MASSACHUSETTS INSTITUTE OF TECHNOLOGY
LINCOLN LABORATORY

CODING FOR CONSTANT-DATA-RATE SYSTEMS

PART I. A NEW ERROR-CORRECTING CODE

RICHARD A. SILVERMAN
MARTIN BALSER

23 OCTOBER 1953

TECHNICAL REPORT NO. 39

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

LINCOLN LABORATORY

CODING FOR CONSTANT-DATA-RATE SYSTEMS

PART I. A NEW ERROR-CORRECTING CODE

Richard A. Silverman

Martin Balser

Group 34

Technical Report No. 39

23 October 1953

ABSTRACT

A study of coding to reduce the frequency of errors in communication systems which transmit data at a constant rate has been started. A new single-error-correcting code (the Wagner code) is described and analyzed. Its performance in a constant-data-rate system is evaluated and compared with Hamming's single-error-correcting code. The Wagner code is superior for many communication applications. Its successful implementation does not require too precise equipment.

CAMBRIDGE

MASSACHUSETTS

CODING FOR CONSTANT-DATA-RATE SYSTEMS
PART I. A NEW ERROR-CORRECTING CODE

A. INTRODUCTION

In this report, we shall consider a communication system in which data consisting of sequences (known as words) of binary digits are transmitted at a predetermined constant rate. (For our purposes, a binary digit is one of two electrical signals of duration T and bandwidth W.*) The nature of the data and the manner in which they are translated into words are irrelevant to this discussion. For example, the data may be English letters, numbers, etc., reduced to sequences of five binary digits each for use in a teletype system, or they may be conventional symbols representing entire messages.

A basic problem of coding is to reduce the average rate of incorrectly received words as much as possible. Accordingly, additional digits are added to the word for the purposes of error detection or correction. The assumption that the words are sent at a constant rate requires that each binary digit be shortened by such an amount that the coded words (message digits plus check digits) have the same duration as the original uncoded words. This shortening of each digit increases its probability of error and, consequently, the probability of error per word. On the other hand, the coding imposes constraints on the digits composing a word, so that errors may show up as inconsistencies and may in many cases be corrected. This tends to reduce the probability of error per word. The efficacy of a code depends on how much the second effect outweighs the first.

The simplest code of all consists of an extra digit selected to make the sum of all the digits in the coded word even (or odd). If the sum of the digits of the received word has the wrong parity, an odd number of errors is known to have been made. There is, however, no indication of the correct replacement for the mistaken word, unless some dependence between separate words (such as the redundancy of printed English[1]) is exploited.

Hamming[2] has devised a code that corrects all single errors. It consists of adding k suitably chosen check digits to the m message digits.** If another digit is added, double errors can be detected as well as single errors corrected.[2] In this paper, we describe a new single-error-correcting code (the Wagner code) and evaluate its performance in a constant-data-rate system, particularly as compared with that of the Hamming code.

Recently, multiple-error-correcting codes have been constructed.[3,4] An analysis of their performance in constant-data-rate systems will be published in Technical Report No. 48, "Coding for Constant-Data-Rate Systems, Part II – Multiple-Error-Correcting Codes."

B. DESCRIPTION OF THE WAGNER CODE

In this study, we are concerned with communication systems that transmit words consisting of binary digits. A binary digit is one of two electrical signals $x_1(t)$ and $x_2(t)$ of duration T and bandwidth W. Let $p(x_i/y)$ be the (a posteriori) probability that if y is received, x_i was

*These two requirements are, strictly speaking, incompatible (see Appendix).

**For m = 2 to 4, k = 3; for m = 5 to 11, k = 4; etc.[2]

sent, and let Δp be $p(x_1/y) - p(x_2/y)$. In the absence of any constraints on the digits composing a word or of dependence between the words themselves, the receiver can compute only $p(x_1/y)$ and $p(x_2/y)$, and for each digit choose x_1 or x_2, depending on whether $p(x_1/y)$ or $p(x_2/y)$ is the larger (or, equivalently, whether Δp is positive or negative). The error-correcting code that we shall describe (named the Wagner code after C. A. Wagner of this laboratory, who suggested the basic idea) enables us to use some of the information presented by the magnitudes of the Δp's* by introducing a constraint on the digits composing a word. This information is ignored by more conventional codes.

In the Wagner code, a transmitted word consists of a sequence of m message digits and an additional digit used as a parity check. As each of the perturbed digits y arrives at the receiver, the a posteriori probabilities $p(x_1/y)$ and $p(x_2/y)$ are calculated. Each digit of the received sequence is tentatively identified as x_1 or x_2, depending on whether $p(x_1/y)$ or $p(x_2/y)$ is the larger, and the values of the a posteriori probabilities are stored in a memory for the duration of a word. The sequence thus obtained is checked for parity. If the parity is correct, the word is printed as received. If the parity check fails, the digit for which the difference Δp between a posteriori probabilities is the smallest is considered the digit most in doubt, and the word is printed with this digit altered. The receiver then clears the stored values of the probability differences from the memory and proceeds to the next word.

Thus we may characterize the Wagner code as one which probably corrects single errors. (Multiple errors are always printed incorrectly.) However, as we shall see, it can be more effective in a constant-data-rate system than a code that corrects all single errors (such as the Hamming code).

