the range of rates $R_c \leq R \leq C$. Whether the two exponents are everywhere identical remains an interesting open question.

We should mention in this connection that, for the case of an IPC with noiseless feedback, Lapidoth [7] obtained a lower bound on the reliability exponent which coincides with the sphere-packing exponent $E_{sp}(R)$ of an ETC, as given by (10), (11). (He also showed that if the average power is constrained to $\leq e^{-1}\mu$, then this lower bound is tight.)

Lapidoth's result is particularly relevant here in that an IPC with feedback is equivalent to a "telephone signaling channel" (TSC) with customer recall, another timing channel model discussed in [1]. The basic TSC (without customer recall) is an ETC with feedback. A TSC with customer-recall capability is an ETC with feedback and with the option of recalling a customer in service any time the encoder wishes.

All variants of the timing channel and the ideal photon channel previously mentioned have the same value $e^{-1}\mu$ for capacity. Ex-

planations for the insensitivity of the capacity under variations including feedback and customer recall have been given in [1], [9], and [8]. Except for the IPC (without feedback), the reliability exponents of these channels are not known fully.

This completes the summary of our results. In Sections II and III, we derive the random-coding exponent $E_r(R)$ and the sphere-packing exponent $E_{sp}(R)$, respectively. Section IV is a discussion of results.

II. RANDOM-CODING BOUND

This section is based on Gallager's results on computing the reliability exponent for channels with continuous inputs and outputs [5, Ch. 7]. We regard $\mu > 0$ as a fixed parameter throughout. Initially, we also fix $0 < \lambda < \mu$ and $R > 0$ and for each $n \geq 1$ consider an ensemble of (n, M, T) codes with $T = n/\lambda + \sqrt{n}$, $M = \lceil e^{TR} \rceil$. A code in the ensemble with codewords $x^n(1), \ldots, x^n(M)$ is assigned the probability $\prod_{m=1}^{M} q(x^n(m))$ where

$$q(x^n) = \alpha^{-1}\phi(x^n) \prod_{i=1}^{n} e_\lambda(x_i) \qquad (12)$$

with

$$\phi(x^n) = \begin{cases} 1, & \text{for } \frac{1}{\lambda} - \frac{1}{\sqrt{n}} < \frac{1}{n}\sum_{i=1}^{n} x_i < \frac{1}{\lambda} + \frac{1}{\sqrt{n}} \\ 0, & \text{otherwise.} \end{cases} \qquad (13)$$

The term α is the normalizing factor and it equals the probability that a codeword chosen at random from the distribution $\prod_{i=1}^{n} e_\lambda(x_i)$ satisfies the "shell constraint" $\phi(x^n) = 1$. A straightforward application of the central limit theorem shows that $\alpha = e^{-o_n(1)}$.[1]

Gallager's bound [5, Ch. 7] on the ensemble average of the probability of ML decoding error, when the mth codeword is transmitted, is given by

$$\overline{P}_{e,m} \leq M^\rho \int dy^n \, dy_0$$
$$\times \left[\int dx^n \, du_0 q(x^n) r(u_0) p(y^n, y_0 | x^n, u_0)^{1/(1+\rho)} \right]^{1+\rho} \qquad (14)$$

where ρ is an arbitrary parameter within the interval $0 \leq \rho \leq 1$. The choice of $r(u_0)$ does not affect the bound asymptotically but a convenient choice facilitates its evaluation.

For our purposes, it turns out that a weaker version of this bound (which amounts to assuming that the observation of u_0 is unavailable to the decoder) is sufficient. Using Minkowsky's inequality [5, p. 523], we obtain from (14)

$$\overline{P}_{e,m} \leq M^\rho \int dy^n \left[\int dx^n \, du_0 q(x^n) r(u_0) \right.$$
$$\left. \times \left[\int dy_0 p(y^n, y_0 | x^n, u_0) \right]^{1/(1+\rho)} \right]^{1+\rho}$$
$$= M^\rho \int dy^n \left[\int dx^n \, du_0 q(x^n) r(u_0) \right.$$
$$\left. \times p(y^n | x^n, u_0)^{1/(1+\rho)} \right]^{1+\rho}. \qquad (15)$$

[1]We write $o_z(1)$ to denote an unspecified positive-valued function of z that goes to zero as z goes to infinity. We write $o(z)$ to denote a function such that $o(z)/z = o_z(1)$.

Next, we note that for any $0 \leq s < \lambda$

$$q(x^n) \leq \alpha^{-1} e^{s(\sum_i x_i - n/\lambda + \sqrt{n})} \prod_{i=1}^n e_\lambda(x_i)$$

$$= \alpha^{-1} e^{-sn/\lambda + s\sqrt{n}} \left(\frac{\lambda}{\lambda - s} \right)^n \prod_{i=1}^n e_{\lambda-s}(x_i)$$

and also observe that

$$p(y^n | x^n, u_0)^{1/(1+\rho)} = \left[\frac{1+\rho}{\mu^{\rho/(1+\rho)}} \right]^n \prod_{i=1}^n e_{\mu/(1+\rho)}(y_i - w_i). \tag{16}$$

