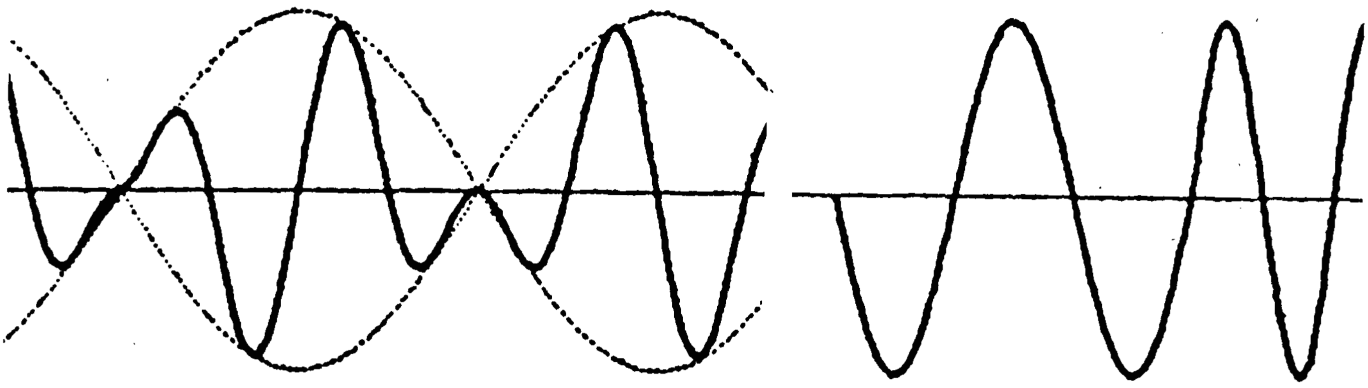


MINIMUM SHIFT KEYING: A SPECTRALLY EFFICIENT MODULATION

Subbarayan Pasupathy

Compact power spectrum, good error rate performance, and easy synchronization make MSK an attractive digital modulation technique.

The ever increasing demand for digital transmission channels in the radio frequency (RF) band presents a potentially serious problem of spectral congestion and is likely to cause severe adjacent and cochannel interference problems. This has, in recent years, led to the investigation of a wide variety of techniques for solving the problem of spectral congestion. Some solutions to this problem include: 1) new allocations at high frequencies; 2) better management of existing allocations; 3) the use of frequency-reuse techniques such as the use of narrow-beam antennas and dual polarizing systems; 4) the use of efficient source encoding techniques; and 5) the use of spectrally efficient modulation techniques [1]. This article will consider the last approach and analyze, in particular, a modulation scheme known as minimum shift keying (MSK). The MSK signal format will be explained and its relation to other schemes such as quadrature phase shift keying (QPSK), offset QPSK (OQPSK), and frequency shift keying (FSK) pointed out. The main attributes of MSK, such as constant envelope, spectral efficiency, error rate performance of binary PSK, and self-synchronizing capability will all be explained on the basis of the modulation format.

The author is with the Department of Electrical Engineering, University of Toronto, Toronto, Ont., Canada.

SPECTRAL EFFICIENCY AND MSK

In any communication system, the two primary communication resources are the transmitted power and channel bandwidth. A general system-design objective would be to use these two resources as efficiently as possible. In many communication channels, one of the resources may be more precious than the other and hence most channels can be classified primarily as power-limited or band-limited. (The voice-grade telephone circuit, with approximately 3 kHz bandwidth, is a typical band-limited channel, whereas space communication links are typically power limited). In power-limited channels, coding schemes would be generally used to save power at the expense of bandwidth, whereas in band-limited channels "spectrally efficient modulation" techniques would be used to save bandwidth.

The primary objective of spectrally efficient modulation is to maximize the bandwidth efficiency, measured in bits/s/Hz.

The primary objective of spectrally efficient modulation is to maximize the *bandwidth efficiency*, defined as the ratio of data rate to channel bandwidth (in units of bits/s/Hz). A secondary objective of such modulation

0163-6804/79/0700-0014$00.75 © 1979 IEEE

schemes may be to achieve this bandwidth efficiency at a prescribed average bit error rate with minimum expenditure of signal power. Some channels may have other restrictions and limitations which may force other constraints on the modulation techniques. For example, communication systems using certain types of nonlinear channels call for an additional feature, namely, a constant envelope, which makes the modulation impervious to such impairments. This is needed because a memoryless nonlinearity produces extraneous sidebands when passing a signal with amplitude fluctuations. Such sidebands introduce out-of-band interference with other communication systems.

A typical example where such considerations are appropriate is in time-division multiple access (TDMA) satellite communication, where the traveling wave tube (TWT) amplifiers are operated near power saturation for high efficiency. For frequency division multiple access (FDMA) application also, constant envelope properties are useful at each ground terminal, if the high power amplifier is operated near power saturation like a class "C" device, where the response would be nonlinear. In this article, we will be concerned with these and similar applications which call for a constant envelope, bandwidth-efficient, digital-modulation technique. Recent investigations into signaling schemes for such applications have centered upon MSK.

Modulation studies during the late 1960's led to the development of MSK [2], [3]. MSK was used by the Data Transmission Co. (Datran) for its proposed data network in 1972 [4]. Other applications which have considered and/or used MSK since then include a proposed AT&T domestic satellite system [5],[6], military tactical radio [7], extremely low frequency (ELF) underwater communication systems [8], and Canadian communications technology satellite (CTS) experiments [9]. We will examine the major attributes of MSK which make it a suitable candidate for such applications. We begin with a brief review of some related modulation techniques such as FSK, PSK, QPSK, and OQPSK.

FSK and PSK

The constraint of a constant envelope feature for the modulation scheme narrows the search to two major signaling techniques, namely, FSK and PSK. Consider binary communication—transmitting a pulse every T seconds (at the signaling rate of $1/T$ baud) to denote one of two equally likely information symbols, +1 or −1. FSK denotes the two states by transmitting a sinusoidal carrier at one of two possible frequencies, whereas binary PSK (BPSK) uses the two opposite phases of the carrier (i.e., 0 and 180°). Fig. 1 shows typical signals in the two types of modulation. Note that BPSK is also equivalent to amplitude modulating the carrier with the information bit stream, i.e., multiplication with +1 or −1. The two schemes can be compared on the basis of their bit error rate (BER) performance (i.e., the average number of errors in transmitting a long bit stream) through an ideal channel. The ideal channel is taken to be a linear all-pass channel, corrupted only by additive white Gaussian noise with a constant (one-sided) power spectral density of N_0 W/Hz. The required ratio, E_b/N_0, of signal energy per bit (E_b) and noise level N_0 to achieve a given BER (such as 1 error in 10^5 bits) is the quantity of interest.

