On identification capacity of infinite alphabets or continuous time channels *

Marat V. Burnashev †

Abstract—Two kinds of channels are considered: 1) discrete time with additive noise channel, and 2) Poisson and white Gaussian (i.e. continuous time) channels. For the type 1 chanel there are given some sufficient conditions when Shannon and identification capacities coincide. It is shown that identification capacity of Poisson and Gaussian channels without bandwidth constraint is infinite. Contrary, in the case of white Gaussian channel with bandwidth constraint, its identification capacity coincides with Shannon capacity.

Index Terms—Identification, capacity, coding theorem, Poisson channel, Gaussian channel, bandwidth.

I. Introduction

We consider two kinds of channels. The first is a channel with independent additive noise

$$z_i = u_i + \xi_i \,, \quad i = 1, \ldots, n, \tag{1}$$

where $\mathbf{u} = u^n = (u_1, \ldots, u_n)$ and $\mathbf{z} = z^n = (z_1, \ldots, z_n)$ are channel input and output blocks, respectively, and $\xi_1, \ldots, \xi_n$ are i.i.d.r.v.'s with density $f(x)$, $x \in R^1$, with respect to Lebesque measure dx on R^1. It is assumed that channel input u^n satisfyes energy constraint

$$\sum_{i=1}^{n} u_i^2 \leq na^2 \,, \tag{2}$$

with a given $a > 0$. Denote that channel by $W(f, a)$ and by $C(f, a)$ its Shannon capacity. The second is a continuous time channel W^T when we are given some set $\mathcal{S}_T = \{S(t)\,,\ 0 \leq t \leq T\}$ (usually infinite) of possible input signals. According to properties of a channel W^T considered each input signal $S \in \mathcal{S}_T$ generates some probability distribution (measure) $Q_S = W^T S$ on the output space $\mathcal{X}_T$ of the channel. We limit ourselves to traditional white Gaussian and Poisson channels and investigate some their properties when $T \to \infty$.

*The research described in this publication was made possible in part by Grant N 98–01–04108 from the Russian Fund for Fundamental Research.

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

Notice that in both models input and output alphabets are infinite (even noncountable).

If $\mathbf{u} = u^n$ is channel's (1) input then by $Q_{\mathbf{u}} = W^{(n)}\mathbf{u}$ we denote the generated probability measure on the channel output R^n. Similarly, if in both models P is some probability distribution on channel input then by $Q_P = W^{(n)}P$ (or $Q_P = W^{(T)}P$) we denote the generated probability measure on the channel output R^n (or on $\mathcal{X}_T$).

Remind now some notions (cf. [1]).

Definition 1. *A collection $(\mathbf{u}_i, \mathcal{D}_i, i = 1, \ldots, M)$ of codeblocks $\mathbf{u}_i \in R^n$ and regions $\mathcal{D}_i \subseteq R^n$ for channel (1) is called an (M, n, δ) – dID–code (deterministic) if the following conditions are satisfied:*

$$Q_{\mathbf{u}_i}(\mathcal{D}_i) \geq 1 - \delta \quad \textit{and} \quad Q_{\mathbf{u}_i}(\mathcal{D}_j) \leq \delta \quad \textit{for any} \quad i \neq j\,.$$

A collection $(S_i, \mathcal{D}_i, i = 1, \ldots, M)$ of signals $S_i \in \mathcal{S}_T$ and regions $\mathcal{D}_i \subseteq \mathcal{X}_T$ for channel W^T is called an (M, T, δ) – dID–code (deterministic) if the following conditions are satisfied:

$$Q_{S_i}(\mathcal{D}_i) \geq 1 - \delta \quad \text{and} \quad Q_{S_i}(\mathcal{D}_j) \leq \delta \quad \text{for any} \quad i \neq j\,.$$

Definition 2. *A collection $(P_i, \mathcal{D}_i, i = 1, \ldots, M)$ of probability measures P_i on R^n (or on $\mathcal{S}_T$) and regions $\mathcal{D}_i$ on R^n (or on $\mathcal{X}_T$) is called an (M, n, δ) (or (M, T, δ)) – ID–code (randomized) if the following conditions are satisfied:*

$$Q_{P_i}(\mathcal{D}_i) \geq 1 - \delta \quad \text{and} \quad Q_{P_i}(\mathcal{D}_{S_j}) \leq \delta \quad \text{for any} \quad i \neq j\,.$$

The rate R of the dID (or ID)–code is $(1/n)\ln\ln M$ for channel (1) and $(1/T)\ln\ln M$ for channel W^T.

For channel W with Shannon capacity $C(W) < \infty$ deterministic identification capacity (dID–capacity) $C_{\mathrm{dID}}(W, \delta)$ and (randomized) identification capacity (ID–capacity) $C_{\mathrm{ID}}(W, \delta)$ are defined for any $0 < \delta < 1$ as maximal achievable rate R (when $n \to \infty$ or $T \to \infty$). It follows from those definitions that

$$C_{\mathrm{dID}}(W, \delta) \leq C_{\mathrm{ID}}(W, \delta)\,; \quad 0 \leq \delta \leq 1\,.$$

Moreover, they satisfy relations

$$C_{\mathrm{ID}}(W, \delta) \geq C(W)\,; \quad 0 < \delta < 1/2\,, \tag{3}$$

$$C_{\mathrm{dID}}(W, \delta) = C_{\mathrm{ID}}(W, \delta) = \infty\,; \quad 1/2 \leq \delta \leq 1\,.$$

Although usually proof of relation (3) is given for a finite discrete memoryless channel W [1, 14, 8], the same proof remains valid for any channel W with well–defined Shannon capacity $C(W)$. Proof of the last relation [1] does not depend on channel W. Moreover, due to the last relation only the case $0 < \delta < 1/2$ is interesting and only it will be considered below.

