Research Report

81-1

THE VITERBI RECEIVER FOR CORRELATIVE ENCODED DIGITAL FM

BY

P.J. McLane

Final Report Part One

D.S.S. Contract OSU80-00110

MARCH 1981

Queen's University at Kingston
Department of Electrical Engineering

THE VITERBI RECEIVER FOR CORRELATIVE ENCODED DIGITAL FM

by

P.J. McLane

Research Report 81-1
Final Report Part One
D.S.S. Contract OSU80-00110

Department of Electrical Engineering
Queen's University
Kingston, Ontario, Canada

March 1981

The research reported herein was supported by the Communications Research Center of the Canadian Department of Communications under D.S.S. Contract OSU80-00110 and by Grants from the Natural Sciences and Engineering Research Council.

THE VITERBI RECEIVER FOR CORRELATIVE ENCODED DIGITAL FM

P.J. McLane

ABSTRACT

The Viterbi Receiver for correlative encoded digital FM guarantees that the sequence error probability is minimal. Herein the receiver is completely specified. A state selection technique is presented which gives the minimal number of state variables for the realization of the Viterbi Algorithm which is the "heart" of the Viterbi receiver. A coherent version of this receiver is then analyzed for bit error rate using the transfer function bound technique. In this work we concentrate on the MSK and duobinary MSK modulation formats. We show that duobinary MSK is no more sensitive in bit errors than MSK. From earlier studies, it is known to have much better spectral efficiency than MSK.

Table of Contents

THE VITERBI RECEIVER FOR CORRELATIVE ENCODED DIGITAL FM

P.J. McLane

1. INTRODUCTION

Adam Lender introduced the concept of correlative encoding in data transmission in 1963 [1,2]. Recently a nice chapter on correlative encoding by Lender appeared in the textbook by Feher [3]. Actually Lender's introduction [1] of correlative encoding concerned digital FM. Since then it has been widely applied in many digital modulation formats. Pasupathy has written a nice tutorial paper [4] on correlative encoding techniques. Deshpande and Wittke [5,6] have done an extensive study on the use of correlative encoding in digital FM to achieve good bandwidth occupancy. Actually our work is a continuation of that of Deshpande and Wittke as some demodulators are developed for the modulation signals they have introduced.

Correlative encoding has application in constant envelope modulation techniques. It is used in the celebrated tamed FM modulation technique [7]. As pointed out by Galko and Pasupathy [8], the encoding polynomial used in [7] is optimal in the sense of bandwidth efficiency in the class of first and second order encoding polynomials. This is also clear from the earlier work in [5,6] but the connection to

tamed FM was not made clear.† Correlative encoding has also found application in non-constant envelope carrier modulations. For instance, it has been incorporated in a high-rate digital radio relay design by Anderson and Barber [9].

Correlative encoding is just one technique in the quest to achieve modulations that are jointly bandwidth and power efficient. That is, the modulations that have good bandwidth occupancy and a good bit error rate. In correlative encoding the input source bits to the modulator are correlated by an encoder. The idea is to have the transmitted signals correlated in time and thus hopefully reduce their bandwidth. A reasonable measure for bandwidth occupancy of a signal is its 99% power bandwidth. That is, the frequency band that contains 99% of the power of the modulated signal. In this way, only 1% of the power will be radiated out-of-band to interfere with other signals in a multi-user, frequency-division-multiplexed, radio transmission environment. In other applications, a different definition of bandwidth may be practical and will lead to the selection of different efficient modulations than those selected in radio applications.

Correlative techniques have so far led to good bandwidth occupancy in digital FM. However, their power efficiency is not as good as that obtained with other FM modulations. For instance, Aulin and Sundberg [10,11] and Anderson and Taylor [12]

† Reference [25] presents a nice explanation of tamed FM modulation and presents some applications for this modulation technique.

have introduced more complicated modulations that are more power efficient than even QPSK but have a smaller band efficiency than the best correlative encoded digital FM modulations of [5,6]. A nice table for comparing bandwidth and power efficiencies of digital FM modulations is given in [5,6].

Less research has been carried out on receivers for some of the efficient modulation formats introduced over the past few years. The receiver for MSK was presented by deBuda [13]. This receiver exploits the inherently differential nature of MSK modulation. A similar situation exists for tamed FM [7] whose receiver used many of deBuda's ideas. For tamed FM, however, this receiver may not be optimal. Our work concentrates on Viterbi Algorithm based receivers for correlative encoded digital FM and this receiver minimizes the sequence error probability. Earlier Schonhoff et al [14] developed this receiver for multi-level digital FM but without correlative encoding. Schonhoff tested his receiver by simulation. Our approach is to use the transfer function bound approach used for working out the bit error rate of low constraint length, convolutional codes in the recent text by Viterbi and Omura [15, p. 227-252]. (It is interesting that, when our analysis for digital modulations is compared with that for convolutional codes, in reality, only the "distance" measure changes. In this sense coding and modulation are really equivalent concepts.) Earlier Omura and Jackson [16] and Aulin [17] have used the transfer

function bounding technique to assess the bit error rate of various modulation formats. This technique was originally introduced by Viterbi [19]. It was also used by Forney [20] to work out the bit error rate of correlative encoded PAM, a problem earlier considered by Kobyashi [21].

In summary, our aim is to completely specify the Viterbi Receiver for Duobinary MSK and to analyze its performance using transfer function bounding methods. Other higher order, correlative polynomials for digital FM could be considered but both the specification and analyses becomes complicated. Some sort of reduced memory receiver [22] seems to be in order for these cases.

2. MAXIMUM LIKELIHOOD SEQUENCE ESTIMATOR

A block diagram for the generation of a continuous phase, FM signal is shown in Fig. 1. A representation of the output signal is

$$x(t) = A \cos(2\pi f_c t + 2\pi\Delta\int_{-\infty}^{t} m(\tau)d\tau) \qquad (1)$$

where A is the carrier amplitude, f_c is the carrier frequency, Δ is the peak deviation and m(t) is the message. We shall assume coherent system operation throughout this work so that any phase offset terms for x(t) are ignored. The message signal is

$$m(t) = \sum_{k=-\infty}^{\infty} d_k \, p(t-kT) \qquad (2)$$

where d_k is a correlative encoding of the data bits, a_k and p(t) is a rectangular pulse of unit height and duration T. For instance, for duobinary encoding, the encoding polynomial is $f(D) = (1+D)/2$, and $d_k = (a_k+a_{k-1})/2$ in (2). The data rate is 1/T bps as we shall consider only binary signalling where $a_k = \pm 1$.

Note from (1) that the signal energy is

$$E_b = \int_0^T s^2(t)dt = \frac{A^2T}{2} .$$

Furthermore, if f_cT is an integer, we can combine (1) and (2) to give the chip waveform

$$x_k(t) = \sqrt{\frac{2E}{T}} \cos(2\pi f_c t + \frac{\pi h}{T}(t-(k-1)T)d_k + \theta_{k-1})$$

$$\text{for } (k-1)T \leq t \leq kT \qquad (3)$$

where $h = 2\Delta T$ is the modulation index and $x(t)$ is constructed from putting all the chip waveforms together contiguously in time. Note that h is the ratio of the one-sided peak deviation, Δ, to the Nyquist bandwidth, $1/2T$.

Equation (3) can be used to specify a phasor state diagram for MSK which is the case $d_k = a_k$, i.e., no correlative encoding and $h = 0.50$. This is presented in Fig. 2 and will be used in our analyses. The phase-tree diagram for MSK is shown in Fig. 3. MSK has four possible phase states and the phases are connected by straight lines of slope $\pi/2T$. The corresponding diagrams for duobinary MSK, i.e., $2\,d_k = a_k + a_{k-1}$, are given in Figs. 4 and 5. Comparing Figs. 2 through 5 it is clear that duobinary MSK will have better bandwidth occupancy than MSK as the phase variation is less and in both cases the phases are continuous and connected with straight lines. However, the phasor state diagram is more populated in the duobinary case resulting in lower minimum free distance,

$$d_f = \frac{1}{2E_b} \min_{\substack{(i,j) \\ i \neq j}} \int_0^T [s_i(t) - s_j(t)]^2 dt, \qquad (4)$$

and hence higher bit error rate. In (3) E_b is the common energy of any two signals $s_1(t)$ and $s_2(t)$ selected from the phase trees in Figs. 3 and 5. A simple calculation gives $d_f = 2$ for MSK and $d_f = 1.73$ for duobinary MSK. This is a loss in dB of 0.63 dB for duobinary MSK relative to MSK.

However, Deshpande and Wittke [5,6] give the bandwidth efficiency, R/B b/s/Hz, of MSK as 0.85 and that of duobinary MSK as 1.09 where R = 1/T is the data rate and B is the 99% RF bandwidth. Thus, duobinary MSK is less power efficient but more band efficient, than MSK.†

The channel we consider will be the additive white Gaussian noise channel. Since the energy is the same for all signals in the phase tree, the log likelihood ratio test for deciding which signal in the tree was actually received is to maximize the correlation

$$\psi = \sum_{k=1}^{\infty} \int_{(k-1)T}^{kT} r(t)\, s_{k\ell}(t)\,dt \tag{5}$$

where $s_{k\ell}(t)$ is ℓ-th waveform in the phase tree for $(k-1)T \leq t \leq kT$. Since this functional is additive

$$\psi_k^* = \max_{\theta_{k-1} \to \theta_k} \left\{ \psi_{k-1}^* + V(\theta_k, \theta_{k-1}) \right\} \tag{6}$$

where ψ^* is the optimal value of ψ, the path metric is

$$V(\theta_k, \theta_{k-1}) = \int_{(k-1)T}^{kT} r(t)\, s_{k\ell}(t)\,dt \tag{7}$$

and the optimization at stage k is performed over all initial phases θ_{k-1} that can reach θ_k. For instance, in MSK θ_k could be $\pi/2$ and then the communicating states, θ_{k-1}, are π and 0 as shown in Fig. 2. For 0, we must have $a_k = d_k = 1$

† Reference [26] compares the spectra for duobinary MSK, QPSK and MSK. From these results, duobinary has a passband similar to QPSK and a much better out-of-band rolloff than MSK. Thus it should be a good modulation in a restricted bandwidth application, a property not enjoyed by MSK as its passband is wider than QPSK. It would be interesting to see if duobinary MSK is more robust to a restricted bandwidth than the non-constant envelope scheme of [27] whose error performance has been presented in [28].

to reach $\pi/2$ and for π we must have $a_k = -1$. The signals $s_{k\ell}$ are then derived from (3) and are marked in Fig. 3 in the phase tree for MSK.

The Dynamic Programming equation (6) is just a representation of the Viterbi Algorithm for finding the most likely received sequence of data bits. It might appear [6] that 4 states are required for MSK and 8 states for duobinary MSK. Actually only half this number are required for the Viterbi Receiver. We show this in the next section. Also, the basic forms of the correlation waveforms, $s_{k\ell}(t)$ in (7) must be specified. We so specify these waveforms also in the following section.

3. STATE-SELECTION PROCEDURE

The Viterbi Algorithm must operate with as few states as possible. Indeed, in most practical applications [22], some sort of reduced state operation is necessary. At first glance it may appear [6] that MSK requires 4 states (see Fig. 2) and duobinary MSK requires 8 states. Actually only half of this number is required.

