Reprinted from: ICC'80 Conf. Record
pp. 26.5.1-26.5.5, Seattle,
June 8-12, 1980.

A GENERALIZED FORMULATION OF MINIMUM SHIFT KEYING MODULATION

JAMES L. MASSEY

Department of System Science, University of California
Los Angeles, CA 90024

ABSTRACT

A general binary modulator is considered consisting of a single-input two-output sequential transducer whose outputs $x(i)$ and $y(i)$ control the transmitted signal in the manner $s(t)=x(i)p(t)+y(i)q(t)$ for $iT \le t < iT + T$, where T is the modulation bit duration. The transducer operates on the ± 1 modulation sequence $\{d(i)\}$ in the manner $x(i)=a(i)[d(i)+d(i-1)]/2$ and $y(i)=b(i)[d(i)-d(i-1)]/2$ where $\{a(i)\}$ and $\{b(i)\}$ are arbitrary ± 1 sequences. It is shown that, provided only that in each symbol interval $p(t)$ and $q(t)$ have the same energy, the optimum demodulator (either soft decision or hard decision) for the additive white Gaussian noise channel need process the received signal over only two symbol intervals. It will further be shown that particular choices of $\{a(i)\}$, $\{b(i)\}$, $p(t)$ and $q(t)$ yield differential minimum shift keying and differential staggered quadriphase shift keying. The generalization of the approach to more general sequence transducers will be discussed.

INTRODUCTION

We consider the modulator structure shown in Fig. 1 consisting of a single-input two-output sequential transducer whose output sequences control the modulated signal in the manner

$$s(t) = x(i)p(t)+y(i)q(t), \quad iT \le t < iT + T \tag{1}$$

where T is the modulation bit duration and where $x(i)$ and $y(i)$ take values in $\{+1, -1, 0\}$.

In the next section, we describe the particular sequential transducer of interest in this paper and characterize its behaviour by means of a "trellis". In the following section we show that, provided only that the "carriers" $p(t)$ and $q(t)$ have equal energy in every bit interval, the optimum demodulator for the additive white Gaussian noise (AWGN) channel need process the received signal over only two bit intervals in making each data bit decision (which can either be a "hard" or "soft" decision). Finally, we show that particular choices of the sequential transducer yield differential minimum shift keying and differential staggered quadriphase shift keying, and we discuss some possible generalizations of our approach.

SEQUENCE TRANSDUCER AND TRELLIS

Let $d(i)$, $-\infty < i < \infty$, be the data sequence that constitutes the modulator input, where each $d(i)$ has value +1 (corresponding to a binary 0) or -1 (corresponding to a binary 1). The heart of the modulator structure is the sequence transducer, shown in Fig. 2, whose output sequence $\{x(i)\}$ and $\{y(i)\}$ are given by

$$x(i) = a(i)[d(i) + d(i-1)]/2 \tag{2a}$$

and

$$y(i) = b(i)[d(i) - d(i-1)]/2 \tag{2b}$$

for $-\infty < i < +\infty$, where $\{a(i)\}$ and $\{b(i)\}$ are arbitrary ± 1 sequences. From (2), we see that the components of the output sequences $\{x(i)\}$ and $\{y(i)\}$ take values in the set $\{-1,0,+1\}$, i.e., these sequences are ternary-valued. Moreover, for each i, either $x(i)=0$ corresponding to $d(i) \ne d(i-1)$ or $y(i)=0$ corresponding to $d(i) = d(i-1)$, but not both. Thus $x(i)$ and $y(i)$ can (and soon will) be used to amplitude-modulate the "quadrature components" of a carrier in such a way that one and only one component will be present in each bit interval.

