INFORMATION AND CONTROL **3**, 351–359 (1960)

Paradoxes Related to the Rate of Transmission of Information*

PETER SWERLING

Engineering Division, The RAND Corporation, Santa Monica, California

This paper discusses paradoxes related to the possibility of infinite information capacity of certain types of channels. First, a paradox of this type is derived which shows that such paradoxes are not necessarily dependent on the assumption of Gaussian statistics.

Next, in the case where signal and noise are assumed to be Gaussian, a different example of this type of paradox is derived; also, a necessary and sufficient condition for the avoidance of this form of the paradox is derived. This condition is shown to be satisfied in a class of plausible physical situations.

In several recent papers Good and Doog (1958, 1959, 1960) have discussed a paradox which arises in connection with the information capacity of certain types of channels. Roughly speaking, the paradox is that an arbitrarily large amount of information can be transmitted in an arbitrarily small time.

Our purpose here is, first, to derive a paradox similar to the Doog-Good paradox, utilizing the Karhunen expansion of random processes (Grenander, 1950); this approach shows that the assumption of Gaussian statistics is not essential for the occurrence of this type of paradox. Next, a necessary and sufficient condition that the paradox be avoided in the Gaussian case is derived; this condition is closely analogous to well known conditions for nonsingularity in the detection and estimation of nonstochastic signals in noise (Grenander, 1950; Kelly *et al.*, 1960; Swerling, 1959). Finally, this condition is shown to be satisfied in a class of plausible physical situations. This provides an alternative approach to the one proposed by Good (1960) for avoiding the paradox in the Gaussian case.

Suppose we have the following type of communication channel:

The input to the channel is a function of time, the "signal," which we

* The views expressed in this paper are not necessarily those of the Corporation.

shall denote by $S(t)$, $0 \leqq t \leqq T$. The channel adds noise $N(t)$ to the signal, so that the output is

$$R(t) = S(t) + N(t), \qquad 0 \leqq t \leqq T \tag{1}$$

$N(t)$ is assumed to be a random process having zero ensemble mean for each t, covariance function $\psi_N(t, t')$, which is continuous in the mean over $0 \leqq t \leqq T$. The process $N(t)$ is not assumed to be necessarily Gaussian.

The capacity of such a channel will be shown to be infinite. More precisely, it will be shown that an arbitrarily large number of bits can be transmitted with arbitrarily small probability of error, using signal wave forms having arbitrarily small power, and with arbitrarily small T.

The Karhunen expansion of the noise process $N(t)$ states that we may represent $N(t)$ as

$$N(t) = \sum_{\nu=1}^{\infty} N_\nu \sqrt{\lambda_\nu}\, \phi_\nu(t) \tag{2}$$

where the N_ν are uncorrelated random variables with mean zero and unit variance, given by

$$N_\nu = \frac{1}{\sqrt{\lambda_\nu}} \int_0^T N(t)\phi_\nu(t)\, dt \tag{3}$$

and where $\phi_\nu(t)$ and λ_ν are the orthonormal eigenfunctions and eigenvalues associated with the kernel ψ_N :

$$\int_0^T \psi_N(t, t')\phi_\nu(t')\, dt' = \lambda_\nu \phi_\nu(t) \tag{4}$$

Now, suppose we have a message to be transmitted, which for the sake of definiteness we will assume to be an infinite sequence $x_1, x_2, \cdots$, where each x_i may be either zero or one.

The transmitted signal $S(t)$ is constructed as follows: first a sequence $x_1^*, x_2^*, \cdots$, is constructed, with

$$\begin{aligned} x_\nu^* &= x_1, \qquad \nu = 1, \cdots, n \\ &= x_2, \qquad \nu = n + 1, \cdots, 2n \end{aligned} \tag{5}$$

and so forth.

The transmitted signal is then defined to be

$$S(t) = \alpha \sum_{\nu=1}^{\infty} x_\nu^* \sqrt{\lambda_\nu}\phi_\nu(t) \tag{6}$$

where α is any positive number.

