On sphere–packing bound, capacity and related results for Poisson channel *

Marat V. Burnashev †and Yury A. Kutoyants ‡

Abstract—A new approach for getting upperbounds for the capacity and the error exponent of the Poisson channel is proposed. It is based on a new sphere–packing lowerbound for the error probability. Strong converse for the coding theorem is also established.

Index Terms—Poisson channel, sphere–packing bound, linear bound, capacity, strong converse of coding theorem.

*The research described in this publication was made possible in part by Grants N 95–01–00136a from the Russian Fund for Fundamental Research and INTAS–94–469.

†M.V. Burnashev is with the Institute for Problems of Information Transmission, Russian Academy of Sciences, 19 Bolshoi Karetnii, 101447 Moscow, Russia.

‡Y.A. Kutoyants is with Laboratoire de Statistique et Processus, Universite du Maine, B.P. 535, 72017 Le Mans, France.

I. Introduction

We consider the Poisson (direct–detection photon) channel [5], [15] which is described as follows. The channel input is a waveform $S(t)$, $0 \le t \le T$, which satisfies

$$0 \le S(t) \le A < \infty\,, \tag{1.1}$$

where the parameter A is the peak power. The waveform $S(\cdot)$ defines a Poisson counting process $x(t)$ with "intensity" equal to $S(t)$ (we assume here that the "dark current" $\lambda_0 = 0$). Thus $x(t)$, $0 \le t \le T$, is a random process with independent increments such that $x(0) = 0$ and for any $0 \le t_1 \le t_2 \le T$

$$\Pr\,\{x(t_2) - x(t_1) = j\} = \frac{e^{-\Lambda}\Lambda^j}{j!}\,, \quad j = 0, 1, 2, \ldots\,,$$

where

$$\Lambda = \int_{t_1}^{t_2} S(\tau)\,d\tau\ .$$

Let $\mathcal{X}_T$ denote the space of right continious (i.e. $x(t+) - x(t) = 0$ for any $0 \le t \le T$) nondecreasing step functions $x(t)$, $0 \le t \le T$, with $x(0) = 0$ and $x(t) - x(t-) = 0$ or $x(t) - x(t-) = 1$. Then with probability 1 any path (trajectory) $x(t)$, $0 \le t \le T$, of a Poisson process with finite intensity belongs to $\mathcal{X}_T$.

A code with parameters (M, T, σ, ϵ) is defined by the following:

a) a set of M waveforms $S_m(t)$, $0 \le t \le T$, which satisfy the "peak power constraint" (1.1) and "the average power constraint"

$$\frac{1}{T}\int_0^T S_m(t)\,dt \le \sigma A\,, \quad m = 1, \ldots, M \tag{1.2}$$

(of course, $0 \le \sigma \le 1$);

b) a "decoder" mapping $D : \mathcal{X}_T \to \{1, 2, \ldots, M\}$;

c) the average error probability P_e satisfies condition

$$P_e = \frac{1}{M}\sum_{i=1}^{M} \Pr\left\{D(x_0^T) \ne m \,|\, S_m(.)\right\} \le \varepsilon\,.$$

The rate R of the code (in nats per second) is $(1/T)\ln M$.

Let A is fixed and $\sigma, R > 0$ be given. The channel capacity $C = C(\sigma)$ is the supremum of rates R such that for any $\varepsilon > 0$ there exists (for T sufficiently large) a code with parameters $(M, T, \sigma, \varepsilon)$ with $M \geq e^{RT}$. Define $P_e^*(M, \sigma, T)$ as the infimum of those P_e for which there exists a code with parameters (M, T, σ, P_e). For $0 \leq R < C$ define the reliability function of the channel by

$$E(R) = E(R, \sigma) = \limsup_{T \to \infty} \frac{1}{T} \ln \frac{1}{P_e^*([e^{RT}], \sigma, T)}. \tag{1.3}$$

The channel capacity C was found by Kabanov [5] (and generalized for the case $\lambda_0 > 0$ by Davis [2]). Certainly, the most important contribution in that area was made by Wyner in [15]–[16], where he found the *exact* form of the reliability function $E(R)$ (which gives as a corollary the capacity C).

In order to find $E(R)$ Wyner showed that in some sense it is possible to "reduce" the Poisson channel to a discrete channel with two inputs and two outputs (Z–type channel). After that applying known formulas for the discrete channel he found $E(R)$. Such *indirect* approach created some problems in the proof and the most difficult one was to show that in such reduction we loose negligibly in accuracy (the paper [16] is fully devoted to that question). Very lengthy proof in [16] (generally correct) contains also some inaccuracies (but correctable).

It should be noted that Poisson channels (and processes) became now quite popular due to the relation with optical channels and similar devices (see, for example, also [3], [8], [9]). For that reason it seems us that the Poisson channel deserves also some *direct* methods of investigation. It especially concerns the "sphere–packing" and similar *nonexistance* results.

It turned out that such *direct* methods are quite simple and give more accurate results (but need to use a bit more advanced tools than were used in [15]–[16]). In the paper, without reduction to a discrete channels, we directly get the upperbound for the reliability function (Ch. V, Theorem 1), which coincides with the lowerbound from [15] and appeared already in ([15]. In particular, we get first a traditional "sphere–packing" upperbound $E(R) \leq E_{sp}(R)$, which turns out to be exact when $R_{crit} \leq R \leq C$ (all values are defined below). Probably, it would be correct to call it a "volume–packing" bound, since there are no natural "spheres" here, but we just follow historic traditions. Clearly, it immediately gives the correct upperbound for the capacity C. Moreover, it allows us also to prove the strong converse of coding theorem (Ch. VI, Theorem 2). In other words, we show that if rate $R > C$ then the minimal possible error probability $P_e^*(M = e^{RT}, \sigma, T) \to 1$ with $T \to \infty$ (and even exponentially rapid with T).

For $0 \leq R < R_{crit}$ a better "straight–line" upperbound $E(R) \leq E_{sl}(R)$ is obtained using a well–known procedure [13]. We need to find only the value $E(R = +0)$ which in the Poisson case is very simple to do.

The sphere–packing and straight–line upperbounds establish together the final upperbound for the reliability function $E(R)$.

We intentionally consider here only the "no dark current" (i.e. $\lambda_0 = 0$) case. It allows us to concentrate essentially on methods avoiding too long expressions. The dark current case makes no difference in approach and will be considered (probably) separately.

The paper is organized as follows. In Section II we give some brief background (probably, not widely known) on Poisson processes and on distinguishability of probability measures. In Section III we show that optimal codes attain only extreme values 0 and A (that important fact was noticed by Wyner [15]–[16], but proof presented here is more general and much simpler). The sphere–packing bound is obtained in Section IV. The straight–line upperbound and final upperbound for the reliability function are presented in Section V. The strong converse for coding theorem is proved in Section VI.

Throughout the paper it is meant $0^0 = 1$.

II. Background

A. Radon–Nikodim derivative and expectations for Poisson process

As was mentioned already, any path of a Poisson process on $[0,T]$ belongs to $\mathcal{X}_T$. Moreover, providing the space $\mathcal{X}_T$ by some natural metrics, we get a measurable space $(\mathcal{X}_T, \mathcal{B}_T)$. Any Poisson process generates its own probability measure on $(\mathcal{X}_T, \mathcal{B}_T)$.