A convenient way to display the input/output relationship of the sequence transducer of Fig. 2 is by means of the trellis shown in Fig. 3. The two nodes of this trellis at each stage correspond to the two possible states of the sequence transducer at the corresponding time instant i. From (2), we see that this state can be chosen as the value of $d(i-1)$; the nodes in Fig. 3 have been labeled with this choice of state. The branches leaving each state at time i correspond to the possible state transitions, the upper branch being the transition when the input $d(i)$ equals -1 and the lower branch being the transition when the input $d(i)$ equals +1; the branches are labeled with the value of $(x(i),y(i))$ corresponding to the output pair for that transition. We see, for instance, that if the transducer is in state +1 at time instant 0, then the input $d(0)=-1$ will cause a transition to state -1 at time instant 1 and the accompanying outputs will be $(x(0),y(0))=(0,-b(0))$; similarly, the input $d(1)=-1$ will also cause a transition from state +1 at time 1 to state -1 at time 2, but the accompanying outputs will be $(x(1),y(1))=(0,-b(1))$. Thus, each path through the entire trellis corresponds to a particular input sequence $\{d(i)\}$, and the labels on this path specify the resulting output sequences $\{x(i)\}$ and $\{y(i)\}$.

The trellis of Fig. 3 is, of course, closely akin to the "trellis" introduced by Forney to represent convolutional codes (Ref. 1). The only essential difference is that the "sections" between the nodes at each stage are identical in convolutional code trellises because the encoder is a time-invariant sequential transducer. The sequence transducer of Fig. 1, however, is time-varying in general because of the effect of the sequences $\{a(i)\}$ and $\{b(i)\}$.

WAVEFORM TRELLIS

Because of (1), we see that the trellis of Fig. 3 can be modified, as shown in Fig. 4, to show the input/output structure of the modulator of Fig. 1. The trellis in Fig. 4 differs from that in Fig. 3 only in that the node depth is labeled by the time t=iT instead of by the time index i, and that the transitions are now labeled with the value of s(t) in the bit interval $iT \le t < (i+1)T$ rather than with (x(i),y(i)). Thus, each path through the entire trellis of Fig. 4 still corresponds to a particular input data sequence $\{d(i)\}$, but the labels on the path now specify the resulting modulated signal s(t) from the modulator of Fig. 1.

MSK OPERATION

We show now that the modulator of Fig. 1 realizes differential binary minimum shift keying (MSK) modulation (Ref. 2) when the sequences $\{a(i)\}$ and $\{b(i)\}$ are chosen as

$$a(i) = 1, \quad \text{all } i \tag{3a}$$

$$b(i) = (-1)^{i+1} \tag{3b}$$

and when the carriers are selected in the manner

$$p(t) = A \sin\left[\left(\omega_o + \frac{\Delta\omega}{2}\right)t + \theta\right] \tag{4a}$$

and

$$q(t) = A \sin\left[\left(\omega_o - \frac{\Delta\omega}{2}\right)t + \theta\right], \tag{4b}$$

where A and θ are an arbitrary amplitude and arbitrary phase, respectively, where ω_o is the carrier center frequency, and where

$$(\Delta\omega)T = \pi. \tag{5}$$

In Fig. 5, we show the waveform trellis corresponding to the specific choice of $\{a(i)\}$ and $\{b(i)\}$ given by (3a) and (3b).

The phase of s(t) is certainly continuous at t=iT if s(t) does not switch between the two carriers at time iT, since then s(t) will be the same carrier waveform in the two bit intervals adjacent at time iT. It remains to show that the phase is continuous when s(t) switches from one carrier to the other at t=iT.

Suppose first that i is even. We see then from Fig. 5 that the only such carrier-switching transitions possible at t=iT for s(t) are:

(i) from -p(t) to -q(t),
(ii) from +p(t) to +q(t),
(iii) from -q(t) to -p(t), or
(iv) from +q(t) to +p(t).

But we see from (4) and (5) that the phase difference at time t=iT between p(t) and q(t) is $(\Delta\omega)iT = i\pi$, which is a multiple of 2π when i is even; thus the phase of s(t) is continuous for all four of the above transitions. A similar argument applies for the case where i is odd. Thus the modulated waveform has continuous phase. Moreover, from Fig. 5 we see that the carrier frequency in the bit interval $iT \le t < iT + T$ is $\omega_o + \Delta\omega/2$ when d(i) = d(i-1) so that the state at times iT and iT+T coincide, but is $\omega_o - \Delta\omega/2$ when $d(i) \ne d(i-1)$. Hence the modulation is indeed differential MSK.

