around $\epsilon = 0$ where the probability density is greater in the former case $[1/(2\sigma_\epsilon^2)^{1/2}]$ than in the latter $[1/(2\pi\sigma_\epsilon^2)^{1/2}]$.

There are still problems remaining to be solved, such as transmission errors, examination of the validity of the visual model, simulation experiments with actual images, and reexamination of experimental results in view of this analysis. The most important of these is the effect of the transmission errors [6], [7]. These effects become larger if the prediction coefficient is brought nearer to 1.0 in order to obtain good picture quality.

ACKNOWLEDGMENT

The author would like to express his gratitude to H. Yoshine for his kind guidance.

REFERENCES

[1] J. B. O'Neal, Jr., "Predictive quantizing systems (differential pulse code modulation) for the transmission of television signals," *Bell Syst. Tech. J.*, pp. 689–721, May–June 1966.

[2] J. C. Candy and R. H. Bosworth, "Methods for designing differential quantizers based on subjective evaluations of edge busyness," *Bell Syst. Tech. J.*, vol. 51, pp. 1495–1516, Sept. 1972.

[3] W. F. Schreiber, "Picture coding," *Proc. IEEE*, vol. 55, pp. 320–330, Mar. 1967.

[4] J. Max, "Quantizing for minimum distortion," *IRE Trans. Inform. Theory*, vol. IT-6, pp. 7–12, Mar. 1960.

[5] P. F. Panter and W. Dite, "Quantization distortion in pulse-count modulation with nonuniform spacing of levels," *Proc. IRE*, vol. 39, pp. 44–48, Jan. 1951.

[6] K. Chang and R. W. Donaldson, "Analysis, optimization, and sensitivity of differential PCM systems operating on noisy communication channels," *IEEE Trans. Commun. Technol.*, vol. COM-20, pp. 338–350, June 1972.

[7] T. Fukinuki *et al.*, "Degradation of TV signals by transmission error in D-PCM" (in Japanese), in *Rec. Tech. Group Commun. Syst.*, Inst. Electron. Commun. Eng. Japan, CS70-113, Mar. 1971.

[8] T. Fukinuki, "Optimization of D-PCM for TV signals with consideration of visual properties" (in Japanese), in *Rec. Tech, Group Commun. Syst.*, Inst. Electron. Commun. Eng. Japan. CS72-138, Jan. 1973.

★

Takahiko Fukinuki was born in Osaka, Japan, on December 26, 1936. He received the B.E. and M.E. degrees in electronics engineering from Kyoto University, Kyoto, Japan, in 1959 and 1961, respectively.

Since 1961 he has been with the Central Research Laboratory, Hitachi, Ltd., Kokubunji, Tokyo, Japan, working on signal processing for digital communications and its applications. From 1969 to 1970 he was a Visiting Scientist at the Cognitive Information Processing Group, Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge. Currently he is a Senior Researcher of the Fifth Department, and a Leader of the Image Processing Group, Central Research Laboratory, Hitachi, Ltd.

Mr. Fukinuki is a member of the Institute of Electronics and Communication Engineers of Japan and the Institute of Television Engineers of Japan.

Optimum Run Length Codes

H. MEYR, HANS G. ROSDOLSKY, AND THOMAS S. HUANG, MEMBER, IEEE

***Abstract*—To realize the full redundancy reducing potential of run length coding over a collection of picture segments with varying run length statistics, adaptive coding techniques have been proposed. This paper compares results for a previously proposed *A*-code, for which the block length was varied adaptively, with those obtained using a fixed block length *B*-code first proposed here. Both codes use variable length codewords integrally related to the code block length. For all pictures analyzed, the fixed block length *B*-code performed nearly as well as the adaptive *A*-code. For both, the bit rates were close to the entropy bound. The reasons for these results are discussed. It is shown that, due to the prevalence of exceptionally long runs, the *A*-code, which is nearly optimal for exponentially distributed run lengths, performs poorly for actual pictures unless the block length is varied adaptively. The strategy used to implement the *A*-code adaptively is described. The optimal block length of the *B*-code, on the other hand, is shown to be largely independent of the picture statistics and need therefore not be varied adaptively. The hardware implementation of both coding techniques, the influence of channel disturbances on image quality, and the problems of error correction and line synchronization are discussed.**

Paper approved by the Associate Editor for Data Communications of the IEEE Communications Society for publication after presentation at the 1973 International Conference on Communications, Seattle, Wash. Manuscript received October 3, 1973.

H. Meyr is with the Research Division, Hasler Ltd., Berne, Switzerland.

H. G. Rosdolsky is with the Scientific Control Systems Ltd. Essen, West Germany.

T. S. Huang is with the Department of Electrical Engineering, Purdue University, Lafayette, Ind.

I. INTRODUCTION

THE convenience of picture transmission by facsimile over voice grade channels is severely limited by the transmission times required. Due to the strong statistical correlation between picture elements, transmission times may be reduced considerably by appropriate coding techniques. The correlation has both horizontal and vertical components and is difficult to exploit fully in transmission for which picture scanning is of a sequential one-dimensional nature. The logic and buffer space required for two-dimensional schemes such as contour coding [1] and algebraic picture transformation [2] are prohibitive in low cost systems. Run length coding [3], [4] utilizes only the horizontal component, the correlation between successive points on a line. Its use for two-tone pictures in low cost systems is almost dictated by the point by point, line by line, scanning mode commonly used.

All evidence indicates that successive run lengths are nearly independent. If successive characters generated by a source are statistically independent, an optimal source code may be constructed by Huffman's method [5]. The distinguishing feature of a Huffman code is that the codeword lengths are related to the frequencies of the source characters, the more frequent characters being assigned the shorter codewords. The size of the source alphabet consisting of all possible run lengths rules out the use of the pure Huffman code for run length coding. Fortunately, it turns out that there exist variable word length codes which perform nearly as well and which are easily implemented in hardware.

Here we investigate two classes of run length codes which we call the *A*-codes and the *B*-codes respectively. The *A*-codes have been described elsewhere by Huang [3], particular *A*-codes have been implemented by Schoenfelder *et al.* in Braunschweig [6]. The *B*-codes are to our knowledge first described here; we claim they are very nearly optimal for the run length distributions occuring in practice.

The codewords for both the *A* and the *B*-codes are multiples of fixed length blocks; particular *A* or *B*-codes are characterized by the block length. The *A*-codes are nearly optimal for exponentially distributed run lengths. For this distribution there exists a simple relation between the optimal block length and the mean run length. The block lengths 6 and 3 used by the Braunschweig group for white and black respectively are typical for the optimal block lengths for printed material.

