EUROPEAN ORGANIZATION FOR NUCLEAR RESEARCH

THE ENTROPY OF DELTA-CODED SPEECH

B.G. Taylor

(to be published in Proceedings IEE)

EP Division

Geneva,
March 1976

THE ENTROPY OF DELTA-CODED SPEECH

B.G. Taylor, B.Sc., Ph.D., C.Eng., M.I.E.E.

Indexing terms: Delta modulation, Entropy, Speech transmission

Abstract

The paper presents an analysis of the information properties of delta-coded speech with channel encoding applications. Predictive coding techniques for reducing the entropy of the average distribution of the signal elements are described, and evaluated in comparison with computations of the relative entropy of Markov process approximations to the message-generating process for orders up to 9. The optimal digital fixed-structure predictors are established for orders up to 7, and also the optimal group codes for block lengths of 2 to 6 elements.

The redundancy is shown to be typically about one half, and a predictor success probability of 0.9 is attainable with a practical 6th-order discrete structure. The entropies of the sequences generated by modulo-2 addition of the predictions and source elements are found to be much closer to the process entropies than the corresponding performance characteristics for group encoding.

For encoding of the predictor error sequence to achieve message compression, 5-element group encoding, attaining a compression factor of 0.5 with a 300-word buffer, is found to be superior to run-length encodings. The combination transformation is much more efficient than direct exact coding of blocks of source elements.

List of principal symbols

B_i = element block

Δ = quantization interval

E = predictor element

f_c = channel data rate per source

f_s = sampling frequency

F_N = conditional entropy

G_N = group entropy

H = process entropy

H_g = geometric distribution entropy

H_{gr} = geometric distribution relative entropy

H_r = relative entropy for run-length encoding

H_T = mod-2 adder sequence entropy

η = transmitter power saving

L_g = geometric distribution sequence length

N = Markov process order

N_a = average codeword length

N_b = average queue length

N_c = 5-element error sequence group encoding word length

N_i = codeword length

N_m = number of sources multiplexed

N_q = buffer capacity

P_b = element error probability

P_{gr} = geometric distribution run-length probability

P_{ov} = buffer overflow probability

P_r = run probability

P_s = predictor success probability

r = run-length

R_B = group encoder compression factor

R_M = upper bound for group encoding

S = source element

S_i = source state vector

T = transformed element

TG_N = error sequence group entropy

TH = error sequence entropy

TH_T = error sequence mod-2 adder output entropy

TR_B = error sequence group encoding compression factor

TR_M = upper bound for error sequence group encoding

Abbreviations

ΔM = delta modulation

f.m. = frequency modulation

p.c.m. = pulse-code modulation

s.n.r. = signal-to-noise ratio

t.d.m. = time-division multiplex

1 Introduction

Delta modulation (ΔM)[1-4] is an attractively economical source-encoding technique for digital speech communication which can be readily implemented[5,6] by integrated semiconductor technology. For high transmitted speech quality, however, the channel capacity requirements of the lowest cost (single integration, non-adaptive) basic configuration are greater than for encoding methods such as logarithmic companded pulse-code modulation (p.c.m.)[7]. With a sufficiently high sampling frequency f_s, and appropriate choice of quantizing interval Δ, the basic delta modulator is capable of any desired frequency response, dynamic range and signal-to-quantizing-noise ratio (s.n.r.); but since ΔM channel bit-rate and sampling frequency are equal, these performance characteristics are bought at the expense of digital transmission system bandwidth. De Jager[8] and Abate[2] have shown that ΔM s.n.r. improves about 9 dB per octave increase of f_s, which compares unfavourably with the exponential relation between s.n.r. and channel bit-rate for p.c.m.

Since the initial conception of ΔM by Deloraine et al[9], an extensive theoretical and experimental research effort has therefore been applied to the synthesis of more elaborate variants of the basic encoder. Adaptive quantizing[2,10,11], overload prediction[12], and syllabic companding by envelope detection[13,14] or digital control[15,16] are among the numerous modification techniques which have been studied to attain performance improvements at the expense of increased encoder complexity rather than higher sampling frequency.

While communication networks of a local nature are often integrated with expensive long-distance circuits such as satellite relays, troposcatter links or submarine cables, the traffic volume carried by these links is normally only a small proportion of the total originated. Furthermore, the transmission media which it is anticipated will be available to future overland communication systems are characterized by wide bandwidth capability at much lower cost than those of today. H_{01} mode circular waveguide, for example, can provide a channel bandwidth of 40 GHz[17], while the inherent capacity of a single laser beam exceeds that required for the total of present day telephone, radio and television world communications.

As a result, it may not prove cost-effective to equip every message source with a complex encoder whose greater efficiency is justified only during infrequent communication events involving high-cost channel capacity. As an alternative approach, minimum-cost ΔM may be used for the individual source encoders where these are required in large numbers, while bandwidth-compressing channel encoders can be associated with the terminal equipment which processes signals for transmission over the expensive major links.

It is possible to filter a ΔM signal and re-encode the message approximant by p.c.m., and Goodman[18] has shown that this can result in a reduction in the channel bit-rate. The different quantizing and overload characteristics of the two techniques result in an s.n.r. degradation which can be made small if sufficient samples are filtered[19], but which would be cumulative if the transcoding were repeated several times over a multi-circuit communications route.

The communications system of a large naval vessel is another example of a situation in which a local network, within which the cost and reliability of the source encoders are more important factors than the line bit-rate, is to be interfaced to a channel for which there is greater emphasis on efficiency of bandwidth utilization. Here, a local line network may be used to interconnect the many on-board crew stations using very reliable simple delta-coders, while it is required on occasions to route communications from any of them over a more bandwidth-sensitive radio link to supporting ships and aircraft.

This paper presents an analysis of the information properties of the signal sequences generated by ΔM speech with a derivation of information preserving (IP) transformations[20] (exact coding schemes) suitable for channel encoder applications. These transformations are such that the entropies of the ΔM operand sequence and its transform are the same, but the output transform sequences, from which the original ΔM signals can be reconstructed exactly by an inverse operation in the channel decoders, are generated at a lower bit-rate and hence require less channel capacity for transmission.

1.1 ΔM signal processor

The magnitude of the compression attainable by an IP transformation in such a channel encoder is determined by the extent to which the source encoder sequences exhibit redundancy as a result of the nature of the encoding scheme and constraints on the properties of the message-generating process. Because of the difficulty of modelling speech, either sine-wave excitation or stationary random (usually Gaussian) signal sources of specified rational power spectral density have been considered extensively in analytic, experimental and computer simulation studies of ΔM, such as those of Van de Weg[21], Zetterberg[22], Halijak and Tripp[23], Abate[2], Goodman[24], Aaron et al[25], Iwersen[26], Gersho[27], Slepian[28] and Greenstein[29].

To study the information properties of the signal sequences generated by a delta modulator when the exciting waveform is an actual speech message, we have implemented a hybrid signal processor with dual analog computers interfaced to a digital minicomputer. An f.m. magnetic tape system is used to scale message spectra appropriately for the analog computers by frequency division by a factor of 64. The test source message[30], which has been subjected to detailed study for speech synthesizer applications, is low-pass filtered by a 7th-order Butterworth active filter with a cut-off frequency of 3.85 kHz. The message corresponds to a ΔM sequence of about 4×10^6 bits.

A precision delta modulator with a resolution of 1 part in 10^5 of the maximum echelon signal range is implemented with an integrator and error amplifier of an analog computer, outboard differential comparator and accurate increment/decrement pulse generators. Thus the analyses are not biased by the implementation-dependent hysteresis, integrator drift or asymmetric idle channel noise[31] properties of more practical delta modulators.

The digital computer interface transfers sequences of ΔM code to a cyclic data buffer by direct memory access and synchronises program execution with the ΔM sampling frequency by interrupts. The results of the on-line analyses are recorded in a format which allows them to be read by Fortran programs for subsequent off-line analysis.

In the discussion which follows, the constituent 0 or 1 symbols of the binary delta-coder output are termed the elements of the message, while the state i of the source at any time is defined by a multi-dimensional vector whose components are a number of the symbols which have been generated prior to that time. Nth-order Markov process approximations to the source are studied for which an element is considered to have a statistical dependence limited to the preceding N symbols.

2 Relative entropy

Three entropy measures* are of interest. There is first the entropy of the process

$$H = -\sum_{i}\sum_{j} p(i)\ p(j|i) \log p(j|i) \tag{1}$$

in which $p(j|i)$ is the conditional probability of occurrence of state j following state i.

Second, there is the entropy of blocks B_i of N elements of the signal sequence,

$$G_N = -\frac{1}{N}\sum_{i} p(B_i) \log p(B_i) \tag{2}$$

And third, for an Nth-order approximation to the source in which the N-element block B_i determines the conditional distribution from which the next symbol S_j is drawn, the entropy

$$F_N = -\sum_{i}\sum_{j} p(B_i,S_j) \log p(S_j|B_i) \tag{3}$$

By adoption of the finite state ($1 \leq i \leq 2^N$) representation of the signal source, F_N may be equated with the process entropy H and evaluated by (1) which involves the conditional rather than the joint probabilities. For this analysis, the on-line software generates a set of 2^N square state transition probability matrices for N = 0 to 9, together with listings of the occurrence probabilities of the appropriately masked operand states. As the channel capacity required for direct transmission of a ΔM message

---

*Apart from explicit exceptions, entropy measures are per element of the signal sequence.

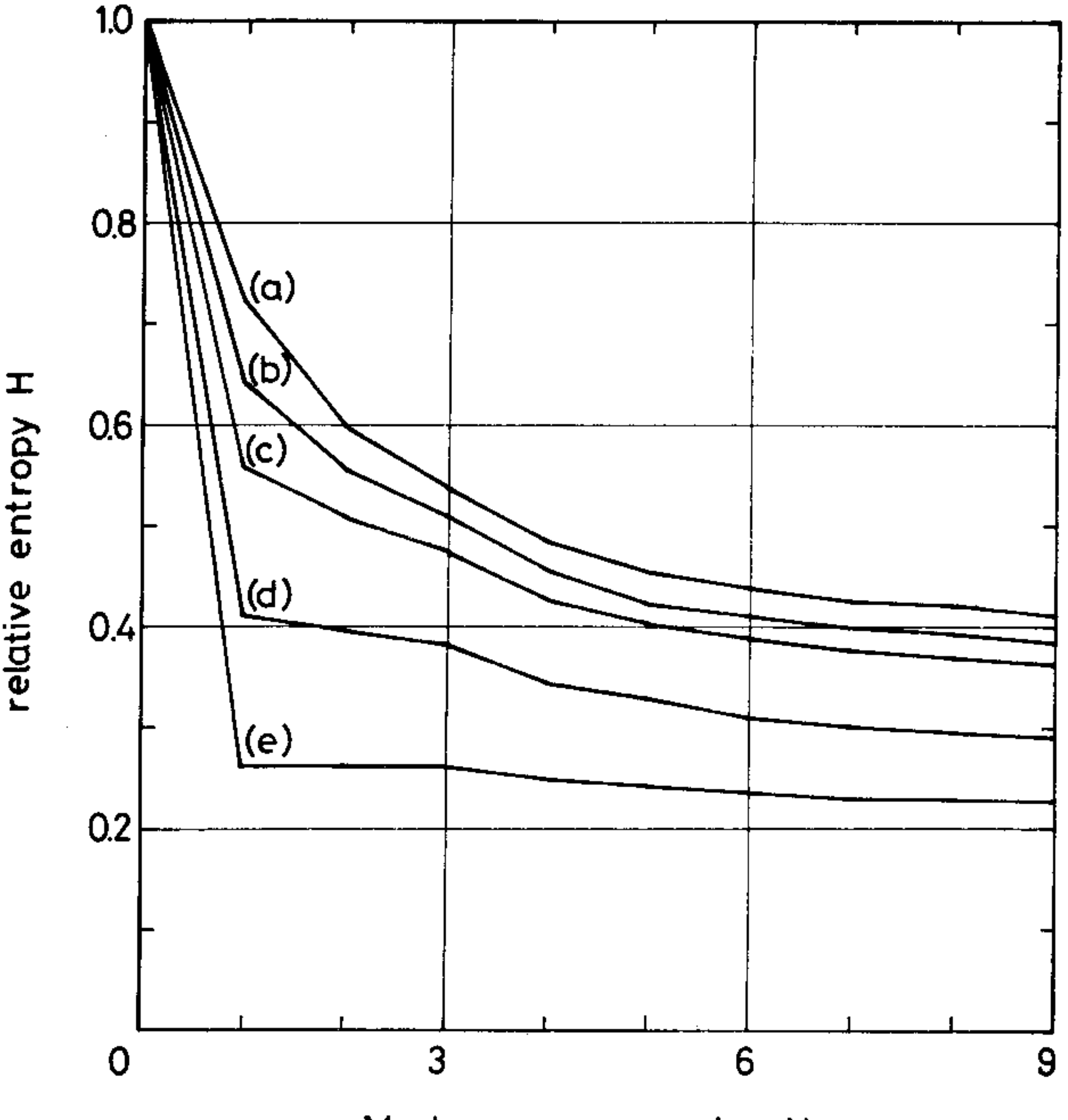


Fig. 1

Relative entropy characteristics for ΔM speech

Sampling frequency f_s = 48 kHz
Pk-pk message amplitude (a) 70Δ (b) 56Δ (c) 42Δ (d) 28Δ (e) 14Δ