Let us consider now two Poisson processes on $[0,T]$ with intensities $p(t)$ and $q(t)$, generating on $(\mathcal{X}_T, \mathcal{B}_T)$ measures P and Q, respectively. Then the following formula for the Radon–Nikodim derivative is known (see, e.g. [8])

$$\ln \frac{dQ}{dP}(x_0^T) = \int_0^T \ln \frac{q(t)}{p(t)}\, dx(t) - \int_0^T [q(t) - p(t)]\, dt \,, \tag{2.1}$$

where the integral in the right side is the ordinary Stilties integral. ¿From (2.1) we can get [8] for any integrable function $f(t)$ the useful "expectation" formula

$$E_P \exp\left\{\int_0^T \ln f(t)\, dx(t)\right\} = \exp\left\{\int_0^T [f(t) - 1] p(t)\, dt\right\} \,, \tag{2.2}$$

where E_P means expectation with respect to measure P (or intensity $p(t)$).

In particular, we can get from (2.2) the formulas

$$\mu(s) = \ln E_P \left(\frac{dQ}{dP}\right)^s =$$

$$= \ln E_P \exp\left\{ s\int_0^T \ln \frac{q(t)}{p(t)}\, dx(t) - s\int_0^T [q(t) - p(t)]\, dt\right\} =$$

$$= \int_0^T \left\{\left[\left(\frac{q(t)}{p(t)}\right)^s - 1\right] p(t) - s\,[q(t) - p(t)]\right\} dt \,, \tag{2.3}$$

$$\rho(P,Q) = E_P \left(\frac{dQ}{dP}\right)^{1/2} = \exp\left\{-\frac{1}{2}\int_0^T \left(\sqrt{q(t)} - \sqrt{p(t)}\right)^2 dt\right\} \,.$$

B. On distinguishability of probability measures

Let on a measurable space $(\mathcal{X}, \mathcal{A})$ there are given two probability measures P and Q with densities $p(x)$, $q(x)$ with respect to a σ–finite measure $\mu(x)$. Denote

$$\|P-Q\| = \int_{\mathcal{X}} |p(x)-q(x)|\, d\mu(x)\,;\;\; \rho(P,Q) = \int_{\mathcal{X}} \sqrt{p(x)q(x)}\, d\mu(x)\,;$$

$$h^2(P,Q) = \frac{1}{2}\int_{\mathcal{X}} \left(\sqrt{p(x)} - \sqrt{q(x)}\right)^2 d\mu(x) = 1 - \rho(P,Q)\,.$$

The following relations are known [10], [11] :

$$h^2(P,Q) \le \frac{1}{2}\|P-Q\| \le h(P,Q)\sqrt{2-h^2(P,Q)} = \left[1-\rho^2(P,Q)\right]^{1/2}\,. \qquad (2.4)$$

Moreover, if we wish to test between P and Q using one observation then for the optimal decision rule minimizing the sum of error probabilities $\alpha+\beta$ we have [7] (see also [10], [11], [1])

$$\alpha + \beta = 1 - \frac{1}{2}\|P-Q\|\,. \qquad (2.5)$$

Therefore from (2.4)–(2.5) we get the bounds (the last one is interesting only for values of ρ close to 1, as we will have later)

$$\alpha+\beta \ge 1 - \sqrt{1-\rho^2} \ge 1 - \sqrt{2\ln\frac{1}{\rho}}\,. \qquad (2.6)$$

III. On approximation of Poisson process

We start this section with the following result.

Proposition 1. *Let $\{S_i(t),\, i=1,\ldots,M\}$ be some given set of signals, satisfying only conditions (1.1) and (1.2) and having some average error probability P_e. Then for any $\delta>0$ there exists some $\gamma=\gamma(\delta)>0$ such that for any set of points $\tau_0=0<\tau_1<\ldots<\tau_N=T$ with $\max_j(\tau_j-\tau_{j-1})\le\gamma$ (for example, equally spaced) we can construct another set of step signals, say $\{s_i(t),\, i=1,\ldots,M\}$, satisfying condition (1.2) and taking on only extreme values 0 and A on each interval (τ_{j-1},τ_j); $j=1,\ldots,N$, that will have the average error probability $\hat{P}_e\le P_e+\delta$.*

Remark. This important fact, significantly simplifying the analysis of Poisson channel, was noticed by Wyner [15], [16] (Theorem 2.1). But, probably, Aaron's strong desire to use only elementary methods made his proof extremely long and full of all possible tricks. As will be shown below that important fact rather easy follows from some more general viewpoint.

Proof. Our proof consists of two simple steps. First, we show that without essential loss of error performance we can "approximate" each signal $S_i(t)$ by some step (i.e. piecewise constant) signal $s_i(t)$ using any "rather dense" set of points $\tau_0=0<\tau_1<\ldots<\tau_N=T$. And second, we show that without loss of error performance we may assume that these step signals take on only extreme values 0 and A.

1) Clearly we consider only Lebesgue–measurable functions $0\le S(t)\le A$; $t\in[0,T]$. It follows from the standard theory of real functions [12] that for any $\varepsilon>0$ there exists some $\gamma=\gamma(\varepsilon)>0$ such that for any set of points $\tau_0=0<\tau_1<\ldots<\tau_N=T$ with $\max_j(\tau_j-\tau_{j-1})\le\gamma$ we can construct a step function, say $s^{(\varepsilon)}(t)$, (with $0\le s^{(\varepsilon)}(t)\le A$; $t\in[0,T]$) taking on only constant values on each interval (τ_{j-1},τ_j); $j=1,\ldots,N$, such that the following conditions will be satisfied

$$\int_0^T\left(\sqrt{S(t)}-\sqrt{s^{(\varepsilon)}(t)}\right)^2dt\le\varepsilon^2\,, \tag{3.1}$$

and

$$\int_0^T s^{(\varepsilon)}(t)\,dt\le\int_0^T S(t)\,dt\,. \tag{3.2}$$

Indeed, due to Luzin's theorem ([12], Ch. IV.4) for any $\delta > 0$ there exists a continious function $0 \le g^{(\delta)}(t) \le A\,;\; t \in [0,T]$, such that

$$\text{mes}\,\{t : S(t) \ne g^{(\delta)}(t)\} < \delta\,.$$

Moreover, the proof of that theorem can be modifyed in such a way that inequality similar to (3.2) will be fullfilled (in fact, we can always add ε to the right side of (3.2) and in such a "weaker" form inequality (3.2) will still be sufficient for our purposes).

Therefore it is sufficient to establish the existence of function $s^{(\varepsilon)}(t)$ for a continious function $S(t)$ and it is quite obvious.

Clearly, if we have any finite number M of signals $S_i(t)$, $i = 1, \ldots, M$, then for any $\varepsilon > 0$ there exists some $\gamma = \gamma(\varepsilon) > 0$ such that for any set of points $\tau_0 = 0 < \tau_1 < \ldots < \tau_N = T$ with $\max_j(\tau_j - \tau_{j-1}) \le \gamma$ we can construct step functions, say $s_i^{(\varepsilon)}(t)$, (with $0 \le s_i^{(\varepsilon)}(t) \le A\,;\; t \in [0,T]$) taking on only constant values on each interval (τ_{j-1}, τ_j); $j = 1, \ldots, N$, such that conditions similar to (3.1)–(3.2) will be satisfied simultaneously for all $i = 1, \ldots, M$.

Assume now that we observe the Poisson process $x(t)$, $0 \le t \le T$, which has the intensity equal to $S_i(t)$ or to $s_i^{(\varepsilon)}(t)$ and based on our observations we wish to decide which of intensities has really taken place. It follows from (2.6)and (2.3) that when testing these two hypotheses the minimal possible sum of error probabilities $\alpha + \beta$ satisfies the following lower bound

$$\alpha + \beta \ge 1 - \left[\int_0^T \left(\sqrt{S_i(t)} - \sqrt{s_i^{(\varepsilon)}(t)}\right)^2 dt\right]^{1/2} \ge 1 - \epsilon\,. \tag{3.3}$$

Therefore for small values of integral in (3.3) the sum $\alpha + \beta$ becomes close to 1 which means that it is "almost impossible" to distinguish between signals $S_i(t)$ and $s_i^{(\varepsilon)}(t)$.