Consider first the case for MSK. In Fig. 2, if state i and j can communicate, but a 1 in row i and column j of its connection matrix, A. Such matrices are used to specify graphs in graph theory. For MSK the connection matrix is shown in Fig. 6. Clearly only two sets of rows are unique. Thus states (1,3) and (2,4) can be used as states in alternating baud intervals. Note that (1,3) can only communicate with (2,4) and vice-versa. Consequently the trellis for Viterbi Algorithm based detection in MSK is as shown in Fig. 7. Of course, the initial state at time 0 must be known but this is simply handled in a short training period for the detector. It is a simple exercise to show that the two-state trellis in Fig. 7 can have merges and a 4-state trellis, using all four states in Fig. 2 cannot.

The result in the previous paragraph could have been deduced without resorting to the connection matrix. However, the case for duobinary MSK is not so straight forward. The connection matrix for this case is shown in Fig. 8 and it is

based on the phasor state diagram in Fig. 4. For state selection choose any two states with the same row with all their non-zero entries to the left of the line of symmetry. This gives states (1,4). For the remaining two states choose states with identical rows and having all non-zero entries to the right of the line of symmetry. This gives states (5,8). Then states (1,4,5,8) can be used at one baud time and the remaining states (2,3,6,7) at the next. The resulting trellis for this case is shown in Fig. 9. Once again, the state for system start-up must be known, but this is easily established in a short initialization procedure.

The basis of the Viterbi receiver for the two cases considered is the trellis diagrams in Figs. 7 and 9. Note that we have gone from a state diagram to a trellis diagram to specify the detector. This is the same as that done in setting up decoders for short constraint length convolutional codes (see Viterbi and Omura [15, pp. 227-235].) All that remains to specify the detector is an algorithm for computing the path metrics for, say, the two possible communicating states into state 2 in Fig. 9. Namely from states 4 and 1. This is computed from the correlation integral in (7). Considering all (two) state transitions into given states and choosing the maximum represents a solution to the functional equation in (6), and the whole procedure is the Viterbi algorithm. The procedure for computing the path

metrics which, following the parlance of Schonhoff et al [14], is called the branch metric calculator, will be specified below in terms of some basic correlations.

4. PATH METRIC CALCULATOR

We will work out the path metric calculator for duobinary MSK. This calculator for MSK is easily derived using the procedure presented below.

The path metric calculator is needed for the correlation integral in (7) which in turn is needed for the path metrics in the Viterbi Algorithm solution of (6). The general form of the correlating signal, $s_{k\ell}(t)$, in (7) is given in (3). In (3) we set $d_k = (a_k+a_{k-1})/2$. Thus $d_k = \pm 1$ or 0 and hence there are 3 instantaneous frequencies associated with $x_k(t)$ in (3). They are $f_c + \Delta$, $f_c-\Delta$ and f_c where $\Delta = 1/4T$ as $h = 2\Delta T = 0.50$. Using signal space concepts we know that in-phase and quadrature carriers form a basis for the signal space of all carrier signals of fixed frequency. Thus every path metric should be obtainable from an in-phase and quadrature correlation system for each of our 3 instantaneous frequencies. This is as shown in Fig. 10. An integrate-and-dump correlation system is shown.

It still remains to specify the path metrics in terms of the correlator outputs in Fig. 10. This is the path metric calculator diagram. The trellis for the Viterbi detection of duobinary MSK is shown in Fig. 9. The two classes of state transactions are labelled as "type A" and "type B". The path metrics, in terms of the correlator outputs from Fig. 10, are shown in Fig. 11. Also these metrics for type B

transitions are shown in Fig. 12. To illustrate the process, the path metric for the state transition 1→2 and 4→2 will be derived. Referring back to Fig. 4 one notes that the state transition 1→2 is specified by having $a_{k-1} = 1$, $a_k = -1$, or $d_k = o$ and $\theta_{k-1} = 0$. Using this in (3) for h = 1/2 we find that[†]

$$s_{k\ell}(t) = \cos(2\pi f_c t).$$

Thus from (7) and Fig. 10, the path metric is d_c. Similarily for the state transition 4→2, $a_{k-1} = -1$, $a_k = -1$, giving $d_k = -1$, and $\theta_{k-1} = \pi/2$. Thus in (3)

$$s_{k\ell}(t) = -\sin[2\pi f_c t = \frac{\pi}{2}(t-(k-1)T)]$$

$$= -\sin(2\pi(f_c-\Delta)t)$$

as $h = 2\Delta T = 0.50$. Hence the path metric is $-d_{-s}$ in terms of the notation in Fig. 10. All the other path metrics in Figs. 11 and 12 are derived in a similar manner.

It is worth noting that as h = 1/2 the carrier phase and symbol timing signals can be easily derived from the input to the Viterbi receiver depicted in Fig. 10. This follows deBuda's idea [12,24] of squaring and filtering the input signal to get a digital FM signal at frequency $2f_c$ with h = 1. This is then Sunde's FSK and for duobinary spectral lines will occur at frequencies $2f_c + \Delta'$ and $2f_c - \Delta'$ where $\Delta' = 2\Delta = 1/2T$. Thus the symbol timing clock is derived from a signal at the difference of these two

[†]The amplitude of the signal in (3) has been removed in specifying the correlator signals.

frequencies. The carrier can be derived from a signal at the sum of these two frequencies. The carrier signals at $2f_c + \Delta'$ and $2f_c - \Delta'$ are also needed for the receiver in Fig. 10. Duobinary MSK also has an instantaneous frequency at f_c. Thus, as Rhodes [26] observed, the squaring and filtering operations mentioned above produce a biphase carrier at $2f_c$. This follows as duobinary MSK can have a QPSK component as staying at the same phase is possible from baud to baud. The carrier signal can then also be recovered from this biphase signal using standard techniques.

Consequently timing and carrier recovery is quite simple for MSK and duobinary MSK. This mainly follows on using deBuda's idea mentioned above and this is a consequence of having $h = 1/2$. DeBuda's idea is also used in the receiver for tamed FM [7].

We now consider the bit error rate analysis for a perfectly coherent version of the Viterbi Receiver illustrated in Fig. 10.

5. TRANSFER FUNCTION BOUNDS: LINEAR MODULATIONS

As mentioned earlier a nice analysis using transfer function bounds to obtain the bit error rate for Viterbi decoding of convolutional codes is given in Chapter 4 of the book by Viterbi and Omura [15]. From equation (4.5.8) of this book the bit error rate for a Viterbi decoder is upper bounded by

$$P_b \leq Q\left(\sqrt{\frac{E_b d_f}{N_o}}\right) \exp\left(\frac{E_b d_f}{2N_o}\right) \frac{\partial T(D,L,I)}{\partial I} \Bigg|_{\substack{D = Z \\ L = I = 1}} \qquad (8)$$

where d_f is the free distance, or minimum distance for the code, $Z = \exp(-E_b/N_0)$, $N_0/2$ is the spectral height for the AWGN channel and $T(D,L,I)$ is the transfer function between states in the trellis where error paths rejoin the true path after some arbitrary diversion point from the true path. This is applied to the modulations we consider by just using Euclidean distance instead of say, Hamming distance, for a discrete memoryless channel and some coding technique

Consider the trellis for MSK as shown in Fig. 7. As the modulation is linear we can compute P_b for any sequence and it is the overall P_b for the receiving system. In Fig. 7 the "dotting" sequence is taken as the transmitted sequence.

The state diagram for paths initially leaving the true path and joining it at some later time is shown in Fig. 13. Here L is the generic variable for the path segment, I present denotes a bit error and the exponent of D denotes the distance between the true and erroneous signal for the path segment L.

The distances for the path segments are computed using the Euclidean distance in (4) and the signal representation in (3). For instance for the transition 1→4

$$s_1(t) = \sqrt{\frac{2E}{T}} \cos\left(2\pi f_c t + \frac{\pi}{2}\frac{t}{T}\right)$$

and

$$s_2(t) = \sqrt{\frac{2E}{T}} \cos\left(2\pi f_c t - \frac{\pi t}{2T}\right)$$

which both follow from the phasor state diagram in Fig. 2. From (4) the distance for this path diversion is unity and the corresponding power of D in Fig. 13 is one. Thus in Fig. 13 the minimum distance path has $d_f = 2$. Each other path cycles at least once in the self loop at node b. This corresponds to not crossing back over to the true path in the trellis in Fig. 7. The error state graph represents all such paths.

It is quite easy to solve for T(D,L,I) in Fig. 13. Namely,

$$\varepsilon_b = D^2 L \varepsilon_b + DLI$$

giving

$$\varepsilon_b = \frac{DLI}{1-D^2L}$$

and finally

$$T(D,L,I) = DLI\varepsilon_b$$

$$= \frac{D^2L^2I^2}{1-D^2L} \qquad (9)$$

Now expanding the denominator in (9) as a geometric series there follows

$$T(D,L,I) = D^2L^2I^2 + D^4L^3I^2$$
$$+ + D^{2k}L^{k+1}I^2 + +$$

Thus the minimum distance path has length 2, has 2 bit errors and distance 2. The kth error path has distance 2k, length k+1 and also contains just two bit errors. These facts are easily verified from the trellis in Fig. 7.

To do our bit error rate analysis only (9) is required. Clearly,

$$\frac{\partial T(D,L,I)}{\partial I} = \frac{2ID^2L^2}{1-D^2L} .$$

Substitution of this result into equation (8) and simplification produces the following upper bound for P_b:

$$P_b \leq 2Q\left(\sqrt{\frac{2E_b}{N_0}}\right)/[1-\exp(-E_b/N_0)] \qquad (10)$$

which was also given by Aulin (17,18]. The factor of 2 denotes the 2 bit errors for any diverging path in the trellis for MSK. Also the denominator is insignificant for most SNR's. It contributes a factor of 2 when E_b/N_0 is as small as $10 \log_{10}(\log 2)$ dB = -1.5 dB. The union bound agreement used in upper bounding P_b will weaken the bound at a higher SNR than this. At a high SNR the bit error rate is $2\,Q(\sqrt{E_b/N_0})$ which is easily predicted from the trellis. Until an improved theory is developed, the SNR where the bound becomes weak must be found by simulation.

6. TRANSFER FUNCTION BOUNDS: NONLINEAR MODULATIONS

In section 5 we derived transfer function bounds for a linear modulation technique. A major simplification with such modulation methods is that the bit error rate is the same no matter what the transmitted sequence. This is not the case for nonlinear modulations. Omura and Jackson [16] and Aulin [17,18] have presented state diagrams based on error function bounds from these diagrams. These diagrams actually present a representation of error sequences based on averaging over all input data sequences. Often, however, these diagrams are quite complicated and for error analysis require computer assisted numerical operations on the transition matrix for these graphs. Even the error graphs themselves are not easily specified and probably some automated approach should be developed for constructing them. We show, at least for duobinary MSK, that error graphs can be simplified and closed form transfer function bounds can be derived from these simplified error state graphs.

As an introduction we first repeat the analysis for MSK based on the error state graph technique [16,17,18]. From Fig. 2 it is clear that error is just one in phase and that it varies between 0 and π radians. This is true for any input sequence. Thus the error graph is as shown in Fig. 14. The distances for the exponents of D in this graph must be worked out using Fig. 2, the signal representation in (3)

and the Euclidean distance formula in (4). Clearly the error graphs of Figs. 13 and 14 are identical, as the modulation is linear, and hence the transfer function is as given previously in (9). Thus the error bound to the bit error rate is as in (10).

We now consider a nonlinear modulation, duobinary MSK. We note that Aulin [18] gives another example in his thesis by considering continuous phase FSK with h = 2/3 and no correlative encoding.

