Let the values of $S_i$ (see Fig. 2) for the register producing $A_1, A_2, \cdots A_n$ be $S_1, S_2, \cdots S_{k-1}$. Then we must have

$$b_i = a_{i-1} \qquad i = 2, 3, \cdots k-1 \tag{19}$$

$$b_1 = \sum_{i=1}^{k-1} S_i a_i \tag{20}$$

and

$$C_i = b_{i-1} \qquad i = 2, 3, \cdots k-1 \tag{21}$$

$$C_1 = \sum_{i=1}^{k-1} S_i b_i. \tag{22}$$

Furthermore, we have from (16) and (17)

$$d_i = a_i + b_i \tag{23}$$

$$e_i = b_i + c_i. \tag{24}$$

Then

$$\begin{aligned} e_1 &= b_1 + c_1 \\ &= \sum_{i=1}^{k-1} S_i a_i + \sum_{i=1}^{k-1} S_i b_i \\ &= \sum_{i=1}^{k-1} S_i (a_i + b_i) \\ &= \sum_{i=1}^{k-1} S_i\, d_i. \end{aligned} \tag{25}$$

Furthermore, for $i = 2, 3, \cdots k-1$

$$\begin{aligned} e_i &= b_i + c_i \\ &= a_{i-1} + b_{i-1} \\ &= d_{i-1}. \end{aligned} \tag{26}$$

Now, since the arguments given will apply to any two successive $B_i$, (25) and (26) show that the sequence of the $B_i$ may be obtained from the same shift register as the $A_i$. Finally, since this is a maximal-length shift register, the sequence of the $B_i$ must just be a shifted version of the sequence of the $A_i$, and (18) must hold.

ACKNOWLEDGMENT

The author is indebted to Dr. B. Elspas for several illuminating discussions on maximal-length feedback shift registers and their generalizations.

BIBLIOGRAPHY

[1] R. W. Hamming, "Error correcting and error detecting codes," *Bell Sys. Tech. J.*, vol. 29, pp. 147–160; April, 1950.
[2] D. Slepian, "A class of binary signaling alphabets," *Bell Sys. Tech. J.*, vol. 35, pp. 203–234; January, 1956.
[3] I. S. Reed, "A class of multiple-error-correcting codes and the decoding scheme," IRE TRANS. ON INFORMATION THEORY, vol. IT-4, pp. 38–49; September, 1954.
[4] J. H. Green, Jr. and R. L. San Soucie, "An error correcting encoder and decoder of high efficiency," PROC. IRE, vol. 46, pp. 1741–1744; October, 1958.
[5] N. Zierler, "Several Binary-Sequence Generators," Lincoln Lab., M.I.T., Lexington, Mass., Tech. Rept. No. 95; September 12, 1955.
[6] S. W. Golomb, "Sequences With Randomness Properties," Glenn L. Martin Co., Baltimore, Md., Final Rept. on Contract No. W 36-039SC-54-36611; June 14, 1955.
[7] B. Elspas, "Theory of autonomous linear sequential networks," IRE TRANS. ON CIRCUIT THEORY, vol. CT-6, pp. 45–60; March, 1959.
[8] R. W. Marsh, "Table of Irreducible Polynomials Over GF(2) Through Degree 19," National Security Agency, Washington, D. C., October 24, 1957.

# A Probabilistic Model for Run-Length Coding of Pictures*

JACK CAPON†

*Summary*—**A first-order Markoff process representation for pictures is proposed in order to study the picture coding system known as run-length coding (differential-coordinate encoding). A lower bound for the saving in channel capacity is calculated on the basis of this model, and is compared with the results obtained by previous investigators. In addition, this representation is shown to yield an insight into the run-length coding system which might not otherwise be obtained. The application of this probabilistic model to an "elastic" system of run-length coding is also discussed.**

* Manuscript received by the PGIT, December 22, 1958. This work was done when the author was employed by the Ford Instrument Company, Long Island City, N. Y., during the summer of 1958.

† Federal Scientific Corporation, New York, N. Y. Formerly with the Dept. Elec. Engrg., Columbia University, New York, N. Y.

## INTRODUCTION

It has been recognized that a saving in either time or bandwidth is possible by exploiting the large-scale redundancies in pictures. One of the first proposals advanced for taking advantage of these redundancies is the variable sweep system of Cherry and Gouriet.[1] The authors found that theoretically a saving of approximately seven in either time or bandwidth is possible with their system.

