PULSE CODE MODULATION RECORDING FOR TELEMETRY APPLICATIONS

WILLIAM N. WAGGENER
EMR-Telemetry

Summary. The problems of signal design and detection and multitrack synchronization are examined for pulse code modulation high-density digital recording. Theoretical bit error probability results are compared with experimental data taken for bandlimited conditions typical of tape recorders. The problem of synchronizing multiple-track data is considered and the maximum likelihood synchronizer is obtained. Variations on the maximum likelihood synchronizer are considered.

Introduction. As digital data systems increasingly appear in telemetry applications, the pulse code modulation (PCM) user must carefully consider the problem of data recording. Real-time recording forms the first line of defense against catastrophic system failure. But, as data rates have soared, the ability to store the data on standard instrumentation recorders has diminished. This problem has spawned the development of high-density digital recorder (HDDR) systems. Currently, the major tape recorder manufacturers offer a variety of HDDR systems. For many users, these standard HDDR systems satisfy their requirements. The current systems do suffer, however, from a lack of compatibility from system to system. For some users, the standard HDDR systems do not, for one reason or another, fulfill their requirements. Many users, for example, have a considerable investment in standard wideband instrumentation recorders and cannot afford to reinvest in the new-generation HDDR systems.

This paper is directed toward the users who wish to consider alternate approaches to high-density recording other than those offered in standard HDDR systems. The intent here is not to provide a design concept for a new system, but rather to examine the key issues of detection and synchronization for multi-track digital recording. The question of signal design and detection for high-density recording has received considerable attention in the past two years and has been the subject of some controversy. Very little effort has been directed toward the synchronization problems associated with multi-track recording. The signal design and detection problem will be reviewed with a brief discussion of some supporting experimental data on the relative performance of the most-common code formats. Some previously unpublished theoretical results dealing with the problem of multitrack synchronization will also be presented.

The Tape Recorder as a Communication Channel. The design of an HDDR system depends largely upon a characterization of the magnetic recording process. The tape recorder can be viewed as a noisy, nonlinear, bandlimited communication channel. The more gross characteristics of the channel can be inferred from an examination of the recording process. An analytical approximation to the linear transfer function of the digital recording system can be obtained. In the simplest model of a recording system, the voltage induced in the read head is proportional to the rate of change of tape magnetization. At low frequencies, this leads to a differentiator model for the recording process.

If the tape head is thought of as having an aperture which views the recorded magnetization pattern, the aperture would have to be infinitely narrow in order to pass all frequencies on the tape. Because of the head geometry and head construction, the head has an effective aperture width which causes a degradation in response at high frequencies. If the head aperture were approximately rectangular, the head frequency response would roll off at high frequencies as a $\frac{\sin x}{x}$ response. The composite recording system transfer function is then typically of the form shown in Figure 1. Experimental measurements verify this general transfer characteristic. The characteristic is non-linear in that the transfer function depends upon record level and other parameters. Nonetheless, Figure 1 represents the general nature of the tape recorder response. Clearly, one important aspect of the recording system as a communication link is the lack of dc response. The response is considerably modified by the characteristics of the tape. Tape imperfections create dropout periods during which the signal amplitude can diminish considerably and so the transmission channel is, in general, time varying. The tape is also a primary contributor to system noise. The optimum design of an HDDR system depends crucially upon considering the bandlimiting characteristics of tape recorder media. When recording on multiple tracks, the recorder can be considered to be a communication channel with multiple communication paths. Each path has a different fixed time delay as well as a random delay component. It should also be recognized that no two heads on a stack are exactly identical and, therefore, the transfer functions of the individual tracks may differ.

Signal Design and Detection. The performance of an HDDR system is fundamentally a function of the channel encoder/decoder. The primary function of the channel encoder is to accept serial input data and convert this data into a sequence of symbols to be fed to the channel modulator. The channel symbols in this paper will be considered to be the recorded PCM code. The decoder, in turn, must extract timing information from the received signal (synchronization) and make optimum symbol decisions in the presence of noise and signal distortion.

Although the typical tape recorder transfer function is of the form shown in Figure 1, most recorder users do not have access to the basic read head output. In most cases, the user receives an equalized signal at the recorder output which has compensated for the differentiation effect at low frequencies. In the direct record mode, the recorder manufacturer has equalized the frequency response so as to be essentially flat from an audio low-frequency cutoff to the inherent high-frequency record limit. In this paper, the effective transfer function of the recorder will be assumed to be that of the equalized response. Other investigators[1, 2] have considered the more fundamental case of the design of signals for the actual unmodified recorder response. The recorder is, therefore, considered to be a channel with roughly a low-pass frequency characteristic having a lower frequency cutoff in the range of several hundred Hertz. In designing the signal for recording, the telemetry user naturally looks to the standard PCM code formats as suggested by the Inter-Range Instrumentation Group (IRIG). This represents a logical choice of signals since standard product synchronizers and detectors exist for these codes.

Of the PCM codes given in the IRIG standards, non-return-to-zero (NRZ). bi-phase, and delay modulation (DM) codes have been the most widely used for tape recorder applications. Where high packing density is not a prime requirement, the bi-phase code is very attractive for use on the tape recorder. The bi-phase code requires no dc response and offers one symbol transition each symbol period. The bi-phase code, while offering the same bit error probability performance of NRZ, requires a transmission channel bandwidth of approximately twice that of NRZ. Thus, for bandlimited channels, bi-phase code is only about half as efficient as NRZ code for the same level of error performance.