Substituting these into (15), we obtain

$$\overline{P}_{e,m} \leq cM^\rho \int dy^n \left[\int dx^n \, du_0 r(u_0) \prod_{i=1}^n e_{\lambda-s}(x_i) \right.$$
$$\left. \times \prod_{i=1}^n e_{\mu/(1+\rho)}(y_i - w_i) \right]^{1+\rho} \tag{17}$$

where for shorthand we put

$$c = \alpha^{-(1+\rho)} e^{s(1+\rho)\sqrt{n}} \left[\frac{e^{-s/\lambda}\lambda(1+\rho)}{(\lambda - s)\mu^{\rho/(1+\rho)}} \right]^{n(1+\rho)}.$$

The bracketed integral in (17) gives the distribution of the departure times y^n in an $M/M/1$ queue with service rate $\mu/(1+\rho)$, arrival rate $\lambda - s$, and initial load u_0. We now specify $r(u_0)$ as the exponential density with parameter $\mu/(1+\rho) - (\lambda - s)$, and restrict s and λ further so that $\mu/(1+\rho) - (\lambda - s) \geq 0$. The $M/M/1$ queue now starts in equilibrium, and by Burke's theorem, the interdeparture times $y_1, \ldots, y_n$ are independent, exponential random variables with parameter $\lambda - s$ (see [1] for a discussion of this point). Hence, (17) can be rewritten in this case as

$$\overline{P}_{e,m} \leq cM^\rho \int dy^n \left[\prod_{i=1}^n e_{\lambda-s}(y_i) \right]^{1+\rho}$$
$$= cM^\rho \left[(1+\rho)^{-1}(\lambda - s)^\rho \right]^n$$
$$= e^{o(n)} M^\rho \left[\frac{e^{-s(1+\rho)/\lambda}\lambda^{(1+\rho)}(1+\rho)^\rho}{(\lambda - s)\mu^\rho} \right]^n \tag{18}$$

where $e^{o(n)} = \alpha^{-(1+\rho)} e^{s(1+\rho)\sqrt{n}}$.

We now put $s = \lambda\rho/(1+\rho)$ (this choice minimizes the bound (18) if we ignore the $e^{o(n)}$ term) to obtain

$$\overline{P}_{e,m} \leq e^{o(n)} M^\rho \left[\frac{e^{-\rho}\lambda^\rho(1+\rho)^{1+\rho}}{\mu^\rho} \right]^n$$
$$= \exp\{-T[E_0(\rho, \lambda) - \rho R] + o(T)\} \tag{19}$$

which is valid for any $0 < \lambda < \mu$. In writing (19), we defined

$$E_0(\rho, \lambda) = -\lambda \log \frac{e^{-\rho}\lambda^\rho(1+\rho)^{1+\rho}}{\mu^\rho} \tag{20}$$

and made use of the facts that $n = \lambda T - \sqrt{n}$ and $o(n) = o(T)$. We now set λ equal to

$$\lambda_\rho = \frac{\mu}{(1+\rho)^{(1+\rho)/\rho}} \tag{21}$$

(this choice maximizes the function $E_0(\rho, \lambda)$ over $\lambda \geq 0$ for a fixed $\rho > 0$) and define

$$E_0(\rho) = E_0(\rho, \lambda_\rho) = \frac{\rho\mu}{(1+\rho)^{(1+\rho)/\rho}} \tag{22}$$

to conclude that for the ensemble with $\lambda = \lambda_\rho$, (19) holds with $E_0(\rho)$ in place of $E_0(\rho, \lambda)$.

To obtain the tightest possible bound for a fixed R, a further maximization is required over $0 \leq \rho \leq 1$. So, we define

$$E_r(R) = \max_{0 \leq \rho \leq 1} [E_0(\rho) - \rho R]. \tag{23}$$

It is straightforward to check that the function $E_0(\rho) - \rho R$ is a differentiable concave function of $\rho \geq 0$ for any given $R \geq 0$. Following the method in [5, Sec. 5.6], the value of $E_r(R)$ is determined as stated in Proposition 1. In particular, we note that the right-hand side of (9) equals $\partial E_0(\rho)/\partial\rho$. Now, for the ensemble with $\lambda = \lambda_\rho$

$$\overline{P}_{e,m} \leq \exp[-TE_r(R) + o(T)]. \tag{24}$$

The next step is to convert this bound on an ensemble average to a bound that applies to a specific code. For this one considers an ensemble of codes with $2M$ codewords and uses expurgation (see, e.g., [5, Sec. 5.6]). The final result is as follows.

Proposition 2: Let $\mu > 0$ and $R > 0$ be given numbers. For any $n \geq 1$, there exists an (n, M, T) code with $M = \lceil e^{TR} \rceil$ such that the probability of ML decoding error on a $\cdot/M/1$ channel with service rate μ satisfies

$$P_{e,m} \leq \exp[-TE_r(R) + o(T)] \tag{25}$$

for each message m, $1 \leq m \leq M$.

III. SPHERE-PACKING BOUND

The sphere-packing bound on the probability of ML decoding error for a given code on a $\cdot/M/1$ timing channel is obtained by considering the use of the same code on an alternate "noisier" channel. Although the underlying idea of the proof to be presented in this section is standard, the continuous and nonmemoryless nature of the channel presents some special difficulties; so, we give a detailed proof.