The a posteriori probabilities $p(x_1/y)$ and $p(x_2/y)$ are functions of the random received waveform y(t) and therefore are themselves random variables. The calculation of their distributions is, in general, very difficult. For simplicity, we shall consider the case where the two transmitted signals have equal energy and equal a priori probabilities and are perturbed by the addition of white Gaussian noise. It has been shown[5] that for this case

$$p(x_1/y) = \beta \exp\left[\gamma \int_0^T x_1(t)\, y(t)\, dt\right] \tag{1}$$

and

$$p(x_2/y) = \beta \exp\left[\gamma \int_0^T x_2(t)\, y(t)\, dt\right] ,$$

where β and γ are constants. Thus the transmitted signal with the larger correlation has the larger a posteriori probability. Equivalently, we may write

$$\frac{p(x_1/y)}{p(x_2/y)} = \exp\left[\gamma(z_1 - z_2)\right] , \tag{2}$$

where

$$z_1 = \int_0^T x_1(t)\, y(t)\, dt \qquad \text{and} \qquad z_2 = \int_0^T x_2(t)\, y(t)\, dt . \tag{3}$$

* The signs of the Δp's are used in making the tentative identification of the transmitted word.

From Eq.(2), we see that the smaller the difference $\Delta z = z_1 - z_2$ between the correlation integrals, the closer to unity the ratio of a posteriori probabilities and, consequently, the smaller the difference Δp between the two probabilities. Thus, if the parity check fails, the digit that should be changed (as the one most in doubt) is the one for which Δz is the smallest.

It is shown in the Appendix that z_1 and z_2 are normally distributed random variables (with means c_1 and c_2, and variances σ_1 and σ_2), so that calculations are especially simple for the correlation detector.* Moreover, under the assumptions made above, the correlation detector is equivalent to the probability detector. Therefore, it is assumed in what follows that detection is by correlation.

C. ANALYSIS OF THE WAGNER CODE

1. Probability of Error Per Digit

As noted above, the correlation integrals z_1 and z_2 (corresponding to the signal that was sent and the signal that was not sent, respectively) are random variables with probability densities

$$W(z_1) = \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left[-\frac{(z_1 - c_1)^2}{2\sigma_1^2}\right] , \tag{4}$$

and

$$W(z_2) = \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left[-\frac{(z_2 - c_2)^2}{2\sigma_2^2}\right] . \tag{5}$$

It is shown in the Appendix that if x_1 and x_2 are suitably chosen, then z_1 and z_2 may be regarded as statistically independent random variables. Accordingly, the probability density of finding a separation $\Delta z = z_1 - z_2$ is

$$W(\Delta z) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(\Delta z - \Delta c)^2}{2\sigma^2}\right] , \tag{6}$$

where $\Delta c = c_1 - c_2 > 0$ and $\sigma^2 = \sigma_1^2 + \sigma_2^2$. Here we make use of a well-known theorem on the distribution of the sum of independent, normally distributed random variables.[6] If Δz is negative, then selecting as the transmitted signal the signal giving the larger correlation integral will result in an error. Thus the probability of error per digit is

$$p(a) = \int_{-\infty}^{0} W(\Delta z)\, d\Delta z = \frac{1}{2}(1 - \operatorname{erf} a) , \tag{7}$$

where $a = \Delta c/\sqrt{2}\sigma$.** The parameter a, which is proportional to the signal-to-noise ratio of the

---

*For simplicity of notation, we shall always use the subscript "one" for the signal that was transmitted. Thus $c_1 > c_2$, and $z_1 < z_2$ results in an error.

**The existence of the simple expression (7) for $p(a)$ was pointed out by Dr. I. S. Reed, who evaluated the equivalent form

$$p(a) = \frac{1}{2\pi\sigma_1\sigma_2}\int_{-\infty}^{\infty} \exp\left[-\frac{(s - c_2)^2}{2\sigma_2^2}\right]\int_{-\infty}^{s} \exp\left[-\frac{(t - c_1)^2}{2\sigma_1^2}\right] dt\, ds ,$$

by a rotation of 45° in the (s,t) plane.

correlator difference, is the significant parameter in the calculations that follow.

Since we do not know which digit was actually sent, we do not know the sign of Δz. From Eq.(6), the joint probability that the correlator difference lies between $|\Delta z|$ and $|\Delta z| + d|\Delta z|$, and that the larger correlation integral corresponds to the transmitted signal, is $d|\Delta z|$ times

$$W(|\Delta z|,\text{right}) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(|\Delta z| - \Delta c)^2}{2\sigma^2}\right] \quad . \tag{8}$$

On the other hand, the joint probability that the correlator difference lies between $|\Delta z|$ and $|\Delta z| + d|\Delta z|$, and that the larger correlation integral does <u>not</u> correspond to the transmitted signal, is $d|\Delta z|$ times

$$W(|\Delta z|,\text{wrong}) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(|\Delta z| + \Delta c)^2}{2\sigma^2}\right] \quad . \tag{9}$$

2. <u>Probability of Correcting a Single Error</u>

Suppose that the received word has n digits. Since the parity check fails if a single error is made, and since then the Wagner code changes the digit with the smallest $|\Delta z|$, the probability $\Pi_n(a)$ that a single error is made and is corrected by the Wagner code is just the probability that the digit with the smallest $|\Delta z|$ is incorrect and that the $n-1$ other digits are correct. $\Pi_n(a)$ can be calculated as follows.