When DM was first proposed, the initial reaction of many people was that DM would be an ideal baseband code for high-density recording applications. This conclusion was based upon a pronounced concentration of spectral energy in a band from about 0.2 bit rate to 0.6 bit rate. The small low-frequency content seemed ideal for the tape recorder channel. The apparent advantage of DM codes over NRZ codes was based solely on a comparison of the energy distribution in the frequency spectrum. Conclusions drawn from the signal frequency spectrum can be very misleading. If a careful examination is made of the performance of DM, it is found that the optimum detector for non-bandlimited DM requires the use of a matched filter whose bandwidth is twice that of the matched filter for NRZ code. Consequently, the optimum receiver for non-bandlimited DM admits 3 dB more noise power into the decision circuitry and, thus, the performance of DM is at least 3 dB poorer than NRZ. Using the optimum single bit period detector, the performance of DM is approximately 3.5dB poorer than NRZ. The additional 0.5dB degradation can only be removed by multiple bit processing. The optimum single bit period detection technique can be inferred from the nature of the DM code. In DM-Mark, a “one” is represented by a midbit period transition provided that a “zero” followed by a zero causes a transition at the beginning of the first zero. This code format suggests the optimum detector. A one is always characterized by a change in the middle of a bit period while a zero never contains a change. By comparing the energy in the first half of the bit period with the energy in the second half of the bit period, ones and zeros can be distinguished. Even though non-bandlimited DM is 3 dB poorer in bit error performance than NRZ, proponents of DM still pointed to the desirable features of the code. In particular, the small dc energy component was cited as well as the sharply concentrated energy at frequencies less than the bit rate. The implication was clearly drawn that DM offers a more-suitable code for the bandlimited tape recorder than does NRZ. In order to examine this conjecture, the effect of bandlimiting on performance of DM and NRZ must be considered.

First, let us consider the effect of low-pass filtering on the performance of NRZ. For ease of testing, a third-order Butterworth low-pass filter has been used to simulate the bandlimiting characteristics of the transmission channel. The bit error probability as a function of signal-to-noise ratio is shown in Figure 2 for bandlimiting NRZ. The bandlimiting is expressed by the ratio of the filter 3-dB bandwidth to the bit rate. Without bandlimiting, the bit error probability for a practical detector is approximately within 1 dB of the theoretical performance. With the data bandlimited to a bandwidth equal to the bit rate, an additional 1.5 to 2 dB degradation is incurred. This degradation is due to the intersymbol interference introduced by the bandlimiting.

Next, consider the effect of bandlimiting on the performance of DM. Figure 3 shows experimental data as measured with an EMR Model 720 Bit Synchronizer for bandlimited DM using the same low-pass filter. With no bandlimiting, the performance of DM is also within 1 dB of the theoretical performance curve. It should be noted that the theoretical performance is 3.5 dB poorer than that of non-bandlimited NRZ. Bandlimiting DM to a bandwidth equal to the bit rate degrades the error performance by an additional 3 dB over the unfiltered code performance. Thus, if one were to compare the performance of filtered DM to the performance of filtered NRZ, the DM code performance would be approximately 4.5 dB poorer than NRZ when bandlimited to a cutoff frequency equal to the bit rate. The performance degradation can be clearly seen by examining the eye patterns for DM and NRZ. For the filtered case, as shown in Figure 4, not only is the DM eye closed more by filtering, but the timing tolerance is one-half that of NRZ.

With a tape recorder channel, high-frequency bandlimiting represents only one aspect of the data recording problem. As indicated, the recorder response is normally equalized to only a low-frequency cutoff in the order of a few hundred Hertz. With the random NRZ data, the signal energy extends to dc and, with a finite probability, long strings of ones or zeros can occur. The lack of dc response seriously degrades the performance of NRZ coding. Under these conditions, the energy near dc is for DM coding small but finite -- with a transition at least every two bit periods. In order to utilize NRZ, it is necessary to ensure a minimum transition density in the code. In some cases, this, can be accomplished by inserting parity bits, while in other applications pseudo-random sequence scrambling can be employed. Group coding or adding parity bits to the data has been employed by some tape recorder manufacturers while others have pseudo-randomly scrambled the data.

In order to realistically compare the performance of NRZ data with DM data for the tape recorder application, a minimum transition density NRZ code must be employed.