[1] E. C. Cherry and G. G. Gouriet, "Some possibilities for the compression of television by recoding," *Proc. IEE*, vol. 100, pt. III, pp. 9–18; January, 1953.

Another scheme which has been proposed is the run-length coding, or differential-coordinate encoding system.[2] Although this system is applicable to half-tone pictures[3] we shall, for simplicity, consider its implementation only for black and white pictures. The basic idea in this scheme is to transmit only the lengths of the black and white runs in a picture as they occur in successive scanning lines. The first implementation of this system appears to be that of Treuhaft.[4] However, he gives no results concerning the possible reduction in either time or bandwidth for a run-length coding type system. Deutsch[5] has measured the probability distribution of runs in a section of typewritten material, and concluded that a saving of approximately two is possible in a run-length coding system which uses an optimum code. More extensive studies of the run-length probability distributions for typewritten material has been made by Michel[6] who concluded from his findings that theoretically a saving of approximately ten is possible with a run-length coding system which employs an optimum code.

These results[5,6] are sufficiently encouraging to warrant further investigation of the run-length coding system. The present work stems from a desire to predict more generally the saving that is possible with a run-length coding system for various types of black and white pictures, and to gain an insight into this system. The analysis is carried out only for black and white pictures, and for a binary digital transmission channel.

## The Description of a Picture in Terms of a First-Order Markoff Process

The process of scanning reduces a picture from a two-dimensional array of cells (resolution elements) to a one-dimensional sequence of cells. In the case of a black and white picture such a sequence would consist of a succession of black and white cells. A section of this sequence might appear as below.

$\cdots$ BBBWWBBBBBBWWWWBWBBWBWW $\cdots$

Thus, in our subsequent discussion, when we use the word "picture" we shall, in reality, be referring to a one-dimensional sequence of cells which results from scanning a picture. (*Note*: Since there are many ways to scan a picture, this representation is not unique. However, for our purposes, this is unimportant.)

The assumption, on which the analysis presented herein is based, can now be stated.

The transition in intensity from a given cell to the immediately succeeding cell is determined solely by the intensity of the given cell.

This dependence is a probabilitistic one which is given in terms of the transition probabilities $p_w(w)$, $p_w(b)$, $p_b(b)$, and $p_b(w)$, where[7]

$p_w(w)$ is the probability that a cell is white, given that the immediately preceding cell is white;

$p_w(b)$ is the probability that a cell is black, given that the immediately preceding cell is white

$$= 1 - p_w(w);$$

$p_b(b)$ is the probability that a cell is black, given that the immediately preceding cell is black;

$p_b(w)$ is the probability that a cell is white, given that the immediately preceding cell is black

$$= 1 - p_b(b).$$

Two other parameters which enter into the analysis are are $p(w)$, and $p(b)$, where[8]

$p(w)$ is the probability of obtaining a white cell;

$p(b)$ is the probability of obtaining a black cell

$$= 1 - p(w).$$

The transition probabilities can be related to each other, by noting first that

$$p(w; b) = p(b; w)$$

where

$p(w; b) =$ the joint probability of obtaining a white cell followed by a black cell;

$p(b; w) =$ the joint probability of obtaining a black cell followed by a white cell.

Hence, since $p(w; b) = p(w)p_w(b)$, and $p(b; w) = p(b)p_b(w)$, we obtain

$$p(w)p_w(b) = p(b)p_b(w).$$

It is easily seen that only two of the six probabilities $p(w)$, $p(b)$, $p_w(w)$, $p_w(b)$, $p_b(b)$, $p_b(w)$ are independent; for example, an independent set comprises $p(w)$ [or $p(b)$], and any one of the four transition probabilities.

At this point we note that the assumption is equivalent to stating that a picture can be represented by a first-order Markoff process with stationary transition probabilities.[9] This model for an information source has been proposed previously by Oliver,[10] among others.

[2] A. E. Laemmel, "Coding Processes for Bandwidth Reduction in Picture Transmission," Microwave Res. Inst., Polytechnic Inst. of Brooklyn, Brooklyn, N. Y., Rept. No. R 246–51; August, 1951.

[3] W. F. Schreiber and C. F. Knapp, "TV bandwidth reduction by digital coding," 1958 IRE National Convention Record, pt. 4, pp. 88–99.

