ATR

AUSTRALIAN TELECOMMUNICATION RESEARCH

Volume 13, Number 1, 1979

ATR is published twice a year (in May and November) by the Telecommunication Society of Australia. In addition special issues may be published.
ATR publishes papers relating to research into telecommunications in Australia.

CONTRIBUTIONS: The editors will be pleased to consider papers for publication. Contributions should be addressed to the Secretary, ATR, c/- Telecom Australia Research Laboratories, 22 Winterton Rd., Clayton, Vic., 3168.

RESPONSIBILITY: The Society and the Board of Editors are not responsible for statements made or opinions expressed by authors of articles in this journal.

SUBSCRIPTIONS: Subscriptions for ATR may be placed with the General Secretary, Telecommunication Society of Australia, Box 4050, G.P.O., Melbourne, Victoria, Australia, 3001. The subscription rates are detailed below. All rates are post free. Remittances should be made payable to the Telecommunications Society of Australia, in Australian currency and should yield the full amount free of any bank charges.

The Telecommunication Society of Australia publishes the following:

1. **The Telecommunication Journal of Australia** (3 issues per year)

Subscription — Free to Members of the Society* resident in Australia
Non-members in Australia $6
Non-members or Members Overseas $9

2. **ATR** (2 issues per year)

Subscription — To Members of the Society* resident in Australia $4
Non-members in Australia $9
Non-members or Members Overseas $12
Single Copies — To Members of the Society resident in Australia $3
Non-members within Australia $6
Non-members or Members Overseas $7

* Membership of the Society $3

All overseas copies are sent post-free by surface mail.

Enquiries and Subscriptions for all publications may be addressed to:
The General Secretary, Telecommunication Society of Australia, Box 4050, G.P.O. Melbourne, Victoria, Australia, 3001.

ATR

AUSTRALIAN TELECOMMUNICATION RESEARCH

ISSN 0001-2777

Volume 13, Number 1, 1979

Contents

Challenge . . .

The impact of new technology is very much a fashionable topic at present. It is a topic which is discussed in all walks of life and is no longer the preserve of the technologists and those who are familiar with the issues involved.

The potential of new technology for assisting the developing nations will be discussed at a meeting in Vienna under United Nations' sponsorship in August this year, and papers will be submitted by participating government organisations. It is interesting to note the changing *sponsorship* of new technology and, particularly its application. In the past, this has been almost entirely from the commercial and economic sectors, whereas it is now being sponsored by both the societal and political sectors. This must necessarily raise interesting questions about the future of the application of new technology; if there is to be a societal or political technological push, what will be its commercial base? How will it be funded and who will benefit? Is the societal push due to national social consciences concerning this limit in the developmental scale which nations are trying to appease? Will the *benefits* of new technology be real or illusory? Action which runs counter to the normal commercial trends will surely direct a significant financial penalty to the nations sponsoring the application, but with what benefit to itself or to the intended beneficiary of its munificence? Will technological charity really pay off?

There are benefits to be gained from a technological self help club in which nations can share experiences and the availability of new technology, but introduce new technology at their own rate and under their own conditions, but is there benefit in going much further?

One of the big challenges of the 80's will surely be our handling of the introduction of new technology on the national scene; the issues involved with the international scene will be considerably more complex and will require the exercise of consideration and wisdom.

Kalman Filter Equalization for a Time-Varying Communication Channel

G. NICHOLSON
Telecom Australia Research Laboratories

J. P. NORTON
Department of Electrical Engineering
University of Tasmania

This paper investigates the application of the discrete Kalman filter to equalization of time-varying digital communication channels. The resultant equalizer, called the message estimator, yields an unconditionally unbiased, linear, minimum mean-square error estimate of the message sequence. The message estimator models the channel as a vector, whose elements are the sampled channel impulse response. An adaptive Kalman filter is used in a decision-directed arrangement to estimate on line the channel vector, and so adapt the message estimator to a time-varying channel.

The relation between mean-square error and error rate for the message estimator is discussed. The message estimates are shown to have non-zero conditional bias for individual symbols and non-zero residual intersymbol interference, both of which can be calculated from the covariance of the message estimates produced by the equalizer. Computer simulation of the proposed adaptive message estimator indicates that it should be particularly effective when the channel is rapidly time-varying or the training period of the equalizer is restricted.

1. INTRODUCTION

A digital transmission system can be modelled for equalization purposes by an equivalent baseband channel (Fig.1). Time-domain equalization techniques make use of the discrete-time nature of the channel, combining received sample values to obtain an estimate of the transmitted message sequence. A large portion of the communication theory literature has addressed itself to the equalization problem (Ref.1). The most widely used form of equalizer is a tapped delay line (TDL) whose tap weights are adjusted to minimize a performance criterion involving the equalizer output. Proakis and Miller (Ref.2) employ a steepest-descent gradient method to adjust the tap weights. The method is easily realisable and requires very little computing, although convergence of the tap weights is slow. Two TDL filters can be used in a decision feedback arrangement with a threshold decision device, to often achieve significant improvements over linear TDL

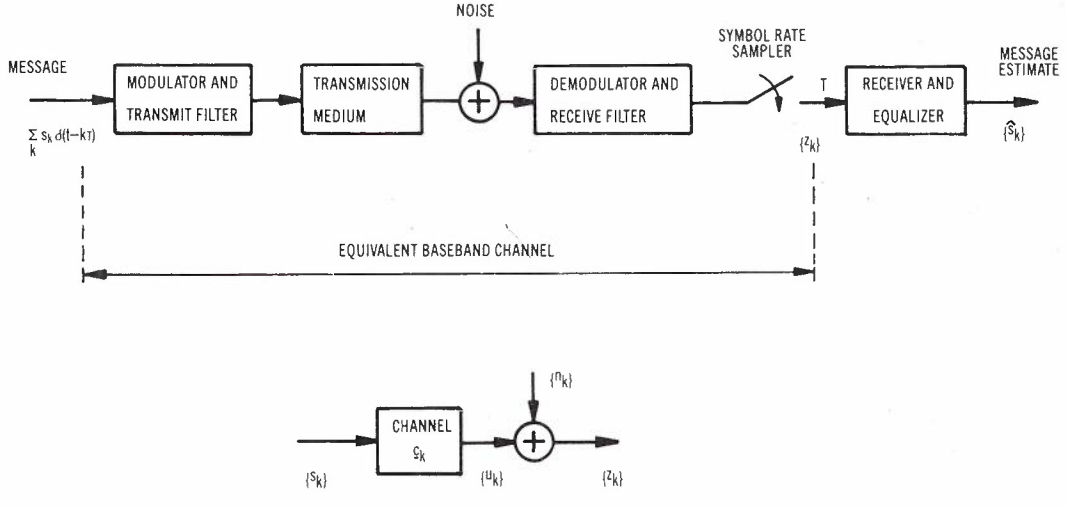


Fig.1 - Digital communication system.

equalizers. The steepest-descent method of tap weight adjustment can be similarly applied to this decision feedback equalizer.

Whilst linear TDL and decision feedback equalizers have dominated the applications of equalizers to digital data transmission, with the advent of cheap digital processing capabilities and the need to rapidly equalize data modems operating in the switched telephone network or multipoint leased networks, the use of more computationally-orientated equalization schemes has become attractive. Computational techniques offer much better convergence rates and error rates, at the expense of more complexity. At one extreme, the Viterbi algorithm (Ref.3) provides a conceptually simple computational estimator for signal sequences transmitted over dispersive and noisy channels, but its computing demands are excessive. The algorithm has a lower error rate than linear TDL or decision feedback equalizers, but cannot readily adapt to a time-varying channel.

Alternatively, the problem of equalizing a digital communication channel in the face of intersymbol interference and noise can be framed as a linear state-estimation problem for which the Kalman filter provides a recursive solution. Lawrence and Kaufman (Ref.4) and other (Refs. 5,6) have taken this view. The approach is pursued in this paper to yield a minimum mean-square error (MSE) estimator of the transmitted message sequence. A second, adaptive Kalman filter is used in a decision-directed arrangement to estimate the sampled channel impulse response and so recursively adapt the message estimator to a time-varying channel (Ref.7).

The adaptive message estimator is compared in performance with linear TDL and decision feedback equalizers, whose tap weights are adjusted by either a steepest-descent method (Ref.2) or using recursive least squares (Refs.7,8). A steepest-descent equalizer requires much less computing for each received sample than the adaptive message estimator. The equations for the recursive least-squares equalizer are similar to those for a Kalman filter, so that its complexity is comparable to that of the adaptive message estimator.

The major advantages of the adaptive message estimator are its fast initial convergence in training and good adaptation properties. Several equalization schemes have been reported which claim fast convergence of the tap weights. Gitlin and Magee (Ref.9) have investigated many of these schemes and concluded that Godard's algorithm (Ref,8) (i.e. a recursive least-squares equalizer) has the quickest convergence. The equalizers investigated in this paper are based on data values sampled at the baud rate. A recent development in equalizer design, fractional-tap spacing (Ref.10), has not been considered, although the convergence speed of such equalizers has been shown to be less sensitive to timing-phase errors.

2. THE CHANNEL MODEL AND EQUALIZER

Synchronous baseband digital transmission may be represented by a finite-memory time-varying channel with sampled impulse response at time kT

$$\underline{c}_k^T = [c_0 c_1 \ldots c_{N-1}]_k \quad , \tag{1}$$

operating on an input message sequence $\{s_k\}$. The channel output z_k at sample time kT, affected by noise n_k of variance σ^2, is

$$z_k = \underline{c}_k^T . \underline{x}_k + n_k \quad , \tag{2}$$

where $\underline{x}_k$ is a state vector composed of N successive signal symbols

$$\underline{x}_k^T = [s_k s_{k-1} \ldots s_{k-N+1}] \quad . \tag{3}$$

The noise is usually assumed to be zero-mean, white, Gaussian, stationary and uncorrelated with the input. The zero-mean unity-variance input sequence $\{s_k\}$ is taken as binary and uncorrelated, although the results of this paper are easily extended to a multilevel uncorrelated input sequence. The channel vector $\underline{c}_k$ is normalized so that the signal-to-noise ratio (SNR) of the channel is $1/\sigma^2$. The input sequence may be described by a state equation

$$\underline{x}_{k+1} = \Psi . \underline{x}_k + \underline{\gamma} . s_{k+1} \quad , \tag{4}$$

where Ψ is the NxN unit downward shift matrix

$$\Psi = \begin{bmatrix} 0 & . & . & . & . & 0 \\ 1 & 0 & . & . & . & . \\ 0 & 1 & 0 & . & . & . \\ . & 0 & 1 & 0 & . & . \\ . & . & 0 & 1 & 0 & . \\ . & . & . & 0 & 1 & 0 \end{bmatrix}$$

and

$$\underline{\gamma} = [\, 1 \; 0 \; . \; . \; . \; 0 \,]^T .$$

A discrete Kalman filter based on the observation equation (2) and state equation (4) gives an unbiased minimum-variance linear estimate $\hat{\underline{x}}_k$, the last element of which is the minimum MSE estimate of s_{k-N+1}. On receiving the sampled output z_k, the filter computes

$$\hat{\underline{x}}_k = \Psi . \hat{\underline{x}}_{k-1} + \underline{g}_k . (z_k - \underline{c}_k^T . \Psi . \hat{\underline{x}}_{k-1}), \tag{5a}$$

where

$$\underline{g}_k = P_{k|k-1} . \underline{c}_k \; / \; (\underline{c}_k^T . P_{k|k-1} . \underline{c}_k + \sigma^2), \tag{5b}$$

$$P_{k|k-1} = \Psi . P_{k-1|k-1} . \Psi^T + \underline{\gamma} . \underline{\gamma}^T \qquad (5c)$$

and subsequently

$$P_{k|k} = P_{k|k-1} - \underline{g}_k . \underline{c}_k^T . P_{k|k-1} \qquad (5d)$$

Here $P_{i|j}$ is the error covariance matrix of state estimate $\hat{\underline{x}}_i$ based on observations up to time jT, i.e.

$$P_{i|j} = E\left[(\underline{x}_i - \hat{\underline{x}}_{i|j})(\underline{x}_i - \hat{\underline{x}}_{i|j})^T\right].$$

This equalization scheme was first considered by Lawrence and Kaufman (Ref.4). More recently, Benedetto and Biglieri (Ref.5) have discussed the Kalman-filter equalizer in detail and compared it with tapped-delay-line equalizers. They show that the system defined by (2) and (4) with $\underline{c}_k$ fixed is uniformly completely controllable and observable. Provided $P_{0|0} > 0$, the filter (5) is then uniformly asymptotically stable.

The structure of the equalizer outlined above is shown in Fig.2. A threshold device operates on the last element $\hat{x}_N = \hat{x}_{k-N+1|k}$ of $\hat{\underline{x}}_{k|k}$ to give a binary estimate of the message symbol s_{k-N+1}, + 1 if $\hat{x}_N$ is positive and -1 otherwise. There is a delay of N-1 samples in producing the final estimate of s_{k-N+1}.

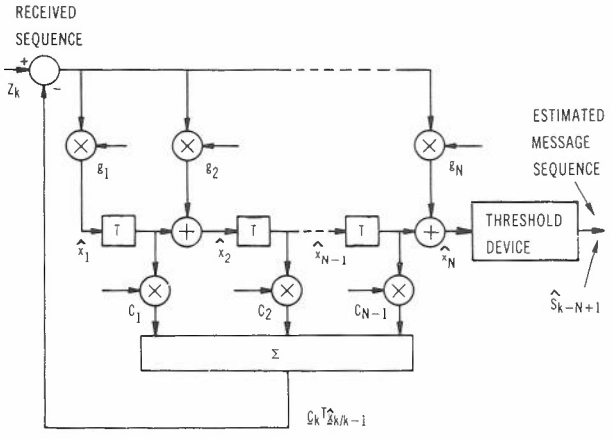


Fig.2 - The message estimator.

In practice the channel coefficients $\underline{c}_k$ and noise variance σ^2 required in (5) are not known accurately, and will usually be time-varying. An estimator of $\underline{c}_k$ is required to track the time variations. It will be assumed that σ^2 is known. At realistic SNR the performance of the message estimator (5) is relatively insensitive to the assumed value of σ^2 (Ref.7).

The channel coefficients $\underline{c}_k$ will be estimated by an adaptive Kalman filter in a decision-directed arrangement with the message estimator (Fig. 3). This method, but without adapting the Kalman filter, was recently proposed for quadrature phase-shift keyed systems by Lee and Cunningham (Ref.6).

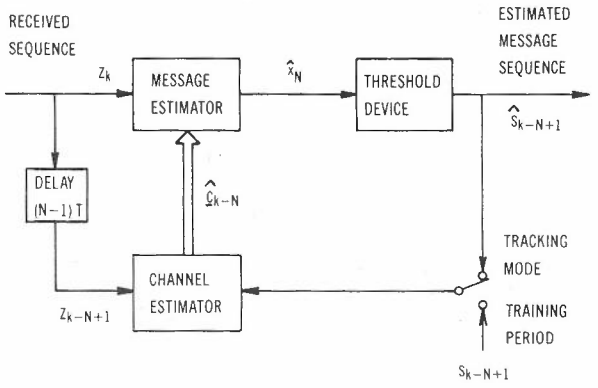


Fig.3 - The adaptive message estimator.

Lawrence and Kaufman (Ref.4) suggest an extended Kalman filter to estimate $\underline{x}_k$ and $\underline{c}_k$ simultaneously, augmenting the state vector to include the channel coefficients. Since the observation equation (2) is nonlinear in $\underline{x}_k$ and $\underline{c}_k$, it is linearized about nominal values for $\underline{x}_k$ and $\underline{c}_k$. This method requires approximately twice the computation per step of the decision-directed method proposed here. Furthermore it is unsuitable for adapting to a rapidly time-varying channel because of the local linearization required.

A variety of identification techniques could have been implemented for the channel estimator. The simplest to use is a steepest-descent gradient method. However, for fast convergence of the equalizer a more elaborate technique must be used and for this reason an adaptive Kalman filter was chosen. Kalman filters are well known for their ability to adapt to time-varying systems. The method of adapting the estimator of $\underline{c}_k$ is discussed in Section 4.

When fast initial convergence is not important, and the time-variations of the channel are slow, a more attractive scheme may be that proposed by Luvison and Pirani (Ref.11). Both the channel vector $\underline{c}_k$ and gain $\underline{g}_k$ for the message estimator equation (5a) are estimated by steepest-descent methods. The advantage is that the amount of computing increases only linearly with N. They report convergence times for the gain $\underline{g}_k$ similar to those for the tap weights in a steepest-descent equalizer (Ref.2).

The Kalman-filter channel estimator represents time variation of the channel by

$$\underline{c}_{k+1} = \underline{c}_k + \underline{w}_{k+1} , \qquad (6)$$

where the covariance Q_k of the "noise" $\underline{w}_k$ is chosen to reflect the expected size of the changes in channel response between samples. The observation equation for $\underline{c}_k$ is taken at time kT as

$$z_k = \underline{r}_k^T . \underline{c}_k + n_k \qquad (7)$$

In the training period the input sequence is known and $\underline{r}_k^T = [s_k\ s_{k-1}\ ..\ s_{k-N+1}]$. In the tracking mode $\underline{r}_k$ consists of the binary estimates $\hat{s}_k$ to $\hat{s}_{k-N+1}$, all assumed correct. In practice, the system error rate must be small enough for any incorrect message estimates in $\underline{r}_k$ to have a negligible effect when estimating $\underline{c}_k$. This assumption usually presents no problem as error rates for data communication are typically less than 10^{-4}. A training period is required initially to establish an error rate of this order. The channel-response estimate will be N-1 samples in arrears to allow for the decision delay in estimating the transmitted sequence. The extra error due to this delay will not be significant unless the delay is a substantial proportion of the convergence time of the equalizer.

After each updating of $\underline{r}_k$, a new estimate of the channel is made, namely

$$\hat{\underline{c}}_k = \hat{\underline{c}}_{k-1} + \underline{g}_k^c \cdot \nu_k \tag{8a}$$

where the innovation (one-step-ahead prediction error) is

$$\nu_k = z_k - \underline{r}_k^T \cdot \hat{\underline{c}}_{k-1} \tag{8b}$$

and

$$\underline{g}_k^c = P^c_{k|k-1} \cdot \underline{r}_k \,/\, (\underline{r}_k^T \cdot P^c_{k|k-1} \cdot \underline{r}_k + \sigma^2) \tag{8c}$$

$$P^c_{k|k-1} = P^c_{k-1|k-1} + Q_k \tag{8d}$$

$$P^c_{k|k} = P^c_{k|k-1} - \underline{g}_k^c \cdot \underline{r}_k^T \cdot P^c_{k|k-1} \tag{8e}$$

As the message estimator (5) uses an estimate of the channel response from (8), replacing $\underline{c}_k$ by $\hat{\underline{c}}_{k-N}$ in (5), $\hat{\underline{x}}_k$ is no longer exactly optimal. However, the insensitivity of the message estimator performance to σ^2 and the whiteness of the observation noise encourage the belief that small errors in $\hat{\underline{c}}_k$ will not affect $\hat{\underline{x}}_k$ significantly.

3. ERROR RATE PERFORMANCE

A MSE performance criterion was used to derive the message estimator. For digital communication systems an accepted measure of performance is the probability of error P_E, which is the probability that a particular message symbol s_k is incorrectly estimated at the receiver. The relation between MSE and the probability of error depends on the amplitude probability distribution of the message estimate $\hat{x}_N$. This relationship is difficult to establish by computer simulation at typical SNR's, because the low error rates imply impractically long runs.

The error covariance matrix $P_{k|k}$, calculated recursively by the message estimator, can be used to provide information on the amplitude probability distribution of $\hat{x}_N$. Consider the message estimator at time kT, based on a known and time-invariant channel $\underline{c}_k$ modelled exactly in (2). Denoting the element in row i and column j of the NxN matrix $P_{k|k}$ by $[P_{k|k}]_{ij}$, if $P_{k|k}$ is the true error covariance then

$$E\left[(s_{k-N+1} - \hat{x}_N)^2\right] = [P_{k|k}]_{NN}. \tag{9}$$

In general, if the vector $\underline{c}_k$ is a good representation of the channel, the calculated $P_{k|k}$ given by (5d) should be a good approximation to the actual error covariance. An estimate of the MSE of $\hat{x}_N$ is provided by the last principal-diagonal element of $P_{k|k}$.

The error covariance $P_{k|k}$ can be used to give further information on the amplitude probability distribution of $\hat{x}_N$. As $\hat{x}_N$, i.e. $\hat{x}_{k-N+1|k}$, is a linear function of received samples $\{z_k\}$, themselves linear functions of the transmitted symbols, and additive noise, it can be written as

$$\hat{x}_N = \hat{x}_{k-N+1|k} = \sum_{m=0}^{k-1} b_{k-m}\, s_{k-m} + \sum_{m=0}^{k-1} a_{k-m}\, n_{k-m}, \tag{10}$$

taking the initial state estimate $\hat{\underline{x}}_{0|0}$ as zero. The coefficients b_{k-m}, which are independent of the message sequence $\{s_k\}$ and noise $\{n_k\}$, represent the residual intersymbol interference affecting message estimate $\hat{x}_N$ after equalization. It can be shown (Ref.12) that

$$b_{k-i} = -[P_{k|k}]_{N,i+1} \tag{11a}$$

for $i = 0,1,2,.....,N-2$ and

$$b_{k-N+1} = 1 - [P_{k|k}]_{NN}. \tag{11b}$$

The second term in (10) is the noise in $\hat{x}_N$ attributable to the channel noise $\{n_k\}$. As a conse-

quence of (11b), the bias of the message estimate at time kT is given by $[P_{k|k}]_{NN}\, s_{k-N+1}$. From (9), the bias is equal in magnitude to the MSE of the message estimate.

The error covariance of the message estimator tends to a nonzero limit as the number of received samples increases, because the new message symbol acts as process noise driving the state equation (4). This is true even for no noise $\{n_k\}$. The effects of both bias and residual intersymbol interference on the amplitude probability distribution of $\hat{x}_N$ are significant at typical SNR's, (Refs. 7,12). A further implication from (9) is that the MSE approaches a nonzero limit as the SNR increases, although the probability of error approaches zero, so the MSE is not always a good indicator of the probability of error.

The probability of error for a time-invariant message estimator was calculated for the example channel of Fig.4. The channel impulse response is the impulse response of an equivalent baseband channel, based on the average measurements reported in Ref.13 for the Bell System switched telecommunications network. DSB-AM transmission at 2400 baud is assumed for a carrier frequency of 1700 Hz and a transmit filter with a 50% raised-cosine-rolloff amplitude characteristics. The sampled channel impulse response is modelled by nine terms in $\underline{c}$, with terms less than 0.02 being ignored.

The probability of error was calculated directly from the linear description (10) of $\hat{x}_N$. The message estimator performance shown in Fig.5 is compared with three curves for time-invariant TDL equalizers. The configuration of each TDL equalizer is recorded as (n,l,D), where n is the total number of feedforward and feedback taps, l is the number of feedback taps and D is the decision delay in sample intervals. The decision delay D has been chosen to make the probability of error close to the minimum possible over the range of SNR, given a fixed number of equalizer taps. The TDL equalizers are of similar complexity to the message estimator. Examining Fig.5, the message estimator compares favourably with a transversal equalizer, but has a significantly higher probability of error than a decision feedback equalizer.

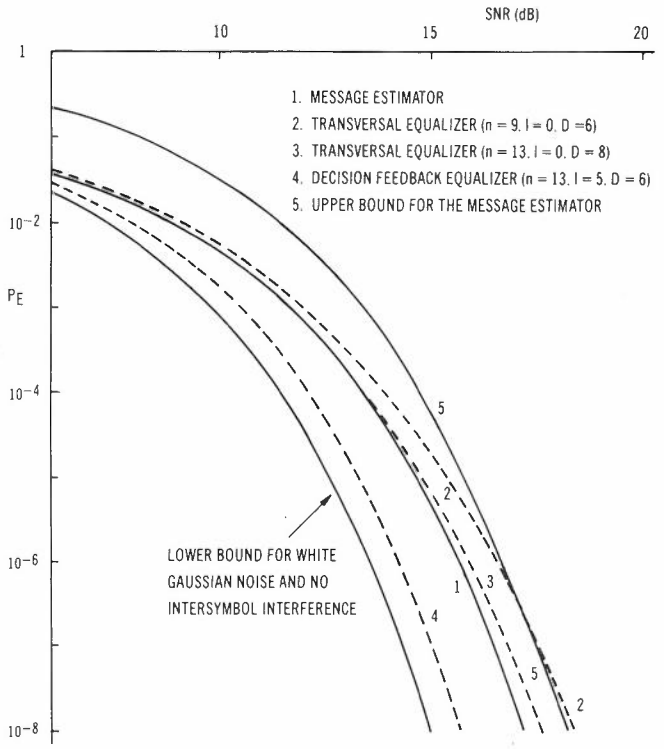


Fig.5 - Probability of error against SNR.

An upper bound on the probability of error for the message estimator, which is simple to determine and only uses information available from the error covariance matrix, can be found by applying Chernoff inequalities. The bound is based on the upper bound developed by Saltzberg (Ref. 14). The bound, a derivation of which is given in (Ref.7), is

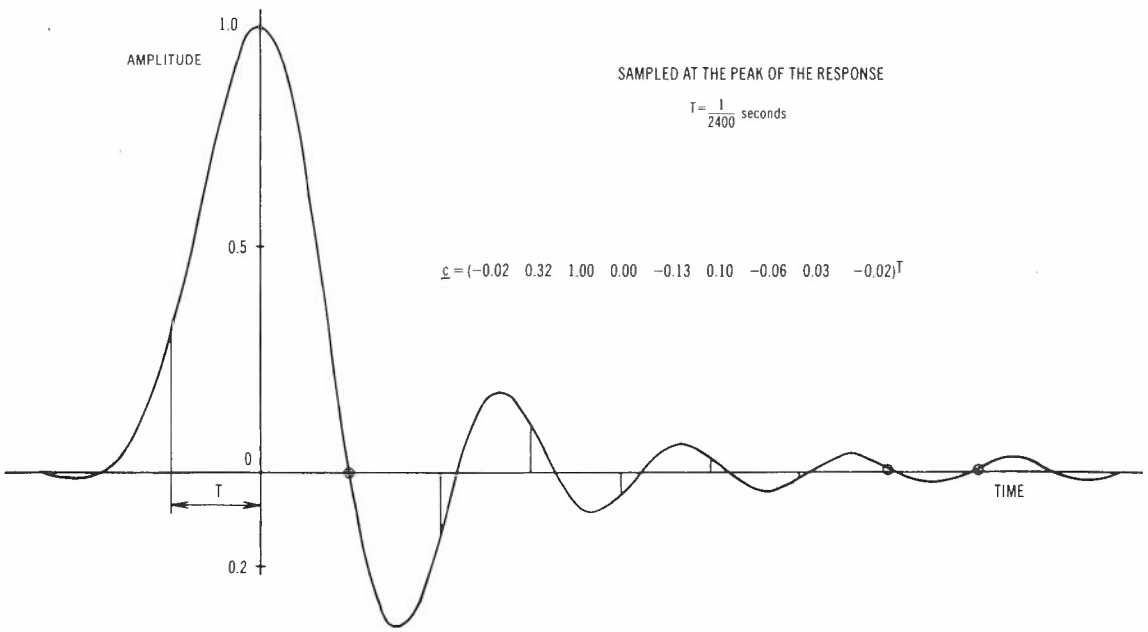


Fig.4 - The channel impulse response.

$$P_E < \exp\left[-\frac{(1-[P_{k|k}]_{NN} - \sum_{i\varepsilon U_N} |b_i|)^2}{2([P_{k|k}]_{NN} - [P_{k|k}]^2_{NN} - \sum_{i\varepsilon U_N} b_i^2)}\right], \qquad (12)$$

where U_N is the set of coefficients b_i ($k \leq i \leq k-N+2$) minimizing (12). The method for obtaining the set U_N is the same as that described in Saltzberg's paper (Ref.14) for the set U. It relies on testing the b_i ($k \leq i \leq k-N+2$) in order of decreasing magnitude.

The upper bound is compared with the message estimator performance in Fig.5. The bound is about an order of magnitude larger than the actual probability of error, the worst discrepancy being at the highest SNR. Although this may be too pessimistic in some cases, it does provide a quick indication of performance. Tighter upper bounds for the probability of error have been reported in the communication theory literature (Ref.15). However, these bounds are sometimes more involved to calculate and moreover require knowledge of all the significant b_i coefficients and the noise power in (10), whereas the upper bound (12) only uses the b_i coefficients ($k \leq i \leq k-N+2$) available from $P_{k|k}$.

4. ADAPTING THE EQUALIZER

The filter equations (8) take account of changes in the channel impulse response from sample to sample through the covariance Q_k. For much of the time, the channel will be relatively time-invariant and a value of zero is appropriate. During abrupt changes a non-zero value of Q_k, chosen if possible from previous experience of channel behaviour, must be used. If Q_k is kept too small or zero, filter divergence is possible through the computed error covariance $P^c_{k|k}$ becoming unrealistically small. As a result the gain $\underline{g}^c_k$ is too small and insufficient weight is given to the innovation v_k in (8a), despite $\hat{\underline{c}}_{k-1}$ being a poor estimate. Divergence may also result from model deficiencies such as truncation of the channel impulse response, a poor value for σ^2 or an oversimplified noise model. Short-word-length computation may contribute additional inaccuracy. A mechanism for adjusting the filter to prevent divergence and accommodate channel time-variations without unduly impairing the steady-state performance is therefore needed.

Leondes and Pearson (Ref.16) classify methods for improving Kalman filter performance in an uncertain environment into bounding techniques, multiple-criteria techniques and adaptive estimation. Bounding techniques attempt to lower the estimation error below an acceptable value without necessarily minimizing it. Multiple-criteria techniques design the filter on a performance index which includes both estimation error and filter sensitivity. Adaptive estimation involves either simultaneous estimation of the uncertain parameters and the state, or adjustment of the process-noise covariance, error covariance or filter gain to compensate for modelling errors. Adaptive estimation has the advantage over bounding and multiple-criteria techniques that it is designed for minimum estimation error.