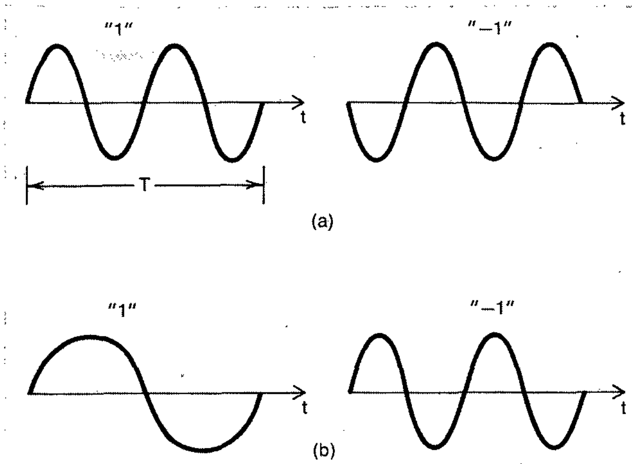


Fig. 1. (a) Binary phase shift keying; (b) Frequency shift keying.

It can be shown that optimum receivers for binary signals in such channels call for matched filters (also known as correlator-detectors) with perfect carrier phase reference available at the receiver. For such coherent receivers, which base their bit decisions after observing the signal over T seconds, there exists a class of signals, of which PSK is one, which turns out to be optimum in the sense of requiring the minimum amount of E_b/N_0 for a specified BER. This optimum class of signals is called "antipodal," i.e., the two signals denoting the two possible information symbols have exactly the same shape but opposite polarity. On the other hand, from the viewpoint of simpler receiver implementation, noncoherent detection schemes, which do not make use of the carrier phase reference, can also be used. For example, a noncoherent detector for FSK can use two bandpass filters tuned to the two frequencies, followed by envelope detectors and bit-rate samplers, and base the binary decision on which of the two sampled envelopes is the larger. It can be shown that for such noncoherent detection, the optimum class of signals is the class of noncoherently orthogonal signals. (Orthogonal signals are those that do not interfere with one another in the process of detection. If the demodulation process is coherent, they can be called coherently orthogonal and if noncoherent detection is employed, noncoherently orthogonal. For example, in noncoherent FSK signaling using envelope detectors, the two FSK signals at frequencies f_1 and f_2 are said to be noncoherently orthogonal, if, when a tone at f_1 is transmitted, the sampled envelope of the output of the receiving filter tuned to f_2 is zero, i.e., no cross talk.) In the case of FSK, the minimum separation between the two frequencies equals the signaling rate $1/T$ for noncoherent orthogonality. The performance of such a noncoherently orthogonal FSK scheme is much poorer than the coherent antipodal PSK scheme. Even if coherently orthogonal FSK is used and detected by coherent methods in an effort to improve performance, it is still poorer by 3-dB (in terms of E_b/N_0) than PSK.

The poorer performance of FSK has been responsible for restricting the use of FSK mainly to low-data-rate low-efficiency applications, while PSK has been the preferred scheme for efficient higher-data-rate applications. We will show later that this is not a fair assessment of FSK. In fact, MSK, a coherently orthogonal FSK modulation scheme, requires only $1/2T$ Hz frequency separation and achieves a performance equivalent to that of BPSK, when a coherent receiver bases its decision after observing the signal over 2 bit periods ($2T$ seconds) rather than just one. Before we view MSK as a particular case of FSK, it is helpful to understand several other PSK schemes, such as QPSK and OQPSK, and to view MSK as a particular variation of OQPSK signaling.

QPSK and OQPSK

The optimum E_b/N_0 performance achievable with BPSK led to a search for mechanisms to improve the bandwidth efficiency of PSK schemes without any loss of performance. It was found that since $\cos 2\pi f_c t$ and $\sin 2\pi f_c t$ (where f_c is the carrier frequency) are coherently orthogonal signals, two binary bit streams modulating the two carrier signals in quadrature can be demodulated separately. (In analog communication, this idea has been used for a long time to multiplex two signals on the same carrier, so as to occupy the same bandwidth, e.g., the two chrominance signals in color television are modulated onto the color subcarrier this way). Such a modulation scheme, increasing the bandwidth efficiency of binary PSK by two, is known as QPSK and is shown in Fig. 2 (a).

The input binary bit stream $\{a_k\}$, ($a_k = \pm 1$) $k = 0,1,2, \cdots$ arrives at a rate of $1/T$ baud and is separated into two streams $a_I(t)$ and $a_Q(t)$ consisting of even and odd bits, respectively, as shown in the example waveforms of Fig. 2(a). The two pulse trains modulate the inphase and quadrature components of the carrier and the sum $s(t)$, the modulated QPSK signal, can be represented as

$$s(t) = \frac{1}{\sqrt{2}} a_I(t) \cos\left(2\pi f_c t + \frac{\pi}{4}\right) + \frac{1}{\sqrt{2}} a_Q(t) \sin\left(2\pi f_c t + \frac{\pi}{4}\right). \quad (1)$$

The two terms in (1) represent two binary PSK signals and can be detected independently due to the orthogonality of $\cos(2\pi f_c t + \frac{\pi}{4})$ and $\sin(2\pi f_c t + \frac{\pi}{4})$. Using a well-known trigonometric identity, (1) can also be written as

$$s(t) = \cos(2\pi f_c t + \Theta(t)) \quad (2)$$

where, as shown in Fig. 3, $\Theta(t) = 0°, \pm 90°$ or $180°$ corresponding to the four combinations of $a_I(t)$ and $a_Q(t)$.

The OQPSK signaling can also be represented by (1) and (2) and the difference between the two modulation techniques is only in the alignment of the two bit streams. The odd and even bit streams, transmitted at the rate of $1/2T$ baud, are synchronously aligned in QPSK [as shown in Fig. 2(a)] such that their transitions coincide. OQPSK modulation is obtained by a shift or offset in the relative alignments of $a_I(t)$ and $a_Q(t)$ data streams by an amount equal to T. Fig. 2(b) shows the offset. (OQPSK is also sometimes referred to as staggered QPSK.)