The low-frequency cutoff of the tape recorder response is normally a very small fraction of the high-frequency cutoff. In order to provide a conservative comparison, the relative performance of DM and NRZ was examined for a channel whose low-frequency cutoff was 0.5% of the upper frequency cutoff. A low-transition-density NRZ signal was used having one data transition every eight bits. Under these conditions the performance of the NRZ maintains a 4-dB performance advantage over DM. With bandpass filtering, the performance of NRZ is dependent upon the data pattern and on the performance of the receiver dc restoration circuitry. Performance of DM on the other hand is essentially independent of code pattern with one very important exception. If the data sequence does not contain the “1 0 1” data sequence or contains it only occasionally, the DM synchronizer is unable, to resolve the synchronization ambiguity. Under these circumstances, the synchronizer is completely unable to detect DM. Consider, for example, a data stream which contains mostly ones or zeros with isolated bit changes occurring relatively infrequently. Under these circumstances, the DM synchronizer is unable to resolve the ambiguity inherent in the DM synchronization process. Thus, contrary to popular belief, DM is, in fact, data-pattern dependent. When pseudo-random scrambling is used, NRZ performance suffers an additional degradation because of the finite probability of a lower transition density. This reduces the margin of performance of NRZ over DM. It is found in general that DM can tolerate a much higher low-frequency cutoff than can NRZ for the same performance level. This is important for communication channels in which the low-frequency cutoff is a significant fraction of the high-frequency cutoff. For tape recorder applications, this can sometimes occur at very low recording speeds. At higher recording speeds, however, the performance advantage of NRZ over DM is clear, provided that NRZ is coded to ensure minimum transition density. Recently King [3] has published experimental measurements comparing NRZ PCM/FM and DM PCM/FM indicating that NRZ maintains a 3 dB advantage over DM for the PCM/FM case as well as for baseband signaling.

In the experimental tests which have been performed, the detection filter used for both NRZ and DM is optimized for unfiltered data. It is theoretically possible to recover most of the performance degradation due to filtering by use of an optimum detection technique such as maximum likelihood estimation (MLSE) as discussed by Forney [4] and others. Using MLSE, it is theoretically possible to reduce the degradation of DM with respect to NRZ to about 3 dB as is the case for unfiltered data. To accomplish this, however, requires a fairly complex processor which processes sequences of data samples rather than performing bit-by-bit decisions.

Signal design theory has been utilized to design specific codes for the HDDR requirement. One technique which has been proposed [5] codes four data bits into six channel symbols. Through this coding, a transmission signal is developed which minimizes the signal energy at low frequencies to more closely match the channel response.

A fundamental premise of communication theory is that the ability to distinguish between two symbols in the presence of noise is dependent only on the difference in the energy of the two symbols. The greater the energy difference between the symbols, the more noise-immune the signaling. If the symbol duration is limited, the greatest energy difference between binary symbols is obviously achieved with NRZ coding. An energy difference equal to that of binary coding is achieved by bi-phase codes although the bi-phase codes require twice the transmission bandwidth of NRZ. As pointed out recently by Pelchat and Geist [6], the bandwidth of a signal is not a reliable indication of the required channel bandwidth and, in fact, the relevant bandwidth is the bandwidth of the energy difference signal. Pelchat and Geist have shown that the power spectra of two-level codes can be made as narrow as desired for a constant bit rate. Paradoxically, the decrease in power spectrum bandwidth requires an increase in effective channel bandwidth. A good example of this is the case of DM. Even though the signal power spectrum is narrow, the minimum distance between symbols is equivalent to an RZ pulse which has, in fact, twice the bandwidth of NRZ. The fact that the minimum distance between symbols in DM coding is one-half that of NRZ coding explains why the performance of DM is at least 3 dB poorer than NRZ. Similar comments can be made relative to signal designs based on binary waveforms. Even though a signal may be designed with a greater energy difference than NRZ, the improvement is achieved at the expense of bandwidth.

Multi-track Bit Synchronization. The problem of establishing bit synchronization for multiple data tracks is crucial to the implementation of an HDDR system. Typically, the problem of establishing synchronization for multiple tracks has been solved in a brute-force fashion. Individual, independent synchronizers have been employed for each track. This provides a most-straightforward solution to the synchronization problem, but does have a number of faults. Since the synchronizers are completely independent, a synchronization failure on one synchronizer completely destroys information from that track. For many systems, such a failure would be nearly catastrophic since it would represent a sizeable loss of data. The problem of multi-track synchronization is analogous to several other related problems which have been investigated. One of these is the synchronization of multiple power networks. A second similar problem is the synchronization of a large digital communication network. Some of the results of these apparently unrelated investigations offer insight into methods by which multi-track synchronization can be improved. In particular, optimum control studies of power network synchronization show that the best control over network synchronization is provided with interacting controls between individual portions of the network as opposed to independent control over each separate part of the network. In a like manner, one of the techniques which has been investigated for controlling the synchronization of a large digital communication network has been a frequency-averaging system whereby timing for any individual subsystem in the network is achieved not only from its own timing sources but from averages of the timing of other subsystems. These results suggest that multi-track synchronization can be improved by providing some sort of coupling between the individual track synchronizers. Interconnecting the individual synchronizers from each track provides a measure of immunity to synchronization failure on any individual track.