Let $|\Delta z_i|$ be the correlator difference for the i-th digit. Since the $|\Delta z_i|$ are independent random variables, the joint probability that the first digit, with correlator difference between $|\Delta z_1|$ and $|\Delta z_1| + d|\Delta z_1|$, is wrong, and that all the other digits, with correlator differences between $|\Delta z_2|$ and $|\Delta z_2| + d|\Delta z_2|$, ..., $|\Delta z_n|$ and $|\Delta z_n| + d|\Delta z_n|$, are right, is

$$W(|\Delta z_1|,\text{wrong}) \prod_{i=2}^{n} W(|\Delta z_i|,\text{right}) \prod_{i=1}^{n} d|\Delta z_i| \quad . \tag{10}$$

Thus the joint probability that the first digit is wrong, that the $n-1$ other digits are correct, and that $|\Delta z_1| < |\Delta z_2| < \ldots < |\Delta z_n|$ is

$$\begin{aligned} &\int_0^\infty W(|\Delta z_n|,\text{right})\, d|\Delta z_n| \int_0^{|\Delta z_n|} W(|\Delta z_{n-1}|,\text{right})\, d|\Delta z_{n-1}| \\ &\ldots \int_0^{|\Delta z_3|} W(|\Delta z_2|,\text{right})\, d|\Delta z_2| \int_0^{|\Delta z_2|} W(|\Delta z_1|,\text{wrong})\, d|\Delta z_1| \quad . \end{aligned} \tag{11}$$

Since there are in all n! orderings of the n correlator differences, the joint probability $\Pi_n(a)$ that a single error is made in a word of n digits and that it is corrected by the Wagner code is given by

$$\begin{aligned} \Pi_n(a) = n! \int_0^\infty \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2\sigma^2}(|\Delta z_n| - \Delta c)^2\right] d|\Delta z_n| \\ \int_0^{|\Delta z_n|} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2\sigma^2}(|\Delta z_{n-1}| - \Delta c)^2\right] d|\Delta z_{n-1}| \end{aligned}$$

$$\cdots \int_0^{|\Delta z_3|} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2\sigma^2}\,(|\Delta z_2| - \Delta c)^2\right] d|\Delta z_2|$$

$$\int_0^{|\Delta z_2|} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2\sigma^2}\,(|\Delta z_1| + \Delta c)^2\right] d|\Delta z_1|$$

$$= \frac{n!}{(\sqrt{\pi})^n} \int_0^{\infty} \exp\left[-(x_n - a)^2\right] dx_n \int_0^{x_n} \exp\left[-(x_{n-1} - a)^2\right] dx_{n-1}$$

$$\cdots \int_0^{x_3} \exp\left[-(x_2 - a)^2\right] dx_2 \int_0^{x_2} \exp\left[-(x_1 + a)^2\right] dx_1 \quad , \tag{12}$$

where $a = \Delta c/\sqrt{2}\sigma$ as in Eq.(7). The conditional probability that a single error will be corrected by the Wagner code, assuming that a single error has been made in an n-digit word, is given by

$$\frac{\Pi_n(a)}{n q^{n-1}(a) p(a)} \tag{13}$$

where $p(a)$ is the probability of error per digit

$$p(a) = \frac{1}{2}\,(1 - \operatorname{erf} a) \quad , \tag{7}$$

and

$$q(a) = 1 - p(a) = \frac{1}{2}\,(1 + \operatorname{erf} a) \quad . \tag{14}$$

Equation (13) follows immediately from the definition of conditional probability and the fact that $n q^{n-1}(a) p(a)$ is the probability of a single error in an n-digit word.

It is desirable to have $\Pi_n(a)$ in a form more suitable for numerical computations. Thus we reduce the n-fold multiple integral (12) to a convenient recurrence formula. Let

$$F_n(a) \equiv \frac{(\sqrt{\pi})^n}{n!}\, \Pi_n(a) \quad , \tag{15}$$

and

$$f_n(x,a) \equiv \int_0^{x} dx_n \exp\left[-(x_n - a)^2\right] \int_0^{x_n} dx_{n-1} \exp\left[-(x_{n-1} - a)^2\right]$$

$$\cdots \int_0^{x_3} \exp\left[-(x_2 - a)^2\right] dx_2 \int_0^{x_2} \exp\left[-(x_1 + a)^2\right] dx_1 \quad . \tag{16}$$

From Eqs.(15) and (16) it follows that

$$f_n(\infty,a) = \Pi_n(a); \quad f_n(0,a) = 0 \quad , \tag{17}$$

and

$$F_n(a) = \int_0^{\infty} dx_n \exp\left[-(x_n - a)^2\right] f_{n-1}(x_n,a) \quad . \tag{18}$$