The run length statistics of 30 pictures were evaluated by us and the bit rates for the *A* and *B*-codes were determined. The redundancy (defined as the ratio of redundant to essential information) of the optimal block length *A*-codes was still about 50 percent for all pictures analyzed. In an effort to reduce this figure, the pictures were encoded adaptively on a line to line basis. Each line was prescanned to determine the mean lengths of the white and black runs separately and was then coded using the corresponding optimal block length. The average redundancy for this adaptive *A*-code was 20 percent.

The experimental run length distribution was in marked disagreement with the exponential model. The experimental run length probabilities appear to decrease as a negative power of the run length—the longer runs are considerably more frequent than in the exponential model.

We demonstrate that the *B*-codes are nearly optimal for the experimental run length distributions. The bit rate turns out to be rather insensitive to the block length. The code B_1 with the block length 2, which is particularly easily implemented, appears to perform universally well for all pictures. The average redundancy obtained using B_1 for both white and black was 30 percent which compares favorably with the adaptive *A*-code.

This paper closes with a description of the facsimile system currently being developed by the Hasler research laboratories. A continuous data stream is guaranteed by a dual pair of buffers at both the transmitting and receiving ends. At any instant one buffer is connected to the scanning or printing unit and the other to the channel. At the start of a line the buffer roles are interchanged. This has an advantage. When the receiver is to print the line last transmitted, the previously printed line is still stored in one of the two buffers. In the case of a transmission error, the buffer roles are simply not interchanged with the result that the previous line is duplicated. Thus the vertical component of the correlation which cannot be reduced by run length coding, is used in a remarkably effective manner for error correction.

II. THE SOURCE

Description of the Source

A run is characterized by two parameters: its color and its length, i.e., by a pair of integers (C,L) where C is a code for the grey tone and L is the run length. The source is characterized by the ensemble of sequences

$$(C_1,L_1),\ (C_2,L_2),\ (C_3,L_3)\cdots \tag{1}$$

with the restriction that the grey tones of successive runs differ, i.e., that $C_{i+1} \neq C_i$.

If the length of a run depends only on the grey tone of the run, and if the grey tone is statistically dependent only on the grey tone of the previous run, it can be shown that the entropy of a typical sequence of n runs is

$$H(C_1,L_1,\cdots C_n,L_n) = \sum_{k=1}^{n} H(L_k \mid C_k) + \sum_{k=1}^{n} H(C_k \mid C_{k-1}). \tag{2}$$

This displays the entropy as the sum of the entropies associated with the run lengths and the grey tones separately. The terms in the first sum on the right are the entropies of the run lengths conditioned on the grey tones, the terms in the second are the entropies of the grey tones conditioned on the grey tones of the preceding runs.

For two-tone pictures, the value of C_k is known with certainty from that of C_{k-1}; the conditional entropies in the second sum vanish identically. The entropy of a sequence of n white and n black runs is

$$n[H(L \mid w) + H(L \mid b)]. \tag{3}$$

Here $H(L \mid w)$ and $H(L \mid b)$ are the entropies associated with run length distributions for white and black respectively. These entropies are related to the run length probabilities by the definition

$$H(L) = -\sum_{k=1}^{\infty} P(k)\log_2 P(k). \tag{4}$$

The mean number of picture elements contained in n white and n black runs is

$$n[E(L \mid w) + E(L \mid b)] \tag{5}$$

where $E(L \mid w)$ and $E(L \mid b)$ are the mean run lengths for black and white respectively. The entropy per picture element is the quotient of (3) and (5)

$$h(L) = \frac{n[H(L \mid w) + H(L \mid b)]}{n[E(L \mid w) + E(L \mid b)]} = \alpha \cdot h(L \mid w) + \beta \cdot h(L \mid b). \quad (6)$$

Here

$$h(L \mid w) = \frac{H(L \mid w)}{E(L \mid w)}; \qquad h(L \mid b) = \frac{H(L \mid b)}{E(L \mid b)} \quad (6a)$$

are the entropies per picture element for white and black respectively and

$$\alpha = \frac{E(L \mid w)}{E(L \mid w) + E(L \mid b)}; \qquad \beta = \frac{E(L \mid b)}{E(L \mid w) + E(L \mid b)} \quad (6b)$$

are the "whiteness" and "blackness" degrees of the picture.

By the well known "Source Coding Theorem" of Information Theory, (6) is the lower bound on the number of bits per picture element required by any code under the assumption that successive runs are independent. All experimental evidence indicates that successive runs are indeed practically independent. R. Arps has shown [9] that the entropy per pel is only about 10 percent lower if a statistical dependence of run length in printed matter is taken into account. In order to be able to compute the entropy, the run length distributions of the white and black runs must be known. If the transition probability for a color change is independent of the run length (Markov model) it can be shown that the run lengths are exponentially distributed [7]. The probability that a white (black) run length exceeds r is

$$P(L > r) = \exp[-r\gamma/E(L)]$$
$$\gamma = E(L) \cdot \log_e [1 - 1/E(L)]. \quad (7)$$

This distribution may also be shown to maximize the entropy per picture element for fixed $E(L)$. The entropy per pel for the distribution of (7) is

$$h(L) = \frac{1}{E(L)} \{E(L) \log_2 E(L) - [E(L) - 1] \log_2 [E(L) - 1]\}. \quad (8)$$

This is plotted (solid curve) as a function of $E(L)$ in Fig. 1.

Experimental Results

Thirty picture segments, digitalized and recorded on magnetic tape, were available for analysis at the Federal Institute of Technology in Zurich. The statistics of a sample of these picture segments are given in Table I. Only two pictures, a technical drawing (number 25) and a type-written letter (number 30), both originally in the A4 format, were representative of typical white–black facsimile pictures. (Note, however, the different resolutions used in the spatial digitalization of both pictures.) The remaining 28 pictures, samples of various printing and typewriter types and also several handwriting specimens, were segments of smaller high-contrast areas and contained relatively few larger areas of white.