Since for small ε we "can not" distinguish between signals $S_i(t)$ and $s_i^{(\varepsilon)}(t)$, in our problem of finding the best (M,T,σ)–code (which has the minimal average error probability) we may choose (loosing only ε in error performance) any "rather dense" set of points $\tau_0 = 0 < \tau_1 <, \ldots, < \tau_N = T$ and assume that all signals $S_i(t)$, $i = 1, \ldots, M$, take on constant values on each interval (τ_{j-1}, τ_j).

2) The next step of our approximation is quite simple. When we test between step signals $S_i(t)$, $i = 1, \ldots, M$, which take on constant values on

each interval (τ_{j-1}, τ_j) it is sufficient to observe only differences $x_j = x(\tau_j) - x(\tau_{j-1}),\ j = 1, \ldots, N$ (clearly, they form a sufficients statistics for testing those signals). So we may assume that we observe only "discrete-time" values $x(\tau_j),\ j = 0, 1, \ldots, N$. But in that case only the integral values

$$\int_{\tau_{i-1}}^{\tau_i} S_i(t)\, dt$$

are important and it is irrelevant how they are realized. In particular, we can always get the same integral values using only extreme values 0 and A for signals $S_i(t)$, $i = 1, \ldots, M,$. That completes the proof of Proposition 1. $\Box$.

IV. Capacity and sphere–packing bound

Let the signal $S_i(t)$ has the total energy $\mu_i AT$, $i = 1, \ldots, M$. As was shown already, we may assume that all signals (intensities) $S_i(t)$ take on only extreme values 0 and A. Denote

$$m_i = \{t \in [0,T] : S_i(t) = A\} \,, \quad m_{ij} = m_i \cap m_j \,, \tag{4.1}$$

and let $mes\,\{m\}$ denote the Lebesgue measure of the set m (all our sets of interest are measurable). Obviously, $mes\,\{m_i\} = \mu_i T$, $i = 1, \ldots, M$.

Clearly, we can assume that the observation space $\mathcal{X}$ is

$$\mathcal{X} = \sum_{n=0}^{\infty} \mathcal{X}_n \,,$$

where the set $\mathcal{X}_n$, $n \geq 1$, consists of all sequences $0 \leq x_1 \leq x_2 \ldots \leq x_n \leq T$ and each x_i represents the photon arriving moment.

Let some code of cardinality M and some decoding rule are chosen. Then to each message θ_i; $i = 1, \ldots, M$, correspond some regions $\mathcal{D}_{in} \subseteq \mathcal{X}_n$, $n = 1, 2, \ldots$ of making the decision in its favor. Surely,

$$\sum_{i=1}^{M} \mathcal{D}_{in} = \mathcal{X}_n \quad \text{and} \quad \mathcal{D}_{in} \cap \mathcal{D}_{jn} = \emptyset \,,\ i \neq j \ .$$

Clearly, if the message θ_i is transmitted then photons can arrive only during the time interval m_i. Moreover, the conditional distribution density $p(x_1, \ldots, x_n | \theta_i)$ is constant on $\mathcal{X}_n \cap m_i$ for any $n = 1, 2, \ldots$. Indeed, let $m_i = [0, mes\,\{m_i\}]$. Then each arriving moment x_{j+1} provided previously given moments $x_1, \ldots, x_j$ has the density

$$p(x_{j+1} | \theta_i, x_1, \ldots, x_j) = A e^{-A(x_{j+1} - x_j)} \ .$$

Therefore for any $(x_1, \ldots, x_n) \in \mathcal{X}_n \cap m_i$ we have

$$p(x_1, \ldots, x_n | \theta_i) = A^n \exp\left\{ -A \sum_{j=1}^{n} (x_{j+1} - x_j) - A(T - x_n) \right\} = A^n e^{-AT} \ .$$

We can introduce the "volume" $|\mathcal{X}_n|$ of the space $\mathcal{X}_n$

$$|\mathcal{X}_n| = \int \cdots \int_{0 \leq x_1 \leq \ldots \leq x_n \leq T} dx_1 \ldots dx_n = \frac{T^n}{n!} = v(n,T) \ . \tag{4.2}$$

Since $mes\{m_i\} = \mu_i T$ and the density $p(x_1, \ldots, x_n|\theta_i)$ is concentrated and constant on the set $\mathcal{X}_n \cap m_i$, a similar volume $|\mathcal{X}_n \cap m_i|$ is equal to

$$|\mathcal{X}_n \cap m_i| = v(n, \mu_i T) = \mu_i^n v(n, T) \ .$$

Denote by $|\mathcal{D}_{in}|$ the "volume" of decision set $\mathcal{D}_{in}$. Surely

$$\sum_{i=1}^{M} |\mathcal{D}_{in}| = |\mathcal{X}_n| = v(n, T) \ .$$

Therefore if our observation belongs to the set $\mathcal{X}_n$ and the message θ_i was transmitted then the conditional error probability $P(e|\theta_i, \mathcal{X}_n)$ can be lowerbounded as

$$P(e|\theta_i, \mathcal{X}_n) \geq \left[1 - \frac{|\mathcal{D}_{in}|}{v(n, \mu_i T)}\right]_+ = \left[1 - \frac{|\mathcal{D}_{in}|}{\mu_i^n v(n, T)}\right]_+ \ ,$$

where $[z]_+ = \max\{0, z\}$. Since

$$P(\mathcal{X}_n|\theta_i) = \frac{(\mu_i AT)^n}{n!} e^{-\mu_i AT} \ , \ n = 0, 1, \ldots \ ,$$

we get for the conditional error probability $P(e|\theta_i)$

$$P(e|\theta_i) \geq e^{-\mu_i AT} \sum_{n=0}^{\infty} \frac{(AT)^n}{n!} \left[\mu_i^n - \frac{|\mathcal{D}_{in}|}{v(n, T)}\right]_+ \ .$$

Then the average error probability $P_e(M, \{\mu_i\}, \{\mathcal{D}_{in}\})$ is lowerbounded by

$$P_e(M, \{\mu_i\}, \{\mathcal{D}_{in}\}) \geq \frac{1}{M} \sum_{i=1}^{M} e^{-\mu_i AT} \sum_{n=0}^{\infty} \frac{(AT)^n}{n!} \left[\mu_i^n - \frac{|\mathcal{D}_{in}|}{v(n, T)}\right]_+ =$$

$$= \frac{1}{M} \sum_{n=0}^{\infty} \frac{(AT)^n}{n!} \sum_{i=1}^{M} e^{-\mu_i AT} \left[\mu_i^n - \frac{|\mathcal{D}_{in}|}{v(n, T)}\right]_+ \ .$$

Denoting $d_{in} = |\mathcal{D}_{in}|/v(n, T)$ we get

$$P_e(M, \{\mu_i\}, \{\mathcal{D}_{in}\}) \geq \frac{1}{M} \sum_{n=0}^{\infty} \frac{(AT)^n}{n!} \sum_{i=1}^{M} e^{-\mu_i AT} \left[\mu_i^n - d_{in}\right]_+ \ , \tag{4.3}$$

where values d_{in} satisfy constraints

$$\begin{array}{lll} 1) & d_{in} & \leq \mu_i^n\,;\ i = 1,\dots,M;\ n = 0,1,\dots \\ 2) & \sum_{i=1}^{M} d_{in} & \leq 1\,;\ n = 0,1,\dots \end{array}$$