Since

$$\frac{d}{dx}\, f_n(x,a) = \exp\left[-(x - a)^2\right] f_{n-1}(x,a) \tag{19}$$

and

$$\exp\left[-(x-a)^2\right] = \frac{\sqrt{\pi}}{2}\,\frac{d}{dx}\,\mathrm{erf}\,(x-a) \quad , \tag{20}$$

we can reduce Eq.(18) by repeated integration by parts to

$$\begin{aligned} F_n(a) = {} & \frac{\sqrt{\pi}}{2}\,F_{n-1}(a) - \left(\frac{\sqrt{\pi}}{2}\right)^2 \frac{1}{2}\,F_{n-2}(a) + \left(\frac{\sqrt{\pi}}{2}\right)^3 \frac{1}{3!}\,F_{n-3}(a) + \ldots \\ & + (-1)^{n-1}\left(\frac{\sqrt{\pi}}{2}\right)^{n-2} \frac{1}{(n-2)!}\,F_2(a) \\ & + (-1)^n \left(\frac{\sqrt{\pi}}{2}\right)^{n-1} \frac{1}{(n-1)!}\left\{\sqrt{\pi}\,p(a) - \int_0^\infty \mathrm{erf}^{n-1}(x-a)\exp\left[-(x+a)^2\right]\,dx\right\} \quad . \end{aligned} \tag{21}$$

Finally, by the definition [Eq.(15)],

$$\begin{aligned} \Pi_n(a) = {} & \frac{1}{2}\binom{n}{n-1}\Pi_{n-1}(a) - \frac{1}{2^2}\binom{n}{n-2}\Pi_{n-2}(a) + \ldots + \frac{(-1)^{n-1}}{2^{n-2}}\binom{n}{2}\Pi_2(a) \\ & + \frac{(-1)^n}{2^{n-1}}\binom{n}{1}\left[p(a) - \frac{1}{2}\,I_n(a)\right] \quad . \end{aligned} \tag{22}$$

where

$$I_n(a) = \frac{2}{\sqrt{\pi}}\int_0^\infty \mathrm{erf}^{n-1}(x-a)\exp\left[-(x+a)^2\right]\,dx \quad . \tag{23}$$

$I_n(a)$ and $\Pi_n(a)$ have been computed for values of n from 2 to 9; the results are tabulated in Tables I and II for selected values of a.

TABLE I

Values of $I_n(a)$ [see Eq. (23)].

n \ a	1.0	1.5	2.0	3.0
2	−0.09943	−0.03062	−0.004604	−0.00002210
3	0.07161	0.02797	0.004533	0.00002210
4	−0.05208	−0.02571	−0.004466	−0.00002210
5	0.03932	0.02375	0.004401	0.00002210
6	−0.02979	−0.02201	−0.004339	−0.00002209
7	0.02304	0.02046	0.004279	0.00002209
8	−0.01782	−0.01906	−0.004221	−0.00002209
9	0.01400	0.01779	0.004165	0.00002209

TABLE II

Values of $\Pi_n(a)$, the joint probability that a single error is made in an n-digit word and that it is corrected by the Wagner code [see Eq. (22)].

n \ a	1.0	1.5	2.0	3.0
2	0.128	0.0323	0.00464	0.0000221
3	0.160	0.0462	0.00691	0.0000331
4	0.181	0.0588	0.00914	0.0000442
5	0.193	0.0704	0.01134	0.0000552
6	0.199	0.0811	0.01350	0.0000663
7	0.200	0.0908	0.01564	0.0000773
8	0.199	0.0997	0.01774	0.0000884
9	0.195	0.1079	0.01982	0.0000994

It can be shown from the original integral forms [Eqs.(23) and (12)] that

$$I_n(0) = \frac{1}{n} \quad , \tag{24}$$

and

$$\Pi_n(0) = \frac{1}{2^n} \quad . \tag{25}$$

Equation (25) may also be derived by the following argument. Since $\alpha = 0$ corresponds to a mean value of zero for the correlator differences, any digit is equally likely to be right or wrong. The probability of one incorrect digit and $n - 1$ correct ones is thus $n/2^n$, and the probability that the incorrect digit has the smallest correlator difference is just $1/n$ of this, giving Eq.(25).

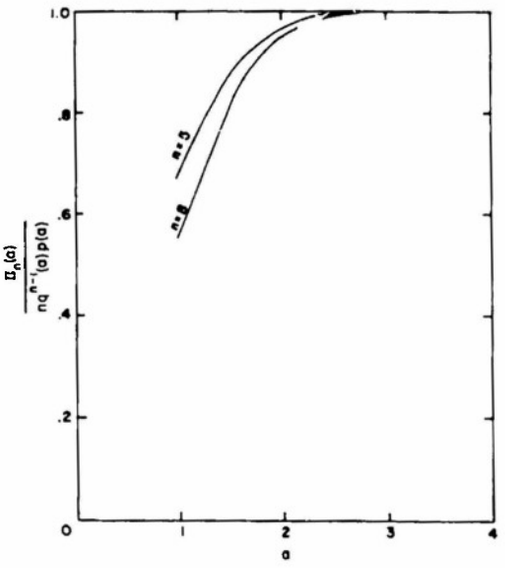

Fig. 1. Conditional probability of correction for Wagner-coded words of 5 and 8 digits.

For large α, we obtain the asymptotic forms

$$I_n(\alpha) \sim (-1)^{n-1} (1 - \operatorname{erf} \alpha) \quad , \tag{26}$$

and

$$\Pi_n(\alpha) \sim n \frac{1 - \operatorname{erf} \alpha}{2} = np(\alpha) \quad . \tag{27}$$

Equation (27) is the form to be expected. For large α, $q(\alpha) \sim 1$, so that $np(\alpha) q^{n-1}(\alpha)$, the probability of a single error in an n-digit word, is approximately $np(\alpha)$. Moreover, as we shall see below, for large α the wrong digit is almost certainly corrected by the Wagner code. Therefore, for large α

$$\Pi_n(\alpha) \sim np(\alpha) \quad . \tag{28}$$

From inspection of Tables I and II, it is seen that the asymptotic forms [Eqs.(26) and (27)] are excellent approximations for $\alpha \geq 3$.