In the case of a finite discrete memoryless channel W we have also [1, 14, 8]

$$C_{\mathrm{ID}}(W, \delta) = C(W)\,; \quad 0 < \delta < 1/2\,. \tag{4}$$

$$C_{\mathrm{dID}}(W,\delta) = 0\,; \quad 0 < \delta < 1/2\,. \tag{5}$$

Remarkable relation (4) (it appeared first in [1]) creates a natural question: does it remain valid for a wider class of channels W (and, in particular, for channels with infinite alphabets) ? We will show that the answer is "Yes" and "No".

In order to get relation (4) for some channel W it is necessary to establish the converse of inequality (3). It should be noted that method of "types" used for such purpose in a finite discrete channel case [8, 14] essentially exploits the fact that input and output channel alphabets are both finite. When dealing with infinite alphabets channels we need some more accurate than in [8, 14] methods. One such method was developed in [5] and essential part of the paper is based on some results from it.

It turns out that for some classical continuous time channels W (i.e. for white Gaussian and Poisson channels without bandwidth constraints) with finite capacities $C(W)$ both deterministic and randomized identification capacities $C_{\mathrm{ID}}(W,\delta)$ and $C_{\mathrm{dID}}(W,\delta)$ are infinite. The reason of such singularity is that their input alphabets are too "rich". Nevertheless, formulas (4)–(5) remain valid for those channels if there is a bandwidth constraint.

In Section II some definitions and auxiliary results are presented. In Section III discrete time channels with additive noise are considered. Section IV contains two examples of channels with infinite ID–capacities. In Section V we consider white Gaussian channel with bandwidth constraint.

II. Auxiliary results

For any two probability measures P and Q on a measurable space $(\mathcal{X},\mathcal{B})$ variational distance (or L_1–distance) is

$$\|P-Q\| = \int_{\mathcal{X}} |dP - dQ|\,.$$

Let $\{P_\alpha,\, \alpha\in\mathcal{A}\}$ (where $\mathcal{A}$ is an arbitrary index set) be some collection of probability measures on a measurable space $(\mathcal{X},\mathcal{B})$. *Convex hull* $\mathrm{conv}\{\mathrm{P}_\alpha\}$ of the family $\{P_\alpha, \alpha\in\mathcal{A}\}$ is the set of all possible finite convex linear combinations of measures from $\{P_\alpha\}$. In other words, any measure $P\in\mathrm{conv}\{P_\alpha\}$ has the form $P=\sum_{i=1}^n c_{\alpha_i}Q_{\alpha_i}$ with $c_{\alpha_i}\geq 0$, $\sum_{i=1}^n c_{\alpha_i} = 1$, $\alpha_i\in\mathcal{A}$.

If P is some channel $W^{(n)}$ input distribution then $W^{(n)}P$ denotes its output distribution (and similarly for channel $W^{(T)}$). Next notions were introduced in [8] (see also [6]). We give them only for channel $W^{(n)}$, but their analogs for channel $W^{(T)}$ are straightforward.

Definition 3. *A collection $(P_i, i=1,\ldots,M)$ of probability measures on R^n is called an (M,n,δ,W) – pairwise separated collection if the following condition is satisfied:*

$$\|W^{(n)}P_i - W^{(n)}P_j\| \geq 2(1-\delta)\,, \quad i\neq j\,.$$

Definition 4. *A collection $(P_i, i=1,\ldots,M)$ of probability measures on R^n is called an (M,n,δ,W) – completely separated collection if the following condition is satisfied:*

$$\|W^{(n)}P_i - \mathrm{conv}\{W^{(n)}P_j,\ j\neq i\}\| \geq 2(1-\delta)\,, \quad i\neq j\,.$$

There is a simple relation between ID–code and completely separated collection of measures [8].

Proposition 1. *For any (M, n, δ, W) – completely separated collection $\{P_i\}$ it is possible to define regions $\{\mathcal{D}_i \subseteq R^n\}$ such that $\{P_i, \mathcal{D}_i\}$ will be (M, n, δ, W) – ID–code. Conversely, any (M, n, δ, W) – ID–code is $(M, n, 2\delta, W)$ – completely separated collection .*

Denote by $M_{ID}(\delta, W^{(n)})$, $M_p(\delta, W^{(n)})$, $M_c(\delta, W^{(n)})$ maximal possible cardinalities of corresponding ID–code, pairwise and completely separated families, respectively, for channel $W^{(n)}$ (and similar, for channel $W^{(T)}$). From Proposition 1 we have for any $0 < \delta < 1/2$

$$M_c(\delta, W^{(n)}) \le M_{ID}(\delta, W^{(n)}) \le M_c(2\delta, W^{(n)}) \le M_p(2\delta, W^{(n)})\,. \tag{6}$$

Due to bounds (6) instead of upperbounding the maximal cardinality $M_{ID}(\delta, W^{(n)})$ of ID–code it turns out to be sufficient to upperbound the maximal cardinality $M_p(\delta, W^{(n)})$ of pairwise separated family of measures at the channel output.

Notice also that any channel W (or $W^{(n)}$, or $W^{(T)}$) acts like a "compressing operator" in the following sense [8, 6]

Lemma 1. *For any channel W, any pair of input distributions P_1, P_2 and corresponding pair of output distributions Q_1, Q_2 the following inequality holds*

$$\|Q_1 - Q_2\| = \|WP_1 - WP_2\| \le \|P_1 - P_2\|\,. \tag{7}$$

III. Discrete time channels with additive noise

We consider first the channel $W(f, a)$ from (1)–(2). Concerning its noise density $f(x)$ we assume that there exist some constants K, K_1, γ, α such that

$$\int_{-\infty}^{\infty} \left(\max_{|t-x|\le u} \sqrt{f(t)} - \min_{|t-x|\le u} \sqrt{f(t)} \right)^2 dx \le K u^{\gamma}\,, \quad u > 0\,, \quad 1 < \gamma \le 2\,; \tag{8}$$

$$\int\limits_{|x|\ge z} f(x)\, dx \le K_1 z^{-\alpha}\,, \quad z > 0\,, \quad \alpha > 2\,. \tag{9}$$

$$1/\gamma + 1/\alpha < 1\,. \tag{10}$$

It should be mentioned that if condition (8) is satisfyed with some $\gamma > 2$ then function $f(x)$ is identically zero for all $x \in R^1$.

