A Generalization of Binary Minimum Shift Keying and Staggered Quadriphase Shift Keying Modulation

J. L. Massey
Consultant to
Communications Systems Research Section

A generalized modulation scheme, which includes minimum shift keying (MSK) and staggered quadriphase shift keying (SQPSK) as special cases, is analyzed. The general modulator can be realized as a one-input, two-output sequence transducer whose outputs select the carrier signal for each baud. This form of the modulator has the practical advantage of not requiring any RF filtering since there is no actual mixing of the carriers with the modulating signals.

It is shown that the optimum demodulator (whether hard-decision or soft-decision) always can make its decisions from the received waveform over two bauds when the interference is additive white Gaussian noise, thus generalizing a well-known result for hard-decision demodulation of MSK and SQPSK signals. The power spectra of MSK and SQPSK signals are derived to isolate the role played by coherency between the modulating signals and the carriers.

I. Introduction

In this report, we study a generalized modulation scheme that includes minimum shift keying (MSK) and staggered quadriphase shift keying (SQPSK) modulation as special cases. The generalized modulator, whose form has certain practically advantageous features, is introduced in Section II as a specific one-input, two-output sequence transducer whose outputs select the carrier signal for each baud. The resulting modulated signal is described by means of a waveform "trellis" associated with the "trellis" of the sequence transducer. The particular choice of carrier signals to obtain SQPSK and MSK modulation are specified.

In Section III, we use the waveform trellis of the modulator in a "magic genie" argument to prove that the optimum demodulator, whether hard-decision or soft-decision, can restrict its observations to two consecutive bauds when the interference is additive white Gaussian noise. This generalizes a well-known fact about MSK and QPSK modulation.

In Section IV, we derive the power spectra for both MSK and SQPSK signals, paying special attention to the required relationships between the carriers and the modulating signals. Finally, in Section V, we make some concluding remarks and give some suggestions for future research.

II. The Modulator

A. Sequence Transducer and Trellis

Let $\mathbf{d} = \cdots d_{-1}, d_0, d_1, d_2, \cdots$ 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 our modulator structure will be the *sequence transducer*, shown in Fig. 1, whose output sequences **x** and **y** are given by

$$x_i = \frac{1}{2}(d_i + d_{i-1}) \tag{1a}$$

and

$$y_i = \frac{1}{2} a_i (d_i - d_{i-1}) \tag{1b}$$

for $-\infty < i < +\infty$, where **a** is the *alternating sequence* defined by

$$a_i = (-1)^{i+1} \tag{2}$$

From Eqs. (1) and (2), we see that the components of the output sequences **x** and **y** 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 $d_i \neq d_{i-1}$) 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-moderate the "quadrature components" of a carrier in such a way that one and only one component will be present for each i.

A convenient way to display the input/output relationship of the sequence transducer of Fig. 1 is by means of the *trellis* shown in Fig. 2. The nodes of this trellis at depth i correspond to the possible states of the sequence transducer at time instant i. From Eq. (1), we see that this *state* can be chosen as the value of d_{i-1}; the nodes in Fig. 2 have been labeled with this choice of state. The branches leaving each state at depth 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 $(x_0, y_0) = (0, +1)$; 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 $(x_1, y_1) = (0, -1)$. Thus, each path through the entire trellis corresponds to a particular input sequence **d**, and the labels on this path specify the resulting output sequences **x** and **y**.

The trellis of Fig. 1 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 depth are identical in convolutional code trellises because the encoder is a time-invariant transducer. The sequence transducer of Fig. 1, however, is *time-varying* because of the effect of the alternating sequence **a**, which has period 2. In fact, the trellis of Fig. 2 is *periodic* with period 2 in the sense that the section beginning at time instant i is the same as that beginning at time instant $i + 2$ for all i.

We remark that the sequence transducer shown in Fig. 3 is equivalent to that in Fig. 1, as is easily seen from the facts that $d'_i = a_i d_i$ and that $d'_{i-1} = a_{i-1} d_{i-1} = -a_i d_{i-1}$, since $a_i = -a_{i-1}$. Note that if **d** is a *random data sequence*, i.e., if the random variables d_i are statistically independent and identically distributed (i.i.d.) with $P(d_i = +1) = P(d_i = -1) = 1/2$, then $\mathbf{d}'$ is also a random data sequence. From Fig. 3, we can see that the sequences **x** and **y** will have identical statistics in this case.

B. Modulator and Trellis

We are now ready to introduce our proposed modulator, whose structure is shown in Fig. 4. Here, $c_x(t)$ and $c_y(t)$ are *carrier waveforms* that will be specified later for specific modulation schemes. The *baud length* will be denoted as T. The function of the "RF selector switch" in Fig. 4 is to select one of the carrier inputs for transmission in each baud according to the following rule:

$$s(t) = \left.\begin{cases} +c_x(t), & \text{if } x_i = +1 \\ -c_x(t), & \text{if } x_i = -1 \\ +c_y(t), & \text{if } y_i = +1 \\ -c_y(t), & \text{if } y_i = -1 \end{cases}\right\}, \quad iT \leqslant t < iT + T \tag{3}$$

We emphasize that no addition of RF signals and *no RF filtering is required* in the modulator, since there is no actual mixing of the carriers with modulating signals.

Because of Eq. (3), we see that the trellis of Fig. 2 can be modified, as shown in Fig. 5, to show the input/output structure of the modulator of Fig. 4. The trellis in Fig. 5 differs from that in Fig. 4 only in that the node depth is labeled by the time iT instead of by the time instant i, and that the transitions are now labeled with the value of $s(t)$ in

the baud $iT \leq t < (i+1)T$ rather than with (x_i, y_i). Thus, each path through the entire trellis of Fig. 5 still corresponds to a particular input data sequence **d**, but the labels on the path now specify the resulting *modulated signal* $s(t)$ from the modulator of Fig. 4.

C. MSK Operation

We show now that the modulator of Fig. 4 realizes binary *minimum shift keying* (MSK) modulation (Ref. 2) when the carriers are selected in the manner

$$c_x(t) = A \sin \left[\left(\omega_0 + \frac{\Delta\omega}{2}\right) t + \theta\right] \tag{4a}$$

and

$$c_y(t) = A \sin \left[\left(\omega_0 - \frac{\Delta\omega}{2}\right) t + \theta\right] \tag{4b}$$

where A and θ are an arbitrary amplitude and arbitrary phase, respectively, where ω_0 is the carrier center frequency, and where

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

Since, in each baud, the modulated signal $s(t)$ will be one of $+c_x(t)$, $-c_x(t)$, $+c_y(t)$ and $-c_y(t)$, it suffices to show that the phase of $s(t)$ is continuous at the transitions between bauds (i.e., at times $t = iT$ for all i and for all choices of the data sequence **d**).