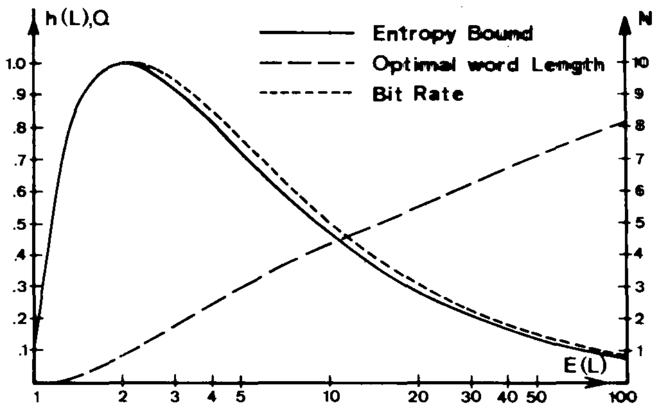


Fig. 1. Entropy $h(L)$ per pel, word length N of optimal A-code and the corresponding bit rate Q as a function of the mean run length $E(L)$ for an exponential distribution.

TABLE I
NUMBER OF WHITE (BLACK) RUNS $N_w(N_b)$, MEAN RUN LENGTH $E(L \mid w)$ AND $E(L \mid b)$, BLACKNESS α, ENTROPY PER PEL FOR WHITE (BLACK) $h(L \mid w)$ $(h(L \mid b)$

Pict.-Nr.	Lines	Points/Line	N_w	N_b	E(L|w)	E(L|b)	Black-ness	h(L|w)	h(L|b)	Remarks
1	80	235	1006	927	16.2	2.7	.132	.227	.868	type
2	100	340	1848	1748	15.2	3.4	.175	.262	.743	type
5	135	356	2133	1998	18.6	4.2	.172	.198	.544	type
10	70	190	665	585	17.9	2.7	.119	.216	.905	type
15	112	224	340	235	68.5	7.6	.072	.088	.424	signature
20	270	189	3256	2986	14.2	1.6	.091	.295	.876	type
25	1348	1060	14114	15460	95.3	5.4	.059	.062	.616	blueprint
30	746	854	11327	10582	53.8	2.6	.042	.089	.856	letter

The entropy per pel was computed under the assumption of independent runs using (6). The probabilities $P(L > r)$ for the white run lengths of pictures 2 and 30 are plotted in Fig. 2. In these figures the probabilities $P(L > r)$ for the exponential model with the same mean run length as for the experimental distributions are shown as dashed curves. For all 30 pictures analyzed there was marked disagreement between the exponential model and the experimental data. The run length probabilities $P(L > r)$ fall off less rapidly than predicted by the exponential model, except in the region of small run lengths where the experimental probabilities decrease faster.

III. EFFICIENT SOURCE CODING

Since successive runs are assumed to be independent, an optimal code could be constructed using Huffman's method [5]. The distinguishing feature of a Huffman code is that the codeword lengths are related to the frequencies of the source characters, the more frequent characters being assigned the shorter codewords. The size of the source alphabet consisting of all possible run lengths rules out the use of the pure Huffman code for run length coding. Fortunately, there exist variable word length codes which perform nearly as well and which are easily

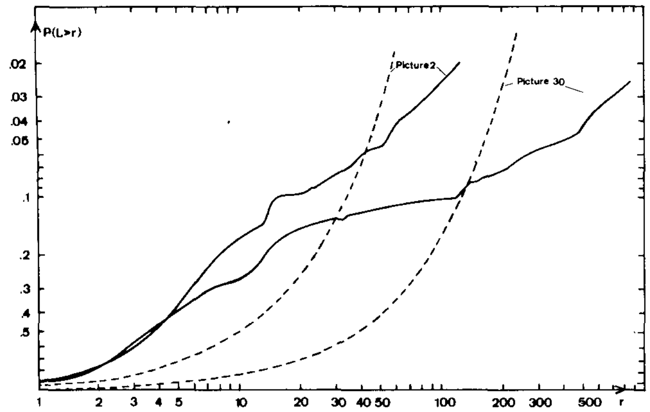


Fig. 2. Experimental run length probabilities $P(L > r)$ for white runs. Corresponding distribution for exponential model with the same mean run length.

implemented in hardware. In the sequel we study variable word length codes under the following constraints.

1) The code word lengths are multiples of a fixed length block of N-bits. The block length for white and black need not be the same.

2) The optimal block length can, if necessary, be adapted to the statistics of a given line (line-adaptive coding).

It is useful to introduce the concept of the *ideal source* of a given code. As is well known a code is optimal if the probability of each source character assigned a code word of n bits is $\frac{1}{2}^n$. Conversely for each code there exists a source (ideal source) for which the code is optimal, i.e., a codeword of n bits has a probability of $\frac{1}{2}^n$. Therefore, the ideal source of a near optimum code must be matched to the actual source as closely as possible. Since, by assumption, the run lengths are independent, this must be achieved for white and black runs independently. There is no possible trade-off between the two colors, e.g., a suboptimal code for one color cannot be compensated by a more efficient coding of the other color.

The A-Codes

The A-codes are a class of codes for which each run length is assigned a binary codeword consisting of one or more fixed length blocks. For A_N the blocks have N bits. The possible codeword lengths are multiples of this number (Fig. 3).

There are exactly

$$M = 2^N - 1 \tag{9}$$

positive integers whose binary representation requires N or fewer bits. The run lengths in the interval

$$0 < L \leq M$$

may be represented by a single block. Since there are no runs of length zero the block consisting entirely of zeroes is used to indicate overflow. In general, run lengths in the interval

$$nM < L \leq (n+1)M \tag{10}$$

are encoded using $n + 1$ blocks of which the first n consist entirely of zeroes and the last one contains the binary representation of the number $L - nM$. Decoding is done by summing the binary numbers represented by the individual blocks; the contribution of the blocks consisting entirely of zeroes is the same as that of a block consisting entirely of 1's, namely M. A codeword's end is marked by the first nonzero block.

A_3 Code

Run Length	Codeword
1	001
2	010
3	011
4	100
5	101
6	110
7	111
8	000001
9	000010
⋮	⋮
14	000111
15	000000001

Fig. 3. Coding example: the A_3-code.