Assume first that all signals have the same total energy μAT, i.e. $\mu_1 = \mu_2 = \ldots = \mu_M = \mu$. Then clearly optimal d_{in} are

$$d_{in} = \min\{1/M,\ \mu^n\}\,;\ i = 1,\dots,M;\ n = 0,1,\dots\,,$$

and we get from (4.3)

$$P_e(M,\mu,T) \geq \frac{1}{M} e^{-\mu AT} \sum_{n=0}^{\infty} \frac{(AT)^n}{n!} (M\mu^n - 1)_+ \tag{4.4}$$

Consider now the general case when $\mu_1 \leq \mu_2 \leq \ldots \leq \mu_M$. Then for given n the minimum over $\{d_{in}\}$ in the rightside of (4.3) is attained when the total mass of $\{d_{in}\}$ is distrubuted among the minimal μ_i's. Indeed, each small mass Δ given to d_{in} takes away the value $e^{-\mu_i AT}\Delta$ from the rightside of (4.3). Therefore it is the best to distribute the total mass among the minimal μ_i's. Denote

$$n_0 = \max\{n : \sum_{i=1}^{M} \mu_i^n > 1\}\,; \tag{4.7}$$

$$k(n) = \min\{k : \sum_{i=1}^{k} \mu_i^n > 1\}\,,\ n = 0,1,\dots,n_0\,. \tag{4.8}$$

Obviously, $k(n)$ is nondecreasing function of n and $k(n_0) \leq M$. Then optimal $\{d_{in}\}$ for $n = 0,1,\dots,n_0$, are

$$d_{in} = \begin{cases} \mu_i^n\,;\ i = 1,\dots,k(n)-2; \\ 1 - \sum_{i=1}^{k(n)-2} \mu_i^n\,;\ i = k(n)-1; \\ 0;\ i = k(n),\dots,M; \end{cases}$$

and we get from (4.3)

$$P_e(M,\{\mu_i\},\{\mathcal{D}_{in}\}) \geq$$

$$\geq \frac{1}{M} \sum_{n=0}^{n_0} \frac{(AT)^n}{n!} \left\{ e^{-\mu_{k(n)-1}AT} \left(\sum_{i=1}^{k(n)-1} \mu_i^n - 1 \right) + \sum_{i=k(n)}^{M} e^{-\mu_i AT} \mu_i^n \right\} \geq$$

$$\geq \frac{1}{M}\sum_{n=0}^{n_0}\frac{(AT)^n}{n!}\sum_{i=k(n)}^{M} e^{-\mu_i AT}\mu_i^n .$$

We formulate obtained result as

Proposition 2 (general sphere–packing bound). *1) The average error probability $P_e(M,\{\mu_i\},T)$ satisfyes the lowerbound*

$$P_e(M,\{\mu_i\},T) \geq \frac{1}{M}\sum_{n=0}^{n_0}\frac{(AT)^n}{n!}\sum_{i=k(n)}^{M} e^{-\mu_i AT}\mu_i^n , \tag{4.9}$$

where n_0 and $k(n)$ are defined in (4.7)–(4.8).
2) Moreover, if $\mu_1 = \mu_2 = \ldots = \mu_M = \mu$, then

$$P_e(M,\mu,T) \geq e^{-\mu AT}\sum_{n=0}^{\infty}\frac{(\mu AT)^n}{n!}\left(1-\frac{1}{M\mu^n}\right)_+ \tag{4.10}$$

Investigating the lowerbound (4.10) for various values of $M = e^{RT}$ we will get below the sphere–packing bound for reliability function and strong converse for coding theorem (see Chapter VI).

Reliability function. We investigate first the case when $\ln M = RT = rAT$ and moreover all signals have the same total energy μAT, $0 < \mu < 1$. Denote $P_e(R,\mu,T)$ the minimal possible average error probability for such M signals. Define also n_1 as

$$n_1 = \max\left\{n : M\mu^n \geq 2\right\} .$$

Obviously, n_1 is a well–defined value provided $M \geq 2$ and it satisfies the bounds

$$\ln\frac{\mu M}{2} \leq n_1 \ln\frac{1}{\mu} \leq \ln\frac{M}{2} .$$

Then taking in (4.10) only the n_1–th term we have

$$P_e(R,\mu,T) \geq \frac{1}{2}e^{-\mu AT}\frac{(\mu AT)^{n_1}}{n_1!} . \tag{4.11}$$

Using also Stirling formula

$$n! \leq \sqrt{2\pi n}\left(\frac{n}{e}\right)^n e^{1/(12n)} ,$$

we get from (4.11)

$$\lim_{T\to\infty} \frac{1}{T} \ln \frac{1}{P_e(R,\mu,T)} \le A\, f(r,\mu)\,, \tag{4.12}$$

where

$$f(r,\mu) = \mu - \frac{r}{\ln(1/\mu)} \ln \frac{e\mu \ln(1/\mu)}{r}\,. \tag{4.13}$$

Consider now the case when each of M signals may have its own total energy $\mu_i AT \le \sigma AT$, $i = 1,\ldots,M$. The reliability function $E(R,\sigma)$ for $R > 0$ does not depend on what kind of error probability (average or maximal) we do consider [4]. For that reason when we deal with function $E(R,\sigma)$ for $R > 0$ we may assume that all signals have the same total energy μAT with some $\mu \le \sigma$. Then we get from (4.12)–(4.13)

$$E(R = rA, \sigma) \le E_{sp}(R = rA,\sigma) = A \max_{0\le\mu\le\sigma} f(r,\mu)\,. \tag{4.14}$$

For a given $0 < \mu < 1$ it is easy to check that $f(r = \mu\ln(1/\mu),\mu) = 0$; $f(r,\mu) > 0$ for $0 \le r < \mu\ln(1/\mu)$; and $f(r,\mu) < 0$ for $r > \mu\ln(1/\mu)$.

Since obviously we should have $E(R,\sigma) > 0$ for any $R < C(\sigma)$, we may conclude that $C(\sigma) \le A\max_{0\le\mu\le\sigma} \mu\ln(1/\mu)$. Opposite inequality for $C(\sigma)$ is also valid [5], [15]. For that reason

$$C(\sigma) = A \max_{0\le\mu\le\sigma} \mu\ln(1/\mu) = \begin{cases} A\sigma\ln(1/\sigma)\,, & \text{if} \quad \sigma \le 1/e\,; \\ A/e\,, & \text{if} \quad \sigma > 1/e\,; \end{cases} \tag{4.15}$$

and therefore we consider here only values $0 \le R < C(\sigma)$.

Notice that if there are no constraints on the choice of μ then maximum of $f(r,\mu)$ is attained for $\mu \ge 1/e$. Indeed, there are always two values of μ giving the same value of $\mu\ln(1/\mu)$ and it is easy to see that among them we should choose the larger one.

In order to find the maximum of $f(r,\mu)$ from (4.14) it is convinient to use a familiar parametrization relating r and μ through the parameter $\rho \ge 0$

$$r = \mu^{1+\rho} \ln\frac{1}{\mu}\,,\ \ \rho \ge 0\,. \tag{4.16}$$

Then we have

$$f(r,\mu) = f(r,\rho) = \mu - \mu^{1+\rho} - \rho r\,,$$

where $\rho = \rho(r,\mu)$ is defined by relation (4.16). Moreover

$$\frac{df(r,\mu)}{d\mu} = \frac{\partial f(r,\mu)}{\partial \mu} + \frac{\partial f(r,\mu)}{\partial \rho}\rho'_\mu = 1-(1+\rho)\mu^\rho\,, \tag{4.17}$$

$$\rho'_\mu = \frac{(1+\rho)\ln(1/\mu)-1}{\mu\ln^2\mu}\,.$$

Notice that if $\sigma \le 1/e$ then

$$\frac{df(r,\mu)}{d\mu} = 1-(1+\rho)\mu^\rho \ge 1-(e\mu)^\rho \ge 0$$

Therefore if $\sigma \le 1/e$, then maximum of $f(r,\mu)$ is attained for $\mu = \sigma$.