The conditional probability that if one error is made it will be corrected by the Wagner code is given by Eq.(13). Curves of this probability as a function of α for words of 5 and 8 digits are shown in Fig. 1. The conditional probability approaches unity for large α, i.e., for large signal-to-noise ratio of the correlator differences, almost all errors are corrected. Below values of α of about 1.5, the conditional probability drops sharply, becoming $1/n$ for $\alpha = 0$ [see Eq.(25)]. As is to be expected, the conditional probability for a given α decreases with increasing n.

D. PROBABILITY OF ERROR FOR WAGNER-CODED WORDS – COMPARISON WITH UNCODED AND HAMMING-CODED WORDS

The probability of error per word for a Wagner-coded word containing m message digits ($n = m + 1$ digits in all) is

$$P_W = 1 - q^{m+1}(\alpha) - \Pi_{m+1}(\alpha) \quad , \tag{29}$$

that is, the probability of error is one minus the sum of the probability that the word is received

correctly and the probability that a single error is made and then corrected. We wish to compare P_W with P_U, the probability of error per word if no code is used, and P_H, the probability of error per word if the Hamming single-error-correcting code is used. Since we are concerned with constant-data-rate systems, the duration of the transmitted signals must be altered if coded words (message digits plus error-correcting digits) are to have the same duration as differently coded or uncoded words. Changing the signal duration changes the variance of the correlator difference and consequently the value of the parameter α and the probability of error per digit [see Eq.(7)]. For large TW (the only case of practical interest), it can be shown[7] that the variance of the output of any averaging device is proportional to T^{-1}. Thus σ_1, σ_2 and $\sigma = \sqrt{\sigma_1^2 + \sigma_2^2}$ are all proportional to $T^{-\frac{1}{2}}$, so that $\alpha \left(= \Delta c/\sqrt{2}\sigma\right)$ is proportional to $T^{\frac{1}{2}}$.* Using this result, we find that, for the same value of α used in Eq.(29),

$$P_U = 1 - q^m\left(\sqrt{\frac{m+1}{m}}\,\alpha\right) \tag{30}$$

and

$$P_H = 1 - q^{m+k}\left(\sqrt{\frac{m+1}{m+k}}\,\alpha\right) - (m+k)\,q^{m+k-1}\left(\sqrt{\frac{m+1}{m+k}}\,\alpha\right) p\left(\sqrt{\frac{m+1}{m+k}}\,\alpha\right) , \tag{31}$$

where k is the number of check digits required by the Hamming code (see above). The three quantities [Eqs.(29), (30) and (31)] have been computed for values of m from 4 to 8 and for selected values of α. The results are given in Table III.

TABLE III

Probabilities of error per word for uncoded, Hamming-coded, and Wagner-coded words containing m message digits

m	α	P_U	P_H	P_W
4	1.0	0.209	0.191	0.143
	1.5	0.0349	0.0248	0.0115
	2.0	0.00313	0.00145	0.00030
	3.0	42×10^{-7}	6×10^{-7}	$< 10^{-7}$
5	1.0	0.269	0.310	0.190
	1.5	0.0493	0.0513	0.0164
	2.0	0.00486	0.00375	0.00045
	3.0	84×10^{-7}	25×10^{-7}	$< 10^{-7}$
6	1.0	0.325	0.335	0.236
	1.5	0.0641	0.0530	0.0220
	2.0	0.00673	0.00346	0.00062
	3.0	138×10^{-7}	17×10^{-7}	$< 10^{-7}$
7	1.0	0.377	0.362	0.282
	1.5	0.0789	0.0552	0.0281
	2.0	0.00871	0.00330	0.00082
	3.0	201×10^{-7}	12×10^{-7}	$< 10^{-7}$
8	1.0	0.425	0.388	0.326
	1.5	0.0937	0.0580	0.0347
	2.0	0.01075	0.00322	0.00103
	3.0	272×10^{-7}	9×10^{-7}	$< 10^{-7}$

*The result that σ is proportional to $T^{-\frac{1}{2}}$ requires that the integrals in Eq. (3) be normalized by dividing them by T. The result that α is proportional to $T^{\frac{1}{2}}$ is independent of the normalization.

There is only a certain range of values of the signal-to-noise ratio of the correlator difference for which it is worth the effort to implement either of the error-correcting codes. For high signal-to-noise ratio, very few errors are made, and additional equipment is generally not justifiable. On the other hand, for low signal-to-noise ratio, multiple errors become too frequent, and single-error-correcting codes are of little use. Thus single-error-correcting codes are of considerable value for values of a from about 1.0 to 3.0.

As shown in Table III, in this range of values the probability of error of Wagner-coded words is considerably less than that of Hamming-coded words. For increasing word length, the advantage of the Wagner code diminishes from two causes: (1) the ratio k/m decreases with increasing m so that the length of the digits in the Hamming-coded word approaches those of the Wagner-coded words, thus narrowing the gap between the corresponding signal-to-noise ratios; and (2) the conditional probability of correcting a single error decreases with increasing m. Nonetheless, even for m = 8 and a = 2, we may expect only 103 errors per 100,000 words using the Wagner code, as compared with 322 errors per 100,000 Hamming-coded words and 1,075 errors per 100,000 uncoded words.