Now we shall describe some result from [5], concerning of approximation of channel's $W(f, a)$ input by a discrete input. For that purpose we introduce briefly some quantization of the input alphabet $\mathcal{A} = [-a\sqrt{n}, a\sqrt{n}]$ of channel $W(f, a)$ (see details in [5]).

First we represent input alphabet $\mathcal{A}$ as follows (below we use symbol "+" instead of "$\cup$" in order to emphasize that we have union of disjoint sets)

$$\mathcal{A} = [-a\sqrt{n}, a\sqrt{n}] = \mathcal{A}_1 + \mathcal{D}_1^- + \mathcal{D}_1^+\,, \quad \mathcal{A}_1 = [-ak(n), ak(n)]\,,$$

$$\mathcal{D}_1^- = [-a\sqrt{n}, -ak(n))\,, \quad \mathcal{D}_1^+ = (ak(n), a\sqrt{n}]\,, \quad k(n) = n^{(\gamma-1)/(2\gamma)}\,.$$

We quantize part $\mathcal{A}_1$, choosing in it the lattice $\mathcal{L}_1$ with span $\delta_1 = \left(\varepsilon^2/(nK)\right)^{1/\gamma}$. Then cardinality L_1 of lattice $\mathcal{L}_1$ satisfyes inequality

$$L_1 \le 2an^{(\gamma+1)/(2\gamma)}(K\varepsilon^{-2})^{1/\gamma} + 2\,.$$

We quantize part $\mathcal{D}_1^+$, choosing in it some increasing sequence of points $ak(n) = x_1 < \ldots < x_{L_2} = a\sqrt{n}$ and denote that set of points $\mathcal{L}_2$. In the alphabet $\mathcal{D}_1^-$ we choose symmetric to $\mathcal{L}_2$ set of points $\mathcal{L}_3 = -\mathcal{L}_2$. Union of three disjoint sets we denote $\mathcal{L} = \mathcal{L}_1 + \mathcal{L}_2 + \mathcal{L}_3$. Any input $u^n = (u_1, \ldots, u_n) \in \mathcal{A}^n$ we approximate by other input $u' = (u'_1, \ldots, u'_n) \in \mathcal{L}^n$ such that each u'_i is the closest to u_i element of alphabet $\mathcal{L}$ with $|u'_i| \le |u_i|$.

We choose sequence $\{x_i\}$ as follows

$$x_{i+1} = a(1+\mu)^i k(n)\,, \quad \mu = \left(\varepsilon^2/(nKa^2)\right)^{1/\gamma}\,, \quad i = 1, \ldots, L_2 - 1.$$

Then for cardinality L_2 of set $\mathcal{L}_2$ we have $L_2 \le \left(a^2K\,\varepsilon^{-2}n\right)^{1/\gamma}\ln n$, and for total cardinality L of discrete set $\mathcal{L}$ we get

$$L \le 2an^{(\gamma+1)/(2\gamma)}(K\varepsilon^{-2})^{1/\gamma} + 2\left(a^2K\,\varepsilon^{-2}n\right)^{1/\gamma}\ln n + 2\,. \tag{11}$$

The following result (Theorem 2b from [5]) describes some properties of thus obtained discrete set $\mathcal{L} = \mathcal{L}(n,\varepsilon)$.

Theorem 1. *Let for noise density $f(x)$ of channel $W(f,a)$ conditions (8)–(10) are fullfilled. Then finite input alphabet $\mathcal{L} = \mathcal{L}(n,\varepsilon) \in [-a\sqrt{n}, a\sqrt{n}]$, $\varepsilon > 0$, $n \ge n_0(\varepsilon)$, has cardinality L, satisfying inequality (11) and, moreover, for any output measure $Q_\pi^{(n)}$ on R^n there exist input blocks $x(i) \in \mathcal{L}^n$, $i = 1, \ldots, M$, satisfying constraint (2), with $n^{-1}\ln M \le C(f,a) + \varepsilon$, such that for their generated measures $\{W^{(n)}(.|x(i))\}$ at channel output the following inequality holds true*:

$$\left\| Q_\pi^{(n)} - \operatorname{conv}\left\{W^{(n)}(\cdot|x(i)),\, i = 1, \ldots, M\right\}\right\| \le \varepsilon\,. \tag{12}$$

Proposition 2. *Let for noise density $f(x)$ of channel $W(f,a)$ conditions (8)–(10) are fullfilled. Then for any $0 < \delta < 1/2$ the following inequalities hold:*

$$C_{ID}(W,\delta) \le \lim_{n\to\infty} \frac{\ln\ln M_p(2\delta, W^{(n)})}{n} \le C(f,a), \quad 0 < \delta < 1/2\,. \tag{13}$$

Proof. Since $M_{ID}(\delta, W^{(n)}) \le M_p(2\delta, W^{(n)})$, left of inequalities (13) follows from definition of $C_{ID}(W,\delta)$. Therefore it remains to prove the second of inequalities (13). For that purpose we fix some small $\varepsilon > 0$ such that $2(\delta+\varepsilon) < 1$ and find corresponding finite input alphabet $\mathcal{L} = \mathcal{L}(n,\varepsilon)$ described above. Let $\{Q_i,\, i = 1, \ldots, M\}$ be some 2δ–pairwise separated collection of output measures. By virtue of Theorem 1 each output measure Q_i can be ε–approximated (see (12)) by another output measure Q'_i generated by some $N = e^{n(C+\varepsilon)}$ input blocks from alphabet $\mathcal{L}^n$. Then collection $\{Q'_i,\, i = 1, \ldots, M\}$ is $2(\delta+\varepsilon)$–pairwise separated. Notice that using N input blocks even on noiseless channel the maximal number $M_p(\mu, N)$ of μ–pairwise separated measures that we can get is upper bounded [8] by (see more accurate estimates in [2])