Proposition 3: Let $R > 0$ be fixed. Let $\{C_i\}$, $i = 1, 2, \ldots$, be a sequence of codes for the $\cdot/M/1$ timing channel with service rate μ. Let (n_i, M_i, T_i) denote the parameter of C_i, R_i its rate, and $P_e(C_i)$ the probability of ML decoding error starting with initial load $u_0 = 0$. Suppose the code lengths T_i increase to infinity and that the rates R_i converge to R. Then

$$\limsup_{i \to \infty} -\frac{1}{T_i} \log P_e(C_i) \leq E_{sp}(R) \tag{26}$$

where $E_{sp}(R)$ is given by (10), (11).

The initial load u_0 has no effect on asymptotic probabilities; it is taken equal to zero here to simplify calculations. The rest of this section consists of the proof of Proposition 3. We fix a sequence of codes $\{C_i\}$ satisfying the hypothesis. We break the proof into several steps.

A. Converse Coding Theorem

For the proof, we require a converse coding theorem. The appropriate form of the converse is given by [1, Theorem 4] which states that the capacity of a $\cdot/M/1$ channel with service rate μ and departure rate λ is given by $C(\mu, \lambda) = \lambda \log(\mu/\lambda)$. To interpret this result for our purposes, fix any $\epsilon > 0$. Consider an (n, M, T) code with rate R. Let λ be the average departure rate of customers, observed over the interval $[0, T]$. The theorem states that if $R > C(\mu, \lambda) + \epsilon$, then the probability of ML decoding error P_e based on observation of the time window $[0, T]$ is bounded away from zero by a function $\alpha(\epsilon, \mu, \lambda)$ which is independent of the code length T. For the types of codes under consideration (window codes), this statement remains true even if the decoder is allowed to observe the channel output after time T, because as we noted earlier, such observations do not provide any relevant information to an ML decoder. For a uniform lower bound on the probability of error, we define

$$\alpha(\epsilon, \mu) = \min_{0 \leq \lambda \leq \mu} \alpha(\epsilon, \mu, \lambda).$$

This function is strictly positive for any $\epsilon > 0$.

An initial use of this converse will be to dispose of a simple special case of Proposition 3 in which $\{\mathcal{C}_i\}$ operate above channel capacity. To be precise, let λ_i denote the departure rate of code $\mathcal{C}_i$ under service rate μ. Suppose the code sequence is such that, for some $\epsilon > 0$, $R_i > C(\mu, \lambda_i) + \epsilon$ for infinitely many i. Then, $\limsup_i P_e(\mathcal{C}_i) \geq \alpha(\epsilon, \mu)$ is bounded away from zero, and $\limsup_i [-T_i^{-1} \log P_e(\mathcal{C}_i)] = 0$. In this case, (26) holds trivially since $E_{sp}(R)$ is a nonnegative function by definition. This completes the proof for such sequences of codes.

The rest of the proof is devoted to the complementary case, namely, the case for which the code sequence $\{\mathcal{C}_i\}$ is such that $R_i \leq C(\mu, \lambda_i)$ for all i sufficiently large. To develop the desired results, we shall consider a generic code $\mathcal{C}$ with parameters (n, M, T) and rate R. We let λ denote the departure rate of $\mathcal{C}$ under service rate μ and assume that $R \leq C(\mu, \lambda)$. The asymptotic nature of (26) allows us to assume that T is as large as we please as we go through the proof.

B. Departure Rates of Codewords

Here, we consider the mean departure rate λ' for an individual codeword as a function of the service rate μ', which we regard as variable. For this it will be convenient to think of the work brought to the queue by each customer as exponentially distributed with mean 1 s. The service time distribution under service rate μ' (customers per second) will then be exponential with mean $1/\mu'$ seconds, as desired.

Fix a codeword x^n in the code $\mathcal{C}$. Let n' denote the expected number of customer departures in the time interval $[0, T]$ under service rate μ'. The departure rate (under service rate μ') is defined as $\lambda' = n'/T$. Let us note the obvious fact that $\lambda' \leq \mu'$, as the departure rate cannot exceed the service rate.

Let W' denote the average work done by a rate-μ' server on the codeword x^n in the time interval $[0, T]$. Clearly, $n' + 1 > W' \geq n'$. Dividing by T, we have $\lambda' + (1/T) > W'/T \geq \lambda'$.

Consider another server at rate $\mu'' > \mu'$. Let n'', λ'', W'' be the corresponding quantities for the rate-μ'' server. Clearly, $n' \leq n''$ and, equivalently, $\lambda' \leq \lambda''$. This is because for any

given realization of the service times of customers brought by x^n, the faster server will serve at least as many customers as the slower one.

For a converse inequality, we begin with $W'/\mu' \geq W''/\mu''$. This can be seen by interpreting W'/μ' as the length of time that the server with rate μ' is busy in the interval $[0, T]$. Since the faster server is never busy when the slower one is idle for any realization of the service times, the inequality follows. Combining this with the previous inequality, we have $\lambda' + (1/T) > W'/T \geq (\mu'/\mu'')\lambda''$. In particular, if $\mu'' = \mu' + \Delta$, we obtain

$$0 \leq \lambda'' - \lambda' < (\lambda''/\mu'')\Delta + (1/T) \leq \Delta + (1/T). \quad (27)$$

This fact will be used in the sequel.