Consider now the case $\sigma > 1/e$ (and therefore $C(\sigma) = 1/e$). Assume first that $\sigma = 1$ (i.e. there are no constraints on the choice of μ). Then it is easy to check that for a given $0 < r \le 1/e$ parameter ρ varies between $\rho = 0$ (in that case $r = \mu\ln(1/\mu)\,, \mu > 1/e,$) and $\rho = 1/(re) - 1$ (in that case $\mu = e^{-er}$). In both extreme cases $f(r,\rho) = 0$. Therefore function $f(r,\mu) = f(r,\rho)$ attains its maximum in the "internal" point when

$$1-(1+\rho)\mu^\rho = 0$$

or

$$\mu = \frac{1}{(1+\rho)^{1/\rho}}\,. \tag{4.18}$$

Let now $1/e < \sigma < 1$. Then it is clear that optimal $\mu = \mu(\rho)$ is given by formula

$$\mu(\rho) = \min\left\{\sigma\,,\ \frac{1}{(1+\rho)^{1/\rho}}\right\}\,. \tag{4.19}$$

In a result, we get that if $\sigma > 1/e$, then for any r optimal value of $\mu = \mu(\rho)$ is defined as the solution of equations (4.16) and (4.18). In order to be accurate it is left to show that those equations have the unique solution. In order to do that it is sufficient to show that $r'_\rho < 0,\ \rho > 0$. Indeed,

$$\mu'_\rho = \frac{\mu}{\rho^2(1+\rho)}\left((1+\rho)\ln(1+\rho)-\rho\right)$$

Therefore

$$\mu^{-(1+\rho)}\frac{dr}{d\rho} = -\ln^2\mu + \mu^{-1}[(1+\rho)\ln(1/e)-1]\mu'_\rho =$$

$$= -\ln^2\mu + \frac{1}{\rho^2(1+\rho)}[(1+\rho)\ln(1/\mu) - 1][(1+\rho)\ln(1+\rho) - \rho] =$$

$$= -\frac{\ln^2(1+\rho)}{\rho^2} + \frac{1}{\rho^3(1+\rho)}[(1+\rho)\ln(1+\rho) - \rho]^2 =$$

$$= \frac{(1+\rho)[\rho - \ln(1+\rho)]^2 - \rho^3}{\rho^3(1+\rho)} .$$

Now since

$$0 < \rho - \ln(1+\rho) < \frac{\rho^2}{1+\rho} , \quad \rho > 0 ,$$

we get finally

$$\frac{dr(\rho)}{d\rho} < 0 , \quad \rho > 0 .$$

That completes our analysis of the maximum of $f(r,\mu)$ from (4.14) and we can combine obtained results as

Proposition 3 (sphere–packing bound). *Let* $M = e^{RT}$, $0 \le R < C(\sigma)$. *Then*

$$E(R,\sigma) \le E_{sp}(R,\sigma) = A\mu - A\mu^{1+\rho} - R\rho , \tag{4.20}$$

where:
1) if $\sigma \le 1/e$ *then*

$$\mu = \sigma , \ \rho = \frac{1}{\ln(1/\sigma)} \ln\frac{A\ln(1/\sigma)}{R} - 1 ;$$

2) if $\sigma > 1/e$ *then* ρ *and* $\mu = \mu(\rho)$ *are defined as unique solutions of equations (4.16) and (4.18).*

V. Straight–line upperbound

As usual the sphere–packing upperbound $E(R) \leq E_{sp}(R)$ from Ch. IV gives optimal result when $R_{crit} \leq R \leq C$, where the critical rate R_{crit} is defined by the value $\rho = 1$, i.e. due to (4.16) and (4.19)

$$R_{crit} = A\mu^2(1) \ln(1/\mu(1)) \,, \quad \mu(1) = \min\{\sigma, 1/2\} \,. \tag{5.1}$$

In the interval $0 \leq R < R_{crit}$ a better "straight–line" upperbound can be obtained through another well–known procedure ([13], Theorem 4).

We find first the value $E(R = +0)$. After that we draw a linear function from $E(+0)$ tangent to $E_{sp}(R)$ (that linear function has the slope equal to -1). Then we have

$$E(+0) - R = E_{sl}(R) < E_{sp}(R) \,; \quad 0 \leq R < R_{crit} \,. \tag{5.2}$$

Remark. In order to be accurate before applying Theorem 4 from [13] we should check that our "sphere–packing" bound has the same meaning as "sphere–packing" bound from [13] (i.e. it can be used for lowerbounding the error probability for list decoding). Clearly, it is so since our "sphere–packing" bound is based purely on "volume–packing" arguments.

In order to find $E(+0)$ we use the "minimal signal distance" arguments. Since $M \to \infty$ when $T \to \infty$ (but $(\ln M)/T \to 0$) and we are interested in logarithmic asymptotics of $P_e(M, T)$, we may assume that all signals have the same energy μAT, $\mu \leq \sigma$.

Assume that we test between two equiprobable signals having energy μAT and supports m_i, m_j, respectively. Then when making a decision we have "uncertainty" only if there are no photons on $m_i \setminus m_j$ and therefore for sum of error probabilites we have

$$2P_e(m_i, m_j) = P(e|\, m_i) + P(e|\, m_j) \geq$$

$$\geq \Pr\{\text{no photons on}\, m_i \setminus m_j |\, m_i\} =$$

$$= e^{-mes\,\{m_i \setminus m_j\} AT} = e^{-\mu AT + mes\,\{m_{ij}\} AT} \,.$$

When we have M signals then for the average error probability $P_e(M,T)$ we have

$$P_e(M,T) \geq \frac{1}{M(M-1)} \sum_{i=1}^{M} \sum_{j \neq i} P_e(m_i, m_j) \geq$$

$$\geq \frac{1}{2M(M-1)} \sum_{i=1}^{M} \sum_{j \neq i} \exp\{-\mu AT + \mathrm{mes}\{m_{ij}\} AT\} \geq$$

$$\geq \frac{1}{2} \exp\left\{-\mu AT + \frac{A}{M(M-1)} \sum_{i=1}^{M} \sum_{j \neq i} \mathrm{mes}\{m_{ij}\}\right\} . \tag{5.3}$$

Now we use the elementary result.

Lemma 1. *Let* $\{m_1, \ldots, m_M\}$ *be a collection of measurable sets on the interval* $[0,T]$*, such that* $\mathrm{mes}\{m_i\} = \mu_i T$; $i = 1, \ldots, M$ *. Then*

$$\sum_{i=1}^{M} \sum_{j \neq i} \mathrm{mes}\{m_{ij}\} \geq T \left(\sum_{i=1}^{M} \mu_i\right)^2 - T \left(\sum_{i=1}^{M} \mu_i\right) .$$

In particular, if $\mu_1 = \ldots = \mu_M = \mu$ *, then*

$$\frac{1}{M(M-1)} \sum_{i=1}^{M} \sum_{j \neq i} \mathrm{mes}\{m_{ij}\} \geq \mu^2 T - \frac{\mu(1-\mu)T}{M-1} . \tag{5.4}$$