SQPSK OPERATION

We show now that the modulator of Fig. 1 realizes differential staggered-quadriphase shift-keying (SQPSK) modulation (Ref. 3), also called offset-keyed quadriphase shift-keying (OKQPSK), when the sequences $\{a(i)\}$ and $\{b(i)\}$ are selected as in (3) but the carriers are selected in the manner

$$p(t) = A \cos(\omega_o t + \theta) \tag{6a}$$

and

$$q(t) = A \sin(\omega_o t + \theta), \tag{6b}$$

where again A and θ are an arbitrary amplitude and arbitrary phase, respectively.

By definition, SQPSK is four-phase modulation in which the phase can change by either 0 or $\pm\pi/2$ radians, but never by $\pm\pi$. This restricted phase-changing is obviously achieved by the modulator of Fig. 1 since, as we see from Fig. 5, s(t) can never make a transition between +p(t) and -p(t) or between +q(t) and -q(t), which are the only transitions according to (6) for which the phase of s(t) would change by $\pm\pi$. Moreover, we see from Fig. 5 that the phase in the bit interval $(i-1)T \le t < iT$ differs from that in $iT \le t < iT+T$ if and only if $d(i) \ne d(i-1)$ so that the modulation is differential.

OPTIMUM DEMODULATION

We now consider demodulation for the modulator of Fig. 1 when the received signal is

$$r(t) = s(t) + n(t), \tag{7}$$

where n(t) is additive white Gaussian noise (AWGN). We will show the rather remarkable fact that the hard-decision demodulator shown in Fig. 6 is optimum (in the sense of minimizing the probability of error in the decision $\hat{d}(i)$ for d(i), for all i) regardless of the choice of p(t) and q(t) (i.e., the two carriers need have no special orthogonality properties) provided only that, for every i, p(t) and q(t) have the same energy in the bit interval $iT \le t < iT+T$ (but the energy could depend on i) and that $\{d(i)\}$ is a random data sequence. Henceforth, we assume that this energy condition and data condition are satisfied.

To demonstrate the optimality of the hard-decision demodulator of Fig. 6, we exploit the "magic genie" approach of Wozencraft and Jacobs (ref. 4, p. 419). Suppose we wish to estimate d(i). Suppose further that the genie is kind enough to tell us both the state, $\sigma(i)$, of the modulator at time instant i and also the state, $\sigma(i+2)$, at time instant i+2. If the genie says $\sigma(i) = +1$ and $\sigma(i+2) = +1$, for instance, we see from the trellis of Fig. 4 that $d(i) = +1$ would imply $s(t) = a(i)p(t)$ for $iT \le t < iT+T$ and $s(t) = a(i+1)p(t)$ for $iT+T \le t < iT+2T$; whereas $d(i) = -1$ would imply $s(t) = -b(i)q(t)$ for $iT \le t < iT+T$ and $s(t) = b(i+1)q(t)$ for $iT+T \le t < iT+2T$. Moreover, any permissible choice of s(t), for $t < iT$ and for $t \ge iT+2T$, when $d(i) = +1$ is also permissible when $d(i) = -1$, since the only requirement is that the state be +1 at time instant i and again +1 at time instant i+2. Thus, the decision problem for d(i) (with genie's help) reduces to deciding which of two equiprobable and equal energy signals was transmitted in the interval $iT \le t < iT+2T$ in the presence of AWGN. The well-known (Ref. 4, pp. 238-239) optimum decision rule is: Choose $\hat{d}(i) = +1$ if and only if

$$\begin{aligned} &a(i)X(i)+a(i+1)X(i+1) \ge \\ &-b(i)Y(i)+b(i+1)Y(i+1) \end{aligned} \tag{8}$$

where

$$X(i) = \int_{iT}^{iT+T} r(t)p(t)dt \tag{9a}$$

and

$$Y(i) = \int_{iT}^{iT+T} r(t)q(t)dt. \tag{9b}$$

But now suppose instead that the genie had told us that $\sigma(i) = -1$ and $\sigma(i+2) = +1$. We see from the trellis of Fig. 4 that $d(i) = +1$ corresponds to $s(t) = b(i)q(t)$ for $iT \le t < iT+T$ and $s(t) = a(i+1)p(t)$ for $iT+T \le t < iT+2T$. Similarly, $d(i) = -1$ corresponds to $s(t) = -a(i)p(t)$ for $iT \le t < iT+T$ and $s(t) = b(i+1)q(t)$ for $iT+T \le t < iT+2T$. Thus, the optimum (genie-aided) decision rule is: Choose $\hat{d}(i) = +1$ if and only if

$$\begin{aligned} &b(i)Y(i)+a(i+1)X(i+1) \ge \\ &-a(i)X(i)+b(i+1)Y(i+1), \end{aligned} \tag{10}$$

which we see is precisely the same condition as (8)!