C. Selecting a Subcode

We now seek to find a subcode $\mathcal{C}_f$ of $\mathcal{C}$ such that $\mathcal{C}_f$ has essentially the same rate R but for which each codeword has approximately the same departure rate λ' as a function of μ', for $0 \leq \mu' \leq \mu$. This step of the proof is the counterpart of finding a fixed-composition subcode in the case of discrete channels.

The service rate–departure rate pairs of interest here lie in the triangle $0 \leq \mu' \leq \mu$, $0 \leq \lambda' \leq \mu'$. We visualize this triangle in a Cartesian coordinate system with μ' marking the horizontal axis, and λ' the vertical. We place a grid on the triangle by subdividing its horizontal and vertical sides into K intervals of length $\Delta = \mu/K$, so that the endpoints of the intervals are $k\Delta$, $k = 0, \ldots, K$. By *bin* (k, j) of this grid, we shall mean the set of (μ', λ') pairs such that $\mu' = k\Delta$ and $(j-1)\Delta < \lambda' \leq j\Delta$, $j \leq k$.

We measure the departure rate of each codeword in $\mathcal{C}$ under service rate $K\Delta = \mu$, and assign a codeword with departure rate λ' to bin (K, j) if $(j-1)\Delta < \lambda' \leq j\Delta$. After assigning all codewords into the appropriate bins in the Kth column of the grid, we pick the bin in that column with the largest number of codewords. We denote this bin by $(K, j(K))$. Clearly, bin $(K, j(K))$ contains at least M/K codewords.

We discard all codewords other than the ones in bin $(K, j(K))$ and repeat the above procedure this time with service rate $(K-1)\Delta$. We obtain another bin $(K-1, j(K-1))$ with at least M/K^2 codewords. This procedure is continued for each k from K to 0 and in the end we obtain a code $\mathcal{C}_f$ with at least M/K^K codewords. (For $k = 0$ all codewords go into bin $(0, 0)$.) We let $\mathcal{C}_f$ denote the code obtained at the end of this procedure.

$\mathcal{C}_f$ is an (n, M_f, T) code with rate R_f in the interval

$$R \geq R_f \geq R - (K/T) \log(K).$$

To be specific, for the rest of the proof we take $K = \sqrt{T}$, so that we may write $R_f = R - o_T(1)$. This choice allows the grid resolution $\Delta = \mu/\sqrt{T}$ to become finer and finer as T increases and at the same time ensures that $\mathcal{C}_f$ has rate close to R.

In the rest of the proof we focus on obtaining a lower bound on $P_e(\mathcal{C}_f)$, the probability of ML decoding error for $\mathcal{C}_f$ (regarded not as a subcode but as a code in its own right) on the channel with rate μ. Any lower bound on $P_e(\mathcal{C}_f)$ gives a lower bound on $P_e(\mathcal{C})$ through the inequality

$$P_e(\mathcal{C}) \geq (M_f/M)P_e(\mathcal{C}_f). \quad (28)$$

Since (M_f/M) does not alter the exponential order of the bounds as a function of T, it will be sufficient to lower-bound $P_e(\mathcal{C}_f)$.

For brevity, we put $\mu_k = k\Delta$. We define λ_k as the average departure rate for $\mathcal{C}_f$ at service rate μ_k; this is computed by taking the average of the departure rates of all codewords in $\mathcal{C}_f$. Clearly, the rate pair (μ_k, λ_k) falls in the same bin $(k, j(k))$ as individual codewords. Finally, we let $C_k = C(\mu_k, \lambda_k)$.

D. Selecting a Critical Channel

Fix ϵ such that $0 < \epsilon < R_f/2$. Let k_0 be the largest integer k such that

$$R_f > C_k + \epsilon. \tag{29}$$

There exists such an integer because at $k = 0$, we have $R_f > C_0 + \epsilon = \epsilon$. Also, $k_0 < K$ since $R_f \leq R$ and

$$R \leq C(\mu, \lambda) = C(\mu_K, \lambda_K)$$

by assumption. We define $\mu_c = \mu_{k_0}$, $\lambda_c = \lambda_{k_0}$, and $C_c = C_{k_0}$. The service rate μ_c defines the "critical channel."

We need to establish for later use two technical points, namely, that

$$R = C_c + \epsilon + o_T(1) \tag{30}$$

and that λ_c is bounded away from zero for T large enough. We begin with the latter.

Clearly, $C(\mu, \lambda_{k_0+1}) \geq C_{k_0+1} \geq R_f - \epsilon > R_f/2$. Let $\lambda_{\min}(R')$ denote the smallest number λ' such that $C(\mu, \lambda') = R'$. Then, $\lambda_{k_0+1} > \lambda_{\min}(R_f/2)$ and λ_{k_0+1} is bounded away from zero by a quantity that depends only on R_f. On the other hand, since $\mu_{k_0+1} - \mu_c = \Delta$, we may use (27) to write

$$\lambda_c \geq \lambda_{k_0+1} - \Delta - (1/T) > \lambda_{\min}(R_f/2) - o_T(1)$$

which bounds λ_c away from zero for T large enough.

