CCIT Report # 399 October 2002

Sequential signaling under peak-power constraint in the Poisson regime

Amir Dabbagh * Shraga I. Bross †

October 15, 2002

Abstract

Sequential signaling over the single-user and two-user multiple-access Poisson channel subject to peak-power constraint is considered. Specifically, we analyze first the performance of a sequential signaling scheme similar to the one proposed by Schalkwijk and Barron in [1], for the AWGN channel, in the Poisson regime. Assuming that information is transmitted in blocks, and the channel is corrupted by additive constant dark current, the error exponent for the single-user case in the low power regime is determined. Then an extension of the Schalkwijk-Barron scheme for the two-user case is suggested and in this case an attainable exponent in the low power regime is established.

Index Terms – Poisson channel, feedback error exponent, sequential signaling.

*Department of Electrical Engineering Technion, Haifa 32000, Israel email:dabbagh@techunix.technion.ac.il
†Department of Electrical Engineering Technion, Haifa 32000, Israel email:shraga@ee.technion.ac.il

I. Introduction

A coding scheme for the infinite bandwidth additive white Gaussian noise (AWGN) channel with noiseless feedback has been considered in [1]. Specifically, the authors describe a peak-power constrained sequential signaling scheme that improves the one-way coding error exponent using a causal noiseless feedback link. Whereas the reliability function for the infinite bandwidth AWGN channel is expressed as [2, 3]

$$E(R) = \begin{cases} C/2 - R, & 0 \leq R \leq C/4 \\ (\sqrt{C} - \sqrt{R})^2, & C/4 \leq R \leq C \end{cases} \tag{1}$$

the attainable error exponent for the signaling scheme of [1] in the presence of feedback is larger and given by

$$E_f(\kappa, R) = \left(\sqrt{\kappa C - R} + \sqrt{C - R}\right)^2 , \tag{2}$$

where $\kappa \triangleq \hat{P}/P_{av}$ is the permissible peak-to-average power ratio.

The Schalkwijk-Barron scheme consists of two modes. First an orthogonal M-ary transmission is performed for a fixed duration T. By the termination of this mode the decoder decides on the most likely message out of the M hypotheses, denoted as $\hat{m}$. The second mode executes sequential binary signaling wherein the two antipodal signals $\hat{m}$ and $-\hat{m}$ are used in order to confirm or reject, respectively, the decoded result $\hat{m}$. Since during this mode their decoding scheme uses Viterbi's sequential decision feedback [26], the time required for the receiver to make a decision varies from one transmission to another and admits random duration T_2. It is well known that feedback does not increase the capacity of a discrete memoryless channel as shown by Shannon [5] and later generalized to continuous-time memoryless channels by Kadota *et al.* [6]. Thus, the Schalkwijk-Barron result as well as [7, 10] demonstrate that whereas unbounded random transmission time codes cannot enlarge the capacity they can improve significantly the attainable error exponent.

The model for the two-user continuous-time Poisson multiple access channel (MAC) studied here is described as follows. Two independent users generate inputs $\lambda_{m_i}(t)$, $i = 1, 2$, $0 \leq t \leq \infty$, that determine the rates of two corresponding doubly stochastic Poisson processes $d_i(t)$. Specifically, $d_i(t)$ corresponds to the number of counts registered by a direct detection device in the interval $[0, t]$, in reaction to the input $\lambda_{m_i}(t)$. The observation is

$$\nu(t) = \sum_{i=1}^{2} d_i(t) + D(t) ,$$

which is also a Poisson process with instantaneous rate $\lambda(t) = \lambda_0 + \sum_{i=1}^{2} \lambda_{m_i}(t)$. The dark current represented by $D(t)$ is a homogeneous Poisson process of rate λ_0. Thus, conditional on the input $\{\lambda(t)\}$, the signal $\{\nu(t)\}$ is a counting process with independent increments and

$$\Pr\left(\nu(t) - \nu(s) = k | \{\lambda(t)\}\right) = e^{-\int_s^t \lambda(\tau) d\tau} \frac{\left(\int_s^t \lambda(\tau) d\tau\right)^k}{k!} , \ t \geq s , \ k \in \mathbb{Z}^+ .$$

Here $\mathbb{Z}^+$ denotes the set of nonnegative integers. It is further assumed that a causal feedback link informs both encoders at every time t of the channel output $\nu(t)$ at all times prior to t. Our main focus herein will be the case where $\lambda_0 \gg 0$ which henceforth will be referred as the non-ideal Poisson MAC. We shall consider as well the single-user Poisson channel with noiseless feedback, the model of which follows straightforward from the above two-user setting.

The capacity of the single-user Poisson channel was found by Kabanov [8] and Davis [9]. The effect of feedback on the capacity of the single-user Poisson channel has been studied by Frey [12] who showed that as long as the dark current is deterministic, capacity is not increased by feedback, whereas feedback may improve the capacity in the case of a random dark current. The reliability function (i.e. the exponential behavior of the probability of error for the best code, as the coding delay increases while the transmission rate is held fixed) of the one-way single-user Poisson channel was determined by Wyner in his seminal contribution [4]. In addition, the impact that feedback has on the reliability of the single-user Poisson channel when $\lambda_0 = 0$ was found by Lapidoth in [10]. The capacity of the Poisson MAC has been obtained by Lapidoth and Shamai in [11], while the reliability function for this channel was found in [13].

In contrast to the single-user channel, feedback can enlarge the capacity region of the discrete-time memoryless MAC [15, 16]. In particular, the feedback capacity of the discrete-time two-user AWGN MAC was determined by Ozarow in [17].

Our interest here is in random transmission time schemes, of the Schalkwijk-Barron type, which satisfy the peak and average-power constraints $(A, \sigma A)$ while improving the one-way coding error exponent of the (single-user and two-user) Poisson channel when $\lambda_0 \neq 0$. Specifically, we shall consider coding schemes which meet the peak and average-power constraints

$$\lambda_{m_i}(t) \leq A \quad , \quad \frac{1}{\overline{T}_t}\int_0^{\overline{T}_t} \lambda_{m_i}(t)dt \leq \sigma A \quad , \quad i = 1,2 \quad 0 \leq t \leq T_t \ , \tag{3}$$

where $\overline{T}_t$ denotes the mean transmission time. Except for these input constraints, it is assumed throughout that the channel is unlimited in bandwidth.

The paper is organized as follows. In section II we propose a random transmission-time code for the single-user case, compute the attainable exponent and establish its optimality subject to the assumption that information is transmitted in blocks. Akin to the Schalkwijk-Barron result (2) wherein $4(C-R)$ is the optimal error exponent when $\kappa = 1$, it is shown that in the Poisson low power regime $4(C-R)$ is the optimal error exponent, yet it is achievable with $\kappa \geq 2$.

In section III a two-user random transmission-time code, based on a sequential procedure for multi-hypothesis testing, is considered. Analyzing the performance of this scheme a lower bound for the two-user attainable feedback exponent is established. Our analysis gives rise to a lower bound on $E_f(R_{sum})$ in the low power regime,

$$E_f(R_{sum}) \geq 2(C_{sum} - R_{sum}) \ ,$$

where R_{sum} and C_{sum} are the two-user rate-sum and rate-sum capacity respectively. Again, this exponent is achievable with communication cost of $\kappa \geq 2$.

II. A single-user random transmission-time code

We start by recalling Wyner's construction of a family of exponentially optimal one-way single-user codes [4]. To this end define a code with parameters (M, T, σ, P_e) by

- A set of waveforms $\{\lambda_m(t)\}, 1 \leq m \leq M, 0 \leq t \leq T$, which satisfy the peak and the average power constraints,

$$0 \leq \lambda(t) \leq A \quad , \quad \frac{1}{T}\int_0^T \lambda_m(t)dt \leq \sigma A \quad , \quad 0 < \sigma \leq 1 \; .$$

- A decoder mapping $D: \; \nu_0^T \to \{1, 2, \ldots, M\}$ such that its average probability of error is given by

$$P_e = \frac{1}{M}\sum_{m=1}^{M} \Pr\left\{D(\nu_0^T) \neq m | \lambda_m(\cdot)\right\} \; .$$

The code is defined as follows: Let $0 < q \leq \sigma < 1$ be fixed. For any positive integer l such that lq is an integer, let $\mathcal{A}^{(l,q)}$ denote the binary matrix of dimension $l \times \binom{l}{lq}$, the columns of which are the $\binom{l}{lq}$ distinct binary l-vectors with exactly lq ones. Subdivide the interval $I_T = [0, T]$ into $\binom{l}{lq}$ disjoint half open, closed on the right, equi-length subintervals. Set for $m \in \{1, \ldots, M\}$ and t in subinterval j of I_T

$$\lambda_m(t) = A\left(\mathcal{A}^{(M,q)}\right)_{m,j} \quad , \quad j = 1, 2, \ldots, \binom{l}{lq} \; . \tag{4}$$