is equal to the average element entropy†, relative entropies are referred to the $H = 1$ bit per element value for a source for which the S_j are equiprobable and independent of preceding elements. Fig. 1 presents the results of five test runs with $f_s = 48$ kHz and values of Δ increment ranging from coarse quantization to that causing frequent slope overloading. We find that in all cases, the redundancy of the ΔM message exceeds one half for $N \geq 4$. The characteristics define the lower bounds of the channel capacity required for transmission of the ΔM signal after IP transformation by an encoder processing N contiguous elements only.

Information properties found to be subject to wide variation with choice of ΔM design parameters would entail elaborate presentation and be of limited application. Fortunately the relative entropy characteristics prove to be substantially independent of changes in sampling frequency and Δ increment provided their choice is appropriate (ie. they are altered in inverse proportion). This invariance is shown by the constant overload probability groups of Fig. 2, which contain members for $f_s = 32$, 48 and 96 kHz, and comparison of the corresponding transition probability matrices shows that this property extends to the individual state transitions also. In general, parameter values $f_s = 96$ kHz, peak-peak message amplitude 128Δ, are considered, these being typical for good communications quality with infrequent overloading. Agreement between repeat test runs is better than 0.1%, and adequate test message duration is indicated by differences of less than 0.5% between results for the complete passage and its halves processed separately.

3 ΔM Predictive coding

Procedures for redundancy reduction by predictive coding[32] lower the average element entropy by a transformation utilizing such information

---

† 'Average element entropy' is used concisely throughout with the meaning 'entropy of the average distribution of the elements' and should be distinguished from the average of the entropies of the distributions of the elements, which is the entropy of the process.

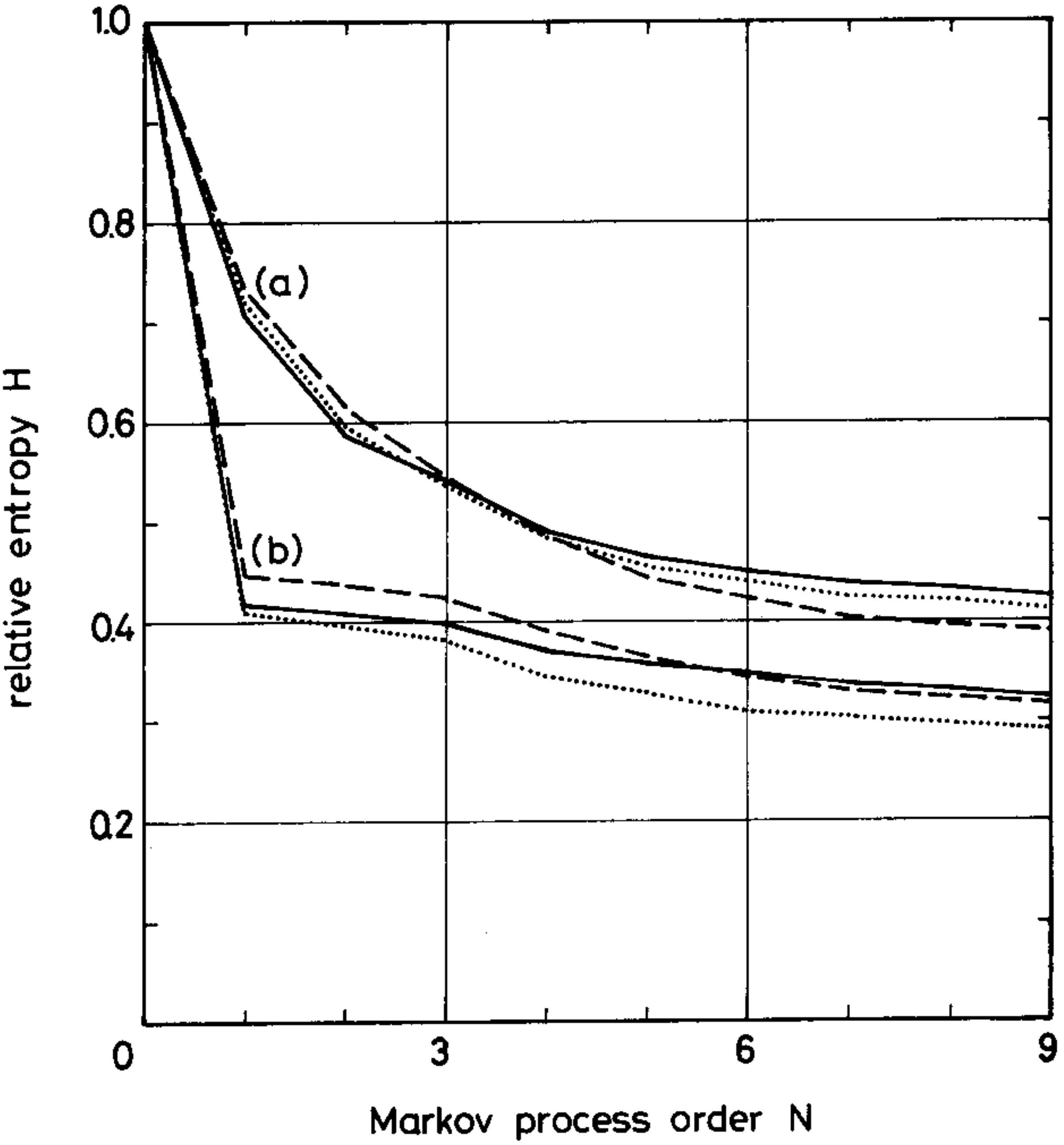


Fig. 2

Process entropy characteristics for constant overload probability

Sampling frequency f_s (——— 32 / (········· 48 kHz / (— — — 96

Constant $\Delta . f_s$ = Pk message amplitude × (1.39 group (a) / (3.48 " (b) about the conditional probability $p(S_j|B_i)$ as it is economically viable to compute at sender and receiver for limited size blocks B_i of past elements. For a ΔM message, an estimator processing N previous elements can make a binary prediction of the next element in sequence and the transformation effected as shown in Fig. 3 by modulo-2 addition of the element and its prediction.

$$T = \overline{S}\cdot E + S\cdot\overline{E} \tag{4}$$

At the receiver, the same prediction is generated by an identical estimator processing the same elements and the new element is determined by processing inversely.

For $\qquad \overline{T} = (\overline{E} + S\cdot E)\cdot(\overline{S} + S\cdot E)$

So that $\qquad \overline{T}\cdot E + T\cdot\overline{E} = S \qquad (5)$

The binary symbols which in conventional ΔM have the significance:

$$\text{(a)} \begin{cases} 0: & \text{Decrement approximant by } \Delta \\ 1: & \text{Increment approximant by } \Delta \end{cases}$$

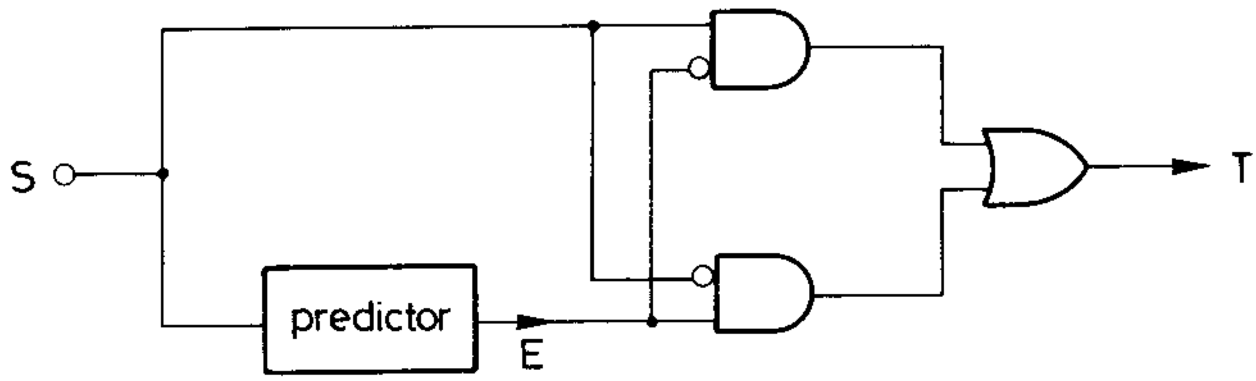
assume in this predictive coding scheme the indirect significance:

0: Predictor output correct, interpret it as (a)
1: Predictor output wrong, complement it and interpret as (a)

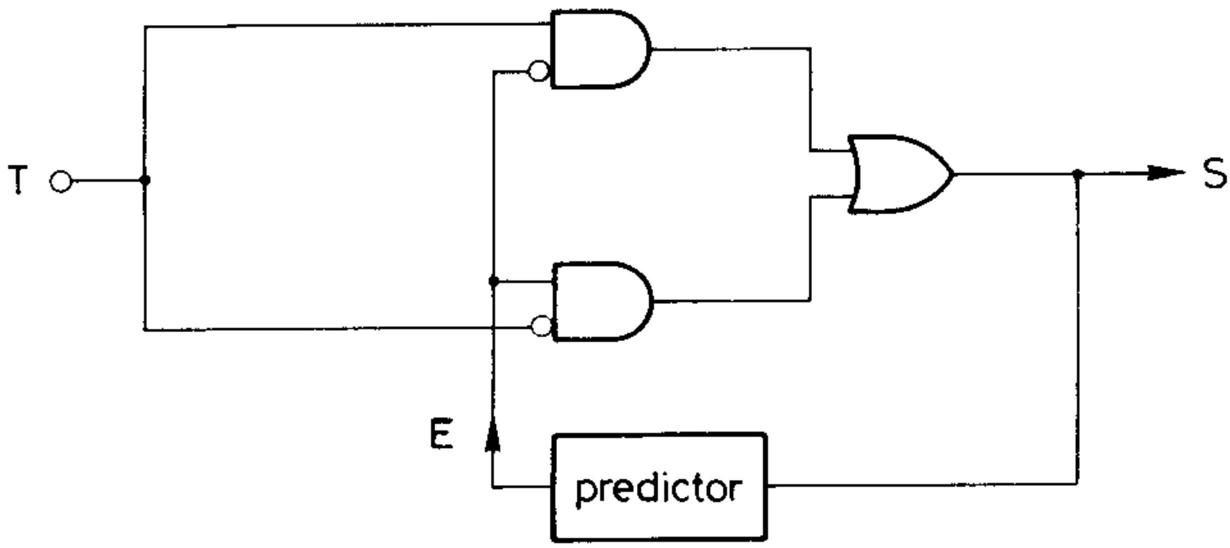
As the transformations are one-one the process entropy computed by (1) for $N \to \infty$, the weighted average of the entropies of the conditional distributions for all possible states, is the same for operand and transform sequences, but for a predictor success probability $P_s > 0.5$, the entropy

$$H_T = -\left[P_s \log P_s + (1-P_s)\log(1-P_s)\right] \tag{6}$$

of the average distribution of the elements of the new sequence is lower. While schemes for encoding the transformed sequence to achieve a reduction in channel bit-rate are described later (Section 5), it is noted here that its direct transmission in an application in which binary 1 is signalled by an



Transformation



Inversion

Fig. 3

ΔM predictive coding principle

output pulse, binary 0 by pulse absence, results in a transmitter power economy of

$$\eta = 100(2P_s - 1)\% \qquad (P_s \geq 0.5) \qquad (7)$$

3.1 Predictor structures

In generating matrix element listings for all vector source states and a range of values of N, the system implicitly defines the optimal fixed-structure predictors operating on blocks of a corresponding number of signal elements. For, on identification of the current state, the next element is merely predicted such that the subsequent state is that for which the transition probability is known to be a maximum.

