ATR

AUSTRALIAN TELECOMMUNICATION RESEARCH

Volume 11, Number 2, 1977

A.T.R. is issued normally twice a year (in May and November) by the Telecommunication Society of Australia. Each volume comprises two regular numbers issued in one calendar year.

Subscriptions for A.T.R. 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.

Contributions to A.T.R. should be addressed to: Secretary of Editorial Board, Australian Telecommunication Research, Telecom Australia Research Laboratories, 59 Little Collins Street, Melbourne, Australia 3000.

ATR

AUSTRALIAN TELECOMMUNICATION RESEARCH

Volume 11, November 2, 1977

Contents

Challenge . . .

The deep interest taken by the public in the mining and use of uranium for power generation is focusing considerable attention on methods of electric power generation. Methods which do not cause pollution, such as solar and wind power, are now receiving increasing attention.

There are several good reasons for this; one is the avoidance of pollution, and the other is the low cost and the availability of the basic natural resource. The use of solar and wind energy does not, at first glance, cause atmospheric pollution in the same way as fossil fuels; however, if large quantities of energy are to be extracted, particularly from the wind, there will inevitably be resulting effects on weather patterns, in the same way that agglomerations of large city buildings are now modifying weather patterns.

It is claimed by advocates of solar and wind power that these resources are available free. All resources are available free; the cost of utilising them lies in paying people to make them available at the input of the energy converter; with the current attitude of unions to seek ever increasing weekly pay for continually reducing hours of work, it is obvious that a considerable escalation will take place in the cost of power generated from fossil fuels. This, apart from capital cost considerations, must inevitably swing the economic balance towards the use of solar and wind energy.

Technology for the effective utilisation of solar and wind energy is still in the early stages. In Australia, the C.S.I.R.O. has been investigating the use of solar energy for low grade heat in domestic applications, but has not applied significant effort to electric power generation in commercial quantities. Telecom Australia has used wind power for several decades in suitable remote locations and is now commencing the use of solar derived power; these applications, however, are really only flea power when considered against the commercial usage of electric power. Some universities are taking an interest in solar power, but are hampered by lack of support funds.

Australia is a country in which there is a vast area to be covered with the availability of electric power, and a country with costly distribution problems coupled with a very high and rapidly increasing wage structure. Consequently, it is a country in which the utilisation of solar and wind energy would be more attractive than it would in other countries. This utilisation can only come after considerable technological R and D, which in turn requires skilled people and funds. The skilled people are leaving for more enlightened countries and the funds are not forthcoming. It is time for the Australian Government to give a firm and positive lead in providing the funds and opportunities for the necessary R and D effort, which would form the basis of cheaper power availability and would also form technology which Australia could export to the remainder of the world.

Refractive Index Profile Determination in Optical Waveguides

P. V. H. SABINE,
Telecom Australia Research Laboratories.

An adequate characterisation of optical fibres, and of waveguides for integrated optical circuits, demands a precise quantitative description of the refractive index variation over a cross-section of the particular component. The first part of this paper reviews available techniques for measuring this refractive index profile (RIP). The second part of this paper presents practical results for RIP measurements in stress-induced channel optical waveguides.

1. INTRODUCTION

In recent years, improved fabrication techniques have decreased dramatically the transmission loss of optical fibres (Refs.1,2). Attention has now been directed towards the control of other fibre parameters. One of the most significant of these parameters is the refractive index profile (RIP) measured over a cross-section of the fibre. Both step-index and graded refractive index (GRIN) fibres are of current interest. An ideal step-index fibre has a uniform core of higher refractive index than the cladding. An ideal GRIN fibre has a RIP which varies smoothly from a maximum on the axis of the core to a constant lower value at the core-cladding interface. Practical low-loss fibres have core perturbations (Ref.3) due to the chemical vapour deposition (CVD) process (Ref.4) by which they are manufactured. A step-index fibre is a suitable transmission medium for applications requiring requiring only moderate bandwidths (Ref.5). Theoretical considerations (Ref.6) show that there exists an optimum near-parabolic RIP for which the transit time of all modes of the GRIN fibre is very nearly equalised, thus affording maximum bandwidth utilization. However, the RIP must be controlled accurately to achieve this optimum condition.

The accomplishment of low fibre transmission loss has also stimulated research into integrated optical circuits (IOCs) (Ref.7). These miniature circuits, if realised, would perform a variety of signal processing operations at the terminals of an optical fibre communication link. The basic building-block of the IOC is the channel optical waveguide, which can confine a light beam in both transverse directions. A detailed characterisation of these waveguides demands a precise knowledge of the RIP over the guide cross-section.

Thus, in both the optical fibre and IOC areas, the task of accurate RIP determination is extremely important. The first part of this paper surveys available techniques for measuring the RIP of an optical waveguide. The second part describes practical results for RIP measurements of stress-induced channel optical waveguides, which are fabricated by an embossing process.

2. SURVEY OF RIP MEASUREMENT TECHNIQUES

Many different techniques have been developed for measuring the RIP of optical waveguides. The more important techniques are now described briefly, more complete details being available in the literature references cited.

2.1 Index Matching

An adaption (Ref.8) of the oil immersion method, long used in minerology and crystallography, allows a direct measure of the refractive indices of the core and cladding of a step-index waveguide. A length of the waveguide is immersed in an oil of known, but variable, refractive index. The refractive index of the oil is changed until the oil-solid interface is no longer visible. At this point the refractive indices of the oil and solid must be identical.

2.2 Electron Microprobe Analyser

For waveguides fabricated by doping the guiding region with some index-raising material, an electron microprobe analyser can reveal the dopant concentration over a cross-section of the guide (Refs.9,10). Only when the dependence of refractive index upon material composition is known quantitatively can the RIP be derived from the dopant distribution data.

2.3 Reflection

In this technique, the reflected light intensity is monitored as a focused laser beam is scanned, at normal incidence, over a polished end-face of the waveguide (Refs.10,11). From this data and the Fresnel formula (Ref.11) the RIP may be simply calculated. But the practical problems of accurately measuring the relatively low reflected intensity, and the small variations that occur as the waveguide face is scanned,

present difficulties. Indeed an elaborate experimental arrangement is needed to isolate unwanted light signals from the detection system and to achieve the necessary sensitivity. One disadvantage of the method is that it determines the RIP only at a polished surface. The polishing process generally creates a thin damage layer in which the RIP is likely to differ significantly from the rest of the guiding region.

2.4 Interferometry - Normal to Waveguide Axis

This mode of interferometry can yield direct quantitative results for guides of simple symmetry (Ref.12). But its shortcomings in a more general case can be illustrated conveniently by considering the measurement of a circularly-symmetric optical fibre (Ref.13). The fibre must be immersed in a cell with optically flat walls and filled with index-matching oil. This transforms the fibre into a quasi-transparent object suitable for examination with a conventional interferometer. Unfortunately, as a result of the circular guide geometry, the observed fringe pattern is not a direct representation of the RIP. The RIP can be deduced only after mathematical manipulation of the interferometric data, usually by way of a digital computer.

2.5 Interferometry - Parallel to Waveguide Axis

For this particular technique (Refs. 12,14, 15,16) a transverse slice is taken from the waveguide and the faces are ground and polished until they are flat and parallel. The sample is then viewed in an interferometer, using a microscope to examine the fringe pattern.

Either a reflected (Ref.12) or a transmitted-light (Refs. 14,15,16) interferometer can be employed. A typical reflected-light measurement system, involving a Michelson interferometer, is shown in Fig.1. One face of the sample is

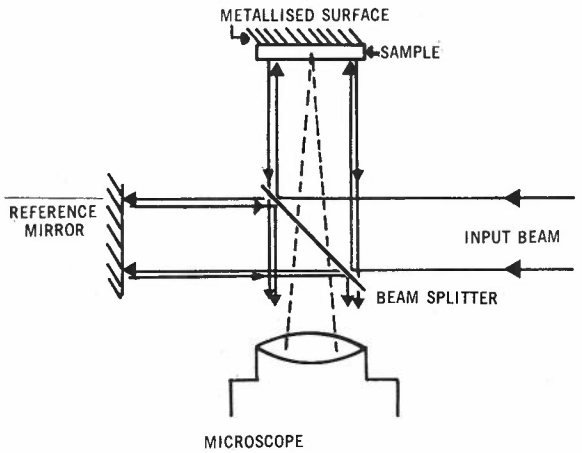


Fig.1 - A Reflected-Light (Michelson) Interferometer.

metallised to provide a mirror for the sample beam. The microscope objective is focussed through the sample onto the metallised back face. Thus the observed fringe pattern is that resulting from only a single light-pass through the sample. The Zeiss Model 1 interference microscope is a dedicated instrument designed for this type of measurement. Fig.2 depicts a typical transmitted-light system, based upon a

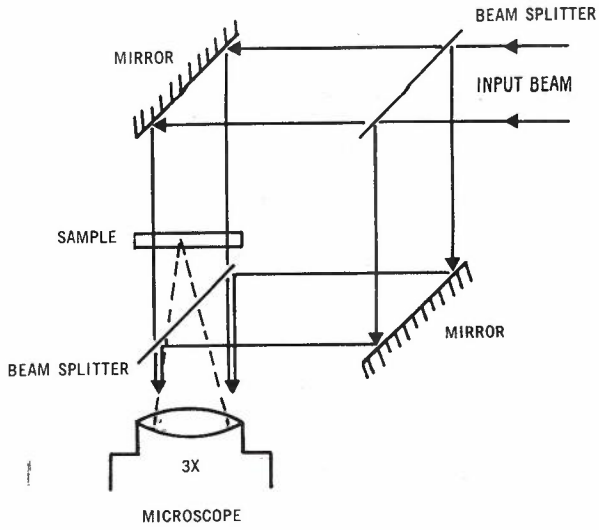


Fig.2 - A Transmitted-Light (Mach-Zehnder) Interferometer.

Mach-Zehnder interferometer. Here the microscope objective is focussed upon the nearer face of the sample. The Leitz interference microscope is a dedicated instrument that allows transmitted-light fringe patterns to be observed at magnifications up to 1000X.

With both reflected and transmitted-light interferometers there are two ways in which the interference fringe pattern can be viewed. By adjusting the tilt of the reference mirror, the sample can be observed against a background field of parallel fringes, so that the fringe displacement is directly proportional to differences in refractive index. Alternatively, the fringes can be spread apart until the entire field of view encompasses a single fringe. In this flat-field mode of operation, points of equal refractive index are connected by fringes of equal intensity, so that a contour map of refractive index is seen.

The difference in refractive index, δn, between two areas of a flat, parallel-sided sample of thickness t is

$$\delta n = q\lambda/t \qquad (1)$$

where q is the number of fringe displacements and λ is the illuminating wavelength. From a photomicrograph of the fringe display, an accurate value of the maximum δn for the waveguide can be calculated and an approximate RIP can be plotted visually or with the aid of a densitometer.

As well as non-uniform sample thickness, several other factors can contribute errors in the interpretation of fringe patterns. These include rapid spatial changes in refractive index, boundary reflections caused by skewed wavefront illumination, and waveguide effects associated with multiple reflections between boundaries. Such difficulties are minimised, generally, by using thin sections of highly multi-mode waveguides.

Two further problems (Ref.15), characteristic of GRIN media, can arise in interferometric measurements. Firstly, it has been assumed in the preceding discussion, that the

bending of rays traversing the sample is negligible and that the observer sees essentially a plane wave. But it is well known that GRIN media have focussing properties (Refs. 15,17) so that obviously ray bending can be significant.

By way of example, consider an axially symmetric GRIN optical fibre whose RIP is defined by

$$n(r) = n(o)\,[1 - \Delta n_m\,(r/a)^2] \qquad r < a \qquad (2)$$

$$= n(o)\,[1 - \Delta n_m] \qquad r \geq a$$

As illustrated in Fig.3, this describes a parabolic RIP across the fibre core of diameter 2a. In equation (2), r is a variable distance measured from the centre of the core, and Δn_m is the maximum relative refractive index difference between the core and cladding. Guided rays follow a sinusoidal path of periodic length (Refs. 15,17)

$$L = 2\pi a/\sqrt{2\Delta n_m} \qquad (3)$$

and are periodically refocussed, as shown in Fig.3. The focal length is

$$f = L/4 \qquad (4)$$

For fibres of small core diameter and large core-cladding refractive index differences, f can be quite small - eg. a = 50 μm and Δn_m = 0.01 determines that f = 555 μm. Clearly, if ray-bending effects are to be neglected, samples of thickness t << f are required. In some instances, these thin samples will produce too few fringe displacements for accurate measurements. An alternative approach (Ref.15) must be introduced to extract details of the RIP from the fringe pattern.

The second problem that arises with GRIN media is the further assumption that the entire fringe pattern is in focus when viewed through the microscope. But for GRIN samples the emergent wavefront is curved, so that the depth of focus of the microscope objective lens must be adequate to encompass the entire pattern of localised fringes in focus simultaneously. It is a simple matter to calculate the required depth of focus in any particular situation (Ref.15). This problem becomes acute when viewing GRIN waveguides of small cross-sectional dimensions at high magnifications.

The above discussion and, in particular, equations (2)-(4) relate to focussing effects in GRIN media of parabolic RIP. Similar results can be expected for GRIN media of near-parabolic profile.

Interferometric measurements made parallel to the waveguide axis offer the advantage that the fringe pattern provides a direct indication of the RIP. For this reason the technique has been used extensively in studies of both step-index and GRIN waveguides. But the technique has two serious disadvantages. Firstly, removing a transverse slice is destructive to a long fibre. Secondly, the sample preparation (Ref.16) is tedious and demands considerable care.

2.6 Near-Field Scan

Gloge and Marcatili (Ref.6) have shown mathematically that in an optical waveguide, with all modes equally excited, there exists a close resemblance between the near-field intensity distribution and the RIP. Their analysis is concerned with the particular class of circularly-symmetric RIPs defined by

$$n(r) = n(o)\,[1 - \Delta n_m\,(r/a)^{\alpha}] \qquad r < a \qquad (5)$$

$$= n(o)\,[1 - \Delta n_m] \qquad r \geq a$$

where $1 < \alpha < \alpha$ and all other symbols retain their previous definitions. Note that $\alpha = 2$ describes a guide of parabolic RIP, while $\alpha = \infty$ describes a step-index guide. Δn_m is assumed small, as is the case in practical waveguides.

When a flat end-face of such a guide is illuminated by an incoherent Lambertian source (exciting all modes uniformly) the power accepted at a radial distance r, expressed as a fraction of the power accepted at the centre of the guiding region is

$$\frac{P(r)}{P(o)} = \frac{n(r)^2 - n(a)^2}{n(o)^2 - n(a)^2} \qquad (6)$$

If all modes propagate without coupling and with equal attenuation, the same power distribution appears at the waveguide output face. This relationship, equation (6), between the RIP and the near-field power plot, was used by Burrus et al. (Ref.9) to obtain qualitative information about the RIPs of several different optical fibres. Payne et al. (Refs. 18,19) have recently refined

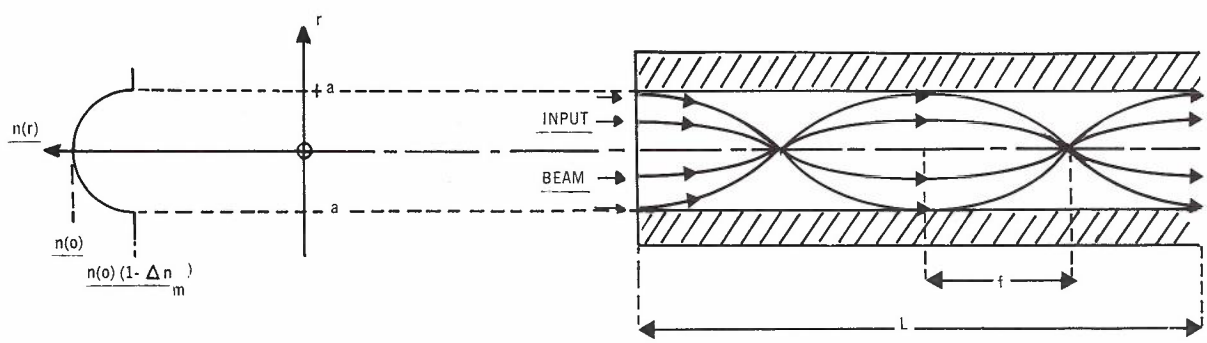


Fig.3 - Light Focussing in a GRIN Fibre with a Parabolic RIP.

this technique and, by taking account of the propagation characteristics of practical fibres, have developed a near-field scanning procedure for measuring RIPs quantitatively.

To reduce the effects of mode conversion and different mode attenuations, short lengths of optical fibres are studied. But this emphasises errors due to the presence of tunnelling leaky modes. These modes contribute additional power to the observed near-field intensity distribution, resulting in an error in the inferred RIP. The magnitude of this error decreases with fibre length as the leaky modes attenuate, but may still be significant after 100m. To eliminate this inaccuracy Payne et al. (Ref.19) introduce a length-dependent correction factor, C(r,z), to equation (6) to determine that

$$\frac{n(r) - n(a)}{n(o) - n(a)} \simeq \frac{n(r)^2 - n(a)^2}{n(o)^2 - n(a)^2} \simeq \frac{P(r)}{P(o)} \frac{1}{C(r,z)} \qquad (7)$$

$$\text{i.e.} \quad n(r) - n(a) \simeq n(o)\, \Delta n_m \frac{P(r)}{P(o)} \frac{1}{C(r,z)} \qquad (8)$$

A set of normalised correction curves, facilitating the conversion of near-field measurements to RIPs, has been published (Refs. 20,21). These curves of C(r,z) are applicable to a range of RIPs and are plotted as a function of a single fibre normalisation parameter. Calculation of this parameter for a particular fibre requires a knowledge of the fibre length, core diameter and

$$\text{numerical aperture} \simeq n(o) \sqrt{2\Delta n_m} \qquad (9)$$

But several general properties of the correction factor are worthy of note. Firstly, for an infinitely long fibre only guided modes remain at the output, so that

$$C(r,\infty) = 1 \qquad (10)$$

and the near-field intensity distribution is a direct representation of the RIP. Secondly, for an infinitely short fibre all leaky modes are present unattenuated, and

$$C(r,z) \to [1 - (r/a)^2]^{-\frac{1}{2}}, \quad z \to 0 \qquad (11)$$

In this case the RIP must be calculated from the near-field pattern. Finally, C(r,z) does not vary greatly with the precise form of the RIP - i.e. with the value of α in equation (5).

Substituting for the form of the RIP from equation (5) into equation (7) shows that

$$\frac{P(r)}{P(o)} = 1 - (r/a)^{\alpha} \quad z \to \infty \qquad (12)$$

$$\frac{P(r)}{P(o)} = \frac{1 - (r/a)^{\alpha}}{\sqrt{1-(r/a)^2}} \quad z \to 0 \qquad (13)$$

These two results are plotted in Fig.4(a) for the case $\alpha = 2$. Equations (12) and (13) indicate that a graph of corrected relative near-field intensity (in dB) versus normalised radius (log scale) is a straight line of slope α. This is illustrated in Fig.4(b).

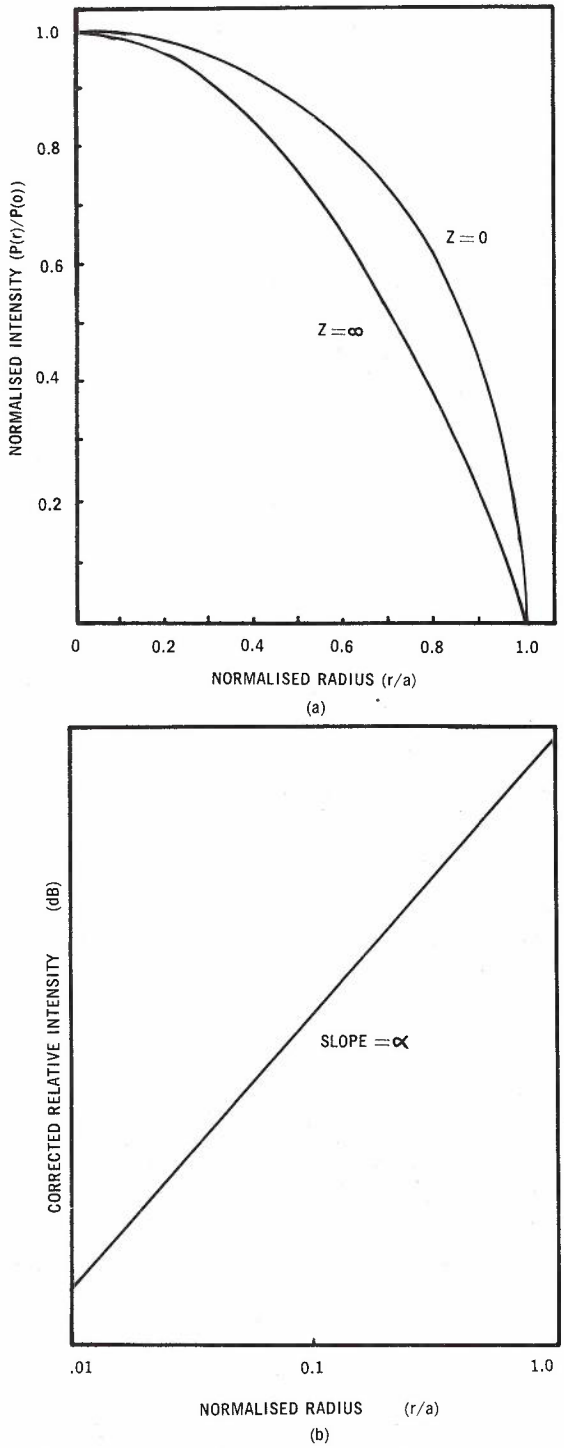


Fig.4 - Calculated Near-Field Intensity Distributions for GRIN Fibres. [After Payne et al. (Refs. 18,19)]

(a) For a fibre of parabolic RIP ($\alpha = 2$) and of length z. The curve for infinite length is equivalent to the index profile.

(b) For a general index profile parameter α.

The near-field scanning method has several attractive features. Sample preparation is minimal; a suitable length of waveguide with flat, polished end-faces is required. Most practical measurements are made with short lengths of fibre - <1m typically. The method leads to a direct description of the RIP (meaning the functional variation of refractive index over the guide cross-section). As well, a direct quantitative measure of the parameter α results. It is this parameter that must be closely controlled if practical GRIN fibres exhibiting low modal dispersion are to be fabricated (Ref.6). The most serious disadvantage is that the technique does not yield a quantitative measurement of either n(o), n(a) or Δn_m. Hence, in order to characterise a waveguide completely, and indeed to evaluate C(r,z) in any given situation, the near-field scanning technique must be complemented by another measurement that determines quantitatively one or more of these parameters.

2.7 Etching and SEM Examination

Workers at Bell Laboratories (Refs. 3,14) have developed a technique for examining glass fibres and preforms fabricated by doping the core region with some index-raising material. Use is made of the fact that glasses of different compositions tend to etch at different rates. A cleanly broken end of the fibre is etched in hydrofluoric acid for a few minutes. The etched surface is metal-coated and then studied either by reflected-light interferometry or with a scanning electron microscope (SEM). Surface etching has proved a valuable aid in the detection of small-scale inhomogeneities and asymmetry. Unfortunately the technique does not produce quantitative information regarding the RIP. Structural properties other than glass composition can effect etch rates and create misleading results. One such structural property is a region of stress.

3. COMPARISON OF RIP MEASUREMENT TECHNIQUES

In each situation the particular technique, or combination of techniques, that provides all of the required refractive index data must be selected. To facilitate the selection process, the most important features of each measurement procedure are summarised in Table 1. In the table a technique is deemed destructive if it is necessary to expose one or more faces of the waveguide to measure the RIP at any given point along the length of the guide.

4. STRESS-INDUCED OPTICAL WAVEGUIDES

Channel optical waveguides can be fabricated by a variety of techniques (Ref.22) the simplest of which is embossing (Refs. 23,24). In this process a die is used to emboss a groove into the surface of a thermo-plastic substrate. The groove is filled with another material, of higher refractive index than the substrate, to form an optical waveguide. Recently we have observed low-loss light-guiding in the vicinity of unfilled embossed grooves (Refs. 25,26,27).

These novel light guides are prepared by a room-temperature embossing process (Refs. 26,27), generally using a 50 x 50 x 3 mm^3 substrate of commercially available poly (methyl methacrylate) (PMMA) and a 254 µm diameter copper wire as the embossing die. The grooves thus formed are approximately 330 µm wide and 35 µm deep. Fig.5 shows the method of waveguide excitation. The output of a helium neon laser (λ = 0.63 µm) is focussed to a spot directly beneath the groove at one end-face. Careful alignment of the focussed input beam allows guiding to be established individually in any one of three regions, A, B or C, surrounding the embossed groove. As sketched in Fig.5, two of these regions, A and C, are located at the corners of the groove.