The decoder mapping D is defined as follows: For $1 \leq m \leq M$ let $v_m([0, T]) = \{t \in [0, T] : \lambda_m(t) = A\}$. The decoder observes the counting process $\{\nu(t)\}$ and computes,

$$\psi_m = \int_{v_m([0,T])} d\nu(t) = \text{number of arrivals in } v_m([0, T]) \; , \; 1 \leq m \leq M \; .$$

Then $D(\nu_0^T) = m^*$ if,

$$\begin{array}{ll} \psi_m < \psi_{m^*} \; , & 1 \leq m < m^* \\ \psi_m \leq \psi_{m^*} \; , & m^* < m \leq M \end{array} \tag{5}$$

Wyner proved that for this family of codes with parameter $M = e^{RT}$, the error probability decays according to $P_e = \exp\{-E(R)T + o(T)\}$, $T \to \infty$ and the error exponent is given by

$$E(R) = \begin{cases} \max_{\substack{0 \leq \rho \leq 1 \\ 0 \leq q \leq \sigma}} A\left[q + s - s(1 + \tau q)^{1+\rho}\right] - \rho R & R_c \leq R \leq C \\ Aq^\star(1 - q^\star)\left(\sqrt{1+s} - \sqrt{s}\right)^2 - R & 0 \leq R \leq R_c \end{cases}$$

Here $\tau = \tau(\rho) \triangleq \left(1 + \frac{1}{s}\right)^{\frac{1}{1+\rho}} - 1$, $s \triangleq \lambda_0/A$, $q^\star = \min\{\sigma, 1/2\}$, while the channel capacity and critical rate are given by

$$\begin{aligned} C &= A\left[q^*(1+s)\ln(1+s) + (1-q^*)s\ln s - (q^*+s)\ln(q^*+s)\right] \\ q^* &= \min\left\{\sigma, \frac{(1+s)^{1+s}}{s^s e} - s\right\} \\ R_c &= As\left\{(q^\star\tau(1)+1)\frac{q^\star}{2}\sqrt{1+\frac{1}{s}}\ln\left(1+\frac{1}{s}\right) - (q^\star\tau(1)+1)^2\ln\left(q^\star\tau(1)+1\right)\right\} \; . \end{aligned} \tag{6}$$

In the range $R_c \leq R \leq C$ the pair $(R, E(R))$ is expressed parametrically in the variable $\rho \in [0,1]$ via (see [4, equation 1.10])

$$\begin{aligned} R(\rho) &= As\left\{\frac{[1+q(\rho)\tau]^{\rho}}{1+\rho} q(\rho)\left(1+\frac{1}{s}\right)^{\frac{1}{1+\rho}} \ln\left(1+\frac{1}{s}\right) - [1+q(\rho)\tau]^{1+\rho}\ln[1+q(\rho)\tau]\right\} \\ q(\rho) &= \min\left\{\sigma\ ,\ \frac{1}{\tau}\left[\left(\frac{1}{s(1+\rho)\tau}\right)^{\frac{1}{\rho}} - 1\right]\right\}\ . \end{aligned}$$

Particularizing (6) to $s \to \infty$, i.e., the low power (high-noise) regime, Wyner showed that

$$C_{s\to\infty} = \frac{Aq^{\star}(1-q^{\star})}{2s} + O\Big(\frac{1}{s^2}\Big)\ , \tag{7}$$

while $E(R)_{s\to\infty}$ coincides with the "very noisy channel" error exponent (1).

We shall now proceed to exhibit a random transmission time signaling scheme with peak and average-power constraints $(A, qA), q \leq \sigma$. Similar to [1] the scheme has two modes. The first mode is an M-ary signaling mode wherein the sender uses Wyner's codebook in order to transmit a message $m \in \{1, \ldots, M\}$ over a duration T interval, where $M = e^{R_1 T}$ (the subscript 1 refers to the first transmission mode). Let the decoder perform as per Wyner's rule (5), then for rates $R_1 \to C$ and $s \to \infty$, we have from (1)

$$-\ln P_1(e) \approx T\left(\sqrt{C} - \sqrt{R_1}\right)^2 \quad , \quad C/4 \leq R_1 \leq C \tag{8}$$

substituting $R_1 = (1-\epsilon)C$ in (8) we have for $\epsilon > 0$ small,

$$-\ln P_1(e) \approx TC\epsilon^2/4 + O(\epsilon^3)\ . \tag{9}$$

We improve on this performance by continuing with a binary signaling mode. Given the decoder's maximum likelihood decision $\hat{m}$ (which is known to the sender via the feedback link) the encoder transmits sequentially, in N periods each of duration T_b (where the number of periods N is random), the waveform (assuming that time zero refers to the beginning of the first mode) $\lambda_b\left(t - \lfloor\frac{t-T}{T_b}\rfloor T_b - T\right)$ where

$$\lambda_b(t) = \begin{cases} \lambda_{\hat{m}}(t), & \text{if } \ \hat{m} = m \\ \lambda_{\hat{m}}^c(t), & \text{if } \ \hat{m} \neq m \end{cases}$$

Here $\lambda_m(t)$ is the Wyner codeword corresponding to message m, for transmission over the interval $[0, T_b]$, and $\lambda_m^c(t)$ denotes its complement codeword over this interval. Thus, we associate the *a posteriori* most probable message $\hat{m}$ with the waveform $\lambda_{\hat{m}}(t)$, while the alternative (i.e., any message $m' \in \{1, \ldots, M\} \setminus \hat{m}$) is associated with the waveform $\lambda_{\hat{m}}^c(t)$. If the decoder's decision, by the termination of the second transmission phase of random duration $T_2 = NT_b$, is upon $\lambda_{\hat{m}}(t)$, $\hat{m}$ is the final decision. If, however, $\lambda_{\hat{m}}^c(t)$ is decided upon, the encoder-decoder pair start over with a new M-ary signaling period, and so forth.

Following the formulation of Lapidoth in [10] a channel use is a pair of probability measures, P^0 and P^m, defined on a measurable space $(\Omega, \mathcal{F})$ on which the "input" process $\lambda(t)$, and the "output" process $\nu(t)$, are defined. The probability measure P^0 serves as a reference probability measure in the sense that $\nu(t)$ is a homogeneous Poisson process of intensity λ_0 with respect to its natural history $\mathcal{F}_t^\nu$ and the measure P^0. We shall use the notation that $\nu(t)$ admits the $(P^0, \mathcal{F}_t^\nu)$-intensity λ_0. It is further assumed that P^m is absolutely continuous with respect to P^0, i.e., $P^m \ll P^0$, and with respect to the measure P^m the output process $\nu(t)$ admits the predictable $(P^m, \mathcal{F}_t^\nu)$-intensity $\lambda_m(t)$.

Thus, a convenient description of P^m is expressed via its Radon-Nykodim derivative with respect to the reference probability P^0. Denote by P_t^v the restriction of P^v to $\mathcal{F}_t^\nu, v \in \{0, m\}$, and define the $(P^m, \mathcal{F}_t^\nu)$ martingale L_t by

$$L_t = \frac{dP_t^m}{dP_t^0} \ .$$

By [23] it follows that

$$L_t = \left[\prod_{n \geq 1} \frac{\lambda_{T_n}}{\lambda_0} \chi_{\{T_n \leq t\}} \right] \cdot \exp \left\{ \int_0^t [\lambda_0 - \lambda_m(s)] \, ds \right\} \ , \tag{10}$$

where T_n denotes the stopping time when the nth jump occurred, and $\chi_{\{G\}}$ denotes the indicator function of the set G.

As the processes $\lambda_m(t), m \in \{1, \ldots, M\}$ defined by Wyner's construction are all adapted to $\mathcal{F}_t^\nu$, they are $\mathcal{F}_t^\nu$ predictable. The existence of a measure P^m which induces a predictable $\mathcal{F}_t^\nu$-intensity given by $\lambda_m(t)$ follows now by [23].

To this end let the decoder observe $\nu_k = \{\nu(t) : t \in T_k = [T + (k-1)T_b, T + kT_b]\}$, and consider the sequence of independent identically distributed observations $\{\nu_k; k = 1, 2, \ldots\}$ where

$$\begin{aligned} H_0 &: \nu_k \sim P^{\hat{m}} \ , \ k = 1, 2, \ldots \\ H_1 &: \nu_k \sim P^{\hat{m}^c} \ , \ k = 1, 2, \ldots \end{aligned}$$

where $P^{\hat{m}}$ induces the predictable $\mathcal{F}_t^{\nu_k}$-intensity $\lambda_{\hat{m}}(t), t \in T_k$ while $P^{\hat{m}^c}$ induces the predictable $\mathcal{F}_t^{\nu_k}$-intensity $\lambda_{\hat{m}}^c(t), t \in T_k$. According to (10) a sequential decision rule can be defined via the Radon-Nykodim derivative

$$L_{T_k}(\lambda_{\hat{m}}, \lambda_{\hat{m}}^c) = \frac{dP_t^{\hat{m}}}{dP_t^{\hat{m}^c}} = \left[\prod_{n \geq 1} \frac{\lambda_{T_n}^{\hat{m}}}{\lambda_{T_n}^{\hat{m}^c}} \chi_{\{T_n \leq t\}} \right] \cdot \exp \left\{ \int_0^t [\lambda_{\hat{m}}^c(s) - \lambda_{\hat{m}}(s)] \, ds \right\} \ .$$

Consequently, the decoder maps each observation $\nu_k \mapsto L_{T_k} \in \mathbb{R}$. This is done by sampling the sufficient statistics $y_k = (n_{0k}, n_k)$, where n_{0k} denotes the number of arrivals which fall in $v_{\hat{m}}(T_k)$, and n_k denotes the total number of arrivals in T_k.

