An Information-Theoretic Model for the Ground Communications Facility Line

O. Adeyemi
Communications Systems Research Section

This article presents a three-state Markov chain as a model for the errors occurring on the Ground Communications Facility (GCF). An analytic expression for the capacity of the channel in terms of the model parameters is obtained, and comparison is made with the capacity of a binary symmetric channel (BSC) with the same bit-error rate. For a better understanding of the channel's intrinsic behavior and for use in estimating the performances of different error-detecting and error-correcting codes, we obtain analytic expressions for three sets of channel parameters. These are the bit error statistics, the block error statistics, and the distribution of burst lengths and error-free (gap) intervals.

I. Introduction and the Model

This article will report the progress made so far in constructing a theoretical model for the Ground Communications Facility (GCF) high-frequency (50-kbps) wideband error distribution using the results of the test runs reported by J. P. McClure (Refs. 1 and 2). In those reports, the "bursty" nature of the errors is clearly noticeable; there are error-free gaps of up to 3-min duration (10,000,000 bits), followed by up to ½ s of sputtering errors.

The data from the test runs, stored in twenty-five tape reels, consisted of records of relative positions of consecutive errors in the sequence of noise digits $z = \{z_n\}$ (in which $z_n = 1$ if the nth digit is in error and equal to 0 otherwise). Because of certain problems with the recordings, only twelve of these reels were found usable, with a total of about 5.73×10^8 bits and an average bit error rate of less than 10^{-4}.

Figure 1 is the histogram for the (error-free) gap lengths X, for $X \geqq 10^3$ bits (or a transmission time of 1/50 s or more). The ordinate represents the number of times a gap of length X occurred in the twelve tape reels. Not shown on the histogram are frequencies for the gap lengths of 100 bits or less. The frequencies for this range of gap lengths are much higher, showing again that the errors occur in bursts. The modal frequencies at $X = 10^3$ and $X = 10^6$ are denoted by $\bar{A}$ and $\bar{B}$, respectively.

Pioneered by Gilbert (Ref. 3), many attempts have been made at using Markov processes as theoretical models for burst-noise channels. For example, Elliot (Ref. 4) and Berkovits, Cohen, and Zierler (Ref. 5) have successfully generalized the original Gilbert model (Fig. 2) for fitting data from telephone networks. Gilbert's model consisted of one perfectly good state G and one error state B; i.e., an error bit can occur only in state B.

In the generalization, an error bit can occur in either the good state G or the bad state B, but with different probabilities. In Fig. 2, transitions between the states occur in the direction of the arrows, and if $P(j|i)$ denotes the one-step transition probability of going from state i to state j, then

$$P(G|B) = p, \qquad P(B|G) = P$$

and $q = 1 - p$, $Q = 1 - P$ are the probabilities of remaining in B and G, respectively. If we let k and h denote the probabilities of correct reception of a bit when the channel is, respectively, in states G and B, then $k' = 1 - k$, $h' = 1 - h$ are the respective probabilities of a bit being in error in those states.

The two-state model is not adequate, however, for the type of gap distribution obtained on the GCF channel (see Fig. 1). For very short gap lengths ($X \leq 10^3$) and medium to larger values of the (error-free) gaps ($10^3 < X \leq 10^5$), one may successfully characterize the communication channel by the generalized Gilbert model. But no two-state Markov chain would suffice to represent the whole range of the gap distribution (especially to include the other modal frequency at $\bar{B}$ for gap lengths of about 10^6 bits or more). We are therefore forced to consider a model with more than two states to account for the (error-free) gaps of length $\geq 10^6$, even though the analysis of such a model may be mathematically unwieldy.

The following is the suggested model for the GCF. A Markov chain with three states B, G_1, and G_2 will be used to generate the gap lengths (Fig. 3). Because of the low bit error rate recorded in the data, we take both states G_1 and G_2 to be perfectly good, i.e., $z_n = 0$ always in G_1 or G_2; the error bursts are produced in state B. Since actual bursts contain good digits interspersed with errors, we let the probability of having a good digit in this state equal $h > 0$. In other words,

$$P(z_n = 1 \mid B) = 1 - h < 1$$

and we write $1 - h = h'$. As explained for Fig. 2,

$$P(G_2|B) = p$$
$$P(B|G_2) = b$$
$$P(G_1|G_2) = g$$
$$P(G_2|G_1) = P$$

are the one-step transition probabilities: $q = 1 - p$, $r = 1 - g - b$, $Q = 1 - P$ are the probabilities of remaining in B, G_2, and G_1, respectively. For example, for a bit error rate $P_1 = 4 \times 10^{-5}$, which is of the order of the average bit error recorded in the GCF data, if we assume $h' = 0.5$, calculations using the data give

$$p = 0.20; \qquad (q = 0.80)$$
$$b = 0.36; \qquad r = 0.21 \times 10^{-2}; \qquad g = 0.6379$$
$$P = 0.29 \times 10^{-4}; \qquad Q = 0.999971$$

Let $\overleftrightarrow{TU}$ denote transitions between states T and U, including sojourns in T and U; and let TUV denote a particular (transition) path between states T and V passing through state U; $T, U, V = B, G_1$, or G_2. Then the following transitions between the states can produce the pattern of gap distributions observed in the data:

(1) A long sojourn in state B (with probability $q = 0.8$ of remaining in B at any instant of time, once having got there) accounts for the very short gaps (long bursts of errors, including cases of consecutive errors).

(2) $\overleftrightarrow{BG_2}$ ($\equiv BG_2G_2, \cdots, BBG_2, \cdots, BG_2B, \cdots, G_2BB, \cdots, G_2BG_2, \cdots$) accounts for gap lengths of about 10^2–10^3 bits.

(3) $BG_2G_1, \cdots, G_1G_2B, \cdots$ account for gap lengths of 10^3–10^4 bits.

(4) $\overleftrightarrow{G_1G_2}$ (including sojourns in G_1 and G_2) accounts for gap lengths of 10^5 bits or more.

Unfortunately, the twelve reels on which our model is based suffer from certain recording problems which have the effect of increasing the gap lengths at the end of each test run, thereby lowering the measured bit error rate. It is assumed that the current test run being conducted on the same channel but at a much lower rate (4.8 kbps) will provide reliable data from which to estimate the parameters of this model. It is even likely that we may have to modify our model to fit the new data. In such an eventuality, however, our method of analysis here is general enough to handle an increase in the number of states of the Markov chain.