From the trellis of Fig. 5, we see that 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 one of $+c_x(t)$, $-c_x(t)$, $+c_y(t)$, or $-c_y(t)$ in the two bauds adjacent at time iT. It remains to show that the phase is continuous when $s(t)$ switches from either $+c_x(t)$ or $-c_x(t)$ to either $+c_y(t)$ or $-c_y(t)$, or from either $+c_y(t)$ or $-c_y(t)$ to either $+c_x(t)$ or $-c_x(t)$, 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 $-c_x(t)$ to $-c_y(t)$,

(ii) from $+c_x(t)$ to $+c_y(t)$,

(iii) from $-c_y(t)$ to $-c_x(t)$, or

(iv) from $+c_y(t)$ to $+c_x(t)$.

But we see from Eqs. (4) and (5) that the phase difference at time $t = iT$ between $c_x(t)$ and $c_y(t)$ 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.

Suppose on the other hand that i is odd. We see now from Fig. 5 that the only carrier-switching transitions that are possible at $t = iT$ for $s(t)$ are:

(i) from $-c_x(t)$ to $+c_y(t)$,

(ii) from $+c_x(t)$ to $-c_y(t)$,

(iii) from $-c_y(t)$ to $+c_x(t)$, or

(iv) from $+c_y(t)$ to $-c_x(t)$.

But, from Eqs. (4) and (5), we see that the phase difference at time $t = iT$ between $c_x(t)$ and $-c_y(t)$ is $(\Delta\omega)iT + \pi = (i+1)\pi$, which is a multiple of 2π when i is odd; thus again the phase of $s(t)$ is continuous for all four possible transitions. This completes the demonstration that the modulation is indeed binary MSK.

More precisely, the above argument shows that the modulated signal $s(t)$ is the same as for MSK modulation. However, in what is commonly considered to be MSK modulation, the data sequence controls the modulated waveform in the manner that when the i^{th} data digit is +1 or −1, then the modulated signal in the i^{th} baud is $\pm A \sin [(\omega_0 + \Delta\omega/2)t + \theta]$ or $\pm A \sin [(\omega_0 - \Delta\omega/2)t + \theta]$, respectively, when the appropriate sign is chosen to maintain phase continuity. In other words, the data digit directly controls the baud frequency. We see from Eq. (1a), however, that x_i (which controls the sinusoid of frequency $\omega_0 + \Delta\omega/2$) is nonzero if and only if $d_i = d_{i-1}$; and that y_i (which controls the sinusoid of frequency $\omega_0 - \Delta\omega/2$) is nonzero if and only if $d_i \neq d_{i-1}$. But the i^{th} bit in the first difference of the binary data stream is 0 (corresponding to +1) when $d_i = d_{i-1}$ and is 1 (corresponding to −1) when $d_i \neq d_{i-1}$. Thus, the modulator of Fig. 4 actually realizes *differential* MSK since the differences in the data sequence form the sequence that actually controls the baud frequency. However, we shall continue to say simply "MSK modulation," rather than the more precise "differential MSK modulation," to describe the modulation performed by the modulator of Fig. 4 when the carriers are specified as in Eq. (6).

D. SQPSK Operation

We show now that the modulator of Fig. 4 realizes *staggered-quadriphase shift-keying* (SQPSK) modulation (Ref. 3), also called offset-keyed quadriphase shift-keying (OKQPSK), when the carriers are selected in the manner

$$c_x(t) = A\cos(\omega_0 t + \theta) \tag{6a}$$

and

$$c_y(t) = A\sin(\omega_0 t + \theta) \tag{6b}$$

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 $\pm\pi/2$ between bauds, but never by $\pm\pi$. (This is usually accomplished by modifying an ordinary QPSK modulator for a baud of length $2T$ in which two independent data sequences binary-antipodally modulate the quadrature carrier components so that one data sequence makes its transitions mid-way between the other's transitions – the advantage over ordinary QPSK is a reduction of cochannel interference due to nonideal effects in demodulation.) But this restricted phase-changing is obviously achieved by the modulator of Fig. 4 since, as we see from Fig. 5, $s(t)$ can never make a transition between $+c_x(t)$ and $-c_x(t)$ or between $+c_y(t)$ and $-c_y(t)$, which are the only transitions according to Eq. (6) for which the phase of $s(t)$ would change by $\pm\pi$.

Again, we remark that we have really shown here only that the modulated signal $s(t)$ is the same as for SQPSK modulation. Just as was the case for MSK modulation, however, when the correspondence between the data sequence and the modulated signal is considered, one finds that the modulator of Fig. 4 actually realizes *differential* SQPSK.

III. The Demodulator

We now consider demodulation for the modulator of Fig. 4 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 $c_x(t)$ and $c_y(t)$ (i.e., these two carriers need have no special orthogonality properties in each baud) provided only that, for every i, $c_x(t)$ and $c_y(t)$ have the *same energy* in the baud $iT \leq t < iT + T$ (but the energy could depend on i) and that $\mathbf{d}$ 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 where i is even. Suppose further that the genie is kind enough to tell us both the state, σ_i, of the modulator at time instant i and also the state, σ_{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. 5 that $d_i = +1$, would imply $s(t) = +c_x(t)$ for $iT \leq t < iT + 2T$, whereas $d_i = -1$ would imply $s(t) = +c_y(t)$ for $iT \leq t < iT + 2T$. Moreover, any permissible choice of $s(t)$, for $t < iT$ and for $t \geq iT + 2T$, 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 only whether $c_x(t)$ or $c_y(t)$ was transmitted in the interval $iT \leq t < iT + 2T$. But this is the classical problem of deciding between two equally-likely equal-energy signals in the presence of AWGN, and the well-known (Ref. 4, pp. 238-239) optimum decision rule is: Choose $\hat{d}_i = +1$ if and only if

$$\int_{iT}^{iT+2T} r(t)c_x(t)\,dt \geq \int_{iT}^{iT+2T} r(t)c_y(t)\,dt \tag{8}$$

Defining

$$X_i = \int_{iT}^{iT+T} r(t)c_x(t)\,dt \tag{9a}$$

and

$$Y_i = \int_{iT}^{iT+T} r(t)c_y(t)\,dt \tag{9b}$$

we see that Eq. (8) can be written as

$$X_i + X_{i+1} \geq Y_i + Y_{i+1} \tag{10}$$

But now suppose instead that the genie had told us that $\sigma_i = -1$ and $\sigma_{i+2} = +1$. Recalling that i is even, we see from the trellis of Fig. 5 that $d_i = +1$ corresponds to $s(t) = -c_y(t)$ for $iT \leq t < iT + T$ and $s(t) = c_x(t)$ for $iT + T \leq t < iT + 2T$. Similarly, $d_i = -1$ corresponds to $s(t) = -c_x(t)$ for $iT \leq t \leq iT + T$ and $s(t) = c_y(t)$ for $iT + T \leq t < iT + 2T$. Thus, the optimum (genie-aided) decision rule is: Choose $\hat{d}_i = +1$ if and only if

$$-\int_{iT}^{iT+T} r(t)c_y(t)\,dt + \int_{iT+T}^{iT+2T} r(t)c_x(t)$$

$$\geq -\int_{iT}^{iT+T} r(t)c_x(t)\,dt + \int_{iT+T}^{iT+2T} r(t)c_y(t)\,dt \quad (11)$$

Using Eq. (9), we can write Eq. (11) as

$$-Y_i + X_{i+1} \geq -X_i + Y_{i+1} \quad (12)$$

which we see is precisely the same condition as Eq. (10)!

