THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 54, No. 6, July–August 1975
Printed in U.S.A.

A Geometric Derivation of Forney's Upper Bound

By J. E. MAZO

(Manuscript received January 17, 1975)

Effective analyses of performance for detection schemes that optimally decode digital data in the presence of intersymbol interference have been slow in coming. Recently, however, Forney has given an upper bound on the bit error probability for maximum-likelihood sequence estimation. Starting from a standard geometrical framework, we give a much simplified derivation of this upper bound. Our derivation places the validity of this important bound more in evidence in that the concepts of whitened matched filter and error event are not introduced.

Let a_j, $j = 1, 2, \cdots, N$, be independent, equilikely binary random variables taking values ± 1. Data transmission usually involves estimating the a_i from a pulse sequence of the form

$$\sum_{j=1}^{N} a_j h(t - jT), \quad -\infty < t < \infty, \tag{1}$$

which is observed in white gaussian noise of (two-sided) spectral density $N_0/2$. In (1), the minimum assumption put on the pulse waveform $h(t)$ is that it be L_2. One possible detection procedure is to decide, on the basis of the received noisy signal, which one of the 2^N equilikely signals given in (1) was "most likely" (maximum-likelihood sequence detection) and use the sequence $\{a_n\}$ which is associated with that sequence as the detected symbols. As N grows large, the probability of deciding incorrectly on the sequence approaches unity; however, the real question revolves about the *bit*-error probability for maximum-likelihood *sequence* detection. Important work on this problem was done recently by Forney,[1,2] who showed that, under certain conditions on $h(t)$, the bit-error probability for large signal-to-noise ratios goes exponentially to zero as

$$P_e \approx (\text{coeff.}) \exp\left(-\frac{d_{\min}^2}{4N_0}\right), \quad N_0 \to 0, \tag{2}$$

where $d_{\min}^2$ is the minimum distance between the signal sequences in (1); i.e., if we add a superscript to distinguish among sequences thus

$$s^{(i)}(t) = \sum_{j=1}^{N} a_j^{(i)} h(t - jT), \; -\infty < t < \infty \quad i = 1, 2, \cdots, 2^N, \tag{3}$$

then

$$d_{\min}^2 = \lim_{N\to\infty} \min_{\substack{i,k \\ i\neq k}} \int_{-\infty}^{\infty} |s^{(i)}(t) - s^{(k)}(t)|^2 dt. \tag{4}$$

Forney's demonstration consists of two steps. First, a lower bound on P_e of the form (2) is established, valid for any $h(t)$. Second, if $H(\omega)$ denotes the Fourier transform of $h(t)$ and

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} |H(\omega)|^2 e^{ik\omega} d\omega = 0 \text{ for integer } k, \; |k| > \nu, \; \nu \text{ integer}, \tag{5}$$

then an upper bound for P_e can be given which is convergent for large signal-to-noise ratios and which, furthermore, also has an asymptotic form given by (2).

In our opinion, Forney's discussion of the upper bound is sufficiently complicated that some question remains as to how firmly the result is established. We shall give a much simpler derivation, but first let us review the situation when $\nu = 0$, i.e., when there is no intersymbol interference. Using the reduction to the standard geometrical picture,[3] the signal points (sequences) received in the absence of noise are as shown in Fig. 1. These signal points are to be regarded as being perturbed by spherically symmetric, N-dimensional, zero mean gaussian noise; the variance of each component of the noise is $N_0/2$. Thus any point in the N-dimensional space may be received and the maximum-likelihood decoder chooses the unperturbed sequence nearest to the received point as the transmitted one. The decoding regions are shown in Fig. 2 by dashed lines, and are labeled $R_{i,j}$ in an obvious way.