Let us mention here that in 1970, Daniel Stern (Ref. 6) of GSFC computed a number of useful empirical probabilities from high-speed data collected from selected NASCOM channels. Thus, we have some data in hand with which to compare the results of the current test run.

II. Summary of Results

For a better understanding of the channel's intrinsic behavior and for use in estimating the performance of different error-detecting and -correcting codes on the channel, analytic expressions in terms of the model parameters are obtained for three groups of channel parameters. These are the bit error statistics, the block error statistics, and the distributions of the burst lengths and the gap intervals.

Section III contains the first group of statistics, consisting of (1) the distribution of distances between consecutive errors, (2) the autocorrelation of bit errors, and (3) the probability of a sequence of consecutive errors. Distribution of distances between consecutive errors will be used to estimate the number of gaps of a given length. For estimating significant error patterns, we require the autocorrelation of bit errors, and the distribution of a sequence of consecutive errors will show how clustered the errors are.

In Section IV we evaluate the capacity of the channel (the maximum rate at which arbitrarily reliable information can be transmitted over the channel). This capacity is then compared with that of a binary symmetric channel (BSC) with the same bit error rate. The capacity of the three-state channel must always be larger than that of the equivalent BSC; ironically, forward error correction is, nevertheless, more difficult.

A very important group of statistics for estimating the performance of error-correcting codes on the channel are the block error distributions; a *block* is defined as a sequence of n bits for a fixed integer n. This group of statistics, discussed in Section V, consists of

(1) The probability of an error block

(2) The distribution of the number of errors in a block

(3) The distribution of the number of errors in the information digits of an interleaved code with a given constant of interleaving (A block code C is said to be interleaved with constant of interleaving t if successive letters of individual code words are separated on the channel by t time units.)

(4) The distribution of distances between extreme errors (i.e., between the first and the last errors) in a block, which statistic is to be used to estimate the performance of codes that can correct error bursts in a block, provided they are confined to a given length

(5) The distribution of symbol errors in an n-symbol word where a symbol is a fixed number, say m, of consecutive bits

Section VI presents the last group of the channel parameters: the distribution of burst lengths. For this analysis, a burst will be defined as a sequence (1) beginning and ending with an error, (2) separated from the nearest preceding and following error by a gap of no less than a given number, say $\mathcal{G}$, called the *guardspace*, and (3) containing within it no gap equal to or greater than the given guardspace.

Since burst correction codes can generally correct bursts up to some fraction of the guardspace, say 1/3 (see Ref. 7), it is important to note what percentage of the burst lengths falls below the given fraction of a guardspace. The optimum value of the guardspace is therefore that which has the highest percentage of bursts less than this correctable fraction (1/3) of $\mathcal{G}$. This group also contains the distribution of the number of errors in a burst of a given length. This statistic is significant because for low error densities within a burst, burst-correcting codes can cope with occasional errors in what should be an error-free guardspace.

III. Bit Error Statistics

To shorten the length of this article, detailed proofs of every proposition will not be presented. Henceforth all references to model parameters will be to those presented in Fig. 3.

The fractions of times spent in states B, G_2, and G_1 are given, respectively, by

$$P(s_0 = B) = \frac{bP}{w_0}$$

$$P(s_0 = G_2) = \frac{pP}{w_0}$$

$$P(s_0 = G_1) = \frac{pg}{w_0} \tag{1}$$

where $w_0 = pg + pP + bP$.

Since errors occur only in state B, and then just with probability h', the bit error probability $P(1)$ is given by

$$P(1) = \frac{bPh'}{w_0}$$

Let $v(k)$ denote the probability of getting k error-free bits between a given error and the one immediately following it (i.e., a gap of length k):

$$v(k) = P(0^k 1 | 1) \tag{2}$$

where $\{0^k 1 | 1\}$ is the event that a given initial error is followed by a gap of length k. It follows that

$$\begin{aligned} v(k) &= P(0^k | 1) - P(0^{k+1} | 1) \\ &= u(k) - u(k+1) \end{aligned} \tag{3}$$

where we have denoted $P(0^k | 1)$ by $u(k)$ and $\{0^k | 1\}$ is the event that a given error is followed by k error-free bits. Therefore, it suffices to find expressions only for $u(k)$ in terms of the model parameters.

Let

$$\begin{aligned} G_1(k) &= P(0^k, s_k = G_1 | 1) \\ G_2(k) &= P(0^k, s_k = G_2 | 1) \\ B(k) &= P(0^k, s_k = B | 1) \end{aligned} \tag{4}$$

where s_k is the state of the channel at time k. Then

$$u(k) = G_1(k) + G_2(k) + B(k)$$

and

$$\begin{aligned} G_1(k+1) &= P(0^{k+1}, s_k = G_1, s_{k+1} = G_1 | 1) \\ &\quad + P(0^{k+1}, s_k = G_2, s_{k+1} = G_1 | 1) \\ &\quad + P(0^{k+1}, s_k = B, s_{k+1} = G_1 | 1) \\ &= G_1(k) Q + G_2(k) g \end{aligned} \tag{5}$$

Similarly,

$$G_2(k+1) = G_2(k) r + G_1(k) P + B(k) p$$

$$B(k+1) = B(k) qh + G_2(k) bh$$

In matrix form,

$$(G_1(k+1), G_2(k+1), B(k+1)) = (G_1(k), G_2(k), B(k)) M \tag{6}$$

where

$$M = \begin{pmatrix} Q & P & 0 \\ g & r & bh \\ 0 & p & qh \end{pmatrix}$$

with

$$\begin{aligned} G_1(0) &= P(s_0 = G_1 | 1) = 0 \\ G_2(0) &= P(s_0 = G_2 | 1) = 0 \\ B(0) &= P(s_0 = B | 1) = 1 \end{aligned} \tag{7}$$

Now Eqs. (6) and (5) give

$$\begin{aligned} &(G_1(k+1), G_2(k+1), B(k+1)) \\ &\quad = (G_1(0), G_2(0), B(0)) M^{k+1} \\ &\quad = (0, 0, 1) M^{k+1} \end{aligned} \tag{8}$$

For the purpose of deriving

$$u(k) = G_1(k) + G_2(k) + B(k)$$

one may use Eq. (8) and a computer after estimating the model parameters. The results of such computer calculations and estimates will be presented in a fuller report. But for use in the next section in finding expression for the capacity of the channel, we give detailed solution to (8).