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 Eq. (10) is satisfied. But the four cases considered exhaust the possible values for σ_i and σ_{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 Eq. (10) is satisfied. Note that this is precisely the rule for the demodulator in Fig. 6 since, for i even, $-a_i = +1$ so that $A_i = X_i + X_{i+1}$ and $B_i = Y_i + Y_{i+1}$, where A_i and B_i are defined as

$$A_i = X_i + X_{i+1} \quad (13a)$$

$$B_i = -a_i(Y_i + Y_{i+1}) \quad (13b)$$

Next, suppose i is odd and the genie tells us that $\sigma_i = +1$ and $\sigma_{i+2} = +1$. From Fig. 5, we see that $d_i = +1$ corresponds to $s(t) = c_x(t)$ for $iT \leq t < iT + 2T$, whereas $d_i = -1$ corresponds to $s(t) = -c_y(t)$ for $iT \leq t < iT + 2T$. The same argument as before now shows the optimum genie-aided decision rule to be: Chose $\hat{d}_i = +1$ if and only if

$$X_i + X_{i+1} \geq -Y_i - Y_{i+1}$$

Analysis of the three remaining choices of σ_i and σ_{i+2} shows that this same decision rule is also optimum. Thus, the genie can again be exorcised. We note further that this optimum decision rule is the rule for the demodulator in Fig. 6 since, for i odd, $-a_i = -1$ so that $A_i = X_i + X_{i+1}$ and $B_i = -Y_i - Y_{i+1}$.

This completes our proof that the hard-decision demodulator in Fig. 6 is optimum in the sense of minimizing $Pr(\hat{d}_i \neq d_i)$ when the noise is AWGN, when **d** is a random data sequence and, when for each i, the carriers $c_x(t)$ and $c_y(t)$ have the same energy in the baud $iT \leq t < iT + T$. Note that the optimum decision $\hat{d}_i$ for d_i depends only on $r(t)$ over the two-baud interval $iT \leq t < iT + 2T$; this is a well-known fact for both binary MSK and SQPSK, but we have now demonstrated that this "two-baud optimality" is independent of whether $c_x(t)$ and $c_y(t)$ are orthogonal over each baud 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_0, 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_0, 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, we 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_0 (and also C) unless and only unless this likelihood ratio is preserved (Ref. 5).

We begin by letting **r**, **s** and **n** be the vector representations of $r(t)$, $s(t)$ and $n(t)$, respectively, in some appropriate Euclidean space ("signal space"). Given that **r** is received and that a magic genie informs the receiver that $d_j = \delta_j$ for all $j \neq i$, the likelihood ratio for the decision on d_i is

$$\Lambda_i = \frac{p(\mathbf{r} \mid d_i = +1, d_j = \delta_i \text{ for } j \neq i)}{p(\mathbf{r} \mid d_i = -1, d_j = \delta_j \text{ for } j \neq i)} \quad (14)$$

where p is a conditional probability density function. Equation (14) can be rewritten as

$$\Lambda_i = \frac{p(\mathbf{r} \mid \mathbf{s}_{+1})}{p(\mathbf{r} \mid \mathbf{s}_{-1})} \quad (15)$$

where $\mathbf{s}_{+1}$ and $\mathbf{s}_{-1}$ are the signal space representations of the waveforms $s_{+1}(t)$ and $s_{-1}(t)$ assumed by $s(t)$ for the data sequence $d_j = \delta_j$ for $j \neq i$ with d_i equal to +1 and −1, respectively. Since $\mathbf{r} = \mathbf{s} + \mathbf{n}$ and the noise $n(t)$ is AWGN, Eq. (15) becomes

$$\Lambda_i = \frac{e^{-\frac{1}{N_0}|\mathbf{r} - \mathbf{s}_{+1}|^2}}{e^{-\frac{1}{N_0}|\mathbf{r} - \mathbf{s}_{-1}|^2}} \quad (16)$$

where N_0 is the one-sided noise power spectral density. The assumption that, in each baud, $c_x(t)$ and $c_y(t)$ have the same energy implies that $s_{+1}(t)$ and $s_{-1}(t)$ have the same energy or, equivalently, $|\mathbf{s}_{+1}|^2 = |\mathbf{s}_{-1}|^2$. Thus, Eq. (16) becomes

$$\Lambda_i = e^{\frac{2}{N_0}[\mathbf{r} \cdot (\mathbf{s}_{+1} - \mathbf{s}_{-1})]} \tag{17}$$

By the correspondence of "dot product" in signal space to correlation of waveforms,

$$\mathbf{r} \cdot (\mathbf{s}_{+1} - \mathbf{s}_{-1}) = \int_{-\infty}^{+\infty} r(t)\,[s_{+1}(t) - s_{-1}(t)]\,dt \tag{18}$$

Because the data sequences yielding $s_{+1}(t)$ and $s_{-1}(t)$ differ only in the value of d_i, we see from the trellis of Fig. 5 that $s_{+1}(t) = s_{-1}(t)$ for $t < iT$ and $t \geqslant iT + 2T$ so that Eq. (18) can be written as

$$\mathbf{r} \cdot (\mathbf{s}_{+1} - \mathbf{s}_{-1}) = \int_{iT}^{iT+2T} r(t)\,[s_{+1}(t) - s_{-1}(t)]\,dt \tag{19}$$

It remains to evaluate Eq. (19). From Eqs. (2) and (9) and the trellis of Fig. 5, we find that, for all four possible values of d_{i-1} and d_{i+1},

$$\int_{iT}^{iT+T} r(t)\,[s_{+1}(t) - s_{-1}(t)]\,dt = X_i + a_i Y_i$$

and

$$\int_{iT+T}^{iT+2} r(t)\,[s_{+1}(t) - s_{-1}(t)]\,dt = X_{i+1} + a_{i+1}\,Y_{i+1}$$

$$= X_{i+1} - a_i\,Y_{i+1}$$

This, together with Eq. (13), implies that Eq. (19) can be written as

$$\mathbf{r} \cdot (\mathbf{s}_{+1} - \mathbf{s}_{-1}) = A_i - B_i \tag{20}$$

(which shows again that we can exorcise the genie, since A_i and B_i can be formed without the genie's help.) Finally, substitution of Eq. (20) into Eq. (17) gives

$$\Lambda_i = e^{\frac{2}{N_0}(A_i - B_i)} \tag{21}$$

which shows that the output $A_i - B_i$ does indeed preserve the value of the likelihood ratio, as was to be shown.