The difference in time alignment in the bit streams does not change the power spectral density and hence both QPSK and OQPSK spectra have the same $(\sin 2\pi fT/2\pi fT)^2$ shape, associated with the rectangular pulse used for signaling. However, the two modulations respond differently when they undergo bandlimiting and

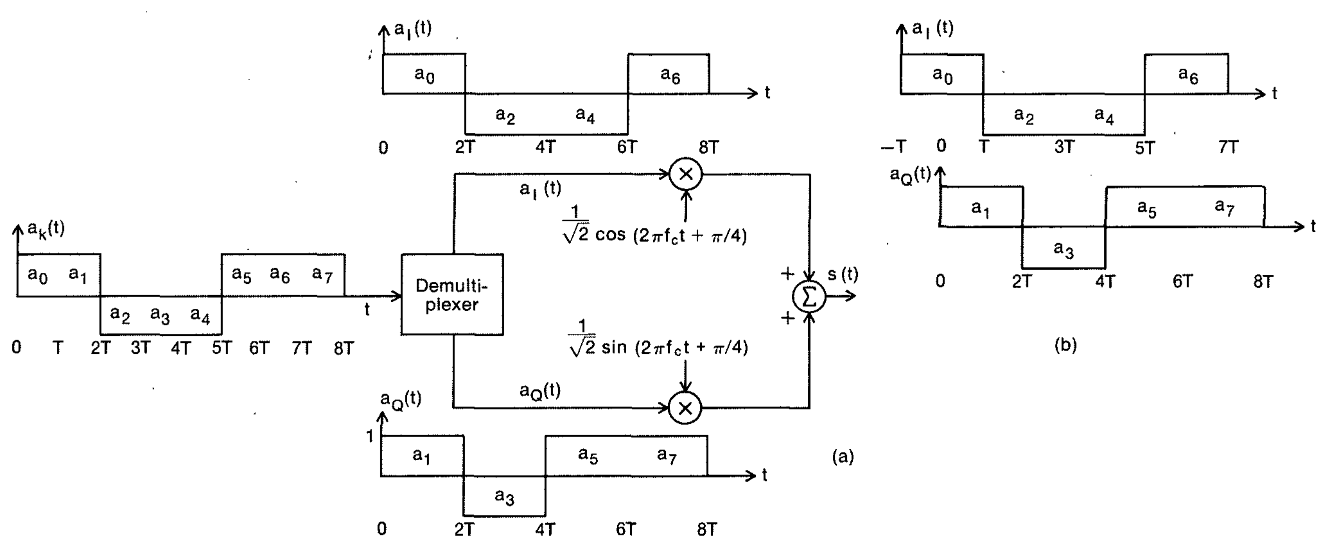


Fig. 2. (a) QPSK modulator; (b) Staggering of data streams in OPQSK.

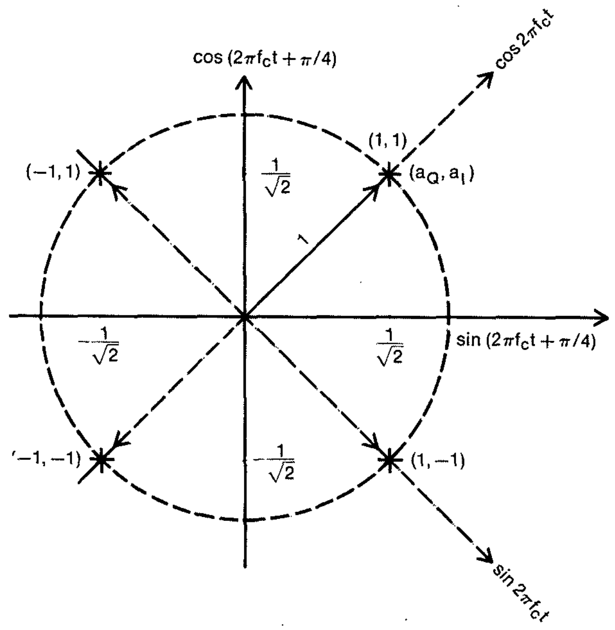


Fig. 3. Signal space diagram for QPSK and OPQSK.

hardlimiting operations, encountered in applications such as satellite communications. The difference in the behavior of the two modulations can be understood by a study of the phase changes in the carrier in the two modulations.

In QPSK, due to the coincident alignment of $a_I(t)$ and $a_Q(t)$ streams, the carrier phase can change only once every $2T$. The carrier phase over any $2T$ interval is any one of the four phases shown in Fig. 3 depending on the values of $\{a_Q(t), a_I(t)\}$. In the next $2T$ interval, if neither bit stream changes sign, the carrier phase remains the same. If one component ($a_I(t)$ or $a_Q(t)$) changes sign, a phase shift of ±90° occurs. A change in both components results in a phase shift of 180°. Fig. 4(a) shows a typical QPSK signal waveform, for the sample sequence $a_I(t)$ and $a_Q(t)$ shown in Fig. 2(a). In a satellite communication system, the modulated signal shown in Fig. 4(a) is bandlimited by a bandpass filter so as to conform to out-of-band spectral emission standards. The bandlimited QPSK will no longer have constant envelope and in fact, the occasional phase shifts of π rad in the carrier will make the envelope go to zero [10]. At the satellite repeater, this signal will undergo hardlimiting which, while restoring the constant envelope to the signal, will at the same time restore essentially all the frequency sidelobes back to their original level prior to filtering. These undesired sidebands negate the bandlimiting of the QPSK signal carried out at the transmitter and introduce out-of-band radiation on the satellite downlink that may interfere with other communication systems. On the other hand, bandlimiting and hardlimiting operations do not seem to produce the same deleterious effect on an OQPSK signal.

In OQPSK, the binary components cannot change states simultaneously. One component has transitions in the middle of the other symbol and hence only one component can switch at a time. This eliminates the possibility of 180° phase changes and phase changes are limited to 0°, ± 90° every T seconds. Fig. 4(b) shows a typical OQPSK waveform for the example bit streams in Fig. 2(b). When a OQPSK signal undergoes bandlimiting, the resulting intersymbol interference (smearing of adjacent pulses on one another) causes the envelope to droop slightly in the region of $\pi/2$ rad phase transitions. Since phase shifts of 180° have been avoided, the envelope does not go to zero as it does in the bandlimited QPSK case. When the bandlimited OQPSK goes through a hard limiter, the slight envelope droop is removed by the limiting process. However, limiting affects only the envelope and the phase is preserved. Consequently, the absence of rapid phase shifts (and hence high frequency content) in the region of a $\pi/2$ phase change means that limiting will not regenerate the high frequency components originally removed by the bandlimiting filter. Thus, out-of-band interference is avoided.

Tests [10] indicate that unlike QPSK, the spectrum of OQPSK after limiting remains essentially unchanged and seems to retain its bandlimited nature in almost its entirety. OQPSK signals also seem to perform better than QPSK in the presence of phase jitter associated with noisy reference carriers [11]. Furthermore, the offset of T s between the 2 bit streams in OQPSK has been shown to be optimum in terms of phase jitter immunity in the presence of additive Gaussian noise [12]. All these advantages possessed by OQPSK stem mainly from the fact that OQPSK avoids the large phase change of 180° associated with the QPSK format. This suggests that further suppression of out-of-band interference in bandlimiting-hardlimiting applications can be obtained, if the OQPSK signal format can be modified to avoid phase transitions altogether. This can be thought of as an obvious motivation for designing constant envelope modulation schemes with continuous phase. MSK is one such scheme, as will be discussed next.