Write the matrix M in Eq. (6) as:

$$M = \begin{pmatrix} 1-P & P & 0 \\ g & 1-g-b & bh \\ 0 & p & (1-p)h \end{pmatrix} \tag{9}$$

The characteristic equation associated with M is

$$\begin{aligned} &\lambda^3 + \lambda^2 [P + ph + g + b - 2 - h] \\ &\quad + \lambda [bP + pPh + phg + 2h + 1 - h(P + p + g + b) \\ &\quad - (P + ph + g + b)] \\ &\quad + h[P + p + g + b - (pP + pg + bP) - 1] = 0 \end{aligned}$$

with roots $\lambda_1, \lambda_2, \lambda_3$ given by

$$\lambda_1 = -\left[(A + \sqrt{A^2 + B^3})^{1/3} + (A - \sqrt{A^2 + B^3})^{1/3} + \frac{\hat{b}}{3}\right]$$

$$\lambda_2 = \frac{1}{2}[(A + \sqrt{A^2 + B^3})^{1/3} + (A - \sqrt{A^2 + B^3})^{1/3}] + \frac{i\sqrt{3}}{2}[(A + \sqrt{A^2 + B^3})^{1/3} - (A - \sqrt{A^2 + B^3})^{1/3}] - \frac{\hat{b}}{3}$$

$$\lambda_3 = \frac{1}{2}[(A + \sqrt{A^2 + B^3})^{1/3} + (A - \sqrt{A^2 + B^3})^{1/3}] - \frac{i\sqrt{3}}{2}[(A + \sqrt{A^2 + B^3})^{1/3} - (A - \sqrt{A^2 + B^3})^{1/3}] - \frac{\hat{b}}{3}$$

where $\lambda_1 < \lambda_2 < \lambda_3$ are real roots and

$$\left.\begin{aligned}
&A = \frac{\hat{q}}{2}; \qquad B = \frac{\hat{p}}{3} \\
&\hat{p} = \hat{c} - \frac{\hat{b}^2}{3}; \qquad \hat{q} = \hat{d} - \frac{\hat{b}\hat{c}}{3} + \frac{2\hat{b}^3}{27} \\
&\hat{b} = P + ph + g + b - 2 - h \\
&\hat{c} = bP + pPh + pgh + 2h + 1 \\
&\qquad - h(P + p + g + b) - (P + ph + g + b) \\
&\hat{d} = h[P + p + g + b - (bP + pP + pg) - 1]
\end{aligned}\right\} \tag{10}$$

To find M^n is in general complicated, but using a method described in Ref. 8 (p. 385), we get

$$M^{(n)} = c_1\lambda_1^n M_1 + c_2\lambda_2^n M_2 + c_3\lambda_3^n M_3 \tag{11}$$

where

$$M_1 = \begin{pmatrix} \dfrac{gP}{[\lambda_1 - (1-P)]^2} & \dfrac{P}{\lambda_1 - (1-P)} & \dfrac{bPh}{[\lambda_1 - (1-P)][\lambda_1 - (1-p)h]} \\ \dfrac{g}{\lambda_1 - (1-P)} & 1 & \dfrac{bh}{\lambda_1 - (1-p)h} \\ \dfrac{bg}{[\lambda_1 - (1-P)][\lambda_1 - (1-p)h]} & \dfrac{p}{\lambda_1 - (1-p)h} & \dfrac{pbh}{[\lambda_1 - (1-p)h]^2} \end{pmatrix}$$

$$M_2 = \begin{pmatrix} \dfrac{gP}{[\lambda_2(1-P)]^2} & \dfrac{P}{\lambda_2 - (1-P)} & \dfrac{bPh}{[\lambda_2 - (1-P)][\lambda_2 - (1-p)h]} \\ \dfrac{g}{\lambda_2 - (1-P)} & 1 & \dfrac{bh}{\lambda_2 - (1-p)h} \\ \dfrac{pg}{[\lambda_2 - (1-P)][\lambda_2 - (1-p)h]} & \dfrac{p}{\lambda_2 - (1-p)h} & \dfrac{bph}{[\lambda_2 - (1-p)h]^2} \end{pmatrix}$$

$$M_3 = \begin{pmatrix} \dfrac{gP}{[\lambda_3 - (1-P)]^2} & \dfrac{P}{[\lambda_3 - (1-P)]} & \dfrac{bPh}{[\lambda_3 - (1-P)][\lambda_3 - (1-p)h]} \\ \dfrac{g}{\lambda_3 - (1-P)} & 1 & \dfrac{bh}{\lambda_3 - (1-p)h} \\ \dfrac{pg}{[\lambda_3 - (1-P)][\lambda_3 - (1-p)h]} & \dfrac{p}{\lambda_3 - (1-p)h} & \dfrac{bph}{[\lambda_3 - (1-p)h]^2} \end{pmatrix}$$

and

$$c_r = \left\{\frac{gP}{[\lambda_r - (1-p)]^2} + 1 + \frac{bph}{[\lambda_r - (1-p)h]^2}\right\}^{-1} \qquad r = 1, 2, 3 \tag{12}$$

Using the fact that

$$[G_1(k), G_2(k), B(k)] = (0, 0, 1)M^k$$

and

$$u(k) = G_1(k) + G_2(k) + B(k)$$

we obtain

$$G_1(k) = c_1\lambda_1^k \frac{pg}{[\lambda_1 - (1-P)][\lambda_1 - (1-p)h]} + c_2\lambda_2^k \frac{pg}{[\lambda_2 - (1-P)][\lambda_2 - (1-p)h]} + c_3\lambda_3^k \frac{pg}{[\lambda_3 - (1-P)][\lambda_3 - (1-p)h]}$$

$$G_2(k) = c_1\lambda_1^k \frac{p}{\lambda_1 - (1-p)h} + c_2\lambda_2^k \frac{p}{\lambda_2 - (1-p)h} + c_3\lambda_3^k \frac{p}{\lambda_3 - (1-p)h}$$

$$B(k) = c_1\lambda_1^k \frac{bph}{[\lambda_1 - (1-p)h]^2} + c_2\lambda_2^k \frac{bph}{[\lambda_2 - (1-p)h]^2} + c_3\lambda_3^k \frac{bph}{[\lambda_3 - (1-p)h]^2}$$

and hence,

$$\begin{aligned} u(k) = {} & c_1\lambda_1^k \left\{\frac{pg}{[\lambda_1 - (1-P)][\lambda_1 - (1-p)h]} + \frac{p}{\lambda_1 - (1-p)h} + \frac{bph}{[\lambda_1 - (1-p)h]^2}\right\} \\ & + c_2\lambda_2^k \left\{\frac{pg}{[\lambda_2 - (1-P)][\lambda_2 - (1-p)h]} + \frac{p}{\lambda_2 - (1-p)h} + \frac{bph}{[\lambda_2 - (1-p)h]^2}\right\} \\ & + c_3\lambda_3^k \left\{\frac{pg}{[\lambda_3 - (1-P)][\lambda_3 - (1-p)h]} + \frac{p}{\lambda_3 - (1-p)h} + \frac{bph}{[\lambda_3 - (1-p)h]^2}\right\} \end{aligned} \tag{13}$$

Next we find the autocorrelation of bit errors, denoted by $r(k), k = 1, 2, \cdots$. By definition then,

$$r(k) = P(z_k = 1 | z_0 = 1); \qquad k = 1, 2, \cdots$$

which is the probability of having an error at time k following a given initial error. The autocorrelation $r(k)$ measures the correlation of error positions, and it is useful for estimating significant error patterns.