The first thing that must be worked out is to decide if the modulation is nonlinear. A nice test for this is as follows. If one can find a transmitted sequence that does not have a minimum distance error path, the modulation is nonlinear.† This assertion is easily verified. Recall that the bit error rate is completely specified by the expression $Q(\sqrt{d_f E_b/N_o})$ for large enough E_b/N_o. However, if our assertion is not true it is specified by $Q(\sqrt{dE_b/N_o})$ with $d > d_f$ and this is a contradiction. It follows that duobinary MSK is nonlinear since the all one's transmitted sequence has a smallest distance error path of distance $3-2/\pi = 2.7$ and $d_f = 1.727$ for this modulation method. The latter point is easily verified by computing the Euclidean distance between the two signal paths in the phase tree in Fig. 5.

†This test was suggested to the author by Stan Simmons of Queen's University.

The test for nonlinearity of a modulation given above suggests an error bounding method. Namely, choose a sequence with a minimum distance error path and then base a transfer function bound for this path and its neighbours by following the techniques in section 5. The doting sequence, 101010 etc., is such a sequence for duobinary MSK. If this is done for the doting sequence one finds that the nearest neighbour in distance to the minimum distance path has distance 4. The trellis for this case is given in Fig. 16. One notes that the smallest distance path in Fig. 15, i.e. for the all one's sequence, is missed in trellis of Fig. 16. Such are the consequences of nonlinearities.

The error rate state graph for duobinary MSK is shown in Fig. 17. This graph is constructed from all bit error sequences that could occur in Fig. 4, the state diagram for this modulation technique. The Euclidean distances between error path is determined, as before, using the signal representation in (3) and the distance measure in (4). The trellis diagrams in Figs. 15 and 16 are an aid in constructing this graph.

Recall from our earlier comments that an error analysis based on Fig. 15 contains a minimum distance path but not a neighbouring path with distance closest to this. Fig. 16 contains this neighbouring path but not a minimum distance path. It would be nice to develop a graph that shows how these two error paths interact. This is available in the error state graph in Fig. 17. Any path through ε_e, i.e. phase

state π, has distance at least distance 4. Also the two smallest distance paths are 0, ε_a, ε_b, T(D,L,I) and 0, ε_a, ε_d, T(D,L,I), which have distances 1.727 and 2.7 respectively. Thus if we draw a graph with state ε_e removed, there is really no loss, as paths of distance at least 4 are deleted. With this deletion, our graph is simplified and closed form results are possible. The original graph in Fig. 17 is just too unwieldy and, in any case, dropping state ε_e represents no practical loss.

Our reduced error state graph is shown in Fig. 18. Only state ε_e from Fig. 17 has been dropped. In our reduced graph

$$T(D,L,I) = LD^x(\varepsilon_b + \varepsilon_d) \tag{11}$$

and from the graph in Fig. 17 the following equations follow:

$$2\varepsilon_a = D^xLI + DLI\varepsilon_b + D^xLI\varepsilon_d \tag{12}$$

$$2\varepsilon_b = DLI\varepsilon_a + D^yLI\varepsilon_c \tag{13}$$

$$2\varepsilon_c = D^xLI + DLI\varepsilon_d + D^xLI\varepsilon_b \tag{14}$$

and

$$2\varepsilon_d = DLI\varepsilon_c + D^yLI\varepsilon_a \tag{15}$$

Solving (12)-(15) for $(\varepsilon_b+\varepsilon_d)$ and substitution into (11) gives

$$T(D,L,I) = \frac{L^3I^2D^{2x}(D + D^y)}{2\left[1 - \frac{L^2I^2}{4}(D+D^x)(D+D^y)\right]} \tag{16}$$

Our bit error rate, P_b, analysis can now be based on (16) and the basic bound on P_b given in (8), which is taken from Viterbi and Omura [15]. We have

$$T(D,1,I) = \frac{(D^{2x+1}+D^{2x+y})I^2}{2[1-\phi(D)I^2]} \tag{17}$$

where

$$\phi(D) = \frac{(D+D^x)(D+D^y)}{4} \tag{18}$$

Differentiation of T(D,1,I) with respect to I and substitution into (8) yields, after some simple manipulations,

$$P_b \leq \frac{Q\left(\sqrt{\frac{d_f E_b}{N_0}}\right)\left(1 + \exp\left[-\frac{(2x+y-d_f)E_b}{2N_0}\right]\right)}{\left(1 - \phi\left[\exp\left(-\frac{E_b}{2N_o}\right)\right]\right)^2} \tag{19}$$

where $d_f = 1+2x = 1.727$ and x, y are defined in Fig. 17. Further simplification using the definitions of x and y give

$$P_b \leq \frac{Q\left(\sqrt{d_f \frac{E_b}{N_0}}\right)\left(1 + \exp\left[-\frac{E_b}{\pi N_0}\right]\right)}{\left(1 - \phi\left[\exp\left(-\frac{E_b}{2N_0}\right)\right]\right)^2} \tag{20}$$

with $d_f = 1.727$ and ϕ as defined in (18).

The result in (20) is our error bound for duobinary MSK. It includes the two shortest Euclidean distance paths for this modulation method. It thus, should be a good error estimate for moderate-to-high SNR. We now compare this with our corresponding result for MSK in (10).

The error rate bound in (20) for duobinary MSK and that in (10) for MSK are plotted for a range of SNR's in Fig. 19. At high SNR the degradation of duobinary MSK relative to MSK is 0.30 dB. It is not just the 0.63 dB loss in free distance due to the more complete error bounds given in (20) and (10). In fading environments where lower SNR's may occur the two schemes perform about the same. At some point for

low E_b/N_o it is easy to show that the error bound for duobinary MSK in (20) is twice that given for MSK in (10). However, at the outset in using (8) from [15] a union bound assumption has been made. For the resulting error bound to be reasonably accurate a moderate-to-high SNR is required. Thus the doubling in error rate noted above is probably not a practical reality.

7. CONCLUSIONS

It is clear from the foregoing that a Viterbi algorithm receiver for correlative encoded MSK can be clearly specified for first order encoding polynomials. Further, for these polynomials, the transfer function bounding method can be applied to give bit error rate estimates at moderate to high SNR levels. For these SNR values analysis shows that duobinary MSK is no more sensitive in bit error rate to SNR fluctuations than is MSK.

To consider FM modulation formats with higher order encoding polynomials one just repeats the methodology presented herein. For instance, this could be done for the tamed FM polynomial, $(1 + 2D + D^2)/4$. In addition, performance impairments such as phase coherence loss, sampling jitter, adjacent channel interference, nonlinear amplification (i.e., in satellite channels) and restricted bandwidth loss could be assessed using the transfer function bound technique. In many applications duobinary MSK looks very attractive as it has good spectral rolloff and a similar passband to QPSK. However, the relative worth of this modulation format will not be really known until the performance analyses noted above have been carried out.†

Acknowledgement

I wish to thank Gillian Woodruff and Jacques Vaisey of Queen's University for their independent derivations of the error state diagram for duobinary MSK given in Fig. 17.

†In addition, simple I and Q demodulators for duobinary MSK should be specified, analyzed and put in perspective with similar receivers for MSK [13,24] and tamed FM [7,25].

REFERENCES

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

2. A Lender, "Correlative Digital Communication Techniques", IEEE Trans. Comm. Tech., Vol. 12, pp. 128-135, December 1964.

3. K. Feher, "Digital Communications: Microwave Applications", Prentice Hall, Englewood Cliffs, New Jersey, 1981.

4. S. Pasupathy, "Correlative Coding: A Bandwidth Efficient Signalling Scheme", IEEE Comm. Soc. Magazine, Vol. 17, pp. 4-11, July 1977.

5. P.H. Wittke and G.S. Deshpande, "Band Efficient Digital Angle Modulation Signals", Dept. Elect. Engng., Queen's University, Kingston, Ontario, Canada, Research Report #80-1, August 1979.

6. G.S. Deshpande and P.H. Wittke, "Correlative Encoded Digital FM", IEEE Trans. on Comm., Vol. COM-29, pp. 156-162, February 1981.

7. G. deJager and C.B. Dekker, "Tamed Frequency Modulation: A Novel Method to Achieve Spectrum Economy in Digital Transmission", IEEE Trans. on Comm., Vol. COM-26, pp. 523-528, May 1978.

8. P. Galko and S. Pasupathy, "On a Class of Generalized MSK", to be presented, International Conference on Communications, June 14-18, 1981, Denver, Colorado.

9. C.W. Anderson and S.G. Barber, "Modulation Considerations for a 91 M bit/s Digital Radio", IEEE Trans. Comm., Vol. COM-26, pp. 523-528, May 1978.

10. T. Aulin and C.-E. Sundberg, "Continuous Phase Modulation: Part I(Full Response Signalling)", to appear, IEEE Trans. on Comm., March 1981.

11. ibid, "Continuous Phase Modulation: Part II (Partial Response Signalling)", to appear, IEEE Trans. on Comm., March 1981.

12. J.B. Anderson and D.P. Taylor, "A Bandwidth Efficient Class of Signal Space Codes", IEEE Trans. on Inform. Theo., Vol. IT-24, pp. 703-712, November 1978.

13. R. deBuda, "Coherent Demodulation of Frequency-Shift Keying with Low Deviation Ratio", IEEE Trans. on Comm., Vol. COM-20, pp. 429-435, June 1972.

14. T.A. Schonhoff, H.E. Nichols and H.M. Gibbons, "Use of the MLSE Algorithm to Demodulate CPFSK", Proc. Int. Conf. on Comm., pp. 25.4.1-25.4.5, Toronto, Canada, June 1978.

15. A.J. Viterbi and J.K. Omura, "Principles of Digital Communication and Coding", McGraw-Hill, New York, 1979.

16. J.K. Omura and D. Jackson, "Cutoff Rates for Channels Using Bandwidth Efficient Modulations", Proc. Nat. Telecomm. Conf., pp. 14.1.1-14.1.11, Houston, Texas, December 1980.

17. T. Aulin, "Viterbi Detection of Continuous Phase Modulated Signals", Proc. Nat. Telecomm. Conf., pp. 14.2.1-14.2.7, Houston, Texas, December 1980.

18. T. Aulin, "Three Papers on Continuous Phase Modulation", Ph.D. Theses, University of Lund, Sweden, 1979.

19. A.J. Viterbi, "Error Bounds for Convolutional Codes and an Asyptotically Optimum Decoding Algorithm", IEEE Trans. Info. Theo., pp. 260-269, Vol. IT-13, April 1967.

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

21. H. Kobayashi, "Correlative Level Coding and Maximum-Likelihood Decoding", IEEE Trans. Info. Theo., Vol. IT-17, pp. 586-594, September 1971.

22. P.J. McLane, "A Residual Intersymbol Interference Error Bound for Truncated State Viterbi Detectors", IEEE Trans. Info. Theo., Vol. IT-26, pp. 548-554, September 1980.

23. J.M. Wozencraft and I.M. Jacobs, "Principles of Communication Engineering", John Wiley and Sons, New York, 1965.

24. R. deBuda, "Fast FSK Signals and Their Demodulation", Canadian Elect. Engng. Journal, Vol. 1, pp. 28-34, 1976.

25. D. Muilwijk, "Tamed Frequency Modulation - A Bandwidth-Soning Digital Modulation Method, Suited for Mobile Radio", Phillips Telecomm. Rev., vol. 37, pp. 35-49, March 1979.

26. S.A. Rhodes, "FSOQ, A New Modulation Technique That Yields a Constant Envelope", Proc. Nat. Telecomm. Conference, pp. 51.1.1 - 51.1.7, Dec. 1980, Houston, Texas.