In a survey Mehra (Ref.17) divides the methods of adaptive estimation into four categories: Bayesian, maximum likelihood, correlation and covariance matching. Of these, covariance matching is chosen for the channel estimator because it avoids specifying which parameters of the model are in error, and what their prior probability distributions are, by lumping all errors into the process-noise covariance or error covariance of the filter. The basic idea behind covarience-matching techniques is to make the filter residuals consistent with their theoretical covariances.

Consider the mean-square value of the filter innovation v_k at time kT, from (8b)

$$E[v_k^2] = E[(z_k - \underline{r}_k^T \hat{\underline{c}}_{k-1})^2]$$

$$= E[(\underline{x}_k^T \underline{c}_k + n_k - \underline{r}_k^T \hat{\underline{c}}_{k-1})^2].$$

If all the estimated message symbols in $\underline{r}_k$ are correct,

$$E[v_k^2] = E[(\underline{r}_k^T \underline{c}_k - \underline{r}_k^T \hat{\underline{c}}_{k-1} + n_k)^2]$$

$$= E[\underline{r}_k^T (\underline{c}_k - \hat{\underline{c}}_{k-1})(\underline{c}_k - \hat{\underline{c}}_{k-1})^T \underline{r}_k]$$

$$+ 2\,E[n_k \underline{r}_k^T (\underline{c}_k - \hat{\underline{c}}_{k-1})] + E[n_k^2].$$

Since $\{n_k\}$ and $\{s_k\}$ are independent sequences,

$$E[v_k^2] = E[\underline{r}_k^T \tilde{\underline{c}}_{k|k-1} \tilde{\underline{c}}^T_{k|k-1} \underline{r}_k] + \sigma^2, \qquad (13)$$

where the estimation error is $\tilde{\underline{c}}_{k|k-1} = \underline{c}_k - \hat{\underline{c}}_{k-1}$.

An estimate of $\tilde{\underline{c}}_{k|k-1} \tilde{\underline{c}}^T_{k|k-1}$ is the computed error covariance $P^c_{k|k-1}$ from the channel estimator. The denominator of the gain equation (8c) provides an estimate of the innovation variance (13). This property is the basis of covariance matching for adaptive Kalman filters. The process-noise covariance Q_k, or the error covariance $P^c_{k|k-1}$, is adjusted to produce consistency between the actual variance of v_k and the variance predicted from the denominator of the gain equa-

tion (8c). In this manner filter divergence is prevented and the channel estimator can adapt to a time-varying channel.

Representative covariance-matching techniques have been presented by Jazwinski (Ref.18), Nahi and Schaefer (Ref.19), and Quigley (Ref.20). Jazwinski computes a positive-definite diagonal Q_k at each step, determined to produce consistency between the innovations and their statistics. To enhance the statistical significance of the estimate of Q_k, a smoothing operation can be performed. Nahi and Schaefer increase or reduce $P^c_{k|k-1}$ when ν_k^2 is beyond lower or upper thresholds. The thresholds are chosen according to a Neyman-Pearson criterion, based on the predicted innovation variance. However, this technique is not applicable as it stands if the probability of the channel changing in a given sample interval is small. Quigley merely increases Q_k by a fixed amount whenever ν_k^2 is more than a specified multiple of the variance $\underline{r}_k^T \cdot P^c_{k|k-1} \cdot \underline{r}_k + \sigma^2$ given by the filter equation (8c).

Of these three decision-directed methods, that of Quigley was chosen for its simplicity. Computer simulation on time-varying channels shows that the method performs adequately. At sample instant kT, the adjustment algorithm examines

$$J_k = \nu_k^2 \,/\, (\underline{r}_k^T \cdot P^c_{k|k-1} \cdot \underline{r}_k + \sigma^2) \qquad (14)$$

and sets Q_k to q1 instead of zero in the covariance-updating equation (8d) if J_k exceeds a threshold T_H. The threshold may be chosen to give a specified probability of adding q1 unnecessarily when the filter is accurate, if the distribution of ν_k^2 is assumed. Alternatively it is adjusted empirically, as is q. With a sensible choice of q and T_H, the adaptive channel estimator can track a slowly time-varying channel with MSE close to the minimum possible. Yet, because of the decision-directed nature of Quigley's method, it can rapidly adapt to abrupt time-variations. In the following results the threshold is 5, corresponding to a 2.5% probability of increasing $P^c_{k|k-1}$ unnecessarily.

5. RESULTS OF EQUALIZER SIMULATION

Some results of computer simulation of the adaptive message estimator are reported in this section. Emphasis is on the initial convergence, and adaptation to a time-varying channel. A more comprehensive set of results for the adaptive message estimator is given in Ref.7. The results are for a SNR of 30 dB, a typical figure on the switched telephone network (Ref.13), and the channel vector of Fig.4. The MSE of the message estimates is used as a performance measure of initial convergence and adaptation of the equalizer. For this particular channel the MSE has been shown to be a good indication of error rate (Ref.7).

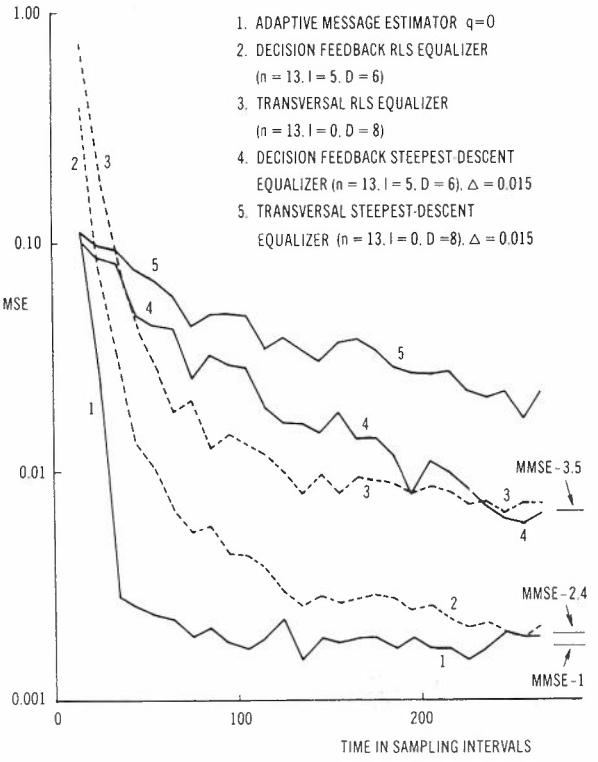


Fig.6 - Initial convergence.

Figure 6 shows the initial MSE convergence in training of the adaptive message estimator. The initial conditions are $\hat{\underline{x}}_{0|0} = 0$, $P_{0|0} = I$, $P^c_{0|0} = I$ and $\hat{\underline{c}}_0 = [0....010....0]^T$, where the one corresponds to the largest coefficient in the sampled channel impulse response. This reflects little prior knowledge of the impulse response other than the length N for modelling and the position of the peak of the impulse response. In practice, system measurements may provide a better initial estimate of the channel, and so improve the initial convergence over that reported here. Since transmission starts at time T, the channel estimator does not begin estimating the channel vector until time NT.

The initial convergence is compared with that of a TDL equalizer, adjusted by recursive least squares (Refs.7,8) or a steepest-descent method (Ref.2). Recursive least squares (RLS) for these TDL configurations is comparable with the adaptive message estimator in the amount of computing per sample interval. The results in Fig.6 were obtained by averaging the MSE of 25 independent runs at intervals of 10 samples. In comparing results for different equalizers, a similar MSE does not necessarily imply a similar error rate. The minimum mean-square error (MMSE) of each equalizer configuration is shown on the figure. The adaptive message estimator converges faster than either the RLS equalizer or the steepest-descent equalizer. If convergence to within about 5% of the minimum MSE is considered, the adaptive message estimator has converged in 90 sample intervals, less than half the time taken by the RLS equalizers. The steepest-descent equalizer performs poorly in comparison with the other equalization schemes.

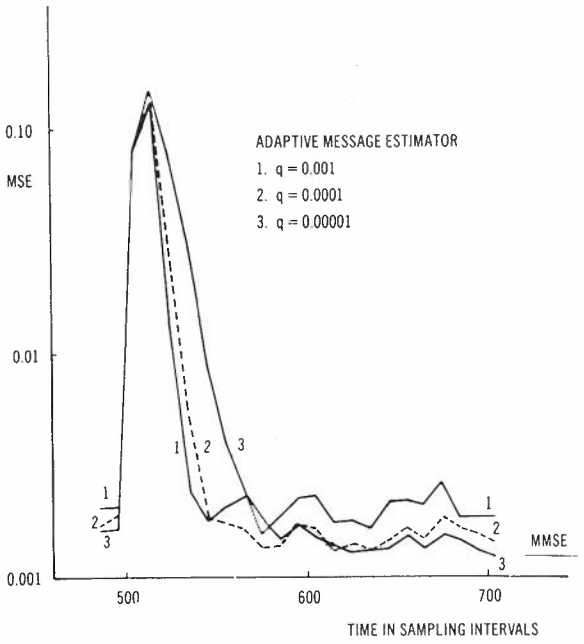


Fig.7 - Recovery after a shift of 0.2T in the sampling instant.

Figure 7 shows the effect on MSE of a change in the sampled channel impulse response, after 500 sample intervals in the training period, equivalent to a sudden shift of 0.2T in the sampling instant. This example illustrates the adaptive message estimator's ability to recover from abrupt changes in the channel vector. The recovery time decreases with increasing q, but the ultimate MSE increases. Selection of q represents a trade-off between these two factors. The range of q tested suggests that the equalizer recovery changes only slowly with the choice of q. A q of 10^{-4} offers the best compromise of the three values used, with convergence to within 5% of the minimum MSE some 80 sample intervals after the timing shift.

To investigate the sensitivity of the adaptive message estimator's performance to q, the channel model is simulated with timing jitter. The timing jiter is modelled by a random walk, with reflecting barriers, in sampling the channel impulse response to give the channel vector (Ref.7). The timing jitter is constrained within limits to avoid a loss of synchronization of the received data $\{z_k\}$. In practice timing jitter is dealt with by a suitable timing recovery circuit and does not usually present a serious impairment to data transmission. However, it is suitable for testing the equalizer's performance when there are frequent, small changes in the channel vector. The results of the simulation with timing jitter as well as those for a time-invariant channel (Fig.8), confirm that the MSE of the adaptive message estimator is relatively insensitive to the choice of q. With timing jitter, there is an increase of about 10% in MSE for q an order of magnitude different from the optimum q.

6. CONCLUSIONS

The discrete Kalman filter has been applied to equalization of time-varying communication channels. A message estimator is used to yield an unbiased linear minimum-MSE estimate of the message sequence $\{s_k\}$. An adaptive Kalman filter in a decision-directed arrangement estimates the channel vector $\underset{\sim}{c}_k$ for use by the message estimator. Any changes in $\underset{\sim}{c}_k$ are modelled as a white process noise, whose covariance is adjusted to produce consistency between the observed innovations and their variance as computed by the filter. This adaptive structure has the advantage that the equalizer can respond quickly to channel time-variations without the penalty of a greatly increased MSE.

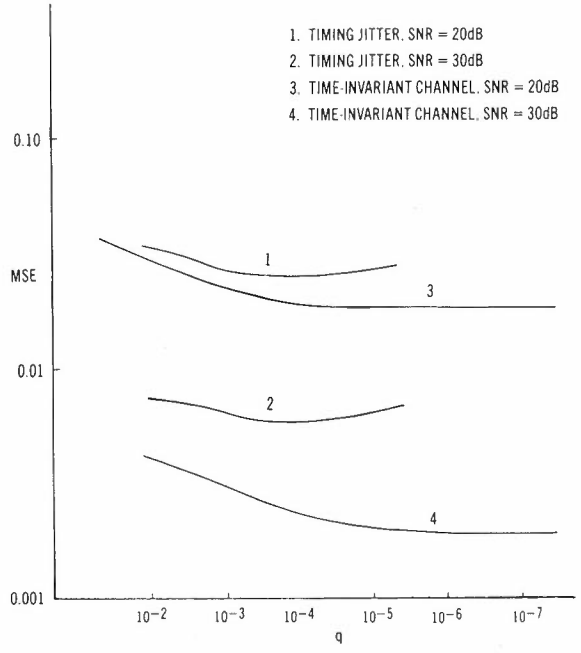


Fig.8 - MSE of the adaptive message estimator against q.

The message symbol estimate $\hat{x}_N$ has been related, through the error covariance matrix, to the message symbols and noise samples, allowing bias, intersymbol interference and an upper bound on the probability of error to be calculated. The probability of error for the message estimator compares favourably with a transversal equalizer, but is significantly higher than that for a decision feedback equalizer.

The performance of the adaptive message estimator has been compared with that of recursive least-squares and steepest-descent equalizers on an example channel. The adaptive message estimator initially converges in about half the time taken by the recursive least-squares equalizer, both equalizers being significantly quicker than the steepest-descent equalizer. The adaptive message estimator has a fast recovery time from significant time-variations in the channel. Its recovery time, and MSE when the channel is time-invariant, are both relatively insensitive to the covariance of the process noise Q_k used in the channel estimator. These conclusions, based on one example channel, are supported by computer simulation on other channels (Ref.7). The convergence rate of the adaptive message estimator, and of the other equalizers, depends to a large extent on the particular channel.

The number of numerical operations per sample interval for the adaptive message estimator is

dominated by a term in N^2, where N is the number of channel coefficients being estimated. The recursive least-squares equalizer has a similar amount of computation. In comparison, the amount of computation increases linearly with the number of taps for the steepest-descent equalizer. A topic of future research is the possible application to the message estimator of the 'fast Kalman estimation' techniques recently developed by Falconer and Ljung (Ref.21). These techniques offer the same convergence rate as conventional Kalman filter implementation, but with fewer numerical operations per sample interval.

In summary, the adaptive message estimator is suitable when the channel is rapidly time-varying or a short training period is essential, therefore justifying the amount of computation required.

7. ACKNOWLEDGEMENT

The work described was supported by a grant from the Radio Research Board.

8. REFERENCES

1. Lucky, R.W., "A Survey of the Communication Theory Literature : 1968-1973", IEEE Trans. Inform. Theory, Vol.IT-19, November 1973, pp.725-739.

2. Proakis, J.G. and Miller, J.H., "An Adaptive Receiver for Digital Signalling through Channels with Intersymbol Interference", IEEE Trans. Inform. Theory, Vol.IT-15, July 1969, pp.484-497.

3. Forney, G.D., Jr., "Maximum-likelihood Sequence Estimation of Digital Sequences in the Presence of Intersymbol Interference", IEEE Trans. Inform. Theory, Vol.IT-18, May 1972, pp.363-378.

4. Lawrence, R.E. and Kaufman, H., "The Kalman Filter for the Equalization of a Digital Communications Channel", IEEE Trans. Commun. Tech., Vol.COM-19, December 1971, pp.1137-1141.

5. Benedetto, S. and Biglieri, E., "On Linear Receivers for Digital Transmission Systems", IEEE Trans. Comm., Vol.COM-22, September 1974, pp.1205-1215.

6. Lee, T.S. and Cunningham, D.R., "Kalman Filter Equalization for QPSK Communications", IEEE Trans. Comm., Vol.COM-24, March 1976, pp.361-364.

7. Nicholson, G., "Recursive Estimation in Time-varying Communication Systems", Master of Engineering Science Thesis, University of Tasmania, March 1978.

8. Godard, D., "Channel Equalization using a Kalman Filter for Fast Data Transmission", IBM J. Res. Develop., Vol.18, May 1974, pp.267-273.

9. Gitlin, R.D. and Magee, F.R., Jr., "Self-orthogonalizing Adaptive Equalization Algorithms", IEEE Trans. Comm., Vol.COM-25, July 1977, pp.666-672.

10. Ungerboeck, G., "Fractional Tap-spacing Equalizer and Consequences for Clock Recovery in Data Modems", IEEE Trans. Comm., Vol.COM-24, August 1976, pp.856-864.

11. Luvison, A. and Pirani, G., "A State-variable Approach to the Equalization of Multilevel Digital Signals", CSELT Rapporti Tecnici, No.3, April 1974, pp.3-22.

12. Nicholson, G. and Norton, J.P., "Bias and Residual Intersymbol interference of Minimum-variance Equalisers for Digital Communication", Electronics Letters, Vol.14, No.20, 28th September 1978, pp.678-679.

13. Duffy, F.P. and Thatcher, T.W., Jr., "1969-1970 Connection Survey : Analog Transmission Performance on the Switched Telecommunications Network", Bell System Tech.J., Vol.50, April 1971, pp.1311-1347.

14. Saltzberg, B.R., "Intersymbol Interference Error Bounds with Application to Ideal Band-limited Signalling", IEEE Trans. Inform. Theory, Vol.IT-14, July 1968, pp.563-568.

15. Jenq, Y.C., Liu, B. and Thomas, J.B., "Probability of Error in PAM Systems with Intersymbol Interference and Additive Noise", IEEE Trans. Inform. Theory, Vol.IT-23, September 1977, pp.575-582.

16. Leondes, C.T. and Pearson, J.O., "Kalman Filtering of Systems with Parameter Uncertainties - A Survey", Int. J. Control, Vol. 17, 1973, pp.785-801.

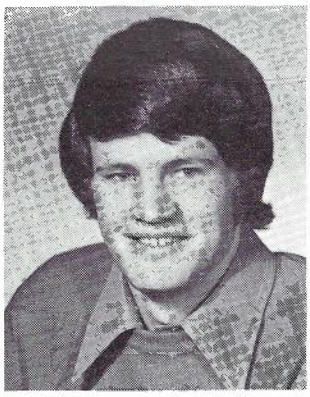
17. Mehra, R.K., "Approaches to Adaptive Filtering", IEEE Trans. Autom. Control, Vol.AC-17, October 1972, pp.693-698.

18. Jazwinski, A.H., "Adaptive Filtering", Automatica, Vol.5, 1969, pp.475-485.

19. Nahi, N.E. and Schaefer, B.M., "Decision-directed Adaptive Recursive Estimators : Divergence Prevention", IEEE Trans. Autom. Control, Vol.AC-17, February 1972, pp.61-68.

20. Quigley, A.L.C., "An Approach to the Control of Divergence in Kalman Filter Algorithms", Int. J. Control, Vol.17, 1973, pp.741-746.

21. Falconer, D.D. and Ljung, L., "Application of Fast Kalman Estimation to Adaptive Equalization", IEEE Trans. Comm., Vol.COM-26, October 1978, pp.1439-1446.



BIOGRAPHIES

GRANT NICHOLSON was educated at the University of Tasmania, Hobart Tasmania, where he was awarded the degrees of Bachelor of Engineering with honours (Electrical) in 1976 and Master of Engineering Science in 1979. He commenced work as an Engineer with Telecom Research Laboratories, Line and Data Systems Section in March 1978. His work in that section has been mainly concerned with PCM Digital Transmission Systems.

J.P. NORTON graduated from the University of Cambridge in 1962. After a year with English Electric-Leo Computers in London he spent three years at Imperial College, London, working for a Ph.D. in Electrical Engineering, then four years at the National Physical Laboratory and Warren Spring Laboratory in the U.K. as a research fellow. In 1971 he joined the Department of Electrical Engineering at the University of Tasmania, where he is currently a Senior Lecturer.

Linearity of Light Emitting Diodes for Analogue Optical Fibre Links

R. W. A. AYRE
Telecom Australia Research Laboratories

Optical fibres have frequently been proposed as a transmission medium for wideband services, and a variety of techniques is available to modulate a wideband analogue signal onto an optical carrier. The simplest of these is direct intensity modulation of the light source, and can be achieved with Light Emitting Diodes (LEDs) or Laser Diodes by modulating their operating current. However, these devices have nonlinear characteristics and are limited in their available output power. This paper describes the results of a series of tests on some commercially available LEDs to determine what levels of distortion and noise can be achieved on analogue links using these devices. Also included is an introduction to the sources of nonlinearity in LEDs and to the dominant sources of noise on analogue optical fibre links.

It has been found that with all of the LEDs tested, it is possible to meet the performance standards for nonlinearity and noise defined for a Main Video Link over a fibre length of 2 km without needing to resort to distortion compensation schemes.

1. INTRODUCTION

One of the applications frequently proposed for optical fibres is as a transmission medium for wideband services, such as television signals. There is a wide variety of modulation techniques that could be used to modulate an analogue signal onto an optical carrier; all of those currently described involve intensity modulation of the optical carrier, usually by an intermediate electrical carrier signal. This intermediate signal may be an amplitude modulated carrier (Ref.1), a frequency modulated carrier (Ref.2), a pulse signal with information carried by modulation of the pulse frequency (Refs.3,4), pulse width (Ref.4), interval between pulses (Ref.5), or one of a variety of similar techniques. All of these involve a more complex and expensive coding or modulation scheme when compared with the simplest scheme, which is direct intensity modulation of the optical carrier by the information signal. This is readily achieved with solid state Light Emitting Diode (LED) or Laser Diode (LD) sources by modulating the current flowing in the source device.

One objection to this technique in the past has been the poor linearity of the devices available, which introduces distortion components to the transmitted optical signal. A second has been the poor overall signal to noise ratio arising from a low level of transmitted optical power or sub-optimum receiver design. This paper describes the results of a series of measurements made to determine what levels of distortion and noise arise in analogue links using modern LED sources operating with linear transmitting and receiving circuits. Distortion compensation schemes have not been considered. Measurements are presented on a number of devices manufactured by Bell Northern Research. An introduction to the causes of nonlinearity in LEDs and to the dominant sources of noise on optical fibre links is included to help illustrate the design tradeoffs that must be made in the design of such links.

In discussing the linearity of LEDs (i.e., linearity in the relationship between LED current and optical power), it is important to make a distinction between the linearity of an LED alone, the linearity of an LED with a short fibre tail attached, and the linearity of a complete LED-fibre link. Various mechanisms within the LED give rise to nonlinearities in the relationship between the LED current and the total radiated power, and other mechanisms cause the spatial distribution of the LED output power to vary, resulting in level-dependent variations in the LED-to-fibre coupling efficiency.

Thus a measurement of the LED linearity made with a large-area detector mounted close to the LED will yield different results to a measurement made through a short length of fibre. Furthermore in a short length of fibre some optical power is carried in modes that are rapidly attenuated, so it is possible for linearity measurements on an LED coupled to a short length of fibre to yield different results to measurements made with a long length of fibre. From a systems point of view, it is the long-fibre results that are important.

Thus all of the device linearity measurements presented in this paper were made through a long (2 km) length of fibre. In the theoretical discussion in the next section the mechanisms that affect the linearity of the LED itself are discussed separately from the structural effects

that cause nonlinearity through varying the coupling efficiency.

The results of linearity and noise measurements should be interpreted in the light of a particular system application for which an appropriate performance specification is defined. The commonly suggested application for analogue fibre links is the point-to-point transmission of video signals over distances of a few km for non-broadcast services. A set of performance specifications for such services in Australia has yet to be decided. Accordingly, the set of performance specifications for a Main Video Link (Ref.17) has been used as a reference in this paper; these specifications are quite stringent and are expected to be typical of the highest grade of service required in such applications. There is no suggestion that the Main Video Link performance specifications could be met by analogue intensity modulated fibre systems over the distances for which these specifications normally apply - distances typically of several hundred km.

2. THEORETICAL BACKGROUND

2.1 Nonlinearity of LED Sources

The generation of optical power in an LED is inherently a linear process, in that a single photon is generated for each radiative recombination of an injected carrier. However, several factors affect the proportion of recombining carriers that undergo radiative recombination and the proportion of generated photons that are absorbed in the LED material, causing nonlinearities in the relationship between the total LED output power and the LED current. The proportion of the total LED output power that is coupled into the fibre also varies with the LED current as a result of spatial or geometric effects, causing further nonlinearities in the relationship between the LED current and the output power received at the end of the fibre.

One factor, the variation of the diode diffusion admittance with injection level and with temperature can be eliminated by proper design of the LED driver. A long list of factors combine to give superlinearity in the transfer characteristics (i.e., the optical power increases at a faster rate than the LED drive current) at low drive currents and sublinearity (i.e., the optical power increases at a slower rate than the drive current) at high currents; these include

- variation of injection efficiency with injection level
- variation of the proportion of carriers recombining in the depletion region
- variation of the internal quantum efficiency with both injection level and temperature
- variation of the internal photon absorption rate with temperature

These affect the static linearity and differential gain of the device and are discussed in detail in (Ref.6).

The variation of carrier lifetime with injection level alters the modulation bandwidth of the device, and thus can cause both differential gain and differential phase distortion in the device, (Refs.6,7,8). This effect is minimised if the modulation bandwidth of the LED is substantially higher than the bandwidth of the signal. Unfortunately, when the device parameters of an LED are chosen to optimise the modulation bandwidth, a lower efficiency and lower optical power output are obtained.

Finally, there are structural factors that affect both the static linearity and the differential gain of complete links through variations in the LED-to-fibre coupling efficiency. The design of high radiance LEDs is usually such that the PN junction is quite large compared with the active area of the diode, with a large annular contact to the N side on the top surface of the chip surrounding and above the active area. A small area contact, typically 50μm diameter is made to the P side base of the chip and defines the emitting area. (Ref.9). At low current levels, the entire junction emits light, but at higher currents the injected current, and hence the optical power developed crowds to form a bright spot directly above the base contact. (Ref.10). Thus, if an optical fibre is properly aligned over the base contact, the optical power coupled to the fibre will increase at a faster rate than the total optical power generated as the diode current crowds towards the region above the contact. (Ref.11). This accentuates the superlinear characteristic of the diode, and also makes the linearity of a diode-fibre combination somewhat dependent on the alignment of the diode and fibre. It is often difficult to obtain repeatable linearity characteristics for a diode-fibre combination unless great care is taken in re-aligning the fibre.

2.2 Noise Sources

Noise in optical fibre systems has been extensively analysed (Refs.12,13,14). The analogue systems we are considering require moderate bandwidth (less than 10 MHz) and a high signal-to-noise ratio (> 45 dB). For such applications it is well established (Refs.13,14) that the optimum receiver has an FET input receiving amplifier, although the new super-beta bipolar transistors have been claimed to perform as well (Ref.15).

In these applications, the total system noise is dominated by shot noise on the receiving photodiode current, with a small contribution from the thermal noise of the FET and the receiving amplifier feedback and biasing resistors. For such a system, the signal to noise ratio is approximately

$$\text{SNR}\left(\frac{\text{p-p signal}}{\text{rms noise}}\right) = \frac{\text{ml}}{\sqrt{2q\text{I}B + \frac{4kTB}{R} + \frac{32\pi^2 kTC^2B^3}{99_m}}} \qquad (1)$$

where I = DC photodetector current, A

m = modulation index (i.e. ml = p-p signal current)

q = 1.6×10^{-19} coulomb

B = system bandwidth, Hz, (brickwall filter assumed)

k = Boltzmann's constant 1.38×10^{-23} joule/^{0}k

T = absolute temperature, ^{0}K

R = parallel resistance at amplifier input, ohm, (including feedback R)

C = total input capacitance, F

g_m= FET transconductance, A/V

The first denominator term is an approximation in that the shot noise actually depends on the instantaneous photocurrent, rather than the average photocurrent. For small modulation indices, or in situations where we require an estimate of SNR without a rigorous specification of the test signal it is a reasonable approximation to use and the errors are small. The second term is the thermal noise on the biasing and feedback resistors at the input of the amplifier, and the third is the FET channel thermal noise referred to the amplifier input as a noise current. (The remaining sources of FET noise are negligible for practical silicon FETs at these bandwidths). Examination of the denominator terms in equation (1) reveals that the spectral density of the shot noise and resistor noise components are flat, while that of the FET increases with frequency.

In a typical application, we might require

SNR = 50 dB unweighted and bandwidth = 10 MHz.

We might use a FET with g_m = 10 mA/V, a 1 mΩ feedback resistor, and m = 50%, and after summing all contributions (FET, photodiode, etc.) in practice obtain a total input circuit capacitance of 10pF. On the assumption that the link is fairly short (a few km) and that no peaking is needed in the receiver to equalise the fibre dispersion or LED bandwidth, the respective noise contributions are then:

Shot noise	2.27 nA
Resistor noise	0.41 nA
FET noise	1.21 nA
for a total noise current of	2.6 nA

DC photocurrent = 1.61 μA

and signal current = 0.81 μA p-p

for an unweighted SNR of 49.9 dB. (p-p signal/rms noise).

Using a PIN photodiode, this photocurrent represents 3.6 μW of optical power, or ∿ - 25 dBm Combinations for other bandwidths or signal-to-noise ratios are readily calculable from equation (1).

From this equation, it is obvious that signal to-noise ratio improves as we increase the modulation index, or as we increase the total input power (and hence the photocurrent I), and as we reduce the bandwidth of the receiver. If the received optical power is high, the shot noise term dominates and the signal-to-noise ratio falls linearly with the received average power. With lower levels of received power, the amplitier noise terms begin to contribute, and the signal-to-noise ratio falls at a faster rate. Figure 1 shows the results of a typical calculation using equation (1) assuming a PIN diode photodetector, and is a useful illustration of the caution one must use in extrapolating short-link SNR figures to longer links.

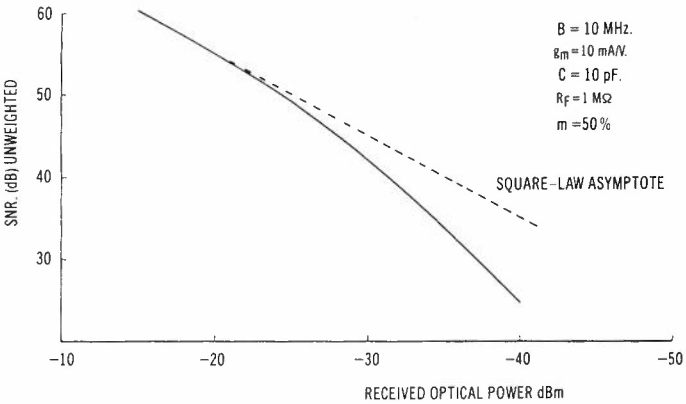


Fig.1 - Example of signal to noise ratio calculated from equation (1).

