

DOI: 10.1070/QEL16223

Evaluation of the capacity of communication lines with nonlinear finite memory

E.G. Shapiro, D.A. Shapiro

Abstract. **We propose a method for calculating the capacity of a finite-memory channel up to the square of the nonlinear memory parameter. A comparison with a regular Gaussian model is performed, in which the Kerr nonlinearity is considered as an additional Gaussian noise. The estimate by the regular Gaussian model is shown to yield a greater capacity as compared with the estimate obtained by the nonlinear memory model. The optimal signal powers, providing the maximum mutual information, are found to be equal.**

Keywords: *fibre-optic communication lines, capacity, nonlinear noise, Shannon limit.*

1. Introduction

Despite the active development of fibre optics since the 1960s, the problem of assessing the capacity of optical communication lines still remains unsolved. The complexity of the problem is due to the influence of nonlinearity, which is inevitably present during the propagation of pulses through fibre-optic systems. The main reasons of optical signal degradation are the Kerr nonlinearity and amplifier noise. The Kerr nonlinearity prevents the achievement of an arbitrarily large signal-to-noise ratio, in contrast to the case of a linear channel. This fact is often called 'nonlinear Shannon limit'.

In statistical communication theory, a link, wherein each symbol of an output sequence depends not only on the respective signal at the input but also on transmitted input and detected output signals, is called a channel with memory. If the probability of detection depends on $2L + 1$ symbols, i.e. on the signal itself, on L symbols before and L symbols after, then one speaks of a channel with finite memory. A regular channel is a limiting case at $L \to \infty$.

With selecting an optimal distribution of input symbols, one can in principle reduce the probability of an error during the signal detection at the receiver. This search of the alphabet would allow one to partially compensate for the intersymbol interference, or pattern effect. However, to date, due to the high computational complexity of the problem, the calculation methods of the capacity of a channel with memory are absent. In this paper, we have developed and implemented such a method in the simplest Gaussian noise model. Neighbouring bits are considered to be an additional random noise. The channel nonlinearity is treated as a conditional probability: the distribution dispersion increases as the cube of an average power of the signal. The signal detection probability is most strongly influenced by two neighbouring bits, and therefore the calculations are performed in the approximation of memory of three symbols in length ($L = 1$).

In this paper, we have proposed an estimate of the capacity of a communication link with nonlinear finite memory with an accuracy of $O(\mu^2)$, where μ is the memory parameter. We have also compared numerically the dependences of the capacity of a finite-memory channel and a channel with average nonlinear noise on the signal power. It is shown that the channel with nonlinear averaged noise has a higher bandwidth capacity in the region of the optimal signal power. In this case, the optimal signal powers for both models are identical.

2. Regular Gaussian-noise channel and finite-memory channel

Mutual information [1] of a discrete time-invariant channel with memory is given by the expression

$$I(X;Y) = \lim_{N\to\infty}\frac{1}{N}I(X_1,\dots,X_N;Y_1,\dots,Y_N), \qquad (1)$$

where $(X_1,\dots,X_N)$ и $(Y_1,\dots,Y_N)$ are the input and output symbol sequences, respectively. For each index i, the value of X_i is a random variable taking values from a set of numbers, called the input alphabet. The values measured at the end of the link are also the values of the random variable Y_i and called the output alphabet.

The authors of papers [2, 3] consider the nonlinear signal-distorting interaction as the Gaussian noise, which cubically depends on the signal power. Mathematically, this is written as follows:

$$Y_k = X_k + Z_k, \quad Z_k = \tilde{Z}_k\sqrt{\sigma^2_{\rm ASE} + \mu P^3}. \qquad (2)$$

Here, X_k is the symbol transmitted in a time interval with the number k; Y_k is the value detected by the receiver; $\tilde{Z}_k$ is the random Gaussian variable with a zero mean and variance equal to unity; $\sigma^2_{\rm ASE}$ and μ are the positive constants; and P is the average signal power. The constant $\sigma^2_{\rm ASE}$ corresponds to the amplified spontaneous emission noise.