Let $\nu_k(T_b) \triangleq \nu(T + kT_b) - \nu(T + (k-1)T_b)$ and let

$$p_j(y_k) = \Pr(\psi_{m,k} = n_{0k}, \nu_k(T_b) = n_k | H = H_j) \ , \ j = 0, 1$$

be the conditional density of y_k given $H_j, j = 0, 1$. Assuming that $j = 0$, the Poisson process $\nu_k(T_b)$ is the sum of two independent Poisson processes with parameters $\Lambda_{00} = (1+s)qAT_b$ and $\Lambda_{10} = (1-q)sAT_b$, therefore $\nu_k(T_b)$ is a Poisson process with parameter $\Lambda_0 = \Lambda_{00} + \Lambda_{10} = (q+s)AT_b$. Furthermore, conditional on $\nu_k(T_b) = n_k$, $\psi_{m,k}$ has the binomial distribution

$$\Pr\{\psi_{m,k} = n_{0k} | \nu_k(T_b) = n_k, H_0\} = \binom{n_k}{n_{0k}} \pi_0^{n_{0k}} (1-\pi_0)^{n_k - n_{0k}} ,$$

where $\pi_0 \triangleq \Lambda_{00}/(\Lambda_{00} + \Lambda_{10}) = q(1+s)/(q+s)$. Thus the joint density $p_0(y_k)$ is

$$\begin{aligned} p_0(y_k) &= \Pr(\psi_{m,k} = n_{0k}, \nu_k(T_b) = n_k | H = H_0) \\ &= \frac{e^{-\Lambda_0}\Lambda_0^{n_k}}{n_k!} \binom{n_k}{n_{0k}} \pi_0^{n_{0k}} (1-\pi_0)^{n_k - n_{0k}} . \end{aligned} \tag{11}$$

Similarly, assuming that $j = 1$, the Poisson process $\nu_k(T_b)$ is the sum of two independent Poisson processes with parameters $\Lambda_{01} = sqAT_b$ and $\Lambda_{11} = (1-q)(1+s)AT_b$. Thus, $\nu_k(T_b)$ is a Poisson process with parameter $\Lambda_1 = \Lambda_{01} + \Lambda_{11} = (1-q+s)AT_b$, and the joint density $p_1(y_k)$ is

$$\begin{aligned} p_1(y_k) &= \Pr(\psi_{m,k} = n_{0k}, \nu_k(T_b) = n_k | H = H_1) \\ &= \frac{e^{-\Lambda_1}\Lambda_1^{n_k}}{n_k!} \binom{n_k}{n_{0k}} \pi_1^{n_{0k}} (1-\pi_1)^{n_k - n_{0k}} , \end{aligned} \tag{12}$$

where $\pi_1 \triangleq \Lambda_{01}/(\Lambda_{01} + \Lambda_{11}) = qs/(1-q+s)$.

Combining (11) and (12) we conclude that

$$L_{T_k}(\lambda_{\hat{m}}, \lambda_{\hat{m}}^c) = \frac{p_0(y_k)}{p_1(y_k)} = e^{\Lambda_1 - \Lambda_0} \left(\frac{\Lambda_0}{\Lambda_1}\right)^{n_k} \left(\frac{\pi_0}{\pi_1}\right)^{n_{0k}} \left(\frac{1-\pi_0}{1-\pi_1}\right)^{n_k - n_{0k}} . \tag{13}$$

Since during the second mode we have a binary hypothesis testing problem it is assumed that the decoder performs a Sequential Probability Ratio Test (SPRT). The properties of the SPRT have been thoroughly investigated in the literature [18]-[22]. It is particularly suitable for our purpose due to its optimality property; for specified levels of error probability, the SPRT is the test with the minimum expected stopping time [18, 21].

Let the observation space $\Omega = \mathbb{R}$ be the set of all real numbers. Following [22], a *sequential decision rule* is a pair of sequences $(\underline{\phi}, \underline{\delta})$ where $\underline{\phi} = \{\phi_j; j = 0, 1, 2, \ldots\}$, $\phi_j : \mathbb{R}^j \to \{0, 1\}$ being a *stopping rule*, while $\underline{\delta} = \{\delta_j; j = 0, 1, 2, \ldots\}$ being a *terminal decision rule*. Here δ_j is a decision rule on $(\mathbb{R}^j, \mathcal{B}^j)$ for each $j \geq 0$, where $\mathcal{B}$ denotes the σ-algebra on $\mathbb{R}$. The rule $(\underline{\phi}, \underline{\delta})$ makes the decision $\delta_N(y_1, y_2, \ldots, Y_N)$, where N is the stopping time defined by $N = \min\{n : \phi(y_1, y_2, \ldots, y_n) = 1\}$.

The *sequential probability ratio test* SPRT(A,B), with boundaries A and B satisfying $0 < A \leq 1 \leq B < \infty$, is defined by

$$\begin{aligned} \omega_n(y_1, \ldots, y_n) &= \prod_{k=1}^{n} L_{T_k} \\ \phi_n(y_1, \ldots, y_n) &= \begin{cases} 0, & \text{if } A < \omega_n(y_1, \ldots, y_n) < B \\ 1, & \text{otherwise} \end{cases} \end{aligned}$$

and

$$\delta_n(y_1, \ldots, y_n) = \begin{cases} H_0, & \text{if } \omega_n(y_1, \ldots, y_n) \geq B \\ H_1, & \text{if } \omega_n(y_1, \ldots, y_n) \leq A \end{cases}$$

Thus, the expected number of successive transmissions during the second mode can be expressed as

$$\mathbb{E}[N] = (1 - P_1(e))\mathbb{E}[N|H_0] + P_1(e)\mathbb{E}[N|H_1] \ ,$$

since $1 - P_1(e)$ and $P_1(e)$ are the priors of the hypotheses H_0 and H_1 respectively. Furthermore, it is evident that A determines the probability of initiating a new M-ary signaling period, whereas B determines the final message error probability.

Let η and α be *miss* and *false alarm* probabilities, defined by

$$\begin{aligned} \eta &= \Pr\left(\delta_N(y_1, \ldots, y_N) = H_1 | H_0\right) \triangleq P_M(\underline{\phi}, \underline{\delta}) \\ \alpha &= \Pr\left(\delta_N(y_1, \ldots, y_N) = H_0 | H_1\right) \triangleq P_F(\underline{\phi}, \underline{\delta}) \ , \end{aligned}$$

and suppose that N is sufficiently large, in which case one can ignore the "excess over the boundaries" and assume that $\omega_N(y_1, \ldots, y_N) \cong A$ or $\omega_N(y_1, \ldots, y_N) \cong B$. In this case Wald's approximations for the SPRT(A,B) yield $B \cong (1 - \eta)/\alpha$, $A \cong \eta/(1 - \alpha)$ and by [22, Proposition III.D.3]

$$\begin{aligned} \mathbb{E}[N|H_0] &= \frac{1}{\xi_0}\left[\eta \ln \frac{\eta}{1-\alpha} + (1-\eta)\ln\frac{1-\eta}{\alpha}\right] \approx \frac{1}{\xi_0}\left(\eta \ln \eta - \ln \alpha\right) \\ \mathbb{E}[N|H_1] &= \frac{1}{\xi_1}\left[(1-\alpha)\ln\frac{\eta}{1-\alpha} + \alpha \ln \frac{1-\eta}{\alpha}\right] \approx \frac{1}{\xi_1}\left(\ln \eta - \alpha \ln \alpha\right) \ , \end{aligned} \tag{14}$$

with $\xi_j \triangleq \mathbb{E}[\ln L_{T_k}\left(\lambda_{\hat{m}}, \lambda^c_{\hat{m}}\right)|H_j]$, $j = 0, 1$, and the approximations on the r.h.s. of (14) are valid for small α and η.