Similar analyses for the case $\sigma(i) = +1$ and $\sigma(i+2) = -1$ and for the case $\sigma(i) = -1$, $\sigma(i+2)=-1$ show that the optimum genie-aided decision rules are again: Choose $\hat{d}(i) = +1$ if and only if (8) is satisfied. But the four cases considered exhaust the possible values for $\sigma(i)$ and $\sigma(i+2)$. We conclude that we can exorcise the genie; we have no use for his information since the optimum decision rule for d(i), when i is even, is independent of his information and is just: Choose $\hat{d}(i) = +1$ if and only if (8) is satisfied. Note that this is precisely the rule for the demodulator in Fig. 6.

Note that the optimum decision $\hat{d}(i)$ depends only on r(t) over the two-symbol interval $iT \le t < iT+2T$; this is a well-known fact for both binary MSK and SQPSK, but we have now demonstrated that this "two-symbol optimality" is independent of whether p(t) and q(t) are orthogonal over each symbol interval as they are in both MSK and SQPSK.

We have argued elsewhere (Refs. 5 and 6) that demodulators should be designed to maximize the cut-off rate, R_o, of the discrete channel created by the modulator, waveform channel and demodulator, rather than to minimize bit error probability for a hard-decision demodulator. Thus, we find it much more satisfactory than showing that the demodulator of Fig. 6 is the optimum hard-decision demodulator to show that:

> When A(i) - B(i) is taken as the output, the demodulator of Fig. 6 is optimum in the sense of maximizing the cut-off rate, R_o, of the discrete channel between the modulator input and demodulator output (and also in the sense of maximizing the capacity, C, of this channel).

To prove this claim, one must show that the demodulator with output A(i) - B(i) preserves the likelihood ratio for the decision on d(i), since any operation on r(t) reduces R_o (and also C) unless and only unless this likelihood ratio is preserved (Ref. 5). The likelihood ratio $\Lambda(i)$ for the decision about d(i) is by definition

$$\Lambda(i)=p(\underline{r}|d(i)=+1)/p(\underline{r}|d(i)=-1) \tag{11}$$

where $\underline{r}$ is the representation of r(t) in some appropriate Euclidean space. We omit the details, but it is straightforward to show that

$$\Lambda(i) = \exp 2(A(i)-B(i))/N_o \tag{12}$$

where N_o is the one-sided noise power spectral density. Equation (12) shows that the likelihood ratio is indeed preserved if the demodulator output is taken as A(i)-B(i) in Fig. 6.

GENERALIZATIONS

Many generalizations of the above work are possible, both in the choice of the sequential transducer and in the choice of the various carriers. The unit-sample response of the sequential transducer in Fig. 2, when the two outputs are taken as the outputs of the adders, are just the vectors [1, 1] and [1, -1] (followed of course by all zeroes.) But these two vectors form the rows of a 2 x 2 Hadamard matrix, and this fact seems responsible for the rather remarkable properties of the resultant modulation system. One suspects that sequential transducers characterized by higher order Hadamard matrices might also have interesting properties.

ACKNOWLEDGMENT

The author is grateful to Dr. James R. Lesh and Dr. Marvin K. Simon, both of the Jet Propulsion Laboratory in Pasadena, CA, for several useful suggestions that were incorporated into this paper. The research reported here was supported in part by NASA Contract NAS7-100.