Agrell et al. [4] proposed a finite-memory Gaussian channel model, in which the average signal power P in (2) is replaced by the empirical power, i.e. by the averaged power of

E.G. Shapiro, D.A. Shapiro Institute of Automation and Electrometry, Siberian Branch, Russian Academy of Sciences, prosp. Akad. Koptyuga 1, 630090 Novosibirsk, Russia; Novosibirsk State University, ul. Pirogova 2, 630090 Novosibirsk, Russia; e-mail: elena.shap@gmail.com, shapiro@iae.nsk.su

Received 30 September 2016; revision received 21 October 2016
Kvantovaya Elektronika **46** (12) 1117–1120 (2016)
Translated by I.A. Ulitkin

the X_k symbol and $2L$ symbols around. Mathematically, it is given by the expression

$$Z_k = \tilde{Z}_k\sqrt{\sigma_{ASE}^2 + \mu\left(\frac{1}{2L+1}\sum_{i=k-L}^{k+L}|X_i|^2\right)^3}. \qquad (3)$$

According to the Shannon theorem, the capacity of a channel without memory is given by the formula

$$C = \sup I(X, Y), \qquad (4)$$

where the maximisation of $I(X, Y)$ is performed over all distributions p_X of the input alphabet; in this case, $\int |x|^2 p_X \mathrm{d}x = P$.

The throughput capacity of a regular channel with Gaussian noise (2) is given by

$$C = \lg\left(1 + \frac{P}{\sigma_{\mathrm{ASE}}^2 + \mu P^3}\right) \qquad (5)$$

in the case of a complex channel and complex Gaussian noise. If the Gaussian noise in the Z_k channel and the value of X_k are real, then

$$C = \frac{1}{2}\lg\left(1 + \frac{P}{\sigma_{\mathrm{ASE}}^2 + \mu P^3}\right), \qquad (6)$$

and the maximum mutual information is reached on the distribution

$$p_X = \frac{1}{\sqrt{2\pi P}}\exp\left(-\frac{x^2}{2P}\right). \qquad (7)$$

The lg function in equations (5) and (6) is the base 2 logarithm.

Infinite input and output alphabets cannot be realised in the numerical calculation. Next, we describe how to switch from a continuous channel to a discrete one. Instead of a real infinite alphabet $x \in (-\infty, \infty)$, we consider the finite input alphabet:

$$x_i = \Delta x(i-1) - \frac{L_x}{2}, \qquad (8)$$

where $i = 1, \ldots, n$; $\Delta x = L_x/(n-1)$; and L_x is a given interval. Let us set $x_0 = -\infty$ and $x_{n+1} = \infty$. The output alphabet can be given by the formulas

$$y_i = \Delta y(j-1) - \frac{L_y}{2}. \qquad (9)$$

Here, $j = 1, \ldots, m$; $\Delta y = L_y/(m-1)$; $y_j \in [-L_y/2, L_y/2]$; $y_0 = -\infty$; and $y_{m+1} = \infty$. Let us denote by q_i the symbol probability x_i. If we select

$$q_i = \int_{\frac{1}{2}(x_i + x_{i-1})}^{\frac{1}{2}(x_i + x_{i+1})} \frac{1}{\sqrt{2\pi P}}\exp\left(-\frac{\tau^2}{2P}\right)\mathrm{d}\tau, \qquad (10)$$

then in a discrete channel with a probabilistic rule

$$P(Y = y_j \mid X = x_i) = p_{ji}$$
$$= \frac{1}{2}\left(\mathrm{erf}\frac{y_j + y_{j+1} - 2x_i}{2\sqrt{\sigma_{\mathrm{ASE}}^2 + \mu P^3}} - \mathrm{erf}\frac{y_j + y_{j-1} - 2x_i}{2\sqrt{\sigma_{\mathrm{ASE}}^2 + \mu P^3}}\right) \qquad (11)$$

mutual information virtually coincides with the capacity of a continuous channel (6) for sufficiently large m, n, L_x and L_y. Here, $\mathrm{erf}(x) = (2/\sqrt{\pi})\int_0^x \mathrm{e}^{-t^2}\mathrm{d}t$ is the error integral.

Figure 1 shows the dependence of the continuous channel capacity (6) on the signal power P (solid curve) and mutual information of discrete channels (10), (11) (points) at $m = n = 51$, $L_x = L_y = 18$, $\sigma_{\mathrm{ASE}}^2 = 1$ and $\mu = 0.00675$. One can see that these dependences are very close.