Thus,

$$R(t) = S(t) + N(t) = \sum_{\nu=1}^{\infty} [\alpha x_\nu^* + N_\nu]\sqrt{\lambda_\nu}\phi_\nu(t) = \sum_{\nu=1}^{\infty} R_\nu\sqrt{\lambda_\nu}\phi_\nu(t) \quad (7)$$

The quantities $R_\nu = \alpha x_\nu^* + N_\nu$ can be recovered from $R(t)$ by

$$R_\nu = \alpha x_\nu^* + N_\nu = \frac{1}{\sqrt{\lambda_\nu}} \int_0^T R(t)\phi_\nu(t)\, dt \quad (8)$$

The sequence $\{R_\nu\}$ consists of n-fold repetitions of the symbols x_i in the original message, multiplied by α and perturbed by the noise variates N_ν. Since the noise variates are uncorrelated, with zero mean and unit variance, it is easily shown that for any value of α, it is possible to choose n sufficiently large so that the original sequence $x_1, x_2, \cdots$ can be recovered from the sequence $\{R_\nu\}$ with arbitrarily small probability of error.

This will be illustrated for the symbol x_1:
Let

$$\tilde{x}_1 = \frac{1}{n\alpha} \sum_{\nu=1}^{n} R_\nu = x_1 + \frac{1}{n\alpha} \sum_{\nu=1}^{n} N_\nu \quad (9)$$

$$\begin{aligned} \hat{x}_1 &= 0 \quad \text{if} \quad |\tilde{x}_1| < |\tilde{x}_1 - 1| \\ &= 1 \quad \text{if} \quad |\tilde{x}_1 - 1| < |\tilde{x}_1| \end{aligned} \quad (10)$$

then,

$$\text{expected value of } \tilde{x}_1 = x_1 \quad (11)$$

$$\text{variance of } \tilde{x}_1 = \frac{1}{\alpha^2 n} \quad (12)$$

By Tchebycheff's inequality, for any $\delta > 0$,

$$\text{Prob}[\,|\tilde{x}_1 - x_1| > 1/4] < \delta \quad (13)$$

for sufficiently large n. But

$$\text{Prob}[\hat{x}_1 \neq x_1] < \text{Prob}[\,|\tilde{x}_1 - x_1| > 1/4]. \quad (14)$$

This completes the proof.

We will now derive another example of this type of paradox, assuming now that both signal and noise are Gaussian random processes with zero mean, continuous in the mean, and having covariance functions $\psi_S(t, t')$ and $\psi_N(t, t')$.

Define

$$\begin{aligned} N_\nu &= \frac{1}{\sqrt{\lambda_\nu}} \int_0^T N(t)\phi_\nu(t)\,dt \\ S_\nu &= \frac{1}{\sqrt{\lambda_\nu}} \int_0^T S(t)\phi_\nu(t)\,dt \end{aligned} \tag{15}$$

where λ_ν and $\phi_\nu(t)$ are the eigenvalues and eigenfunctions associated with the noise covariance function ψ_N.

Also define, for $\mu,\nu = 1, \cdots, m$, the matrices

$$\begin{aligned} \mathbf{N}^{(m)} &= \overline{N_\mu N_\nu} = \mathbf{I}^{(m)} \\ \mathbf{S}^{(m)} &= \overline{S_\mu S_\nu} \end{aligned} \tag{16}$$

where the bar denotes expected value, and $\mathbf{I}^{(m)}$ is the m by m identity matrix.

Denote by $E^{(m)}$ the expected information in $R_1, \cdots, R_m$ relative to $S_1, \cdots, S_m$. Then,

$$E^{(m)} = \tfrac{1}{2} \log \left| \mathbf{I}^{(m)} + \mathbf{S}^{(m)} \right| \tag{17}$$

If the covariance functions of signal and noise are proportional:

$$\psi_S = \alpha \psi_N,$$

then

$$E^{(m)} = \frac{m}{2} \log (1 + \alpha). \tag{18}$$

This approaches infinity as m approaches infinity. In this form, the paradox will not arise, as m goes to infinity, provided $E^{(m)}$ remains bounded. It is also not hard to show that the boundedness of the sequence $E^{(m)}$ is sufficient to insure that the Doog-Good paradox will not arise for the case where $R(t)$ is sampled at m discrete times, and then m is made to go to infinity (as in Good and Doog, 1958).

Now, from (17),

$$E^{(m)} = \tfrac{1}{2} \sum_{\nu=1}^{m} \log[1 + \chi_\nu^{(m)}], \tag{19}$$

where $\chi_\nu^{(m)}$, $\nu = 1, \cdots, m$, are the eigenvalues of the matrix $\mathbf{S}^{(m)}$.