To compute the bitrate Q, we consider the white and black runs from a typical picture separately. Each white (black) run length in the range (10) contributes $(n + 1)N$ bits to the coded text. The number of white (black) runs in this range is, by the law of large numbers, approximately

$$P[nM < L \leq (n+1)M]K = \{P[L > nM] - P[L > (n+1)M]\}K$$

where K is the total number of white (black) runs. The total number of bits in the coded text corresponding to white (black) runs is then

$$N_{\text{out}} = \sum_{n=0}^{\infty} (n+1)P[nM < L \leq (n+1)M]NK.$$

Transforming this by partial summation one obtains

$$N_{\text{out}} = NK \sum_{n=0}^{\infty} P(L > nM).$$

The number of picture elements is the product of the mean run length $E(L)$ and the number of runs K

$$N_{\text{in}} = E(L)K = K \sum_{n=0}^{\infty} P(L > n).$$

The *bit rate* Q, by definition the average number of bits per picture element, is the quotient

$$Q = \frac{N_{\text{out}}}{N_{\text{in}}} = N\Big[\sum_{n=0}^{\infty} P(L > nM) \Big/ \sum_{n=0}^{\infty} P(L > n)\Big]. \tag{11}$$

For exponentially distributed run lengths (7) these sums may be evaluated explicitly; the result being

$$Q = \frac{N}{E(L)} \cdot \frac{1}{1 - \exp\left[-\gamma M/E(L)\right]}$$

$$\gamma = -E(L) \log_e \left[1 - 1/E(L)\right]. \tag{12}$$

In Fig. 1 the N that minimizes (12) and the corresponding optimal bit rate are plotted as functions of the mean run length $E(L)$. The entropy per picture element for exponentially distributed run lengths is also plotted. It can be seen that the A-codes are well suited for exponentially distributed run lengths.

The dependence of the bit rate Q on the block length N is illustrated by Fig. 4 where (12) is plotted (solid curves) for $E(L) = 15.2$ and $E(L) = 53.8$. In both cases there exists a unique block length for which the bit rate Q is minimal.

The run length distribution of the ideal source for the A_N-code is given by

$$P(L > r = mM + k) = \frac{1}{2^{mN}}\left(1 - \frac{k}{2^N}\right)$$

$$m \geq 0; \quad 0 \leq k < M. \tag{13}$$

Evaluating the mean run length for this distribution and rearranging terms one obtains the expression

$$N = 1 + \log\left[E(L) - 1\right]/\log 2 \tag{14}$$

for the optimal block length as a function of the mean run length that was used by the authors in their adaptive A-code. For distribution (13), the entropy per picture element and the bit rate Q (12) are easily shown to be identical.

$$h(L) = Q = \frac{N}{E(L)} \cdot \frac{1}{1 - 2^{-N}}. \tag{15}$$

The optimal block length calculated from (14) and the bit rate Q (Fig. 5) are in remarkable agreement with the corresponding results for the exponential distribution plotted in Fig. 1.

Experimental Results for the A-Codes

The bit rates for the A-codes were calculated from the experimental run length distribution using the theoretical result (11). The bit rates thus calculated for the white runs for Pictures 2 and 30 are plotted (dashed curves) as functions of the code's block length in Fig. 4. The experimental bit rates are in good agreement with the corresponding values for the exponential model (solid curves) except in the region of the minimum where the experimental bit rates are larger by between 10 percent and 20 percent. Also shown in Fig. 4 is the entropy per picture element for the white runs calculated from the experimental run length distribution using the definition (4) (dashed horizontal line) and the corresponding result (solid line) for exponentially distributed run lengths. The experimental entropy is about 20 percent less than predicted by the exponential model. Due to both effects the experimental gap between optimal bit rate for the A-codes and the entropy bound is several times larger than in the exponential model. This fact indicates that the A-codes are not optimal for the run length distributions occurring in practice.

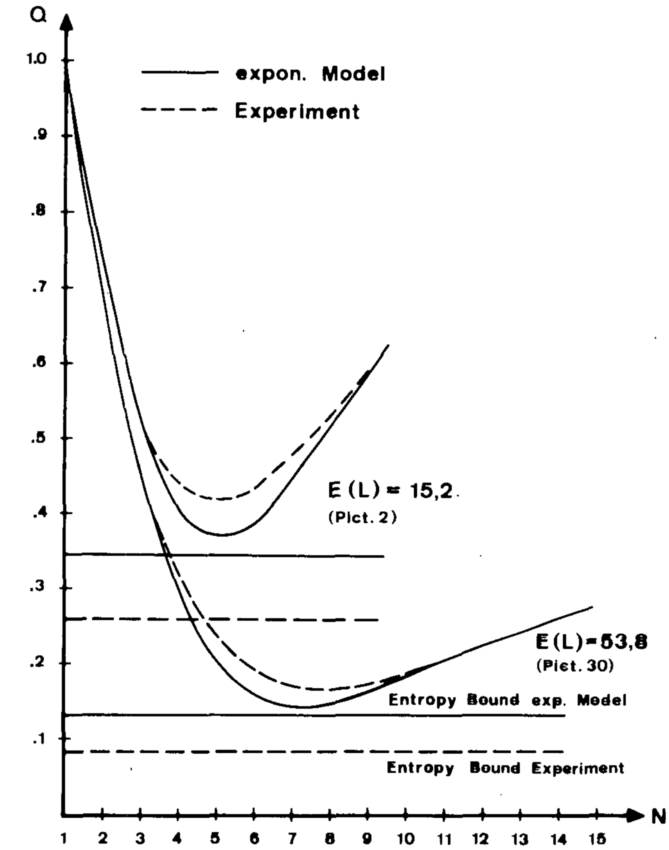


Fig. 4. Experimentally derived bit rate Q of A-codes for white runs (dashed curves) and corresponding curves for the exponential model with the same mean run length. Experimental entropy bounds and the corresponding values for the exponential model also indicated.

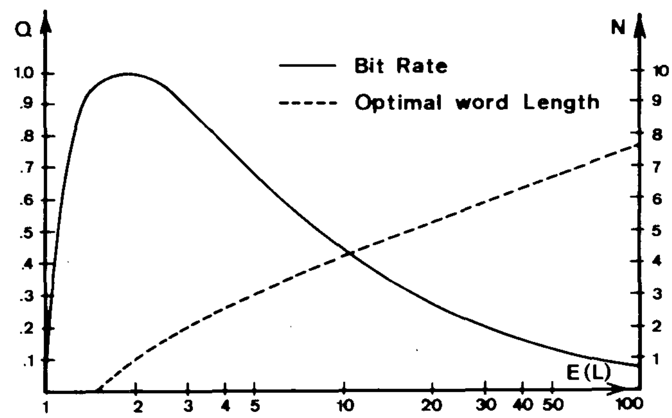


Fig. 5. Block length N of optimal A-code and the corresponding bit rate Q as functions of $E(L)$ for the distribution of the ideal source.