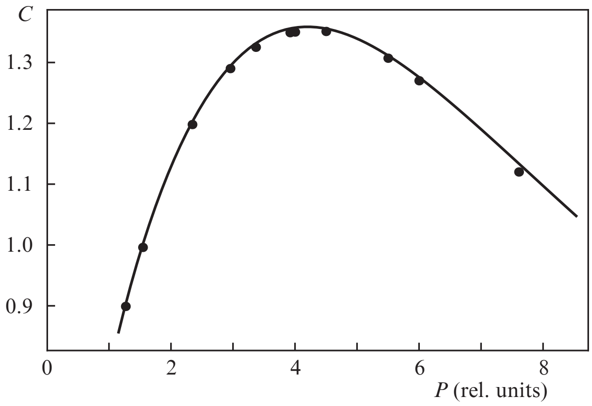

Figure 1. Dependence of the capacity on the signal power for a Gaussian channel (solid curve) and a discrete channel (points).

3. Lower limit of the capacity

A direct calculation of mutual information (1) is a technically challenging task, and so it is convenient to use an approximate, but close evaluation of mutual information. To estimate the lower limit, we used the capacity of the following auxiliary channel. Consider a fixed bit interval with number k. The value transmitted in this bit interval is a random variable, as well as a recorded value. Let q_i mean the probability of an event $X_k = x_i$. In the finite-memory model (3), the detection probability $Y_k = y_j$ depends on the nearest left L bits $(X_{k-L}, \ldots, X_{k-1})$ and right L bits $(X_{k+1}, \ldots, X_{k+L})$. Then, $L = 1$.

Let us denote by $p_{ji}(x_t, x_r)$ the event probability $Y_k = y_j$, $X_k = x_i$, $X_{k-1} = x_t$, $X_{k+1} = x_r$. Then, the conditional probability Q_{ji} of detection $Y_k = y_j$ for the transmitted value of $X_k = x_i$ is given by the expression

$$Q_{ji} = \sum_{t,r} p_{ji}(x_t, x_r) q_t q_r. \qquad (12)$$

Consider the function

$$F(q_1, \ldots, q_n) = \sum_{j,i} Q_{ji} q_i \lg\frac{Q_{ji}}{\sum_{i'} Q_{ji'} q_{i'}}, \qquad (13)$$

determining mutual information in the auxiliary channel. We show that mutual information of the auxiliary channel with an accuracy of $O(\mu^2)$ coincides with mutual information (1) of a channel with memory (3).

We introduce the function $v(\mu)$, which is given by the formula

$$v(\mu) = \sum_{j=1}^{J}\sum_{i=1}^{K} f_{ji}(\mu) Pr_i \lg\frac{f_{ji}(\mu)}{\sum_{i'=1}^{K} f_{ji'}(\mu) Pr_{i'}}, \qquad (14)$$

where $Pr_i \geqslant 0$; $\sum_i Pr_i = 1$; $f_{ji}(\mu)$ are the positive smooth functions of μ; and the equalities $\sum_{j=1}^{J} f_{ji}(\mu) = 1$ are fulfilled for all i, such that $Pr_i > 0$. Then,

$$\frac{\mathrm{d}v(\mu)}{\mathrm{d}\mu} = \sum_{j=1}^{J}\sum_{i=1}^{K}\frac{\mathrm{d}f_{ji}(\mu)}{\mathrm{d}\mu}Pr_i \lg\frac{f_{ji}(\mu)}{\sum_{i'=1}^{K} f_{ji'}(\mu)Pr_{i'}}. \quad (15)$$

To prove (15), we will use the equality

$$\sum_{j=1}^{J}\frac{\mathrm{d}f_{ji}(\mu)}{\mathrm{d}\mu} = 0.$$

Next, we show that mutual information (1) in the case of finite input and output alphabets can be represented in the form of (14). Let $\{x_i, i = 1, \ldots, n\}$ and $\{y_j, j = 1, \ldots, m\}$ be input and output alphabets, respectively. In total, there are n^N different input sequences and m^N different output sequences. We enumerate the input sequence by numbers $i = 1, \ldots, K = n^N$, and the output sequence by numbers $j = 1, \ldots, J = m^N$. Denoting by $f_{ji}(\mu)$ the probability of detecting a sequence with the number j at an initial sequence with the number i, we obtain the desired representation of (14).