We note that, since the demodulator of Fig. 6 with $A_i - B_i$ taken as the output preserves the likelihood ratio (and hence *all* the statistical information about d_i), the *optimum* (in the sense of maximizing R_0 or C) *soft-decision modulator* with any *specified number of demodulator output levels* can be obtained simply by appropriately quantizing the output Z_i (cf. Refs. 5 and 7 for details of this procedure).

IV. Modulation Spectra

A. General Considerations

We now examine the spectral properties of the signal produced by the modulator of Fig. 4.

With the sequence $\mathbf{x}$ of Eq. (1a), we associate the pulsed waveform $x(t)$ defined by

$$x(t) = x_i \quad iT \leqslant t < (i+1)T \tag{19}$$

In the same way, we associate pulsed waveforms $y(t)$, $d(t)$, and $a(t)$ with the sequences $\mathbf{y}$, $\mathbf{d}$, and $\mathbf{a}$, respectively. Note that $a(t)$ is a "square wave" of period $2T$.

For purposes of spectral analysis, we need now to randomize the time origin so that all time signals under consideration become stationary random processes. We accomplish this by replacing all time functions [e.g., $x(t)$ and $s(t)$] with their translations by t_0 seconds [e.g., $x(t + t_0)$ and $s(t + t_0)$] where t_0 is a random variable uniformly distributed over the interval $0 \leqslant t_0 < 2T$ and statistically independent of the data pulse values. We will write $R_x(\tau)$ and $R_s(\tau)$, for instance, to denote the resulting autocorrelation functions

$$R_x(\tau) = E[x(t + t_0 + \tau)\,x(t + t_0)]$$

$$= E[x(t_0 + \tau)\,x(t_0)]$$

and

$$R_s(\tau) = E[s(t + t_0 + \tau)\,s(t + t_0)]$$

$$= E[s(t_0 + \tau)\,s(t_0)]$$

respectively. It should be noted that, since *all* time signals are translated by t_0 seconds, this randomization of the time origin neither introduces nor removes any "coherence" between the modulating signals and the carriers.

Note that the data signal $d(t + t_0)$ is the familiar telegraph wave (with baud length T and amplitude ±1). Thus, $R_d(\tau)$ is the triangular autocorrelation function shown in Fig. 7; and the power spectral density (the Fourier transform of $R_d(\tau)$) is the familiar

$$S_d(f) = T\,\mathrm{sinc}^2\,(fT) \tag{20}$$

where

$$\mathrm{sinc}\,(z) = \frac{\sin(\pi z)}{\pi z}$$

The spectrum $S_d(f)$ is also shown in Fig. 7.

Recalling Eqs. (1a) and (1b), we see that

$$x(t) = \frac{1}{2}\,[d(t) + d(t - T)] \tag{21a}$$

and

$$y(t) = \frac{1}{2}\,a(t)\,[d(t) - d(t - T)] \tag{21b}$$

From Eqs. (21a) and (21b), we find by a simple calculation that

$$R_x(\tau) = R_y(\tau) = \frac{1}{4}R_z(\tau + T) + \frac{1}{2}R_z(\tau) + \frac{1}{4}R_z(\tau - T)$$

$$= \frac{1}{2}R_\Delta\left(\frac{\tau}{2}\right) \tag{22}$$

which further implies

$$S_x(f) = S_y(f) = S_d(2f) = T\,\mathrm{sinc}^2\,(2fT) \tag{23}$$

These autocorrelation functions and spectra are shown in Fig. 8. Inasmuch as $x(t)$ and $y(t)$ are ternary-valued (0, +1 or −1), it is a bit surprising that their autocorrelation functions are of the same form as the binary-valued (±1) data signal $d(t)$.

Recall from our discussion in Section II about the sequences **x** and **y** that, for each i, one of x_i and y_i will be 0 and the other will be +1 or −1. Thus, the random processes $x(t + t_0)$ and $y(t + t_0)$ are highly dependent. It is quite surprising then to find that *the random processes* $x(t + t_0)$ and $y(t + t_0)$ *are uncorrelated*, i.e., that

$$R_{xy}(\tau) = E\,[x(t + t_0 + \tau)\,y(t_0 + t)] \equiv 0 \tag{24}$$

as can be verified by a direct (but somewhat tedious) calculation from Eqs. (21a) and (21b).

We now turn to the main task at hand, finding the spectrum of the modulated signal, i.e., of the random process $s(t + t_0)$. From Eq. (3), we see that

$$s(t) = x(t)\,c_x(t) + y(t)\,c_y(t) \tag{25}$$

which is our starting point. To proceed further, we must consider specific carrier signals. One general note of caution is in order, however, at this point. Because of Eq. (24), it might seem that $s(t)$ in Eq. (25) can be treated as the sum of two independently modulated signals. This is not true in general, however, because some types of coherence between $c_x(t)$ and $c_y(t)$ can cause $x(t + t_0)\,c_x(t + t_0)$ to be correlated with $y(t + t_0)\,c_y(t + t_0)$, even though $x(t + t_0)$ and $y(t + t_0)$ are uncorrelated. We shall soon see that this type of coherence between $c_x(t)$ and $c_y(t)$ does not occur with SQPSK modulation, but does occur with MSK modulation.