$$M_p(N,\mu) \le \left(\frac{2}{1-\mu}\right)^{N-1}, \quad 0<\mu<1\,.$$

Since the channel $W^{(n)}$ is a compressing operator (Lemma 1), the maximal number of $2(\delta+\varepsilon)$–pairwise separated measures, generated by every collection of $N = e^{n(C+\varepsilon)}$ input blocks, is upperbounded by the same formula (with $2(\delta+\varepsilon)$ instead of μ). Total number of collections of cardinality N on the alphabet of cardinality L^n does not exceed L^{nN}. Therefore for number $M_p(2(\delta+\varepsilon), W^{(n)})$ we get

$$M_p(2(\delta+\varepsilon), W^{(n)}) \le \left(\frac{2}{(1-2\delta-2\varepsilon)}\right)^{N-1} L^{nN} \le$$

$$\le \exp\left\{\left(\ln\frac{2}{(1-2\delta-2\varepsilon)} + n\ln L\right)e^{n(C+\varepsilon)}\right\}.$$

Since ε can be chosen arbitrary small, we get formula (13). $\triangle$.

Combining now formulas (3) and (13) we get

Proposition 3. *If for noise density $f(x)$ of channel $W(f,a)$ conditions (8)–(10) are fullfilled then*

$$C_{ID}(W,\delta) = C(f,a), \quad 0<\delta<1/2\,. \tag{14}$$

It is clear that among conditions (8)–(10) (that were used in Theorem 1 and Proposition 3) only condition (8) looks a bit complicated for use. We demonstrate that, in fact, it is not so difficult and that condition characterizes some "smoothness" of function $f(x)$. For that purpose we consider some examples and the first of them is, probably, the most natural case when condition (8) is fullfilled.

Example 1. Assume that density $f(x)$ is absolutely continuous and there exists Fisher information $I(f) = \int f'^2/f dx < \infty$. Notice that

$$\left(\max_{|t-x|\le u}\sqrt{f(t)} - \min_{|t-x|\le u}\sqrt{f(t)}\right)^2 \le$$

$$\le \left[\int_{x-u}^{x+u}\frac{|f'(v)|}{2\sqrt{f(v)}}\,dv\right]^2 \le \frac{1}{2}u\int_{x-u}^{x+u}\frac{f'^2(v)}{f(v)}\,dv\,.$$

Therefore

$$\int_{-\infty}^{\infty}\left(\max_{|t-x|\le u}\sqrt{f(t)} - \min_{|t-x|\le u}\sqrt{f(t)}\right)^2 dx \le \frac{1}{2}u\int_{-\infty}^{\infty}\int_{x-u}^{x+u}\frac{f'^2(v)}{f(v)}\,dv\,dx =$$

$$= \frac{1}{2} u \int_{-\infty}^{\infty} \frac{f'^2(v)}{f(v)} \int_{v-u}^{v+u} dx\, dv = u^2 I(f)\,.$$

Therefore condition (8) is fullfilled with $\gamma = 2$, $K = I(f)$. Then condition (9) (i.e. $1/\gamma + 1/\alpha < 1$) will be fullfilled if $\alpha > 2$.

Example 2. Assume that: 1) density $f(x)$ is absolutely continuous for $x \in R^1$;
2) there are k "irregular" points $-\infty < x_1 < \ldots < x_k < \infty$ such that outside of small vicinities of those points Fisher integral converges;
3) in small vicinities of "irregular" points $x_1, \ldots, x_k$ density $f(x)$ behaves like $|x - x_i|^{\gamma_i}$, $\gamma_i > 1$.
Then condition (8) is fullfilled with $\gamma = \min_i\{\gamma_i\}$.

Example 3. Assume that density $f(x)$ has points of discontinuities (e.g. $f(x) = 1$ for $|x| \le 1/2$ and $f(x) = 0$ for
$|x| > 1/2$). Then condition (8) is fullfilled with $\gamma = 1$ and therefore condition (9) (i.e. $1/\gamma + 1/\alpha < 1$) can not be fullfilled. It means that we can not claim that Theorem 1 (and therefore Proposition 3) remain valid for such density.

IV. Two examples of channels W with infinite $C_{\mathrm{dID}}(W)$

a) White Gaussian noise channel without bandwidth constraint

Consider first white Gaussian noise channel (of unit intensity). If $S(t)$, $0 \le t \le T$, is its input signal then output observation $X_0^T = \{X(t),\ 0 \le t \le T\}$ is described through stochastic differentials

$$dX(t) = S(t)dt + dW_t\,, \quad 0 \le t \le T\,,$$

where W_t is a standard Wiener process. It is assumed that input signals $S(t)$, $0 \le t \le T$, must satisfy only the energy constraint

$$\int_0^T S^2(t)\, dt \le AT\,,$$

where $A > 0$ is some prescribed constant (see Remark 1 below about peak power constraint).
For Shannon capacity $C(A)$ of such channel we have [12]: $C(A) = A$.
Contrary, we shall prove that deterministic dID–capacity
$C_{\mathrm{dID}}(A)$ of that channel is infinite. For that purpose it is sufficient to show that for any $\delta > 0$ there exists an arbitrary large system of signals $\{S_1, \ldots, S_M\}$ that serves as deterministic (M, T, δ) – dID–code.

Let $\{s_i(t),\ 0 \le t \le T;\ i = 1, \ldots, M\}$, be an arbitrary orthonomal system of functions, i.e.

$$(s_i, s_j) = \int_0^T s_i(t) s_j(t)\, dt = 0 \quad \text{for any } i \ne j\,; \quad (s_i, s_i) = 1 \quad \text{for any } i\,.$$

We put

$$S_i(t) = \sqrt{AT}\, s_i(t),\ 0 \le t \le T; \quad i = 1, \ldots, M\,. \tag{15}$$

In order to describe now decision making process at the channel output we fix some parameter $z > 0$ and introduce the following system of sets $\{\mathcal{D}_i(z)\}$

$$\mathcal{D}_i(z) = \left\{ X_0^T : \int_0^T S_i(t)\, dX(t) \ge z \right\} , \quad i = 1, \ldots, M \,.$$

We make a decision in favor of signal S_i if at the channel output we have $X_0^T \in \mathcal{D}_i(z)$.