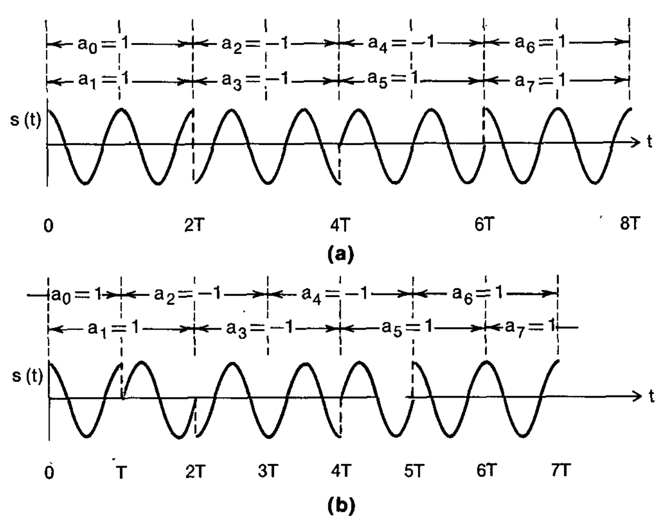


Fig. 4. (a) QPSK waveform; (b) OPQSK waveform.

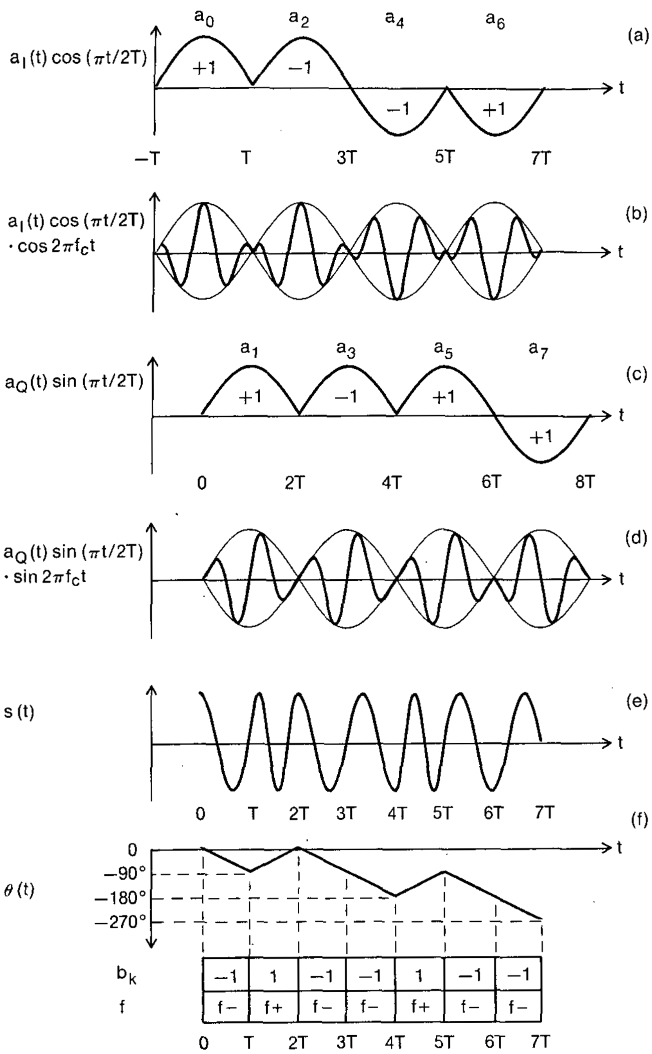


Fig. 5. MSK waveforms.

MSK can be viewed as a form of offset quadrature phase shift keying with a half-sinusoidal rather than rectangular weighting.

MSK

MSK can be thought of as a special case of OQPSK with sinusoidal pulse weighting [13],[14]. Consider the OQPSK signal, with the bit streams offset as shown in Fig. 2(b). If sinusoidal pulses are employed instead of rectangular shapes, the modified signal can be defined as MSK and equals

$$s(t) = a_I(t)\cos\left(\frac{\pi t}{2T}\right)\cos 2\pi f_c t + a_Q(t)\sin\left(\frac{\pi t}{2T}\right)\sin 2\pi f_c t. \quad (3)$$

Fig. 5 shows the various components of the MSK signal defined by (3). Fig. 5(a) shows the modified in-phase bit stream waveform, for the sample $\{a_I\}$ stream shown in Fig. 2(b). The corresponding values of the even-bits are shown as ± 1 inside the waveform. The in-phase carrier [the first term in (3)], obtained by multiplying the waveform in Fig. 5(a) by $\cos 2\pi f_c t$, is shown in Fig. 5(b). Similarly, the sinusoidally shaped odd-bit stream and the quadrature carrier are shown in Fig. 5(c) and (d). The composite MSK signal $s(t)$, the addition of Fig. 5(b) and (d), is shown in Fig. 5(e). The waveform in Fig. 5(e) can be better understood if we use a well-known trigonometric identity to rewrite (3) as

$$s(t) = \cos[2\pi f_c t + b_k(t)\frac{\pi t}{2T} + \phi_k] \quad (4)$$

where b_k is $+1$ when a_I and a_Q have opposite signs and b_k is -1 when a_I and a_Q have the same sign and ϕ_k is 0 or π corresponding to $a_I = 1$ or -1. Note that $b_k(t)$ can also be written as $-a_I(t)a_Q(t)$.

From Fig. 5(e) and (4), we deduce the following properties of MSK.

1) It has constant envelope.

2) There is phase continuity in the RF carrier at the bit transition instants.

3) The signal is an FSK signal with signaling frequencies $f_+ = f_c + 1/4T$ and $f_- = f_c - 1/4T$. Hence the frequency deviation equals half the bit rate, i.e., $\Delta f = f_+ - f_- = 1/2T$.
This is the minimum frequency spacing which allows the two FSK signals to be coherently orthogonal, in the sense discussed in the section on FSK and PSK; hence the name "minimum shift" keying. Since the frequency spacing is only half as much as the conventional $1/T$ spacing used in noncoherent detection of FSK signals, MSK is also referred to as Fast FSK [3].[1]

4) The excess phase of the MSK signal, referenced to the carrier phase, is given by the term

$$\Theta(t) = b_k(t)\frac{\pi t}{2T} = \pm\frac{\pi t}{2T}$$

in (4), which increases or decreases linearly during each bit period of T seconds. A bit b_k of $+1$ corresponds to an increase of the carrier phase by 90° and corresponds to an FSK signal at the higher frequency f_+. Similarly, $b_k = -1$ implies a linear decrease of phase by 90° over T s, corresponding to the lower frequency f_-. (In order to make the phase continuous at bit transitions, the carrier frequency f_c should be chosen such that f_c is an integral multiple of $1/4T$, one-fourth the bit rate.) The excess phase $\Theta(t)$ is shown in Fig. 5(f), with the corresponding frequencies and values of b_k shown below.

Thus MSK can be viewed either as an OQPSK signal with sinusoidal pulse weighting or as a continuous phase (CPFSK) signal with a frequency separation equal to one-half the bit rate.