3. THE TEST METHOD

Linearity and noise tests were made on a complete video link comprising an LED driver, LED, 2 km fibre, PIN photodetector and FET input receiving amplifier. The relevant parameters of the fibre are summarised in Table 1. The linearity measurements were confined to measurements of differential gain and differential phase, these being the most critical parameters in previous tests of optical fibre links carrying colour video signals.

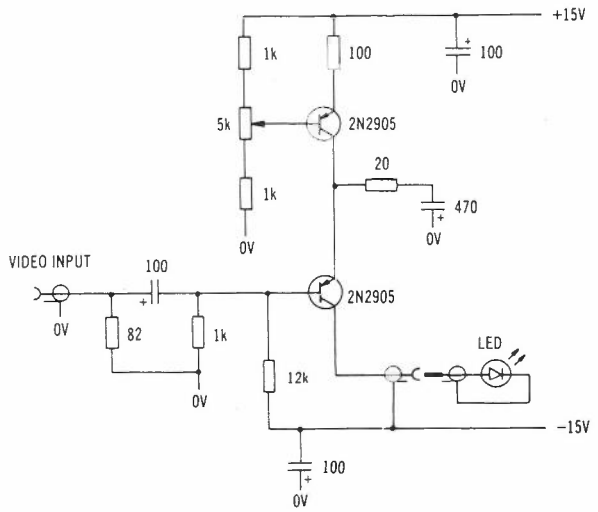


Fig.2 - LED driver circuit.

Each diode was supplied with bias from a constant current source, and a current representing the modulating signal was negatively modulated onto the bias current. The modulator circuit is shown in Fig.2 and Fig.3 illustrates the location of the bias point and the signal excursion relative to the diode characteristic for a typical diode.

Notice that the modulator is AC coupled. This is reasonable in a test circuit where the average picture level does not vary. In a driver circuit intended to carry realistic video signals with varying average picture level, some form of DC clamping is necessary to restrict the signal excursion to the most linear region of the diode characteristic.

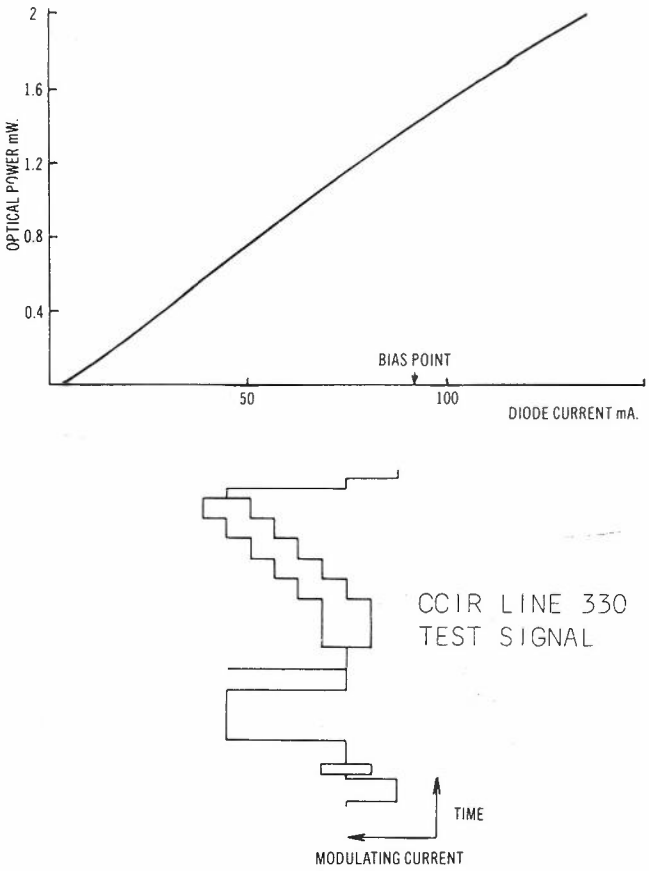


Fig.3 - Location of bias point and signal excursion on diode characteristic.

Each LED was mounted on a heatsink and was initially driven with a peak-to-peak signal current of 30 mA superimposed on a bias current of 95 mA. The fibre was carefully aligned to maximise the optical power detected at the receiving end of the fibre.

The LED was then driven with a fixed signal current of 30 mA peak-to-peak and the differential gain and phase of the overall link measured at various bias currents to determine the optimum bias for each device. Figure 5 is an example of the results obtained in this test; for most of the devices a broad minimum between 70 mA and 110 mA is found.

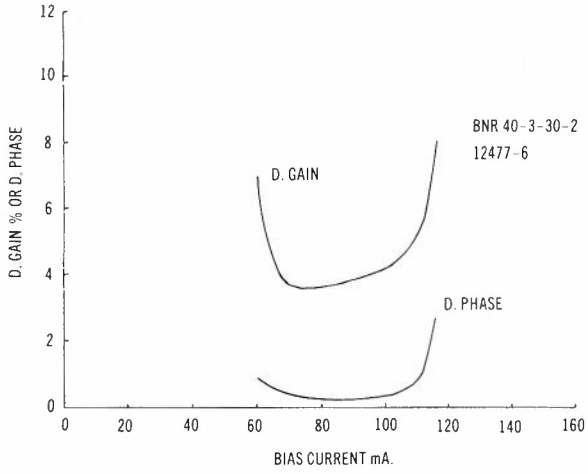


Fig.5 - Example of distortion versus bias current results.

The receiver circuit is shown in Fig.4. The overall differential gain and differential phase of the LED driver and receiver cascaded are approximately 0.4% and 0.25^0 at 2 x normal signal level respectively; these figures are substantially lower than those of the LEDs measured.

Each LED was biased at or slightly above its optimum bias current thus determined, and driven with various signal currents. The output signal voltage, differential gain, and differential phase were then measured as the modulation depth was varied. Finally, with no modulation the lum-

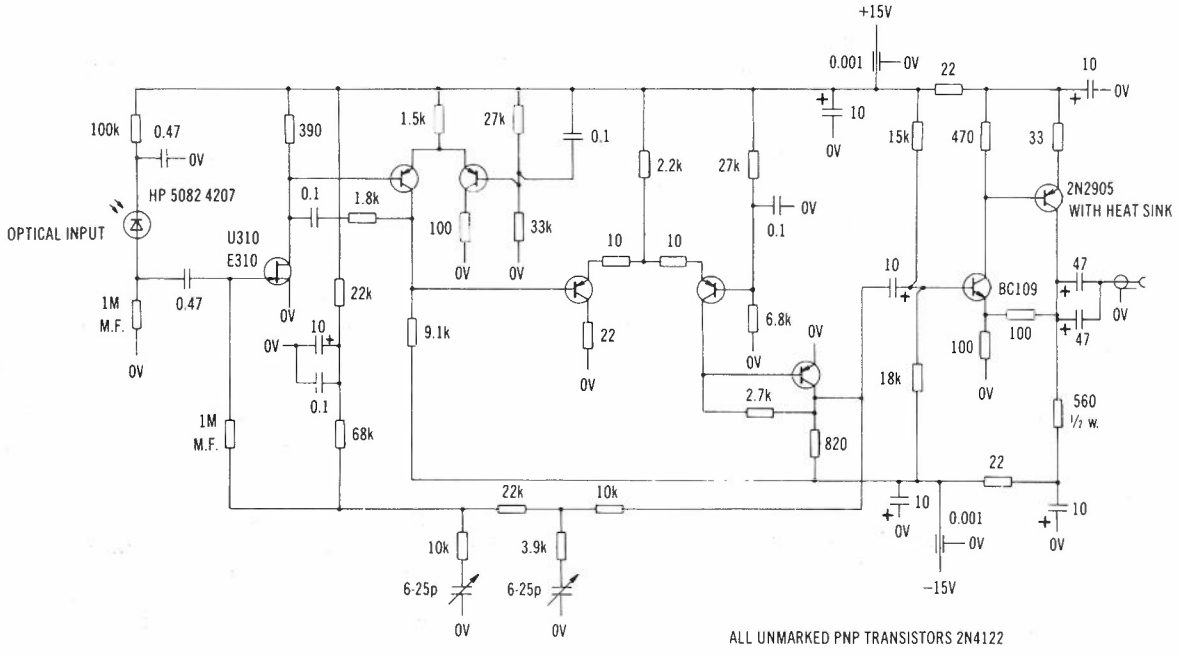


Fig.4 - Receiving amplifier.

inance channel noise and chrominance channel noise were measured so that the SNRs for each modulation depth could be calculated.

For some of the devices, the source-end of the fibre was intentionally misaligned in each of four directions to induce an extra coupling loss of 3 dB, and the differential gain measurement repeated.

The test signal used was the standard Line 330 video signal nominated by the CCIR (Ref.16). This signal is composed of a 10 µS bar, followed by a 200nS HAD pulse of sine-squared shape, followed by a five-step staircase with superimposed signal at colour subcarrier frequency. The bar and staircase both reach the maximum video signal amplitude, 100 units, while the superimposed subcarrier signal level is constant at 40 units peak-to-peak. Differential gain was measured as the change in the amplitude of subcarrier signal at the receiver as the staircase signal rises. Differential phase was measured as the change in the phase of the subcarrier signal as the staircase rises. In both cases, the peak-to-peak variation was recorded. Both measurements were made on a Tektronix 521 A Vectorscope using a full field of the test signal.

4. RESULTS

Curves of optical power versus current for a typical LED are presented in Figs.6,7, and 8. Figure 6 is the power-current characteristic for the total emitted power measured with a large area detector mounted close to the LED. Figure 7 is the same characteristic measured with a finer current resolution and plotted on log-log scales. Figure 8 represents the power coupled to a short length of fibre.

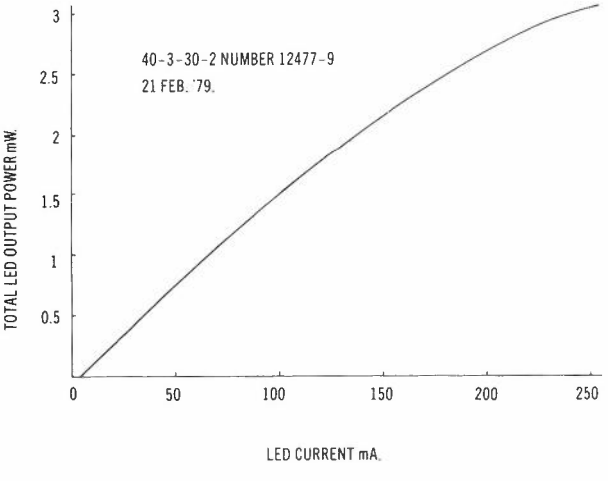


Fig.6 - LED power-current characteristic.

The saturation effect at high currents is quite obvious in Figs.6 and 8. The deviation from linearity at low currents is not obvious in these figures, but is quite apparent in Fig.7, where it is seen as a departure from unity slope at low currents.

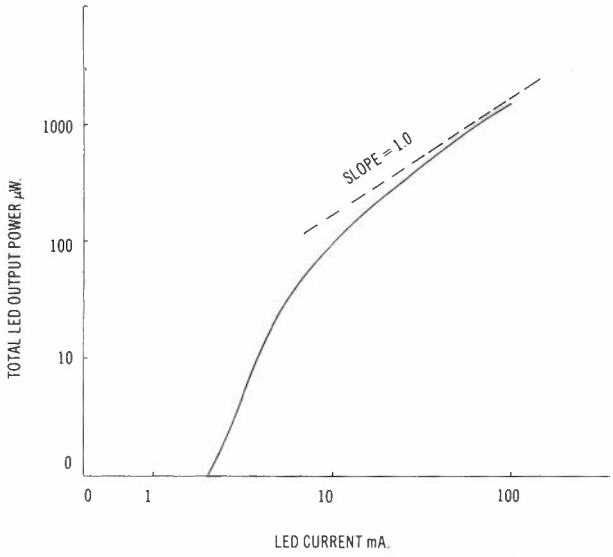


Fig.7 - LED power-current characteristic.

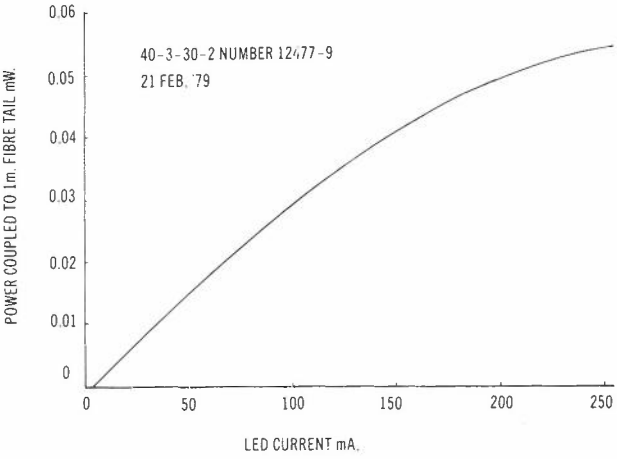


Fig.8 - LED + fibre power-current characteristic.

Figure 5 was discussed earlier in the paper; it illustrates the change in differential gain and differential phase as the LED bias current is varied with constant modulating current; a broad minimum between 70 and 110 mA is usually found.

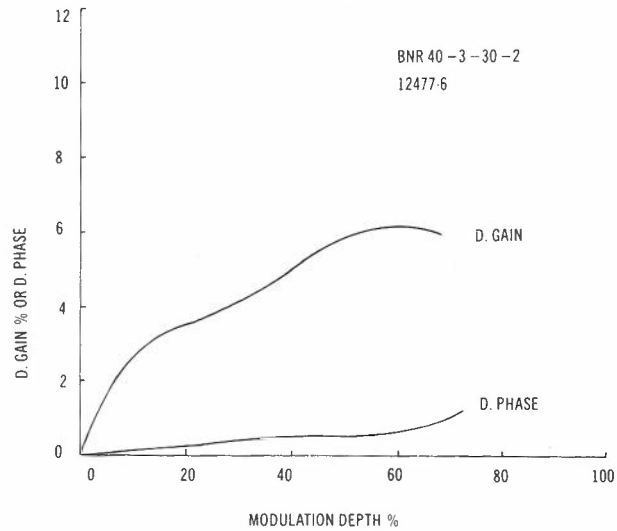


Fig.9 - D. gain and D. phase measurements c LED No.12477-6.

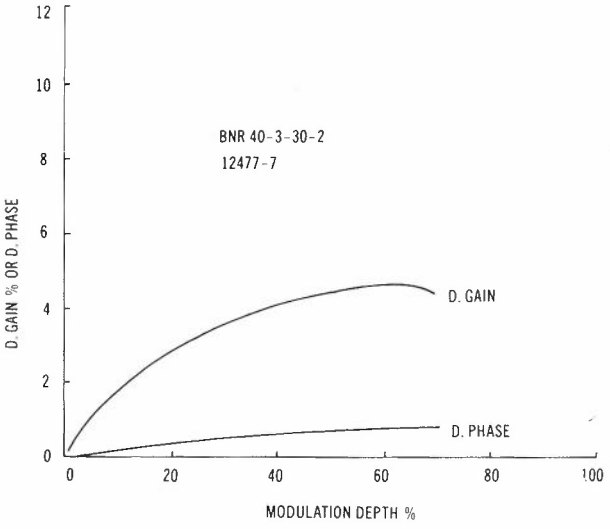


Fig.10 - D. gain and D. phase measurements on LED No.12477-7.

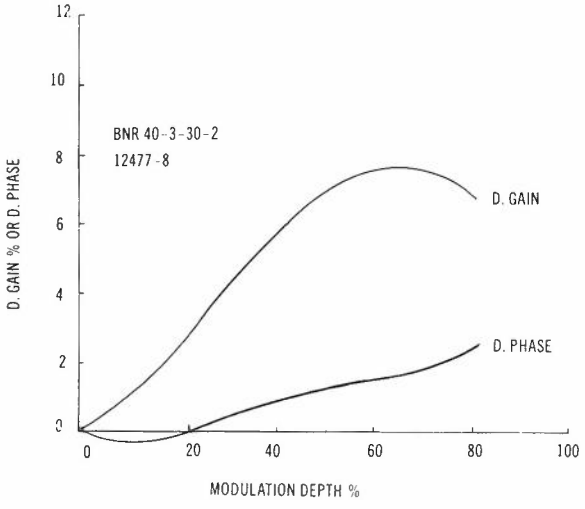


Fig.11 - D. gain and D. phase measurements on LED No.12477-8.

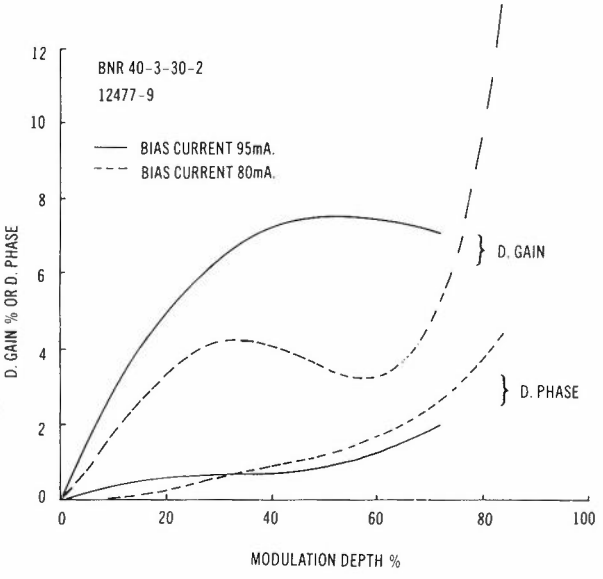


Fig.12 - D. gain and D. phase measurements on LED No.12477-9.

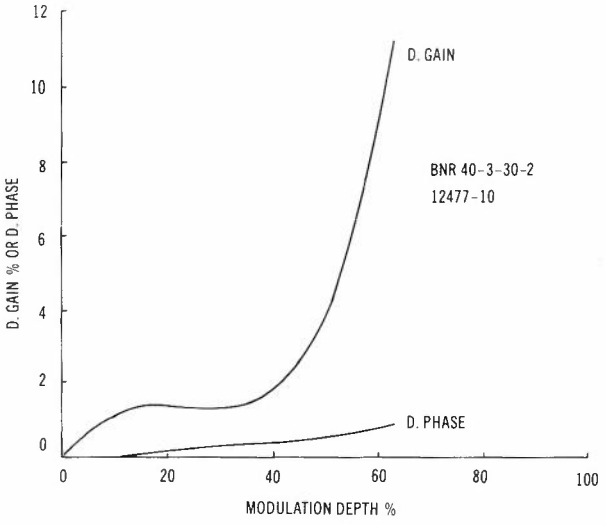


Fig.13 - D. gain and D. phase measurements on LED No.12477-10.

Graphs of differential gain and differential phase versus current for 2.1 km links using each of the BNR LEDs so far tested are presented in Figs.9 to 13. Each LED was tested at its optimum bias current for these graphs, this bias current being listed in Table 2. However, in Fig.12, the effect of operating at a current other than the correct bias current is demonstrated. For this particular LED the optimum bias current was found to be 80 mA. If it is operated at 95 mA, which was the optimum current for most of the LEDs tested, a substantial increase in differential gain is encountered.

There is a wide spread in the performance obtained from different LEDs, particularly in the differential gain performance. However, it is possible with each LED to meet the performance objectives for a Main Video Link (Ref.17) i.e., 5% differential gain and 4^0 differential phase for modulation depths up to 30% and most can be driven harder.

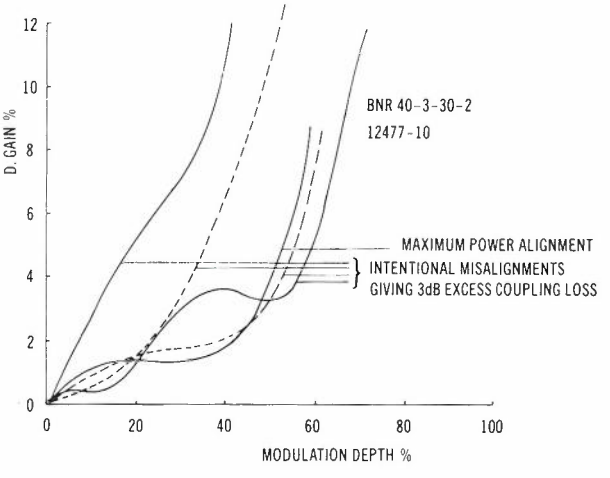


Fig.14 - Effect of diode fibre misalignment.

Figure 14 shows the effect of slight misalignment between the diode and the fibre on the differential gain performance of the link. In each case, the misalignment caused an extra coupling loss of 3 dB relative to the peak coupled power. The range of differential gain values ob-

tained at any particular bias current and modulation depth illustrates the need for very precise alignment between diode and fibre.

Table 2 lists the signal-to-noise ratios in the luminance channel and chrominance channel obtained for each of the tested diodes over the same 2.1 km link. The differences between the units are due to differences in the power coupled to the fibre. Two devices just fail to meet the luminance channel SNR requirement of 58 dB for a Main Video Link (Ref.17) at the modulation depth quoted, but would meet this requirement at a higher modulation depth. All devices meet the chrominance channel SNR requirement with a small margin.

5. CONCLUSIONS

The results presented in this paper can only be interpreted in the light of a particular system application for which an appropriate performance specification has been devised. The LEDs that have been tested give a sufficiently high power output with low distortion to be used in a link meeting Main Video Link performance specifications over a distance of 2 km with currently available fibres. There is little or no margin for LED degradation, but the fibre used has a smaller core diameter (and therefore lower coupled power) and slightly higher loss than one that would be selected for video distribution applications.

The Main Video Link specification is quite stringent, and signals meeting this standard would be given a very high rating in subjective viewing tests (Ref.18). It is likely that a lesser standard will be set for non-broadcast video distribution services. If this eventuates, the LEDs can be used at higher modulation depths to obtain longer link spans.

A quite separate point also arises from these experiments. The nonlinearity of a given LED depends quite critically on its bias point, and each diode should be carefully measured to determine its proper bias point before being put into service. Similarly, the alignment of the diode with its fibre has a strong influence on the nonlinearity of the link. One way of ensuring this alignment would be to use diodes with short fibre tails properly aligned in the laboratory or workshop and permanently attached to the diode cases, and to join these fibre tails to the long lengths of installed fibre.

A variety of techniques have been proposed to improve the linearity of analogue fibre links, including the use of optical feedback, complementary predistortion, and feed-forward compensation (Ref.19). Such techniques can reduce differential gain and differential phase distortions by a factor of at least four in ideal situations. However, the strong dependence of the link linearity on the alignment of the fibre with the LED can be expected to limit the performance improvement and the applicability of such schemes in practical analogue fibre links.

6. REFERENCES

1. Hara, E.H. and Ozeki, T., "Optical Video Transmission by FDM Analogue Modulation", Canadian Cable Television Convention, 1976.

2. Bark, P., et al., "Installation of an Experimental Optical Cable Link", Third European Conference on Optical Fibre Communications, Munich, 1977, pp.243-245.

3. Timmerman, C.C., "SNR of a Video Signal Transmitted by a Fibre-Optic System using Pulse Frequency Modulation", IEEE Trans. on Broadcasting BC23, March 1977, pp.12-16.

4. Takasaki, Y., et al., "New Fibre Optic Analogue Baseband Transmission Plan for Colour TV Signals", IEEE Trans. on Communications COM26, June 1978, pp.902-907.

5. Ueno, Y. and Yasugi, T., "Optical Fibre Communications Systems using Pulse Interval Modulation", NEC Research and Development, January 1978, pp.45-52.

6. Asatani, K. and Kimura, T., "Analyses of LED Nonlinear Distortions", IEEE Trans. on Electron Devices ED25, February 1978.

7. Namizaki, H., et al., "Current Dependence of Spontaneous Carrier Lifetimes in GaAs-GaAlAs Double Heterostructure Lasers", Applied Physics Letters, Vol.24, May 1976, pp.486-487.

8. Lee, T.P., "The Nonlinearity of Double Heterostructure LEDs for Optical Communications", Proc. IEEE, Vol.65, September 1977, pp.1408-1410.

9. Burrus, C.A. and Miller, B.I., "Small Area Double Heterostructure Aluminium-Gallium-Arsenide Electroluminescent Diodes for Optical Fibre Transmission Lines", Optics Communications, Vol.4, December 1971, pp.307-309.

10. Williamson, W.J., "Variations in the Surface Brightness Distribution of Light Emitting Diodes for Optical Communication", IREE International Electronics Convention, Sydney, 1975, Digest pages 110-111.

11. Teede, N.F., "Linearity of High Radiance Light Emitting Diodes", Telecom Australia Research Laboratories Report 7086, 1977.

12. Personick, S.D., "Receiver Design for Digital Fibre Optic Communication Systems", Bell System Technical Journal, Vol.52, No.6, July 1973, pp.843-875.

13. Goell, J.E., "Input Amplifiers for Optical PCM Receivers", Bell System Technical Journal, Vol.53, No.9, November 1974, pp.1771-1793.

14. Boisrobert, C.Y., et al., "Low Noise Photodetection - Preamplifier Analysis", Proceedings Second European Conference on Optical Fibre Communications, Paris, 1976, pp.293-297, (in French).

15. Knorr, S.G., et al., "Low Noise Fibre Optics Receiver with Super-Beta Bipolar Transistors", Fiber and Integrated Optics, Vol.1, No.4, pp.369-382.

16. CCIR Recommendation 473-1, Geneva, 1974.

17. Telecom Australia, "Television Relay Facilities - Operational Performance Objectives", Engineering Instruction Planning 08010, 1973.

18. Macdiarmid I.F. and Allnatt, J.W., "Performance Requirements for the Transmission of the PAL Coded Signal", Proceedings of the IEEE, Vol.125, No.6, June 1978, pp.571-580.

19. Strauss, J., "Linearised Transmitters for Analogue Fiber Links", Laser Focus, October 1978, pp.54-61.

TABLE 1 - Fibre Parameters

AWA Step Index Fibre	
Outer Diameter	130μM
Core Diameter	52μM
N.A.	0.2
Attenuation at 840nM	4.7dB/km
Length	2.1km

TABLE 2 - Signal to Noise Ratios for Various Diodes *

Diode No.	Luminance SNR dB	Chrominance SNR dB	Optim Bias Cur
12477 -6	58.0	53.1	95
-7	57.8	53.5	95
-8	57.8	52.7	90
-9	60.4	56.4	80
-10	62.0	57.1	95

* These results are for 2.1 km link with optical loss of 9.9 dB. The LEDs were modulated at 30% modulation depth and biased at their optimum bias current as derived from linearity measurements.

Modulation Depth $m = \frac{I(\text{signal})P\text{-}P}{I(\text{Bias})}$

BIOGRAPHY

ROBERT AYRE graduated from the George Washington University, Washinton DC with the degree of B.Sc. in 1967, and later from Monash University with the degrees of B.E. and M.Eng.Sc. in 1969 and 1972 respectively. His postgraduate work included studies of noise and base resistance in bipolar transistors.

In 1972 he joined the Visual Communications section of Telecom Research Laboratories, where he worked on specialised aspectes of television transmission, including time multiplexed transmission of video and audio signals. The combined backgrounds in low noise amplifiers and video signal transmission led to his becoming involved with the Guided Media section's optical fibre transmission project. Since 1977, he has been working in both sections, and his current projects include the characterization of Laser & LED optical sources and the design of terminal equipment for optical fibre links.

Measurement of Group Path Variations of an Ionospherically Propagated HF Signal from its Modulation Envelope

D. W. CORNELIUS *

Division of Theoretical and Space Physics
Latrobe Univeristy
Melbourne

E. A. ESSEX

Division of Theoretical and Space Physics
Latrobe University
Melbourne

An experiment was conducted using the 1 KHz amplitude modulation on the 4.5 MHz VNG standard frequency and time transmission to study differences occurring in the group path (delay) of the ionospherically propagated signals. Propagation modes involved both F-region reflection and scattering. The experiment seeks to distinguish between F-region scatter and ground backscatter by the differences in relative propagation delays of the modes. The method was found to be limited by frequency selective fading which distorts the modulation envelope. The results indicate however that direct F-region scatter is the main scatter mode for the given experimental conditions.

1. INTRODUCTION

H.F. Standard frequency and time transmissions are used to compare, at long distances, the stability of other oscillators (using the carrier signal), and to synchronise clocks (using the modulation). Propagation through the ionosphere limits the accuracy of the received carrier because of phase path changes introduced by the ionosphere. The phase path is related to the transit time of a surface of constant phase in the medium (Davies, (Ref.1)). These phase path changes are due to temporal variations in the underlying ionization as well as changes in the reflection height.

Phase path variations have been measured using the VNG signals by Joyner (Ref.2). Cornelius et al (Ref.3) show Doppler frequency shifts (time rate of change of phase path changes) obtained from the VNG signals. The limits of accuracy of the received carrier imposed by ionospheric propagation for the VNG 12 MHz signal received at Townsville is discussed by Heron and Rose (Ref.4). It was concluded that averaging for about 1 hour gives a 0.1 Hz accuracy, depending on the time of day.

The amplitude modulation is also affected by ionospheric propagation. The group path in the ionosphere is now important. This is related to the transit time of a wave packet (modulation). In an ionized medium the group path for a signal is longer than the true path as the pulse is slowed or retarded in that medium. The group path is the distance covered in free space if the wave packet had travelled with the free space velocity, (Davies, (Ref.1)). The theorem of Breit and Tuve (e.g. Davies (Ref.1)), can often be easily used to estimate the time of flight or group path of an ionospherically propagated signal.

This paper describes a simple experiment designed to measure the propagation delay of the modulation on the 4.5 MHz VNG standard frequency transmission. The purpose of the experiment was to test for the existence of a direct F-region scatter reflection mode proposed by Cornelius et al (Ref.3), using the modulation delay to distinguish between ground backscatter and direct F-region scatter.