27. M.C. Austin and N.U. Chang, "Quadrature Overlapped Raised-Cosine Modulation", To appear, IEEE, Trans. on Comm. March 1981.

28. D. Divsalar and M.K. Simon, "Performance of Quadrature Overlapped Raised-Cosine Modulation over Nonlinear Satellite Channels", To appear, Proc. Int. Conf. on Communications, Denver, Colorado, June 14-16, 1981.

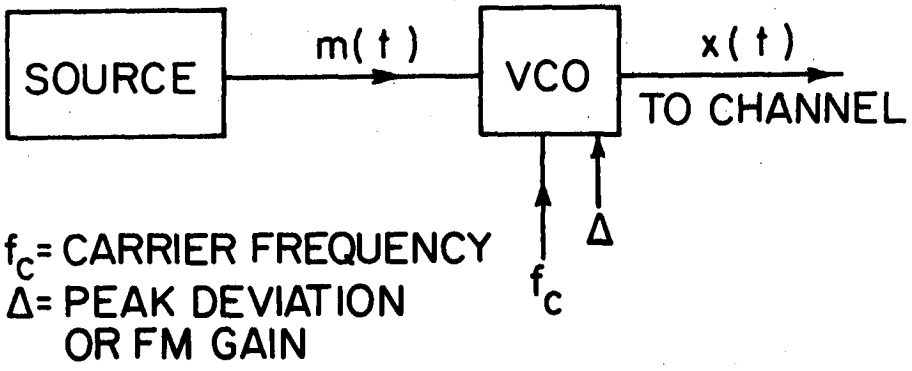

Figure 1 The continuous phase, FM modulator

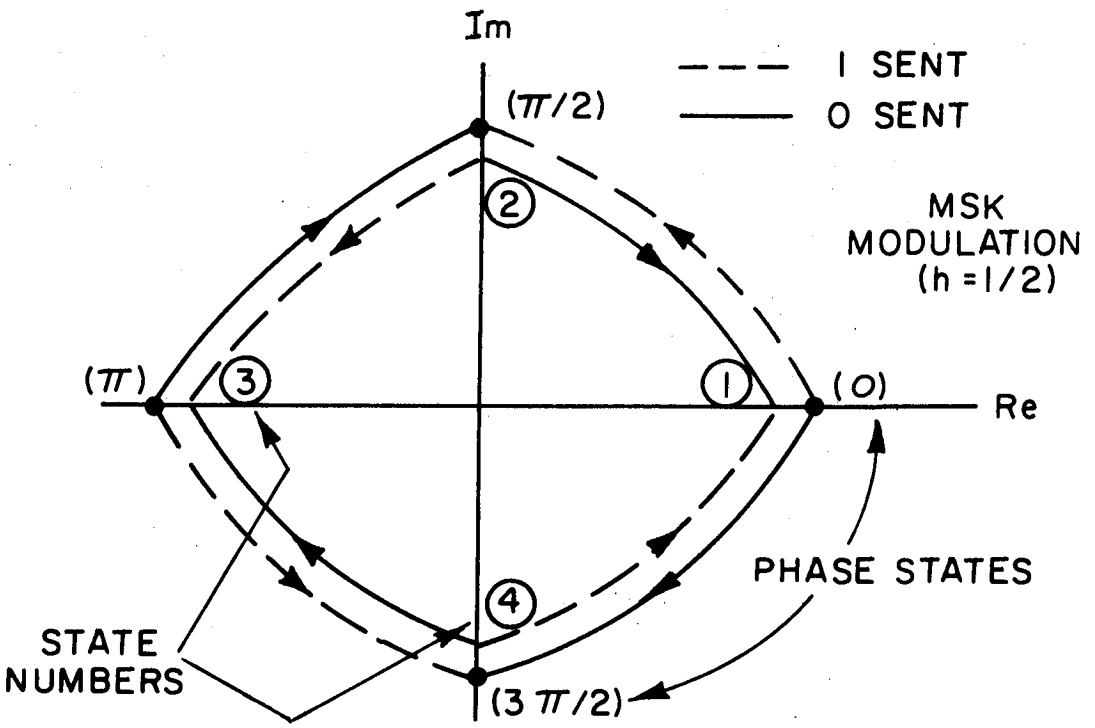