The optimal bit rates for the A-codes for white and black separately are tabulated for a sample of the pictures in Table II (columns NAW and NAB: optimal block lengths, columns QAW and QAB: the corresponding bit rates). The gap between optimal bit rate and entropy bound ($h(L \mid w)$ and $h(L \mid b)$ of Table I) is particularly marked for the white runs. Since 80 percent or more of the points in

TABLE II
OPTIMAL BLOCK LENGTH FOR WHITE AND BLACK RUNS AND CORRESPONDING BIT RATES FOR OPTIMAL A_N AND B_N-CODES

Pict.-No.	NAW	QAW	NAB	QAB	NBW	QBW	NBB	QBB
1	5	.413	2	.968	1	.271	0	1.000
2	5	.427	2	.817	1	.303	1	.975
5	6	.379	2	.783	1	.240	2	.890
10	5	.431	2	.995	1	.253	0	1.000
15	8	.117	3	.628	3	.124	1	.511
20	5	.445	1	1.000	1	.350	0	1.000
25	9	.102	3	.751	2	.071	1	.707
30	8	.168	2	.965	2	.098	0	1.000

TABLE III
REDUNDANCIES OF THE ORIGINAL PICTURES AND OF VARIOUS CODES APPLIED TO THESE PICTURES

Pct. No.	Original	Optimal A-Code	Adaptive A-Code	$B_{1,1}$	$B_{1,0}$
1	2.20	.56	.25	.20	.18
2	1.88	.43	.22	.21	.23
5	2.88	.74	.33	.45	.43
10	2.46	.61	.23	.17	.15
15	7.92	.78	.15	.46	.79
20	1.88	.43	.09	.31	.18
25	9.53	.47	.07	.20	.38
30	7.54	.72	.11	.21	.20

typical pictures are white (the picture "Blackness" is tabulated in Table I and the average blackness of all 30 pictures analyzed was 0.122) and since the entropy per point is less for white by a factor of roughly 4 (columns $h(L \mid w)$ and $h(L \mid b)$ of Table I; the entropies per picture element were 0.22 and 0.74 for white and black respectively) the bit rate for a picture as whole is dependent largely on the bit rate for white.

The redundancies of a sample of the original pictures and of various codes applied to these pictures are tabulated in Table III. The redundancy R is defined as the ratio of redundant to essential information

$$R = \frac{Q - h(L)}{h(L)}. \tag{16}$$

The redundancies of the same pictures coded using generally distinct optimal block length A-codes for black and white are tabulated in the second column of Table III. The average redundancy after coding of all 30 pictures was 0.46 (roughly 1 redundant bit in every 3); a redundancy reduction by a factor of about 7.

The A-code was also applied adaptively using a program written for this purpose. Each line is first scanned to determine the mean lengths of the black and white runs separately. The block length to be used for each color is then calculated using formula (14). The line is then coded using these block lengths. The bit rate of the picture as a whole is the average of the bit rate of the constituent lines.

The redundancies of the adaptive A-code are tabulated for some of the pictures in the third column of Table III. In all cases the adaptive A-code performed significantly better than the optimal fixed block length A-code. The average redundancy of all 30 pictures analyzed was 0.19 (1 redundant bit out of 6), an improvement by a factor of roughly 2.

As we will see, similar savings are possible using the fixed length block B-codes. This is due to the fact that the distribution of the ideal source of the B-code is a much better approximation to the actual distribution than the one of the A-codes.

The B-Codes

As can be seen from (10), the codeword length for the A-codes is roughly proportional to the run length. The A-code would be nearly optimal if the run lengths were exponentially distributed. The experimental data, however, show that the long runs are more prevalent than predicted by the exponential model (Fig. 2). The ratio of the codeword length to the run length should therefore be a decreasing function of the latter.

The B-codes are a class of codes for which the codeword length increases roughly as the logarithm of the run length. They are optimal if the information content per run satisfies

$$\log [1/P(L)] \approx \log L,$$

i.e., if, at least for large L

$$P(L) \approx (1/L)^{\alpha} \tag{17}$$

for some positive constant α. This distribution decreases less rapidly with L than the exponential distribution, i.e., large runs are more prevalent. The mean of this distribution in fact diverges.

As is the case for the A-codes, the codewords for the B-codes consist of one or more fixed length blocks, see Fig. 6. The block length for B_N is $N + 1$ bits. The first bit of each block is a "continuation bit" (an extension to grey tones by using more than one continuation bit is conceivable) and the following N are "information bits." For a codeword consisting of n blocks there are 2^{nN} possible combinations of the information bits. Within a codeword the

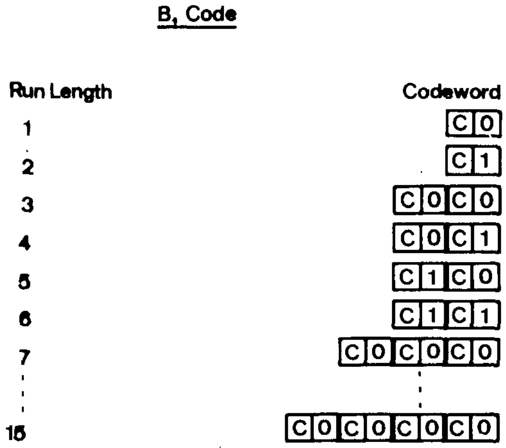


Fig. 6. Coding example: the B_1-code.

continuation bits are equal. The number of possible codewords consisting of n blocks is therefore 2^{nN}.

The 2^N run lengths in the range

$$1 \leq L < 1 + 2^N$$

are assigned codewords consisting of a single block; the 2^{2N} run lengths in the range

$$1 + 2^N \leq L < 1 + 2^N + 2^{2N} = \frac{2^{3N} - 1}{2^N - 1}$$

are assigned the codewords consisting of two blocks. In general the 2^{nN} run lengths in the range

$$\frac{2^{nN} - 1}{2^N - 1} \leq L < \frac{2^{nN} - 1}{2^N - 1} + 2^{nN} \qquad (18)$$

are assigned the codewords consisting of n blocks. The binary number z coded in the nN information bits of the codeword is

$$z = L - \frac{2^{nN} - 1}{2^N - 1}. \qquad (18a)$$

The code is uniquely decipherable; though not instantaneously (each codeword is the prefex of infinitely many others). If the continuation bit of a block is the same as that of the preceding block the block belongs to the same codeword as its predecessor; if not it is the start of a new codeword. (A new codeword is started at each color change.) The limiting B_0-code is equivalent to direct point by point transmission.