The determination of group path is possible however, from phase path measurements alone. The group path can be found from measurements of the phase path variations of a carrier and its side bands (Toman, (Ref.5)). This method is currently being developed and should prove to be a more accurate method for determining the group path.

2. EXPERIMENTAL

The 4.5 MHz VNG signal was received at La Trobe University (37^0 43'S, 145^0 3.3'E, geographic coordinates) at a ground distance of about 40 km, resulting in near vertical reflection. The VNG signal is double side-band amplitude modula-

* Current address
Ionospheric Prediction Service,
P.O. Box 702,
Darlinghurst, N.S.W. 2010.

ted each second (except for the 59th second each minute) with a 1 kHz signal whose length is controlled as follows. For seconds 1 to 55 each minute, the modulation is 50 ms long, for seconds 55 to 58 each minute the modulation is 5 ms long, and the second 0 or 60 has modulation 500 ms long. The 1 kHz modulation on the 4.5 MHz signal was detected and the variations in the delay found by comparing the time the received modulation commenced with the edge of a one second pulse from a local standard oscillator. The one second pulses generated from the local standard oscillator were not synchronized with the transmitted tone so relative delay only was measured. The length of the modulation bursts were also measured each second. These values were punched directly onto paper tape for computer analysis.

Apparent errors in the modulation delay due to variations in the RF signal level need to be reduced as much as possible. In order to achieve this the receiver incorporated an AGC with a time constant of 4.7 ms which can hold the IF output level constant for signal levels down to 1 microvolt.

The receiving antenna was a tuned loop, duplicating the experimental conditions discussed by Cornelius et al (Ref.3). Doppler spectrum and signal level were also recorded. Local ionograms were not available due to a malfunction. The experiment was carried out over several nights, the results from night to night not being significantly different.

3. INTERPRETATION OF POSSIBLE RESULTS

Measurement of the modulation delay and length under single-hop F-region reflection conditions provides a reference for the same measurements under scatter conditions. The case of the signal returned by direct F-region backscatter is considered first.

The exact location of the scattering irregularities in the F-region is uncertain, however the irregularities are assumed here to be near overhead and near the peak of the layer. If this is the case, then as the reflection point approaches the peak of the layer, the modulation delay would increase as the group retardation increases. When scattering commenced the delay would be little different. As the critical frequency of the layer further fell the modulation delay would gradually return to a value near that corresponding to F-region reflection before group retardation became large. This delay and subsequent values depend on the nature and location of the irregularities and the critical frequency of the layer.

If the signal were returned via ground backscatter involving two ionospheric reflections then the following sequence of tone delay would be expected. As the reflection point approached the peak of the layer the modulation delay would increase as before. When the group path approached infinity (peak reached) the ground backscattered signal, scattered from the edge of the skip zone, would be the first received return. That is, reflection from overhead is no longer possible, but an oblique ray can be reflected. This ray hits the ground at a minimum range called the skip distance (minimum distance for trans-ionospheric communication under these circumstances). At this point some of the energy is scattered in all directions and some energy may return to the receiver via the same path (Dieminger, (Ref.6)). When this occurs the path length has doubled or increased by at least twice the height of the layer. This corresponds, for a 300 km reflection height and no group retardation, to an increase in delay of at least 2 ms. However for more oblique backscatter, from ground ranges further from the skip zone edge, reflection height and retardation fall, but range increases. As the critical frequency of the layer falls the backscatter return will be received from greater ranges corresponding to larger minimum zenith angles of reflection. At no time would the delay approach that due to a single F-region return.

A combination of direct F-region and ground backscatter would give delay figures as expected for F-region scatter as these are returned first However the modulation length would be increased by the time difference of the first F-region and most distant ground return. This pulse length increase could be up to 10-15 ms (e.g. Earl and Bourne, (Ref.7)) corresponding to the delay from low angle backscatter.

4. RESULTS

The experimental results of one morning only will be discussed, that of 28 October, 1975. The signal strength record indicates a substantial drop in level occurred between 0242 and 0245 hours (the signal level is shown in Figs.1 and 2). The Doppler spectrum shows a transition from

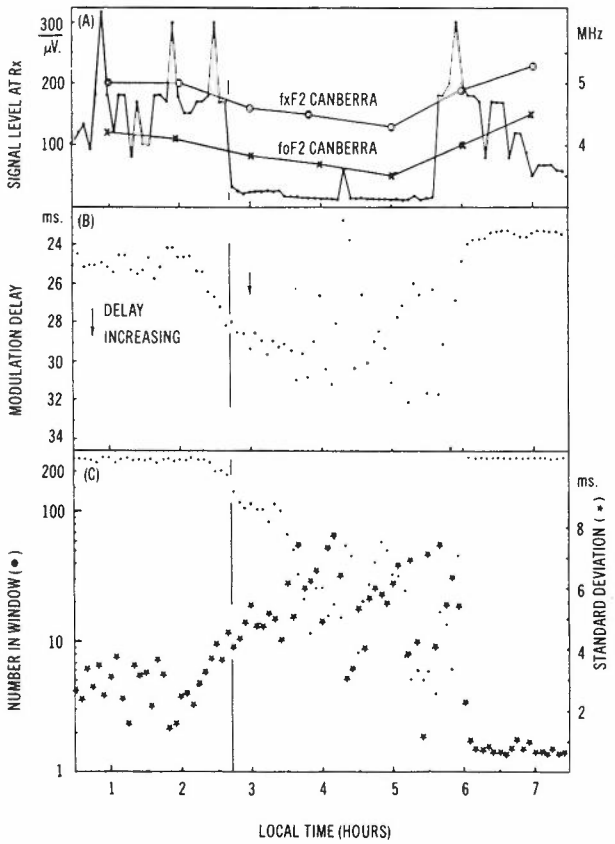


Fig.1 - Modulation delay data for 28 October, 1975, 4.5 MHz. The vertical line indicates the time of transition to scatter reflection.

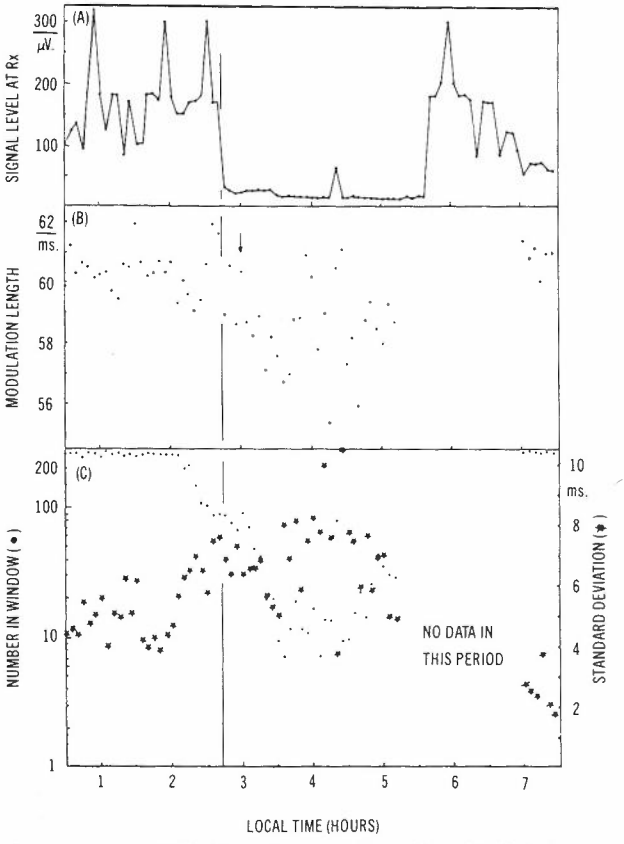


Fig.2 - Modulation length data for 28 October, 1975, 4.5 MHz.

a narrow spectrum (one or few reflection points) to a broad spectrum (indicating many reflection points) at 0243 AEST, and that this broad spectrum was received until 0536 hours. This kind of spectral record is indicative of a scatter mode (Cornelius, (Ref.3)). The ionosonde data from Canberra is shown in Fig.1 (a). It is expected that the ionospheric parameters for Beveridge would be lower than those at Canberra due to a north-south gradient in electron density. Cornelius et al (Ref.3), shows Canberra and Beveridge data for October 1973. It is seen that the variations were similar but the Canberra values were consistently larger. To reduce the Canberra fxF2 data for Beveridge as shown in Fig.1 (a) approximately 0.25 MHz should be subtracted. If this is done the two scatter transition points correspond well to the fall and increase in fxF2. These conditions correspond to the onset of a scatter reflection mode as discussed by Cornelius et al (Ref.3). Figs.1, 2, 3, 4 and 5 show the modulation delay, length and related data grouped into five minute periods with the average delay and length shown twice for easy comparisons.

The raw numbers from the counter were processed via a window in the computer program such that only those numbers within a certain range of the expected delay or length were accepted and included in the average. The computer window applied to the delay data allowed numbers between 25 ms longer and 10 ms shorter (compared to the value during reflection). This is long enough to include signals resulting from even distant ground backscatter (up to 15 ms absolute delay). The number of values satisfying this criterion in each five minute interval is shown in Fig.1 (c), together with the standard deviation of the numbers within the five minute period.

The maximum number of counts possible using the leading edge of all length tone bursts was 59 per minute or 295 in five minutes. Most totals were well under this maximum due to triggering of the counter by unwanted signals.

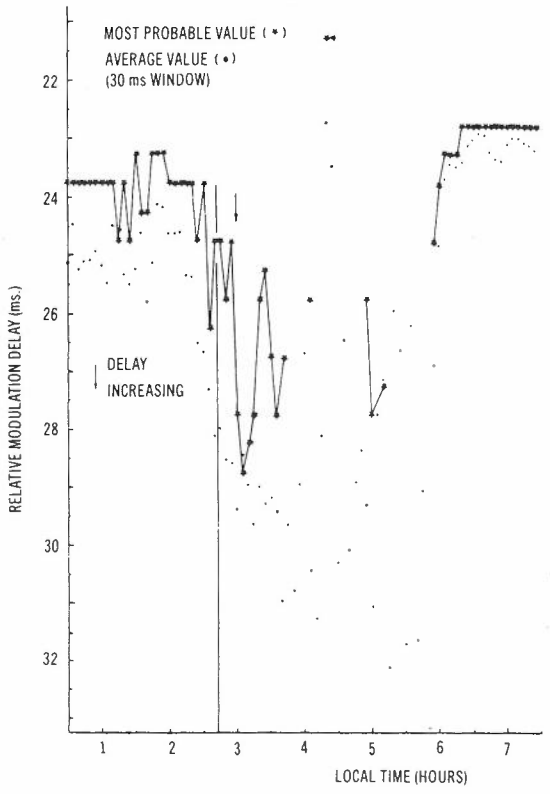


Fig.3 - Modulation delay detail for 28 October, 1975, 4.5 MHz.

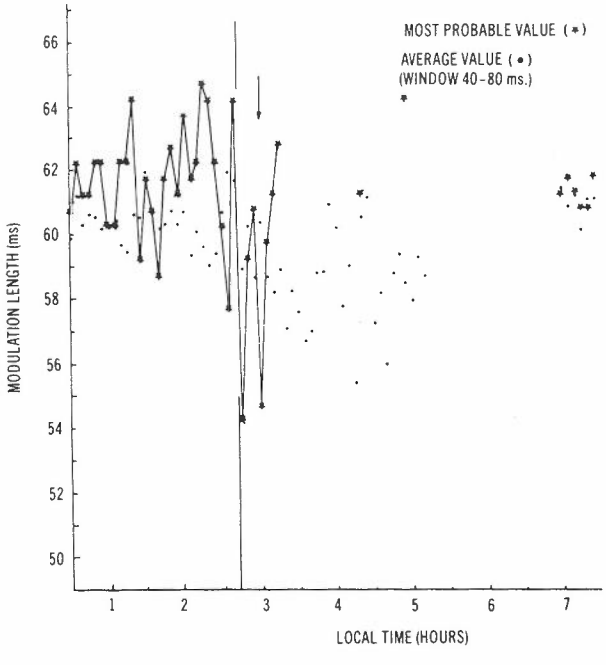


Fig.4 - Modulation length detail for 28 October, 1975, using a window of 40 to 80 ms.

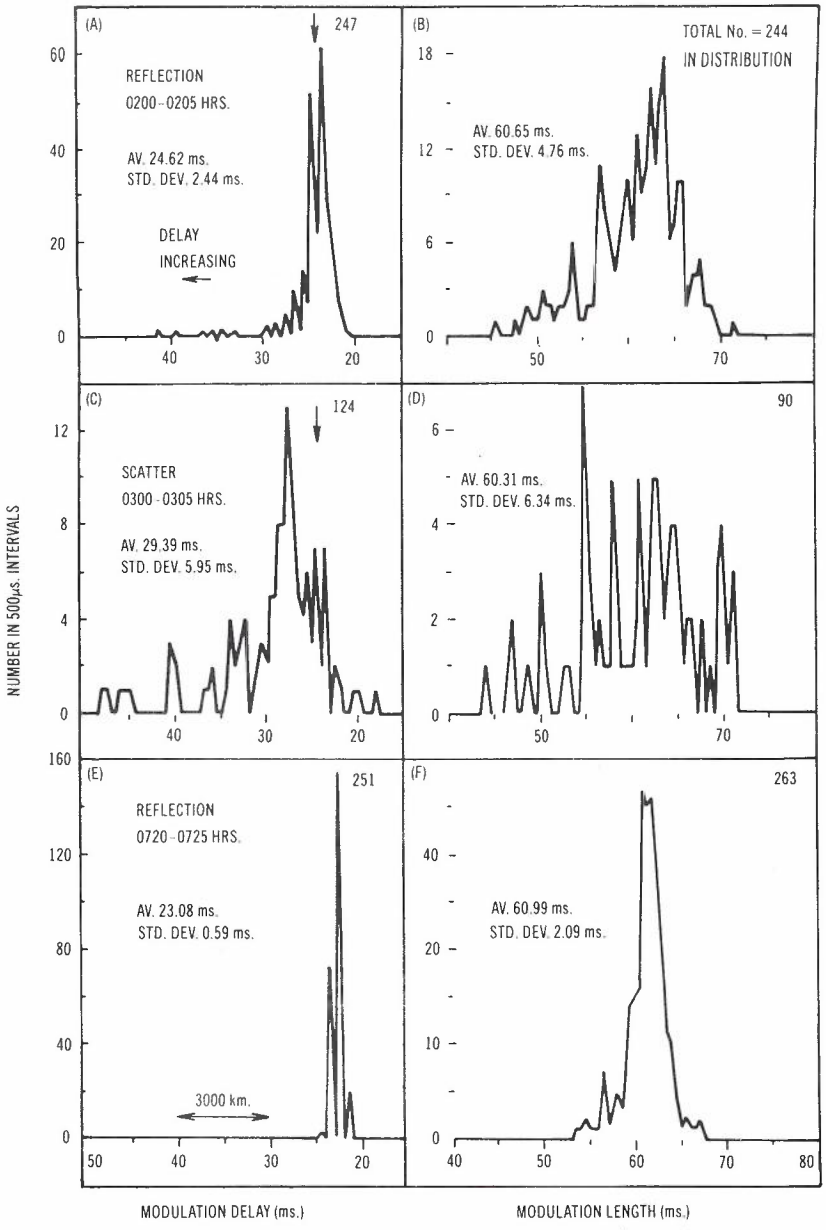


Fig.5 - Delay and length data distributions for three 5 minute intervals.

The window for the modulation length allowed lengths between 40 and 80 ms to be included in the average. Again this allows for lengths due to a combination of F-region and ground scatter effects. The number of values satisfying this criterion in each five minute period are shown in Fig.2 (c). The maximum possible number of 50 ms counts per minute is 55, or 275 in five minutes.

The standard deviations were often as large or larger than the effect that is being looked for. The standard deviation is considerably larger during the scatter times and the smallest when reflection recommences at about 0530 hours. As some of the distributions are non-Gaussian the most probable value is also shown plotted in Figs.3 and 4. Figs.1 and 2 show only the average of each five minute distribution. When the scatter condition reverts back to reflection at about 0530 hours, the signal strength was very small. The number of values of delay falling in the window during this time was too small and their deviation too large to be useful for study. It is possible that under these conditions the receiver AGC is responding to unwanted signals within the receiver bandwidth rather than the VNG signal. At times before and after the transition at 0243 hours however, sufficient values fall within the window to permit study.

At 0243 hours no increase in delay by 2 ms occurs in either the average or most probable value, (see Fig.3). A step increase in delay would have been expected if a simple transition from F-region reflection to ground backscatter occurred.

For direct F-region scattering near the peak it would be expected that the modulation delay would return to a value showing less group delay than the value when fxF2 was very close to 4.5 MHz. It is seen that the average delay continues to increase, showing about 5 ms more apparent delay than the reflection case.

The most probable value shows a large increase in delay well after scattering commenced, then returns to a value near that soon after

scattering commenced (0325 hours). On the basis of the average delay (up to 0330 hours only being considered; after this time there are too few numbers in the window) showing a large increase in range, ground backscatter could be proposed. However, the delay can show apparent variations if the modulation depth changes from second to second for relatively long times. If the modulation depth is small then the triggering of the counter occurs later as it takes a longer time before the tone envelope voltage reaches the fixed trigger value of the comparator. This causes an apparent increase in delay.

Visual observations with an oscilloscope of the filtered audio showed that this was often the case: the audio level could remain low for all the modulation period. The effect is not simply related to broad band fading as the AGC could cope with even fast changes in signal level. It is well known, however (e.g. Terman, (Ref.8) ; Davies, (Ref.9)), that selective frequency fading can cause apparent distortion of a double side band signal due to the different frequency components fading independently. This effect would be more noticeable under scatter conditions where many 'reflection' points exist and wave interference due to differing phases can result in frequency-selective fading being observed.

If the carrier fades with respect to both side bands then over-modulation can occur. This was also observed from time to time but it was not as general as 'undermodulation'. On some occasions the modulation was barely visible. The conclusiveness of this experiment is ultimately limited by this effect.

It is useful to observe the distribution of the one second counts within each five minute period. A selection of these are shown in Fig.5. The histograms show the number of counts within each 500 μs interval between the limits of the window. The vertical scale representing the number of counts is a convenient size for each distribution. The distributions for modulation delay and modulation length for three times during the night are shown representing reflection, scatter, and reflection conditions again later in the morning. The data shown in Fig.5 (c) and (d) was recorded approximately twenty minutes after the scatter propagation commenced, corresponding to the arrowed time in Figs.1 and 3.

An effect of the 1 kHz modulation is that the counter tends to trigger preferentially at 1 ms intervals. This can be clearly seen in Fig.5 (e) where the number of counts was zero in a ½ ms interval, yet the number of counts in the intervals on either side was 156 and 74. This effect was due to the period of the modulation being 1 ms. The voltage comparator triggers at a fixed voltage level, however the voltage only increases each 1 ms. The comparator therefore is most likely to trigger from second to second at different 1 ms intervals as the audio amplitude varies from second to second,

During reflection conditions Fig.5 (a) shows a non-Gaussian distribution with mode 23.75 ms and average of 24.62 ms (50 to 15 ms window or 24.39 ms for a 35 to 20 ms window). This type of distribution is common (see Fig.3) where the most probable values (mode) have less delay than the average. The side of the distribution representing minimum delay for the distribution has a sharper edge. This perhaps represents delays suffering minimal fading and hence confined to a narrow interval determined by the group range. The drawn out values representing apparent increased delay (to the left) show the effects of varying amounts of fading. The mode of the distribution would be as meaningful as the average for this reason. Note the very sharp distribution recorded after reflection recommenced at about 0600, the corresponding modulation length data is also comparatively narrow. This indicates an absence of multiple echoes and resultant interferences.

The reason for the tone length distributions not being centred at 50 ms was due to the decay time constant of the envelope smoothing circuit. Fig.5 (d) shows the effect of selective frequency fading in that the average and mode have been shifted to a smaller value. This is due to undermodulation reducing the timed length because of the finite rise time of tone. On this basis the peak in Fig.5 (c) for tone delay is also showing greater apparent delay than that just due to group path.

Referring to Fig 5 (c), when reception is via a scatter mode, although the main peak has moved towards longer delays compared to the distribution at 0200 many returns have the same time delay where most of the returns were received in Fig.5 (a) (arrowed) corresponding to direct F-region reflection. This suggests that even if the main peak at 0300 is due to backscatter that some returns appear to come directly from the F-region. It could be considered that some direct F-region returns were recorded with a majority of ground backscatter. Alternatively the increased delay of the peak might be due to fading as described above.

The possibility that the peak of the distribution shown in Fig.5 (c) corresponds to a ground backscattered echo will now be considered. The expected virtual range of such an echo can be determined using Snell's Law of refraction and an expression for the group delay in the ionosphere (neglecting the magnetic field) obtained by integrating the group refractive index along the ray path. An exact expression of group delay can be found for an assumed parabolic layer (Croft and Hoogasian, (Ref.10)). Snell's Law is used to compute the angle of incidence to the parabolic layer necessary for the ray to become horizontal, whereby it can be propagated out of the layer and strike the ground at some distant point. The range to the base of the layer is calculated and the contribution to the group path of the ray in the ionosphere found using the method of Croft and Hoogasian, (Ref.10). Twice the group path to the point the ray strikes the ground near the edge of the skip zone gives the backscatter group path.

In order to determine the group path the height of the peak of the layer and the layer thickness must be determined. As ionograms were not available from the local station the Canberra ionograms were used, it being assumed that the main difference between the ionosphere above Canberra and near Melbourne was the critical or plasma frequency. The plasma frequency was determined using the time when fxF2 dropped below 4.5 MHz and assuming it decreased at about the same rate as the Canberra value. The plasma frequency was taken to be 3.45 ± 0.05 MHz at 0300 hrs in the calculation. The Canberra 0300 hrs ionogram was reduced to a true height profile, that is

converted from virtual or group height vs. frequency to "true height" or electron density vs. frequency using the method described by Titheridge (Refs.11 and 12). This resulted in a profile that could be well approximated for the calculation by a parabolic F-region of peak height 294 km and thickness 140 km.

The calculations show that the ray cannot become horizontal in the layer and hence propagate to the ground unless the zenith angle of the ray is 43 degrees. The minimum group path occurs when the zenith angle is 58 degrees giving a total backscatter path of 1880 km. This corresponds to a ground scatter at a distance of 770 km.

In order to test if the distribution of Fig. 5 (c) could be due to ground backscatter we must now determine the path length represented by the peak of the distribution. This would normally be best done by noting the group path under reflection conditions then adding the extra range under scatter conditions or by having the local clock supplying the 1 second pulses synchronized with the transmitted tone burst. Neither of these methods were possible. In this case we can use the delay due to reflection from a sporadic-E layer which was present during part of the night (also confirmed by Canberra ionograms) and which invariably has a group height between 100 and 120 km. Reflection from sporadic-E occurred for about 10 minutes near 0420 hrs giving returns from a much reduced delay i.e. much lower than F-region (see Fig.2) and showing increased signal levels. There is little doubt that these signals were from a sporadic-E (Es) layer passing overhead. Assuming the peaks of the distributions from the Es layer correspond to approximately 110 km group path then the delay in Fig. 5 (c) can be estimated from the increased delay over the Es delay plus that of the Es path itself. The extra delay is 6.25 ms corresponding to a group path increase of 1875 km. The total path at 0300 hrs becomes 1875 + 2 x 110 = 2095 km.

The calculations show that a minimum of 1880 km could be expected for ground backscatter, therefore the distribution could be representative of ground backscatter. However, as previously discussed, the distributions have been shifted to greater delay by frequency selective fading and even if this were by only 1 ms then the path falls back to 1795 km, less than the minimum backscatter range. Further, many of the points representing five minutes of data in Fig.3 up to 0330 show delay of 2 ms less than the delay at 0300 which appears to be relatively high. Even if there were no fading at these other times the ranges indicated are near 1500 km, much less than the minimum backscatter range.

It is concluded therefore that the extra delay of the modulation after scattering commences is not due to ground backscatter.

5. CONCLUSION

No jump in delay has been observed when the specular to scatter transition occurs. It has also been observed that some echoes were recorded that had a delay characteristic of direct F-region scatter, and that the extra delay does not appear to be due to ground backscatter. Unfortunately the conclusiveness of this experiment is limited by the widths of the distribution shown in Fig.5 which are due ultimately to wave interference resulting from multiple reflection points. Direct F-region scatter appears to be observed in preference to ground backscatter, for the given experimental conditions.

6. ACKNOWLEDGEMENTS

The Canberra ionosonde data were supplied by the Ionospheric Prediction Service of the Australian Department of Science and the Environment. We would like to thank Dr. L.F. McNamara of the Ionospheric Prediction Service for assistance with the true height determination.

7. REFERENCES

1. Davies, K., "Ionospheric Radio Waves", 1969, Blaisdell Publishing Co., MA., U.S.A.

2. Joyner, K.H., "Phase Height Measurements of the Ionosphere", Ph.D. Thesis Latrobe University, 1975.

3. Cornelius, D.W., Joyner, K.H., Dyson, P.L. and Butcher, E.C., "Scattering from F-region Irregularities at 4.5 and 7.5 MHz", J. Atmos. Terr. Phys., 1975, Vol.37, pp.769-776.

4. Heron, M.L. and Rose, R.J., "The Accuracy of Frequency Standards Disseminated by H.F. Radio Paths through the Ionosphere", Aust. J. Phys. 31, pp.333-345, 1978.

5. Toman, K., "Ionospheric Phase and Group Path", J. Atmos. Terr. Phys., 1967, Vol.29, pp.1019-1023.

6. Dieminger, W., "Geophysical Investigations by Scattered Radio Waves", Radio Science 10, pp.681-691, 1975.

7. Earl, G.F. and Bourne, I.A., "Spectral Characteristics of HF Ground Backscatter", J. Atmos. Terr. Phys., 1975, Vol.37, pp.1339-1347.

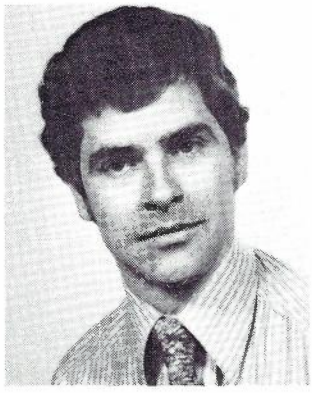
8. Terman, F.E., "Electronic and Radio Engineering", 1955, McGraw-Hill, N.Y.

9. Davies, K., "Ionospheric Radio Propagation", Dover 1966, (National Bureau of Standards Monograph 80, 1965, U.S.A.).

10. Croft, T.A. and Hoogasian, H., "Exact Ray Calculations in a Quasi-Parabolic Ionosphere with no Magnetic Field", Radio Science 3, pp.69-74, 1968.

11. Titheridge, J.E., "The Overlapping Polynomial Analysis of Ionograms", Radio Science 2, pp.1169-1125, 1967.

12. Titheridge, J.E., "The Generalized Polynomial Analysis of Ionograms", Technical Report 78/1 Radio Research Centre, Univ., of Auckland, New Zealand, 1978.



BIOGRAPHIES

DAVID W. CORNELIUS completed a Ph.D. degree at Latrobe University in 1976 studying ionospherically induced frequency shifts on the VNG Standard Frequency Transmissions. Following two years in private industry, he joined the Department of Science and Environment (Ionospheric Prediction Service) to become the senior Australian scientist at the Learmonth Solar Observatory in Western Australia.

ELIZABETH A. ESSEX was born in Grafton, N.S.W. on April 21, 1940. She received the BSc with honours and PhD degrees in Physics from the University of New England in 1961 and 1966 respectively. While working for the PhD degree she held an ICI ANZ fellowship.

After completing the doctoral degree, she continued to carry out research in experimental ionospheric physics. From 1966 to 1968 she held the position of lecturer in Physics at the University of West Indies. In 1969 she joined the Division of Theoretical and Space Physics at La Trobe University, Melbourne, where she now holds the position of senior lecturer. During 1974 she spent six months at the Air Force Cambridge Research Laboratories at Boston, U.S.A. investigating ionospheric gravity wave models. At present she is on leave in the U.S.A. as a NRC/NAS senior resident research associate at the Air Force Geophysics Laboratory, Massachussetts. Currently her interests include ionospheric variability and its investigation by wave propagation techniques. Dr. Essex is a member of the Institute of Physics, Australian Institute of Physics, American Association of Physics Teachers and the American Geophysical Union. She is married with two sons.

New Hardware Realisation of a Transversal Filter

K. S. ENGLISH
Telecom Australia Research Laboratories

Methods of implementing transversal filters for processing continuous signals at audio frequencies are reviewed and a new implementation is described. The proposed structure achieves a hardware simplification by slightly separating the sampling interval from the unit delays and is particularly suitable for applications such as adaptive filtering where the filter coefficients must be digitally programmable.

1. INTRODUCTION

The term transversal filter is applied to any filter in which the output is generated from a weighted sum of delayed versions of the input. Transversal filters are particularly useful in applications requiring synthesis of arbitrary time-domain responses. Some representative applications in the communications field are pulse-forming and equalizer circuits, in frequency bands ranging from sub-audio through to UHF. Transversal filters are also fundamental building blocks in adaptive filters, where simple correlation techniques are used to adapt the filter coefficients so as to track the response of an external circuit. Many diverse applications exist for adaptive filters, including adaptive equalization of data channels (Ref.1) or room acoustics, echo cancellation (Ref.2) and noise suppression (Ref.3). Adaptive filters have been widely studied and technical feasibility of such systems has been established. However, many elegant applications for adaptive filters have failed to achieve commercial status solely on economic grounds. Contributions from L.S.I. technology will help to reverse this trend, but there is still a need for simple structures tailored to particular applications without requiring customised components.

The present paper describes a new structure* for a low-cost transversal filter, meeting the following broad specifications:

(i) operation to be at audio or sub-audio frequencies,

(ii) filter coefficients must be digitally programmable,

(iii) the output must be a continuous (i.e. not sampled) signal.