[4] M. A. Treuhaft, "Description of a System for Transmission of Line Drawings with Bandwidth-Time Compression," Microwave Res. Inst., Polytechnic Inst. of Brooklyn, Brooklyn, N. Y., Rept. No. R 339-53; September 4, 1953.

[5] S. Deutsch, "Some Statistics Concerning Typewritten or Printed Material," Microwave Res. Inst., Polytechnic Inst. of Brooklyn, Brooklyn, N. Y., Rept. No. R-526; October 31, 1956.

[6] W. S. Michel, "Statistical encoding for text and picture communication," *Commun. and Electronics*, vol. 35, pp. 33–36; March, 1958.

[7] It is well to emphasize that the transition probabilities are not unique, since they depend on the particular scanning direction which is used. However, for a given scanning direction the transition probabilities are unique.

[8] $p(w)$ and $p(b)$ are unique for a particular picture, since they do not depend on the scanning direction.

[9] W. Feller, "An Introduction to Probability Theory and Its Applications," John Wiley and Sons, Inc., New York, N. Y., vol. 1, 2nd ed.; 1957.

[10] B. M. Oliver, "Efficient coding," *Bell Sys. Tech. J.*, vol. 31, pp. 724–750; July, 1952.

Since a transition from a given intensity to any other intensity is possible in a sufficiently large number of steps, it follows that our Markoff process representation for pictures is transitive, and thus ergodic. This fact will be of importance to us in our subsequent work.

It should be mentioned that our model does not take into account all of the dependence which exists among the various cells in the picture. Hence we can only expect that our results concerning information rates and savings in channel capacity are first-order approximations to the true values.

In particular, our results for the saving in channel capacity will at best be crude lower bounds. However, it should be borne in mind that a model which takes into account more of the dependences among the cells in the picture than ours will not be as amenable to analysis as our model.

Elias[11] has calculated the possible saving in channel capacity of a run-length coding system on the basis of a model in which successive cells in the picture are considered to be independent.

## Number of Bits Required to Code a Picture by Means of a Nonstatistical Coding System

The present-day picture transmission method which specifies the intensity of each cell in a picture independently of the intensity of any other cell will be termed a nonstatistical coding system. This system does not exploit the redundancies which exist in a picture; as a consequence, the system is inefficient in the sense that it requires a larger number of bits to specify the picture than is actually necessary.

We now proceed to calculate the number of bits $N_1$ required by a nonstatistical coding system to specify a picture. Since the nonstatistical coding system specifies the intensity of each cell in a picture independently of all the other cells, it is clear that one bit is required for each cell in the picture. If we let $N$ represent the total number of cells in the picture, then it is apparent that $N_1$ is equal to $N$.

It is advantageous to pause at this point to investigate the number $N$. If we consider that an average picture is $8\frac{1}{2}$ by 11 inches and that the resolution is 100 lines per inch in the vertical and horizontal directions, then $N$ will be $(8.5)(11)(100)^2$, or approximately one million. Hence, we observe that $N$ is very large, and indeed for our purposes may be considered to be infinite. We shall make use of this fact in our subsequent work by investigating the behavior of coding systems as $N$ approaches infinity. Since $N$ is very large, such a limiting behavior will tend to specify the true situation quite closely. In addition, taking limits as $N$ approaches infinity enables us to simplify considerably the expressions which will be encountered.

[11] P. Elias, "Predictive coding," IRE Trans. on Information Theory, vol. IT-1, pp. 16–33; March, 1955.

## Minimum Number of Bits Required to Code the Markoff Source Representation of a Picture

We now consider the minimum possible number of bits, $N_{\min}$, required by a statistical coding system to code the Markoff source representation of a picture. It is well known[12,13] that if the Markoff process is ergodic, and is described by the transition probabilities introduced previously, then the statistical coding system which achieves this minimum is one which codes the intensity of each cell on the basis of a knowledge of the intensity of the immediately preceding cell. If a binary alphabet is used, then it can be shown that $N_{\min}/N$ is equal to the conditional information content (conditional entropy) of the Markoff process; *i.e.*,[14]

$$\frac{N_{\min}}{N} = -p(b)p_b(b)\log p_b(b) - p(b)p_b(w)\log p_b(w) - p(w)p_w(w)\log p_w(w) - p(w)p_w(b)\log p_w(b). \quad (1)$$

We emphasize again that any coding system, whether statistical or nonstatistical, cannot code a picture with a fewer number of bits than $N_{\min}$. However, it is possible for a coding system to achieve the lower bound $N_{\min}$. In fact, as we shall see subsequently, the run-length coding system does achieve this lower bound as $N$ approaches infinity.