To establish (30), it suffices to show that

$$C_{k_0+1} - C_c = o_T(1) \tag{31}$$

as we have by definition $C_c + \epsilon < R_f \leq C_{k_0+1} + \epsilon$ and $R_f = R - o_T(1)$ by construction. Since $\mu_{k_0+1} - \mu_c = \Delta$ and $\lambda_{k_0+1} - \lambda_c \leq \Delta + (1/T)$, the distance between the points $(\mu_{k_0+1}, \lambda_{k_0+1})$ and (μ_c, λ_c) is $o_T(1)$, and both points are bounded away from the origin for T large enough. Since the function $C(\mu, \lambda)$ is analytic away from the origin, (31) follows. This completes the proof of (30).

We shall also need the observation that, for each codeword in $\mathcal{C}_f$, the average number of departures n' for that codeword under service rate μ_c satisfies $|n'/T - \lambda_c| \leq \Delta$. This is because the average departure rate n'/T for any codeword and the departure rate λ_c for the entire code both lie in the same bin $(k_0, j(k_0))$ of size Δ.

E. Sphere-Packing Argument

The remaining part of the proof consists of the standard steps of the sphere-packing bound. We begin by computing the divergence between the transition probabilities of the actual channel with service rate μ and an imaginary alternate channel. Here, we take the alternate channel also as a $\cdot/M/1$ channel but with

a service rate μ_c as found in the previous subsection. We actually assume that the alternate channel gives service at rate μ_c to those customers who enter service in the interval $[0, T]$ and at rate μ to those who enter after time T. The reason for switching the service rate from μ_c to μ is to keep the divergence between the actual and alternate channels as small as possible. This point will be clear in the following development.

To be precise, the transition probability density function for the alternate channel is defined as

$$p'(y^n|x^n) = \prod_{i=1}^{n} e_{\mu_i}(y_i - w_i) \tag{32}$$

where $\mu_i = \mu_c$ if $y_1 + \cdots + y_{i-1} < T$ and $\mu_i = \mu$ otherwise.

For a given channel input x^n, we define the *divergence random variable* by

$$z(y^n) = z(x^n, y^n) = \log \frac{p'(y^n|x^n)}{p(y^n|x^n)} \tag{33}$$

where y^n is regarded as a random vector with distribution $p'(y^n|x^n)$. It is easy to see that we can write

$$z(y^n) = \sum_{i=1}^{n} z_i(y^i) \tag{34}$$

where

$$z_i(y^i) = z_i(x^i, y^i) = \log \frac{p'(y_i|x^i, y^{i-1})}{p(y_i|x^i, y^{i-1})}. \tag{35}$$

The increment $z_i(y^i)$ is conditionally independent of previous increments $z_j(y^j)$, $1 \leq j < i$, given y^{i-1}. The increments become zero after the service rate is switched from μ_c to μ. Using these observations, we calculate the expectation of the increment $z_i(y^i)$ as

$$\mathbf{E}'[z_i(y^i)] = \int_{y^n} p'(y^n|x^n) \log \frac{p'(y_i|x^i, y^{i-1})}{p(y_i|x^i, y^{i-1})} \, dy^n$$
$$= D(\mu_c\|\mu)P'\left[\sum_{j=1}^{i-1} y_j < T \Big| x^n\right] \tag{36}$$

where we have defined

$$D(\mu_c\|\mu) = -1 + \frac{\mu}{\mu_c} + \log \frac{\mu_c}{\mu} \tag{37}$$

and

$$P'\left[\sum_{j=1}^{i-1} y_j < T \Big| x^n\right] = \int_{\sum_{j=1}^{i-1} y_j < T} p'(y^{i-1}|x^{i-1}) \, dy^{i-1} \tag{38}$$

is the probability that customer i enters service before time T and receives service at rate μ_c, conditional on x^n being transmitted.

The expectation of the sum $z(y^n)$ is then given by

$$\mathbf{E}'[z(y^n)] = D(\mu_c\|\mu) \sum_{i=1}^{n} P'\left[\sum_{j=1}^{i-1} y_j < T \Big| x^n\right]. \tag{39}$$

It is straightforward to see that the above sum equals the expected number of customers who enter service before time T on the alternate channel, given that x^n is transmitted and the

service rate is μ_c. Denoting this number by $n' = n'(x^n)$, we have

$$\mathbf{E}'[z(y^n)] = n'D(\mu_c\|\mu). \tag{40}$$

The variance of $z(y^n)$ is also needed in the following. This variance turns out to be the sum of the variances of the increments $z_i(y^i)$ because of the conditional independence property mentioned above. By a similar calculation, we obtain that

$$\mathrm{var}[z(y^n)] = n'(1 - \mu/\mu_c)^2. \tag{41}$$

Now, fix $\delta > 0$ and define

$$A_{x^n} = \{y^n\colon |z(y^n) - n'D(\mu_c\|\mu)| \le n'\delta\}. \tag{42}$$

By Chebyshev's inequality, we have the bound

$$P'(A_{x^n}) \ge 1 - \frac{(1 - \mu/\mu_c)^2}{n'\delta^2}. \tag{43}$$

Following these preliminary calculations, we are now ready to compute the sphere-packing bound. Let D_m, $1 \le m \le M_f$, denote the ML decoding regions for $\mathcal{C}_f$ with respect to the actual channel that has service rate μ. Put $n'_m = n'(x^n(m))$ and $A_m = A_{x^n(m)}$. Then, the probability of decoding error for the mth codeword is bounded as follows:

$$\begin{aligned} P_{e,m} &= \int_{D_m^c} p(y^n|x^n(m))\,dy^n \\ &\ge \int_{D_m^c \cap A_m} p(y^n|x^n(m))\,dy^n \\ &\ge e^{-n'_m[D(\mu_c\|\mu)+\delta]} \int_{D_m^c \cap A_m} p'(y^n|x^n(m))\,dy^n \\ &\ge e^{-n'_m[D(\mu_c\|\mu)+\delta]} \left[\int_{A_m} p'(y^n|x^n(m))\,dy^n \right. \\ &\qquad\qquad \left. - \int_{D_m} p'(y^n|x^n(m))\,dy^n\right] \\ &\ge e^{-n'_m[D(\mu_c\|\mu)+\delta]} \left[1 - \frac{(1-\mu/\mu_c)^2}{n'_m\delta^2} - (1-P'_{e,m})\right] \end{aligned}$$

where $P'_{e,m}$ is the probability of decoding error on the alternate channel using decoding region D_m.

Since for each codeword of $\mathcal{C}_f$ we have

$$T(\lambda_c - \Delta) \le n'_m \le T(\lambda_c + \Delta)$$

we can write

$$P_{e,m} \ge e^{-T\lambda_c[D(\mu_c\|\mu)+\delta]-o(T)} \left[P'_{e,m} - \frac{(1-\mu/\mu_c)^2}{T(\lambda_c - \Delta)\delta^2}\right] \tag{44}$$

where $o(T) = T\Delta[D(\mu_c\|\mu) + \delta]$ and we assume T is large enough so that $\lambda_c - \Delta > 0$ (recall that λ_c is bounded away from zero for T large). Averaging over all codewords, we obtain

$$P_e(\mathcal{C}_f) \ge e^{-T\lambda_c[D(\mu_c\|\mu)+\delta]-o(T)} \left[P'_e(\mathcal{C}_f) - \frac{(1-\mu/\mu_c)^2}{T(\lambda_c - \Delta)\delta^2}\right] \tag{45}$$

where $P'_e(\mathcal{C}_f)$ denotes the probability of decoding error on the alternate channel.

Let $\delta = T^{-1/3}$. Then, $1/[T(\lambda_c - \Delta)\delta^2]$ is $o_T(1)$ since $\lambda_c > \lambda_{\min}(R_f/2) - o_T(1)$. Also note that $P'_e(\mathcal{C}_f) > \alpha(\epsilon, \mu_c)$. So, we may write

$$P_e(\mathcal{C}_f) \ge e^{-T\lambda_c D(\mu_c\|\mu)-o(T)}[\alpha(\epsilon, \mu_c) - o_T(1)] \tag{46}$$

where the $T\delta$ term in the exponent in (45) has been absorbed into the $o(T)$ term.

The choice of μ_c and λ_c in the bound has been dictated by the procedure of selecting $\mathcal{C}_f$. These parameters depend, among other things, on the internal composition of the code. This makes the computation of the term $\lambda_c D(\mu_c\|\mu)$ practically impossible. However, the lower bound (46) would remain valid if $\lambda_c D(\mu_c\|\mu)$ were replaced by a more readily computable upper bound on it. Such an upper bound is given by

$$\max\{\lambda_c D(\mu_c\|\mu)\colon 0 \le \mu_c \le \mu,\ 0 \le \lambda_c \le \mu_c\} \tag{47}$$

such that

$$R' = \lambda_c \log(\mu_c/\lambda_c) \tag{48}$$

where we put $R' = R - \epsilon - o_T(1)$. The constraint (48) corresponds to the condition (30) which holds true due to the way the critical channel has been selected.

To carry out this optimization, we find it convenient to eliminate μ_c in favor of a new parameter $\rho \ge 0$ by setting $\mu_c = \mu/(1 + \rho)$. The desired upper bound now becomes

$$\sup\{\lambda_c[\rho - \log(1+\rho)]\colon \rho \ge 0,\ 0 \le \lambda_c \le \mu/(1+\rho)\} \tag{49}$$

such that

$$R' = \lambda_c \log \frac{\mu}{(1+\rho)\lambda_c}. \tag{50}$$

We find that at a stationary point of the Lagrangian for this optimization problem, the following relation must hold:

$$\lambda_c = \frac{\mu}{(1+\rho)^{(1+\rho)/\rho}} \tag{51}$$

which is a necessary condition for optimality. Substituting this into the objective function, the problem is reduced to a one-dimensional search

$$\sup_{\rho \ge 0} \left[\frac{\rho\mu}{(1+\rho)^{(1+\rho)/\rho}} - \rho R'\right]. \tag{52}$$

Setting the derivative with respect to ρ equal to zero, we obtain that the supremum is given by the parametric equations (10), (11), and equals $E_{sp}(R')$.