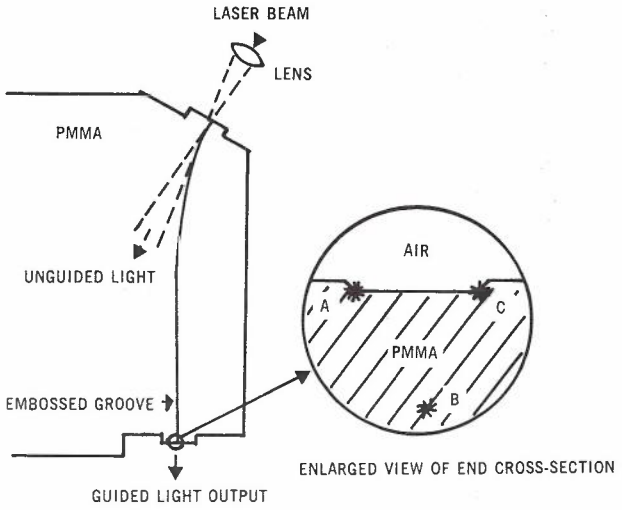


Fig.5 - Method of Waveguide Excitation and Definition of Guiding Regions.

The third region, B, is embedded at a depth of about 250 µm beneath the middle of the groove. Fig.6 illustrates the two different types of light guide. Detailed investigations (Refs. 26, 27) reveal that guiding is attributable to residual stresses in the PMMA substrate beneath the groove.

The embedded waveguide provides a guided light beam with considerable immunity from substrate surface damage and seems to offer the greater potential for optical communications applications. Therefore more recent studies have been aimed at characterising the light-guiding properties of the embedded region B. An average attenuation of 0.045 dB/cm has been recorded, at a wavelength of 0.63 µm, over a 22cm long waveguide (Refs. 28,29). Two attractive features of the stress-induced waveguide are the very simple fabrication process and the low optical attenuation. Short lengths of these waveguides could become a basis for the construction of practical IOCs. A "rolled embossing" process developed at Telecom Australia Research Laboratories (Refs. 28,29,30) is a convenient technique for manufacturing long waveguides, which could find applications as optical data buses.

To date, the RIP has remained the most significant unknown characteristic of the stree-induced waveguide.

5. RIP MEASUREMENTS OF STRESS-INDUCED WAVEGUIDES

It is intuitively obvious that the embedded stress-induced optical waveguide must be a region of GRIN material. Two techniques, listed

TABLE 1 - Comparison of Techniques for Measuring RIP of Optical Waveguides. (δn = refractive index variation).

TECHNIQUE (and references)	WAVEGUIDE SAMPLE		RESULTS					CONTENTS
	Preparation	Destructive?	Direct Measure?		Quantitative?	Resolution		
			δn	RIP		δn	Spatial	
INDEX MATCHING (Ref.8)	Nil	No	Yes	No	Yes	High	-	Simple technique. Applicable only to step-index type waveguides (w/gs).
ELECTRON MICROPROBE ANALYSER (Refs.9,10)	Nil	Yes	No	No	No	-	Good	Applicable only when guiding region formed by doping host medium with index-raising material.
REFLECTION (Refs.10,11)	Minimal	Yes	Yes	Yes	Yes	Low-Medium	High	Elaborate system needed to achieve required sensitivity. Surface measurement only.
INTERFEROMETRY normal to w/g axis (Refs.12,13)	Nil	No	No	No	Yes	Good	Medium	Digital computer generally required to extract RIP from interferometric data.
INTERFEROMETRY parallel to w/g axis (Refs.12,14,15,16)	Tedious and Exacting	Yes	Yes	Yes	Yes	High	Good	Extensively used for step-index and GRIN w/gs. Dedicated microscopes for transmission and reflection interferometry available.
NEAR-FIELD SCAN (Refs.9,18,19)	Little	Yes	No	Yes	No	High	High	Rapid technique. Simple equipment. W/g length-dependent correction factor must be applied to data to deduce accurate RIP. Finer δn and spatial resolution than interferometry. Particularly applicable to CVD fibres.
ETCHING and SEM EXAMINATION (Refs.3,14)	Little	Yes	No	No	No	-	-	Applicable to glass w/gs formed by doping host medium with index-raising material. Rapid technique for detecting small scale inhomogeneities or asymmetry.

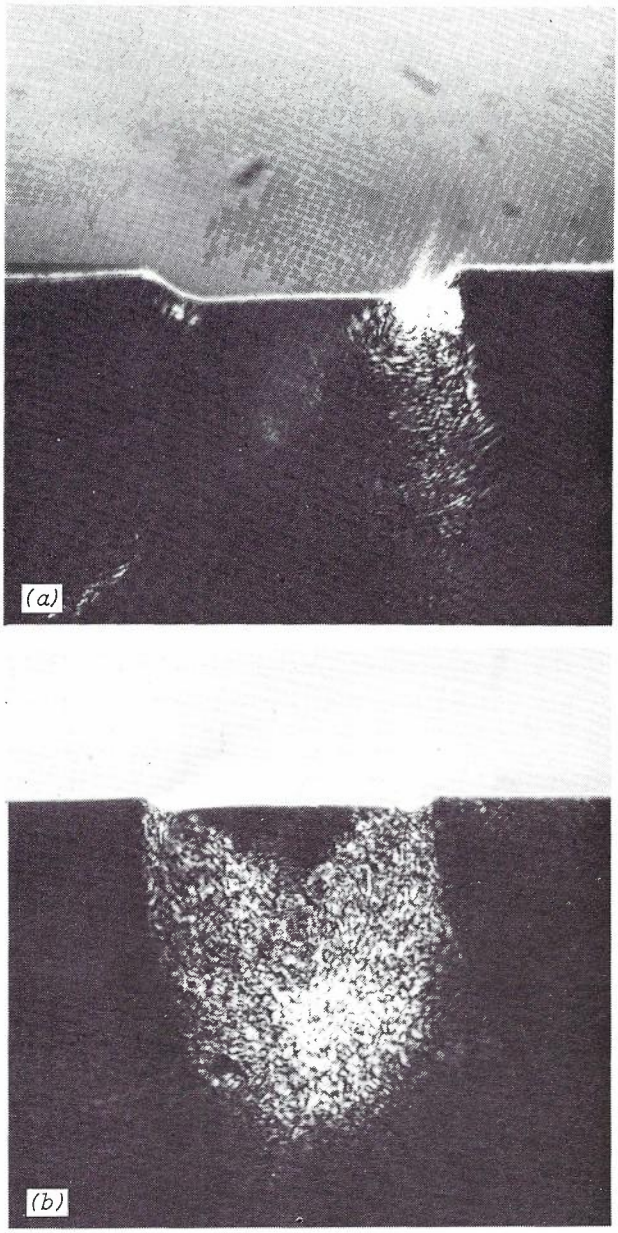


Fig.6 - Light Guiding in a Stress-Induced Waveguide.
(b) Guiding in Region B.
(a) Guiding in Region C.

in Table 1, have been used widely to measure RIPs of GRIN waveguides:- interferometry, parallel to the waveguide axis, and near-field scanning. Both techniques were tried in turn.

5.1 Interferometric Measurements

Far-field radiation pattern studies (Ref.29), which enabled a crude estimate of Δn_m for the embedded waveguide, suggested that t = 800 μm was a desirable sample thickness for interferometric examination. The problems associated with obtaining a flat, parallel-sided, optically-polished transverse slice 800 μm thick, from a waveguide fabricated in a PMMA substrate, were formidable. A realization that heat generated during the sample preparation process would partially or completely anneal the stress region, thus distorting the RIP, made these problems more acute. Suitable samples were prepared as follows.

A straight section of waveguide was embossed, with a 254 μm diameter copper wire die, on one major face of a 50 x 50 x 3 mm^3 PMMA sheet. The light-guiding characteristics of this waveguide were verified. Three transverse slices, each about 3mm thick, were slowly and carefully cut from the guide using a jeweller's saw. Each segment was trimmed to a length of 15 mm. The three segments were glued into a framework, made from 3mm thick PMMA to form a single disc of 30mm diameter, as sketched in Fig.7. All joints were filled completely with an adhesive whose hardness was similar to that of PMMA. The PMMA disc was mounted in a mechanical polishing machine. Both sides were polished slowly in a slurry of water and successively finer abrasive powders. Heat generation was minimised by keeping the sample totally immersed in the slurry.

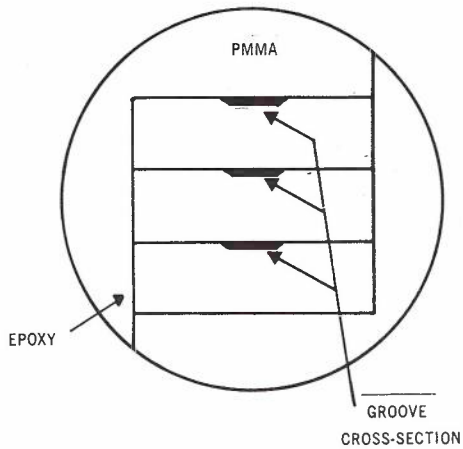


Fig.7 - Stress-Induced Waveguide Samples for Interferometric Examination.

When the disc had been polished to its final thickness, both surfaces were examined with a Michelson interferometer to determine their flatness and parallelism. Shortly after polishing a flatness of better than .05 μm could be measured over the region of interest surrounding each groove. After several hours, small lumps began to appear in the PMMA beneath the grooves. Presumably these lumps were caused by the slow extrusion of PMMA from the stressed region. Storing the polished samples at low temperatures (-5°C) retarded lump formation. But in order to obtain accurate RIP data, the samples were studied as soon as possible after final polishing.

The samples were studied by transmitted-light interferometry, using a Mach-Zehnder interferometer in the arrangement shown in Fig.2. Experience soon showed that, because of the birefringence of the guiding regions, the most meaningful fringe patterns (Ref.31) could be obtained by:

(1) using a polarised light source (helium neon laser, λ = 0.63 μm),

and

(2) including a polariser in the sample beam of the interferometer. The polariser was

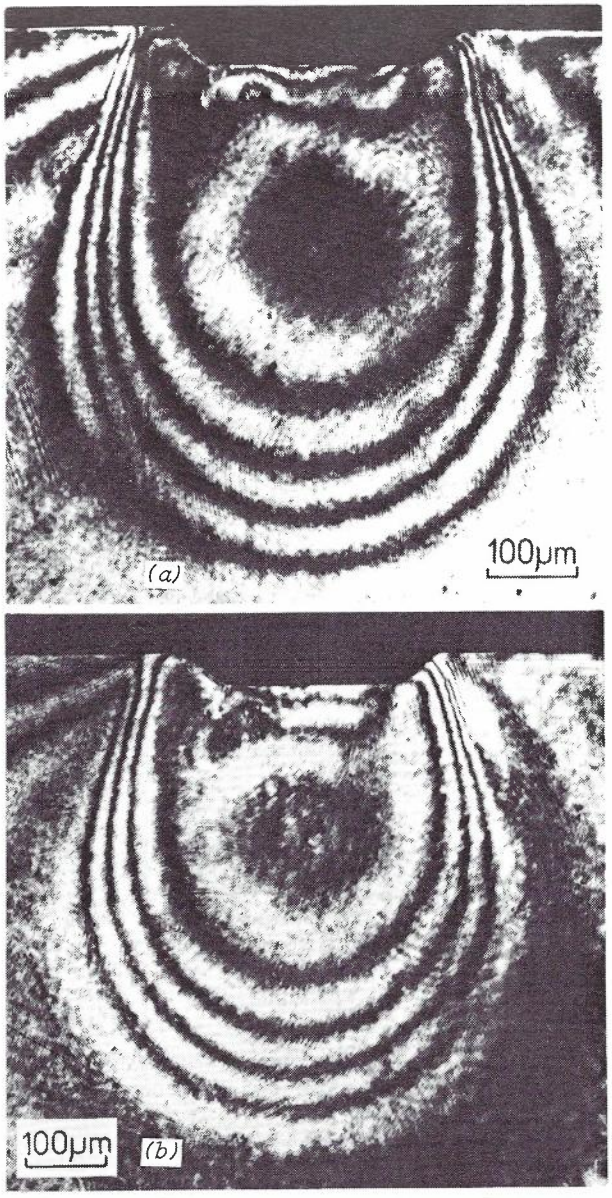


Fig.8 - Interferograms of Stress-Induced Waveguide Sample - t = 790 μm.

(a) Incident light polarised parallel to PMMA surface.

(b) Incident light polarised normal to PMMA surface.

aligned for maximum transmission of the incident polarised beam, and was placed immediately beneath the sample.

The photomicrographs of Fig.8 picture the observed fringe pattern for a sample of thickness t = 790 μm. By rotating the sample, fringe patterns for various orientations of the input beam polarization were studied.

The photographs shown in Fig.8 were recorded with a single fringe background, so that the fringes describe contours of equal refractive index. Several important features may be noted. Firstly, there is an embedded high refractive index region directly beneath the middle of the embossed groove. Secondly, this high-index region is surrounded by a GRIN region of lower refractive index, and of quasi-circular symmetry. A comparison of Fig.8 with the fringe pattern for a GRIN fibre of parabolic RIP (Refs. 12,15) highlights some similarities. Thirdly, the embedded high refractive index region has "lobes" that extend up towards the corners of the grooves. Careful alignment of the interferometer permitted one small circular fringe to be formed at the corners of the groove. These fringes, which indicate localised high refractive index regions, are evident in the photomicrographs. A further comparison of Figs.6 and 8 clearly shows that the guiding regions observed in practice do indeed correspond to the high refractive index regions.

From Fig.8, the number of fringe displacements is q = 5.5 for the sample of thickness t = 790 μm. Equation (1) can be rewritten in the form

$$\Delta n_m \simeq q\lambda/t\, n(o) \qquad (14)$$

Substituting n(o) ≃ 1.492, which is the refractive index of bulk PMMA, determines that the maximum relative refractive index difference for the embedded waveguide region is $\Delta n_m = 0.3\%$.

Fig.8 shows also that the diameter of the embedded waveguide is 2a = 590 μm. The focal length of the guide can be estimated from equations (3) and (4), which strictly apply only for parabolic RIPs, as f ≃ 6 mm.

One disturbing feature of the results presented in Fig.8 is that, in each photomicrograph, two supposedly separate dark fringes join together. A detailed stress analysis (Ref.31) proves that this phenomenon is due to the characteristics of the localised stress-induced birefringence. Because of the ambiguity introduced by the joining fringes, and to obtain added information about the RIP, a near-field scan measurement was undertaken.

5.2 Near-Field Scan Mesurements

The near-field scanning system is illustrated in Fig.9. Light from a tungsten filament lamp passed through a diffuser and was focussed onto the end-face of the waveguide. The waveguide sample consisted of a straight embossed section, 3.5 cm long, with flat, polished end-faces. A magnified image of the waveguide output face was formed in the plane of an apertured silicon PIN photodiode, which could be scanned automatically in two dimensions. The photodiode output was amplified and fed into a ten-level comparator. This comparator controlled the pen-down function of the X-Y recorder. In this mode of operation the X-Y recorder plotted directly an intensity contour map of the near-field pattern.

A typical uncorrected near-field intensity contour plot of a stress-induced waveguide, formed by embossing with a 254 μm diameter copper wire die, appears in Fig.10(a). The general similarity of this plot and the fringe patterns of Fig.9 is apparent. The near-field plot clearly shows the central high intensity (and hence high refractive index) region of the embedded guide, and two high-intensity regions near the corners of the groove, corresponding to the corner guides.

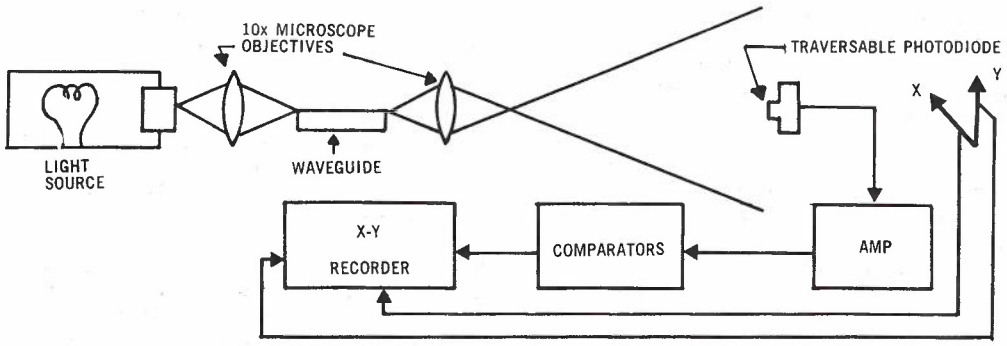
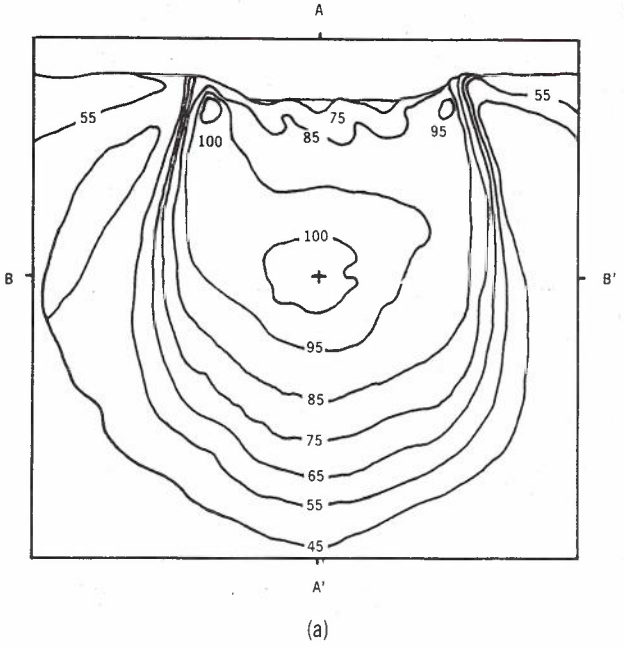
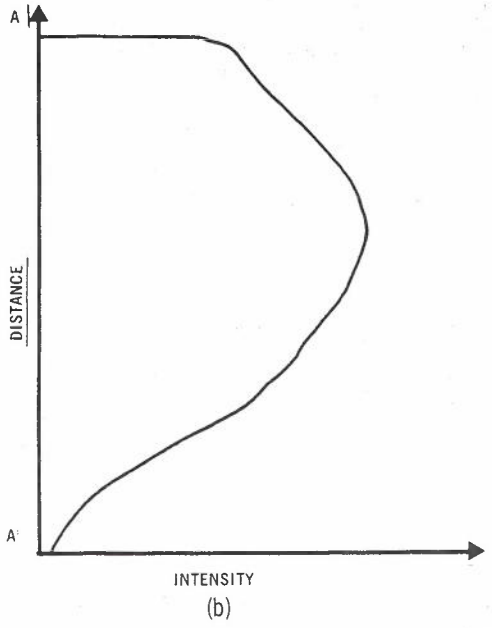


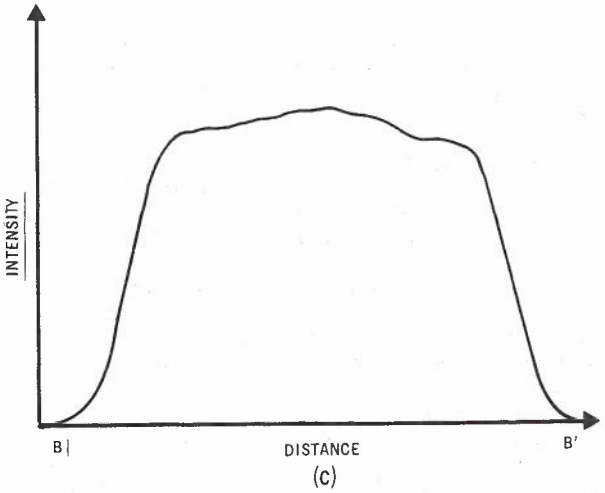
Fig.9 - The Experimental Near-Field Scanning System.



(a)



(b)



(c)

Fig.10 - Near-Field Scan Results for a Stress-Induced Waveguide.

(a) Near-Field Intensity Contour Plot, uncorrected for presence of Leaky Modes.

Numbers indicate relative intensity.

(b) Uncorrected intensity scan normal to embossed surface.

(c) Uncorrected intensity scan parallel to embossed surface.

When the comparator circuitry was by-passed and the amplified photodiode output connected directly to the Y input of the recorder, a one-dimensional intensity plot could be drawn. Figs. 10(b) and (c) are two uncorrected orthogonal scans of intensity, each scan passing through the centre of the embedded high refractive index region. These scans reveal that the RIP normal to the embossed surface is characteristic of a GRIN medium and could be approximated by a parabolic function, while the RIP parallel to the surface is more characteristic of a step-index variation.

6. CONCLUSION

The first part of this paper has reviewed available techniques for measuring RIPs of optical waveguides. Unfortunately, the most convenient and informative techniques all require that at least one face of the waveguide be exposed at the point of measurement. This makes the technique destructive for long lengths of optical waveguides and fibres. At present, the best optical fibres are being made by CVD techniques (Ref.4). In this process the RIP of the fibre is virtually determined during the preform fabrication. The outer diameter is the only parameter that can be controlled at the fibre drawing stage. There is a need to develop a rapid, non-destructive method for evaluating

RIP data, such as the thickness and refractive index of each deposited layer, during preform fabrication. This would allow some degree of on-line feedback control of RIPs to be introduced to the fibre manufacturing process.

The second half of the paper has described RIP measurements in a particular type of stress-induced optical waveguide. We have demonstrated previously (Refs. 26,28) that these waveguides can exhibit sufficiently low attenuation to be of use in the construction of practical optical devices. Furthermore the fabrication process is very simple and is amenable to mass production. The measurements reported here show that the stress-induced guide is a GRIN waveguide possessing light focussing properties. This suggests two additional potential areas of application. The first is in the development of image transmission devices. The second is in the fabrication of simple optical circuit components, such as power splitters, that are compatible with GRIN fibres. An important area of present research is to discover the shape and material composition of dies needed to emboss waveguides of specified RIPs.

ACKNOWLEDGEMENTS

The technical assistance of Mr. P. Francis is gratefully acknowledged. Discussions with Mr. D. Payne, of Southampton University, and with colleagues Drs. W. Williamson, N. Teede and L. Cahill, were most helpful. Dr. L. Cahill and Mr. L. Lorrain must be thanked for assistance with the scanning circuitry.

REFERENCES

1. Payne, D.N. and Gambling, W.A., "New Silica-Based Low-Loss Optical Fibre", Electron. Lett., July 1974, Vol. 10, No.15, pp. 289-290.

2. Horiguchi, M. and Osanai, H., "Spectral Losses of Low-OH-Content Optical Fibres", Electron. Lett., June 1976, Vol.12, No.12, pp. 310-312.

3. Presby, H.M., Standley, R.D., MacChesney, J.B. and O'Connor, P.B., "Material-Structure of Germanium-Doped Optical Fibres and Preforms", Bell Syst. Tech. J., Dec. 1975, Vol.54, No.10, pp. 1681-1692.

4. Gambling, W.A., Payne, D.N., Hammond, C.R. and Norman, S.R., "Optical Fibres Based on Phosphosilicate Glass" Proc. IEEE, June 1976, Vol.123, No.6, pp. 570-576.

5. Miller, S.E., Marcatili, E.A.J., and Li, T., "Research Toward Optical-Fibre Transmission Systems. Part I : The Transmission Medium", Proc. IEEE, Dec. 1973, Vol.61, No.12, pp. 1703-1726.

6. Gloge, D. and Marcatili, E.A.J., "Multimode Theory of Graded-Core Fibres", Bell Syst. Tech. J., Nov. 1973, Vol.52, No.9, pp.1563-1578.

7. Tamir, T. (Ed.), "Integrated Optics", Springer-Verlag, Berlin, 1975.

8. Liu, Y.S., "Direct Measurement of the Refractive Indices for a Small Numerical Aperture Cladded Fibre: A Simple Method", Appl. Opt., June 1974, Vol.13, No.6, pp. 1255-1256.

9. Burrus, C.A., Chinnock, E.L., Gloge, D., Holden, W.S., Li, T., Standley, R.D., and Keck, D.B., "Pulse Dispersion and Refractive-Index Profiles of Some Low-Noise Multimode Optical Fibres", Proc. IEEE, Oct. 1973, Vol.61, No.10, pp. 1498-1499.

10. Eickhoff, W. and Weidel, E., "Measuring Method for the Refractive Index Profile of Optical Glass Fibres", Optical and Quantum Electronics, March 1975, Vol.7, No.2, pp. 109-113.

11. Ikeda, M., Tateda, M. and Yoshikiyo, H., "Refractive Index Profile of a Graded Index Fibre: Measurement by a Reflection Method", Appl. Opt., April 1975, Vol.14, No.4, pp. 814-815.