PULSE SHAPING AND POWER SPECTRA

The power spectra of QPSK, OQPSK, and MSK (shifted to baseband) can all be expressed by the magnitude squared of $P(f)$, the Fourier transform of the symbol shaping function $p(t)$. Thus, for QPSK and OQPSK [see (1)]

[1]There are several forms of MSK described in the literature; however, all of them are spectrally equivalent to the version described here. Most of them differ in details, such as bit-to-symbol precoding and the use of differential encoding.

$$p(t) = \begin{cases} \frac{1}{\sqrt{2}} & |t| \leq T \\ 0 & \text{elsewhere.} \end{cases} \quad (5)$$

and for MSK [see (3)]

$$p(t) = \begin{cases} \cos\left(\frac{\pi t}{2T}\right) & |t| \leq T. \\ 0 & \text{elsewhere.} \end{cases} \quad (6)$$

Thus the spectral density $G(f)$ for QPSK and OQPSK (normalized to have the same power) is given by

$$\frac{G(f)}{T} = 2\left(\frac{\sin 2\pi fT}{2\pi fT}\right)^2 \quad (7)$$

and for MSK is given by

$$\frac{G(f)}{T} = \frac{16}{\pi^2}\left(\frac{\cos 2\pi fT}{1\text{-}16f^2T^2}\right)^2. \quad (8)$$

The spectra are sketched in Fig. 6.

The difference in the rates of falloff of these spectra can be explained on the basis of the smoothness of the pulse shape $p(t)$. The smoother the pulse shape, the faster is the drop of spectral tails to zero. Thus, MSK, having a smoother pulse, has lower sidelobes than QPSK and OQPSK. A measure of the compactness of a modulation spectrum is the bandwidth B which contains 99 percent of the total power. For MSK, $B \cong (1.2/T)$ while for QPSK and OQPSK, $B \cong (8/T)$ [13]. This indicates that in relatively wide-band satellite links (where, for example, filtering is not used after the nonlinearities), MSK may be spectrally more efficient than QPSK or OQPSK. However, as can be seen from Fig. 6, the MSK spectrum has a wider mainlobe than QPSK and OQPSK. This suggests that in narrow-band satellite links, MSK may not be the preferred method. Computer simulations [15], which take into account all relevant parts of typical wide-band and narrow-band TDMA satellite links, tend to support the above conclusions. Assuming the transmitter and satellite amplifiers to be operating at power saturation, results [15] show that MSK gives superior performance to QPSK when α, the product of channel spacing and symbol duration exceeds 1.8 and to OQPSK only when α exceeds 2.3. OQPSK is shown to have superior performance to QPSK except when α is less than 1.4. However, it should be pointed out that in realistic system applications [13], [15], the difference in the required (E_b/N_0) for the three schemes seems to be less than 1 dB and the choice of modulation method depends on other, less obvious, criteria.

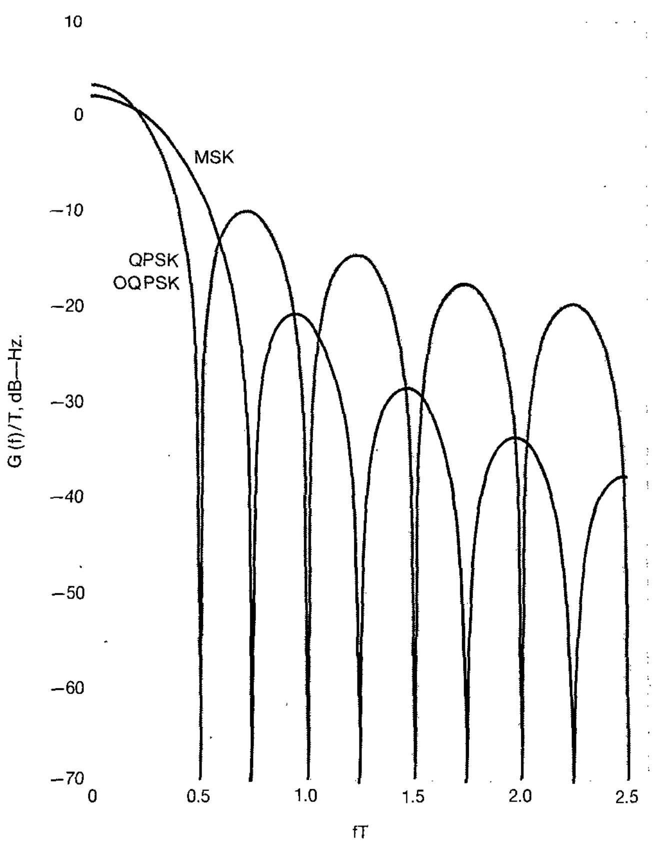


Fig. 6. Spectral density of QPSK, OPQSK, and MSK.

MSK has some excellent special properties that make it an attractive alternative when other channel constraints require bandwidth efficiencies below 1.0 bit/s/Hz. For example, the continuous phase nature of MSK makes it highly desirable for high-power transmitters driving highly reactive loads [8]. Since intersymbol switching occurs when the instantaneous amplitude of $p(t)$ is zero, the finite rise and fall times and data asymmetry inevitably present in practical situations have a minimal effect on the MSK performance [16]. In addition, as we shall see below, MSK has simple demodulation and synchronization circuits.

MSK TRANSMITTER AND RECEIVER

A typical MSK modulator is shown in Fig. 7 [6],[9],[16]. The multiplier produces two phase coherent signals at frequencies f_+ and f_-. The advantage of this method of forming the binary FSK signals is that the signal coherence and the deviation ratio are largely unaffected by variations in the data rate [9]. The binary FSK signals, after being separated by means of narrow bandpass filters, are properly combined to form the in-phase and quadrature carrier components. These carriers are multiplied with the odd and even bit streams $a_I(t)$ and $a_Q(t)$ [which are offset by T, as in Fig. 2(b)] to produce the MSK modulated signal $s(t)$ [as defined in (3)].