Now assume we transmit (1, 1) and ask for the probability that the first bit is in error. This is the same as the probability that the received signal point is in $R_{-1,1} \cup R_{-1,-1}$, or equivalently that the received signal point is to the left of the line labeled S in Fig. 1. In N dimensions it would be the probability that the received signal point is on the opposite side of an $(N - 1)$ dimensional hyperplane. This is clearly a simple one-dimensional gaussian problem having the well-known Q-

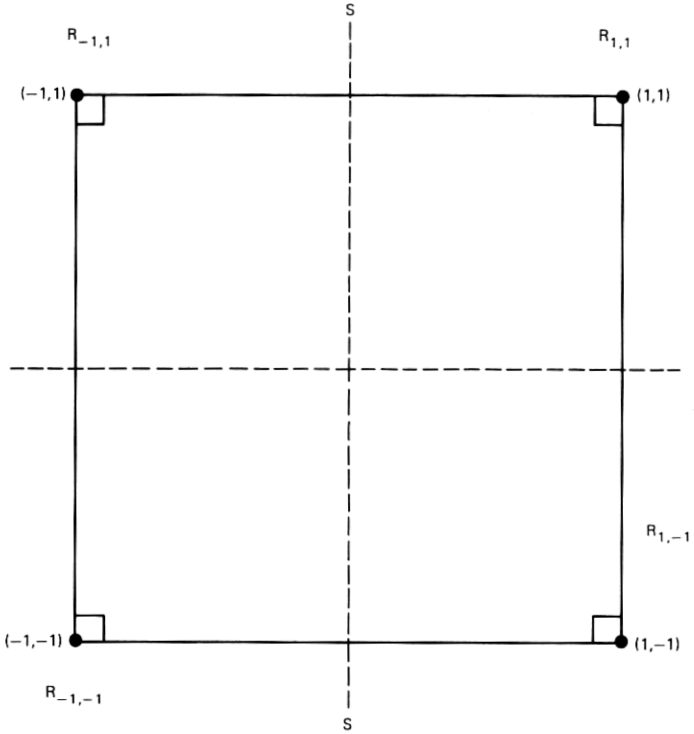

Fig. 1—Four signal points corresponding to sequences for $N = 2$ in the absence of intersymbol interference.

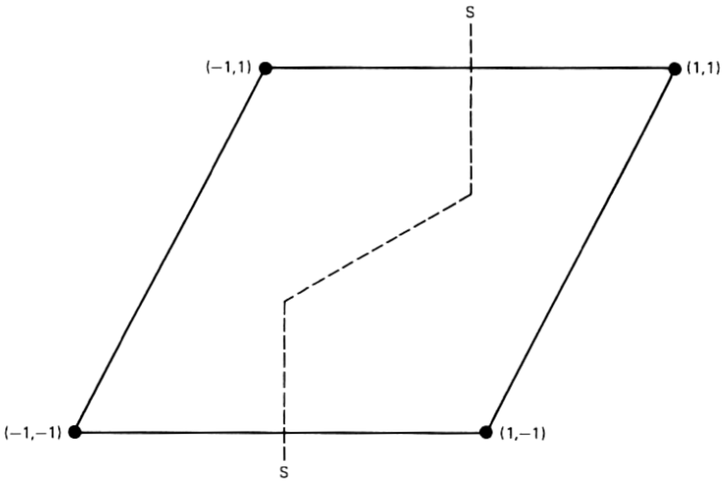

Fig. 2—Four signal points with intersymbol interference.

function for an answer, independent (because of the simple geometry) of the dimension or of which bit in the sequence is transmitted.

When intersymbol interference is present, the error probability for the kth bit may well depend on k and N. In addition, the surface S which separates the sequences that have $a_k = +1$ from those which have $a_k = -1$ is no longer a hyperplane, although it is made up of segments which are hyperplanes. Finally, and perhaps a bit vaguely, the "shape" of the surface may depend on k. An example for $N = 2$ is given in Fig. 2 showing the separating surface for the first bit.

Our goal is to derive Forney's upper bound by geometrical arguments about as simple as those used in the discussion of Fig. 1.