Consider the function $V(N, \mu) = I(X_1, \ldots, X_N; Y_1, \ldots, Y_N)/N$. If $i = 0$, the equality $V(N, \mu, q_1, \ldots, q_n) = F(q_1, \ldots, q_n)$ is met. In addition, by using (15) and the relation $\sum_{j=1}^{m} p_{ji}(x_t, x_r) = 1$, which is valid for any fixed i, t and r, it is easy to obtain the equality $\mathrm{d}V/\mathrm{d}\mu = \mathrm{d}F/\mathrm{d}\mu$ at $\mu = 0$, $N \to \infty$.

Then we use the Taylor series expansion of the functions and find that mutual information in the finite-memory channel (3) coincides to an accuracy of $O(\mu^2)$ with $F(q_1, \ldots, q_n)$. Note that (13) can be formally regarded as an auxiliary discrete channel without memory. Essiambre et al. [5] showed that mutual information in the channel (13) is a lower boundary of mutual information (1). However, the accuracy of this estimate was not considered in the literature.

4. Numerical experiment

We have found the dependence of the auxiliary channel capacity (13) on the signal power $P = \sum_i x_i^2 q_i$ at $\mu = 0.00675$ and 0.027. The mutual information of the auxiliary channel was maximised by using the distribution of the input alphabet; the results of the procedure are shown in Fig. 2. Figures 2a and 2b show the dependences of the auxiliary channel capacity on the signal power at $\mu = 0.00675$ and 0.027, respectively. Other parameters are as follows: $m = n = 41$, $L_x = L_y = 8$ and $\sigma^2_{\mathrm{ASE}} = 1$. In both figures the upper curve is an estimate of the capacity by a regular Gaussian-noise channel (6) and the lower curve describes the auxiliary channel capacity. Note that the optimal values of the signal power for both channels are identical in Figs 2a and 2b. However, if the memory parameter μ increases, the difference between the estimates also increases.

For the finite-memory model, we used a relaxation modification [6]. The method of finding the optimal distribution is a generalisation of the Arimoto method [7] for a channel with the probabilistic rule Q_{ji}, independent of q_i, and involves three steps.

Step 0: we choose an initial set of probabilities $q_1^{(0)}, \ldots, q_n^{(0)}$ and assume $k = 0$.

Step 1: using the probability $q_1^{(k)}, \ldots, q_n^{(k)}$, we calculate the matrix

$$W_{ij}^{(k)} = \frac{q_i^{(k)} Q_{ji}}{\sum_l Q_{jl} q_l^{(k)}}.$$

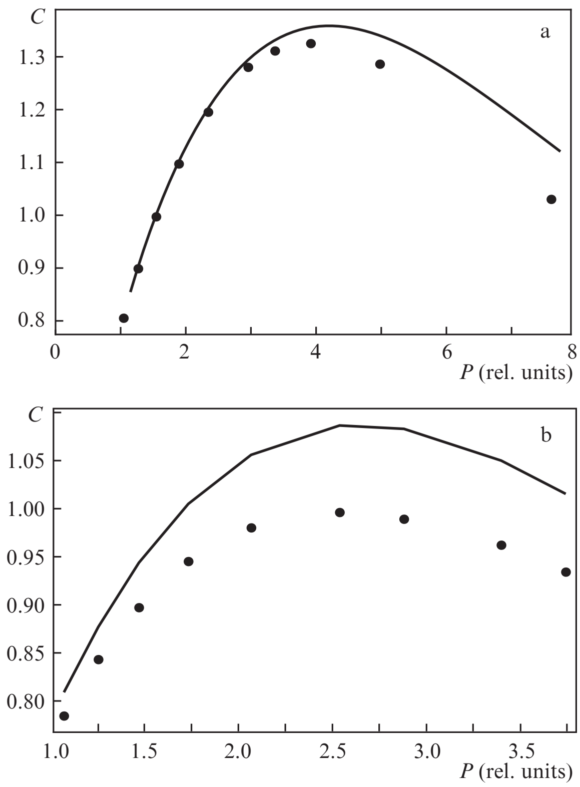

Figure 2. Dependences of the capacity on the signal power for a Gaussian channel (solid curve) and a channel with memory (points) at μ = (a) 0.00675 and (b) 0.027.

Step 2: we find a new set of probabilities $q_1^{(k+1)}, \ldots, q_n^{(k+1)}$, at which we achieve a maximum

$$I(q_1, \ldots, q_n, W^{(k)}) = \sum_{j,i} Q_{ji} q_i \lg\frac{W_{ij}^{(k)}}{q_i}.$$

Steps 1 and 2 are repeated as long as $|q_i^{(k+1)} - q_i^{(k)}|$ is less than the given ε.