E. IDENTIFICATION OF THE SMALLEST CORRELATOR DIFFERENCE

Until now we have assumed that the receiver can pick out the smallest correlator difference with infinite precision. Suppose, however, that the equipment used in implementing the Wagner code is such that the smaller of two correlator differences within $\epsilon\,\Delta c$ of each other cannot be identified with certainty. Let $P_{W,\epsilon}$ be the revised probability of error per word, assuming that single errors remain uncorrected whenever any correlator difference lies within $\epsilon\,\Delta c$ of the smallest. In this notation, the P_W used above is $P_{W,0}$.*†

Suppose that the first digit has the smallest correlator difference $|\Delta z_1|$ and is known to be the only one in error. Then the probability that $|\Delta z_i|$, the correlator difference of the i-th digit ($i \neq 1$), lies within $\epsilon\,\Delta c$ of $|\Delta z_1|$ is given by

$$\text{prob}\,(|\Delta z_1| < \Delta z_i < |\Delta z_1| + \epsilon\,\Delta c / \Delta z_i > |\Delta z_1|)$$

$$= \frac{1}{\text{prob}(\Delta z_i > |\Delta z_1|)} \int_{|\Delta z_1|}^{|\Delta z_1|+\epsilon\Delta c} W(\Delta z_i)\, d\Delta z_i \qquad (32)$$

$$= \int_{|\Delta z_1|}^{|\Delta z_1|+\epsilon\Delta c} W(\Delta z)\, d\Delta z \Big/ \int_{|\Delta z_1|}^{\infty} W(\Delta z)\, d\Delta z \quad ,$$

---

* The receiver may also have difficulty in determining the sign of small correlator differences. The percentage increases in P_U, P_H and P_W produced by this difficulty are of the order of magnitude of

$$(P_{W,\epsilon} - P_{W,0})/P_{W,0} \quad ,$$

and hence, as we shall see, very small.

† For simplicity of calculation, we have chosen $P_{W,\epsilon}$ as the probability of error per word, assuming that single errors remain uncorrected whenever two or more correlator differences lie within $\epsilon\Delta c$ of each other. In practice, a smaller probability of error can be obtained by the simple expedient of arbitrarily changing the first digit of every set of digits with correlator differences within $\epsilon\Delta c$ of each other. In this way, we correct almost half of the errors which remain uncorrected in calculating $P_{W,\epsilon}$.

where $W(\Delta z)$ is given by Eq.(6). [Since the correlator difference is always positive for a correct digit, the absolute value sign is not needed on $|\Delta z_i|$. The factor $\text{prob}(\Delta z_i > |\Delta z_1|)$ in the denominator renormalizes the distribution $W(\Delta z_i)$ from one in which all values of Δz_i are possible to one which only allows values of $\Delta z_i > |\Delta z_1|$.]

The probability [Eq.(32)] is written for a given $|\Delta z_1|$. Let $J(\epsilon,a)$ be the probability that Δz_i will lie within $\epsilon\,\Delta c$ of $|\Delta z_1|$, no matter what the value of $|\Delta z_1|$. It is clear that $J(\epsilon,a)$ is independent of i. Using Eqs.(6) and (9), we find

$$J(\epsilon,a) = \int_0^\infty W(|\Delta z_1|, \text{wrong})\ \text{prob}\,(|\Delta z_1| < \Delta z_i < |\Delta z_1| + \epsilon\,\Delta c/\Delta z_i > |\Delta z_1|)\, d|\Delta z_1|$$

$$= \frac{1}{\sqrt{2\pi}\,\sigma}\int_0^\infty \exp\left[-\frac{(|\Delta z_1| + \Delta c)^2}{2\sigma^2}\right] \frac{\displaystyle\int_{|\Delta z_1|}^{|\Delta z_1|+\epsilon\,\Delta c} \exp\left[-\frac{(\Delta z - \Delta c)^2}{2\sigma^2}\right] d\Delta z}{\displaystyle\int_{|\Delta z_1|}^{\infty} \exp\left[-\frac{(\Delta z - \Delta c)^2}{2\sigma^2}\right] d\Delta z}\, d|\Delta z_1| \tag{33}$$

$$= \frac{1}{\sqrt{\pi}}\int_0^\infty \exp\left[-(x+a)^2\right] \frac{\text{erf}(x - a + \epsilon\,a) - \text{erf}(x-a)}{1 - \text{erf}(x-a)}\, dx \quad .$$

The probability that Δz_i does not lie within $\epsilon\,\Delta c$ of $|\Delta z_1|$, assuming that $\Delta z_i > |\Delta z_1|$, is thus $1 - J(\epsilon,a)$, and the probability that none of the $n-1$ correct digits has a Δz within $\epsilon\,\Delta c$ of $|\Delta z_1|$ is $[1 - J(\epsilon,a)]^{n-1}$. Thus we find

$$P_{W,\epsilon} = 1 - q^{m+1}(a) - [1 - J(\epsilon,a)]^m\, \Pi_{m+1}(a) \quad . \tag{34}$$

Calculated values of $J(0.1,a)$ for selected values of a are given in Table IV, together with a comparison of $P_{W,0.1}$ and $P_{W,0}$ for words of six message digits.