These specifications are representative of many problems in the fields of communications and control. The intended application for the proposal is a new low-cost adaptive echo canceller, which will be described at a later date.

* Covered by Australian Provisional Patent App. No.PD7273/79.

2. REVIEW OF IMPLEMENTATION TECHNIQUES

Transversal filters based on tapped analogue delay lines Fig.1 implement an impulse response of the form:

$$h(\tau) = \sum_{i=0}^{N-1} h_i \, \delta(\tau - iT) \qquad (1)$$

where $\delta(\tau)$ is the Dirac delta function, T is the delay between tapped outputs and h_i are the weighting coefficients.

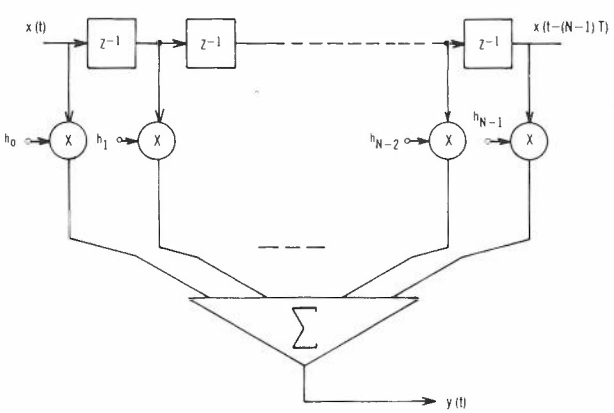


Fig.1 - Basic transversal filter using continuous delay line.

Good continuous delay lines are difficult to construct at audio frequencies and it is customary to sample the input signal and implement the delays using either analogue or digital shift registers. Lowpass filtering is then required prior to sampling to avoid aliasing, and a lowpass filter is usually necessary at the output to restore a linear time-invariant input/output relationship.

Denote the input signal by x(t) and the output from the filter by y(t). In a sampled-data implementation x(t) is sampled at intervals T

generating an input sequence $x_k = x(kT)$, and a sequence of output samples y_k according to:

$$y_k = \sum_{i=0}^{N-1} h_i \, x_{k.i} \qquad (2)$$

The output $y(t)$ is obtained by lowpass filtering a signal $z(t)$ which comprises piecewise constant segments of duration T as given by:

$$z(t) = y_k \quad , \; kT \leq t < (k+1).T \qquad (3)$$
$$\quad , \; k = -\infty, \ldots . \infty$$

In order to derive the overall response, a convolution representation of waveform $z(t)$ is more convenient:

$$z(t) = \emptyset_T(t) * \hat{y}(t) \qquad (4)$$

where

$$\hat{y}(t) = \sum_{n=-\infty}^{\infty} y_n \, \delta(t-nT) \qquad (5)$$

$$\emptyset_T(t) = \begin{cases} 1 & , \; 0 \leq t < T \\ 0 & , \; \text{otherwise} \end{cases} \qquad (6)$$

Using a similar impulse representation $\hat{x}(t)$ of the input sequence, we have from (2) and (5):

$$\hat{y}(t) = \sum_{i=0}^{N-1} h_i \, \hat{x}(t-iT) \qquad (7)$$

where

$$\hat{x}(t) = \sum_{n=-\infty}^{\infty} x_n \, \delta(t-nT) \qquad (8)$$

Let $q(\tau)$ denote the impulse response of an ideal lowpass filter with cutoff frequency $\frac{\pi}{T}$.

The sampling theorem provides the identity:

$$q(t) * \hat{x}(t) = q(t) * x(t) \qquad (9)$$

provided that $x(t)$ contains no frequency components above $\frac{\pi}{T}$. Using $q(\tau)$ to represent the lowpass filtering of $z(t)$, from (4), (7) and (9) we obtain:

$$y(t) = q(t) * z(t)$$

$$= q(t) * \emptyset_T(t) * \sum_{i=0}^{N-1} h_i \, \hat{x}(t-iT)$$

$$= q(t) * \emptyset_T(t) * \sum_{i=0}^{N-1} h_i \, x(t-iT) \qquad (10)$$

Thus the usual sampled data implementation based on discrete convolution realises an impulse response of the form:

$$h(\tau) = q(\tau) * \emptyset_T(\tau) * \sum_{i=0}^{N-1} h_i \, \delta(\tau-iT) \qquad (11)$$

Equation (11) remains substantially true when the ideal lowpass response $q(\tau)$ is replaced by a physically realisable approximation. The zero-order hold term $\emptyset_T(\tau)$ contributes a group delay of T/2 and 3.9 dB roll-off over the passband $(0, \frac{\pi}{T})$ which can be absorbed by a complementary modification to $q(\tau)$.

Charge-coupled device (CCD) and bucket-brigade device (BBD) shift registers are very attractive for implementing transversal filters. Both technologies are comparatively new (post 1970) and avoid the complex analogue/digital and digital/analogue conversion processes in digital implementations. If the filter coefficients are altered infrequently then the weighted sum function can be realized very simply using resistive coupling from each delayed output to the summing junction of an operational amplifier Fig.2. The overall impulse response will be quite close to that given in (11). Some progress has also been reported in realizing CCD transversal filters with programmable coefficients (Refs.4,5).

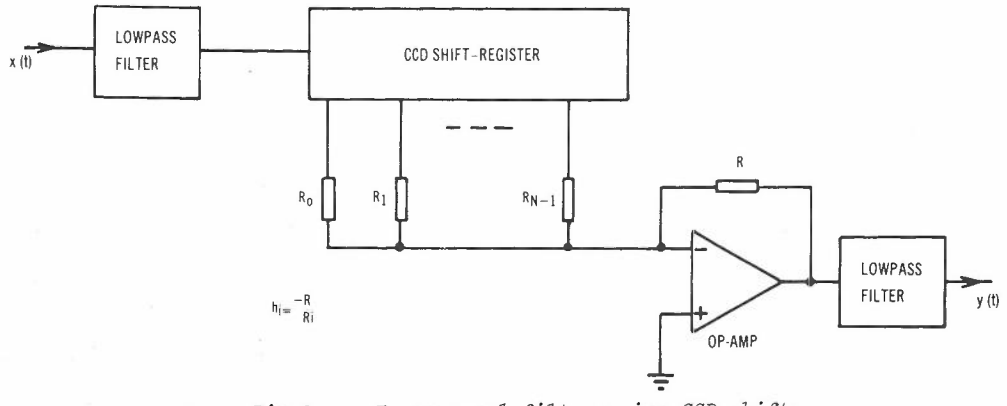


Fig.2 - Transversal filter using CCD shift register and resistive coupling.

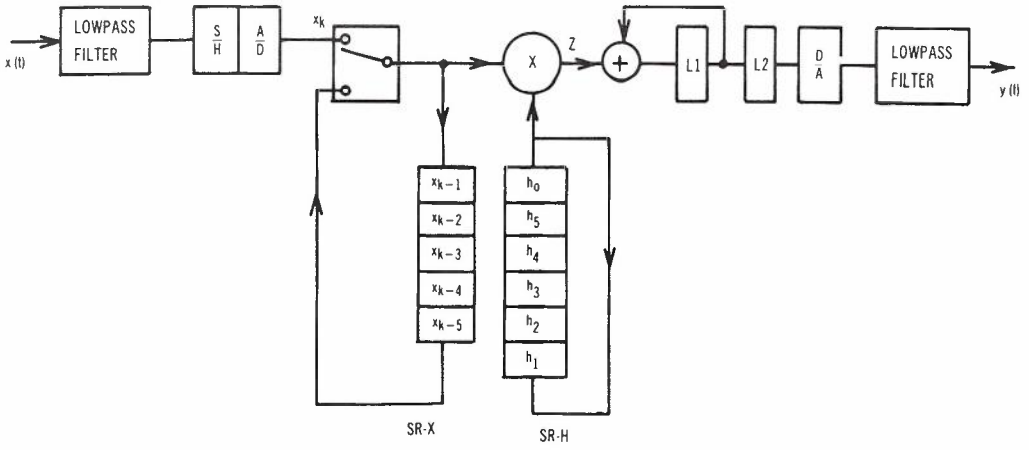


Fig.3 - Fully digital transversal filter with six coefficients.

Digital techniques are generally preferred when the filter coefficients must be programmable. The dominant cost of a digital implementation is the N-fold multiplication evident in equation (2). A well-known word-serial structure requiring only one multiplier and a single accumulator is illustrated in Fig.3. Recirculating shift registers store the input and coefficient samples, and are clocked at a rate N times greater than the sample rate. The digital multiplier must be capable of performing N multiplications in one sample interval. Figure 3 shows the register contents at the instant of acquiring a new sample x_k, which replaces the oldest sample in the SR-X register. Subsequently, the multiplexer switches to the output of SR-X and registers SR-X and SR-H are recirculated once before the next sample is taken. The output y_{k+1} is accumulated in latch L1 over this period then transferred to latch L2 for a further period T. The implementation in Fig.3 adds a calculation delay of at least T/N to the response of (11). Details of the method of setting the coefficients in SR-H have been omitted from Fig.3. If the coefficients are fixed then shift register SR-H could be replaced by a programmable read-only memory (PROM) programmed with the filter coefficients. When the transversal filter formed part of an adaptive filter, a digital adder would be inserted in the SR-H recirculation loop so that small corrections can be added into the stored coefficients during processing (Refs.2,3). By altering the bit-level operations of this filter numerous variations can be constructed offering a range of speed/complexity tradeoffs.

3. SIMPLIFIED STRUCTURE

An essential feature of the sampled-data implementations reviewed in the preceding section was equality between the input sampling interval and the unit delays realised in the overall impulse response (or perhaps an integer multiple of the sampling interval). This was necessary for the sampling operation to properly transform the problem from continuous-time to discrete-time. The proposed structure acquires input samples at intervals T but realises an impulse response, analogous to (11), of the form:

$$h(\tau) = q(\tau) * \phi_{\frac{T}{N}}(\tau) * \sum_{i=0}^{N-1} h_i \, \delta(\tau - iT') \qquad (12)$$

where

$$T' = T\left(1 - \frac{1}{N}\right) \qquad (13)$$

and

$$\phi_{\frac{T}{N}}(\tau) = \begin{cases} 1 & , \quad 0 \le \tau < \frac{T}{N} \\ 0 & , \quad \text{otherwise} \end{cases} \qquad (14)$$

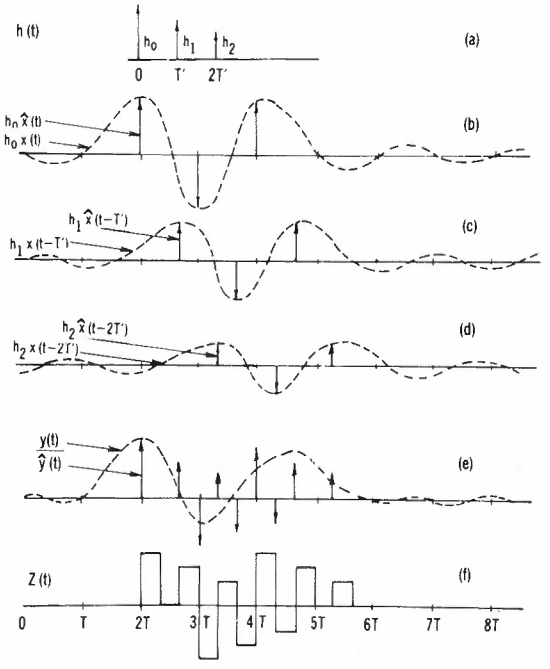


Fig.4 - Illustrative waveforms from simplified implementation:

(a) impulse response to be synthesized,

(b), (c), (d) delayed and scaled versions of input waveform x(t) and sampled waveform $\hat{x}(t)$;

(e) desired output waveform y(t) and sampled version $\hat{y}(t)$;

(f) a physically realisable approximation to $\hat{y}(t)$.

Ignoring the fixed filtering, for a given set of coefficients h_i, equation (12) describes a slightly time-compressed version of the impulse-response in (11). Whereas a band-limited impulse response $g(\tau)$ is best approximated in (11) by choosing $h_i \simeq g(iT)$, the modified response (12) requires $h_i \simeq g(iT')$. For typical values of N (32 to 256) this slight compression of the impulse response is virtually undetectable in the overall response, but it leads to a favourable simplification in the hardware implementation. Figure 4 illustrates the manner in which this is achieved. Consider the problem of realising the 3 coefficient impulse response shown in Fig.4a, for input signals $x(t)$ bandlimited to π/T. Figure 4b illustrates a representative waveform $x(t)$ (shown by dashed line) and also $\hat{x}(t)$, the impulse representation of the result of sampling $x(t)$ at unit intervals T, both waveforms scaled by coefficient h_0. Figures 4c and 4d show delayed and scaled replicas of both $x(t)$ and $\hat{x}(t)$. In each of the Figures 4b,c,d, the continuous waveform $h_i x(t-iT')$ can be reconstructed from the sampled version $h_i x(t-iT')$ by lowpass filtering with cutoff frequency π/T.

The desired output $y(t)$ (shown dashed in Fig.4e) results from superimposing the continuous waveforms in 4b,c,d. Superimposing the sampled waveforms in 4b,c,d obtains the waveform $\hat{y}(t)$, also shown in Fig.4e.

$$y(t) = \sum_{i=0}^{2} h_i \, x(t-iT') \tag{15}$$

$$\hat{y}(t) = \sum_{i=0}^{2} h_i \, \hat{x}(t-iT') \tag{16}$$

The output $y(t)$ can be generated by lowpass filtering the physically unrealisable waveform $\hat{y}(t)$. Figure 4f illustrates a physically realisable form of waveform $\hat{y}(t)$, denoted by $z(t)$ and represented by the convolution of $\hat{y}(t)$ and $\phi_{T/3}(t)$ as defined in (14). Figure 4e differs from a similar representation of the conventional implementation in that the component impulses all occur at separate time instants. This temporal separation enables $z(t)$ to be generated by presenting each product $h_i x_{k-i}$ at the output for a duration of T/3, without any explicit summation operation.

For general values of N the desired output can be generated by lowpass filtering a waveform $z(t)$ with the composition:

$$z(t) = \phi_{\frac{T}{N}}(t) * \hat{y}(t) \tag{17}$$

$$\hat{y}(t) = \sum_{i=0}^{N-1} h_i \, \hat{x}(t-iT(1-\tfrac{1}{N})) \tag{18}$$

Substituting from (8) and making a change of variable:

$$z(t) = \sum_{i=0}^{N-1} h_i \phi_{\frac{T}{N}}(t) * \sum_{n=-\infty}^{\infty} x_n \delta(t-iT(1-\tfrac{1}{N})-nT)$$

$$= \sum_{i=0}^{N-1} \sum_{k=-\infty}^{\infty} h_i x_{k-i} \phi_{\frac{T}{N}}(t-kT+\tfrac{iT}{N}) \tag{19}$$

The desired waveform $z(t)$ must comprise piecewise constant segments of duration $\frac{T}{N}$ as given by:

$$z(t) = h_i x_{k-i} \quad , \quad kT - \tfrac{iT}{N} \leq t < kT - \tfrac{iT}{N} + \tfrac{T}{N} \tag{20}$$

$$, \quad k = -\infty, ---, \infty$$

$$, \quad i = 0, ---, N-1$$

Examination of Fig.3 will reveal that a digital representation of waveform $z(t)$ is actually generated at the multiplier output (labelled "z") Thus the accumulator L1 and latch L2 can be omitted without substantially altering the overall response. This represents a small saving in hardware but the major cost is still in the digital multiplier.

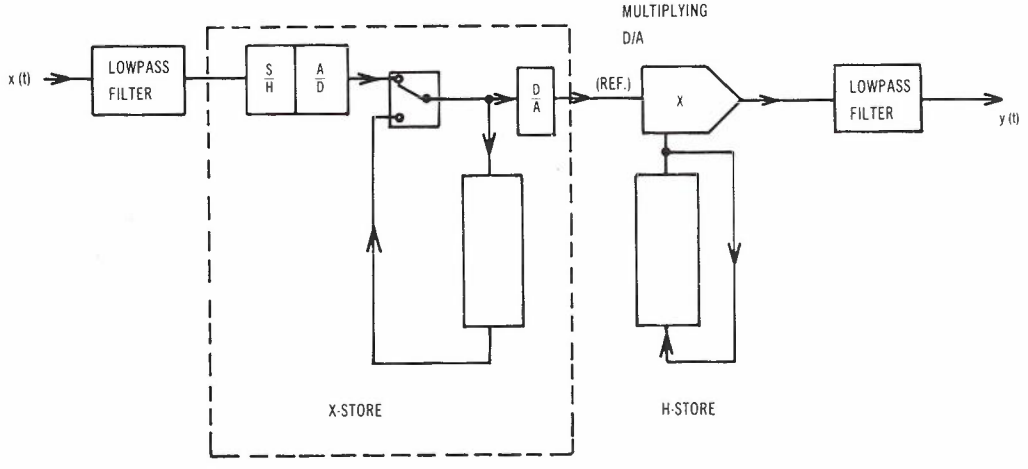


Fig.5 - Transversal filter using the proposed structure.

Practical utility of the proposal derives from the ease with which the simplified structure can be further modified to eliminate the digital multiplier. Figure 5 shows a variation where the multiplication and D/A conversion functions are combined into a multiplying D/A with a second D/A driving the reference input. Furthermore, both input and output of the block labelled "x-store" are analogue signals. The whole of the function internal to this block can be performed by a single analogue integrated circuit such as the Reticon SAM-128V device. With the current availability of good quality multiplying D/A convertors at quite reasonable cost, Fig.5 provides the basis for a practical low-cost hardware implementation of a transversal filter. A prototype filter with 128 coefficients and 8 kHz sample rate has been constructed and verified the theoretical basis for the method.

4. CONCLUSION

The primary advantage of the proposed realisation is the structural simplification it achieves. Separation of the unit delays from the sampling interval enables the summation operation to be absorbed within the smoothing action of the output lowpass filter. Eliminating the accumulator enables the multiplication operation to be efficiently performed using a multiplying D/A convertor. No approximations are introduced other than the usual idealised lowpass filters tacitly assumed in sampled-data theory. Performance of the output filter is made slightly more critical by the virtual absence of any zero-order hold at the output, but not prohibitively so. In some applications this output filter will be an existing sharp cutoff filter contained within subsequent transmission equipment.

The fields of application for the proposed scheme cover the low frequency applications where the filter coefficients must be digitally programmable. Where the latter feature is not essential, the proposal is unlikely to be competitive with a CCD implementation using resistively coupled output Fig.2. The limiting factor on the sampling rate is the output D/A convertor settling time. Bearing in mind that low cost was the principle justification for the new structure, if too much upgrading of the D/A becomes necessary then it is probably expedient to revert to the fully digital implementation Fig.3 where the settling time can be N times slower. Nevertheless, the rates achievable with the new structure encompass a number of practical applications.

5. REFERENCES

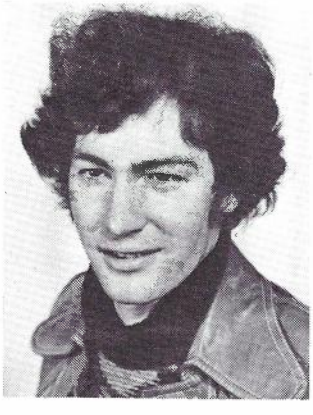
1. Lucky, R., "Automatic Equalization for Digital Communication", Bell Syst. Tech. J., Vol.44, pp.547-588, April 1965.

2. Demytko, N. and English, K.S., "Echo Cancellation on Time-Variant Circuits", Proc. IEEE, Vol.65, No.3, pp.444-453, March 1977.

3. Widrow, B., et al., "Adaptive Noise Cancelling: Principles and Applications", Proc. IEEE, Vol.63, No.12, pp.1692-1716, Dec. 1975.

4. Tiemann, J.J., et al., "Charge-transfer Devices Filter Complex Communications Signals", Electronics, Nov.14, 1974, pp.113-116.

5. Denyer, P.B., Mavor, J. and Arthur, J.W., Miniature Programmable Transversal Filter Using CCD/MOS Technology", Proc. IEEE, Vol.67, No.1, pp.42-50, January 1979.



BIOGRAPHY

KEVIN S. ENGLISH was born in Melbourne in 1952 and received the B.E. (Hons) degree from Monash University in 1974. He has worked in the Computer Applications Section of the Telecom Australia Research Laboratories since 1974, involved in studies of adaptive echo cancellation and computer analysis of experimental data.

Theory of Coupled Transmission Lines and its Application to Optical Fibres

A. E. KARBOWIAK
School of Electrical Engineering
University of New South Wales
Kensington N.S.W. 2033

D. H. IRVING
School of Electrical Engineering
University of New South Wales
Kensington N.S.W. 2033

Coupled transmission line theory is examined with particular reference to modelling signal propagation along multimode optical fibres. Among others, it is shown that the model is a simple way of obtaining results pertaining to optical fibres. The usefulness of the method is illustrated by application to the solution of a number of specific problems.

An examination of the properties of the model shows that by connecting one transmission line system to another having complementary characteristics it is possible to realize a system with improved bandwidth handling characteristics. This is in support of the observation that when an undercompensated graded index fibre is spliced alternately with an overcompensated fibre a system of superior quality results.

1. INTRODUCTION

Any waveguiding system be it a coaxial cable, a multimode waveguide or an optical fibre can be modelled exactly by an appropriate set of coupled transmission lines (Ref.1). Such a representation has the merit of simplicity and is a powerful aid in the solution of more complex problems, such as those encountered with the trunk waveguide system (Refs.2,3). The exact representation involves a set of n-transmission lines (one for each degree of freedom) with the equivalent voltages and currents on each of the lines. The lines are coupled with each other using "voltage" or "current" terms as appropriate to the model. When dealing with transmission systems having a very large number of degrees of freedom a simpler and often adequate representation in terms of "power waves" is preferred. This is the case with many problems concerning the Trunk Waveguide System (Ref.4) or an optical fibre.

In this paper we present a simple theory of coupled transmission lines with a view to modelling optical fibres. We shall see that this model can be used to solve a number of problems equivalent to those encountered with multimode optical fibres and, because the model is easy to visualize, the method is a powerful aid in suggesting solutions to complex problems encountered with optical fibres. We illustrate the use of the model by solving a number of specific problems. Among others, the model suggests means by which the bandwidth handling capacity of sub-optimal graded index fibres can be significantly increased.

2. DESCRIPTION OF THE COUPLED TRANSMISSION LINE MODEL

Depending on applications we may consider a discrete or a continuous model. With the discrete model we consider a set of n-transmission lines of total length $z = L$. In this application we shall consider "power coupling" from line to line which is defined by the matrix of the coupling functions $\underline{\underline{c}}(x)$. With this model, if the amplitudes on lines i and j are A(i) and A(j) respectively then c(i,j,z) describes the amount of coupling between A(i) and A(j) as a function of distance z, along the line. More specifically, if $A^2(z)$ and $B^2(z)$ describe the power in lines i and j respectively at a point then $A^2(z).c^2(z)\,dz$ is the amount of power coupled from line i to line j in distance dz, and $B^2(z).c^2(z)\,dz$ describes the amount of power coupled from line j to line i. A(i) and A(j) can therefore be considered to be quantities proportional to square root of power in the respective lines.

In some applications we can consider the coupling functions to be constants and independent of the modes considered. We shall then find that the results so obtained are in agreement with those obtained from statistical considerations (Ref.5) where we set the constant coupling coefficient equal in value to the rms value of the assumed random coupling.

In other cases solutions for non-constant coupling is required. The model permits such solutions but in many cases closed form solu-

tions cannot be obtained. Computer solutions, however, are still relatively simple to obtain.

If we wish to model a multimode fibre with a continuum of rays then the appropriate model is one consisting of a continuum of coupled transmission lines. This is obtained as a limit of the discrete model. The impulse response of such models has all the features of impulse responses obtained with optical fibres.

3. SIGNAL RESPONSE OF AN UNCOUPLED TRANSMISSION LINE SYSTEM

The uncoupled transmission line model corresponds to a uniform graded optical fibre free from statistical imperfections.

We consider a continuum of transmission lines in variable x, $x \varepsilon (0,1)$. The initial condition (or launching condition, or illumination) at the origin $z = 0$ can be described by a normalised illumination function $I(x)$, such that $\int_0^1 I(x).dx = 1$. If a unit step signal is applied at the input end ($z = 0$) to the transmission line system, at time $t = 0$, then each elemental transmission line $(x, x+dx)$ will be excited by a fraction $I(x).dx$ of the input energy.

The system of transmission lines is characterised by an attenuation profile $\alpha(x)$ and a velocity profile $u(x)$, a single valued function of x. Clearly, $\alpha(x)$ is the attenuation of the elemental line x and $u(x)$ is the corresponding signal velocity.

It should be noted that the propagation process is modelled by means of one parameter, x, only. The model is quite general and can be used to represent propagation in optical fibres where it is customary to use the parameter I and β (see Reference 6). In general, the rays/modes can be ordered in decreasing velocity by mapping into x on a one-to-one correspondence and in the process the whole range of I and β values will be covered. Clearly, therefore, $u(x)$ is always a monotonic single valued function.

The general problem can now be stated as follows: Given the illumination function $I(x)$, for a system with given $\alpha(x)$ and $u(x)$ profiles, determine the output signal if the system is L metres long and the input signal at $x = 0$ is given. To solve the general problem it will be sufficient to determine the impulse response or the response to a unit step signal, the Edge Distortion Function (EDF). We proceed to determine EDF.

The time of arrival of the head of the unit step function along line x is clearly

$$t(x) = L/u(x) = Lf(x) \qquad (1)$$

The range of $t(x)$, corresponding to the extreme values of $u(x)$, is $(t_0, t_1) = (L/u_0, L/u_1)$.

This relation can be inverted to read

$$x(t) = F_u(t) \qquad (2)$$

Equation (2) describes the EDF in the interval (t_0, t_1) for the case $I(x) = 1$ and $\alpha(x) = 0$ for all x. To obtain EDF, in general, $x(t)$ in (2) must be weighted by $I(x)$ and $\exp - \alpha(x)$. That is in the interval (t_0, t_1) the EDF is given by

$$h_u(t) = \int_0^{F_u(t)} I(x) \,.\, \exp{-\alpha x}.dx \qquad (3)$$

which reduces to (2) for $I = 1$ and $\alpha = 0$. Equation (3) is a minor modification of (2), and for the purposes of illustration it will be sufficient to consider (2).

In absence of attenuation and for $I = 1$ the EDF is, therefore,

$$x_u(t) = 0 \text{ for } t < L/u_0$$

$$= F_u(t), \ \frac{L}{u_0} \leq t \leq \frac{L}{u_1}$$

$$= 1 \text{ for } t > L/u_1 \qquad (4)$$

The impulse response is, of course, given by

$$x_i(t) = x_u'(t) \qquad (5)$$

Two simple examples will help to clarify the relations.

Example 1:

Consider the velocity profile (Fig.1)

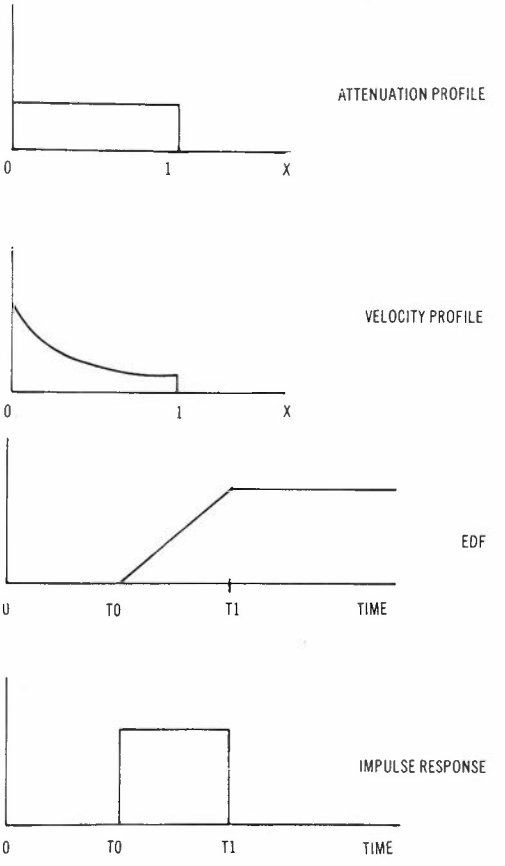


Fig.1 - Fibre response example 1: $\alpha(x) = const.$, $u(x) = a/(b+x)$

$$u = \frac{a}{b+x} \quad , \quad 0 < x < 1 \qquad (6)$$

Substituting this in (1) and inverting, gives for the unit step response (EDF).

$$x_u(t) = 0, \text{ for } t < \frac{L}{u_0}$$

$$= \frac{t/L - 1/u_0}{1/u_1 - 1/u_0} \text{ , for } \frac{L}{u_0} \leq t \leq \frac{L}{u_1}$$

$$= 1, \text{ for } t > \frac{L}{u_1} \qquad (7)$$

Thus the EDF response is a ramp.

The impulse response is clearly

$$x_i(t) = \frac{L}{L(1/u_1 - 1/u_0)} \text{ , } \frac{L}{u_0} \leq t \leq \frac{L}{u_1}$$

$$= 0 \quad \text{otherwise} \qquad (8)$$

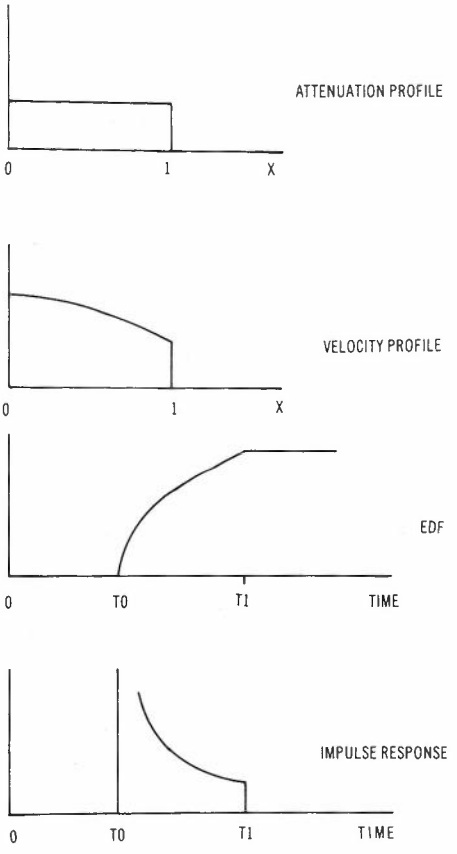


Fig.2 - Fibre response example 2: $\alpha(x) = const.$, $u(x) = u_0 \cos Bx$

Example 2:

Consider the velocity profile (Fig.2)

$$u(x) = u_0 \cdot \cos Bx \qquad (9)$$

This corresponds to a one-dimensional step index fibre, a slab waveguide, having half-acceptance angle of B radians.