## Number of Bits Required by the Run-Length Coding System to Code the Markoff Process Representation of Pictures

Unlike the statistical coding scheme described above, which codes individual cells, the run-length coding system codes groups of cells.[15] Each group consists of an all-white or all-black linear array of cells. Such arrays or "runs," are readily found along the customary scanning lines. The run-length coding system counts the number of cells in each run and suitably encodes this number for transmission. If a binary alphabet is used in the coding process, then it can be shown[12] that the minimum number of bits required to code the run-lengths is given by the information content of the probability distribution of the run-lengths. Thus, in order to calculate the minimum number of bits required by the run-length coding system, we shall digress and find the probability distribution of the run-lengths. It should be pointed out that no specific codes are proposed (*e.g.*, Shannon-Fano-Huffman codes), but that only the minimum number of bits required to code an information source is computed.

[12] C. E. Shannon, "A mathematical theory of communication," *Bell Sys. Tech. J.*, vol. 27, pp. 379–473, July, 1948; pp. 623–656, October, 1948.

[13] A. I. Khinchin, "Mathematical Foundations of Information Theory," Dover Publications, Inc., New York, N. Y., pp. 13–28; 1957.

[14] All logarithms are to the base two.

[15] W. S. Michel, W. O. Fleckenstein, and E. R. Kretzmer, "A coded facsimile system," 1957 WESCON Convention Record, pt. 2, pp. 84–93.

A white run of length $x$ ($= 1, 2, \cdots$) is defined as a set of $x$ consecutive white cells preceded by a black cell and succeeded by a black cell. A black run of length $x$ is defined analogously.

Let $R_w$ and $R_b$ be the random variables which represent, respectively, the length of a white and a black run. In view of the assumption we have made, we see that as $N$ approaches infinity, the probability of obtaining a white run of length equal to $x$ is the same as the probability that $(x - 1)$ transitions from white to white have taken place, and that this is followed by a transition from white to black. Thus[16]

$$\text{prob}\,(R_w = x) = p_w(w)^{x-1}p_w(b), \qquad x = 1, 2, \cdots \quad (N \to \infty). \tag{2}$$

Similarly, it is obtained that the probability of obtaining a black run of length $x$ is

$$\text{prob}\,(R_b = x) = p_b(b)^{x-1}p_b(w), \qquad x = 1, 2, \cdots \quad (N \to \infty). \tag{3}$$

If we let $H_w$ denote the information content of the white run-length probability distribution, then

$$\begin{aligned} H_w &= -\sum_{x=1}^{\infty} \text{prob}\,(R_w = x) \log \text{prob}\,(R_w = x) \\ &= -\log\,[(1 - p_w(w))/p_w(w)] \\ &\qquad - [1 - p_w(w)]^{-1} \log p_w(w) \end{aligned} \tag{4}$$

where (4) is obtained by making use of the summation formulas

$$\sum_{x=0}^{\infty} a^x = (1 - a)^{-1} \qquad |\,a\,| < 1 \tag{5}$$

$$\sum_{x=0}^{\infty} xa^x = a(1 - a)^{-2} \qquad |\,a\,| < 1. \tag{6}$$

Similarly, if we let $H_b$ denote the information content of the black run-length probability distribution, we obtain

$$\begin{aligned} H_b &= -\sum_{x=1}^{\infty} \text{prob}\,(R_b = x) \log \text{prob}\,(R_b = x) \\ &= -\log\,[[1 - p_b(b)]/p_b(b)] \\ &\qquad - [1 - p_b(b)]^{-1} \log p_b(b). \end{aligned} \tag{7}$$

At this point we note that one of the consequences of the assumption that $N$ approaches infinity is that the run-lengths are independent. If $N$ is finite then any set of $i$ run-lengths must satisfy the inequality that the sum of their lengths be less than or equal to $N$; hence these run-lengths are dependent. If we allow $N$ to approach infinity then any subset of $k$ ($k < i$) run-lengths will satisfy the inequality independently of the other $k - i$ run-lengths; *i.e.*, the set of $k$ run-lengths is independent of the set of $k - i$ run-lengths. Since $i$ and $k$ are arbitrary it follows that all of the run-lengths will be independent as $N$ approaches infinity. This fact will be used in our subsequent discussion.