Now if signal S_i was transmitted then for the probability that received signal X_0^T will not belong to the decision set $\mathcal{D}_i(z)$ we have

$$P\{ X_0^T \notin \mathcal{D}_i(z) \big| \, S_i \} = P\left\{ \int_0^T S_i(t)\, dX(t) < z \middle| \, S_i \right\} =$$

$$= P\left\{ AT + \int_0^T S_i(t)\, dW_t < z \right\} = \Phi\left(\frac{z - AT}{\sqrt{AT}} \right) ; \quad i = 1, \ldots, M \,.$$

On the other hand, if signal S_i is transmitted then for the probability to make also decision in favor of signal S_j we have for any $j \ne i$

$$P\{ X_0^T \in \mathcal{D}_j(z) \big| \, S_i \} = P\left\{ \int_0^T S_j(t)\, dX(t) \ge z \middle| \, S_i \right\} =$$

$$= P\left\{ \int_0^T S_j(t)\, dW_t \ge z \right\} = \Phi\left(-\frac{z}{\sqrt{AT}} \right) .$$

If we put now $z = AT/2$ then for any $j \ne i$ and any $AT \ge 8\ln(1/\delta)$ we get

$$P\{ X_0^T \notin \mathcal{D}_i(z) \big| \, S_i \} = P\{ X_0^T \in \mathcal{D}_j(z) \big| \, S_i \} = \Phi\left(-\frac{\sqrt{AT}}{2} \right) \le e^{-AT/8} \le \delta \,.$$

Since on $[0, T]$ there exist arbitrary large number M of orthonomal functions $\{s_i(t)\}$, we get

Proposition 4. *If $\delta > 0$ and $AT \ge 8\ln(1/\delta)$ then for white Gaussian channel the maximal cardinality M of deterministic dID–code is not limited, and therefore $C_{\rm dID}(A) = C_{\rm ID}(A) = \infty$.*

Remarks. 1) It is clear that if there is an additional peak power constraint (like $|S(t)| \le K \,,\ K > 0$), we will still have $C_{\rm dID}(A) = \infty$.

2) From that example it may seem that any system of signals $\{S_1, \ldots, S_M\}$ will be a good dID–code, if any pair S_i and S_j of those signals is "well-separated" (i.e. can be tested with small error probabilities). But it is not true. Indeed, compliment the system above by zero–signal $S_0(t) \equiv 0$. Then in the new system $\{S_0, S_1, \ldots, S_M\}$ any two signals are still "well-separated". That new system $\{S_0, S_1, \ldots, S_M\}$ will be $(M+1, \delta, T)$–dID–code if we can test (with small error probabilities) zero–signal S_0 (as simple hypothesis) against all remaining orthogonal signals $\{S_1, \ldots, S_M\}$ (as composite hypothesis) [8]. But it is possible only if the number M is not too large (more exactly, only if $M \le e^{AT/4}$, see details in [4, 7]).

b) Poisson channel

The Poisson (direct–detection photon) channel [16], [17] is described as follows. The channel input is a waveform $S(t)$, $0 \le t \le T$, which satisfies

$$0 \le S(t) \le A < \infty \,, \tag{16}$$

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

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

where

$$\Lambda = \int_{t_1}^{t_2} (\lambda_0 + S(\tau))\, d\tau \;.$$

Let $\mathcal{X}_T$ denote the space of right continuous (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 $\mathcal{X}_T$. Denote $\mathcal{S}_T$ the set of all Lebesque–measurable functions $S(t)$, $0 \le t \le T$, satisfying the "peak power constraint" (16).

It is known that the Shannon capacity $C(A, \lambda_0)$ of such channel is finite and it monotone decreases with λ_0. Moreover, [16], $C(A, 0) = A/e$ (for $\lambda_0 > 0$ case see explicit formulas in [11, 17]).

Contrary, dID–capacity $C_{\mathrm{dID}}(A, \lambda_0)$ of such channel is infinite for any $\lambda \ge 0$. For proof it is sufficient to show that for any $\delta > 0$ and $T \ge T_0(A, \delta)$ there exists a system of signals $\{S_1, \ldots, S_M\}$ that serves as (M, T, δ) – dID–code and, moreover, its cardinality M can be made arbitrary large.

For that purpose we construct the following system of step signals $\{S_1, \ldots, S_M\}$. We devide $[0, T]$ on N equal segments Δ of length T/N. Each signal $S_i(t)$ on every segment Δ take on only extreme values 0 and A. Denote

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

and let $mes\,\{m\}$ denotes the Lebesque measure of the set m. Support m_i of each signal S_i consists of εN segments Δ and therefore, $mes\,\{m_i\} = \varepsilon T$, where the value $\varepsilon < \delta$ will be chosen later. Therefore total energy of each signal S_i is $\mu = \varepsilon A T$.

We demand that the system of supports $\{m_1, \ldots, m_M\}$ satisfyes condition

$$mes\,(m_i \cap m_j) \le \delta \varepsilon T \qquad \text{for any } i \ne j \,.$$

Lemma 2. *Maximal possible cardinality M of such system of supports $\{m_1,\ldots,m_M\}$ satisfyes the lower bound*

$$M \geq \sqrt{\frac{2(\delta-\varepsilon)}{\delta(N+2)}}\exp\left\{\frac{\varepsilon(\delta-\varepsilon)^2 N}{4\delta(1-\delta)}\right\}\,, \quad 0<\varepsilon\leq\delta\,, \tag{17}$$

Remark. Similar lower bounds (but in a weaker form) was proved already in [1] (Statement 1) and [2] (Theorem 1) by an "exhaustive" method. For the sake of variety we prove the inequality (17) by random choice arguments.