Substituting this upper bound into (46), we obtain

$$-\frac{1}{T} \log P_e(\mathcal{C}_f) \le E_{sp}[R - \epsilon - o_T(1)] + o_T(1). \tag{53}$$

Applying this result to the sequence of codes $\{\mathcal{C}_i\}$, we have

$$\limsup_{i \to \infty} -\frac{1}{T_i} \log P_e(\mathcal{C}_i) \le E_{sp}(R - \epsilon). \tag{54}$$

The bound (26) is obtained by taking the infimum of $E_{sp}(R-\epsilon)$ over $\epsilon > 0$. This completes the proof of Proposition 3.
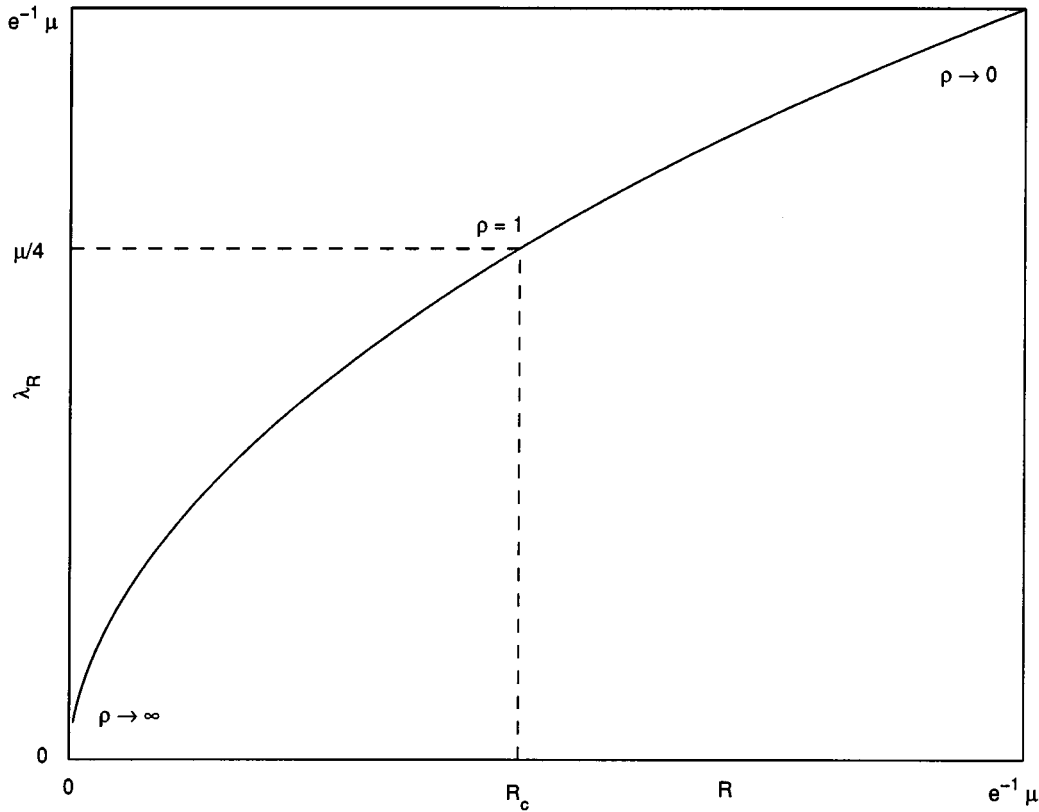
Fig. 2.   The function λ_R versus R.

IV. DISCUSSION OF RESULTS

We conclude the paper by discussing some issues related to the above results.

A. Coding Implications

The calculations that lead to the random-coding and sphere-packing exponents provide some insight on composition of optimal codes and dominant error events. To discuss this, let λ_R be the function of R, $0 \leq R \leq C = e^{-1}\mu$, defined parametrically by

$$\lambda_R = \frac{\mu}{(1+\rho)^{(1+\rho)/\rho}} \tag{55}$$

$$R = \frac{\mu}{(1+\rho)^{(1+\rho)/\rho}} \log(1+\rho)^{1/\rho}. \tag{56}$$

This function is depicted in Fig. 2.

The function λ_R appears in Section II as the arrival rate that achieves the random-coding exponent $E_r(R)$ for $R_c \leq R \leq C$. It also appears in Section III as the critical departure rate that yields the sphere-packing exponent $E_{sp}(R)$ for $0 \leq R \leq C$. These results together imply that for rates $R_c \leq R \leq C$, λ_R is the optimal arrival–departure rate; i.e., any deviation from λ_R results in an error exponent inferior to $E(R)$. For $0 \leq R < R_c$, the results in this paper provide no basis to claim that λ_R is optimal in this sense.

Notice that the ratio $\lambda_R/\mu = 1/(1+\rho_R)^{(1+\rho_R)/\rho_R}$, which represents the average fraction of time the server is busy, ranges from $1/4$ to e^{-1} as R goes from R_c to C. Also notice that

for a code operating at arrival rate λ_R, the fraction of available channel capacity that is being used is given by $R/C(\mu, \lambda_R) = 1/(1+\rho_R)$.

Another conclusion we can draw is the following. Consider a code with rate $R_c \leq R \leq C$ and arrival rate λ_R. The derivation of the sphere-packing bound shows that, for any such code, a decoding error is most likely to occur when the empirical mean of the service time realization approximately equals $1/\mu_c = (1+\rho_R)/\mu$. Thus, when a decoding error occurs for a code optimized for its arrival rate, the server should be expected to be busy $(1+\rho_R)$ times more than usual.