Noting that $\mathbb{E}(n_k|H_j) = \Lambda_j$ and $\mathbb{E}(n_{0k}|H_j) = \Lambda_{0j}$ we obtain using (13)

$$\begin{aligned} \xi_0 &= AT_b\left[1 - 2q + (s(2q-1) + q)\ln\left(\frac{1+s}{s}\right)\right] \\ \xi_1 &= AT_b\left[1 - 2q + (s(2q-1) - (1-q))\ln\left(\frac{1+s}{s}\right)\right] \ . \end{aligned}$$

Using the Taylor series expansion $\ln(1 + x) = x - x^2/2 + x^3/3 + \ldots$ for $x \to 0$, we obtain for the low power regime $(s \to \infty)$

$$\begin{aligned} \xi_0\Big|_{s\to\infty} &= AT_b\left[1 - 2q + (s(2q-1) + q)\left(\frac{1}{s} - \frac{1}{2s^2} + \frac{1}{3s^3} + O\left(\frac{1}{s^4}\right)\right)\right] \\ &= AT_b\left[\frac{1}{2s} + \frac{q-2}{6s^2} + O\left(\frac{1}{s^3}\right)\right] \\ \xi_1\Big|_{s\to\infty} &= AT_b\left[1 - 2q + (s(2q-1) - (1-q))\left(\frac{1}{s} - \frac{1}{2s^2} + \frac{1}{3s^3} + O\left(\frac{1}{s^4}\right)\right)\right] \\ &= -AT_b\left[\frac{1}{2s} + \frac{1-q}{6s^2} + O\left(\frac{1}{s^3}\right)\right] \ . \end{aligned} \tag{15}$$

For simplicity we assume that $\alpha = \eta$, in which case as $s \to \infty$

$$\begin{aligned} \mathbb{E}[N|H_0] &\approx -\frac{\ln \alpha}{\xi_0} \\ \mathbb{E}[N|H_1] &\approx \frac{\ln \alpha}{\xi_1} \ , \end{aligned} \tag{16}$$

while at the same time (14) and (15) imply that up to first order terms (i.e. $O(\frac{1}{s})$) $\mathbb{E}[N|H_1] = \mathbb{E}[N|H_0]$. Now let $\epsilon \to 0$ in such a way that $\epsilon^2 T \to \infty$, then $R_1 = (1-\epsilon)C \to C$ and $P_1(e) \to 0$ by (9) whence $\mathbb{E}[N] = \mathbb{E}[N|H_0] = \mathbb{E}[N|H_1]$.

The second mode fails (i.e., it terminates with an irreversible erroneous decision) if and only if the decoder's decision by the end of the first mode was $\hat{m} \neq m$ and then during the second mode a false alarm has occurred. Consequently, the message error probability for our scheme is $P(e) = P_1(e)P_F(H_0|H_1) = P_1(e)\alpha$. We shall henceforth assume that the exponent of α is significantly larger than the exponent of $P_1(e)$.

Let $\overline{T}_m$ denote the mean time for transmission of a single message via our coding scheme, then the error exponent for our random transmission time scheme can be expressed as

$$E_f(R) = -\frac{\ln P(e)}{\overline{T}_m} = -\frac{\ln \alpha}{\overline{T}_m} .$$

The mean time for the termination of the second phase follows from (16),

$$\overline{T}_2 = T_b \mathbb{E}[N] = T_b \mathbb{E}[N|H_1] \approx T_b \frac{\ln \alpha}{\xi_1} . \tag{17}$$

The mean time for transmission of a single message can be written in the following way,

$$\overline{T}_m = (T + \overline{T}_2)\Big\{1 + P_1(e) + [P_1(e)]^2 + \cdots\Big\} = \frac{T + \overline{T}_2}{1 - P_1(e)} \tag{18}$$

Define the information rate R of the system in terms of R_1, the information rate during the first mode, as

$$R\overline{T}_m = \ln M = R_1 T . \tag{19}$$

For rate R_1 approaching the capacity $R_1 = (1-\epsilon)C \to C$, and $P_1(e) \to 0$, we get from (17) (18) and (19), $\overline{T}_m = T + \overline{T}_2 = \frac{R}{R_1}\overline{T}_m + \overline{T}_2$, whence

$$\frac{R_1 - R}{R_1}\overline{T}_m = \overline{T}_2 \implies \frac{C - R}{C}\overline{T}_m = T_b \frac{\ln \alpha}{\xi_1} .$$

Therefore, the error exponent can be expressed as,

$$\begin{aligned} E_f(R) = -\frac{\ln \alpha}{\overline{T}_m} &= -\frac{\xi_1}{T_b} \cdot \frac{C - R}{C} \\ &= A\left[\frac{1}{2s} + \frac{1-q}{6s^2} + O\left(\frac{1}{s^3}\right)\right]\left(\frac{C-R}{C}\right) \end{aligned} \tag{20}$$

where the last step follows from (15).

Equation (20) reveals that, as $s \to \infty$, up to first order terms (i.e. $O(\frac{1}{s})$) $E_f(R)$ is insensitive to the choice of q during the orthogonal phase, whereas the second order term dictates that q should be as small as possible. This is due to the fact that when s gets large the coefficient of $\frac{1}{s}$ in ξ_1 is obtained by subtraction of the Poisson process intensities over the orthogonal intervals thus eliminating q. On the other hand it is obvious from (20) that $E_f(R)$ depends on C and hence sensitive to the choice of the average power during the M-ary signaling period.

Suppose next that an average-power constraint $\sigma < 1/2$ is imposed on our signaling scheme. Let the average signaling power during the M-ary phase be βA, while that during the sequential binary transmission phase be γA, where $\beta \geq \gamma, \gamma < 1/2$. As a result

$$-\ln P_1(e) \approx TC(\beta)\epsilon^2/4 ,$$

for $R_1 = (1-\epsilon)C(\beta)$ and $\epsilon > 0$ small, where $C(\beta) \triangleq A\beta(1-\beta)/(2s)$.

Provided that $P_1(e) \to 0$, we can use during the second phase the signaling pair $(\lambda_{\hat{m}}, \lambda_{\hat{m}}^c)$ as follows. If the decoder decodes $\hat{m} = m$, we transmit $\lambda_{\hat{m}}$ with average-power γA, while if the decoder is wrong we transmit $\lambda_{\hat{m}}^c$ with average-power $(1-\gamma)A$. Since this event has vanishing probability, it does not affect the average power. Furthermore, by (20) the exponent $E_f^{(\beta,\gamma)}(R)$ is given by

$$E_f^{(\beta,\gamma)}(R) \;=\; A\left[\frac{1}{2s} + \frac{1-\gamma}{6s^2} + O\left(\frac{1}{s^3}\right)\right]\left(\frac{C(\beta)-R}{C(\beta)}\right) .$$

Consequently, one needs to maximize $E_f^{(\beta,\gamma)}(R)$ subject to the average-power constraint

$$\beta T + \gamma \overline{T}_2 \leq \sigma \overline{T}_m \quad , \quad P_1(e) \to 0$$

which implies $\gamma \leq (\sigma C(\beta) - \beta R)\,/\,(C(\beta) - R)$.

It can be easily verified that

$$\begin{aligned} \frac{\partial}{\partial\beta}\left(\frac{C(\beta)-R}{C(\beta)}\right) &= \frac{\tilde{R}(1-2\beta)}{\beta^2(1-\beta)^2} \\ \frac{\partial^2}{\partial\beta}\left(\frac{C(\beta)-R}{C(\beta)}\right) &= \frac{-2\tilde{R}\left[\beta(1-\beta)+(1-2\beta)^2\right]}{\beta^3(1-\beta)^3} , \end{aligned}$$

with $\tilde{R} \triangleq R/\frac{A}{2s}$.

Consequently, on the interval $[0,1]$, $E_f^{(\beta,\gamma)}(R)$ is a concave function of β. Furthermore, on $[0,1/2]$ it is an increasing function of β, and it attains its global maximum at $\beta = 1/2$. Taking into account the average-power constraint σ, we conclude that

$$\begin{array}{lll} \beta = \frac{1}{2} \;, & \gamma = \epsilon \ll 1 & \text{if } \; R/\left(\frac{A\sigma}{4s}\right) < 1 \\ \beta = 1 - \frac{R}{2}/\left(\frac{A\sigma}{4s}\right) \;, & \gamma = \epsilon \ll 1 & \text{if } \; R/\left(\frac{A\sigma}{4s}\right) \geq 1 \end{array}$$

It is instructive to note that our analysis implies that $E_f^{(\beta,\gamma)}(R)$ is maximized by picking γ as small as possible. The same conclusion has been reached by Schalkwijk-Barron with regard to the average signaling power during the second stage of the transmission in [1, section III].

To summarize our results so far, as long as $R/\left(\frac{A\sigma}{4s}\right) < 1$, similar to $C_{s\to\infty}$, the error exponent $E_f(R)_{s\to\infty}$ is maximized with $\beta = 1/2$, in which case

$$\begin{aligned} \max_{\beta} E_f(R)_{s\to\infty} = E_f(R)_{s\to\infty}\Big|_{\beta=1/2} &= A\left[\frac{1}{2s} + O\left(\frac{1}{s^2}\right)\right]\left(\frac{C(1/2)-R}{C(1/2)}\right) \\ &\cong 4(C_{s\to\infty} - R) . \end{aligned} \tag{21}$$

In order to show that $4(C-R)$ is the maximum achievable exponent for the Poisson single-user channel with constant dark current we follow the same steps as in [1, Appendix],

- The expected time to decision for the optimum sequential test for a given message error probability $P(e)$ is not smaller than the two-step test which upon reaching the value $1-\pi_+, \pi_+ > P(e) > 0$ for some message m_π joins together all other messages and continues with an optimal binary test.
- For the first transmission phase it follows by the same arguments as in [1, Appendix] that $\overline{T} \geq (\ln M)/C$.
- Following a similar procedure as in [33] one can show that in order to communicate a binary message over the Poisson feedback channel, in the presence of constant dark current, with a peak-power and intensity-modulating-waveform constraints, the optimal signals are orthogonal. It follows then by our analysis herein that asymptotically

  $$\overline{T}_b \approx -T_b \frac{\ln P(e)}{\xi_0} = -\frac{\ln P(e)}{4C} ,$$

  where the last step follows from (16), (7) and the approximation of ξ_0 as $s \to \infty$.
- The above steps establish $4(C-R)$ as an upper bound on the attainable exponent for the optimal sequential test.