While the asymptotic process entropies for the ΔM source signal and the modulo-2 adder output sequence of Fig. 3 are the same, the entropy of a zero-order approximation to the latter, computed for three values of Δ increment and shown by the characteristic of Fig. 4, is much less than that of a zero-order approximation to the former. For a range of N and Δ, the average element entropy of the transformed sequence for N-element prediction does not greatly exceed the process entropy of an Nth-order approximation to the ΔM message source and, for the typical median characteristic, an $N = 6$ predictor achieves $P_s = 0.9$, corresponding to $\eta = 80\%$. The average element entropy is within 14% of the 6th-order process entropy, and the redundancy in both cases exceeds one half. The optimal predictors for this characteristic and $N = 1$ to 7 are given in Table 1.

While increase of N by one order doubles the number of distinguishable states, the optimal predictor structure and performance changes only if the transition probability distributions for two or more of the states previously considered together have maxima for complementary next elements. In the case of characteristic (a) for $N = 3$-4 and (c) for $N = 1$-2 in Fig. 4, for example, this does not occur so that the predictors operating on the shorter signal sequences are as effective as those processing the

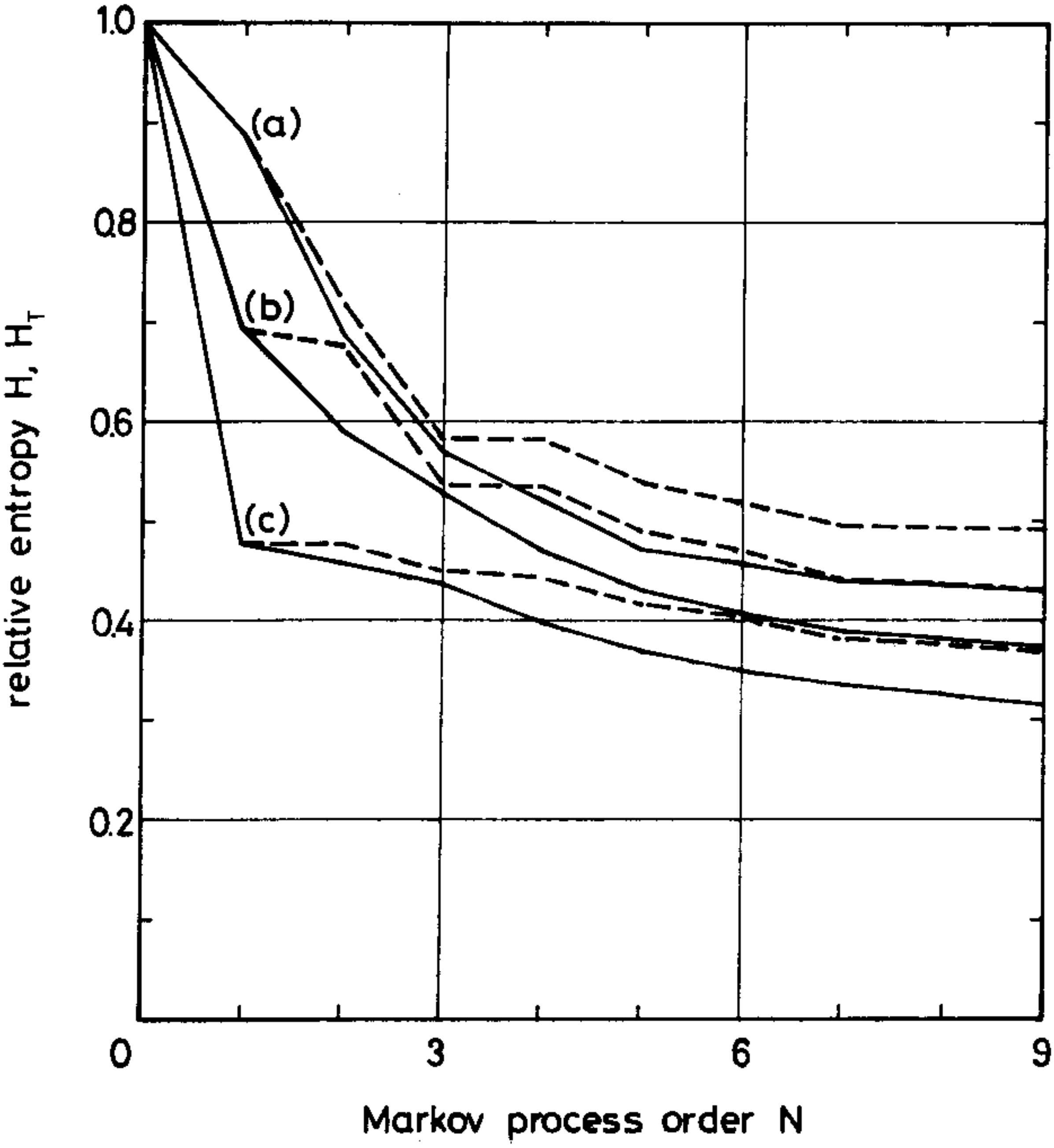


Fig. 4

Predictive coding sequence entropy (zero-order)

Sampling frequency $f_s = 96$ kHz

Pk-pk message amplitude (a) 256Δ (b) 128Δ (c) 64Δ

———) Relative entropy of process (order N)

- - - -) Mod-2 adder output sequence entropy (zero-order) with N-element optimum predictor

Table 1

OPTIMAL PREDICTORS FOR DELTA-CODED SPEECH

Sampling frequency 96 kHz   Pk-pk message amplitude 128Δ

(Vector states in octal)

N = 1

STATE	PRED
0	1
1	0

N = 2

0	0
1	0
2	1
3	1

N = 3

0	0	4	1
1	0	5	0
2	1	6	1
3	0	7	1

N = 4

00	0	10	1
01	0	11	0
02	1	12	1
03	0	13	0
04	1	14	1
05	0	15	0
06	1	16	1
07	0	17	1

N = 5

00	0	10	1	20	0	30	0
01	0	11	0	21	0	31	0
02	0	12	1	22	1	32	1
03	1	13	0	23	0	33	0
04	1	14	1	24	1	34	0
05	0	15	0	25	0	35	1
06	1	16	1	26	1	36	1
07	1	17	1	27	0	37	1

N = 6

00	0	10	1	20	0	30	1	40	0	50	1	60	0	70	0
01	0	11	0	21	0	31	0	41	0	51	0	61	0	71	0
02	0	12	1	22	1	32	1	42	0	52	1	62	1	72	1
03	1	13	0	23	0	33	0	43	0	53	0	63	0	73	1
04	0	14	1	24	1	34	1	44	1	54	1	64	1	74	0
05	0	15	0	25	0	35	1	45	0	55	0	65	0	75	1
06	1	16	1	26	1	36	1	46	1	56	1	66	1	76	1
07	0	17	1	27	0	37	1	47	0	57	1	67	0	77	1

N = 7

000	0	020	0	040	0	060	0	100	0	120	0	140	0	160	0
001	0	021	0	041	0	061	0	101	0	121	0	141	0	161	0
002	0	022	0	042	0	062	1	102	0	122	1	142	0	162	0
003	1	023	0	043	0	063	0	103	1	123	0	143	0	163	0
004	0	024	1	044	1	064	1	104	0	124	1	144	0	164	0
005	0	025	0	045	0	065	0	105	0	125	0	145	0	165	0
006	1	026	1	046	1	066	1	106	1	126	1	146	1	166	1
007	1	027	1	047	0	067	0	107	0	127	0	147	0	167	1
010	0	030	1	050	1	070	1	110	1	130	1	150	0	170	0
011	0	031	0	051	0	071	0	111	0	131	0	151	0	171	0
012	1	032	1	052	1	072	1	112	1	132	1	152	1	172	1
013	1	033	1	053	0	073	1	113	0	133	0	153	0	173	1
014	1	034	1	054	1	074	0	114	1	134	1	154	1	174	0
015	1	035	1	055	0	075	1	115	0	135	1	155	1	175	1
016	1	036	1	056	1	076	1	116	1	136	1	156	1	176	1
017	1	037	1	057	1	077	1	117	1	137	1	157	1	177	1