The problem of synchronizing multiple channels can be approached theoretically. If all tracks have the same fixed delay, * the signals from the individual tracks are of the form:

$$\tilde{y}_1 = g_1 \sum_{-\infty}^{+\infty} \tilde{a}_{1k}\, p\,(t - kT + \tau) + \tilde{n}_1 \tag{1}$$

$$\tilde{y}_2 = g_2 \sum_{-\infty}^{+\infty} \tilde{a}_{2k}\, p\,(t - kT + \tau) + \tilde{n}_2$$

$$\vdots$$

The task of the synchronizer is to examine the signals and to extract an estimate of τ in some “optimum” manner. The approach taken in this paper is to make a maximum likelihood estimate (MLE) of τ from the signals. For ease of statement, the input signal is written in vector notation:

$$\underline{\tilde{y}} \triangleq \begin{bmatrix} \tilde{y}_1 \\ \tilde{y}_2 \\ \tilde{y}_M \end{bmatrix} \qquad \begin{array}{l} \text{M x 1 vector} \\ \text{(M = number of channels)} \end{array} \tag{2}$$

The MLE approach is to find a value of τ such that the following function is maximized:

$$\max \qquad \ln p\,(\underline{\tilde{y}}\,|\,\tau) \tag{3}$$

$$\text{with } p\,(\underline{\tilde{y}}\,|\,\tau) = \text{joint conditional density of } \underline{\tilde{y}}$$

Defining the log likelihood ratio as:

$$\Lambda\,(\tau) \triangleq \ln p\,(\underline{y}\,|\,\tau) \tag{4}$$

* This assumption can be removed and does not affect the basic results.

After some rather lengthy mathematics, the log likelihood ratio can be shown to be of the form:

$$\Lambda(\tau) = \sum_{1}^{M} \sum_{k} \ln \cosh \frac{2}{N_o} \quad g_j \, z_{jk}(\tau) \tag{5}$$

where:

$$z_{jk}(\tau) \triangleq \int y_j(t) \, p(t - kT + \tau) \, dt$$

This rather formidable expression represents the answer to the MLE problem. The synchronizer should choose the value of τ which maximizes Equation (5). For some particular data sequences, $z_k(\tau)$ is periodic and, hence, $\wedge(\tau)$ does not have a single maximum. The synchronizer must find a local maximum of $\wedge(\tau)$ over a symbol period. To find such a maximum, the derivative of Equation (5) is taken:

$$\frac{\partial \Lambda(\tau)}{\partial \tau} = \sum_{1}^{M} \sum_{k} \frac{2}{N_o} \quad g_j \, \dot{z}_{jk}(\tau) \, \tanh\left[\frac{2}{N_o} \quad g_j \, z_{jk}(\tau)\right] \tag{6}$$

The synchronizer must now find a value of τ such that a $\partial\wedge/\partial\tau$ approaches zero. This suggests a synchronizer structure with feedback as shown in Figure 5. The loop filters provide the summation (over k) function. The order of summations in Equation (5) can be reversed so that the individual signals can be summed first and a single loop filter used. This realization is shown in Figure 6. It should be noted that the signals are summed in accordance with their relative amplitudes, g_j. The realizations shown in Figures 5 and 6 pose some practical implementation problems. Consequently, some suboptimum approximations to the MLE synchronizer are in order.

Rather than burden the reader with the details of approximations to the MLE synchronizer, suffice it to say that the same methods are employed and the same results obtained as with single-channel synchronizers. It can also be shown that the multi-track synchronizer, using the master VCO, is exactly equivalent to an interconnected group of individual synchronizers.

Several variations of the basic multiple-track synchronizer have been considered. The first technique consists of a master VCO which provides the timing reference to all tracks. Each track has a local phase shifter which adds a phase-correction term to the master VCO input and applies the resulting signal to a local phase comparator. The output of the local phase comparator is filtered and is used to both drive the local phase shifter and to provide a control signal to the master VCO as shown in Figure 7. For this system, the VCO provides a master clock which is controlled by the composite signal from all tracks. The phase shifter which is associated with each track provides a vernier control over the local track phase error so as to minimize the tracking phase error. In essence, the idea is that the master VCO maintains an average frequency with each individual track minimizing its local phase error. If a signal drops out on any individual track, the effect on synchronization is minimized due to the coupling between tracks through the master VCO.

The second technique which has been considered is to provide individual synchronizers for each track, but to provide some form of coupling between synchronizers so that if a signal drops out on any individual track, the synchronizer continues to receive an average frequency error signal from the other track synchronizers.

In order to fully understand the dynamics of a complete synchronizer, a simulation was performed of a six-track synchronizer. With the many tracks involved, it is difficult to examine all of the combinations and permutations of the various loop parameters. Consequently, all loops were set up to have the same gains and type of filtering and an effort was concentrated on understanding the behavior of the multi-track synchronizer loops with relatively simple combinations of input signals and gain changes. The overall loop was originally set up such that with no local control the loop was equivalent to a simple phase-lock loop having a 1-radian-per-second natural frequency. Next, local gain was applied to the individual loops. In comparing the response with individual track feedback with the response of the system having no local control, the local phase control rapidly reduces the phase error in the individual tracks. This is to be expected since the loop dynamics are now determined primarily by the local loop characteristics and less by the master VCO characteristics. The local loop has an effectively wider bandwidth than the overall master loop and allows the local track to respond much more rapidly than is permitted using the master VCO alone. With only proportional gain in the local loops, it can be shown that a static phase error exists in the local loops. This is, of course, undesirable and integral control should be applied on the local loops to eliminate steady state phase error. Figure 8 shows the phase error response in loops 1 and 2 with the addition of integral control on each loop for a 1 -radian step input on track 1 and an initial phase error of 0. 5 radian on track 2.