Substituting this expression in (1) and solving for x, we obtain for the EDF

$$x_m(t) = 0, \quad t < \frac{L}{u_0}$$

$$= \frac{\cos^{-1}(L/u_0 t)}{\cos^{-1}(u_1/u_0)} \text{ , } \frac{L}{u_0} < t < \frac{L}{u_1}$$

$$= 1, \quad t > \frac{L}{u_1} \qquad (10)$$

The impulse response is of course the derivative of the above and has the familiar form shown in Fig.2.

In the above we assumed a constant attenuation profile. In general $\alpha = \alpha(x)$, and consequently to obtain EDF the expression (3) has to be used.

Usually the attenuation profile has a relatively small effect on the fibre response, the major effect comes from the velocity profile and the initial conditions at $t = 0$ which in the above examples we have taken as $x(0) = 1$, $x \varepsilon (0,1)$.

4. STEADY STATE POWER DISTRIBUTION PROFILES

In absence of coupling the power distribution profile is determined by the launching conditions modified by the attenuation profile $\alpha(x)$. However, inter-line coupling affects the power distribution profile profoundly as the following analysis demonstrates.

Consider the discrete line model. Let V_i be the voltage (as defined previously) on the i-th line. Further, we assume for the moment that the coupling and attenuation profiles are constants, c and α, respectively.

After a sufficient distance the coupling will bring about a power distribution profile which is peculiar to the particular selection of c and α, independently of launching conditions.* For this to be so the difference equation

$$V(i)(\alpha+2c) = c\left[V(i-1)+V(i+1)\right] \qquad (11)$$

must be satisfied. Therefore, the required power distribution profile V(i) is a solution to (11).

---

* This is a consequence of the equipartition theorem of statistical mechanics (see for example: R.C. Tolman, "The Principles of Statistical Mechanics", Oxford University Press, 1959, p.95).

The general solution to (11) is **

$$V(i) = A_1 a_1^i + B_2 a_2^i \qquad (12)$$

where a_1, a_2 are the two solutions of the characteristic equation. These are

$$a_{1,2} = \frac{\alpha+2c \pm \sqrt{\alpha^2 + 4c\alpha}}{2c} \qquad (13)$$

The constants of integration need to be determined from the boundary conditions. These might be chosen as follows. The lines 1 to N-1 might represent the modes of propagation while line N models the energy scattered by radiation and therefore lost by the system, in which case we set V(n) = 0. To obtain a normalised plot of V(i) we choose V(0) = 1. These conditions determine the constants A_1 and A_2 as follows:

$$A_1 = 1/\left[(a_1/a_2)^N - 1\right] \qquad A_2 = 1/\left[1-(a_2/a_1)^N\right] \qquad (14)$$

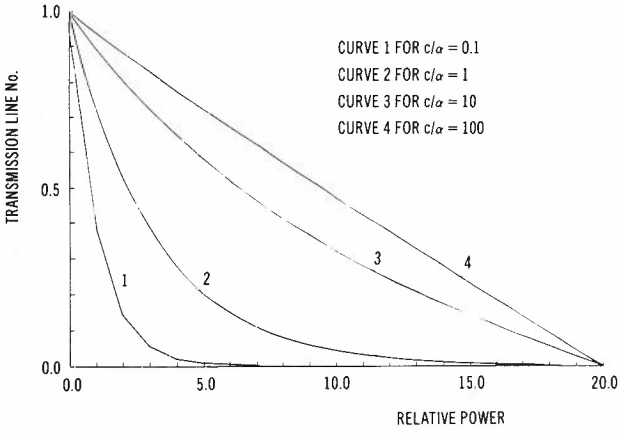


Fig.3 - Voltage distribution profiles for a 20-transmission line system with constant attenuation and coupling.

Figure 3 shows plots of voltage distribution profiles for a 20-transmission line model and a range of c/α ratios. We note that once the ratio c/α is larger than 10, then the voltage profile is substantially independent of this ratio.

If the coupling and attenuation profiles are not constants (independent of i) then equation (11) becomes a difference equation with variable coefficients.

$$c(i)\left[V(i-1)+V(i+1)\right]-V(i)\left[\alpha(i)+2c(i)\right] = 0 \qquad (15)$$

While in general a closed form solution is not possible a computer solution for any α(i) and c(i) profiles is easily obtained. A simple algorithm (though not a particularly quickly convergent one) is to assume a reasonable V(i) profile and use the left-hand side of (15) iteratively as a correction to V(i) vector until a solution to the required accuracy is obtained.

** See for example: P.M. Morse and H. Feshbach, "Methods of Theoretical Physics", (McGraw-Hill, 1953, p.692).

Figure 4 shows a selection of results for a range of α(i) and c(i) profiles for a 10-transmission line system. Once again the results are very much dependent on the choice of the coupling profile.

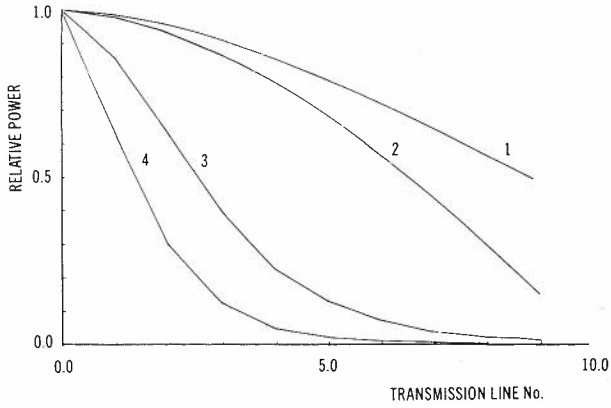


Fig.4 - Voltage distribution profiles for lines having mode dependent attenuation and coupling.

Curve 1 α = 0
c = (0.3, 0.3....0.3, 0.1)

Curve 2 α = 0
c = (0.3, 0.3....0.3)

Curve 3 α = (0.01, 0.02, 0.05, 0.1, 0.15, 0.15, 0.2, 0.2, 0.25)
c = as for curve 2

Curve 4 α = as for curve 3
c = (0.05, 0.07, 0.1, 0.12, 0.15, 0.17, 0.2, 0.22, 0.25, 0.05)

5. IMPULSE RESPONSE OF UNIFORMLY COUPLED LINES

A complete treatment of uniformly coupled transmission lines is possible and various solutions under a variety of assumptions, can be obtained. Here we show a very simple derivation, from first principles, leading to a solution of wide applicability.

In the first instance we consider a two-line model.

Let the velocities and attenuations on lines 1 and 2 be u_1, u_2 and α_1, α_2 respectively. Then the impulse response of the system L metres long will be confined to the interval $L/u_1 - L/u_0$.

With optical fibres a significant amount of mode mixing may take place in a critical distance of say 100 or 1000 m. Let us therefore assume that in a distance l_c a 50% power transfer takes place from mode 1 to mode 2 and vice versa. We thus have the discrete model shown in Fig.5.

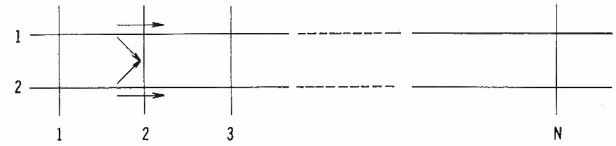


Fig.5 - Discrete model of uniformly coupled lines.

The solution for large $n(=L/l_c)$ will, clearly, be independent of the initial condition: we arbitrarily assume an impulse input to line 1 only and examine the scattering at the n points along the line. The energy will travel along the line in a multitude of ways but every contribution arriving at z = L along any particular path will have the amplitude $(0.5)^n$, which even for n = 100 is microscopically small.

A typical path will include k intervals in line 2 and n-k in line 1. The energy arriving along this path will be delayed by $k.\delta$ where

$$\delta = l_c(1/u_1 - 1/u_0) \tag{16}$$

In all, there are C_k^n such paths. Therefore, the total contribution to amplitude at time

$$t_k = L/u_0 + k.\delta \tag{17}$$

is

$$h(k) = (0.5)^n.C_k^n \tag{18}$$

Similarly

$$h(k+1) = (0.5)^n.C_{k+1}^n \tag{19}$$

The solution for the impulse response is therefore contained in the following difference equation

$$h(k+1)-h(k) = (0.5)^n(C_{k+1}^n-C_k^n) \tag{20}$$

But

$$C_{k+1}^n - C_k^n = (\frac{n-k}{k+1} -1)\ C_k^n \tag{21}$$

Let

$$k = n_0 + i \tag{22}$$

with $n_0 = n/2$

then (20) transforms into the equivalent form

$$h(i+1) - h(i) = -h(i)(\frac{2i+1}{n_0+i+1})$$

For reasons which will become apparent we are interested in the solution for which $n_0 >> i+1$. The equation simplifies to

$$h(i+1) - h(i) = -h(i)(\frac{2i+1}{n_0}) \tag{23}$$

One way of obtaining a solution to (23) is to proceed to the limit: the difference equation then passes into a differential equation

$$\frac{dh}{dx} = -h(x)\ .\ \frac{2x}{n_0} \tag{24}$$

The solution is

$$h(x) = A\ e^{-x^2/n_0} \tag{25}$$

which can be more conveniently, expressed as a normalised function of t

$$h(t) = \frac{1}{\sigma\sqrt{2\pi}}\ e^{-\frac{1}{2}\ (\frac{t}{\sigma})^2} \tag{26}$$

where

$$\sigma = \frac{1}{2}\ (\ \frac{1}{u_1} - \frac{1}{u_0}\ )\sqrt{L\ l_c} \tag{27}$$

We recall L is the total length of the transmission line system and l_c is the length required for 50% mode conversion.

The solution exhibits two important characteristics: the width of the impulse response increases (1) with $\sqrt{L}$, and (2) with $\sqrt{l_c}$. Figure 6 shows results of calculations for a system with l_c = 100 m, L = 10 km, having velocities $u_0 = c$ and $u_1 = c(1-10^{-3})$. The rms width of the impulse response is clearly 1.67 μsec which represents 10 fold improvement in bandwidth handling capacity in absence of coupling.

Therefore to narrow down the impulse response under uniform coupling conditions, it is necessary to reduce l_c, that is to increase the degree of mode coupling.

Attenuation on the whole has a relatively minor influence on the shape of the impulse response. This can be shown as follows.

Let α_1 and α_2 be the attenuations in line 1 and 2 respectively then at the mid-point of the time interval $(L/u_1,\ L/u_0)$ the contribution to the impulse response will be attenuated by the factor exp - $L(\alpha_1 + \alpha_2)/2$. This must be so because the particular contribution would have been due to that part of the signal that has travelled half the distance in line 1 and the other half in line 2. At other times the impulse response shown in (25) needs to be multiplied by the factor exp - $t(\alpha_2 - \alpha_1)\Delta u$. With this modification the impulse response becomes

$$h_\alpha(t) = \frac{E}{\sigma\sqrt{2\pi}}\ .\ \exp - \frac{1}{2}(\frac{t-t_\alpha}{\sigma}) \tag{28}$$

where

$$\tau_\alpha = \frac{\Delta\alpha}{4}\left(\frac{1}{u_1} - \frac{1}{u_0}\right)^3 \qquad (29)$$

The factor $E = \exp - L(\alpha_1 + \alpha_2)/2$ and $\Delta\alpha = \alpha_2 - \alpha_1$.

This shows that attenuation has no effect on the width of the impulse response. There is however a trade-off between loss and bandwidth as discussed elsewhere.*

Finally, if the transmission system consists of n-lines rather than just two, then instead of the term C_k^n in equation (18) we have $C^n_{k1,\ k2\ \cdots\cdots kn}$. That is we are considering combinations of k_1 sites with delay δ_i, k_2 sites with delay δ_2, etc., in n, with the constraint that the total delay ($\Sigma k_i\delta_i$) is constant. Clearly, therefore, the solution is essentially of the same form as that shown in equation (26).

We note that equation (27) agrees precisely with that derived by Personik (Ref.5).

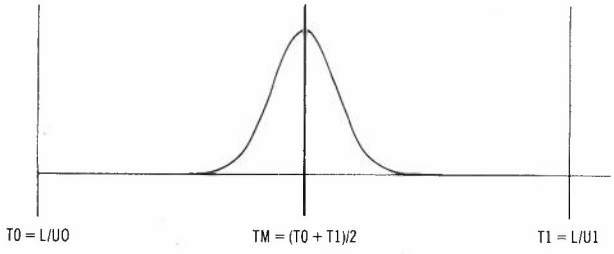


Fig.6 - Impulse response of a 2-line system with uniform coupling: L = 10 km, l_c = 100 m, $u_0 = c$, $u_1 = u_0(1-10^{-3})$.

6. IMPULSE RESPONSE OF OPTIMALLY COUPLED LINES

We have seen in the last section that uniformly coupled lines have, always, a better impulse response than uncoupled lines. Moreover, it is not necessary for the lines to be uniformly coupled: The nature of the solution will be similar if instead of uniform coupling we have periodic coupling, with the same average intensity.

The remaining question is: Is it possible to couple lines in a specific manner and reduce the impulse response further? We have shown that with mode mixing this is not possible.

However, given a set of transmission lines i, having a velocity profile u(i) and an attenuation profile α(i) it is possible, in principle, to couple lines in specific ways (mix modes selectively) at frequent intervals, by means such as line transposition. For example, given a linear velocity profile, we can transpose periodically lines (1 to n) with indices adding up to n+1 resulting in a distortion-free system. In practice perfect compensation is unlikely, but a considerable improvement in impulse response is certainly possible with approximate complementary matching.

We can therefore confidently conjecture that by splicing alternately fibres which on average complement one another velocity profiles a system having wider bandwidth will result. We shall show in a subsequent paper that this indeed is true of at least one class of optical fibres.

7. ACKNOWLEDGEMENTS

The work described has been carried out as part of a research programme of the Department of Communications supported by a research grant from the Radio Research Board.

8. REFERENCES

1. Schelkunoff, S.A., "Electromagnetic Waves", McGraw-Hill, N.Y., 1941.

2. Miller, S.E., "Coupled Wave Theory and Waveguide Application", BSTJ, Vol.33, p.661, 1954.

3. Karbowiak, A.E., "Statistical Aspects of Signal Distortion in Trunk Waveguide Communication", Proc. IEE, Vol.119, No.11, pp.1538-1548, Nov.1972.

4. Karbowiak, A.E., "Trunk Waveguide Communication", Chapman & Hall, 1965.

5. Personick, S.C., "Time Dispersion in Dielectric Waveguide", Bell System Tech. Journal, Vol.50, pp.843-859, March 1971.

6. Love, J.D. and Snyder, A.W., "Ray Analysis of Multimode, Optical Fibres", Annales des Telecom., Vol.32, Nos.3-4, March-April 1977, pp.109-113.

* See D. Marcuse, "High Order Loss Processes and the Loss Penalty of Multimode Operation", BSTJ, Vol.51, p.1819, 1972.

BIOGRAPHIES

A.E. KARBOWIAK graduated from London University with B.Sc. (Eng) in 1949 and a Ph.D. degree in 1953.

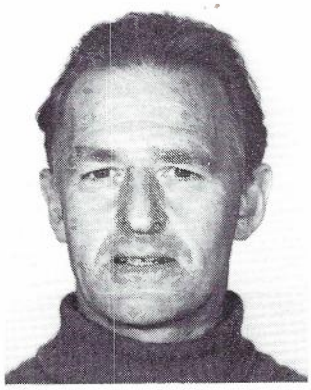
From 1953 he was with Standard Telecommunications Laboratories Limited, England, carrying out research in microwaves and waveguides. In 1958 he became Head of the Microwave Department, with the research program covering various fields, including Long Distance Communication by waveguide using millimetric waves. In 1962 he became Head of the Optical Systems Group where he directed research in Optical Communications.

In 1964 he was appointed to the Chair of Electrical Engineering (Communications) at the University of New South Wales where he is at present Head of the Department of Communications. For his contributions in the field of communications, Professor Karbowiak was awarded the D.Sc. (Eng) degree from London University in 1968.

In 1971 he was consultant with the British Post Office Research Laboratories, Martlesham, East Anglia, U.K. concerning future communications systems including Trunk Waveguide Systems and High-speed Digital Communication systems.

Professor Karbowiak has published some 80 technical articles and papers dealing with various aspects of research and two books: "Trunk Waveguide Communication" (Chapman & Hall, 1965) and "Theory of Communications" (Oliver & Boyd, 1965). He is a co-editor of a text book: "Information, Computers, Machines and Man" (John Wiley & Sons, 1971). He is the author of a number of patents and has been honoured by various industrial and professional awards in U.K.

He was elected Fellow of the Royal Society of Arts (U.K.) in 1976, and is a Foundation Fellow of the Australian Academy of Technological Sciences.



DOUGLAS IRVING obtained his first class Hons, B.E. degree at the University of New South Wales, where he has since been employed as a Professional Officer in the Department of Communications. He has been concerned with the propagation characteristics of coaxial cable, circular waveguide and optical fibre systems with particular reference to digital communication. He is currently developing high speed optical fibre systems and studying pulse distortion in optical fibres.

Telephone Mouldings for the Australian Environment *

B. A. CHISHOLM
Telecom Australia Research Laboratories

G. FLATAU
Telecom Australia Research Laboratories

H. J. RUDDELL
Telecom Australia Research Laboratories

In 1974, Telecom Australia commenced a test programme to evaluate alternative materials to ABS for future generation telephones. Seventeen different plastics derived from six basic polymer types were chosen for study. Moulded test specimens were exposed in a glasshouse at Melbourne for up to two years. Samples were withdrawn at various intervals of sunshine hours and subjected to a variety (twelve in all) of mechanical, thermal, colour and burning behaviour tests. It was found that no one plastic possessed all the desired properties. It was concluded that one specific grade of ASA was the most likely alternative to ABS.

1. INTRODUCTION

Up to the early 1960s, thermosetting plastics, phenolics and melamines were used for the case and handset of Australian produced telephone instruments. With these compression moulded plastics the range of possible colours was strictly limited and it was not until 1964, when a decision was taken to introduce ABS utilising the injection moulding process, that a wide choice of colours became possible. The initial range comprised only five colours (grey, green, ivory, red and mustard), but in the succeeding years this has been extended to nine colours.

Whilst it is known that oxidation of the chemically unsaturated elastomeric component of ABS, Butadiene, can cause embrittlement and surface dulling, this has not proved to be a real problem in our operational experience. The major disadvantage found with ABS has been the ease with which the surface may be blemished by scratching or staining and the difficulty of removing such blemishes. Marks from ball-point pens have proved particularly difficult in reconditioning operations owing to the deep score line made by the ball-point and the deposit of ink at the base of the indentation. But overall, the performance of ABS has justified its selection for the production of Australian telephones.

With the exception of Japan, where PVC has been used for many years, the majority of telecommunication organizations still employ ABS even though the range of available thermoplastics has grown enormously over the last 10-15 years. In 1970, Telecom Australia decided that a comprehensive investigation of polymeric materials suitable for the injection moulding of telephone parts should be undertaken in order to assess their relative performance, and by comparison, determine whether any of the new materials would offer advantages over ABS. A great deal of pertinent information was already available from studies conducted by other telecommunication organizations, and this assisted greatly in reducing the number of potential polymers to be investigated. However, because of the varied and severe climatic conditions under which some telephones may have to operate in Australia, a final selection could not be made until those polymers considered to have the most promise had been evaluated under local environmental conditions.

This paper is based on experimental studies carried out by R.J. Boast and D.J. Adams during 1974 to 1977 in the Telecom Australia Research Laboratories.

2. SPECIAL CLIMATIC CONSIDERATIONS

Australia lies between latitudes 10-44°S. Slightly more than 40% of the continent lies within the tropics. The remaining 60% is in what is usually described as the temperate zone, which ranges from sub-tropical southern Queensland to cooler areas such as Victoria, and extends into the mountainous, often snowbound, Tasmanian highlands.

---

* This paper was presented at the 2nd International Conference "Plastics in Telecommunication", London, September 1978. It is published with the permission of the Plastics and Rubber Institute, U.K.

In the summer month of January average daily temperatures vary from about 29°C in the far north of the continent to around 18°C in the far south. Average mid-winter (July) temperatures range from 24°C in the north to 10°C in the south During exceptional summers in the northern inland shade temperatures sometimes rise above 43°C and afternoons hotter than 38°C are frequently experienced. The most persistently hot areas are inland parts of Western Australia. For instance, at Marble Bar 160 consecutive days of 38°C or over have been recorded. Sydney's mean annual temperature of 17,4°C compares to London's 9.8°C, New York's 11°C, Rome's 15.4°C, Berlin's 9°C, Tokyo's 13.8°C, and Bombay's 27°C.

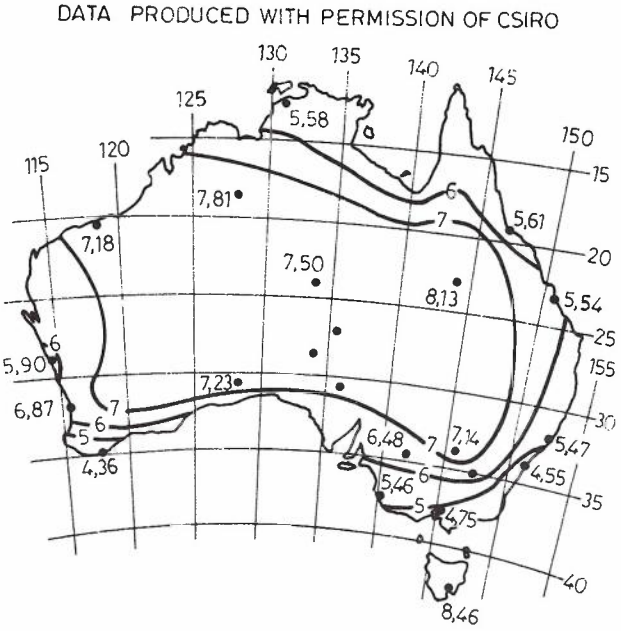


Fig.1 - Ultraviolet radiation map of Australia.

Australia's fame as a land of sunshine is no legend. Throughout the year the average daily hours of sunshine of the capital cities is 6.7 h. This average is considerably greater in some inland towns. The map in Fig.1 (Ref.1), based on a concept developed by Martin (Ref.2) of the Commonwealth Scientific and Industrial Research Organization (CSIRO) shows continental ultraviolet (UV) radiation levels in the form of a solar weathering index, providing an indication of the relative degrading influence of sunlight on exposed materials at any given site within Australia. The mean intensity of UV radiation in Melbourne is approximately twice that of London (Ref.2). Brisbane and Perth are 50% higher again.

Such climatic data explains why the choice of a polymer for the Australian environment cannot be solely based on considerations or data from other sources.

3. EXPERIMENTAL CONSIDERATIONS

3.1 Choice of Polymers

Seventeen polymers were selected as a result of information gathered from overseas sources and

TABLE 1 - Polymers chosen for Test Programme

Polymer Type	Abbreviation and Designation of Various Grades Used	Comments
Acrylonitrile-Butadiene-Styrene	ABS (T) ABS (DM)	Control Materials
	ABS (III)	Flame retardant added
	ABS (IV)	ABS/PVC Blend
Acrylonitrile-Chlorinated Polyethylene-Styrene	ACS (I) ACS (II)	Flame Retardant added
Acrylonitrile-Styrene-Acrylic Elastomer	ASA (I) ASA (II)	
Poly(vinylchloride)	PVC (I) PVC (II) PVC (III)	Moulding problems. Rejected.
Polypropylene	PP (I) PP (II)	
Polycarbonate	PC (I) PC (II)	
Poly(butylene terephthalate)	PBTP (I) PBTP (II)	
Poly(ethylene terephthalate)	PETP	Moulding problems. Rejected.
Polyoxymethylene	ACETAL	

local polymer suppliers, as well as from theoretical considerations. These seventeen materials derived from 6 basic types (see Table 1) were considered to be the most likely alternatives to ABS for telephone mouldings. Also included, for comparative purpose, were the two ABS resins, "Cycolac" T and "Cycolac" DM, which have been sequentially the standard resins for Telecom Australia telephones since 1964.

The total number of polymers included in our test programme was therefore nineteen. The colour was restricted to "appliance white", to optimize discolouration comparisons.

3.2 Selection of Test Methods

The shape of a moulded telephone case makes it difficult to ensure uniform ageing over the entire surface. It is not possible to cut standard size test pieces from the moulding. An evaluation based purely on artificial ageing was rejected because of the lack of known correlation factors between artificial and natural ageing. The project was therefore divided into two parts. The first test series, on the full number of likely polymers, was conducted with conventional test pieces. Standard test methods were used to screen these materials and establish an order of merit. The most promising mat-

TABLE 2 - Details of Test Programme

Name of Test	Method	Exposure Period, Sunshine Hours					
		Unexposed	125	250	1000	2300 (1 Year)	4600 (2 Years)
Tensile Strength at yield at 0°C and 23°C	ASTM D1708 - 70 Testing Speed 1 mm/min	x	x	x	x	x	x
Tensile Strength at break at 0°C and 23°C		x	x	x	x	x	x
Elongation at break at 0°C and 23°C		x	x	x	x	x	x
Flexural Yield Strength at 0°C and 23°C	ASTM D790 - 71 Method 1, Procedure B	x	x	x	x	x	x
Flexural Modulus at 0°C and 23°C	Testing Speed, 10 mm/min. Weathered surfaces in tension	x	x	x	x	x	x
Izod Impact Strength 0°C and 23°C	AS 1146 - 1972 Part 1 Specimen thickness 3.40 mm	x	x	x	x	x	x
Falling Dart Impact strength	ASTM D3029 - 72, Procedure B. Weathered surface in tension	x	x	x			x
Heat Deflection Temperature	AS 1369 - 1973	x					x
Vicat Softening Point	AS 1368 - 1973	x					x
Rockwell Hardness	ASTM D785 - 65 Procedure A. R & L scales	x					x
Barcol Hardness	ASTM D2583 - 67 Barcol Impressors 935 and 936	x					x
Pencil Hardness	(Ref.5)	x					
Mar Resistance	ASTM D673 - 70	x					
Stain Resistance	ASTM D2299 - 68	x					
Specular Gloss	ASTM D523 - 67	x	x	x	x	x	x
Colour	ASTM D1925 - 70	x	x	x	x	x	x
Minimum Oxygen Concentration	AS 2122.2 - 1978 Part 2	x					
Smoke Generation	ASTM 2843 - 70	x					

x = Test performed

erials from this initial test series were then subjected to further tests, in the form of moulded parts, to arrive at a final decision.

The test programme was designed on the assumption that all the required information should be obtainable from the results of 12 specific tests. The parameters of interest were; Tensile Strength, Elongation, Flexural Strength, Heat Distortion, Softening Point, Mar Resistance, Colour Retention, Gloss Retention, Minimum Oxygen Concentration, Smoke Generation, Impact Resistance and Hardness. Some of these parameters were obtained by more than one test method, e.g. impact strength by both Izod and Falling Dart, at two temperatures, 0°C and 23°C. For a listing of the full range of tests, see Table 2.

3.3 Preparation of Test Specimens for Initial Test Series

A survey of the specimen sizes required for each of the tests listed in the section above showed that the entire test programme could be accommodated by three different sized test pieces, and that these pieces could be injection moulded in a single family mould. Two hundred lifts from this mould, for each polymer, were required to cover replicate testing at the various ageing intervals shown in Table 2, resulting in 10 200 moulded pieces from which 14 500 test specimens were obtained.

The injection moulding machine used to produce the mouldings was a five ounce (140 g) Johns Hydraulic, with an in-line screw and air circulating hopper dryer. A portable air circulating dryer was used to dry the polymer pellets before transfer to the hopper. Circular cross-sectioned runners and maximum size balanced gates were incorporated in the mould to minimize stresses in the finished moulding.

Each material was assessed on its moulding performance in the three cavity mould, and this experience was utilised to predict the likelihood of future success in moulding telephone cases. Two materials were rejected on this basis: a PVC sample which failed to fill the mould cavity and could not be moulded without degradation, and a polyester which exhibited both amorphous and crystalline regions after moulding, see Table 1.

3.4 Exposure Conditions

Evaluation of the physical properties and appearance of a polymer, in the as-moulded condition, provides insufficient data on which to base life-time predictions, since there is no assurance of the retention of these parameters over the operating life of ten or more years.

Life expectancy predictions under normal operating conditions require a very long test exposure determined by the estimated service life of the article under test. Hence, to shorten the test period, it is customary to increase the test stresses (e.g. temperature, UV irradiation, environmental cycles). Correlation of the results of accelerated ageing with those from natural exposure is difficult, but ageing tests can be used to fairly rapidly separate materials into "probable" and "improbable" categories. It was considered that the latter step was unnecessary in this study as it has already been covered by several overseas organizations. Also the great number of specimens required for a full correlation experiment would have exceeded the capacity of the artificial weathering chambers available to us.

It was therefore decided that the exposure tests should be conducted in a manner similar to that used by the British Post Office (BPO), that is, exposure to natural light under glass. Although it could be argued that this condition is more severe than the average condition under which a telephone operates, it does simulate the worst condition for a telephone in a northern window exposure inside an Australian home or office.