Let

$$\begin{aligned} \bar{G}_1(k) &= P(s_k = G_1 | 1) \\ \bar{G}_2(k) &= P(s_k = G_2 | 1) \\ \bar{B}(k) &= P(s_k = B | 1) \end{aligned} \tag{14}$$

Then it follows that

$$r(k) = (1-h)\bar{B}(k)$$

To find $\bar{B}(k)$ we need the following facts:

$$\begin{aligned} \bar{B}(k+1) &= \bar{B}(k)q + \bar{G}_2(k)b \\ \bar{G}_2(k+1) &= \bar{G}_2(k)r + \bar{G}_1(k)P + \bar{B}(k)p \\ \bar{G}_1(k+1) &= \bar{G}_1(k)Q + \bar{G}_2(k)g \end{aligned} \tag{15}$$

which are easy to derive.

Here also, as in (8), we have

$$(\bar{G}_1(k+1), \bar{G}_2(k+1), \bar{B}(k+1)) = (\bar{G}_1(k), \bar{G}_2(k), \bar{B}(k))\bar{M}$$

where

$$\bar{M} = \begin{pmatrix} Q & P & 0 \\ g & r & b \\ 0 & p & q \end{pmatrix}$$

with

$$\bar{G}_1(0) = 0 = \bar{G}_2(0); \qquad \bar{B}(0) = 1$$

Hence,

$$(\bar{G}_1(k+1), \bar{G}_2(k+2), \bar{B}(k+1)) = (0, 0, 1)\bar{M}^{k+1} \tag{16}$$

In this case also, as in (8), we can use a computer to find the vector

$$(\bar{G}_1(k), \bar{G}_2(k), \bar{B}(k)); \qquad k = 1, 2, \cdots$$

But it is also valuable to give an explicit solution in terms of the model parameters by calculating M^k for any k. We shall do this in Section V.

The distribution of consecutive sequence of errors is easily obtained. For example, the probability of a sequence of k errors following a given error, $P(1^k | 1)$, is given by

$$P(1^k | 1) = (qh')^k$$

while the probability of k consecutive errors following a given good bit, $P(1^k | 0)$, is given by

$$P(1^k | 0) = bh'(qh')^{k-1} + (qh')^k; \qquad k = 1, 2, \cdots$$

IV. Capacity of the Channel

For estimating the maximal rate for which reliable transmission over the channel is possible, we shall find an analytic expression for the capacity of the channel in terms of the model parameters.

Let H denote the entropy of the noise sequence $z = \{z_n\}$; then

Proposition 1

The capacity C of the burst-noise channel is given by

$$C = 1 - H$$

and

$$H = -\lim_{n\to\infty} \sum_{z_i = 0 \text{ or } 1} P(z_1, \cdots, z_{n+1}) \log P(z_{n+1} | z_1, \cdots, z_n) \tag{17}$$

Proof:

The proof is by classical information-theoretic arguments. The mutual information of the n-extension of the channel is

$$I(X^n, Y^n) = H(Y^n) - H(Y^n | X^n) \tag{18}$$

where $X(Y)$ is the input (output) and $H(\cdot)$ is the entropy function. The transmission rate is then

$$R = \lim_{n\to\infty} \frac{R(X^n, Y^n)}{n} \tag{19}$$

the capacity C is given by

$$\begin{aligned} C &= \max_{p(x)} R \\ &= \max_{p(x)} \lim_{n\to\infty} \left\{ \frac{H(Y^n)}{n} - \frac{H(Y^n | X^n)}{n} \right\} \end{aligned} \tag{20}$$

and $p(x)$ is an input distribution.

Now for additive noise–i.e., for $Y = X + Z$, Z the error sequence–we can easily show that

$$H(Y^n | X^n) = H(z_1, \cdots, z_n) \tag{21}$$

independent of X^n and that for $p(x_1, \cdots, x_n) = 2^{-n}$,

$$\frac{H(Y^n)}{n} = 1 \tag{22}$$

But

$$\frac{H(z_1, \cdots, z_n)}{n} = H(z_n | z_1, \cdots, z_{n-1})$$

So by Eqs. (20), (21), and (22), we have

$$\begin{aligned} C &= 1 - \lim_{n\to\infty} H(z_n | z_1, \cdots, z_{n-1}) \\ &= 1 - H \end{aligned} \tag{23}$$

where H denotes

$$\lim_{n\to\infty} H(z_n | z_1, \cdots, z_{n-1})$$

and

$$\begin{aligned} &H(z_n | z_1, \cdots, z_{n-1}) \\ &\quad = -\sum_{z_i = 0 \text{ or } 1} p(z_1, \cdots, z_n) \log p(z_n | z_1, \cdots, z_{n-1}) \end{aligned}$$

This completes the proof of Proposition 1.

We note that a uniformly distributed, zero-memory binary source achieves this capacity (Ref. 3).