In the course of the synchronizer investigation the possibility of using a discrete phase shifter for each track was considered. The discrete phase shifter effectively picks the phase of the master VCO which is closest to the phase error of the local loop. Thus, the phase shifter in each loop can be realized utilizing digital logic. The addition of the nonlinear quantization in the individual loops greatly complicates the analysis of the synchronizer. However, these quantization loops are readily simulated.

Figure 9 shows the phase error in loops 1 and 2 with a 0.1 radian quantization in the local phase comparator. Each of these cases considers an initial phase error of 1 radian in loop 1 and 0. 5 radian in loop 2. The behavior of the loop is more or less as might be expected. The error is reduced to the point at which the local phase error cycles back and forth between limits of the same magnitude as the phase shifter quantization.

The performance of the MLE synchronizer has been estimated using the Cramer-Rao (C-R) bounds and the mean square error computed for a periodic input signal. The performance analysis yields the following interesting results:

1. The root-mean square (rms) timing error is reduced by the square root of the number of channels combined.

2. The rms error increases with the reciprocal of the transition density.

3. The C-R performance bound is approached by the squaring loop synchronizer.

4. By virtue of result (1), a multi-channel synchronizer is relatively immune to the loss of a single channel.

5. The inputs to the loop filter should be weighted in accordance with the signal amplitude.

6. If the individual channel signal amplitudes vary, the amplitude should be measured and the gain into the loop adoptively changed.

7. The individual channel sampling should be adjusted to compensate for any fixed delay between channels.

Conclusions. The two major considerations in the design of an HDDR system are (1) signal design and detection, and (2) synchronization. Signal design and detection for HDDR has previously received considerable attention. The relative performance of the more-common codes are verified by experimental measurements and the effects of bandlimiting are examined. The problem of multi-track synchronization has been investigated both theoretically and by simulation. The MLE multi-track synchronizer configuration is obtained and has been shown to be equivalent to coupled independent synchronizers. The performance of the MLE synchronizer shows it to be quite tolerant of individual track dropouts. Individual track phase error feedback can be employed to track local phase variations with the master VCO providing average frequency control.

References

(1) Kobayashi, H., and Tang, D. T., “Application of Partial-response Channel Coding to Magnetic Recording Systems,” IBM J. Res. Develop., July 1970.

(2) Kobayashi, H., “Application of Probabilistic Decoding to Digital Magnetic Recording Systems,” IBM J. Res. Develop., Jan. 1971.

(3) King, D. A. , “ Telemetry System Parameters and Bit Error Performance of NRZ and DM PCM/FM,” Technical Publication TP-76-66, Pacific Missile Test Center, Point Mugu. Calif., Mar. 1976.

(4) Forney, G. D., Jr.. “Maximum -Likelihood Sequence Estimation of Digital Sequences in the Presence of Intersymbol. Interference,” IEEE Trans. on Information Theory, Vol. IT-18, No. 3, May 1972.

(5) Halpern, P. H. , U. S. Patent 3, 921, 210, Nov. 1975.

(6) Pelchat, M.G. and Geist, J. M., “Surprising Properties of Two-Level ‘Bandwidth Compaction’ Codes, “ IEEE Trans. on Communications, Sept. 1975.