We see that the restriction to intensity modulated pulses penalizes the communication cost in the sense that, while $4(C-R)$ is an achievable exponent for the AWGN channel with $\kappa = 1$, the same exponent is achievable in the Poisson regime, via a similar coding approach, yet with $\kappa \geq 2$ (equality holds when $\sigma = 1/2$).

III. A two-user random transmission time code

In this section we propose a random transmission time scheme that incorporates the noiseless feedback in order to improve the attainable exponent for the two-user Poisson MAC.

We begin by recalling some known results about the two-user Poisson MAC without feedback. The capacity region for this model was determined by Lapidoth-Shamai in [11] and it is described as follows. For a fixed peak power A and average-power constraint $(q_1 A, q_2 A)$ on the two users the set

$$\{\boldsymbol{R} = (R_1, R_2) \ : \ R_1 \leq Ag(q_2, q_1), R_2 \leq Ag(q_1, q_2), R_1 + R_2 \leq Af(q_1, q_2)\} \tag{22}$$

describes a "typical" pentagon of attainable rates under the $(q_1 A, q_2 A)$ average-power constraint, where

$$\begin{aligned} f(q_1, q_2) &\triangleq -(s + q_1 + q_2)\ln(s + q_1 + q_2) + q_1 q_2 (2+s)\ln(2+s) \\ &\quad +(1-q_1)(1-q_2)s\ln s + (q_1 + q_2 - 2q_1 q_2)(1+s)\ln(1+s) \\ g(q_1, q_2) &\triangleq -(1-q_1)(s+q_2)\ln(s+q_2) - q_1(1+s+q_2)\ln(1+s+q_2) \\ &\quad +q_1 q_2(2+s)\ln(2+s) + (1-q_1)(1-q_2)s\ln s \\ &\quad +(q_1+q_2-2q_1q_2)(1+s)\ln(1+s) \ . \end{aligned}$$

The pentagon's size and shape depend on the parameters (q_1, q_2, s), and in particular as $s \to \infty$ the pentagon becomes a rectangle. It is shown in [11, Lemmas 3,4] that one of the vertices of the pentagon (22) intersects the boundary of the capacity region provided that

$$\frac{\partial f(q_1,q_2)}{\partial q_1} \cdot \frac{\partial g(q_2,q_1)}{\partial q_2} - \frac{\partial f(q_1,q_2)}{\partial q_2} \cdot \frac{\partial g(q_2,q_1)}{\partial q_1} = 0 \ .$$

Furthermore, the rate-sum capacity is attainable with symmetric average-power constraint [11].

Computing the limit of $f(q,q)$ as $s \to \infty$, assuming average-power constraint σA on both users, one obtains

$$C_{sum} = A\frac{q^*(1-q^*)}{s} + O\left(\frac{1}{s^2}\right) \quad , \quad q^* = \min\{\sigma, 1/2\} \; . \tag{23}$$

A two-user code with parameters $(M_1, M_2, T, \sigma, P_e)$ for this model is defined by

- Two sets of waveforms $\{\lambda_{m_1}(t)\}$ and $\{\lambda_{m_2}(t)\}$, $1 \le m_1 \le M_1$, $1 \le m_2 \le M_2$, $0 \le t \le T$, which satisfy the peak and average-power constraints (3).
- A decoder mapping $D : \nu_0^T \to \{1, 2, \ldots, M_1\} \times \{1, 2, \ldots, M_2\}$ such that its average probability of error is given by

$$P_e = \frac{1}{M_1 M_2} \sum_{m_1=1}^{M_1} \sum_{m_2=1}^{M_2} \Pr\left\{D(\nu_0^T) \neq (m_1, m_2) | \lambda_{m_1}(\cdot), \lambda_{m_2}(\cdot)\right\} \; .$$

A code construction of a family of exponentially optimal one-way two-user codes is exhibited in [13], and it is defined as follows. When $q_1 = q_2 = q \le \sigma$ construct first $\mathcal{A}^{(M_1+M_2,q)}$, then subdivide the interval $I_T = [0, T]$ into $\binom{M_1+M_2}{(M_1+M_2)q}$ disjoint half open, closed on the right, equi-length subintervals. Partition arbitrarily the set of $M_1 + M_2$ rows of $\mathcal{A}^{(M_1+M_2,q)}$ into two subsets of sizes M_1 and M_2 respectively, thereby obtaining two submatrices denoted as $\mathcal{A}_{\mathcal{M}_1}^{(M_1+M_2,q)}$ and $\mathcal{A}_{\mathcal{M}_2}^{(M_1+M_2,q)}$, respectively. Set for $m_1 \in \{1, \ldots, M_1\}, m_2 \in \{1, \ldots, M_2\}$ and t in subinterval j of I_T

$$\begin{aligned} \lambda_{m_1}(t) &= A\left(\mathcal{A}_{\mathcal{M}_1}^{(M_1+M_2,q)}\right)_{m_1,j} \\ \lambda_{m_2}(t) &= A\left(\mathcal{A}_{\mathcal{M}_2}^{(M_1+M_2,q)}\right)_{m_2,j} \; . \end{aligned}$$

When $q_1 \neq q_2$ assume, without loss of generality, that $q_1 < q_2$. In this case we generate first the code $\left(\mathcal{A}_{\mathcal{M}_1}^{(M_1+M_2,q)}, \mathcal{A}_{\mathcal{M}_2}^{(M_1+M_2,q)}\right)$ and modify $\mathcal{A}_{\mathcal{M}_1}^{(M_1+M_2,q)}$ replacing it with another matrix that conforms with q_1. To this end define $q = q_1/q_2 < 1$. Next, construct $\mathcal{A}^{(M_1,q)}$, let $\zeta = \binom{M_1}{M_1 q}$ and modify the jth row of $\mathcal{A}_{\mathcal{M}_1}^{(M_1+M_2,q)}$ as follows

- Replace any 1 with the jth row of $\mathcal{A}^{(M_1,q)}$.
- Replace any 0 with a ζ-repetition of itself.

As a result we get a pair of simplex systems, corresponding to the parameters q_1, q_2, where it is implicitly assumed that the new matrix $\mathcal{A}_{\mathcal{M}_1}$ corresponds now to a "refinement" of the original partitioning of I_T.

This code is shown in [13] to exhibit the properties of a "random code". That is, as we let χ_i denote the support $\chi_i \triangleq \{t : \lambda_i(t) = A\}$ then if χ_i and $\chi_j, i \neq j$ belong both either to $\mathcal{A}_{\mathcal{M}_1}$ or $\mathcal{A}_{\mathcal{M}_2}$ then one of the following holds (here $\mu(B)$ denotes Lebesgue measure of a measurable set B)

$$\left|\frac{1}{T}\mu\{\chi_i\} - q_1\right| \le \epsilon \quad \left|\frac{1}{T}\mu\{\chi_i \cap \chi_j\} - q_1^2\right| \le \epsilon \; ,$$

or

$$\left|\frac{1}{T}\mu\{\chi_i\} - q_2\right| \leq \epsilon \quad \left|\frac{1}{T}\mu\{\chi_i \cap \chi_j\} - q_2^2\right| \leq \epsilon \ ,$$

while otherwise we have

$$\left|\frac{1}{T}\mu\{\chi_i \cap \chi_j\} - q_1 q_2\right| \leq \epsilon \ , \ \chi_i \in \mathcal{A}_{\mathcal{M}_1} \ , \ \chi_j \in \mathcal{A}_{\mathcal{M}_2} \ .$$

A suitable mapping D for the decoder is defined as follows. For any hypothesis $m = (m_1, m_2)$ the decoder computes

$$\begin{aligned}
\psi_m^{(1)} &= \int_{\chi_{m_1} \cap \chi_{m_2}} d\nu(t) = \text{number of arrivals in } \chi_{m_1} \cap \chi_{m_2} \\
\psi_m^{(0)} &= \int_{\chi_{m_1} \cup \chi_{m_2}} d\nu(t) - \psi_m^{(1)} \ .
\end{aligned}$$

Then $D(\nu_0^T) = m^* = (m_1^*, m_2^*)$ iff

- $\psi_{m^*}^{(1)} + \psi_{m^*}^{(0)} \geq \psi_m^{(1)} + \psi_m^{(0)} \quad \forall \ m \neq m^*$
- If $\psi_{m^*}^{(1)} + \psi_{m^*}^{(0)} = \psi_m^{(1)} + \psi_m^{(0)}$ then $\psi_{m^*}^{(1)} > \psi_m^{(1)}$
- If two pairs m and m^* exist such that $\psi_{m^*}^{(1)} = \psi_m^{(1)}$ and $\psi_{m^*}^{(0)} = \psi_m^{(0)}$ and both maximize $\psi^{(1)} + \psi^{(0)}$ an error is declared.