Now let us write H as:

$$H = \lim_{n\to\infty} \sum_{\{z_1, \cdots, z_n\}} P(z_1, \cdots, z_n)\, h(z_1, \cdots, z_n)$$

with

$$\begin{aligned} &h(z_1, \cdots, z_n) \\ &\quad = -\sum_{z_{n+1} = 0 \text{ or } 1} P(z_{n+1} | z_1, \cdots, z_n) \log P(z_{n+1} | z_1, \cdots, z_n) \end{aligned} \tag{24}$$

And if we assume that our model contains only one error state B, we can show easily (see Ref. 3) that $h(z_1, \cdots, z_n)$ can assume only $(n + 1)$ values:

$$h(0^n), h(10^{n-1}), h(10^{n-2}), \cdots, h(10), h(1) \tag{25}$$

where $\{10^k\}$ is the event that an error is followed by k error-free bits. Using Eqs. (25) and (24), we have

$$H = \sum_{k=0}^{\infty} P(10^k)\, h(10^k) \tag{26}$$

In terms of $u(k)$ and $v(k)$ (see Eqs. 2 and 3), $P(10^k)$ can be written as

$$P(10^k) = P(1)\, u(k)$$

and hence,

$$P(0 | 10^k) = \frac{u(k+1)}{u(k)} \tag{27}$$

Equation (24) then becomes

$$h(10^k) = -\frac{u(k+1)}{u(k)}\log\frac{u(k+1)}{u(k)} - \left(1-\frac{u(k+1)}{u(k)}\right)\log\left(1-\frac{u(k+1)}{u(k)}\right) \tag{28}$$

Hence, by Eq. (26), H becomes

$$H = -P(1)\sum_{k=0}^{\infty} u(k)\left\{\frac{u(k+1)}{u(k)}\log\frac{u(k+1)}{u(k)} + \left(1-\frac{u(k+1)}{u(k)}\right)\log\left(1-\frac{u(k+1)}{u(k)}\right)\right\} \tag{29}$$

The entropy H can also be written in terms of $v(k)$, using the fact that

$$v(k) = u(k) - u(k+1)$$

We see from (26) that H then becomes

$$H = -P(1)\sum_{k=0}^{\infty} v(k)\log v(k) \tag{30}$$

For actual computations, however, it is convenient to use Eq. (29) because, as shown in (12),

$$u(k) = c_1\lambda_1^k + c_2\lambda_2^k + c_3\lambda_3^k \tag{31}$$

where c_1, c_2, c_3 are constants. And since $\lambda_1 < \lambda_2 < \lambda_3$, with λ_1 and λ_2 very much less than 1, we have

$$\frac{u(k+1)}{u(k)} \to \lambda_3 \tag{32}$$

By (29), then,

$$h(10^k) \to -\lambda_3\log\lambda_3 - (1-\lambda_3)\log(1-\lambda_3) = h_0, \text{ say} \tag{33}$$

The convergence in (32) is very fast. For example, for $P(1) = 4 \times 10^{-5}$, $h' = 0.5$, the transition probabilities are

$$p = 0.2, \qquad q = 0.8$$

$$b = 0.356, \qquad r = 0.002, \qquad g = 0.642$$

$$P = 0.0003, \qquad Q = 0.99997$$

and $[u(k+1)]/[u(k)] = 0.999991$ (to six decimal places) for $k \geqq 21$, while $\lambda_3 = 0.999989$.

Thus, in general the approximation $h(10^k) = h_0$ is good for all $k \geqq k_0$, some k_0 large enough. Therefore, using Eq. (33) in (29), we get

$$H = -P(1)\sum_{k=0}^{k_0-1} u(k)\left\{\frac{u(k+1)}{u(k)}\log\frac{u(k+1)}{u(k)} + \left(1-\frac{u(k+1)}{u(k)}\right)\log\left(1-\frac{u(k+1)}{u(k)}\right)\right\} + h_0P(1)\sum_{k=k_0}^{\infty} u(k) \tag{34}$$

But

$$P(1)\sum_{k=k_0}^{\infty} u(k) = \sum_{k=k_0}^{\infty} P(10^k) = 1 - \sum_{k=0}^{k_0-1} P(10^k) = 1 - P(1)\sum_{k=0}^{k_0-1} u(k)$$

Hence, the capacity of the channel is given by

$$C = 1 - H$$

$$\simeq 1 + P(1)\sum_{k=0}^{k_0-1} u(k)\left\{\frac{u(k+1)}{u(k)}\log\frac{u(k+1)}{u(k)} + \left(1-\frac{u(k+1)}{u(k)}\right)\log\left(1-\frac{u(k+1)}{u(k)}\right)\right\} - h_0\left[1 - P(1)\sum_{k=0}^{k_0-1} u(k)\right] \tag{35}$$

Comparison between the model capacity C and that of an equivalent binary symmetric channel C (BSC), for two sets of parameter values, is shown in Table 1.

V. Block Error Statistics

We turn now to computing the group of parameters which has the most direct application to block codes. Specifically, for estimating the performances of burst-correcting codes we find (A) the probability of getting an error block; (B) the distribution of the number of errors in a block, denoted by $P(k,n)$; (C) the probability, denoted by $P_t(k,n)$, of k information-bit errors in n-information-bit word in an interleaved code with constant of interleaving t; (D) the distribution of symbol errors in an n-symbol word; and (E) the distribution of distances between extreme errors in a block.

A. Proportion of Blocks in Error

As in Ref. 5, let

$$TOU(n) = P(z_1 = \cdots = z_n = 0, s_n = U \mid s_0 = T); \quad U, T = B, G_1, G_2 \tag{36}$$

Thus, $TOU(n)$ is the probability, starting from a given initial state T, of having a gap of length n and ending at a final state U at time n. Then

$$\begin{aligned} BOB(1) &= qh \\ BOG_2(1) &= p \\ BOG_1(1) &= 0 \\ G_2OB(1) &= bh \\ G_2OG_2(1) &= r \\ G_2OG_1(1) &= g \\ G_1OB(1) &= 0 \\ G_1OG_2(1) &= P \\ G_1OG_1(1) &= Q \end{aligned}$$

and

$$\begin{aligned} BOB(n) &= BOB(n-1)\,qh + BOG_2(n-1)\,bh \\ BOG_2(n) &= BOB(n-1)\,p + BOG_2(n-1)\,r \\ &\quad + BOG_1(n-1)\,P \\ BOG_1(n) &= BOG_2(n-1)\,g + BOG_1(n-1)\,Q \\ G_2OB(n) &= G_2OB(n-1)\,qh + G_2OG_2(n-1)\,bh \\ G_2OG_2(n) &= G_2OB(n-1)\,p + G_2OG_2(n-1)\,r \\ &\quad + G_2OG_1(n-1)\,P \\ G_2OG_1(n) &= G_2OG_2(n-1)\,g + G_2OG_1(n-1)\,Q \\ G_1OB(n) &= G_1OB(n-1)\,qh + G_1OG_2(n-1)\,bh \\ G_1OG_2(n) &= G_1OB(n-1)\,p + G_1OG_2(n-1)\,r \\ &\quad + G_1OG_1(n-1)\,P \\ G_1OG_1(n) &= G_1OG_2(n-1)\,g + G_1OG_1(n-1)\,Q \end{aligned} \tag{37}$$