Figure 2 Phasor state diagram for MSK modulation

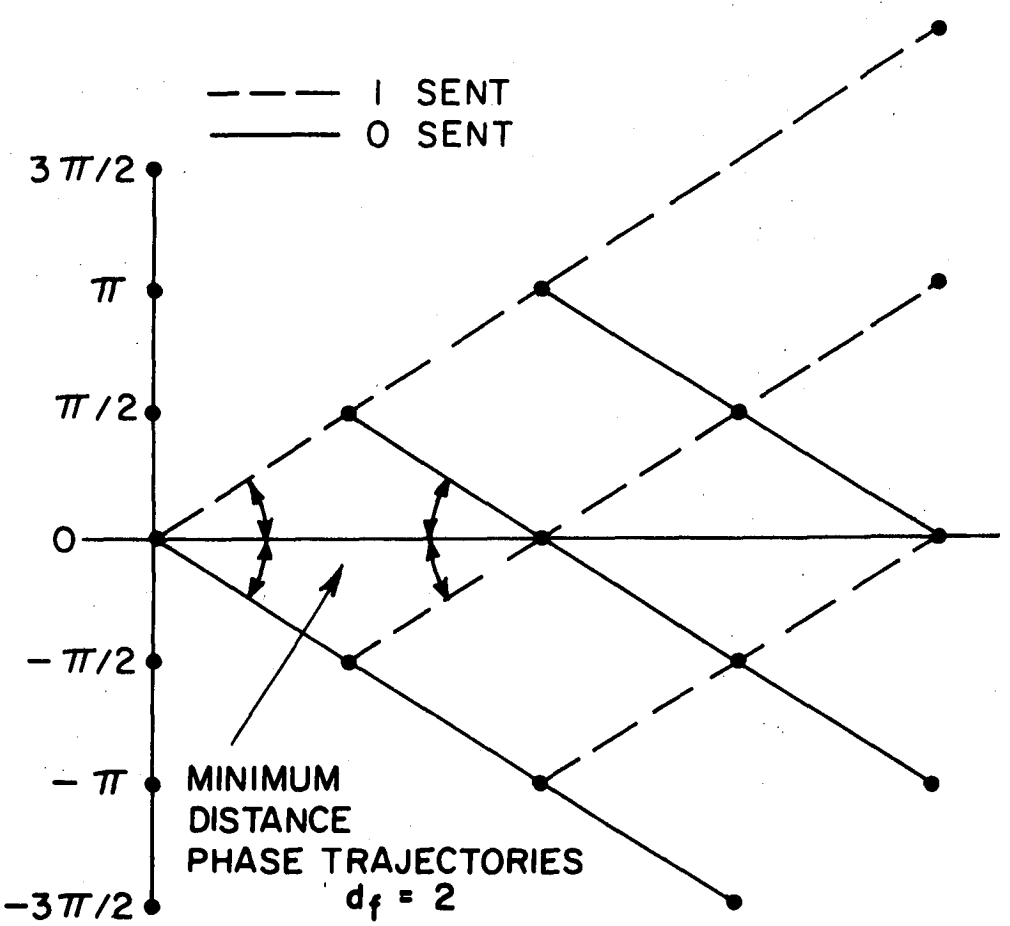

Figure 3 Phase tree for MSK modulation

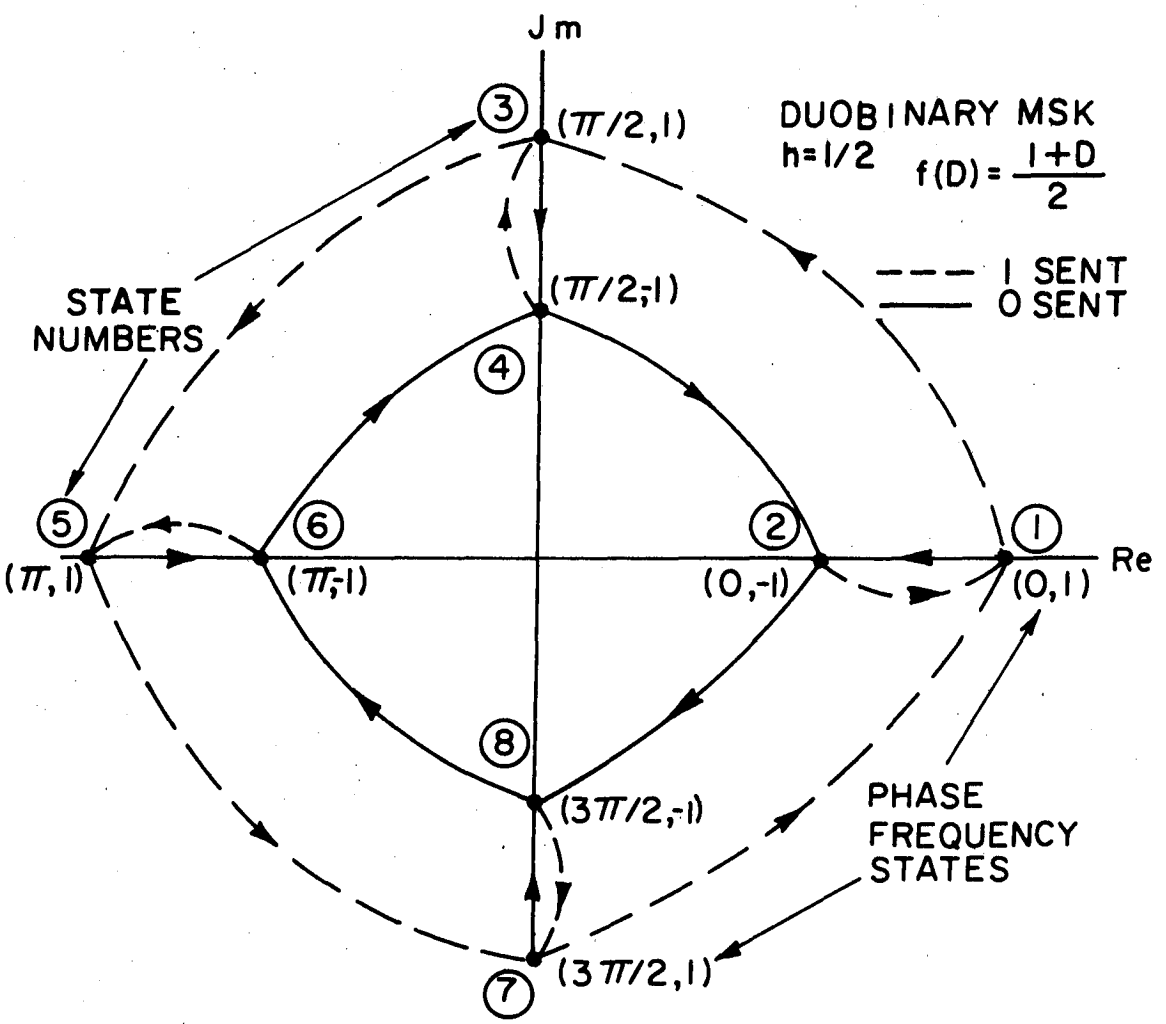

Figure 4 Phasor state diagram for duobinary MSK

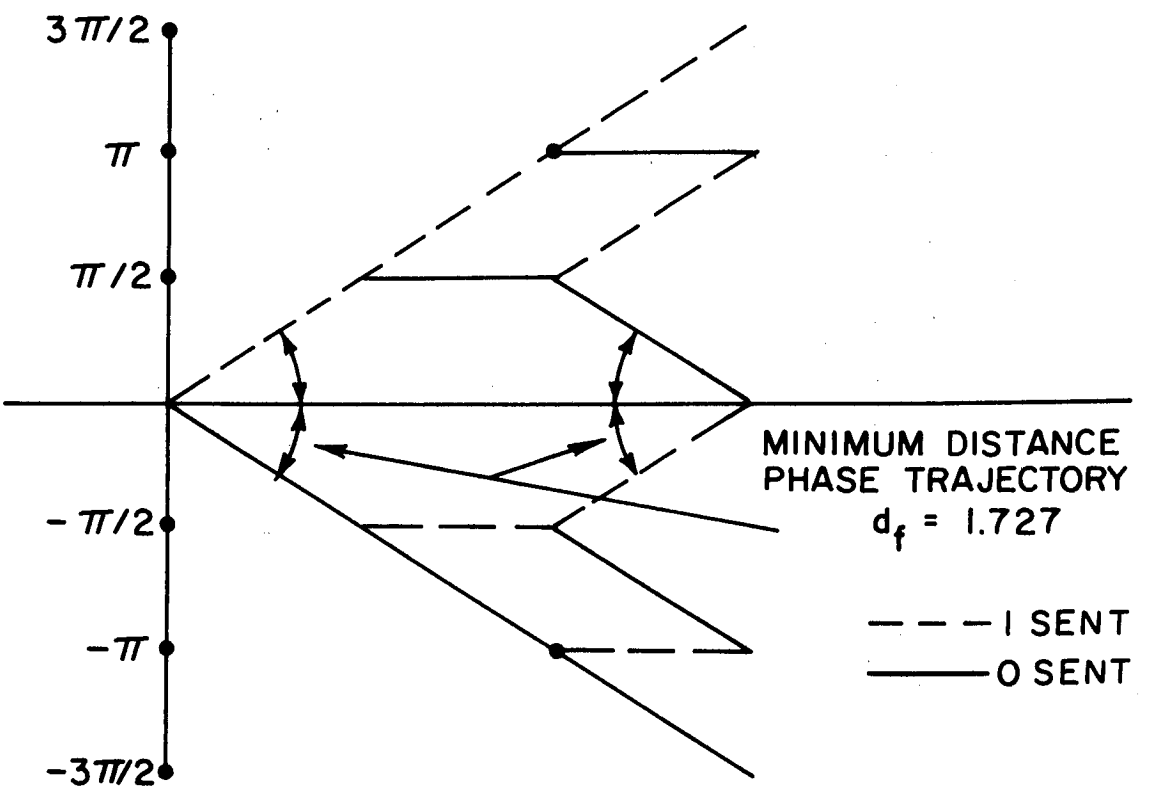

Figure 5 Phase tree for duobinary MSK

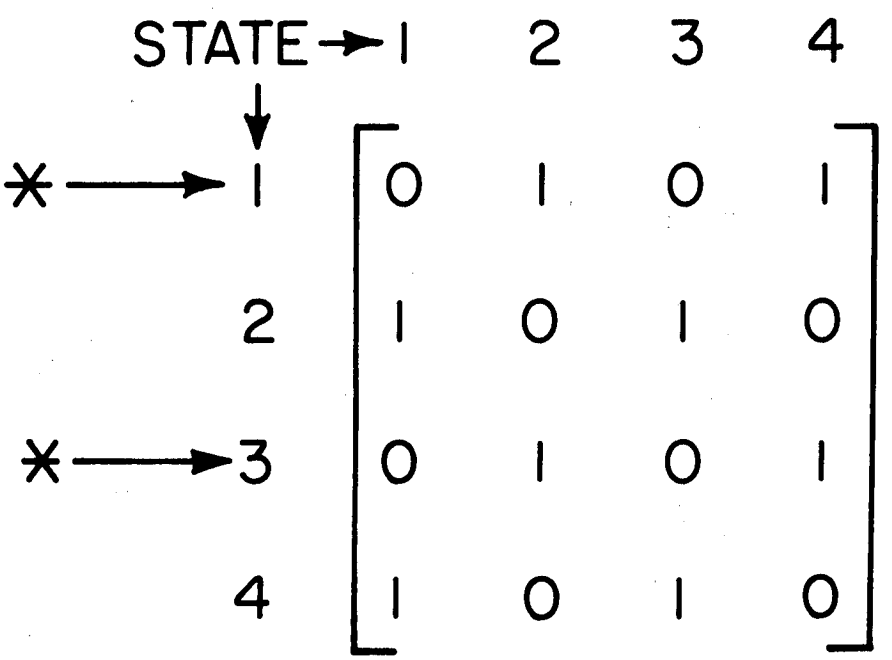

Figure 6 Connection matrix for state selection for MSK

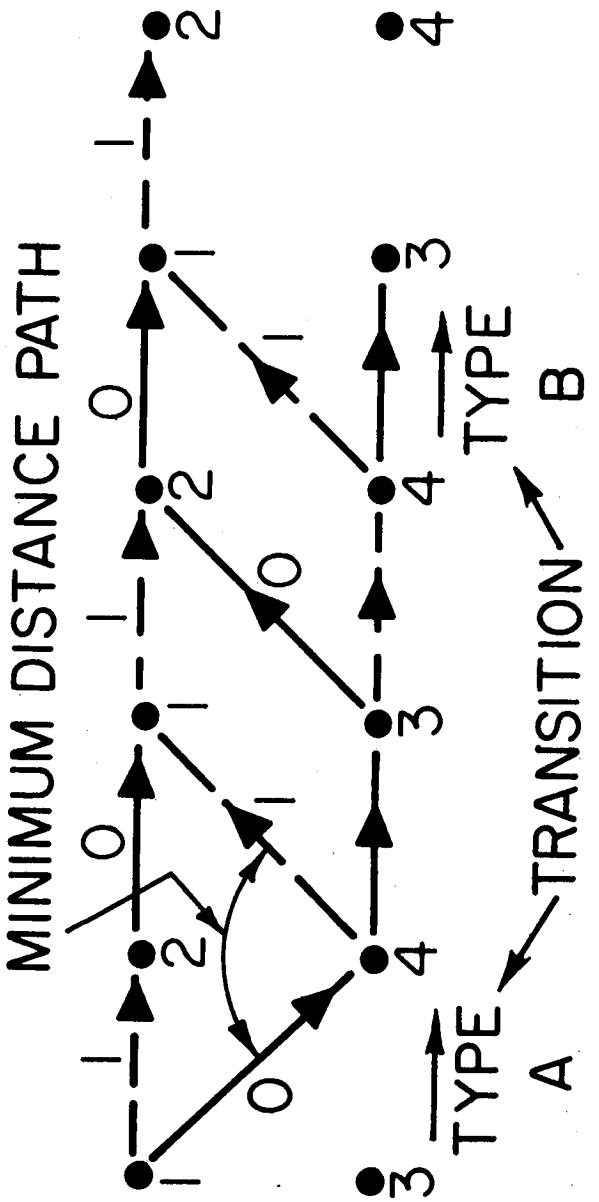

Figure 7 Trellis for MSK demodulation using the Viterbi detector