B. SQPSK Spectrum

As we saw in Subsection II.D., the signal $s(t)$ is an SQPSK modulated signal when the carriers are specified by Eqs. (6a) and (6b). From these equations and Eq. (25), we find in this case that

$$s(t + t_0) = A\,x(t + t_0)\cos\,[\omega_0\,(t + t_0) + \theta]$$

$$+ A\,y(t + t_0)\sin\,[\omega_0\,(t + t_0) + \theta] \tag{26a}$$

It follows from Eq. (26a) that

$$R_s(\tau) = E\,[s(t + t_0 + \tau)\,s(t + t_0)]$$

$$= \frac{1}{2}A^2\,R_x(\tau)\cos\,(\omega_0\tau) + \frac{1}{2}A^2\,R_y(\tau)\cos\,(\omega_0\tau)$$

$$+ A^2\,R_{xy}(\tau)\sin\,(\omega_0\tau) \tag{27}$$

(Where the vanishing of the double carrier frequency terms under expectation strictly speaking requires T to be an integer multiple of the carrier period $2\pi/\omega_0$, but is virtually the case whenever $\omega_0 >> 1/T$.) By virtue of Eq. (24), we see that the term resulting from the cross-correlation of the two carriers vanishes (i.e., Eq. (26) represents independent modulation of orthogonal carriers) so that Eq. (27) becomes

$$R_s(\tau) = A^2 R_x(\tau) \cos(\omega_0 \tau) \tag{28}$$

where we have also made use of Eq. (22). From Eqs. (28) and (23), it follows that

$$S_s(f) = \frac{1}{2} A^2 T \{\text{sinc}^2 [2(f - f_0) T] + \text{sinc}^2 [2(f + f_0) T]\} \tag{29}$$

where $f_0 = \omega_0/2\pi$ is the carrier frequency in hertz. Equation (29) gives, of course, the well-known spectrum of SQPSK, which coincides with that of ordinary QPSK. The purposes of our deriving this result here are to emphasize the role played by the vanishing of $R_{xy}(\tau)$ and to point out that, because the phase angle θ in Eq. (26) is arbitrary, no special phase relationship is required between the carriers and the modulating signals. The spectrum $S_s(f)$ of Eq. (29) is sketched in Fig. 9.

C. MSK Spectrum

When the carriers in Eq. (25) are specified as in Eqs. (4a) and (4b), $s(t + t_0)$ becomes the MSK signal

$$s(t + t_0) = A\, x(t + t_0) \sin\left[\left(\omega_0 + \frac{\Delta\omega}{2}\right)(t + t_0) + \theta\right] + A\, y(t + t_0) \sin\left[\left(\omega_0 - \frac{\Delta\omega}{2}\right)(t + t_0) + \theta\right] \tag{26b}$$

Applying familiar trigonometric identities, we can rewrite Eq. (26b) as

$$s(t + t_0) = A\,[x(t + t_0) + y(t + t_0)] \cdot \cos\left[\frac{\Delta\omega}{2}(t + t_0)\right] \sin[\omega_0 (t + t_0) + \theta] + A\,[x(t + t_0) - y(t + t_0)] \cdot \sin\left[\frac{\Delta\omega}{2}(t + t_0)\right] \cos[\omega_0 (t + t_0) + \theta] \tag{26c}$$

We are thus led to define the signals

$$d_1(t) = x(t) + y(t) \tag{30a}$$

and

$$d_2(t) = x(t) - y(t) \tag{30b}$$

With the aid of Eqs. (21a) and (21b), we see from Eqs. (30a) and (30b) that

$$d_1(t) = \begin{cases} d(t), & iT \leqslant t < (i+1)T \text{ and } i \text{ odd} \\ d(t - T), & iT \leqslant t < (i+1)T \text{ and } i \text{ even} \end{cases} \tag{30c}$$

and

$$d_2(t) = \begin{cases} d(t - T), & iT \leqslant t < (i+1)T \text{ and } i \text{ odd} \\ d(t), & iT \leqslant t < (i+1)T \text{ and } i \text{ even} \end{cases} \tag{30d}$$

In other words, $d_1(t)$ and $d_2(t)$ are *independent amplitude random data signals* of baud length $2T$, *staggered* so that the transitions in $d_1(t)$ occur at times iT with i odd whereas the transitions in $d_2(t)$ occur at times iT with i even. The data values (+1 or −1) in $d_1(t)$ correspond to those of $d(t)$ in the bauds $iT \leqslant t < (i + 1)T$ with i odd, while the data values in $d_2(t)$ correspond to those of $d(t)$ in the bauds $iT \leqslant t < (i + 1)T$ with i even. In particular, it follows that

$$R_{d_1}(\tau) = E[d_1(t + t_0 + \tau)\, d_1(t + t_0)] = R_{d_2}(\tau) = R_d(\tau/2) \tag{31}$$

Substituting Eqs. (30a) and (30b) into Eq. (26c), we find

$$s(t + t_0) = A\, d_1(t + t_0) \cos\left[\frac{\Delta\omega}{2}(t + t_0)\right] \sin[\omega_0 (t + t_0) + \theta] + A\, d_2(t + t_0) \sin\left[\frac{\Delta\omega}{2}(t + t_0)\right] \cos[\omega_0 (t + t_0) + \theta] \tag{32}$$

It follows from the statistical independence (and zero means) of the amplitudes of $d_1(t + t_0)$ and $d_2(t + t_0)$ that the cross-correlation between the two terms on the right in

Eq. (32) vanishes. Moreover, each of the two terms clearly has the same autocorrelation function so that

$$R_s(\tau) = E[s(t+t_0+\tau)\, s(t+t_0)]$$

$$= A^2 \cos(\omega_0 \tau) R_m(\tau) \tag{33}$$

where $R_m(\tau)$ is the autocorrelation function of the random process

$$m(t+t_0) = d_1(t+t_0) \cos\left[\frac{\Delta\omega}{2}(t+t_0)\right] \tag{34}$$

We now proceed to find

$$R_m(\tau) = E[m(t+t_0+\tau)\, m(t+t_0)] \tag{35}$$

$$= E[m(t_0+\tau)\, m(t_0)] \tag{36}$$

where we have also used the fact that $(\Delta\omega)T = \pi$. Taking Fourier transforms (recalling that $R_m(\tau) = R_m(-\tau)$) in Eq. (38) yields

$$S_m(f) = \frac{T}{2}\left[\text{sinc}\left(2fT - \frac{1}{2}\right) + \text{sinc}\left(2fT + \frac{1}{2}\right)\right]^2$$

$$= \frac{T}{2}\left\{\text{sinc}\left[2\left(f - \frac{\Delta f}{2}\right)T\right] + \text{sinc}\left[2\left(f + \frac{\Delta f}{2}\right)T\right]\right\}^2 \tag{39}$$

where

$$\Delta f = \frac{\Delta\omega}{2\pi} = \frac{1}{2T} \tag{40}$$

Equation (39) is our desired expression for the spectrum of $m(t)$.