As in (3), we consider signal points identified by their respective data sequences $\{a_i\}^N$ and label them with a superscript. We focus on the kth bit being in error, and define sets A and B:

$$A = \{\mathbf{a}^{(i)} | a_k^{(i)} = +1\}, \quad B = \{\mathbf{a}^{(i)} | a_k^{(i)} = -1\} \equiv \{\mathbf{b}^{(i)}\}. \tag{6}$$

We are mainly interested in the chance that the maximum-likelihood decoder selects a point $\mathbf{b} \in B$, given that a particular sequence $\mathbf{a}^{(1)}$ (say), $\mathbf{a}^{(1)} \in A$, was transmitted. One upper bound on this is the union bound

$$P_e(k) \leqq \sum_{\mathbf{b} \in B} Q\left(\frac{d(\mathbf{a}^{(1)}, \mathbf{b})}{\sqrt{2N_0}}\right), \tag{7}$$

where $d(\mathbf{a}^{(1)}, \mathbf{b})$ is the euclidean distance between $\mathbf{a}^{(1)}$ and $\mathbf{b}$; i.e., if l labels the particular $\mathbf{b}$ sequence,

$$\begin{aligned} d^2(\mathbf{a}^{(1)}, \mathbf{b}) &= \int_{-\infty}^{\infty} |s^{(1)}(t) - s^{(l)}(t)|^2 dt \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} |H(\omega)|^2 \left| \sum_{j=1}^{N} (a_j^{(1)} - b_j^{(l)}) e^{ij\omega} \right|^2 d\omega. \end{aligned} \tag{8}$$

On writing (8), and henceforth, we set $T = 1$. Equation (7) is a bad bound because it includes too many terms on the right-hand side. This is easily seen by applying it to the N-dimensional hypercube (no intersymbol interference), for which we obtain (ignoring unessential coefficients)

$$P_e(k) \leqq 2^N e^{-(d^2/4N_0)}, \tag{9}$$

where d^2 is the length of a hypercube edge. Thus, the bound, for any fixed N_0, approaches ∞ as the length of the sequence N increases.

Our next step will be to make some simple observations about the geometry of the received signal points [when (5) is true] that will allow us to delete most of the terms on the right-hand side of (7). Following

Forney, we are motivated to define another set B_1 of signal points which is a subset of B. A vector $\mathbf{b} \in B_1$ if $\{\mathbf{b}_j - \mathbf{a}_j^{(1)}\}_{j=1}^N$ (after deleting all zeros which begin the sequence and end the sequence of coefficients $\{\mathbf{b}_j - \mathbf{a}_j^{(1)}\}$) does not contain ν or more consecutive zeros either to the right of the kth position or to the left of the kth position. Forney's upper bound then reads

$$P_e(k) \leqq \sum_{\mathbf{b} \in B_1} Q\left(\frac{d(\mathbf{a}^{(1)}, \mathbf{b})}{\sqrt{2N_0}}\right). \tag{10}$$

To see why no further terms need be included in (10), select an arbitrary signal point b^* not included in the sum in (10); i.e., $b^* \in B - B_1 \equiv B^*$. From the way things have been defined, we may write

$$\mathbf{a}^{(1)} = (\boldsymbol{\alpha}^{\mathrm{L}}, 1, \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3) \tag{11}$$

$$\mathbf{b}^* = (\boldsymbol{\beta}^{\mathrm{L}}, -1, \boldsymbol{\beta}_1, \boldsymbol{\alpha}_2, \boldsymbol{\beta}_3), \tag{12}$$

where $\boldsymbol{\alpha}_1 - \boldsymbol{\beta}_1$ does not contain ν consecutive zeros in its coefficient sequence and $\boldsymbol{\alpha}_3 - \boldsymbol{\beta}_3 \neq 0$. The 1 in (11) and the (-1) in (12) occur in the kth position, and $\boldsymbol{\alpha}_2$ is at least of dimension ν, corresponding to the ν (or more) positions where $\mathbf{a}^1$ and $\mathbf{b}^*$ are to agree. Now $\boldsymbol{\alpha}_{\mathrm{L}} - \boldsymbol{\beta}_{\mathrm{L}}$ may or may not contain ν consecutive zeros,† and we distinguish these two cases in our discussion. First assume that $\boldsymbol{\alpha}_{\mathrm{L}} - \boldsymbol{\beta}_{\mathrm{L}}$ does not contain ν consecutive zeros. Then note that

$$\mathbf{b}^{(1)} \equiv (\boldsymbol{\beta}_{\mathrm{L}}, -1, \boldsymbol{\beta}_1, \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3) \in B_1 \tag{13}$$

$$\mathbf{a}^{(2)} \equiv (\boldsymbol{\alpha}_{\mathrm{L}}, 1, \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\beta}_3) \in A. \tag{14}$$