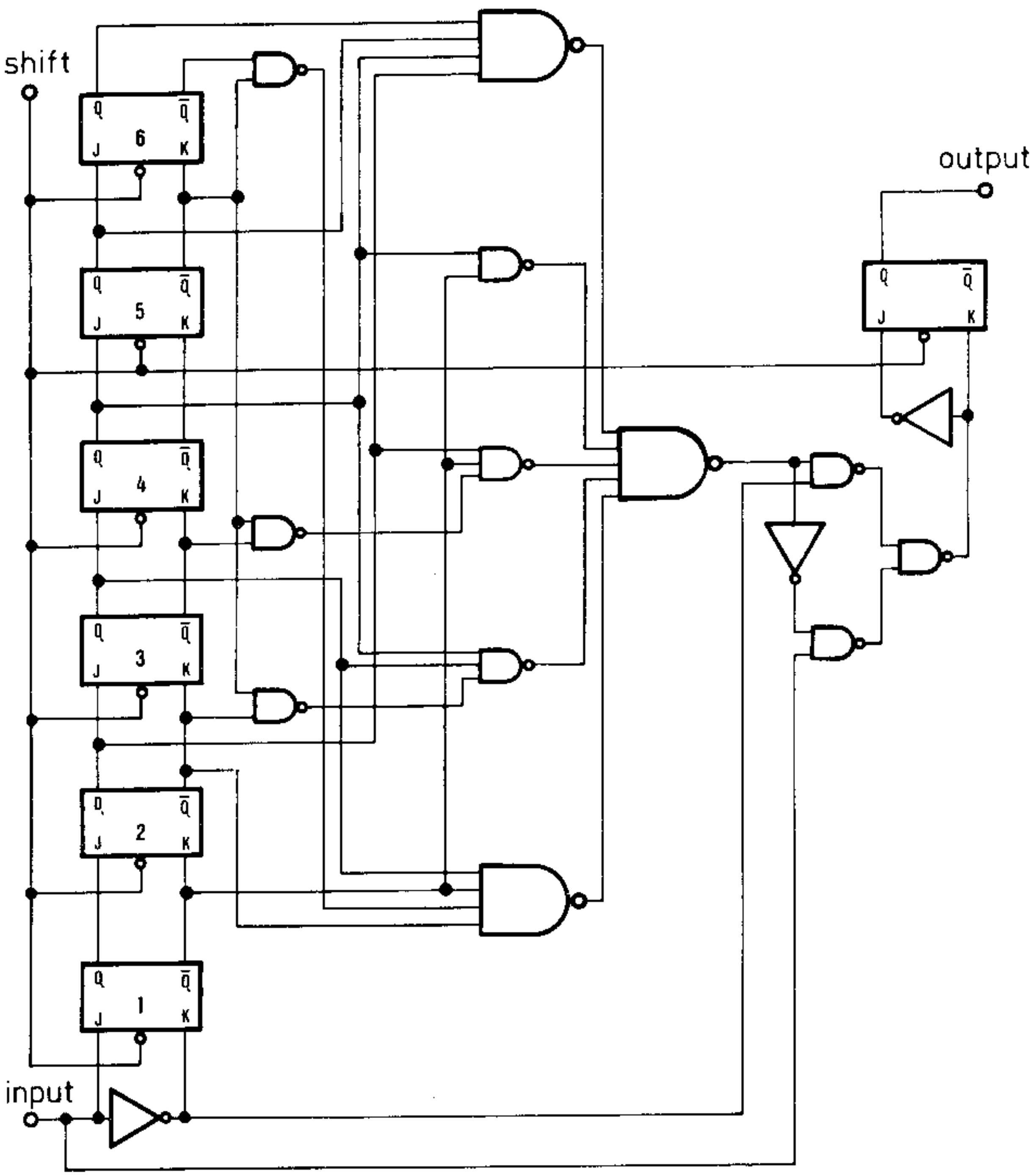


Fig. 5

6th-order predictive coder configuration for ΔM speech longer. Digital predictors of this type may be realised by a $2^N \times 1$ bit read-only memory addressed by the current state vector and programmed with the next element prediction. Read-mostly memory[33] or even random access memory may be used if adaptation of the predictor function to the source statistics encountered in service is envisaged.

3.2 6th-order predictive coder

As a consequence of the asymptotic equipartition property the source state ensemble for N large comprises a 2^{NH} member set of high probability and a remaining set of low probability. While a distinct set boundary does not occur for N = 6, H = 0.41, the on-line analysis confirms that a number of the 64 states of the ensemble have quite low probabilities. Eight states have occurrence probabilities of less than 10^{-4}, and in a chain of over 4×10^6 states, there are no occurrences at all of 34_8, all transitions to 70_8 and 71_8 being from 74_8. This property may be exploited to derive practically convenient predictor functions which represent worthwhile simplifications at the expense of negligible performance degradation from that of the optimal structure.

States with low occurrence probabilities, and those with high occurrence probabilities but near equiprobable transition probabilities (high state entropies) may be assigned arbitrary predictions where logic minimisation techniques indicate that this results in a hardware simplification. Veitch-Karnaugh mapping of the 6th-order predictor defined in Table 1, for example, reveals that the introduction of four such 'don't care' predictions allows the 1's factoring to be carried out in groups of four or fewer variables by combining four or more elements per term to yield the sub-optimal predictor function

$$a \cdot b \cdot c \cdot e + c \cdot \overline{f} + e \cdot \overline{f} \cdot (d + b) + c \cdot d \cdot (e + b) + d \cdot \overline{e} \cdot \overline{f} \cdot (a + b) \qquad (8)$$

in which the order of generation of the sequence is ...a...f...
A complete predictive coder for this example is shown in Fig. 5. The output of the modulo-2 adder conditions the encoder output flip-flop on

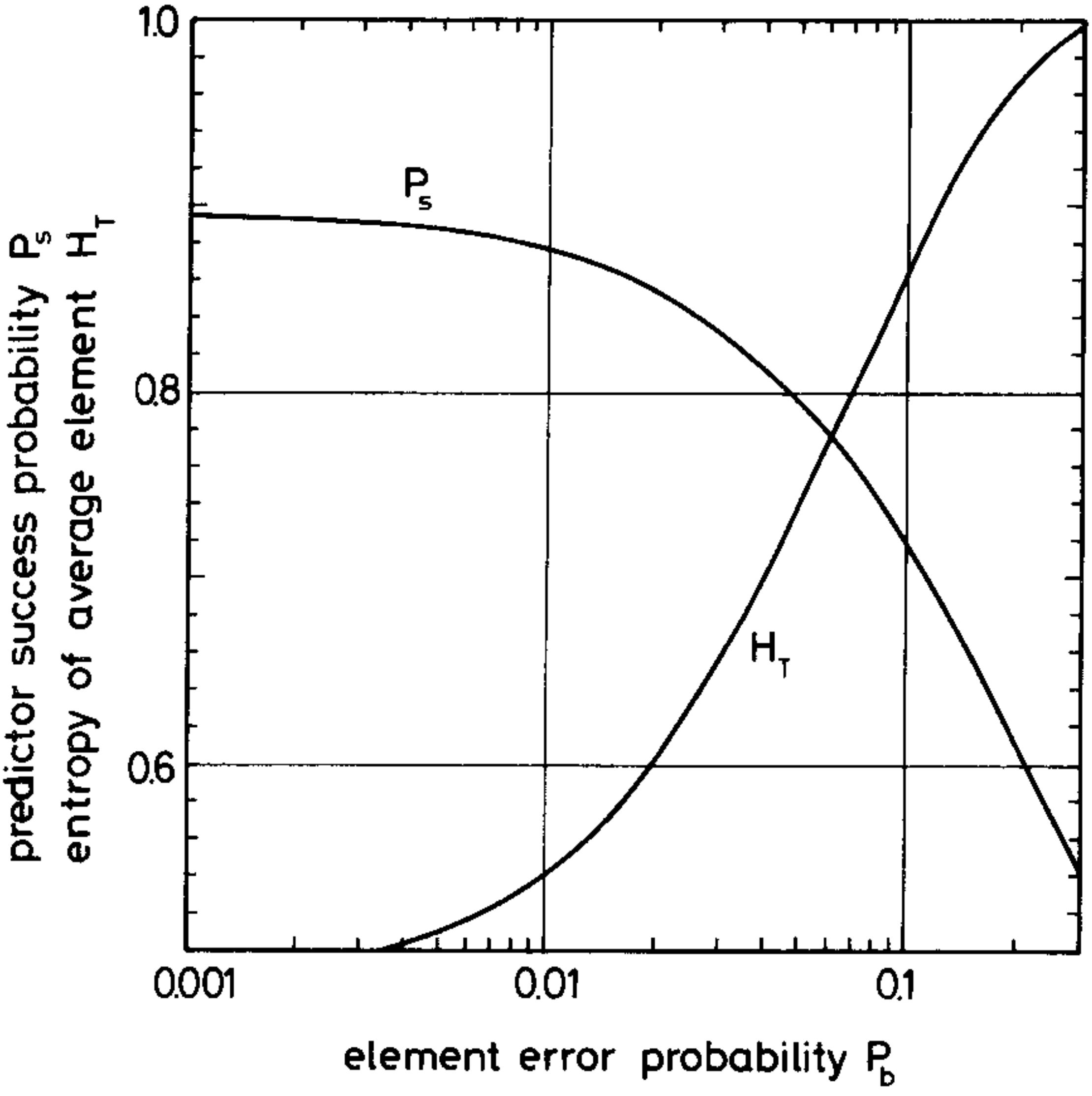


Fig. 6

6th-order predictor performance degradation produced by channel errors the arrival of the shift pulse which clocks into the first position of the shift register the most recent signal element with which the prediction has been compared. Obvious modifications of the configuration form the corresponding decoder, in which the input signal drives the mod-2 adder but the output sequence enters the shift register.

By prefixing the analysis system by this coder, transition matrices for the predictive coded message are generated which can be processed in the same way as those for the ΔM signal direct. The results show that P_s for the former case and $N = 0$ is less than for the latter case and $N = 6$ by 0.14%, which only slightly exceeds the variation between repeat test runs. So the performance of the practical predictor described is little degraded from that of the optimal structure by the simplifications allowed in its derivation.

3.3 Data errors

In direct ΔM transmission, signal detection errors due to channel noise result in an output s.n.r. reduction which Braun et al[34] have reported ranges from 1 dB for error probability $P_e = 0.001$ to 15 dB for $P_e = 0.1$. When the redundancy of the signals generated by an information source is reduced, the susceptibility of the message to mutilation by noise is increased. For the predictive coding of ΔM speech, incorrectly detected elements cause immediate signal reconstruction errors on reception, but in addition P_s is reduced by the impairment of the decoder predictor's assessment of the subsequent N states and the increased entropy of the data sequence.

The sensitivity of the 6th-order predictor performance to data errors is determined by a series of signal processor runs in which random element errors are introduced with a range of controlled probabilities P_b. During each sampling interval, the software generates a random number from a uniform distribution over 0-4095, and causes the hardware to introduce an error when the numbers are less than a threshold parameter which is P_b as a 12-bit binary fraction. Fig. 6 shows the degradation of P_s with increasing

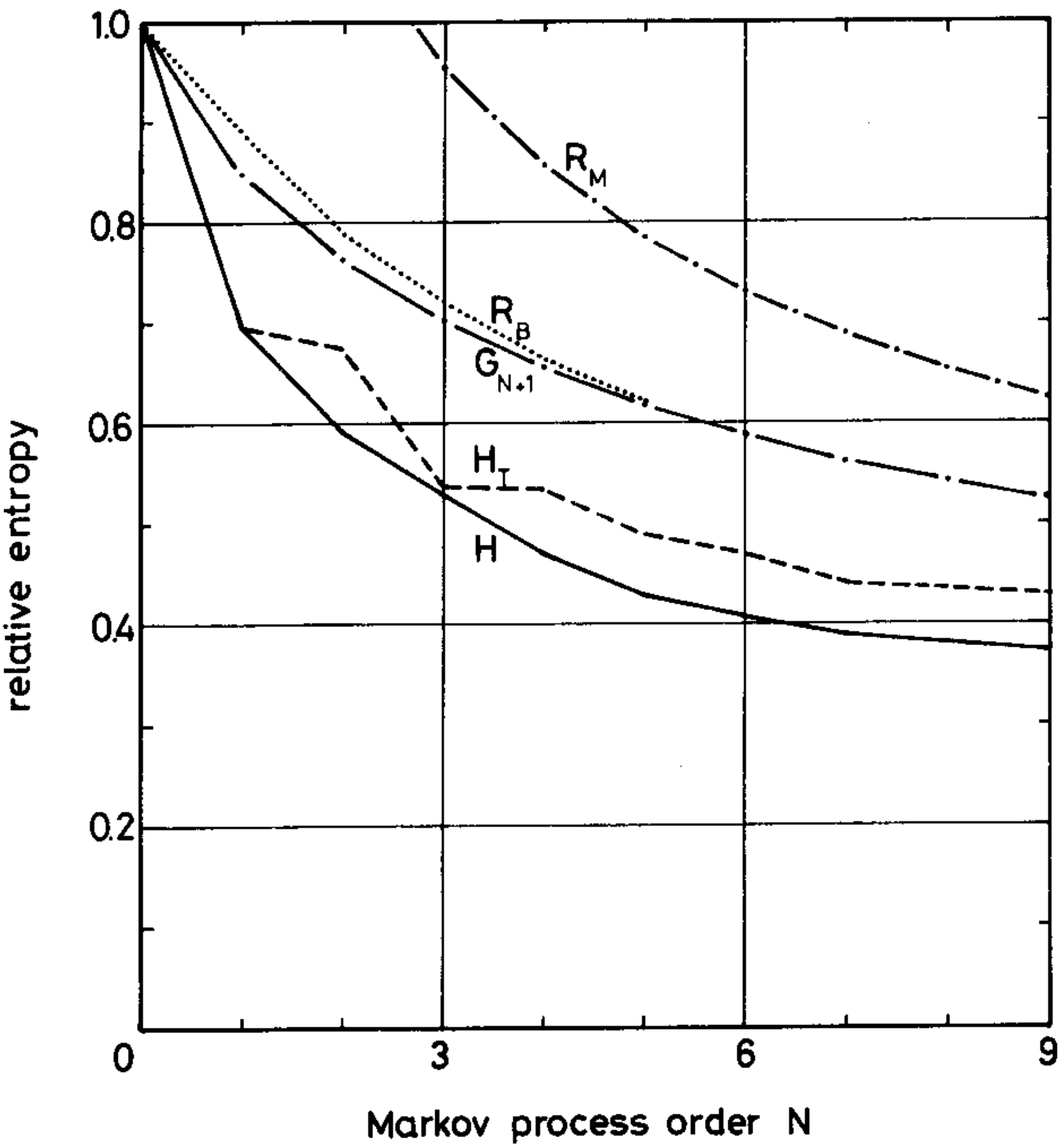


Fig. 7

Delta-coded speech entropies

R_M	Upper bound for group encoding	R_B	Huffman encoder characteristic
G_{N+1}	N+1 element group entropy	H_T	Mod-2 adder sequence entropy
H	Process entropy		

data error rate, with the corresponding rise in H_T as the sequence properties change from the constrained statistics of ΔM speech to those of binary random noise. The predictor remains effective for error rates up to that causing an appreciable reduction in output s.n.r. However, as with all differential transmission schemes, it is advisable to inhibit error accumulation by the periodic clearing of encoder and decoder memories, in this case to the quiescent states 12_8 or 25_8.