Proof. Let $\chi_i(t)$ be the indicator function of the set m_i. Then

$$\mathrm{mes}\{m_{ij}\} = \int_0^T \chi_i(t)\, \chi_j(t) dt = (\chi_i, \chi_j) ,$$

and

$$\sum_{i=1}^{M} \sum_{j \neq i} \mathrm{mes}\{m_{ij}\} = \sum_{i=1}^{M} \sum_{j=1}^{M} (\chi_i, \chi_j) - T \sum_{i=1}^{M} \mu_i =$$

$$= \int_0^T \sum_{i=1}^{M} \sum_{j=1}^{M} \chi_i(t) \chi_j(t) dt - T \sum_{i=1}^{M} \sigma_i = \int_0^T \left[\sum_{i=1}^{M} \chi_i(t)\right]^2 dt - T \sum_{i=1}^{M} \mu_i \geq$$

$$\geq \frac{1}{T} \left[\int_0^T \sum_{i=1}^{M} \chi_i(t) dt\right]^2 - T \sum_{i=1}^{M} \mu_i = T \left(\sum_{i=1}^{M} \mu_i\right)^2 - T \left(\sum_{i=1}^{M} \mu_i\right) ,$$

from where Lemma 1 follows. $\square$

Now we get from (5.3)–(5.4)

$$P_e(M,T) \geq \frac{1}{2}\exp\left\{-\mu(1-\mu)AT\left(1+\frac{1}{M-1}\right)\right\}$$

and therefore $E(+0) \leq A\sup_{\mu\leq\sigma}\ \mu(1-\mu)$. Using well–known "random coding with expurgation" arguments [4] it is easy to obtain the opposite inequality as well. Therefore

$$E(+0) = A\sup_{\mu\leq\sigma}\ \mu(1-\mu) = A\left\{\begin{array}{cl} \sigma(1-\sigma)\,, & \text{if } 0<\sigma\leq 1/2\,; \\ 1/4\,, & \text{if } 1/2\leq\sigma\leq 1\,. \end{array}\right. \tag{5.5}$$

We can combain results from Ch. IV–V in

Theorem 1. *The reliability function $E(R)$ satisfyes the upperbound*

$$E(R) \leq \left\{\begin{array}{c} E(+0)-R\,,\ +0\leq R\leq R_{crit}\,; \\ E_{sp}(R)\,,\ R_{crit}\leq R\leq C\,, \end{array}\right. \tag{5.6}$$

where C, R_{crit}, $E(+0)$ and $E_{sp}(R)$ are defined in (4.15), (5.1), (5.5) and (4.20), respectively.

Remark. We should emphasize here once more that upperbound (5.6) appeared already in [15] and we present here it's independent *direct* proof. Moreover, due to [15] this bound is *exact* (i.e. there is *equality* in (5.6)).

VI. Strong converse of coding theorem

We assume that the rate R is close or greater than $C(\sigma)$ and establish the strong converse of coding theorem. For that purpose we use the convinient analytic result (see proof in Appendix).

Lemma 2. *For any $n \geq 1$ and $\lambda > 0$ the following inequality holds*

$$e^{-\lambda} \sum_{k=0}^{n} \frac{\lambda^k}{k!} \geq \Phi\left(\frac{n-\lambda}{\sqrt{n}}\right), \tag{6.1}$$

where $\Phi(x) = (2\pi)^{-1/2} \int_{-\infty}^{x} e^{-u^2/2} du$.

We consider first the case when all signals have the same total energy μAT. Then from (4.10) and (6.1) we have for the error probability of any such code

$$P_e(R,\mu,T) \geq \max_n \left\{ \left(1 - \frac{1}{M\mu^n}\right) \Phi\left(\frac{n-\mu AT}{\sqrt{n}}\right) \right\}. \tag{6.2}$$

Assume now that $\ln M = \mu \ln(1/\mu) AT + a\sqrt{AT}$; $0 < \mu < 1$, where a is an arbitrary constant. Then we put (constant b is chosen below)

$$n = \frac{\mu \ln(1/\mu) AT + b\sqrt{AT}}{\ln(1/\mu)},$$

and get from (6.2)

$$P_e(R,\mu,T) \geq \left(1 - e^{-(a-b)\sqrt{AT}}\right) \Phi\left(\frac{b}{\sqrt{\mu \ln^2(1/\mu) - b(AT)^{-1/2} \ln \mu}}\right). \tag{6.3}$$

Now choosing $b = a - (AT)^{-1/4}$ we get from (6.3)

$$\liminf_{T\to\infty} P_e(R,\mu,T) \geq \Phi\left(\frac{a}{\sqrt{\mu \ln^2(1/\mu)}}\right). \tag{6.4}$$

Assume now that $R = rA > A\mu \ln(1/\mu)$; $0 < \mu < 1$. Then we can continue (6.2) as follows

$$P_e(R,\mu,T) \geq 1 - \min_n \left\{ \frac{1}{M\mu^n} + \Phi\left(\frac{\mu AT - n}{\sqrt{n}}\right) \right\} \geq$$

$$\geq 1 - \min_n \left[\exp\left\{ -rAT + n \ln \frac{1}{\mu} \right\} + \exp\left\{ -\frac{(n-\mu AT)^2}{2n} \right\} \right], \tag{6.5}$$

where we used the known inequality

$$\Phi(-x) \leq 0.5\, e^{-x^2/2}, \quad x \geq 0 .$$

Now we put

$$n = \frac{aAT}{\ln(1/\mu)}, \quad \mu \ln(1/\mu) \leq a < r,$$

and choose a such that both exponents in the right side of (6.5) become equal. Then we have

$$P_e(R, \mu, T) \geq 1 - 2\, e^{-AT(r-a)},$$

where

$$a = \frac{\ln(1/\mu)}{1 + 2\ln(1/\mu)} \left[\mu + r + \sqrt{(\mu+r)^2 - \mu^2 \left(1 + 2\ln(1/\mu)\right)} \right].$$

After some standard algebra we can get

$$r - a \geq \frac{\left(r - \mu \ln(1/\mu)\right)^2}{2r \ln(e/\mu)}.$$

In a result we get from (6.5)

$$P_e(R, \mu, T) \geq 1 - 2\exp\left\{ -\frac{\left(R - A\,\mu \ln(1/\mu)\right)^2 T}{2R\, \ln(e/\mu)} \right\}, \; R > A\mu \ln(1/\mu). \tag{6.6}$$

Remark. Inequalities like (6.6) have a sense only when the total energy μAT is sufficiently large.

We can formulate just obtained results in the following form.

Proposition 4. *Let all signals have the same total energy μAT.*
1) If $M = \exp\{\mu \ln(1/\mu) AT + a\sqrt{AT}\}$; $0 < \mu < 1$, then lowerbound (6.4) holds.
2) If $M = e^{RT}$ with $R > A\,\mu \ln(1/\mu)$ then lowerbound (6.6) holds.

Consider now the general case of nonequal energy $M = e^{RT}$ signals with $R > C(\sigma)$. We try first a very simple approach. Enlarging the time interval $[0, T]$ to $[0, T(1+\sigma)]$ we can reduce the problem to the case with equal energy σAT signals, but on the interval $[0, T(1+\sigma)]$. That case was already considered in (6.5)–(6.6), where we should replace r by $r/(1+\sigma)$ and μ by $\sigma/(1+\sigma)$. Then for the minimal possible average error probability $P_e^*(R, \sigma, T)$ and $R > A\,\sigma \ln((1+\sigma)/\sigma)$ we get the lowerbound

$$P_e^*(R, \sigma, T) \geq 1 - 2\exp\left\{-\frac{(R - A\,\sigma\ln((1+\sigma)/\sigma))^2\, T}{2R\,\ln\left(e(1+\sigma)/\sigma\right)}\right\}.$$

We will use this bound when $R \geq C(\sigma) + 2A\,\sigma\ln(1+\sigma)$. In that case it can be reduced to the form

$$P_e^*(R, \sigma, T) \geq 1 - 2\exp\left\{-\frac{(R - C(\sigma))^2\, T}{8R\ln(e(1+\sigma)/\sigma)}\right\}. \tag{6.7}$$

Consider now the case $C(\sigma) < R < C(\sigma) + 2A\,\sigma\ln(1+\sigma)$. We repeat essentially the same arguments, but a bit more accurately. Choose on $[0, \sigma]$ some points $0 = \nu_0 < \nu_1 < \ldots < \nu_N = \sigma$ and denote $\nu_i - \nu_{i-1} = \delta_i$; $i = 1, \ldots, N$. Let M_i be the number of messages in the interval $(\nu_{i-1}, \nu_i]$; $i = 1, \ldots, N$. Each M_i messages we consider as a separate code and we will lower-bound the average error probability for those N codes assuming that i–th code has prior probability M_i/M. Moreover, without loss of accuracy, we can assume that all M_i messages have the same total energy $\nu_i AT$, but on the interval $[0, T(1+\delta_i)]$; $i = 1, \ldots, N$.