The implementation of B_1 which, as analysis of the experimental data has shown, is nearly always optimal, is particularly simple. In this case equations (18) can be written in the form

$$2^n \leq L + 1 < 2^{n+1} \qquad (19)$$

and

$$z = L + 1 - 2^n. \qquad (19a)$$

The first of these relations implies that the binary representation of $L + 1$ is of the form

$$L + 1 = 1 \cdot 2^n + a_{n-1} \cdot 2^{n-1} + \cdots + a_0, \qquad (20)$$

i.e., that the most significant 1 is in the 2^n's place. The second relation implies that the binary number coded in the n information bits is

$$z = a_{n-1} \cdot 2^{n-1} + a_{n-2} \cdot 2^{n-2} + \cdots a_0, \qquad (20a)$$

i.e., that the information bits are obtained by truncating the most significant 1 from the binary representation of $L + 1$. The coder, therefore, consists essentially of a counter which is preset to one at the beginning of each run. The n information bits are obtained by truncating the most significant bit of the state of the counter at the next color change. Decoding is similarly simple. The number of the consecutive continuation bits of the same value determines the position of the most significant bit of the binary number $L + 1$ while the n information bits determine the other bits of $L + 1$. The number $L + 1$ presets a down counter. Each counting pulse delivers then a picture element of the respective color. (Note that the continuation bits can also be used to determine the color of a run.)

An expression for the bit rate in terms of the run length probabilities may be derived in the same manner as the corresponding result, (11), for the A-codes. For B_N the bit rate is

$$Q = (N + 1) \cdot \frac{\sum_{n=1}^{\infty} P[L \geq (2^{nN} - 1)/(2^N - 1)]}{\sum_{n=0}^{\infty} P(L > n)}. \qquad (21)$$

The run length distribution of the ideal source for the B_N-codes is for r satisfying

$$r = \frac{2^{mN} - 1}{2^N - 1}$$

given by

$$P(L > r) = \frac{2}{[1 + r(2^N - 1)]^{1/N}}. \qquad (22)$$

In the intervals between these particular values of r, the distribution is linear in r. Distribution (22) is analogous to the exponential distribution for the A-codes. From (22) it follows that the asymptotic behavior of the pdf is

$$P(L = r) \approx (1/r)^{(1+1/N)}.$$

As already pointed out this implies that the average information per run increases as the logarithm of the run length. The probabilities (22) are plotted for several values of N in Fig. 8.

Experimental Results for the B-Codes

The optimal bit rates for the B-codes are tabulated in Table II (columns NBW and NBB represent the optimal number of information bits per block for white and black respectively; columns QBW and QBB the corresponding bit rates). In Fig. 7 the bit rates of B_N for white in Pic-

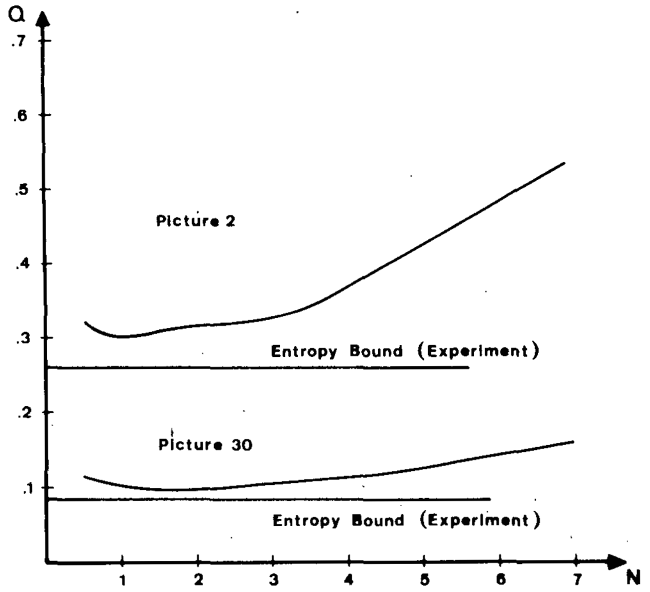


Fig. 7. Experimental bit rate for the B-codes for white runs. Experimental entropy bound indicated.

tures 2 and 30 are plotted against N. The bit rate for white, which as already pointed out is the decisive color, is nearly always better than for the corresponding optimal A-code. Comparing Fig. 7 and Fig. 4, one sees that the gap between bit rates and experimental entropy per picture element is now largely closed. The index of the optimal B_N-code is even less dependent on the mean run length than it was for the A-codes. Furthermore, the minima in Fig. 7 are flatter than for the A-codes; the penalty for using a suboptimal block length is smaller.

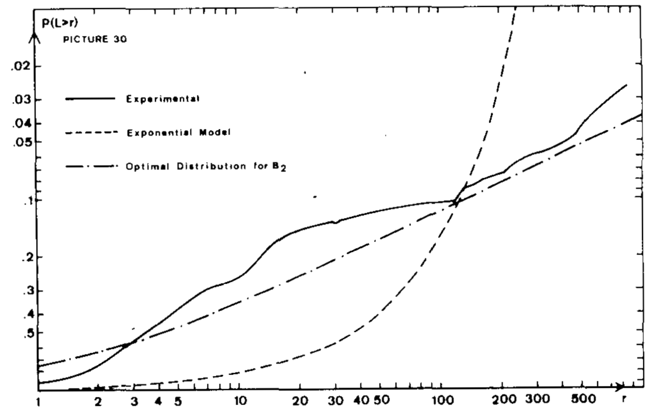
That the B_N-codes are well suited for the run length distributions occuring in practice can also be seen by inspection of Fig. 8. The distribution for the ideal source of the B_N-code is a very good approximation to the distribution of the actual source; this is not true for the ideal source of the A_N-code, shown as dashed curves.

For black, the optimal bit rates for the B-codes were less optimal than for the A-codes. For most pictures B_0 (direct transmission) was optimal, for the others B_1. The penalty for using B_1 for black when B_0 was optimal was small in all cases. Since B_1 was always optimal (cf. Picture 2) or very nearly optimal (cf. Picture 30) for white this implies that B_1 may be used for both white and black.