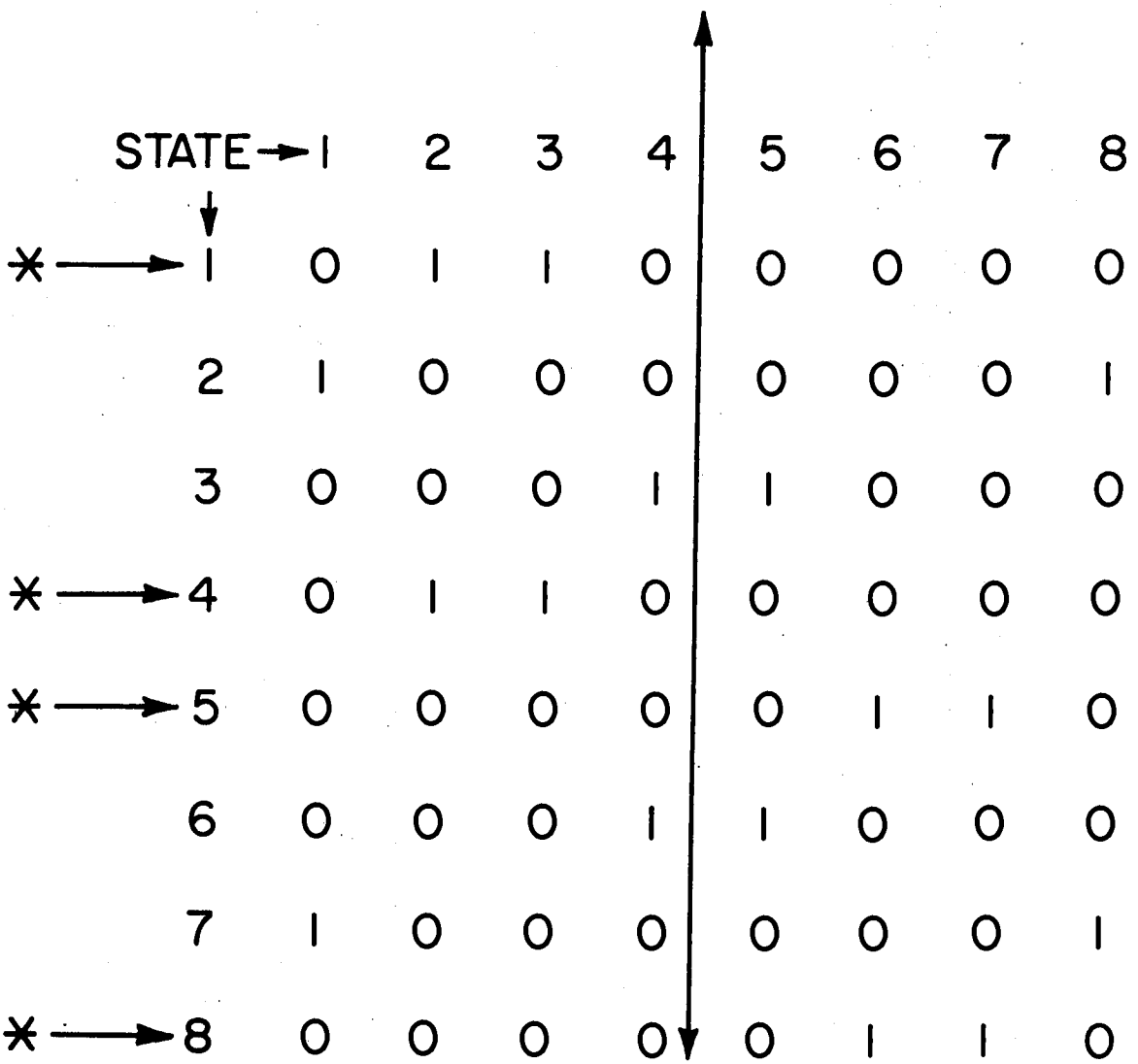

Figure 8 State connection matrix for duobinary MSK

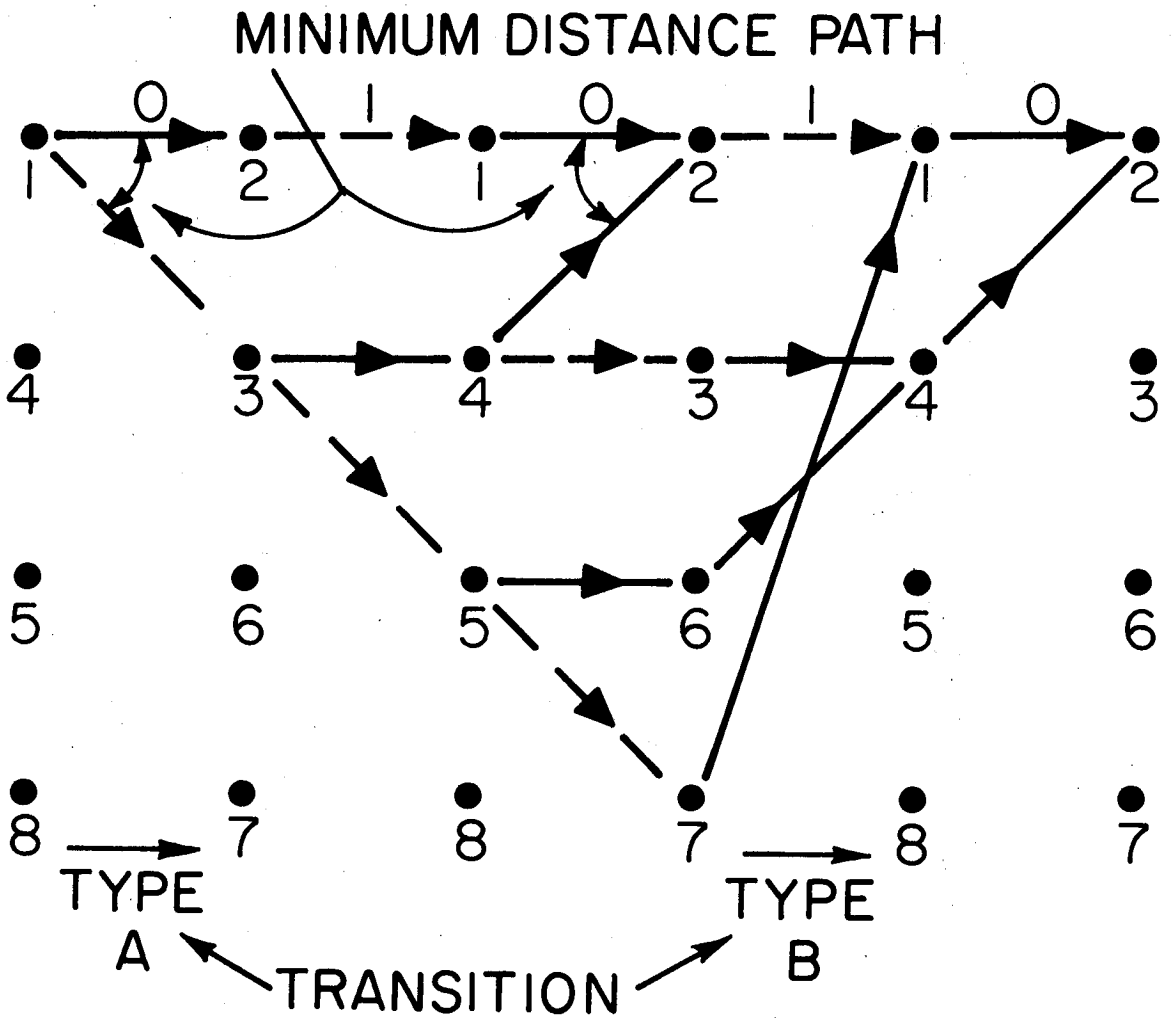

Figure 9 Trellis for duobinary MSK demodulation using the Viterbi detector

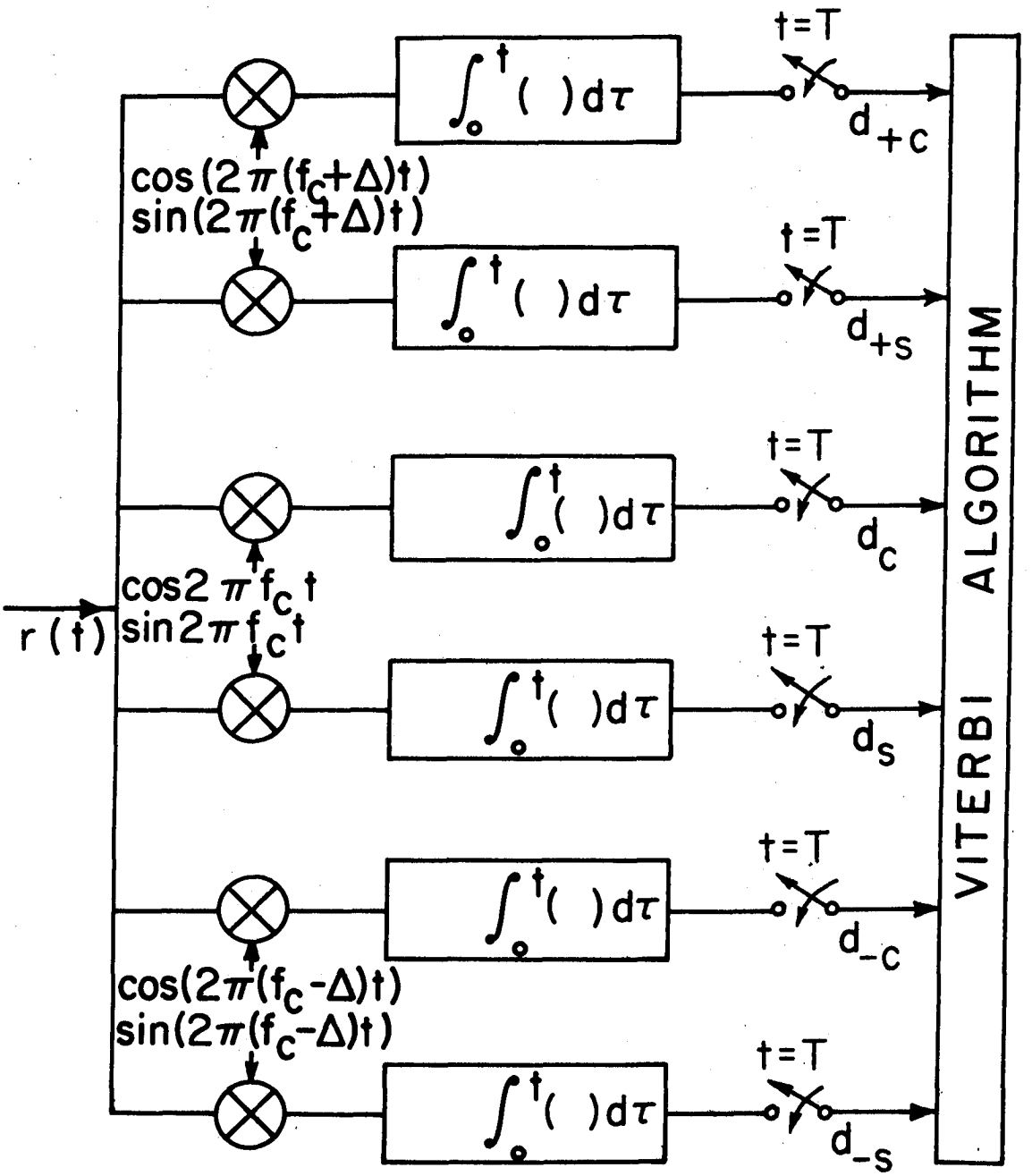

Figure 10 Receiver structure for the Viterbi Receiver

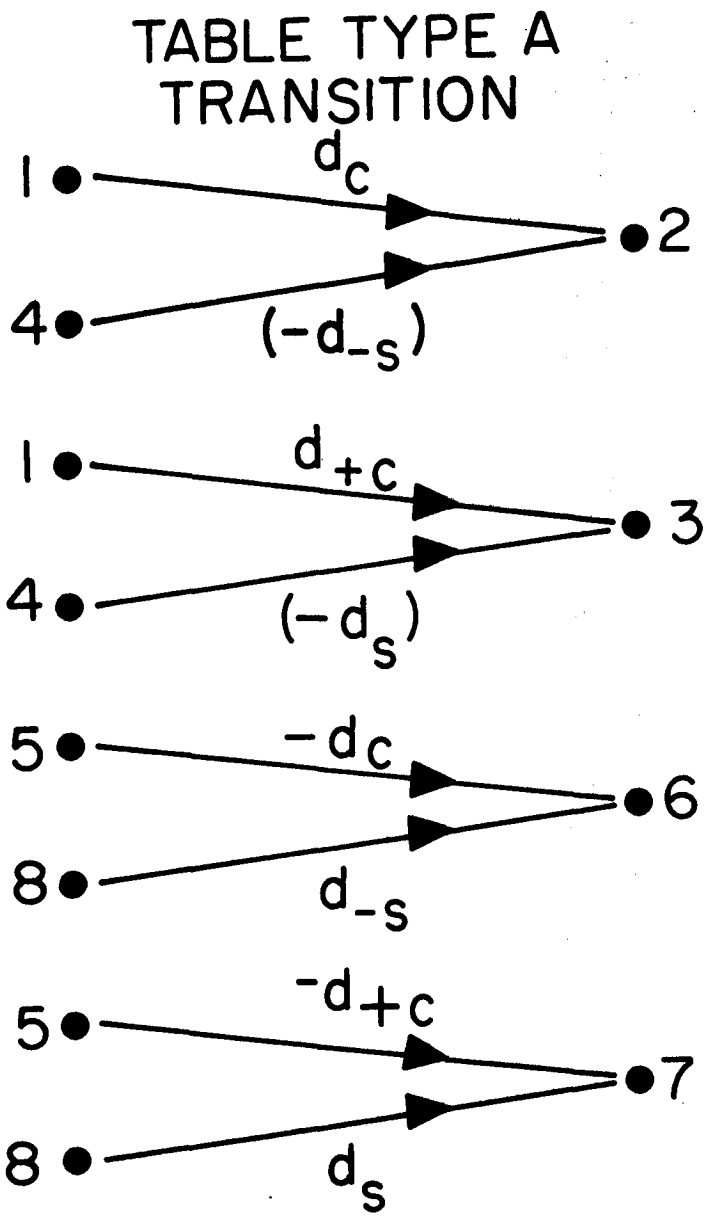

Figure 11 Path metrics for type A transitions in the trellis for duobinary MSK (see Fig. 9)