The block diagram of a typical MSK receiver is shown in Fig. 8. The received signal (equal to $s(t)$ of (3) in the absence of noise and intersymbol interference) is multiplied by the respective in-phase and quadrature carriers $x(t)$ and $y(t)$ followed by integrate and dump circuits. The multiplier-integrator constitutes correlation detection or matched filtering, an optimum coherent receiver in the absence of intersymbol interference. Note the integration interval of $2T$ s. The demodulation works as follows: if $s(t)$ of (3) is multiplied by $x(t)$ (= $\cos\pi t/2T \cos 2\pi f_c t$), the low frequency component of the result

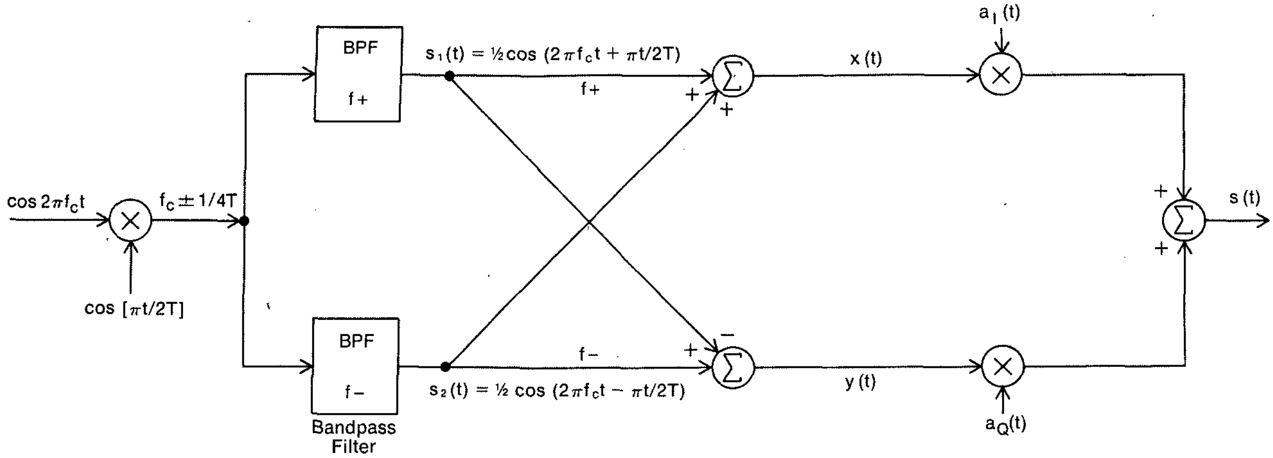


Fig. 7. MSK modulator: $x(t) = \cos(\pi t/2T)\cos 2\pi f_c t$; $y(t) = \sin(\pi t/2T)\sin 2\pi f_c t$.

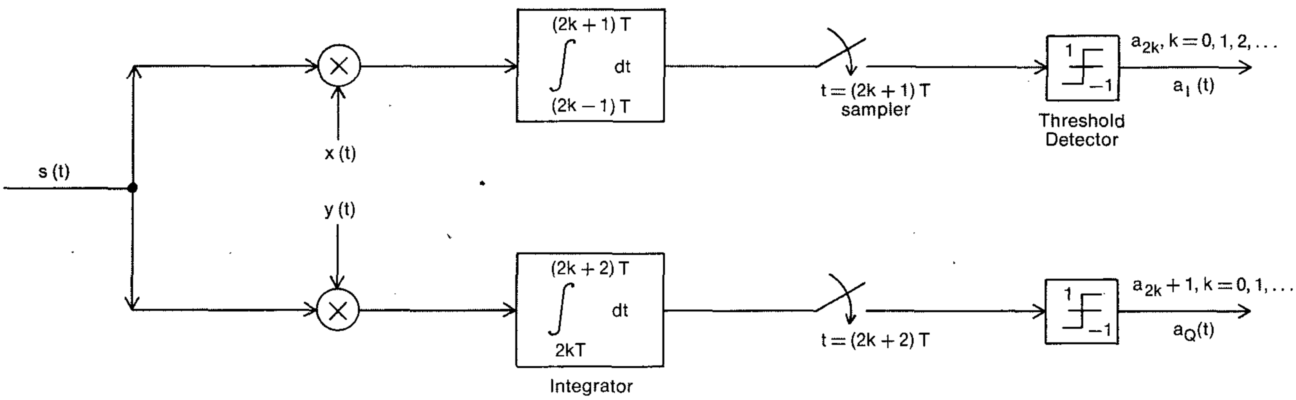


Fig. 8. MSK receiver: $x(t) = \cos(\pi t/2T)\cos 2\pi f_c t$; $y(t) = \sin(\pi t/2T)\sin 2\pi f_c t$.

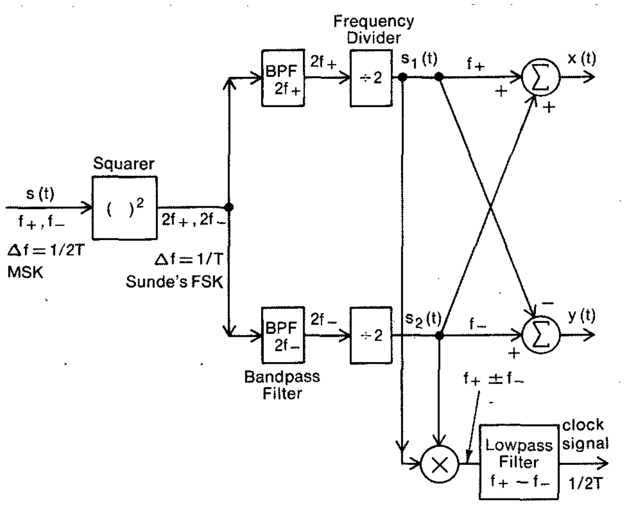


Fig. 9. Synchronization circuits for MSK: $f_+ = f_c + 1/4T$; f_- $f_c - 1/4T$.

(the output of the integrator) equals $a_I(t)(1 + \cos\pi t/T)/4$ and hence the polarity of the sampler output determines the value of $a_I(t)$. The operation on the quadrature channel is similar and determines the values of the odd-bits.

The reference waveforms $x(t)$ and $y(t)$ and the clock signal at ½ the bit rate needed at the samplers are recovered from $s(t)$ as shown in Fig. 9. Although the MSK signal $s(t)$ has no discrete components which can be used for synchronization, it produces strong discrete spectral components at $2f_+$ and $2f_-$ when passed through a squarer. (The squarer, in effect, doubles the modulation index and produces an FSK signal with $\Delta f = 1/T$, known as Sunde's FSK [17]. Sunde's FSK has 50 percent of its total power in line components at the two transmitter frequencies.) These components are extracted by bandpass filters (in practice, by phase-locked loops) and then frequency division circuits produce the signals

$$s_1(t) = \tfrac{1}{2}\cos(2\pi f_c t + \frac{\pi t}{2T}) \qquad (9)$$

and

$$s_2(t) = \tfrac{1}{2}\cos(2\pi f_c t - \frac{\pi t}{2T}), \qquad (10)$$

respectively. The sum and difference $s_1 + s_2$ and $s_2 - s_1$ produce the reference carriers $x(t)$ and $y(t)$, respectively.

If $s_1(t)$ and $s_2(t)$ are multiplied and low-pass filtered, the output is $\frac{1}{4} \cos 2\pi t/2T$ (a signal at ½ the bit rate) which is the desired timing waveform. Thus the MSK format lends itself to very easy self-synchronization.

The natural 180° ambiguity produced by the divide-by-two circuits [i.e., whether the outputs are $\pm s_1(t)$, $\pm s_2(t)$] can be removed by differential encoding and other techniques [14],[18]. (Differential encoding is a technique in which changes or transitions in bit streams, rather than the absolute values themselves, are encoded as +1 and −1. At the receiver, a differential decoder will be employed to generate the original bit stream.)