Since all the black and white runs which constitute the picture have the same probability distribution, it is obvious that the total number of bits required by the run-length coding system is obtained by summing the information contents $H_w$ and $H_b$ over the entire picture. We denote the total number of runs by $n$; since $n$ is very large we can assume, without loss of generality, that $n$ is even so that there are $n/2$ white and $n/2$ black runs in the picture. Hence the total number of bits $N_R$ required to specify the Markoff process representation of the picture by run-length coding is

$$\begin{aligned} N_R &= (n/2)(H_w + H_b) \\ &= (n/2)(-\log\,((1 - p_w(w))/p_w(w)) \\ &\qquad - (1 - p_w(w))^{-1} \log p_w(w) \\ &\qquad - \log\,((1 - p_b(b))/p_b(b)) \\ &\qquad - (1 - p_b(b))^{-1} \log p_b(b)). \end{aligned} \tag{8}$$

We are now in a position to prove the following theorem.

*Theorem*: In the limit, as $N$ approaches infinity, the number of bits $N_R$ required to code the black and white runs in a first-order Markoff chain is equal to the theoretical lower bound $N_{\min}$.

*Proof*: From (1) and (8) we obtain

$$N_R/N_{\min} = \lim_{N\to\infty} (n/2N)((1 - p_w(w))^{-1} + (1 - p_b(b))^{-1}). \tag{9}$$

We now digress to evaluate the $\text{limit}_{N\to\infty}\,(n/2N)$. In order to do this we note that the sum of the lengths of the white and black run-lengths is equal to the number of cells in the picture. That is,

$$\sum_{i=1}^{n/2} (R_w^i + R_b^i) = N \tag{10}$$

where $R_w^i$ and $R_b^i$ are defined as the random variables which represent, respectively, the length of the $i$th white and $i$th black runs.

Eq. (10) can be rewritten as follows:

$$(2/n) \sum_{i=1}^{n/2} (R_w^i + R_b^i) = 2N/n. \tag{11}$$

However, $(2/n) \sum_{i=1}^{n/2} R_w^i$ is a sum of independent and identically distributed random variables; hence, by the Kolmogoroff Strong Law of Large Numbers[17] (as $N \to \infty$), this sum converges with probability one to the expected value of the white run lengths, $E(R_w)$ [if $E(R_w) < \infty$]. Similarly, $(2/n) \sum_{i=1}^{n/2} R_b^i$ converges to $E(R_b)$ [if $E(R_b) < \infty$], with probability one, as $N$ approaches infinity. Thus,

$$\lim_{N\to\infty} (2N/n) = E(R_w) + E(R_b) \tag{12}$$

[16] The probability distribution of $R_w$, as well as that of $R_b$, is known as a geometric distribution.[9]

[17] M. Loeve, "Probability Theory," D. Van Nostrand Co., Inc., Princeton, N. J.; 1955.

which, by (2) and (3) and the summation formula in (6), becomes

$$= [[1 - p_w(w)]^{-1} + [1 - p_b(b)]^{-1}]. \tag{13}$$

Substituting (13) into (9), we obtain

$$N_R/N_{\min} = 1$$

which proves the theorem.

The result stated by our theorem is not surprising, since both $N_R$ and $N_{\min}$ represent the number of bits required to specify $N$ cells in the first-order Markoff chain. The quantity $N_{\min}$ is calculated on the basis of individual cells and $N_R$ is calculated on the basis of groups of cells which become independent as $N$ approaches infinity. Thus $N_R$ and $N_{\min}$ must be the same as $N$ approaches infinity. The fact that the groups of cells are independent is important; this will be seen to be true when we consider a run-length coding system which uses the same code for black and white runs of the same length.

## Saving in Channel Capacity

We now turn our attention to the calculation of the saving in channel capacity $S$ of the run-length coding system with respect to the non-statistical coding scheme. It is apparent that $S$ is given by

$$S = N/N_R. \tag{14}$$

In view of (1) and the theorem which we have just proved, (14) becomes

$$S = N/N_R = N/N_{\min}$$
$$= (-p(b)p_b(b) \log p_b(b) - p(b)p_b(w) \log p_b(w)$$
$$- p(w)p_w(w) \log p_w(w) - p(w)p_w(b) \log p_w(b))^{-1}. \tag{15}$$

Since $N_{\min}$ is less than or equal to $N$ for all values of $p(w)$ and of transition probabilities, it follows that $S$ is always greater than or equal to unity, for all types of pictures. Thus, our probabilistic model predicts that the saving in channel capacity is always greater than or equal to unity for all possible pictures. For a given transmission facility the saving in channel capacity becomes a saving in time. For a given transmission time, the saving in channel capacity becomes a saving in bandwidth.