Proof. Indeed, we consider a binary code of length N and constant weight εN. Assume that we choose all M codeblocks randomly. Then the probability P that there will be at least two codeblocks with number of common ones $> \varepsilon\delta N$ satisfyes the inequality

$$P \leq M(M-1)\sum_{i>\epsilon\delta N}\binom{\epsilon N}{i}\binom{N-\epsilon N}{\epsilon N - i}\left[2\binom{N}{\epsilon N}\right]^{-1}.$$

Notice that for $i > \epsilon\delta N$ we have

$$\binom{\epsilon N}{i+1}\binom{N-\epsilon N}{\epsilon N-i-1}\left[\binom{\epsilon N}{i}\binom{N-\epsilon N}{\epsilon N-i}\right]^{-1} \leq \frac{\epsilon(1-\delta)^2}{\delta(1-2\epsilon+\delta\epsilon)} \leq \frac{\epsilon}{\delta}\;.$$

Therefore replacing that sum by the geometrical progression we get

$$P \leq M(M-1)\delta\binom{\epsilon N}{\epsilon\delta N}\binom{N-\epsilon N}{\epsilon N-\epsilon\delta N}\left[2(\delta-\epsilon)\binom{N}{\epsilon N}\right]^{-1} <$$

$$< \frac{\delta(N+2)M^2}{2(\delta-\varepsilon)}\exp\{Nf(\delta,\varepsilon)\}\;; \tag{18}$$

$$f(\delta,\varepsilon) = \varepsilon h(\delta) + (1-\varepsilon)h\left(\frac{\varepsilon(1-\delta)}{1-\varepsilon}\right) - h(\varepsilon)\,, \quad 0<\varepsilon\leq\delta\,,$$

where we used inequality [10]

$$\frac{1}{n+2} \leq \binom{n}{k}\exp\{-nh(k/n)\} \leq 1\,.$$

Now for function $f(\delta,\varepsilon)$ we have

$$f'_\delta = \varepsilon\ln\frac{1-\delta}{\delta} + \varepsilon\ln\frac{\varepsilon(1-\delta)}{(1-2\varepsilon+\varepsilon\delta)}\,,$$

$$f''_{\delta^2} = -\frac{2\varepsilon}{(1-\delta)} - \frac{\varepsilon}{\delta} - \frac{\varepsilon^2}{(1-2\varepsilon+\varepsilon\delta)}\,.$$

Notice also that $f(\delta,\delta) = f'_\delta(\delta,\delta) = 0$. Therefore from Taylor formula we get the upperbound

$$f(\delta,\varepsilon) \le -\frac{\varepsilon(\delta-\varepsilon)^2}{2\delta(1-\delta)}\,, \quad 0<\varepsilon\le\delta\,. \tag{19}$$

Now from (18)–(19) for the probability P we have

$$P < \frac{\delta(N+2)M^2}{2(\delta-\varepsilon)}\exp\left\{-\frac{\varepsilon(\delta-\varepsilon)^2 N}{2\delta(1-\delta)}\right\}\,, \quad 0<\varepsilon\le\delta\,,$$

from where lower bound (17) follows. $\square$.

In order to describe the decision making method at the channel output $\mathcal{X}_T$, we consider first a simpler case of no "dark current" (i.e. when $\lambda_0 = 0$).

Introduce the following system of decision sets

$$\mathcal{D}_i = \left\{x_0^T \in \mathcal{X}_T : \text{there is at least one photon (event) on time set } m_i\right\}\,, \quad i=1,\ldots,M\,.$$

We make a decision in favor of signal S_i if $x_0^T \in \mathcal{D}_i$ (i.e. if observation process $x(t)$ has at least one jump at some point belonging to m_i).

Now if signal S_i was transmitted then for the probability that received signal x_0^T will not belong to the decision set $\mathcal{D}_i$ we have

$$P\{x_0^T \notin \mathcal{D}_i \big| S_i\} = P\{\text{no photons on } m_i | S_i\} = e^{-\varepsilon AT}\,; \quad i=1,\ldots,M\,.$$

On the other hand, if signal S_i was transmitted then for the probability to make also decision in favor of signal S_j we have for any $j \neq i$

$$P\{x_0^T \in \mathcal{D}_j \big| S_i\} = P\{\text{there are photons on } m_i\cap m_j\} \le 1-e^{-\varepsilon\delta AT} \le \varepsilon\delta AT\,.$$

We put now $\varepsilon AT = \ln(2/\delta)$ and then for any $i \neq j$ we get

$$\max\left\{P\{x_0^T \notin \mathcal{D}_i \big| S_i\};\, P\{x_0^T \in \mathcal{D}_j \big| S_i\}\right\} \le \delta\,\max\{1/2, \ln(2/\delta)\} = \delta\ln(2/\delta)\,.$$

Now if $AT \ge 2\delta^{-1}\ln(2/\delta)$ then $\varepsilon \le \delta/2$ and we get from (17) that the number of maximal possible signals M satisfyes lower bound

$$M \ge \frac{1}{\sqrt{(N+2)}}\exp\left\{\frac{N\delta\ln(2/\delta)}{16AT}\right\}\,, \quad AT \ge 2\delta^{-1}\ln(2/\delta)\,. \tag{20}$$

Since the number N may be chosen arbitrary large, we get from (20)

Proposition 5a. *If $AT \ge 2\delta^{-1}\ln(2/\delta)$ then the cardinality M of (M,T,δ)–dID–code is not limited, and therefore for Poisson channel with $\lambda_0 = 0$ we have $C_{\rm dID}(A) = C_{\rm ID}(A) = \infty$.*

Consider now the case when the "dark current" $\lambda_0 > 0$. We fix some parameter $z > 0$ and introduce the following system of decision sets

$$\mathcal{D}_i = \left\{ x_0^T \in \mathcal{X}_T : \text{the number of photons on } m_i \text{ is } \geq (\lambda_0 + A - z)\varepsilon T \right\}, \quad i = 1, \ldots, M\,.$$

We make a decision in favor of signal S_i if $x_0^T \in \mathcal{D}_i$.