$$\begin{aligned} P(z_1 = 0 = \cdots = z_n = 0) &= P(s_0 = B)\,[BOB(n) \\ &\quad + BOG_2(n) + BOG_1(n)] \\ &\quad + P(s_0 = G_2)\,[G_2OB(n) \\ &\quad + G_2OG_2(n) + G_2OG_1(n)] \\ &\quad + P(s_0 = G_1)\,[G_1OB(n) \\ &\quad + G_1OG_2(n) + G_1OG_1(n)] \end{aligned}$$

and

$$P(\text{block error}) = 1 - P(z_1 = \cdots = z_n = 0) \tag{38}$$

B. P(k, n)

Following Ref. 5 again, let $TU(n,k) = P(k \text{ bit errors in } n\text{-bit words}, s_n = U \mid s_0 = T)$, which as before is the probability, given initial state T, of getting k errors in n-bit word and ending at a final state U. Then

$$\begin{aligned} BB(1,0) &= BOB(1) = qh \\ BG_2(1,0) &= BOG_2(1) = p \\ BG_1(1,0) &= BOG_1(1) = 0 \\ G_2B(1,0) &= G_2OB(1) = bh \\ G_2G_2(1,0) &= G_2OG_2(1) = r \\ G_2G_1(1,0) &= G_2OG_1(1) = g \\ G_1B(1,0) &= G_1OB(1) = 0 \\ G_1G_2(1,0) &= G_1OG_2(1) = P \\ G_1G_1(1,0) &= G_1OG_1(1) = Q \\ BB(1,1) &= B1B(1) = qh' \\ BG_2(1,1) &= B1G_2(1) = 0 \\ BG_1(1,1) &= B1G_1(1) = 0 \\ G_2B(1,1) &= G_21B(1) = bh' \\ G_2G_2(1,1) &= G_21G_2(1) = 0 \\ G_2G_1(1,1) &= G_21G_1(1) = 0 \\ G_1B(1,1) &= G_11B(1) = 0 \\ G_1G_2(1,1) &= G_11G_2(1) = 0 \\ G_1G_1(1,1) &= G_11G_1(1) = 0 \end{aligned} \tag{39}$$

In general, for $k = 0, 1, \cdots, n$, we have

$$\begin{aligned} BB(n,k) &= BB(n-1,k)\,qh + BG_2(n-1,k)\,bh \\ &\quad + BB(n-1,k-1)\,qh' \\ &\quad + BG_2(n-1,k-1)\,bh' \\ BG_2(n,k) &= BB(n-1,k)\,p + BG_2(n-1,k)\,r \\ &\quad + BG_1(n-1,k)\,P \end{aligned}$$

$$BG_1(n,k) = BG_2(n-1,k)\,g + BG_1(n-1,k)\,Q$$

$$\begin{aligned} G_2B(n,k) &= G_2B(n-1,k)\,qh + G_2G_2(n-1,k)\,bh \\ &\quad + G_2B(n-1,k-1)\,qh' \\ &\quad + G_2G_2(n-1,k-1)\,bh' \end{aligned}$$

$$\begin{aligned} G_2G_2(n,k) &= G_2B(n-1,k)\,p + G_2G_2(n-1,k)\,r \\ &\quad + G_2G_1(n-1,k)\,P \end{aligned}$$

$$G_2G_1(n,k) = G_2G_2(n-1,k)\,g + G_2G_1(n-1,k)\,Q$$

$$G_1G_1(n,k) = G_1G_2(n-1,k)\,g + G_1G_1(n-1,k)\,Q$$

$$\begin{aligned} G_1B(n,k) &= G_1B(n-1,k)\,qh + G_1G_2(n-1,k)\,bh \\ &\quad + G_1B(n-1,k-1)\,qh' \\ &\quad + G_1G_2(n-1,k-1)\,bh' \end{aligned}$$

$$\begin{aligned} G_1G_2(n,k) &= G_1B(n-1,k)\,p + G_1G_2(n-1,k)\,r \\ &\quad + G_1G_1(n-1,k)\,P \end{aligned} \tag{40}$$

$$\begin{aligned} P(k \text{ errors in } n\text{-word}) &= P(s_0 = B)\,[BB(n,k) \\ &\quad + BG_2(n,k) + BG_1(n,k)] \\ &\quad + P(s_0 = G_2)\,[G_2B(n,k) \\ &\quad + G_2G_2(n,k) + G_2G_1(n,k)] \\ &\quad + P(s_0 = G_1)\,[G_1B(n,k) \\ &\quad + G_1G_2(n,k) + G_1G_1(n,k)] \end{aligned} \tag{41}$$

where $P(s_0 = B)$, $P(s_0 = G_2)$, $P(s_0 = G_1)$ are as given in (1).

C. Distribution of Symbol Errors

To find P(k-symbol errors in n-symbol word), in which we take a symbol to be in error if at least one of its m bits is in error, we note that the algorithm above works here also if we replace n by nm. We omit the details.

D. $P_t(k, n)$

It is clear that $P_t(k,n)$ is the same as the probability of getting k errors in a block of length tn on our channel sampled at every tth step. That is, if, as in (15), we put

$$\bar{M} = \begin{pmatrix} Q & P & 0 \\ g & r & b \\ 0 & p & q \end{pmatrix} = \begin{pmatrix} 1-P & P & 0 \\ g & 1-g-b & b \\ 0 & p & 1-p \end{pmatrix} \tag{42}$$

to calculate $P_t(k,n)$, all we need is the t-step transition matrix $M^{(t)}$; we use the $M^{(t)}$ entries as our transitions instead of M and get $P_t(k,n)$ using the algorithm (Eqs. 39–41) which gave us $P(k,n)$.