B. Zero-Rate Exponent

The major open problem that this work brings forth is the determination of $E(R)$ for rates $0 \leq R < R_c$. It is, therefore, of particular interest to find bounds on the zero-rate exponent $E(0)$ defined by (6), as this would give an indication on how $E(R)$ behaves at low rates. At present, the best bound that we are aware of is $\mu/4 \leq E(0) \leq \mu$, which is obtained in this paper. The source of difficulty here appears to be the memory in the channel which makes the evaluation of pairwise error probabilities intractable. Some further observations related to this problem are as follows.

We first observe that for any (n, M, T) code, there is a probability $\geq e^{-\mu T}$ that no customer departure occurs in the interval $[0, T]$. This gives the trivial upper bound $E(R) \leq \mu$ on the reliability exponent, which holds even if there is instantaneous, noiseless feedback from channel output to the transmitter. This upper bound coincides with $E_{sp}(0) = \mu$.

We have defined $E(0)$ as the supremum of $E(R)$ over $R > 0$, thus restricting attention to the best attainable performance of code sequences for which the number of codewords grows exponentially with T. If we consider code sequences for which the number of codewords is fixed or grows sufficiently slowly with T, then the exponent $E_{sp}(0) = \mu$ is achievable, as we demonstrate next.

Let $\mathcal{C}_M$ be an $(M - 1, M, T)$ code such that the mth codeword brings $m - 1$ customers at time 0 and $M - m$ customers at time T, $1 \leq m \leq M$. Consider a decoder for this code that decides that message m was sent if it observes $m - 1$ customer departures in $[0, T]$. Then, the probability of decoding error for the mth codeword is given by

$$P_{e,m} = P(s_1 + \cdots + s_{m-1} > T) \qquad (57)$$

where $\{s_i\}$ are the service times, which are independent and identically distributed (i.i.d.) exponential with mean $1/\mu$. Clearly, $P_{e,m}$ is monotonically increasing in m. So, the average probability of error for the code satisfies $P_e(\mathcal{C}_M) \leq P_{e,M}$.

A straightforward application of the Chernoff bound [5, p. 126] yields

$$P_{e,M} \leq e^{-rT}[\mu/(\mu - r)]^{(M-1)} \qquad (58)$$

for any $0 \leq r < \mu$. Thus,

$$-\frac{1}{T} \log P_e(\mathcal{C}_M) \geq r - [(M-1)/T] \log[\mu/(\mu - r)]. \quad (59)$$

It follows that if $M/T = o_T(1)$, then asymptotically as $T \rightarrow \infty$, the right-hand side of (59) can be made as close to μ as desired. Thus, the exponent $E_{sp}(0) = \mu$ can be achieved arbitrarily closely by such codes.

Next, we show that it is possible to have M grow linearly with T while still maintaining an error exponent better than $E_r(0) = \mu/4$. For this, let $\beta = (M - 1)/(\mu T)$ and $r = \theta\mu$ to write the right-hand side of (59) as

$$\mu[\theta + \beta \log(1 - \theta)]. \qquad (60)$$

A simple analysis shows that for any $0 \leq \beta < 1$, (60) is maximized over $0 \leq \theta < 1$ by $\theta = 1 - \beta$, yielding

$$-\frac{1}{T} \log P_e(\mathcal{C}_M) \geq \mu[1 - \beta + \beta \log(\beta)]. \qquad (61)$$

Thus, for β less than about 0.382, M can grow as $\beta\mu T$ with a resultant error exponent greater than $E_r(0)$.

This shows that the random-coding argument of Section II fails to produce the best achievable exponent for the range of code parameters just considered. It would be of interest, therefore, to study possible extensions of the above code construction to nonzero rates by allowing multiple but discrete arrival points, as in the models studied in [3].

As a final remark, we note that for the IPC considered in Section I, the best attainable error exponent for codes with $M = 3$ codewords is $\mu/2$, whereas for the ETC the same number is μ. So, at least for codes with fixed number of codewords, a difference emerges between the relative error-correcting capabilities of the two channels. Whether this is true at any nonzero rate remains open.

REFERENCES

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

[2] E. Arikan, "An upper bound on the cutoff rate of sequential decoding," *IEEE Trans. Inform. Theory*, vol. 34, pp. 55–63, Jan. 1988.

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

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

[5] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.

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

[7] A. Lapidoth, "On the reliability function of the ideal Poisson channel with noiseless feedback," *IEEE Trans. Inform. Theory*, vol. 39, pp. 491–503, Mar. 1993.

[8] R. Sundaresan, "Point process channel and the capacity of the exponential server channel," in *Proc. Int. Symp. Information Theory*, Sorrento, Italy, June 25–30, 2000.

[9] R. Sundaresan and S. Verdú, "Robust decoding for timing channels," *IEEE Trans. Inform. Theory*, vol. 46, pp. 405–419, Mar. 2000.

[10] ——, "Sequential decoding for the exponential server timing channel," *IEEE Trans. Inform. Theory*, vol. 46, pp. 705–709, Mar. 2000.

[11] A. D. Wyner, "Capacity and error exponent for the direct detection photon channel—Part I," *IEEE Trans. Inform. Theory*, vol. 34, pp. 1449–1461, Nov. 1988.

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