Now if signal S_i was transmitted then for the probability that received signal x_0^T will not belong to the decision set $\mathcal{D}_i$ we have

$$P\{ x_0^T \notin \mathcal{D}_i \big| S_i\} = P\{\text{the number of photons on } m_i \text{ is } < (\lambda_0 + A - z)\varepsilon T | S_i\} \leq$$

$$\leq e^{-(\lambda_0+A)\varepsilon T} \sum_{n=0}^{(\lambda_0+A-z)\varepsilon T} \frac{((\lambda_0+A)\varepsilon T)^n}{n!} \leq$$

$$\leq e^{-z\varepsilon T} \left(\frac{\lambda_0 + A}{\lambda_0 + A - z} \right)^{(\lambda_0+A-z)\varepsilon T} \leq \exp\left\{ -\frac{\varepsilon z^2 T}{2(\lambda_0 + A)} \right\}, \tag{21}$$

where we used simple inequalities $(1-x)\ln(1-x) + x - x^2/2 \geq 0\,; 0 \leq x < 1$, and

$$\sum_{k=0}^{n} \frac{b^k}{k!} \leq \left(\frac{be}{n} \right)^n \;; \quad b \geq n\,.$$

On the other hand, if signal S_i is transmitted then for the probability to make also decision in favor of signal S_j we have for any $K > \lambda_0$ and $z + K \leq (1-\delta)(\lambda_0 + A)$

$$P\{ x_0^T \in \mathcal{D}_j \big| S_i\} = P\{\text{the number of photons on } m_j \text{ is } \geq (\lambda_0 + A - z)\varepsilon T |\, S_i\} \leq$$

$$\leq P\{\text{the number of photons on } m_{ij} \text{ is } \geq (\lambda_0 + A - z - K)\varepsilon T |\, S_i\} +$$

$$+ P\{\text{the number of photons on } m_j \setminus m_{ij} \text{ is } \geq K\varepsilon T |\, S_i\} \leq$$

$$\leq e^{-(\lambda_0+A)\varepsilon\delta T} \sum_{n=(\lambda_0+A-z-K)\varepsilon T}^{\infty} \frac{((\lambda_0+A)\varepsilon\delta T)^n}{n!} +$$

$$+ e^{-\lambda_0\varepsilon(1-\delta)T} \sum_{n=K\varepsilon T}^{\infty} \frac{(\lambda_0\varepsilon(1-\delta)T)^n}{n!} \leq$$

$$\leq e^{-((\lambda_0+A)(1-\delta)-z-K)\varepsilon T} \left(\frac{(\lambda_0+A)\delta}{\lambda_0 + A - z - K} \right)^{(\lambda_0+A-z-K)\varepsilon T} +$$

$$+ e^{((K-\lambda_0(1-\delta))\varepsilon T} \left(\frac{\lambda_0(1-\delta)}{K} \right)^{K\varepsilon T} \leq$$

$$\leq \exp\left\{ -\frac{\varepsilon T \left[(\lambda_0+A)(1-\delta) - z - K \right]^2}{2(\lambda_0 + A - z - K)} \right\} + \exp\left\{ -\frac{\varepsilon T \left[K - \lambda_0(1-\delta) \right]^2}{2K} \right\} \leq$$

$$\leq \exp\left\{-\frac{\varepsilon T\left[(\lambda_0+A)(1-\delta)-z-K\right]^2}{2(\lambda_0+A)}\right\}+\exp\left\{-\frac{\varepsilon T\left[K-\lambda_0(1-\delta)\right]^2}{2(\lambda_0+A)}\right\}, \qquad (22)$$

where we used simple inequalities $\ln(1+x) \leq x - x^2/2\,;\ -1 < x \leq 0$, and

$$\sum_{k=n}^{\infty} \frac{b^k}{k!} \leq \left(\frac{be}{n}\right)^n \;; \quad b \leq n\,.$$

Choose now $z = A(1-\delta)/3,\ K = \lambda_0(1-\delta)+z$. Then all exponents in right sides of (21) and (22) will be equal and we get for any $i \neq j$

$$\max\left\{P\{x_0^T \notin \mathcal{D}_i \big| S_i\};\, P\{x_0^T \in \mathcal{D}_j \big| S_i\}\right\} \leq 2\exp\left\{-\frac{\varepsilon(1-\delta)^2A^2T}{18(\lambda_0+A)}\right\}.$$

Choosing finally

$$\varepsilon = \frac{18(\lambda_0+A)\ln(2/\delta)}{(1-\delta)^2A^2T}\,,$$

we get

$$\max\left\{P\{x_0^T \notin \mathcal{D}_i \big| S_i\};\, P\{x_0^T \in \mathcal{D}_j \big| S_i\}\right\} \leq \delta\,.$$

Notice that only the case $\delta < 1/2$ is interesting (otherwise, the number of possible signals M is obviously infinite). Therefore if for $\delta < 1/2$ we have also $T \geq 144A^{-2}(\lambda_0+A)\delta^{-1}\ln(2/\delta)$ then $\varepsilon \leq \delta/2$, and we get from (17) that the number of possible signals M satisfyes lower bound

$$M \geq \frac{1}{\sqrt{(N+2)}}\exp\left\{\frac{N\delta\ln(2/\delta)}{AT}\right\}, \quad \text{if} \quad \frac{A^2T}{(\lambda_0+A)} \geq \frac{144\ln(2/\delta)}{\delta}\,. \qquad (23)$$

Since the number N in (23) may be chosen arbitrary large, we can generalize Proposition 5a as follows

Proposition 5. *If $T \geq 144A^{-2}(\lambda_0+A)\delta^{-1}\ln(2/\delta)$ then the cardinality M of (M,T,δ)–dID–code is not limited, and therefore for Poisson channel with $\lambda_0 \geq 0$ we have $C_{\mathrm{dID}}(A,\lambda_0) = C_{\mathrm{ID}}(A,\lambda_0) = \infty$.*

V. White Gaussian noise channel with bandwidth constraint

When we showed in Section IV a) that for white Gaussian noise channel without bandwidth constraint $C_{\mathrm{dID}}(A) = \infty$, it was crucial that the set $\mathcal{S}_T$ of possible input signals has infinite (more exactly, very fast growing with T) dimension. Now we shall show that if that dimension grows linearly with T then $C_{\mathrm{ID}}(A,\delta) = C(A)$, $0 < \delta < 1/2$. Moreover, we obviously have $C_{\mathrm{dID}}(A,\delta) = 0$, $0 < \delta < 1/2$.