By the method of Ref. 8 (p. 385) used in Section III, we get

$$\bar{M}^{(t)} = \frac{1}{pP + bP + pg} M_1 + \frac{1}{D_1}\left(\frac{2-A}{2}\right)^t M_2 + \frac{1}{D_2}\left(\frac{2-B}{2}\right)^t M_3$$

where

$$M_1 = \begin{pmatrix} pg & pP & bP \\ pg & pP & bP \\ pg & pP & bP \end{pmatrix}$$

$$M_2 = \begin{pmatrix} 4(2p-A)^2 Pg & 2(2p-A)^2(2P-A)P & 4(2P-A)(2p-A)bP \\ 2(2p-A)^2(2P-A)g & (2p-A)^2(2P-A)^2 & 2(2P-A)^2(2p-A)b \\ 4(2P-A)(2p-A)pg & 2(2P-A)^2(2p-A)p & 4(2P-A)^2 pb \end{pmatrix}$$

$$M_3 = \begin{pmatrix} 4(2p-B)^2 Pg & 2(2p-B)^2(2P-B)P & 4(2P-B)(2p-B)Pb \\ 2(2p-B)^2(2P-B)g & (2p-B)^2(2P-B)^2 & 2(2P-B)^2(2p-B)b \\ (2P-B)(2p-B)4pg & 2(2P-B)^2(2p-B)p & 4(2P-B)^2 pb \end{pmatrix}$$

$$D_1 = 4Pg\,[2p - A]^2 + (2P - A)^2 (2p - A)^2 + 4pb\,(2P - A)^2$$

$$D_2 = 4Pg\,[2p - B]^2 + (2P - B)^2 (2p - B)^2 + 4pb\,(2P - B)^2$$

$$A = X + \sqrt{X^2 - 4Y}$$

$$B = X - \sqrt{X^2 - 4Y}$$

$$X = P + p + g + b$$

$$Y = pP + bP + pg$$

For example, the $P(s_t = G_1 | s_1 = B)$ of being in state G_1 after t steps starting from state B is given by

$$P(s_t = G_1 | s_1 = B) = \frac{pg}{pg + pP + bP} + \frac{4(2P - A)(2p - A)}{D_1}\left(\frac{2 - A}{2}\right)^t + \frac{4(2P - B)(2p - B)}{D_2}\left(\frac{2 - B}{2}\right)^t$$

E. Distances Between Extreme Errors in a Block

Denote by P_k the probability of k bits between extreme errors in a block of length n, given there are at least two errors in the block. Then, by definition,

$$P_k = \frac{P\,(k \text{ bits between extreme errors and} \geqq 2 \text{ errors in the block}}{P\,(\geqq 2 \text{ errors in the block})}$$

By definition, the numerator is equal to

$$\sum_{x=0}^{N-k-2} P\,(0^x 1 \underset{k}{\leftrightarrow} 10^y); \qquad y = N - k - 2 - x$$

$$= \sum_{x=0}^{N-k-2} P\,(0^x 1)\, P\,(z_{k+1} = 1 | z_0 = 1)\, P\,(0^y | 0^x 1 \underset{k}{\leftrightarrow} 1)$$

$$= \sum_{x=0}^{N-k-2} P\,(0^x 1)\, r\,(k + 1)\, P\,(0^y | 1)$$

$$= P\,(1)\, r\,(k+1) \sum_{x=0}^{N-k-2} P\,(0^x | 1)\, P\,(0^{N-k-2-x} | 1)$$

$$= P\,(1)\, r\,(k+1) \sum_{x=0}^{N-k-2} u\,(x)\, u\,(N - k - 2 - x)$$

where we have used the fact that

$$u\,(x) = P\,(0^x | 1) \text{ and } u\,(N - k - x - 2) = P\,(0^{N-k-2-x} | 1)$$

for which expressions in terms of the model parameters were found in Section III.

Also,

$$(P \geqq 2 \text{ errors in block}) = 1 - P\,(0, n) - P\,(1, n)$$

Hence,

$$P_k = \frac{P\,(1)\, r\,(k+1) \sum_{x=0}^{N-k-2} u\,(x)\, u\,(N - k - 2 - x)}{1 - P\,(0, n) - P\,(1, n)} \tag{43}$$

VI. The Burst Statistics

A. Distribution of Burst Lengths

We start with the distribution of burst lengths. Denote the guardspace by $\mathcal{G}$ and the probability of a burst of length n, for $n = 1, 2, \cdots$, by $L\,(n)$. Then, by definition, it follows that

$$L\,(n) = \sum_{k=0}^{\min(\mathcal{G}-1,\, n-2)} P\,(0^k 1 0^l 1 0^m 1 \cdots 1 0^t | 1) \tag{44}$$

over all $l, m, \cdots$, such that

$$0 \leqq l, m, \cdots, \leqq \mathcal{G} - 1$$
$$t \geqq \mathcal{G} \tag{45}$$

and

$$0^k 1 0^l 1 0^m 1 \cdots\cdots 1 0^t$$

is a sequence of $(n - 1) + t$ bits.

Proposition 2

$$L(n) = \begin{cases} 0; & n \leq 0 \\ u(\mathcal{G}) = P(0^{\mathcal{G}} | 1); & n = 1 \\ \displaystyle\sum_{k=0}^{\min(\mathcal{G}-1, n-2)} v(k) L(n-k-1); & n \geq 2 \end{cases} \quad (46)$$

Proof:

Equation (46) is almost obvious by inspection, but let us sketch a proof. Let us note first that

(1) $L(0) = 0$

(2) $L(1) = P(0^{\mathcal{G}} | 1) = u(\mathcal{G})$

since a burst must start with an error, implying that the least length a burst can have is 1.

From Eq. (44),

$$L(n) = \sum_{k=0}^{\min(\mathcal{G}-1, n-2)} P(0^k 1 0^l 1 0^m 1 \cdots 1 0^t | 1); \qquad t \geq \mathcal{G}$$

$$= \sum_{k \geq 0} P(0^k 1 | 1) P\{0^l 1 0^m 1 \cdots 1 0^t | 1\} \quad (47)$$

But

$$P\{0^l 1 0^m 1 \cdots 1 0^t | 1\} = L(n-k-1)$$

and

$$P(0^k 1 | 1) = v(k)$$

Now substitute these expressions in Eq. (47) to obtain (46).

B. Distribution of Errors Within a Burst

The last parameter of interest is the distribution of errors within a burst of a given length: P (k errors in a burst of length N). We state this in the following proposition.

Proposition 3

$$P(k \text{ errors in a given burst of length } N) = \frac{Q(k, N)}{L(N)}$$

where

$$Q(k, N) = \begin{cases} 0 \text{ if } 0 = k = N; \text{ or } 0 \leq k \leq 1, N \geq 2; \text{ or } k > N \\ u(\mathcal{G}) = P(0^{\mathcal{G}} | 1) \qquad \text{if } N = k = 1 \\ \displaystyle\sum_{x=0}^{\min(\mathcal{G}-1, n-k)} v(x) Q(k-1, N-x-1); \\ \qquad 2 \leq k \leq N \end{cases} \quad (48)$$

Here $Q(k, N)$ is the probability of getting a burst of length N that contains k errors.