If we put $\nu_i' = \nu_i/(1+\delta_i)$ then we come to the case with M_i equal energy signals on $[0, T(1+\delta_i)]$ with common ν_i'. Then we get similar to (6.2)

$$P_e^*(R, \sigma, T) \geq \sum_{i=1}^{N} \frac{M_i}{M} \max_n \left\{\left(1 - \frac{1}{M_i \nu_i'^n}\right)\Phi\left(\frac{n - \nu_i' AT(1+\delta_i)}{\sqrt{n}}\right)\right\} =$$

$$= \sum_{i=1}^{N} \frac{M_i}{M} \max_n \left\{\left(1 - \frac{(1+\delta_i)^n}{M_i \nu_i^n}\right)\Phi\left(\frac{n - \nu_i AT)}{\sqrt{n}}\right)\right\}.$$

Choose n_i, $i = 1, \ldots, N$, such that

$$\frac{n_i - \nu_i AT}{\sqrt{n_i}} \geq 2\sqrt{BAT},$$

where constant $B \geq 0$ will be chosen later. Then

$$\sqrt{n_i} \geq \left(\sqrt{B+\nu_i}+\sqrt{B}\right)\sqrt{AT}$$

and we have

$$P_e^*(R,\sigma,T) \geq \frac{\Phi(2\sqrt{BAT})}{M}\sum_{i=1}^{N}\left[M_i - \left(\frac{1+\delta_i}{\nu_i}\right)^{n_i}\right]_+ \geq$$

$$\geq \Phi(2\sqrt{BAT})\left[1-\frac{1}{M}\sum_{i=1}^{N}\left(\frac{1+\delta_i}{\nu_i}\right)^{n_i}\right]_+ \geq$$

$$\geq 1-\left\{\Phi(-2\sqrt{BAT})+\frac{1}{M}\sum_{i=1}^{N}\left(\frac{1+\delta_i}{\nu_i}\right)^{n_i}\right\}.$$

Therefore for $B \geq 0$ we get

$$P_e^*(R,\sigma,T) \geq 1-\min_{B\geq 0}\left\{\frac{1}{2}e^{-2BAT}+\frac{1}{M}\sum_{i=1}^{N}\left(\frac{1+\delta_i}{\nu_i}\right)^{n_i}\right\}. \tag{6.8}$$

We choose now points $0=\nu_0<\nu_1<\ldots<\nu_N=\mu$ in such a way that the following condition will be satisfyed

$$\left(\frac{1+\delta_i}{\nu_i}\right)^{n_i} \leq Me^{-2BAT}\,; \quad i=1,\ldots,N,$$

or equaly

$$\left(\sqrt{B+\nu_i}+\sqrt{B}\right)^2 \ln\frac{1+\delta_i}{\nu_i} \leq r-2B\,; \quad i=1,\ldots,N\,. \tag{6.9}$$

Then we have from (6.8)

$$P_e^*(R,\sigma,T) \geq 1-(N+1)\,e^{-2BAT}\,, \tag{6.10}$$

and it is left to evaluate possible constants $B>0$ and N.

We will not try to find the maximal possible coefficient $B=B(R,\sigma)$ and will prefer to get a compact significant expressions. Then using in (6.9) simple inequalities $\sqrt{B+\nu} \le \sqrt{B} + \sqrt{\nu}$ and $\sqrt{r-2B} \ge \sqrt{r} - \sqrt{2B}$, we get that it is sufficient to have fullfilled inequalities

$$\sqrt{B} \le \frac{r - \nu_i \ln((1+\delta_i)/\nu_i)}{4\sqrt{r \ln(e(1+\delta_i)/\nu_i)}}\,; \quad i = 1, \ldots, N\,.$$

Keeping in mind (6.7) we choose

$$\sqrt{B} = \frac{r - C(\sigma)/A}{4\sqrt{2r \ln(e(1+\sigma)/\sigma)}}\,. \tag{6.11}$$

Therefore we should show that it is possible to choose points $0 = \nu_0 < \nu_1 < \ldots < \nu_N = \sigma$ such that the following conditions will be satisfyed for $i = 1, \ldots, N$,

$$\frac{r - C(\mu)/A}{\sqrt{2\ln(e(1+\sigma)/\sigma)}} \le \frac{r - \nu_i \ln((1+\delta_i)/\nu_i)}{\sqrt{\ln(e(1+\delta_i)/\nu_i)}}\,. \tag{6.12}$$

We start with the case $i = 1$ and put $\nu_1 = \sigma/5$. Then it is easy to check that for any $0 < \sigma \le 1$ we have

$$r - \nu_1 \ln \frac{1+\nu_1}{\nu_1} \ge r - C(\sigma)\,.$$

Moreover for any $\nu \ge \sigma/5$ we have

$$2\ln(e(1+\sigma)/\sigma) \ge \ln(e(1+\nu)/\nu)\,, \quad \nu \ge \sigma/5\,.$$

Therefore condition (6.12) is fullfilled for $i = 1$ and moreover we left to satisfy only condition

$$\frac{r - C(\sigma)/A}{\sqrt{2}} \le r - \nu_i \ln \frac{1+\delta_i}{\nu_i}\,, \quad i = 2, \ldots, N\,,$$

or, if we denote $r - C(\sigma)/A = \varepsilon$,

$$\nu_i \ln \frac{1+\delta_i}{\nu_i} - C(\sigma)/A \le \frac{2-\sqrt{2}}{2}\varepsilon\,, \quad i = 2, \ldots, N\,.$$

Since $C(\sigma) \geq \nu_i \ln(1/\nu_i)$ we get that it is sufficient to satisfy conditions

$$\nu_i \ln(1+\delta_i) \leq \frac{2-\sqrt{2}}{2}\varepsilon\,, \quad i = 2, \ldots, N\,.$$

Therefore we can put

$$\delta_i = \delta = \frac{(2-\sqrt{2})\varepsilon}{2\sigma}\,, \quad i = 2, \ldots, N\,.$$

Then all conditions (6.12) will be satisfyed and

$$N \leq \frac{\sigma}{\delta} + 2 = \frac{(2+\sqrt{2})\sigma^2}{\varepsilon} + 2\,.$$

Finally we get from (6.10)

$$P_e^*(R,\sigma,T) \geq$$
$$\geq 1 - \left(\frac{4\sigma^2 A}{(R - C(\sigma))} + 3\right) \exp\left\{-\frac{(R-C(\sigma))^2 T}{32 R \ln(e(1+\sigma)/\sigma)}\right\}\,. \tag{6.13}$$

We can formulate just obtained result in the following form.

Theorem 2 (strong converse of coding theorem).
If $R > C(\sigma)$ then for the minimal possible error probability $P_e^*(R,\sigma,T)$ lowerbound (6.13) holds.