It is convenient to take first the conditional expectation in Eq. (36) for a fixed t_0, which gives

$$E\{m(t_0+\tau)\, m(t_0) \mid t_0\} = E\left\{d_1(t_0+\tau)\, d_1(t_0) \cos\left[\frac{\Delta\omega}{2}(t_0+\tau)\right] \cos\left[\frac{\Delta\omega}{2} t_0\right] \middle| t_0\right\}$$

$$= \begin{cases} \cos\left[\frac{\Delta\omega}{2}(t_0+\tau)\right] \cos\left[\frac{\Delta\omega}{2} t_0\right], & 0 \leqslant t_0 < T-\tau \text{ and } T \leqslant t_0 < 2T \\ 0, & T-\tau \leqslant t < T \end{cases} \tag{37}$$

Recalling that t_0 is uniformly distributed over $0 \leqslant t_0 < 2T$, we find that with the aid of Eq. (37) that

$$R_m(\tau) = \frac{1}{2T}\int_0^{2T} E\{m(t_0+\tau)\, m(t_0) \mid t_0\}\, dt_0$$

$$= \begin{cases} \frac{1}{2}\left(1 - \frac{\tau}{2T}\right)\cos\left(\frac{\Delta\omega}{2}\tau\right) + \frac{1}{2\pi}\sin\left(\frac{\Delta\omega}{2}\tau\right), & 0 \leqslant \tau < 2T \\ 0, & \tau \geqslant 2T \end{cases} \tag{38}$$

From Eqs. (33) and (39), we have finally

$$S_s(f) = \frac{1}{4} A^2 T \left\{ \text{sinc} \left[2\left(f - f_0 - \frac{\Delta f}{2}\right) T \right] + \text{sinc} \left[2\left(f - f_0 + \frac{\Delta f}{2}\right) T \right] \right\}^2 + \frac{1}{4} A^2 T \left\{ \text{sinc} \left[2\left(f + f_0 - \frac{\Delta f}{2}\right) T \right] + \text{sinc} \left[2\left(f + f_0 + \frac{\Delta f}{2}\right) T \right] \right\}^2 \quad (41)$$

which is the well-known spectrum of MSK, but in a form perhaps more transparent than the usual one. This spectrum is sketched in Fig. 9 for comparison with that of SQPSK.

We remark here that the phase angle θ in Eq. (26c) is arbitrary. Thus, the MSK spectrum of Eq. (41) does not require any special phase relationship between the modulating signal and the carriers $c_x(t)$ and $c_y(t)$.

The MSK spectrum does, of course, depend on the coherency between the modulating signals, $x(t)$ and $y(t)$, and the carriers, $\sin[(\omega_0 + \Delta\omega/2)\, t + \theta]$ and $\sin[(\omega_0 - \Delta\omega/2)\, t + \theta]$, in Eq. (25). If there were no cross-correlation between the two terms on the right in Eq. (26b), then the autocorrelation function of $s(t + t_0)$ would be simply

$$\frac{1}{2} A^2 R_x(\tau) \cos\left[\left(\omega_0 + \frac{\Delta\omega}{2}\right)\tau\right] + \frac{1}{2} A^2 R_y(\tau) \cos\left[\left(\omega_0 - \frac{\Delta\omega}{2}\right)\tau\right]$$

$$= \frac{1}{2} A^2 R_x(\tau) \left\{ \cos\left[\left(\omega_0 + \frac{\Delta\omega}{2}\right)\tau\right] + \cos\left[\left(\omega_0 - \frac{\Delta\omega}{2}\right)\tau\right] \right\}$$

and hence the power spectrum of $S(t + t_0)$ would be

$$\frac{1}{4} A^2 \left[S_x\left(f - f_0 - \frac{\Delta f}{2}\right) + S_x\left(f - f_0 + \frac{\Delta f}{2}\right) + S_x\left(f + f_0 + \frac{\Delta f}{2}\right) + S_x\left(f + f_0 - \frac{\Delta f}{2}\right) \right]$$

$$= \frac{1}{4} A^2 T \left\{ \text{sinc}^2 \left[2\left(f - f_0 - \frac{\Delta f}{2}\right) T \right] + \text{sinc}^2 \left[2\left(f - f_0 + \frac{\Delta f}{2}\right) T \right] \right\}$$

$$+ \frac{1}{4} A^2 T \left\{ \text{sinc}^2 \left[2\left(f + f_0 + \frac{\Delta f}{2}\right) T \right] + \text{sinc}^2 \left[2\left(f + f_0 - \frac{\Delta f}{2}\right) T \right] \right\} \quad (42)$$

Comparing Eqs. (41) and (42), we see that the cross-correlation between the two terms on the right in Eq. (27) has the effect of *changing the spectrum from a sum of the squares of two sinc functions* (as in Eq. (42)) to the square of the sum of the same two sinc functions. Moreover, outside the central lobe of $S_x(f)$ (i.e., outside the range $f_0 - 3/4T < f < f_0 + 3/4T$), one of the sinc functions in Eq. (41) is positive and the other is negative for all f, as follows from the fact that their arguments differ by $2(\Delta f)\, T = 1$. Thus, one of these sinc functions always tends to cancel the other outside the central lobe, which is the reason that the MSK spectrum drops off much more rapidly than that of SQPSK outside the central lobe.

We have already mentioned that the phase angle θ between the carriers and the modulating signals has no effect on the MSK spectrum. In fact, more is true. The phase angle θ in Eq. (26b) can even be time-varying, i.e., a random process $\theta(t + t_0)$, and the above analysis goes through unchanged provided only that $\theta(t)$ not change appreciably over one baud (say, by 0.1 radians or less.) This latter condition is required to insure the vanishing of the expectation of the double carrier frequency terms when computing $R_s(\tau)$.

V. Concluding Remarks

It is well-known (Ref. 8) that MSK modulation can be viewed as "offset QPSK" modulation in which the modulating signals are sinusoidal pulses in each band rather than rectangular pulses – this follows immediately from Eq. (32) above. Simon (Ref. 9) has offered a generalization of MSK, based on this fact, by considering more general pulse shaping.

By expanding on the left and using the usual trigonometric identities, one easily proves the identity

$$\left[\text{sinc}\left(Z - \frac{1}{2}\right) + \text{sinc}\left(Z + \frac{1}{2}\right) \right]^2 = \frac{1}{2\pi^2} \frac{1 + \cos(2\pi Z)}{\left(Z^2 - \frac{1}{4}\right)^2} \quad (43)$$

With the aid of Eq. (43), Eq. (39) can be rewritten as

$$S_m(f) = \frac{4T}{\pi^2} \frac{1 + \cos(4\pi f T)}{(16 f^2 T^2 - 1)^2} \quad (44)$$

which is the usual baseband form of the spectrum of MSK and shows that the spectrum drops off as $1/f^4$ for large f (in contrast to that of QPSK which drops off only as $1/f^2$).

Our aim in this report has been to study MSK in particular from a new standpoint with the hope of isolating the fundamental structure that accounts for its spectral improvement over QPSK. The "sequence transducer" approach that we have taken appears amenable to new generalizations of MSK signalling by different choices of the sequence transducer and carriers. This seems to us to warrant further investigation.