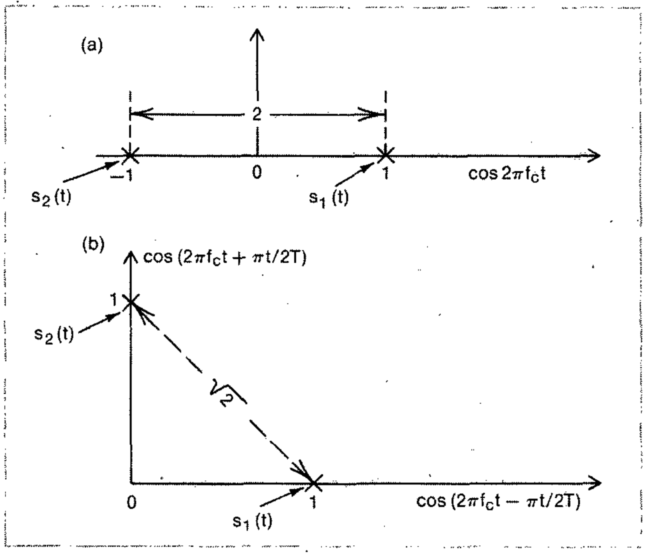


Fig. 10. Signal space diagrams for (a) BPSK (antipodal); and (b) FSK (orthogonal).

ERROR RATE PERFORMANCE

In addition to its bandwidth efficiency and self-synchronizing capabilities, the error rate performance of MSK in an ideal channel, as defined in the section on FSK and PSK is also of interest. It is relevant at this point to summarize the performance capabilities of related binary and quadrature modulations. Binary PSK (BPSK) with $\pm \cos 2\pi f_c t$ representing "1" and "−1," is, as we mentioned before, an example of antipodal signaling and can be represented as shown in Fig. 10(a) with a normalized distance of two between them. In QPSK, the two carriers used ($\cos(2\pi f_c t + \pi/4)$ and $\sin(2\pi f_c t + \pi/4)$) are orthogonal and hence the two bit streams $a_I(t)$ and $a_Q(t)$ can be demodulated independently. Hence for the same (E_b/N_0), the error probabilities of coherently detected BPSK and QPSK are the same. Since staggering the bit streams does not change the orthogonality of the carriers, OQPSK has the same performance as BPSK and QPSK.

MSK (3) uses antipodal symbol shapes ($\pm \cos \frac{\pi t}{2T}$ and $\pm \sin \frac{\pi t}{2T}$) over $2T$ to modulate the two orthogonal carriers just as was the case in QPSK. Also, the energy in the two shapes $p(t)$ in (5) and (6) are the same. Thus, when matched filtering is used to recover the data (as was the case in Fig. 8), MSK has the same performance as BPSK, QPSK, and OQPSK.

Note, however, that it is the detection of the MSK signal on the basis of observation over $2T$ seconds that results in this good performance. If MSK is coherently detected as an FSK signal with bit decision made over an observation interval of T seconds, MSK would be poorer than BPSK by 3 dB. This is because the coherent performance of equal energy binary signals in white Gaussian noise depends only on the "distance" between the two signals in the signal space—the larger the distance, the less the probability of error. (This is intuitively appealing since the larger the distance, the less the possibility of mistaking one signal for the other.) As Fig. 10(b) shows, MSK, viewed as an example of orthogonal FSK signaling, has only a distance of $\sqrt{2}$ between the two signals. This decrease in the distance in FSK as compared to BPSK translates into an (E_b/N_0) increase of 3 dB needed to maintain the same error rate as in a BPSK scheme. This 3 dB disadvantage of FSK over BPSK vanishes in the case of MSK, where bit decisions are made after observing the waveform over two bit periods.

Even though T is the duration of one bit and one decision per transmitted bit is needed, better performance may be obtained by observing the received waveform over a longer period, thus giving us more knowledge about the underlying signal and/or noise process. In MSK, the phases are related over 2 bit periods and hence by observing over $2T$ seconds and using the continuous phase nature of MSK, we know more about the signal. In fact, as is evident from the OQPSK format of MSK (3), we know that over $2T$ seconds, it is one of two antipodal signals which has been transmitted and hence the equivalence in the performance of MSK and BPSK.

Since MSK is a type of FSK, it can also be noncoherently detected (e.g., by means of a discriminator) whereas QPSK systems require either a fully coherent or differentially coherent detection scheme. This possibility of noncoherent detection permits inexpensive demodulation of MSK when the received (E_b/N_0) is adequate and provides a low-cost flexibility feature in some systems.

EXTENSIONS AND GENERALIZATIONS

MSK or continuous phase FSK (CPFSK) may be generalized to include other values of Δf, the frequency separation, and a longer bit memory before the decision has to be made. For larger observation intervals such as $3T$ or $5T$, a maximum improvement of 0.8 dB has been reported for $\Delta f = 0.715/T$ [18]-[20]. However, the complexity of the circuits involved does not seem to favor these schemes over the simple yet efficient MSK modulation.

Similarly, while retaining the advantage of good bit error rate performance, the spectral properties of MSK can be improved by shaping the data pulses further. Note that in MSK, the symbol pulse shape $p(t)$ is $\cos[\pi t g(t)/2T]$ where $g(t) = 1$, $0 \leqslant t \leqslant T$. Other choices of $g(t)$ are possible with the spectral falloff rate depending on the end-point behavior of the shape chosen [21],[22]. For example, a function such as $g(t) = \sin(2\pi t/T)/(2\pi t/T)$ (known as sinusoidal frequency shift keying [23]) results in a much smoother $p(t)$ and produces an asymptotic spectral falloff that is twice as fast as in MSK. Unfortunately, all these generalizations tend to produce a broader (main lobe) spectrum than MSK thus worsening the performance at low bandwidth/bit rate values. MSK has been extended to multiple level pulses, known as multiple amplitude MSK (MAMSK) [24]. Other recent works [25],[26] indicate that application of an efficient baseband coding scheme such as correlative coding [27] to MSK may be the answer to further spectral economy and good performance.

SUMMARY

The MSK scheme was shown to be a special case of continuous phase FSK signaling with frequency deviation equal to ½ the bit rate. MSK can also be viewed as a form of offset QPSK signaling in which the symbol pulse is a half-cycle sinusoid rather than the usual rectangular form. It combines in one modulation format many attractive attributes such as constant envelope, compact spectrum, the error rate performance of BPSK, and simple demodulation and synchronization circuits.

MSK is an excellent modulation technique for digital links when bandwidth conservation and the use of efficient amplitude-saturating transmitters are important requirements.

These features make MSK an excellent modulation technique for digital links in which bandwidth conservation and the use of efficient transmitters with nonlinear (amplitude saturated) devices are important design criteria.