12. Martin, W.E., "Refractive Index Profile Measurements of Diffused Optical Waveguides", Appl. Opt., Sept. 1974, Vol.13, No.9, pp. 2112-2116.

13. Marhic, M.E., Ho, P.S., and Epstein, M., "Non-destructive Refractive-Index Profile Measurements of Clad Optical Fibres", Appl. Phys. Lett., May 1975, Vol.26, No.10, pp. 574-575.

14. Burrus, C.A. and Standley, R.D., "Viewing Refractive-Index Profiles and Small-Scale Inhomogeneities in Glass Optical Fibres: Some Techniques", Appl. Opt., Oct. 1974, Vol.13, No.10, pp. 2365-2369.

15. Stone, J. and Burrus, C.A., "Focussing Effects in Interferometric Analysis of Graded-Index Optical Fibres", Appl. Oct., Jan. 1975, Vol.14, No.1, pp. 151-155.

16. Presby, H.M., Mammel, W. and Derosier, R.M., "Refractive Index Profiling of Graded Index Optical Fibres", Rev. Sci. Instrum., March 1976, Vol.47, No.3, pp. 348-352.

17. Matsushita, K. and Ikeda, K., "Newly Developed Glass Devices for Image Transmission", Proc. Soc. Photo-optical Instrumentation Engineers, Oct. 1972, Vol.31, pp. 23-35.

18. Payne, D.N., Sladen, F.M.E. and Adams, M.J., "Index Profile Determination in Graded Index Fibres", Proc. First European Conference on Optical Fibre Communication", Sept. 1975, pp. 43-45.

19. Sladen, F.M.E., Payne, D.N. and Adams, M.J., "Determination of Optical Fibre Refractiv Index Profiles by a Near-Field Scanning Technique", Appl. Phys. Lett., March 1976, Vol.28, No.5, pp. 255-258.

20. Adams, M.J., Payne, D.N. and Sladen, F.M.E "Length-Dependent Effects due to Leaky Modes or Multimode Graded-Index Optical Fibres", Opt. Commun., May 1976, Vol.17, No.2, pp. 204-209.

21. Adams, M.J. Payne, D.N. and Sladen, F.M.E "Correction Factors for the Determination of Optical-Fibre Refractive Index Profiles by the Near-Field Scanning Technique", Electron. Lett May 1976, Vol.12, No.11, pp. 281-283.

22. Pavlopoulos, T.G. and Crabtree, K., "Fabrication of Channel Optical Waveguides in Glass by CW Laser Heating", J. Appl. Phys. Nov. 1974, Vol.45, No.11, pp. 4964-4968.

23. Ulrich, R., Weber, H.P., Chandross, E.A., Tomlinson, W.J., and Franke, E.A., "Embossed Optical Waveguides", Appl. Phys. Lett., March 1972, Vol.20, No.6, pp. 213-215.

24. Aumiller, G.D., Chandross, E.A., Tomlinson, W.J. and Weber, H.P., "Submicrometer Resolution Replication of Relief Patterns for Integrated Optics", J. Appl. Phys., Oct. 1974, Vol. 45, No.10, pp. 4557-4562.

25. Sabine, H., "Loss Measurements in Embossed Optical Waveguides", IREE International Electronics Convention Digest, Aug. 1975, pp. 104-105.

26. Sabine, P.V.H., "Stress-Induced Light Guiding in Embossed Waveguides", Electron, Lett., Oct. 1975, Vol.11, No.21, pp. 501-502.

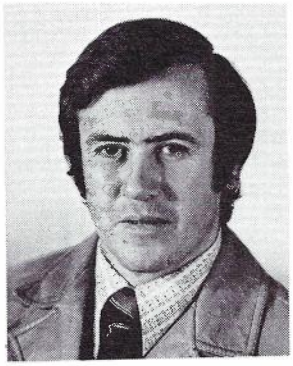
27. Sabine, P.V.H., "Stress-Induced Channel Optical Waveguides Fabricated by an Embossing Process", Telecom Australia, Res. Lab. Report No. 7022, 1976.

28. Sabine, P.V.H., "Loss Measurements in Stress-Induced Optical Waveguides Fabricated by an Embossing Process", Electron, Lett., March 1976, Vol.12, No.5, pp. 115-116.

29. Sabine, P.V.H., "Measurement of Attenuation in Stress-Induced Optical Waveguides Fabricated by an Embossing Process", Telecom Australia, Res.Lab. Report No.7023, 1976.

30. Sabine, P.V.H. and Francis, P.S. : Australian Provisional Patent PC4499, 1976.

31. Sabine, P.V.H., to be published as Telecom Australia, Res. Lab. Report.



BIOGRAPHY

HARVEY SABINE received the degrees of B.Sc. and B.E., with first class honours, from the University of Adelaide in 1966 and 1968, respectively. The same University awarded him a Ph.D. in 1973 for research in the field of surface acoustic wave devices. In 1972 he joined the Solid State and Quantum Electronics Section of the Research Laboratories of Telecom Australia, where currently he holds the position of Senior Engineer. His present interests include laser sources and modulators, terminal devices for optical fibre communication links and components for integrated optical circuits.

Line Coding for Digital Data Transmission

N. Q. DUC,
B. M. SMITH
Telecom Australia Research Laboratories

This paper presents a broad survey of codes for baseband digital line transmission. The various codes are incorporated in a framework which should assist in understanding their characteristics vis-a-vis other codes and also allowing any new codes to be readily classified.

Only a brief description of each code and its properties is given, together with applications where appropriate; references are cited which give a fuller description of the codes and their comparative behaviour. To assist the reader an index is provided, together with a list of codes that have equivalent names.

1. INTRODUCTION

In an earlier paper (Ref.1) a survey of baseband coding techniques for digital line transmission was carried out. The review was mainly confined to techniques found in and/or applicable to cable pulse code modulation (PCM) transmission systems. In this paper the survey is expanded to cover a broader class of line coding methods and not just those commonly found in PCM applications.

The paper begins with a discussion on why a line code is used in baseband digital systems. This is followed by an index of the line codes referred to in the paper, and to further assist the reader, we have prepared a list of codes that have two or more equivalent names. In Section 3 a description of linear codes is given. These encompass partial-response (or correlative-level) coding and two-phase (or frequency) coding. Non-linear codes which include alphabetic and non-alphabetic codes are discussed in Section 4.

2. PRELIMINARIES

In baseband digital line transmission it is highly desirable that the transmitted signal has the following two properties :

a. It must carry sufficient timing information which can be recovered at the end of a transmission section. This clock information may be extracted from the received line signal itself, or it may be an added timing component which has been inserted into the transmitted line signal.

b. The transmitted signal must have no DC component and may contain only a small amount of low-frequency (LF) components. This arises from the requirement that the digital signal is normally AC coupled to the transmission medium. Furthermore, in some applications (eg. PCM systems) DC power is fed to the remote regenerative repeaters over the same cable pair as the digital signal.

To satisfy the above transmission requirements some form of line coding is adopted. This code can be defined as an encoding procedure which converts a data sequence into another sequence satisfying the previous two requirements for baseband line transmission. The line signal then consists of encoded pulses which may be shaped to give a more desirable time and/or frequency response, eg. rectangular pulse shape (with a certain duty ratio), raised cosine or gaussian time or frequency response depending on the type of application. In this paper we often find it convenient to separate the coding aspects from the pulse shaping of the line signal as the structure and properties of a line code are best described and understood in this way. However the examples of the various line codes that are given use the pulse shaping that is commonly associated with the particular line code being considered.

Although some form of redundancy in the line signal is not essential (see Sect. 3.2) to satisfy the above requirements, it is a characteristic of all the line codes in this paper with the exception referred to in Sect. 3.2. The redundancy can be introduced by:

a. Increasing the signalling rate (or symbol or line rate) with respect to the input data rate, and/or

b. Increasing the number of levels in the transmitted signal.

2.1 Index to Line Coding Techniques Described In This Survey

Line Code/Signal	Section
Alphabetic Codes	4.1
Alternative-Mark-Inversion (AMI) Code	3.3
Binary Alphabetic Codes	4.1(a)
" Non-Alphabetic Codes	4.2(a)
" Signal	3.2
Biphase	3.4
Biphase-L	3.4
Biphase-M	3.4
Bipolar Code	3.3
Biternary Code	3.3
B6ZS Code	4.2(b)
Carter Code	4.2(a)
CHDBn Code	4.2(b)
Class 4 Signal	3.3
Conditioned Diphase	3.4
Correlative-Level Coding	3.3
Delay Modulation	4.2(a)
Dicode	3.3
Differential Encoding	3.1
Diphase	3.4
Duobinary Signal	3.3
Feedback Balanced Codes	4.2(b)
Filled Bipolar or AMI Codes	4.2(b)
FOMOT Code	4.1(b)
Frequency Doubling Coding	3.4
" Modulation	3.4
Generalized Partial Response Coding	3.3
Harvey Code	4.1(a)
HDBn Code	4.2(b)
Higher-Order Bipolar Codes	3.3
Interran Pulse Position Coding	3.5
Linear Codes	3
L742 Code	4.1(b)
Manchester Codes	3.4
Miller Code	4.2(a)
Modified Duobinary Signal	3.3
" Pulse Doublet Code	3.5
" Two-Phase Modulation	3.5
MS43 Code	4.1(b)
Multilevel Alphabetic Codes	4.1(b)
" Bipolar Codes	3.3
" Non-Alphabetic Codes	4.2(b)
Multiple-Response Coding	3.3
Neu A-Code	4.1(a)
Neu B-Code	4.1(a)
Non-Alphabetic Codes	4.2
Non-Linear Codes	4
Non-Return-to-Zero-Inverse Signal (NRZ-I)	3.1
Non-Return-to-Zero-Level Signal(NRZ-L)	3.1
" " -Mark " (NRZ-M)	3.1
Paired-Selected Ternary (PST) Code	4.1(b)
Partial-Response (PR) Coding	3.3
" " Class 4 Signal	3.3
Phase Modulation	3.4
Polarity Pulse Coding - Binary	4.2(a)
Polarity Pulse Coding - Multilevel version	4.2(b)
Precoded Diphase	3.4
Precoding	3.1,3.3
Pulse Doublet Coding	3.5
Pulse Position Coding	3.5
Time Polarity Control (TPC) Coding	4.2(b)
Transparent Interleaved Bipolar (TIB) Code	4.2(b)
Twinned Binary Code	3.3
Two-Level AMI Class I Code	3.5
" " " " II Code	3.4
Two-Phase (Or Frequency) Modulation	3.4
Unit-Disparity Code	4.1(a)
VL43 Code	4.1(b)
WAL-1 Code	3.4
WAL-2 Code	3.4
Zero-Disparity Code	4.1(a)
2B-3B Code	4.1(a)
3B-2Q Code	4.1(b)
4B-3T Codes	4.1(b)
6B-3QI Code	4.1(b)
6B-4T Code	4.1(b)
90^{o} Carrier Diphase	3.4
10B-7T Code	4.1(b)

2.2 Equivalent Code Names

Partial-Response Coding
Multiple-Response Coding
Correlative-Level Coding

Alternative-Mark-Inversion (AMI)
Bipolar

Twinned Binary
Dicode

Class 4 Partial Response
Modified Duobinary

Conditioned Diphase
Precoded Diphase
Biphase-M
Frequency-Doubling Coding
2-Level AMI Class II

90^{o} Carrier Diphase
WAL-2

Delay Modulation
Miller Code

3. LINEAR CODES

A line code is said to be linear if the encoded sequence it produces can be completely derived from the input data (or its precoded version) using a linear relationship. The basic binary signal discussed in Section 3.2 is the only line signal in this class not containing any form of redundancy. On the other hand, correlative-level (Ref.6) or partial-response (Ref.7) coding uses an increase in the number of transmitted levels, while two-phase modulation coding employs an increase in signalling rate.

3.1 Precoding

In some line codes, it is desirable to precode or transform the basic data sequence into another sequence of the same number of levels to achieve a certain decoding advantage over the unprecoded data. The advantage of precoding is described in a subsequent section. However because this technique is encountered several times in this paper, it is convenient to discuss it here. It should be noted that providing the original data symbols are equally likely and statistically independent, precoding does not alter the statistics of the data sequence. (Ref.19).

To illustrate the concept of precoding, the following very common example is appropriate. Consider the block diagram of a simple binary precoder illustrated in Fig.1. The precoded sequence $\{b_n\}$ is obtained by modulo-2 addition of the input data sequence $\{a_n\}$ and the sequence $\{b_n\}$ delayed by a single period, i.e.:

$$b_n = a_n + b_{n-1} \quad \text{Mod 2} \qquad (1)$$

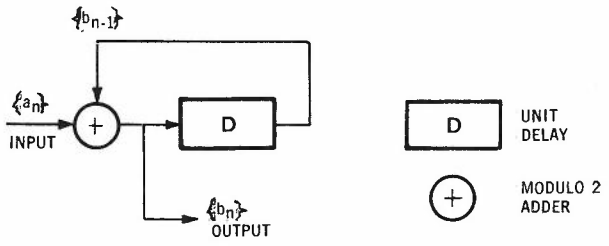


Fig.1 - A simple example of binary precoding.

The above precoding technique is in fact the differential encoding method (Ref.2) that is widely used in digital magnetic recording. Given a binary data sequence the differentially encoded waveform consists of transitions corresponding to the "ones" and no transitions otherwise. This binary waveform is also known as Non-Return-to-Zero-Mark (NRZ-M) or Non-Return-to-Zero-Inverse (NRZ-I) while the one-mark and zero-space representation is known as Non-Return-to-Zero-Level (NRZ-L). From Fig.2 one can readily see that the NRZ-M representation is identical to the NRZ-L representation of the precoded data sequence.

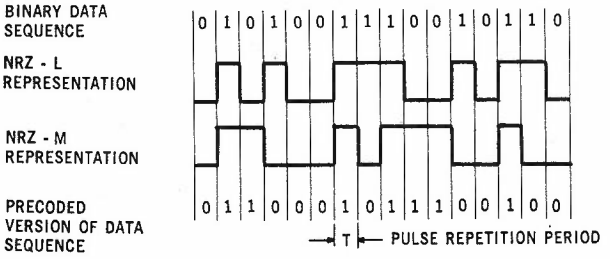


Fig.2 - NRZ-L and NRZ-M representations of a binary data sequence.

Obviously the modulo-2 operation can be generalized to modulo-m operation for multilevel signals, while the more general form of precoding is described in Section 3.3.

3.2 Binary Signal

The simplest form of line signal is the binary signal and it may be suitably modified to satisfy the two requirements as stated in Section 2.

- The binary signal may be scrambled to ensure sufficient transitions from which timing information can be recovered from the received signal.
- To satisfy the second requirement, the DC and LF components are suppressed and filtered before transmission. At the receiver these may be reinstated by local generation known as DC restoration or quantized feedback (Ref.3).

Two examples of applications of the above technique are the baseband transmission of the groupbandwidth data signal (Ref.4) and the Bell T-4M 274 Mbit/s digital transmission system over coaxial cables (Ref.5).

3.3 Partial-Response or Correlative-Level Coding

In this transmission technique, intersymbol interference is allowed in precisely prescribed integral amounts and at precisely determined time instants, as opposed to Nyquist's ideal non-intersymbol interference transmission (see, for example, Ref.8). The transmission channel does not respond fully within one symbol interval at the sampling instants, but only responds partially. That is each sampled impulse response extends over more than one interval thus causing the amplitude levels to be correlated in the transmitted signal. Partial-response or correlative-level coding systems are also called multiple-response coding systems (Ref.9). For a given number of input data levels, the number of transmitted signal levels is increased, leading to a higher required signal-to-noise ratio (SNR) for a given error rate[1]. On the other hand, partial-response (PR) coding techniques provide an expedient means of shaping the signal power spectral density into a convenient format for transmission. In some circumstances this advantage may outweigh the disadvantage of the increased number of levels.

A block diagram of the PR transmitter is illustrated in Fig.3. The PR encoder transfer function H(D) can be expressed as :

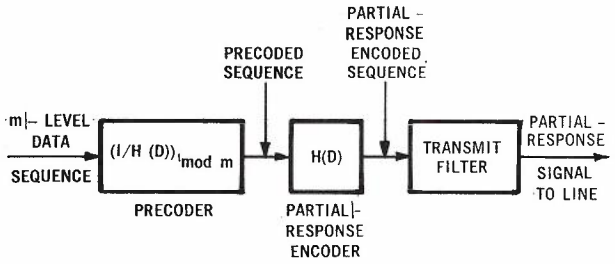


Fig.3 - Block diagram of a partial response transmitting system.

$$H(D) = \sum_{k=0}^{N-1} h_k D^k \qquad (2)$$

where the coefficients h_k are integers which may be positive or negative, h_o and the number of input data levels, m, are relatively prime, and D is the unit-delay operator. By choosing an appropriate function for H(D), nulls can be produced in the signal spectral density function at zero frequency and at integral multiples of one-half the line frequency (or Nyquist frequency). In passing it should be noted that all the commonly used PR codes have only two non-zero h_k terms.

---

[1] This penalty in SNR however may be recovered by making use of the inherent redundancy which exists in the PR signal (Refs. 10-11).

Since the levels in the transmitted PR signal are correlated, a decoding error at the receiver can yield additional subsequent errors. To avoid this error propagation phenomenon, the input data sequence is converted into another sequence by the precoder whose transfer function P(D) is the inverse over modulo m of the PR encoder transfer function, viz.:

$$P(D) = 1/H(D) \quad \text{mod } m \qquad (3)$$

It can be shown that only a simple modulo-m operation is required to recover the original data sequence. (Ref.10).

In the following paragraphs, we shall look at some commonly used line codes belonging to this class of PR signalling.

Bipolar or Alternate-Mark-Inversion (AMI) Code

This widely used method of line coding is usually described as follows:

In the bipolar encoded sequence a binary input "zero" is uncoded and is represented by a space or absence of a pulse. On the other hand, a binary input "one" is represented by a mark, or pulse which alternates in polarity.

This code can also be described in terms of PR coding with binary input data. The PR coding becomes (Fig.4):

$$H(D) = 1 - D \qquad (4)$$

$$\text{and} \quad P(D) = 1/(1-D) \quad \text{mod } 2 \qquad (5)$$

This code has found numerous applications, notably in junction cable primary level (1.544 - 2.048 Mbit/s) PCM systems (Ref.12). It possesses several characteristics that are desirable in baseband pulse transmission. It has no DC component and contains only a small amount of low-frequency (LF) components (Fig.5). Timing information can be easily recovered from the received line signal by simple full-wave rectification followed by a narrow bandpass filter or

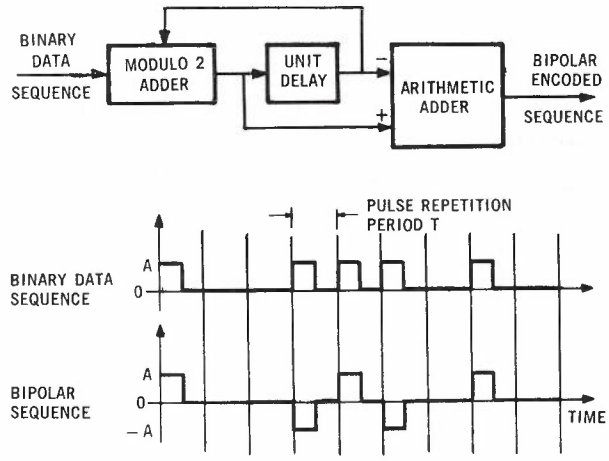


Fig.4 - A bipolar or AMI encoder and its associated waveforms (with half-width rectangular pulses).

a phase-locked loop. Errors in the received signal can also be detected and monitored without reference to the transmitted information using the pulse polarity alternation rule. Of course, errors that occur successively in even numbers and with alternating polarities cannot be detected. A forerunner of the bipolar code is the twinned binary code or dicode which is in fact the non-precoded version.

As mentioned earlier in Section 2, it is convenient to separate the coding aspects from the pulse shaping of the line signal. For random binary input data, both the AMI and twinned binary codes yield the same normalized power spectral density characteristics (Fig.5) with nulls at zero frequency and at multiples of the signalling frequency, namely,

$$W_N(f) = \sin^2 \pi fT \qquad (6)$$

where T is the pulse repetition period. For a given transmitted pulse shape, the actual power spectral density is then given by:

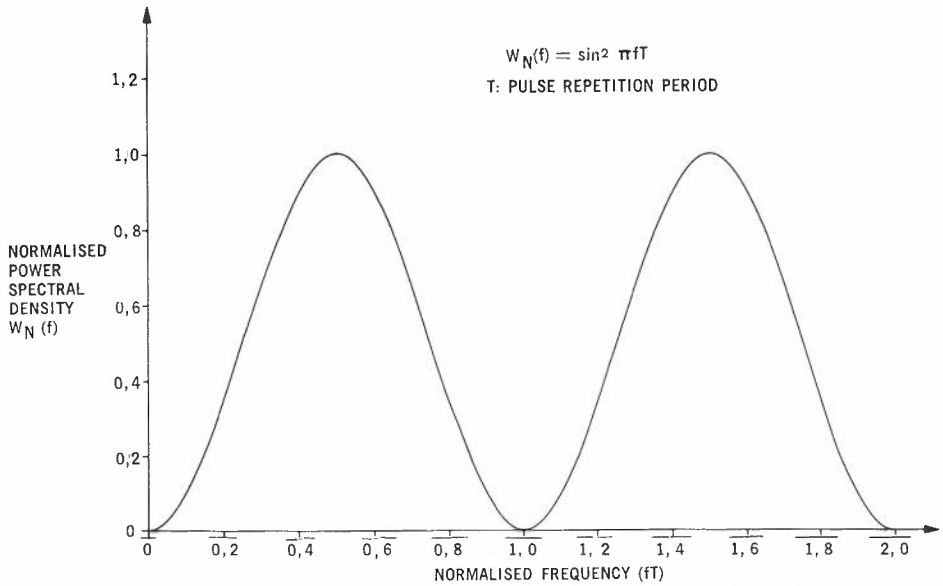


Fig.5 - Normalized power spectral density of random bipolar sequence (negative frequency component not shown)

$$W(f) = |\, G(f)\,|^2 .\ W_N(f)/T \qquad (7)$$

where G(f) is the Fourier transform of the transmitted pulse shape and $W_N(f)$ is the normalized spectral density. For the half-width rectangular pulse, the power spectral density of the random AMI line signal is shown in Fig.6.

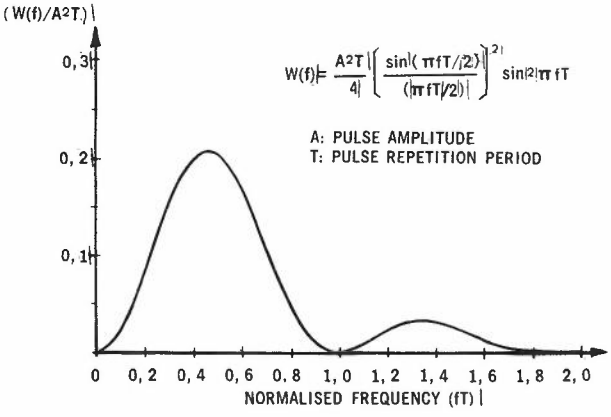


Fig.6 - Power spectral density of a random bipolar sequence with half-width rectangular pulses (negative frequency component not shown).

Multilevel or higher order bipolar codes can be similarly generated using the same expression for H(D) = 1-D, and for P(D) = 1/(1-D) the latter being operated over modulo m where m is the number of information levels of the input data.

Duobinary Code

This code devised by Lender (Ref.13) is obtained by using the following relations for H(D) and P(D) in Fig.3 :

$$H(D) = 1 + D \qquad (8)$$

and $$P(D) = 1/(1+D) \bmod m \qquad (9)$$

The presence of (1+D) in the partial-response expression H(D) yields nulls at odd multiples of the Nyquist frequency (one half the signalling frequency) in the spectral density characteristic of the encoded sequence (Fig.7(b)). The application of duobinary coding is more or less restricted by the high values of the signal spectrum at low frequencies, although it is used in frequency-modulated systems where this disadvantage is not relevant.

Modified Duobinary or PR Class 4 Code

In this code, independently introduced by Lender (Refs. 14,15), and van Gerwen (Refs. 16, 17)[2], the expressions for H(D) and P(D) are respectively :

$$H(D) = 1 - D^2 \qquad (10)$$

and $$P(D) = 1/(1 - D^2) \bmod m \qquad (11)$$

[2] The classification of this code as PR class 4 was made by Kretzmer (Ref.7).

Equation (10) can be factorized into

$$H(D) = (1 - D)\,(1 + D), \qquad (12)$$

implying that nulls exist at zero frequency and at integral multiples of the Nyquist frequency. The presence of these nulls is particularly useful as it allows the insertion of pilot-tones and the application of single-sideband (SSB) transmission in carrier systems (Refs. 17,18).