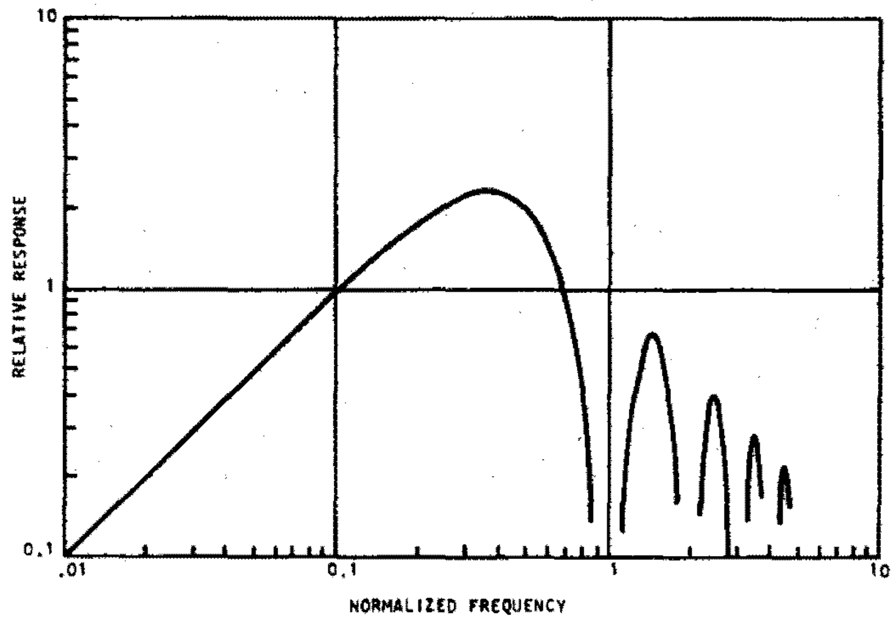

Fig. 1 - Theoretical Frequency Response for Ideal Read Head

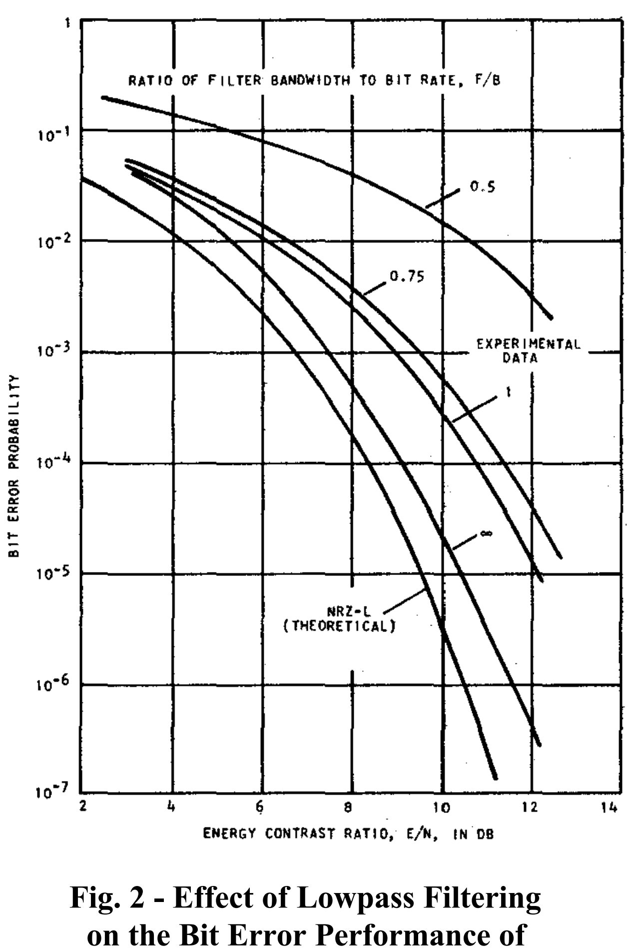

Fig. 2 - Effect of Lowpass Filtering on the Bit Error Performance of NRZ-Level (Third-order Butterworth Lowpass Filter)

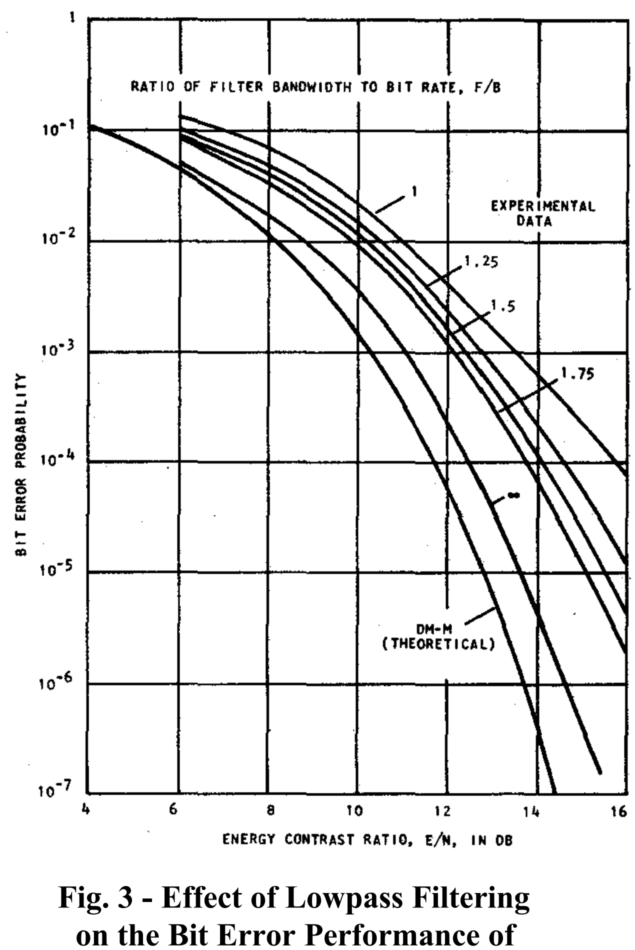

Fig. 3 - Effect of Lowpass Filtering on the Bit Error Performance of DM-Mark (Third-order Butterworth Lowpass Filter)