Finally, we remark that several recent papers, e.g., (Refs. 10 and 11), have considered a "phase trellis" approach to MSK and other forms of continuous phase modulation. (Our approach is closely related to these but somewhat more general since we can admit arbitrary "carrier signals" when converting the trellis of the sequential transducer to a "waveform trellis.") Anderson and Taylor (Ref. 11) in particular show substantial power improvements (2 to 4 dB) over QPSK together with a smaller bandwidth. However, their modulation scheme introduces "convolutional code type" constraints into the modulated signal so that demodulation over a fixed number of bauds is not optimal (as it is for MSK). At this writing, it is not clear whether these schemes actually exploit the potential "coding gain" of simpler modulation formats and thus cannot be further encoded efficiently, or whether in fact they actually utilize bandwidth more efficiently than simpler modulation schemes. It seems to us that this is an important question to settle.

Acknowledgment

The author is indebted to Dr. James R. Lesh for several helpful suggestions in the course of the research described here.

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.

7. L. Lee, "On Optimum Soft-Decision Modulation," *IEEE Trans. Info. Th.*, Vol. IT-22, pp. 437-444, July 1976.

8. S. A. Gronemeyer and A. L. McBride, "MSK and Offset QPSK Modulation," *IEEE Trans. Comm.*, Vol. COM-24, pp. 809-820, Aug. 1976.

9. M. K. Simon, "A Generalization of Minimum-Shift-Keying (MSK)-type Signalling Based Upon Input Data Symbol Pulse Shaping, *IEEE Trans. Comm.*, Vol. COM-24, pp. 845-856, Aug. 1976.

10. T. Aulin and C. E. Sundberg, *Binary CPFSK Type of Signalling with Input Data Symbol Pulse Shaping – Error Probability and Spectrum*, Technical Report TR-99, TTT, Dept. of Elec. Engr., Lund University, Sweden, April 1978.

11. J. B. Anderson and D. P. Taylor, "A Bandwidth-Efficient Class of Signal-Space Codes," *IEEE Trans. Info. Th.*, Vol. IT-24, pp. 703-712, Nov. 1978.

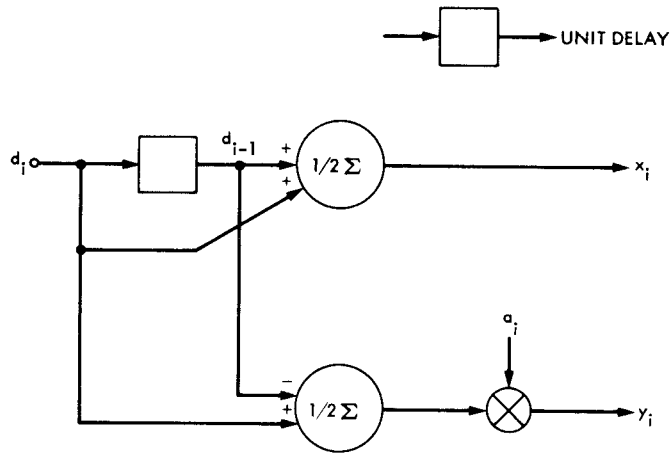


Fig. 1. The sequence transducer

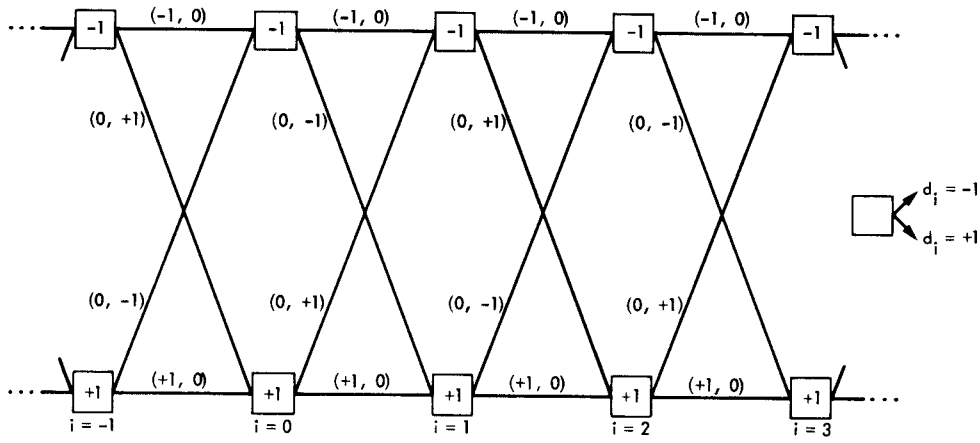


Fig. 2. Trellis for the sequence transducer of Fig. 1

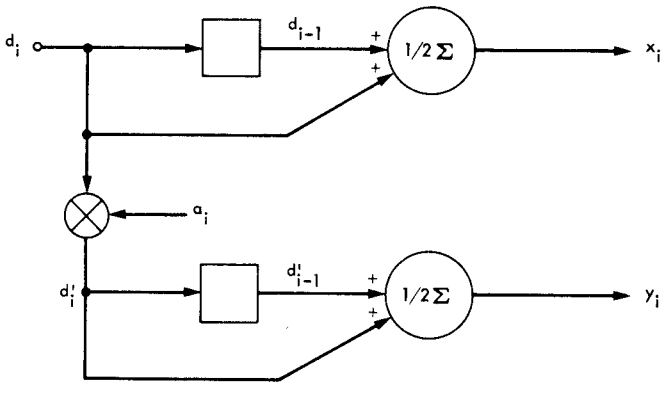
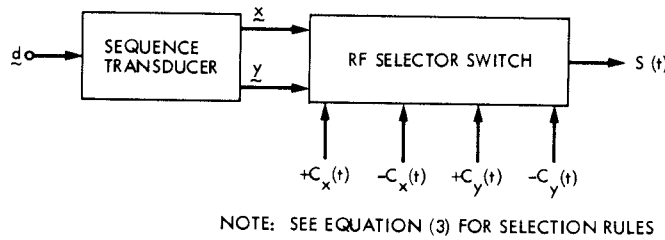