A comparison of the power spectral density functions of random data sequences encoded according to the previous three PR schemes is illustrated in Fig.7.

Other Partial-Response Coding Schemes

As discussed previously, the PR encoder transfer function H(D) can be designed to yield certain desirable characteristics. A variety of schemes have been proposed (see, for example, Refs. 7,19). With multilevel input data, partial-response signalling can be possible only if the multilevel data can be precoded. A condition which ensures that the above requirement is satisfied has been derived by Gerrish and Howson (Ref.20). This is previously stated in the definition of the PR encoder transfer function (cf. eq.2), namely, the coefficient h_o and the number of input levels m must be relatively prime.

The previous PR systems are based on delays equal to multiples of the symbol interval T. However it is also possible to use different values of delay. When the delay is halved, and therefore equal to T/2, the duobinary code degenerates into what is known as the biternary code (Ref.21). Correspondingly, the null at the Nyquist frequency in the duobinary case is now shifted to twice this value, ie. to the signalling frequency.

Generalized Partial-Response Coding

In the majority of data transmission schemes where the signal has been distorted channel, it is usual to equalize the signal at the receiver. However this may lead to noise enhancement and pre-equalization at the transmitter has been proposed to overcome this effect. But the pre-equalization of the signal may given rise to very large peaks in the signal level which is highly undesirable from the implementation point of view. Generalized partial-response coding (Refs. 22,23) has been proposed to overcome this problem but at the penalty of an increased number of levels at the receiver.

In this generalized technique analogue intersymbol interference (as opposed to integral amounts in the ordinary PR case) is deliberately introduced before the signal is transmitted. In other words, the coefficients h_k in Eq.(2) need not be all integers, and the precoder is now operating over a field of real numbers.

3.4 Two-Phase (or Frequency) Modulation Coding

In this class of line coding, transmission redundancy is introduced by using a signalling rate equal to twice the input binary data rate.

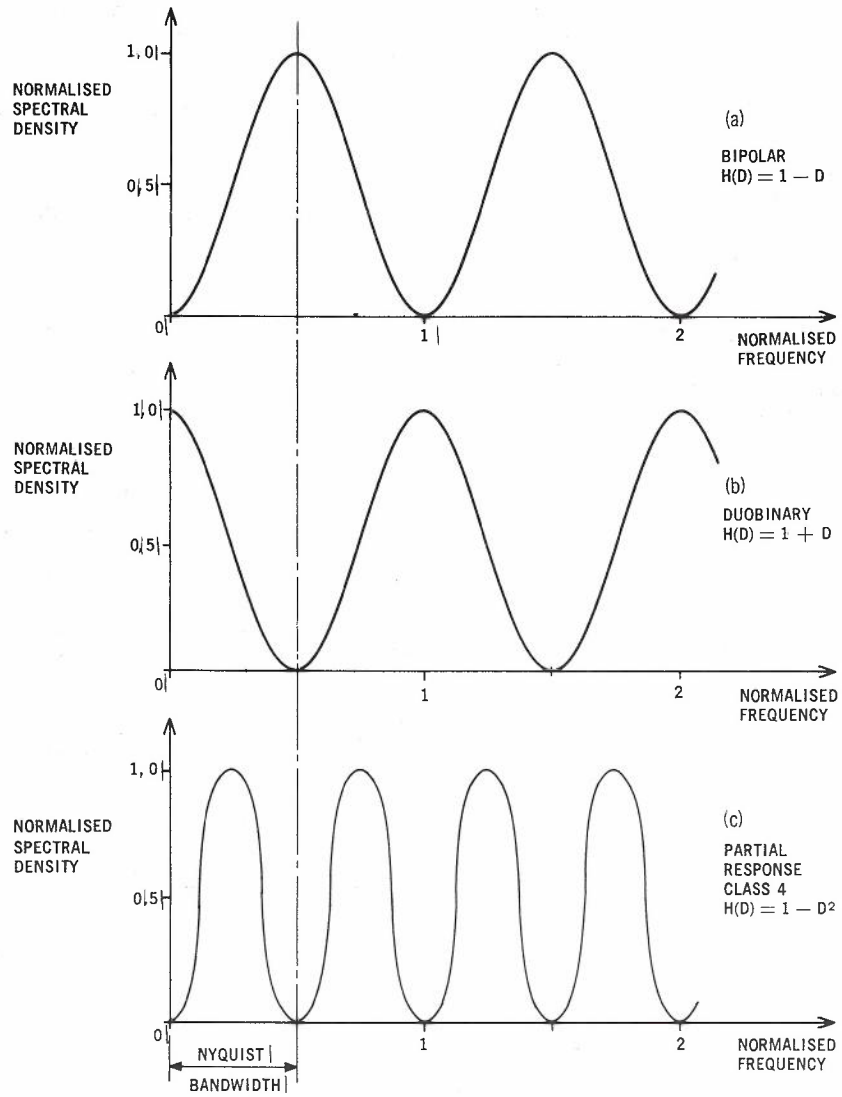


Fig.7 - Comparison of normalized spectral densities of random data encoded according to bipolar, duobinary and partial response class 4 schemes.

The line signal is DC free and contains a large number of transitions from which timing information can be recovered.

Diphase

In its simplest form, a two-phase modulation line signal is derived by allocating respectively two complementary phases to represent the two states in the binary data to be transmitted. The two basic elements of diphase are shown in Fig.8(a); the resulting line signal is shown in Fig.8(b) and is known as diphase or biphase-L (Ref.24). Obviously the allocation of the basic elements could be reversed.

On the other hand, if the binary data is first precoded according to the structure given in Fig.1 and then transmitted as above, precoded diphase is obtained (Fig.8(b)). The latter signal is also known as biphase - M(Ref.24), or frequency - doubling coding (Ref.25), or conditioned diphase (Ref.26) or two-level AMI Class II coding (Ref.27). A common alternative description of precoded diphase is that there is always a signal transition in the middle of each symbol period and an additional transition at the beginning where there is a binary "zero" in the original data sequence.

Both diphase and precoded diphase, sometimes collectively known as Manchester codes (Ref.24), can also be interpreted as Walsh-Type 1 (WAL-1) modulations of a NRZ binary sequence and of its precoded version, respectively. Since precoding

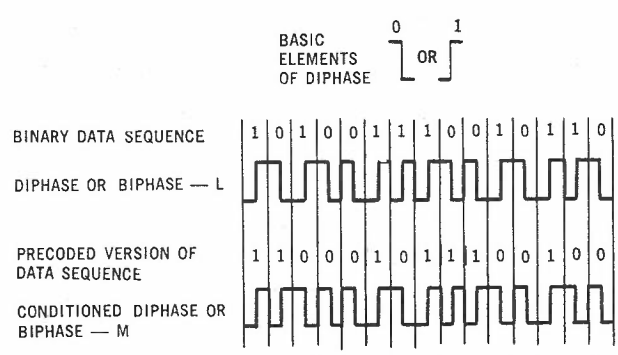


Fig.8 - Examples of diphase and conditioned diphase line codes.

does not alter the statistics of a random binary data (Section 3.1) the spectral density function is the same for both diphase and precoded diphase. Using the techniques described in Ref.2 (Chapter 19) the two-sided spectral density can be shown to be :

$$W(f) = A^2T \left[\frac{\sin(\pi fT/2)}{(\pi fT/2)}\right]^2 \sin^2(\pi fT/2) \quad (13)$$

where A is one half of the peak-to-peak amplitude and T the pulse repetition period.

In summary, diphase and conditioned diphase can be described in terms of transitions, or as a form of modulation, viz. either phase reversal modulation or binary frequency shift modulation or amplitude modulation (suppressed carrier) with the carrier frequency equal to the data rate.

90° Carrier Diphase

In some applications, it is advantageous to shift the transitions in the original diphase by 90° (Ref.28). The resultant line signal can be then interpreted as Walsh - Type 2 (WAL-2) modulation of a NRZ binary sequence. The two-sided spectral density function when transmitting a random binary data sequence is then given by :

$$W(f) = 4A^2T \left[\frac{\sin(\pi fT/2)}{(\pi fT/2)}\right]^2 \sin^4(\pi fT/4) \quad (14)$$

where A and T are as previously defined. From the plot in Fig.9, of the two spectral densities of diphase (WAL-1 modulation) and 90° carrier diphase (WAL-2 modulation) it can be seen that the line signal energy of the latter code is concentrated at higher frequencies than the former. As a consequence, the latter code has a self-equalizing property in cable transmission when coherent detection is used, ie. when the WAL-2 diphase line signal is regarded as a double sideband amplitude modulation signal with suppressed carrier (DSBAM-SC). (Ref.28).

3.5 Modified Two-Phase Modulation Coding

Modified versions of the previous two-phase modulation coding techniques have also been proposed. Basically, they are linear combinations of diphase, partial-response coding and a timing component, and they may or may not involve an increase in transmitted levels. The waveforms of the schemes described below are shown in Fig.10.

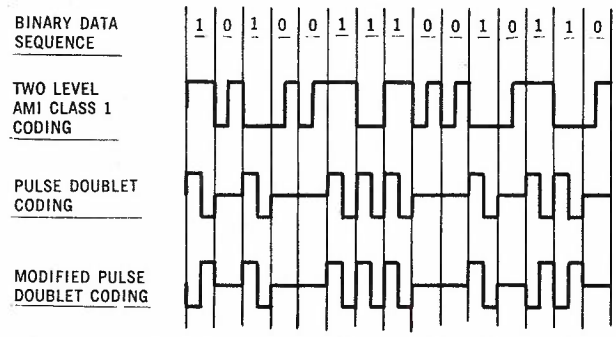


Fig.10 - Waveforms of some modified two-phase modulation schemes.

Two-Level AMI Class I Coding

This line signal together with the two-level AMI Class II signal has been proposed for use in digital fibre optic communication system (Ref.27),

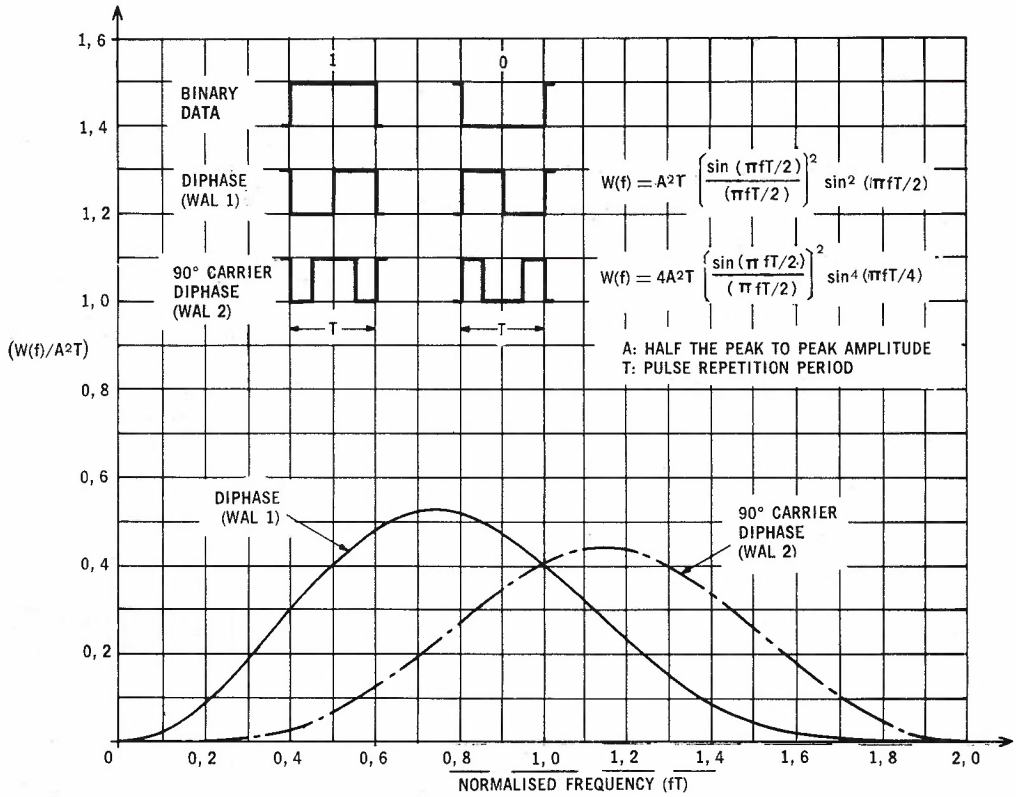


Fig.9 - Power spectral densities of diphase and 90° carrier diphase (negative frequency component not shown).

which encounter transmission requirements similar to those in cable systems. As mentioned in Section 3.4, the two-level AMI Class II is identical to precoded or conditioned diphase. As for the two-level AMI Class 1 signal, it is in fact a linear combination of three components, namely, a bipolar or AMI full-width signal, a diphase signal and a timing component at the binary data rate. Its power spectral density for random data can be readily derived from Figs. 6 and 9.

Pulse Doublet Coding

This line signal was first proposed by Marko et al. (Ref.29) for application in high-speed digital coaxial cable systems (of the order of 280 Mbit/s and above). The encoding procedure consists of transmitting a pair of contiguous half-width pulses of opposite polarities for every binary "one" in the data sequence, while binary "zeros" are not encoded. From this procedure, the signal can be readily decomposed into a diphase signal and a timing component at the binary data rate. Its power spectral density can be thus obtained from Fig.9. Since "zeros" are not encoded and "ones" encoded into pulse doublets of opposite polarities, the signal is basically a two-level scheme and presents several advantages in system implementation (Ref.29).

A slightly modified pulse doublet coding scheme (Ref.30) is proposed for application in data above voice (DAV) transmission over symmetric pair cables. Binary "ones" in the data input are encoded as pulse doublets and "zeros" are not encoded. However, the polarity of the doublets is controlled by both the alternate-mark-inversion rule and the even-odd time slot rule. A simpler explanation of the above encoding rule is as follows : the encoded signal can be regarded as the result of a double-sideband modulation with suppressed carrier of an AMI signal. The carrier frequency is half the bit rate and its phase is shifted by half a symbol period relative to the AMI signal. Its power spectral density can be thus readily derived from the AMI spectrum. Note that in Ref.30, a half-width AMI signal is used and the line signal therefore consists of quarter-width pulses.

Intertran Pulse Position Coding

This line signal used in the digital loop equipment of the Canadian Dataroute network (Ref.31) is basically a linear combination of a 90° carrier diphase signal and a timing component at twice the binary data rate. It thus has the self-equalizing property described in Section 3.4 if synchronous detection is used and its spectral density can be easily obtained from Fig.9. Since the discrete timing component is located at twice the data rate which is a null in the spectral density of 90° carrier diphase, timing information can be directly recovered from the received line signal.

4. NON-LINEAR CODES

As their name indicates, the encoded sequences produced by non-linear codes cannot be completely derived from the input data by means of a linear relationship. Non-linear codes can be divided into two classes, namely, alphabetic and non-alphabetic. Again, as in the case of linear coding, redundancy in the line signal is in the form of either increase in the number of transmitted levels or increase in signalling rate or a combination of these two.

4.1 Alphabetic Codes

An alphabetic code is defined as one in which x K-level digits are converted into words of y L-level digits using a translation table. Such a code is conveniently denoted as an xK-yL code, although this notation is not always adopted. The ratio of signalling or line rate to the input data rate is then equal to :-

$$\frac{B_L}{R} = \frac{y}{x} \qquad (15)$$

For the conversion to be possible, the following relationship must be satisfied:

$$K^x \leq L^y$$

or

$$x \leq y \log_K L \qquad (16)$$

Since the encoding procedure of an alphabetic code requires the conversion of input data into individual blocks or words, these have to be properly framed or synchronized at the receiving terminal before the information can be correctly recovered, even if no error had occurred during transmission. Synchronization is normally achieved by monitoring the various inbuilt properties of the code, such as sum of transmitted digits (or accumulated disparity or running digital sum), permissible patterns, state transitions, etc.

(a) Binary Alphabetic Codes

There exists a variety of codes belonging to this category.

Zero-Disparity Code

The valid codewords in this scheme are 2n-bit words which must contain n "ones" and n "zeros". To keep the redundancy to a low level, n must be made large.

Unit-Disparity Code (Ref.32)

In this method, a two-state ("one" and "zero") line signal is transmitted but the only valid codewords are digit groups in which the number of "ones" differs by only 1 from the number of "zeros" used.

Neu Codes (Ref.37)

In this technique, binary data is converted into another binary sequence via binary - ternary - binary translation. In other words, the binary data is first converted into a ternary stream using an appropriate conversion procedure (eg. 3 binary - into - 2 ternary, 3B-2T). The resultant sequence is then translated back into binary using one of the rules given in Table 1.

TABLE 1 - Neu Code Translation

Ternary Symbol	A-Code	B-Code
+1	01	10
0	00	00 and 11 alternating
-1	10	01
	(11 not used)	

The B-code has the following properties. On the average, the number of "ones" is equal to the number of "zeros" and the line signal can be said to be balanced. The maximum number of consecutive "ones" or "zeros" is equal to 4. Error monitoring can also be carried out using the alternating rule in the translation. The previous three properties have been attained through an increase in signalling rate. The ratio of the latter rate to the input data rate is 4/3 using 3B-2T as the intermediate binary-to-ternary conversion.

Harvey Code (Ref.38)

The generation of this code is similar to the previous one. An intermediate binary-to-ternary conversion is used and the ternary sequence is translated back into binary using the rule given in Table 2.

TABLE 2 - Harvey Code Translation

Ternary Symbol	Binary Word
+1	1000
0	1100
-1	1110

The start of each codeword is uniquely described by a zero-to-one transition, and each code word can be identified by the location of the one-to-zero transition. These properties have been achieved through a large amount of transmission redundancy. Using a 3B-2T code as the intermediate binary-to-ternary conversion, the ratio of signalling rate to input data rate is 8/3.

2B-3B

This class of coding is where 2 binary digits are converted into 3 binary digits, according to a translation table. A particular example has been proposed for use in digital communications over optical fibres (Ref.56). This code is suited to this application as it is two-level and hence minimizes the effect of the laser diode transmitter non-linearity. It also provides error monitoring, no AGC requirement and a "dc-constrained" signal.

(b) Multilevel Alphabetic Codes

A large number of these codes has been proposed, mainly for application in high-capacity cable PCM systems. Apart from a few recent proposals these codes have been described in a previous paper (Ref.1). In the present work only a brief description is given, and the interested reader is directed to Ref.1 for further details.

Multilevel Zero-Disparity Code

One example of a multilevel equivalent of the binary zero disparity code (see subsection 4.1(a)) can be formed as follows. Assuming that the number of levels L is odd, then the code is generated by adding to a word of n digits, an (n+1)th digit whose magnitude is equal to the digital sum of the previous n digits but of opposite polarity. A valid codeword is obtained when the digital sum of the n-bit word does not exceed (L-1)/2. Codes with even L are slightly more difficult to generate because of the restriction on the amplitude of the added digit. (Ref.54).

Paired-Selected Ternary (PST) Code (Ref.39)

The PST code translation is as shown in Table 3. It consists of two alphabets which alternate each time 01 or 10 is detected in the binary input stream. Note that the ternary pairs ++, -- and 00 are not used, and they can therefore be used to provide framing information.

TABLE 3 - PST Code Translation

Binary Word	Ternary Word		
	Mode A	Mode B	Word Digital Sum
00	-+	-+	0
01	0+	0-	±1
10	+0	-0	±1
11	+-	+-	0

Modes alternate after each 01 or 10.

When the probabilities of occurrence of "ones" and "zeros" differ considerably, a modified version of PST is preferred (Ref.39). Other alphabetic codes without change in signalling rate can be similarly constructed.

4B-3T Codes

There exists a number of proposals which consist of translating blocks of four binary digits into words of three ternary digits.

In the STL[3] version (Ref.40), six four-digit words are converted directly into balanced or zero-disparity 3-digit ternary words. The remaining ten are translated into ternary words of ±1, ±2, and ±3 digital sums. These ternary words are grouped into two alphabets according to the polarity of their digital sum (Table 4). The two modes are then appropriately applied, thus compensating for any disparity in the digi-

3 STL stands for Standard Telecommunication Laboratories Ltd., a subsidiary of International Telephone and Telegraph Corp.

TABLE 4 - STL 4B-3T Code Translation

Binary Word	Ternary Word		
	Mode A	Mode B	Word Digital Sum
0000	+0-	+0-	0
0001	-+0	-+0	0
0010	0-+	0-+	0
0011	+-0	+-0	0
0100	++0	--0	±2
0101	0++	0--	±2
0110	+0+	-0-	±2
0111	+++	---	±3
1000	++-	--+	±1
1001	-++	+--	±1
1010	+-+	-+-	±1
1011	+00	-00	±1
1100	0+0	0-0	±1
1101	00+	00-	±1
1110	0+-	0+-	0
1111	-0+	-0+	0

Mode A: RDS = -1, -2, -3
Mode B: RDS = 0, 1, 2
RDS: Running Digital Sum at End of Ternary Word

tal running sum. Modified versions of this 4B-3T code have also been proposed (Ref.41).

In the Franaszek MS43 code (Ref.42), the translation is derived systematically from a set of permissible states by keeping the running digital sum below a certain value. The ternary words are grouped into three alphabets corresponding to the running digital sums at the end of the words. (Table 5)

TABLE 5 - MS43 Code Translation

Binary Word	Ternary Word		
	Mode A	Mode B	Mode C
0000	+++	-+-	-+-
0001	++0	00-	00-
0010	+0+	0-0	0-0
0100	0++	-00	-00
1000	+-+	+-+	---
0011	0-+	0-+	0-+
0101	-0+	-+0	-0+
1001	00+	00+	--0
1010	0+0	0+0	-0-
1100	+00	+00	0--
0110	-+0	-+0	-+0
1110	+-0	+-0	+-0
1101	+0-	+0-	+0-
1011	0+-	0+-	0+-
0111	-++	-++	--+
1111	++-	+--	+--

Mode A: RDS = 1
Mode B: RDS = 2, 3
Mode C: RDS = 4
RDS: Running Digital Sum at End of Ternary Word

A modified version of MS43, called "FOMOT" (Four Mode Ternary) has also been proposed (Ref.43). It has four modes of operation instead of three.

L742 Code (Ref.44)

This code has been constructed using the technique devised by Franaszek (Ref.24). In this scheme blocks of four binary digits are translated into words of two seven-level symbols grouped into the three alphabets.

VL43 Code (Ref.42)

This an example of systematically constructed codes, but the length of the line codewords may be varied. However, for practical reasons the lengths are taken as multiples of a chosen value. In the VL43 codes, blocks of 4 or 8 binary data digits are converted into words of 3 or 6 ternary symbols grouped into five alphabets. The variable-length characteristic allows a decrease in the range over which the running digital sum can vary. This range is also known as digital sum variation or DSV.

3B-2Q Code (Ref.45)

In this code blocks of three digits are converted into words of two four-level symbols. The latter are grouped into two alphabets and the modes alternate when the first two input data bits are both "zeros" or both "ones".

6B-3Q1 Code (Ref.46)

In this code blocks of six binary input digits are translated into words of three five-level symbols grouped into two alphabets. Mode alternation is controlled by the running digital sum at the end of the quinary word.

Other Alphabetic Codes

For a given number of levels in the required line signal, a nost of alphabetic codes can be readily constructed provided that the condition in Eq.(16) is satisfied. However, sufficient redundancy must be allowed in order to fulfil the requirements for baseband transmission (Section 2). Furthermore, the number of transmitted levels is usually restricted because of implementation difficulties. Ternary line codes still attract great interest, and attempts are made to derive three-level alphabetic codes with efficiency greater than that of the 4B-3T types. One proposal introduces a 10B-7T (10 Binary-into-7 Ternary) code by increasing the number of digits in the word translation (Ref.47). In another proposal 6B-4T (6 Binary-into-4 Ternary) code, the digital sum variation (DSV) is left unbounded; however if the input data sequence is random, for example if a scrambler is used then probabilistic bounds can be placed on the DSV (Ref.48).

4.2 Non-Alphabetic Codes

(a) Binary Non-Alphabetic Codes

Polarity Pulse Coding (Ref.34) or Carter Code (Ref.33)

Polarity pulse coding takes an n-bit data block and transmits it either unchanged or with

the bits inverted, as an (n+1)-bit code word so as to reduce the running digital sum since the beginning of transmission. The extra digit, which serves as polarity digit, is to indicate whether or not the original data word has been inverted. Similar but multilevel techniques known as feedback balanced block coding (Ref.35) are described in a subsequent section.

Miller Code (Ref.49) or Delay Modulation (Ref.50)

This code is a variation of the diphase or biphase-L described earlier in Section 3.4. It is derived by deleting every second transition in the diphase signal. Another description of the encoding procedure consists of the following rules:

A transition from one level to the other is placed at the midpoint of the bit period when the binary data contains a "one". No transition is used for a "zero" unless it is followed by another "zero", in which case the transition is placed at the end of the bit period of the first "zero".