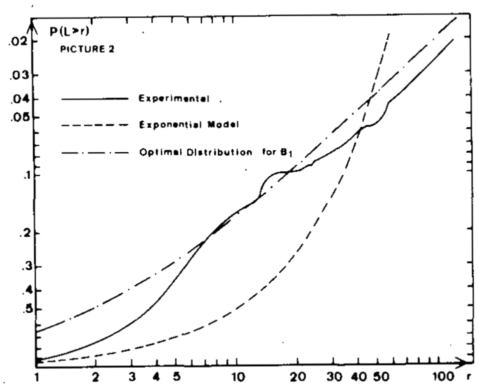
The redundancies of B_1 for both black and white and, by way of comparison, of B_1 for white and B_0 for black are tabulated, for a sample of the pictures analyzed, in the fourth and fifth columns of Table III. The average redundancies for all 30 pictures were 0.30 and 0.31, respectively. This is better than the optimal block length A-code but not quite as good as the adaptive A-code.

IV. INSTRUMENTATION CONSIDERATIONS

The hardware implementation of the Hasler system is illustrated schematically in Fig. 9. The transmitting and receiving units consist of four main blocks: buffer memories (two at each end); control units; coder/decoder; and modulator/demodulator. Each of the two buffers at either



(a)



(b)

Fig. 8. (a) and (b) Run length probabilities $P(L > r)$ for white runs and optimal distributions for B_1 and B_2. Corresponding distribution for the exponential model with experimentally derived $E(L)$ shown dashed.

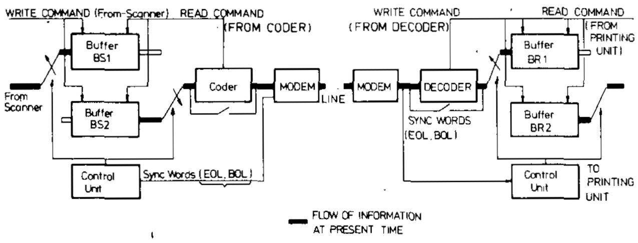


Fig. 9. The Hasler facsimile system: a block diagram.

end can store a full (uncoded) line of 1000 picture points.

The two buffers at each terminal are alternatively switched to the scanning/printing units and the encoding/decoding units. While the current line is written from the scanning unit into one of the two buffers at the transmitting end, the previous line that was stored in the other is being encoded and transmitted. When this is over the buffer roles are exchanged. Similarly, while the printing unit is reading a line from one of the memories at the receiving end, the next line is being decoded and stored in the other memory. The interchange of memory roles at the receiving end is regulated from the transmitter by the transmission of special synchronization codewords, the BOL ("Beginning of Line") and the EOL ("End of Line").

The use of two memories has the advantage that with this relatively simple scheme the channel bit rate may be utilized fully. Given a sufficiently fast channel the system speed is limited only by the speed of the scanning and printing units.

Given sufficiently rapid memory access times and sufficiently rapid logic elements, any form of run length coding may be implemented. Before being coded and transmitted, the line is stored as a whole in one of the transmitter buffers and is therefore available for prescans if exotic coding schemes such as the adaptive A-code are to be used. Conversely, the coding/decoding units may be bypassed altogether ("short circuited") if so desired. The coder used in the Hasler system generates the B_1 run length code for both white and black runs.

V. LINE SYNCHRONIZATION AND ERROR CORRECTION

The penalty for efficient source coding is increased sensitivity to transmission errors. For run length coding, a single transmission error results in a shortening or lengthening of a run length and displacement to the left or the right of all subsequent points in the line. In their erroneous positions the runs contribute nothing to a two-dimensional image and the resulting effects, such as a black run on a white background, can be very disturbing. It turns out to be better not to print out such lines.

In the Hasler system a transmission error generally affects only a single line. The line beginnings and ends are marked by independently decoded syncwords, the BOL ("Beginning of Line") and the EOL ("End of Line"). The bit sequence used for the EOL cannot occur in the output of the run length encoder (it is chosen to correspond to a run length larger than the line length) and may be recognized anywhere within the sequence of bits received.

Since the sum of the run length between a BOL and the following EOL must equal the line length, the independent decoding procedure for the EOL allows for error detection and partial correction. If the run length sum differs from the line length when the EOL is detected, the line received is known to be wrong. In this case the previous line is duplicated.

This strategy turns out to be remarkably effective even at extremely high error rates. This is illustrated by Fig. 10(a)–(c) which show a correctly transmitted original, the same picture incorrectly transmitted (one line in four on the average in error), and the picture resulting from replacing the incorrect lines by their predecessors.

The line error rate is closely related to the block error rate discussed by Balkovic *et al.* [8]. Assuming a bit rate of 0.5, the block length corresponding to the average line is 500 bits. According to the above reference, the error probability for this block length is 10^{-3} or less for 90 percent of all calls at the typical channel bit rate of 2000 bits/s. The fact that errors tend to occur in bursts is advantageous since the several errors in a single burst generally affect only a single line.

Twenty bits are required for an unambiguous EOL and a smaller number for the BOL. Line errors resulting from the incorrect transmission of these syncwords are far less probable (by a factor of roughly 10) than those resulting from the incorrect transmission of a run length. An incorrect BOL results in a single incorrect line. An incorrect EOL is not recognized and is interpreted by as part of the run length code. The result is that one line is skipped and the next is interpreted by the hardware as being incorrect and is replaced by the correct line.

The strategy of duplicating the last correct line is easily realized for the two buffer organization discussed above. When a line has been transmitted and stored in one of the receiver's buffers the previous line is still available in the other. If the new line is in error the buffer roles are simply not interchanged as usual.