4 ΔM Group encoding

Instantaneous group encoding procedures considered by Shannon[35], Fano[36] and Huffman[37] effect redundancy reduction by the assignment, to N-element blocks B_i of the signal sequences generated by a source, of uniquely decipherable N_i-element codewords in a manner determined by their occurrence probability set. For all procedures, the average channel capacity required for transmission of the encoded signal, expressed per element of the source output, approaches G_N given by (2) as N increases, while G_N converges to H for large N, so that the coding schemes are asymptotically completely efficient. But for finite N the Huffman approach yields group codes with optimal efficiency, and it is therefore applied to ΔM speech for comparison with predictive coding.

4.1 Group entropy

The relative entropies of (N+1)-element blocks of signal elements are computed from the transition matrices and the results are shown in Fig. 7. The group entropies are compared with the Markov process and modulo-2 adder output sequence entropies based on the same total number of signal elements, including the immediate elements in the last two cases, direct transmission corresponding to a zero-order approximation or encoding in groups of one element. The characteristic defines a lower bound to the required channel capacity for signal transmission after contiguous element group encoding by any procedure.

By optimal coding it is always possible to establish a transformation

Table 2

OPTIMAL GROUP CODES FOR DELTA-CODED SPEECH

Block length N+1 elements

N = 1

0	1	(3)
1	1	(1)
2	1	(2)
3	0	(3)

N = 2

0	3	(4)
1	02	(5)
2	1	(1)
3	00	(5)
4	03	(5)
5	1	(2)
6	09	(5)
7	2	(4)

N = 3

00	1	(4)
01	01	(6)
02	5	(4)
03	01	(8)
04	6	(4)
05	2	(2)
06	09	(5)
07	03	(6)
10	02	(6)
11	08	(5)
12	3	(2)
13	7	(4)
14	00	(8)
15	2	(4)
16	01	(7)
17	3	(4)

N = 4

00	5	(4)	20	21	(8)
01	20	(8)	21	24	(8)
02	11	(7)	22	06	(5)
03	002	(12)	23	001	(10)
04	08	(5)	24	0D	(5)
05	0C	(5)	25	2	(2)
06	002	(9)	26	01	(5)
07	000	(12)	27	02	(7)
10	13	(7)	30	001	(12)
11	07	(5)	31	001	(9)
12	3	(2)	32	03	(5)
13	02	(5)	33	09	(5)
14	003	(9)	34	003	(12)
15	05	(5)	35	03	(7)
16	25	(8)	36	03	(8)
17	02	(8)	37	07	(4)

N = 5

00	03	(5)	20	01	(8)	40	053	(9)	60	001	(12)
01	052	(9)	21	21	(8)	41	010	(9)	61	2201	(16)
02	04	(8)	22	07	(5)	42	23	(8)	62	083	(10)
03	221	(12)	23	003	(10)	43	11001	(19)	63	000	(12)
04	01	(7)	24	09	(5)	44	03	(6)	64	082	(10)
05	111	(11)	25	2	(2)	45	0D	(5)	65	08	(5)
06	1101	(15)	26	0E	(5)	46	0A1	(10)	66	15	(7)
07	201	(12)	27	001	(11)	47	22000	(20)	67	03	(7)
10	13	(7)	30	08801	(18)	50	003	(11)	70	005	(12)
11	16	(7)	31	0A0	(10)	51	0F	(5)	71	0881	(14)
12	0B	(5)	32	0C	(5)	52	3	(2)	72	101	(11)
13	002	(10)	33	17	(7)	53	0A	(5)	73	05	(7)
14	0441	(13)	34	22001	(20)	54	081	(10)	74	200	(12)
15	089	(10)	35	24	(8)	55	02	(5)	75	09	(8)
16	04401	(17)	36	011	(9)	56	25	(8)	76	045	(9)
17	004	(12)	37	051	(9)	57	05	(8)	77	06	(5)

resulting in an average codeword length $\sum_i p(B_i)N_i$ within one element of the average block total entropy NG_N, so that an upper bound, also shown in Fig. 7, is set by

$$NR_B \leq NG_N + 1 = NR_M \tag{9}$$

in which the ratio of the channel capacity required for the encoded signal to that for direct transmission

$$R_B = \sum_{i=1}^{2^N} \frac{p(B_i)N_i}{N} \tag{10}$$

indicates the redundancy reduction produced.

4.2 Optimal group codes

The optimal group codebooks for 2 to 6 signal elements are given in Table 2. For compactness of presentation, each right-justified codeword is expressed as a hexadecimal number, followed in parentheses by the (decimal) number of binary digits which are valid. Codes for larger blocks are unlikely to be of practical interest because of their complexity and the long word lengths associated with low-probability states. To evaluate the efficiencies of the group codes, the codeword lengths N_i elements, indicated for each block by Table 2, are assigned to the appropriate groups B_i of the source sequence. Each element is considered to be the first of a following group, so that the evaluation for each group length N is the average for N effective test runs with time origins at each of the elements of the leading ΔM code block. Values of R_B are shown in Fig. 7.

In appraising the group encoding results, it is noted that even for small N, the codes are efficient in the sense that R_B is near the lower bound set by the relative entropy of groups of (N+1) elements of the source sequences, and the performance characteristic converges to G_{N+1} much faster than does the upper bound R_M. But the convergence of G_N to the process entropy H is slow, so that large blocks must be encoded to achieve a high overall coding efficiency. In comparison, the relative entropy (zero-order) H_T of the modulo-2 adder output sequence for an N-element optimum predictor remains much closer to H for all N, so that a predictive coding approach is found to be superior to optimal group encoding for ΔM speech. H_T for a 3-element predictor, for example, which has a quite elementary structure (Table 1), is lower than the attainable R_B for blocks of 8 elements, although the group encoder for the latter is complex in form and generates very long codewords.

5 Error sequence coding

We now consider applications to channel encoding and transmitter power economy of the 6th-order near-optimal predictive coder example defined in Section 3.2. The binary signal sequence generated by modulo-2 addition of the ΔM message and predictor output is termed the encoder error sequence, for it has the function of indicating the errors in the decoder predictions. Direct transmission and three methods of variable-length coding the error sequence to achieve channel bandwidth compression are evaluated and compared.

5.1 Direct error transmission

As was noted earlier, a transmitter power saving is possible by direct transmission of the predictive coded sequence instead of the original ΔM message, and for the 6th-order predictor with $P_s = 0.9$ this has value $\eta = 80\%$. In a pack-set application in which the low-level circuits are microminiaturized and the power input to the final transmitter stages is a large proportion of the total, operational life from a primary power source is thus extended by a factor of several times. In addition, use of this mode extends to pulse systems the operational advantages brought to analog communications by suppressed-carrier voice-controlled (VOX) working. For both techniques result in there being no output from a sender except during speech utterances, so that single-channel conversational mode operation becomes possible among the contributing members of a communications net, bit synchronisation being established by the transmitting member and no frame synchronisation being required to decode the transmitted sequences.

The entropy characteristics of Fig. 7, and the corresponding predictor performances, were computed by processing a speech message generated by the continuous reading of a prepared text. In the case of duplex communications, the power saving (and also the detection immunity) compared with direct ΔM are therefore further enhanced by the occurrence of intervals during which the transmitting output of a listening operator is zero, although his talk channel remains active for interruption of the sender at any time. The power saving for this case becomes

$$\eta = 100(1 - 2P_t + 2P_s P_t)\% \qquad (P_s \geq 0.5) \qquad (11)$$

in which P_t is the channel activation probability, so that an extension of operational life by a factor of 10 is indicated in this service for $P_t = 0.5$. At fixed stations for a transmitter limited by output device mean dissipation the encoding gains are more usefully exploited to allow an increase in channel signal power. The thermal time-constants of the dissipating structures of high-power tubes and semiconductors are typically sufficient to smooth the fluctuations during speech and allow close to a factor of $1/\{2(1 - P_s)\}$ increase (7 dB for the 6th-order predictor), but they are not sufficient to average the dissipated power over active and inactive periods in duplex working. For the case of a peak power limited sender, and for other ways of exploiting the redundancy of ΔM speech, we must consider further processing of the encoder error sequence.

5.2 Error sequence entropy

In the channel encoder situation, it is required to translate the attainable average element entropy reduction to a decrease in bit-rate, either to increase the size of the t.d.m. group, or to cascade the operation of removal of the inefficient source redundancy with that of substitution by the check digits of an error-correcting code chosen