Fig.11 illustrates the relationship between diphase and delay modulation. It can also be shown (Ref.51) that delay modulation has its power concentrated at a low frequency (just below half the bit rate), making it particularly useful when bandwidth availability is small.

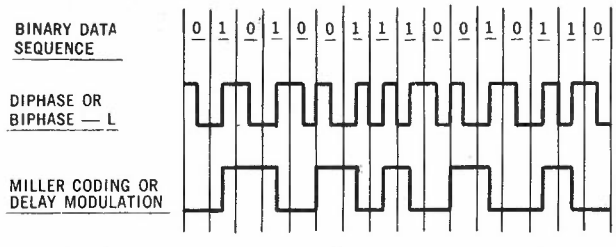


Fig.11 - Waveforms of diphase and delay modulation.

(b) Multilevel Non-Alphabetic Codes

Time Polarity Control Coding (Ref.12)

In this transmission scheme time slots are labelled alternately positive and negative. Binary "ones" in the input data occurring in negatively-labelled slots are transmitted with a negative polarity. "Ones" occurring elsewhere and "zeros" are sent out unaltered. However adverse data patterns can produce low-frequency and even DC components.

Filled Bipolar or AMI Codes

These codes are generally derived from the basic bipolar encoding rule (Section 3.3), namely, successive data "ones" alternate in polarity while binary "zeros" in the data are transmitted uncoded. However the long sequences of "zeros" are undesirable as little or no timing information can be derived for regeneration purposes. This deficiency can be readily corrected either by breaking the long zero sequences by scrambling the input data or by substituting for the sequences of consecutive "zeros" of a certain length some chosen non-zero patterns. At the decoder these are located and removed.

For clarity we shall adopt Croisier's scheme of naming the polarities of ternary line pulses (Refs. 52,57). Spaces are denoted as "0". Positive and negative pulses generated using the bipolar encoding rules are called "B" pulses. On the other hand, pulses (whether inserted or inverted from the original sequence) that violate this rule are denoted as "V" pulses. In other words, a "V" pulse has the same polarity as that of the preceding pulse.

B6ZS Code

With this code the encoded sequence is obtained from the original bipolar stream upon substituting the following pattern:

$$B0VB0V \qquad (17)$$

for any strings of six consecutive zeros.

Other bipolar with n-zero substitution codes (modal or non-modal) can also be derived (Ref.53).

High-Density Bipolar (HDB) and Compatible High-Density Bipolar (CHDB) Codes (Ref.52)

A HDB codes of order n ($n \geq 2$), denoted as HDBn, is one in which (n+1) consecutive "zeros" are replaced by one of the following sequences:

$$B00...V \qquad (18)$$

$$\text{or } \underbrace{000...V}_{(n+1)\text{ digits}} \qquad (19)$$

and where the remaining "ones" or pulses in the original data sequence alternate in polarity, taking also into account the pulses present in the filling patterns. The choice of these sequences (Eqs. 18,19) is carried out in such a way that the number of B pulses between the successive V pulses is always odd. Successive V pulses are therefore alternating in polarity and consequently the digital sum variation (DSV) is only two.

A variation of the substitution patterns leads to substantial advantages. In this case the filling sequences are:

$$00...B0V \qquad (20)$$

$$\text{or } \underbrace{00...00V}_{(n+1)\text{ digits}} \qquad (21)$$

and the codes are called compatible HDBn or simply CHDBn. They are so named because a universal CHDB decoder can be used, irrespective of the order n.

Noting the slight differences in the definitions of n in BnZs, HDBn and CHDBn, it can be seen that B3ZS, HDB2 and CHDB2 codes are identical. Another version of filled bipolar called transparent interleaved bipolar (TIB) code has also been proposed (Ref.52).

Feedback Balanced Codes

There are two types of feedback balanced codes, one involving redundancy in the number of symbols or digits and the other in the number of amplitude levels.

The first scheme is the multilevel version of the binary polarity pulse coding (see subsection 4.2(a) and Ref.34). A polarity pulse usually of magnitude equal to the maximum amplitude of the multilevel signal is used to indicate whether the data word has been inverted. (Ref.35). The polarity pulse may also be referred to as a pilot pulse and can be used to advantage for other purposes e.g. adaptive equalizer adjustment. (Ref.36).

The second scheme involves an increase in the number of possible levels for each digit or symbol. This redundancy is then used to minimize the low frequency components by controlling the polarity of some of the output pulses by negative feedback (Ref.55). This multilevel signal cannot be classified with the partial response codes as the generating algorithm is essentially non-linear[4].

5. CONCLUDING REMARKS

In this survey we have attempted to describe various line coding techniques proposed for applicable in baseband digital transmission. It was not practical to incorporate in the survey all the codes discussed in the literature; rather the aim has been to present the large variety of methods in a unified fashion so that any other proposals may be readily classified and understood.

Comparison of the various properties of the line coding techniques described in this paper has been minimal. Further details may be found in the references cited. It is worth mentioning that some line codes may also be used in digital magnetic recording systems where recording requirements are similar to those for transmission (Section 2).

REFERENCES

1. Duc, N.Q., "Line Coding Techniques for Baseband Digital Transmission", Austral.Telecommun. Res., May 1975, Vol.9, pp.3-17.

2. Bennett, W.R. and Davey, J.R., Data Transmission, New York, N.Y. : McGraw-Hill, 1965.

3. Bennett, W.R., "Synthesis of Active Networks", Proc. Polytech. Inst. Brooklyn Symp. Series, Vol.5, Modern Network Synthesis, April 1955, pp. 45-61.

4. "Data Transmission at 48 kbit/s Using 60-to 108-kHz Group Band Circuits", Recommen. V.35, CCITT Green Book, Data Transmission, Vol.8, ITU, Geneva 1973, pp. 143-151.

5. Waldhauer, F.D., "Quantized Feedback in an Experimental 280-Mb/s Digital Repeater for Coaxial Transmission", IEEE Trans. Commun., Jan. 1974, Vol. COM-22, pp. 1-5.

6. Lender, A., "Correlative Digital Communication Techniques", IEEE Trans. Commun. Tech., Dec. 1964, Vol. COM-12, pp. 128-135.

7. Kretzmer, E.R., "Generalization of a Technique for Binary Data Communication", IEEE Trans. Commun. Tech., Feb. 1966, Vol. COM-14, pp. 67-68.

8. Lucky, R.W., Salz, J. and Weldon, E.J., Jr., Principles of Data Communication, New York, N.Y.: McGraw-Hill, 1968.

9. Ridout, P.N. and Ridout, I.B., "Adaptive Multiple Response Reception and Waveform Correction for Data Transmission via Channels in the FDM Network", 1974 Intl. Zurich Seminar on Digital Commun., Zurich, Switzerland, March 12-15, 1974, Proceedings, pp. H6(1)-H6(7).

10. Kobayashi, H., "Correlative Level Coding and Maximum - Likelihood Decoding", IEEE Trans. Inform. Theory, Sept. 1971, Vol. IT-17, pp. 586-594.

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

12. Aaron, M.R., "PCM Transmission in the Exchange Plant", Bell Syst. Tech. J., Jan. 1962, Vol.41, pp. 99-141.

13. Lender, A., "The Duobinary Technique for High-Speed Data Transmission", IEEE Trans. Comm Electronics, May 1963, Vol.82, pp. 214-218.

14. Lender, A., "Correlative Level Coding for Binary-Data Transmission", IEEE Spectrum, Feb. 1966, Vol.3, pp. 104-115.

15. Lender, A., "Correlative Data Transmission with Coherent Recovery Using Absolute Reference' IEEE Trans. Commun. Tech., Feb. 1968, Vol. COM-16, pp. 108-115.

16. van Gerwen, P.J., "On the Generation and Application of Pseudo-Ternary Codes in Pulse Transmission", Philips Res. Rept., August 1965, Vol.20, pp. 469-484.

17. van Gerwen, P.J., "Efficient Use of Pseudo-Ternary Codes for Data Transmission", IEEE Trans. Commun. Tech., August 1967, Vol. COM-15, pp. 658-660.

18. Becker, F.K., Kretzmer, E.R. and Sheehan, J.R., "A New Signal Format for Efficient Data Transmission", Bell Syst. Tech. J., May-June 1966, Vol.45, pp. 755-758.

19. Kabal, P. and Pasupathy, S., "Partial-Response Signalling", IEEE Trans. Commun., Sept. 1975, Vol. COM-23, pp. 921-934.

20. Gerrish, A.M. and Howson, R.D., "Multilevel Partial-Response Signalling", 1976 IEEE Intl. Conf. on Commun., Minneapolis, Minn., June 12-14, 1967, Digest, p. 186.

21. Brogle, A.P., "A New Transmission Method for Pulse-Code Modulation Communication Systems" IRE Trans. Commun. Syst., Sept. 1960, Vol. CS-8, pp. 155-160.

22. Tomlinson, M., "New Automatic Equalizer Employing Modulo Arithmetic", Electron. Letters, March 1971, Vol.7, pp. 138-139.

4 In Ref.1 this code was mistakenly classified as a line scheme.

23. Harashim, H. and Miyakawa, H., "Matched-Transmission Technique for Channels with Intersymbol Interference", IEEE Trans. Commun., August 1972, Vol. COM-20, pp. 774-780.

24. Spitzer, C.F., "Digital Magnetic Recording of Wideband Analog Signals", Computer Design, Oct. 1973, pp. 83-90.

25. Knoll, A.L., "Spectrum Analysis of Digital Magnetic Recording Waveforms", IEEE Trans. Electronic Computers, Dec. 1967, Vol. EC-16, pp. 732-743.

26. Hanna, J.L., Kirk, B.C. and Moore, E.J., "Digital Data Loops Using Diphase Transmission", IEEE Intl. Conf. on Commun., Philadelphia, Pa., June 19-21, 1972, Record, pp. 13/12 - 13/15.

27. Takasaki, Y. and Tanaka, M., "Line Coding Plan for Fibre Optic Communication Systems", Proc. IEEE, July 1975, Vol.63, pp. 1081-1082.

28. Boulter, R.A., "A 60 kbit/s Data Modem for Use over Physical Pairs", 1974 Intl. Zurich Seminar on Digital Commun., Zurich, Switz. March 12-15, 1974, Proceedings, pp. H3(1) - H3(6).

29. Marko, H., Weisz, R. and Binkert, G., "A Digital Hybrid Transmission System for 280 Mbit/s and 500 Mbit/s", IEEE Trans. Commun., Feb. 1975, Vol. COM-23, pp. 274-282.

30. Pospischil, R., "2Mb/s Long-Haul Data Above Voice Transmission Over Cable", 1975 IEEE Intl. Conf. on Commun., San Francisco, Calif., June 16-18, 1975, Record, pp. 31/12 - 31/16.

31. Delorenzi, A. and Meyer, A., "The Dataroute Digital Loop Equipment", 1974 IEEE Intl. Conf. on Commun., Minneapolis, Minn., June 17-19, 1974, Record, pp. 2C/1 - 2C/5.

32. Cattermole, K.W. "Low-Disparity Codes and Coding for PCM", IEE Conf. on Transmission Aspects of Commun. Networks, London, Feb. 1964.

33. Carter, R.O., "Low-Disparity Binary Coding System", Electron. Letters, May 1965, Vol.1, pp. 67-69.

34. Franklin, J.N. and Pierce, J.R., "Spectra and Efficiency of Binary Codes Without DC", IEEE Trans. Commun., Dec. 1972, Vol. COM-20, pp. 1182-1184.

35. Ogawa, K., Wakahara, T., Endo, T. and Watanabe, K., "Design and Transmission Characteristics of 1.544 Mb/s PCM-FDM Converter", Rev. of Elect. Commun. Labs., Nov-Dec. 1974, Vol.22, pp. 963-970.

36. Hagiwara, S., Hinoshita, S. and Kawashima, M., "PCM-FDM : System Capability and Performance Improvement on Waveform Equalization and Synchronization", IEEE Trans. on Communications, August 1974, Vol. COM-22, pp. 1149-1154.

37. Neu, W. and Kundig, A., "Project for a Digital Telephone Network", IEEE Trans. Commun. Tech. Oct. 1968, Vol. COM-16, pp. 633-647.

38. Harvey, J.T., "Symbol Set for Regenerated Communications", Electron. Letters, August 1973, Vol.9, pp. 384-385.

39. Sipress, J.M., "A New Class of Selected Ternary Pulse Transmission Plans for Digital Transmission Lines", IEEE Trans. Commun. Tech. Sept. 1965, Vol.13, pp. 366-372.

40. Jessop, A., Norman, P. and Waters, D.B., "120 Mbit/s Coaxial System Demonstrated in the Laboratory", Elect. Commun., April 1973, Vol.48, nos.1 and 2, pp. 79-92.

41. Cattermole, K.W., Ellis, R.R. and Wallace, P.W., "A 4B-3T Code with Improved Error Properties", Dept. of Elect. Eng. Sc., Uni. of Essex, Colchester, Telecommun. Syst. Group, Report 77.

42. Franaszek, P.A., "Sequence - State Coding for Digital Transmission", Bell Syst. Tech. J., Jan. 1968, Vol.47, pp. 143-157.

43. Buchner, J.B., "Ternary Line Signals", 1974 Intl. Zurich Seminar on Digital Commun., Zurich, Switz., March 12-15, 1974, Proceedings, pp. F1(1) - F1(9).

44. Benedetto, S., Castellani, V. and De Vincentiis, G., "On a Class of Alphabetic Non-Linear Multilevel Codes for Synchronous Data Transmission", 1973 IEEE Intl. Conf. on Commun., Seattle, Wash. June 1973.

45. Ingram, D.G.W., Bylanski, P. and Dorward, R.M., "Some Factors in Design of High Speed Digital Transmission Systems for Coaxial Cables", 1972 Intl. Zurich Seminar on Digital Comm., Zurich, Switz., March 1972, Proceedings, pp. D5(1) - D5(5).

46. Aratani, T. and Fukinuki, H., "800 Mb/s PCM Multilevel Transmission System over Coaxial Cables", Rev. of Elect. Commun. Lab., Sept-Oct. 1970, Vol.18, pp. 740-753.

47. Cattermole, K.W., "Codes for Digital Line Transmission", Electron. Letters, Oct. 1975, Vol.11, pp. 548-550.

48. Catchpole, R.J., "Efficient Ternary Transmission Codes", Electron. Letters, Oct. 1975, Vol.11, pp. 482-484.

49. Miller, U.S. Patent 3, 108, 261, Oct. 22, 1963.

50. Jacobi, G., U.S. Patent 3,414, 894, Dec. 3, 1968.

51. Hecht, M. and Guida, A., "Delay Modulation", Proc. IEEE, July 1969, Vol.57, pp. 1314-1316; see also Correction in Proc. IEEE, Jan. 1970, Vol.58, p. 182.

52. Croisier, A., "Introduction to Pseudo-ternary Transmission Codes", IBM J. Res. Dev., July 1970, Vol.14, pp. 354-367.

53. Johannes, V.I., Kaim, A.G. and Walzman, T., "Bipolar Pulse Transmission with Zero Extraction" IEEE Trans. Commun. Tech., April 1969, Vol. COM-17, pp. 303-310.

54. Nakagome, Y., Amano, K. and Ota, C., "A Multilevel Code having Limited Digital Sum", Electron. Commun. in Japan, Jan. 1968, Vol.51-A, pp. 37-44.

55. Kaneko, H. and Sawai, A., "Feedback Balanced Code for Multilevel PCM Transmission", IEEE Trans. Commun. Tech., Oct. 1969, Vol. COM-17, pp. 554-563.

56. Y. Takasaki, M. Tanaka, N. Maeda, K. Yamashita andK. Nagano, "Optical Pulse Formats for Fibre Optic Digital Communications", IEEE Trans. Commun., April 1976, Vol. COM-24, pp. 404-413.

57. J.B. Buchner, "Ternary Line Codes", Philips Telecommunication Review, Vol.34, No.2, June 1976, pp. 72-85.

BIOGRAPHY

NGUYEN QUANG DUC was born in Gia Dinh, Vietnam, on June 23, 1947. He came to Australia in November 1964 on a four-year Colombo Plan Scholarship sponsored by the Australian Government. He received the B.E. (Honours) and the Ph.D. degrees in Electrical Engineering from the University of Queensland, Brisbane in 1969 and 1973, respectively. His doctorate project, which was financially supported by the Australian Radio Research Board, dealt with Majority Logic Decoding of Error-Correcting Codes. From 1969 to 1972 he was a tutor in Electrical Engineering and a Part-time Tutor in Pure Mathematics at the University of Queensland. He then joined the Line and Data Systems Section of the Australian Post Office Research Laboratories, Melbourne. His research effort was initially concentrated on the evaluation of line coding techniques, transmission capacity and error control for high-speed digital coaxial cable systems. He was later engaged in the feasibility study of PCM transmission over the Australian rural cable network. Presently, his work is mainly concerned with data communications over the FDM telephone network. Outside the Section he contributes as a consultant on coding techniques and communication theory. From June 1976 to May 1977 he was "Visiting Associate Protessor" at the School of Computer and Information Science, Syracuse University, U.S.A. Dr. Duc is a member of the Institution of Radio and Electronic Engineers (Australia) and the Institute of Electrical and Electronics Engineers (U.S.A.).

BERNARD SMITH was born in Mt. Gambier, South Australia and studied Electrical Engineering at the University of Adelaide where he obtained the degrees of Bachelor of Engineering (Honours) in 1964 and Doctor of Philosophy in 1969. He joined the Research Laboratories of Telecom Australia in 1968 where he is currently an Engineer Class 4 (Acting) in the Transmission Systems Branch. Whilst working in the Research Laboratories, he has been concerned with various problems in line transmission systems and in particular digital data transmission at all speeds in the telephone network.

The Design of Equalisers for Class 4 Partial Response Data Signals.

R. P. COUTTS
B. M. SMITH
Telecom Australia Research Laboratories

High speed data transmission over bandlimited channels such as the FDM telephony network usually requires group delay equalisation to reduce the effects of intersymbol interference. This paper describes an efficient method of calculation of the mean square sampled error for SSBAM (or baseband) signals. This method is then used for the design of a simple equaliser for a Class 4 partial response system to be used for a supermastergroup channel modem. Lastly, the effect of such equalisation for this example is used to compare various criteria of distortion often used in data transmission.

1. INTRODUCTION

The transmission of digital data signals over bandlimited channels such as the Frequency Division Multiplexed (FDM) telephony network is a well established practice; for example, channels ranging from voice up to the supermastergroup (equivalent to 900 voice channels) have been exploited for this purpose. The various channel filters produce a rising group delay distortion at the bandedges which can cause significant amount of intersymbol interference in the received pulses especially at the higher data rates.

To minimise this interference several approaches can be adopted. For example, the traditional approach has been to use a fixed group delay equaliser, often of quite high order, to reduce the group delay distortion below a somewhat arbitrarily chosen bound in the frequency domain. A more valid approach recognises that only the intersymbol interference at the sample instants is relevant to the data receiver's performance. For example, a fixed equaliser could be designed to minimise this sampled interference or more ideally one of several differing types of automatically adaptive equalisers could be used.

When these channels are considered in more detail, one of their notable characteristics is that there is no obviously "best" choice for the modulation technique; rather, there are various possibilities which have similar performance but with varying areas and degrees of difficulty. One technique which offers a relative simplicity of implementation and which has been widely used for data transmission over group and wider bandwidth FDM channels, is the single side-band amplitude modulated (SSBAM) signal with Class 4 partial response (PR) coding. Ref.4

In this paper, we propose to discuss a method of designing a fixed equaliser to minimise the intersymbol interference at the sample instants for the SSBAM Class 4 PR signal. The requirement for this design method arose during the development of a data modem for operation over the supermastergroup FDM channel and where there was insufficient resources to develop an adaptive equaliser for this application.

One measure of the intersymbol interference or distortion at the sampling instants is the Mean Square Error (MSE) and this criterion is used to design the equaliser. This criterion has been used previously to determine the performance of a SSBAM Class 4 PR signal to quadratic group delay distortion. Ref.3 In this paper we expand the method presented by Mazo Ref.5 to give an efficient method of calculation of the MSE and which allows a simple automatic design of the optimum equaliser with a given structure.

An example is then given to illustrate the method and finally a comparison is made with regard to other measures of performance: namely peak distortion and average error probability (for a given signal to noise ratio) in order to assess the usefulness of the MSE criterion.

2. CALCULATION OF MEAN SQUARE ERROR

Fig.(1) illustrates the model of the ideal and actual sampled transmission system where the sampled error sequence is the difference between the two output samples.

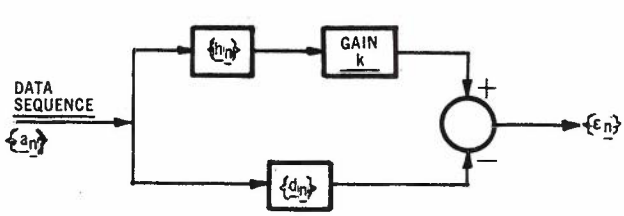


Fig. 1. Model of sampled data error.

In this figure:

$\{a_n\}$ is the (p+1) level input sequence with possible values $\{\pm1,\pm3,\ldots,\pm p\}$

$\{h_n\}$ is the baseband sampled impulse response of the total system

$\{d_n\}$ is the desired response for no distortion.

An interesting feature of this model in Fig. 1 is the gain term k which can be optimised to minimise the mean square error (MSE) and can be interpreted as a single tap linear equaliser.

The error sequence $\{\varepsilon_n\}$ is given by the convolution

$$\{\varepsilon_n\} = \{a_n\} * \{kh_n - d_n\} \tag{1}$$

Thus

$$\varepsilon_n = \sum_{i=-\infty}^{\infty} a_i(kh_{n-i} - d_{n-i}) \tag{2}$$

where for a Class 4 PR signal d_i is given by:

$$d_i = \begin{cases} +1 & i = -1 \\ -1 & i = +1 \\ 0 & i \neq -1,+1 \end{cases} \tag{3}$$

The mean square error (MSE) which is denoted $\langle\varepsilon^2\rangle$ is given by:

$$\langle\varepsilon^2\rangle = E\{\varepsilon_n^2\}$$

$$= E\{\sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} a_i a_j (kh_{n-i} - d_{n-i})(kh_{n-j} - d_{n-j})\}$$

$$= \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} E\{a_i a_j\}(kh_{n-i} - d_{n-i})(kh_{n-j} - d_{n-j}) \tag{4}$$

But if $\{a_n\}$ is an equilikely uncorrelated data sequence, then

$$E\{a_i a_j\} = \delta_{ij} \tag{5}$$

and equation (4) becomes

$$\langle\varepsilon^2\rangle = E\{a_i^2\}(k^2 \sum_{i=-\infty}^{\infty} h_i^2 - 2k(h_{-1} - h_{+1}) + 2) \tag{6}$$

The optimum gain factor k to minimise this equation is determined from

$$\frac{\partial\langle\varepsilon^2\rangle}{\partial k} = 0 \tag{7}$$

which leads to

$$k_{opt} = \frac{(h_{-1} - h_{+1})}{\sum_{i=-\infty}^{\infty} h_i^2} \tag{8}$$

Thus substituting for k in equation (6) and knowing Ref.1 that

$$E\{a_i^2\} = \frac{p(p+2)}{3} \tag{9}$$

The optimum MSE is given by

$$\langle\varepsilon^2\rangle = \frac{p(p+2)}{3}\left[2 - \frac{(h_{-1} - h_{+1})^2}{\sum_{i=-\infty}^{\infty} h_i^2}\right] \tag{10}$$

In order to evaluate equation (10) one must evaluate the total sampled pulse response energy and the term containing the two dominant pulse terms.

2.1 Sampled Pulse Energy

In this paper, we develop a method for the calculation of the total sampled pulse response energy for a SSBAM system which is a simple function of timing and carrier phase offset. The method is based on considering the energy in the sampled frequency domain which is termed the "Equivalent Nyquist Spectrum" Ref.1 and can be interpreted as the resultant spectrum due to "aliasing" because of under-sampling. The energy up to the Nyquist frequency is the energy in the sampled time pulse response.

Specifically, using Parseval's Theorem and the Sampling Theorem the sampled response energy is given by:

$$\sum_{i=-\infty}^{\infty} h_i^2 = \frac{1}{2\pi}\int_{-\pi/T}^{\pi/T} |H_{eq}(\omega)|^2 d\omega \tag{11}$$

where

T is the baud rate

π/T is the Nyquist frequency (rad/s)

$H_{eq}(\omega)$ is the "Equivalent Nyquist Spectrum"

In Appendix 1 the frequency response of the channel is resolved into odd and even components about the Nyquist frequency in order to derive $H_{eq}(\omega)$ where this equivalent baseband response is a function of carrier phase and timing offset. Thus assuming there is no energy outside twice the Nyquist frequency (i.e. only one level of frequency overlap), the sampled pulse response energy is given by (Appendix 1).

$$\sum_{i=-\infty}^{\infty} h_i^2 = \frac{4}{T}\{A + B\cos(\frac{2\pi\zeta}{T} - 2\phi) - C\sin(\frac{2\pi\zeta}{T} - 2\phi)\} \tag{12}$$

where

A, B and C are constants dependent on the channel characteristics (see Appendix 1).