Assume that the set $\mathcal{S}_T$ of possible input signals has dimension $2WT$, where $W > 0$ is a given constant. In other words, we are given a set (basis) $\{s_i(t),\ldots,s_{2WT}\}$ of orthonomal on $[0,T]$ functions and as an input signal we can choose any function of the form

$$S(t) = \sum_{i=1}^{2WT} u_i s_i(t)\,, \quad 0 \leq t \leq T\,,$$

with

$$\sum_{i=1}^{2WT} u_i^2 \le AT\,.$$

Shannon's formula [12] gives capacity of such channel: $C(W,A) = W\ln(1+A/(2W))$. Using standard procedure [12] we can reduce that continuous time channel to a discrete time additive Gaussian noise channel. For that purpose we replace channel output $X(t)$, $0 \le t \le T$, by WT outputs

$$x_i = \int_0^T s_i(t)\,dX(t) = u_i + \xi_i\,, \quad i = 1,\ldots,2WT\,,$$

where $\{\xi_i,\, i = 1,\ldots,2WT\}$ are i.i.d. r.v.'s with $N(0,1)$–Gaussian distribution. We loose no information in that transition (it is sufficient statistics in the problem). Denoting $n = 2WT$ we come to the model when received (output) signal is

$$x_i = u_i + \xi_i\,, \quad i = 1,\ldots,n\,,$$

and input signal $(u_1,\ldots,u_n)$ satisfyes constraint

$$\sum_{i=1}^{n} u_i^2 \le na^2\,, \quad a^2 = A/(2W)\,.$$

Capacity of such channel is given by formula [12]: $C(a) = 1/2\ln(1+a^2)$. Both formulas for capacities differ only by normalization.

Such discrete time additive Gaussian noise channel is a particular case of a more general model considered in Section III. Moreover, Gaussian density $f(x)$ satisfies assumptions of Example 1 with $I(f) = 1$. Therefore Theorem 1 and Proposition 3 are valid for such channel and we get

Proposition 6. *Let dimension of input signals set $\mathcal{S}_T$ linearly grows with T. Then*

$$C_{ID}(W,A,\delta) = C(W,A,\delta), \quad 0 < \delta < 1/2\,. \tag{24}$$

Remark. Main surprising feature of ID–codes for finite channels [1] was that their cardinality $M_{ID}(n,\delta)$ grows as "double" exponent of blocklength n. In the case of white Gaussian channel, let $N(T)$ be dimension of input signals set $\mathcal{S}_T$. Then we have

$$e^{N(T)} \le M_{dID}(T,\delta) \le M_{ID}(T,\delta) \le e^{e^{N(T)}}\,.$$

It is clear that choosing in this case $\ln\ldots\ln N(T) = KT$, we can get any "number of exponents" in $M_{ID}(T,\delta)$.

VI. Concluding remarks

1. Based on paper's results I would like to make the following

Conjecture. For any channel W the following alternative is valid: $C_{ID}(W, \delta) = C(W, \delta)$ or $C_{ID}(W, \delta) = \infty$.

Unfortunately, the author has not been able to prove (or disprove) that conjecture.

2. Prof. Te Sun Han informed me that in his book [13] it was shown by using a very different method that the identification capacity of stationary white Gaussian noise channels coincides with the usual transmission capacity.

References

[1] R. Ahlswede and G. Dueck, "Identification via channels," *IEEE Trans. Inform. Theory*, vol. 35, no. 1, pp. 15 - 29, 1989.

[2] L. A. Bassalygo and M. V. Burnashev, "Estimate for the maximal number of messages for a given probability of succesful deception," *Probl. Inform. Transm.*, vol. 30, no. 2, pp. 42 - 48, 1994.

[3] L. A. Bassalygo and M. V. Burnashev, "Authentication, identification and pairwise separated measures," *Probl. Inform. Transm.*, vol. 32, no. 1, pp. 41 - 47, 1996.

[4] 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.

[5] M. V. Burnashev, "On method of types and approximation of output measures for channels with infinite alphabets," *Probl. Inform. Transm. – to appear*, 2000.

[6] Marat V. Burnashev, "Selected topics of information theory and mathematical statistics," *Preprint 96 - 011, Sonderforschungsbereich 343 "Diskrete Strukturen in der Mathematik", Bielefeld University, Germany*, 1996.

[7] M. V. Burnashev and I. A. Begmatov, "On a signal detection problem leading to stable distributions," *Theory of Prob. and Its Appl.*, vol. 35, no. 3, pp. 1169 - 1172, 1990.

[8] M. V. Burnashev and S. Verdu, "Measures separated in L_1 metrics and ID–codes," *Probl. Inform. Transm.*, vol. 30, no. 3, pp. 3 - 14, 1994.

[9] M. V. Burnashev and Y. A. Kutoyants, "On sphere–packing bound, capacity and related results for Poisson channel, "*Probl. Inform. Transm.*, vol. 35, no. 2, pp. 3 - 22, 1999.

[10] T. M. Cover and J. A. Tomas, *Information Theory.* New York: Wiley, 1991.

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

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

[13] T. S. Han. *Information–Spectrum Methods in Information Theory*, (in Japanese). Tokyo: Baifukan–Press, 1998.

[14] T. S. Han and S. Verdu, "New results in the theory of identification via channels," *IEEE Trans. Inform. Theory*, vol. 38, no. 1, pp. 14 - 25, 1992.

[15] T. S. Han and S. Verdu, "Approximation theory of output statistics," *IEEE Trans. Inform. Theory*, vol. 39, no. 3, pp. 752 - 772, 1993.

[16] 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.

[17] 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.