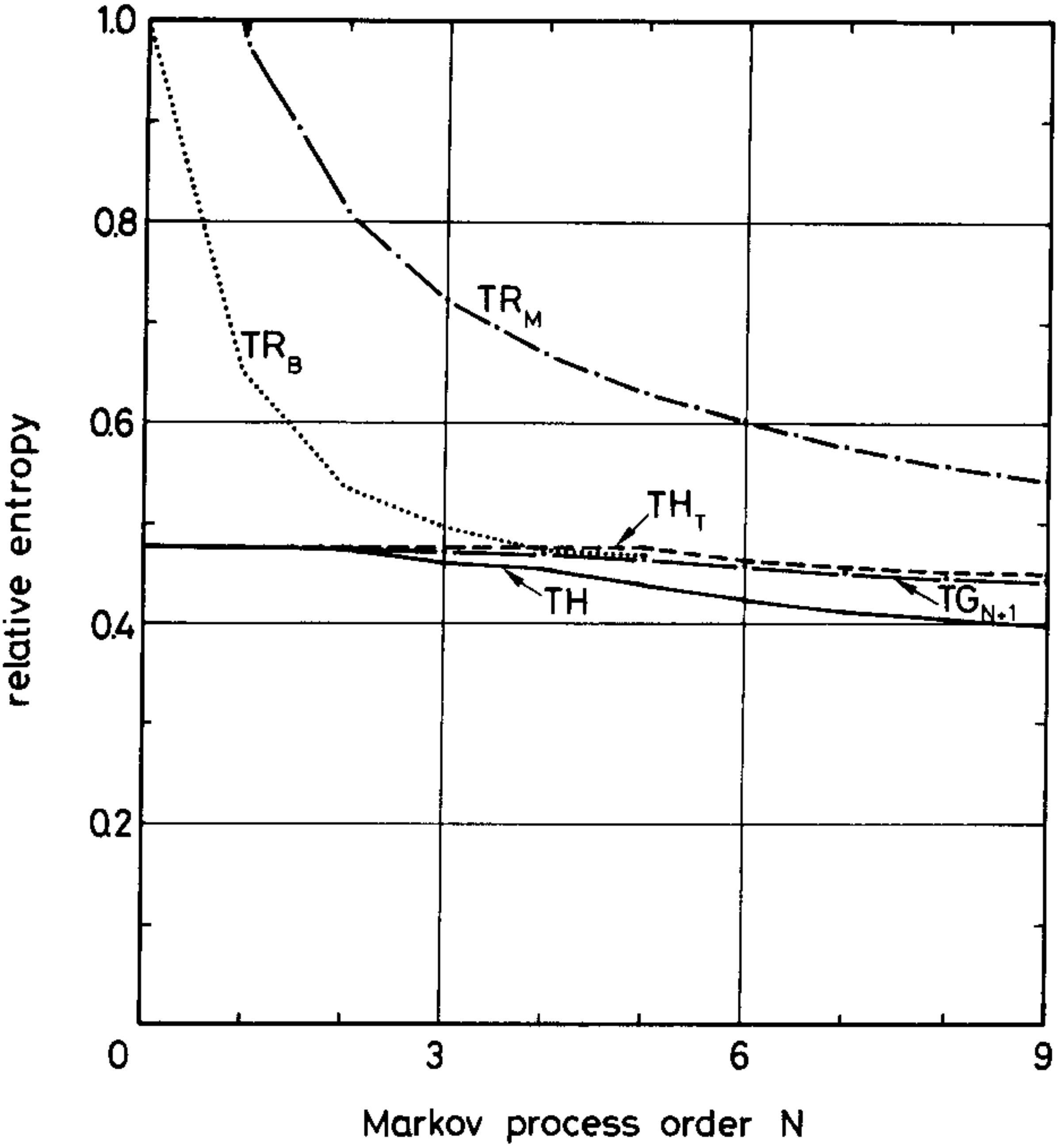


Fig. 8

<u>ΔM predictive coding error sequence entropies</u>

TR_M	Upper bound for group encoding	TR_B	Huffman encoder characteristic
TG_{N+1}	N+1 element group entropy	TH_T	Mod-2 adder sequence entropy
TH	Process entropy		

specifically for the particular channel noise properties encountered.

By processing the transition probability matrices for the encoder of Fig. 5 by the information analysis programs, the error sequence information properties summarized in Fig. 8 are determined. The process entropy characteristic given indicates the reduction of the statistical interdependence of sequential elements resulting from the coding; for the invariant distribution for the zero-order case with element probabilities P_s, $1-P_s$, has entropy not exceeding those computed from the sets of conditional distributions for N = 1 to 9 by more than 16%. The variation with N of TH_T, the average element entropy of the sequences generated by a second application of predictive coding (using optimal predictors of order N) to the error sequence generated by coding with the near-optimal 6th-order predictor, is even less over this range. In fact an extension of the initial predictor order by 1 achieves an entropy reduction which it requires a second application predictor order N > 9 to match.

5.3 Run-length encodings

The zero-order approximation to the error sequence suggested by the entropy results of Fig. 8 implies a geometric probability distribution for the lengths of runs of consecutive 0 symbols in the sequence. The probability of a run of length r ($r \geq 0$, so that every symbol 1 both terminates and starts a run) is

$$P_{gr} = P_s^r (1 - P_s) \qquad (12)$$

and the entropy per selection from the distribution,

$$H_g = \sum_{r=0}^{\infty} P_{gr} \log P_{gr} , \qquad (13)$$

has value 4.66 bits for the 6th-order predictor.
The average sequence length per run,

$$L_g = \sum_{r=0}^{\infty} P_{gr} (r + 1) \qquad (14)$$

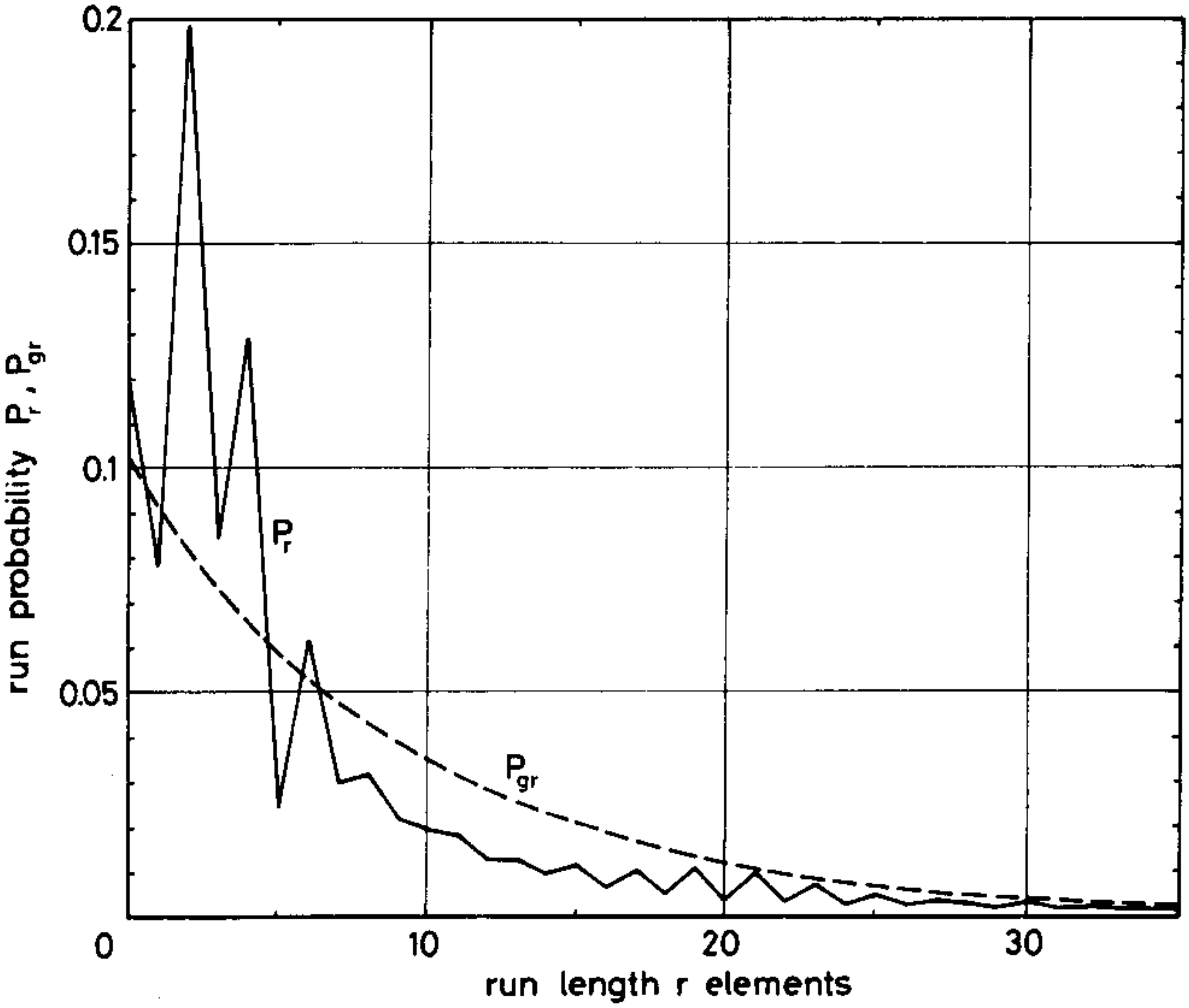


Fig. 9

Run length distributions for ΔM predictive coding

P_r - Run length probability for 6th-order predictive coding

P_{gr} - Geometric distribution for $P_s = 0.9$

is 9.83 elements and the relative entropy per element $H_{gr} = H_g/L_g = 0.474$, corresponding to TH_T for N = 0. Similar form relations derive from the actual run probabilities P_r.

Error sequence run-length distributions for $r < 256$ are determined by the on-line analysis programs for ΔM speech with predictive coding. Fig. 9 compares the run-length distribution for the typical ΔM parameter case with the geometric distribution for the same P_s, and distributions for other cases are similar. Deviations for small r are caused by the previous element dependence of the error element distribution remaining after coding, and for $r \geq 7$, P_r is always less than P_{gr}. From the run probability results, the average entropy per run is computed as 4.31 bits and the average sequence length per run as 9.75 elements so that the relative entropy per element $H_r = 0.442$ is rather lower than for the geometric distribution. While Huffman encoding of the run-lengths in accordance with their occurrence probabilities allows this compression to be closely attained, the codebook is quite large (64 entries to include 98% of the run lengths) and the encoder, although very much simpler than that which would be required to achieve the same compression by a direct application to the source sequence (32K entries), is still unattractively complex.

Considering only comma-free codes for direct transmission through binary channels, a first practical alternative is fixed length binary number coding of the run-lengths, which of course is still a variable-rate encoding procedure overall because of the distribution of element sequence lengths selecting each number. For this approach 6-bit coding yields a compression of 0.615, while encoding only run lengths 0 to 31 (which includes 94% of the total) by 5 bits results in a ratio of 0.513. The occurrence of the longest runs, which are associated with quiet intervals between spoken words, is a function of the system noise and reverberation prior to source encoding and is likely to be less frequent in the average communications situation than for the studio environment in which the test message is prepared and the precision laboratory equipment by which it is processed.

Table 3

RUN LENGTH CODEBOOK

m = 6   geometric distribution

r	codeword	r	codeword
0	00 (3)	16	36 (6)
1	01 (3)	17	37 (6)
		18	38 (6)
2	04 (4)	19	39 (6)
3	05 (4)		
4	06 (4)	20	74 (7)
5	07 (4)	21	75 (7)
6	08 (4)	22	76 (7)
7	09 (4)	23	77 (7)
		24	78 (7)
8	14 (5)	25	79 (7)
9	15 (5)		
10	16 (5)	26	F4 (8)
11	17 (5)	27	F5 (8)
12	18 (5)	28	F6 (8)
13	19 (5)	29	F7 (8)
		30	F8 (8)
14	34 (6)	31	F9 (8)
15	35 (6)		

For a second procedure of intermediate complexity, the set of P_r may be approximated by a geometric distribution which Golomb[38] has noted is favourable to Huffman coding for integer values of

$$m = -\frac{\log 2}{\log P_s} \tag{15}$$

For the 6th-order predictor, m = 6 is appropriate and the initial entries for the run-length codebook are given in Table 3, examination of the structure of which reveals the following encoding rule for a run-length of r elements:

By division of r by 6, obtain 6A + R. (integers $A \geq 0$, $0 \leq R \leq 5$)

Encoding rule: Output A binary 1's, followed by

$$\begin{cases} \text{for } 0 \leq R \leq 1, \text{ the 3-bit binary representation of } R \\ \text{for } 2 \leq R \leq 5, \text{ the 4-bit binary representation of } R+2 \end{cases}$$

An equally straightforward decoding rule applies.

Evaluation of this encoding procedure for the actual run-length distribution of Fig. 9 indicates an average relative entropy (per element of the error sequence) of 0.486, reduced to 0.406 by truncation of the codebook at 32 entries. The approximation of the run-length statistics by a geometric distribution, and of P_s by a value giving m integer, thus yields a practical run-length encoding scheme with performance within 10% of H_r.