τ is the timing offset from the optimum value for a distortionless channel.

ϕ is the receiver carrier phase offset.

This expression for the sampled energy is valid for any SSBAM system and will be used in equation (10) for the case of Class 4 PR signals

2.2 Dominant Pulse Term

For a carrier system, the pulse response is dependent on the carrier phase offset and is given by:

$$h(t) = x(t)\cos\phi + \tilde{x}(t)\sin\phi \qquad (13)$$

where

$x(t)$ is the pulse response of the channel

$\tilde{x}(t)$ is the quadrature pulse response which in the case of SSBAM is the Hilbert transform of $x(t)$ denoted $\hat{x}(t)$

For the case of Class 4 PR signals, one is interested in the term $(h(\tau-T) - h(\tau+T))$ which has been written $(h_{-1} - h_{+1})$.

Thus the dominant pulse term from (13) is given by

$$(h_{-1} - h_{+1})^2 = (a\cos\phi + b\sin\phi)^2 \qquad (14)$$

where $a = (x_{-1} - x_{+1})$

$b = (\tilde{x}_{-1} - \tilde{x}_{+1})$

for the case of Class 4 PR signals.
In equation (12) if B, C < < A which will be the case for such signal shaping, then the total sampled pulse energy is only very weakly dependent on carrier offset ϕ and so the main effect of such offset will be on the dominant pulse term. Thus to find the optimum carrier phase offset

$$\frac{\partial \langle \varepsilon^2 \rangle}{\partial \phi} = 0 \qquad (15)$$

which leads to

$$\frac{\partial (h_{-1} - h_{+1})^2}{\partial \phi} = 0 \qquad (16)$$

so that from substitution of equation (14) ϕ is given by:

$$\phi_{opt} = \tfrac{1}{2} \tan^{-1} \left\{ \frac{2ab}{a^2 - b^2} \right\} \qquad (17)$$

and substitution into equation (14) gives

$$(h_{-1} - h_{+1})^2 = a^2 + b^2 \qquad (18)$$

Thus given a specific timing offset, the optimum phase to minimise the MSE for a SSBAM or VSBAM system is given from (17) and the MSE contribution by (18).

In this particular paper, we are considering SSBAM Class 4 PR signals and the MSE is given by

$$\langle \varepsilon^2 \rangle = \frac{p(p+2)}{3} \left[2 - \frac{(h_{-1} - h_{+1})^2}{\Sigma h_i^2} \right] \qquad (19)$$

where

$$\Sigma h_i^2 = \frac{4}{T} \left\{ A + B\cos\left(\frac{2\pi\tau}{T} - 2\phi_{opt}\right) - C\sin\left(\frac{2\pi\tau}{T} - 2\phi_{opt}\right) \right\} \qquad (20)$$

$$\phi_{opt} = \tfrac{1}{2} \tan^{-1} \left\{ \frac{2ab}{a^2 - b^2} \right\} \qquad (21)$$

and

$$(h_{-1} - h_{+1})^2 = a^2 + b^2 \qquad (22)$$

This value of MSE is dependent on the timing offset τ which can then be optimised numerically.

3. EXAMPLE EQUALISER DESIGN

In the previous section a method has been presented for the calculation of MSE for carrier data transmission systems using SSBAM Class 4 PR signals. This section describes the application of this method to the design of a simple fixed equaliser for a data modem to operate over the FDM supermastergroup (SMG) channel.

3.1 Equivalent Channel Response

In order to compute the amount of linear distortion and thence to design an equaliser, the attenuation and group delay of all the various filters in the system including the actual transmission channel were measured. These responses are then converted to a normalised, equivalent low-pass form and stored sequentially on a computer file along with the resultant product of these responses called the "Equivalent Channel Response".

By normalised it is meant that all bandwidths are normalised to a symbol rate of 1 Baud and thus the Nyquist bandwidth is 0.5Hz in contrast to 2.9013MHz as actually used in the system operating in the supermastergroup channel Secondly, all bandpass filters are transformed to equivalent low pass form as discussed in Franks Ref.2 . However, for the case considered here where a true SSBAM signal is used, this equivalent low pass response is simply the frequency shifted version of the bandpass frequency response. In the case of a negative frequency shift as occurs in the example treated in this paper, the low-pass equivalent response is the complex conjugate of the negative frequency bandpass response. This effect results in an inversion of the frequency response (i.e. low frequencies become high frequencies) upon transformation.

For the SMG system in question, the normalised frequency response of the channel (not including the PR waveshaping) is shown in Fig. 2

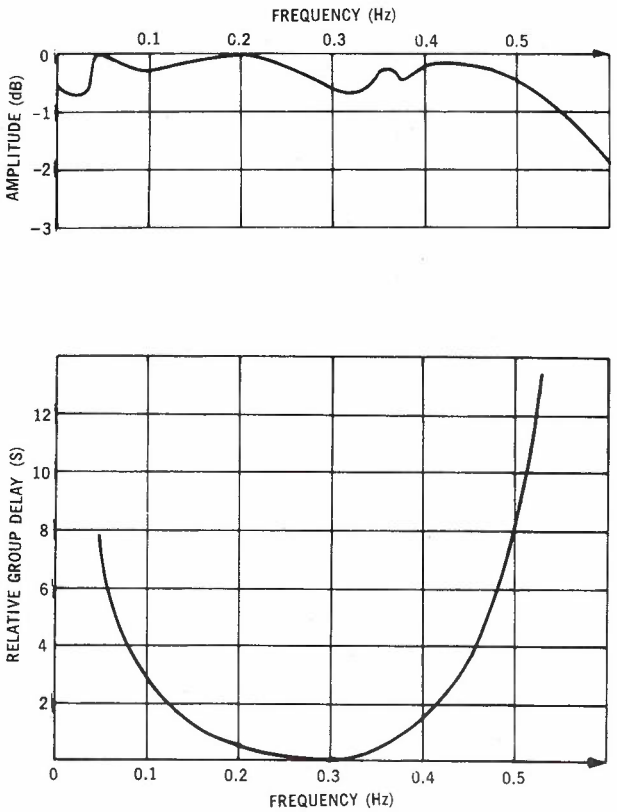


Fig. 2. Equivalent normalised low-pass channel response.

From this response it can be seen that group delay distortion is the dominant source of channel impairment and is primarily due to the SMG channel itself.

In order to equalise this system, it was decided to use several simple equaliser sections in cascade where the section parameters could be optimised numerically using the sampled mean square error as derived in section 2.

3.2 Equaliser Structure

To equalise the group delay response shown in Fig. 2 , an equaliser using several cascaded sections would be used before demodulation at the receiver. The equaliser section suitable is illustrated in Fig. 3 with corresponding group delay response characteristics.

The group delay characteristic illustrated in Fig. 3 is characterised by two parameters α and β and for infinite Q components is given by: [Ref.6].

$$\tau(\omega) = \frac{2\alpha}{\alpha^2+(\omega-\beta)^2} + \frac{2\alpha}{\alpha^2+(\omega+\beta)^2} \qquad (23)$$

which is assumed to be sufficiently accurate for this example.

To equalise the SMG channel, several such sections are required and Fig. 4 shows the

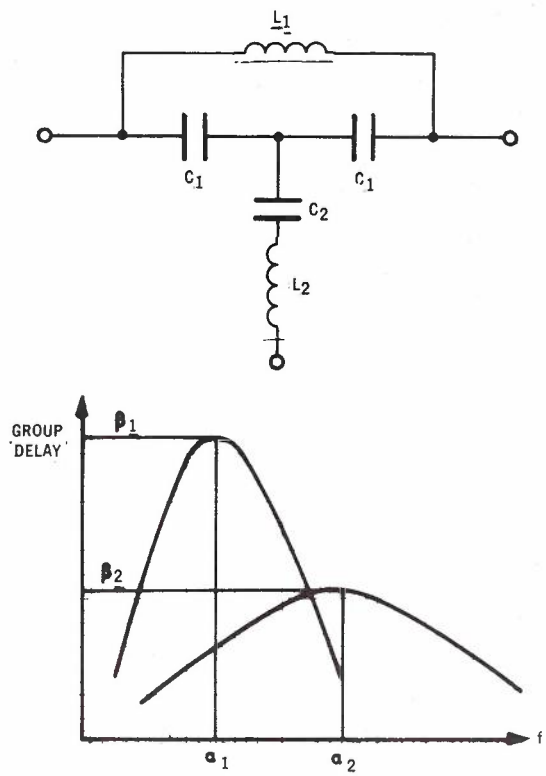


Fig. 3. Equaliser section structure.

decrease in the minimum RMS error for increasing number of equaliser sections.

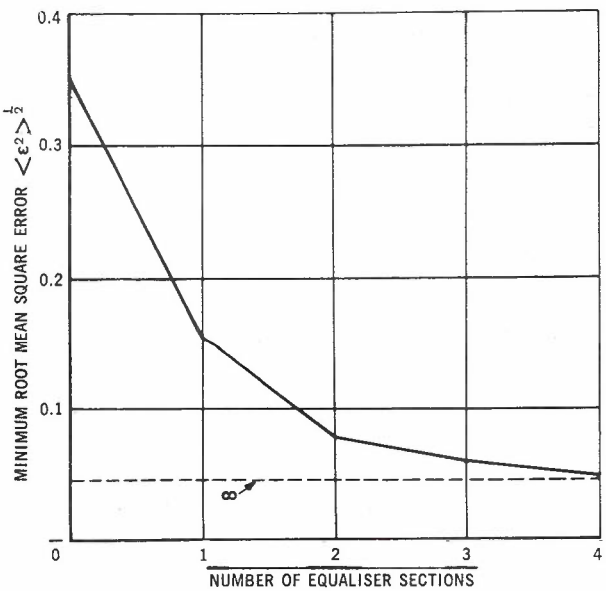


Fig. 4. Variation of minimum $\langle\varepsilon^2\rangle^{\frac{1}{2}}$ with the number of equaliser sections for a 3 level class 4 PR signal.

However, there appears to be little point in using more than three equalizer sections in this example. Thus most of the group delay distortion which effects the sampled mean square error has been removed by three simple delay sections.

The effect of such an equaliser on the actual group delay of the channel is illustrated

in Fig. 5 for one and three section equalisers. From these results it can be seen that the equalizer attempts to minimise the delay distortion in the central 60% of the signal spectrum (see Fig. 5) and outside this band, the actual group delay distortion is quite high.

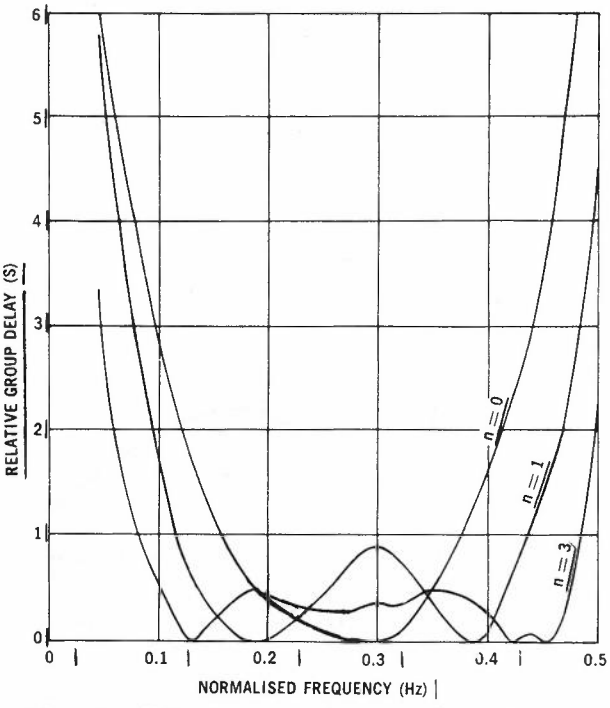


Fig. 5. Effect of equaliser with various number of sections on relative group delay of SMG Channel.

4. COMPARISON OF DESIGN CRITERIA

In this equalizer design example, the MSE criterion was used. However, it is useful to compare this criterion with peak distortion, another often used criterion, and then examine how these relate to average error probability which is perhaps the most desirable measure of the system performance. This comparison is done using the example of the supermastergroup channel with various amounts of equalization.

4.1 Peak Distortion

Firstly, one can define peak distortion for Class 4 PR signals as the fraction of the eye pattern which is closed. This criterion which can be estimated from an eyepattern display on an oscilloscope as illustrated in Fig. 6 where o is the eyeopening and e is the distance between the centres of the distortion regions.

Thus the peak distortion r is given by

$$r = 1 - o/e \tag{24}$$

which for a 3 level signal is given by [Ref.4]

$$r = \frac{2}{(h_{-1}-h_{+1})} \left\{ \sum_{i=-p}^{q}{}'' |h_i| + \frac{|h_{-1} + h_{+1}|}{2} \right\} \tag{25}$$

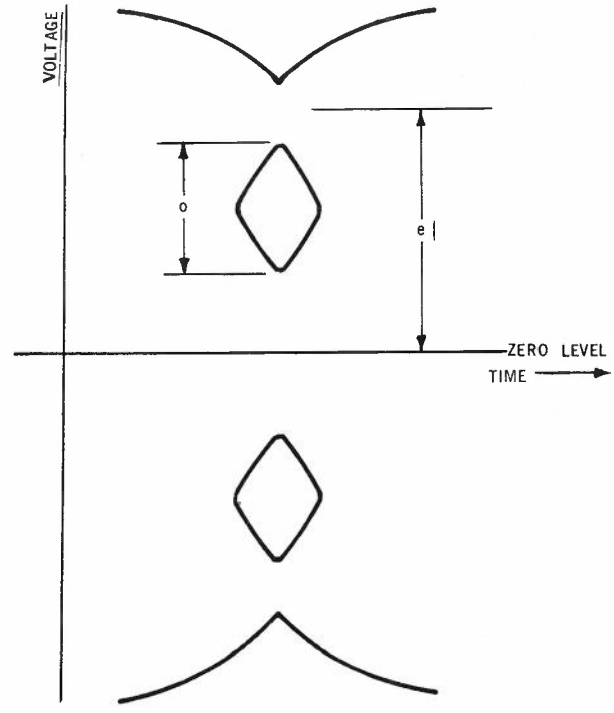


Fig. 6. Eye pattern for a 3 level class 4 PR signal.

where Σ'' means a summation with the omission of the terms h_{-1} and h_{+1} and where there are N non-zero terms in this summation so that

$$N = p + q - 1 \tag{26}$$

This expression for peak distortion is normalised so that it is independent of the gain k mentioned in Section 2.

4.2 Average Error Probability

For a Class 4 PR signal with 3 levels, the average error probability is given by:

$$\bar{P}_e = \frac{1}{2^N} \sum_{i=1}^{2^N} \left\{ \frac{1}{2} P_r \left[X_i + n > (h_{-1} - h_{+1} - k') \right] + \frac{1}{2} P_r \left[X_i + n > (k' + (h_{-1} + h_{+1})) \right] + \frac{1}{2} P_r \left[X_i + n > (k' - (h_{-1} + h_{+1})) \right] \right\} \tag{27}$$

where $X_i = \sum_{j=-p}^{q}{}'' a_j h_j$ is the intersymbol interference due to the i th data sequence $\{a_j\}$ (there are 2^n equiprobable possible sequences), k' is the decision threshold and

$$P_r \{X_i + n > \ell\} = Q \left\{ \frac{\ell - X_i}{\sigma} \right\} \tag{28}$$

is the probability of the intersymbol interference plus additive Gaussian noise n exceeding a level ℓ where σ is the RMS noise value. Ref.1. In equation (27), ℓ has three possible values where the first term corresponds to errors in the outer levels and latter terms to errors in the inner level of the eye. In this paper k' will be taken as that setting which gives the minimum MSE and is given by: (Section 2)

$$k' = \frac{1}{k} = \frac{\sum_{i=-p}^{q} h_i^2}{(h_{-1}-h_{+1})} \qquad (29)$$

For the comparison of performance criteria a Signal to Noise Ratio (SNR) of 17dB at the decision point is used and this is not, in general, the SNR at the receiver input. This decision point SNR is defined as:

$$SNR = \frac{\sum_{i=-p}^{q} h_i^2}{\sigma^2} \qquad (30)$$

In this example, N = 62 is taken to be a sufficient number of terms of the impulse response. However, in order to calculate equation (27), a series expansion method Ref 7 was used to minimise the otherwise time consuming direct calculation of this expression.

Table 1 gives the results for the different criteria for different stages of equalisation of the supermastergroup channel

Table 1 - Comparison of Performance Criteria for Different Stages of Equalization of the SMG Channel.

Description	Root Mean Square Error $\varepsilon^{\frac{1}{2}}$	Peak Distortion r	Average Error Probability $\overline{P}_\varepsilon$ for SNR=17dB
Unequalised	0.332	1.15	5.62×10^{-3}
3 Section	0.061	0.293	1.11×10^{-6}
Ideal Phase Equalisation	0.044	0.105	5.09×10^{-7}
No Distortion	0	0	4.15×10^{-7}

From this table it can be seen that a three section equalizer is adequate to equalise the channel with respect to average error probability. Secondly, MSE is seen to be more relevant with respect to average error probability than peak distortion in this application. This is due to the fact that the intersymbol interference is attributable to a large number of terms so that the sequence giving the peak distortion is reasonably improbable.

5. CONCLUSIONS

High speed data transmission over band-limited channels such as the FDM telephony network usually requires group delay equalisation to reduce the effects of intersymbol interference. One measure of such distortion is MSE and this paper describes an efficient method of calculation of this criterion for a SSBAM (or baseband) signal.

In particular, specific formalae are derived for the calculation of MSE for a Class 4 PR signal for the optimum carrier phase offset where the optimum timing is found numerically. This calculation method is then used in the design of a simple three section group delay equalizer for data transmission over a supermastergroup channel. If such a design were carried out using frequency domain criteria, a much more complex equalizer would have been required.

Lastly, using this specific example of the supermastergroup channel, the chosen criterion of MSE is compared with peak distortion and related to the average error probability. In this situation MSE is seen to be a good criterion for equalizer design.

REFERENCES

1. Lucky, P.W., Salz, J. and Weldon, Jr.,E.J., "Principles of Data Communication". McGraw Hill. 1968.

2. Franks, L. "Signal Theory". Prentice Hill. 1969.

3. Smith, B.M. "The Mean Square Error of SSBAM Class 4 Partial Response Data Signals Distorted by Parabolic Group Delay". IEEE Transactions on Communications. Vol.Com.24. No.11, Nov. 1976, pp.1252-1256.

4. Kabal, P. and Pasupathy, S. "Partial Response Signalling". IEEE Transactions on Communications. Vol.Com-23. No.9, Sept. 1975, pp.921-934.

5. Mazo, J.E. "Optimum Timing Phase for an Infinite Equaliser". B.S.T.J. Vol.54. No.1, Jan.1975, pp.189-201

6. Geffe, P.R. "Simplified Modern Filter Design". John F. Rider. 1963

7. Ho, E.Y., and Yeh, Y.S. "A New Approach for Evaluating the Error Probability in the Presence of Intersymbol Interference and Additive Gaussian Noise". B.S.T.J. Vol.49.Nov.1970.pp.2249-2266.

APPENDIX 1

In this appendix, an expression for the sampled pulse energy for a SSBAM system is derived which is a simple function of carrier phase and timing offsets. The basic method is an extension of the work of Mazo Ref. 5 and assumes similarly that the energy spectrum of the baseband pulse is limited to no more than twice the Nyquist frequency π/T rad/s as illustrated in Fig.7.

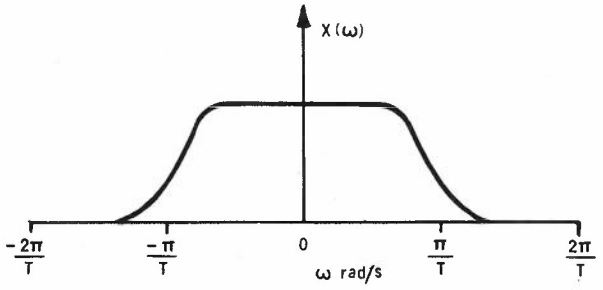


Fig. 7. Bandlimited low pass channel response.

The "Equivalent Nyquist Spectrum" Ref.1 of an equivalent low pass Fourier transform X(ω) of the system is given by

$$X_{eq}(\omega) = \sum_{n=-\infty}^{\infty} X(\omega - n.\frac{2\pi}{T}) \qquad (31)$$

where $|\omega| \leq \pi/T$

$$= \sum_{n=-1}^{1} X(\omega - n.\frac{2\pi}{T}) \qquad (32)$$

$$X(\omega) = 0 \text{ for } |\omega| \geq |\frac{2\pi}{T}|$$

The response X(ω) can be written

$$X(\omega) = R(\omega)e^{j\psi(\omega)} \qquad (33)$$

where R(ω) is the amplitude response and

ψ(ω) is the phase response including the effect of timing offset (i.e. linear phase component).

Following the example of Mazo [Ref.5], the amplitude and phase responses can be resolved into odd and even components about the Nyquist frequency π/T rad/s. Thus using the substitution

$$\mu = \frac{T}{\pi}(\omega - \frac{\pi}{T}) \qquad (34)$$

one can write

$$R(\mu) = R_e(\mu) + R_o(\mu) \qquad (35)$$

$$\psi(\mu) = \psi_e(\mu) + \psi_o(\mu)$$

where

$$\psi_e(\mu) = \psi_e(-\mu)$$

$$\psi_o(\mu) = -\psi_o(-\mu) \qquad (36)$$

$$R_e(\mu) = R_e(-\mu)$$

$$R_o(\mu) = -R_o(-\mu)$$

and for $\omega \in [0, 2\pi/T]$ then $\mu \in [-1, +1]$

Thus considering ω > 0 and substituting for the odd and even components in equation (32) one obtains for the equivalent Nyquist spectrum

$$X_{eq}(\mu) = [R_e + R_o]\, e^{j(\psi_e + \psi_o)} + [R_e - R_o]\, e^{j(-\psi_e + \psi_o)} \qquad (37)$$

where the second term is the contribution due to n = -1 in equation (32).

Expanding equation (37) leads to

$$X_{eq}(\mu) = 2\, e^{j\psi_o}\left[R_e \cos\psi_e + j\, R_o \sin\psi_e\right] \qquad (38)$$

However, for a carrier system one must also consider the effect of carrier phase offset φ, where the equivalent baseband spectrum H(ω) is given by

$$H(\omega) = X(\omega)\cos\phi + \tilde{X}(\omega)\sin\phi \qquad (39)$$

For this analysis, $\tilde{X}(\omega)$ will be taken as the Hilbert transform $\hat{X}(\omega)$ where

$$\hat{X}(\omega) = -j\,\mathrm{sgn}\omega X(\omega) \qquad (40)$$

The equivalent Nyquist spectrum of $\mathrm{sgn}\omega X(\omega)$ is thus given by

$$[\mathrm{sgn}\ \omega\ X(\omega)]_{eq} = [R_e + R_o]e^{j(\psi_e + \psi_o)} + [R_e - R_o]e^{j(-\psi_e + \pi + \psi_o)} \qquad (41)$$

$$= 2e^{j\psi_o}\left[R_o \cos\psi_e + j\, R_e \sin\psi_e\right] \qquad (42)$$

Thus $H_{eq}(\mu)$ is given by

$$H_{eq}(\mu) = X_{eq}(\mu) \cos \phi$$

$$-j\left[X(\omega)\mathrm{sgn}\omega\right]_{eq} \sin\phi \qquad (43)$$

which after substitution of equations (42) and (38) gives

$$H_{eq}(\mu) = 2e^{j\psi_o}\left[R_e \cos(\psi_e - \phi)\right.$$

$$\left. + jR_o \sin(\psi_e - \phi)\right] \qquad (44)$$

Now by using Parseval's and the Sampling Theorems, the sampled pulse energy is given by

$$\sum_{i=-\infty}^{\infty} h_i^2 = \frac{1}{2\pi}\int_{-\pi/T}^{\pi/T} |H_{eq}(\omega)|^2 \, d\omega$$

$$= \frac{1}{T}\int_{-1}^{0} |H_{eq}(\mu)|^2 \, d\mu \qquad (45)$$

If the even phase response can be split as

$$\psi_e = \frac{\pi\tau}{T} + \psi_{ce} \qquad (46)$$

where ψ_{ce} is the even component of the channel phase response

τ is the timing offset.

$$\sum_{i=-\infty}^{\infty} h_i^2 = \frac{4}{T}\{A + B\cos(\frac{2\pi\tau}{T} - 2\phi)$$

$$- C \sin(\frac{2\pi\tau}{T} - 2\phi)\} \qquad (47)$$

where A, B and C are integrals independent of the carrier phase and timing offsets and are given by:

$$A = \int_{-1}^{0} \frac{R_e^2 + R_o^2}{2} \, d\mu \qquad (48)$$

$$B = \int_{-1}^{0} \frac{R_e^2 - R_o^2}{2} \cos 2\psi_{ce} \, d\mu \qquad (49)$$

$$C = \int_{-1}^{0} \frac{R_e^2 - R_o^2}{2} \sin 2\psi_{ce} \, d\mu \qquad (50)$$

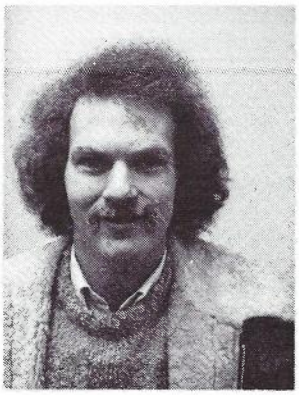
and R_e, R_o and ψ_{ce} are found from the original amplitude and frequency responses. In equations (48) to (50)

$$\frac{R_e^2 + R_o^2}{2} = \tfrac{1}{4}\{R^2(\mu) + R^2(-\mu)\} \qquad (51)$$

$$\frac{R_e^2 - R_o^2}{2} = \tfrac{1}{2} R(\mu) R(-\mu) \qquad (52)$$

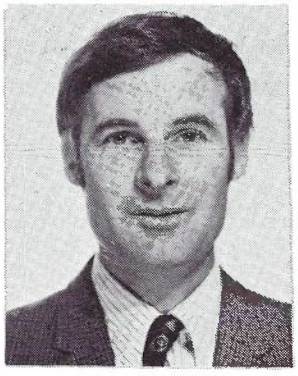
$$\psi_{ce} = \tfrac{1}{2}\{\psi(\mu) + \psi(-\mu)\} \qquad (53)$$

These results are applicable to any SSBAM data transmission system (and baseband systems where ϕ=0). In this paper, however, we are interested in Class 4 PR signals where the amount of spectral energy outside the Nyquist frequency is quite small implying that B and C are very small in comparison with A.