It is a simple matter to prove that a necessary and sufficient condition for the boundedness of $E^{(m)}$ is the boundedness of the sequence trace $\mathbf{S}^{(m)} = \sum_{\nu=1}^{m} \chi_\nu^{(m)}$. But

$$\text{trace } \mathbf{S}^{(m)} = \sum_{\nu=1}^{m} \chi_\nu^{(m)} = \sum_{\nu=1}^{m} \overline{S_\nu^2} \tag{20}$$

Thus, a necessary and sufficient condition for the boundedness of $E^{(m)}$, and hence for the avoidance of the Doog-Good paradox in the Gaussian case, is

$$\sum_{\nu=1}^{\infty} \overline{S_\nu^2} < \infty. \tag{21}$$

If we define

$$s_\nu = \sqrt{\lambda_\nu} S_\nu = \int_0^T S(t)\phi_\nu(t)\, dt \tag{22}$$

then condition (21) becomes

$$\sum_{\nu=1}^{\infty} \frac{1}{\lambda_\nu} \overline{s_\nu^2} < \infty \tag{23}$$

where

$$\overline{s_\nu^2} = \int_0^T \int_0^T \psi_S(t, t')\phi_\nu(t)\phi_\nu(t')\, dt\, dt' \tag{24}$$

In the form (23), this condition is seen to be the direct analog of well known conditions for nonsingularity in the detection and estimation of nonstochastic signals in Gaussian noise (Grenander, 1950; Kelly *et al.*, 1960; Swerling, 1959).

A few additional facts are worth noting: $E^{(m)}$ is a monotone increasing sequence; consequently, if (21) holds, $E^{(m)}$ approaches a finite limit E:

$$E = \lim_{m \to \infty} E^{(m)} < \infty \tag{25}$$

It is reasonable to define E to be the expected information in the random process $R(t)$ relative to $S(t)$. In fact, let R_ν^*, $\nu = 1, \cdots, m$, be any finite collection of random variables obtained by linear operations on the random process $R(t)$, such as, for example, the values of $R(t)$ at m time points; and let S_ν^* be the variables obtained by the same linear operations on the process $S(t)$. Then the expected information in R_ν^*, $\nu = 1, \cdots, m$ relative to S_ν^*, $\nu = 1, \cdots, m$ is not larger than E.

The fulfillment of condition (23) would no doubt preclude the type of signal used in the first part of this paper in our original derivation of the paradox.

It remains to exhibit a physically plausible class of cases in which (21) can be shown to hold. We will give two slightly different examples.

First, we assume that $S(t)$ and $N(t)$ are stationary processes, obtained in the following way: We suppose there is an "original" signal process $h(t)$, having spectral density $H(\omega)$, where $H(\omega)$ has finite integral. Suppose $S(t)$ is obtained by passing $h(t)$ through a linear filter with impulse response function $w(t)$:

$$S(t) = \int_0^\infty w(\tau)h(t - \tau)\, d\tau \tag{26}$$

Let the spectral densities of $S(t)$ and $N(t)$ be denoted by $F_S(\omega)$ and $F_N(\omega)$. Then

$$F_S(\omega) = | W(\omega) |^2 H(\omega) \tag{27}$$

where

$$W(\omega) = \int_0^\infty w(t)e^{-i\omega t}\, dt \tag{28}$$

It will also be assumed that

$$F_N(\omega) = | W(\omega) |^2 \tag{29}$$

This is equivalent to assuming that $N(t)$ is obtained by passing white noise through the same linear filter through which $h(t)$ was passed.

As before, it is assumed that the received signal $R(t)$ is observed over a finite interval, $0 \leqq t \leqq T$.

Theorem. *A sufficient condition that (23) hold is that $w(t)$ vanish outside a finite interval.*

Proof: Suppose $w(t) = 0$ for $t > t_0$ and, of course, for $t < 0$. Then,

$$S(t) = \int_0^{t_0} w(\tau)h(t - \tau)\, d\tau \tag{30}$$

Define

$$\begin{aligned} h^*(t) &= h(t), \qquad -t_0 \leqq t \leqq T \\ &= 0, \quad \text{otherwise} \end{aligned} \tag{31}$$

Then, within the observation interval $0 \leqq t \leqq T$,

$$S(t) = \int_0^{t_0} w(\tau)h^*(t - \tau)\, d\tau, \qquad 0 \leqq t \leqq T \tag{32}$$

Let

$$S^*(t) = \int_0^{t_0} w(\tau)h^*(t - \tau)\, d\tau, \quad \text{all } t \tag{33}$$

then

$$\begin{aligned} S^*(t) &= S(t), \qquad 0 \leqq t \leqq T \\ &= 0, \qquad \mid t \mid > t_0 + T \end{aligned} \tag{34}$$