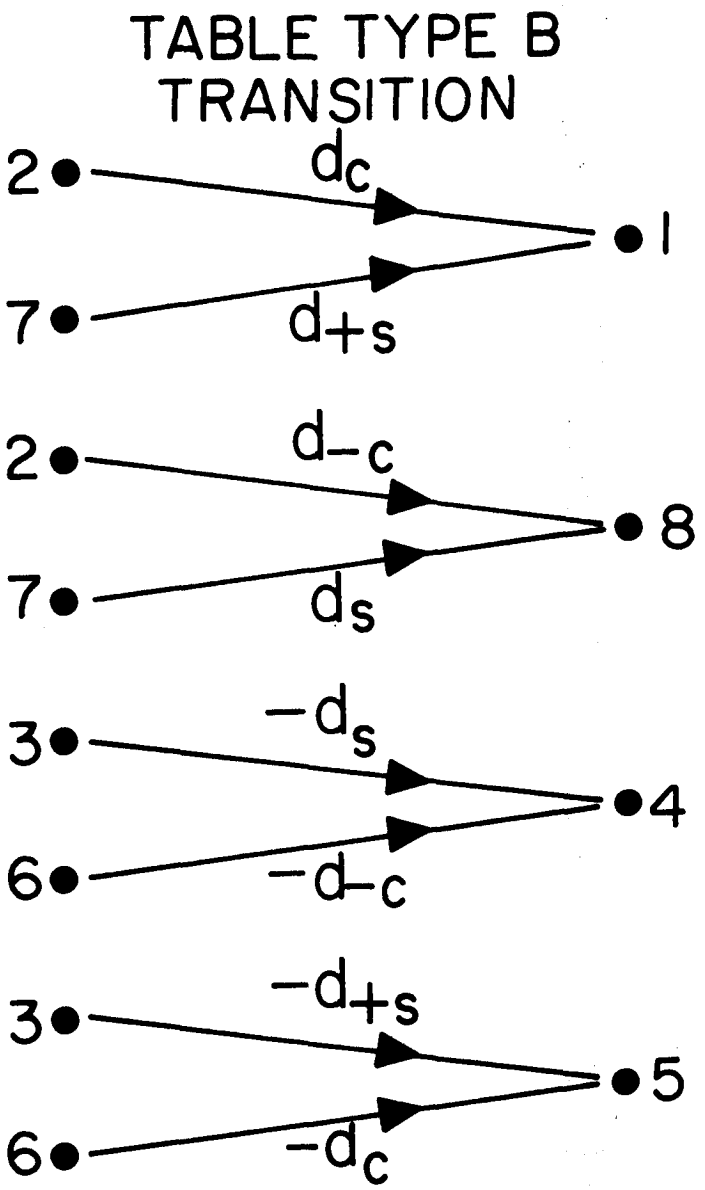

Figure 12 Path metrics for type B transitions in the trellis for duobinary MSK (see Fig. 9)

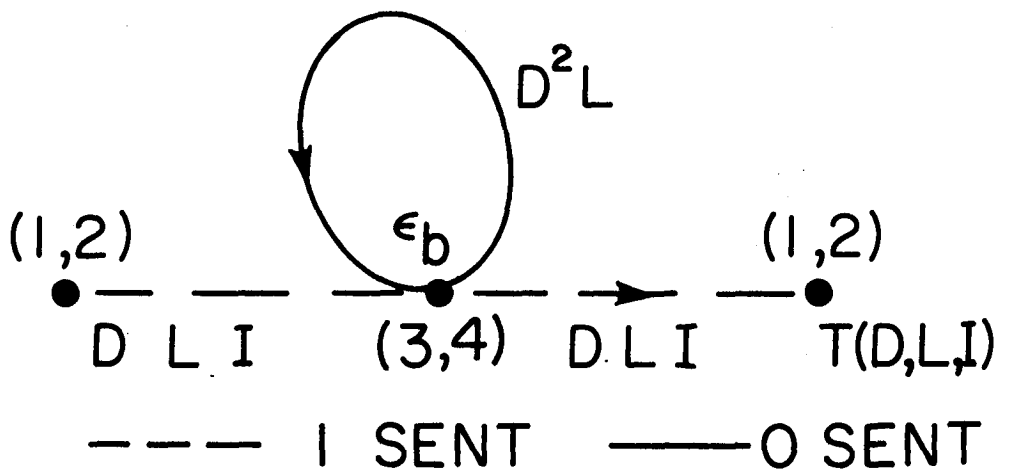

Figure 13 Error graph for MSK modulation and Viterbi reception

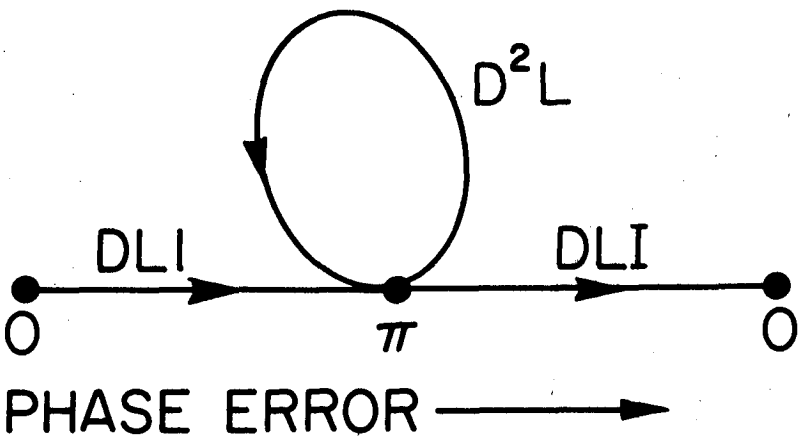

Figure 14 Error graph for MSK modulation based on an arbitrary transmitted sequence

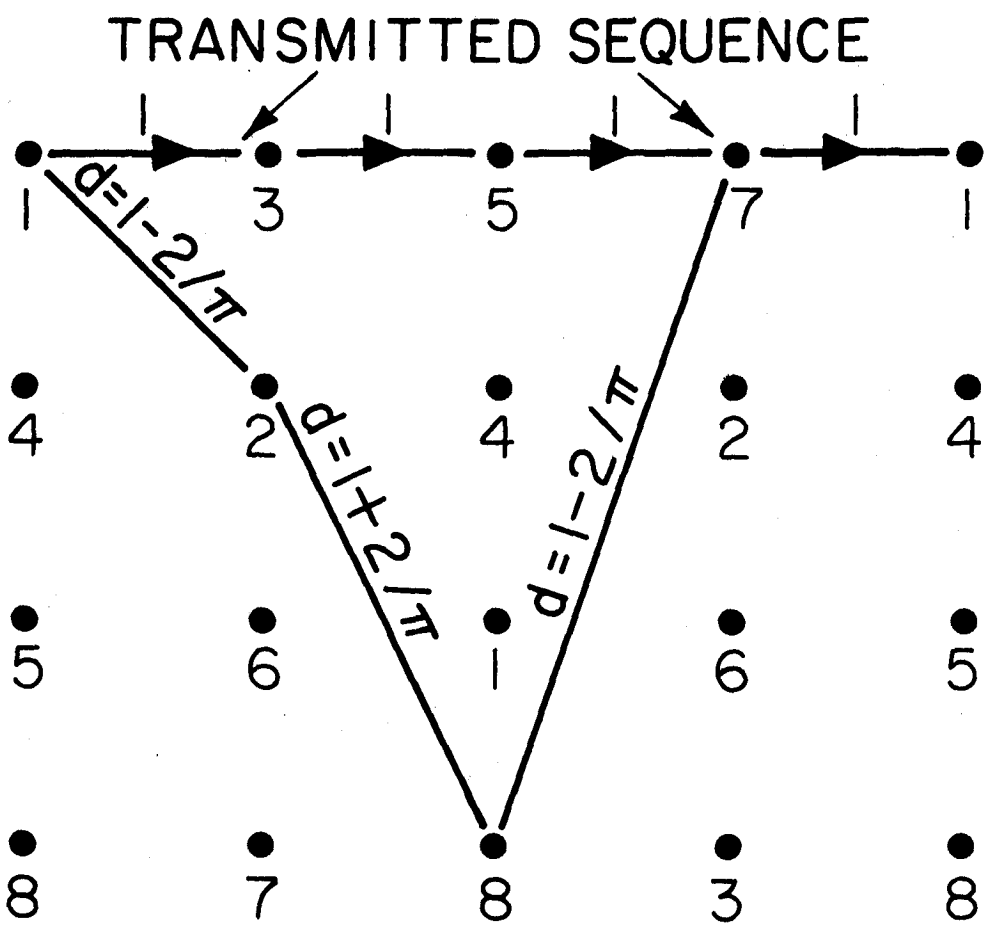

Figure 15 Trellis diagram for the all one's transmitted sequence for duobinary MSK

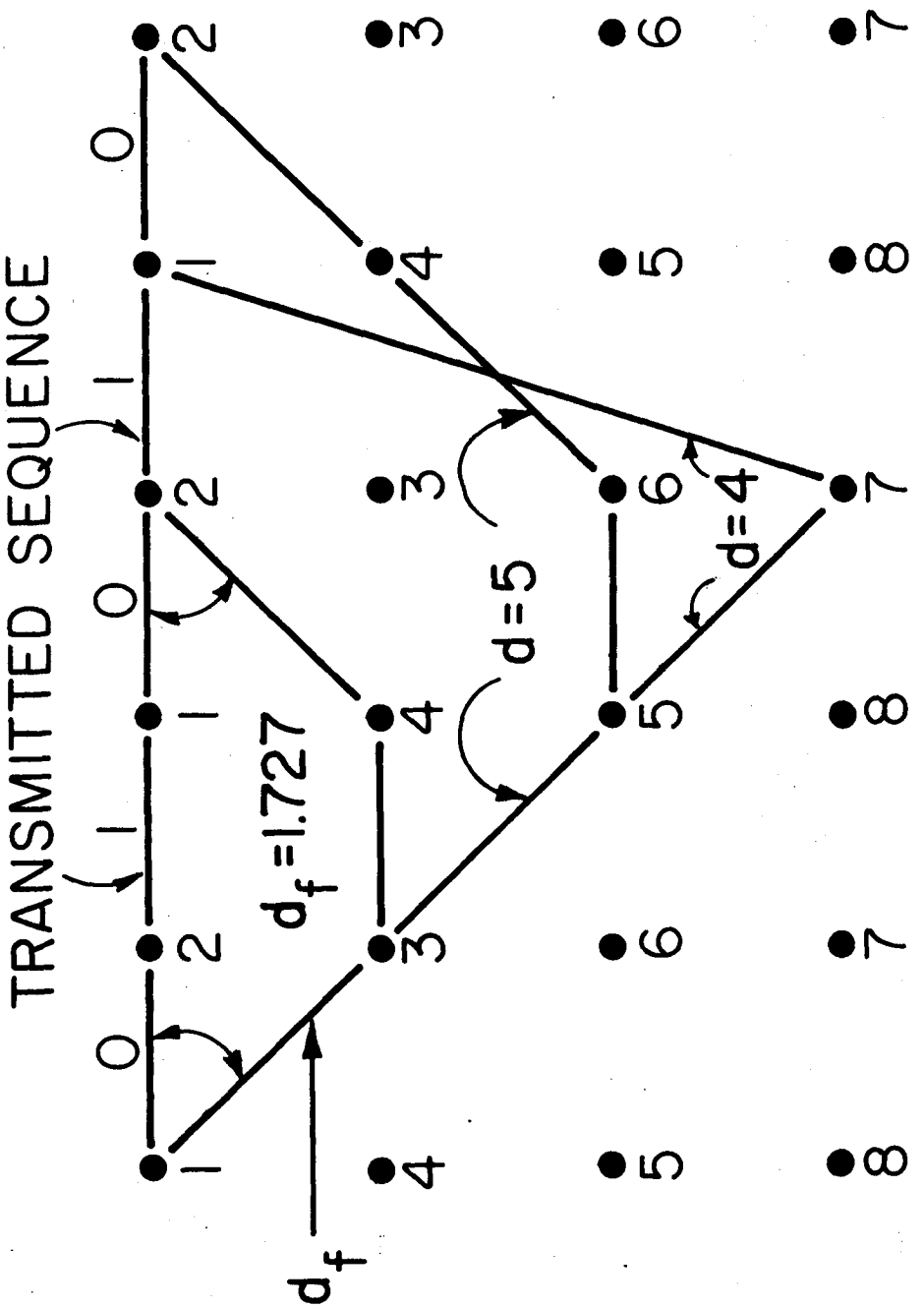

Figure 16 Trellis diagram for the transmitted sequence 0101010 etc. for duobinary MSK