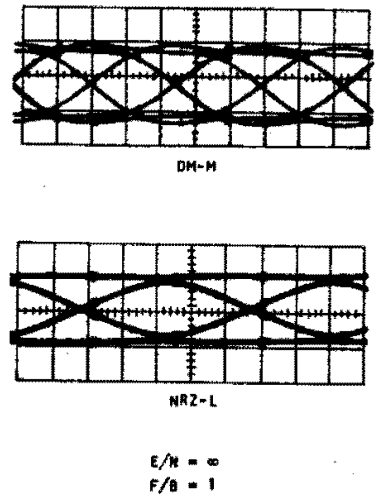

Fig. 4 - Eye Patterns at Matched Filter Output for DM-Mark and NRZ -Level

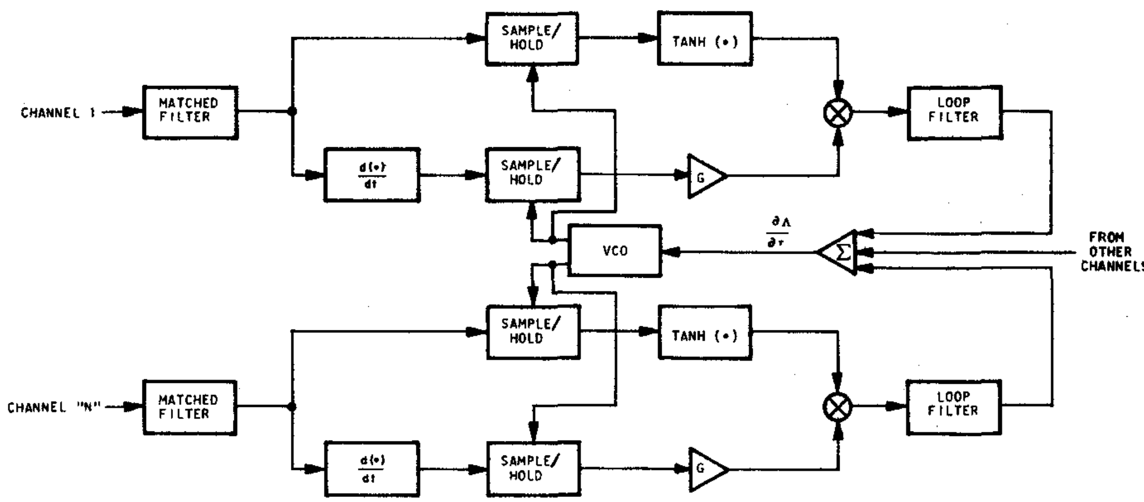

Fig. 5 - Maximum Likelihood Synchronizer for Multiple Channels

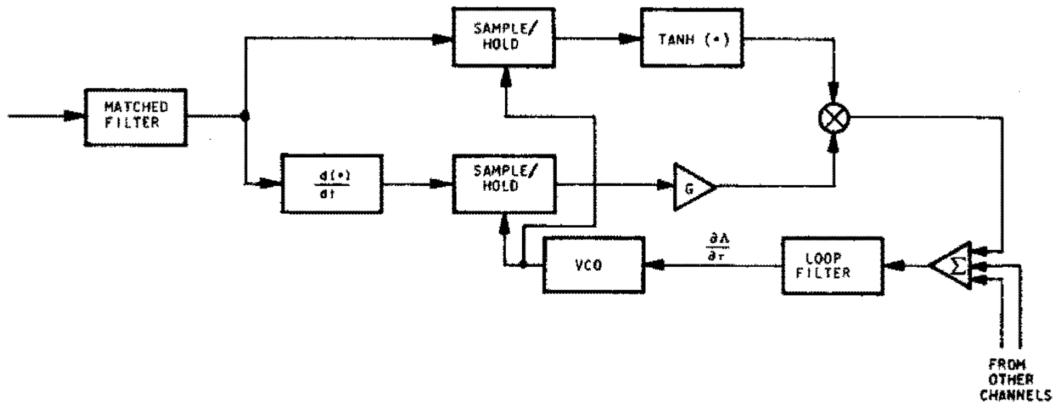

Fig. 6 - Alternate Configuration of Maximum Likelihood Multiple-channel Synchronizer