Fig. 3. An alternative sequence transducer equivalent to that in Fig. 1



NOTE: SEE EQUATION (3) FOR SELECTION RULES

Fig. 4. Proposed modulator

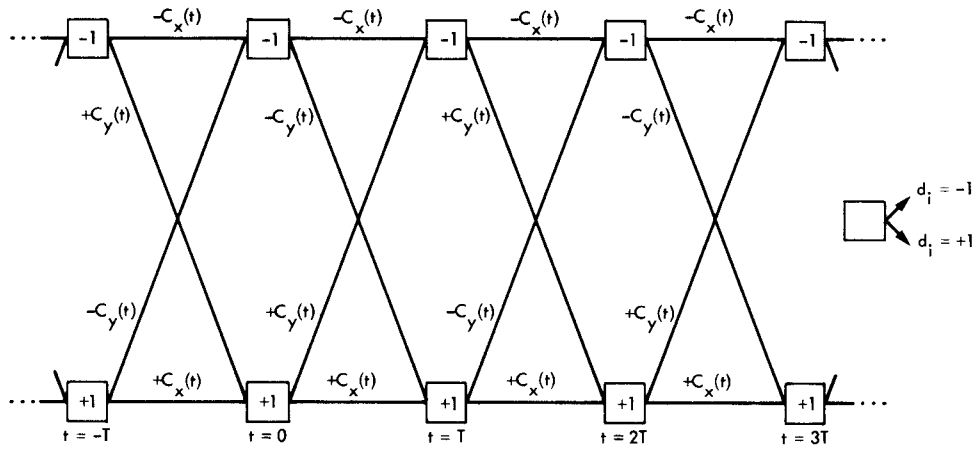


Fig. 5. Waveform trellis for the modulator of Fig. 4

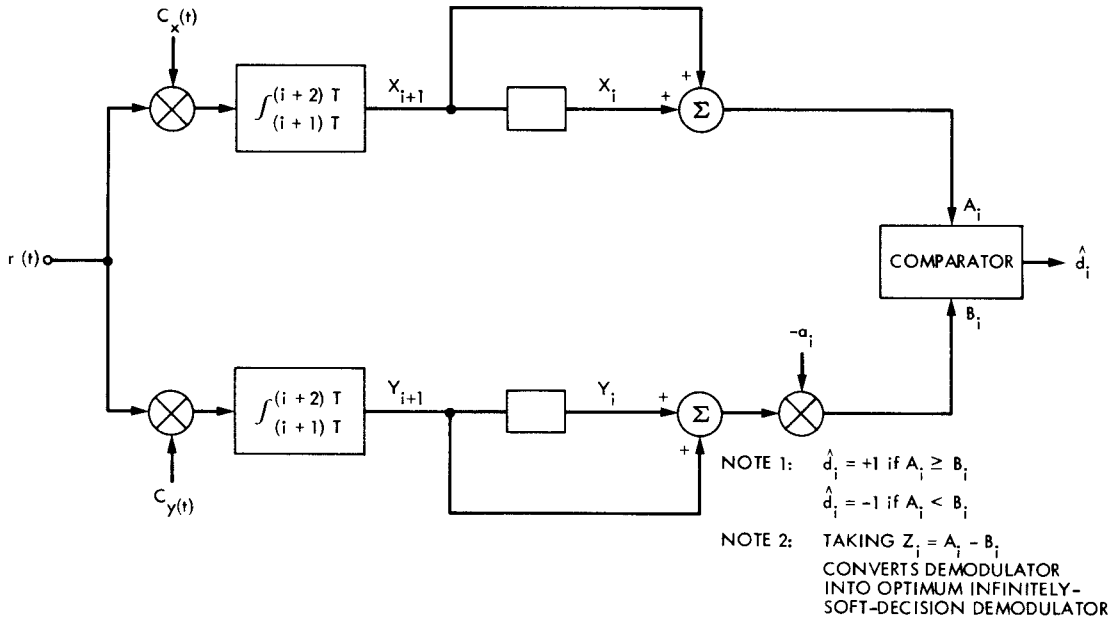


Fig. 6. Optimum hard-decision demodulator for the modulator of Fig. 4

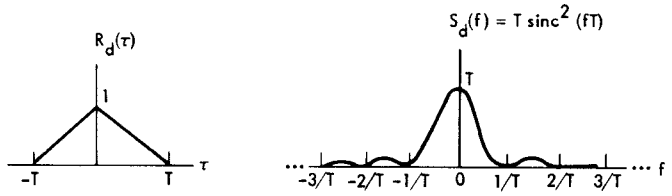


Fig. 7. The autocorrelation function $R_d(\tau)$ and the power spectral density $S_d(f)$ of the random data signal $d(t + t_0)$

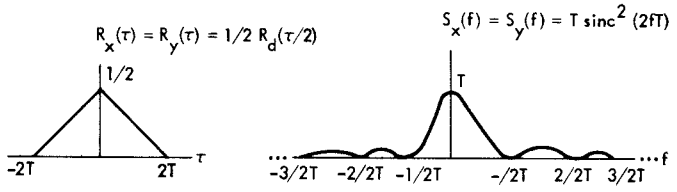


Fig. 8. The autocorrelation functions and power spectra of the outputs $x(t)$ and $y(t)$ of the sequence transducer of Fig. 1

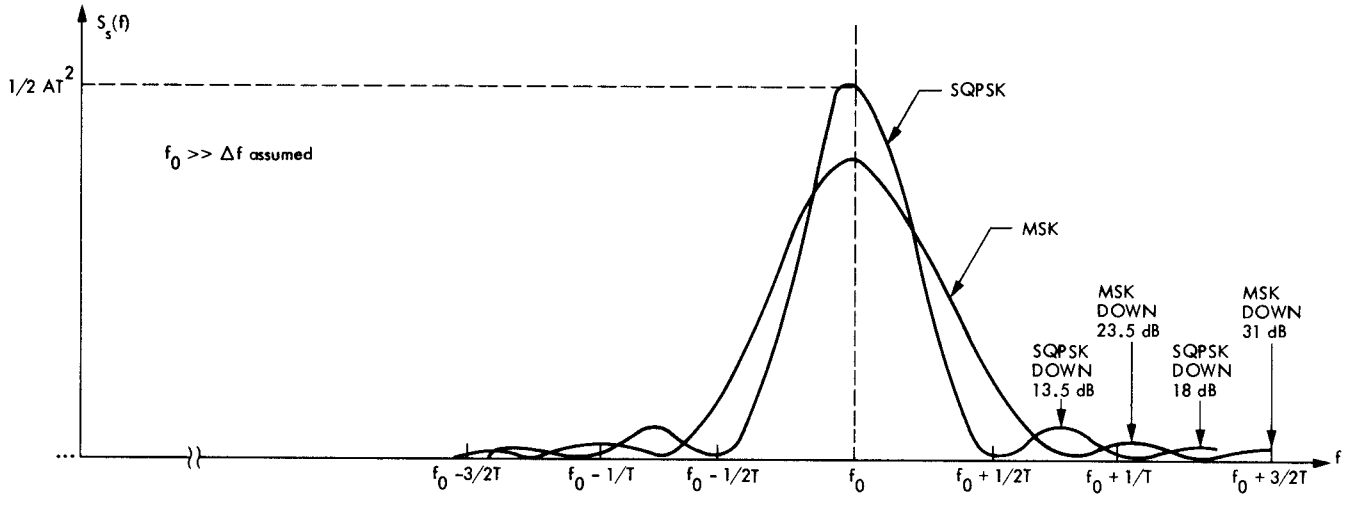


Fig. 9. Power spectra of MSK and SQPSK showing attenuation relative to spectral peak at sidelobe center frequencies