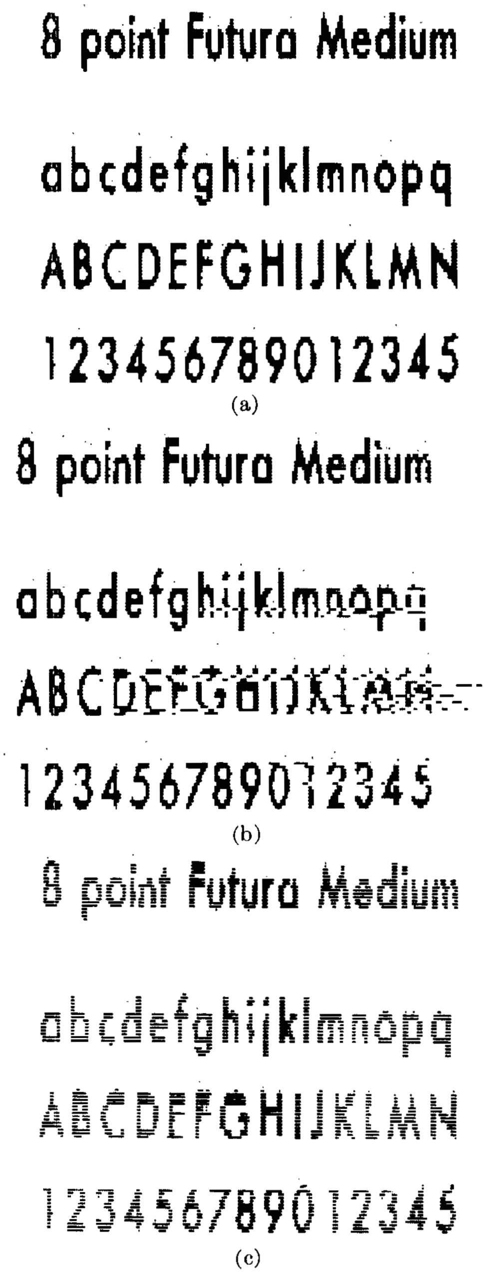


Fig. 10. (a) Correctly transmitted picture. (b) Incorrectly transmitted picture. Probability of a line in error is 0.25. (c) Incorrectly transmitted picture with same line error rate as Fig. 10(b). Line in error replaced by its correctly transmitted predecessor.

ACKNOWLEDGMENT

The authors wish to express their gratitude to Mr. H. Koller of the Swiss Federal Institute of Technology who kindly permitted us to use the pictoral data on which this study is based. T. S. Huang would also like to thank Dr. H. Mey of the Hasler Research Division and Prof. Baumann for making his visit at the Federal Institute of Technology possible. We also thank Dr. E. Hafner, Hasler Research Lab., for review of the paper and helpful discussions.

REFERENCES

[1] W. F. Schreiber, T. S. Huang, and O. J. Tretiak, "Contour coding of images," presented at WESCON, Session 8, 1968.
[2] H. G. Musman, "Ueber lineare Transformationen zur Redundanz-Reduktion," *Nachrichtentech. Fachber.*, vol. 40, pp. 13–27, 1971.
[3] T. S. Huang, "Run-length coding and its extensions," in *Proc Symp. Picture Bandwidth Compression*, Mass. Inst. Tech., Cambridge, Apr. 1969.
[4] D. Preuss, "Redundanzreduzierende Codierung von Faksimile-signalen," *Nachrichtentech. Z.*, vol. 24, pp. 564–568, 1971.
[5] D. A. Huffman, "A method for the construction of minimum-redundancy codes," *Proc. IRE*, vol. 40, pp. 1098–1101, Sept. 1952.
[6] H. Schönfelder, D. Preuss, W. Schlink, and H. Wendt, "Experimentalvorführung zur Quellencodierung," *Nachrichtentech. Fachber.*, vol. 40, pp. 56–71, 1971.
[7] J. Capon, "A probabilistic model for run-length coding of pictures," *IRE Trans. Inform. Theory*, vol. IT-5, pp. 157–163, Dec. 1959.
[8] M. D. Balkovic *et al.*, "Highspeed voiceband data transmission performance on the switched telecommunication networks," *Bell Syst. Tech. J.*, vol. 50, pp. 1349–1385, 1972.
[9] R. Arps, "The statistical dependence of run lengths in printed matter," *Nachrichtentech. Fachber.*, vol. 40, pp. 218–226, 1971.

★

H. Meyr received the Dip. Ing. and Ph.D. degrees from the Swiss Federal Institute of Technology, Zurich, Switzerland, in 1967 and 1973, respectively.

In 1968 he joined the Brown Boveri Corp., Zurich, Switzerland. He was a R+D Engineer involved in the simulation of large scale power control systems on a analog computer. In 1969 he joined the Swiss Federal Institute for Reactor Research as a Research Assistant involved in theoretical studies in the field of correlation analysis applied to neutron time-of-flight experiments. This work pertained to his graduate studies in the field of digital systems and statistical communication theory. Since 1970 he has been a Research Engineer with Hasler AG, Berne, Switzerland, working in the fields of digital facsimile encoding and correlation measurement techniques. Presently, he is a Visiting Assistant Professor with the Department of Electrical Engineering, University of Southern California, Los Angeles.

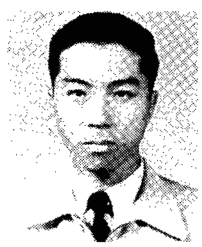
Dr. Meyr has several patents and has published several papers.

★

Hans G. Rosdolsky was born in Vienna, Austria, on January 17, 1943. He received the B.S. and M.S. degrees in mathematics and physics from the University of Michigan, Ann Arbor, in 1963 and 1964, respectively. He received the Ph.D. degree in theoretical high energy physics from the University of Michigan in 1968.

From 1968 to 1970 he was a Research Associate at the University of Oregon, Eugene, working in the field of elementary particle physics. In 1970 he joined Hasler AG, Berne, Switzerland, where he worked as a Programmer participating in the development of a fully automated telex exchange. Also, he did theoretical work on traffic signal control and picture coding with the firms research division. In 1973 he joined the DATUM research firm, Bonn, West Germany, where he was engaged in adapting the TRIPS transportation planning program package to the Siemens 4004 computer used by the German Ministry of Transport. Presently, he is with the Scientific Control Systems, Ltd., Essen, West Germany, where he is working in the field of industrial process control.

★



Thomas S. Huang (S'61–M'63) received the B.S. degree in electrical communication from the National Taiwan University, Taiwan, and the M.S. and Sc.D. degrees in electrical engineering from the Massachusetts Institute of Technology, Cambridge.

From 1963 to 1973, he was on the faculty of the Department of Electrical Engineering, M.I.T. During the academic year 1971–72, he was on sabbatical leave visiting ETH-Zurich, Switzerland, as a Guggenheim Fellow. During the academic year 1972–73, he was again on leave, working at the M.I.T. Lincoln Laboratory. In 1973, he joined Purdue University, Lafayette, Ind., where he is at present a Professor of Electrical Engineering.

Dr. Huang's professional interest lies in the broad area of information and communication technology but especially the transmission and processing of multidimensional signals. He has served as a consultant to numerous industrial firms and government agencies. He is coauthor (with R. R. Parker) of the book *Network Theory: An Introductory Course* (Addison-Wesley) and coeditor (with O. J. Tretiak) of the book *Picture Bandwidth Compression* (Gordon and Breach). He is an editor of the *International Journal Computer Graphic and Image Processing*, and an associate editor of *Pattern Recognition*.