BIOGRAPHY

REGINALD P. COUTTS was born in Adelaide in 1949. He received the B.Sc. and B.E. (Hons.) degrees from the University of Adelaide in 1972 and 1973 respectively. From 1973 to 1976 he worked towards a Ph.D degree at the University of Adelaide in the area of data communications in a fading environment. In 1976, he joined the Line and Data Systems Section of the Telecom Australia Research Laboratories where he worked on high speed data transmission. Early in 1977, he joined the Datel Design and Provisioning Section of the Engineering Department working on Baseband Data Transmission techniques. Dr. Coutts is a member of the Institute of Electrical and Electronics Engineers.



BERNARD SMITH was born in Mt. Gambier, South Australia and studied Electrical Engineering at the University of Adelaide where he obtained the degrees of Bachelor of Engineering (Honours) in 1964 and Doctor of Philosophy in 1969. He joined the Research Laboratories of Telecom Australia in 1968 where he is currently an Engineer Class 4 (Acting) in the Transmission Systems Branch. Whilst working in the Research Laboratories, he has been concerned with various problems in line transmission systems and in particular digital data transmission at all speeds in the telephone network.

A Technique for the Measurement of Random Timing Jitter

J. A. BYLSTRA
Telecom Australia Research Laboratories

A brief review is given of jitter measurement techniques. A technique is presented for measuring jitter of a wide range of jitter frequencies starting from dc. Rms and peak-to-peak measurements are discussed as well as limitations and refinements of the measurement techniques. Finally, a novel jitter calibration method is presented. This method offers: ease of application, wide range and the possibility of high accuracy.

1. INTRODUCTION

Jitter, the unintended displacement of the transitions of a digital signal from their ideal timing positions, is one of the basic impairments in digital transmission systems. It is often required to measure random jitter. This paper describes a method for the measurement of random jitter with jitter frequency components down to dc. Low frequency jitter (close to dc) is often of special interest as this type of jitter is hard to attenuate. High frequency jitter can be attenuated with rather simple means by using a jitter reducer which incorporates an elastic memory and a simple phase-locked loop (Ref.1). On the other hand high frequency jitter, rather than low frequency, is of interest to the operation of digital regenerators as high frequency jitter affects the error performance of a regenerator.

Reference 2 describes a technique which is particularly effective for the evaluation of the timing performance of digital regenerators. The method given in reference 2 is applicable in cases where the jitter to be measured is dependant on the transmitted signal pattern, full control over the transmitted pattern is possible and the jitter to be measured has a simple periodic nature.

The method to be described in this paper is applicable in cases where the jitter to be measured is not pattern dependent and does not have a simple periodic nature, or where control over the transmitted pattern is not possible. For example, the method was used to measure the output jitter of digital multiplexes.

Although in this paper reference is made to measuring jitter on digital signals, the measurement principles discussed apply equally well to measuring inherent jitter on analogue channels, using sinewave test tones, as is of relevance to data transmission using carrier type modems.

A novel jitter calibration method is also described in this paper.

2. JITTER MEASUREMENT TECHNIQUE

2.1 Alternative Methods

A review of phase jitter measurement techniques is given in Refs. 3 and 4.

The simplest jitter measurement technique is to display the jittered signal on an oscilloscope. The reference signal acts as the triggering signal to compare the phase of the jittered signal. If the original signal clock (timing) source is used as a trigger signal, peak-to-peak jitter will be measured down to dc. Calibration is obtained by adjusting the time base for a display of one reference clock period in 10 horizontal display units. This well known method is most suitable for simple periodic jitter. However, for complex waveforms it is often hard to see the extremities of the jitter and this reduces the accuracy of the measurement.

Reference 5 describes a digital jitter measurement technique. This technique may enable a high degree of accuracy. However, such a digital technique requires circuit techniques capable of operation at, dependent on the required accuracy, around 100 times the bit rate of the measurement signal. For signals at about 2Mbit/s or higher, as are common in PCM systems, such a digital technique is difficult to implement (circuit operation required at 200 Mbit/s or higher).

2.2 Technique Proposed

The proposed jitter measurement technique is illustrated in Fig.1. This technique is a slight modification of a technique described in Refs.3 and 4. A latch type phase comparator (also called a sawtooth comparator; (Ref.6)) compares the associated clock signal of the jittered signal to be measured with an, ideally, jitterfree reference clock signal. The frequency of the reference clock signal must equal the average clock frequency of the measured

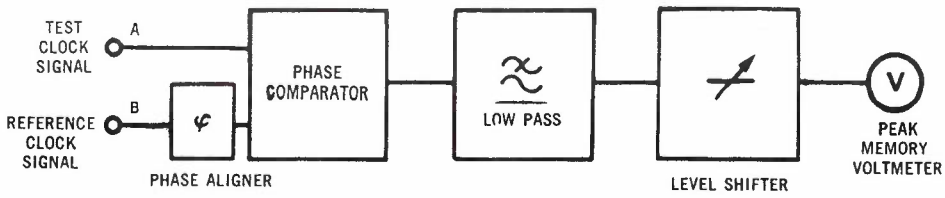


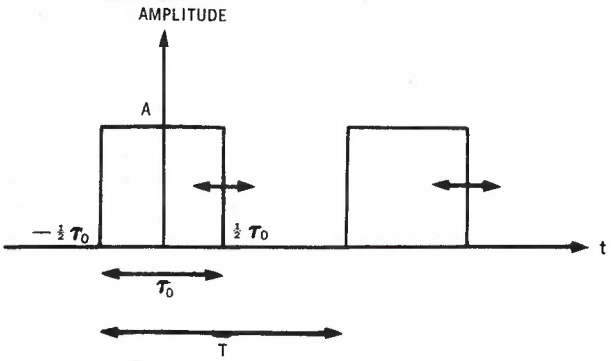
Fig.1 - Jitter Measuring Set.

digital signal. The output of the phase comparator is a pulse width modulated signal. The spectrum of this signal contains the required baseband jitter signal, and at higher frequencies, components at the reference clock frequency and lower and upper sidebands.

Normally, a jittered clock signal is a signal which exhibits unwanted phase or frequency (angular) modulation. This jitter is contained in sidebands around the clock frequency and its harmonics. Basically, the phase comparator is used as a synchronous detector for the jitter signal.

Other forms of phase jitter (see Ref.4), e.g. due to the addition of a signal (e.g. cross-talk) or due to noise, causing the transitions of the digital signal to move, will also be measured using a set in accordance with Fig.1.

A simple causal relation to explain the occurrence of the baseband jitter signal can be obtained by calculating the Fourier series of an unmodulated pulse train. Using the symbols given in Fig.2 we find:



$\omega_r = \frac{2\pi}{T}$ = REFERENCE CLOCK FREQUENCY

Fig.2 - Quasi-Static Pulse Train.

$$f(t) = A\frac{\tau_o}{T} + \frac{2A}{\pi}\sum_{n=1}^{\infty}\frac{1}{n}\sin\left(n\pi\frac{\tau_o}{T}\right)\cos n\,\omega_r\,t \qquad (1)$$

baseband signal | filtered out

where $\omega_r = \frac{2\pi}{T}$

Equation (1) has a component at dc and at the reference clock frequency ω_r. Equation (1) will also be approximately valid for slowly varying width τ, if the variation of τ is slow compared with ω_r (quasi-static approximation).

By filtering out the unwanted frequency components with a low pass filter, the desired baseband jitter signal is obtained. As derived in section 2.3, in most cases the jitter signal may be obtained with very low distortion, by using a low pass filter, if the significant jitter frequency components are below a fourth of the reference clock frequency.

The ability of measuring jitter components down to dc is achieved in two steps. Firstly the original digital signal clock source is taken as the reference clock signal. This is only possible if transmitter and receiver are at the same location and a looped transmission path is used (see also section 2.6). If the reference clock source is jitterfree, absolute jitter will be measured. Otherwise jitter relative to the reference source will be measured.

As a second step, dc coupling is used between the output of the low pass filter and a metering device. To assure the dc level is at a desired level a variable level shifter (see Fig.4) is applied. If peak-to-peak jitter has to be measured, it is only necessary to coarsely adjust the dc level to give both a positive and negative peak reading on a peak reading meter (see Fig.3).

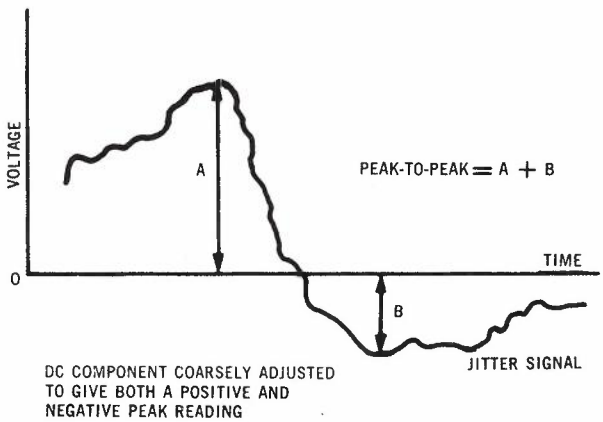


Fig.3 - Method of p-p Measurement.

It is possible to use an alternative to the level shifter. If the jitter signal is quantized using an analogue-to-digital (A-D) convertor, the peak-to-peak jitter can be obtained as the difference between the maximum and minimum quantized signal level, measured over a certain period, using hard wired logic circuitry or an on-line mini or micro computer. This approach is very suitable if the measurements

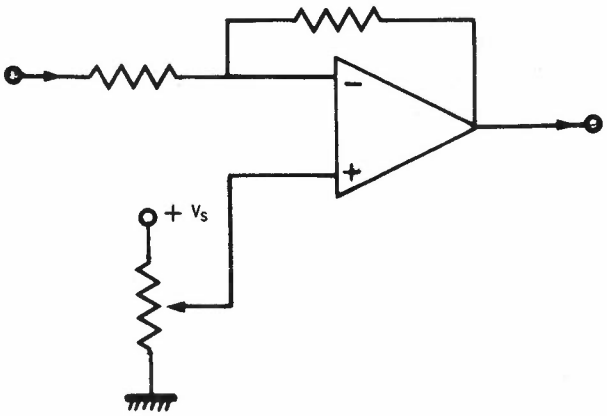


Fig. 4 - Variable Level Shifter.

have to be automated or a simple to operate instrument is required. Using a pulse width-to-pulse amplitude (PW-PA) convertor instead of a low pass filter, as discussed in section 2.4, it is even possible to integrate the PW-PA and the A-D convertor.

type phase comparator. Fig.5 shows the phase characteristic of this phase comparator measured at the output of the low pass filter for input signals having a constant frequency difference. The continuous curve is the ideal phase characteristic. The dotted curve is a practical phase characteristic. The deviation of the ideal curve will be small (e.g. less than 1%) if the logic components of the phase comparator operate fast compared with the signal rates used.

For proper functioning of the jitter measuring set it has to be checked whether the phase comparator is still working in its linear range. An oscilloscope can be used to check whether the transitions of the phase comparator output are not too close in time. Phase shifting, which can be effected simply by inverting the reference clock signal, is used to coarsely align measured and reference signals. A large linear range phase comparator will make the phase alignment less critical. It is possible to automate the phase alignment in a simple way, by using the fact that a latch type phase comparator sharply changes the duty cycle of its output pulse train wherever a non-

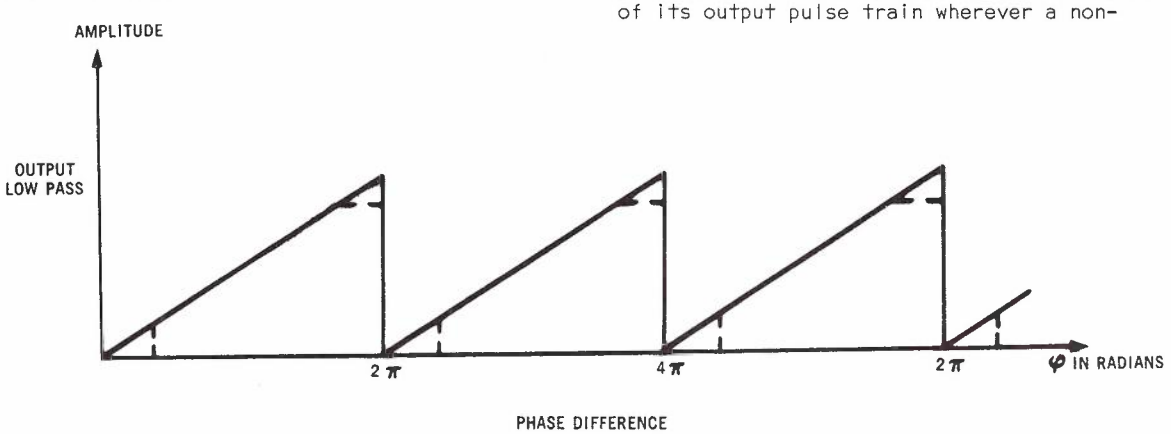


Fig. 5 - Latch Type Phase Comparator; Phase Characteristic.

A latch type phase comparator is preferred as this type has a large linear range of approximately one bit period or 360° maximum. A large linear range will be required if high jitter amplitudes are to be measured. (see also section 3.2). Ref.7 describes possible technical implementations and limitations of a latch

linearity in the phase characteristic occurs. Such an automatic phase alignment will be stable until the jitter frequency and amplitude are so high that no clear difference in output duty cycle due to jitter or due to a non-linearity in the phase characteristic can be observed. A diagram of this automatic phase aligner is shown in Fig.6.

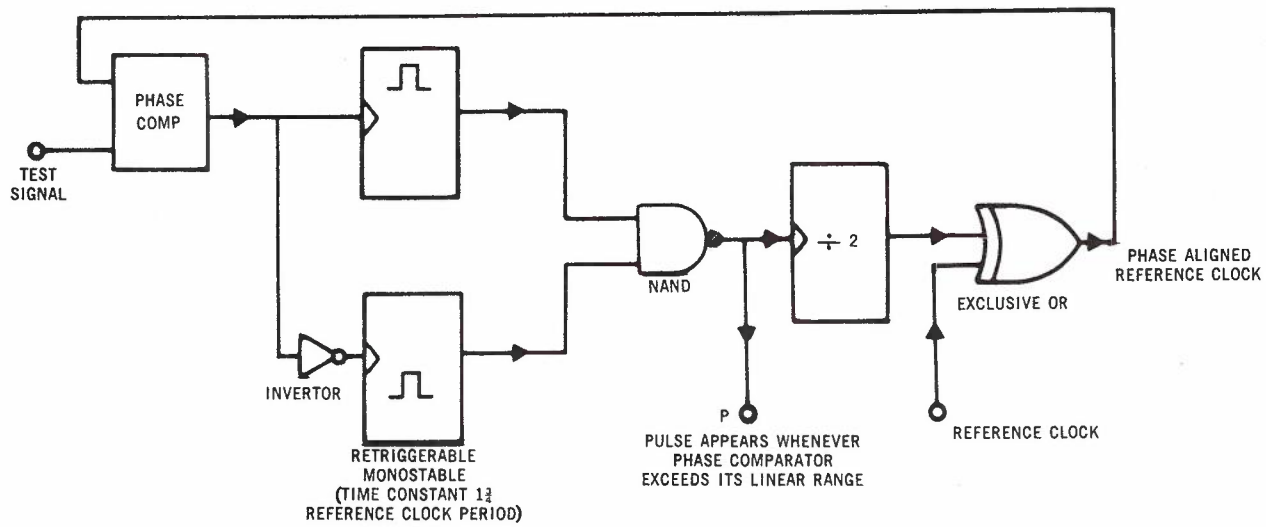


Fig. 6 - Automatic Phase Aligner.

In the next section of this chapter some limitations and refinements of the jitter measuring set of Fig.1 will be discussed.

2.3 Determination of Frequency Range

The possible distortion, due to sidebands, of the baseband jitter signal at the output of the low pass filter in Fig.1 can be evaluated by calculating the spectrum of a pulse width (duration) modulated pulse train for single edge modulation. This spectrum is dependent on whether the modulation is obtained by natural sampling or uniform sampling (Ref.8). In natural sampling the modulated pulses follow the amplitude variation of the sampled wave form during the sampling interval. Using the phase comparator as sampler results in natural sampling. Therefore the following calculation is based on natural sampling.

If the modulating jitter signal is a single sinewave: $\cos\omega_m t$, where ω_m = modulating frequency, we find from Ref.8 that the baseband signal is represented by:

$$Ap + \frac{AM}{2}\cos\omega_m t \qquad (2)$$

where Ap is a dc term,

A = pulse amplitude,

p = a factor equal to the duty cycle of the unmodulated pulse train ($p=\frac{\tau_o}{T}$ in Fig.2), and

M = modulation index = peak-to-peak jitter expressed as a fraction of the unit interval of the reference signal.

The additional frequency components have frequencies at $k\omega_r \pm n\omega_m$ for k,n = 1,2,...∞. Our concern is for components at $\omega_r - n\omega_m$ (k=1) as these components may interfere with the baseband jitter signal. From Ref.8 we find that components at $\omega_r - n\omega_m$ are represented by:

$$\frac{AJ_n(\pi M)}{\pi}\sin\{(\omega_r - n\omega_m)t + \phi(n)\}\ n=1,2...\infty \qquad (3)$$

where J_n = Bessel function of first kind of integer order

and $\phi(n)$ is a constant which depends on n.

Using Eq.2 and Eq.3 we find that the amplitude of each component at $\omega_r - n\omega_m$ compared with the baseband amplitude is:

$$20\log\frac{M\pi/2}{J_n(\pi M)}\quad \text{(in dB)} \qquad (4)$$

Computing Eq.4 with $J_n(x)$ obtained from Ref.9 for n = 1,2 and 3, we find:

TABLE 1

Peak-to-Peak Jitter in % (M)	Amplitude Ratio in dB Compared With Component at ω_m		
	$\omega_r - \omega_m$	$\omega_r - 2\omega_m$	$\omega_r - 3\omega_m$
5	0	-28	-60
10	0	-22	-48
20	0	-16	-36
40	-1.8	-11	-24

Note that, for low modulation levels, the $\omega_r - 2\omega_m$ component follows a linear law with respect to the peak-to-peak jitter M (6dB per octave) and $\omega_r - 3\omega_m$ follows a square law (12dB per octave).

As shown in Table 1, for most practical cases the component at $\omega_r - 3\omega_m$ can be neglected. However, if there are jitter components near 1/3 the reference frequency or higher, the interfering sideband components at $\omega_r - 2\omega_m$ may cause appreciable distortion. E.g. a measured 20% peak-to-peak jitter, using a set according to Fig.1, may be accurate within

$$20 \pm 3\% \text{ due to } \frac{Jn(\pi M)}{M\pi/2} \times 100 = 15\%$$

distortion of the component at $\omega_r - 2\omega_m$.

In conclusion, a jitter measuring set according to Fig.1 can give accurate results for moderate jitter amplitudes (e.g. up to 20%) if the significant jitter components have frequencies below 1/3 the reference clock frequency. For higher jitter amplitudes (e.g. between 20-50%) the significant jitter components should have frequencies below 1/4 the reference clock frequency.

In the next section a method is discussed which gives improved performance for high jitter frequencies.

2.4 Increased Frequency Range

The baseband jitter signal may be obtained precisely, with theoretically no distortion, by converting the pulse width modulated signal at the output of the phase comparator in Fig.1 into a pulse amplitude modulated signal (Refs. 8 and 10). Fig.7 illustrates the modified jitter measurement technique.

Note that for peak-to-peak measurements a low pass filter is no longer required.

For rms measurements, the baseband jitter signal may be obtained by using a low pass filter. As is well-known, the frequency spectrum of a pulse amplitude modulated pulse train has components at ω_m and at $k\omega_r \pm \omega_m$ (k=1,2,...∞) for a single modulating sinewave with frequency

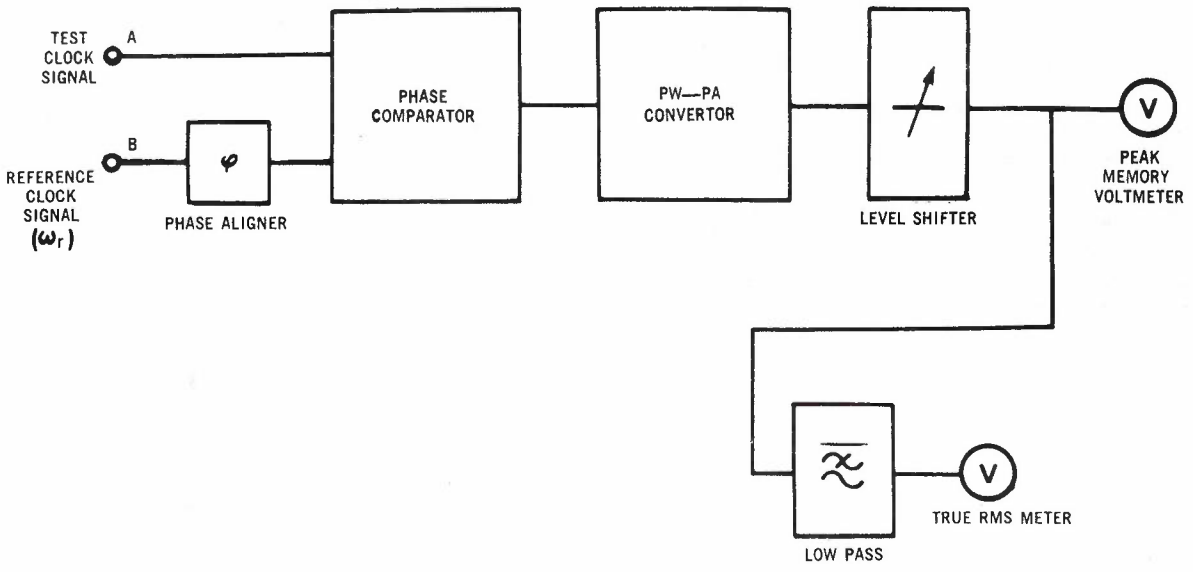


Fig.7 - Jitter Measuring Set with Improved High Frequency Performance.

ω_m. Therefore, a distortion free baseband jitter signal can be obtained if the jitter bandwidth is smaller than $\frac{1}{2}\ \omega_r$.

2.5 Increased Measurement Range

Using a latch type phase comparator, jitter up to a maximum of about 100% peak-to-peak can be measured. This range may be increased, as is well known, by applying dividers at the inputs of the phase comparator as shown in Fig.8. Both dividers must divide by the same number. Jitter up to dx100% can be measured using a divider by d.