The test specimens were exposed on racks inclined at 45°, facing north, inside a glasshouse located on the roof of a four storey building in the centre of the city of Melbourne (37° 51' S). In the dusty, polluted atmosphere of a modern city, it was found necessary to institute a weekly cleaning programme of washing the roof and walls of the glasshouse with water, and a light dusting of the specimen samples.

The exposure period for the specimens was designed to span at least two southern summers, and it was intended to monitor the radiation incident on the test specimens by measuring the time required to fade various blue wool standards of BS1006-1961. These are pieces of dyed wool cloth, varying in their colour fastness to light, graded from 1 to 7, with No.1 being the least stable to light. A blue wool standard is said to have faded when the difference in colour between exposed and unexposed portion is the same as contrast No.4 of the grey scale of BS2662-1961.

However, blue wool standard No.7 faded in a matter of three weeks in trials conducted during autumn (April 1974) in Melbourne, which meant that many of these standards would have to be exposed sequentially to span the expected exposure period of two years. In addition, it was found almost impossible to reproducibly detect and quantify the amount of fading required for replacement of the wool standard, and these difficulties led to the abandonment of the method. The fact that assessment by blue wool standards was found to be unsatisfactory in Australia, whilst apparently suitable in the BPO trials [Harrison and Portwood (Ref.3)], confirms previous observations that the radiation intensity in Australia is much greater than that experienced in the United Kingdom or Europe.

It was therefore decided to use cloud-free sunshine hours, as recorded by the Bureau of Meteorology at Laverton (16 km from Melbourne), as the exposure index. This gives an approximate measure of the amount of radiation received, and provides an effective method of monitoring the prevailing exposure conditions. Figure 2 shows the sunshine parameters encountered during the test period.

Air was circulated within the glasshouse by an exhaust fan mounted on the southern wall. Ambient internal air temperature during the exposure period ranged from 3°C in winter to 48°C in summer.

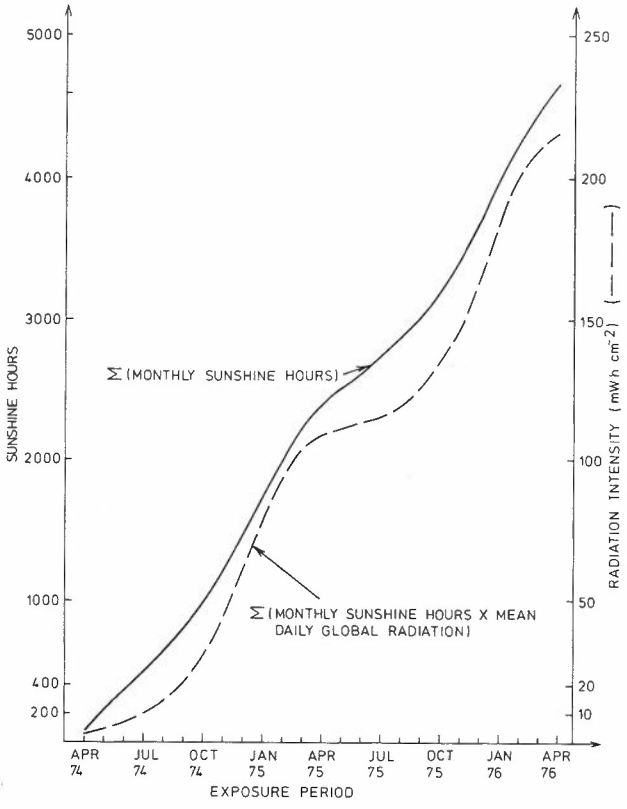


Fig.2 - Exposure parameters.

Several test specimens, especially PVC, showed severe discolouration during exposure. This was considered to be due to thermal degradation (Ref.2), hence, surface temperatures of samples inside the glasshouse were monitored. On days when the shade temperature was only 22°C, surface temperatures of 60°C were noted, rising as high as 66°C on other occasions.

In tests carried out on similar test specimens placed inside a non-airconditioned office, behind a glass window facing north, surface temperatures as high as 52°C were recorded. The temperatures encountered in the glasshouse were therefore considered not to be unrealistic, when the following factors are taken into consideration:

(i) higher temperatures than those which prevailed in Melbourne during these tests do occur in Melbourne and in other parts of Australia,

(ii) it is known that temperatures in public telephone booths can be very high for long periods,

(iii) higher surface temperatures are likely on coloured samples because of their higher absorbance.

Some years ago temperatures as high as 80°C were measured on the surface of black polyethylene insulators, pole mounted at a site 150 km north of Melbourne!

4. TEST RESULTS

In selecting a material for a part as widely used as a subscriber's telephone, many factors must be considered such as economics, mechanical and thermal properties, colour stability, resistance to surface damage, combustibility and moulding behaviour. High resistance to impact damage and maintenance of this property in service is essential. Important, but to a lesser extent, are properties which affect the aesthetic appearance of the telephone. However, here compromises may be necessary unless they adversely affect overall performance.

The test programme generated a considerable quantity of data encompassing all of these factors, except for the economic considerations. A full analysis of the data with detailed explanation of the performance of each material is beyond the scope of this paper. In some cases further experimental work may be needed. Hence, only those tests where the results showed significant trends will be dealt with in some detail. They will be covered under the following headings:

(i) Mechanical Properties,

(ii) Aesthetic Appearance,

(iii) Thermal Characteristics.

Tests performed at 0°C and 23°C showed similar behaviour patterns and therefore only results obtained at 23°C are discussed. In retrospect, testing at 0°C was of little significance and a higher test temperature, e.g. 35°C, may have been more revealing having regard to the temperatures likely to be encountered in service in Australia.

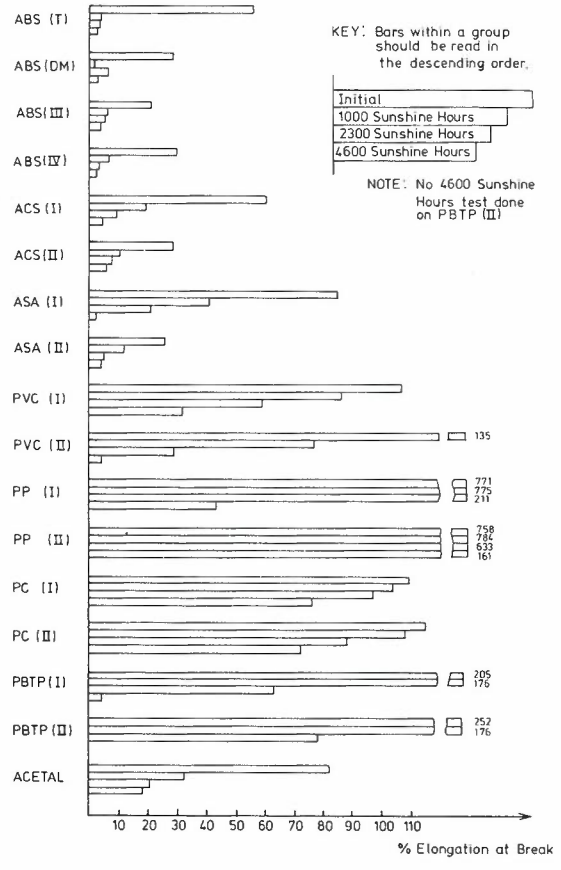


Fig.3 - Elongation at break, at 23°C.

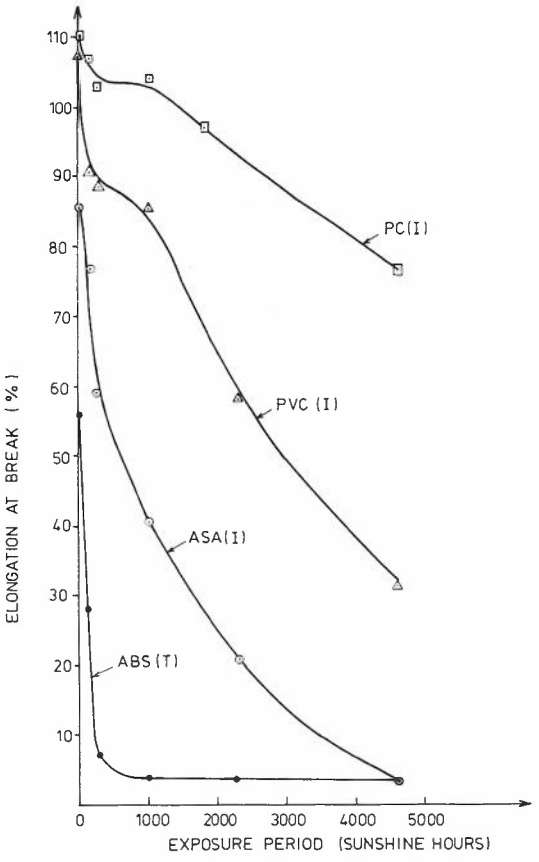


Fig.4 - Elongation at break vs exposure period.

4.1 Mechanical Properties

4.1.1 Elongation at Break This property was used to monitor chain scission and subsequent embrittlement of the plastics following exposure to infra red (IR) and UV radiation and thermal stressing. Overall results are shown in Fig.3. Furthermore, Fig.4 shows the changes in elongation versus exposure period for ABS (T), ASA (I), PVC (I), and PC (I).

The decrease in elongation for ABS (T) was extremely rapid and after 250 sunshine hours exposure this material displayed tensile properties characteristic of hard, brittle material with coincident yield and break points, whereas for ASA (I) and (II), ACS (I) and (II), and PVC (II), this brittle behaviour was not evident until after 4600 sunshine hours.

Figure 5 indicates the tensile yield stress characteristics. As for elongation, ABS (T) showed a rapid decrease compared with other materials such as ASA (I), ACS (I) and (II), PVC (I), and PC (I) and (II).

4.1.2 Impact Strength Two methods were used to measure impact strength; notched Izod and Falling Dart.

Comparison of results taken over the exposure period shows different behaviour patterns for each method, (see Figs.6 and 7).

Few of the materials had an Izod impact strength, prior to exposure, anywhere near the values quoted by raw materials suppliers' data sheets. The ABS materials performed satisfactorily in the Izod tests with only small decreases over the two year exposure period.

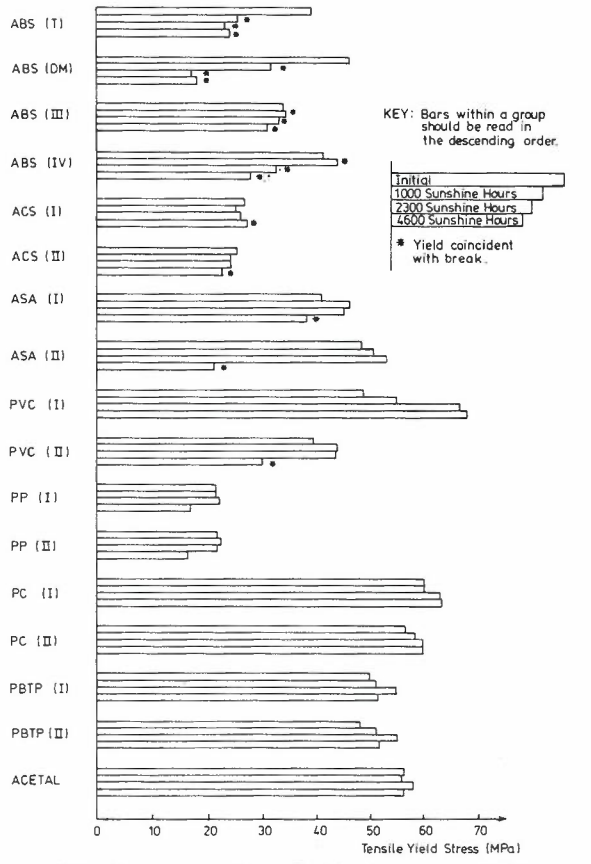


Fig.5 - Tensile yield stress at 23°C.

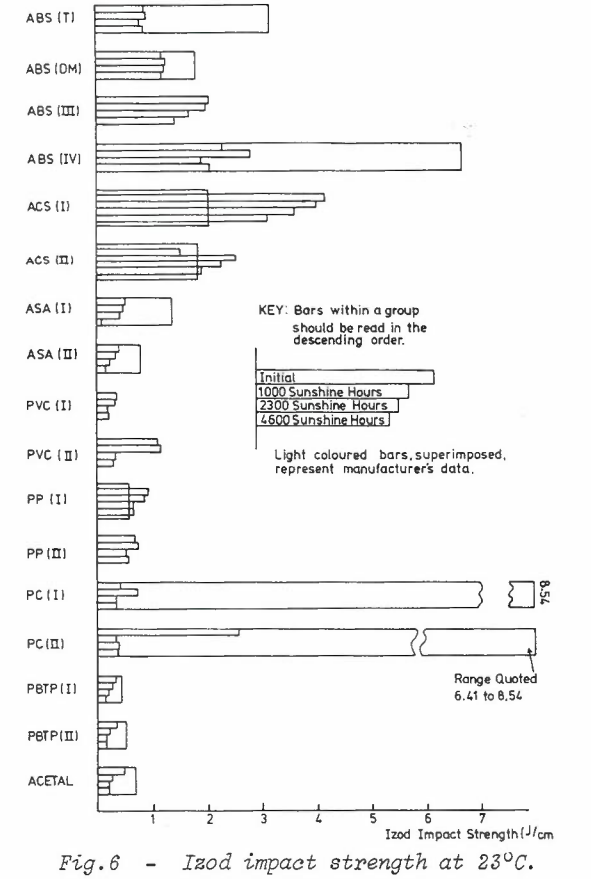


Fig.6 - Izod impact strength at 23°C.

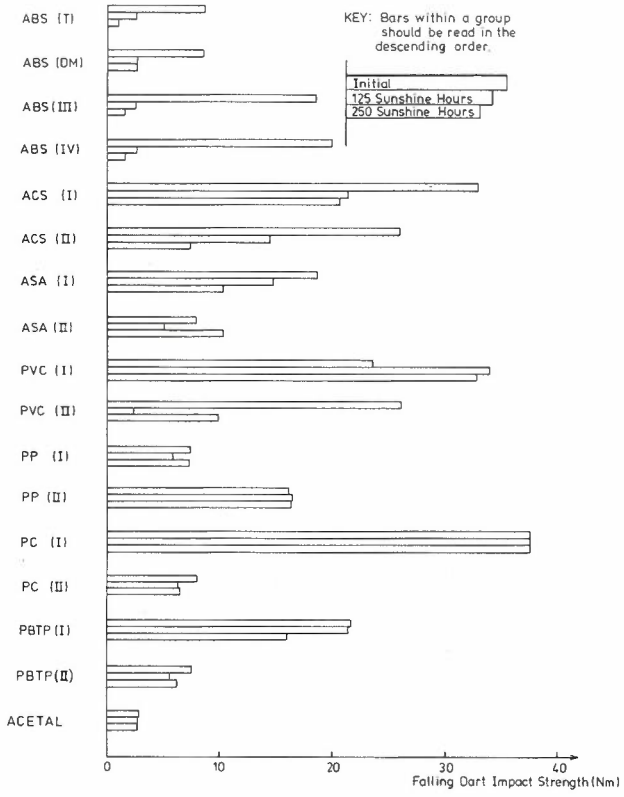


Fig.7 - Falling Dart impact strength at 23°C.

With the exception of PC (II), the initial values of impact strength by the Falling Dart method were of the anticipated magnitudes. For some materials such as ABS (T), impact strength fell dramatically after only 125 sunshine hours, whereas the deterioration was much less for ACS (I), PC (I) and ASA (I). This is clearly shown by Fig.8.

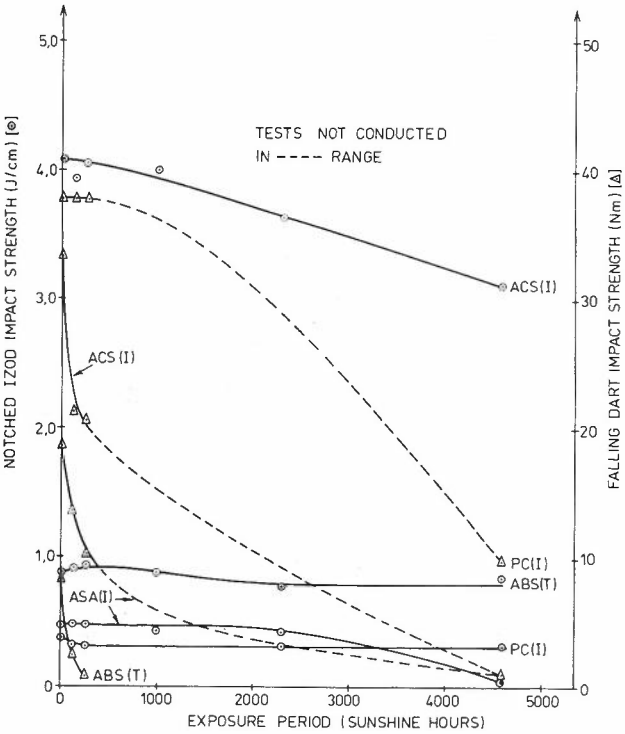


Fig.8 - Izod and Falling Dart impact strengths at 23°C.

Based on these results it appears that Falling Dart Impact is a more effective method than Izod as a screening test for resistance to embrittlement. Of course, the ultimate requirement for impact strength of a telephone in service is its resistance to impact damage after falling from a desk or other item of furniture. To simulate this, a modified Falling Dart impact test has been developed for application to telephone mouldings. This is achieved by a mass equal to a complete handset being dropped from 0.75 m on to a complete telephone housing. The number of such impacts required to crack the case are specified. The telephone is rotated so that any point on the surface of the case can be tested [Adams and Jones (Ref.4)]. This test will be used in future studies to monitor the degradation of telephone mouldings.

Based on a limited survey in Melbourne, we estimate that approximately 3% of telephones returned from subscribers' premises, for various reasons, have cracked cases.

4.2 Aesthetic Appearance

An objective measurement of aesthetic appearance is almost impossible. A number of tests were tried and for convenience have been grouped together as follows:

(i) Resistance to Surface Damage,

(ii) Colour Change,

(iii) Resistance to Staining.

4.2.1 Resistance to Surface Damage This includes such properties as mar resistance, indentation hardness, scratch resistance and some aspects of gloss retention.

The results of mar resistance are not included because of the inability to control such variables as:

(a) the degree of operator dependence,

(b) variations in mass of silicon carbide falling on to the test specimens; grit tended to adhere to the walls of the hopper,

(c) some particles of grit became embedded in the test specimen; these could not be removed without causing surface damage.

Nevertheless, broad indications were that PC, ASA, PBTP and PVC had superior mar resistance to ABS, whilst ACS and PP were inferior.

Measurement of indentation hardness using the Rockwell R scale presented few problems. ACS and PP are softer than ABS and therefore more prone to indentation by such items as ball-point pens or diamond rings. This would be a reason for the non-selection of ACS and PP for telephone mouldings.

Telephones are prone to disfiguration by scratch marks from pencils and ball-point pens. Thus, the pencil hardness test [Ruddell (Ref.5)] appears to be a practical method of assessing the scratch resistance of various plastics. Results are presented in Fig.9. An inadequate aspect of the method is that the steps in hardness between the grades of pencils are really not as evenly

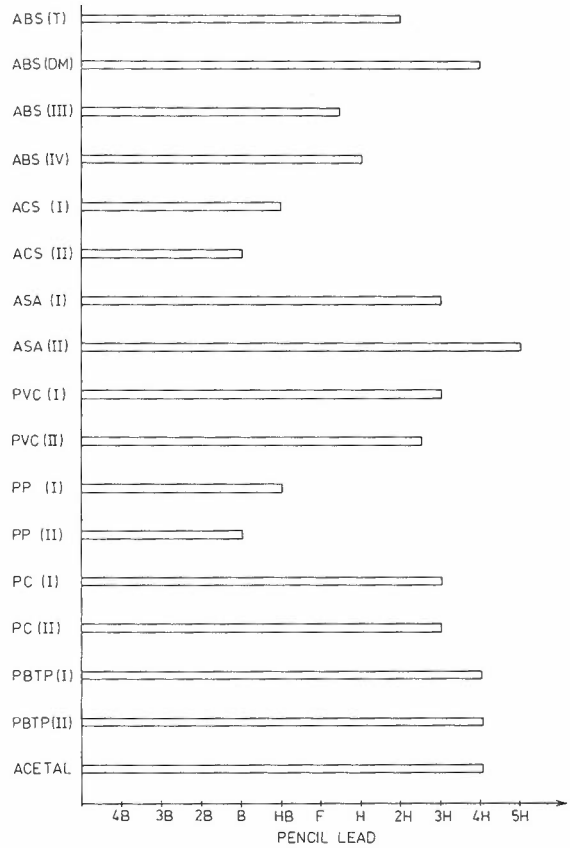


Fig.9 - Pencil hardness measurements for unexposed mouldings.

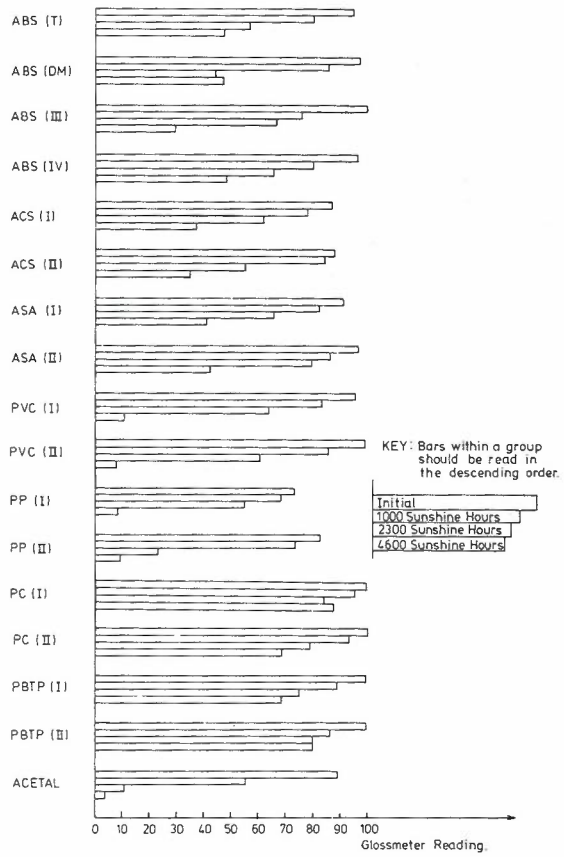


Fig.10 - Specular gloss measurements.

spaced as shown on the horizontal scale. Gouza (Ref.6) suggests that the steps are uneven, and that the hardness of pencils varies according to the formulation of graphite and clay used by different pencil manufacturers. Despite these unknown factors, the test made it possible to discriminate between materials such as ACS and PP, which can be scratched with soft lead pencils, and the harder more scratch resistant materials such as PVC and PC. The results showed trends similar to those from Rockwell hardness tests and confirmed the objections to ACS and PP.

The changes in specular gloss with exposure are shown in Fig.10. Interpretation of this data tends to be subjective. Any material with an initial value of less than 80 may not be pleasing to some people, whilst higher values may be unacceptable to others. At the other end of the scale, after exposure the dull, discoloured surface of PVC or powdery surface of polypropylene are definitely unsatisfactory. Undoubtedly, PC and PBTP have the best gloss and retain this after exposure under glass.

4.2.2 Colour Change CIE colour co-ordinates were measured using a Hunterlab D25D3 colourimeter, and the Yellowness Index (ASTM D1925) was calculated. Values for Yellowness Index agreed closely with visual observations for all materials up until 1000 sunshine hours. This type of measurement became unsuitable after longer exposures as gross discolouration of PVC occurred. Colour difference, ΔE, was calculated using Hunter's L, a, b colour solid co-ordinates where $\Delta E = (\Delta L^2 + \Delta a^2 + \Delta b^2)^{\frac{1}{2}}$. These values are shown in Fig.11 and are in good agreement with visual observations.

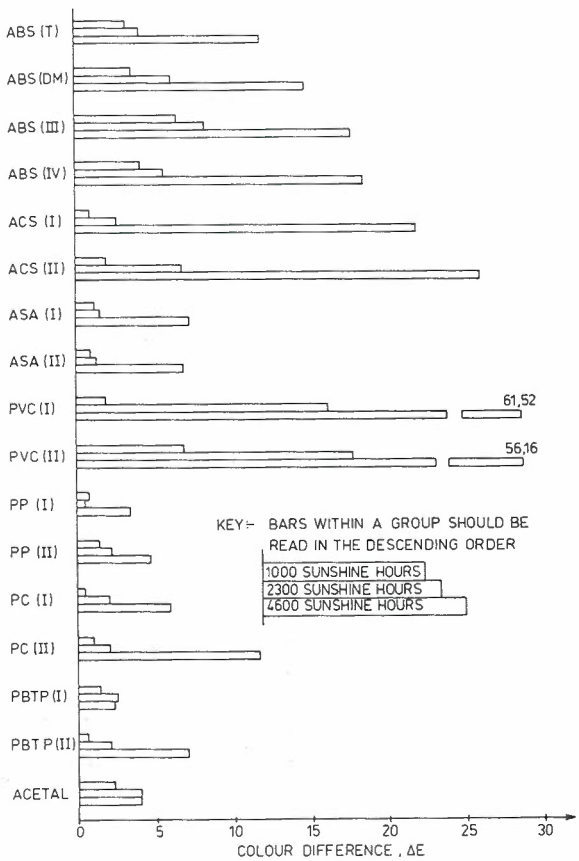


Fig.11 - Colour difference.

The most significant colour change was shown by the two grades of PVC, which changed dramatically from white to yellow to a dark brown after 1800 sunshine hours. After approximately 3500 sunshine hours exposure, all PVC specimens had turned black. These findings confirmed the necessity of repeating tests in the Australian environment irrespective of data already available elsewhere. For instance, similar exposure under glass of these materials in England did not show this degree of discolouration [Ruddell (Ref.7)].

Apart from PVC and ACS, ABS showed greater colour changes than most other polymers. Of the four grades of ABS, ABS (T) had the best resistance to yellowing. Minimal colour changes were met with PC (I), ASA (I) and (II), and PBTP (I).

4.2.3 Resistance to Staining The resistance to staining by common domestic products is shown in Table 3. No plastics were totally immune to staining. The materials with low indentation hardness and low pencil scratch resistance such as ACS and PP were most readily stained. The harder materials such as PVC, PC and PBTP showed the best resistance to staining.

4.3 Thermal Characteristics

4.3.1 Heat Distortion Temperature As mentioned previously, surface temperatures of 52°C have been recorded on white test specimens behind glass inside a non-airconditioned office building. The heat deflection temperatures of the test materials are shown in Fig.12. The low values for PBTP are unsatisfactory for a telephone moulding material.

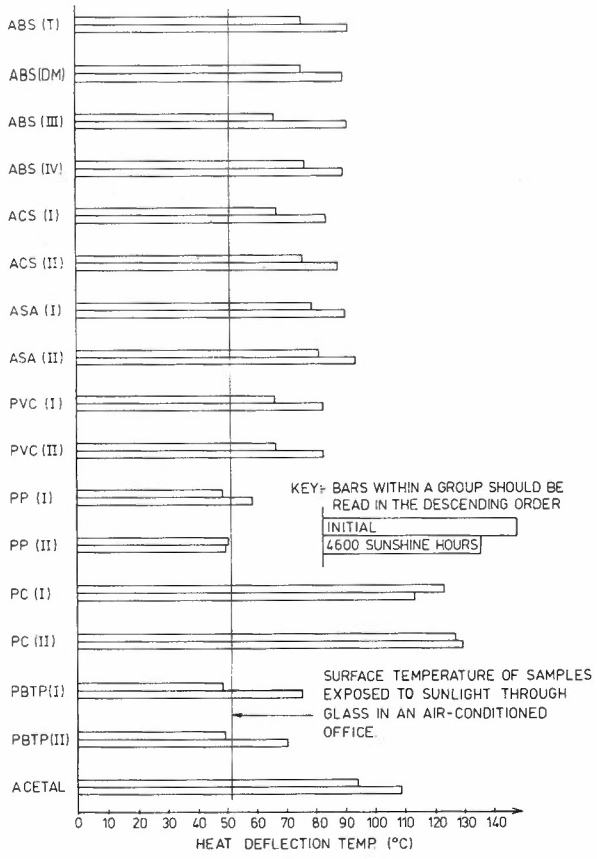


Fig.12 - Heat deflection temperature at 1,8 MPa fibre stress.

4.3.2 Burning Behaviour Because of the complex nature of burning behaviour of plastics, difficulties arise in predicting the exact way in which plastics will behave in a real fire situation. Minimum Oxygen Concentration for flame propagation (MOC) (Ref.8), previously known as Oxygen Index, was chosen as the method most pertinent for a comparative evaluation. The MOC varied over a wide range (see Table 4) with PVC and the fire retarded grades of ABS and ACS significantly superior to ABS, whilst general purpose ACS, ASA, PP and acetal were inferior. Although it is desirable that material for telephone mouldings should have a higher MOC than ABS (T), it is worth noting that telephones made from the latter material have not been a fire hazard or cause of fire during 14 years of service experience.