A graph of the saving is shown in Figs. 1(a) and 1(b), where for the sake of convenience the independent variables are chosen to be $p(w)$ and $p_w(w)$. It is easily found that for a fixed $p(w)$, the minimum value of $S$, denoted by $S_{\min}(p(w))$, occurs when $p_w(w) = p(w)$, and is given by

$$S_{\min}(p(w)) = (-p(w) \log p(w) - p(b) \log p(b))^{-1}. \tag{16}$$

The minimum possible value of $S_{\min}(p(w))$ occurs when $p(w) = \frac{1}{2}$, and is equal to one. Thus, the saving is one for pictures which contain equal black and white areas, and have a transition probability $p_w(w)$ of $\frac{1}{2}$. In such cases, there would be no point in using a statistical coding system, and we would no doubt use a nonstatistical coding system. However, such pictures very rarely occur in practice. In fact, such a picture would be the equivalent of a random pattern, in which there is no dependence among the various cells. Hence our probabilistic model predicts that for pictures which are equivalent to random patterns there is no point in using a run-length coding system, since there is very little to be gained.

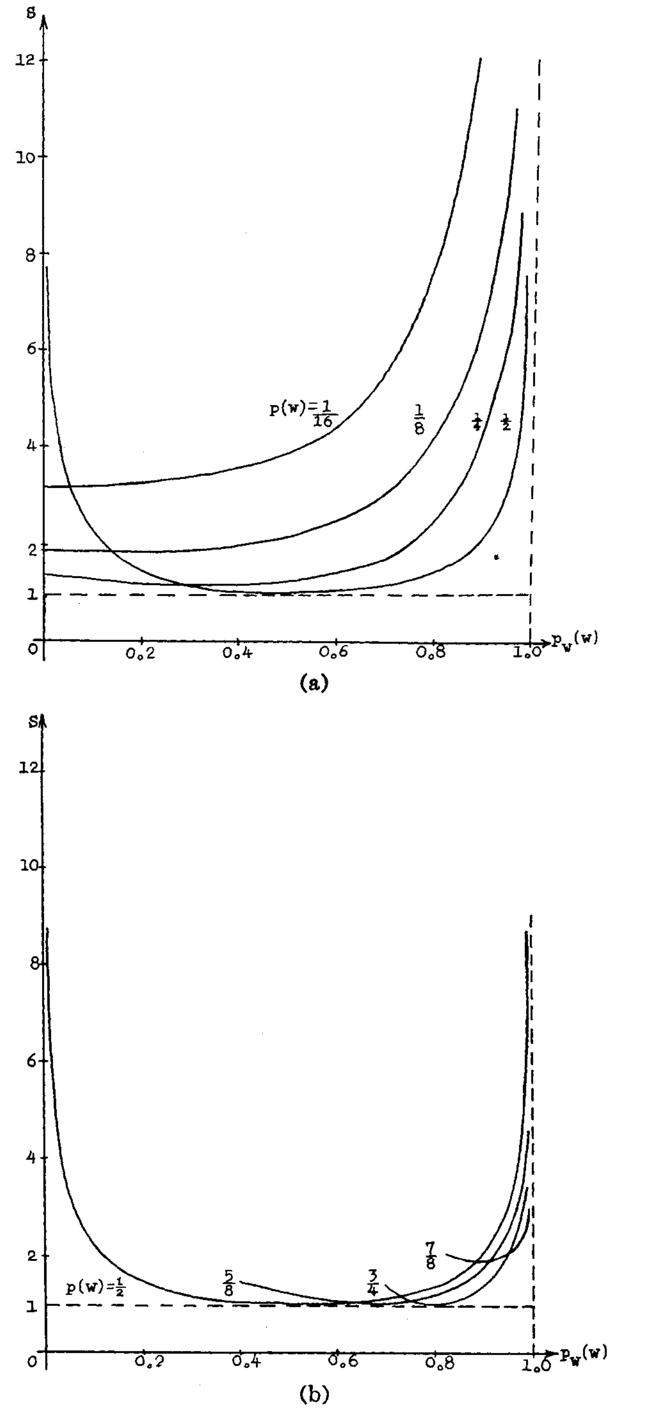


Fig. 1—(a) Saving in channel capacity vs transition probability $p_w(w)$, for various values of $p(w)$ between 0 and $\frac{1}{2}$. (b) Saving in channel capacity vs transition probability $p_w(w)$, for various values of $p(w)$ between $\frac{1}{2}$ and 1.