REFERENCES

[1] J. G. Smith, "Spectrally efficient modulation," in *Proc. IEEE Int. Conf. Commun.* (ICC '77), June 1977, pp. 3.1-37-3.1-41.

[2] M. L. Doelz and E. H. Heald, Collins Radio Co., "Minimum-shift data communication system," U.S. Patent 2 977 417, Mar. 28, 1961.

[3] R. DeBuda, "Coherent demodulation of frequency shift keying with low deviation ratio," *IEEE Trans. Commun.*, vol. COM-20, pp. 429-435, June 1972.

[4] W. A. Sullivan, "High capacity microwave system for digital data transmission," *IEEE Trans. Commun.*, vol. COM-20, pp. 466-470, June 1972.

[5] D. M. Brady, "FM-CPSK: Narrowband digital FM with coherent phase detection," in *Proc. IEEE Conf. Commun.* (ICC '72), June 1972, pp. 44-12-44-16.

[6] ——, "A constant envelope digital modulation technique for millimeter-wave satellite system," in *Proc. IEEE Int. Conf. Commun.* (ICC '74), Minneapolis, MN, June 1974.

[7] E. J. Sass and J. R. Hannum, "Minimum shift keying modem for digitized voice communications," *J. RCA*, pp. 80-84, 1974.

[8] S. L. Bernstein *et al.*, "A signaling scheme for extremely low frequency (ELF) communication," *IEEE Trans. Commun.*, vol. COM-22, pp. 508-528, Apr. 1974.

[9] D. P. Taylor *et al.*, "A high speed digital modem for experimental work on the communications technology satellite," *Can. Elec. Eng. J.*, vol. 2, no. 1, pp. 21-30, 1977.

[10] S. A. Rhodes, "Effects of hardlimiting on bandlimited transmissions with conventional and offset QPSK modulation," in *Proc. Nat. Telecommun. Conf.*, Houston, TX, 1972, pp. 20F/1-20F/7.

[11] ——, "Effect of noisy phase reference on coherent detection of offset QPSK signals," *IEEE Trans. Commun.*, vol. COM-22, pp. 1046-1055, Aug. 1974.

[12] R. D. Gitlin and E. H. Ho, "The performance of staggered quadrature amplitude modulation in the presence of phase jitter," *IEEE Trans. Commun.*, vol. COM-23, pp. 348-352, Mar. 1975.

[13] S. A. Gronemeyer and A. L. McBride, "MSK and offset QPSK modulation," *IEEE Trans. Commun.*, vol. COM-24, pp. 809-820, Aug. 1976.

[14] H. R. Mathwich, J. F. Balcewicz, and M. Hecht, "The effect of tandem band and amplitude limiting on the E_b/N_0 performance of minimum (frequency) shift keying (MSK)," *IEEE Trans. Commun.*, vol. COM-22, pp. 1525-1540, Oct. 1974.

[15] M. Atobe, Y. Matsumoto, and Y. Tagashira, "One solution for constant envelope modulation," in *Proc. 4th Int. Conf. Digital Satellite Commun.*, Montreal, P.Q., Canada, Oct. 1978, pp. 45-50.

[16] R. M. Fielding *et al.*, "Performance characterization of a high data rate MSK and QPSK channel," in *Proc. IEEE Int. Conf. Commun.* (ICC '77), June 1977, pp. 3.2-42-3.2-46.

[17] W. R. Bennett and S. O. Rice, "Spectral density and autocorrelation functions associated with binary frequency shift keying," *Bell Syst. Tech. J.*, vol. 42, pp. 2355-2385, Sept. 1963.

[18] R. DeBuda, "Fast FSK signals and their demodulation," *Can. Elec. Eng. J.*, vol. 1, no. 1, pp. 28-34, 1976.

[19] W. B. Osborne and M. B. Luntz, "Coherent and noncoherent detection of CPFSK," *IEEE Trans. Commun.*, vol. COM-22, pp. 1023-1036, Aug. 1974.

[20] R. DeBuda, "About optimal properties of fast frequency-shift keying," *IEEE Trans. Commun.* (Corresp.), vol. COM-22, pp. 1726-1728, Oct. 1974.

[21] M. K. Simon, "A generalization of minimum-shift-keying (MSK)-type signaling based upon input data symbol pulse shaping," *IEEE Trans. Commun.*, vol. COM-24, pp. 845-856, Aug. 1976.

[22] M. Rabzel and S. Pasupathy, "Spectral shaping in MSK-type signals," *IEEE Trans. Commun.* vol. COM-26, pp. 189-195, Jan. 1978.

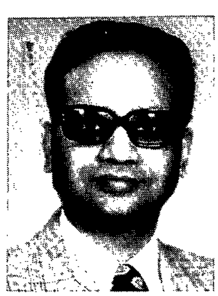
[23] F. Amoroso, "Pulse and spectrum manipulation in the minimum (frequency) shift keying (MSK) format," *IEEE Trans. Commun.* (Corresp.), vol. COM-24, pp. 381-384, Mar. 1976.

[24] W. J. Weber, III, *et al.*, "A bandwidth compressive modulation system using multiamplitude minimum shift-keying (MAMSK)," *IEEE Trans. Commun.*, vol. COM-26, pp. 543-551, May 1978.

[25] G. J. Garrison, "A power spectral density analysis for digital FM," *IEEE Trans. Commun.*, vol. COM-23, pp. 1228-1243, Nov. 1975.

[26] F. de Jager and C. B. Dekker, "Tamed frequency modulation, A novel method to achieve spectrum economy in digital transmission," *IEEE Trans. Commun.*, vol. COM-26, pp. 534-542, May 1978.

[27] S. Pasupathy, "Correlative coding; A bandwidth efficient signaling scheme," *IEEE Commun. Soc. Mag.*, vol. 17, pp. 4-11, July 1977.



Subbarayan Pasupathy was born in Madras, India. He received the B.E. degree in telecommunications from the University of Madras, India, in 1963, the M. Tech. degree in electrical engineering from the Indian Institute of Technology, Madras, India, in 1966, and the M. Phil. and Ph.D. degrees in engineering and applied science from Yale University, New Haven, CT, in 1970 and 1972, respectively.

From 1965-1967, he was a Research Scholar and part-time Lecturer at the Indian Institute of Technology, Madras, India, and he worked as a Teaching Assistant at Yale University, CT, from 1968-1971. From 1972-1973, he was a Postdoctoral Fellow at the University of Toronto, Toronto, Ont., Canada, working in the area of array processing of sonar signals. Since 1973, he has been with the Department of Electrical Engineering, University of Toronto, Toronto, Ont., Canada, where he is now an Associate Professor and Chairman of the Communications Group. His current research interests lie in the areas of digital communications, sonar-radar systems, and communication theory.

Dr. Pasupathy is a registered Professional Engineer in the Province of Ontario, Canada, and a member of the IEEE.