TABLE 3 - Resistance to Staining

Test Mat. / Polymer	Tea	Coffee	Flyspray	Lipstick	Grease	Ball Point Pen
ABS (T)	F	F	P	F	P	F
ABS (DM)	F	F	P	F	P	F
ABS (III)	F	F	P	F	F	F
ABS (IV)	F	F	P	P	F	F
ACS (I)	F	F	F	F	F	F
ACS (II)	F	F	F	F	F	F
ASA (I)	F	F	P	F	P	F
ASA (II)	F	F	P	F	P	F
PVC (I)	P	F	P	P	F	F
PVC (II)	F	F	P	P	P	F
PP (I)	F	P	F	P	F	P
PP (II)	F	P	F	P	F	F
PC (I)	P	F	P	P	P	F
PC (II)	P	P	P	P	P	F
PBTP (I)	F	F	P	P	P	F
PBTP (II)	F	F	P	P	P	P
ACETAL	F	F	P	P	P	P

F = FAIL

P = PASS (unaffected, no colour change and no appreciable change in surface texture)

TABLE 4 - Minimum Oxygen Concentration for Flame Propagation (AS 2122.2 - 1978 Part 2)

Material	Result	Material	Result
	%		%
ABS (T)	19.2	PP (I)	17.7
ABS (DM)	19.7	PP (II)	17.5
ABS (III)	27.8		
ABS (IV)	21.3	PC (I)	26.5
		PC(II)	26.8
ACS (I)	27.6		
ACS (II)	19.9	PBTP (I)	23.6
		PBTP (II)	23.4
ASA (I)	18.7		
ASA (II)	18.4	ACETAL	15.8
PVC (I)	45.8		
PVC (II)	34.6		

4.4 Summary

A survey of all the results obtained, including some not listed in this paper, has led to the conclusion that the following materials were at least equal to ABS (T) in performance: ASA, PC, PBTP and PVC. Of these, ASA (I) has been chosen as the material most likely to replace ABS (T). Telephone cases moulded from ASA (I) are being exposed to sunlight under glass for a two year period together with ABS (T), for comparison. The main parameters being monitored are impact strength and aesthetic appearance.

One advantage of ASA (I) over the other materials is that in any future commercial production of telephone cases the existing tooling, designed for ABS, will be adequate.

Polycarbonate was rejected because of raw material cost and because previous experience in other applications had shown it susceptible to crazing and stress cracking.

The heat distortion temperature of PBTP is so low that it could be encountered in service in Australia, and therefore use of this material would be risky.

The thermal degradation of PVC and subsequent discolouration raises serious doubts regarding its use for telephone handsets in Australia. Nevertheless, PVC is in use in Japan and being considered seriously by two other administrations. A number of cases moulded from PVC have been included in the current evaluations for comparison purposes.

5. CONCLUSION

Examination of moulding properties and extensive physical testing over a two year period has provided comparative data which makes it possible to assess ABS, as currently used for moulding telephones, against seventeen other polymers.

The data has shown that no single polymer is superior in all aspects, confirming findings by other administrations [Portwood and Cottrill (Ref. 9)].

However, the decision to test materials under Australian climatic conditions, and not to rely solely on information published elsewhere, has been proven correct.

ASA is the material considered at this stage most likely to replace ABS, but further investigation is needed before a final decision can be made.

6. ACKNOWLEDGEMENTS

The authors wish to thank the staff of the London Materials Section, BPO for their assistance, frank discussions and co-operation over many years.

The permission of the Director Research, Telecom Australia, to publish this paper is acknowledged.

7. REFERENCES

1. CSIRO, "Rebuild" 2, No.5, (Div. Build. Res., 1977).

2. Martin, K.G., British Polymer J., 5, (1973) p.444.

3. Harrison, J.C. and Portwood, R., Plastics and Polymers, (Dec.1970), p.422.

4. Adams, D.J. and Jones, K.H., "The Development of an Impact Test for Telephone Cases", Telecom Australia Research Labs. Report No. 7228, 1978.

5. Ruddell, H.J., Personal communication, Showa Denko, Japan, 1975.

6. Gouza, J.J., "Testing of Polymers" 2, (Edited J.V. Schmitz), (Interscience Publishers, New York, 1966).

7. Ruddell, H.J., Personal communication, BPO, England, 1976.

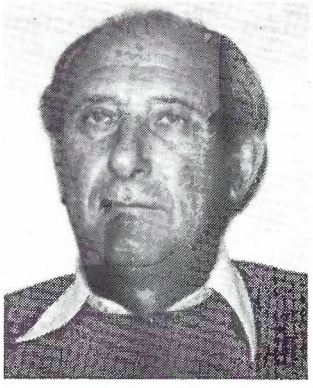
8. Australian Standard, 2122.2 (1978) Part 2.

9. Portwood, R.C. and Cottrill, N.C.W., "Plastics in Telecommunications" Conference, London. 1974.

BIOGRAPHIES

GERALD FLATAU graduated from RMIT in 1951 with a Fellowship Diploma in Applied Physics, and joined the PMG Research Laboratories in the same year. He became Head of the Physics Division in 1960, and the Physics and Polymer Section in 1965, and is currently the Staff Scientist of the Physical Sciences Branch, with responsibilities for the co-ordination of scientific activities of the Branch.

Mr. Flatau's main interests are in the fields of component reliability, materials evaluation and failure analysis, and environmental phenomena. In the last few years he has had extensive involvement in National and International standardisation activities.



HEC RUDDELL obtained the Diploma in Applied Chemistry at the Melbourne Technical College in 1949 and joined the PMG Research Laboratories as a Chemist Class 1 in 1955 after 19 years of technical experience in the oil, rubber, general chemicals and plastics industries. In 1966 he was promoted to the position of Chemist Class 3 to set up the Polymer Division (now Polymer Section) which is responsible for the investigation, development and application of all plastics, rubbers and adhesives in communication equipment. He is now the Principal Chemist in charge of this Section. He was actively associated with the early manufacture and installation of plastics cable including the development of epoxy resins and the epoxy resin field pack, the design of joints for plastics submarine cables laid off the coast of Queensland and the recent development of "filled" cable.

Mr. Ruddell has been an associate member of The Plastics Institute of Australia since 1959 and is currently involved in the preparation of National standards on telecommunications cables, mechanical properties of plastics and flammability test methods.

BRUCE CHISHOLM graduated from the then Bendigo Technical College in 1965 with a Diploma in Applied Chemistry and joined the PMG Research Laboratories in 1966 as a technical officer in the chemical laboratory. He was promoted to the position of Chemist Class 1 in the Polymer Section some 10 years ago and presently holds the position of Senior Polymer Chemist, Class 3. His work has covered a variety of polymer applications, such as sealing methods, insect resistant cables and cable materials.

In 1978 he was awarded a Telecom Australia post graduate scholarship to undertake studies for a MSc in Polymer Technology at the University of Technology, Loughborough, U.K.

Spectra of Baseband Line Codes with Violations

C. T. BEARE
Telecom Australia Engineering Headquarters

Violations of the line code can be used as a means of providing an extra signalling channel. When this is done many of the advantageous properties of the original line code may be lost. This is investigated in this paper with reference to AMI and class 4 partial response line codes. Both random and repetitive patterns of violations are considered.

1. INTRODUCTION

Line coding is used in baseband data transmission to ensure that the signal transmitted to line has desirable properties. Removal of DC and low frequency content in the signal is usually necessary for purposes of AC coupling or DC power feed. On some lines such as loaded cable telephony lines some restriction on high frequencies is also desirable due to the poor transmission quality of lines at these frequencies. Coding is also used to aid timing recovery at the receiver by ensuring sufficient transitions occur in the line signal especially if long strings of ones or zeros occur in the input data.

Numerous codes have been proposed and are in use in different applications. Duc (Ref.1) and Duc and Smith (Ref.2) classify many such codes and summarise their significant characteristics.

As most codes use some form of redundancy to derive the line signal, this redundancy can be used as an extra signalling channel by using violations of the line coding rule. The signals that can be transmitted in this way include low speed clock or framing information and signals to activate or disable test loops at the remote location.

With violation signalling, care must be taken to prevent the violations destroying the desirable features of the particular line code. In particular, if a pattern of violations is sent continuously, the spectrum of the line signal can be altered considerably. If only one or two violations are used to signal the end of data and the beginning of a coded information sequence, the problem of changing the spectrum does not exist. However, interrupting the data flow may not be possible in many situations.

Violations of line codes are used for other purposes also. The B3ZS code is derived by violating the AMI or bipolar code to remove strings of zeros and thus improve timing recovery at the received end.

In this paper, the effect of violations on the spectrum is considered for the AMI or bipolar line code and the partial response class 4 line code.

2. BIPOLAR CODE WITH RANDOM VIOLATIONS

One method of calculating the spectrum of a digital line signal is to Fourier transform the autocorrelation function of the particular signal. Discrete points in the autocorrelation function are given by

$$R(nT) = \langle x(t)\, x(t + nT) \rangle \qquad (1)$$

where $x(t)$ is the transmitted line signal

and $\langle\ \rangle$ denotes time average

For AMI (bipolar) line coding using rectangular pulses of width T three signal levels (-1, 0, +1) are used and in any one pulse period the transmitted line signal is constant (Fig.1). If i and j are used to represent levels in the transmitted line signal and we define:-

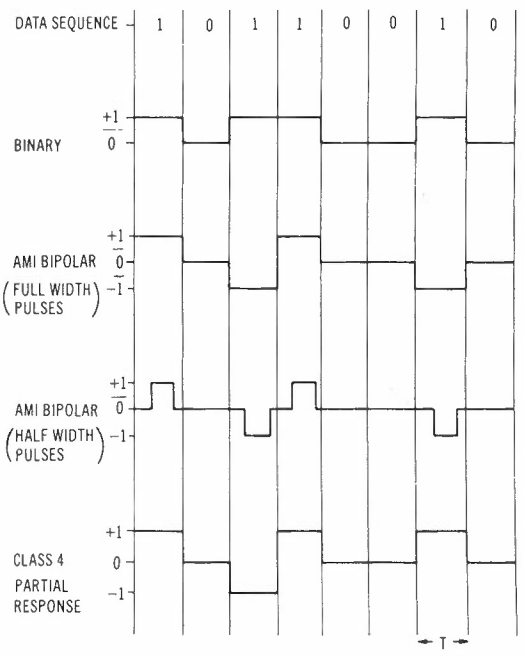


Fig.1 - Line code signal.

P(i) = probability that level i occurs

and

P(j/i,n) = probability that level j occurs given that level i occurred n bit periods previously

then providing the source is ergodic, (1) can be replaced by the ensemble average

$$R(nT) = \sum_{\text{all } i} i\,P(i) \sum_{\text{all } j} j\,P(j/i,n) \qquad (2)$$

This is valid for any multi-level line code using rectangular pulses and by inspection for the AMI line code (Fig.1) becomes

$$R(nT) = P(1)\{P(1/1,n) - P(-1/1,n)\}$$

$$+ P(-1)\{P(-1/-1,n) - P(1/-1,n)\} \qquad (3)$$

and because +1 and -1 pulses are coded in similar ways by the AMI rule

$$P(1) = P(-1)$$
$$P(1/1,n) = P(-1/-1,n)$$
$$P(-1/1,n) = P(1/-1,n)$$

giving

$$R(nT) = 2P(1)\{P(1/1,n) - P(-1/1,n)\} \qquad (4)$$

It can be shown (Appendix 1) that

$$P(1/1,n) = \tfrac{1}{4} \text{ for } n > 1$$

and

$$P(-1/1,n) = \tfrac{1}{4} \text{ for } n > 1$$

thus

$$R(nT) = 0 \text{ for } n > 1 \qquad (5)$$

Also as rectangular pulses are being considered, the autocorrelation function will be linear between the discrete points (Fig.2).

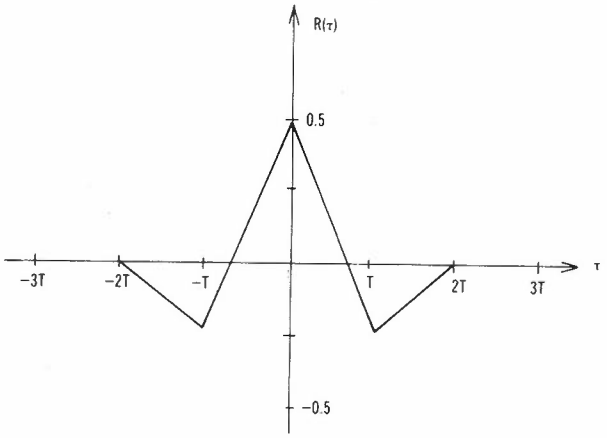


Fig.2 - Autocorrelation function of AMI.

Fourier transforming gives

$$S(\omega) = \int_{-\infty}^{\infty} R(\tau) e^{-j\omega\tau}\, d\tau$$

$$= \frac{1}{\omega^2 T}(1 - \cos \omega T)^2 \qquad (6)$$

or

$$S(\omega) = \frac{T}{2}\,\text{sinc}^2\left(\frac{\omega T}{2}\right)(1 - \cos \omega T) \qquad (7)$$

Two components of the power spectrum can be distinguished. The sinc squared component is due to the square pulse shape used while the second factor is introduced by the AMI coding. In fact, if Nyquist or other signal shaping is used the sinc squared function can be modified to the function corresponding to the particular pulse shaping employed (Ref.3).

Violations can now be introduced to this line code. In this paper a violation of the line code is said to occur when a binary one is transmitted in violation of the coding rule. For example, the binary sequence 1 0 1 1 0 1 may normally be transmitted as +1 0 -1 +1 0 -1. A violation in the second binary 1 would result in transmission of the sequence +1 0 +1 +1 0 -1. Note that a violation is not always defined in this way. Another common definition used leaves the coding rule changed after the violation. For example, the above sequence would then be transmitted as +1 0 +1 -1 0 +1. This type of violation is not considered in this paper.

Consider random violations of the AMI line code. To retain zero DC component it is necessary to keep the running digital sum zero. Thus define

p = probability of violating +1 for -1
= probability of violating -1 for +1

We need to also define subsets of P(i) and P(j/i,n) as follows:-

$\bar{P}(i)$ = probability of level i occurring according to normal coding rule

$\hat{P}(i)$ = probability of level i occurring by violation of normal coding rule

$\bar{P}(j/i,n)$ = probability of level j occurring given that level i occurred according to normal coding rule n bit periods previously

$\hat{P}(j/i,n)$ = probability of level j occurring given that level i occurred by violation n bit periods previously.

Then (4) becomes

$$R(nT) = 2\bar{P}(1)\{\bar{P}(1/1,n) - \bar{P}(-1/1,n)\}$$

$$+ 2\hat{P}(1)\{\hat{P}(1/1,n) - \hat{P}(-1/1,n)\} \qquad (8)$$

In this expression

$$R(0) = 2.\tfrac{1}{4}.(1-p) + 2.\tfrac{1}{4}.p = \tfrac{1}{2}$$

$$R(T) = 2.\tfrac{1}{4}.(1-p)\{\tfrac{1}{2}p - \tfrac{1}{2}(1-p)\}$$

$$+ 2.\tfrac{1}{4}.p.\{\tfrac{1}{2}(1-p) - \tfrac{1}{2}p\} = -\tfrac{1}{4}(1-2p)^2$$

and for $n > 1$

$$\bar{P}(1/1,n) = \hat{P}(1/-1,n) = \hat{P}(-1/1,n) \quad (9)$$

and

$$\bar{P}(-1/1,n) = \hat{P}(-1/-1,n) = \hat{P}(1/1,n) \quad (10)$$

which is evident from the definitions of $\bar{P}(j/i,n)$ and $\hat{P}(j/i,n)$ and because +1 and -1 pulses are coded in a similar way by the AMI rule.

Thus

$$R(nT) =$$

$$2\left[\bar{P}(1) - \hat{P}(1)\right]\left[\bar{P}(1/1,n) - \bar{P}(-1/1,n)\right] \text{ for } n > 1 \quad (11)$$

Again it is possible to show (Appendix 1) that

$$\bar{P}(1/1,n) = \tfrac{1}{4} \text{ for } n > 1$$

and

$$\bar{P}(-1/1,n) = \tfrac{1}{4} \text{ for } n > 1 \quad (12)$$

$$\therefore R(nT) = 0 \text{ for } n > 1$$

If we write $A = (1-2p)^2$ then the Fourier transformed autocorrelation function becomes

$$S(\omega) = \frac{T}{2}\,\mathrm{sinc}^2\left(\frac{\omega T}{2}\right)(1 - A\cos\omega T) \quad (13)$$

where $A = (1-2p)^2$

Note how again it is possible to distinguish the code and pulse shape factors in the spectrum.

The spectrum is plotted for various values of p in Fig.3. As expected, increasing the probability of violation increases the spectrum low frequency content, until at $p = 0.5$ the signal spectrum becomes that of a random binary signal. In normal situations, violations will not be random in nature and this result will not be applicable. However, any violation pattern would be expected to modify the line signal spectrum in a similar manner. In Section 4, the effect on the AMI spectrum of a particular repetitive violation pattern is demonstrated.

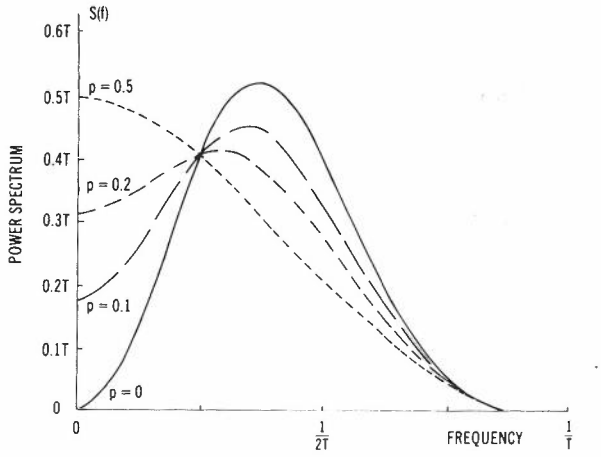


Fig.3 - Power spectrum of AMI with violations.

3. PARTIAL RESPONSE CLASS 4 WITH RANDOM VIOLATIONS

Partial response class 4 line signals can be generated by interleaving two AMI streams with each stream coded from alternate data bits.

Equation (13) is the spectrum of AMI using full width pulses. If half width pulses were used (Fig.1) then the derivation follows as before, but now in addition to R(nT) as before, an extra condition is necessary due to the half width pulses, namely

$$R\left(\frac{2n+1}{2}T\right) = 0 \quad \text{for } n > 0$$

This results in a power spectrum for half width AMI with violations

$$S_1(\omega) = \frac{T}{8}\,\mathrm{sinc}^2\left(\frac{\omega T}{4}\right)(1 - A\cos\omega T) \quad (14)$$

Class 4 partial response is formed by adding together two independent half width AMI streams where each AMI stream is coded at one half the data rate. Thus the power spectrum of class 4 with random violations is obtained from (14) by halving the rate and doubling the amplitude giving

$$S_2(\omega) = \frac{T}{2}\,\mathrm{sinc}^2\left(\frac{\omega T}{2}\right)(1 - A\cos 2\omega T) \quad (15)$$

where

$A = (1-2p)^2$ as before

Note how the pulse shaping and code shaping factors are again evident. The spectrum of the non violated code can be obtained by setting $p = 0$.

The violated code spectrum is plotted for various p in Fig.4. For this line code the violations destroy the frequency null at one half the symbol frequency as well as increasing the low frequency content of the line signal.

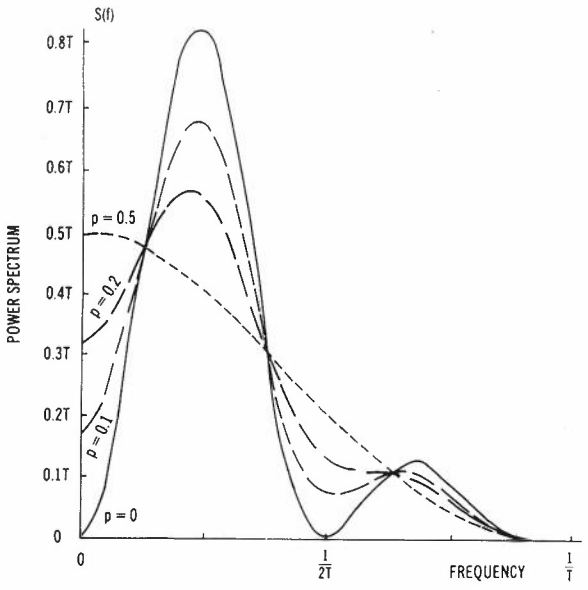


Fig.4 - Power spectrum of Class 4 partial response with violations.

4. BIPOLAR WITH CODE WORD VIOLATION

In Section 3 the effect of random violations on the spectrum of a bipolar coded signal was demonstrated. However, in most practical violation schemes, the violations are applied in a repetitive manner and not randomly. For example, the Bell Digital Data Service (DDS) uses code words with violations (Ref.4). As an illustration of the effect of repetitive violations, the spectrum of the idle code transmitted in the Bell DDS system is calculated.

Each code word is of the form BBXOV

where B is any ± 1 level pulse transmitted according to the normal bipolar rule

X is a system determined pulse (either 0 or B)

0 is a binary 0 pulse

V is any ± 1 level pulse transmitted in violation of the normal bipolar rule

In particular, the idle code (normally all binary ones) is transmitted as

1 -1 1 0 0 1 -1 1 -1 0 0 -1 repetitively
code word 1 code word 2

As this signal is repetitive, the spectrum of the line signal will be a line spectrum and can be calculated from a Fourier Series (Appendix 2). The resulting spectrum is shown in Fig.5 with

$$2C_n^2 \begin{cases} = \frac{8}{n^2\pi^2}\left(2\sin\frac{n\pi}{3} - \sin\frac{n\pi}{6} - \sin\frac{n\pi}{2}\right)^2 & \text{for n odd} \\ = 0 & \text{for n even} \end{cases} \qquad (16)$$

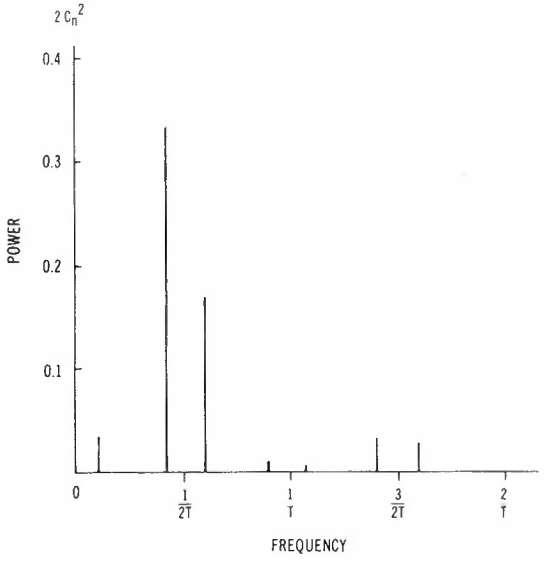


Fig.5 - Line spectrum of DDS violated AMI idle code.

Where $2C_n^2$ is the power of the spectral component at n/12ths of the data rate. The two most significant components appear at 5/12 and 7/12 of the data rate.

If a 2400 bit/s modem strapped for data transmission at 0dBm with random data is used with this idle code format, the power at each frequency would be as shown in Table 1.

TABLE 1 - Power in Idle Code for 2400 bit/s DDS modem

FREQUENCY (Hz)	POWER
200	-11 dBm
1000	-1,7 dBm
1400	-4.6 dBm
2200	-31 dBm
2600	-33 dBm

This compares with a power of +2,1 dBm at 1200 Hz in a modem transmitting binary ones according to the normal bipolar rule.

5. CONCLUSIONS

Before using violations of line codes as a means of achieving an extra signalling channel, some investigation must be undertaken to ensure that such violations do not destroy the useful properties of the original line code. This paper illustrates the effect on the spectrum of bipolar and class 4 partial response line codes of random violations of the coding rule. In these cases, increasing the number of violations increases the low frequency content of the line signal and in the case of class 4 partial response removes the frequency null at half the data rate.

By example using the Bell DDS idle code the effect of repetitive violations in producing a particular line spectrum is shown. In any proposed system of repetitive violations, a few initial calculations can avoid later problems associated with high level components at undesirable frequencies on the line.

6. ACKNOWLEDGEMENTS

The author is grateful to Mr. P. Leung and Mr. J. Billington for helpful discussions and comments concerning the material in this paper.

7. REFERENCES

1. Duc, N.Q., "Line Coding Techniques for Baseband Digital Transmission", Australian Telecommunication Research, Vol.9, pp.3-17, May 1975.

2. Duc, N.Q. and Smith, B.M., "Line Coding for Digital Data Transmission", Australian Telecommunication Research, Vol.11, No.2, pp.14-27, 1977.

3. van Derwen, P.J., "On the Generation and Application of Pseudo-Ternary Codes in Pulse Transmission", Philips Research Report, August 1965, Vol.20, pp.469-484.

4. Bender, E.C. et-al "Digital Data System : Local Distribution System", Bell Sys. Tech. J., Vol.54, pp.919-942, May-June 1975.

APPENDIX 1

Case 1 no violations

From definitions and for $n > 1$

$$P(1/1,n) = \text{prob (binary 1 received)} \times \text{prob (odd number of binary 1s in last n-1 symbols)} = 1/2 \times 1/2 = 1/4$$

$$P(-1/1,n) = \text{prob (binary 1 received)} \times \text{prob (even number of binary 1s in last n-1 symbols)} = 1/2 \times 1/2 = 1/4$$

Case 2 with violations

Firstly define

$$A_{even,even} = \sum_{\substack{m=0 \\ m \text{ even}}}^{n-1} \text{prob (m binary is in last n-1 symbols)} \times \text{prob (even number of violations in m symbols)}$$

Similarly for $A_{odd,odd}$ $A_{even,odd}$ and $A_{odd,even}$

Now from definitions and for $n > 1$

$$\bar{P}(1/1,n) = \text{prob (symbol = 1 by violation)} \times (A_{even,even} + A_{odd,odd}) + \text{prob (symbol = 1 by normal coding)} \times (A_{odd,even} + A_{even,odd})$$

and

$$\bar{P}(-1/1,n) = \text{prob (symbol = 1 by violation)} \times (A_{odd,even} + A_{even,odd}) + \text{prob (symbol = 1 by normal coding)} \times (A_{even,even} + A_{odd,odd})$$

now prob (even number of violations in m symbols) = 1/2

and prob (odd number of violations in m symbols) = 1/2

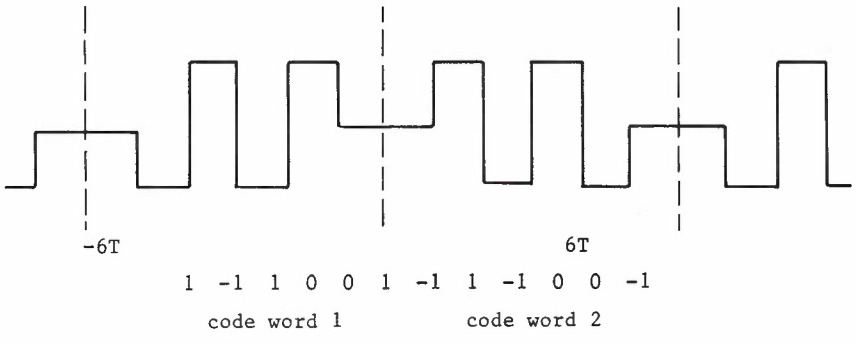
$$\therefore A_{even,even} = \sum_{\substack{m=0 \\ m \text{ even}}}^{n-1} \text{prob (m binary is in n-1 symbols)} \times 1/2 = 1/2 \times \text{prob (even number of binary is in n-1 symbols)} = 1/2 \times 1/2 = 1/4$$

Similarly $A_{odd,odd} = A_{odd,even} = A_{even,odd} = 1/4$

$$\therefore \bar{P}(1/1,n) = \tfrac{1}{2}p\ (1/4+1/4) + 1/2(1-p)\ (1/4+1/4) = 1/4$$

$$\text{and } \bar{P}(-1/1,n) = \tfrac{1}{2}P\ (1/4+1/4) + 1/2(1-p)\ (1/4+1/4) = 1/4$$

APPENDIX 2 - derivation of DDS idle code line spectrum



Because the line signal s(t) is an even periodic function and has no D.C. component, the voltage spectrum can be calculated as follows:

$$s(t) = 2\sum_{n=1} c_n \cos(n\omega_o t) \text{ where } \omega_o = \frac{2\pi}{12T}$$

where

$$c_n = \frac{1}{12T} \int_{-6T}^{6T} s(t) \cos\frac{n\pi t}{6T}\, dt$$

and by symmetry

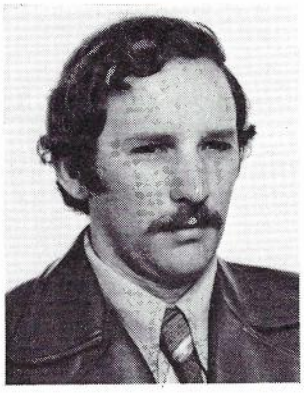
$$c_n = \begin{cases} \frac{1}{3T} \int_{0}^{3T} s(t) \cos\frac{n\pi t}{6T}\, dt & n \text{ odd} \\ 0 & n \text{ even} \end{cases}$$

$$= \begin{cases} \frac{1}{3T} \int_{T}^{2T} \cos\frac{n\pi t}{6T}\, dt - \frac{1}{3T} \int_{2T}^{3T} \cos\frac{n\pi t}{6T}\, dt & n \text{ odd} \\ 0 & n \text{ even} \end{cases}$$

$$= \begin{cases} \frac{2}{n\pi}\left(2\sin\frac{n\pi}{3} - \sin\frac{n\pi}{6} - \sin\frac{n\pi}{2}\right) & n \text{ odd} \\ 0 & n \text{ even} \end{cases}$$

and the power in a particular component is

$$2C_n^2 = \begin{cases} \frac{8}{n^2\pi^2}\left(2\sin\frac{n\pi}{3} - \sin\frac{n\pi}{6} - \sin\frac{n\pi}{2}\right)^2 & n \text{ odd} \\ 0 & n \text{ even} \end{cases}$$



BIOGRAPHY

CHRIS BEARE obtained his honours Engineering degree at the University of Adelaide in 1973 after graduating in science the previous year. That same year he began work as an Experimental Officer at the Defence Research Centre, Salisbury. In 1977 after taking study leave from the Department of Defence, he completed a Ph.D. at the University of Adelaide in the field of high speed data transmission. He joined Telecom Headquarters as a Class 2 Engineer in 1978 and is currently a senior engineer in the Datel Section of Customer Networks Branch, Telecom Headquarters where he is responsible for the design and development of Telecom's Digital Data Network.

ATR AUSTRALIAN TELECOMMUNICATION RESEARCH

ISSN 0001-2777

VOLUME 13, NUMBER 1
1979

Contents*

*Abbreviated Titles