Statement (13) is true by the absence of ν consecutive zeros in $\boldsymbol{\beta}_{\mathrm{L}} - \boldsymbol{\alpha}_{\mathrm{L}}$ and also in $\boldsymbol{\beta}_1 - \boldsymbol{\alpha}_1$; (14) is true because $\mathbf{a}_k^{(2)} = +1$. We may at this point imagine the four signal points $\mathbf{a}^{(1)}, \mathbf{b}^*, \mathbf{b}^{(1)}, \mathbf{a}^{(2)}$ in general position as in Fig. 3. Now focus attention on the triangle $(a^{(1)}, a^{(2)}, b^*)$.‡ We have (letting 0_i denote a string of at least i zeros)

$$\begin{aligned} d^2(a^{(1)}, b^*) &= \|a^{(1)} - b^*\|^2 \\ &= \|\alpha^{\mathrm{L}} - \beta^{\mathrm{L}}, 2, \alpha_1 - \beta_1, 0_\nu, \alpha_3 - \beta_3\|^2 \\ &= \|\alpha^{\mathrm{L}} - \beta^{\mathrm{L}}, 2, \alpha_1 - \beta_1\|^2 + \|\alpha_3 - \beta_3\|^2. \end{aligned} \tag{15}$$

The last step in (15) follows from the Fourier integral expression (8) for the distance and from (5), the requirement that intersymbol interference not extend beyond ν. The right member of (15) can readily be seen from (11, 12) and (13, 14) to be $d^2(a^{(2)}, b^*) + d^2(a^{(1)}, a^{(2)})$.

† Recall that a string of zeros in the beginning does not count.
‡ We drop bold-face notation for vectors here, and also allow ourselves the freedom of writing the subscript L as a superscript.

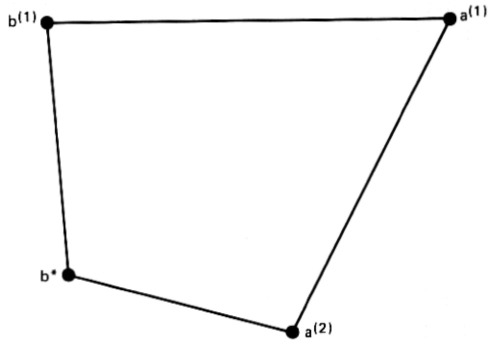

Fig. 3—The four signal points defined in the test illustrated in general position (not necessarily planar).

Hence,

$$d^2(a^{(1)}, b^*) = d^2(a^{(2)}, b^*) + d^2(a^{(1)}, a^{(2)}) \tag{16}$$

and $(a^{(1)}, a^{(2)}, b^*)$ forms a right triangle. In an entirely similar manner, one verifies

$$d^2(a^{(2)}, b^{(1)}) = d^2(a^{(1)}, b^{(1)}) + d^2(a^{(1)}, a^{(2)}), \tag{17}$$

implying that $(a^{(1)}, a^{(2)}, b^{(1)})$ is a right triangle. Since $b^{(1)} - b^* = a^{(1)} - a^{(2)}$, we have

$$d^2(a^{(1)}, a^{(2)}) = d^2(b^{(1)}, b^*) \tag{18}$$

and the additional fact that the four points lie in the same plane. Equations (16), (17), and (18), and planarity, imply that $(a^{(1)}, a^{(2)}, b^{(1)}, b^*)$ form a rectangle as shown in Fig. 4. This demonstration assumes that there are ν consecutive zeros to the right of the kth position and not to the left. If we interchange the words "right" and "left", the same type of demonstration will apply. There remains the case when there are ν consecutive zeros both to the right and to the left of the kth position. In this case we write

$$a^{(1)} = (\alpha_3^L, \alpha_2^L, \alpha_1^L, 1, \alpha_1, \alpha_2, \alpha_3) \tag{19}$$

$$b^* = (\beta_3^L, \alpha_2^L, \beta_1^L, -1, \beta_1, \alpha_2, \beta_3), \tag{20}$$