The reliability function for the one-way two-user Poisson MAC was determined in [13] and it is defined as follows. For any rate pair (R_1, R_2) inside the capacity region, the average probability of error decays according to $P_e = \exp\{-E(R_1, R_2)T + o(T)\}, T \to \infty$ and the error exponent is given by

$$\begin{aligned}
E(R_1, R_2) \ = \ \min\Big\{ & \max_{0 \leq \rho_1 \leq 1} AE_{11}(\rho_1, q_2, q_1) - \rho_1 R_1 \ , \ \max_{0 \leq \rho_2 \leq 1} AE_{11}(\rho_2, q_1, q_2) - \rho_2 R_2 \\
& \max_{0 \leq \rho_3 \leq 1} AE_{12}(\rho_3, q_1, q_2) - \rho_3 (R_1 + R_2) \Big\} \ ,
\end{aligned}$$

where

$$\begin{aligned}
E_{11}(\rho, q_1, q_2) &= s + q_1 + q_2 - (1 - q_1)s[1 + \varpi_0 q_1]^{1+\rho} - q_2(1 + s)[1 + \varpi_1 q_1]^{1+\rho} \\
E_{12}(\rho, q_1, q_2) &= s + q_1 + q_2 \\
& \quad - \left[(1 - q_1)(1 - q_2)s^{\frac{1}{1+\rho}} + (q_1 + q_2 - 2q_1 q_2)(1 + s)^{\frac{1}{1+\rho}} + q_1 q_2 (2 + s)^{\frac{1}{1+\rho}}\right]^{1+\rho}
\end{aligned}$$

and

$$\varpi_0 \triangleq \left(1 + \frac{1}{s}\right)^{\frac{1}{1+\rho}} - 1 \ , \quad \varpi_1 \triangleq \left(1 + \frac{1}{1+s}\right)^{\frac{1}{1+\rho}} - 1 \ .$$

The reliability function $E(R_1, R_2)$ has been obtained in [13] assuming that the decoder is a maximum-likelihood decoder. Nevertheless, it can be shown using the bounding technique proposed in [14] that the aforementioned code construction and decoder mapping achieve $E(R_1, R_2)$ as well.

In what follows we consider a symmetric two-user signaling scheme and analogously to the single-user code, during the first phase of the transmission both senders will communicate at rate-sum approaching the rate-sum capacity. Since the dominating exponent for rate pairs approaching the rate-sum capacity of the pentagon (22) (assuming fixed average-power signaling) is $AE_{12}(\rho, q, q) - \rho(R_1 + R_2)$ (see for example [24, Fig. 8]) we compute the limit of this exponent in the low power regime. To this end note that

$$\begin{aligned} E_{12}(\rho, q) &= s + 2q - s\left[1 + q\tau_3\right]^{(1+\rho)} \\ \tau_3 &\triangleq 2(1-q)\tau_1 + q\tau_2 \\ \tau_1 &\triangleq \left(1 + \frac{1}{s}\right)^{\frac{1}{1+\rho}} - 1 \\ \tau_2 &\triangleq \left(1 + \frac{2}{s}\right)^{\frac{1}{1+\rho}} - 1 \, . \end{aligned}$$

Taking the limit of τ_1 and τ_2 as $s \to \infty$ we get

$$\begin{aligned} \tau_1 &= \frac{1}{(1+\rho)s}\left[1 - \frac{\rho}{2(1+\rho)s} + O\left(\frac{1}{s^2}\right)\right] \\ \tau_2 &= \frac{2}{(1+\rho)s}\left[1 - \frac{\rho}{(1+\rho)s} + O\left(\frac{1}{s^2}\right)\right] \, . \end{aligned}$$

Therefore,

$$\tau_3 = \frac{1}{(1+\rho)s}\left[2 - \frac{(1+q)\rho}{(1+\rho)s} + O\left(\frac{1}{s^2}\right)\right] \, ,$$

and

$$\begin{aligned} E_{12}(\rho, q) &= s + 2q - s\left[1 + (1+\rho)\tau_3 q + \frac{(1+\rho)\rho}{2}(\tau_3 q)^2 + O(\tau_3^3 q^3)\right] \\ &= s + 2q - s\left\{1 + \frac{q}{s}\left[2 - \frac{(1+q)\rho}{(1+\rho)s} + O\left(\frac{1}{s^2}\right)\right] + \frac{(1+\rho)\rho q^2}{2(1+\rho)^2 s^2}\left[4 + O\left(\frac{1}{s}\right)\right]\right\} \\ &= \frac{\rho}{(1+\rho)} \cdot \frac{q(1-q)}{s} + O\left(\frac{1}{s^2}\right) \, . \end{aligned} \tag{24}$$

Combining (23) with (24) we conclude that in the low power regime

$$E(R_{sum}) \approx \max_{0 \le \rho \le 1}\left[\frac{\rho}{1+\rho} C_{sum} - \rho R_{sum}\right] \quad , \quad 0 \le R_{sum} \le C_{sum} \, ,$$

from which it follows as in [25, Section 3.4, pp. 155-157] that

$$E(R_{sum}) = \begin{cases} C_{sum}/2 - R_{sum} & 0 \le R_{sum}/C_{sum} \le \frac{1}{4} \\ \left(\sqrt{C_{sum}} - \sqrt{R_{sum}}\right)^2 & \frac{1}{4} \le R_{sum}/C_{sum} \le 1 \end{cases} \tag{25}$$

This result is expected since the joint decoder utilizes the power of both users. As a result, when considering a symmetric scheme wherein $q_1 = q_2 = q$, the error exponent of the joint decoder improves by a factor of two as compared to the situation of a single-user with average-power q.

We shall now proceed to exhibit a random transmission time signaling scheme with peak and average power constraints $(A, qA, qA), q \leq \sigma$, and as before the scheme consists of two phases. The first phase is a regular two-user M_1M_2-ary signaling phase wherein the senders use the aforementioned codebook in order to communicate a message pair $m = (m_1, m_2), m_i \in \{1, \dots, M_i\}, i = 1, 2$ over a duration T interval, where $M_i = e^{R_i^{(1)}T}$ (the superscript 1 refers to the first transmission phase). Let the decoder use the joint decoding rule as defined above, then for rate-sums $R_{sum} = R_1^{(1)} + R_2^{(1)} = (1-\epsilon)C_{sum}$ and $s \to \infty$, we have from (25)

$$-\ln P_1(e) \approx TC_{sum}\epsilon^2/4 + O(\epsilon^3)\ . \tag{26}$$

We improve on this performance by continuing with a quaternary signaling mode. Given the decoder's decision $\hat{m} = (\hat{m}_1, \hat{m}_2)$ (which is known to both senders via the feedback link) encoder i transmits sequentially, in N periods each of duration T_Q (where again the number of periods N is random), the waveform (assuming that time zero refers to the beginning of the first mode) $\lambda_i^{(Q)}\left(t - \lfloor\frac{t-T}{T_Q}\rfloor T_Q - T\right)$ where for $0 < \gamma < 1/2$

$$\lambda_i^{(Q)}(t) = \begin{cases} A & \text{if } \ t \in [0, \gamma T_Q] & \text{and } \ \hat{m}_i = m_i \\ 0 & \text{if } \ t \in [\gamma T_Q, T_Q] & \text{and } \ \hat{m}_i = m_i \\ 0 & \text{if } \ t \in [0, \gamma T_Q] & \text{and } \ \hat{m}_i \neq m_i \\ A & \text{if } \ t \in [\gamma T_Q, T_Q] & \text{and } \ \hat{m}_i \neq m_i \end{cases}$$

That is, each sender conveys (repeatedly) to the decoder whether or not its corresponding message has been decoded correctly via binary signaling as follows,

- It activates its peak power over $[0, \gamma T_Q]$ to confirm the decoder's decision regarding its message.
- It activates its peak power over $[\gamma T_Q, T_Q]$ to reject the decoder's decision regarding its message.

If the decoder's decision, by the termination of the second transmission phase of random duration $T_2 = NT_Q$, is upon a positive confirmation from both senders, $\hat{m} = (\hat{m}_1, \hat{m}_2)$ is the final decision. If however, the decoder decides that at least one sender conveyed a rejection, the encoders-decoder start over with a new M_1M_2-ary signaling period, and so forth.

This decision protocol is suboptimal in the sense that no partial information is allowed, either both messages are accepted or both are rejected. Thus, it is clear that the resulting error exponent lower bounds the optimal one.