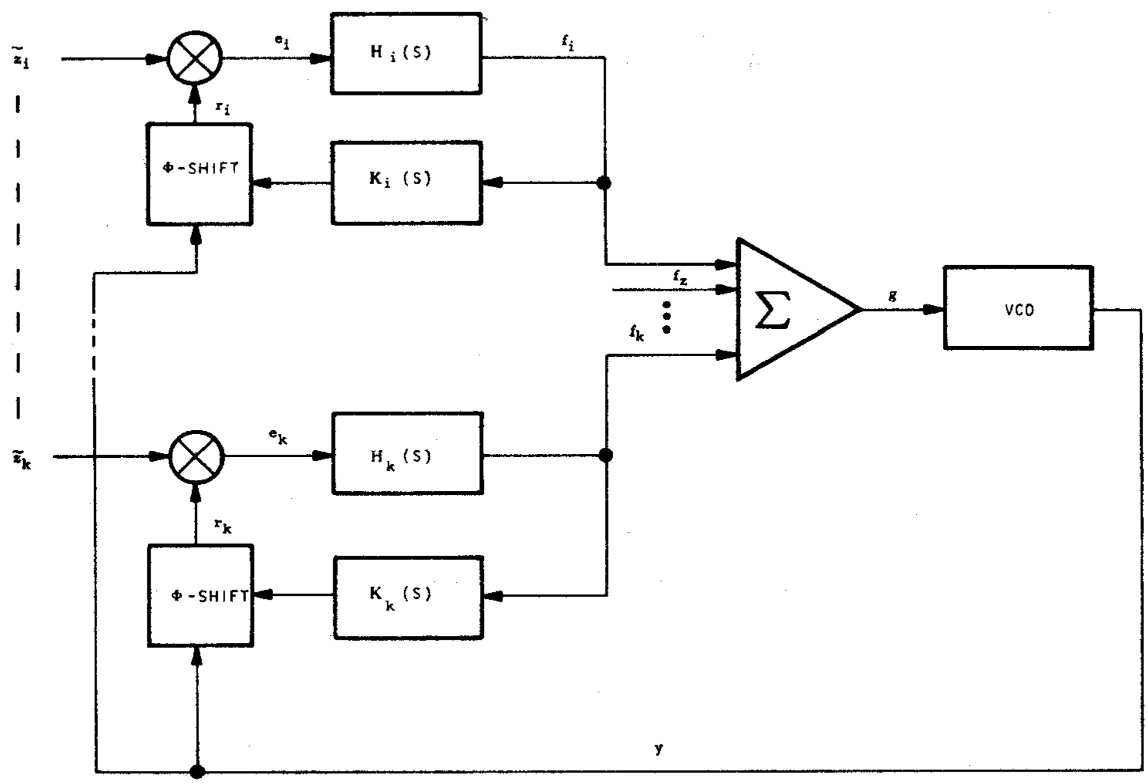

Fig. 7 - Multi-track Bit Synchronizer

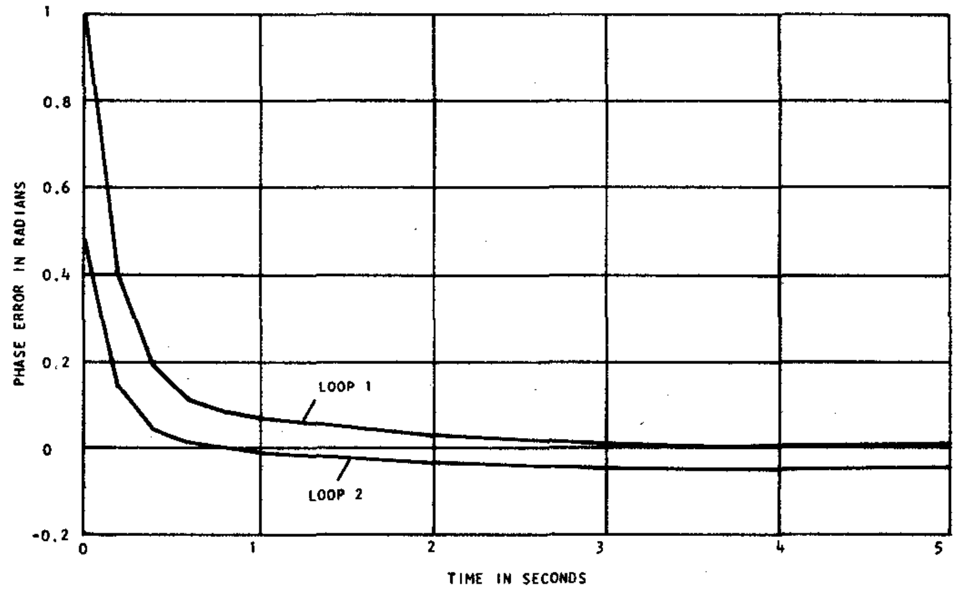

Fig. 8 - Response of Individual Loops with Integral Feedback

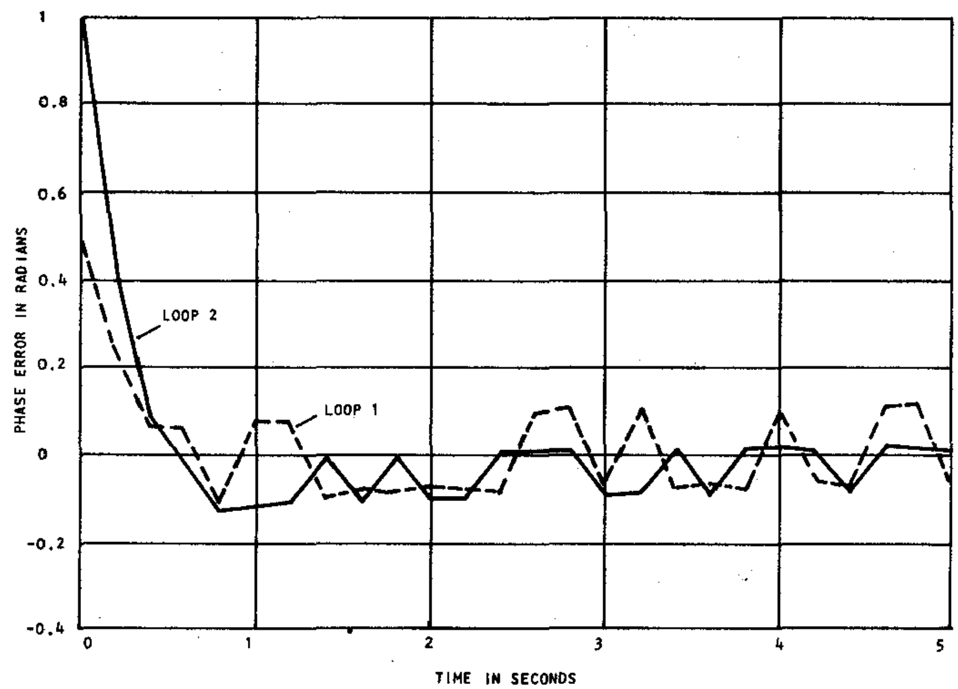

Fig. 9 - Phase Error in Loops 1 and 2 with a 0.1 Radian Quantization