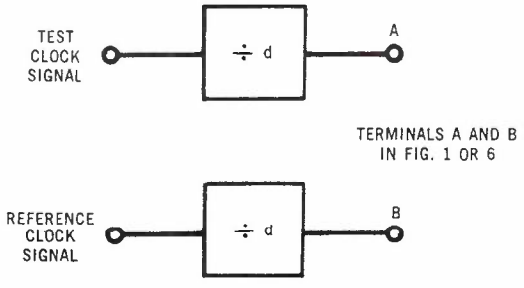


Fig.8 - Extended Measurement Range.

The effect of the divider is that the jitter signal is now sampled at a lower frequency. This means that the maximum jitter frequency which can be measured accurately is decreased by the dividing factor. Eq.3 is applicable if ω_r is replaced by $\omega_r' = \omega_r/d$.

Another effect of the use of a divider is that the resolution of the measurement is reduced. In Eqs. 2 and 4 the modulation index M has to be replaced by M' = M/d.

2.6 Reference Source in Case of Non-looped Transmission

If a looped transmission path cannot be used, it is no longer possible to use the original digital signal clock source as reference clock source. The usual method is to employ a phase-locked loop (P.L.L.) to derive the reference source from the signal to be measured (see Fig.9).

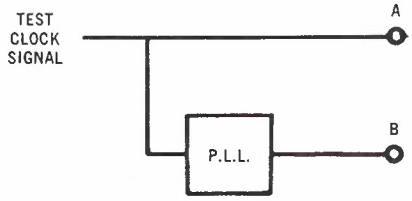


Fig.9 - Reference Signal Derived by P.L.L.

A P.L.L. has a low pass characteristic and any jitter in its pass band will appear as jitter in the reference signal. The cut-off frequency of conventional analogue P.L.L.'s used to derive the reference source is usually about 10-30Hz. Any jitter below this cut-off frequency will not be measured. It is possible to employ a digital type P.L.L. which, according to Ref.11, can be built with time constants as long as 1 day. Using such a digital P.L.L. to derive the reference source, very low frequency jitter can be measured.

In case it is important to measure low frequency jitter, an alternative to the use of a P.L.L. is the use of a highly stable signal clock source at the transmitter side and a highly stable unsynchronized reference clock at the receiver side. A stability of 1 part in 10^{10} per day or 1 part in 10^{11} per month, using crystal or rubidium clocks respectively, is possible. Periodic synchronization between the clock source and reference clock would be required to keep the frequency difference between the two clocks very small.

2.7 RMS Measurements

With the test set of Fig.1 rms jitter can be measured by using an rms meter rather than a peak memory voltmeter. However the static dc

component of the jitter (term Ap in Eq.2) has to be eliminated in order to correctly measure the actual rms jitter. Fig.10 shows this static dc component in relation to the phase characteristic of the phase comparator. Naturally, ac coupling of the rms meter will eliminate this dc component. However if low frequency jitter is to be measured, ac coupling is undesirable. One possible method is to dc couple the rms meter and to shift the dc level, with the level shifter, until a minimum reading from the rms meter is obtained. A dc offset will increase the measured rms value, as:-

$$rms_{measured} = \sqrt{rms^2_{jitter} + dc^2_{offset}},$$

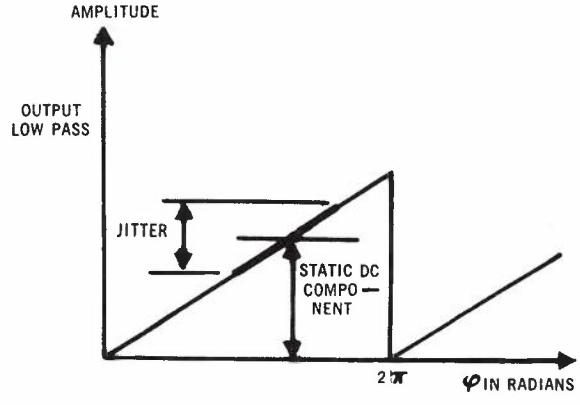


Fig.10 - Jitter in Relation to Phase Comparator Characteristic.

thus the measured rms value is minimum for a zero dc off-set. This method is applicable if the jitter process is stationary.

Alternatively, the mean jitter $\overline{f(t)}$ (which equals the static dc component) may be measured, with an average reading meter, in concurrence with the rms measurement. Both average and rms meter have to be dc coupled to the low pass output. The measured mean jitter is then used to correct the rms reading according to:

$$\text{if } \overline{f(t)} = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} f(t)\,dt, \text{ then}$$

$$rms_{jitter} = \sqrt{\lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} \left[f(t) - \overline{f(t)}\right]^2 dt}$$

$$= \sqrt{\lim_{T\to\infty}\left[\frac{1}{2T}\int_{-T}^{T} \left\{f(t)\right\}^2 dt\right] - \left\{\overline{f(t)}\right\}^2}.$$

(square of measured rms) (square of measured mean)

(5)

where f(t) = output low pass, and
2T = measurement period.

3. CALIBRATION

3.1 General Principle

Although several methods for calibrating phase jitter measurements exist, methods known to the author suffer from range limitations, application difficulties or inaccuracy. The novel method to be described below offers a wide range, ease of application and the possibility of a high accuracy.

Figs. 10 and 11 show the principle of the calibration technique. The technique is based on the fact that the phase difference between 2 clock signals of unequal frequency changes linearly with time. This linearity is exact if the clocks are stable and do not exhibit any jitter. In Fig.11 a clock signal f_1, with frequency f_1, is generated and a second clock signal f_2, with frequency f_2, is derived from signal f_1 by using a rate convertor (e.g. a P.L.L. or a frequency synthesizer may be used). The relation between f_2 and f_1 is:

$$f_2 = \frac{k}{m} f_1, \qquad (6)$$

where k and m are integers.

Signal f_1 is fed into an n stage counter, where n is chosen such that exactly n pulses from f_1 are counted when the phase between signals f_1 and f_2 has accumulated a difference of 2π radians or 1 bit at the rate of $|f_2-f_1|$. As 1 bit at a $|f_2-f_1|$ rate has a duration of

$\frac{1}{|f_2-f_1|}$ seconds, we find that

$$\frac{n}{f_1} = \frac{1}{|f_2-f_1|} \text{ or using Eq.6} \qquad (7)$$

$$n = \left|\frac{m}{k-m}\right| \qquad (8)$$

In order to find a desired integer n, k and m must have appropriate values.

As n pulses of signal f_1 represent a phase difference of 2π radians between signals f_1 and f_2, and the phase difference between f_1 and f_2 changes linearly, then q pulses of f_1 will represent a phase difference of:

$$\frac{q}{n} \times 2\pi \pm i.2\pi \text{ radians } (i=0,1,2,..\infty) \qquad (9)$$

The number q is set by a data word Q offered to a magnitude comparator. When the counter has counted q or more pulses, the magnitude comparator produces a pulse which blocks the output of

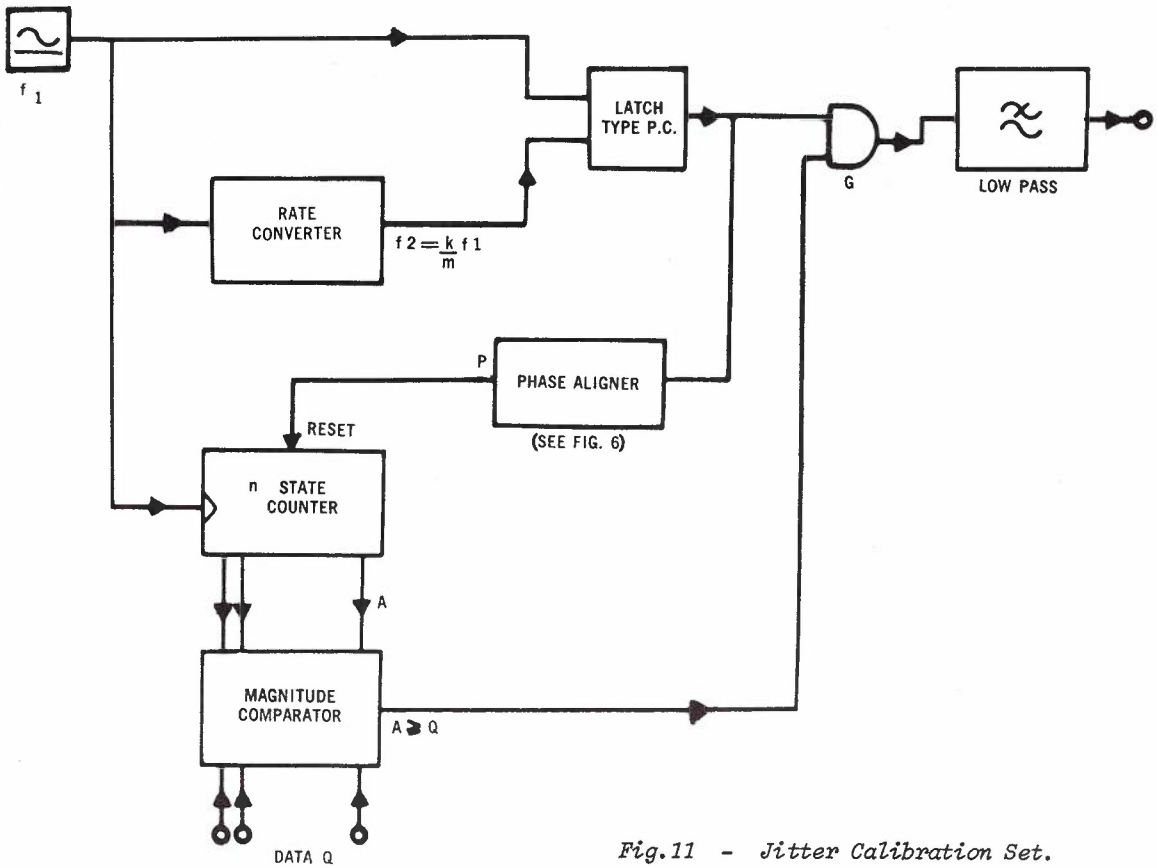


Fig.11 - Jitter Calibration Set.

the phase comparator, comparing f_1 and f_2, through gate G. Relevant waveforms are shown in Fig.12. By changing q a discrete number of phase difference points (up to a maximum of n) can be calibrated. To align a zero (or $\pm i \times 2\pi$ rad.) phase difference (q=0) with minimum output voltage of the low pass output a phase aligner as shown in Fig.6 may be used. Alternatively, a graphical or arithmetic correction may be used.

As an example for operation in the MHz range : suppose that f_1 = 1 MHz, m = 2000, k = 1998 then from Eqs.6 and 8 we find that f_2 = 999 kHz and n = 1000. Another example is f_1 = 1 MHz, m = 200, k = 198, f_2 = 990 kHz and n = 100.

Note that it is only necessary to calibrate the phase jitter meter somewhere in the relevant frequency range of operation, and that it is not required that f_1 equals the reference frequency in the jitter measurements.

Note also that the described calibration method is based on the linearity of the phase difference between two clocks and not on the phase characteristic of the phase comparator used. Any phase comparator could be used. However the method shown for phase aligning the counter (Fig.11) will only work for a phase comparator with a latch type characteristic.

3.2 Peak-to-Peak (p-p) Measurements

Using the calibration technique of section 3.1 it follows from Eq.9 that the calibrated p-p jitter ψ_{p-p} is:

$$\psi_{p-p} = \frac{q}{n} \times 2\pi \text{ radians, or}$$

$$\psi_{p-p} = \frac{q}{n} \times 100\% \qquad (10)$$

The relation between phase difference ψ, as

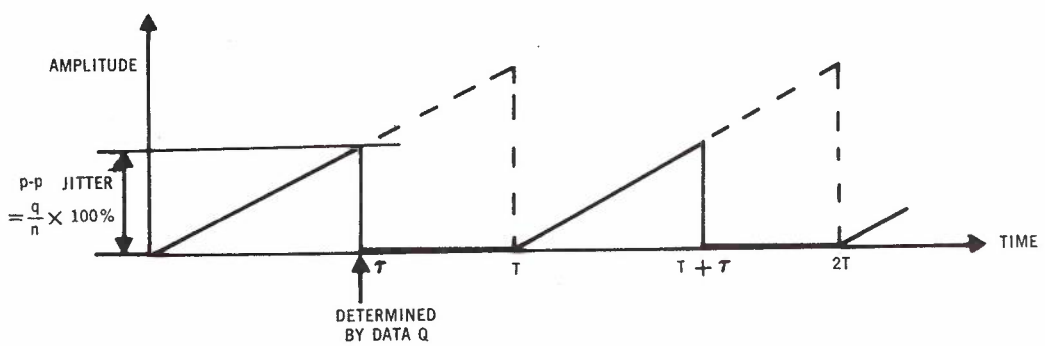


Fig.12 - Calibrated Jitter Waveforms at Low Pass Output.

measured by the phase comparator, and low pass output voltage V is in general:

$$V = C(\psi) \quad \psi \tag{11}$$

For the range of the phase comparator where $C(\psi)=C$ is constant, the relation between the measured peak-to-peak voltage (Vp-p) and ψp-p is

$$V_{p-p} = C\ \psi_{p-p} \tag{12}$$

Therefore in the range where $C(\psi)=C$ applies, the measured ψ_{p-p} is independent of the actual part of the phase comparator characteristic used for the jitter measurement and a dc offset, as required for the method of Fig.3, will not influence the measurement result. If $C(\psi)$ is not constant in the relevant range, as for a sinusoidal comparator, it would be necessary to make a correction for the applied dc offset.

Note that Eq.(10) is only dependent on the ratio of two numbers, which is consistent with expressing jitter as a ratio (e.g. % or degrees). If it is desired to express jitter in time, which for some purposes is useful, it is necessary to measure the frequency of the applied reference clock and multiply its inverse with the measured jitter ratio.

3.3 RMS Measurement

Using Eq.(11) ψ_{rms} is defined as:

$$\psi_{rms} = \frac{1}{T}\int_0^T \psi^2 dt = \frac{1}{T}\int_0^T \frac{V^2}{C^2(\psi)}\ dt$$

where T = measurement period (13)

A true rms meter measures $V_{rms} = \frac{1}{T}\int_0^T V^2 dt$

Therefore, if undue complications are to be avoided, the phase comparator must be used in a range where $C(\psi)$ is constant.

For $C(\psi)$ = C, it follows that

$$V_{rms} = C\ \psi_{rms} \tag{14}$$

As for a certain waveform $V_{rms} = \lambda\ V_{p-p}$, where λ is a constant* $(0 < \lambda \leq 1)$, we find, comparing Eqs.12 and 14, that the ψ_{p-p} calibration curve also applies to ψ_{rms} if a true rms meter is used instead of a p-p voltmeter.

---

* For the waveform of Fig.12 λ can be calculated as

$\lambda = \frac{\tau}{3T}$, and $\psi_{rms} = \sqrt{\frac{\tau}{3T}}\ \psi_{p-p}$.

3.4 A Note on Accuracy

The potential main factor determining the accuracy of the calibration is jitter in signal f_2 (Fig.11). If $f_1 = \sin\omega_1 t$ and $f_2 = \sin \omega_2 t + \psi_2(t)$, where $\psi_2(t)$ represents the jitter in signal f_2, then the low pass output voltage is proportional to:

$$(\omega_1 - \omega_2)t + \psi_2(t) \quad \text{mod } (\tfrac{q}{n} 2\pi) \tag{15}$$

Assume that ψ_{p-p} is to be calibrated. In the worst case, the maximum possible error in any calibrated ψ_{p-p} value equals the p-p phase variation of $\psi_2(t)$ (see Eq.15).

In many cases the contribution of $\psi_2(t)$ will cause a constant error (the p-p variation of $\psi_2(t)$) and this error can be corrected. However, in general, the accuracy of the calibration is clearly higher the lower the jitter in signal f_2.

4. CONCLUDING REMARKS

A method has been presented which allows to measure jitter of very low, as well as medium to high, jitter frequencies. This method is particularly suitable for high bit rate signals.

Techniques have been discussed which enable a simple digital read-out of the measured jitter. In particular, the presented calibration method could be used as part of an analogue-to-digital convertor, using comparison of the calibrated jitter with the measured jitter by successive approximation, to obtain a digital read-out.

5. ACKNOWLEDGEMENTS

The author wishes to acknowledge the valuable experimental assistance given by R.B. Coxhill.

5. REFERENCES

1. Mayo, J.S., "Experimental 224 Mbit/s PCM Terminals", Bell System Techn, J.,Vol.44, No.9. November 1965, pp.1213-1241.

2. Semple, G.J., "A Method for Assessing the Retiming Operation of PCM Regenerators". Aust. Telecom. Research, Vol.8.,No.2, 1974, pp.13-26.

3. Boatwright, J.T., "Phase Jitter Measurement on Voice Frequency Channels", IEEE Int.Conf. on Com. (ICC) 1970, pp.31/20-31/23.

4. C.C.I.T.T., Com. Special D, "Phase Jitter and its Measurements", Source : Sub-working Party for Question 3/1V, Doc.No.136, March 1972.

5. Lubarsky, A., "A digital method of Measuring Phase Jitter", IEEE Trans. on Com. Tech., October 1971, pp.736-737

6. Byrne, C.J., "Properties and Design of the Phase-Controlled Oscillator with a Sawtooth Comparator", B.S.T.J., March 1962, pp.559-602.

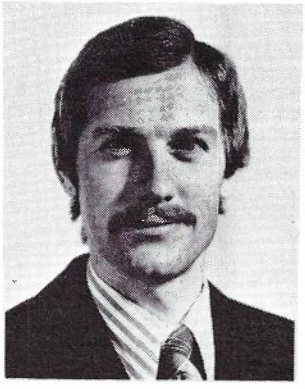
7. Macfarlane, I.P., "Pulse Generator and Flip-Flops", IEEE Trans. on Instrum. and Measurement, Vol.Im-21,No.2, May 1972, pp.148-153.

8. Black, H.S., "Modulation Theory", D. Van Nostrand Comp. Inc., New York, 1953, pp.264-279.

9. Abromowitz, M., Stegun, I.A., editors, "Handbook of Mathematical Functions", Dover Publ. Inc., New York, 1965, pp.390-401.

10. Taub, H., Schilling, D.L., "Principles of Communication Systems", McGraw-Hill, 1971, pp.179-187.

11. Pan, J.W., "Synchronizing and Multiplexing in a Digital Communications Network", Proceedings of the IEEE., Vol.60, No.5, May, 1972, pp.594-601.



BIOGRAPHY

JAN A. BYLSTRA was born in Rotterdam, The Netherlands. He received the M. Eng. Sc. degree in 1970 from the Technical University, Delft, The Netherlands. During 1970 and 1971 he was with the Research Laboratories of the Dutch P.T.T., where he worked on transmission system aspects of digital data networks. In December 1971 he joined the Line and Data Systems Section of the APO Research Laboratories. Initially, he worked on digital multiplexing techniques for PCM networks. At a later stage, he was involved in studies concerned with the introduction of a leased line data network. Recently Mr. Bylstra joined the Performance Objectives Section, Telecom Australia, H.Q., as a senior engineer. Mr. Bylstra is a Member of the Institute of Engineers, Australia and a Member of the Institute of Electrical and Electronics Engineers.

Australian Telecommunication Research

GUIDE TO AUTHORS

June 1977

1. INTRODUCTION

Australian Telecommunication Research is printed by offset lithography direct from camera-ready typescript which is prepared through the journal editors.

Pages when printed have a two-column format, A4 size.

In order to obtain the best possible copy with uniformity, and to reduce time consuming referrals back to you, please observe the few simple rules set out below as general requirements. These are illustrated by the format of this guide. This will reduce publishing time and limit costly mistakes.

2. MANUSCRIPTS

These must be typed, preferably using double spacing and with a single column per page. To improve legibility and layout, mathematical equations may be clearly handwritten, but take note of the restrictions discussed in Section 9.

Number each page in the top right hand corner.

3. TITLE

Type the title of your paper in capital letters at the top of page 1. The title should not exceed two lines of typing.

Insert name(s) of author(s) in capitals, with affiliation(s) in lower case underneath, on the left side of the page below the title. Degrees, decorations etc. should not be included. The following is an example of the format required;

A.B. SMITH,
Telecom Australia,
Melbourne.

C.D. JONES,
Department of Physics,
University of Sydney,
N.S.W.

4. ABSTRACT

An abstract, not exceeding 150 words, indicating the aim, scope and conclusions of the paper should follow below the affiliation of the author(s).

5. HEADINGS

Headings generally make the text more readable. All headings should be preceded by numbers commencing at the left margin. These are of three orders as illustrated by the following example:-

1. FIRST-ORDER HEADINGS

These headings should be typed in capitals and underlined. Each first-order heading should be prefixed by a number which indicates its sequence in the text, followed by a full stop. The first-order heading is preceded by two lines of space and followed by two lines of space.

1.1 Second-Order Headings

These headings should be underlined and typed in lower case letters with initial capitals only. They should be prefixed by numbers separated by a full stop to indicate sequence in the text. They should be preceded by two lines of space and followed by two lines of space.

1.1.1 Third-Order Headings. These should be typed as for second-order headings, underlined and followed by a full stop. The text should continue on the same line as the heading. As with other headings they should be preceded by two lines of space.

Numbering of third-order headings is optional, but where used for clarity, should be of the form 1.1.1 (i.e. two full stops as separators).

6. REFERENCES

A list of references should be given at the end of the paper, typed in close spacing with a line between each reference cited. References must be sequentially numbered in the order in which they are called in the text. They should be called up in the text as (Ref. 1).

6.1 Format of References

The format is: reference number, full stop, 2 spaces, name of author, title of reference in inverted commas, abbreviated name of the journal, month and year of publication, volume and issue number, followed by the page numbers. Do not use any underlining in the list of References. An example is:-

1. Payne D.N. and Gambling W.A., "Silica-Based Low-Loss Optical Fibre", Electron Lett., July 1974, Vol. 10, No. 15, pp. 289-290.

2. Berry L.T.M., "The Gradient Projection Vector for Multicommodity Network Problems", Australian Telecommunication Research, 1975, Vol. 9, No. 1.

7. TABLES

Table headings appear above the table, thus:-

TABLE 5 - Truth Table for Attenuator Correction

System Gain (dB)				Correction in Decimal (dB)	Correction in Binary		
-16	0	+16	+32		B6	B5	B4-B1
1	0	0	0	48	1	1	0
0	1	0	0	32	1	0	0
0	0	1	1	16	0	1	0
0	0	0	1	0	0	0	0

8. FIGURES

Figures, drawings, sketches etc., which are to appear in the text can be supplied in clear unambiguous freehand form. They will be drafted in ATR format through the editors.

All symbols used in either drafted or freehand sketches should conform with the appropriate Australian standard used by Telecom.

When reference to a figure is made in the body of the text, it will take the following form: (Fig. 1).

9. EQUATIONS

Equations are numbered consecutively with Arabic numerals in parentheses, placed at the right hand margin, e.g.:-

$$g(t) = g_1(t) * g_2(t) \qquad (14)$$

Take care when equations are hand written that all subscripts, Greek letters, symbols, etc., are unambiguous and easily recognisable by the typist. Take your time to carefully write equations. Formulate a style of setting out which is close to text book presentation to reduce errors, misinterpretation and reader fatigue. Special instructions to the typist (e.g. an illustration of photo copy of a page of text) may be necessary for your particular requirements.

Fundamental rules include:-

(i) *carefully* form Greek or mathematical symbols

(ii) *set*, where applicable, all equation equal signs (or equivalent) directly under each other

(iii) *correctly* locate summation and integral limit values to avoid confusion as subscripts for other variables in the equation.

(iv) because of reproduction limitations, any symbols used should be confined to the English alphabet or selected from Table I below:-

TABLE I - Range of Symbols

1	2	3	4	5	6	7	8	9	0	=	∠
γ	δ	ε	θ	τ	υ	ξ	ι	ο	ρ	π	
α	σ	φ	ι	λ	η	ϳ	κ	ω	¨	÷	
ζ	χ	ψ	×	β	ν	μ	´	~	ι		
✓	|	{	}	±	⌜	⌠	⌠	⌝	⌡	_	⌡
Γ	Δ	←	Θ	↔	Τ	Ξ	↑	↓	ℓ	Π	
∇	Σ	Φ	<	Λ	¶	>	§	Ω	^	•	
≃	Ξ	Ψ	∝	∞	∿	∂	`	†	∫		

("Δ 12 IBM SYMBOL 12" type ball)

NOTE Combinations of the above symbols will produce larger symbols

e.g. ⌜ and ∠ give ∑

⌠ and ⌡ give ∫

⌡ and ι give { and }

INDEX — VOLUMES 6-10 INCLUSIVE
(Subject, Title, Author Intermixed)

Printed by Standard Newspapers Limited, 10-14 Park Road, Cheltenham, 3192.

AUSTRALIAN
TELECOMMUNICATION
RESEARCH

VOLUME 11, NUMBER 2,
1977

Contents*

*Abbreviated Titles