Denote the four possible message combinations during the sequential phase as
H_0 : both encoders confirm.
H_1 : encoder 1 confirms while encoder 2 rejects.
H_2 : encoder 1 rejects while encoder 2 confirms.
H_3 : both encoders reject.

Next let the decoder observe $\nu_k = \{\nu(t) : t \in T_k = [T + (k-1)T_Q, T + kT_Q]\}$, and consider the sequence of independent identically distributed observations $\{\nu_k; k = 1, 2, \dots\}$ where

$$H_i \ : \ \nu_k \sim f_j \ , \ k = 1, 2, \dots \ , \ j = 0, 1, 2, 3$$

where f_0 induces the predictable $\mathcal{F}_t^{\nu_k}$-intensity

$$\lambda(t) = \begin{cases} 2A & t \in [T+(k-1)T_Q, T+(k-1)T_Q+\gamma T_Q] \\ 0 & t \in (T+(k-1)T_Q+\gamma T_Q, T+kT_Q] \end{cases}$$

f_1 and f_2 induce the predictable $\mathcal{F}_t^{\nu_k}$-intensity $\lambda(t) = A, t \in T_k$, and f_3 induces the predictable $\mathcal{F}_t^{\nu_k}$-intensity

$$\lambda(t) = \begin{cases} 0 & t \in [T+(k-1)T_Q, T+(k-1)T_Q+\gamma T_Q] \\ 2A & t \in (T+(k-1)T_Q+\gamma T_Q, T+kT_Q] \end{cases}$$

To proceed further, it is helpful to denote the Kullback-Leibler distance (referred also as the relative entropy) between the densities f and g by $D(f \parallel g)$, so that

$$D(f \parallel g) \triangleq \mathbb{E}_f \left[\log \frac{f(X)}{g(X)}\right] = \int f(x) \log \frac{f(x)}{g(x)}\, dx \ .$$

A straightforward computation shows that when f_j induces the predictable $\mathcal{F}_t^{\nu_k}$-intensity $\lambda^{(j)}(t)$ and f_l induces the predictable $\mathcal{F}_t^{\nu_k}$-intensity $\lambda^{(l)}(t)$, then (see also [27, pp. 1025, example 2])

$$D(f_j \parallel f_l) = \int_0^{T_Q} \left[\lambda^{(l)}(t) - \lambda^{(j)}(t)\right] dt + \int_0^{T_Q} \lambda^{(j)}(t) \ln \left[\frac{\lambda^{(j)}(t)}{\lambda^{(l)}(t)}\right] dt \ .$$

In our case the relative entropies admit the values

$$\begin{aligned} D(f_0 \parallel f_1) &= D(f_0 \parallel f_2) \\ &= AT_Q \left[1 - 2\gamma + (2+s)\gamma \ln\left(\frac{2+s}{1+s}\right) + s(1-\gamma)\ln\left(\frac{s}{1+s}\right)\right] \\ D(f_0 \parallel f_3) &= AT_Q \left[2(1-2\gamma) + (2+s)\gamma \ln\left(\frac{2+s}{s}\right) + s(1-\gamma)\ln\left(\frac{s}{2+s}\right)\right] \end{aligned}$$

from which one can verify that in the low power regime

$$\begin{aligned} D(f_0 \parallel f_3)\Big|_{s\to\infty} &= AT_Q \left[\frac{2}{s} + \frac{4\gamma - 8}{3s^2} + O\left(\frac{1}{s^3}\right)\right] \\ D(f_0 \parallel f_1)\Big|_{s\to\infty} &= AT_Q \left[\frac{1}{2s} - \frac{\gamma}{3(1+s)^2} - \frac{1}{3s^2} + O\left(\frac{1}{s^3}\right)\right] \ . \end{aligned} \tag{27}$$

Since during the second phase the decoder faces a quaternary hypothesis testing problem, we consider as a test strategy the M-ary sequential probability ratio test (MSPRT), defined in the sequel.

Let $X_1, X_2, \ldots$ be an infinite sequence of i.i.d. distributed random variables with density f, and let H_j be the hypothesis that $f = f_j$ for $j = 0, 1, \ldots, M-1$. It is asumed that $f_l \neq f_j$ almost surely for all $j \neq l$. Assume further that the prior probabilities of the hypotheses are known, and let π_j denote the prior probability of H_j. If only n observations are available, the minimum-probability-of-error test uses as the test statistics the vector of posterior probabilities: $\boldsymbol{p}_n \triangleq \left(p_n^0, \ldots, p_n^{M-1}\right)$ where

$$p_n^j = \Pr\{H = H_j | X_1, \ldots, X_n\} \ .$$

The minimum-probability-of-error test picks the hypothesis with the largest posterior probability, conditioned on the observations [22]. When the observations arrive periodically over time a compromise must be struck between the number of observations and the error probability. A sequential test strikes such a compromise, and it consists of a stopping rule and a final decision rule.

For the cost assignment, we assume that each time step taken by the test costs a positive amount c, and that decision errors in the final decision are penalized through a uniform cost function $W(\delta; H)$, that is,

$$\begin{aligned} W(H_j, H_j) &= 0 \\ W(H_j, H_l) &= 1 \ , \quad \forall j \neq l \ . \end{aligned}$$

The Bayesian optimization problem is then

$$\min_{\mathcal{S}} \mathbb{E}\{cN + W(\delta; H)\} \ ,$$

where $\mathcal{S}$ denotes the set of all admissible sequential tests.

A recursive solution for the optimal multihypothesis sequential test in a Bayesian setting has been obtained in [28]-[30], but this solution is very complex and impractical except for a few special cases. This motivates the interest in alternative *suboptimal* tests that exhibit a simple structure and which are amenable to practical implementation. In this sense the MSPRT appears to be a good choice. Our choice is further justified since we are mainly interested in the asymptotic performance of the M-ary sequential test when the error probabilities are small and the expected stopping time is large, in which case it is shown in [31, section III] that an approximation to the optimal test gives exactly the MSPRT structure. Furthermore, it is shown in [32] that the MSPRT is asymptotically optimal relative not only to the expected sample size but also to any positive moment of the stopping time distribution, when the error probabilities are small.

For the M-ary sequential probability ratio test the stopping time N_A and final decision δ are described as follows [31, section II]:

- $N_A =$ first $n \geq 1$ such that $p_n^l > 1/(1 + A_l)$ for at least one l .
- $\delta = H_m$, where $m = \arg\max_j p_{N_A}^j$.

The parameters A_j are assumed to be positive, and typically they are less than one in which case the condition $p_n^j > 1/(1 + A_j)$ can be satisfied by at most one j. It can be verified that the MSPRT with $M = 2$ reduces to the SPRT with parameters $(\pi_1 A_1/\pi_0, \pi_1/(\pi_0 A_0))$.

Let $\alpha_{j,l} = P_{f_j}(\text{accept } H_l)$ and let

$$\alpha_l \triangleq P(\text{accept } H_l \text{ incorrectly}) = \sum_{j: j \neq l} \pi_j \alpha_{j,l} \ .$$

The asymptotic performance of the MSPRT (i.e. the behavior of the test in the limit as $A_j \to 0$, $j = 1, 2, \ldots$) can be summarized via the following approximations due to Baum and Veeravalli in [31]

$$\begin{aligned} \mathbb{E}_{f_l}[N_A] &\approx \frac{-\ln A_l}{\min_{j: j \neq l} D(f_l \parallel f_j)} \\ \alpha_l &\approx \pi_l A_l \eta_l \ , \end{aligned} \tag{28}$$

where $0 < \eta_l < 1$ for each l. Here the α_l's are the *frequentist* error probabilities, not to be confused with the probabilities of error conditioned on a particular hypothesis.

The second phase fails if and only if the decoder's decision by the end of the first phase was $\hat{m} \neq m$ and then during the second phase a false alarm has occurred. This is due to the fact that any other error during the second phase will not cause a retransmission of the two messages. Consequently, the message error probability for our scheme is

$$\begin{aligned} P(e) &= \sum_{j=1}^{3} \Pr(\text{sent message is } H_j) P_{f_j}(\text{accept } H_0) \\ &= P\left(\text{accept } H_0 \text{ incorrectly}\right) = \alpha_0 \ . \end{aligned}$$

Letting $\overline{T}_m = T + \overline{T}_Q$ denote the mean time for transmission of a message pair $m = (m_1, m_2)$ via our coding scheme, then following similar considerations as in section II we obtain that

$$\begin{aligned} \overline{T}_Q &= T_Q \sum_{i=0}^{3} \pi_i \mathbb{E}_{f_i}[N_A] \approx T_Q \mathbb{E}_{f_0}[N_A] \\ \overline{T}_m &= T_Q \mathbb{E}_{f_0}\left[N_A\right] \frac{C_{sum}}{C_{sum} - R_{sum}} \\ &\approx T_Q \frac{-\ln A_0}{\min_{j:j\neq 0} D(f_0 \parallel f_j)} \cdot \frac{C_{sum}}{C_{sum} - R_{sum}} \ , \end{aligned}$$

where the last step follows from (28). We now assume that the exponent of α_0 is determined by the exponent of A_0, and suppose further that during the first phase of $M_1 M_2$-ary communication both senders use the same average-power β while during the second phase of 4-ary hypothesis testing they use average-power γ to confirm the decoder's decision. It follows that

$$\begin{aligned} E_f^{(\beta,\gamma)}(R_{sum}) &= -\frac{\ln P(e)}{\overline{T}_m} \\ &= \frac{1}{T_Q} \cdot \frac{C_{sum}(\beta) - R_{sum}}{C_{sum}(\beta)} \min_{j:j\neq 0} D(f_0 \parallel f_j) \ . \end{aligned} \tag{29}$$