TABLE IV

Selected values of $J(0.1,a)$ and comparison of $P_{W,0}$ and $P_{W,0.1}$ for words of 6 message digits

a	$J(0.1,a)$	$P_{W,0}$	$P_{W,0.1}$
1.0	0.00379	0.236	0.241
1.5	0.000426	0.0220	0.0222
2.0	0.0000189	0.00062	0.00062+

From Table IV, we see that even for a rather crude receiver, which cannot distinguish correlator differences lying within 0.1 Δc of each other, the percentage change in the probability of error per word is at most about two per cent in the region of interest. Thus the advantage of the Wagner code over the Hamming code does not depend on great precision of the correlators or the memory.

F. SUMMARY AND CONCLUSIONS

Following a suggestion of C. A. Wagner, we have constructed a code that, unlike those previously considered, makes explicit use of the magnitude of the difference of the a posteriori probabilities of the transmitted signals. The code is applied to systems transmitting sequences (words) of binary digits that are corrupted by the addition of white Gaussian noise and then detected by correlation. The simplicity of the calculations described in Secs. C and E derives from

the fact that the a posteriori probability $p(x_i/y)$ that x_i was sent if y is received is proportional (for suitable x_i) to the exponential of the finite-time correlation

$$z_i = \int_o^T x_i \, y dt \quad .$$

Under the assumptions made in Sec. B, z_i is normally distributed, and a pair of signals x_1 and x_2 can be chosen such that their correlations are statistically independent.

Unlike the Hamming code, the Wagner code has only a probability of correcting single errors. However, unless the signal-to-noise ratio of the correlator difference voltage is less than unity or the word is too long, this probability is high. The Wagner code requires only a single check digit, the burden of its operation being placed on a memory which must store the values of the a posteriori probabilities (or their differences) for the duration of a word. The Hamming code, on the other hand, requires several check digits (see Sec. A).

If the words are transmitted at a constant rate, the coded words (message digits and check digits) must have the same duration as the uncoded words (message digits only); therefore, the digits composing the word must be appropriately shortened. This increases the probability of error per digit and diminishes the advantage of the code. Since the digits of a Wagner-coded word are longer than those of the corresponding Hamming-coded word, they are less likely to be incorrectly received. Thus it is natural to ask, first, whether coding is worth the trouble in a constant-data-rate system, and second, whether the Wagner code which shortens the digits less but corrects only a certain percentage of single errors is better than the Hamming code which shortens the digits more but corrects all single errors.

These questions are answered in Sec. D. It is found that both codes are valuable even for very short words, but that the Wagner code makes considerably fewer errors, depending on the length of the word and the signal-to-noise ratio of the correlator difference voltage. (No attempt is made to relate channel and output signal-to-noise ratios.)

The successful operation of the Wagner code does not require great precision of the correlators or memory, as shown in Sec. E.

The numerical computations reported here were done by Mrs. Elizabeth Munro.

APPENDIX
ON THE DISTRIBUTION OF FINITE-TIME CORRELATIONS

Consider a transmitter sending one of two signals $x_1(t)$ and $x_2(t)$ of duration T and bandwidth W. Suppose that they are corrupted by the addition of white Gaussian noise. Thus, if $x_1(t)$ is sent, the received signal is

$$y(t) = x_1(t) + n(t) \quad , \tag{A-1}$$

where n(t) is a representative noise waveform. Let z_1 and z_2 denote the finite-time correlations

$$z_1 = \int_o^T x_1(t)\, y(t)\, dt$$

and

$$z_2 = \int_o^T x_2(t)\, y(t)\, dt \quad . \tag{A-2}$$

In Sec. B it was stated without proof that z_1 and z_2 are normally distributed random variables. In Sec. C it was stated without proof that $x_1(t)$ and $x_2(t)$ can be chosen so that z_1 and z_2 are statistically independent. This appendix is devoted to a justification of these statements.

In order to prove the first statement, we use a well-known theorem of Shannon[8] that any signal of bandwidth W is uniquely determined by its values at sample points spaced 1/2W seconds apart. Thus $x_1(t)$, $x_2(t)$ and n(t) each can be written in the form*

$$f(t) = \sum_{r=1}^{2TW} f(r/2W)\, \mathrm{sinc}\,(2Wt - r) \quad , \tag{A-3}$$

where

$$\mathrm{sinc}\, x = \sin \pi x / \pi x \quad . \tag{A-4}$$

The function sinc x satisfies the orthonormality property

$$\int_{-\infty}^{\infty} \mathrm{sinc}\,(x - r)\, \mathrm{sinc}\,(x - s)\, dx = \delta_{rs} \quad . \tag{A-5}$$

Moreover, if the spectrum of the Gaussian noise n(t) is flat, its values at the 2TW sample points are statistically independent.[8] Combining Eqs.(A-1) to (A-5), we see that z_1 and z_2 are finite sums of statistically independent Gaussian variables, and thus are themselves Gaussian random variables.[9] Moreover, since z_1 and z_2 are derived from the same Gaussian random variables, namely, the sample point values of n(t), their joint distribution is two-dimensional Gaussian.[9]

To prove the second statement, it suffices to choose $x_1(t)$ and $x_2(t)$ such that $\overline{z_1 z_2} = \overline{z_1}\,\overline{z_2}$ (where the averages are over the noise statistics), since linear independence implies statistical independence for two variables whose joint distribution is Gaussian.[10] The averages of