Also, $S^*(t)$ can be written

$$S^*(t) = \int_{-\infty}^{\infty} w(\tau)h^*(t - \tau)\, d\tau \tag{35}$$

Therefore, if

$$U(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S^*(t)e^{-i\omega t}\, dt \tag{36}$$

$$V(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} h^*(t)e^{-i\omega t}\, dt \tag{37}$$

Then,

$$U(\omega) = V(\omega)W(\omega) \tag{38}$$

Also, from (22) and (35),

$$s_\nu = \int_0^T S^*(t)\phi_\nu(t)\, dt \tag{39}$$

Consequently, as has been proved by Kelly *et al.* (1960), Appendix I,

$$\begin{aligned} \sum_{\nu=1}^{\infty} \frac{1}{\lambda_\nu} s_\nu^{\,2} &\leqq \int_{-\infty}^{\infty} \frac{\mid U(\omega) \mid^2}{F_N(\omega)}\, d\omega \\ &= \int_{-\infty}^{\infty} \frac{\mid W(\omega) \mid^2 \mid V(\omega) \mid^2}{W \mid (\omega) \mid^2}\, d\omega \\ &= \int_{-\infty}^{\infty} \mid V(\omega) \mid^2\, d\omega \end{aligned} \tag{40}$$

The proof of (40) by Kelly *et al.* (1960) uses the complex waveform notation, so that in order to apply their proof it is first necessary to set their ω_0 equal to zero; also, their function $R(s)$ is, for $\omega_0 = 0$, twice the covariance function of the real noise process, so that their λ_ν are twice our λ_ν. That is why the integral in our Eq. (40) is extended from $-\infty$ to ∞, while in their Eqs. (1–18) it is extended from 0 to ∞. (In addition, of course, there are a number of other differences in notation.)

By Parseval's theorem,

$$\begin{aligned} \int_{-\infty}^{\infty} | V(\omega) |^2 \, d\omega &= \int_{-\infty}^{\infty} | h^*(t) |^2 \, dt \\ &= \int_{-t_0}^{T} | h^*(t) |^2 \, dt \end{aligned} \tag{41}$$

Thus, from (40) and (41),

$$\begin{aligned} \sum_{\nu=1}^{\infty} \frac{1}{\lambda_\nu} \overline{s_\nu^2} &\leqq \int_{-t_0}^{T} \overline{| h^*(t) |^2} \, dt \\ &= (T + t_0) \int_{-\infty}^{\infty} H(\omega) \, d\omega \end{aligned} \tag{42}$$

It is highly probable that the restriction that $w(t)$ vanish outside a finite interval is not necessary. On the other hand, some restriction on $w(t)$ is necessary; for example, a restriction sufficient to insure that $F_N(\omega)$ does not vanish outside a finite interval.

The second example is as follows: suppose $S(t)$ is obtained from $h(t)$ by Eq. (26), but $h(t)$ is assumed to be a random process having finite total energy between $-\infty$ and T:

$$\int_{-\infty}^{T} \overline{| h(t) |^2} \, dt < \infty \tag{43}$$

(thus, $h(t)$ cannot be stationary). As before, assume $N(t)$ is a stationary process with spectral density $| W(\omega) |^2$.

Then, (23) holds without any restrictions on $w(t)$ other than that $w(t) = 0$ for $t < 0$.

Since $\frac{1}{2} \sum_{\nu=1}^{\infty} (1/\lambda_\nu)\overline{s_\nu^2}$ is actually an upper bound for

$$E = \lim_{m\to\infty} E^{(m)},$$

one half the quantity on the right-hand side of (42) provides an upper bound for E in our first example. Similarly, one half the quantity on the left of (43) provides an upper bound for E in the second example.

RECEIVED: September 27, 1960.

REFERENCES

GOOD, I. J., AND CAJ DOOG, K. (1958). A paradox concerning rate of information, *Information and Control* **1,** 113–126.

GOOD, I. J., AND CAJ DOOG, K. (1959). A paradox concerning rate of information, *Information and Control* **2,** 195–197.

GOOD, I. J. (1960). Effective sampling rates for signal detection, *Information and Control* **2,** 116–140.

GRENANDER, U. (1950). Stochastic processes and statistical inference, *Ark. Mat.* **1,** 195–277.

KELLY, E. J., REED, I. S., AND ROOT, W. L. (1960). The detection of radar echoes in noise. I, *Soc. Ind. Appl. Math.* **8,** 309–341.

SWERLING, P. (1959). Parameter estimation for waveforms in additive Gaussian noise, *J. Soc. Ind. Appl. Math.* **7,** 152–166.