Combining (27) with (29) we conclude that in the low power regime

$$E_f^{(\beta,\gamma)}(R_{sum}) = A\left[\frac{1}{2s} - \frac{\gamma}{3(1+s)^2} - \frac{1}{3s^2} + O\left(\frac{1}{s^3}\right)\right] \frac{C_{sum}(\beta) - R_{sum}}{C_{sum}(\beta)} \ . \tag{30}$$

Again, equation (30) reveals that, as $s \to \infty$, up to first order terms (i.e. $O(\frac{1}{s})$) $E_f^{(\beta,\gamma)}(R_{sum})$ is insensitive to the choice of γ, whereas the second order term dictates that γ should be as small as possible. On the other hand it is obvious from (30) that $E_f^{(\beta,\gamma)}(R_{sum})$ depends on C and hence sensitive to the choice of β during the M-ary signaling period.

Consequently, one needs to maximize $E_f^{(\beta,\gamma)}(R_{sum})$ subject to the average-power constraint

$$\beta T + \gamma \overline{T}_Q \leq \sigma \overline{T}_m \ , \quad P_1(e) \to 0$$

The optimization yields

$$\begin{array}{lll} \beta = \frac{1}{2} \ , & \gamma = \epsilon \ll 1 & \text{if } R_{sum}/\left(\frac{A\sigma}{2s}\right) < 1 \\ \beta = 1 - \frac{R_{sum}}{2}/\left(\frac{A\sigma}{2s}\right) \ , & \gamma = \epsilon \ll 1 & \text{if } R_{sum}/\left(\frac{A\sigma}{2s}\right) \geq 1 \end{array}$$

To summarize our results so far, as long as $R_{sum}/\left(\frac{A\sigma}{2s}\right) < 1$, similar to $C_{sum}|_{s\to\infty}$, the error exponent $E_f(R_{sum})|_{s\to\infty}$ is maximized with $\beta = 1/2$, in which case

$$\begin{aligned} \max_{\beta} E_f(R_{sum})\Big|_{s\to\infty} &= A\left[\frac{1}{2s} + O\left(\frac{1}{s^2}\right)\right]\left(\frac{C_{sum}(1/2) - R_{sum}}{C_{sum}(1/2)}\right) \\ &\cong 2(C_{sum} - R_{sum}) \ . \end{aligned}$$

We see that for the two-user Poisson MAC subject to constant dark current, the attainable feedback exponent in the low power regime is *at least* $2(C_{sum} - R_{sum})$. This is achieved with communication cost of $\kappa = 2/\left(R/\frac{A}{4s}\right) \geq 2$.

References

[1] J. P. M. Schalkwijk and M. E. Barron, "Sequential signaling under a peak power constraint," *IEEE Trans. Inform. Theory.* vol. IT-17, No. 3, pp. 278-282, May 1971.

[2] C. E. Shannon, R. G. Gallager, and E. R. Berlekamp, "Lower bounds to error probability for for coding on discrete memoryless channels, I and II," *Inform. Contr.,* vol. 10, Jan. 1967, pp. 65-103, and May 1967, pp. 522-552.

[3] A. D. Wyner, "On the probability of error for communication in white Gaussian noise," *IEEE Trans. Inform. Theory,* vol. IT-13, No. 1, pp. 86-90, Jan. 1967.

[4] A. D. Wyner, "Capacity and error exponent for the direct detection photon channel – parts I and II," *IEEE Trans. Inform. Theory,* vol. IT-34, No. 6, pp. 1449-1471, November 1988.

[5] C. E. Shannon, "The zero-error capacity of a noisy channel," *IRE Trans. Inform. Theory,* vol. IT-2, No. 5, pp. 8-19, Sep. 1967.

[6] T. T. Kadota, M. Zakai, and J. Ziv, "Capacity of a continuous memoryless channel with feedback," *IEEE Trans. Inform. Theory,* vol. IT-17, No. 4, pp. 372-378, July 1971.

[7] G. L. Turin, "Signal design for sequential detection systems with feedback," *IEEE Trans. Inform. Theory,* vol. IT-11, No. 4, pp. 401-408, July 1965.

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

[9] M. H. A. Davis, "Capacity and cut-off rate for Poisson-type channels," *IEEE Trans. Inform. Theory,* vol. IT-26, No. 6, pp. 710-715, November 1980.

[10] A. Lapidoth, "On the reliability function of the ideal Poisson channel with noiseless feedback," *IEEE Trans. Inform. Theory,* vol. IT-39, No. 2, pp. 491-503, March 1993.

[11] A. Lapidoth and S. Shamai (Shitz), "The Poisson multiple-access channel," *IEEE Trans. Inform. Theory,* vol. IT-44, No. 2, pp. 488-502, March 1998.

[12] M. R. Frey, "Information capacity of the Poisson channel," *IEEE Trans. Inform. Theory,* vol. IT-37, No. 2, pp. 244-256, March 1991.

[13] S. I. Bross, M. V. Burnashev and S. Shamai (Shitz), "Error exponents for the two-user Poisson multiple-access channel," *IEEE Trans. Inform. Theory.* vol. IT-47, No. 5, pp. 1999-2016, July 2001.

[14] S. I. Bross and S. Shamai (Shitz), "Capacity and decoding rules for the Poisson arbitrarily varying channel," *Proc. of ISIT 2001, p. 273, Washington, DC, June 2001,* and submitted to *IEEE Trans. Inform. Theory.*

[15] N. T. Gaarder and J. K. Wolf, "The capacity region of a multiple-access discrete memoryless channel can increase with feedback," *IEEE Trans. Inform. Theory,* vol. IT-21, No. 1, pp. 100-102, January 1975.

[16] T. M. Cover and S. K. Leung, "An achievable rate region for the multiple access channel with feedback," *IEEE Trans. Inform. Theory,* vol. IT-27, No. 3, pp. 292-298, May 1981.

[17] L. H. Ozarow, "The capacity of the white Gaussian multiple-access channel with feedback," *IEEE Trans. Inform. Theory,* vol. IT-30, No. 4, pp. 623-629, July 1984.

[18] A. Wald, *Sequential Analysis.* New-York: Wiley, 1947.

[19] D. Siegmund, *Sequential Analysis: Tests and Confidence Intervals.* New-York: Springer-Verlag, 1985.

[20] M. Woodroofe, *Nonlinear Renewal Theory in Sequential Analysis*, SIAM, 1982.

[21] E. L. Lehmann, *Testing Statistical Hypotheses.* New-York: Wiley, 1959.

[22] H. Vincent Poor, *An Introduction to Signal Detection and Estimation.* New-York: Springer-Verlag, 1988.

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

[24] R. G. Gallager, "A perspective on multiaccess channels," *IEEE Trans. Inform. Theory.* vol. IT-31, No. 2, pp. 124-142, March 1985.

[25] A. J. Viterbi and J. K. Omura, *Principles of Digital Communication and Coding.* New York: McGraw-Hill 1979.

[26] A. J. Viterbi, "The effect of sequential decision feedback on communication over the Gaussian channel," *Inform. and Contr.*, vol. 8, No. 1, pp. 80-92, Feb. 1965.

[27] S. Verdú, "On channel capacity per unit cost," *IEEE Trans. Inform. Theory.* vol. IT-36, No. 5, pp. 1019-1030, September 1990.

[28] A. G. Tartakovsky, "Sequential testing of many simple hypotheses with independent observations," *Probl. Inform. Transm.,* vol. 24, No. 4, pp. 299-309, 1989.

[29] S. Zacks, *The Theory of Statistical Inference.* New York: Wiley, 1971.

[30] D. Blackwell and M. Girschik, *Theory of Games and Statistical Decisions.* New York: Wiley, 1970.

[31] C. W. Baum and V. V. Veeravalli, "A sequential procedure for multi-hypothesis testing," *IEEE Trans. Inform. Theory.* vol. IT-40, No. 6, pp. 1994-2007, November 1994.

[32] V. P. Dragalin, A. G. Tartakovsky and V. V. Veeravalli, "Multi-hypothesis sequential probability ratio tests – part I: asymptotic optimality," *IEEE Trans. Inform. Theory.* vol. IT-45, No. 7, pp. 2448-2461, November 1999.

[33] L. A. Shepp, J. K. Wolf, A. D. Wyner and J. Ziv, "Binary communication over the Gaussian channel using feedback with a peak energy constraint," *IEEE Trans. Inform. Theory.* vol. IT-15, No. 4, pp. 476-478, July 1969.