---

*It will be seen that f(t) as given by (A-3) does not vanish for $t < 0$ and $t > T$. However, for large TW, f(t) is negligibly small outside the finite interval. This justifies the use of idealized signals of duration T and bandwidth W.

z_1 and z_2 are

$$c_1 = \overline{z_1} = \overline{\int_o^T x_1 y\, dt} = \overline{\int_o^T (x_1 + n)\, x_1\, dt}$$

$$= \int_o^T \overline{(x_1^2 + x_1 n)}\, dt = \int_o^T x_1^2\, dt \text{ (since } \overline{x_1 n} = x_1 \bar{n} = 0), \quad , \tag{A-6}$$

and

$$c_2 = \overline{z_2} = \overline{\int_o^T x_2 y\, dt} = \overline{\int_o^T (x_1 + n)\, x_2\, dt} = \int_o^T \overline{(x_1 x_2 + x_2 n)}\, dt$$

$$= \int_o^T x_1 x_2\, dt \text{ (since } \overline{x_2 n} = x_2 \bar{n} = 0). \tag{A-7}$$

The calculation of $\overline{z_1 z_2}$ is more complicated. As a first step, write $\overline{z_1 z_2}$ as the double integral

$$\overline{z_1 z_2} = \overline{\int_o^T x_1 y\, dt \int_o^T x_2' y'\, dt'} = \int_o^T \int_o^T \overline{x_1 y x_2' y'}\, dt\, dt' \tag{A-8}$$

where x_1 and y are written for argument t [as in Eqs.(A-6) and (A-7)], and x_2' and y' for argument t'. Then, substituting $y = x_1 + n$, $y' = x_1' + n'$, and expanding, we obtain

$$\overline{z_1 z_2} = \int_o^T \int_o^T \overline{(x_1^2 x_1' x_2' + n x_1 x_1' x_2' + n' x_1^2 x_2' + n n' x_1 x_2')}\, dt\, dt'$$

$$= \int_o^T x_1^2\, dt \int_o^T x_1' x_2'\, dt' + \int_o^T \int_o^T \overline{n n'}\, x_1 x_2'\, dt\, dt' \text{ (using } \bar{n} = \bar{n}' = 0) \tag{A-9}$$

$$= \overline{z_1}\,\overline{z_2} + \int_o^T \int_o^T R_n(t - t')\, x_1(t)\, x_2(t')\, dt\, dt' \quad ,$$

where R_n is the autocorrelation function of the noise n. Since x_1 and x_2 are negligibly small outside the interval $0 \leq t \leq T$, the last integral in Eq.(A-9) can be approximated by

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} R_n(t - t')\, x_1(t)\, x_2(t')\, dt\, dt' = \int_{-\infty}^{\infty} R_n(\tau)\, R_{12}(\tau)\, dt \quad , \tag{A-10}$$

where

$$R_{12}(\tau) = \int_{-\infty}^{\infty} x_1(t + \tau)\, x_2(t)\, dt \quad . \tag{A-11}$$

We can choose x_1 and x_2 such that $R_{12}(\tau)$ is very small where $R_n(\tau)$ is significant, that is, for $\tau \leq 1/W$.* Indeed, if x_1 and x_2 have nonoverlapping energy spectra within the band W, $R_{12}(\tau) \equiv 0$. Thus, for suitably chosen x_1 and x_2, we may neglect the integral Eq.(A-10), so that Eq.(A-9) becomes

$$\overline{z_1 z_2} \approx \overline{z_1}\,\overline{z_2} \quad . \tag{A-12}$$

*This means that x_1 and x_2 are approximately orthogonal and remain so when time-displaced with respect to each other by an amount $\leq$ the correlation time of the noise.

Equation (A-12) states that z_1 and z_2 are (approximately) linearly independent and consequently (approximately) statistically independent.

The results just derived are equally valid if the band W starts at $W_o = nW$, with n an integer. They are approximately valid in any case if $W_o >> W$, which is usually the case in communication problems.[11]

REFERENCES

1. C. E. Shannon, Bell System Tech. Jour. 30, 50-64 (January 1951).
2. R. W. Hamming, Bell System Tech. Jour. 29, 147-160 (April 1950).
3. C. A. Desoer, "Communication Through Channels in Cascade," ScD Thesis, Department of Electrical Engineering, M. I. T. (February 1953).
4. I. S. Reed, "A Class of Multiple-Error-Correcting Codes and the Decoding Scheme," Technical Report No. 44, Lincoln Laboratory, M. I. T. (9 October 1953), and forthcoming paper in the Journal of the Society of Industrial and Applied Mathematics.
5. P. M. Woodward and I. L. Davies, Proc. I. E. E. 99, III, 37 (March 1952).
6. H. Cramér, *Mathematical Methods of Statistics*, Princeton University Press (1946) p. 212.
7. W. B. Davenport, Jr., R. A. Johnson and D. Middleton, Jour. Appl. Phys. 23, 4, 377-388 (April 1952).
8. C. E. Shannon, Proc. I. R. E. 37, 10-21 (January 1949).
9. H. Cramér, *op. cit.*, pp. 312-313.
10. H. Cramér, *op. cit.*, pp. 310-311.
11. A. Kohlenberg, "Exact Interpolation of Band-Limited Functions," forthcoming paper in Jour. Appl. Phys.