where $\alpha_3 - \beta_3 \neq 0 \neq \alpha_3^L - \beta_3^L$ and neither $\alpha_1 - \beta_1$ nor $\alpha_1^L - \beta_1^L$ contain ν consecutive zeros in their coefficients. Further, we are to assume α_2 and α_2^L each have dimension at least ν. If we define

$$b^{(1)} = (\alpha_3^L, \alpha_2^L, \beta_1^L, -1, \beta_1, \alpha_2, \alpha_3) \in B_1 \tag{21}$$

$$a^{(2)} = (\beta_3^L, \alpha_2^L, \alpha_1^L, 1, \alpha_1, \alpha_2, \beta_3) \in A, \tag{22}$$

the proof that $(a^{(1)}, a^{(2)}, b^{(1)}, b^*)$ is a rectangle can be carried out using the same techniques as earlier.

Figure 4 makes it clear why, if $a^{(1)}$ is transmitted, terms like b^* do not have to be included in the right-hand side of (10). The term

$$Q\left(\frac{d(a^1, b^1)}{\sqrt{2N_0}}\right)$$

in (10) is the probability that, if $a^{(1)}$ is transmitted, the received signal will be on the "wrong side" of the hyperplane H which perpendicularly bisects the line (b^1a^1). Now b^* only needs to be included in (10) if its associated decoding region contributes some set of points of positive measure not accounted for in some other way. Thus, from Fig. 4, this is only the case if these new points were on the same side of H as $a^{(1)}$. But any point on that side is closer to $a^{(2)}$ than b^* and, hence, would never be decoded into b^*. Hence the gaussian measure of the decoding region for b^* is already included in the term

$$Q\left(\frac{d(b^1, a^1)}{\sqrt{2N_0}}\right).$$

If we calculate further upper bounds for (10) by letting $N = \infty$, we obtain a bound independent of k. Averaging this over the possible transmitted symbols gives precisely Forney's upper bound. The fact that the resulting upper bound converges for N_0 small enough (for

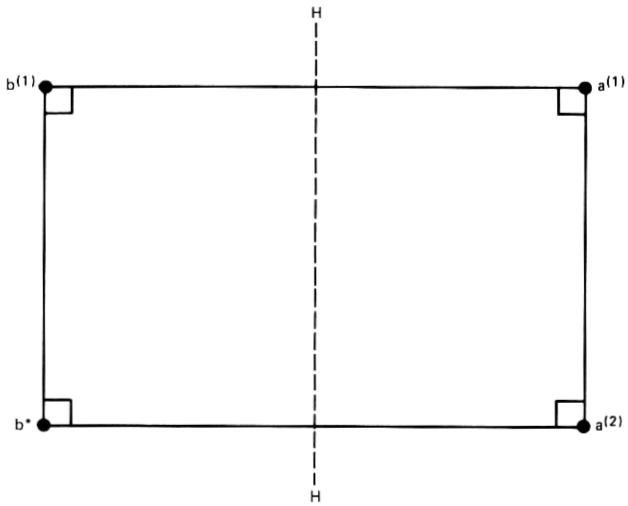

Fig. 4—The actual relationship of the four signal points defined in the text.

$N = \infty$) has been recently discussed by Foschini.[4] This last step is an important one in a full proof, and was overlooked in the initial work.[2]

REFERENCES

1. G. David Forney, "Lower Bounds on Error Probability in the Presence of Large Intersymbol Interference," IEEE Trans. Commun., *COM-20* (February 1972), pp. 76–77.
2. G. David Forney, "Maximum Likelihood Sequence Estimation of Digital Sequences in the Presence of Intersymbol Interference," IEEE Trans. Inform. Theory, *IT-18* (May 1972), pp. 363–378.
3. J. M. Wozencraft and I. M. Jacobs, *Principles of Communication Engineering.* New York: John Wiley, 1965, Ch. 4.
4. G. J. Foschini, "Performance Bound for Maximum Likelihood Reception of Digital Data," IEEE Trans. Inform. Theory, *IT-21*, No. 1 (January 1975), pp. 47–50.