5.4 Group error encoding

While predictive coding has been considered in earlier sections as an IP transformation lowering the source sequence average element entropy, significant reductions are effected in the entropies of blocks of N elements for $N \geq 2$ also. The group entropy characteristic TG_{N+1} and upper bound for the optimal coding TR_M shown in Fig. 8 are obtained for the encoder error sequence by the processing method described in Section 4 for the ΔM signal prior to coding. In the present case, TG_{N+1}

Table 4

OPTIMAL ERROR SEQUENCE GROUP CODES

Block length N+1 elements

N = 1

0	1	(1)
1	1	(2)
2	0	(3)
3	1	(3)

N = 2

0	1	(1)
1	2	(3)
2	1	(3)
3	02	(5)
4	3	(3)
5	00	(5)
6	03	(5)
7	01	(5)

N = 3

00	1	(1)
01	4	(4)
02	3	(3)
03	02	(6)
04	1	(3)
05	05	(6)
06	01	(6)
07	02	(8)
10	5	(4)
11	03	(5)
12	03	(6)
13	01	(8)
14	04	(6)
15	001	(9)
16	03	(8)
17	000	(9)

N = 4

00	1	(1)	20	1	(4)
01	7	(4)	21	24	(7)
02	5	(4)	22	13	(6)
03	40	(8)	23	41	(8)
04	1	(3)	24	03	(7)
05	21	(7)	25	001	(10)
06	01	(7)	26	001	(9)
07	00A	(9)	27	002	(12)
10	6	(4)	30	25	(7)
11	01	(5)	31	010	(10)
12	23	(7)	32	003	(12)
13	00B	(9)	33	0001	(14)
14	22	(7)	34	009	(9)
15	0000	(14)	35	001	(12)
16	011	(10)	36	003	(9)
17	002	(9)	37	0001	(13)

N = 5

00	1	(1)	20	4	(4)	40	03	(5)	60	03	(8)
01	5	(4)	21	13	(8)	41	0B	(6)	61	09	(8)
02	6	(4)	22	03	(6)	42	05	(7)	62	021	(10)
03	021	(9)	23	01	(8)	43	041	(9)	63	0201	(13)
04	3	(4)	24	15	(7)	44	01	(6)	64	203	(12)
05	14	(7)	25	083	(11)	45	081	(10)	65	2001	(16)
06	21	(8)	26	00B	(10)	46	025	(9)	66	10003	(19)
07	091	(11)	27	202	(12)	47	000	(10)	67	20001	(20)
10	7	(4)	30	12	(7)	50	11	(8)	70	011	(9)
11	05	(6)	31	020	(10)	51	001	(9)	71	0401	(13)
12	11	(7)	32	0801	(15)	52	082	(11)	72	201	(12)
13	049	(10)	33	10002	(19)	53	04001	(17)	73	040000	(21)
14	13	(7)	34	081	(11)	54	00A	(10)	74	008	(10)
15	0800	(15)	35	0203	(13)	55	10001	(19)	75	1001	(15)
16	090	(11)	36	009	(10)	56	0801	(14)	76	0803	(15)
17	001	(10)	37	0802	(15)	57	0202	(13)	77	040001	(21)

remains much closer to TH, even for N small, and is actually slightly less than TH_T. As a third approach, the group codes are therefore constructed by the application, to the encoder error sequence transition matrices, of the procedure used for group encoding for ΔM sequences direct, and the optimal codebooks are given in Table 4. Evaluation of these codes yields the characteristic TR_B shown in Fig. 8 for the reduction in required channel capacity. Convergence to the lower bound occurs rapidly, an average bit-rate reduction of one half being obtained with groups of 4 only, and TR_B being less than TH_T for block length ≥ 5. An attractive case, compromising considerations of coding efficiency and hardware complexity, is the 5-element group code which achieves $TR_B = 0.479$.

Comparison of the most practical case performance for each of the three error sequence encoding procedures described is now possible and the compression factors attainable are summarized below.

(a)	6-element binary number coding of run-lengths:	0.615
(b)	Geometric distribution run-length encoding:	0.486
(c)	5-element group encoding of error sequence:	0.479

Method (a) achieves simplicity at the cost of reduced efficiency, while speed considerations will determine whether an implementation of method (c), which attains the best performance with increased memory capacity, is preferable to obtaining a similar compression with greater arithmetic logic by method (b).

6 Channel buffer

In speech communication, as opposed to telegraphy, the rate of reconstruction of the message for the recipient cannot be varied from its rate of generation without distortion of the intelligence conveyed. Where redundancy reduction is effected by a statistical encoding procedure, the required channel capacity is proportional to the information content of the source messages. At channel encoders and decoders, buffer stores are therefore necessary to smooth the fluctuating data rate for transmission and reconstruct its variation with time on reception.

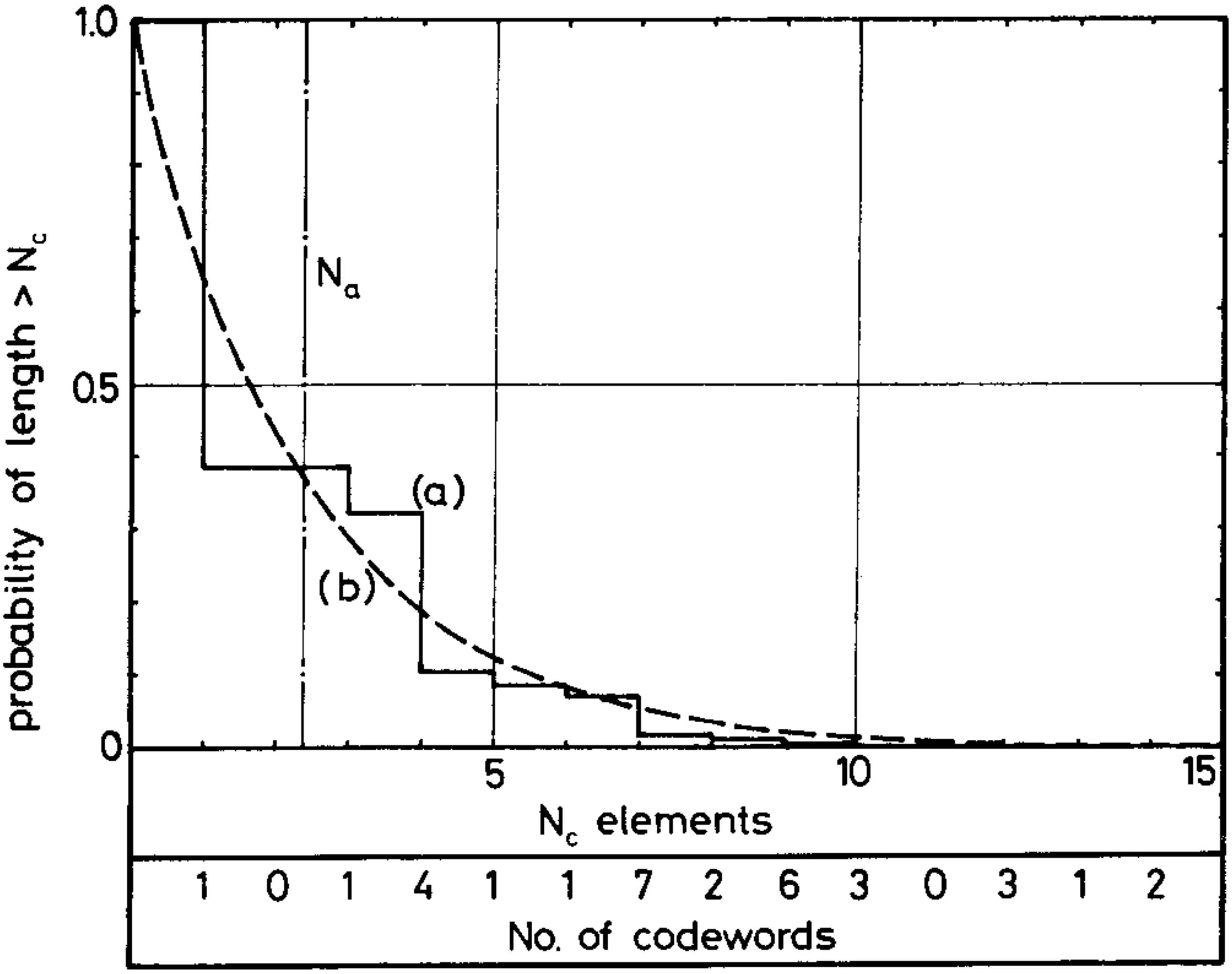


Fig. 10

Codeword length distributions

(a) 6th-order predictive coder followed by 5-element error sequence group encoding

(b) Negative exponential approximation with the same average N_a = 2.4 elements

6.1 Queue organisation

For the specific numerical results, it is assumed that redundancy reduction is achieved by the 6th-order predictor described in Section 3.2 followed by 5-element group encoding of the error sequence by the code given in Table 4 (N = 4). Each ΔM source, with data rate f_s, thus selects for transmission codewords of length N_c = 1 to 14 elements (excluding 2 and 11) at a rate $^{f_s}/5$. The sequence of codewords generated by successive selections by the members of the t.d.m. group of N_m sources is held in the buffer store, of limited capacity N_q words, and this queue is serviced in order by the sender which transmits corresponding signals over the channel at a bit-rate $N_m f_c$. For its complete transmission, each codeword requires a holding time proportional to its length, after which the channel becomes available to service the next codeword in the buffer.

When a codeword is generated, it encounters a number w of codewords ahead of it awaiting or undergoing transmission. With a probability which is small, w is zero and the codeword is processed immediately; while more frequently it experiences a delay which will have duration > τ if w-1 or fewer codeword transmissions are completed during time τ. Buffer capacity and channel data rate require to be chosen such that τ rarely exceeds $5\ ^{N_q}/f_s\ N_m$, so that the probability of buffer overflow is small. Blasbalg and Van Blerkom [20] suggest degrading the source fidelity to maintain transmission when buffer overflow occurs. Short codewords which approximate the longer sequences selected may be sent until the overflow clears. The requirement for the receiver is identical, codewords transmitted by the channel entering the buffer store which receives sequential service from the group decoder.

6.2 Codeword length distribution

From the information analysis of the predictor error sequence described in Section 5 the total occurrence frequency of each 5-element state is found and, by grouping those states for which the corresponding codewords from Table 4 are equal in length, the probabilities of all N_c