We turn our attention now to the question of when there are large savings. From Fig. 1(a) and 1(b) we observe that large savings are possible for the following three cases:

1) If $p_w(w)$ is close to unity, then the saving will be large, regardless of the value of $p(w)$. This corresponds to the situation where the white runs in the picture are very long. In this case a very large saving can be expected.

2) If $p(w)$ is close to zero, or one, then the saving is again very large, regardless of the value of $p_w(w)$. This situation corresponds to a picture which is almost completely black, or completely white, and again it is to be expected that a large saving is to be obtained by coding runs rather than using a nonstatistical transmission method.

3) If $p_w(w)$ is close to zero, and $p(w)$ is close to one-half, then the saving is again very large. The situation is that corresponding to a picture which is a checkerboard pattern. Thus, since it is known that all black and white runs in the picture will be of unit length, the transmitter need send no bits to specify the picture (except possibly for one bit to specify whether the first cell is black or white).

We observe that for given values of $p(w)$ the curves of Fig. 1(b) do not extend beyond certain critical values of $p_w(w)$. This is so, since values of $p_w(w)$ less than these critical values lead to impossible values for the other transition possibilities.

## Comparison with Deutsch's Results

It is of interest to compare the saving in channel capacity predicted by the probabilistic model with the value of saving found by previous investigators who have obtained that saving by actually measuring the probability distribution of the run-lengths in a picture. Towards this end, we compare our results with those of Deutsch,[5] who has measured the run-length probability distribution for the two cases of horizontally and vertically scanning a picture. From his results we obtain that

$$p(w) = 4107/5025,$$

and that the transition probability $p_w(w)$ in the vertical direction is

$$p_w(w) = 3702/4107 \qquad \text{(vertical direction)}.$$

Substituting these values in (15) we obtain

$$S = 1.78 \qquad \text{(vertical direction)}.$$

The corresponding value obtained by Deutsch is 2.02.

The transition probability in the horizontal direction is

$$p_w(w) = 3640/4107 \qquad \text{(horizontal direction)}.$$

The value of $p(w)$ is of course the same for both the vertical and horizontal directions. Substituting these values in (15), we obtain

$$S = 1.67 \qquad \text{(horizontal direction)}.$$

The corresponding value found by Deutsch is 2.14. Thus the saving in channel capacity predicted by our model is in error by approximately 12 per cent in the vertical direction and about 22 per cent in the horizontal direction. In both cases the probabilistic model predicts a saving which is less than the true value. In view of the rather close agreement found for the value of the saving we would feel that the probabilistic model is a good one for predicting the amount of saving possible for the run-length coding system. However, more results of this kind would certainly have to be obtained before any general statement to this effect could be made.

## The Use of the Same Code for Black and White Runs of the Same Length vs Different Code for Black and White Runs of the Same Length

It is certainly simpler to implement a run-length coding system which uses the same code for black and white runs of the same length than one which uses a different code for black and white runs of the same length. However, we have seen that our model predicts that this latter method achieves the lower bound of (1), as $N$ approaches infinity, and thus offers a greater saving than the former method. We now investigate how much more of a saving is possible with this latter method than with the former. In order to do this we must compute the information content of the probability distribution of a run-length which can be either white or black.

Let $R$ denote the random variable which represents the length of either black or white runs. Then, by reasoning which is similar to that used in obtaining (2) we find

$$\text{prob}\,(R = x) = p(w)p_w(w)^{x-1}p_w(b) + p(b)p_b(b)^{x-1}p_b(w) \qquad x = 1, 2, \cdots \qquad (N \to \infty). \tag{17}$$

Denoting the information content of this probability distribution by $H_S$, we find that

$$H_S = \sum_{x=1}^{\infty} (-p(w)p_w(w)^{x-1}p_w(b) - p(b)p_b(b)^{x-1}p_b(w)) \log \cdot (p(w)p_w(w)^{x-1}p_w(b) + p(b)p_b(b)^{x-1}p_b(w)). \tag{18}$$

If the number of bits required to code the picture by this method is denoted by $N_S$, then

$$N_S = nH_S. \tag{19}$$

It is easily seen that if $p(w) = p(b) = \frac{1}{2}$, then $N_S = N_{\min}$. Thus this method of coding achieves the lower bound of (1) for pictures which contain equal black and white areas. Hence for pictures which contain roughly the same number of black and white cells there is no advantage to be gained in coding black and white runs of the same length with different codes, instead of coding them with the same code.