Acknowledgement

The paper was started and partially completed while the first author was visiting the University of Maine (Le Mans, France). The authors wish to thank the University for providing conditions for joint research

Appendix

Proof of Lemma 2. Consider first the case $\lambda \le n$. Integrating by parts we can get the known formula

$$e^{-\lambda} \sum_{k=n+1}^{\infty} \frac{\lambda^k}{k!} = \frac{1}{n!} \int_0^{\lambda} t^n e^{-t}\, dt\,.$$

Notice that for function

$$f(t) = \ln \left(t^n e^{-t} \right) = n \ln t - t$$

we have

$$f'(t) = \frac{n}{t} - 1\,;\ f''(t) = -\frac{n}{t^2}\,;\ f^{(3)}(t) = \frac{2n}{t^3}\,.$$

Then due to Taylor formula we have for $t \le \lambda \le n$

$$\ln(t^n e^{-t}) \le n \ln \left(\frac{n}{e} \right) - \frac{1}{2n}(t-n)^2$$

and

$$e^{-\lambda} \sum_{k=n+1}^{\infty} \frac{\lambda^k}{k!} \le \frac{1}{n!} \left(\frac{n}{e} \right)^n \int_0^{\lambda} \exp\left\{ -\frac{(t-n)^2}{2n} \right\} dt \le$$

$$\le \frac{1}{\sqrt{2\pi}} \int_{\frac{n-\lambda}{\sqrt{n}}}^{\sqrt{n}} e^{-u^2/2}\, du = \Phi\left(\sqrt{n}\right) - \Phi\left(\frac{n-\lambda}{\sqrt{n}} \right)\,.$$

Therefore

$$e^{-\lambda} \sum_{k=0}^{n} \frac{\lambda^k}{k!} = 1 - e^{-\lambda} \sum_{k=n+1}^{\infty} \frac{\lambda^k}{k!} \ge \Phi\left(-\sqrt{n}\right) + \Phi\left(\frac{n-\lambda}{\sqrt{n}} \right)\,,$$

that gives (6.1) for $\lambda \le n$.

Consider now the case $\lambda > n$. Then we use the formula

$$e^{-\lambda} \sum_{k=0}^{n} \frac{\lambda^k}{k!} = \frac{1}{n!} \int_{\lambda}^{\infty} t^n e^{-t}\, dt\,.$$

Now for function

$$f(t) = \ln \left(t^n e^{-t} \right) = n \ln t - t$$

we have

$$f(t) \geq f(\lambda) + \frac{(n-\lambda)}{\lambda}(t-\lambda) - \frac{n}{2\lambda^2}(t-\lambda)^2\,,\ t \geq \lambda$$

Using Stirling formula we get

$$e^{-\lambda}\sum_{k=0}^{n}\frac{\lambda^k}{k!} \geq \frac{\lambda^n e^{-\lambda}}{n!}\int_{\lambda}^{\infty}\exp\left\{-\frac{(\lambda-n)}{\lambda}(t-\lambda) - \frac{n}{2\lambda^2}(t-\lambda)^2\right\}dt =$$

$$= \frac{\lambda^{n+1}e^{-\lambda}}{n!\sqrt{n}}\int_{\frac{(\lambda-n)}{\sqrt{n}}}^{\infty}\exp\left\{-\frac{1}{2}u^2 + \frac{(\lambda-n)^2}{2n}\right\}dt \geq$$

$$\geq \frac{\lambda}{n}e^{-\lambda}\left(\frac{e\lambda}{n}\right)^n\Phi\left(\frac{n-\lambda}{\sqrt{n}}\right)\exp\left\{\frac{(\lambda-n)^2}{2n} - \frac{1}{12n}\right\}.$$

It is easy to check that

$$e^{-\lambda}\left(\frac{e\lambda}{n}\right)^n\exp\left\{\frac{(\lambda-n)^2}{2n}\right\} \geq 1\,,\ \lambda \geq n\,.$$

Therefore we get for $\lambda \geq n$

$$e^{-\lambda}\sum_{k=0}^{n}\frac{\lambda^k}{k!} \geq \frac{\lambda}{n}\Phi\left(\frac{n-\lambda}{\sqrt{n}}\right)e^{-1/(12n)}\,,$$

that gives (6.1) for $\lambda \geq ne^{1/(12n)}$.

Therefore it is left to prove (6.1) for $n < \lambda < ne^{1/(12n)}$. For that purpose we consider the difference

$$g(n,\lambda) = e^{-\lambda}\sum_{k=0}^{n}\frac{\lambda^k}{k!} - \Phi\left(\frac{n-\lambda}{\sqrt{n}}\right)$$

and show that $g'_{\lambda}(n,\lambda) \geq 0$ for $n \leq \lambda \leq ne^{-1/(12n)}$. Since $g(n,n) \geq 0$ it will complete the proof of (6.1).

We have

$$g'_{\lambda}(n,\lambda) = \frac{1}{\sqrt{2\pi n}}\exp\left\{-\frac{(\lambda-n)^2}{2n}\right\} - \frac{\lambda^n}{n!}e^{-\lambda}$$

Notice that using Gauss formula

$$n! \geq \sqrt{2\pi}\left(\frac{n+\frac{1}{2}}{e}\right)^{n+\frac{1}{2}}\exp\left\{-\frac{1}{24n+12}\right\}\,,$$

we can get

$$n! \geq \sqrt{2\pi n}\left(\frac{n}{e}\right)^n \exp\left\{\frac{1}{12n} - \frac{1}{24n^2}\right\} .$$

Using this lowerbound for $\lambda = n(1+x)$ we get that it is sufficient to establish the validity of inequality

$$\frac{x^2}{2} - x + \ln(1+x) - \frac{1}{12n^2} + \frac{1}{24n^3} \leq 0 ; \quad 0 \leq x \leq e^{1/(12n)} - 1 .$$

Since $\ln(1+x) \leq x - x^2/2 + x^3/3$, it is sufficient to check the inequality

$$\left[e^{1/(12n)} - 1\right]^3 \leq \frac{1}{4n^2} - \frac{1}{8n^3} , \quad n \geq 1 ,$$

which is easy to verify. $\square$.

References

[1] M. V. Burnashev, "Minimax detection of inaccurately known signal in the background of white Gaussian noise," *Theory of Prob. and Its Appl.*, vol. 24, No. 1, pp. 106 - 118, 1979.

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

[3] M. R. Frey, "Information capacity of the Poisson channel," *IEEE Trans. Inform. Theory,* vol. IT-37, pp. 244 - 256, March 1991.

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

[5] Y. M. Kabanov, "The capacity of a channel of the Poisson type," *Theory of Prob. and Its Appl.*, vol. 23, No. 1, pp. 143 - 147, 1978.

[6] A. N. Kolmogorov and S. V. Fomin, *Elements of Theory of Functions and Functional Analyses.* Moscow: Nauka, 1968.

[7] C. H. Kraft, "Some conditions for consistency and uniform consistency of statistical procedures," *Univ. California Publ. Statist.*, No. 2, pp. 125 - 142, 1955.

[8] Y. A. Kutoyants, *Parameter Estimation for Stochastic Processes.* Berlin: Heldermann, 1984.

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

[10] L. Le Cam, "Convergence of estimates under dimensionality restrictions," *The Annals of Statistics,* vol. 1, No. 1, pp. 38 - 53, 1973.

[11] L. Le Cam, *Asymptotic Methods in Statistical Decision Theory.* New York: Springer–Verlag, 1986.

[12] I. P. Natanson, *Theory of Real Variable Functions.* Moscow: Nauka, 1974.

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

[14] J. Wolfowitz, "The coding of messages subject to chance errors," *Illinois J. of Math.,* vol. 1, pp. 591 - 606, 1957.

[15] A. D. Wyner, "Capacity and error exponent for the direct detection photon channel - Part I," *IEEE Trans. Inform. Theory,* vol. IT-34, pp. 1449 - 1461, November 1988.

[16] A. D. Wyner, "Capacity and error exponent for the direct detection photon channel - Part II," *IEEE Trans. Inform. Theory,* vol. IT-34, pp. 1462 - 1471, November 1988.