REFERENCES

1. G.D. Forney, Jr., "The Viterbi Algorithm," Proc. IEEE, Vol. 61, pp 268-273, March 1973.
2. M.L. Doelz and E.H. Heald, "Minimum-Shift Data Communications," U.S. Patent No. 2,977,417, March 28, 1961.
3. M.K. Simon and J.G. Smith, "Offset Quadrature Communications with Decision Feedback," IEEE Trans. Comm., Vol. COM-22, pp 1576-1584, Oct.1974.
4. J.M. Wozencraft and I.M. Jacobs, Principles of Communication Engineering. New York: Wiley 1965.
5. J.L. Massey, "Coding and Modulation in Digital Communications," in Proc. 1974 Int. Zurich Seminar on Digital Comm., pp E2(1)-E2(4), March 12-15, 1974.
6. J.L. Massey, "Comparison of Phase Modulation Systems," in The Deep Space Network Progress Report 42-49, Jet Propulsion Laboratory, Pasadena, Calif., Feb. 1979, pp 57-63.

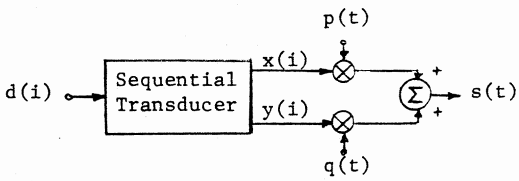

Fig. 1: Modulator Structure

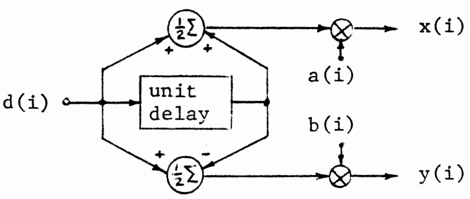

Fig. 2: The Sequential Transducer

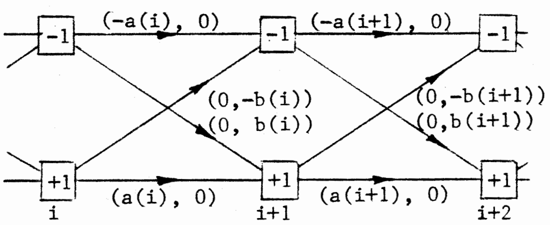

Fig. 3: Sequential Transducer Trellis
(Leave state by upper branch if data is -1)

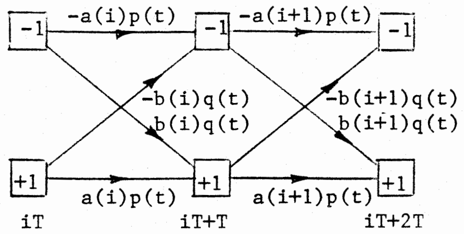

Fig. 4: Modulator Waveform Trellis
(Leave state by upper branch if data is -1)

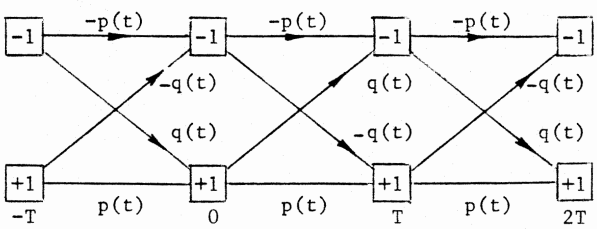

Fig. 5: Modulator Waveform Trellis for Special Case When $a(i) = 1$ and $b(i) = (-1)^{i+1}$
(Leave state by upper branch if data is -1)

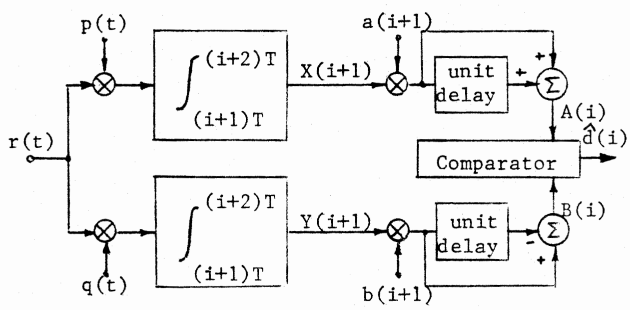

Fig. 6: Optimum Hard-Decision Demodulator
(Comparator output is +1 if $A(i) \geq B(i)$)