Proof:

$$P(k \text{ errors in a given burst of length } N)$$

$$= \frac{P(\text{burst of length } N \text{ with } k \text{ errors})}{P(\text{burst of length } N)} \quad (49)$$

Write the numerator as $Q(k, N)$. Then, by definition of a burst,

$$\begin{aligned} &Q(0,0) = 0 \\ &Q(k, N) = 0 \quad \text{for} \quad 0 \leq k \leq 1, N \geq 2; \quad \text{or} \quad k > N \\ &Q(1,1) = P(0^{\mathcal{G}} | 1) = u(\mathcal{G}) \end{aligned} \quad (50)$$

Now, for $2 \leq k \leq N$,

$$Q(k, N) = P\{0^x \underbrace{1^y 1 \cdots 1 0^t}_{N-1} | 1\}$$

where the $(N-1)$ bits indicated contain $(k-1)$ errors, $0 \leq x, y \cdots \leq \mathcal{G} - 1$; $t \geq \mathcal{G}$. Hence,

$$\begin{aligned} Q(k, N) &= \sum_{x=0}^{\min(\mathcal{G}-1, n-k)} P(0^x \underbrace{1 0^y 1 \cdots 1 0^t}_{N-1} | 1) \\ &= \sum_{x=0}^{\min(\mathcal{G}-1, n-k)} P(0^x 1 | 1) P\{0^y \underbrace{1 \cdots 1 0^t}_{N-x-2} | 1\} \end{aligned} \quad (51)$$

$$0 \leq y \cdots \leq \min(\mathcal{G} - 1, N - x - k); \qquad t \geq \mathcal{G}$$

and the $(N - x - 2)$ bits indicated contain $(k - 2)$ errors.

Therefore,

$$Q(k,N) = \sum_{x=0}^{\min(\mathcal{G}-1,\, n-k)} v(x)\, Q(k-1, N-x-1) \qquad (52)$$

Since P (burst of length n) $= L(N)$, the proposition is proved.

VII. Conclusion

The empirical counterparts of all the probabilities are now being computed. A more detailed analysis of this model, including recommendations as to the optimal error-detecting and -correcting codes to be employed on the channel, will be contained in a more detailed report in preparation.

References

1. McClure, J. P., "GCF 50-kbps Wideband Data Error Rate Test," in *The Deep Space Network Progress Report*, Technical Report 32-1526, Vol. VI, pp. 149–151, Jet Propulsion Laboratory, Pasadena, Calif., Dec. 15, 1971.
2. McClure, J. P., "Madrid-to-JPL 50-kbit/s Wideband Error Statistics," in *The Deep Space Network Progress Report,* Technical Report 32-1526, Vol. IX, pp. 177–180, Jet Propulsion Laboratory, Pasadena, Calif., June 15, 1972.
3. Gilbert, E. N., "Capacity of a Burst-Noise Channel," *Bell Sys. Tech. J.*, Vol. 39, 1960, pp. 1253–1265.
4. Elliott, E. O., "Estimates of Error Rates for Codes on Burst-Noise Channels," *Bell Sys. Tech. J.*, Vol. 42, 1963, pp. 1977–1997.
5. Berkovits, S., Cohen, E. L., and Zierler, N., *A Model for Digital Error Distribution*, Tech. Memo, Mitre Corporation, Bedford, Mass.
6. Stern, Daniel C., *Statistical Analysis of Errors Occurring in the Transmission of High-Speed Digital Data*, Master's Thesis, Dept. of EE., University of Maryland, 1970.
7. Sussman, S. M., "Communication Channel Attributes as Related to Error Control," IEEE Annual Communications Convention, 1965, pp. 5–13.
8. Feller, W., *An Introduction to Problem Theory and its Application*, Vol. I, 2nd ed., John Wiley, New York, 1959.

Table 1. Comparison between capacity of model and that of equivalent BSC

Model parameters	C	C (BSC)	$\frac{C\,(\mathrm{BSC})}{C}$
$P(1) = 4 \times 10^{-5}, h' = 0.5$ $p = 0.20, q = 0.80$ $b = 0.356, r = 0.21 \times 10^{-2}$ $g = 1 - b - r$ $P = 0.289 \times 10^{-4}$ $Q = 0.9999711$	0.999831	0.999551	0.9970
$P(1) = 0.2$ $p = 0.25; q = 0.75$ $b = 0.205$ $r = 0.699$ $P = 0.1511$ $Q = .8489$	0.56007	0.499598	0.8920

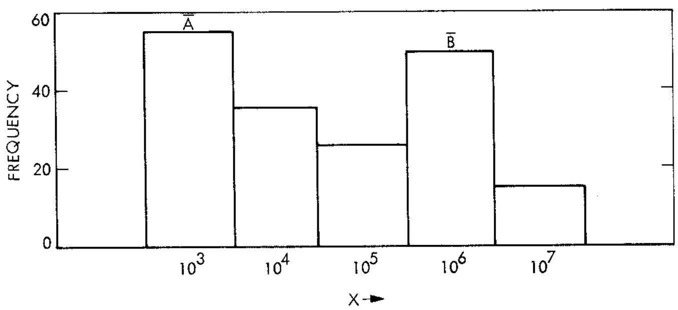

Fig. 1. Histogram for gap distribution

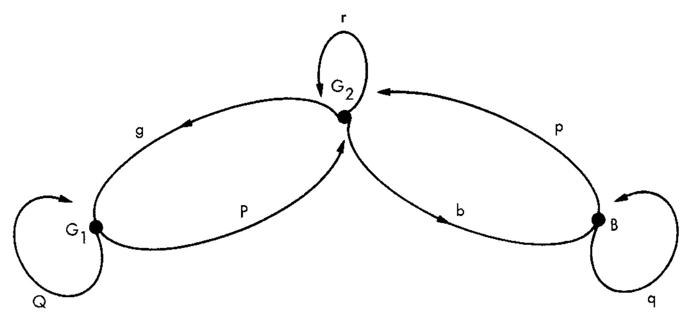

Fig. 3. Transition diagram for the Markov model

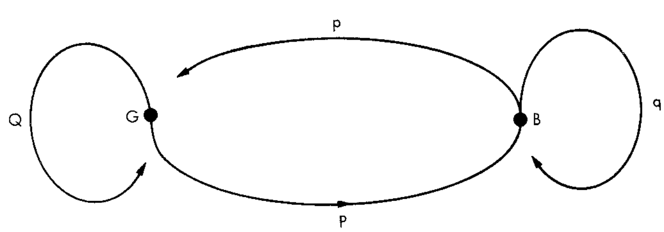

Fig. 2. Gilbert model