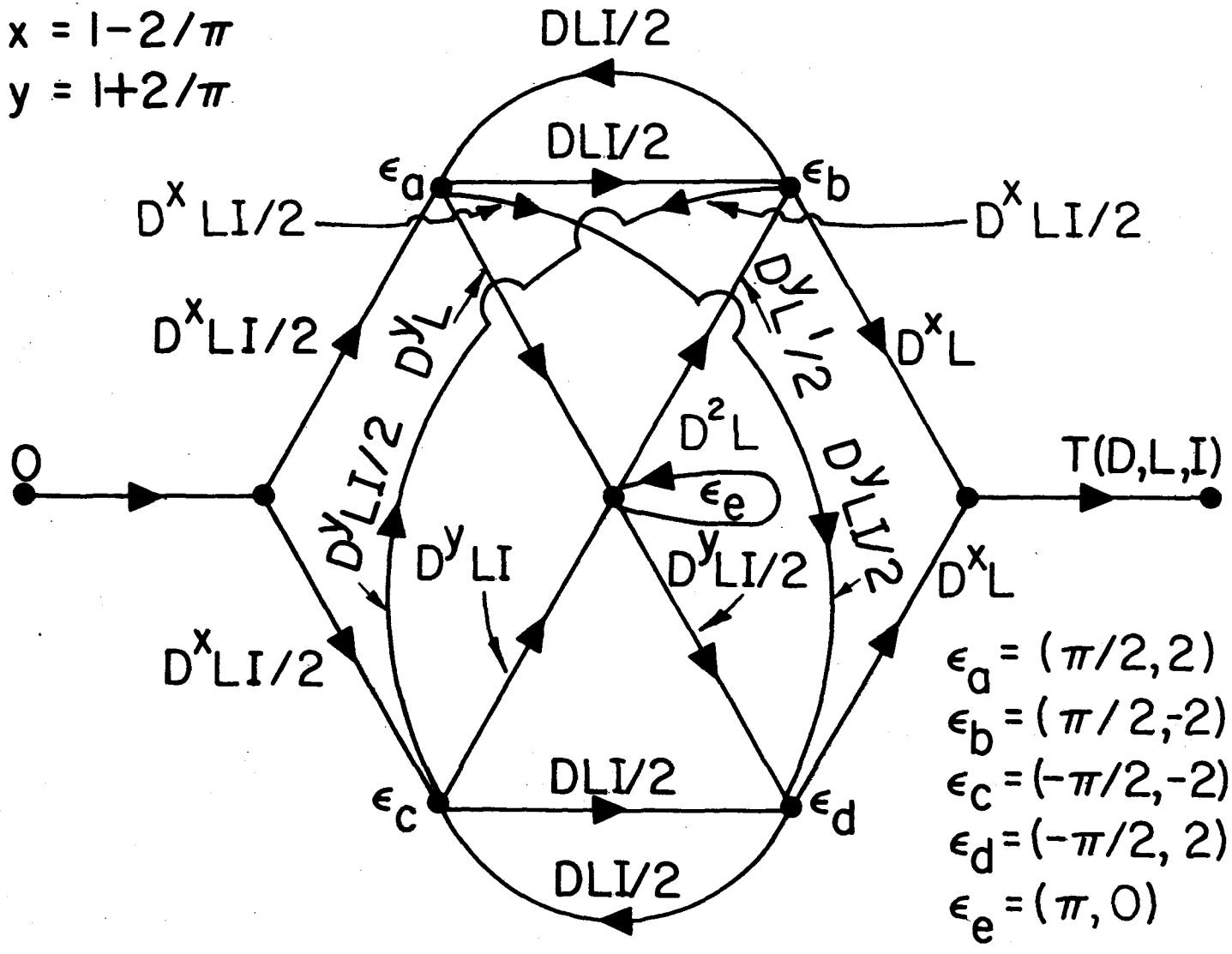

Figure 17 Error state graph for duobinary MSK

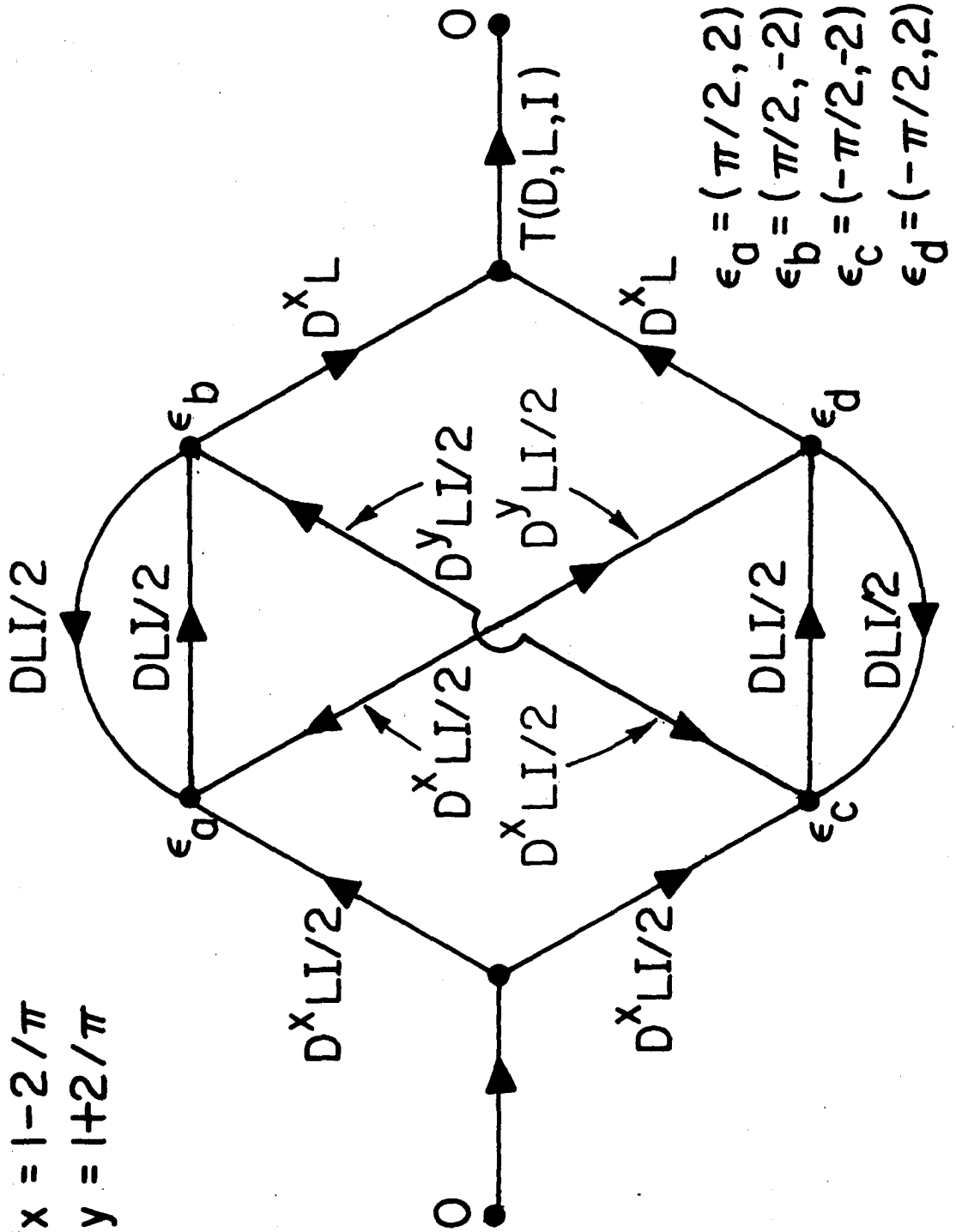

Figure 18 Reduced error state graph for duobinary MSK. Here we have dropped paths of least distance of 4 and 5.

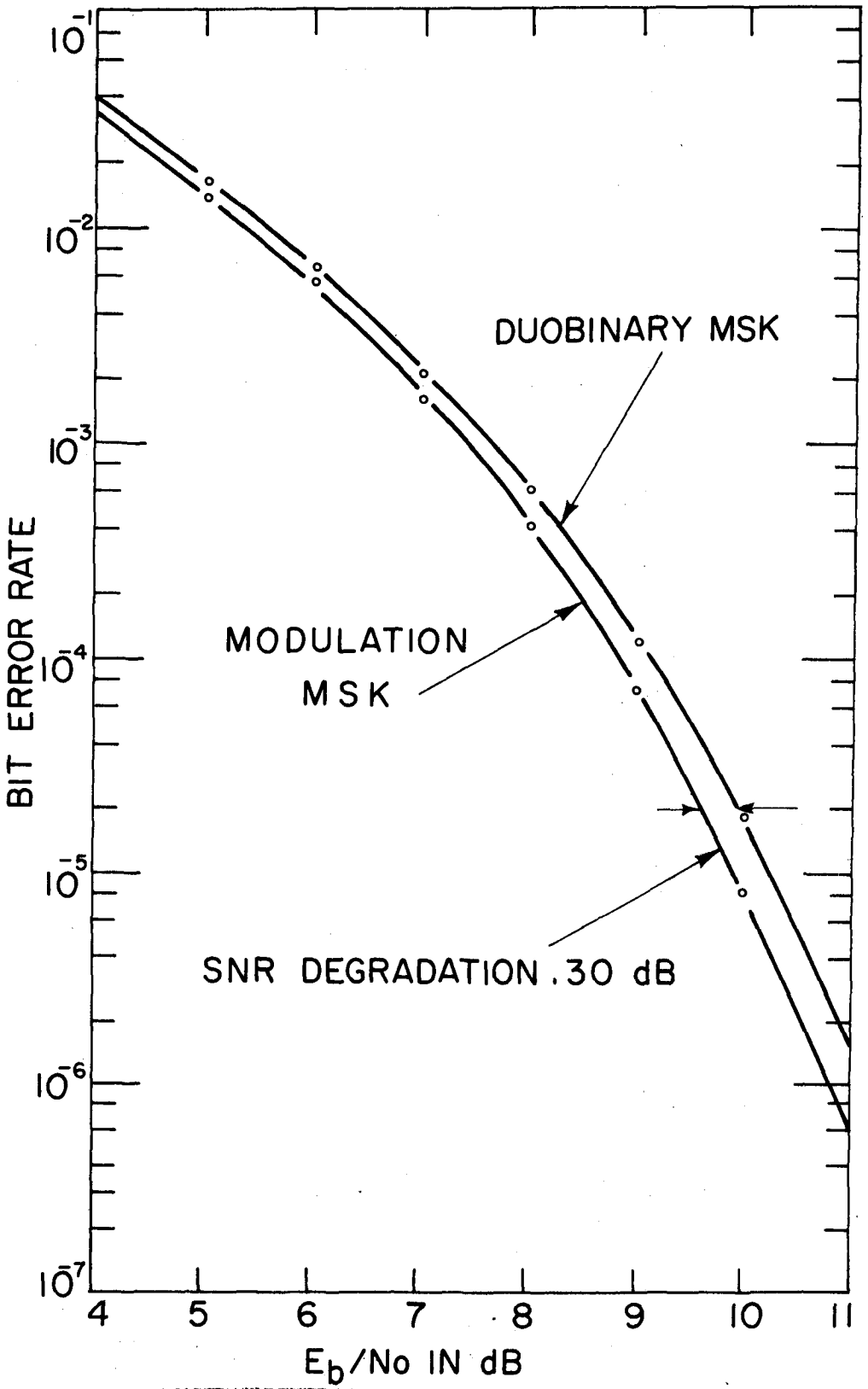

Figure 19 Bit error rate transfer function bound as a function of E_b/N_0 for MSK and duobinary MSK