If $p(w)$ [or $p(b)$] is close to zero, then it is easily seen that $H_S$ is approximately twice as large as $H_{\min}$. Thus in this case there is an advantage to be gained in using a different code for black and white runs of the same length, rather than using the same code.

A straightforward calculation shows that run-lengths in this coding scheme are no longer independent, except when $p(w) = p(b) = \frac{1}{2}$. This accounts for the fact that $N_S$ and $N_{\min}$ are different for those pictures which do not have the same number of black and white cells.

## Exponential Character of the Run-Length Probability Distribution

The probability distribution of the black and white runs [(2) and (3)] as predicted by our probabilistic model for pictures has an exponential character, as shown in Fig. 2. This result also bears some resemblance to the results for the probability distributions measured by other investigators. Michel[6] finds a probability distribution of runs which has a very strong exponential trend. Some of the results obtained by Deutsch[5] also indicate a strong exponential trend. However, some of his results do not indicate such a trend. In particular, the probability distribution that he finds for the white runs in the vertical direction does not indicate an exponential trend, and in fact exhibits no trends.

## Application of the Probabilistic Model to an "Elastic" System of Run-Length Coding

It will be recalled that the saving in channel capacity is calculated for a particular probability distribution of run-lengths. Thus, the saving pertains only to that set of pictures which has the aforementioned probability distribution. If a picture with a different probability distribution is coded by the run-length coding system, then it is possible that there will be no saving in channel capacity. That is, if the code is not "matched" to the probability distribution of run-lengths in a particular picture, there may not be a saving in channel capacity.[6] This is one of the disadvantages of the run-length coding system. One method of solving this problem would be to measure the run-length probability distribution of the picture before the picture is coded, in order to match the code to this distribution. Such a system, which changes its code for different pictures, is known as an "elastic" system. In practice, an elastic system which measures the run-length probability distributions directly is difficult to implement, since the measurement of run-length probability distributions is difficult to make.

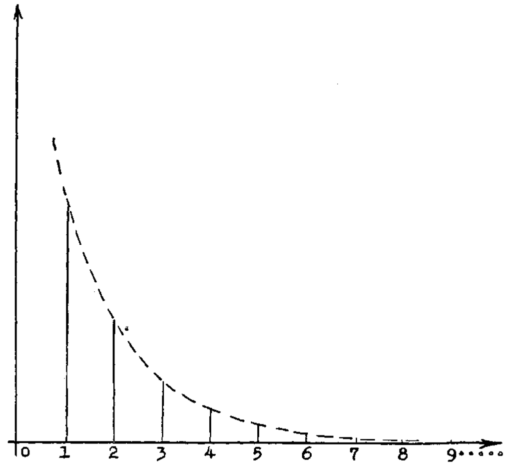


Fig. 2—Exponential character of the probability distribution of run-lengths.

As an alternative to the above method there is the possibility of computing the probability distribution of run-lengths on the basis of the probabilistic model described previously. The outstanding feature of this model is that it is described by parameters which are very simple to measure in practice. The transition probabilities which describe the model can be measured very easily by means of the optical correlator used by Kretzmer[18] and other investigators[19] in their studies concerning the measurements of the statistics of pictures. Once these transition probabilities have been measured, the elastic run-length coding system could match its code to the run-length probability distributions which are calculated from it. In this manner, the elastic run-length coding system could achieve a greater saving in channel capacity, on the average, than would be otherwise possible.

## Acknowledgment

The author would like to thank Professor Lotfi A. Zadeh and Dr. Ernest R. Kretzmer for many helpful discussions and criticisms concerning this material.

[18] E. R. Kretzmer, "Statistics of television signals," *Bell Sys. Tech. J.*, vol. 31, pp. 751–763; July, 1952.

[19] J. Capon, "Bounds to the Entropy of Television Signals," Res. Lab. of Electronics, M.I.T., Cambridge, Mass., Tech. Rept. No. 296; May 25, 1955.