The difficulty is that Q_{ji} depends on variables $q_1, \ldots, q_n$. We have used an iterative procedure [6] to find $q_1^{(k+1)}, \ldots, q_n^{(k+1)}$ in step 2. Let us denote

$$R(q_1, \ldots, q_n) = I(q_1, \ldots, q_n, W^{(k)}) - \lambda\Big(\sum_i q_i - 1\Big), \quad (16)$$

where λ is the Lagrange multiplier. The extremum $R(q_1, \ldots, q_n)$ is defined by a system of equations

$$\frac{\partial R}{\partial q_l} = \sum_{j,i}\frac{\partial Q_{ji}}{\partial q_l} q_i \lg\frac{W_{ij}^{(k)}}{q_i} + \sum_j Q_{jl} \lg W_{lj}^{(k)} - \lg q_l - 1 - \lambda, \quad (17$$

where $\sum_l q_l = 1$; $l = 1, \ldots, n$. Equations (17) can be rewritten as $T_l = 0$, where

$$T_l = \exp\Bigg(\sum_{j,i}\frac{\partial Q_{ji}}{\partial q_l} q_i \lg\frac{W_{ij}^{(k)}}{q_i} + \sum_j Q_{jl} \lg W_{lj}^{(k)} - 1 - \lambda\Bigg) - q_i. \quad (18)$$

Note that representation (18) provides an inequality $q_i \geqslant 0$. To find the desired unknowns $q_1, \ldots, q_n$, we introduce an additional parameter τ. Assuming$q_l(\tau)$ and $\lambda(\tau)$ to be functions of the variable τ, we find the solution to a system of differential equations

$$\frac{\mathrm{d}}{\mathrm{d}\tau} T_l(q_1(t), \ldots, q_n(t)) = -T_l. \tag{19}$$

The solution to a system of equations (19) converges rapidly to the solution to system (17). We will use the found values as $q_1^{(k+1)}, \ldots, q_n^{(k+1)}$. The additional restriction, $\sum_i x_i^2 q_i = P$, only slightly complicates the proposed algorithm.

5. Conclusions

Thus, we have proposed a method for calculating the capacity of a communication line with nonlinear finite memory with an accuracy of $O(\mu^2)$, where μ is the memory parameter. With this method, we have compared the capacity of a nonlinear finite-memory channel and a regular channel in which the Kerr nonlinearity is taken into account as an additional Gaussian noise. In the regular channel model, the signal-distorting noise is Gaussian. In the finite-memory model (3), the noise is a mixture of n^2 Gaussian noises with various dispersions, which makes these models different. We have shown that the optimal signal powers coincide for both methods of mutual information estimation in the communication line. However, the assessment of the capacity in a regular Gaussian-noise channel model is greater than the estimate of the capacity by the auxiliary channel taking into account the temporal distribution of the transmitted signal. With the growth of the memory parameter, the difference in estimates of the information channel capacity using the proposed models increases.

Acknowledgements. This work was supported by the Russian Science Foundation (Project No. 16-11-10133). E.G. Shapiro acknowledges the support the Ministry of Education and Science of the Russian Federation (Grant No. 14.B25.31.0003). D.A. Shapiro acknowledges the support of the RF President's Grants Council for the State Support to Leading Scientific Schools of the Russian Federation (Grant No. NSh-6898.2016.2).

References

1. Shannon C. *Bell Syst. Techn. J.*, **27**, 379; **27**, 623 (1948).
2. Splett A., Kurtzke C., Petermann K., in *Proc. 19-th ECOC* (Montereux, Switzerland, 1993).
3. Poggiolini P., Carena A., Curri V., Bosco G., Forghieri F. *IEEE Photon. Technol. Lett.*, **23**, 742 (2011).
4. Agrell E., Alvarado A., Durisi G., Karlsson G. *J. Lightwave Technol.*, **32**, 2862 (2014).
5. Essiambre R., Kramer G., Winzer P.J., Foschini G.J., Goebel B. *J. Lightwave Technol.*, **28**, 662 (2010).
6. Shapiro E.G., Shapiro D.A., Turitsyn S.K. *Opt. Express*, **23**, 15119 (2015).
7. Arimoto S. *IEEE Trans. Inf. Theory*, **18**, 14 (1972).