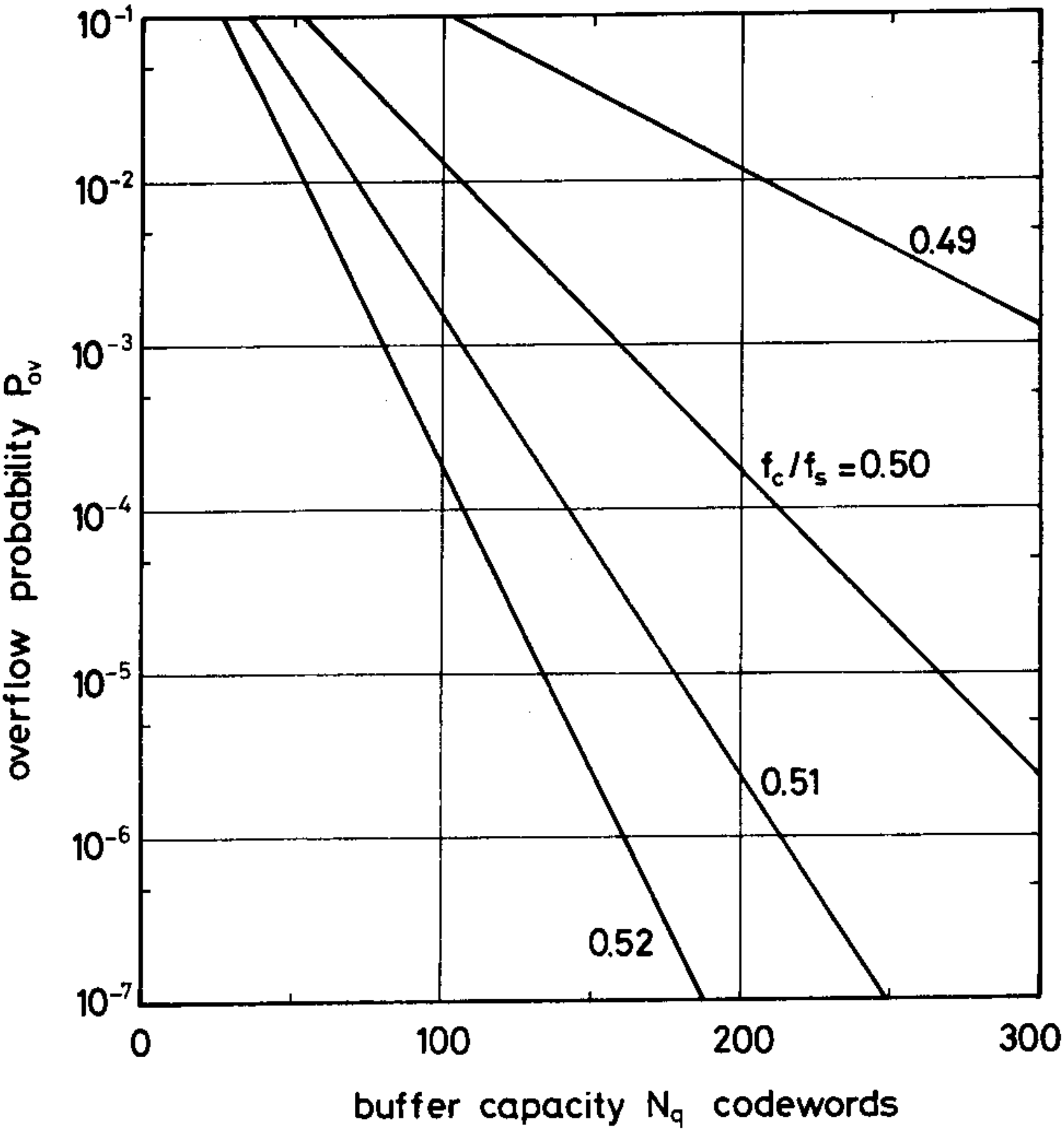


Fig. 11

<u>Buffer characteristics</u>

Compression factor = $^{f_c}/_{f_s}$

are computed. The resultant integral distribution function $p(>N_c)$ is given in Fig. 10 and the average codeword length

$$N_a = \sum_{N_c=1}^{14} p(N_c)\, N_c \qquad (16)$$

is 2.40 elements.

As shown, a negative exponential approximation to the distribution with the same average may be taken, for which

$$p(N_c) = \frac{1}{N_a}\, e^{-N_c/N_a} \qquad (17)$$

For N_m large, successive codewords in the buffer queue may be considered to have lengths independently selected in accordance with this probability set. The average channel holding time is $N_a/N_m f_c$, and the probability that transmission of w-1 codewords is completed during time τ has the Poisson distribution

$$p_{w-1} = \left(\frac{\tau\, N_m\, f_c}{N_a}\right)^{w-1} \frac{1}{(w-1)!}\, e^{-\frac{\tau N_m f_c}{N_a}} \qquad (18)$$

6.3 Buffer capacity

From the recurrence relations of queueing theory (Molina[39]) it is readily shown that, for a total rate of selection of codewords of $N_m f_s/5$, the probability that a particular codeword is delayed in the buffer for a time τ and then the channel becomes available during the interval dt is

$$p_\tau = \left(1 - \frac{f_s N_a}{5 f_c}\right) \frac{N_m f_s}{5}\, e^{-\left(1 - \frac{f_s N_a}{5 f_c}\right)\frac{\tau N_m f_c}{N_a}}\, dt \qquad (19)$$

Hence the probability of a codeword being delayed greater than t before transmission

$$p(>t) = \frac{N_m f_s}{5}\left(1 - \frac{f_s N_a}{5f_c}\right)\int_t^{\infty} e^{-\frac{tN_m f_c}{N_a}(1 - \frac{f_s N_a}{5f_c})} dt$$

$$= \frac{f_s N_a}{5f_c} e^{-\frac{tN_m f_c}{N_a}(1 - \frac{f_s N_a}{5f_c})} \qquad (20)$$

The average number of codewords stored in the buffer is then

$$N_b = \frac{N_m f_s}{5}\int_0^{\infty} t \frac{d}{dt} p(>t)\ dt$$

$$= \frac{(f_s N_a)^2}{(5f_c - f_s N_a)\ 5f_c} \qquad (21)$$

which is independent of the number of sources multiplexed.

While queue length increases without limit as $({}^{f_c}/N_a)/({}^{f_s}/5) \to 1$, it falls rapidly to practical values as f_c is increased above the minimum permitted by TR_B. From (21), the probability of buffer overflow becomes

$$P_{ov} = \frac{f_s N_a}{5f_c} e^{-N_q(\frac{5f_c}{f_s N_a} - 1)} \qquad (22)$$

Fig. 11 gives representative channel buffer characteristics which allow the selection of an appropriate compromise between attainable compression factor ${}^{f_c}/f_s$ and the required buffer capacity for a range of overflow probabilities. For a 2 to 1 reduction in data rate, $P_{ov} < 10^{-5}$ is obtained with $N_q = 270$.

7 Acknowledgments

The author is indebted to Prof. W.E.J. Farvis of the University of Edinburgh for valuable discussions of this work, and to Prof. S. Michaelson and the staff of Edinburgh Regional Computing Centre for their helpful co-operation. The paper presents material from a part of a Ph.D. dissertation entitled 'The Entropy of Delta-coded Speech' submitted by the author in May 1968 to the Faculty of Science of the University of Edinburgh. The research was supported by the Science Research Council and Ferranti Ltd.

8 References

1 STEELE, R.: 'Delta modulation systems', (Pentech Press, 1975)

2 ABATE, J.E.: 'Linear and adaptive delta modulation', Proc. IEEE, 1967, 55, pp. 298 - 308

3 SCHINDLER, H.R.: 'Delta modulation', IEEE Spectrum, 1970, 7, pp. 69 - 78

4 JAYANT, N.S.: 'Digital coding of speech waveforms: PCM, DPCM, and DM quantizers', Proc. IEEE, 1974, 62, pp. 611 - 632

5 LAANE, R.R. and MURPHY, B.T.: 'Delta modulation codec for telephone transmission and switching applications', Bell Syst. Tech. J., 1970, 49, pp. 1013 - 1031

6 BALDWIN, G.L. and TEWKSBURY, S.K.: 'Linear delta modulator integrated circuit with 17 Mbit/s sampling rate', IEEE Trans., 1974, COM-22, pp. 977 - 985

7 SMITH, B.: 'Instantaneous companding of quantized signals', Bell Syst. Tech. J., 1957, 36, pp. 653 - 709

8 DE JAGER, F.: 'Delta modulation, a method of PCM transmission using the 1 - unit code', Philips Res. Rep., 1952, 7, pp. 442 - 466

9 DELORAINE, E.M., VAN MIERO, S. and DERJAVITCH, B.: 'Méthode et système de transmission par impulsions', French patent No. 932 140, 1946

10 WINKLER, M.R.: 'High information delta modulation', IEEE Int. Conv. Rec., 1963, part 8, pp. 260 - 265

11 JAYANT, N.S.: 'Adaptive delta modulation with a one-bit memory', Bell Syst. Tech. J., 1970, 49, pp. 321 - 342

12 NEWTON, C.M.B.: 'Delta modulation with slope-overload prediction', Electron. Lett., 1970, 6, pp. 272 - 274

13 GREEFKES, J.A. and DE JAGER, F.: 'Continuous delta modulation', Philips Res. Rep., 1968, 23, pp. 233 - 246

14 BROLIN, S.J. and BROWN, J.M.: 'Companded delta modulator for telephone transmission', IEEE Trans., 1968, COM-16, pp. 149 - 157

15 GREEFKES, J.A. and RIEMENS, K.: 'Code modulation with digitally controlled companding for speech transmission', Philips Tech. Rev., 1970, 31, pp. 335 - 353

16 KIKKERT, C.J.: 'Digital techniques in delta modulation', IEEE Trans., 1971, COM-19, pp. 570 - 574

17 MILLER, S.E.: 'Waveguide as a communication medium', Bell Syst. Tech J., 1954, 33, pp. 1209 - 1265

18 GOODMAN, D.J.: 'The application of delta modulation to analog-to-PCM encoding', Bell Syst. Tech. J., 1969, 48, pp. 321 - 343

19 GOODMAN, D.J. and GREENSTEIN, L.J.: 'Quantizing noise of ΔM/PCM encoders', Bell Syst. Tech. J., 1973, 52, pp. 183 - 204

20 BLASBALG, H. and VAN BLERKOM, R.: 'Message compression', IRE Trans., 1962, SET-8, pp. 228 - 238

21 VAN DE WEG, H.: 'Quantizing noise of a single integration delta modulation system with an N-digit code', Philips Res. Rep., 1953, 8, pp. 367 - 385

22 ZETTERBERG, L.H.: 'A comparison between delta and pulse code modulation', Ericsson Technics, 1955, 2, pp. 95 - 154

23 HALIJAK, C.A. and TRIPP, J.S.: 'A deterministic study of delta modulation', IEEE Int. Conv. Rec., 1963, part 8, pp. 247 - 259

24 GOODMAN, D.J.: 'Delta modulation granular quantizing noise', Bell Syst. Tech. J., 1969, 48, pp. 1197 - 1218

25 AARON, M.R., FLEISCHMAN, J.S., MCDONALD, R.A. and PROTONOTARIOS, E.N.: 'Response of delta modulation to Gaussian signals', Bell Syst. Tech. J., 1969, 48, pp. 1167 - 1195

26 IWERSEN, J.E.: 'Calculated quantizing noise of single-integration delta-modulation coders', Bell Syst. Tech. J., 1969, 48, pp. 2359 - 2389

27 GERSHO, A.: 'Stochastic stability of delta modulation', Bell Syst. Tech. J., 1972, 51, pp. 821 - 841

28 SLEPIAN, D.: 'On delta modulation', Bell Syst. Tech. J., 1972, 51, pp. 2101 - 2137

29 GREENSTEIN, L.J.: 'Slope overload noise in linear delta modulators with Gaussian inputs', Bell Syst. Tech. J., 1973, 52, pp. 387 - 422

30 Reading in English of 'The north wind and the sun', (attributed to Aesop, ca. 600 B.C.). Transcriptions and documentation from W. Lawrence, Dept. of Phonetics, Univ. of Edinburgh

31 WANG, P.P.: 'Idle channel noise of delta modulation', IEEE Trans., 1968, COM-16, pp. 737 - 742

32 ELIAS, P.: 'Predictive coding', IRE Trans., 1955, IT-1, pp. 16 - 33

33 WILCOCK, J.D.: 'Semiconductor memories', Electronics and Power, 1974, 20, pp. 8 - 11

34 BRAUN, W., ENGELBRECHT, L., MAGNUSKI, H. and STEEL, F.R.: 'Delta-multiplex for long troposcatter links', IEEE Int. Conv. Rec., 1965, part 2, pp. 62 - 76

35 SHANNON, C.E.: 'The mathematical theory of communication', Bell Syst. Tech. J., 1948, 27, pp. 379 - 423, 623 - 656

36 FANO, R.M.: 'The transmission of information', (MIT Press, 1961)

37 HUFFMAN, D.A.: 'A method for the construction of minimum redundancy codes', Proc. IRE, 1952, 40, pp. 1098 - 1101

38 GOLOMB, S.W.: 'Run-length encodings', IEEE Trans., 1966, IT-12, pp. 369 - 401

39 MOLINA, E.C.: 'Application of the theory of probability to telephone trunking problems', Bell Syst. Tech. J., 1927, 6, pp. 461 - 494